

## Study of Regularized Generalized Structured Component Analysis to Overcome Multicollinearity in Component-Based SEM

### Kajian Regularized Generalized Structured Component Analysis untuk Mengatasi Multikolinearitas pada SEM Berbasis Komponen

Fitri Amanah<sup>(1)</sup>, Fitri Rahmawati<sup>(2)</sup>

<sup>(1)</sup> Program Studi Statistika, Universitas Islam Bandung

<sup>(2)</sup> Program Studi Matematika, Universitas Pendidikan Indonesia

Email: <sup>(1)</sup>fitriamanah@unisba.ac.id, <sup>(2)</sup>fitrirahmawati@upi.edu

#### Abstract

Multicollinearity is one of the issues that may arise in the analysis of Structural Equation Modeling (SEM). An indication of multicollinearity is the high correlation between latent variables and the correlation between indicators forming the latent construct. Multicollinearity causes the interpretation of SEM analysis to be inappropriate. In this study, Regularized generalized structured component analysis (RGSCA) is used as a solution to overcome multicollinearity in component-based SEM. The research aims to apply RGSCA to East Java poverty data, which contains multicollinearity. The first step is analyze data using GSCA, however the weights of  $z_2$  and  $z_3$  indicators are not significant, and the three estimated path coefficients are also not significant at the 95% confidence interval. The high correlation value between the  $z_1, z_2, z_3$  indicators further indicates the presence of multicollinearity. Futhermore, the data are analyzed using RGSCA with ridge parameters namely  $\lambda_1 = 2, \lambda_2 = 3, \lambda_3 = 10$  which provides minimum prediction error (CV). The results of the analysis reveal that all estimation of loading factors, weights and path coefficients are significant at 95% confidence intervals. The interpretation of the path coefficient results suggests that education, health, and economy significantly influence poverty, while health and economy also have a significant effect on education, and health additionally exhibits a significant effect on economy. The overall model evaluation results obtained a FIT value of 0.662, indicating that the model can explain about 66.2% of the data variation.

**Keywords:** Regularized GSCA, Multicollinearity, SEM, Poverty



**Abstrak**

Multikolinearitas merupakan salah satu masalah yang dapat terjadi pada analisis Structural Equation Modeling (SEM). Indikasi terjadinya multikolinearitas adalah tingginya korelasi antara variabel laten dan korelasi antara indikator yang membentuk konstruk laten. Multikolinearitas dapat menyebabkan interpretasi hasil analisis SEM menjadi tidak tepat. Pada penelitian ini menerapkan *Regularized generalized structured component analysis* (RGSCA) sebagai solusi mengatasi multikolinearitas pada SEM berbasis komponen. Tujuan penelitian adalah menerapkan RGSCA pada data kemiskinan Jawa Timur 2021 yang mengandung masalah multikolinearitas. Langkah pertama adalah melakukan analisis data menggunakan GSCA namun diperoleh bobot indikator  $z_2$  dan  $z_3$  tidak signifikan, serta tiga estimasi *path koefisien* tidak signifikan pada interval kepercayaan 95%. Nilai korelasi antara indikator  $z_1, z_2, z_3$  sangat tinggi yang mengindikasikan terjadinya multikolinearitas. Selanjutnya, dilakukan analisis data menggunakan RGSCA dengan menerapkan parameter ridge  $\lambda_1 = 2, \lambda_2 = 3, \lambda_3 = 10$  yang menghasilkan kesalahan prediksi (*CV*) minimum. Diperoleh seluruh estimasi loading faktor, bobot dan *path koefisien* signifikan pada interval kepercayaan 95%. Hasil estimasi *path koefisien* menunjukkan variabel pendidikan, kesehatan dan ekonomi berpengaruh signifikan terhadap kemiskinan, variabel kesehatan dan ekonomi berpengaruh signifikan terhadap pendidikan, serta variabel kesehatan juga berpengaruh signifikan terhadap ekonomi. Hasil evaluasi model keseluruhan menunjukkan nilai FIT sebesar 0,662 artinya model mampu menjelaskan sekitar 66,2% variasi dari data.

**Kata kunci:** *Regularized GSCA, Multikolinearitas, SEM, Kemiskinan*

**1. PENDAHULUAN**

*Structural equation modeling* (SEM) adalah teknik analisis statistik yang mengkombinasikan analisis faktor dan analisis regresi (analisis jalur). SEM bertujuan untuk menguji hubungan antar variabel laten dan hubungan antara variabel laten dengan masing-masing indikatornya secara bersamaan [5]. Secara umum SEM memiliki dua komponen model, yaitu model pengukuran dan model struktural. Model pengukuran menggambarkan hubungan antara variabel laten dengan indikator-indikatornya, sedangkan model struktural menggambarkan hubungan antara variabel laten yang berbeda [14]. SEM banyak digunakan dalam berbagai bidang penelitian, seperti ilmu sosial, ekonomi, pendidikan, dan sebagainya. Beberapa penelitian yang menggunakan analisis SEM yaitu [12] penerapan SEM untuk menganalisis kepuasan mahasiswa terhadap sistem informasi akademik Universitas Tanjungpura, [13] analisis SEM untuk mengidentifikasi dampak penerapan *total quality management* pada kinerja karyawan di PT. X melalui peningkatan produktivitas, dan [15] penerapan SEM-PLS untuk mengetahui faktor-faktor yang mempengaruhi keputusan pembelian secara online di Kota Ambon.

