**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

Modelos de Equações Estruturais: um estudo de simulação

Erasnilson Vieira Camilo

Dissertação apresentada para obtenção do título
de Mestre em Ciências. Área de concentração:
Estatística e Experimentação Agronômica

**Piracicaba
2015**

Erasnilson Vieira Camilo
Estatístico

Modelos de Equações Estruturais: um estudo de simulação
versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientadora:
Prof ^a Dr^a **ROSELI APARECIDA LEANDRO**

Dissertação apresentado para obtenção do título de Mestre
em Ciências. Área de concentração: Estatística e Experi-
mentação Agronômica

Piracicaba
2015

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP**

Camilo, Erasnilson Vieira

Modelos de Equações Estruturais: um estudo de simulação / Erasnilson Vieira Camilo. -
versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2015.
98 p. : il.

Dissertação (Mestrado) - - Escola Superior de Agricultura "Luiz de Queiroz".

1. Modelos de equações estruturais 2. Diagramas de caminhos 3. Simulação I. Título

CDD 519.5
C183m

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

DEDICATÓRIA

Aos meus pais, Erasmo (in memoriam) e Margarida

Ao meu irmão Nildo

AGRADECIMENTOS

Agradeço a Deus, por sempre me fortalecer e estar à frente de minhas vitórias.

A minha mãe, Margarida, ao meu irmão, Nildo, e meu pai, Erasmo (in memoriam), por todo apoio e compreensão, nessa fase da minha vida.

Aos professores: Carlos Tadeu, César Gonçalves, Clarice Garcia, Cristian Marcelo, Idemauro Antônio, Renata Alcarde, Sílvio Sandoval, Sônia Maria e Taciana Villela, por seus ensinamentos, orientações e por serem exemplos de seriedade e profissionalismo.

A professora Roseli Aparecida Leandro, por sua orientação, seriedade e confiança ao longo de todo o trabalho.

Aos amigos e familiares de Piracicaba, pela amizade e incentivo.

A todos os funcionários do departamento pela ajuda ao longo desses anos.

Aos professores da minha graduação, por terem acreditado em mim, e incentivado a continuar os estudos.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES, pelo apoio financeiro em forma de bolsa de estudos.

Em geral, agradeço a todos que contribuíram direta e indiretamente para a conclusão deste trabalho.

SUMÁRIO

RESUMO	9
ABSTRACT	11
LISTA DE FIGURAS	13
LISTA DE TABELAS	17
1 INTRODUÇÃO	19
2 REVISÃO BIBLIOGRÁFICA	21
2.1 MEE e o Diagrama de Caminhos	25
2.2 Causa e Correlação	27
2.3 Modelos Recursivos	33
2.4 Abordagem Matricial	35
2.4.1 Mínimos quadrados a dois passos	40
2.4.2 Índices de ajuste do modelo	42
3 MÉTODOS	45
3.1 Identificação do modelo	45
3.2 Método de Máxima Verossimilhança	45
3.3 Ajuste do modelo	46
4 RESULTADOS E DISCUSSÃO	47
5 CONCLUSÃO	85
REFERÊNCIAS	87
APÊNDICE	89

RESUMO

Modelos de Equações Estruturais: um estudo de simulação

Com a necessidade de analisar dados cada vez mais complexos nas mais diversas áreas, surge também a necessidade de novas técnicas e novas modelagens. Nesse contexto, os Modelos de Equações Estruturais são uma evolução da análise de caminhos proposto por Wright (1921), que por sua vez, consegue captar as relações de causa e efeito entre diferentes tipos de variáveis. Este trabalho tem como objetivo abordar uma revisão em torno dessa técnica, com ênfase aos modelos recursivos que utilizam em sua maioria apenas variáveis observadas. A aplicação deste trabalho está fundamentada em um processo de simulação considerando assim, seis cenários com diferentes tipos de relações numa mesma estrutura. As médias das estimativas dos parâmetros no processo de simulação resultam em valores próximos aos valores verdadeiros. Por meio de figuras e gráficos, pode-se observar o comportamento dos parâmetros por meio do erro quadrático médio e boxplot. A análise foi realizada com a utilização do *software* R e programas implementados com pacote **sem** (FOX; WEISBERG, 2012) e programação apresentada no Apêndice deste trabalho.

Palavras-chave: Modelos de equações estruturais; Diagramas de caminhos; Simulação

ABSTRACT

Structural Equation Model: a simulation study

With the increasingly need to analyze complex data in several areas, comes the need for new techniques and new models. In this context, the Structural Equation Modeling is an evolution of the path analysis proposed by Wright (1921), which, can capture the relations of cause and effect between different types of variables. The present work aims to approach a review about this technique with emphasis on recursive models using mostly observed variables only. The application of this work is based on a simulation process considering six scenarios with different types of relationships within the same structure. The mean estimates of the parameters in the simulation result in values that are close to the true values, and by means of figures and graphs, one can observe the behavior of the parameters by means of the mean squared error and boxplot. The analysis was performed using the software *software* R and implemented programs as **sem** package (FOX; WEISBERG, 2012) and the programming is presented in the Appendix of this work.

Keywords: Structural equation modeling; Path diagrams; Simulation

LISTA DE FIGURAS

Figura 1 - Diagrama de Caminho proposto por Duncan (1975)	25
Figura 2 - Diagrama de Caminho proposto por Duncan (1975)	29
Figura 3 - Diagrama de caminhos direcionados Grace (2006)	30
Figura 4 - Diagrama de caminhos multi-direcionados Grace (2006)	31
Figura 5 - Diagrama de Caminho proposto por Duncan (1975)	32
Figura 6 - Diagrama de Caminho proposto por Duncan (1975)	33
Figura 7 - Modelo estrutural simulado	38
Figura 8 - Modelo estrutural simulado (estrutura geral)	48
Figura 9 - Representação gráfica do Cenário I	54
Figura 10 - Representação gráfica do Cenário II	55
Figura 11 - Representação gráfica do Cenário III	56
Figura 12 - Representação gráfica do Cenário IV	57
Figura 13 - Representação gráfica Cenário V	58
Figura 14 - Representação gráfica do Cenário VI	59
Figura 15 - Erro quadrático médio associado aos parâmetros: (a) λ_{11} , (b) λ_{12} , (c) λ_{23} , (d) λ_{24} , (e) $\sigma_{X_1X_1}$, (f) $\sigma_{X_2X_2}$, (g) $\sigma_{X_3X_3}$, (h) $\sigma_{X_4X_4}$, (i) β_{21} , (j) $\sigma_{\epsilon_1\epsilon_1}$ e (k) $\sigma_{\epsilon_2\epsilon_2}$	61
Figura 16 - Gráfico boxplot das estimativas dos parâmetros do Cenário I, com amos- tras de $n = 50$, em mil iterações	62
Figura 17 - Histograma para simulação do Cenário I: (a) λ_{11} , (b) λ_{12} , (c) λ_{23} , (d) λ_{24} (e) $\sigma_{X_1X_1}$, (f) $\sigma_{X_2X_2}$, (g) $\sigma_{X_3X_3}$, (h) $\sigma_{X_4X_4}$, (i) β_{21}	63
Figura 18 - Histograma para simulação do Cenário I: (a) $\sigma_{\epsilon_1\epsilon_1}$ e (b) $\sigma_{\epsilon_2\epsilon_2}$	64
Figura 19 - Erro quadrático médio associado aos parâmetros: (a) λ_{11} , (b) λ_{12} , (c) λ_{23} , (d) λ_{24} , (e) $\sigma_{X_1X_1}$, (f) $\sigma_{X_2X_2}$, (g) $\sigma_{X_3X_3}$, (h) $\sigma_{X_4X_4}$, (i) β_{21} , (j) $\sigma_{X_1X_2}$, (k) $\sigma_{\epsilon_1\epsilon_1}$ e (l) $\sigma_{\epsilon_2\epsilon_2}$	65
Figura 20 - Gráfico boxplot das estimativas dos parâmetros do Cenário II, com amos- tras de $n = 50$, em mil iterações	66
Figura 21 - Histograma para simulação do Cenário II: (a) λ_{11} , (b) λ_{12} , (c) λ_{23} , (d) λ_{24} , (e) $\sigma_{X_1X_1}$, (f) $\sigma_{X_2X_2}$, (g) $\sigma_{X_3X_3}$, (h) $\sigma_{X_4X_4}$ e (i) β_{21}	67
Figura 22 - Histograma para simulação do Cenário II: (a) $\sigma_{X_1X_2}$, (b) $\sigma_{\epsilon_1\epsilon_1}$ e (c) $\sigma_{\epsilon_2\epsilon_2}$	68

Figura 23 - Erro quadrático médio associado aos parâmetros: (a) λ_{11} , (b) λ_{12} , (c) λ_{23} , (d) λ_{24} , (e) $\sigma_{X_1X_1}$, (f) $\sigma_{X_2X_2}$, (g) $\sigma_{X_3X_3}$, (h) $\sigma_{X_4X_4}$, (i) β_{21} , (j) $\sigma_{X_1X_2}$, (k) $\sigma_{\epsilon_1\epsilon_1}$ e (l) $\sigma_{\epsilon_2\epsilon_2}$	69
Figura 24 - Histograma para simulação do Cenário III: (a) λ_{11} , (b) λ_{12} , (c) λ_{23} , (d) λ_{24} , (e) $\sigma_{X_1X_1}$, (f) $\sigma_{X_2X_2}$, (g) $\sigma_{X_3X_3}$, (h) $\sigma_{X_4X_4}$ e (i) β_{21}	70
Figura 25 - Histograma para simulação do Cenário III: (a) $\sigma_{X_1X_2}$, (b) $\sigma_{\epsilon_1\epsilon_1}$ e (c) $\sigma_{\epsilon_2\epsilon_2}$	71
Figura 26 - Gráfico boxplot das estimativas dos parâmetros do Cenário III, com amos- tras de $n = 50$, em mil iterações	71
Figura 27 - Erro quadrático médio associado aos parâmetros: (a) λ_{11} , (b) λ_{12} , (c) λ_{23} , (d) λ_{24} , (e) $\sigma_{X_1X_1}$, (f) $\sigma_{X_2X_2}$, (g) $\sigma_{X_3X_3}$, (h) $\sigma_{X_4X_4}$, (i) β_{21} , (j) $\sigma_{X_3X_4}$, (k) $\sigma_{\epsilon_1\epsilon_1}$ e (l) $\sigma_{\epsilon_2\epsilon_2}$	72
Figura 28 - Histograma para simulação do Cenário IV: (a) λ_{11} , (b) λ_{12} , (c) λ_{23} , (d) λ_{24} , (e) $\sigma_{X_1X_1}$, (f) $\sigma_{X_2X_2}$, (g) $\sigma_{X_3X_3}$, (h) $\sigma_{X_4X_4}$ e (i) β_{21}	73
Figura 29 - Histograma para simulação do Cenário IV: (a) $\sigma_{X_3X_4}$, (b) $\sigma_{\epsilon_1\epsilon_1}$ e (c) $\sigma_{\epsilon_2\epsilon_2}$	74
Figura 30 - Gráfico boxplot das estimativas dos parâmetros do Cenário IV, com amos- tras de $n = 50$, em mil iterações	74
Figura 31 - Erro quadrático médio associado aos parâmetros: (a) λ_{11} , (b) λ_{12} , (c) λ_{23} , (d) λ_{24} , (e) $\sigma_{X_1X_1}$, (f) $\sigma_{X_2X_2}$, (g) $\sigma_{X_3X_3}$, (h) $\sigma_{X_4X_4}$, (i) β_{21} , (j) $\sigma_{X_3X_4}$, (k) $\sigma_{\epsilon_1\epsilon_1}$ e (l) $\sigma_{\epsilon_2\epsilon_2}$	75
Figura 32 - Histograma para simulação do Cenário V: (a) λ_{11} , (b) λ_{12} , (c) λ_{23} , (d) λ_{24} , (e) $\sigma_{X_1X_1}$, (f) $\sigma_{X_2X_2}$, (g) $\sigma_{X_3X_3}$, (h) $\sigma_{X_4X_4}$ e (i) β_{21}	76
Figura 33 - Histograma para simulação do Cenário V: (a) $\sigma_{X_3X_4}$, (b) $\sigma_{\epsilon_1\epsilon_1}$ e (c) $\sigma_{\epsilon_2\epsilon_2}$	77
Figura 34 - Gráfico boxplot das estimativas dos parâmetros do Cenário V, com amos- tras de $n = 50$, em mil iterações	77
Figura 35 - Erro quadrático médio associado aos parâmetros: (a) λ_{11} , (b) λ_{12} , (c) λ_{23} , (d) λ_{24} , (e) $\sigma_{X_1X_1}$, (f) $\sigma_{X_2X_2}$, (g) $\sigma_{X_1X_2}$, (h) $\sigma_{X_3X_3}$, (i) $\sigma_{X_4X_4}$, (j) $\sigma_{X_3X_4}$ e (k) β_{21}	78
Figura 36 - Erro quadrático médio associado aos parâmetros: $\sigma_{\epsilon_1\epsilon_1}$ (a) e $\sigma_{\epsilon_2\epsilon_2}$ (b)	79
Figura 37 - Histograma para simulação do Cenário VI: (a) λ_{11} , (b) λ_{12} , (c) λ_{23} , (d) λ_{24} , (e) $\sigma_{X_1X_1}$, (f) $\sigma_{X_2X_2}$, (g) $\sigma_{X_1X_2}$, (h) $\sigma_{X_3X_3}$ e (i) $\sigma_{X_4X_4}$	80
Figura 38 - Histograma para simulação do Cenário VI: (a) $\sigma_{X_3X_4}$, (b) β_{21} , (c) $\sigma_{\epsilon_1\epsilon_1}$ e (d) $\sigma_{\epsilon_2\epsilon_2}$	81

Figura 39 - Gráfico boxplot das estimativas dos parâmetros do Cenário VI, com amostras de $n = 50$, em mil iterações 82

LISTA DE TABELAS

Tabela 1 - Correlações entre variáveis da Figura 3	30
Tabela 2 - Correlações entre variáveis da Figura 4	31
Tabela 3 - Estimativa dos parâmetros do modelo para os diferentes tamanhos amostrais (Cenário I)	55
Tabela 4 - Estimativa dos parâmetros do modelo para os diferentes tamanhos amostrais (Cenário II)	56
Tabela 5 - Estimativa dos parâmetros do modelo para os diferentes tamanhos amostrais (Cenário III)	57
Tabela 6 - Estimativa dos parâmetros do modelo para os diferentes tamanhos amostrais (Cenário IV)	58
Tabela 7 - Estimativa dos parâmetros do modelo para os diferentes tamanhos amostrais (Cenário V)	59
Tabela 8 - Estimativa dos parâmetros do modelo para os diferentes tamanhos amostrais (Cenário VI)	60

1 INTRODUÇÃO

Diferentes pesquisas e análises estatísticas surgem da necessidade do pesquisador em solucionar fenômenos de natureza quantitativa ou qualitativa. Com a evolução tecnológica e o aprimoramento das ferramentas estatísticas, é possível analisar dados cada vez mais complexos envolvendo diferentes tipos de variáveis, observadas ou latentes. Assim, por meio dos métodos inferenciais, é possível obter respostas a diversos problemas, para isso, estudos envolvendo relações simultâneas ou dependências de causa e efeito podem ser abordados por meio dos Modelos de Equações Estruturais.

O surgimento dessa técnica dá-se com uma evolução da análise de caminhos de Wright (1921) ao relacionar diagramas com correlações entre as variáveis. Posteriormente, trabalhos como os de Jöreskog e Sörbom (1993), deram uma nova visão a análise de caminhos, cuja evolução são os modelos de equações estruturais, no qual é possível relacionar diferentes tipos de variáveis.

Trabalhos retratados por Bollen (1989), Hair Jr. et al. (2010) e Grace (2006), ilustram e exemplificam a aplicabilidade desses modelos em diferentes áreas, como por exemplo: (i) na economia, ao analisar equações simultâneas envolvendo despesas e consumo, ou produção demandada e ofertada (FOX, 2008); (ii) na sociologia ao estudar relações sobre influências dos colegas sobre a aspiração dos estudantes do ensino médio (DUNCAN; HALLER; PORTES, 1968); na genética, ao estudar relações causais entre diferentes genótipos (ROSA et al., 2011). De forma geral, esse tipo de análise utiliza mais de uma equação na modelagem o que a torna mais complexa.

Os modelos de equações estruturais são geralmente compostos por um diagrama de caminhos que relacionam causa e efeito entre as variáveis, dos quais, destacam-se os modelos recursivos por apresentarem diagramas de caminhos unidirecionados, enquanto que os modelos não-recursivos apresentam caminhos de loops e feedback em sua estrutura. Existem, ainda, os que envolvem variáveis latentes como é o caso dos modelos de equações lineares estruturais (LISREL).

Neste trabalho, será abordada uma revisão metodológica sobre o tema dos modelos de equações estruturais, o surgimento, as influências e aplicação por meio de alguns cenários que ilustrarão o processo de simulação do mesmo. Os Conceitos fundamentais, como diagramas de caminhos e distinção entre modelos, serão destacados, com ênfase aos

modelos recursivos, que segundo Duncan (1975), são aqueles em que as variáveis são afetadas de uma única forma.

O objetivo deste trabalho está direcionado em:

- i) revisão metodológica e matemática sobre Modelos de Equações Estruturais;
- ii) aplicação dos modelos recursivos por meio de alguns cenários que ilustrarão o processo de simulação do mesmo;
- iii) estudo de simulação utilizando o pacote estatístico **sem** (FOX; WEISBERG, 2012).

