

**MENGATASI HETEROGENITAS DATA ORDINAL MENGGUNAKAN
FIMIX-PLS**

(Skripsi)

Oleh

Heni Noviyanti



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2017**

ABSTRAK

MENGATASI HETEROGENITASI DATA ORDINAL MENGGUNAKAN FIMIX-PLS

Oleh

Heni Noviyanti

Heterogenitas cenderung terjadi dalam data yang berasal dari persepsi dan evaluasi individu. Mengatasi heterogenitas data dapat dilakukan dengan cara pengelompokan objek ke dalam beberapa kelompok berdasarkan ukuran kemiripan atau ciri umum antar objek. Dalam pengelompokan kadang ditemukan objek yang tidak bisa diukur secara langsung karena tidak mempunyai nilai kuantitatif. Objek tersebut disebut dengan variabel laten. Pengelompokan terhadap variabel laten memerlukan data-data ataupun variabel terobservasi yang digunakan sebagai indikator, yang biasa disebut sebagai variabel manifest. Metode yang dapat digunakan untuk mengelompokkan variabel laten berdasarkan variabel indikator yang keduanya bertipe data kategorik adalah analisis FIMIX-PLS (*Finite Mixture-Partial Least Square*). Tujuan penelitian ini adalah mengatasi heterogenitas data menggunakan metode FIMIX-PLS. FIMIX-PLS mengatasi heterogenitas data dengan memperkirakan probabilitas keanggotaan setiap kelompok dan memperkirakan koefisien jalur. Analisis data ECSI (*European Customer Satisfaction Index*) secara global tanpa mempertimbangkan keheterogenan data menghasilkan kebaikan model yang cukup rendah dengan nilai R-Square sebesar 0,452. Sedangkan jika dianalisis dengan membagi data ke dalam kelompok-kelompok dengan karakteristik tertentu dapat menghasilkan 6 kelompok dan memiliki kebaikan model yang tinggi dengan R-square salah satu kelompok mencapai 0,978.

Kata kunci : Variabel laten, Variabel manifest, FIMIX-PLS (*Finite Mixture - Partial Least Square*), ECSI (*European Customer Satisfaction Index*)

ABSTRACT

OVERCOMING THE HETEROGENEITY OF ORDINAL DATA USING FIMIX-PLS

By

Heni Noviyanti

The heterogeneity tends to occur on data that derived from individual perception and evaluation. Overcoming the heterogeneity data can be done by grouping objects into groups based on similarity measure or common characteristics among the object. In data grouping occasionally found objects that can not be measured directly because it has no quantitative value. The object is mentioned as latent variables. Latent variables grouping requires data or observed variables that are used as indicator, commonly mentioned as manifest variables. Methods that can be used to categorize the latent variables based on indicator variables that are both of type categorical data is FIMIX-PLS (Finite Mixture-Partial Least Square) analysis. The purpose of this study is to overcome data heterogeneity using FIMIX-PLS methods. FIMIX-PLS overcomes data heterogeneity by estimate the probability of membership of each group and by estimate the path coefficients. Data analysis ECSI (European Customer Satisfaction Index) globally without considering data heterogeneity results goodness of fit models that quite low with value of R-Square is 0,452. Meanwhile, when analyzed by dividing data into groups with specific characteristics results 6 groups and has goodness of fit with R-square of one group reached 0.978.

Keywords: Latent variable, Manifest variables, FIMIX-PLS (Finite Mixture - Partial Least Square), ECSI (European Customer Satisfaction Index)

**MENGATASI HETEROGENITAS DATA ORDINAL MENGGUNAKAN
FIMIX-PLS**

Oleh

Heni Noviyanti

Skripsi

**Sebagai Salah Satu Syarat Untuk Memperoleh Gelar
SARJANA SAINS**

Pada

**Jurusan Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam**



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2017**

Judul Skripsi : **MENGATASI HETEROGENITAS DATA
ORDINAL MENGGUNAKAN FIMIX-PLS**

Nama Mahasiswa : **Heni Noviyanti**

No. Pokok Mahasiswa : 1317031039

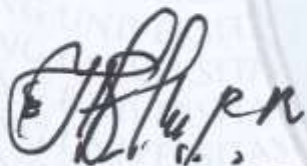
Jurusan : Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

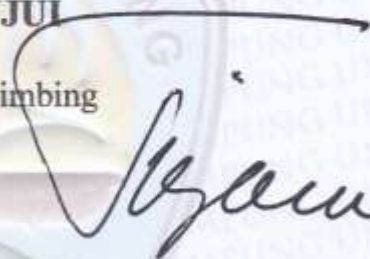


MENYETUJUI

1. Komisi Pembimbing

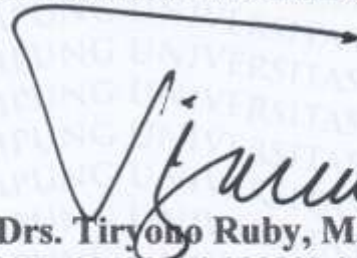


Drs. Eri Setiawan, M.Si.
NIP 19581101 198803 1 002



Drs. Tiryo Ruby, M.Sc., Ph.D.
NIP 19620704 198803 1 002

2. Ketua Jurusan Matematika



Drs. Tiryo Ruby, M.Sc., Ph.D.
NIP 19620704 198803 1 002

MENGESAHKAN

1. Tim Penguji

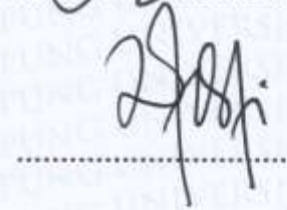
Ketua : **Drs. Eri Setiawan, M.Si.**



Sekretaris : **Drs. Tiryono Ruby, M.Sc., Ph.D.**



Penguji
Bukan Pembimbing : **Widiarti, S.Si., M.Si.**

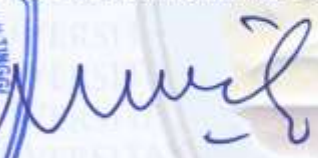


2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



Prof. Warsito, S.Si., D.E.A., Ph.D.

NIP 19710212 199512 1 001



Tanggal Lulus Ujian Skripsi : **24 Januari 2017**

PERNYATAAN SKRIPSI MAHASISWA

Saya yang bertandatangan di bawah ini :

Nama : **Heni Noviyanti**

Nomor Pokok Mahasiswa : **1317031039**

Jurusan : **Matematika**

Dengan ini menyatakan bahwa skripsi ini adalah hasil pekerjaan saya sendiri dan semua tulisan yang tertuang dalam skripsi ini telah mengikuti kaidah karya penulisan ilmiah Universitas Lampung.

Bandar Lampung, Januari 2017

Yang Menyatakan




Heni Noviyanti
NPM. 1317031039

RIWAYAT HIDUP

Penulis dilahirkan di Hargo Pancuran pada 5 Februari 1995, putri pertama dari dua bersaudara buah hati pasangan Bapak Sudarto dan Ibu Sri Hartati.

Penulis menyelesaikan pendidikan Taman kanak-kanak (TK) Aisyiyah Bustanul Athfal Kelaten pada 2001, sekolah dasar di SDN 1 Kelaten pada 2007, sekolah menengah pertama di SMPN 1 Penengahan pada 2010, sekolah menengah atas di SMAN 1 Penengahan pada 2013. Pada tahun yang sama penulis terdaftar sebagai Mahasiswa Program Studi Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung melalui jalur Seleksi Bersama Masuk Perguruan Tinggi Negeri. Selama menjadi mahasiswi, penulis bergabung dalam Himpunan Mahasiswa Jurusan Matematika (HIMATIKA) sebagai anggota bidang Keilmuan periode 2014-2015 hingga periode 2015-2016.

Penulis melaksanakan Kuliah Kerja Nyata (KKN) di Desa Kanoman, Kecamatan Semaka, Kabupaten Tanggamus, Provinsi Lampung pada Januari- Maret 2016 dan penulis melaksanakan Praktik Kerja Lapangan di Badan Pusat Statistik Kalianda, Lampung Selatan pada Juli-Agustus 2016. Penulis pernah menerima beasiswa Peningkatan Prestasi Akademik pada tahun ajaran 2015-2016. Selama masa studi penulis juga pernah menjadi asisten praktikum mata kuliah Eksplorasi Data dan Statistika Dasar.

KATA INSPIRASI

“Maka sesungguhnya disamping kesukaran terdapat pula kemudahan. Sesungguhnya disamping ada kepayahan (jasmani) itu, ada pula kelapangan”
(Al-insyirah 22: 5-6)

“segala sesuatu yang dimulai dari hal yang baik maka akan menuai hasil yang baik”
(Muhammad Sadam Husein)

“Boleh jadi kamu membenci sesuatu padahal ia amat baik bagimu dan boleh jadi (pula) kamu menyukai sesuatu, padahal ia amat buruk bagimu; Allah mengetahui, sedang kamu tidak mengetahui”
(QS. Al-Baqarah 2:216)

“Kerjakan sesuatu sesuai kapasitasmu, terus belajar untuk meningkatkan kualitasmu”



Allhamdulillahirobbil'alamin.....

*Kuhaturkan kepada Allah SWT atas segala rahmat, karunia, dan hidayah-Nya
serta suri tauladanku Nabi Muhammad SAW yang menjadi pedoman hidup
dalam berikhtiar*

*Ibunda yang tercinta dan Ayahanda terbaik terimakasih atas segala doa dan
perjuanganmu yang telah membawaku menuju kesuksesan*

*Mungkin hanya inilah yang mampu kubuktikan kepadamu bahwa aku tak
pernah lupa akan air mata yang jatuh dalam memperjuangkanku, bahwa aku
tak pernah lupa nasihat dan dukunganmu, bahwa aku tak pernah lupa
segalanya dan selamanya*

Saya persembahkan mahakarya yang sederhana ini kepada:

*Ibunda (Sri Hartati), Ayahanda (Sudarto), adikku (Mas Anjang Adi Bijak Sono),
Dosen, serta teman seperjuangan atas waktu, motivasi, dan pengorbanan kalian
yang telah membantuku dalam menyelesaikan skripsi ini*

Serta

*Almamater tercinta yang turut dalam pembentukan pribadi saya menjadi lebih
dewasa dalam berpikir, berucap, dan bertindak*

SANWACANA

Puji syukur kehadirat Allah SWT yang telah melimpahkan berkah dan rahmatNya sehingga penulis dapat menyelesaikan skripsi yang berjudul “**Mengatasi Heterogenitas Data Ordinal Menggunakan FIMIX-PLS**”. Oleh karena itu, penulis ingin mengucapkan terima kasih kepada:

1. Bapak Drs. Eri Setiawan, M.Si. selaku pembimbing pertama. Terima kasih Bapak atas kesediaan waktu, tenaga, pemikiran, motivasi, dukungan, dan pengarahan dalam proses penyusunan skripsi ini.
2. Bapak Drs. Tiryono Ruby, Msc., Ph.D. selaku pembimbing kedua dan selaku Ketua Jurusan Matematika FMIPA Universitas Lampung. Terima kasih Bapak atas kesediaan waktu, tenaga, pemikiran, motivasi, dukungan dan pengarahan yang telah diberikan.
3. Ibu Widiarti, S.Si., M.Si. selaku pembahas. Terima kasih atas kesediaan waktu dan pemikiran Ibu dalam memberikan kritik dan saran yang membangun dalam proses penyusunan skripsi ini.
4. Ibu Dian Kurniasari, S.Si., M.Si. selaku pembimbing akademik yang selalu memberi arahan, memberi nasihat dan meluangkan waktunya kepada penulis selama proses perkuliahan.
5. Bapak Dr. Warsito, S.Si., D.E.A., Ph.D. selaku Dekan FMIPA Universitas Lampung.