Secara umum terdapat dua pendekatan SEM, yaitu SEM berbasis kovarians (*covariance-based SEM*) dan SEM berbasis komponen (*component-based SEM*). Penggunaan SEM berbasis kovarians sangat bergantung pada terpenuhinya asumsi parametrik, diantaranya data berdistribusi normal multivariate, sampel data harus besar, indikator yang digunakan bersifat reflektik, dan teori yang mendukung model harus kuat [16]. Namun seringkali dalam penelitian, data tidak memenuhi asumsi tersebut. Sehingga dikembangkan SEM berbasis komponen untuk mengatasi keterbatasan SEM berbasis kovarians tersebut. Dalam pendekatan SEM berbasis komponen terdapat dua metode yang umum digunakan, yaitu *Partial Least Square* (PLS-SEM) dan *Generalized Structured Component Analysis* (GSCA). Beberapa penelitian terkait SEM berbasis komponen yaitu [1] menerapkan *integrated generalized structured component analysis* untuk menguji hubungan antara dimensi *travel appscape*, manfaat yang dirasakan, keyakinan dalam

memilih dan niat menggunakan aplikasi tersebut oleh pengguna dan [11] menggunakan analisis PLS-SEM untuk mengetahui factor-faktor yang mempengaruhi niat beli konsumen *non-green* terhadap produk kecantikan alami (*green beauty products*).

Salah satu permasalahan yang mungkin terjadi pada analisis SEM adalah multikolinearitas. Gokmen dkk menjelaskan multikolinearitas menyebabkan kesalahan dalam estimasi parameter pada model regresi linear [2]. Sedangkan menurut [9] multikolinearitas menyebabkan interpretasi hasil analisis SEM menjadi tidak tepat. Pada penelitian Hwang dan Takane (2014), multikolinearitas menyebabkan bobot dan *path koefisien* pada SEM-GSCA menjadi tidak signifikan. Pada SEM indikasi terjadinya multikolinearitas dapat dilihat dari tingginya korelasi antar variabel laten dan korelasi antar indikator yang membentuk konstruk variabel laten. Sehingga berdasarkan [8], terdapat dua jenis multikolinearitas pada SEM yaitu multikolinearitas antar variabel laten pada model struktural dan multikolinearitas antar indikator pada model pengukuran. Menurut [10] masalah multikolinearitas harus dipertimbangkan ketika mengevaluasi SEM berbasis komponen, yaitu dengan menganalisis nilai VIF pada semua variabel eksogen. Jika nilai VIF lebih dari 3.3 maka terjadi multikolinearitas.

Hwang dan Takane kemudian memperkenalkan *regularized generalized structured component analysis* (RGSCA) sebagai metode untuk mengatasi multikolinearitas pada SEM model GSCA. GSCA sendiri merupakan SEM berbasis komponen yang mendefinisikan variabel laten sebagai komponen atau komposit tertimbang dari variabel indikatornya [6]. Adapun prosedur estimasi parameter RGSCA di adaptasi dari estimasi parameter regresi *ridge*, yaitu dengan mensubstitusikan parameter *ridge* untuk setiap langkah estimasi GSCA [7]. Pada penelitian [8], penerapan RGSCA menghasilkan bobot dan *path koefisien* menjadi signifikan.

Kemiskinan merupakan salah satu studi kasus yang banyak dibahas pada analisis SEM. Menurut Badan Pusat Statistik, kemiskinan diartikan sebagai ketidakmampuan dari segi ekonomi untuk memenuhi kebutuhan dasar makanan maupun bukan makanan yang diukur dari sisi pengeluaran. Karenanya banyak variabel yang dapat digunakan untuk mengukur faktor-faktor yang mempengaruhi kemiskinan. Namun seringkali ditemukan adanya keterkaitan atau hubungan yang erat antar variabel. Dengan kata lain variabel yang digunakan tidak selalu independen, yang mana dikenal dengan multikolinearitas. Sehingga RGSCA dapat digunakan sebagai metode analisis SEM jika terjadi multikolinearitas pada data. Pada penelitian ini, akan dibahas penerapan RGSCA pada data kemiskinan Jawa Timur 2021 yang mengandung multikolinearitas.

## 2. TINJAUAN PUSTAKA

### 2.2 Generalized Structured Component Analysis (GSCA)

*Generalized structured component analysis* (GSCA) adalah sebuah metode pendekatan SEM berbasis komponen, dimana variabel laten didefinisikan sebagai komponen atau komposit tertimbang dari variabel indikatornya. Metode GSCA diperkenalkan oleh Hwang dan Takane pada tahun 2004 untuk mengatasi kelemahan dari SEM berbasis kovarians dan PLS-SEM, diantaranya GSCA tidak mengasumsikan data berdistribusi normal, ukuran sampel tidak harus besar, indikator yang digunakan dapat reflektif maupun formatif, dan GSCA menyediakan kriteria *goodness of fit* untuk menilai kebaikan model secara keseluruhan [6]. Secara umum GSCA terdiri atas 3 model, yaitu model struktural, model pengukuran dan model pembobotan.

Misalkan  $\gamma$  menyatakan variabel laten ( $\gamma = 1, 2, \dots, P$ ) dan  $z$  menyatakan indikator ( $z = 1, 2, \dots, J$ ), maka persamaan umum GSCA adalah sebagai berikut [8] :

$$V'z = A'W'z + e \quad (2.1)$$

dengan  $V = [I, W']$ ,  $A = [C', B']$ ,  $e = [\epsilon, \zeta]$ ,  $z$  adalah vektor variabel indikator,  $C$  adalah matriks koefisien loading faktor,  $\gamma$  adalah vektor variabel laten,  $\epsilon$  adalah vektor error dari variabel indikator,  $B$  adalah matriks koefisien jalur,  $\zeta$  adalah vektor error dari variabel laten, dan  $W$  adalah matriks bobot.

Parameter GSCA yang akan diestimasi adalah  $W$  dan  $A$  dengan menggunakan algoritma *alternating least square*. Pada tahap pertama  $A$  diperbarui dengan  $W$  dianggap tetap. Persamaan untuk mengestimasi  $A$  adalah sebagai berikut :

$$a = (\Omega' \Omega)^{-1} \Omega' \text{vec}(\Psi) \quad (2.2)$$

dengan  $\Psi = ZV$ ,  $\Gamma = ZW$ ,  $\Omega$  adalah matriks yang dibentuk melalui penghapusan kolom  $I \otimes \Gamma$  yang berkorespondensi dengan elemen nol pada  $\text{vec}(A)$ , dan  $I$  adalah matriks identitas.

