



v.10, n.1

Vitória-ES, jan.-. mar. 2013

p. 131 - 156 ISSN 1807-734X DOI: <http://dx.doi.org/10.15728/bbr.2013.10.1.6>

## Dados em painel em contabilidade e finanças: teoria e aplicação

**Luiz Paulo Lopes Fávero<sup>†</sup>**

*Universidade de São Paulo*

### RESUMO

A utilização de modelos que envolvam dados longitudinais em contabilidade e finanças tem sido recorrente. No entanto verifica-se uma falta de cuidado quanto aos critérios para a adoção de um modelo em detrimento de outro, bem como a ausência de uma discussão mais detalhada sobre os possíveis estimadores a serem estudados em cada situação. Este artigo tem por objetivo apresentar, de forma conceitual e aplicada, os principais estimadores de dados em painel que podem ser utilizados nessas áreas do conhecimento, bem como auxiliar na definição do modelo mais consistente a ser adotado, em função das características dos dados. Para um painel curto, são discutidos os modelos POLS com erros-padrão robustos clusterizados, com estimador *between*, efeitos fixos, efeitos fixos com erros-padrão robustos clusterizados, efeitos aleatórios e efeitos aleatórios com erros-padrão robustos clusterizados. Já para um painel longo, são discutidos os modelos com efeitos fixos, efeitos aleatórios, efeitos fixos com termos de erro AR(1), efeitos aleatórios com termos de erro AR(1), POLS com erros AR(1) e *Pooled* FGLS com erros AR(1). Este artigo também tem por propósito aplicar tais modelos em um caso real, com base nos dados da *Compustat Global*. Ao final, são apresentadas as principais rotinas para a aplicação de cada um dos modelos em Stata.

**Palavras-Chave:** Dados em painel; contabilidade; finanças; método de estimação.

Recebido em 02/05/2011; aceito em 26/02/2012; divulgado em 26/03/2013.

**\*Autor para correspondência:**

<sup>†</sup> Professor Livre-Docente da Universidade de São Paulo.

**Vínculo:** Professor Livre-Docente da Universidade de São Paulo

**Endereço:** Av. Professor Luciano Gualberto, 908, FEA 3, sala 211-  
Cidade Universitária, São Paulo – SP – Brasil

**E-mail:** [lpfaver@usp.br](mailto:lpfaver@usp.br)

**Telefone:** (11)3091-5820

**Nota do Editor:** Esse artigo foi aceito por Bruno Funchal



Este trabalho foi licenciado com uma Licença [Creative Commons - Atribuição 3.0 Não Adaptada](http://creativecommons.org/licenses/by/3.0/).

## 1 INTRODUÇÃO

É

cada vez mais crescente e importante a utilização de modelos que envolvam dados provenientes de várias *cross-sections* ao longo do tempo (dados em painel) em contabilidade e finanças. Como muitos dados de empresas, municípios ou países são divulgados periodicamente, o pesquisador é convidado, naturalmente, a aplicar modelos longitudinais para o estudo de fenômenos que sofrem influência das diferenças entre os indivíduos e da própria evolução temporal.

Segundo Marques (2000), a principal vantagem da utilização de modelos de dados em painel refere-se ao controle da heterogeneidade individual, ou seja, à possibilidade de se medirem separadamente os efeitos gerados por conta de diferenças existentes entre cada observação em cada *cross-section*, além de ser possível avaliar a evolução, para um dado indivíduo, das variáveis em estudo ao longo do tempo.

Por outro lado, ainda de acordo com Marques (2000), os dados em painel providenciam uma maior quantidade de informação, maior variabilidade dos dados, menor colinearidade entre as variáveis, maior número de graus de liberdade e maior eficiência na estimação. A inclusão da dimensão em *cross-section*, num estudo temporal, confere uma maior variabilidade aos dados, na medida em que a utilização de dados agregados resulta em séries mais suaves do que as séries individuais que lhes servem de base. Esse aumento na variabilidade dos dados contribui para a redução de uma eventual colinearidade existente entre variáveis.

Seguindo Lima (2007), desde Balestra e Nerlove (1966), vários estudos têm sido publicados com a utilização de dados em painel, merecendo destaque os trabalhos de Anderson e Hsiao (1982), Bhargava, Franzini e Narendranathan (1982), Bhargava e Sargan (1983), Baltagi e Griffin (1984), Holtz-Eakin, Newey e Rosen (1988), Breusch, Mizon e Schmidt (1989), Maddala (1993), Islam (1995), Ahn e Schmidt (1997), Krishnakumar e Ronchetti (2000), Hsiao (2003), Arellano (2003), Frees (2004), Baltagi (2008), Mátyás e Sevestre (2008), e Wooldridge (2010).

Marques (2000) faz uma importante contribuição, uma vez que, além de apresentar os principais conceitos referentes aos modelos de dados em painel, elabora uma comparação entre os diferentes estimadores utilizados por diversos autores em estudos com dados micro e macroeconômicos.

Em contabilidade e finanças, por outro lado, ainda é incipiente a utilização de modelos que levem em consideração os aspectos longitudinais dos fenômenos contábeis e financeiros. No Brasil, merecem destaque os trabalhos de Pimentel (2006), Souza (2006), Lima, Lima, Fávero e Galdi (2007), Coelho (2007), Nakamura, Martin, Forte, Carvalho Filho, Costa e Amaral (2007), Carneiro e Sherris (2008), Lima (2009), Bastos e Nakamura (2009), Almeida (2010), Malacrida, Lima, Fávero e Lima (2010) e Fávero e Sotelino (2011).

No entanto verifica-se ainda, nessa área, uma falta de cuidado quanto aos critérios para a adoção de um modelo em detrimento de outro, bem como a ausência de uma discussão mais detalhada sobre os possíveis estimadores a serem estudados em cada situação. Em outras palavras, a utilização de dados em painel em contabilidade e finanças é, por vezes, elaborada sem que haja uma preocupação mais profunda com a escolha do melhor modelo a ser utilizado, ou seja, pouco tem sido discutido sobre a adequação do uso da técnica e sobre a definição dos melhores estimadores. Nesse sentido, merecem destaque os recentes trabalhos de Pimentel (2009) e de Jones, Kalmi e Mäkinen (2010).

Este artigo tem como objetivo apresentar, de forma conceitual e estruturada, os principais estimadores de dados em painel que podem ser utilizados em contabilidade e finanças, bem como auxiliar na definição do modelo mais consistente a ser adotado, em função das características dos dados. Ademais, tem por propósito aplicar tais modelos em um caso real, com base nos dados da *Compustat Global*. Ao final, são apresentadas as principais rotinas para a aplicação de cada um dos modelos em Stata, uma vez que se entende que tais procedimentos podem propiciar uma melhor relação entre teoria e prática, além de facilitar a implementação dos modelos em pesquisas futuras.

Dessa forma, o presente estudo não tem por intuito sugerir a aplicação de dados em painel em uma determinada situação, visto que isso depende, fundamentalmente, da questão de pesquisa e dos dados à disposição do pesquisador. O intuito, caso seja esta a técnica a ser utilizada, é de auxiliar na aplicação correta, com vistas à determinação de modelos mais adequados à realidade e voltados à tomada de decisão.

A Seção 1 traz uma revisão conceitual dos principais estimadores de dados em painel e faz uma distinção entre os modelos em painel curto (com um número de indivíduos maior do que o período de análise) e em painel longo (com o número de períodos maior do que o número de indivíduos em estudo). A Seção 2 apresenta uma aplicação dos principais modelos apresentados e discute o procedimento de definição do melhor modelo por meio da análise dos resultados obtidos. Por fim, a Seção 3 destina-se às considerações finais.

## 2 MODELOS DE DADOS EM PAINEL

Existem muitos modelos diferentes que podem ser utilizados para dados em painel. A distinção básica entre eles, segundo Greene (2007), é a existência de efeitos fixos ou aleatórios. O termo “efeitos fixos” dá uma ideia equivocada da modelagem, pois para ambos os casos, os efeitos no nível do indivíduo (firmas, entidades governamentais ou países, por exemplo) são aleatórios. Assim, segundo Cameron & Trivedi (2009), os modelos de efeitos fixos apresentam a complicação adicional de que os regressores sejam correlacionados com os efeitos do nível do indivíduo e, portanto, uma estimação consistente dos parâmetros do modelo requer uma eliminação ou controle dos efeitos fixos. Assim, um modelo que leva em conta os efeitos específicos do indivíduo para uma variável dependente  $y_{it}$  especifica que:

$$y_{it} = \beta_{0i} + x'_{it} \beta_1 + \varepsilon_{it} \quad (1)$$

Em que  $x_{it}$  são regressores,  $\beta_{0i}$  são os efeitos aleatórios específicos de indivíduo e  $\varepsilon_{it}$  representa o erro idiosincrático.