2 REVISÃO BIBLIOGRÁFICA

Em diversas áreas da ciência, problemas que envolvem uma ou mais variáveis são geralmente resolvidos com uso de técnicas estatísticas bem conhecidas. Na análise de regressão linear simples, por exemplo, uma variável Y_i (dependente) pode ser explicada por uma outra variável (explanatória) mais um erro ϵ_i caracterizado por ser independente, normalmente distribuído com média $E(\epsilon_i) = 0$ e variância constante σ^2 . Aplicações envolvendo altura e peso, comprimento e largura, são exemplos clássicos. Quanto à especificação do modelo, os parâmetros são considerados constantes e a variável explanatória (independente) considerada fixa (SOUZA, 1998). Neste caso, a estimação dos parâmetros pode ser realizada por meio do método dos mínimos quadrados ou da máxima verossimilhança (DRAPER; SMITH, 1981), (MONTGOMERY; PECK; VINING, 2012).

De forma análoga à regressão linear simples, a análise de regressão múltipla, destaca-se por ser uma técnica em que a variável dependente Y_i é função linear de duas ou mais variáveis explanatórias. Embora exista uma grande utilização dos métodos mencionados anteriormente, em casos específicos, como na agronomia, por exemplo, em que há interesse em estudar doses de fertilizantes em uma determinada plantação, faz-se necessário utilizar modelos polinomiais, que segundo Rawlings, Pantula e Dickey (1998) não se restringe apenas a uma reta, podendo ser representada também por uma função polinomial das variáveis explanatórias e função linear dos parâmetros.

Métodos como de mínimos quadrados e da máxima verossimilhança são os mais utilizados no processo de estimação dos parâmetros, porém, surge uma outra problemática: “Como selecionar o melhor modelo?”. Esse questionamento é muito frequente na vida do pesquisador e uma prática recorrente é a utilização e comparação do R^2 (coeficiente de determinação) entre modelos, como uma forma de verificar o quanto o modelo explica o comportamento dos dados, porém, é necessário um maior cuidado em sua utilização, pois muitas vezes um R^2 alto não implica em um modelo bem ajustado aos dados e nem com melhores interpretações práticas.

Assim, o processo de avaliação dá-se por meio da análise dos resíduos, verificando, se o mesmo satisfaz as pressuposições estabelecidas, ou seja, a independência, a normalidade e a homoscedasticidade.

Outras formas de comparação entre modelos, bem utilizadas na prática, são os

critérios de informação e os mais conhecidos são: critério de informação de Akaike (AIC) e suas extensões (AKAIKE, 1974); critério de informação Bayesiano (BIC) (SCHWARZ et al., 1978), sendo ambos função do estimador de máxima verossimilhança e penalizam modelos com maior número de parâmetros. Na comparação, escolhe-se aquele que apresente menor valor do critério pré-estabelecido pelo pesquisador, que pode ser menor AIC ou menor BIC.

Enquanto outras técnicas estatísticas não conseguem modelar causa-efeito entre as variáveis, na análise dos Modelos de Equações Estruturais isso é possível. Neles é possível considerar a dependência de uma determinada variável resposta como função das outras variáveis, explicativas ou não. Assim, Surgem nomenclaturas diferentes das usuais, as variáveis respostas serão denominadas de variáveis endógenas e podem aparecer em outra equação do modelo como variável explanatória ou variável exógena.

Para Hair Jr. et al. (2010), Mulaik (2009) e Fox (2008), os modelos de equações estruturais (MEE) não se restringem apenas a uma técnica, mas a uma família de modelos reconhecida por diferentes nomenclaturas, dentre elas, análise de caminho, análise fatorial confirmatória ou explanatória e modelos de equações lineares estruturais (LISREL).

Segundo Pugesek e Tomer (2003), a história dos MEE reside em meio à complexidade da área estatística, metodológica e filosófica. Para tanto, é necessário um conhecimento prévio sobre ideias conceituais; causalidade e algumas características estatísticas, como: modelos de regressão múltiplas, equações normais e alguns métodos de estimação.

A crescente utilização desse método deve-se ao fato de que os MEE englobam diversas características que não estão presentes em outras modelagens. Por exemplo, na análise de regressão, a variável resposta não pode ser variável explicativa no mesmo modelo, enquanto que nos MEE isso é possível por meio das variáveis exógenas e endógenas a serem identificadas na modelagem.

Numa visão mais ampla, Pugesek e Tomer (2003) exibem algumas especificações interessantes sobre os MEE:

- i) os MEE representam traduções de uma série de relações causa e efeito entre as variáveis;
- ii) as relações são descritas por meio de parâmetros que indicam a magnitude do efeito (direto ou indireto) que as variáveis independentes (seja observada ou latente) têm sobre as variáveis dependentes (observadas ou latentes);

- iii) os MEE oferecem aos pesquisadores um método abrangente para a quantificação e teste de modelos teóricos.

Para a utilização dos MEE, é necessário uma gama de informações disponíveis sobre as variáveis envolvidas, sejam elas exógenas ou endógenas, além de uma estrutura pre-estabelecida pelo pesquisador, ou definida por meio de técnicas exploratórias, que auxiliem o processo de busca das relações causais. Assim, Bollen (1989) apresenta três componentes indispensáveis em Modelos de Equações Estruturais: (i) análise de caminhos; (ii) variáveis latentes e modelos de mensuração e (iii) métodos de estimação, componentes estes que servirão de base para análise completa dos modelos de equações estruturais.

Segundo Bollen (1989), Pugesek e Tomer (2003) e Horvath (2011), Sewall Wright foi o precursor da análise de caminhos, sendo esta inicialmente desenvolvida entre o período de 1918 a 1960, com trabalhos relacionados a crescimento da população de coelhos (WRIGHT, 1918). No entanto, Iriondo, Albert e Escudero (2003) afirmam que os trabalhos iniciais de Sewall Wright com enfoque em MEE surgiram entre 1921 a 1934, em estudos de covariâncias entre características de porcos *Guinea*, nos quais Wright desenvolveu algumas formas de transformar as correlações observadas em um sistema de equações que descrevessem as hipóteses sobre relações causais representadas por um diagrama de caminhos.

A análise de caminhos, por sua vez, contribuiu para a construção das equações estruturais que relacionam causa e efeito entre as variáveis endógenas e exógenas, sendo possível estudar efeitos diretos e indiretos de uma variável sobre outra.

Segundo Smith et al. (2014), as variáveis podem ser medidas diretamente, como é o caso da altura das plantas ou da massa dos grãos e para esses tipos de variáveis atribui-se o nome de variáveis manifestas. As variáveis de natureza conceitual ou teórica, que não podem ser mensuradas diretamente, são denominadas de variáveis latentes, como: nível de inteligência, interesse profissional e satisfação do cliente. Nesse cenário, Wang e Fan (2011) afirmam que a presença de variáveis latentes, além de fornecerem previsões mais precisas, também mostram relações importantes sobre os dados. Segundo Palomo, Dunson e Bollen (2007), estudos envolvendo variáveis latentes fornecem uma estrutura parcimoniosa para a modelagem da matriz de covariância.

Os modelos de equações estruturais são geralmente compostos por dois componentes: modelo de mensuração e equação estrutural. Segundo Pugesek e Tomer (2003) e

Lee (2007), para os mensuração, há uma ligação entre as equações por meio de parâmetros estruturais que ligam fatores endógenos às variáveis exógenas, já na equação estrutural, relacionam-se os fatores latentes entre si.

Rosa et al. (2011) e Fox (2008) abordam a utilização desses modelos na genética quantitativa quando há interesse em estudar as relações simultâneas recursivas entre fenótipos em sistemas multivariados. Assim, Fox (2008) descreve as seguintes características para as variáveis endógenas e exógenas:

- i) há uma equação estrutural (equação de regressão) para cada variável endógena;
- ii) uma variável endógena pode, também, ser uma variável explicativa em uma ou mais equações estruturais;
- iii) as variáveis endógenas são quantitativas contínuas;
- iv) as variáveis exógenas, podem ser de natureza contínua, discreta ou categórica.

Fox (2008) destaca ainda, que os erros estruturais ou erros de medição representam a parte não explicada pelo modelo sobre as variáveis endógenas, e esses apresentam algumas características:

- i) existe um erro para cada variável endógena;
- ii) assume-se que os erros possuem esperança zero e que sejam independentes da variável endógena;
- iii) os erros, para diferentes observações, são considerados independentes um dos outros, porém, dependendo da estrutura do modelo, podem ser correlacionados.

De forma prática e de leitura simplificada, Bollen (1989) aborda os Modelos de Equações Estruturais com exemplificações e demonstrações em forma matricial. O auxílio computacional para utilização deste tipo de análise está bem acessível a estudantes e pesquisadores, por meio dos *softwares*: AMOS, EQS, LISREL, Mplus, SAS, e R. Destes, destaca-se o *software* R Core Team (2013) que é de acesso livre e com suporte para os pacotes: *lavaan* (ROSSEEL, 2012) e *sem* (FOX; NIE; BYRNES, 2013).

Em síntese, os MEE são formados por equações múltiplas que representam relações causais, em que a variável endógena em uma determinada equação pode também

ser variável exógena em outra equação estrutural. Uma forma de estimar os parâmetros nos MEE dá-se por meio de alguns métodos, sendo os mais conhecidos: método de mínimos quadrados e método da máxima verossimilhança.

Segundo Mulaik (2009), somente após Bock e Bargmann, em 1966, desenvolverem um método para testar MEE com covariância entre as variáveis é que Karl Jöreskog, em 1969, propôs o modelo de análise fatorial confirmatória geral, descrevendo assim, soluções matemáticas para estimativas de máxima verossimilhança dos parâmetros do modelo.

Segundo Bollen (1989), a abordagem dos MEE está relacionada a procedimentos realizados na matriz de covariâncias, ou seja, ao invés de minimizar a soma de quadrados das diferenças entre valores observados e ajustados como na análise de regressão, nos MEE, a hipótese fundamental está interligada com a matriz de covariância das variáveis observadas. Assim, a hipótese a ser analisada refere-se a seguintes expressão: $\Sigma(\theta) = \Sigma$, em que Σ é a matriz de covariâncias populacional e $\Sigma(\theta)$ é a matriz de covariâncias estimada pelo modelo.

Quanto aos métodos de estimação, pretende-se minimizar a diferença entre a relação da matriz de variância das variáveis observadas e a matriz de variância estimada pelo modelo. Nesse contexto, métodos como o da máxima verossimilhança, mínimos quadrados não ponderados e mínimos quadrados generalizados são comumente utilizados.

2.1 MEE e o Diagrama de Caminhos

Os modelos de equações estruturais são representados graficamente por um diagrama de caminhos que exhibe as relações entre as variáveis por meio de linhas e setas. Para Fox e Weisberg (2012), as linhas do diagrama de caminhos relacionam as variáveis através de parâmetros estruturais em que as setas são identificadas como causa e efeito de uma variável sobre outra, sendo representadas por linhas com uma única seta indicadora, enquanto as setas com dupla indicação determinam presença de correlação entre as variáveis.

Figura 1 - Diagrama de Caminho proposto por Duncan (1975)

O diagrama, ilustrado na Figura 1, será útil em algumas demonstrações e na abordagem para a simulação dos MEE. Considera-se, ainda, que ε possui distribuição normal com média $E(\varepsilon) = 0$ e variância $Var(\varepsilon) = \sigma_{\varepsilon\varepsilon}$, assume-se, também, que $E(X) = 0$, devido ao fato de se trabalhar com variáveis centradas em torno da média, com $Var(X) = \sigma_{XX}$, e portanto, tem-se que $E(Y) = 0$ e $Var(Y) = \sigma_{YY}$.

Na Figura 1, apresenta-se um diagrama de caminhos com duas variáveis: X e Y , representando, respectivamente, a variável exógena e variável endógena do modelo, e ε o erro ao acaso com $E(\varepsilon) = 0$ e $Var(\varepsilon) = \sigma_{\varepsilon\varepsilon}$. O modelo matemático associado ao diagrama de caminho apresentado na Figura 1 é descrito pela Equação (1):

$$Y = \lambda_{YX}X + \varepsilon, \quad (1)$$

em que X é vista como “causa” e Y como “efeito”, ambas centradas em torno da média, e λ_{YX} é o coeficiente estrutural do modelo.

Para a estimação do coeficiente estrutural λ_{YX} , assumem-se que: (i) as variáveis X e Y estão centradas em torno da média, isto é, $E(X) = E(Y) = 0$; (ii) X e ε são não correlacionados, logo, $E(\varepsilon X) = E(\varepsilon Y) = 0$ e tem-se:

$$\begin{aligned} \sigma_{YY} &= E(Y^2) & \sigma_{XX} &= E(X^2) & \sigma_{YX} &= E(YX) \\ \sigma_{X\varepsilon} &= E(X\varepsilon) & \sigma_{\varepsilon\varepsilon} &= E(\varepsilon^2) & \sigma_{Y\varepsilon} &= E(Y\varepsilon), \end{aligned}$$

uma vez que,

$$Cov(YX) = E(YX) - E(Y)E(X) = E(YX).$$

Multiplicando-se a Equação (1) por ε e calculando-se a esperança matemática,

$$E(Y\varepsilon) = \lambda_{YX}E(X\varepsilon) + E(\varepsilon\varepsilon)$$

obtem-se que a variância do erro é $\sigma_{\varepsilon\varepsilon}$, $E(Y\varepsilon) = E(\varepsilon\varepsilon)$, ou seja, $\sigma_{Y\varepsilon} = \sigma_{\varepsilon\varepsilon}$.

De forma análoga, ao multiplicar-se por X e Y a Equação (1) e calcular a esperança matemática, encontra-se a seguinte expressão:

$$E(Y^2) = \lambda_{YX}E(XY) + E(\varepsilon Y)$$

$$\sigma_{YY} = \lambda_{YX}\sigma_{YX} + \sigma_{\varepsilon\varepsilon}$$

e

$$E(YX) = \lambda_{YX}E(X^2) + E(\varepsilon X)$$

$$\sigma_{YX} = \lambda_{YX}\sigma_{XX}$$

$$\lambda_{YX} = \frac{\sigma_{XY}}{\sigma_{XX}}$$

pode-se, ainda, reescrever a variância σ_{YY} em função de X por meio das propriedade de covariância, ou seja,

$$\begin{aligned} Cov(X, Y) &= E(XY) - E(X)E(Y) \\ \sigma_{XY} &= \lambda_{YX}\sigma_{XX} \end{aligned} \quad (2)$$

e a variância de Y dada por:

$$\sigma_{YY} = \lambda_{YX}^2\sigma_{XX} + \sigma_{\varepsilon\varepsilon} \quad (3)$$

em que λ_{YX} (coeficiente estrutural) dado por $\lambda_{YX} = \frac{\sigma_{XY}}{\sigma_{XX}}$. As especificações do modelo serão apresentadas na forma escalar como uma breve introdução ao tema. No decorrer deste trabalho serão apresentadas definições e demonstrações numa abordagem matricial e processos de simulação realizados para a confirmação do mesmo.

2.2 Causa e Correlação

Considerando-se que as variáveis X e Y são centradas sobre as médias e, ao padronizar a Equação (1), ou seja, multiplicar os termos do modelo pelo inverso do desvio padrão da variável dependente Y e com algumas propriedades algébricas, é possível reescrever o modelo em termos de coeficientes de caminhos com variáveis padronizadas, ou seja,

$$\frac{Y}{\sigma_Y} = \lambda_{YX} \frac{\sigma_X}{\sigma_X} \frac{X}{\sigma_Y} + \frac{\sigma_\varepsilon}{\sigma_\varepsilon} \frac{\varepsilon}{\sigma_Y} \quad (4)$$

em que, $\lambda_{YX} = \frac{\sigma_{XY}}{\sigma_{XX}}$ e σ_{XX} é a variância da variável X e σ_X seu desvio-padrão, $\sigma_{\varepsilon\varepsilon}$ é a variância dos erros aleatórios ε e σ_ε seu desvio-padrão.

$$\frac{Y}{\sigma_Y} = \frac{\sigma_{XY}}{\sigma_{XX}} \frac{\sigma_X}{\sigma_X} \frac{X}{\sigma_Y} + \frac{\sigma_\varepsilon}{\sigma_\varepsilon} \frac{\varepsilon}{\sigma_Y}$$

$$\frac{Y}{\sigma_Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \frac{\sigma_X}{\sigma_{XX}} X + \frac{\sigma_\varepsilon}{\sigma_Y} \frac{\varepsilon}{\sigma_\varepsilon}$$

Como $\sigma_{Y\varepsilon} = \sigma_{\varepsilon\varepsilon}$, as correlações dadas por ρ_{YX} e $\rho_{Y\varepsilon}$ são denotadas por,

$$\rho_{Y\varepsilon} = \frac{\sigma_{\varepsilon Y}}{\sigma_{\varepsilon}\sigma_Y} = \frac{\sigma_{\varepsilon\varepsilon}}{\sigma_{\varepsilon}\sigma_Y} = \frac{\sigma_{\varepsilon}}{\sigma_Y}$$

$$\rho_{YX} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y}$$

e assim, o modelo matemático associado ao diagrama de caminhos, apresentado na Figura 1, pode ser reescrito em função das variáveis padronizadas, como:

$$Y^* = \rho_{YX}X^* + \rho_{Y\varepsilon}\varepsilon^* \quad (5)$$

em que $Y^* = \frac{Y}{\sigma_Y}$, $X^* = \frac{X}{\sigma_X}$ e $\varepsilon^* = \frac{\varepsilon}{\sigma_{\varepsilon}}$, e assim,

$$E(Y^*) = E\left(\frac{Y}{\sigma_Y}\right) = 0$$

$$Var(Y^*) = Var\left(\frac{Y}{\sigma_Y}\right)$$

$$Var(Y^*) = \frac{1}{\sigma_Y^2}Var(Y)$$

$$Var(Y^*) = 1$$

De forma similar $E(Y^{*2}) = E(X^{*2}) = E(\varepsilon^{*2}) = 1$ e a $Cov(X^*, Y^*) = E(X^*Y^*)$, expressada por:

$$E(X^*Y^*) = \rho_{YX}E(X^{*2}) + \rho_{Y\varepsilon}E(X^*\varepsilon^*)$$

$$E(X^*Y^*) = \rho_{YX}\frac{1}{\sigma_{XX}}E(X^2) + \rho_{Y\varepsilon}E\left(\frac{X}{\sigma_X}\frac{\varepsilon}{\sigma_{\varepsilon}}\right)$$

$$E(X^*Y^*) = \rho_{YX}\frac{1}{\sigma_{XX}}\sigma_{XX} + \rho_{Y\varepsilon}\frac{1}{\sigma_X\sigma_{\varepsilon}}E(X\varepsilon)$$

$$E(X^*Y^*) = \rho_{YX}$$

de forma análoga,

$$E(Y^{*2}) = \rho_{YX}E(Y^*X^*) + \rho_{Y\varepsilon}E(Y^*\varepsilon^*)$$

$$1 = \rho_{YX}\frac{1}{\sigma_Y\sigma_X}E(YX) + \rho_{Y\varepsilon}\frac{1}{\sigma_Y\sigma_{\varepsilon}}E(Y\varepsilon)$$

$$1 = \rho_{YX}\frac{1}{\sigma_Y\sigma_X}(\lambda_{YX}\sigma_{XX}) + \rho_{Y\varepsilon}\frac{1}{\sigma_Y\sigma_{\varepsilon}}\sigma_{\varepsilon Y}$$

$$1 = \rho_{YX}\frac{1}{\sigma_Y\sigma_X}\left(\frac{\sigma_{YX}}{\sigma_{XX}}\sigma_{XX}\right) + \rho_{Y\varepsilon}\rho_{Y\varepsilon}$$

$$1 = \rho_{YX}\rho_{YX} + \rho_{Y\varepsilon}\rho_{Y\varepsilon}$$

$$1 = \rho_{YX}^2 + \rho_{Y\varepsilon}^2.$$

Considerando-se agora três variáveis padronizadas: X , Y e Z , com $E(X) =$

$E(Y) = E(Z) = E(\varepsilon_1) = E(\varepsilon_2) = 0$, e diagrama de caminhos apresentado na Figura 2, os coeficientes de caminhos a serem estimados são: λ_{YX} , $\lambda_{Y\varepsilon_1}$, β_{ZY} e $\lambda_{Z\varepsilon_2}$.