Pada tahap kedua  $W$  diperbarui dengan  $A$  dianggap tetap. Persamaan untuk mengestimasi  $W$  adalah sebagai berikut :

$$\eta_k = (\Pi' \Pi)^{-1} \Pi' \text{vec}(Z\Delta) \quad (2.3)$$

dengan  $w_p$  adalah kolom ke- $p$  yang tidak diketahui komponen bobotnya dalam  $W$ ,  $\eta_k$  adalah vektor yang dibentuk dengan menghilangkan setiap elemen nol pada  $w_p$ ,  $\Pi$  adalah matriks yang dibentuk melalui penghapusan kolom dari  $\beta \otimes Z$  yang berkorespondensi dengan elemen nol pada  $w_p$ .

### 2.3 Regularized Generalized Structured Component Analysis (RGSCA)

*Regularized Generalized Structured Component Analysis* (RGSCA) pertama kali diperkenalkan oleh Hwang pada tahun 2009 untuk mengatasi multikolinearitas pada SEM GSCA [7]. Menurut Hwang dan Takane [8], parameter RGSCA yang akan diestimasi adalah  $W$ ,  $C$  dan  $B$ . Tujuan dari RGSCA adalah untuk mengatasi multikolinearitas pada parameter yang akan diestimasi menggunakan *regularized least squares optimization*.

Persamaan umum RGSCA adalah sebagai berikut :

$$\emptyset = SS(ZV - ZWA) + \lambda_1 SS(W) + \lambda_2 SS(C) + \lambda_3 SS(B) \quad (2.4)$$

dengan  $W$  adalah matriks bobot,  $B$  adalah matriks koefisien jalur,  $C$  adalah matriks koefisien loading faktor,  $Z$  adalah matriks berukuran  $N \times J$  dengan  $Z = [z_1, z_2, \dots, z_N]$ ,  $V = [I, W']$ , dan  $\lambda_1, \lambda_2, \lambda_3$  adalah nilai parameter ridge. Pada Persamaan (2.4), parameter ridge ditambahkan pada model GSCA untuk mengatasi multikolinearitas yang mungkin terjadi di semua estimasi parameter secara bersamaan.

Selanjutnya Persamaan (2.4) diminimumkan menggunakan *alternating regularized least square* melalui tiga tahap sebagai berikut :

**Tahap pertama**, perbarui  $C$  dengan  $W$  dan  $B$  dianggap tetap.

Misalkan  $A = [C^* + B^*]$ , dengan  $C^* = [C, 0]$  dan  $B^* = [0, B]$  maka Persamaan (2.4) dapat ditulis kembali sebagai :

$$\emptyset = SS(ZV - ZW[C^* + B^*]) + \lambda_1 SS(W) + \lambda_2 SS(C^*) + \lambda_3 SS(B^*) \quad (2.5)$$

Selanjutnya meminimumkan Persamaan (2.5) terhadap  $C$  sama dengan meminimumkan Persamaan (2.6) berikut:

$$\begin{aligned} \emptyset_1 &= SS(ZV - ZWB^* - ZWC^*) + \lambda_2 SS(C^*) \\ &= SS(\text{vec}(T) - (I \otimes ZW) \text{vec}(C^*)) + \lambda_2 SS(\text{vec}(C^*)) \\ &= SS(\text{vec}(T) - \Omega c) + \lambda_2 c' c \end{aligned} \quad (2.6)$$

dengan  $T = ZV - ZWB^*$ ,  $c$  adalah vector yang terdiri dari elemen bukan nol di  $vec(C^*)$ ,  $\Omega$  adalah matriks kolom dari  $I \otimes ZW$  yang berkorespondensi dengan elemen bukan nol pada  $vec(C^*)$ , dan  $I$  adalah matriks identitas.

Maka estimasi *ridge least square* dari  $c$  adalah :

$$\hat{c} = (\Omega' \Omega + \lambda_2 I)^{-1} \Omega' vec(T) \quad (2.7)$$

**Tahap kedua**, perbarui  $B$  dengan  $W$  dan  $C$  dianggap tetap.

Pada tahap kedua, Persamaan (2.4) diminimumkan terhadap  $B$  sama dengan meminimumkan Persamaan (2.8) berikut :

$$\begin{aligned} \phi_2 &= SS(ZV - ZWC^* - ZWB^*) + \lambda_3 SS(B^*) \\ &= SS(vec(Y) - (I \otimes ZW) vec(B^*) + \lambda_3 SS(vec(B^*))) \\ &= SS(vec(Y) - \Phi b) + \lambda_3 SSb'b \end{aligned} \quad (2.8)$$

dengan  $Y = ZV - ZWC^*$ ,  $b$  adalah vector dari elemen bukan nol di  $vec(B^*)$ ,  $\Phi$  adalah matriks kolom dari  $I \otimes ZW$  yang berkorespondensi dengan elemen bukan nol pada  $vec(B^*)$ , dan  $I$  adalah matriks identitas.

Maka estimasi *ridge least square* dari  $b$  adalah :

$$\hat{b} = (\Phi' \Phi + \lambda_3 I)^{-1} \Phi' vec(Y) \quad (2.9)$$

**Tahap ketiga**, perbarui  $W$  dengan  $B$  dan  $C$  dianggap tetap.

Pada tahap ketiga ini, Persamaan (2.4) dituliskan kembali sebagai :

$$\phi_3 = SS(ZV - ZWA) + \lambda_1 SS(W) \quad (2.10)$$

kemudian Persamaan (2.10) diminimumkan terhadap  $W$ .

Misalkan :

$$\Lambda = WA$$

$w_p$  = kolom ke- $p$  pada  $W$  yang dibagi oleh kolom ke- $t$  di  $V$  dengan  $t = J + P$  ( $p = 1, \dots, P$ ).

$V_{(-t)}$  = matriks  $V$  dengan kolom ke- $t$  adalah vector nol.

$V_{(t)}$  = matriks  $V$  dengan semua kolom adalah vector nol kecuali kolom ke- $t$ .