Com o termo do erro sendo  $\mu_{it} = \beta_{0i} + \varepsilon_{it}$  and  $x'_{it}$  correlacionado com o termo de erro invariante no tempo ( $\beta_{0i}$ ), assume-se que  $x'_{it}$  não é correlacionado com o erro idiosincrático  $\varepsilon_{it}$ . O modelo de efeitos fixos implica que  $E(y_{it}|\beta_{0i}, x_{it}) = \beta_{0i} + x'_{it}\beta_1$ , assumindo-se que  $E(y_{it}|\beta_{0i}, x_{it}) = 0$ , de modo que  $\beta_j = \partial E(y_{it}|\beta_{0i}, x_{it})/\partial x_{j,it}$ . A vantagem do modelo de efeitos fixos é que pode ser obtido um estimador consistente do efeito marginal do  $j$ -ésimo regressor de  $E(y_{it}|\beta_{0i}, x_{it})$ , dado que  $x_{j,it}$  varia no tempo.

No modelo de efeitos aleatórios, por outro lado, assume-se que  $\beta_{0i}$  é puramente aleatório, ou seja, não é correlacionado com os regressores. A estimação, portanto, é elaborada com um estimador FGLS (*feasible generalized least squares*). A vantagem do modelo de efeitos aleatórios é que este estima todos os coeficientes, mesmo dos regressores invariantes no tempo, e, portanto, os efeitos marginais. Ademais,  $E(y_{it}|x_{it})$  pode ser estimado. Porém a grande desvantagem é que estes estimadores são inconsistentes se o modelo de efeitos fixos for mais apropriado.

Conforme já discutido, a variável dependente e os regressores podem potencialmente variar simultaneamente ao longo do tempo e entre indivíduos. Enquanto a variação, ao longo do tempo ou para um dado indivíduo, é conhecida por *within variance*, a variação entre indivíduos é chamada de *between variance*. De acordo com Wooldridge (2010), no modelo de efeitos fixos, o coeficiente de um regressor com baixa variação *within* será imprecisamente

estimado e não será identificado se não houver qualquer *within variance*. Assim, é de fundamental importância a distinção entre essas variações para a definição do melhor modelo de dados em painel.

A variação total das observações de um regressor  $x$  em torno da média geral  $\bar{x} = 1/\sum_i T_i \sum_t x_{it}$  no conjunto de dados pode ser decomposta na soma da variação *within* ao longo do tempo para cada indivíduo em torno de  $\bar{x}_i = 1/T \sum_t x_{it}$  e na variação *between* entre indivíduos (para  $\bar{x}_i$  em torno de  $\bar{x}$ ). De acordo com Cameron & Trivedi (2009):

$$\text{Variância Within: } s_{xw}^2 = \frac{1}{\sum_i T_i - 1} \sum_i \sum_t (x_{it} - \bar{x}_i + \bar{x})^2$$

$$\text{Variância Between: } s_{xB}^2 = \frac{1}{N-1} \sum_i (\bar{x}_i - \bar{x})^2$$

$$\text{Variância Geral: } s_{xO}^2 = \frac{1}{\sum_i T_i - 1} \sum_i \sum_t (x_{it} - \bar{x})^2$$

As notações  $N$  e  $\sum_i T_i$  correspondem, respectivamente, ao número de indivíduos e ao número total de observações ao longo do tempo. Quando da apresentação da aplicação de dados em painel neste artigo, serão apresentadas e discutidas as variâncias de cada um dos regressores.

Ainda de acordo com Cameron & Trivedi (2009), os estimadores dos parâmetros  $\beta_1$  de um modelo de efeitos fixos para a equação (1) eliminam os efeitos fixos  $\beta_{0i}$ , ou seja, é elaborada uma transformação *within* pela diferenciação de médias. Dessa forma, uma estimação *within* elabora uma modelagem com os dados diferenciados em torno da média, e não se pode estimar um coeficiente de um regressor sem variação ao longo do tempo. Assim, os efeitos fixos  $\beta_{0i}$  na equação (1) podem ser eliminados pela subtração das médias de cada indivíduo  $\bar{y}_i = \bar{x}_i' \beta_1 + \bar{\varepsilon}_i$  no modelo correspondente, resultando o modelo *within*, ou modelo de diferenças de média:

$$(y_{it} - \bar{y}_i) = (x_{it} - \bar{x}_i)' \beta_1 + (\varepsilon_{it} - \bar{\varepsilon}_i) \quad (2)$$

Em que  $\bar{x}_i = T_i^{-1} \sum_{t=1}^{T_i} x_{it}$  e o estimador *within* é o estimador OLS (*ordinary least squares*) desse modelo. De acordo com Cameron & Trivedi (2009), pelo fato de  $\beta_{0i}$  ter sido

eliminado, o estimador OLS oferece estimativas consistentes de  $\beta_1$ , mesmo se  $\beta_{0i}$  for correlacionado com  $x_{it}$ , como é o caso do modelo de efeitos fixos.

O estimador *between* utiliza somente a variação entre indivíduos (*cross-sections*) e é o estimador OLS de uma regressão de  $\bar{y}_i$  em função de  $\bar{x}_i$ , apresentada a seguir (equação (3)). Por levar em consideração apenas as variações *cross-section* nos dados, o coeficiente de qualquer regressor que seja invariante entre indivíduos não pode ser identificado.

$$\bar{y}_i = \beta_0 + \bar{x}_i' \beta_1 + (\beta_{0i} - \beta_0 + \bar{\varepsilon}_i) \quad (3)$$

A consistência desse estimador requer que o termo do erro  $(\beta_{0i} - \beta_0 + \bar{\varepsilon}_i)$  não seja correlacionado com  $x_{it}$ , o que ocorre quando  $\beta_{0i}$  for um efeito aleatório, mas não quando for um efeito fixo. Segundo Hsiao (2003), esse estimador é raramente utilizado, pois os estimadores de efeitos aleatórios acabam sendo mais consistentes.

O estimador de efeitos aleatórios, por outro lado, é um estimador FGLS na equação (1). Assim, o modelo de efeitos aleatórios é o modelo de efeitos individuais:

$$y_{it} = x'_{it} \beta_1 + (\beta_{0i} + \varepsilon_{it}) \quad (4)$$

Com  $\beta_{0i} \sim (\beta_0, \sigma_\alpha^2)$  e  $\varepsilon_{it} \sim (0, \sigma_\varepsilon^2)$ . Desta forma, o termo de erro  $\mu_{it} = \beta_{0i} + \varepsilon_{it}$  is é correlacionado ao longo do tempo  $t$ , para uma dada observação  $i$ , com correlação:

$$\text{corr}(\mu_{it}, \mu_{is}) = \sigma_\alpha^2 / (\sigma_\alpha^2 + \sigma_\varepsilon^2), \text{ para todo } s \neq t \quad (5)$$

O estimador de efeitos aleatórios é o estimador FGLS de  $\beta$  da equação (4), dadas as correlações dos erros da equação (5).

Segundo Cameron & Trivedi (2009), em modelos com erros heterocedásticos e autocorrelacionados, o estimador GLS (*generalized least squares*) pode ser calculado como um estimador OLS, em um modelo que tenha erros não correlacionados homocedásticos, obtido de (4) por uma transformação linear apropriada. No caso do modelo de efeitos aleatórios da equação (4), tal modelo transformado é dado pela equação (6).

$$(y_{it} - \theta_i \bar{y}_i) = (1 - \theta_i) \beta_0 + (x_{it} - \theta_i \bar{x}_i)' \beta_1 + \{(1 - \theta_i) \beta_{0i} + (\varepsilon_{it} - \theta_i \bar{\varepsilon}_i)\} \quad (6)$$

Um estimador FGLS é obtido substituindo-se  $\theta_i$ , que é dado por uma estimativa consistente indicada por:

$$\theta_i = 1 - \sqrt{\sigma_\varepsilon^2 / (T_1 \sigma_u^2 + \sigma_\varepsilon^2)} \quad (7)$$

O estimador de efeitos aleatórios será consistente e completamente eficiente se o modelo de efeitos aleatórios for apropriado, porém será inconsistente se o modelo de efeitos fixos for apropriado, uma vez que a correlação entre  $x_{it}$  e  $\beta_{0i}$  resulta numa correlação entre os regressores e o termo do erro na equação (6). Da mesma forma, ainda segundo Cameron & Trivedi (2009), se não ocorrerem efeitos fixos, então o estimador de efeitos aleatórios será consistente, mas ineficiente e, portanto, uma estimação com erros-padrão robustos clusterizados deveria ser obtida.