Figura 2 - Diagrama de Caminho proposto por Duncan (1975)

Assumindo-se que $\lambda_{\varepsilon_1 X} = \lambda_{\varepsilon_2 X} = \lambda_{\varepsilon_2 Y} = 0$, o modelo matemático associado ao diagrama de caminhos apresentado na Figura 2 pode ser escrito por meio do seguinte sistema de equações estruturais:

$$\begin{cases} Y = \lambda_{YX}X + \lambda_{Y\varepsilon_1}\varepsilon_1 \\ Z = \beta_{ZY}Y + \lambda_{Z\varepsilon_2}\varepsilon_2 \end{cases} \quad (6)$$

ao substituir a primeira equação do sistema na segunda equação, obtém-se:

$$\begin{aligned} Z &= \beta_{ZY}(\lambda_{YX}X + \lambda_{Y\varepsilon_1}\varepsilon_1) + \lambda_{Z\varepsilon_2}\varepsilon_2 \\ Z &= \beta_{ZY}\lambda_{YX}X + \beta_{ZY}\lambda_{Y\varepsilon_1}\varepsilon_1 + \lambda_{Z\varepsilon_2}\varepsilon_2 \end{aligned} \quad (7)$$

logo, o efeito indireto de X sobre Z é dado da seguinte forma:

$$\begin{aligned} E(ZX) &= \beta_{ZY}\lambda_{YX}E(X^2) + \beta_{ZY}\lambda_{Y\varepsilon_1}E(\varepsilon_1X) + \lambda_{Z\varepsilon_2}E(\varepsilon_2X) \\ E(ZX) &= \beta_{ZY}\lambda_{YX} \\ \rho_{ZX} &= \beta_{ZY}\lambda_{YX} \end{aligned}$$

em que, $E(X^2) = 1$, $E(\varepsilon_i X) = 0$ com $i = 1, 2$.

No estudo dos modelos de equações estruturais, pode-se observar que mesmo quando duas variáveis não estão ligadas por uma seta, como ocorre com X e Z na Figura 2, não necessariamente uma variável deixa de afetar a outra. Assim, Fox (2008) afirma que os efeitos indiretos podem ser identificados com uma espécie de “caminhos compostos” através do diagrama de caminhos.

Grace (2006) classifica os diagramas de caminhos como: diagramas direcionados e não direcionados. De forma simplificada, os diagramas direcionados destacam-se por apresentar setas com uma única indicação (Figura 3), sendo possível identificar a variável

exógena X_1 e as variáveis endógenas Y_1 e Y_2 , e o erro associado a cada variável endógena representado por ε_i com $i = 1, 2$.

Figura 3 - Diagrama de caminhos direcionados Grace (2006)

Considerando-se valores normalizados das variáveis observadas (Tabela 1) e o diagrama de caminhos da Figura 3, o coeficiente de caminhos que relaciona X_1 até Y_1 é representado pela correlação entre X_1 e Y_1 (0,50). De forma análoga, o coeficiente de Y_1 para Y_2 é dado pela correlação entre Y_1 e Y_2 (0,60).

Tabela 1 - Correlações entre variáveis da Figura 3

	X_1	Y_1	Y_2
X_1	1,00		
Y_1	0,50	1,00	
Y_2	0,30	0,60	1,00

Segundo Grace (2006), a relação dada entre X_1 e Y_2 leva a uma regra fundamental na análise de caminhos, que é o produto entre os coeficientes de caminhos ao longo de um percurso (caminhos compostos). Logo, tem-se caminhos compostos a partir de X_1 para Y_2 como o produto de (0,50) vezes (0,60), o que equivale a (0,30).

Na Figura 3, observa-se as relações diretas de uma variável sobre outra, com uma única indicação. No entanto, há casos em que a mesma variável pode indicar mais de uma, como no diagrama apresentado na Figura 4.

Figura 4 - Diagrama de caminhos multi-direcionados Grace (2006)

Neste caso, ao considerar os valores da Tabela 2, observa-se a divergência sobre o efeito direto entre X_1 e Y_2 de (0,50) com o produto do caminho indireto, dado de X_1 a Y_1 e Y_1 a Y_2 (0,30).

Tabela 2 - Correlações entre variáveis da Figura 4

	X_1	Y_1	Y_2
X_1	1,00		
Y_1	0,55	1,00	
Y_2	0,50	0,60	1,00

Portanto, como visto anteriormente, devido ao novo caminho direcionado (X_1 e Y_2), parte-se para uma nova regra, que, segundo Grace (2006), quando duas variáveis estão ligadas por mais de um percurso de causalidade, deve-se utilizar o cálculo dos coeficientes de regressão parciais. Logo, uma regressão parcial ou correlação parcial, entre duas variáveis, explica a influência de variáveis que afetam determinadas variáveis. Os coeficientes de regressão e de correlação parciais surgem quando duas variáveis estão conectadas através de uma via causal.

Assim, quando há correlação entre os erros, pode-se utilizar a correlação parcial para estimar o parâmetro desconhecido. Logo, por meio do sistema de equações dado na Equação (8), tem-se o diagrama na Figura 5.

Figura 5 - Diagrama de Caminho proposto por Duncan (1975)

$$\begin{cases} Y = \lambda_{YX}X + \lambda_{Y\varepsilon_1}\varepsilon_1 \\ Z = \lambda_{ZX}X + \lambda_{Z\varepsilon_2}\varepsilon_2 \end{cases} \quad (8)$$

Ao isolar os termos dos erros da Equação (8) e calcular a esperança do produto dentre eles, tem-se:

$$\varepsilon_1 = \frac{Y - \lambda_{YX}X}{\lambda_{Y\varepsilon_1}} \quad \varepsilon_2 = \frac{Z - \lambda_{ZX}X}{\lambda_{Z\varepsilon_2}}$$

$$\begin{aligned} E(\varepsilon_1\varepsilon_2) &= E \left[\frac{(Y - \lambda_{YX}X)}{\lambda_{Y\varepsilon_1}} \frac{(Z - \lambda_{ZX}X)}{\lambda_{Z\varepsilon_2}} \right] \\ E(\varepsilon_1\varepsilon_2) &= E \left[\frac{(Y - \lambda_{YX}X)}{\lambda_{Y\varepsilon_1}} \frac{(Z - \lambda_{ZX}X)}{\lambda_{Z\varepsilon_2}} \right] \\ E(\varepsilon_1\varepsilon_2) &= \frac{E(YZ) - \lambda_{ZX}E(YX) - \lambda_{YX}E(ZX) + \lambda_{YX}\lambda_{ZX}E(X^2)}{\lambda_{Y\varepsilon_1}\lambda_{Z\varepsilon_2}} \end{aligned}$$

uma vez que, $E(X\varepsilon_1) = E(X\varepsilon_2) = 0$, e

$$E(YX) = \lambda_{YX}E(X^2) + \lambda_{Y\varepsilon_1}E(\varepsilon_1X) \quad (9)$$

$$\rho_{YX} = \lambda_{YX} \quad (10)$$

$$E(ZX) = \lambda_{ZX}E(X^2) + \lambda_{Z\varepsilon_2}E(\varepsilon_2X) \quad (11)$$

$$\rho_{ZX} = \lambda_{ZX} \quad (12)$$

com $1 = \lambda_{YX}^2 + \lambda_{Y\varepsilon_1}^2$ e $1 = \lambda_{ZX}^2 + \lambda_{Z\varepsilon_2}^2$, os parâmetros $\lambda_{Y\varepsilon_1} = \sqrt{1 - \lambda_{YX}^2}$ e $\lambda_{Z\varepsilon_2} = \sqrt{1 - \lambda_{ZX}^2}$ ou $\lambda_{Y\varepsilon_1} = \sqrt{1 - \rho_{YX}^2}$, $\lambda_{Z\varepsilon_2} = \sqrt{1 - \rho_{ZX}^2}$ de acordo com as Equações (10) e (12). Dessa

forma, tem-se:

$$\begin{aligned}
 E(\varepsilon_1 \varepsilon_2) &= \frac{E(YZ) - \lambda_{ZX}E(YX) - \lambda_{YX}E(ZX) + \lambda_{YX}\lambda_{ZX}E(X^2)}{\lambda_{Y\varepsilon_1}\lambda_{Z\varepsilon_2}} \\
 E(\varepsilon_1 \varepsilon_2) &= \frac{\rho_{YZ} - \rho_{ZX}\rho_{YX} - \rho_{YX}\rho_{ZX} + \rho_{YX}\rho_{ZX}}{\sqrt{1 - \rho_{YX}^2}\sqrt{1 - \rho_{ZX}^2}} \\
 E(\varepsilon_1 \varepsilon_2) &= \frac{\rho_{YZ} - \rho_{YX}\rho_{ZX}}{\sqrt{1 - \rho_{YX}^2}\sqrt{1 - \rho_{ZX}^2}} \\
 \rho_{\varepsilon_1 \varepsilon_2} &= \frac{\rho_{YZ} - \rho_{YX}\rho_{ZX}}{\sqrt{1 - \rho_{YX}^2}\sqrt{1 - \rho_{ZX}^2}}
 \end{aligned} \tag{13}$$

finalmente, $\rho_{YZ.X}$ dado na Equação (14), será o mesmo que $\rho_{\varepsilon_1 \varepsilon_2}$ da Equação (13).

$$\rho_{YZ.X} = \frac{\rho_{YZ} - \rho_{YX}\rho_{ZX}}{\sqrt{1 - \rho_{YX}^2}\sqrt{1 - \rho_{ZX}^2}} \tag{14}$$

2.3 Modelos Recursivos

Segundo Duncan (1975), um modelo é dito recursivo se todas ligações causais forem afetadas de uma única forma. Ao considerar quatro variáveis, é possível arranjá-las em 24 tipos de diagramas, no entanto, para uma abordagem inicial, é apresentado o diagrama na Figura 6, com quatro variáveis padronizadas, três erros aleatórios para cada equação estrutural e o sistema de equações dado por:

$$\begin{cases} X_2 = \lambda_{21}X_1 + \lambda_{22}\varepsilon_2 \\ X_3 = \beta_{32}X_2 + \lambda_{31}X_1 + \lambda_{33}\varepsilon_3 \\ X_4 = \beta_{43}X_3 + \beta_{42}X_2 + \lambda_{41}X_1 + \lambda_{44}\varepsilon_4 \end{cases} \tag{15}$$

em que as variáveis exógenas são não correlacionadas com os erros, ou seja, $E(X_1\varepsilon_1) = E(X_2\varepsilon_2) = E(X_3\varepsilon_3) = 0$ e com outras variáveis da equação $E(X_2\varepsilon_3) = E(X_2\varepsilon_4) = E(X_3\varepsilon_4) = 0$.

Figura 6 - Diagrama de Caminho proposto por Duncan (1975)

Para obter as estimativas dos parâmetros, utiliza-se as propriedades de esperança matemática, por conseguinte, a relação causal dada por X_1 em X_2 tem como parâmetro associado λ_{21} , que será obtido da seguinte forma:

$$\begin{aligned} E(X_1 X_2) &= \lambda_{21} E(X_1 X_2) + \lambda_{22} E(X_1 \varepsilon_2) \\ \rho_{12} &= \lambda_{21} \end{aligned}$$

sucessivamente, será identificado todos os parâmetros. Da relação causal indireta, dada de X_1 em X_3 e direta de X_2 em X_3 , com as seguintes expressões:

$$\begin{aligned} E(X_1 X_3) &= \beta_{32} E(X_1 X_2) + \lambda_{31} E(X_1^2) + \lambda_{33} E(X_1 \varepsilon_3) \\ \rho_{13} &= \beta_{32} \rho_{12} + \lambda_{31} \end{aligned} \tag{16}$$

$$\begin{aligned} E(X_2 X_3) &= \beta_{32} E(X_2^2) + \lambda_{31} E(X_1 X_2) + \lambda_{33} E(X_2 \varepsilon_3) \\ \rho_{23} &= \beta_{32} + \lambda_{31} \rho_{12} \end{aligned}$$

resolvendo o sistema $\begin{cases} \rho_{13} = \beta_{32} \rho_{12} + \lambda_{31} \\ \rho_{23} = \beta_{32} + \lambda_{31} \rho_{12} \end{cases}$, tem-se que :

$$\begin{aligned} \rho_{23} &= \beta_{32} + (\rho_{13} - \beta_{32} \rho_{12}) \rho_{12} \\ \rho_{23} &= \beta_{32} + \rho_{13} \rho_{12} - \beta_{32} \rho_{12}^2 \\ \rho_{23} &= \beta_{32} (1 - \rho_{12}^2) + \rho_{13} \rho_{12} \\ \beta_{32} &= \frac{\rho_{23} - \rho_{13} \rho_{12}}{1 - \rho_{12}^2} \end{aligned} \tag{17}$$

ao substituir a Equação (17) na Equação (16), obtém-se:

$$\begin{aligned} \rho_{13} &= \beta_{32} \rho_{12} + \lambda_{31} \\ \rho_{13} &= \left(\frac{\rho_{23} - \rho_{13} \rho_{12}}{1 - \rho_{12}^2} \right) \rho_{12} + \lambda_{31} \\ \rho_{13} &= \left(\frac{\rho_{12} \rho_{23} - \rho_{13} \rho_{12}^2}{1 - \rho_{12}^2} \right) + \lambda_{31} \\ \lambda_{31} &= \rho_{13} - \left(\frac{\rho_{12} \rho_{23} - \rho_{13} \rho_{12}^2}{1 - \rho_{12}^2} \right) \\ \lambda_{31} &= \frac{\rho_{13} - \rho_{12} \rho_{23}}{1 - \rho_{12}^2} \end{aligned}$$

de forma análoga, estima-se os demais parâmetros.

2.4 Abordagem Matricial

A abordagem dos MEE numa forma matricial vem a facilitar a identificação das estruturas do modelo, de modo que é possível organizar seus componentes em vetores e matrizes. Dessa maneira, os parâmetros do modelo são diferenciados quanto ao tipo de variáveis, ou seja, as relações entre variáveis exógenas e endógenas são representadas por parâmetros que formam uma matriz específica que relaciona apenas parâmetros desse tipo de variável, que será denotada pela matriz $\mathbf{\Gamma}$. No entanto, para os parâmetros associados a relação entre variáveis endógenas, atribui-se a matriz $\mathbf{B}$.

Dessa forma, destaca-se o modelo de relações lineares estruturais - LISREL numa abordagem geral, que compreende dois componentes, ou seja, equação de medição e equação estrutural. Para a equação de medição observa-se as equações do tipo:

$$\mathbf{Y} = \mathbf{\Lambda}'_Y \boldsymbol{\eta} + \boldsymbol{\epsilon}_1 \quad (18)$$

$$\mathbf{X} = \mathbf{\Lambda}'_X \boldsymbol{\xi} + \boldsymbol{\epsilon}_2 \quad (19)$$

a equação estrutural, que relaciona os fatores latentes, pode ser dado por:

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \mathbf{\Gamma}\boldsymbol{\xi} + \boldsymbol{\delta} \quad (20)$$

na Equação (18), $\mathbf{\Lambda}_Y$ é a matriz dos parâmetros que representa a influência dos fatores endógenos sobre as variáveis observadas. Na Equação (19), $\mathbf{\Lambda}_X$ representa a matriz dos parâmetros, a influência dos fatores exógenos sobre as variáveis observadas.

Algumas especificações são necessárias para a aplicabilidade do modelo e o desenvolvimento matricial do mesmo. Nesse contexto, têm-se $\mathbf{Y}_{(r \times 1)}$ e $\mathbf{X}_{(s \times 1)}$ como vetores aleatórios das variáveis observadas com indicadores $\boldsymbol{\eta}$ e $\boldsymbol{\xi}$. Os vetores $\boldsymbol{\epsilon}_{1(r \times 1)}$ e $\boldsymbol{\epsilon}_{2(s \times 1)}$ são erros de mensuração.