$\Lambda_{(-p)}$  = perkalian matriks  $W$  (dengan kolom ke- $p$  adalah vector nol) dan matriks  $A$  (dengan baris ke- $p$  adalah vector nol).

$\Lambda_{(p)}$  = perkalian matriks  $W$  (dengan semua kolom nol kecuali kolom ke- $p$ ) dan matriks  $A$  (dengan semua baris nol kecuali baris ke- $p$ ).

$m_{(t)}$  = vektor baris dari matriks  $J + P$  yang semua elemennya nol kecuali elemen ke- $t$  adalah satu.

$a_{(p)}$  = baris ke- $p$  dari matriks  $A$ .

Untuk memperbarui  $w_p$ , Persamaan (2.10) dituliskan kembali sebagai berikut:

$$\phi_3 = \sum_{p=1}^P SS((\beta' \otimes Z) w_p - vec(Z\Delta)) + \lambda_1 \sum_{p=1}^P w_p' w_p \quad (2.11)$$

dengan  $\beta = m_{(t)} - a_{(p)}$  dan  $\Delta = \Lambda_{(-p)} - V_{(t)}$ .

Misalkan :

$\theta_p$  = vektor yang terbentuk dengan mengeliminasi setiap elemen nol dari  $w_p$ .

$E$  = matriks yang terbentuk dengan cara mengeliminasi kolom dari  $\beta' \otimes Z$  yang berkorespondensi dengan elemen nol pada  $w_p$ .

Maka estimasi *ridge least square* dari  $\theta_p$  adalah :

$$\hat{\theta}_p = (\mathbf{E}'\mathbf{E} + \lambda_1 \mathbf{I})^{-1} \mathbf{E}' \text{vec}(\mathbf{Z}\Delta) \quad (2.12)$$

Vector  $\mathbf{w}_p$  baru dibentuk dari  $\hat{\theta}_p$  dan selanjutnya dikalikan dengan  $\frac{\sqrt{N}}{\mathbf{w}_p' \mathbf{Z}' \mathbf{Z} \mathbf{w}_p}$ .

Selanjutnya untuk menentukan nilai parameter  $\lambda_1, \lambda_2, \lambda_3$  dapat menggunakan metode *K-fold cross validation*. Menurut [4] nilai  $K$  yang umumnya digunakan adalah  $K = 5$  atau  $K = 10$ . Pada metode *K-fold cross validation*, seluruh data set ( $\mathbf{Z}$ ) dibagi menjadi  $K$  sub-data. Salah satu dari  $K$  sub-data digunakan untuk data uji (*testing data*), sedangkan sisanya  $K-1$  sub-data digunakan sebagai data latih (*training data*). Misalkan  $\mathbf{Z}^{(k)}$  menyatakan data uji ke- $k$  dari  $\mathbf{Z}$  ( $k = 1, 2, \dots, K$ ),  $\mathbf{Z}^{(-k)}$  menyatakan data latih yang tersisa setelah menghilangkan  $\mathbf{Z}^{(k)}$  dari  $\mathbf{Z}$ . *Regularized* GSCA kemudian diterapkan pada  $\mathbf{Z}^{(-k)}$  untuk mengestimasi parameter.

Misalkan  $\mathbf{W}^{(-k)}$  dan  $\mathbf{A}^{(-k)}$  menyatakan matriks estimasi parameter dari data latih. Maka estimasi *cross validation* dari kesalahan prediksi (CV) dengan nilai  $\lambda_1, \lambda_2, \lambda_3$  tetap adalah sebagai berikut [8] :

$$CV(\lambda_1, \lambda_2, \lambda_3) = \frac{1}{K} \sum_{k=1}^K \mathbf{SS}(\mathbf{Z}^{(k)} \mathbf{V}^{(k)} - \mathbf{Z}^{(k)} \mathbf{W}^{(-k)} \mathbf{A}^{(-k)}) \quad (2.13)$$

Persamaan (2.13) dihitung secara berulang dengan  $\lambda_1, \lambda_2, \lambda_3$  yang berbeda. Kemudian akan dipilih kombinasi nilai  $\lambda_1, \lambda_2, \lambda_3$  yang menghasilkan nilai CV terkecil.

### 3. METODE PENELITIAN

#### 3.1 Data dan Variabel Penelitian

Data yang digunakan adalah data sekunder yang berkaitan dengan kemiskinan, kualitas pendidikan, kualitas kesehatan dan ekonomi di Provinsi Jawa Timur tahun 2021. Jumlah kabupaten/kota yang menjadi unit amatan sejumlah 38. Data tersebut diperoleh dari publikasi online pada situs resmi Badan Pusat Statistik (BPS).

**Tabel 3.1 Variabel Penelitian**

Variabel Laten	Indikator
Kemiskinan	Persentase penduduk miskin ( $z_1$ )
	Indeks kedalaman kemiskinan ( $z_2$ )
	Indeks keparahan kemiskinan ( $z_3$ ).
Pendidikan	Angka melek huruf usia 15 – 55 tahun ( $z_4$ )
	Rata-rata lama sekolah ( $z_5$ )
	Harapan lama sekolah ( $z_6$ )
Kesehatan	Persentase rumah tangga miskin yang menggunakan jamban sendiri/bersama tahun 2021 ( $z_7$ )
	Angka harapan hidup ( $z_8$ )
	Persentase wanita pernah kawin usia 15-49 tahun menurut penolong elahiran tenaga medis / paramedis tahun 2021 ( $z_9$ )
Ekonomi	Persentase penduduk miskin usia 15 tahun ke atas yang tidak bekerja ( $z_{10}$ )
	Persentase rumah tangga yang memiliki status tempat tinggal sewa atau kontrak ( $z_{11}$ )
	Persentase rumah tangga menurut jenis atap terluas yaitu asbes atau seng ( $z_{12}$ )
	Persentase penduduk yang berkerja di sektor formal ( $z_{13}$ )

Variabel Laten	Indikator
	Persentase penduduk yang bekerja bukan pertanian ( $z_{14}$ )
	Rata-rata pengeluaran makanan per kapita ( $z_{15}$ )