6. Bapak dan Ibu Dosen Jurusan Matematika, Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung atas bimbingan, nasihat, dan ilmu yang diberikan selama masa studi.
7. Ibuku tercinta dan Ayahku terbaik atas segala pengorbanan, doa, dorongan, semangat, dan kasih sayang yang tulus serta senantiasa berjuang untuk keberhasilan penulis.
8. Sahabat-sahabat penulis Vinny, Karindha, Eky, Suci, Sinta, Zefni, Tina, Selma, Risa, Silvia, Mba Okni dan Hesti. Terima kasih atas kebersamaannya dan dukungan kalian selama ini. Semoga akan terus berlanjut sampai kapanpun.
9. Kawan-kawan satu bimbingan, tetap semangat dan jangan menyerah karena kita pasti bisa. Terima kasih atas bantuan dan dukungannya dalam menyelesaikan skripsi ini.
10. Teman-teman seperjuangan Matematika angkatan 2013. Terima kasih atas keakraban dan kebersamaan selama ini.
11. Semua pihak yang telah memberikan bantuan dalam menyelesaikan skripsi ini yang tidak dapat penulis sebutkan satu persatu.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna, besar harapan penulis semoga skripsi ini bermanfaat bagi semua pihak yang memerlukannya.

Bandar Lampung, Januari 2017
Penulis,

Heni Noviyanti

DAFTAR ISI

	Halaman
DAFTAR TABEL	iii
DAFTAR GAMBAR	v
I. PENDAHULUAN	
1.1 Latar Belakang dan Masalah	1
1.2 Tujuan Penelitian	4
1.3 Manfaat Penelitian	4
II. TINJAUAN PUSTAKA	
2.1 Analisis Multivariat	5
2.2 SEM (<i>Structural Equation Modeling</i>)	6
2.3 Variabel dalam SEM	7
2.3.1 Model dalam SEM	7
2.3.2 Asumsi yang Mendasari Penggunaan SEM	8
2.4 Uji Asumsi Statistik	13
2.4.1 Uji Multikolinearitas	13
2.5 Metode Pendugaan PLS	15
2.6 Langkah-Langkah PLS-SEM	16
2.7 FIMIX PLS (<i>Finite Mixture Partial Least Square</i>)	22
2.8 Estimasi Parameter Model Kelas Laten	27
2.8.1 Tahap Ekspektasi	29
2.8.2 Tahap Maksimasi	30
2.9 Analisis Finite Mixture	32
III. METODOLOGI PENELITIAN	
3.1 Waktu dan Tempat Penelitian	35

3.2 Data Penelitian.....	35
3.3 Langkah-Langkah Penelitian.....	35

IV. HASIL DAN PEMBAHAAN

4.1 Spesifikasi Model.....	38
4.1.1 Spesifikasi Model Struktural.....	38
4.1.2 Spesifikasi Model Pengukuran	41
4.2 Identifikasi data ECSI.....	43
4.3 Membangun Diagram Jalur dan Estimasi <i>SemPLS</i>	49
4.3.1 Diagram Jalur	49
4.3.2 Estimasi <i>SemPLS</i>	49
4.3.3 Pengujian Validitas dan reliabilitas	51
4.4 Pengujian Model Struktural	58
4.5 Evaluasi Model Secara Agregat	60
4.6 Pembahasan FIMIX-PLS.....	61
4.7 Pengujian Untuk Beberapa Ukuran K Pada FIMIX-PLS	63
4.8 Evaluasi dan Intepretasi Hasil Segmen Spesifik PLS	66

V. KESIMPULAN

DAFTAR PUSTAKA

LAMPIRAN

DAFTAR TABEL

Tabel	halaman
4.1 Jalur Model Struktural.....	39
4.2 Pendekatan Matriks Model Struktural	40
4.3 Jalur Model Pengukuran.	41
4.4 Data ECSI untuk Variabel Laten <i>Image</i>	45
4.5 Data ECSI untuk Variabel Laten <i>Expectation</i>	46
4.6 Data ECSI untuk Variabel Laten <i>Quality</i>	46
4.7 Data ECSI untuk Variabel Laten <i>Value</i>	47
4.8 Data ECSI untuk Variabel Laten <i>Satisfaction</i>	47
4.9 Data ECSI untuk Variabel Laten <i>Complaints</i>	48
4.10 Data ECSI untuk Variabel Laten <i>Loyalty</i>	48
4.11 Hasil Estimasi untuk Model Struktural (Koefisien Jalur).....	50
4.12 Hasil Estimasi untuk Model Pengukuran (<i>Loadings</i>).	50
4.13 Hasil re-estimasi untuk Model Pengukuran.....	52
4.14 <i>Cross Loading</i>	54
4.15 AVE (<i>Average variance extracted</i>) tanpa indikator CUSL2.	55
4.16 <i>Outer Loadings</i> tanpa indikator CUSL2.	56
4.17 AVE (<i>Average variance extracted</i>) tanpa indikator CUSL2 dan IMAG3....	57
4.18 <i>Composite Reliability</i>	58
4.19 R-square dan R Square Adjusted.	59
4.20 <i>Path Coefficient</i> secara agregat.	59
4. 21 Klasifikasi kriteria untuk tiap K.....	64
4.22 Ukuran <i>pk</i> tiap segmen yang berbeda.....	64

4.23 R-square tiap segment untuk $K = 6$	66
4.24 <i>Path Coefficient</i> secara agregat dan pada setiap segmen.	67
4.25 T statistik <i>Path Coefficient</i> secara agregat dan pada setiap segmen.	68

DAFTAR GAMBAR

Gambar	halaman
2.1 <i>Path</i> kelas laten.....	25
3.1 Diagram Alir langkah-langkah penelitian.....	36
3.2 Model Jalur ECSI.....	37
4.1 Model Sebab-Akibat dari ECSI	39
4.2 Diagram Jalur Model ECSI.....	42
4.3 Diagram Jalur Model ECSI Beserta Nilai Koefisien	49
4.4 Diagram Jalur Model ECSI Tanpa Indikator CUSL2	53

I. PENDAHULUAN

1.1. Latar Belakang dan Masalah

Suatu penelitian tidak akan terlepas dari keberadaan data yang merupakan bahan baku informasi untuk memberikan gambaran spesifik mengenai objek penelitian. Berdasarkan bentuk dan sifatnya, data penelitian dapat dibedakan dalam dua jenis yaitu data kualitatif (yang berbentuk kata-kata/kalimat) dan kuantitatif (yang berbentuk angka). Data dapat dikelompokkan dalam empat jenis tingkatan berdasarkan tipe skala pengukuran yaitu: data nominal, data ordinal, data interval dan data rasio. Data nominal memungkinkan peneliti mengelompokkan objek ke dalam kategori tertentu, sedangkan data ordinal tidak hanya menyatakan kategori tetapi juga menyatakan peringkat dari kategori tersebut. Data interval memiliki tingkatan berada di atas nominal dan ordinal, kemudian rasio merupakan data dengan tingkat pengukuran tertinggi. Dalam penelitian sosial ekonomi data ordinal sering kali digunakan untuk mengetahui suatu variabel yang tidak dapat diukur secara langsung yang disebut dengan variabel laten. Menurut Hair (2010) variabel laten adalah suatu konstruk dalam model persamaan struktural yang tidak dapat diukur secara langsung, tetapi dapat direpresentasikan atau ditentukan oleh satu atau lebih variabel indikator.

Sedangkan variabel indikator adalah variabel yang dapat diukur melalui observasi, sehingga variabel laten tersebut dapat diukur secara tidak langsung oleh variabel indikator. Analisis statistik yang dapat digunakan untuk mengukur hubungan variabel laten dan variabel indikator adalah SEM (*Structural Equation Modeling*). Persamaan model struktural ini pada dasarnya merupakan gabungan analisis regresi, analisis korelasi, analisis jalur dan analisis faktor.

Pada umumnya terdapat dua jenis SEM yang sudah dikenal yaitu CB-SEM (*covariance based structural equation modeling*) yang dikembangkan oleh Joreskog (1973) dan PLS-SEM (*partial least square structural equation modeling*) yang dikembangkan oleh Wold (1974). CB-SEM menuntut basis teori yang kuat, ukuran sampel besar memenuhi berbagai asumsi parametrik dan memenuhi uji kelayakan model (*goodness of fit*), sedangkan PLS-SEM tidak mengharuskan ukuran data yang besar dan terpenuhinya asumsi parametrik (Latan, 2012). Metode PLS digunakan karena metode ini tidak memerlukan asumsi kenormalan data. Ketika mengestimasi model persamaan struktural, peneliti umumnya memperlakukan data seolah-olah dikumpulkan dari satu populasi (Muthen, 1989). Masalah yang sering muncul yaitu asumsi yang menganggap bahwa populasi homogen sering kali tidak realistis, sebagai misal pada penelitian di bidang marketing, konsumen yang berasal dari segmen pasar yang berbeda biasanya memiliki struktur yang berbeda pula.

Dugaan heterogenitas terjadi karena sampel yang diambil berasal dari populasi yang tidak sama, sehingga perlu dilakukan segmentasi. Segmentasi merupakan pengelompokan data berdasarkan karakteristik tertentu dengan tujuan untuk

mengidentifikasi heterogenitas yang tidak teramati. Penelitian yang menggunakan data dengan populasi yang telah sesuai dengan cluster atau stratanya, maka tidak akan terjadi heterogenitas dan ini sangat baik ketika dilakukan analisis. Namun sebaliknya, ketika sampel diambil dari populasi yang beragam kemudian dianalisis tanpa dilakukan segmentasi maka kesimpulan menjadi bias dan tidak valid. FIMIX-PLS segmentasi adalah metode untuk mengungkap heterogenitas yang tidak teramati di dalam model struktural (Hahn et. al., 2002, Sarstedt et. al., 2011). Metode ini menangkap heterogenitas dengan memperkirakan probabilitas dari keanggotaan segmen untuk setiap observasi dan secara bersamaan memperkirakan koefisien jalur semua segmen. Dalam SEM yang memuat variabel laten, dapat diatasi dengan metode FIMIX-PLS (*Finite Mixture PLS*) yang dikembangkan Hahn et. al. (2002).

Heterogenitas cenderung terjadi dalam data yang berasal dari persepsi dan evaluasi individu. Misalnya dalam analisis kepuasan pelanggan, konsumen dapat membentuk segmen yang berbeda dan masing-masing memiliki tanggapan yang berbeda. Heterogenitas ini dapat mempengaruhi bagian dari model pengukuran (hubungan antara variabel laten dengan variabel indikator) dan model struktural (hubungan antara variabel laten dengan variabel laten). Peneliti dapat memperoleh estimasi parameter yang berbeda ketika mereka mengabaikan keheterogenan data yang akan dianalisis. Dalam studi kepuasan pelanggan analisis bisa menyesatkan bila ada perbedaan yang signifikan antara perkiraan parameter yang dilakukan analisis secara agregat dan dianalisis dengan membentuk segmen (Jedidi et. al., 1997).