A expressão da estimativa por mínimos quadrados generalizados factíveis de um coeficiente de regressão do modelo (1), supondo-se efeitos aleatórios, torna-se igual à da estimativa do mesmo coeficiente em um modelo de efeitos fixos (estimação *within*) se  $\hat{\theta}_i = 1$ .

## 2.1 Painel Curto

Se não ocorrerem efeitos fixos mas os erros demonstrarem correlação dentro do painel, então o estimador de efeitos aleatórios será consistente mas ineficiente e, portanto, uma estimação com erros-padrão robustos clusterizados deverá ser obtida. Dessa forma, para um painel curto, em que  $T < N$ , uma estimação com erros-padrão robustos clusterizados pode ser obtida considerando-se a premissa de que os erros são independentes entre indivíduos e que  $N \rightarrow \infty$ , ou seja, que  $(\varepsilon_{it}, \varepsilon_{js}) = 0$  para  $i \neq j$ , que  $E(\varepsilon_{it}, \varepsilon_{is})$  não seja restrita e que  $\varepsilon_{it}$  seja heterocedástico.

De acordo com Cameron e Trivedi (2009), o passo inicial para a aplicação de um modelo com dados em painel é a aplicação de um modelo POLS (*pooled ordinary least squares*), que assume que os regressores sejam exógenos e que o termo de erro seja  $\mu_{it}$ , em vez da decomposição  $\alpha_i + \varepsilon_{it}$ . Portanto:

$$y_{it} = \beta_0 + x'_{it} \beta_1 + \mu_{it} \quad (8)$$

Os parâmetros desse modelo são estimados por meio de OLS, mas a inferência requer que haja controle da correlação *within* do erro  $\mu_{it}$  para um dado indivíduo, a ser elaborado utilizando-se erros-padrão robustos com agrupamento no nível do indivíduo.

## 2.2 Painel Longo

Para dados em painel longo, ou seja, com muitos períodos para um número relativamente menor de indivíduos, os efeitos individuais  $\beta_{0i}$  podem ser incorporados em  $x_{it}$  como variáveis *dummy* para cada período, de acordo com o seguinte modelo:

$$y_{it} = \beta_{0i} + \gamma_t + x'_{it}\beta_1 + \varepsilon_{it} \quad (9)$$

de modo a haver muitos efeitos de tempo  $y_t$  (efeitos mensais, trimestrais ou anuais, por exemplo). Um modelo *pooled*, para  $T > N$ , em que os regressores  $x_{it}$  contemplam o intercepto, o efeito temporal e, possivelmente, um vetor de variáveis de indivíduo, pode ser escrito como:

$$y_{it} = x'_{it}\beta_1 + \mu_{it} \quad (10)$$

Como  $T$  é maior do que  $N$ , passa a ser necessária a especificação de um modelo que considere a existência de correlação serial do erro (Beck e Katz, 1995). Dessa forma, para dados em painel longo, os modelos *pooled* com métodos de estimação OLS (POLS) e FGLS passam a ser mais adequados, já que permitem a utilização de um modelo AR(1) para  $\mu_{it}$  ao longo do tempo, em que  $\mu_{it}$  seja heterocedástico (Hoechle, 2007). Assim:

$$\mu_{it} = \rho_i\mu_{i,t-1} + \varepsilon_{it} \quad (11)$$

Em que os termos  $\varepsilon_{it}$  são não correlacionados no tempo, porém com correlação entre indivíduos diferente de zero ( $\text{corr}(\varepsilon_{it}, \varepsilon_{is}) = \sigma_{ts}$ ).

Alternativamente à inclusão de um vetor de variáveis *dummy* para cada período, estima-se, finalmente, um modelo de efeitos individuais com termos de erro AR(1), que representa um modelo melhor do que aquele que considera os termos de erro i.i.d. Assim:

$$y_{it} = \beta_{0i} + x'_{it}\beta_1 + \mu_{it} \quad (12)$$

Logo, segundo Cameron & Trivedi (2009), esse modelo, potencialmente, gerará estimativas dos parâmetros mais eficientes. Nesse caso, dada a estimativa de  $\hat{\rho}$  na equação (11), primeiramente, elimina-se o efeito do erro AR(1) e, na sequência, elimina-se o efeito

individual por meio da aplicação da diferença de médias. Assim, a modelagem pode considerar  $\beta_{0i}$  como um efeito fixo ou um efeito aleatório.

Após a apresentação dos modelos de dados em painel, explicita-se que este trabalho aplicará dez diferentes tipos de modelagens, a fim de propiciar um melhor entendimento dos diversos tipos de estimadores e das suas condições de uso, bem como apresentar modelos para o estudo do comportamento dos retornos dos preços de ações de empresas listadas em bolsas de valores de países da América Latina, em uma perspectiva longitudinal. O quadro 1 apresenta esses dez diferentes tipos de modelos. No apêndice, encontram-se as rotinas para a aplicação de cada um destes modelos no *software* Stata.

Modelo	Descrição
POLS com Erros-Padrão Robustos Clusterizados	$y_{it} = \beta_0 + x'_{it} \beta_1 + \mu_{it}$ Estimação OLS com controle da correlação <i>within</i> do erro $\mu_{it}$ ao longo do tempo.
Modelo com Estimador <i>Between</i>	$\bar{y}_i = \beta_0 + x_i^r \beta_1 + (\beta_{0i} - \beta_0 + \varepsilon_i^-)$ O estimador <i>between</i> somente utiliza a variação das <i>cross-sections</i> e é o estimador OLS de uma regressão de $\bar{y}_i$ em função de $x_i^-$ . A consistência desse estimador requer que o termo de erro $(\beta_{0i} - \beta_0 + \varepsilon_i^-)$ não seja correlacionado com $x_{it}$ .
Efeitos Fixos	$y_{it} = \beta_{0i} + x'_{it} \beta_1 + \varepsilon_{it}$ Os parâmetros $\beta_{0i}$ podem ser correlacionados com os regressores $x_{it}$ , o que permite uma forma limitada de endogeneidade. Assume-se que $x_{it}$ não é correlacionado com o erro idiossincrático $\varepsilon_{it}$ .
Efeitos Fixos com Erros-Padrão Robustos Clusterizados	$y_{it} = \beta_{0i} + x'_{it} \beta_1 + \varepsilon_{it}$ Os termos $\beta_{0i}$ podem ser correlacionados com os regressores $x_{it}$ , o que permite uma forma limitada de endogeneidade. Assume-se que os erros sejam independentes entre indivíduos e que $\varepsilon_{it}$ seja heterocedástico.

