

# Modelação de um conjunto de dados financeiros em painel por um modelo de equações simultâneas

*SPE 2019 - Amarante*

**A. Rocha\*, C. Miranda\*, E. Vieira\*\* e M. S. Miranda\*\*\***  
**\*CIDMA e ISCA, \*\*GOVCOP e ISCA, \*\*\*CIDMA**  
**Universidade de Aveiro**



# Sumário

- 1 Introdução
- 2 O SEM
- 3 O modelo e os dados
- 4 Análise dos resultados
- 5 Comentários finais
- 6 Referências

- Estimar modelo relativo a um conjunto de variáveis financeiras.
- Considerar interdependência entre variáveis (endogeneidade).
- Dados reais com observações atípicas.
- Dados em painel ou dados longitudinais - dados de um conjunto de empresas com várias observações anuais para cada empresa.

# SEM - *Simultaneous Equations Model*

## Modelo de equações simultâneas - SEM:

- é um modelo estatístico caracterizado por um sistema de equações interdependentes;
- generaliza um modelo de regressão multivariado - erros correlacionados com regressores e heterocedásticos;
- inclui dois tipos de variáveis:
  - endógenas - dependem de outras variáveis do sistema;
  - exógenas - determinadas exteriormente ao sistema;
- permite descrever problemas reais em que existem relações de dependência entre variáveis explicativas.

# SEM - *Simultaneous Equations Model*

## Forma estrutural (FE)

$$\mathbf{Y}\Gamma + \mathbf{X}\mathbf{B} + \mathbf{E} = \mathbf{0},$$

$\mathbf{Y}$  e  $\mathbf{X}$ — matrizes das observações das variáveis endógenas e exógenas;  
 $\Gamma$  e  $\mathbf{B}$ — matrizes de parâmetros estruturais;  
 $\mathbf{E}$ — matriz dos erros aleatórios.

## Pressupostos

Os erros são gerados por processo multivariado estacionário, são **não correlacionados ao longo do tempo**, não correlacionados com as variáveis exógenas e seguem um modelo normal multivariado.

# SEM - *Simultaneous Equations Model*

Introdução

O SEM

O modelo e  
os dados

Análise dos  
resultados

Comentários  
finais

Referências

Outra representação do SEM:

$$\mathbf{y} = \mathbf{Z}\boldsymbol{\delta} + \mathbf{e},$$

$$\mathbf{y} = \text{vec}(\mathbf{Y}), \mathbf{Z} = \text{diag}[\mathbf{Z}_1 \quad \cdots \quad \mathbf{Z}_M], \text{ com } \mathbf{Z}_i = [\mathbf{Y}_i \quad \mathbf{X}_i].$$

Variáveis  $\mathbf{Z}_i$  - variáveis explicativas do SEM, incluem variáveis endógenas, as quais são correlacionadas com os erros.

# Estimação do SEM

- **Estimação tradicional** - Estimadores *2SLS* (*Two stages least squares*) e *3SLS* (*Three stages least squares*)- estimadores com boas propriedades sob um conjunto de pressupostos do modelo, mas podem sofrer grandes perturbações quando há desvios em relação ao modelo ou na presença de observações atípicas na amostra.
- **Estimação robusta** é pouco sensível a ligeiros afastamentos dos pressupostos assumidos para o modelo.
- Neste trabalho usámos uma adaptação do estimador robusto proposto em Rocha (2010) - **estimador RGMM**.

# Estimação robusta do SEM

Resumo do algoritmo de cálculo do estimador RGMM:

- **P.1.** Obter valores iniciais dos resíduos - estimativas 2SLS.
- **P.2.** Estimar a matriz de covariâncias dos erros - estimador robusto Fast MCD (*Minimum Covariance Determinant*), com base nos resíduos obtidos em P.1.
- **P.3.** Resolver o problema de minimização - função de Huber com resíduos ponderados pelas estimativas das covariâncias obtidas em P.2.