### 3.2 Tahapan Analisis Data

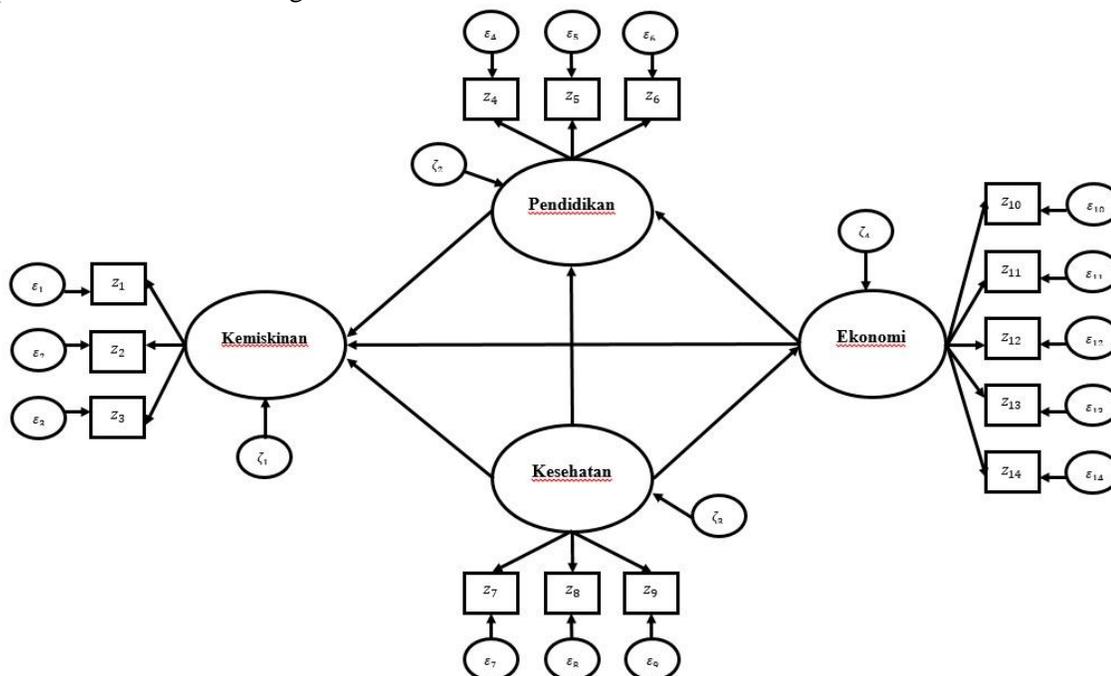
Teknik analisis data yang digunakan dalam penelitian ini adalah GSCA dan RGSCA dengan bantuan *software GSCA Pro*. Adapun tahapan analisis data adalah sebagai berikut :

1. Menggambar diagram jalur faktor-faktor yang mempengaruhi kemiskinan.
2. Melakukan analisis data menggunakan metode GSCA untuk mendapatkan estimasi parameter loading faktor, bobot, dan *path* koefisien.
3. Melihat nilai korelasi antar indikator dan korelasi antar variabel laten jika terdapat indikasi multikolinearitas pada langkah pertama.
4. Melakukan analisis data menggunakan metode RGSCA untuk mengatasi multikolinearitas pada langkah kedua dan untuk mendapatkan estimasi parameter loading faktor, bobot, dan *path* koefisien.
5. Membuat kesimpulan

## 4. HASIL DAN PEMBAHASAN

### 4.1 Diagram Jalur

Diagram jalur faktor-faktor yang mempengaruhi kemiskinan yang digunakan dalam penelitian ini adalah sebagai berikut :



Gambar 4.1. Grafik Model GSCA

### 4.2 Hasil Analisis GSCA

Analisis GSCA dilakukan untuk mendapatkan estimasi parameter loading faktor, bobot, dan *path* koefisien. Menurut Hwang dan Takane (2014), loading faktor digunakan untuk mengetahui

## JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI

Fitri Amanah, Fitri Rahmawati

hubungan antara indikator dan variabel laten, bobot digunakan untuk mengetahui kontribusi masing-masing indikator kepada variabel latennya, sedangkan *path* koefisien untuk mengetahui hubungan antar variabel laten. Hasil perhitungan loading faktor dan *path koefisien* diperoleh dari Persamaan (2.2), sedangkan hasil perhitungan bobot diperoleh dari Persamaan (2.3). Perhitungan dilakukan dengan bantuan *software GSCA Pro*. Hasil dari estimasi parameter disajikan pada tabel berikut :

**Tabel 4.1. Estimasi Loading Faktor dan Bobot menggunakan GSCA**

Laten	Indikator	Loading Faktor		Bobot	
		Estimasi	95% CI	Estimasi	95% CI
KM	$z_1$	0.966	0.94-0.98	0.472	0.18-0.70
	$z_2$	0.990	0.97-0.99	<b>0.350</b>	<b>-0.23-1.12</b>
	$z_3$	0.934	0.88-0.96	<b>0.212</b>	<b>-0.34-0.64</b>
PD	$z_4$	0.827	0.74-0.91	0.287	0.19-0.34
	$z_5$	0.980	0.96-0.99	0.601	0.49-0.71
	$z_6$	0.813	0.59-0.89	0.213	0.10-0.28
KS	$z_7$	0.723	0.54-0.92	0.317	0.16-0.47
	$z_8$	0.878	0.75-0.96	0.526	0.39-0.64
	$z_9$	0.696	0.34-0.87	0.444	0.17-0.59
EKO	$z_{10}$	0.714	0.60-0.85	0.181	0.10-0.27
	$z_{11}$	0.900	0.84-0.94	0.236	0.16-0.40
	$z_{12}$	0.670	0.43-0.81	0.134	0.07-0.20
	$z_{13}$	0.824	0.65-0.92	0.187	0.09-0.30
	$z_{14}$	0.952	0.91-0.97	0.273	0.11-0.40
	$z_{15}$	0.870	0.62-0.96	0.178	0.06-0.31