Efeitos Aleatórios	$y_{it} = x'_{it} \beta_1 + (\beta_{0i} + \varepsilon_{it})$ <p>Os parâmetros <math>\beta_{0i}</math> e os termos de erro idiossincrático <math>\varepsilon_{it}</math> são independentes e identicamente distribuídos (i.i.d.). O estimador de efeitos aleatórios é o FGLS de <math>\beta_1</math>, dado que <math>\text{corr}(\mu_{it}, \mu_{is}) = \sigma_{\alpha}^2 / (\sigma_{\alpha}^2 + \sigma_{\varepsilon}^2)</math>.</p>
Efeitos Aleatórios com Erros-Padrão Robustos Clusterizados	$y_{it} = x'_{it} \beta_1 + (\beta_{0i} + \varepsilon_{it})$ <p>Se não houver efeitos fixos, mas os erros apresentarem correlação <i>within</i>, o estimador de efeitos aleatórios é consistente, porém ineficiente. Portanto, erros-padrão robustos clusterizados precisam ser obtidos.</p>
Efeitos Fixos com Termos de Erro AR(1)	$y_{it} = \beta_{0i} + x'_{it} \beta_1 + \mu_{it}$ <p>Com <math>\mu_{it} = \rho_i \mu_{i,t-1} + \varepsilon_{it}</math>. Considera-se <math>\beta_{0i}</math> como sendo um efeito fixo.</p>
Efeitos Aleatórios com Termos de Erro AR(1)	$y_{it} = \beta_{0i} + x'_{it} \beta_1 + \mu_{it}$ <p>Com <math>\mu_{it} = \rho_i \mu_{i,t-1} + \varepsilon_{it}</math>. Considera-se <math>\beta_{0i}</math> como sendo um efeito aleatório.</p>
<i>Pooled</i> com Método de Estimação OLS e Termos de Erro AR(1)	$y_{it} = \beta_{0i} + \gamma_t + x'_{it} \beta_1 + \varepsilon_{it}$ <p>Com <math>\mu_{it} = \rho_i \mu_{i,t-1} + \varepsilon_{it}</math>, em que os <math>\varepsilon_{it}</math> são serialmente não correlacionados, mas com correlação entre indivíduos igual a <math>\text{corr}(\varepsilon_{it}, \varepsilon_{is}) = \sigma_{\alpha}^2 \neq 0</math></p>
<i>Pooled</i> com Método de Estimação FGLS e Termos de Erro AR(1)	$y_{it} = \beta_{0i} + \gamma_t + x'_{it} \beta_1 + \varepsilon_{it}$ <p>Similar ao modelo <i>pooled</i> com método de estimação OLS, mas com estimador FGLS.</p>

Quadro 1: Modelos de Dados em Painel a Serem Estimados

### 3 UMA APLICAÇÃO

Após uma discussão acerca dos principais estimadores de dados em painel, apresenta-se uma aplicação em contabilidade financeira.

Como muitos dos dados contábeis e financeiros apresentam periodicidade de divulgação mensal, trimestral ou anual, é comum que os estudos nessas áreas utilizem modelos de dados em painel curto, já que o número de indivíduos (empresas, por exemplo), ultrapassa o número de períodos de divulgação dos dados. Por outro lado, nada impede que o pesquisador baseie seu estudo numa amostra de empresas de apenas um determinado setor, ou utilize dados com frequência de divulgação maior (diária, por exemplo), o que poderia ocasionar uma modelagem com dados em painel longo. De qualquer maneira, é fundamental que a identificação dessa característica da base de dados seja feita de forma anterior à modelagem propriamente dita. No presente artigo, serão utilizadas duas bases de dados, sendo a primeira em painel curto e a segunda em painel longo.

Inicialmente, uma base da *Compustat Global* contendo dados sobre a rentabilidade de ações de 473 companhias provenientes de 7 países da América Latina (Argentina, Brasil, Chile, Colômbia, México, Peru e Venezuela), ao longo de 118 meses (1998-2007), totalizando 28.257 observações, será utilizada para o estudo de um painel curto. Na sequência, um estrato da base anterior, sorteado aleatoriamente, será utilizado, com dados de apenas 40 empresas, também no período de 118 meses, totalizando 4.720 observações, com vistas ao estudo do painel longo.

No período analisado, muitas companhias apresentaram taxas de crescimento expressivas nos preços das ações em um ou mais meses. Foram considerados, portanto, apenas retornos mensais inferiores a 100%. Por meio do gráfico 1, é possível verificar que esses retornos apresentam comportamentos similares ao longo do tempo, ainda que existam diferenças nas médias e nas inclinações entre séries individuais.

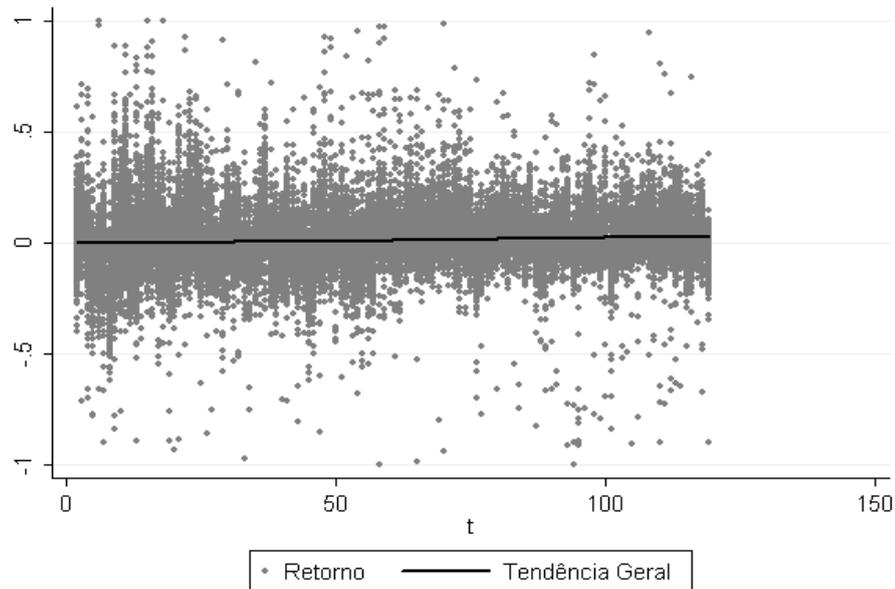


Gráfico 1: Evolução dos Retornos Mensais dos Preços das Ações na América Latina

Cada ponto no gráfico 1 representa um par retorno do preço da ação-mês. Esse comportamento sugere a elaboração de modelos longitudinais, já que os regressores podem variar entre companhias e ao longo do tempo, conforme será apresentado e discutido adiante. Enquanto o gráfico 2 apresenta a variação dos retornos mensais dos preços das ações ao longo do tempo para cada companhia, ou seja, mostra os desvios do retorno em relação à média individual de cada companhia (*within variation*), o gráfico 3 apresenta a variação dos retornos mensais entre as empresas, ou seja, mostra os desvios dos retornos mensais dos preços das ações das empresas em relação à média geral para cada instante de tempo (*between variation*).

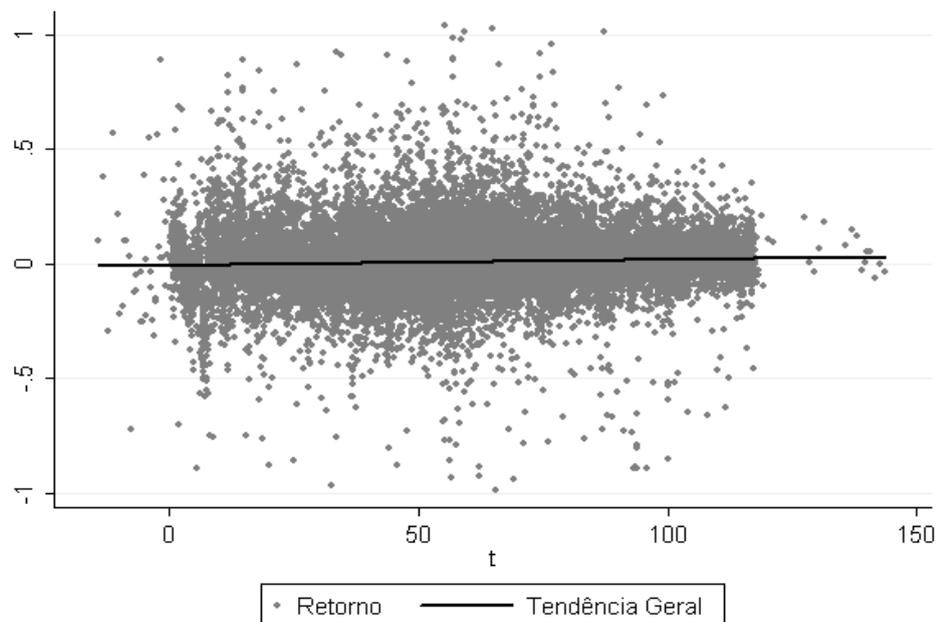


Gráfico 2: Desvios dos Retornos Mensais em Relação à Média de Cada Empresa ao Longo do Tempo (*Within Variation*)