Na Equação (20), $\boldsymbol{\eta}_{(q_1 \times 1)}$ é um vetor aleatório endógeno da variável latente e $\boldsymbol{\xi}_{(q_2 \times 1)}$ é um vetor aleatório exógeno da variável latente; $\mathbf{B}_{(q_1 \times q_2)}$ e $\mathbf{\Gamma}_{(q_1 \times q_2)}$ são coeficientes de regressão que representam o efeito causal sobre $\boldsymbol{\eta}$ e $\boldsymbol{\xi}$, e $\boldsymbol{\delta}_{(q_1 \times 1)}$ representa o vetor de erros aleatórios da estrutura dos fatores latentes. Assume-se, ainda, que $(\mathbf{I} - \mathbf{B})$ é não singular, $\boldsymbol{\xi}$ é não correlacionado com $\boldsymbol{\delta}$.

Ao isolar a variável latente endógena na Equação (20), a expressão é dada

por:

$$\begin{aligned}
\boldsymbol{\eta} &= \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\delta} \\
\boldsymbol{\eta} - \mathbf{B}\boldsymbol{\eta} &= \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\delta} \\
(\mathbf{I} - \mathbf{B})\boldsymbol{\eta} &= \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\delta} \\
(\mathbf{I} - \mathbf{B})^{-1}(\mathbf{I} - \mathbf{B})\boldsymbol{\eta} &= (\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\delta}) \\
\boldsymbol{\eta} &= (\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\delta})
\end{aligned} \tag{21}$$

por simplificação e considerando a independência entre os erros, (visto anteriormente) as demonstrações das variâncias são simplificadas da seguinte forma:

$$\begin{aligned}
Var(\boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\delta}) &= Var(\boldsymbol{\Gamma}\boldsymbol{\xi}) + Var(\boldsymbol{\delta}) \\
&\boldsymbol{\Gamma}Var(\boldsymbol{\xi})\boldsymbol{\Gamma}' + Var(\boldsymbol{\delta}) \\
&\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}' + \boldsymbol{\Phi}_{\delta}.
\end{aligned} \tag{22}$$

A demonstração de cada elemento da matriz da $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ dada na Equação (27) é vista a seguir:

$$\begin{aligned}
\boldsymbol{\Sigma}_{YY}(\boldsymbol{\theta}) &= Var(\boldsymbol{\Lambda}'_Y\boldsymbol{\eta} + \boldsymbol{\epsilon}_1) \\
&Var(\boldsymbol{\Lambda}'_Y\boldsymbol{\eta}) + Var(\boldsymbol{\epsilon}_1) \\
&\boldsymbol{\Lambda}_Y Var(\boldsymbol{\eta})\boldsymbol{\Lambda}'_Y + \boldsymbol{\Phi}_{\epsilon_1} \\
&\boldsymbol{\Lambda}_Y Var[(\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\delta})]\boldsymbol{\Lambda}'_Y + \boldsymbol{\Phi}_{\epsilon_1} \\
&\boldsymbol{\Lambda}_Y(\mathbf{I} - \mathbf{B})^{-1}Var[\boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\delta}][(\mathbf{I} - \mathbf{B})^{-1}]'\boldsymbol{\Lambda}'_Y + \boldsymbol{\Phi}_{\epsilon_1} \\
&\boldsymbol{\Lambda}_Y(\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}' + \boldsymbol{\Phi}_{\delta})[(\mathbf{I} - \mathbf{B})^{-1}]'\boldsymbol{\Lambda}'_Y + \boldsymbol{\Phi}_{\epsilon_1}
\end{aligned} \tag{23}$$

$$\begin{aligned}
\boldsymbol{\Sigma}_{XX}(\boldsymbol{\theta}) &= Var(\boldsymbol{\Lambda}'_X\boldsymbol{\xi} + \boldsymbol{\epsilon}) \\
&Var(\boldsymbol{\Lambda}'_X\boldsymbol{\eta}) + Var(\boldsymbol{\epsilon}_2) \\
&\boldsymbol{\Lambda}_X Var(\boldsymbol{\eta})\boldsymbol{\Lambda}'_X + \boldsymbol{\Phi}_{\epsilon_2} \\
&\boldsymbol{\Lambda}_X\boldsymbol{\Phi}\boldsymbol{\Lambda}'_X + \boldsymbol{\Phi}_{\epsilon_2}
\end{aligned} \tag{24}$$

$$\begin{aligned}
Cov(\boldsymbol{\eta}, \boldsymbol{\xi}') &= Cov[(\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\delta}), \boldsymbol{\xi}'] \\
&= (\mathbf{I} - \mathbf{B})^{-1}Cov[(\boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\delta}), \boldsymbol{\xi}'] \\
&= (\mathbf{I} - \mathbf{B})^{-1}Cov(\boldsymbol{\Gamma}\boldsymbol{\xi}, \boldsymbol{\xi}') + Cov(\boldsymbol{\delta}, \boldsymbol{\xi}') \\
&= (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma}Cov(\boldsymbol{\xi}, \boldsymbol{\xi}') + Cov(\boldsymbol{\delta}, \boldsymbol{\xi}') \\
&= (\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma}\boldsymbol{\Phi}
\end{aligned} \tag{25}$$

$$\begin{aligned}
\boldsymbol{\Sigma}_{YX}(\boldsymbol{\theta}) &= Cov(\mathbf{Y}, \mathbf{X}') \\
&= Cov(\boldsymbol{\Lambda}'_Y\boldsymbol{\eta} + \boldsymbol{\epsilon}_1, \boldsymbol{\xi}'\boldsymbol{\Lambda}_X + \boldsymbol{\epsilon}'_2) \\
&= Cov(\boldsymbol{\Lambda}'_Y\boldsymbol{\eta}, \boldsymbol{\xi}'\boldsymbol{\Lambda}_X) + Cov(\boldsymbol{\Lambda}'_Y\boldsymbol{\eta}, \boldsymbol{\epsilon}'_2) + Cov(\boldsymbol{\epsilon}_1, \boldsymbol{\xi}'\boldsymbol{\Lambda}_X) + Cov(\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}'_2) \\
&= \boldsymbol{\Lambda}'_YCov(\boldsymbol{\eta}, \boldsymbol{\xi}')\boldsymbol{\Lambda}'_X + \boldsymbol{\Lambda}'_YCov(\boldsymbol{\eta}, \boldsymbol{\epsilon}'_2) + Cov(\boldsymbol{\epsilon}_1, \boldsymbol{\xi}')\boldsymbol{\Lambda}_X + Cov(\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}'_2) \\
&= \boldsymbol{\Lambda}'_YCov(\boldsymbol{\eta}, \boldsymbol{\xi}')\boldsymbol{\Lambda}'_X + \boldsymbol{\Lambda}'_YCov(\boldsymbol{\eta}, \boldsymbol{\epsilon}'_2) + Cov(\boldsymbol{\epsilon}_1, \boldsymbol{\xi}')\boldsymbol{\Lambda}_X + Cov(\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}'_2) \\
&= \boldsymbol{\Lambda}'_Y(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Lambda}'_X
\end{aligned} \tag{26}$$

Logo, a matriz de covariância estimada do modelo será denotada por meio da seguinte estrutura matricial:

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\Sigma}_{YY}(\boldsymbol{\theta}) & \boldsymbol{\Sigma}_{YX}(\boldsymbol{\theta}) \\ \boldsymbol{\Sigma}_{XY}(\boldsymbol{\theta}) & \boldsymbol{\Sigma}_{XX}(\boldsymbol{\theta}) \end{bmatrix} \tag{27}$$

A partir da Equação (23), (24) e (26), a matriz dos estimadores será composta da expressão vista na Equação (28), de estrutura complexa, porém que envolve tanto variáveis observadas como variáveis latentes, ou seja, um modelo mais geral.

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{bmatrix} \boldsymbol{\Lambda}'_Y(\mathbf{I} - \mathbf{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}' + \boldsymbol{\Phi}_\delta)[(\mathbf{I} - \mathbf{B})^{-1}]'\boldsymbol{\Lambda}'_Y + \boldsymbol{\Phi}_{\epsilon_1} & \boldsymbol{\Lambda}'_Y(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Lambda}'_X \\ \boldsymbol{\Lambda}_X\boldsymbol{\Phi}\boldsymbol{\Gamma}'[(\mathbf{I} - \mathbf{B})^{-1}]\boldsymbol{\Lambda}'_Y & \boldsymbol{\Lambda}_X\boldsymbol{\Phi}\boldsymbol{\Lambda}'_X + \boldsymbol{\Phi}_{\epsilon_2} \end{bmatrix} \tag{28}$$

De forma genérica, não se tem a matriz de covariâncias populacional, porém, utiliza-se a matriz de covariância amostral $\mathbf{S}$ como base para encontrar as estimativas dos parâmetros. Segundo Bollen (1989) e Schumacker e Lomax (2010), é preciso resolver o sistema $\mathbf{S} = \boldsymbol{\Sigma}(\boldsymbol{\theta})$ e verificar alguns critérios de avaliação e ajustes do modelo.

A abordagem matricial, envolvendo variáveis latentes e observadas, representa uma forma geral para os MEE (BOLLEN, 1989). No entanto, o foco deste trabalho está em estudar os MEE considerando-se apenas variáveis observadas. Assim, ao considerar

MEE com variáveis observadas, Bollen (1989) expõe de forma clara a construção para esses modelos. Uma vez que não há presença das variáveis latentes, o modelo denotado na Equação (20) será reescrito considerando $\mathbf{Y} = \boldsymbol{\eta}$, $\mathbf{X} = \boldsymbol{\xi}$ e $\boldsymbol{\delta}$ como vetor de erros para relação $\mathbf{X}$ e $\mathbf{Y}$.

$$\mathbf{Y} = \mathbf{B}\mathbf{Y} + \boldsymbol{\Gamma}\mathbf{X} + \boldsymbol{\delta} \quad (29)$$

Assim, $\mathbf{Y}_{p \times 1}$ e $\mathbf{X}_{q \times 1}$ são variáveis observadas, $\mathbf{B}_{p \times p}$ e $\boldsymbol{\Gamma}_{p \times q}$ são parâmetros do modelo em que a matriz de covariância é dada pela Equação (30), contendo a mesma informação vista na Equação (27). Os termos p e q representam respectivamente o número de equações no sistema ou número de variáveis endógenas e o número de variáveis exógenas.

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \begin{bmatrix} (\mathbf{I} - \mathbf{B})^{-1} (\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}' + \boldsymbol{\Phi}_{\delta}) [(\mathbf{I} - \mathbf{B})^{-1}]' & '(\mathbf{I} - \mathbf{B})^{-1} \boldsymbol{\Gamma}\boldsymbol{\Phi} \\ \boldsymbol{\Phi}\boldsymbol{\Gamma}' [(\mathbf{I} - \mathbf{B})^{-1}]' & \boldsymbol{\Phi} \end{bmatrix} \quad (30)$$

Por conseguinte, considerando-se o diagrama de caminhos apresentado na Figura 7, é possível deduzir um sistema de equações, Equação (31), e por meio deste escrever a notação matricial para que seja possível utilizar outras técnicas de estimação e validação do modelo. Nesse contexto, Bollen (1989) e Schumacker e Lomax (2010) expõem as matrizes do modelo e algumas características sobre as mesmas.

Figura 7 - Modelo estrutural simulado

$$\begin{cases} Y_1 = \lambda_{11}X_1 + \lambda_{12}X_2 + \varepsilon_1 \\ Y_2 = \beta_{21}Y_1 + \lambda_{23}X_3 + \lambda_{24}X_4 + \varepsilon_2 \end{cases} \quad (31)$$

Na Figura 7, há duas variáveis endógenas, em que Y_1 é vista como causa e Y_2 como efeito e matricialmente associado ao sistema de MEE na Equação (31), a matriz $\mathbf{B}$ com dimensão 2×2 contém os parâmetros que relacionam as variáveis endógenas no

modelo.

$$\mathbf{B} = \begin{bmatrix} 0 & 0 \\ \beta_{21} & 0 \end{bmatrix}$$

Dessa maneira, a especificação de $\mathbf{\Gamma}$ de dimensão 2×4 relaciona os efeitos entre as variáveis exógenas sobre as endógenas do modelo.

$$\mathbf{\Gamma} = \begin{bmatrix} \lambda_{11} & \lambda_{12} & 0 & 0 \\ 0 & 0 & \lambda_{23} & \lambda_{24} \end{bmatrix}$$

Pela definição, até então, é possível exemplificar o modelo descrito na Equação (29) e posteriormente obter conclusões sobre sua estrutura e estimativas, porém, na prática, utiliza-se a estrutura apresentada pelo pesquisador e por meio da análise dos MEE, verifica-se se a mesma é satisfatória ou não.

Assim, faz-se necessário identificar outros componentes matriciais que são importantes para a estimação dos parâmetros. Nessa estrutura, a matriz de parâmetros $\mathbf{\Phi}_\delta$ representa a matriz de covariância dos erros associados ao modelo:

$$\mathbf{\Phi}_\delta = \begin{bmatrix} \delta_{11} & 0 \\ 0 & \delta_{22} \end{bmatrix}$$

e $\mathbf{\Phi}$, representa a matriz de covariância entre as variáveis exógenas.

$$\mathbf{\Phi} = \begin{bmatrix} \phi_{11} & 0 & 0 & 0 \\ 0 & \phi_{22} & 0 & 0 \\ 0 & 0 & \phi_{33} & 0 \\ 0 & 0 & 0 & \phi_{44} \end{bmatrix}$$

É possível identificar, por meio de um diagrama de caminhos, o sistema de equações, como também, associar suas estruturas matriciais. Para o modelo em questão, atribui-se a seguinte expressão matricial de acordo com a Equação (29).

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \beta_{21} & 0 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} + \begin{bmatrix} \lambda_{11} & \lambda_{12} & 0 & 0 \\ 0 & 0 & \lambda_{23} & \lambda_{24} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix} \quad (32)$$

Segundo Schumacker e Lomax (2010), alguns parâmetros podem ser fixo ou livres, a distinção entre eles dá-se como: os parâmetros livres são aqueles que não são conhecidos, que precisam ser estimados por algum método; quanto aos fixos, são parâmetros não livres, especificados por um determinado valor como 0 ou 1.

A análise dos MEE pode ser dada com base em diversos métodos e formas de verificação do melhor ajuste do modelo, um dos métodos bastante utilizado nesse tipo de metodologia é o método de mínimos quadrados a dois passos.

2.4.1 Mínimos quadrados a dois passos

Uma maneira de compreender o método de mínimos quadrados a dois passos dá-se por meio da reorganização das matrizes do modelo na Equação (32), na seguinte forma:

$$\mathbf{y}_j = \mathbf{Y}_j \boldsymbol{\beta}_j + X_j \boldsymbol{\lambda} + \boldsymbol{\varepsilon}_j \quad (33)$$

em que j representa a j -ésima equação estrutural, $\mathbf{y}_j$ de dimensão $(n \times 1)$ é o vetor da variável resposta da j -ésima equação, $\mathbf{Y}_j$ $(n \times q_j)$ é a matriz de q_j variáveis endógenas explanatórias na equação j , $\boldsymbol{\beta}_j$ de dimensão $(q_j \times 1)$ representa o vetor de parâmetros estruturais das variáveis endógenas explanatórias, X_j de dimensão $(n \times m_j)$ é a matriz de m_j variáveis explanatórias na equação j , $\boldsymbol{\lambda}$ de dimensão $(m_j \times 1)$ é o vetor de parâmetros associados as variáveis exógenas explanatórias e $\boldsymbol{\varepsilon}_j$ $(n \times 1)$ são erros associados a j -ésima equação.

Segundo Fox (2008), o método de mínimos quadrados a dois passos consiste em realizar uma regressão das variáveis endógenas explanatórias (variáveis que mesmo sendo endógenas aparecem em outro momento da equação como explanatória de outra endógena) com todas as variáveis exógenas no primeiro passo, obtendo assim, os coeficientes de regressão e a predição da variável endógena explanatória.

A notação matricial especificada na Equação (33), pode ser dada como:

$$\mathbf{y}_j = \begin{bmatrix} \mathbf{Y}_j & X_j \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_j \\ \boldsymbol{\lambda}_j \end{bmatrix} + \boldsymbol{\varepsilon}_j$$

de uma forma mais reduzida, como:

$$\mathbf{y}_j = \mathbf{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}_j.$$

Nesse primeiro passo de estimação, ajusta-se regressões que envolvam apenas variáveis endógenas explanatórias com as demais variáveis explanatórias da equação, logo, é possível obter o estimador de mínimos quadrados ordinários:

$$\mathbf{P}_j = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_j$$

e os valores ajustados por:

$$\widehat{\mathbf{Y}}_j = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_j. \quad (34)$$

No segundo passo do método de mínimos quadrados ordinários, aplicam-se X_j e $\widehat{Y}_j$ como instrumentos da equação estrutural, sendo possível estimar os parâmetros associados entre variáveis endógenas ($\mathbf{y}_j$) e variáveis endógenas explicativas ($\mathbf{Y}_j$), como também, parâmetros da relação variáveis endógenas ($\mathbf{y}_j$) e exógenas ($\mathbf{X}_j$).