# Modelo de dados em painel

Para comparação dos resultados, procedemos à estimação dos parâmetros pela metodologia usual para modelos de dados em painel (**considerando a correlação temporal**).

Um modelo de **dados em painel** pode ser descrito pela equação:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + (\mu_i + \nu_{it}), \quad (1)$$

$i$  - indivíduos, empresas, países, etc.;  $t$  - o tempo;  $\boldsymbol{\beta}$  - vetor de parâmetros do modelo;  $\mathbf{x}_{it}$  -  $i$ -ésima observação relativa a  $K$  variáveis explicativas;  $\mu_i$  - efeito não observável relativo aos indivíduos;  $\nu_{it}$  - restante variação aleatória não explicada.

# Estimação do modelo de dados em painel

Vamos considerar um modelo de dados em painel com efeitos aleatórios, no qual se supõe:

- $\mu_j$  são iid com média zero e variância  $\sigma_\mu^2$ ,
- $\nu_{it}$  são iid com média zero e variância  $\sigma_\nu^2$ ,
- $\mu_j$  e  $\nu_{it}$  independentes,
- $x_{it}$  independentes de  $\mu_j$  e de  $\nu_{it}, \forall i, t$ .

A estimação deste modelo é usualmente feita pelo Método dos Mínimos Quadrados Generalizado.

## O modelo e os dados

- A Estrutura de Capital e a Política de Dividendos são dois tópicos clássicos das Finanças Empresariais.
- Os dois tópicos têm sido analisados de forma isolada, não sendo considerada a sua interdependência. No entanto, de acordo com a teoria, é de esperar que exista interdependência: a necessidade de financiar a empresa condiciona a distribuição de dividendos; por outro lado, se distribuirmos, a empresa poderá necessitar de mais empréstimos para investir.
- Os resultados publicados em diferentes estudo empíricos não são consensuais.

# O modelo e os dados

Objetivo: Estimar o modelo que descreve a variação das variáveis, considerando a sua interdependência:

- Estrutura de capital (Endividamento)- **LEV**  
Peso do capital alheio, e portanto, que tem de ser reembolsado, sobre a totalidade das origens de capital da empresa (próprias e alheias).
- Política de dividendos (Dividendos)- **DPO**  
Representa a parte dos resultados obtidos na atividade da empresa que são distribuídos aos acionistas, como fonte de remuneração.
- Propriedade de capital (Capital)- **MGO**  
Capital da empresa que pertence aos membros do conselho de administração, ou seja, àqueles que definem a estratégia das empresas.

## O modelo e os dados

- A propriedade de capital (MGO), a estrutura de capital (LEV) e a política de dividendos (DPO) podem ser vistas como três forças internas capazes de minimizar os custos de agência.
- Os custos de agência resultam de conflitos entre vários agentes das empresas, como por exemplo, entre acionistas (que detêm o capital da empresa) e os gestores (que tomam decisões): os gestores devem tomar decisões que favoreçam os primeiros, aumentando a sua riqueza, mas nem sempre o fazem, e tomam decisões em favor deles próprios (por exemplo, gastos em mordomia). Assim, se uma empresa pagar dividendos, ou tiver mais empréstimos, fica com menos dinheiro para os gestores gastarem em seu próprio proveito. Por outro lado, quanto maior a participação acionista, maior será a sua participação na gestão, e nas tomadas de decisão.

## O modelo

Propomos a modelação que se segue.