Pada penelitian ini akan dikaji heterogenitas tidak teramati dari data ECSI (*European Customer Satisfaction Index*) atau data indeks kepuasan pelanggan Eropa terhadap *provider* telepon genggam dengan sampel yang berjumlah 250. Data ECSI diindikasikan heterogen karena setiap orang memiliki tanggapan yang berbeda-beda atas kepuasan dari *provider* telepon genggam. Data ini akan dianalisis secara agregat kemudian akan dianalisis menggunakan FIMIX-PLS. Jumlah segmen terbaik hasil dari FIMIX-PLS dipilih berdasarkan nilai kriteria dari AIC (*Akaike Information Criterion*), CAIC (*Consistent AIC*) dan EN (*Normed Entropy*). Setelah diperoleh segmen terbaik maka akan diketahui probabilitas keanggotaan setiap segmen. *Software* yang digunakan yaitu program *SmartPLS* dan FIMIX-PLS merupakan fitur khusus pada *software SmartPLS*.

1.2. Tujuan Penelitian

Adapun tujuan dilakukannya penelitian ini adalah:

1. Mengatasi heterogenitas data dengan menggunakan FIMIX-PLS.
2. Mengetahui jumlah segmen terbaik dari hasil FIMIX-PLS berdasarkan nilai kriteria dari AIC (*Akaike Information Criterion*), CAIC (*Consistent AIC*) dan EN (*Normed Entropy*).

1.3. Manfaat Penelitian

Manfaat dilakukannya penelitian ini yaitu dapat menambah wawasan keilmuan dalam menerapkan metode SEM dengan FIMIX-PLS (*Finite Mixture Partial Least Square*).

II. TINJAUAN PUSTAKA

2.1 Analisis Multivariat

Analisis Multivariat adalah metode yang digunakan untuk menganalisis data yang terdiri dari lebih dari satu peubah secara simultan. Sering kali data yang dikumpulkan dalam suatu penelitian adalah dari sejumlah unit objek yang besar dan pada setiap objek banyak variabel yang diukur. Untuk menganalisis data semacam ini, statistik univariat tidak lagi dapat menyelesaikan masalah secara baik, sehingga diperlukan statistik multivariat.

Suatu matriks acak $X = (X_1, X_2, \dots, X_p)$ berderajat p dikatakan berdistribusi normal multivariat dengan vektor nilai tengah $\underline{\mu}$ dan matriks kovarian Σ dituliskan :

$$X \sim N_p(\underline{\mu}, \Sigma) \quad (2.1)$$

Misalkan $X = (X_1, X_2, \dots, X_p)$ variabel acak dari distribusi normal multivariat dengan vektor nilai tengah $\underline{\mu}$ dan matriks kovarian Σ , penduga $\underline{\mu}$ diberikan oleh :

$$\underline{\mu} = E(x) = \begin{bmatrix} E(X_1) \\ E(X_2) \\ \vdots \\ E(X_p) \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}$$

dengan :

$$\mu = \frac{1}{n} (\sum_{i=1}^n x_i) = \frac{X_1, X_2, \dots, X_n}{n} \quad (2.2)$$

sedangkan penduga Σ diberikan oleh :

$$\hat{\Sigma} = \frac{1}{n-1} [\sum_{r=1}^n (X_{ir} - \bar{X}_I)(X_{jr} - \bar{X}_J)] \quad (2.3)$$

Konsep kovarian dirangkum dalam suatu matriks yang memuat varian dan kovarian sebagai berikut :

$$\Sigma = \begin{bmatrix} var(X_1) & cov(X_1, X_2) & \dots & cov(X, X_1) \\ cov(X_2, X_1) & var(X_2) & \dots & cov(X_2, X_p) \\ \vdots & \vdots & \ddots & \vdots \\ cov(X_p, X_1) & cov(X_p, X_2) & \dots & var(X_p) \end{bmatrix} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{bmatrix}$$

(Sartono, 2003).

2.2 SEM (*Structural Equation Modeling*)

SEM merupakan gabungan dari dua metode statistik yang terpisah yaitu analisis faktor (*factor analysis*) yang dikembangkan di ilmu psikologi dan psikometri dan model persamaan simultan (*simultaneous equation modeling*) yang dikembangkan di ekonometrika (Ghozali, 2005).

Model persamaan structural atau SEM (*Structure Equation Model*) memainkan berbagai peranan penting, antara lain sebagai system persamaan simultan, analisis kausal linear, analisis lintasan (*path analysis*), *analysis covariance structure*, dan model persamaan struktural. Meskipun demikian ada beberapa hal yang membedakan SEM dengan analisis regresi biasa maupun teknik multivariat yang lain, karena SEM membutuhkan lebih dari sekedar perangkat statistik yang didasarkan atas regresi biasa dan analisis varian. SEM terdiri dari 2 bagian yaitu model variabel laten dan model variabel pengukuran (Wijanto, 2008).

2.3 Variabel dalam SEM

a. Variabel Laten

Variabel kunci yang menjadi perhatian di dalam SEM adalah variabel laten, dimana variabel laten merupakan konsep abstrak, seperti perilaku orang, sikap, perasaan, dan motivasi. Variabel laten dapat diamati secara tidak langsung dan tidak sempurna melalui efeknya pada variabel teramati. SEM mempunyai 2 jenis variabel laten, yaitu eksogen dan endogen. SEM membedakan kedua jenis variabel ini berdasarkan keikutsertaan variabel sebagai variabel terkait pada persamaan-persamaan dalam model. Variabel laten eksogen sebagai variabel bebas pada persamaan yang ada dalam model. Sedangkan variabel endogen merupakan variabel terikat pada persamaan yang ada dalam model.

b. Variabel Teramati (*Indicator Variable*)

Variabel teramati atau terukur adalah variabel yang dapat diamati atau dapat diukur secara empiris dan sering disebut indikator. Variabel teramati merupakan efek atau ukuran dari variabel laten. Variabel teramati yang berkaitan atau merupakan efek dari variabel laten eksogen (ξ) diberi notasi matematik dengan label X, sedangkan yang berkaitan dengan variabel laten endogen (η) diberi label Y. Simbol diagram lintasan dari variabel teramati adalah bujur sangkar (Wijanto, 2008).

2.3.1. Model dalam SEM

SEM memiliki model-model antara lain model struktural dan model pengukuran, berikut ini gambaran kedua model.

a. Model struktural

Model struktural adalah model yang menggambarkan hubungan-hubungan diantara variabel-variabel laten. Variabel-variabel laten dibagi menjadi dua kelas, yaitu variabel eksogen dan variabel endogen. Hubungan-hubungan ini umumnya linear meskipun perluasan SEM memungkinkan untuk mengikutsertakan hubungan yang non-linear. Model struktural dapat dibuat dalam notasi sederhana

$$Y = YB + Z \quad (2.4)$$

Dimana variabel Y merupakan matriks variabel laten, baik variabel eksogen maupun endogen. Nilai rentang error Z diasumsikan menjadi terpusat dengan $E[Z] = 0$.

b. Model Pengukuran

Model pengukuran memodelkan hubungan antara variabel laten dengan variabel indikator. Hubungan tersebut bersifat reflektif dari variabel laten terkait. Pada diagram PLS, 1 variabel indikator hanya dapat dihubungkan pada 1 variabel laten. Seluruh indikator yang terhubung dengan satu variabel laten disebut blok. Jadi, masing-masing variabel laten memiliki bloknnya sendiri. Blok tersebut dapat berhubungan secara formatif dan reflektif.

2.3.2. Asumsi yang Mendasari Penggunaan SEM

Asumsi-Asumsi yang mendasari penggunaan SEM adalah sebagai berikut:

a. Distribusi Normal Multivariat

Masing-masing indikator mempunyai nilai yang berdistribusi normal terhadap

masing-masing indikator lainnya. Karena permulaan yang kecil normalitas multivariat dapat menuntun kearah perbedaan yang besar dalam pengujian chi-square, dengan demikian akan melemahkan kegunaannya. Secara umum, pelanggaran asumsi ini menaikkan chi-square dan di dalam kondisi tertentu akan menurunkannya.

Selanjutnya penggunaan pengukuran ordinal atau nominal akan menyebabkan adanya pelanggaran normalitas multivariat. Perlu diperhatikan bahwa normalitas multivariat diperlukan untuk MLE, yang merupakan metode dominan dalam SEM yang akan digunakan untuk membuat estimasi koefisien-koefisien jalur struktur. Khususnya, MLE membutuhkan variabel-variabel endogen yang berdistribusi normal. Secara umum, sebagaimana ditunjukkan dalam suatu studi simulasi menunjukkan bahwa dalam kondisi data yang sangat tidak normal, pendugaan parameter SEM seperti misalnya estimasi jalur, masih dianggap akurat tetapi koefisien-koefisien signifikansi yang bersangkutan akan menjadi terlalu tinggi, sehingga nilai chi-square akan meningkat. Perlu diingat bahwa untuk uji keselarasan chi-square dalam model keseluruhan, nilai chi-square tidak harus signifikan jika ada keselarasan model yang baik. Semakin tinggi nilai chi-square, semakin besar perbedaan model yang diestimasi dan matriks kovarian sesungguhnya. Namun, keselarasan model akan semakin tidak baik. Chi-square yang tinggi dapat mengarahkan peneliti berpikir bahwa model-model yang sudah dibuat memerlukan modifikasi. Kurangnya normalitas multivariat biasanya menaikkan statistik chi-square. Misalnya, statistik keselarasan chi-square secara keseluruhan untuk model yang bersangkutan akan bias kearah kesalahan tipe 1, yaitu menolak suatu model yang seharusnya

diterima. Pelanggaran terhadap normalitas multivariat juga cenderung menurunkan (*deflate*) kesalahan-kesalahan standar mulai dari menengah sampai ke tingkat tinggi. Kesalahan-kesalahan yang lebih kecil dari yang seharusnya terjadi mempunyai makna jalur-jalur regresi dan kovarian-kovarian faktor / kesalahan didapati akan menjadi signifikan secara statistik dibandingkan dengan seharusnya yang terjadi.

b. Linearitas

SEM mempunyai asumsi adanya hubungan linear antara variabel-variabel indikator dan variabel-variabel laten, serta antara variabel-variabel laten sendiri. Sekalipun demikian, sebagaimana halnya dengan regresi, peneliti dimungkinkan untuk menambah transformasi eksponensial, logaritma, atau non-linear lainnya dari suatu variabel asli ke dalam model yang dimaksud.

c. Pengukuran Tidak Langsung (*Indirect measurement*)

Secara tipikal, semua variabel dalam model merupakan variabel-variabel laten.

d. Indikator Jamak

Beberapa indikator harus digunakan untuk mengukur masing-masing variabel laten dalam model. Regresi dapat dikatakan sebagai kasus khusus dalam SEM dimana hanya ada satu indikator di setiap variabel laten. Kesalahan pemodelan dalam SEM membutuhkan adanya lebih dari satu pengukuran untuk masing-masing variabel laten.

e. Rekursifitas

Suatu model disebut rekursif jika semua anak panah menuju satu arah, tidak ada factor pengulangan (*feedback looping*), dan faktor gangguan (*disturbance terms*) atau kesalahan residual untuk variabel-variabel endogenous yang tidak dikorelasikan. Dengan kata lain, model-model rekursif merupakan model-model dimana semua anak panah mempunyai satu arah tanpa putaran umpan balik dan peneliti dapat membuat asumsi kovarian-kovarian gangguan kesalahan semua 0. Dapat diartikan bahwa semua variabel yang tidak diukur yang merupakan determinan dari variabel-variabel endogenous tidak dikorelasikan satu dengan lainnya sehingga tidak membentuk *feedback loops*. Model-model dengan gangguan kesalahan yang berkorelasi dapat diperlakukan sebagai model rekursif hanya jika tidak ada pengaruh-pengaruh langsung diantara variabel-variabel endogen.