Ao aplicar mínimos quadrados novamente,

$$\mathbf{y}_j = \begin{bmatrix} \widehat{\mathbf{Y}}_j & X_j \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_j \\ \boldsymbol{\lambda}_j \end{bmatrix} + \boldsymbol{\varepsilon}_j$$

por substituição da Equação (34), obtém-se:

$$\mathbf{y}_j = \begin{bmatrix} \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_j & X_j \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_j \\ \boldsymbol{\lambda}_j \end{bmatrix} + \boldsymbol{\varepsilon}_j \quad (35)$$

desenvolvendo o método de mínimos quadrados, tem-se:

$$\begin{bmatrix} \widehat{\boldsymbol{\beta}}_j \\ \widehat{\boldsymbol{\lambda}}_j \end{bmatrix} = \begin{bmatrix} \mathbf{Y}_j'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_j & \mathbf{Y}_j'\mathbf{X}_j \\ \mathbf{X}_j'\mathbf{Y}_j & \mathbf{X}_j'\mathbf{X}_j \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{Y}_j'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_j \\ \mathbf{X}_j'\mathbf{y}_j \end{bmatrix}$$

A estimativa da variância é dada por:

$$\hat{V} \begin{bmatrix} \hat{\beta}_j \\ \hat{\lambda}_j \end{bmatrix} = s_{e_j}^2 \begin{bmatrix} \mathbf{Y}_j' \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Y}_j & \mathbf{Y}_j' \mathbf{X}_j \\ \mathbf{X}_j' \mathbf{Y}_j & \mathbf{X}_j' \mathbf{X}_j \end{bmatrix}^{-1}$$

em que,

$$s_{e_j}^2 = \frac{\mathbf{e}' \mathbf{e}}{n - q_j - m_j}$$

e

$$\mathbf{e}_j = \mathbf{y}_j - \mathbf{Y}_j \hat{\beta} - \mathbf{X}_j \hat{\gamma}_j$$

A verificação do ajuste e confirmação da hipótese inicial pode ser considerada quando se utiliza alguns critérios para verificar o quão bom o modelo está. Logo, segundo Bollen (1989) são diversos os critérios e índices de verificação.

2.4.2 Índices de ajuste do modelo

O teste χ^2 é uma estatística de ajuste do modelo que verifica se a hipótese $\Sigma = \Sigma(\theta)$ é aceitável ou não:

$$(n - 1)F \sim \chi^2 \left(\frac{1}{p} (p + q) (p + q + 1) - t \right)$$

em que F é a função de máxima verossimilhança, t é o número de parâmetros livres do modelo, p e q são as quantidades de variáveis endógenas e exógenas respectivamente. Espera-se que os níveis descritivos para esse teste sejam superiores a 0,05, ou seja, deve-se aceitar a hipótese de nulidade, e assim, aceitar a estrutura proposta. No entanto, esse teste é sensível à presença de assimetria no conjunto de dados.

A raiz do quadrado médio residual aproximado (RMSEA) é um índice que quantifica como o modelo está mal ajustado, dessa forma, um valor próximo a zero indica bom ajuste. Kline (2011) descreve RMSEA como:

$$RMSEA = \sqrt{\frac{\chi^2 - df}{df(N - 1)}}$$

em que, df é o grau de liberdade do modelo e N é número de observações na amostra. Para um bom ajuste, o valor do RMSEA deve ser inferior a 0,05.

Outro índice utilizado nessa metodologia é o de bom ajuste do modelo (GFI).

Considera-se que valores próximos a 1 indicam bom ajuste do modelo (KLINE, 1998).

$$GFI = 1 - \frac{C_{res}}{C_{tot}} \quad (36)$$

Na Equação (36), C_{res} e C_{tot} são, respectivamente, a soma de quadrado da matriz de variâncias e covariâncias residuais e matriz residual; e a soma de quadrado da matriz dos dados observados.

3 MÉTODOS

3.1 Identificação do modelo

Segundo Pugesek e Tomer (2003), é importante notar como os parâmetros são especificados no modelo, uma vez que os mesmos são combinados para formar uma matriz de variâncias e covariâncias implícitas no modelo a ser estimada, para então ser testada com a matriz de variâncias e covariâncias das variáveis observadas.

A partir da matriz de variâncias e covariâncias observada, determina-se como o modelo será identificado ou não. Este procedimento deve ser efetuado antes que algum modelo possa ser testado (PUGESEK; TOMER, 2003). Existem três níveis de identificação para MEE: (i) *under-identified* quando um ou mais parâmetros não podem ser estimados a partir da matriz de variância-covariância observada; (ii) *just-identified* quando todos os parâmetros são estimáveis resultando em ajuste perfeito (PUGESEK; TOMER, 2003); (iii) *over-identified*, modelo este, que possui o número de informações sobre a matriz de variância-covariâncias superior ao número de parâmetros a ser estimado. Considera-se, ainda, que a diferença entre o número de elementos conhecidos da matriz de variância-covariância e o número de parâmetros do modelo a ser estimado é conhecido, como os graus de liberdade df do modelo.

3.2 Método de Máxima Verossimilhança

A utilização desse método tem por suposição que $\mathbf{X}$ e $\mathbf{Y}$ seguem uma distribuição normal multivariada e $\mathbf{S}$ é uma matriz de covariância amostral, p e q representam, respectivamente o número de variáveis endógenas e exógenas do modelo.

$$FML = \log |\Sigma(\boldsymbol{\theta})| + tr(\mathbf{S}\Sigma^{-1}(\boldsymbol{\theta})) - \log |\mathbf{S}| - (p + q) \quad (37)$$

Em condições perfeitas, quando $\Sigma(\boldsymbol{\theta}) = \mathbf{S}$ pode-se, então, substituir $\mathbf{S}$ por $\hat{\Sigma}$ na Equação (38), resultando em $tr(\mathbf{I})$ sendo o traço da matriz identidade com o mesmo valor de $(p + q)$ (BOLLEN, 1989). Logo, a função de máxima verossimilhança será nula, ocasionando um ajuste perfeito com $FML = 0$.

$$FML = \log |\Sigma(\boldsymbol{\theta})| + tr(I) - \log |\mathbf{S}| - (p + q) \quad (38)$$

Segundo Grace (2006), algumas propriedades fundamentais dos MEE são derivadas do processo de estimação simultâneo associado ao método da máxima verossimilhança, que busca estimativas para os coeficientes estruturais, tal que, a diferença entre as matrizes $\Sigma(\theta)$ e S sejam mínimas.

Em síntese, o método da máxima verossimilhança resulta em informações completas sobre os estimadores, uma vez que todos os requisitos estatísticos sejam cumpridos e o modelo corretamente especificado. Os estimadores são assintoticamente imparciais, eficientes e consistentes (KLINE, 2011).

3.3 Ajuste do modelo

Em geral, na verificação do ajuste, existem diversas formas ou índices de comparação para aceitação do modelo proposto. Alguns dos índices mais utilizados são descritos na Seção 2.4.2, embora existam algumas modificações desses e outros índices, que podem ser vistos em Kline (1998) e Bollen (1989).

Neste trabalho, ao considerar o processo iterativo, será utilizado o critério χ^2 relativo, dado por χ^2/gl em substituição ao critério χ^2 . Assim, a cada iteração, será verificado se o valor do χ^2 relativo é inferior a três, sendo este o critério de seleção para aceitação do modelo.

4 RESULTADOS E DISCUSSÃO

Para verificar a eficiência dos modelos de equações estruturais utilizando variáveis observadas, foi realizado um estudo de simulação de acordo com a metodologia descrita na Seção 2.1. O processo de simulação e estimação dos parâmetros sucedeu-se por meio de programas implementados e utilização do pacote **sem** (FOX; NIE; BYRNES, 2013) no software (R Core Team, 2013). Essa função tem como objetivo executar as análises estatísticas envolvendo modelos de equações estruturais, sendo imprescindíveis algumas informações especificadas e necessárias para aplicação dessa função.

Argumentos para utilização da função **sem** do pacote **sem** (FOX; NIE; BYRNES, 2013) para o software (R Core Team, 2013):

- **ram**: descrição do modelo em códigos que representem a relação de causa e efeito entre as variáveis;
 - os coeficientes estruturais são representados por uma seta direcionada, $->$;
 - as variâncias e covariâncias da estrutura do modelo são representadas por setas bidirecionadas, $<->$.
- **S**: matriz de variâncias e covariâncias das variáveis observadas no modelo, descrita na forma de matriz triangular inferior ou matriz simétrica.
- **N**: o número de observações em que a matriz de variâncias e covariâncias foi gerada.

Para uma simulação inicial, considerou-se o **Algoritmo I**, descrito abaixo, como também o diagrama de caminhos da Figura 9 e Equação (39), em que a covariância entre X_1 e X_2 , e X_3 e X_4 será definida como nula.

$$\begin{cases} Y_1 = \lambda_{11}X_1 + \lambda_{12}X_2 + \epsilon_1 \\ Y_2 = \beta_{21}Y_1 + \lambda_{23}X_3 + \lambda_{24}X_4 + \epsilon_2. \end{cases} \quad (39)$$

Figura 8 - Modelo estrutural simulado (estrutura geral)

1 Algoritmo I.

- 1.1 Gerou-se uma amostra de tamanho n da variável X_1 , com $X_1 \sim N(0, \sigma_{X_1X_1})$ representando uma variável centrada em torno da média.
- 1.2 Para inserir a covariância entre as variáveis X_1 e X_2 , considera-se X_2 como combinação linear de X_1 mais um erro aleatório $\varepsilon_2 \sim N(0, \sigma_{\varepsilon_2\varepsilon_2})$, conforme metodologia abordada na Seção 2.1.
- 1.3 A variável endógena Y_1 é descrita como combinação linear das variáveis exógenas (geradas nos itens anteriores) pré-multiplicadas pelos respectivos parâmetros associados ao diagrama da Figura 8 mais um erro $\varepsilon_Y \sim N(0, \sigma_{YY})$.
- 1.4 O processo de simulação para Y_2 dá-se de forma similar a Y_1 , porém, com a inclusão da variável endógena Y_1 pré-multiplicada por β_{21} .

Como descrito no **Algoritmo I**, obtêm-se a variável exógena X_2 por meio da expressão:

$$X_2 = \frac{\sigma_{X_1X_2}}{\sigma_{X_1X_1}} X_1 + \varepsilon_2$$

em que $E(X_2) = 0$ e a variância dada por:

$$E(X_2^2) = \frac{(\sigma_{X_1X_2})^2}{\sigma_{X_1X_1}} + \sigma_{\varepsilon_2\varepsilon_2}.$$

A variável X_4 , por sua vez, é gerada de forma análoga a X_2 , ou seja:

$$X_4 = \frac{\sigma_{X_3X_4}}{\sigma_{X_3X_3}} X_3 + \varepsilon_4$$

em que ε_4 provém de uma distribuição normal com $E(\varepsilon_4) = 0$ e $E(\varepsilon_4^2) = \sigma_{\varepsilon_4\varepsilon_4}$, logo, $X_4 \sim N(0, \sigma_{X_4X_4})$, e a variância denotada por:

$$E(X_4^2) = \frac{(\sigma_{X_3X_4})^2}{\sigma_{X_3X_3}} + \sigma_{\varepsilon_4\varepsilon_4}.$$

Para definir valores dos parâmetros estruturais para simulação e revendo a estrutura do modelo descrito anteriormente, usa-se:

```
library(sem)
lambda11=1; lambda12=1.5; lambda23=2.5; lambda24=3; beta21=3
sigma.x1x1=1
sigma.x1x2=0; sigma.x3x3=3
sigma.x3x4=0
sig.epsx2=4; sig.epsx4=5
sigma.x2x2=((sigma.x1x2^2)/sigma.x1x1) + sig.epsx2
sigma.x4x4=((sigma.x3x4^2)/sigma.x3x3) + sig.epsx4
sigma.eps1eps1=1; sigma.eps2eps2=2; sigma.eps1eps2=0

modelo=specifyModel()
x1 -> y1, lambda11, NA
x2 -> y1, lambda12, NA
x3 -> y2, lambda23, NA
x4 -> y2, lambda24, NA
y1 -> y2, beta21, NA
x1 <-> x1, sigma.x1x1, NA
x2 <-> x2, sigma.x2x2, NA
x3 <-> x3, sigma.x3x3, NA
x4 <-> x4, sigma.x4x4, NA
y1 <-> y1, sigma.eps1eps1, NA
y2 <-> y2, sigma.eps2eps2, NA
```

Após definir o modelo e os parâmetros da simulação, gera-se as variáveis e os erros aleatórios como descrito nos passos anteriores e no **Algoritmo I**. No entanto, a função **sem** utiliza a matriz de variâncias e covariâncias para adotar o método da máxima verossimilhança e estimar os parâmetros e outras estatísticas do modelo. Logo, é fundamental criar matrizes com as características necessárias para implementação dessa função.

Fixando a uma semente qualquer (set.seed) e definindo o tamanho para cada amostra (n), têm-se as variáveis por simulação.

```

set.seed(3) # fixar semente
n=100      # tamanho das amostras

x1 = rnorm(n,sd=sqrt(sigma.x1x1))
x2 = (sigma.x1x2/sigma.x1x1)*x1+rnorm(n,sd=sqrt(sig.epsx2))
x3 = rnorm(n,sd=sqrt(sigma.x3x3))
x4 = (sigma.x3x4/sigma.x3x3)*x3+rnorm(n,sd=sqrt(sig.epsx4))
e1 = rnorm(n,sd=sqrt(sigma.eps1eps1))
e2 = (sigma.eps1eps2/sigma.eps1eps1)*e1+
     rnorm(n,sd=sqrt(sigma.eps2eps2-((sigma.eps1eps2^2)/sigma.eps1eps1)))
y1 = lambda11*x1+lambda12*x2+e1
y2 = beta21*y1+lambda23*x3+lambda24*x4+e2

```

Através da função `data.frame` é possível organizar os vetores de amostras para obtenção da matriz de variâncias e covariâncias.

```

X = data.frame(x1,x2,x3,x4) # matriz das variáveis exógenas
Y = data.frame(y1,y2)      # matriz das variáveis endógenas
p = dim(Y)[[2]] # número de variáveis endógenas
q = dim(X)[[2]] # número de variáveis exógenas
var_obs = data.frame(X,Y) # matriz com variáveis endógenas e exógenas

S = cov(var_obs) # matriz de variâncias e covariâncias observada

```

Ao definir a matriz `S` como matriz de variâncias e covariância; o tamanho amostral `n`; e o número de iterações, é possível utilizar a função `sem` para verificar a hipótese $S = \Sigma(\theta)$ e alguns critérios de avaliação. No quadro abaixo, pode-se observar que há indícios de que o modelo está bem ajustado, considerando-se o nível descritivo alto, ou seja, a hipótese de igualdade entre as matrizes é satisfeita. Além disso, todos os parâmetros são significativos e próximos ao valor verdadeiro.

```

mod =sem(modelo1,S,n,maxiter=10000)
>summary(mod)

Model Chisquare = 8.059973   Df = 10 Pr(>Chisq) = 0.6229788
AIC = 30.05997
BIC = -37.99173

Normalized Residuals

```

```

      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.8586 -0.6458 -0.3156 -0.1769  0.0000  2.3980

R-square for Endogenous Variables
      y1      y2
0.9038 0.9924

Parameter Estimates
      Estimate Std Error  z value  Pr(>|z|)
lambda11      1.1961646 0.13063262  9.156706 5.350543e-20 y1 <--- x1
lambda12      1.4810304 0.05089788 29.098077 3.796735e-186 y1 <--- x2
lambda23      2.6540013 0.06723640 39.472687 0.000000e+00 y2 <--- x3
lambda24      2.9828700 0.05053124 59.030213 0.000000e+00 y2 <--- x4
beta21         2.9436281 0.03300300 89.192741 0.000000e+00 y2 <--- y1
sigma.x1x1      0.7328675 0.10416525  7.035624 1.983717e-12 x1 <--> x1
sigma.x2x2      4.8275793 0.68616224  7.035624 1.983717e-12 x2 <--> x2
sigma.x3x3      3.1022115 0.44092914  7.035624 1.983717e-12 x3 <--> x3
sigma.x4x4      5.4923767 0.78065244  7.035624 1.983717e-12 x4 <--> x4
sigma.eps1eps1  1.2381235 0.17597921  7.035624 1.983717e-12 y1 <--> y1
sigma.eps2eps2  1.3884028 0.19733898  7.035624 1.983717e-12 y2 <--> y2

Iterations = 0

```

Valores como Estatística χ^2 e nível descritivo podem ser calculados de forma matricial, por meio das estimativas encontradas anteriormente.

```

# Matriz lambda
Lambda = rbind(
  c(lambda11,lambda12,0,0),
  c(0,0,lambda23,lambda24))
# dim(n.endog,n.exóg)

# Matriz. betas
Beta=rbind(
  c(0,0),
  c(beta21,0))
# dim(n.endog,n.endog)

#Matriz de Cov xx
Sigma.xx=rbind(
  c(sigmax1x1,0,0,0),
  c(0,sigmax2x2,0,0),
  c(0,0,sigmax3x3, 0),
  c(0,0,0,sigmax4x4))

# Matriz de Cov. epsilon
Sigma.eps=rbind(
  c(sigma.e1e1,0),

```

```

c(0,sigma.e2e2))

Sighat=matrix(NA,nrow=q+p,ncol=q+p)
dimnames(Sighat)=dimnames(S)
Sighat[1:q,1:q]=Sigma.xx
Sighat[c((q+1):(q+p)),c((q+1):(q+p))]=Sighat.yy
Sighat[c((q+1):(q+p)),1:q]=Sighat.yx
Sighat[1:q,c((q+1):(q+p))]=t(Sighat.yx)

```

Com as matrizes estimadas, pode-se calcular os demais critérios, dessa forma, tem-se:

$$\widehat{\Phi} = \begin{bmatrix} 0,733 & 0,000 & 0,000 & 0,000 \\ 0,000 & 4,828 & 0,000 & 0,000 \\ 0,000 & 0,000 & 3,102 & 0,000 \\ 0,000 & 0,000 & 0,000 & 5,492 \end{bmatrix}$$

$$\widehat{\Phi}_{\delta} = \begin{bmatrix} 1,238 & 0,000 \\ 0,000 & 1,388 \end{bmatrix}$$

$$\widehat{\mathbf{B}} = \begin{bmatrix} 0,000 & 0,000 \\ 2,944 & 0,000 \end{bmatrix}$$

$$\widehat{\Gamma} = \begin{bmatrix} 1,196 & 1,481 & 0,000 & 0,000 \\ 0,000 & 0,000 & 2,654 & 2,983 \end{bmatrix}$$