Equação referente ao **endividamento**:

$$LEV = \gamma_{21}DPO + \gamma_{31}MGO + \beta_{11}x_1 + \beta_{21}ROA + \beta_{31}TANG + e_1. \quad (2)$$

Equação referente aos **dividendos**:

$$DPO = \gamma_{12}LEV + \gamma_{32}MGO + \beta_{12}x_1 + \beta_{42}SIZE + \beta_{52}GRTH + e_2. \quad (3)$$

Equação referente ao **capital**:

$$MGO = \gamma_{23}DPO + \beta_{13}x_1 + \beta_{23}ROA + \beta_{63}FCF + e_3. \quad (4)$$

# Variáveis

- **LEV - *Endividamento***, razão entre o valor contabilístico da dívida total e o valor contabilístico dos ativos totais;
- **DPO - *Dividendos***, razão entre o dividendo por ação e o lucro por ação.
- **MGO - *Capital***, razão entre o número de ações detidas por membros da Direção e o número total de ações;
- **ROA - *Rendibilidade***, razão entre o lucro antes de juros e imposto, e o total de ativos;
- **SIZE - *Dimensão***,  $\ln(\text{total ativo})$ ;
- **TANG - *Ativo não corrente***, razão entre o total do ativo não corrente e o total do ativo;
- **GRTH - *Crescimento das vendas***, razão entre a diferença de vendas num período e no anterior e as vendas no período anterior;
- **FCF - *Fluxos de caixa disponíveis***, razão entre o resultado da atividade principal da empresa, corrigido de gastos que não implicam pagamentos ao exterior, e o total do ativo.

# Os dados

O estudo incidiu sobre o **mercado ibérico** - do ponto de vista financeiro estas empresas têm características semelhantes:

- **Propriedade de capital concentrada** - O capital da empresa está nas mãos de um número reduzido de acionistas.
- **Baixa proteção de acionistas minoritários** - A legislação não protege os interesses dos acionistas que detêm pequenas percentagens de capital, apenas faz recomendações.
- **Mercado de capitais pouco desenvolvido** - O mercado de capitais tem poucas empresas cotadas e as ações disponíveis para circulação não são muitas. Por outro lado, há dias em que algumas ações não são transacionadas.



# Os dados

- Empresas da Península Ibérica, cotadas em bolsa (Euronext e Bolsa de Madrid), excluindo empresas das áreas financeira e desportiva (regras contabilísticas diferentes).
- Valores retirados da base de dados SABI<sup>1</sup> e da Comissão do Mercado de Valores Mobiliários (CMVM<sup>2</sup>).
- **Variáveis endógenas** - LEV, DPO e MGO.
- **Variáveis exógenas** - ROA, SIZE, TANG, GRTH, FCF e  $x_1$  (termo constante).

---

<sup>1</sup>Base de dados financeiros sobre empresas de Portugal e Espanha, da responsabilidade da Bureau Van Dijk, subsidiária da Moody's

<sup>2</sup><http://www.cmvm.pt>

# Os dados

A partir de um conjunto inicial de 223 empresas, ficaram 76 empresas e trabalhámos com os anos disponíveis para cada uma delas (entre 7 e 24 anos), resultando num conjunto formado por 1284 observações (**dados em painel não balanceados**).

Dividimos este conjunto em duas subamostras:

- **Amostra para estimação** - obtida selecionando aleatoriamente 51 empresas - **amostra com 816 observações** - usada para estimar os parâmetros;
- **Amostra para teste** - formada pelas restantes 25 empresas - **amostra com 468 observações** - usada para calcular os resíduos e testar os resultados.

Da análise exploratória das amostras obtiveram-se os resultados:

|                  | Amostra-E |            | Amostra-T |            |
|------------------|-----------|------------|-----------|------------|
|                  | Mínimo    | Máximo     | Mínimo    | Máximo     |
| Amplitude        | 0.98      | 5 629 879  | 0.94      | 3 666 619  |
| sd               | 0.13      | 235 397.80 | 0.12      | 246 161.50 |
| MAD <sup>3</sup> | 0         | 19 439.84  | 0         | 27 233.88  |
| $\bar{x}$        | 0.07      | 44 882.68  | 0.03      | 100 786.20 |
| Med              | 0         | 10 584.01  | 0         | 18 683     |

---

<sup>3</sup>MAD = Med(| $x_i$  - Med( $x_i$ )|)

# Os dados

- Procedemos à standardização dos dados antes de estimar os parâmetros do modelo, dadas as grandes disparidades existentes nos valores observados das diferentes variáveis.
- Confirmámos a identificabilidade de todas as equações do modelo antes de proceder à estimação.
- Todos os procedimentos relativos ao tratamento e análise dos dados foram efetuados com o programa R (R Core Team 2017).