f. Ketepatan yang Tinggi

Apakah data berupa data interval atau ordinal, data-data tersebut harus mempunyai jumlah nilai yang besar. Jika variabel-variabel mempunyai jumlah nilai yang sangat kecil, maka masalah-masalah metodologi akan muncul pada saat peneliti membandingkan varian dan kovarian, yang merupakan masalah sentral dalam SEM.

g. Residual-Residual Acak dan Kecil

Rata-rata residual-residual atau kovarian hasil penghitungan yang diestimasi minus harus sebesar 0, sebagaimana dalam regresi. Suatu model yang sesuai akan

hanya mempunyai residual-residual kecil. Residual-residual besar menunjukkan kesalahan spesifikasi model, sebagai contoh, beberapa jalur mungkin diperlukan untuk ditambahkan ke dalam model tersebut.

h. Gangguan Kesalahan yang Tidak Berkorelasi (*Uncorrelated Error Terms*)

Seperti di dalam regresi, maka gangguan kesalahan diasumsikan saja. Sekalipun demikian, jika memang ada dan dispesifikasi secara eksplisit dalam model oleh peneliti, maka kesalahan yang berkorelasi (*correlated error*) dapat diestimasi dan dibuat modelnya dalam SEM.

i. Kesalahan Residual yang Tidak Berkorelasi (*Uncorrelated Residual Error*)

Kovarian nilai-nilai variabel tergantung yang diprediksi dan residual-residual harus sebesar 0.

j. Multikolinearitas yang Lengkap

Multikolinearitas diasumsikan tidak ada, tetapi korelasi antara semua variabel bebas dapat dibuat model secara eksplisit dalam SEM. Multikolinearitas yang lengkap akan menghasilkan matriks kovarian tunggal, yang mana peneliti tidak dapat melakukan penghitungan tertentu, misalnya inversi matriks karena pembagian dengan 0 akan terjadi.

k. Ukuran Sampel

Ukuran sampel tidak boleh kecil, karena SEM bergantung pada pengujian-pengujian yang sensitif terhadap ukuran sampel dan magnitude perbedaan-perbedaan matriks kovarian. Secara teori, untuk ukuran sampelnya berkisar antara

200 - 400 untuk model-model yang mempunyai indikator antara 10 - 15. Satu survei terhadap 72 penelitian yang menggunakan SEM didapatkan median ukuran sampel sebanyak 198. Sampel di bawah 100 akan kurang baik hasilnya jika menggunakan SEM (Sarwono, 2007).

2.4. Uji Asumsi Statistik

Sebelum dilakukan proses lebih lanjut yaitu mengestimasi model, terlebih dulu harus dilakukan uji asumsi statistik, salah satunya ialah uji multikolinearitas. Uji multikolinearitas perlu dilakukan dalam model persamaan struktural agar proses estimasi dapat dilakukan dengan baik dan output yang dihasilkan tidak bersifat bias. Uji ini dibutuhkan sebagai syarat yang harus dipenuhi dalam pengolahan SEM.

2.4.1 Uji Multikolinearitas

Dalam model persamaan struktural, asumsi secara empiris yang tidak boleh dilanggar adalah multikolinearitas, karena multikolinearitas dapat memberikan efek yang fatal yaitu model menjadi *non identified* yang berarti parameter dalam model tidak dapat diestimasi dan keluaran dalam bentuk diagram jalur tidak dapat ditampilkan atau jika parameter berhasil diestimasi dan output diagram jalur berhasil ditampilkan, tetapi hasilnya dapat bias (Wijanto, 2008). Ada beberapa cara untuk mengetahui ada tidaknya multikolinearitas, diantaranya :

a. Nilai korelasi (korelasi antar variabel bebas)

Uji multikolinearitas bertujuan untuk menguji apakah ditemukan adanya korelasi

antar variabel bebas (*independen*) pada model. Asumsi multikolinearitas mengharuskan tidak adanya korelasi yang sempurna atau besar diantara variabel-variabel independen. Analisis koefisien korelasi bertujuan untuk mempelajari apakah ada hubungan antara dua variabel. Koefisien korelasi antar variabel independen haruslah lemah (dibawah 0.5). Jika korelasi kuat, terjadilah problem multikolinearitas. Koefisien korelasi dirumuskan sebagai berikut :

$$\begin{aligned}
 r_{xy} &= s_{xy} / s_x s_y \\
 &= \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{[\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2]^{1/2}} \\
 &= \frac{n \sum_{i=1}^n X_i Y_i - (\sum_{i=1}^n X_i)(\sum_{i=1}^n Y_i)}{\sqrt{[n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2][n \sum_{i=1}^n Y_i^2 - (\sum_{i=1}^n Y_i)^2]}} \quad (2.5)
 \end{aligned}$$

dimana x dan y adalah variabel bebas (*independen*) pada model, sedangkan n adalah banyaknya sampel yang digunakan (Santoso, 2012).

b. VIF (*Variance Inflation Faktor*) dan *Tolerance*

Metode untuk menguji adanya multikolinearitas dapat dilihat pada *tolerance value* atau VIF (*variance inflation factor*). Kedua ukuran ini menunjukkan setiap variabel independen manakah yang dijelaskan oleh variabel independen lainnya. *Tolerance* mengukur variabilitas variabel independen yang terpilih yang tidak dijelaskan oleh variabel independen lainnya. Nilai *tolerance* yang rendah sama dengan nilai VIF yang tinggi.

Nilai VIF dapat diperoleh dengan rumus berikut :

$$VIF = \frac{1}{Tolerance}$$

Batas *tolerance value* adalah 0,10 atau nilai VIF adalah 10. Jika $VIF > 10$ dan nilai *tolerance* < 0.10 , maka terjadi multikolinearitas tinggi antar variabel bebas

dengan variable bebas lainnya. Jika $VIF < 10$ dan nilai *tolerance* > 0.10 , maka dapat diartikan tidak terdapat multikolinearitas pada penelitian tersebut. Regresi yang baik memiliki VIF disekitar angka 1 (satu) dan mempunyai angka *Tolerance* mendekati 1 (Santoso, 2012).

Cara yang dapat dilakukan untuk menanggulangi jika terjadi adalah dengan mengeluarkan salah satu variabel bebas yang memiliki korelasi yang tinggi dari model regresi dan identifikasi variabel lainnya untuk membantu prediksi.

2.5. Metode Pendugaan PLS

Metode pendugaan yang digunakan pada penelitian ini adalah metode kuadrat terkecil parsial atau PLS (*Partial Least Square*). PLS-SEM merupakan sebuah pendekatan kausal yang bertujuan memaksimalkan variansi dari variabel laten kriteria yang dapat dijelaskan (*explained variance*) oleh variabel laten prediktor (Solihin dan Ratmono, 2013).

Menurut Fornell dan Bookstein (1982) *Partial Least Square* metode analisis yang powerful oleh karena tidak mengasumsikan data harus dengan pengukuran skala tertentu, jumlah sample kecil. PLS dapat juga digunakan untuk konfirmasi teori. Dibandingkan dengan *covariance based SEM* (yang diwakili oleh *software LISREL, EQS atau AMOS*) *component based PLS* mampu menghindari dua masalah besar yang dihadapi oleh *covariance based SEM* (CBSEM) yaitu tidak akan terjadi masalah matriks *singularity (inadmissible solution)* (Ghozali, 2014).

PLS dapat dianggap sebagai model alternatif dari *covariance based SEM* yang lebih cocok untuk tujuan prediksi. Dengan pendekatan PLS diasumsikan bahwa

semua ukuran *variance* adalah *variance* yang berguna untuk dijelaskan. Oleh karena pendekatan untuk mengestimasi variabel laten dianggap sebagai kombinasi linear dari indikator maka menghindarkan masalah *indeterminacy* dan memberikan definisi yang pasti dari komponen skor. PLS memberikan model umum yang meliputi teknik korelasi kanonikal, *redundancy analysis*, regresi berganda, MANOVA (*multivariate analysis of variance*) dan analisis komponen utama (*principle component analysis*) (Ghozali, 2014).

2.6. Langkah-Langkah PLS-SEM

1. Spesifikasi Model

Model analisis jalur semua variabel laten dalam PLS terdiri dari tiga set hubungan: (1) *inner model* yang menspesifikan hubungan antar variabel laten (*structural model*), (2) *outer model* yang menspesifikan hubungan antara variabel laten dengan indikator atau variabel manifestnya (*measurement model*), dan (3) *weight relation* dalam mana nilai kasus dari variabel laten dapat diestimasi. Tanpa kehilangan generalisasi, dapat diasumsikan bahwa variabel laten dan indikator atau manifest variabel di skala *zero means* dan unit *variance* (nilai *standardized*) sehingga parameter lokasi (parameter konstanta) dapat dihilangkan dalam model.

2. Merancang Model Struktural (*inner model*)

Inner model adalah persamaan struktural yang menggambarkan spesifikasi hubungan antar variabel laten berdasarkan teori substantif penelitian, disebut juga dengan *inner relation*. Diasumsikan bahwa variabel laten dan indikator atau variabel manifest adalah pada skala *zero means* atau nilai rata-rata

sama dengan nol dan unit varians sama dengan satu, tanpa menghilangkan sifat umumnya. Sehingga parameter lokasi yaitu parameter konstanta dapat dihilangkan dari model. Persamaan matematis model struktural secara umum adalah sebagai berikut:

$$\eta = \mathbf{B}\eta + \Gamma\xi + \zeta \quad (2.6)$$

$$(\eta - \mathbf{B}\eta) = \Gamma\xi + \zeta \quad (2.7)$$

$$(\mathbf{I} - \mathbf{B})\eta = \Gamma\xi + \zeta \quad (2.8)$$

$$\eta = (\mathbf{I} - \mathbf{B})^{-1}\Gamma\xi + (\mathbf{I} - \mathbf{B})^{-1}\zeta \quad (2.9)$$

dengan :

η = vektor variabel endogen (dependen)

ξ = vektor variabel laten eksogen

ζ = vektor residual (*unexplained variance*)

Pada model rekursif maka hubungan antar variabel laten, dispesifikan sebagai berikut:

$$\eta_j = \sum_i \mathbf{B}_{ji} \eta_i + \sum_i \gamma_{jb} \xi_b + \zeta_j \quad (2.10)$$

Pada SEM perancangan model adalah berbasis teori, akan tetapi pada PLS dapat berupa:

- a. Teori
- b. Hasil penelitian empiris
- c. Analogi, hubungan antar variabel pada bidang ilmu lain
- d. Normatif, misal peraturan pemerintah, undang-undang, dan lain sebagainya
- e. Rasional (PLS: bisa eksplorasi hubungan antar variabel)

3. Merancang Model Pengukuran (*outer model*)

Persamaan umum *outer model* dengan indikator refleksif adalah :

$$\mathbf{x} = \Lambda_x \boldsymbol{\xi} + \boldsymbol{\delta} \text{ dan } \mathbf{y} = \Lambda_y \boldsymbol{\eta} + \boldsymbol{\varepsilon} \quad (2.11)$$

dengan :

Λ_y = matrik koefisien yang merupakan pengaruh variabel η terhadap variabel indikator \mathbf{y}

Λ_x = matrik koefisien yang merupakan pengaruh variabel ξ terhadap variabel indikator \mathbf{x}

Pada SEM semua bersifat refleksif, model pengukuran tidak penting. Namun pada PLS perancangan *outer model* sangat penting yaitu reflektif atau formatif.