**Tabel 4.2. Estimasi Path Koefisien menggunakan GSCA**

	Estimasi	SE	95%CI
PD → KM	-0.080	0.483	-1.18-0.73
KS → KM	-0.460	0.349	-0.90-0.42
EKO → KM	-0.354	0.276	-0.86-0.12
KS → PD	0.508	0.062	0.36-0.59
EKO → PD	0.571	0.056	0.49-0.73
KS → EKO	0.486	0.114	0.23-0.67

Berdasarkan Tabel 4.1 diketahui bahwa seluruh estimasi loading faktor signifikan pada interval kepercayaan 95%. Artinya hubungan indikator dan variabel latennya signifikan secara statistik. Sedangkan hasil estimasi bobot menunjukkan terdapat dua bobot indikator yang tidak signifikan pada interval kepercayaan 95%, yaitu  $z_2$  (indeks kedalaman kemiskinan) dan  $z_3$  (indeks keparahan kemiskinan). Sehingga dua indikator tersebut dianggap tidak memberikan kontribusi secara signifikan kepada variabel latennya.

Hasil estimasi *path* koefisien pada Tabel 4.2 menunjukkan tiga estimasi *path* koefisien tidak signifikan pada interval kepercayaan 95%, yaitu hubungan variabel pendidikan dengan variabel kemiskinan, variabel kesehatan dengan variabel kemiskinan, dan variabel ekonomi dengan variabel kemiskinan. Artinya variabel pendidikan, kesehatan dan ekonomi tidak berpengaruh signifikan terhadap kemiskinan pada interval kepercayaan 95%. Hal ini senada dengan penelitian Hwang (2014) yang menunjukkan adanya bobot dan *path* koefisien yang tidak signifikan pada

model GSCA. Hal tersebut mengindikasikan adanya multikolinearitas yang menyebabkan hasil estimasi tidak akurat. Karenanya perlu dilihat nilai korelasi antar variabel untuk memastikan adanya multikolinearitas pada model GSCA tersebut.

#### 4.2 Matriks Korelasi

Pada penelitian Hwang dan Takane (2014), multikolinearitas dilihat dari tingginya nilai korelasi antara indikator dengan variabel laten dan korelasi antar variabel laten. Menurut [3] tingginya korelasi antara variabel independen yang dimaksud yaitu saat nilai  $r > 0,8$  Pada penelitian ini, nilai korelasi antar indikator dan nilai korelasi antar variabel laten diperoleh dari *software GSCA Pro*.

**Tabel 4.3. Matriks Korelasi 15 Indikator**

		KM			PD			KS			EKO					
		$z_1$	$z_2$	$z_3$	$z_4$	$z_5$	$z_6$	$z_7$	$z_8$	$z_9$	$z_{10}$	$z_{11}$	$z_{12}$	$z_{13}$	$z_{14}$	$z_{15}$
KM	$z_1$	1														
	$z_2$	<b>0.92</b>	1													
	$z_3$	<b>0.81</b>	<b>0.96</b>	1												
	$z_4$	-0.26	-0.31	-0.28	1											
PD	$z_5$	-0.58	-0.50	-0.40	0.60	1										
	$z_6$	-0.65	-0.72	-0.74	0.19	0.36	1									
	$z_7$	-0.57	-0.51	-0.41	0.70	0.72	0.39	1								
KS	$z_8$	-0.81	-0.70	-0.57	0.50	0.72	0.53	0.73	1							
	$z_9$	-0.66	-0.51	-0.38	0.22	0.48	0.28	0.45	0.78	1						
	$z_{10}$	-0.64	-0.52	-0.46	0.03	0.38	0.48	0.40	0.60	0.38	1					
	$z_{11}$	-0.64	-0.48	-0.37	0.27	0.41	0.25	0.46	0.78	0.78	0.49	1				
EKO	$z_{12}$	-0.43	-0.36	-0.30	0.11	0.27	0.20	0.26	0.46	0.41	0.29	0.66	1			
	$z_{13}$	-0.61	-0.48	-0.37	0.26	0.37	0.34	0.54	0.68	0.42	0.62	0.60	0.38	1		
	$z_{14}$	-0.72	-0.58	-0.47	0.21	0.40	0.37	0.55	0.84	0.66	0.63	0.85	0.55	0.76	1	
	$z_{15}$	-0.58	-0.30	-0.11	0.23	0.50	0.12	0.49	0.75	0.65	0.48	0.77	0.57	0.66	0.80	1

**Tabel 4.4. Matriks Korelasi Variabel Laten**

	KM	PD	KS	EKO
KM	1			
PD	-0.730	1		
KS	-0.694	0.786	1	
EKO	-0.642	<b>0.818</b>	0.486	1

Berdasarkan Tabel 4.3 diketahui korelasi antara  $z_1, z_2, z_3$  sangat tinggi, yaitu  $z_1$  dan  $z_2$  ( $r = 0,92$ ),  $z_1$  dan  $z_3$  ( $r = 0,81$ ), serta  $z_2$  dan  $z_3$  ( $r = 0,96$ ). Artinya terjadi multikolinearitas antara  $z_1, z_2, z_3$  pada variabel laten kemiskinan. Selanjutnya berdasarkan Tabel 4.4, korelasi tinggi juga terjadi pada variabel laten pendidikan dan variabel laten ekonomi ( $r = 0,818$ ). Langkah ini sejalan dengan [1] yang mana salah satu tahapan yang dilakukan adalah mengecek ada tidaknya multikolinearitas pada model GSCA. Karna pada penelitian ini, data terindikasi multikolinearitas maka selanjutnya data akan dianalisis menggunakan *regularized structured*

*component analysis* (RGSCA). Metode ini merupakan perluasan dari metode GSCA yang dapat digunakan untuk mengatasi multikolinearitas.