Procedemos à estimação dos parâmetros do modelo:

- com o estimador robusto do SEM, **ignorando a correlação temporal**.
- com o estimador do modelo de dados em painel com efeitos aleatórios, **ignorando a endogeneidade** - *package plm*<sup>4</sup>.

Não há nenhum *package* que permita incorporar os procedimentos robustos descritos nas três fases do processo de estimação. O *package systemfit*<sup>5</sup> permite estimar o SEM apenas pelos métodos tradicionais.

---

<sup>4</sup><https://cran.r-project.org/web/packages/plm/plm.pdf>

<sup>5</sup><https://cran.r-project.org/web/packages/systemfit/systemfit.pdf>

# Análise dos resultados

Introdução

O SEM

O modelo e os dados

Análise dos resultados

Comentários finais

Referências

Estimativas dos coeficientes do modelo - amostra para estimação:

|         | LEV   | DPO    | MGO    | Cte    | ROA    | TANG  | SIZE  | GRTH   | FCF   |
|---------|-------|--------|--------|--------|--------|-------|-------|--------|-------|
| RGMM.1  |       | 0.440  | -0.360 | -0.001 | -0.037 | 0.133 |       |        |       |
| ALEAT.1 |       | 0.002  | -0.163 | 0.000  | -0.042 | 0.121 |       |        |       |
| RGMM.2  | 0.400 |        | 0.284  | -0.009 |        |       | 0.127 | -0.003 |       |
| ALEAT.2 | 0.001 |        | 0.068  | 0.000  |        |       | 0.129 | -0.002 |       |
| RGMM.3  |       | -0.697 |        | 0.002  | 0.127  |       |       |        | 0.288 |
| ALEAT.3 |       | 0.035  |        | 0.000  | 0.152  |       |       |        | 0.134 |

# Análise dos resultados

A comparação das duas abordagens baseou-se numa análise de resíduos:

- Calculámos os valores dos resíduos relativos a cada uma das equações, para cada uma das metodologias.
- Calculámos a raiz do erro quadrático médio (*RMSE*) dos vetores dos resíduos, para cada uma das metodologias, de acordo com (5)

$$RMSE(\mathbf{e}_i) = \sqrt{\frac{1}{n}[(\mathbf{e}_i - \bar{\mathbf{e}}_i)'(\mathbf{e}_i - \bar{\mathbf{e}}_i)]}. \quad (5)$$

# Análise dos resultados

- Calculámos a distância de Mahalanobis robusta ( $RMD$ ) dos vetores dos resíduos, de acordo com (6)

$$RMD_i = \sqrt{(\mathbf{e}_i - T(\mathbf{E}))' \mathbf{C}(\mathbf{E})^{-1} (\mathbf{e}_i - T(\mathbf{E}))}, \quad (6)$$

$T(\mathbf{E})$  e  $\mathbf{C}(\mathbf{E})$  - estimadores MCD de localização e escala.

- Representámos os valores das  $RMD$  em boxplot e calculámos algumas estatísticas descritivas.