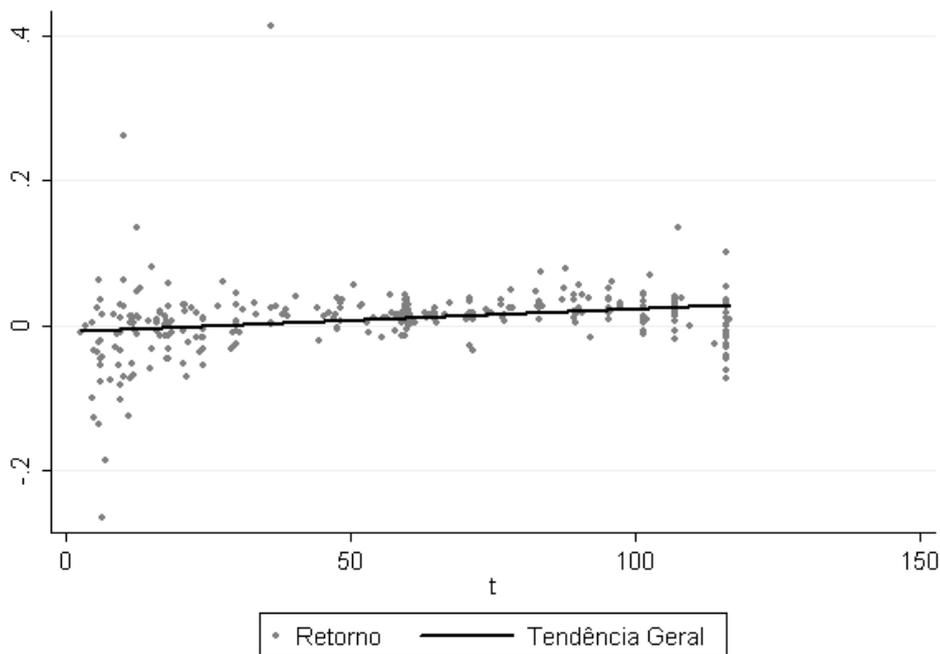


Gráfico 3: Desvios dos Retornos Mensais em Relação à Média Geral para Cada Instante de Tempo (*Between Variation*)

Nessa aplicação, a utilização da base da *Compustat Global* é utilizada com o intuito de verificar se a razão preço-fluxo de caixa é mais significativa do que a razão preço-lucro por ação para influenciar os retornos mensais dos preços das ações de companhias na América Latina ao longo do tempo. De acordo com Kennon (2010), como alguns investidores preferem utilizar o fluxo de caixa a fazer uso do lucro por ação para a avaliação dos preços das ações, uma vez que argumentam que, enquanto o primeiro não é facilmente manipulado, o mesmo não pode ser dito em relação ao segundo, essa aplicação propicia uma investigação sobre o assunto, sob uma perspectiva longitudinal e com o uso de diversos estimadores.

Como discutido, 10 diferentes modelos de dados em painel serão elaborados, com diferentes considerações sobre os estimadores e os termos de erro. O modelo geral é dado por:

$$\text{retorno}_{it} = \beta_{0i} + \beta_1 \cdot (\text{pcf})_{it} + \beta_2 \cdot (\text{pe})_{it} + \varepsilon_{it} \quad (13)$$

Em que  $\beta_1$  e  $\beta_2$  representam as mudanças no retorno do preços da ação quando uma unidade da razão preço-fluxo de caixa (pcf) ou da razão preço-lucro por ação (pe) ocorre, respectivamente, *ceteris paribus*.

A seguir, serão discutidos os resultados das modelagens, tanto para um painel curto, quanto para um painel longo.

### 3.1 Modelos para Dados em Painel Curto

Como a amostra, neste caso, oferece dados de 473 empresas em 118 meses, o painel pode ser considerado curto ( $T < N$ ).

A tabela 1 apresenta a decomposição de variância para cada um dos regressores.

**Tabela 1: Estatísticas do Painel Curto e Decomposição de Variância**

Variável	Decomposição	Média	Desvio-Padrão	Mínimo	Máximo	Observações
id (empresa)	geral					N.T = 28.257 N = 473
	<i>between</i>					
	<i>within</i>		0,000			
t (mês)	geral		35,492	1,000	118,000	N.T = 28.257 N = 473
	<i>between</i>		33,315	1,500	118,000	
	<i>within</i>		26,489	-5,398	145,188	
var_y (retorno)	geral		0,141	-0,999	1,000	N.T = 28.257 N = 473
	<i>between</i>	0,013	0,036	-0,134	0,414	
	<i>within</i>		0,139	-0,986	1,027	
var_x1 (pcf)	geral		246,040	-19.841,38	10.594,7	N.T = 20.224 N = 361
	<i>between</i>	7,246	212,442	-2.168,16	3.363,81	
	<i>within</i>		221,678	-17.665,97	7.238,13	
var_x2 (pe)	geral		212,550	-5.790,53	12.251,82	N.T = 27.816 N = 463
	<i>between</i>	12,965	228,033	-2.872,58	3.889,48	
	<i>within</i>		176,068	-4.521,99	8.375,30	

De acordo com a tabela 1, nota-se que a ação é invariante no tempo e, portanto, apresenta variação *within* igual a zero. Por outro lado, a variável referente ao tempo (mês) não é invariante entre empresas, já que se trata de um painel desbalanceado e, portanto, a sua variação *between*, mesmo sendo menor do que a *within*, não é igual a zero. Das demais variáveis, apenas a *pcf* apresenta maior variação entre indivíduos (*between*) do que ao longo do tempo (*within*), porém ainda não é possível afirmar que a estimação *within* resultará numa perda de eficiência, já que a proporção entre as variâncias *within* e *between* de cada variável é diferente e ainda não se conhecem as significâncias estatísticas de cada um delas nos modelos. A tabela 1, todavia, oferece um maior embasamento para a adoção dos modelos de dados em painel e a aplicação de diversos estimadores. As colunas “Mínimo” e “Máximo” apresentam,

respectivamente, os valores mínimos e máximos de  $x_{it}$  para a linha “geral”,  $\bar{x}_i$  para a linha “between” e  $(x_{it} - \bar{x}_i + \bar{x})$  para a linha “within”.

A tabela 2 apresenta os resultados dos modelos considerando 6 diferentes estimadores.

**Tabela 2: Modelos de Dados em Painel Curto**

Variável	POLS com Erros-Padrão Robustos Clusterizados	Estimador <i>Between</i>	Efeitos Fixos	Efeitos Fixos com Erros-Padrão Robustos Clusterizados	Efeitos Aleatórios	Efeitos Aleatórios com Erros-Padrão Robustos Clusterizados
pcf	1,52x10 <sup>-5*</sup> (5,93x10 <sup>-6</sup> )	1,24x10 <sup>-5</sup> (1,32x10 <sup>-5</sup> )	2,15x10 <sup>-5*</sup> (4,96x10 <sup>-6</sup> )	2,15x10 <sup>-5*</sup> (5,74x10 <sup>-6</sup> )	1,68x10 <sup>-5*</sup> (4,62x10 <sup>-6</sup> )	1,68x10 <sup>-5*</sup> (5,19x10 <sup>-6</sup> )
pe	-1,09x10 <sup>-5</sup> (6,19x10 <sup>-6</sup> )	-5,59x10 <sup>-7</sup> (1,1x10 <sup>-5</sup> )	-1,28x10 <sup>-5</sup> (6,05x10 <sup>-6</sup> )	-1,28x10 <sup>-5</sup> (5,16x10 <sup>-6</sup> )	-1,16x10 <sup>-5</sup> (5,13x10 <sup>-6</sup> )	-1,16x10 <sup>-5</sup> (4,54x10 <sup>-6</sup> )
constante	0,013* (1,04x10 <sup>-3</sup> )	0,009* (2,11x10 <sup>-3</sup> )	0,013* (1,03x10 <sup>-3</sup> )	0,013* (5,23x10 <sup>-5</sup> )	0,011* (1,59x10 <sup>-3</sup> )	0,011* (1,17x10 <sup>-3</sup> )
N.T	20.224	20.224	20.224	20.224	20.224	20.224
R <sup>2</sup>	6,0x10 <sup>-4</sup>					
R <sup>2</sup> (geral)		4,0x10 <sup>-4</sup>	6,0x10 <sup>-4</sup>	6,0x10 <sup>-4</sup>	6,0x10 <sup>-4</sup>	6,0x10 <sup>-4</sup>
R <sup>2</sup> (between)		3,1x10 <sup>-4</sup>	1,0x10 <sup>-3</sup>	1,0x10 <sup>-3</sup>	1,0x10 <sup>-3</sup>	1,0x10 <sup>-3</sup>
R <sup>2</sup> (within)		8,0x10 <sup>-4</sup>	2,1x10 <sup>-3</sup>	2,1x10 <sup>-3</sup>	1,5x10 <sup>-3</sup>	1,5x10 <sup>-3</sup>
F	5,12	0,73	9,57	7,10		
sig. F	0,006	0,483	0,000	0,001		
Wald $\chi^2$					14,06	10,83
sig. $\chi^2$					0,001	0,004

Obs.: Erros-Padrão entre parênteses.