Das matrizes acima, é possível calcular os valores dos elementos da matriz descrita na Equação (30), ou seja:

$$\Sigma(\theta)_{YY} = (\mathbf{I} - \mathbf{B})^{-1} (\Gamma \Phi \Gamma' + \Phi_{\delta}) [(\mathbf{I} - \mathbf{B})^{-1}]'$$

$$\Sigma(\theta)_{YY} = \begin{bmatrix} 1,000 & 0,000 \\ 2,944 & 1,000 \end{bmatrix} \begin{bmatrix} 11,638 & 0,000 \\ 0,000 & 70,720 \end{bmatrix} \begin{bmatrix} 1,000 & 0,000 \\ 2,944 & 1,000 \end{bmatrix}'$$

$$\Sigma(\theta)_{YY} = \begin{bmatrix} 12,876 & 37,901 \\ 37,901 & 183,676 \end{bmatrix}$$

$$\Sigma(\theta)_{YX} = (\mathbf{I} - \mathbf{B})^{-1} \Gamma \Phi$$

$$\Sigma(\theta)_{YX} = \begin{bmatrix} 1,196 & 1,481 & 0,000 & 0,000 \\ 3,521 & 4,360 & 2,654 & 2,983 \end{bmatrix} \begin{bmatrix} 0,733 & 0,000 & 0,000 & 0,000 \\ 0,000 & 4,828 & 0,000 & 0,000 \\ 0,000 & 0,000 & 3,102 & 0,000 \\ 0,000 & 0,000 & 0,000 & 5,492 \end{bmatrix}$$

$$\Sigma(\theta)_{YX} = \begin{bmatrix} 0,877 & 7,150 & 0,000 & 0,000 \\ 2,580 & 21,046 & 8,233 & 16,383 \end{bmatrix}$$

A matriz de variâncias e covariâncias estimada pelo modelo pode ser dada por:

$$\Sigma(\theta) = \begin{bmatrix} 0,733 & 0,000 & 0,000 & 0,000 & 0,877 & 2,580 \\ 0,000 & 4,828 & 0,000 & 0,000 & 7,150 & 21,046 \\ 0,000 & 0,000 & 3,102 & 0,000 & 0,000 & 8,233 \\ 0,000 & 0,000 & 0,000 & 5,492 & 0,000 & 16,383 \\ 0,877 & 7,150 & 0,000 & 0,000 & 12,876 & 37,901 \\ 2,580 & 21,046 & 8,233 & 16,383 & 37,901 & 183,676 \end{bmatrix}$$

e a matriz de variância e covariância observada

$$S(\theta) = \begin{bmatrix} 0,733 & -0,132 & -0,043 & -0,102 & 0,681 & 1,578 \\ -0,132 & 4,828 & -0,077 & -0,444 & 6,992 & 19,102 \\ -0,043 & -0,077 & 3,102 & 0,995 & -0,222 & 10,547 \\ -0,102 & -0,444 & 0,995 & 5,492 & -0,682 & 17,016 \\ 0,681 & 6,992 & -0,222 & -0,682 & 12,408 & 33,899 \\ 1,578 & 19,102 & 10,547 & 17,016 & 33,899 & 179,923 \end{bmatrix}$$

```
> # FML
> Sighat_inv=solve(Sighat)
> no.rows.S=dim(S)[[1]]
> (fit.crit = log(det(Sighat)/det(S))+sum(S*t(Sighat_inv))-(p+q))
[1] 0.08141387

> # Ajuste qui-quadrado
> (crit_ajust = mod$criterion)
[1] 0.08141387

> (model.chisq = (n-1)*crit_ajust)
[1] 8.059973
```

```

> t=11    # número de par. desconhecidos
> # gl
> (df=((q+p)*(q+p+1)/2)-t)
[1] 10

> # p-valor do modelo ajustado
> (pvalue=1-pchisq(model.chisq,df))
[1] 0.6229788

```

A simulação descrita acima foi realizada em apenas uma iteração, ou seja, foram geradas as amostras, porém, faz-se necessário, também analisar um estudo completo para obter mais informações do modelo, além dos critérios usuais. Assim, a próxima etapa deste trabalho dá-se com mil processos de iterações, ou seja, serão mil iterações, considerando-se, para cada, uma simulação similar ao que foi visto anteriormente em seis cenários, com amostras de tamanho: $n = 50$, $n = 250$ e $n = 500$.

Devido à grande quantidade de critérios para o ajuste do modelo, optou-se por selecionar apenas os parâmetros cujo modelo apresentasse a razão da estatística qui-quadrado (χ^2) e graus de liberdade inferior a 3 (KLINE, 1998). Deste modo, todo o processo de simulação será representado com base no sistema de equações dado na Equação (39) e diagrama de caminhos na Figura 8, dos quais são geradas amostras e estimados os parâmetros de acordo com as especificações utilizadas em cada cenário.

Para o primeiro cenário, considerou-se a ausência da covariância entre X_1 e X_2 , e X_3 e X_4 , como ilustrado na Figura 9. As estimativas para diferentes amostras são observadas na Tabela 3.

Figura 9 - Representação gráfica do Cenário I

Tabela 3 - Estimativa dos parâmetros do modelo para os diferentes tamanhos amostrais (Cenário I)

Parâmetros	$n = 50$		$n = 250$		$n = 500$		Valor verdadeiro
	Média	Mediana	Média	Mediana	Média	Mediana	
λ_{11}	0,9967	0,9980	0,997	0,997	1,0003	1,0000	1,0000
λ_{12}	1,5012	1,5100	1,5002	1,5000	1,4996	1,5000	1,5000
λ_{23}	2,4978	2,5000	2,4970	2,5000	2,4992	2,5000	2,5000
λ_{24}	3,0051	3,0100	3,0006	3,0000	2,9991	3,0000	3,0000
β_{21}	3,0012	3,0000	2,9998	3,0000	3,0014	3,0000	3,0000
$\sigma_{X_1X_1}$	0,9983	0,9880	0,9967	0,9920	1,0035	1,0040	1,0000
$\sigma_{X_2X_2}$	3,9889	3,9500	3,9859	3,9600	3,9919	3,9900	4,0000
$\sigma_{X_3X_3}$	2,9895	2,9400	3,0081	2,9900	2,9937	2,9900	3,0000
$\sigma_{X_4X_4}$	4,9668	4,9400	4,9849	4,9700	5,0043	5,0000	5,0000
$\sigma_{\epsilon_1\epsilon_1}$	0,9674	0,9530	0,9905	0,9880	0,9962	0,9950	1,0000
$\sigma_{\epsilon_2\epsilon_2}$	1,8861	1,8580	1,9770	1,9700	1,9928	1,9900	2,0000

Para o segundo cenário, considerou-se a presença de covariância positiva entre X_1 e X_2 , logo, como no diagrama apresentado na Figura 10, foram obtidas as estimativas dos parâmetros apresentadas na Tabela 4.

Figura 10 - Representação gráfica do Cenário II

Tabela 4 - Estimativa dos parâmetros do modelo para os diferentes tamanhos amostrais (Cenário II)

Parâmetros	$n = 50$		$n = 250$		$n = 500$		Valor verdadeiro
	Média	Mediana	Média	Mediana	Média	Mediana	
λ_{11}	0,9947	0,9890	0,9994	0,9980	0,9970	0,9980	1,0000
λ_{12}	1,4961	1,5000	1,4979	1,5000	1,5010	1,5000	1,5000
λ_{23}	2,5064	2,5000	2,4997	2,5000	2,5006	2,5000	2,5000
λ_{24}	3,0025	3,0000	3,0021	3,0000	2,9996	3,0000	3,0000
β_{21}	3,0001	3,0000	3,0009	3,0000	2,9999	3,0000	3,0000
$\sigma_{X_1X_1}$	1,0091	0,9930	0,998	0,9940	1,0009	0,9990	1,0000
$\sigma_{X_2X_2}$	5,0364	4,9900	5,0062	4,9800	4,9937	5,0000	5,0000
$\sigma_{X_1X_2}$	1,0148	0,9890	1,0005	0,9950	0,9949	0,9920	1,0000
$\sigma_{X_3X_3}$	2,9628	2,9300	2,9970	2,9800	3,0044	3,0000	3,0000
$\sigma_{X_4X_4}$	5,0573	4,9800	5,0108	4,9700	4,9863	4,9800	5,0000
$\sigma_{\epsilon_1\epsilon_1}$	0,9679	0,9570	0,9929	0,9920	0,9948	0,9950	1,0000
$\sigma_{\epsilon_2\epsilon_2}$	1,9010	1,8690	1,9782	1,9600	1,9862	1,9800	2,0000

Para o terceiro cenário, observam-se as estimativas obtidas na Tabela 5, pelo diagrama da Figura 11, considerando-se, agora, a existência de covariância negativa entre as variáveis X_1 e X_2 . É possível observar que as médias dos parâmetros estimados têm valores próximos aos valores verdadeiros considerados para este cenário.

Figura 11 - Representação gráfica do Cenário III

Tabela 5 - Estimativa dos parâmetros do modelo para os diferentes tamanhos amostrais (Cenário III)

Parâmetros	$n = 50$		$n = 250$		$n = 500$		Valor verdadeiro
	Média	Mediana	Média	Mediana	Média	Mediana	
λ_{11}	0,9950	0,9930	1,0031	1,007	1,0008	0,9990	1,0000
λ_{12}	1,4984	1,5000	1,4987	1,5000	1,5003	1,5000	1,5000
λ_{23}	2,5052	2,5000	2,5004	2,5000	2,499	2,5000	2,5000
λ_{24}	3,0005	3,0000	2,9995	3,0000	2,9982	3,0000	3,0000
β_{21}	3,0026	3,0000	2,9991	3,0000	3,0008	3,0000	3,0000
$\sigma_{X_1X_1}$	1,0038	0,9920	1,0020	0,9990	0,9966	0,9930	1,0000
$\sigma_{X_2X_2}$	5,0619	5,0100	4,9885	4,9700	4,9915	4,9900	5,0000
$\sigma_{X_1X_2}$	-1,0151	-0,9840	-0,9984	-0,9870	-0,9957	-0,9960	-1,0000
$\sigma_{X_3X_3}$	3,0212	2,9800	2,9773	2,9700	3,0101	3,0000	3,0000
$\sigma_{X_4X_4}$	5,0309	4,9900	5,0046	5,0000	4,9877	4,9800	5,0000
$\sigma_{\epsilon_1\epsilon_1}$	0,9500	0,9400	0,9976	1,0000	0,9973	0,9950	1,0000
$\sigma_{\epsilon_2\epsilon_2}$	1,8882	1,8550	1,9728	1,9700	1,9884	1,9900	2,0000

O quarto cenário será desenvolvido a partir do surgimento da covariância positiva entre as variáveis exógenas X_3 e X_4 . Nessa etapa, o processo de simulação é semelhante ao cenário 1 e é representado pela Figura 12 e as estimativas dos parâmetros são dadas na Tabela 6.

Figura 12 - Representação gráfica do Cenário IV

Tabela 6 - Estimativa dos parâmetros do modelo para os diferentes tamanhos amostrais (Cenário IV)

Parâmetros	$n = 50$		$n = 250$		$n = 500$		Valor verdadeiro
	Média	Mediana	Média	Mediana	Média	Mediana	
λ_{11}	0,9926	0,9950	0,9997	0,9990	0,9987	1,0010	1,0000
λ_{12}	1,4978	1,5000	1,5021	1,5000	1,5003	1,5000	1,5000
λ_{23}	2,5074	2,5100	2,5031	2,5000	2,5011	2,5000	2,5000
λ_{24}	3,0006	3,0000	3,0003	3,0000	2,9981	3,0000	3,000
β_{21}	2,9992	3,0000	3,0000	3,0000	3,0001	3,0000	3,0000
$\sigma_{X_1X_1}$	0,9975	0,9890	1,0032	1,0010	0,9962	0,9950	1,0000
$\sigma_{X_2X_2}$	4,0209	3,9900	4,0078	4,0100	4,0061	4,0100	4,0000
$\sigma_{X_3X_3}$	3,0442	2,9900	2,9871	2,9700	3,0057	3,0000	3,0000
$\sigma_{X_4X_4}$	5,3984	5,3000	5,3188	5,3100	5,3364	5,3400	5,3333
$\sigma_{X_3X_4}$	1,0083	0,9810	1,0075	1,0080	1,0019	0,9980	1,0000
$\sigma_{\epsilon_1\epsilon_1}$	0,9562	0,9360	0,9862	0,9820	0,9953	0,9940	1,0000
$\sigma_{\epsilon_2\epsilon_2}$	1,8818	1,8300	1,9714	1,9600	1,9901	1,9800	2,0000

Para o quinto cenário, considerou-se a covariância negativa entre as variáveis exógenas X_3 e X_4 . Verificando assim, as estimativas obtidas na Tabela 7, pelo diagrama da Figura 13.

Figura 13 - Representação gráfica Cenário V

Tabela 7 - Estimativa dos parâmetros do modelo para os diferentes tamanhos amostrais (Cenário V)

Parâmetros	$n = 50$		$n = 250$		$n = 500$		Valor verdadeiro
	Média	Mediana	Média	Mediana	Média	Mediana	
λ_{11}	0,9996	1,0060	1,0018	0,9990	1,002	1,0010	1,0000
λ_{12}	1,5017	1,5000	1,5008	1,5000	1,5009	1,5000	1,5000
λ_{23}	2,4997	2,5000	2,5044	2,5000	2,5028	2,5000	2,5000
λ_{24}	2,9989	3,0000	3,0014	3,0000	3,0011	3,0000	3,0000
β_{21}	3,0042	3,0000	3,0003	3,0000	3,0005	3,0000	3,0000
$\sigma_{X_1X_1}$	0,9994	0,9840	0,9973	0,9960	1,0021	1,0010	1,0000
$\sigma_{X_2X_2}$	3,9463	3,9200	4,0072	3,9900	4,0062	4,0000	4,0000
$\sigma_{X_3X_3}$	3,0061	2,9500	2,9960	2,9800	2,9942	2,9900	3,0000
$\sigma_{X_4X_4}$	5,3409	5,2600	5,3346	5,3200	5,3357	5,3300	5,3333
$\sigma_{X_3X_4}$	-1,0152	-0,9890	-1,0043	-1,0140	-0,9939	-0,9940	-1,0000
$\sigma_{\epsilon_1\epsilon_1}$	0,9533	0,9450	0,9961	0,9930	0,9947	0,9930	1,0000
$\sigma_{\epsilon_2\epsilon_2}$	1,8543	1,8310	1,9696	1,9700	1,9857	1,9900	2,0000

Para o ultimo cenário a ser representado nesse processo de simulação, levou-se em consideração apenas covariâncias positivas entre X_1 e X_2 , assim como, para X_3 e X_4 , resultando em estimativas dadas na Tabela 8 e representada pela Figura 14.

Figura 14 - Representação gráfica do Cenário VI

Tabela 8 - Estimativa dos parâmetros do modelo para os diferentes tamanhos amostrais (Cenário VI)

Parâmetros	$n = 50$		$n = 250$		$n = 500$		Valor verdadeiro
	Média	Mediana	Média	Mediana	Média	Mediana	
λ_{11}	0,9926	0,9950	0,9997	0,9990	0,9987	1,0010	1,0000
λ_{12}	1,4978	1,5000	1,5021	1,5000	1,5003	1,5000	1,5000
λ_{23}	2,5074	2,5100	2,5031	2,5000	2,5011	2,5000	2,5000
λ_{24}	3,0006	3,0000	3,0003	3,0000	2,9981	3,0000	3,0000
β_{21}	2,9992	3,0000	3,0000	3,0000	3,0001	3,0000	3,0000
σ_{X1X1}	0,9975	0,9890	1,0032	1,001	0,9962	0,9950	1,0000
σ_{X2X2}	4,0209	3,9900	4,0078	4,0100	4,0061	4,0100	4,0000
σ_{X3X3}	3,0442	2,9900	2,9871	2,9700	3,0057	3,0000	3,0000
σ_{X4X4}	5,3984	5,3000	5,3188	5,3100	5,3364	5,3400	5,3333
σ_{X3X4}	1,0083	0,9810	1,0075	1,0080	1,0019	0,998	1,0000
$\sigma_{\epsilon_1\epsilon_1}$	0,9562	0,9360	0,9862	0,9820	0,9953	0,9940	1,0000
$\sigma_{\epsilon_2\epsilon_2}$	1,8818	1,8300	1,9714	1,9600	1,9901	1,9800	2,0000

Ao utilizar o processo de simulação, faz-se necessário verificar o comportamento do erro quadrático médio como forma de observar a convergência das estimativas dos parâmetros. A seguir, são ilustrados os gráficos referentes aos erros quadrático médios em diferentes tamanhos amostrais, assim, consideraram-se os valores simulados cujo modelo tenha apresentado uma razão da estatística qui-quadrado pelos graus de liberdade menor que três. Também, são apresentados os gráficos boxplot como ilustração do comportamento das estimativas. Os pontos indicam as médias das estimativas dos parâmetros da simulação em cada cenário. Os valores verdadeiros atribuídos aos parâmetros são: $\lambda_{11} = 1$, $\lambda_{12} = 1,5$, $\lambda_{23} = 2,5$, $\lambda_{24} = 3$, $\beta_{21} = 3$, $\sigma_{X1X1} = 1$, $\sigma_{X2X2} = 4$, $\sigma_{X3X3} = 3$, $\sigma_{X4X4} = 5$, $\sigma_{X3X4} = 1$, $\sigma_{\epsilon_1\epsilon_1} = 1$, e $\sigma_{\epsilon_2\epsilon_2} = 2$.