- a. Kontruksi diagram jalur.
- b. Konversi diagram jalur ke bentuk persamaan.
- c. Estimasi bobot

Pada model reflektif , bobot w_{jh} adalah koefisien regresi dari z_j dalam regresi sederhana x_{jh} pada estimasi *inner model* z_j , dengan z_j adalah variabel yang distandarisasi :

$$x_{jh} = w_{jh}z_j + e_{jh} \quad (2.12)$$

dari persamaan (2.12) diperoleh $e_{jh} = x_{jh} - w_{jh}z_j$, kemudian estimasi bobot model reflektif diperoleh dengan metode OLS diperoleh dengan cara meminimumkan jumlah kuadrat e_{jh} . Jumlah kuadrat e_{jh} diturunkan terhadap w_{jh} dan diperoleh:

$$\hat{w}_{jh} = \frac{cov(x_{jh}, z_j)}{var(z_j^2)} \quad (2.13)$$

d. Estimasi Jalur yang menghubungkan antar variabel laten (koefisien jalur) dan antara variabel laten dengan indikatornya (*loading*).

- Estimasi *outer* model

Estimasi *outer* model y_j dari standarisasi variabel laten ($\xi_j - m_j$) dengan rata-rata = 0 dan standar deviasi =1, diestimasi dengan kombinasi linear dari pusat variabel indikator melalui persamaan berikut:

$$\hat{y}_{jh} = \sum_{h=1}^J \hat{w}_{jh}(x_{jh} - \bar{x}_{jh}) \quad (2.14)$$

dengan w_{jh} dan \bar{w}_{jh} keduanya adalah pembobot *outer* model.

- Estimasi *inner* model

Dengan mengikuti algoritma dari Wold (1985), maka estimasi *inner* model dari *standardized* variabel laten ($\xi_j - m_j$) didefinisikan dengan

$$z_j \propto \sum e_{ji}y_i \quad (2.15)$$

dimana bobot *inner* model e_{ji} dapat dipilih melalui 3 skema yaitu: skema jalur, skema centroid dan skema faktor.

- Estimasi rata-rata dan lokasi parameter

Pada tahap ini estimasi didasarkan pada matriks data asli dan hasil estimasi bobot pada tahap pertama dan koefisien jalur pada tahap kedua, tujuannya adalah untuk menghitung rata-rata dan lokasi parameter untuk indikator dan variabel laten. Estimasi untuk ξ_j dan η_j sebagai berikut:

$$\widehat{\xi}_J = \sum_{h=1}^J \widehat{w}_{jh} x_{jh} \quad (2.16)$$

dan

$$\widehat{\eta}_J = \sum_{h=1}^J \widehat{w}_{jh} y_{jh} \quad (2.17)$$

e. Evaluasi kecocokan model.

f. *Outer Model* refleksif.

Untuk model penelitian yang menggunakan *outer model* refleksif dievaluasi berdasarkan *convergent*, *discriminant validity*, *composite reliability*. Nilai *convergent* dilihat dari nilai *loading*, nilai tersebut dianggap cukup antara 0.5 sampai 0.6 untuk jumlah variabel laten antara 3 sampai 7. Nilai *discriminant validity* dilihat berdasarkan nilai AVE, nilai AVE tersebut > 0.5. Nilai *composite reliability* yang masih dapat diterima adalah ≥ 0.7 .

g. *Outer Model* formatif

Untuk model penelitian yang menggunakan *outer model* formatif dievaluasi berdasarkan pada *substantive content*-nya yaitu dengan melihat signifikansi dan *weight*.

h. Inner Model *GOF*

Diukur menggunakan *Q-square predictive relevance*.

Rumus Q-Square:

$$Q^2 = 1 - (1 - R_1^2)(1 - R_2^2) \dots (1 - R_p^2)$$

Dimana $R_1^2, R_2^2 \dots R_p^2$ adalah R square variabel endogen dalam model.

Interpretasi Q^2 sama dengan koefisien determinasi total dalam analisis jalur (mirip dengan R^2 pada regresi).

i. Uji Hipotesis

Hipotesis statistik untuk *outer model*:

$H_0: i = 0, \text{ vs } H_1: i \neq 0$

Hipotesis statistik untuk *inner model*: Variabel eksogen terhadap endogen:

$H_0 : \gamma_i = 0, \text{ vs } H_1 : \gamma_i \neq 0$

Hipotesis statistik untuk *inner model*: Variabel endogen terhadap endogen:

$H_0 : \beta_i = 0, \text{ vs } H_1 : \beta_i \neq 0$

j. Statistik uji

t-test; p-value $\leq 0,05$ (alpha 5%); signifikan

Outer model signifikan: indikator bersifat valid

Inner model signifikan: terdapat pengaruh signifikan

PLS tidak mengasumsikan data berdistribusi normal: menggunakan teknik resampling dengan metode *bootstrap*.

PLS sebagai model prediksi tidak mengasumsikan distribusi tertentu untuk mengestimasi parameter dan memprediksi hubungan kausalitas. Oleh karena itu, teknik parametrik untuk menguji signifikansi parameter tidak diperlukan dan model evaluasi untuk prediksi bersifat non-parametrik. Evaluasi model PLS dilakukan dengan mengevaluasi outer model dan inner model. Outer model merupakan model pengukuran untuk menilai validitas dan reliabilitas model. Melalui proses iterasi algoritma, parameter model pengukuran (validitas konvergen, validitas diskriminan, *composite reliability* dan *crombach's alpha*) diperoleh, termasuk nilai R² sebagai parameter ketepatan model prediksi. Inner model merupakan model struktural untuk memprediksi hubungan kausalitas antar variabel laten. Melalui proses *bootstrapping*, parameter uji T statistik diperoleh untuk memprediksi adanya hubungan kasualitas (Jogiyanto dan Abdillah, 2009).

2.7. FIMIX PLS (*Finite Mixture Partial Least Square*)

Heterogenitas sering hadir dalam penelitian empiris, peneliti harus selalu mempertimbangkan potensi sumber heterogenitas, misalnya, dengan membentuk kelompok-kelompok data berdasarkan karakteristik diamati seperti demografi (misalnya, usia atau jenis kelamin). Ketika struktur data yang heterogen dapat ditelusuri kembali ke karakteristik diamati, situasi ini dianggap sebagai heterogenitas diamati. Sayangnya, sumber heterogenitas dalam data tidak pernah bisa sepenuhnya diketahui. Akibatnya, masalah validitas mungkin muncul terkait dengan heterogenitas yang tidak teramati sehingga estimasi untuk model jalur PLS menjadi kurang akurat (Becker et. al., 2013).

Peneliti perlu menerapkan suatu teknik segmentasi yang memungkinkan untuk mengidentifikasi dan mengatasi masalah heterogenitas yang tidak teramati. FIMIX-PLS segmentasi adalah metode untuk mengungkap heterogenitas yang tidak teramati di dalam model struktural. Metode FIMIX-PLS mengidentifikasi heterogenitas data dengan memperkirakan probabilitas dari keanggotaan segmen untuk setiap observasi dan secara bersamaan memperkirakan koefisien jalur semua segmen. Pendekatan ini bergantung pada konsep model campuran yang terbatas, yang mengasumsikan bahwa populasi keseluruhan adalah fungsi kepadatan campuran dari kelompok tertentu (Hahn et. al., 2002, Sarstedt et. al., 2011).

Fungsi distribusi model campuran merupakan kombinasi linear dari dua atau lebih fungsi kepadatan probabilitas (fkp). Kegunaan mendasar dari model campuran adalah dapat menggambarkan fkp yang rumit atau kompleks. Fungsi distribusi

segmen spesifik didefinisikan sebagai berikut, dengan asumsi bahwa η_i

berdistribusi sebagai FIMIX bersyarat normal multivariate $f_{i|k}(\cdot)$ dengan K

($K < \infty$) segmen

$$\eta_i \sim \sum_{k=1}^K \rho_k f_{i|k}(\eta_i | \xi_i, B_k, \Gamma_k, \Psi_k) \quad (2.18)$$

Subtitusikan $f_{i|k}(\eta_i | \xi_i, B_k, \Gamma_k, \Psi_k)$ ke dalam persamaan berikut:

$$\eta_i = \sum_{k=1}^K \rho_k \left[\frac{1}{(2\pi)^{M/2} \sqrt{|\Psi_k|}} \right] \exp \left\{ -\frac{1}{2} \left((I - B_k) \eta_i + (-\Gamma_k) \xi_i \right)' \Psi_k^{-1} \left((I - B_k) \eta_i + (-\Gamma_k) \xi_i \right) \right\} \quad (2.19)$$

Likelihood data yang diamati yaitu:

$$L = \prod_{i=1}^I \left[\sum_{k=1}^K \rho_k \left[\frac{B_k}{(2\pi)^{M/2} \sqrt{|\Psi_k|}} \exp \left\{ -\frac{1}{2} (B_k \eta_i + \Gamma_k \xi_i)' \Psi_k^{-1} (B_k \eta_i + \Gamma_k \xi_i) \right\} \right] \right] \quad (2.20)$$

Metode estimasi maksimum *likelihood* adalah metode klasik yang dapat

digunakan secara praktis untuk mendapatkan estimator yang tidak bias dan

bervariansi minimum atau *uniformly minimum variance unbiased estimator*

(UMVUE). Tetapi, dalam kasus statistik dengan permasalahan data yang akan

dicari nilai estimasinya tidak memuat informasi yang dibutuhkan secara lengkap.

Metode estimasi maksimum *likelihood* tidak bisa digunakan secara langsung.

Solusi untuk permasalahan tersebut salah satunya adalah dengan algoritma EM.

Algoritma EM digunakan untuk memaksimalkan kemungkinan dalam model ini

dan untuk memastikan konvergensi. Suatu karakteristik utama dari algoritma EM

adalah melakukan perhitungan secara iteratif (berulang-ulang) untuk mendapatkan

estimator dengan permasalahan data tidak lengkap. Setiap iterasi dari algoritma

EM terdiri dari dua tahap yaitu tahap ekspektasi (*E-step*) dan tahap maksimisasi

(*M-step*). Tahap ekspektasi yaitu dicari ekspektasi dari fungsi *likelihood* data

berdasarkan data terobservasi untuk mengganti keanggotaan setiap individu pada

setiap kelas laten yang tidak diketahui. Pada tahap maksimisasi dicari nilai estimator yang dapat memaksimumkan fungsi yang telah didefinisikan pada tahap ekspektasi.