\*sig. < 0,05.

Como pode ser observado, os coeficientes estimados variam de modelo para modelo, o que reflete a existência de resultados diferentes se as variações *within* ou *between* forem utilizadas.

Primeiramente, verifica-se, em relação à adequação dos modelos, que o vetor de regressores apresenta significância estatística em todos os casos, à exceção do modelo com estimador *between* (sig. F para os modelos POLS, *between* e com efeitos fixos e sig. Wald  $\chi^2$

para os modelos com efeitos aleatórios). Como as estatísticas  $R^2$  são consideravelmente baixas, esses modelos não são relativamente adequados para propósitos de previsão. Um resultado importante, porém, relaciona-se com a existência de maiores valores para os  $R^2$  *within* em todos os modelos em que essa estatística é calculada. Tendo-se por base as expressões de cada uma das estatísticas  $R^2$

$$R^2 \text{ geral: } \rho^2\{(y_{it} - \bar{y}_i), (x'_{it} \hat{\beta} - \bar{x}'_i \hat{\beta})\}$$

$$R^2 \text{ between: } \rho^2(\bar{y}_i, \bar{x}'_i \hat{\beta})$$

$$R^2 \text{ within: } \rho^2(y_{it}, x'_{it} \hat{\beta})$$

Em que  $\rho^2(x,y)$  representa correlação quadrática entre  $x$  e  $y$ . Nota-se que os estimadores *within* explicam melhor a variação *within* em todos os modelos, mesmo aqueles com efeitos aleatórios.

Nota-se ainda que a variável *pe* não é estatisticamente significativa (sig. > 0,05) nos modelos apresentados, na presença da variável *pcf*. Esta última, com exceção do modelo com estimador *between*, é significativa para explicar o comportamento dos retornos dos preços das ações (sig. < 0,05), confirmando o argumento de alguns analistas em favor do uso dessa variável.

Com relação à variável *pcf*, verifica-se que os erros-padrão nos modelos de efeitos fixos e efeitos aleatórios com erros-padrão robustos clusterizados são maiores do que os respectivos modelos sem esta consideração. Os regressores estimados nos modelos POLS e *between* oferecem erros-padrão ainda maiores, mesmo com a variável *pcf* sendo estatisticamente significativa (sig. < 0,05) no modelo POLS.

O Teste LM de Breusch-Pagan, aplicado após a modelagem de efeitos aleatórios, auxilia na rejeição da hipótese nula de que há adequação do modelo POLS em relação ao modelo de efeitos aleatórios, já que  $\chi^2 = 70,7$  (sig.  $\chi^2 = 0,000$ ). Na sequência, por meio do teste F de Chow, a hipótese nula de que há igualdade de interceptos e inclinações para todas as companhias (POLS) é rejeitada. Portanto, esses parâmetros diferem daqueles obtidos por meio do modelo de efeitos fixos, já que  $F = 2,34$  (sig.  $F = 0,000$ ). Finalmente, o teste de Hausman para efeitos fixos auxilia na rejeição da hipótese nula de que o modelo de efeitos aleatórios oferece estimativas dos parâmetros mais consistentes, já que, para este caso,  $\chi^2 = 17,07$  (sig.  $\chi^2 = 0,000$ ).

Segundo Islam (1995), a principal utilidade da modelagem de dados em painel é a sua habilidade em permitir que diferenças ocorram entre países, o que faz com que os resultados sejam significativamente diferentes daqueles obtidos por meio de regressões isoladas para cada país. Na tabela 3, os coeficientes de regressão são apresentados para cada um dos países da amostra.

**Tabela 3: Coeficientes por País**

País	pcf	constante
Argentina	$-1,97 \times 10^{-5}$	0,006
Brasil	$-1,63 \times 10^{-5}$	0,018
Chile	$2,08 \times 10^{-5}$	0,007
Colômbia	$9,07 \times 10^{-6}$	0,015
México	$-2,66 \times 10^{-5}$	0,007
Peru	$8,75 \times 10^{-5}$	0,016
Venezuela	$1,32 \times 10^{-4}$	0,010

Obs.: Variável Dependente: Retorno Mensal do Preço da Ação.

Embora a razão preço-fluxo de caixa seja mais significativa para explicar os retornos dos preços de ações nos países da América Latina, a tabela 3 revela a existência de diferentes influências. Os diferentes coeficientes e sinais da variável *pcf* e da constante expressam a importância de se considerar a modelagem de dados em painel e propiciam a formulação de pesquisas futuras sobre as razões econômicas pelas quais os países apresentam diferentes comportamentos nos preços das ações de suas companhias ao longo do tempo.

### 3.2 Modelos para Dados em Painel Longo

Para este caso, como a amostra oferece dados de 40 empresas ao longo de 118 meses, o painel pode ser considerado longo ( $T > N$ ). Dessa forma, como a influência temporal é bastante importante em séries longas, modelos de efeitos fixos e aleatórios serão também aplicados com a consideração de componentes autorregressivos (AR(1)) para os resíduos, o que pode resultar em parâmetros com estimativas mais eficientes para painéis longos.

Assim como elaborado para o painel curto, a tabela 4 apresenta a decomposição de variância para cada um dos regressores do painel longo.

**Tabela 4: Estatísticas do Painel Longo e Decomposição de Variância**

Variável	Decomposição	Média	Desvio-Padrão	Mínimo	Máximo	Observações
id (empresa)	geral					N.T = 4.720 N = 40
	<i>between</i>					
	<i>within</i>		0,000			
t (mês)	geral		34,066	1,000	118,000	N.T = 4.720 N = 40
	<i>between</i>		0,000	60,500	60,500	
	<i>within</i>		34,066	1,000	118,000	
var_y (retorno)	geral		0,144	-0,662	0,899	N.T = 4.720 N = 40
	<i>between</i>	0,012	0,005	0,005	0,022	
	<i>within</i>		0,144	-0,656	0,896	
var_x1 (pcf)	geral		22,261	-264,68	107,37	N.T = 4.720 N = 40
	<i>between</i>	5,312	5,106	-3,262	10,906	
	<i>within</i>		21,682	-270,274	101,776	
var_x2 (pe)	geral		66,362	-536,13	735,86	N.T = 4.720 N = 40
	<i>between</i>	14,488	12,089	1,365	40,972	
	<i>within</i>		65,279	-537,894	734,095	

Aqui, propositadamente, foram escolhidas 9 empresas de modo que a variável referente ao tempo (mês) fosse invariante, ou seja, que o painel fosse balanceado e, portanto, que sua variação *between* fosse igual a zero. Todas as demais variáveis apresentaram menor variação entre indivíduos (*between*) do que ao longo do tempo (*within*), porém também não é possível afirmar que a estimação *between* resultará numa perda de eficiência.

Da mesma forma que o realizado para o painel curto, a tabela 5 apresenta os resultados dos modelos, considerando também 6 diferentes estimadores.