# Análise dos resultados

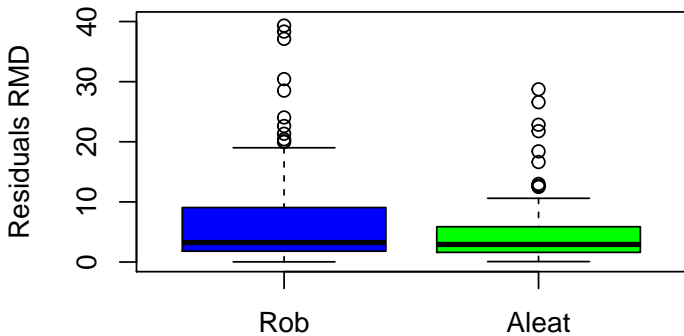
Valores do RMSE dos resíduos de cada equação:

|       | RMSE  |       |       |
|-------|-------|-------|-------|
|       | Eq.1  | Eq.2  | Eq.3  |
| RGMM  | 0.683 | 0.868 | 1.039 |
| ALEAT | 0.602 | 0.880 | 0.763 |

- Os valores do RMSE dos resíduos são semelhantes para os dois métodos, para as equações 1 e 2.
- Apenas para a equação 3 se obtém maior RMSE dos resíduos pelo método robusto.
- De acordo com o RMSE, os dois métodos de estimação apresentam idêntica qualidade de ajustamento.

# Análise dos resultados

Distâncias de Mahalanobis Robustas (RMD) dos resíduos



# Análise dos resultados

Estatísticas descritivas das RMD dos resíduos:

|       | RMD       |       |         |       |
|-------|-----------|-------|---------|-------|
|       | $\bar{x}$ | Med   | sd      | MAD   |
| RGMM  | 65.634    | 3.265 | 258.860 | 2.645 |
| ALEAT | 74.609    | 2.927 | 431.637 | 2.423 |

- A média e o desvio padrão foram fortemente inflacionadas pelos resíduos atípicos (boxplot). **Método robusto é menos afetado.**
- **A mediana e o MAD foram menos influenciados pelos resíduos atípicos** e têm valores semelhantes para ambas as metodologias.
- O método robusto conduziu a menores RMD dos resíduos (localização e dispersão menor ou igual).

## Comentários finais

O erro quadrático médio mostrou que as duas metodologias conduzem a resultados análogos.

A distância de Mahalanobis robusta permitiu diferenciar melhor as duas abordagens.

O método robusto teve melhor desempenho: valores menores das distâncias de Mahalanobis robustas e valores idênticos do erro quadrático médio.

# Comentários finais

Na análise efetuada, o facto de não se considerar a correlação temporal não prejudicou o resultado da estimação e permitiu simplificar a modelação.

Será interessante aplicar métodos robustos também no processo de estimação de dados em painel e procurar identificar qual a modelação e metodologia de estimação mais adequada.

## Obrigada.

As autoras foram apoiadas pela Fundação para a Ciência e a Tecnologia (FCT), através do CIDMA - Centro de Investigação de Matemática Aplicada e do GOVCOPP - Unidade de Investigação em Governança, Competitividade e Políticas Públicas.

- [1] Ashley Vo, D. H., Nguyen, V. T., Managerial Ownership, Leverage and Dividend Policies: Empirical Evidence from Vietnam's Listed Firms Models. *International Journal of Economics and Finance*, 6(5):274-284, 2014.
- [2] Baltagi, B.H., *Econometric analysis of panel data*, Wiley, 2013.
- [3] Farinha, J., Dividend policy, corporate governance and the managerial entrenchment hypothesis: An empirical analysis, *Journal of Business Finance and Accounting*, Vol. 30 No. 9 e 10, pp. 1173-1209, 2003.
- [4] Lee, C., Liang, W., Lin, F., Yang, Y. Applications of simultaneous equations in finance research: methods and empirical results. *Review of Quantitative Finance and Accounting*, 47(04): 943-971, 2016.
- [5] Greene, W., *Econometric Analysis*, Seventh edition, Prentice Hall, 2011.
- [6] Myers, S.C., The capital structure puzzle, *Journal of Finance*, Vol. 39 No. 3, pp. 575-592, 1984.
- [7] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2017). <https://www.R-project.org/>.
- [8] Rousseeuw, P. and Van Driessen, K.. A Fast Algorithm for the Minimum Covariance Determinant Estimator. *Technometrics*, Vol.41 No. 3, pp. 212-223, 1999.