Persamaan (2.21) dan (2.22) merupakan suatu perumusan fungsi kemungkinan EM (*expectation-maximization*) dan loglikelihood yang ($\ln L$) sebagai fungsi tujuan yang sesuai untuk memaksimalkan

$$L = \prod_i \prod_k [\rho_k f(\eta_i | \xi_i, B_k, \Gamma_k, \Psi_k)]^{z_{ik}} \quad (2.21)$$

$$\ln L = \sum_i \sum_k z_{ik} \ln(f(\eta_i | \xi_i, B_k, \Gamma_k, \Psi_k)) + \sum_i \sum_k z_{ik} \ln(\rho_k) \quad (2.22)$$

Nilai harapan pada persamaan 2.22 dihitung berdasarkan *E-step*, dimana z_{ik} adalah 1 jika subjek i masuk ke kelas k (0 jika tidak). Ukuran segmen k (ρ_k) dengan parameter $P_{ik}, B_k, \Gamma_k, \Psi_k$ dari fungsi probabilitas bersyarat yang diberikan dan dugaan sementara (nilai yang diharapkan) untuk z_{ik} dihitung sesuai teorema Bayes sebagai berikut :

$$E(z_{ik}) = P_{ik} = \frac{\rho_k f_{i|k}(\eta_i | \xi_i, B_k, \Gamma_k, \Psi_k)}{\sum_{k=1}^K \rho_k f_{i|k}(\eta_i | \xi_i, B_k, \Gamma_k, \Psi_k)} \quad (2.23)$$

Persamaan (2.22) dimaksimalkan di *M-step*. Awalnya, proporsi pencampuran baru ρ_k dihitung dengan rata-rata P_{ik} yang dihasilkan dari *E-step* sebelumnya :

$$\rho_k = \frac{\sum_{i=1}^I P_{ik}}{I} \quad (2.24)$$

dengan :

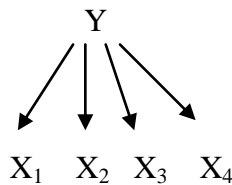
$f_{i|k}(\cdot)$ = probability ke- i pada kelas k dan parameter (\cdot)

ρ_k = proporsi campuran dari laten pada kelas k

η_i = vektor ke- i dari variabel endogen dalam inner model

- ξ_i = vektor ke-i dari variabel eksogen dalam inner model
- B_k = matriks koefisien $M \times M$ model *inner* kelas k yang mengukur hubungan antar variabel dependen (endogen)
- Γ_k = matriks koefisien $M \times J$ model *inner* kelas k yang mengukur hubungan antara variabel laten eksogen dan laten endogen
- Ψ_k = matriks $M \times M$ varian peubah latent kelas ke k
- P_{ik} = probabilitas keanggotaan i pada kelas k
- I = banyaknya observasi

Model dengan satu variabel laten (Y) dan 4 variabel indikator (X) diilustrasikan pada Gambar 2.1.



Gambar 2.1. *Path* kelas laten

Dimisalkan terdapat J variabel indikator dengan X_j adalah variabel indikator ke- j ($j = 1, 2, \dots, J$) dan satu variabel laten sebanyak K kelas. Probabilitas individu pada variabel X adalah

$$\begin{aligned}
 P(X=x) &= P(X = x \cap Y = 1) \cup (X = x \cap Y = 2) \cup \dots \cup (X = x \cap Y = k) \\
 &= \sum_{k=1}^K P(X = x \cap Y = k)
 \end{aligned} \tag{2.25}$$

dan

$$\begin{aligned}
 P(X = x \cap Y = k) &= P(Y = k)P(X = x|Y = k) \\
 &= P(Y = k)P(X_1=x_1 \cap X_2 = x_2 \cap \dots \cap X_j = x_j | Y = k)
 \end{aligned}$$

$$= P(Y = k)P(\cap_{j=1}^J X_j = x_j | Y = k) \quad (2.26)$$

dengan $P(Y = k)$ adalah probabilitas kelas laten k ($k = 1, 2, 3, \dots, K$).

Ide dasar dari kelas laten yaitu variabel indikator independen dengan syarat variabel laten, sehingga probabilitas variabel indikator dengan syarat variabel laten adalah :

$$P(X = x | Y = k) = \prod_{j=1}^J P(X_j = x_j | Y = k) \quad (2.27)$$

Persamaan (2.27) substitusikan ke persamaan (2.26) dan diperoleh

$$P(X = x \cap Y = k) = P(Y = k) \prod_{j=1}^J P(X_j = x_j | Y = k) \quad (2.28)$$

Persamaan (2.28) disubstitusikan ke persamaan (2.25) diperoleh

$$P(X = x) = \sum_{k=1}^K P(Y = k) \prod_{j=1}^J P(X_j = x_j | Y = k)$$

dengan

$$\sum_{k=1}^K P(Y = k) = 1 \quad (2.29)$$

Persamaan (2.25) menyatakan bahwa individu-individu diklasifikasikan dalam K kelas laten yang *mutually exclusive* dan persamaan (2.28) menyatakan variabel indikator *mutually independent*.

Dimisalkan setiap X_j terdapat T_j kemungkinan *outcome*. X_{ijt} adalah nilai terobservasi dari variabel indikator j dengan X_{ijt} bernilai 1 jika individu i berasal dari respon t variabel indikator j dan 0 untuk yang lain. Terdapat variabel laten Y sebanyak K kelas. Probabilitas individu dengan variabel indikator X berpola x_i ($i = 1, 2, \dots, n$) berada pada kelas laten k ($k = 1, 2, \dots, K$) adalah

$$f(X_i; P_k) = P(X = x_i | Y = k) = \prod_{j=1}^J \prod_{t=1}^{T_j} (P_{jtk})^{x_{ijt}} \quad (2.30)$$

$$P_{jtk} = P(X_j = t | Y = k) \quad (2.31)$$

$$\sum_{t=1}^{T_j} P_{jtk} = 1$$

Fungsi kepadatan probabilitas untuk semua kelas adalah

$$\begin{aligned} P(X = x_i) &= \sum_k^K P_k f(X_i; P_k) \\ &= \sum_k^K P_k \prod_{j=1}^J \prod_{t=1}^{T_j} (P_{jtk})^{x_{ijt}} \end{aligned} \quad (2.32)$$

dengan

$$P_k = P(Y = k) \quad (2.33)$$

2.8. Estimasi Parameter Model Kelas Laten

Beberapa parameter statistik seperti rata-rata dan standar deviasi dapat dengan mudah diestimasi dengan menyelesaikan suatu persamaan yang dikenal dengan solusi *close-form*. Tetapi untuk model statistik yang kompleks seperti model kelas laten, penurunan secara *close-form* tidak bisa dicapai sehingga diperlukan modifikasi data untuk mendapatkan nilai parameter yang diinginkan.

Didefinisikan X adalah variabel indikator dengan $x = (x_1, x_2, \dots, x_n)$ adalah data terobservasi dari variabel indikator dan Y adalah variabel laten dengan $y = (y_1, y_2, \dots, y_n)$ adalah vektor data tidak teramati. Data lengkap didefinisikan sebagai $Z = (X, Y)$ dan x_i berpasangan satu-satu dengan y_i dengan $i = 1, 2, \dots, n$. Data x menjadi tidak lengkap karena y tidak tersedia. Adanya permasalahan data tidak lengkap tersebut dapat diatasi dengan algoritma EM untuk menyelesaikan estimasi maksimum *likelihood*. Seluruh data terobservasi adalah campuran dari beberapa kelas laten. Oleh karena itu, persamaan (2.32) dapat dipandang sebagai model campuran dengan P_k sebagai proporsi campuran dan $f(X_i; P_k)$ sebagai fungsi kepadatan multinomial dengan satu kali percobaan

dengan bentuk fkp pada persamaan (2.30). Sehingga algoritma EM dapat digunakan melalui pendekatan model campuran.

Langkah awal yang dilakukan adalah menentukan fungsi *likelihood* dari data terobservasi, yaitu:

$$L(\theta) = \prod_{i=1}^n P(X = x_i)$$

untuk mempermudah perhitungan digunakan fungsi log *likelihood* sebagai

$$\begin{aligned} \log L(\theta) &= \log \prod_{i=1}^n P(X = x_i) \\ &= \sum_{i=1}^n \log P(X = x_i) \\ &= \sum_{i=1}^n \log \sum_{k=1}^K P_k \prod_{j=1}^J \prod_{t=1}^{T_j} (P_{jtk})^{x_{ijt}} \end{aligned} \quad (2.34)$$

Dimana θ adalah estimator parameter θ . Terdapat dua masalah dalam penentuan nilai maksimum fungsi log *likelihood* pada persamaan (2.34) yaitu adanya bentuk logaritma penjumlahan sebanyak K mengakibatkan penurunan secara *close form* tidak dapat dicapai dan jumlah kelas tidak diketahui, sehingga digunakan fungsi log *likelihood* data lengkap. Fungsi log *likelihood* data lengkap didefinisikan sebagai berikut:

$$\log L_{lengkap} = \sum_{i=1}^n \sum_{k=1}^K x_{ik} \log P_k \prod_{j=1}^J \prod_{t=1}^{T_j} (P_{jtk})^{x_{ijt}} \quad (2.35)$$

dengan x_{ik} adalah vektor indikator yang merepresentasikan keanggotaan individu pada kelas laten, x_{ik} bernilai 1 jika individu berasal dari kelas k dan 0 untuk yang lain.

Algoritma EM dimulai dengan pemilihan nilai awal untuk P_k dan P_{jtk} yang kemudian melalui tahap ekspektasi dan maksimasi secara berulang-ulang hingga dicapai P_k dan P_{jtk} yang baru dan konvergen.

2.8.1 Tahap Ekspektasi

Fungsi Q diperoleh dengan menentukan ekspektasi dari persamaan (2.35)

berdasarkan variabel Y dengan syarat variabel X . Fungsi Q ditentukan sebagai

$$\begin{aligned} Q(\theta, \theta^l) &= E_{Y|X, \theta^{l-1}} \left[\sum_{i=1}^n \sum_{k=1}^K x_{ik} \log P_k \prod_{j=1}^J \prod_{t=1}^{T_j} (P_{jtk})^{x_{ijt}} \right] \\ &= \sum_{i=1}^n \sum_{k=1}^K E_{Y|X, \theta^{l-1}} [x_{ik}] \log P_k \prod_{j=1}^J \prod_{t=1}^{T_j} (P_{jtk})^{x_{ijt}} \end{aligned} \quad (2.36)$$

karena nilai dari x_{ik} biner yaitu 0 dan 1, maka ekspektasinya adalah hanya pada saat x_{ik} bernilai 1 yaitu ketika x_i berasal dari kelas k sebagai

$$E_{Y|X, \theta^{l-1}} [x_{ik}] = P(x_{ik} = 1|X) = P(Y = k|X = x_i) \quad (2.37)$$

dengan teorema Bayes diperoleh

$$\begin{aligned} E_{Y|X, \theta^{l-1}} [x_{ik}] &= \frac{P(Y = k)P(X = x_i|Y = k)}{\sum_{k=1}^K P(Y = k)P(X = x_i|Y = k)} \\ &= \frac{P(Y=k) \prod_{j=1}^J \prod_{t=1}^{T_j} P(X_{j=t}|Y=k)^{x_{ijt}}}{\sum_{k=1}^K P(Y=k) \prod_{j=1}^J \prod_{t=1}^{T_j} P(X_{j=t}|Y=k)^{x_{ijt}}} \end{aligned} \quad (2.38)$$

persamaan (2.31) dan persamaan (2.32) disubstitusikan ke persamaan (2.33)

diperoleh

$$\begin{aligned} E_{Y|X, \theta^{l-1}} [x_{ik}] &= \frac{P_k \prod_{j=1}^J \prod_{t=1}^{T_j} (P_{jtk})^{x_{ijt}}}{\sum_{k=1}^K P_k \prod_{j=1}^J \prod_{t=1}^{T_j} (P_{jtk})^{x_{ijt}}} \\ &= \frac{P_k f(X_i; P_k)}{\sum_{k=1}^K P_k f(X_i; P_k)} \end{aligned} \quad (2.39)$$

Substitusikan nilai P_k dan P_{jtk} awal pada persamaan (2.39) diperoleh nilai

probabilitas variabel Y pada kelas laten k dengan syarat variabel X dengan pola x_i .