Tabela 5: Modelos de Dados em Painel Longo

Variável	Efeitos Fixos	Efeitos Aleatórios	Efeitos Fixos com Erros AR(1)	Efeitos Aleatórios com Erros AR(1)	Pooled OLS com Erros AR(1)	Pooled FGLS com Erros AR(1)
pcf	-5,68x10 <sup>-5</sup> (9,74x10 <sup>-5</sup> )	-7,55x10 <sup>-5</sup> (9,45x10 <sup>-5</sup> )	-5,77x10 <sup>-5</sup> (4,32x10 <sup>-5</sup> )	-6,49x10 <sup>-5</sup> (4,88x10 <sup>-5</sup> )	-7,13x10 <sup>-5</sup> (2,36x10 <sup>-4</sup> )	-1,49x10 <sup>-4</sup> (1,97x10 <sup>-4</sup> )
pe	-8,42x10 <sup>-5</sup> * (3,24x10 <sup>-5</sup> )	-9,11x10 <sup>-5</sup> * (3,17x10 <sup>-5</sup> )	-7,88x10 <sup>-5</sup> * (1,70x10 <sup>-5</sup> )	-8,58x10 <sup>-5</sup> * (2,32x10 <sup>-5</sup> )	-8,90x10 <sup>-5</sup> (6,90x10 <sup>-5</sup> )	-8,63x10 <sup>-5</sup> (6,07x10 <sup>-5</sup> )
constante	0,014* (2,23x10 <sup>-3</sup> )	0,014* (2,22x10 <sup>-3</sup> )	0,013* (2,20x10 <sup>-3</sup> )	0,014* (2,21x10 <sup>-3</sup> )	0,014 (8,22x10 <sup>-3</sup> )	0,013 (7,24x10 <sup>-3</sup> )
N.T	4.720	4.720	4.680	4.720	4.720	4.720
R <sup>2</sup>					1,8x10 <sup>-3</sup>	
R <sup>2</sup> (geral)	1,9x10 <sup>-3</sup>	1,9x10 <sup>-3</sup>	1,9x10 <sup>-3</sup>	1,9x10 <sup>-3</sup>		
R <sup>2</sup> (between)	0,494	0,504	0,462	0,499		
R <sup>2</sup> (within)	1,5x10 <sup>-3</sup>	1,5x10 <sup>-3</sup>	1,2x10 <sup>-3</sup>	1,5x10 <sup>-3</sup>		
F	3,59		2,88			
sig. F	0,027		0,050			
Wald $\chi^2$		9,00		7,87	1,72	2,57
sig. $\chi^2$		0,011		0,048	0,422	0,276

Obs.: Erros-Padrão entre parênteses.

\*sig. < 0,05.

De acordo com a tabela 5, é possível verificar que os coeficientes estimados também variam entre os modelos. Primeiramente, nota-se a existência de erros-padrão bem superiores nos modelos de efeitos fixos e aleatórios (mais de 100%) em comparação com aqueles reportados pelos respectivos modelos com efeitos AR(1) nos termos de erro. Esse fato talvez tenha ocorrido pela natureza do painel em análise, ou seja, pelo fato de ser longo.

Porém, mesmo permitindo-se que os termos de erro sejam correlacionados entre companhias, nota-se que não houve, nesse caso, uma redução dos erros-padrão dos modelos *pooled* com estimadores OLS e FGLS em comparação com aqueles obtidos, anteriormente, por meio dos modelos de efeitos fixos e aleatórios com termos de erro AR(1).

Em relação à adequação dos modelos propriamente ditos, nota-se a significância estatística do conjunto de variáveis nos casos em que foram considerados efeitos fixos ou

aleatórios, com ou sem termos de erro AR(1). Como apresentado quando da elaboração dos modelos para o painel curto, embora haja relativa importância das estatísticas  $R^2$  para efeitos de predição, seus valores não são consideravelmente elevados nos modelos em análise.

Os modelos de efeitos fixos e aleatórios oferecem uma alternativa, para dados em painel longo, em que são considerados os efeitos individuais com termos de erro AR(1), e representam melhores modelos do que aqueles que consideram os termos de erro i.i.d., o que poderá gerar estimativas dos parâmetros mais eficientes. De fato, os modelos de efeitos fixos e aleatórios com termos de erro AR(1) apresentam erros-padrão da ordem de 30% a 50% menores do que aqueles obtidos pelos respectivos modelos sem a consideração de termos de erro AR(1).

O teste de Hausman aplicado aos modelos de efeitos fixos e aleatórios com termos de erro AR(1) auxilia na rejeição da hipótese nula de que o modelo com efeitos aleatórios oferece estimativas consistentes dos parâmetros, já que, nesse caso,  $\chi^2 = 10,50$  (sig.  $\chi^2 = 0,005$ ).

Por fim, é importante mencionar que, nesse caso, os resultados parecem estar contrariando aqueles obtidos para o painel curto, ou seja, a variável *pcf* não é estatisticamente significativa (sig. > 0,05) na presença da variável *pe*, que se apresenta com sinal negativo. Porém, como as empresas consideradas neste caso são provenientes apenas de Argentina, Brasil e México, uma investigação mais detalhada sobre as razões econômicas subjacentes a esse fenômeno precisa ser elaborada. Como os sinais negativos dos parâmetros dos regressores estão condizentes com aqueles apresentados na tabela 3 para os referidos países, ressalta-se ainda mais a importância da aplicação correta dos modelos em painel para o estudo das diferenças existentes entre indivíduos e ao longo do tempo para um determinado fenômeno.

#### 4 CONCLUSÕES E RECOMENDAÇÕES

Modelos de dados em painel possibilitam que o pesquisador avalie a relação entre alguma variável de desempenho e diversas variáveis preditivas, permitindo que se elaborem inferências sobre as eventuais diferenças entre indivíduos e ao longo do tempo sobre a evolução daquilo que se pretende estudar. Dadas as suas características, é natural que muitas pesquisas em contabilidade e finanças venham a fazer uso de tais modelos, uma vez que muitos dados são publicados com determinada periodicidade para empresas, municípios, estados ou países.

Para tanto, é necessário, assim como para qualquer outra técnica econométrica, que a aplicação venha acompanhada de rigor metodológico e de certos cuidados quando da análise dos resultados, principalmente se estes tiverem como objetivo a elaboração de previsões. A adoção de determinado estimador, em detrimento de outro considerado viesado ou inconsistente, pode auxiliar o pesquisador na escolha do melhor modelo, valorizando a sua pesquisa e propiciando novos estudos sobre o tema escolhido.

No presente artigo, procurou-se elaborar seis diferentes modelos para um específico painel curto, e outros seis para um painel longo. Os dados, em ambos os casos, foram provenientes da *Compustat Global*. Segundo Makino, Isobe e Chan (2004), trabalhos prévios têm utilizado modelagens longitudinais para se estudar a variabilidade de variáveis de desempenho entre firmas ou países ao longo do tempo, com pesquisas empíricas desenvolvidas em diversas áreas do conhecimento. Porém não é comum que se encontrem estudos aplicados a mercados de ações, considerando diferenças entre economias emergentes.

A análise da contribuição da razão preço-fluxo de caixa e da razão preço-lucro por ação sobre os retornos mensais de ações de países emergentes possibilita que seja incrementada a discussão sobre como diferem os comportamentos dos mercados em países considerados emergentes. Porém essa abordagem foi adotada apenas como exemplo dentro de um objetivo maior, que foi o de apresentar como os diferentes estimadores podem gerar resultados inconsistentes, quando da elaboração de modelos de dados em painel, e o de auxiliar o pesquisador para a escolha do modelo mais adequado, tanto no caso de um painel curto, quanto no de um painel longo.

O presente artigo pretende contribuir apenas com uma parte das inúmeras pesquisas que podem surgir. Modelos cujas variáveis de desempenho estejam apresentadas na forma de *dummy*, com dados censurados ou com dados de contagem apresentam estimadores consistentemente diferentes e, conseqüentemente, rotinas específicas em *softwares* como o Stata. Tais modelos não foram discutidos neste artigo. Espera-se, portanto, que essa discussão seja inicial e ganhe importância em contabilidade e finanças, dada a vastidão de possibilidades de pesquisa.

## REFERÊNCIAS

AHN, S. C.; SCHMIDT, P. Efficient estimation of dynamic panel data models: alternative assumptions and simplified estimation. **Journal of Econometrics**, v. 76, n. 1-2, p. 309-321, 1997.