Para o cenário I, observa-se os seguintes gráficos associados ao diagrama da Figura 9:

Figura 15 - Erro quadrático médio associado aos parâmetros: (a) λ_{11} , (b) λ_{12} , (c) λ_{23} , (d) λ_{24} , (e) $\sigma_{X_1X_1}$, (f) $\sigma_{X_2X_2}$, (g) $\sigma_{X_3X_3}$, (h) $\sigma_{X_4X_4}$, (i) β_{21} , (j) $\sigma_{\epsilon_1\epsilon_1}$ e (k) $\sigma_{\epsilon_2\epsilon_2}$

Figura 16 - Gráfico boxplot das estimativas dos parâmetros do Cenário I, com amostras de $n = 50$, em mil iterações

Figura 17 - Histograma para simulação do Cenário I: (a) λ_{11} , (b) λ_{12} , (c) λ_{23} , (d) λ_{24}
 (e) σ_{X1X1} , (f) σ_{X2X2} , (g) σ_{X3X3} , (h) σ_{X4X4} , (i) β_{21}

Figura 18 - Histograma para simulação do Cenário I: (a) $\sigma_{\epsilon_1 \epsilon_1}$ e (b) $\sigma_{\epsilon_2 \epsilon_2}$

Cenário 2: Ilustração dos erros quadráticos médios e boxplot das estimativas.

Figura 19 - Erro quadrático médio associado aos parâmetros: (a) λ_{11} , (b) λ_{12} , (c) λ_{23} , (d) λ_{24} , (e) $\sigma_{X_1X_1}$, (f) $\sigma_{X_2X_2}$, (g) $\sigma_{X_3X_3}$, (h) $\sigma_{X_4X_4}$, (i) β_{21} , (j) $\sigma_{X_1X_2}$, (k) $\sigma_{\epsilon_1\epsilon_1}$ e (l) $\sigma_{\epsilon_2\epsilon_2}$

Figura 20 - Gráfico boxplot das estimativas dos parâmetros do Cenário II, com amostras de $n = 50$, em mil iterações

Figura 21 - Histograma para simulação do Cenário II: (a) λ_{11} , (b) λ_{12} , (c) λ_{23} , (d) λ_{24} , (e) σ_{X1X1} , (f) σ_{X2X2} , (g) σ_{X3X3} , (h) σ_{X4X4} e (i) β_{21}

(a)

(b)

(c)

Figura 22 - Histograma para simulação do Cenário II: (a) σ_{X1X2} , (b) $\sigma_{\epsilon_1\epsilon_1}$ e (c) $\sigma_{\epsilon_2\epsilon_2}$

Cenário 3: Ilustração dos erros quadráticos médios e boxplot das estimativas.

Figura 23 - Erro quadrático médio associado aos parâmetros: (a) λ_{11} , (b) λ_{12} , (c) λ_{23} , (d) λ_{24} , (e) $\sigma_{X_1X_1}$, (f) $\sigma_{X_2X_2}$, (g) $\sigma_{X_3X_3}$, (h) $\sigma_{X_4X_4}$, (i) β_{21} , (j) $\sigma_{X_1X_2}$, (k) $\sigma_{\epsilon_1\epsilon_1}$ e (l) $\sigma_{\epsilon_2\epsilon_2}$

Figura 24 - Histograma para simulação do Cenário III: (a) λ_{11} , (b) λ_{12} , (c) λ_{23} , (d) λ_{24} , (e) $\sigma_{X_1X_1}$, (f) $\sigma_{X_2X_2}$, (g) $\sigma_{X_3X_3}$, (h) $\sigma_{X_4X_4}$ e (i) β_{21}

Figura 25 - Histograma para simulação do Cenário III: (a) $\sigma_{X_1X_2}$, (b) $\sigma_{\epsilon_1\epsilon_1}$ e (c) $\sigma_{\epsilon_2\epsilon_2}$

Figura 26 - Gráfico boxplot das estimativas dos parâmetros do Cenário III, com amostras de $n = 50$, em mil iterações

Cenário 4: Ilustração dos erros quadráticos médios e boxplot das estimativas.

Figura 27 - Erro quadrático médio associado aos parâmetros: (a) λ_{11} , (b) λ_{12} , (c) λ_{23} , (d) λ_{24} , (e) $\sigma_{X_1X_1}$, (f) $\sigma_{X_2X_2}$, (g) $\sigma_{X_3X_3}$, (h) $\sigma_{X_4X_4}$, (i) β_{21} , (j) $\sigma_{X_3X_4}$, (k) $\sigma_{\epsilon_1\epsilon_1}$ e (l) $\sigma_{\epsilon_2\epsilon_2}$

Figura 28 - Histograma para simulação do Cenário IV: (a) λ_{11} , (b) λ_{12} , (c) λ_{23} , (d) λ_{24} , (e) σ_{X1X1} , (f) σ_{X2X2} , (g) σ_{X3X3} , (h) σ_{X4X4} e (i) β_{21}

Figura 29 - Histograma para simulação do Cenário IV: (a) $\sigma_{X_3X_4}$, (b) $\sigma_{\epsilon_1\epsilon_1}$ e (c) $\sigma_{\epsilon_2\epsilon_2}$

Figura 30 - Gráfico boxplot das estimativas dos parâmetros do Cenário IV, com amostras de $n = 50$, em mil iterações

Cenário 5: Ilustração dos erros quadráticos médios e boxplot das estimativas.

Figura 31 - Erro quadrático médio associado aos parâmetros: (a) λ_{11} , (b) λ_{12} , (c) λ_{23} , (d) λ_{24} , (e) $\sigma_{X_1X_1}$, (f) $\sigma_{X_2X_2}$, (g) $\sigma_{X_3X_3}$, (h) $\sigma_{X_4X_4}$, (i) β_{21} , (j) $\sigma_{X_3X_4}$, (k) $\sigma_{\epsilon_1\epsilon_1}$ e (l) $\sigma_{\epsilon_2\epsilon_2}$

Figura 32 - Histograma para simulação do Cenário V: (a) λ_{11} , (b) λ_{12} , (c) λ_{23} , (d) λ_{24} , (e) $\sigma_{X_1X_1}$, (f) $\sigma_{X_2X_2}$, (g) $\sigma_{X_3X_3}$, (h) $\sigma_{X_4X_4}$ e (i) β_{21}

Figura 33 - Histograma para simulação do Cenário V: (a) σ_{X3X4} , (b) $\sigma_{\epsilon_1\epsilon_1}$ e (c) $\sigma_{\epsilon_2\epsilon_2}$

Figura 34 - Gráfico boxplot das estimativas dos parâmetros do Cenário V, com amostras de $n = 50$, em mil iterações

Cenário 6: Ilustração dos erros quadráticos médios e boxplot das estimativas.

Figura 35 - Erro quadrático médio associado aos parâmetros: (a) λ_{11} , (b) λ_{12} , (c) λ_{23} , (d) λ_{24} , (e) $\sigma_{X_1X_1}$, (f) $\sigma_{X_2X_2}$, (g) $\sigma_{X_1X_2}$, (h) $\sigma_{X_3X_3}$, (i) $\sigma_{X_4X_4}$, (j) $\sigma_{X_3X_4}$ e (k) β_{21}

Figura 36 - Erro quadrático médio associado aos parâmetros: $\sigma_{\epsilon_1 \epsilon_1}$ (a) e $\sigma_{\epsilon_2 \epsilon_2}$ (b)

Figura 37 - Histograma para simulação do Cenário VI: (a) λ_{11} , (b) λ_{12} , (c) λ_{23} , (d) λ_{24} , (e) $\sigma_{X_1X_1}$, (f) $\sigma_{X_2X_2}$, (g) $\sigma_{X_1X_2}$, (h) $\sigma_{X_3X_3}$ e (i) $\sigma_{X_4X_4}$

(a)

(b)

(c)

(d)

Figura 38 - Histograma para simulação do Cenário VI: (a) σ_{X3X4} , (b) β_{21} , (c) $\sigma_{\epsilon_1\epsilon_1}$ e (d) $\sigma_{\epsilon_2\epsilon_2}$

Figura 39 - Gráfico boxplot das estimativas dos parâmetros do Cenário VI, com amostras de $n = 50$, em mil iterações

Considerando-se os valores médios da simulação para cada um dos parâmetros presentes em cada um dos seis cenários considerados, verifica-se que as estimativas médias estão bastante próximas dos valores verdadeiros (Tabelas I a VI). O critério indicado por Kline (1998) para seleção dos modelos e obtenção das estimativas para parâmetros, indica

distribuições simétricas para os parâmetros estruturais, enquanto que as variâncias e covariâncias apresentam-se de forma assimétrica. Assim, através desse estudo, pode-se pensar em utilizar outras formas de estimação dos parâmetros, como é o caso dos métodos de estimação bayesianos, uma vez que os parâmetros estruturais possuem, de alguma forma, comportamentos similares aos das distribuições de probabilidades conhecidas, como no caso da função de distribuição normal e da distribuição gama.

5 CONCLUSÃO

Neste trabalho, a abordagem dos Modelos de Equações estruturais trouxe uma revisão e evolução sobre o tema, com destaque aos modelos recursivos que utilizam apenas variáveis observadas. Na aplicabilidade dos MEE, observou-se a existência dessa abordagem em diferentes áreas da ciência, como também, foi possível distinguir as diferentes formas e conceitos das variáveis: observadas, latentes, endógenas e exógenas. Ao longo do trabalho foram encontrados alguns problemas, dos quais, a dificuldade na obtenção dos dados reais e ausência de referencial teórico com abordagem matemática mais aprofundada, no entanto, foi possível utilizar estudos de simulação em diferentes cenários para o desenvolvimento do mesmo.

No estudo de simulação, optou-se em trabalhar com o *software* R Core Team (2013) e o pacote `sem` de (FOX; WEISBERG, 2012) para obtenção das estimativas dos parâmetros por meio do método de máxima verossimilhança. Consequentemente, este trabalho teve como foco, a simulação de uma estrutura de modelos recursivos considerando vários cenários, dos quais, considerou-se a ausência e existência de covariâncias entre as variáveis. Por meio dos diagramas de caminhos associados aos sistemas de equações simuladas, foi possível estimar os valores dos parâmetros, obtendo assim, valores bem próximos aos dos originais. Verificou-se, também, os erros quadráticos médios como uma forma de observar o comportamento dessas estimativas em grandes amostras, visto que o decaimento desses valores indica que, assintoticamente, os valores são próximos aos valores reais.

REFERÊNCIAS

- AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on automatic Control**, Piscataway, v. 19, n. 6, p. 716–723, 1974.
- BOLLEN, K. A. **Structural equation with latent variable**. New York: Wiley series in probability and mathematical statistics, 1989. 514p.
- DRAPER, N. R.; SMITH, H. Applied regression analysis 2nd ed. New York. John Wiley and Sons, 1981. 706p.
- DUNCAN, O. D. **Introduction to structural equation models**. London: academic press, 1975. 180p.
- DUNCAN, O. D.; HALLER, A. O.; PORTES, A. Peer influences on aspirations: A reinterpretation. **American Journal of Sociology**, Chicago, v. 74, n. 2, p. pp. 119–137, 1968. Disponível em: <<http://www.jstor.org/stable/2775916>>. Acesso em: 08 jul. 2014.
- FOX, J. **An Introduction to Structural Equation Modeling**. Rio de Janeiro. 2008. Disponível em: <<http://socserv.mcmaster.ca/jfox/Courses/Brazil-2008/SEMs-notes.pdf>>. Acesso em: 01 mar. 2014.
- FOX, J.; NIE, Z.; BYRNES, J. **sem: Structural Equation Models**. 2013. R package version 3.1-3. Disponível em: <<http://CRAN.R-project.org/package=sem>>. Acesso em: 01 jun. 2014.
- FOX, J.; WEISBERG, S. Structural equation modeling in r with the sem package. **An Appendix to An R Companion to Applied Regression**, 2012. Disponível em: <<http://cran.r-project.org/doc/contrib/Fox-Companion/>>. Acesso em: 07 jul. 2014.
- GRACE, J. B. **Structural equation modeling and natural systems**. New York: Cambridge University Press, 2006. 365p.
- Hair Jr., J. F.; BLACK, W. C.; BABIN, B. J.; ANDERSON, R. E. **Multivariate data analysis**. 7. ed. New York: Prentice Hall, 2010. 761p.
- HORVATH, S. **Weighted network analysis. Applications in Genomics and Systems Biology**. Los Angeles: Springer, 2011. 421p.
- IRIONDO, J. M.; ALBERT, M. J.; ESCUDERO, A. Structural equation modelling: an alternative for assessing causal relationships in threatened plant populations. **Biological Conservation**. Madrid, v. 113, n. 3, p. 367–377, 2003.
- JÖRESKOG, K. G.; SÖRBOM, D. **LISREL 8: Structural equation modeling with the SIMPLIS command language**. New York: Scientific Software International, 1993. 210p.
- KLINE, R. B. **Principles and practice of structural equation modeling**. New York: Guilford, 1998. 427p.
- KLINE, R. B. **Principles and practice of structural equation modeling**. New York: Guilford press, 2011. 360p.

- LEE, S. **Structural equation modeling: A Bayesian approach**. Hong Kong: John Wiley & Sons, 2007. 573p.
- MONTGOMERY, D. C.; PECK, E. A.; VINING, G. G. **Introduction to linear regression analysis**. New Jersey: John Wiley & Sons, 2012. 450p
- MULAIK, S. A. **Linear Causal Modeling with Structural Equations**. New York: Chapman and Hall/CRC, 2009. 428p.
- PALOMO, J.; DUNSON, D. B.; BOLLEN, K. **Bayesian structural equation modeling. Handbook of latent variable and related models**, New York: Elsevier, 2007. 435p.
- PUGESEK, B. H.; TOMER, A. **Structural equation modeling: applications in ecological and evolutionary biology**. New York: Cambridge Univ Pr, 2003. 409p.
- R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2013. Disponível em: <<http://www.R-project.org/>>. Acesso em: 20 jul. 2014.
- RAWLINGS, J. O.; PANTULA, S. G.; DICKEY, D. A. **Applied regression analysis: a research tool**, New York: Springer. 1998. 380p.
- ROSA, G. J. M; VALENTE, B. D.; CAMPOS, G.; WU, X.; GIANOLA, D.; SILVA, M. A. Inferring causal phenotype networks using structural equation models. **Genetics Selection Evolution**, Madison, v. 43, n. 6, p. 1-13. 2011.
- ROSSEEL, Y. lavaan: An R package for structural equation modeling. **Journal of Statistical Software**, v. 48, n. 2, p. 1–36, 2012. Disponível em: <<http://www.jstatsoft.org/v48/i02/>>. Acesso em: 30 ago. 2014.
- SCHUMACKER, R. E.; LOMAX, R. G. **A beginner's guide to structural equation modeling**. London: Routledge. 2010. 510p.
- SCHWARZ, G. Estimating the dimension of a model. **The annals of statistics**, New York. v. 6, n. 2, p. 461–464, 1978.
- SMITH, R.; DAVIS, A. S.; JORDAN, N. R.; ATWOOD, L. W.; DALY, A. B.; GRANDY, A. S.; HUNTER, M. C.; KOIDE, R. T.; MORTENSEN, D. A.; EWING, P. Structural equation modeling facilitates transdisciplinary research on agriculture and climate change. **Crop Science**, New York, v. 54, n. 2, p. 475–483, 2014.
- SOUZA, G. S. **Introducao aos modelos de regressao linear e nao-linear**. Brasilia: Embrapa, 1998. 489p.
- WANG, Y.; FAN, T. A bayesian analysis on time series structural equation models. **Journal of Statistical Planning and Inference**, New York, v. 141, n. 6, p. 2071 – 2078, 2011. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0378375810005756>>. Acesso em: 15 ago. 2014.
- WRIGHT, S. On the nature of size factors. **Genetics**, Washington, v. 3, n. 4, p. 367, 1918.
- WRIGHT, S. Correlation and causation. **Journal of agricultural research**, Washington, v. 20, n. 7, p. 557–585, 1921.