Dimisalkan terdapat dua kelas laten, nilai parameter P_k dan P_{jtk} awal

disubstitusikan ke persamaan (2.39). Jika $P(Y = 1|X = x_i)$ mendekati nilai 1 dan $P(Y = 2|X = x_i)$ mendekati nilai 0 maka dapat disimpulkan data dengan pola x_i berasal dari kelas laten pertama. Jadi pada tahap ekspektasi ditentukan dari mana asal masing-masing data yang terobservasi, apakah dari kelas pertama, kedua, dan seterusnya. Fungsi Q pada tahap ekspektasi didefinisikan sebagai berikut:

$$Q(\theta, \theta^l) = \sum_{k=1}^K \sum_{i=1}^N P(Y = k|X = x_i) \log P_k \prod_{j=1}^J \prod_{t=1}^{T_j} (P_{jtk})^{x_{ijt}}$$

dalam penelitian ini P_k biasanya dinotasikan dengan ρ_k dan $f(X_i; P_k)$ pada persamaan (2.18) dapat ditulis dengan $f_{i|k}(\eta_i | \xi_i, B_k, \Gamma_k, \Psi_k)$. Sehingga fungsi Q dapat ditulis juga seperti

$$Q(\theta, \theta^l) = \sum_{k=1}^K \sum_{i=1}^N P(Y = k|X = x_i) \log \rho_k f_{i|k}(\eta_i | \xi_i, B_k, \Gamma_k, \Psi_k)$$

2.8.2 Tahap Maksimasi

Dari persamaan (2.33) diketahui $\sum_{k=1}^K P(Y = k) = 1$, sehingga pemaksimalan fungsi Q dapat dilakukan menggunakan metode pengali Lagrange dengan kendala $\sum_{k=1}^K P(Y = k) - 1 = 0$. Fungsi Lagrangennya adalah

$$\begin{aligned} \Phi(\theta) &= Q(\theta, \theta^l) + \lambda \left(\sum_{k=1}^K P(Y = k) - 1 \right) \\ &= \sum_{k=1}^K \sum_{i=1}^N P(Y = k|X = x_i) \log P_k \prod_{j=1}^J \prod_{t=1}^{T_j} (P_{jtk})^{x_{ijt}} \\ &\quad + \lambda (\sum_{k=1}^K P_k - 1) \end{aligned} \quad (2.40)$$

dengan λ adalah pengali Lagrange.

Berikut ini dicari nilai maksimum untuk P_k dengan cara menurunkan persamaan (2.40) terhadap P_k dan λ kemudian menyamakannya dengan 0.

$$\frac{\partial \Phi(\theta)}{\partial P_k} = \frac{1}{P_k} \sum_{i=1}^n P(Y = k|X = x_i) + \lambda = 0 \quad (2.41)$$

$$\frac{\partial \Phi(\theta)}{\partial \lambda} = \sum_{k=1}^K P_k - 1 = 0 \quad (2.42)$$

dari persamaan (2.41) diperoleh

$$P_k = -\frac{1}{\lambda} \sum_{i=1}^n P(Y = k|X = x_i) \quad (2.43)$$

Persamaan (2.43) disubstitusikan ke persamaan (2.42) diperoleh

$$\sum_{i=1}^n \sum_{k=1}^K P(Y = k|X = x_i) = -\lambda \quad (2.44)$$

Karena $\sum_{k=1}^K P(Y = k) = 1$ maka dari persamaan (2.44) diperoleh $-\lambda = n$.

Dengan mensubstitusikan $-\lambda = n$ ke persamaan (2.43) diperoleh P_k baru sebagai estimator dari P_k sebagai

$$P_k \text{ baru} = \frac{1}{n} \sum_{i=1}^n P(Y = k|X = x_i) \quad (2.45)$$

$P_{jtk \text{ baru}}$ sebagai estimator dari P_{jtk} diperoleh dengan cara menyelesaikan fungsi

Lagrange dengan kendala $\sum_{t=1}^{T_j} P_{jtk} = 1$ sebagai

$$\begin{aligned} \Phi(\theta) &= Q(\theta, \theta^l) + \lambda \sum_{t=1}^{T_j} P_{jtk} - 1 \\ &= \sum_{k=1}^K \sum_{i=1}^n P(Y = k|X = x_i) \log P_k \prod_{j=1}^J \prod_{t=1}^{T_j} (P_{jtk})^{X_{ijt}} \\ &\quad + \lambda \left(\sum_{t=1}^{T_j} P_{jtk} - 1 \right) \end{aligned} \quad (2.46)$$

nilai maksimum untuk P_{jtk} diperoleh dengan cara menurunkan persamaan (2.46)

terhadap P_{jtk} dan λ dan menyamakannya dengan 0.

$$\frac{\partial \Phi(\theta)}{\partial P_{jtk}} = \frac{1}{P_{jtk}} \sum_{i=1}^n P(Y = k|X = x_i) X_{ijt} + \lambda = 0 \quad (2.47)$$

$$\frac{\partial \Phi(\theta)}{\partial \lambda} = \sum_{t=1}^{T_j} P_{jtk} - 1 = 0 \quad (2.48)$$

dari persamaan (2.48) diperoleh

$$P_{jtk} = -\frac{1}{\lambda} \sum_{i=1}^n P(Y = k|X = x_i) X_{ijt} \quad (2.49)$$

persamaan (2.49) disubstitusikan ke persamaan (2.48) diperoleh

$$\sum_{t=1}^{T_j} \sum_{i=1}^n P(Y = k|X = x_i) X_{ijt} = -\lambda \quad (2.50)$$

Karena X_{ijt} bernilai 1 jika individu i berasal dari respon t variabel indikator j dan 0 untuk yang lain, maka $\sum_{t=1}^{T_j} X_{ijt} = 1$. Persamaan (2.50) menjadi

$$-\lambda = \sum_{i=1}^n P(Y = k|X = x_i) \quad (2.51)$$

Persamaan (2.51) disubstitusikan ke persamaan (2.49) diperoleh estimator untuk P_{jtk} seperti berikut

$$P_{jtk_{baru}} = \frac{\sum_{i=1}^n P(Y=k|X=x_i) X_{ijt}}{\sum_{i=1}^n P(Y=k|X=x_i)} \quad (2.52)$$

Pemilihan nilai awal dan kompleksitas model kelas laten kadang menyebabkan fungsi log *likelihood* hanya mencapai maksimum lokal. Oleh karena itu lebih baik menjalankan algoritma lebih dari satu nilai awal yang berbeda untuk memastikan fungsi log *likelihood* telah mencapai maksimum global.

2.9. Analisis Finite Mixture

Analisis FIMIX-PLS digunakan untuk mendeteksi apakah data yang dianalisis memiliki heterogenitas sehingga terjadi segmentasi responden. Untuk itu akan dilakukan analisis dengan membagi data ke dalam beberapa segmen misal $k=2$, $k=3$, $k=4$, setelah itu menentukan segmen terbaik dengan melihat kriteria AIC, CAIC dan EN. Dalam menentukan jumlah segmen terbaik akan dibandingkan nilai AIC (*Akaike Information Criterion*), CAIC (*Consistent AIC*) dan EN (*Normed Entropy*) pada berbagai jumlah segmen. Berbagai kriteria ini digunakan untuk menilai perhitungan FIMIX-PLS dan kualitas segmentasi. Tujuan utama dari analisis untuk menangkap heterogenitas pengelompokan data dalam *inner*

path. Sedangkan EN (*Normed Entropy*) statistik (Ramaswamy et al., 1993) merupakan kriteria untuk menganalisis hasil spesifik kelas dari FIMIX-PLS. Besarnya nilai EN berkisar dari 0 sampai 1, dengan meningkatnya nilai EN menunjukkan kualitas pemisahan kelas segmentasi yang semakin baik (Ghozali, 2014).

Salah satu pendekatan yang digunakan untuk mendeteksi heterogenitas yaitu dengan melihat nilai AIC (*Akaike Information Criterion*) yang didefinisikan sebagai berikut:

$$AIC_G = -2 \ln L_G + 2 M_G \quad (2.53)$$

dengan :

M_G = jumlah parameter dalam model statistik

L_G = nilai maksimal dari *likelihood function* untuk estimasi model

Selain AIC terdapat kriteria lain yang dapat digunakan untuk mendeteksi heterogenitas yaitu CAIC (*Consistent AIC*). Pendekatan CAIC merupakan modifikasi dari AIC yang bertujuan mengatasi permasalahan ketika AIC gagal menetapkan estimasi parameter secara tepat. CAIC lebih kokoh jika dibandingkan AIC khususnya untuk ukuran sampel yang besar. Nilai *Consistent AIC* didefinisikan sebagai berikut:

$$CAIC = -2 \ln L_G + M_G (\ln N + 1) \quad (2.54)$$

dengan :

M_G = jumlah parameter dalam model statistik

L_G = nilai maksimal dari *likelihood function* untuk estimasi model

Pendekatan lain yang dapat digunakan yaitu BIC (*Bayesian Information Criterion*) yang didefinisikan sebagai berikut:

$$BIC = -2 \ln L_G + M_G \ln N \quad (2.55)$$

dengan :

M_G = jumlah parameter dalam model statistik

L_G = nilai maksimal dari *likelihood function* untuk estimasi model

Pendekatan BIC mirip dengan CAIC dalam menentukan estimasi parameter.

Kriteria klasifikasi yang didasarkan pada statistik entropi menunjukkan tingkat pemisahan antara segmen, hal ini dapat membantu untuk menilai apakah analisis menghasilkan cluster yang terpisah dengan baik. Kriteria ini menunjukkan tingkat klasifikasi semua pengamatan dan memperkirakan probabilitas keanggotaan segmen atas dasar kasus per kasus dan kemudian mengungkapkan jumlah yang paling tepat dari segmen. Meningkatnya nilai EN menunjukkan kualitas pemisahan kelas segmentasi yang semakin baik. Meningkatnya nilai EN menunjukkan kualitas pemisahan kelas segmentasi yang semakin baik

$$EN = 1 - \frac{\sum_{i=1}^I \sum_{k=1}^K -P_{ik} \ln(P_{ik})}{I \ln(K)} \quad (2.56)$$

dengan:

P_{ik} = Probabilitas keanggotaan i pada kelas k

(Ramaswamy et. al., 1993).

III. METODOLOGI PENELITIAN

3.1. Waktu dan Tempat Penelitian

Penelitian ini dilaksanakan pada semester ganjil tahun ajaran 2016/2017 bertempat di Jurusan Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung.

3.2. Data Penelitian

Penelitian ini menggunakan data ordinal *European Customer Satisfaction Index* (ECSI) yang termuat dalam *package smartPLS 3*.