#### 4.3 Hasil Analisis RGSCA

Persamaan umum RGSCA mengandung  $\lambda_1, \lambda_2, \lambda_3$  yang merupakan parameter ridge. Nilai  $\lambda_1, \lambda_2, \lambda_3$  diperoleh dari Persamaan (2.13) yang menghasilkan nilai *cross validation error* paling minimum. Sedangkan nilai K yang dipilih adalah  $K = 5$ . Selanjutnya dengan bantuan *software GSCA Pro* diperoleh kombinasi nilai  $\lambda_1, \lambda_2, \lambda_3$  dari 0 sampai 10 sebagai berikut :

**Tabel 4.5. Nilai CV dengan Berbagai Kombinasi  $\lambda_1, \lambda_2, \lambda_3$**

$\lambda_1$	$\lambda_2$	$\lambda_3$	CV
0	0	0	9.034765
0	0	1	8.957632
0	0	2	8.918208
...	...	...	...
2	3	8	8.754296
2	3	9	8.748593
<b>2</b>	<b>3</b>	<b>10</b>	<b>8.743974</b>
2	4	0	8.946887
2	4	1	8.887862
...	...	...	...
10	10	8	9.074786
10	10	9	9.069339
10	10	10	9.065059

Berdasarkan Tabel 4.5 diperoleh  $\lambda_1 = 2, \lambda_2 = 3, \lambda_3 = 10$  yang menghasilkan estimasi *cross validation error* paling minimum. Sehingga nilai parameter ridge tersebut yang digunakan dalam Persamaan (2.5) untuk mengestimasi parameter RGSCA, yaitu :

$$\phi = SS(\mathbf{ZV} - \mathbf{ZWA}) + 2 SS(\mathbf{W}) + 3 SS(\mathbf{C}) + 10 SS(\mathbf{B}) \quad (4.1)$$

Perhitungan estimasi loading faktor menggunakan Persamaan (2.7), estimasi *path* koefisien menggunakan Persamaan (2.9), dan estimasi bobot menggunakan Persamaan (2.12). Perhitungan dilakukan dengan bantuan *software GSCA Pro*. Adapun hasil dari estimasi parameter disajikan pada tabel berikut :

**Tabel 4.6. Estimasi Loading Faktor dan Bobot menggunakan RGSCA**

Laten	Indikator	Loading Faktor		Bobot	
		Estimate	95% CI	Estimate	95% CI
KM	$z_1$	0.892	0.87-0.91	0.451	0.40-0.49
	$z_2$	0.919	0.91-0.92	0.345	0.30-0.38
	$z_3$	0.87	0.85-0.90	0.239	0.18-0.28
PD	$z_4$	0.768	0.69-0.84	0.309	0.27-0.36

**JURNAL MATEMATIKA, STATISTIKA DAN KOMPUTASI**  
Fitri Amanah, Fitri Rahmawati

Laten	Indikator	Loading Faktor		Bobot	
		Estimate	95% CI	Estimate	95% CI
KS	z <sub>5</sub>	0.904	0.89-0.91	0.547	0.44-0.63
	z <sub>6</sub>	0.763	0.65-0.82	0.256	0.21-0.33
	z <sub>7</sub>	0.686	0.46-0.83	0.340	0.19-0.47
	z <sub>8</sub>	0.817	0.68-0.89	0.524	0.41-0.61
	z <sub>9</sub>	0.628	0.41-0.81	0.422	0.25-0.58
EKO	z <sub>10</sub>	0.660	0.50-0.80	0.180	0.11-0.23
	z <sub>11</sub>	0.834	0.79-0.86	0.235	0.19-0.27
	z <sub>12</sub>	0.625	0.31-0.76	0.140	0.07-0.19
	z <sub>13</sub>	0.764	0.63-0.85	0.193	0.11-0.27
	z <sub>14</sub>	0.880	0.85-0.90	0.253	0.18-0.32
	z <sub>15</sub>	0.810	0.65-0.89	0.191	0.11-0.30

**Tabel 4.7. Estimasi Path Koefisien menggunakan GSCA**

	Estimate	SE	95%CI
PD → KM	-0.225	0.089	-0.39-(-0.08)
KS → KM	-0.311	0.113	-0.46-(-0.05)
EKO → KM	-0.237	0.069	-0.35-(-0.12)
KS → PD	0.436	0.041	0.36-0.51
EKO → PD	0.477	0.036	0.41-0.55
KS → EKO	0.382	0.086	0.21-0.52

Berdasarkan Tabel 4.6 diperoleh seluruh estimasi loading faktor signifikan pada interval kepercayaan 95%. Artinya hubungan antara indikator dan variabel latennya signifikan secara statistik. Selanjutnya hasil estimasi bobot menunjukkan seluruh bobot indikator signifikan pada interval kepercayaan 95%. Artinya seluruh indikator memberikan kontribusi yang sama baiknya kepada variabel latennya. Hasil estimasi *path* koefisien menggunakan metode RGSCA disajikan pada Tabel 4.7 juga menunjukkan seluruh estimasi *path* koefisien signifikan pada interval kepercayaan 95%. Hasil penelitian ini sejalan dengan penelitian Hwang dan Takane (2014), yaitu penerapan RGSCA mampu membuat bobot dan *path* koefisien menjadi signifikan.

Hasil interpretasi *path* koefisien pada penelitian ini adalah sebagai berikut :

1. Pendidikan berpengaruh signifikan terhadap kemiskinan, yaitu semakin baik tingkat pendidikan akan dapat mengurangi kemiskinan.
2. Kesehatan berpengaruh signifikan terhadap kemiskinan, yaitu semakin baik kualitas kesehatan akan dapat mengurangi kemiskinan.
3. Ekonomi berpengaruh signifikan terhadap kemiskinan, yaitu semakin baik kondisi ekonomi akan dapat mengurangi kemiskinan.
4. Kesehatan berpengaruh signifikan terhadap pendidikan, yaitu semakin baik kualitas kesehatan akan dapat meningkatkan tingkat pendidikan.
5. Ekonomi berpengaruh signifikan terhadap pendidikan, yaitu semakin baik kondisi ekonomi akan dapat meningkatkan tingkat pendidikan.
6. Kesehatan berpengaruh signifikan terhadap ekonomi, yaitu semakin baik kualitas kesehatan akan dapat meningkatkan kondisi ekonomi.