APÊNDICE

Script para simulação Cenário I

```

rm(list=ls(all=TRUE))
require(semPlot)          # Diagrama de caminhos
require(xtable, quietly = TRUE) # tabelas em latex

# Geração de tabelas em tex
Ftable = function(b,titulo,linha) {
  tab = xtable(b, caption=titulo,label="dados",digits = 4,sep=",")
  tabel =print.xtable(tab,table.placement = "H",
    include.rownames = linha,caption.placement="top",
    format.args = list(big.mark = ".", decimal.mark = ","))
  return(tabel)
}

# Geração de matrizes em tex
Fmatrix = function(a) {
  matable = xtable( a, align=rep("", ncol(a)+1) ,sep=",",digits = 3,label="matriz",)
  mattex = print(matable, floating=FALSE, tabular.environment="bmatrix",
    print.results=FALSE,hline.after=NULL, include.rownames=FALSE,
    include.colnames=FALSE,format.args =list(big.mark = ".", decimal.mark = ","))
  return(mattex)
}

# Função Erro quadrático médio em diferentes tamanhos amostrais
EQM=function(I,CausalModelRAM,N){
  n.par = nrow(CausalModelRAM)
  theta.erro = theta.n = matrix(numeric(I*n.par),n.par,dimnames=list(CausalModelRAM[,2]))
  for(i in 1:I){
    mod=MEE(CausalModelRAM,n=N*i,I=I,v.orig=valor,par=parametros)
    theta.erro[,i]=round(rowMeans((mod$Média$theta-valor)^2),4)
    theta.n[,i] = N*i
  }
  return(list("eqm"=theta.erro,"n"=theta.n,"I"=I))
}

# função remove valores "NA" dos EQM
f=function(DAD){
  I=DAD$I
  A=B=matrix(numeric(0),I)
  for(i in 1:I){
    if(!is.na(DAD$eqm[1,i])){B[i] = i}else
    {A[i] = i }
    col=na.omit(A)
    EQM = DAD$eqm[,-col]
    N = DAD$n[,-col]
  }
  return(list("EQM"=EQM,"n"=N))
}

require(sem)
MEE = function(CausalModelRAM,n,I,v.orig,par){
  n.par = nrow(CausalModelRAM) #Parâmetros desconhecidos
  theta = matrix(numeric(I*n.par),n.par,dimnames=list(CausalModelRAM[,2]))
  Pr = theta
  Erro = theta
  for(i in 1:I){

```

```

x1 = rnorm(n,sd=sqrt(sigma.x1x1))
x2 = (sigma.x1x2/sigma.x1x1)*x1+rnorm(n,sd=sqrt(sig.epsx2))
x3 = rnorm(n,sd=sqrt(sigma.x3x3))
x4 = (sigma.x3x4/sigma.x3x3)*x3+rnorm(n,sd=sqrt(sig.epsx4))
e1 = rnorm(n,sd=sqrt(sigma.eps1eps1))
e2 = (sigma.eps1eps2/sigma.eps1eps1)*e1+
      rnorm(n,sd=sqrt(sigma.eps2eps2-((sigma.eps1eps2^2)/sigma.eps1eps1)))
y1 = lambda11*x1+lambda12*x2+e1
y2 = beta21*y1+lambda23*x3+lambda24*x4+e2

datX = data.frame(x1,x2,x3,x4) # Variáveis exógenas
datY = data.frame(y1,y2)      # Variáveis endógenas
p = dim(datY)[[2]]            # Número de variáveis endógenas
q = dim(datX)[[2]]            # Número de variáveis endógenas
dat.obs=data.frame(datX,datY)

S = cov(dat.obs) # SIGMA observado
sem.mod = sem(CausalModelRAM, S=S, N=n,maxiter=1000000)
resumo = summary(sem.mod)      # estat. resumo

if((resumo$chisq/resumo$df)<3)      # critério menor que 3
{theta[,i] = resumo$coeff[,1]; Erro[,i] = resumo$coeff[,2]; Pr[,i] = resumo$coeff[,4]} else
{theta[,i] = NA; Erro[,i] = NA; Pr[,i]= NA }
}
var = cbind(x1,x2,x3,x4,y1,y2)
result.mean = list(
'MeanParâmetros'=rowMeans(theta,na.rm=T), 'theta'=theta, 'P.valor'=Pr,
'Desv.Pd.val'=Erro, "V.original"=v.orig, "I"=I,
'n.par'=nrow(CausalModelRAM))
Resumo = summary.matrix(t(theta),dimnames=F)
result = cbind('Parâmetros'=par, 'Estimativa'=round(rowMeans(theta,na.rm=T),4),
'Mediana'=Resumo[3,], 'valor.original'=v.orig)
return(list( "Média" = result.mean, "tab.resumo" = result,
"path" = sem.mod, "Variáveis" = var))
}

# sigma.x1x2=0 e sigma.x3x4=0
lambda11=1; lambda12=1.5; lambda23=2.5; lambda24=3; beta21=3
sigma.x1x1=1;sigma.x1x2=0;sigma.x3x3=3;sigma.x3x4=0;
sigma.eps1eps1=1; sigma.eps2eps2=2;sigma.eps1eps2=0
sig.epsx2=4;sig.epsx4=5
sigma.x2x2=((sigma.x1x2^2)/sigma.x1x1) + sig.epsx2
sigma.x4x4=((sigma.x3x4^2)/sigma.x3x3) + sig.epsx4
v.orig= c(lambda11,lambda12,lambda23,lambda24,beta21,sigma.x1x1,sigma.x2x2,
sigma.x1x2,sigma.x3x3,sigma.x4x4,sigma.x3x4,sigma.eps1eps1,
sigma.eps2eps2,sigma.eps1eps2)
par = c(";*lambda!(11);",";*lambda!(12);",";*lambda!(23);",";*lambda!(24);",
";*beta!(21);",";*sigma!(x1x1);",";*sigma!(x2x2);",";*sigma!(x1x2);",
";*sigma!(x3x3);",";*sigma!(x4x4);",";*sigma!(x3x4);",
";*sigma!(*epsilon!1*epsilon!1);",";*sigma!(*epsilon!2*epsilon!2);",
";*sigma!(*epsilon!1*epsilon!2);")

modelo=readLines(con = textConnection('
x1 -> y1, lambda11, NA
x2 -> y1, lambda12, NA
x3 -> y2, lambda23, NA
x4 -> y2, lambda24, NA
y1 -> y2, beta21, NA

```

```

x1 <-> x1, sigma.x1x1, NA
x2 <-> x2, sigma.x2x2, NA
x3 <-> x3, sigma.x3x3, NA
x4 <-> x4, sigma.x4x4, NA
y1 <-> y1, sigma.eps1eps1, NA
y2 <-> y2, sigma.eps2eps2, NA
'))
cat(modelo, file = 'model.spec', sep = '\n')
modelo1 <- specifyModel(file = 'model.spec')

# removendo parâmetros para simulação do cenário 1
valor = v.orig[-c(8,11,14)]
parametros = par[-c(8,11,14)]

mod=MEE(CausalModelRAM=modelo1,n=50,I=1000,v.orig=valor,par=parametros)
mod.1=MEE(CausalModelRAM=modelo1,n=250,I=1000,v.orig=valor,par=parametros)
mod.2=MEE(CausalModelRAM=modelo1,n=500,I=1000,v.orig=valor,par=parametros)
M = cbind(mod$tab.resumo[,1],mod$tab.resumo[,c(2,3)],
          mod.1$tab.resumo[,c(2,3)],mod.2$tab.resumo[,c(2,3)],
          mod.2$tab.resumo[,c(4)])

# valores muito pequenos geram "NA", utilizar a função para removê-los!
EQM.matrix = EQM(I=10,CausalModelRAM=modelo1,N=50)

# Renomear título, e corrigir nomenclatura dos parâmetros em Latex
Ftable(M,titulo="Modelo simulado para amostras de tamanho $n=$
        com $I=$ iterações para variáveis",linha=F)

# Armazenando dados simulados em arquivo extensão .RData
save.image("x1x2=0.RData")

# Gráficos dos EQM e Boxplot
load("/home/erasnilson/Dropbox/Meus arquivos/simulacao/Algorit.DInsert/cov_x1x2/Dados/x1x2=0.RData")

for(i in 1:nrow(modelo1)){
  plot(erro$n[i,],erro$eqm[i,],xlab="Amostra(n)",lty=20,col="seagreen",#cex.lab=2,
       pch = 20,cex=1.5,ylab=expression(paste(EQM)))
}

cont = na.omit(mod$Estimativas$theta[1,])
(Aceitação=(length(cont)/1000)*100)

cont = na.omit(mod.1$Estimativas$theta[1,])
(Aceitação=(length(cont)/1000)*100)

cont = na.omit(mod.2$Estimativas$theta[1,])
(Aceitação=(length(cont)/1000)*100)

# Boxplot para estimativas dos parâmetros
# mudar "mod", para diferentes tamanhos amostrais
box = t(mod$Estimativas$theta)
head(box)
boxplot.matrix(box[,1:4],col="lightcyan ",
               names=c(expression(lambda[11]),expression(lambda[12]),
                       expression(lambda[23]),expression(lambda[24]))),
               points(c(1,2,3,4),valor[1:4],col="red",cex=.8,pch=19)

boxplot(box[,5],col="lightcyan ",xlab=expression(beta[21])) )

```

```

points(c(1),valor[5],col="red",cex=.8,pch=19)

boxplot.matrix(box[,6:9],col="lightcyan ",
               names=c(expression(sigma[x1x1]),expression(sigma[x2x2]),
                       expression(sigma[x3x3]),expression(sigma[x4x4]))),
points(c(1,2,3,4),valor[6:9],col="red",cex=.8,pch=19)

boxplot.matrix(box[,10:11],col="lightcyan ",
               names=c(expression(sigma[epsilon[1]*epsilon[1]]),
                       expression(sigma[epsilon[2]*epsilon[2]]))),
points(c(1,2),valor[10:11],col="red",cex=.8,pch=19)

```

Script para o Cenário II

```

load("/home/erasnilson/Dropbox/Meus arquivos/simulacao/Algorit.DISSERT/cov_x1x2/Dados/x1x2=1.RData")
for(i in 1:nrow(modelo3)){
  plot(erro$n[i,],erro$eqm[i,],xlab="Amostra(n)",lty=10,col="seagreen",
       pch = 20,cex=1.5,ylab=expression(paste(EQM)))
}

cont = na.omit(mod$Estimativas$theta[1,])
(Aceitação=(length(cont)/1000)*100)

cont = na.omit(mod.1$Estimativas$theta[1,])
(Aceitação=(length(cont)/1000)*100)

cont = na.omit(mod.2$Estimativas$theta[1,])
(Aceitação=(length(cont)/1000)*100)

box = t(mod$Estimativas$theta)
head(box)
boxplot.matrix(box[,1:4],col="lightcyan ",
               names=c(expression(lambda[11]),expression(lambda[12]),
                       expression(lambda[23]),expression(lambda[24]))),
points(c(1,2,3,4),valor[1:4],col="red",cex=.8,pch=19)

boxplot(box[,5],col="lightcyan ",xlab=expression(beta[21]))
points(c(1),valor[5],col="red",cex=.8,pch=19)

boxplot.matrix(box[,6:10],col="lightcyan ",
               names=c(expression(sigma[x1x1]),expression(sigma[x2x2]),
                       expression(sigma[x1x2]),
                       expression(sigma[x3x3]),expression(sigma[x4x4]))),
points(c(1,2,3,4,5),valor[6:10],col="red",cex=.8,pch=19)

boxplot.matrix(box[,11:12],col="lightcyan ",
               names=c(expression(sigma[epsilon[1]*epsilon[1]]),
                       expression(sigma[epsilon[2]*epsilon[2]]))),
points(c(1,2),valor[11:12],col="red",cex=.8,pch=19)

for(i in 1:nrow(modelo3)){
  hist(box[,i],ylab="Frequência",xlab="",main="")
  abline(v=valor[i],col="red",cex=100,lwd=4)
}

```

Script para o Cenário III

```
rm(list=ls(all=TRUE))
load("/home/erasnilson/Dropbox/Meus arquivos/simulacao/Algorit.DISSERT/cov_x1x2/Dados/
x1x2=-1.RData")
for(i in 1:nrow(modelo3)){
  plot(erro$n[i,],erro$eqm[i,],xlab="Amostra(n)",lty=10,col="seagreen",
       pch = 20,cex=1.5,ylab=expression(paste(EQM)))
}

cont = na.omit(mod$Estimativas$theta[1,])
(Aceitação=(length(cont)/1000)*100)

cont = na.omit(mod.1$Estimativas$theta[1,])
(Aceitação=(length(cont)/1000)*100)

cont = na.omit(mod.2$Estimativas$theta[1,])
(Aceitação=(length(cont)/1000)*100)

box = t(mod$Estimativas$theta)
head(box)
boxplot.matrix(box[,1:4],col="lightcyan ",
               names=c(expression(lambda[11]),expression(lambda[12]),
                       expression(lambda[23]),expression(lambda[24]))),
               points(c(1,2,3,4),valor[1:4],col="red",cex=.8,pch=19)

boxplot(box[,5],col="lightcyan ",xlab=expression(beta[21]))
points(c(1),valor[5],col="red",cex=.8,pch=19)

boxplot.matrix(box[,6:10],col="lightcyan ",
               names=c(expression(sigma[x1x1]),expression(sigma[x2x2]),
                       expression(sigma[x1x2]),
                       expression(sigma[x3x3]),expression(sigma[x4x4]))),
               points(c(1,2,3,4,5),valor[6:10],col="red",cex=.8,pch=19)

boxplot.matrix(box[,10:11],col="lightcyan",
               names=c(expression(sigma[epsilon[1]*epsilon[1]]),
                       expression(sigma[epsilon[2]*epsilon[2]]))),
               points(c(1,2),valor[10:11],col="red",cex=.8,pch=19)

for(i in 1:nrow(modelo3)){
  hist(box[,i],ylab="Frequência",xlab="",main="")
  abline(v=valor[i],col="red",cex=100,lwd=4)
}
```

Script para o Cenário IV

```
load("/home/erasnilson/Dropbox/Meus arquivos/simulacao/Algorit.DISSERT/Cov_x3_x4/
Dados_x3_x4/x3x4=1.RData")
for(i in 1:nrow(modelo4)){
  plot(erro$n[i,],erro$eqm[i,],xlab="Amostra(n)",lty=20,col="seagreen",
       pch = 20,cex=1.5,ylab=expression(paste(EQM)))
}

cont = na.omit(mod$Estimativas$theta[1,])
(Aceitação=(length(cont)/1000)*100)
```



```

cont = na.omit(mod.1$Estimativas$theta[1,])
(Aceitação=(length(cont)/1000)*100)

cont = na.omit(mod.2$Estimativas$theta[1,])
(Aceitação=(length(cont)/1000)*100)

box = t(mod$Estimativas$theta)
head(box)
boxplot.matrix(box[,1:4],col="lightcyan ",
               names=c(expression(lambda[11]),expression(lambda[12]),
                           expression(lambda[23]),expression(lambda[24]))),
points(c(1,2,3,4),valor[1:4],col="red",cex=.8,pch=19)

boxplot(box[,5],col="lightcyan ",xlab=expression(beta[21]) )
points(1,valor[5],col="red",cex=.8,pch=19)

boxplot.matrix(box[,6:10],col="lightcyan ",
               names=c(expression(sigma[x1x1]),expression(sigma[x2x2]),
                           expression(sigma[x3x3]),expression(sigma[x4x4]),
                           expression(sigma[x3x4]))),
points(c(1,2,3,4,5),valor[6:10],col="red",cex=.8,pch=19)

boxplot.matrix(box[,10:11],col="lightcyan ",
               names=c(expression(sigma[epsilon[1]*epsilon[1]]),
                           expression(sigma[epsilon[2]*epsilon[2]]))),
points(c(1,2),valor[10:11],col="red",cex=.8,pch=19)

for(i in 1:nrow(modelo4)){
  hist(box[,i],ylab="Frequência",xlab="",main="")
  abline(v=valor[i],col="red",cex=100,lwd=4)
}

```

Script para o Cenário V

```

load("/home/erasnilson/Dropbox/Meus arquivos/simulacao/Algorit.DISSERT/Cov_x3_x4/
Dados_x3_x4/x3x4=-1.RData")
for(i in 1:nrow(modelo5)){
  plot(erro$n[i,],erro$eqm[i,],xlab="Amostra(n)",lty=10,col="seagreen",
       pch = 20,cex=1.5,ylab=expression(paste(EQM)))
}

cont = na.omit(mod$Estimativas$theta[1,])
(Aceitação=(length(cont)/1000)*100)

cont = na.omit(mod.1$Estimativas$theta[1,])
(Aceitação=(length(cont)/1000)*100)

cont = na.omit(mod.2$Estimativas$theta[1,])
(Aceitação=(length(cont)/1000)*100)

box = t(mod$Estimativas$theta)
head(box)
boxplot.matrix(box[,1:4],col="lightcyan ",
               names=c(expression(lambda[11]),expression(lambda[12]),
                           expression(lambda[23]),expression(lambda[24]))),
points(c(1,2,3,4),valor[1:4],col="red",cex=.8,pch=19)

```

```

boxplot(box[,5],col="lightcyan ",xlab=expression(beta[21])) )
points(1,valor[5],col="red",cex=.8,pch=19)

boxplot.matrix(box[,6:10],col="lightcyan ",
               names=c(expression(sigma[x1x1]),expression(sigma[x2x2]),
                        expression(sigma[x3x3]),expression(sigma[x4x4]),
                        expression(sigma[x3x4]))))
points(c(1,2,3,4,5),valor[6:10],col="red",cex=.8,pch=19)

boxplot.matrix(box[,10:11],col="lightcyan ",
               names=c(expression(sigma[epsilon[1]*epsilon[1]]),
                        expression(sigma[epsilon[2]*epsilon[2]])))
points(c(1,2),valor[10:11],col="red",cex=.8,pch=19)

for(i in 1:nrow(modelo5)){
  hist(box[,i],ylab="Frequência",xlab="",main="")
  abline(v=valor[i],col="red",cex=100,lwd=4)
}

```

Script para o Cenário VI

```

load("/home/erasnilson/Dropbox/Meus arquivos/simulacao/Algorit.DISSERT/cov_x1x2 e cov_x3_x4/
Dados_x12_x34/x12_x34=1.RData")
for(i in 1:nrow(modelo6)){
  plot(erro$n[i,],erro$eqm[i,],xlab="Amostra(n)",lty=20,col="seagreen",
       pch = 20,cex=1.5,ylab=expression(paste(EQM)))
}
cont = na.omit(mod$Estimativas$theta[1,])
(Aceitação=(length(cont)/1000)*100)

cont = na.omit(mod.1$Estimativas$theta[1,])
(Aceitação=(length(cont)/1000)*100)

cont = na.omit(mod.2$Estimativas$theta[1,])
(Aceitação=(length(cont)/1000)*100)

box = t(mod$Estimativas$theta)
head(box)
boxplot.matrix(box[,1:4],col="lightcyan ",
               names=c(expression(lambda[11]),expression(lambda[12]),
                        expression(lambda[23]),expression(lambda[24]))))
points(c(1,2,3,4),valor[1:4],col="red",cex=.8,pch=19)

boxplot(box[,5],col="lightcyan ",xlab=expression(beta[21])) )
points(1,valor[5],col="red",cex=.8,pch=19)

boxplot.matrix(box[,6:11],col="lightcyan ",
               names=c(expression(sigma[x1x1]),expression(sigma[x2x2]),
                        expression(sigma[x1x2]),
                        expression(sigma[x3x3]),expression(sigma[x4x4]),
                        expression(sigma[x3x4]))))
points(c(1,2,3,4,5,6),valor[6:11],col="red",cex=.8,pch=19)

boxplot.matrix(box[,10:11],col="lightcyan ",

```

```
names=c(expression(sigma[epsilon[1]*epsilon[1]]),  
          expression(sigma[epsilon[2]*epsilon[2]]))  
points(c(1,2),valor[10:11],col="red",cex=.8,pch=19)  
  
for(i in 1:nrow(modelo6)){  
  hist(box[,i],ylab="Frequência",xlab="",main="")  
  abline(v=valor[i],col="red",cex=100,lwd=4)  
}
```