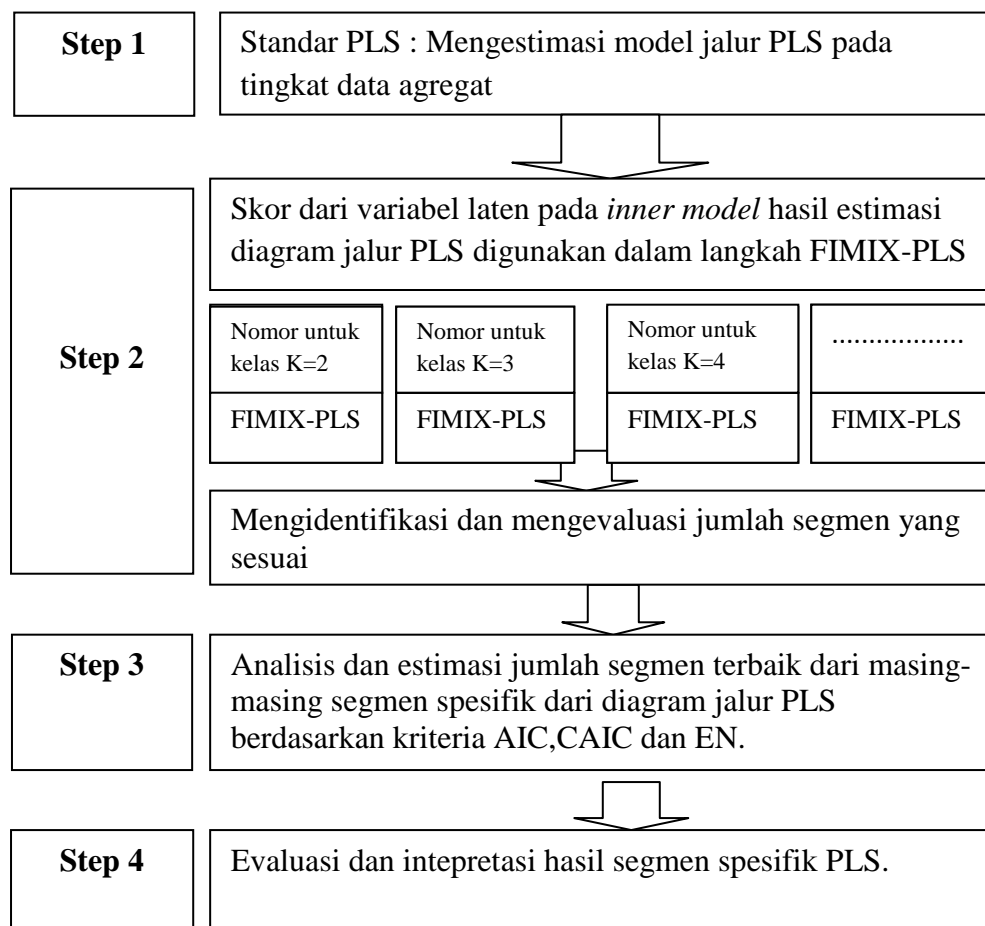
3.3. Langkah-Langkah Penelitian

Dengan menggunakan perangkat bantuan *software smartPLS*, langkah-langkah penelitian yang dilakukan adalah sebagai berikut:

1. Spesifikasi model struktural dan model pengukuran.
2. Identifikasi data *European Customer Satisfaction Index* (ECSI).
3. Mengestimasi model jalur PLS pada tingkat data agregat.
4. Skor dari variabel laten pada *inner model* hasil estimasi jalur PLS digunakan dalam langkah FIMIX-PLS.

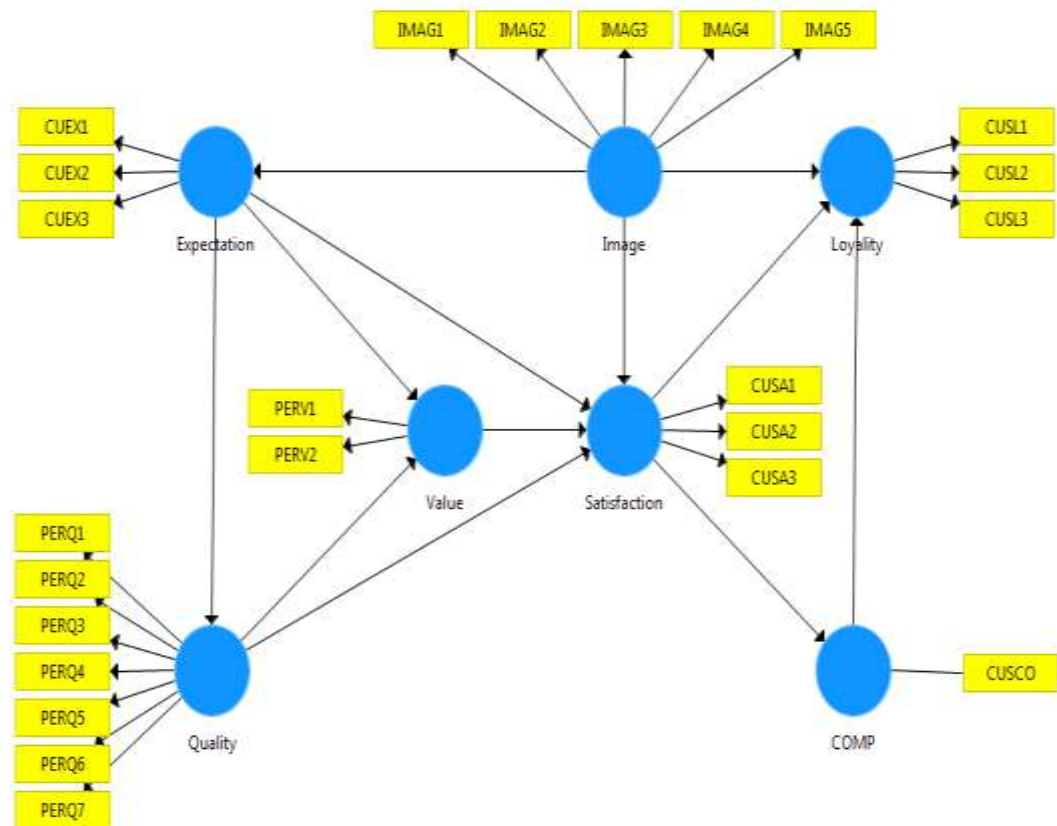
5. Lakukan pengujian untuk beberapa ukuran K pada FIMIX-PLS.
6. Mengidentifikasi dan mengevaluasi jumlah segmen yang sesuai.
7. Analisis dan estimasi jumlah segmen terbaik dari masing-masing segmen spesifik dari diagram jalur PLS berdasarkan nilai kriteria dari AIC (*Akaike Information Criterion*), CAIC (*Consistent AIC*) dan EN (*Normed Entropy*).
8. Evaluasi dan intepretasi hasil segmen spesifik PLS.

Secara garis besar langkah-langkah penelitian yang akan dilakukan dapat tersaji sebagai berikut:



Gambar 3.1 Diagram Alir langkah-langkah penelitian.

Dalam penelitian ini digunakan model jalur seperti gambar 3.2. Model awal menggunakan 7 variabel laten dan 24 variabel indikator. Variabel laten terdiri dari 1 variabel laten eksogen (ξ) yaitu *image* (merk dari suatu barang) dan 6 variabel laten endogen (η) yaitu *expectation* (harapan konsumen dari pengalaman menggunakan barang atau jasa sebelumnya), *quality* (kualitas produk atau kualitas layanan), *value* (harga suatu barang), *satisfaction* (kepuasan pelanggan), kemudian ada konsekuensi dari kepuasan yaitu adanya variabel laten *complaints* (keluhan pelanggan) dan variabel dependent utama dari model ini yaitu *loyalty* (kesetiaan pelanggan). Dan 24 variabel indikator merupakan jumlah dari masing-masing indikator yang dimiliki variabel laten.



Gambar 3.2. Model Jalur ECSI.

V. KESIMPULAN

Berdasarkan analisis data ECSI yang telah dilakukan maka dapat disimpulkan bahwa:

1. Data ECSI merupakan data yang heterogen, analisis secara agregat terhadap *loyalty* (kesetiaan pelanggan) hanya menghasilkan nilai R-Square sebesar 0,452 dan FIMIX-PLS mengatasi heterogenitas data dengan membuat segmen yang salah satu segmennya dapat menghasilkan R-Square terhadap *loyalty* sebesar 0,978.
2. Jumlah segmen terbaik data ECSI berdasarkan kriteria AIC, CAIC dan EN yaitu 6 segmen.

DAFTAR PUSTAKA

- Becker, J.-M., Rai, A., Ringle, C. M., and Völckner, F. 2013. *Discovering Unobserved Heterogeneity in Structural Equation Models to Avert Validity Threats*. *MIS Quarterly*, 37(3): 665-694.
- Fornell, C. dan Bookstein, F. 1982. *Two Structural Equation Models: LISREL and PLS Applied to Customer Exit-Voice Theory*. *Journal of Marketing Research*. 19. 440-452.
- Ghozali, I. 2005. *Model Persamaan Struktural*. Badan Penerbit Universitas Diponegoro, Semarang.
- Ghozali, I. 2014. *Structural Equation Modeling Metode Alternatif dengan Partial Least Square (PLS)*. Badan Penerbit Universitas Diponegoro, Semarang.
- Hahn, C.H., Johnson, M.D., Herrmann, A. and Huber, F. 2002. *Capturing customer heterogeneity using a finite mixture PLS approach*. *Schmalenbach Business Review*, Vol. 54 No. 3, pp. 243-69.
- Hair, J.F., Black, W.C., Babin, B.J., dan Anderson, R.E. 2010. *Multivariate Data Analysis*. Prentice Hall, Englewood Cliffs.
- Jedidi, K., Jagpal, H.S. and DeSarbo, W.S. 1997a. *Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity*. *Marketing Science*, Vol. 16 No. 1, pp. 39-59.
- Jogiyanto dan Abdillah, W. 2009. *Konsep dan Aplikasi PLS untuk Penelitian Empiris*. Fakultas Bisnis Universitas Gajah Mada, Yogyakarta.
- Joreskog, K.G. 1973. *A general Method for Estimating a Linear Structural Equation System*. In A. S. Goldberger & O.D. Duncan (Eds). *Structural Equation Models in the Social Science* (pp.85-112). New York. Academic Press.
- Latan, H. 2012 . *Structural Equation Modeling :Konsep dan Aplikasi Menggunakan Program LISREL 8.80*. Alfabeta, Bandung .
- Monecke, A. Dan Leisch, F. 2012. *semPLS: Structural Equation Modeling Using Partial Least Square*. *Journal of Statistical Software*.

- Muthen, Bengt O. 1989. *Laten Variable Modeling in Heterogenous Population*. Psychometrika. 54. 557-585.
- Ramaswamy, V., DeSarbo, W.S., Reibstein, D.J. and Robinson, W.T. 1993, *An empirical pooling approach for estimating mix elasticities with PIMS data*, Marketing Science, 12, 103-24.
- Sarwono, J. 2007. *Analisis Jalur untuk Riswt Bisnis dengan SPSS*. ANDI, Yogyakarta.
- Solihin, M. Dan Ratmono, D. 2013. *Analisis SEM-PLS dengan WarPLS 3.0*. ANDI, Yogyakarta.
- Santoso, S. 2012. *Analisis SPSS pada Statistik Parametrik*. PT Elex Media Komputindo, Jakarta.
- Sarstedt, M., Becker, J.-M., Ringle, C. M., & Schwaiger, M. 2011. *Uncovering and treating unobserved heterogeneity with FIMIX-PLS: Which model selection criterion provides an appropriate number of segments?*, Schmalenbach Business Review, 63(1).
- Sarstedt, M. (2008b). *A review of recent approaches for capturing heterogeneity in partial least squares path modelling*. Journal of Modelling in Management, Vol. 3 No. 2, pp. 140-161.
- Sartono, B. 2003. *Analisis Peubah Ganda*. Buku Ajar Statistika FMIPA IPB, Bogor.
- Wedel, M. and Kamakura, W.A. 2000. *Market Segmentation: Conceptual and Methodological Foundations*. Kluwer, Boston, MA.
- Wijanto, S.H. 2008. *Structural Equation Modeling dengan Lisrel 8.8: Konsep dan Tutorial*. Graha Ilmu, Yogyakarta.
- Wold, H. (1974). *Causal flows with latent variables: Partings of the ways in light of NIPALS modeling*. European Economic Review, 5(1), 67–86.