Hasil evaluasi model keseluruhan menunjukkan nilai FIT sebesar 0,662 artinya model mampu menjelaskan sekitar 66,2% variasi dari data.

Secara umum multikolinearitas dapat ditemui pada analisis SEM. Pada penelitian [1], salah satu tahapan yang dilakukan adalah mengecek multikolinearitas dengan *Variance Inflation Factors (VIF)*. Sedangkan pada penelitian Hwang dan Takane (2014), multikolinearitas dilihat dari tingginya nilai korelasi antara indikator dengan variabel laten dan korelasi antar variabel laten. Hasil penelitian ini sejalan dengan penelitian Hwang dan Takane (2014), yang mana metode RGSCA berhasil mengatasi efek multikolinearitas. Hal tersebut dapat dilihat dari perbandingan signifikansi bobot dan *path* koefisien antara metode GSCA dan RGSCA yang telah dijelaskan pada Tabel 4.1 dan Tabel 4.6.

#### 4. KESIMPULAN

Berdasarkan hasil penelitian, diketahui terdapat indikasi multikolinearitas pada data faktor-faktor yang mempengaruhi kemiskinan Jawa Timur 2021. Terlihat dari tingginya korelasi antara  $z_1, z_2, z_3$  yaitu  $z_1$  dan  $z_2$  ( $r = 0,92$ ),  $z_1$  dan  $z_3$  ( $r = 0,81$ ), serta  $z_2$  dan  $z_3$  ( $r = 0,96$ ), serta tingginya korelasi antara variabel laten pendidikan dan variabel laten ekonomi ( $r = 0,818$ ). Penerapan RGSCA dengan parameter ridge  $\lambda_1 = 2, \lambda_2 = 3, \lambda_3 = 10$  menghasilkan nilai estimasi loading faktor, bobot, dan *path* koefisien signifikan pada interval 95%. Interpretasi hasil estimasi *path* koefisien adalah variabel pendidikan, kesehatan, dan ekonomi berpengaruh signifikan terhadap tingkat kemiskinan. Selain itu, variabel kesehatan dan ekonomi juga berpengaruh signifikan terhadap pendidikan, serta variabel kesehatan berpengaruh signifikan terhadap ekonomi. Hasil evaluasi model keseluruhan menunjukkan nilai FIT sebesar 0,662 artinya model mampu menjelaskan sekitar 66,2% variasi dari data..

#### DAFTAR PUSTAKA

- [1] Fakfare, P., Promsivapallop, P., dan Manosuthi, N., 2023 . Applying Integrated Generalized Structured Component Analysis to Explore Tourists' Benefit Consideration and Choice Confidence toward Travel Appscape. *Technological Forecasting and Social Change*. 188. 122321.
- [2] Gokmen, S., Dagalp, R., dan Kilickaplan, S., 2020. Multicollinearity in Measurement Error Models. *Communications in Statistics – theory ang Methods*. 51. 474-485.
- [3] Gujarati, D. N., 2004. *Basic Econometrics*, Fourth Edition. The McGraw Hill Companies Inc. New York.
- [4] Hastie, T., Tibshirani, R., & Friedman, J., 2001. *The elements of statistical learning: data mining, inference, and prediction*. Springer. New York.
- [5] Hox, J. dan Bechger, T., 1999. An Introduction to Strutural Equation Modeling. *Family Science Review*, 11, 354-373
- [6] Hwang, H. dan Takane, Y., 2004. Generalized Structured Component Analysis. *Psychometrika*, 69, 81-99.
- [7] Hwang, H., 2009. Regularized Generalized Structured Component Analysis. *Psychometrika* 74, 517–530.
- [8] Hwang, H. dan Takane, Y., 2014. *Generalized Structured Component Analysis: A Component-Based Approach to Structural Equation Modeling*. CRC Press, New York.
- [9] Marsh, H. W., Dowson, M., Pietsch, J., & Walker, R. 2004. Why Multicollinearity Matters: A Reexamination of Relations Between Self-Efficacy, Self-Concept, and Achievement. *Journal of Educational Psychology*, 96(3), 518–522.

- [10] Manosuthi, N., Lee, J.S., dan Han, H., 2020. An Innovative Application of Composite-Based Structural Equation Modeling in Hospitality Research with empirical Example. *Cornell Hospitality Quarterly*. 62(1). 139-156
- [11] Munerah, S., Koay, K.Y., dan Thambiah, S., 2021. Factors Influencing No-green Consumers' Purchase Intention : A Partial Least Squares Structural Equation Modeling (PLS-SEM) Approach. *Journal of Cleaner Production*. 280. 124192.
- [12] Nuhaliza, W., Kusnandar, D., & Perdana, H., 2022. Penerapan Structural Equation Modeling pada Analisis Kepuasan Mahasiswa Terhadap Sistem Informasi Akademik Universitas Tanjungpura. *Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster)*. Volume 11, No. 3 (2022), hal 513 – 522.
- [13] Putriama, J., Arina, F., & Ekawati., R., 2014. Pengaruh Penerapan Total Quality Management Melalui Produktivitas Karyawan Terhadap Kinerja Perusahaan Dengan Metode Structural Equation Modeling (SEM) di PT. X. *Jurnal Teknik Industri Untirta*. Volume 2, No. 2 (2014).
- [14] Raykov, T. dan Marcoulides, G.A., 2006. *A First Course in Structural Equation Modeling*, Second Edition, Lawrence Erlbaum Associates, London.
- [15] Siwalette, R., dkk, 2022. Analisis Faktor-Faktor Yang Berpengaruh Terhadap Pembelian Secara *Online* di Kota Ambon Menggunakan Metode *Structural Equation Modeling - Partial Least Square* (SEM-PLS). *Journal of statistics and Its Application*. Volume 4, No. 1 (2022).
- [16] Tenenhaus, M., 2014. Component-based Structural Equation Modeling. *Total quality management and business Excellence*. 19. 871-886.