- ALMEIDA, J. E. F. **Qualidade da informação contábil em ambientes competitivos**. 2010. 174 f. Tese (Doutorado em Controladoria e Contabilidade) – Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo.
- ANDERSON, T. W.; HSIAO, C. Formulation and estimation of dynamic models using panel data. **Journal of Econometrics**, v. 18, n. 1, p. 47-82, 1982.
- ARELLANO, M. **Panel data econometrics: advanced texts in econometrics**. New York: Oxford University Press, 2003.
- BALESTRA, P.; NERLOVE, M. Pooling cross section and time series data in the estimation of a dynamic model: the demand for natural gas. **Econometrica**, v. 34, n. 3, p. 585-612, 1966.
- BALTAGI, B. H. **Econometric analysis of panel data**. 4. ed. New York: John Wiley and Sons, 2008.
- BALTAGI, B. H.; GRIFFIN, J. M. Short and long run effects in pooled models. **International Economic Review**, v. 25, n. 3, p. 631-645, 1984.
- BASTOS, D. B.; NAKAMURA, W. T. Determinantes da estrutura de capital das companhias abertas no Brasil, México e Chile no período 2001-2006. **Revista Contabilidade e Finanças**, v. 20, n. 50, p. 75-94, 2009.
- BECK, N.; KATZ, J. N. What to do (and not to do) with time-series cross-section data. **American Political Science Review**, v. 89, n. 3, p. 634-647, 1995.
- BHARGAVA, A.; FRANZINI, L.; NARENDRANATHAN, W. Serial correlation and the fixed effects model. **Review of Economic Studies**, v. 49, n. 4, p. 533-549, 1982.
- BHARGAVA, A.; SARGAN, J. D. Estimating dynamic random effects models from panel data covering short time periods. **Econometrica**, v. 51, n. 6, p. 1635-1659, 1983.
- BREUSCH, T. S.; MIZON, G. E.; SCHMIDT, P. Efficient estimation using panel data. **Econometrica**, v. 57, n. 3, p. 695-700, 1989.
- CAMERON, A. C.; TRIVEDI, P. K. **Microeconometrics using Stata**. College Station: Stata Press, 2009.
- CARNEIRO, L. A. F.; SHERRIS, M. Corporate interest rate risk management with derivatives in Australia: empirical results. **Revista Contabilidade e Finanças**, v. 19, n. 46, p. 86-107, 2008.
- COELHO, A. C. D. **Qualidade informacional e conservadorismo nos resultados contábeis no Brasil**. 2007. 240 f. Tese (Doutorado em Controladoria e Contabilidade) – Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo.
- FÁVERO, L. P. L.; SOTELINO, F. B. Elasticities of stock prices in emerging markets. In: BATTEN, J. A.; SZILAGYI, P. G. (Ed.) *The impact of the global financial crisis on emerging financial markets*. Emerald Group Publishing Limited, 2011. (*Contemporary Studies in Economic and Financial Analysis, Vol. 93*, p. 473-493).

FREES, E. W. **Longitudinal and panel data: analysis and applications in the social sciences**. Cambridge: Cambridge University Press, 2004.

GREENE, W. H. **Econometric analysis**. 6. ed. Upper Saddle River: Prentice Hall, 2007.

HOECHLE, D. Robust standard errors for panel regressions with cross-sectional dependence. **Stata Journal**, v. 7, n. 3, p. 281-312, 2007.

HOLTZ-EAKIN, D.; NEWEY, W.; ROSEN, H. S. Estimating vector auto regressions with panel data. **Econometrica**, v. 56, n. 6, p. 1371-1395, 1988.

HSIAO, C. **Analysis of panel data**. 2. ed. Cambridge: Cambridge University Press, 2003.

ISLAM, N. Growth empirics: a panel data approach. **The Quarterly Journal of Economics**, v. 110, n. 4, p. 1127-1170, 1995.

JONES, D. C.; KALMI, P.; MÄKINEN, M. The productivity effects of stock option schemes: evidence from Finnish panel data. **Journal of Productivity Analysis**, v. 33, n. 1, p. 67-80, 2010.

KENNON, J. **Price to cash flow ratio**. Disponível em: <<http://beginnersinvest.about.com/od/financialratio/a/pricetocashflow.htm>>. Acesso em: 07 ago. 2010.

KRISHNAKUMAR, J.; RONCHETTI, E. (Ed.) **Panel data econometrics: future directions**. Amsterdam: North Holland, 2000.

LIMA, G. A. S. F. Nível de evidenciação x custo da dívida das empresas brasileiras. **Revista Contabilidade e Finanças**, v. 20, n. 49, p. 95-108, 2009.

LIMA, G. A. S. F. **Utilização da teoria da divulgação para avaliação da relação do nível de disclosure com o custo da dívida das empresas brasileiras**. 2007. 98 f. Tese (Doutorado em Controladoria e Contabilidade) - Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo.

LIMA, G. A. S. F. et al. Influência do *disclosure* voluntário no custo de capital de terceiros. In: CONGRESSO USP DE CONTROLADORIA E CONTABILIDADE, 7., 2007, São Paulo. **Anais...** Universidade de São Paulo, 2007.

MADDALA, G. S. **The econometrics for panel data**. Brookfield: Elgar, 1993.

MAKINO, S.; ISOBE, T.; CHAN, C. M. Does country matter? **Strategic Management Journal**, v. 25, n. 10, p. 1027-1043, 2004.

MALACRIDA, M. J. C. et al. An accounting accruals model to predict future operating cash flows: evidence from Brazil. **International Journal of Management**, v. 27, n. 3, p. 562-578, 2010.

MARQUES, L. D. **Modelos dinâmicos com dados em painel: revisão da literatura**. Série *Working Papers* do Centro de Estudos Macroeconômicos e Previsão (CEMPRE) da Faculdade de Economia do Porto, Portugal, n. 100, 2000.

MÁTYÁS, L.; SEVESTRE, P. (Ed.). **The econometrics of panel data: fundamentals and recent developments in theory and practice**. 3. ed. New York: Springer, 2008.

NAKAMURA, W. T. et al. Determinantes de estrutura de capital no mercado brasileiro: análise de regressão com painel de dados no período 1999-2003. **Revista Contabilidade e Finanças**, v. 18, n. 44, p. 72-85, 2007.

PIMENTEL, R. C. **Accounting earnings properties and determinants of earnings response coefficient in Brazil**. 2009. 162 f. Tese (Doutorado em Controladoria e Contabilidade) – Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo.

PIMENTEL, R. C. **O mercado de eurobonds e as captações brasileiras: uma abordagem empírico-descritiva**. 2006. 178 f. Dissertação (Mestrado em Controladoria e Contabilidade) – Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo.

SOUZA, M. S. **Fluxo de caixa por regime de competência**. 2006. 93 f. Dissertação (Mestrado em Controladoria e Contabilidade) – Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo.

WOOLDRIDGE, J. M. **Econometric analysis of cross section and panel data**. 2. ed. Cambridge: MIT Press, 2010.

**APÊNDICE: ROTINAS EM STATA**

Definição do Painel:

```
xtset id t
```

Elaboração do Gráfico 1:

```
graph twoway (scatter var_y t) (lfit var_y t)
```

Elaboração do Gráfico 2:

```
preserve
xtdata, fe
graph twoway (scatter var_y t) (lfit var_y t)
restore
```

Elaboração do Gráfico 3:

```
preserve
xtdata, be
graph twoway (scatter var_y t) (lfit var_y t)
restore
```

Elaboração das Tabelas de Decomposição de Variância do Painel (Tabelas 1 e 4):

```
xtsum id t var_y var_x1 var_x2
```

Elaboração dos Modelos de Dados em Painel Curto:

- POLS com Erros-Padrão Robustos Clusterizados:

```
regress var_y var_x1 var_x2, vce(cluster id)
```

- Modelo com Estimador *Between*:

```
xtreg var_y var_x1 var_x2, be
```

- Efeitos Fixos:

```
xtreg var_y var_x1 var_x2, fe
```

- Efeitos Fixos com Erros-Padrão Robustos Clusterizados:

```
xtreg var_y var_x1 var_x2, fe vce(cluster id)
```

- Efeitos Aleatórios:

```
xtreg var_y var_x1 var_x2, re
```

- Efeitos Aleatórios com Erros-Padrão Robustos Clusterizados:

```
xtreg var_y var_x1 var_x2, re vce(cluster id)
```

Elaboração da Tabela de Comparação dos Estimadores para os Modelos em Painel Curto (Tabela 2):

```
quietly regress var_y var_x1 var_x2, vce(cluster id)
```

```
estimates store POLS_rob
```

```
quietly xtreg var_y var_x1 var_x2, be
```

```
estimates store BE
```

```
quietly xtreg var_y var_x1 var_x2, fe
```

```
estimates store FE
```

```
quietly xtreg var_y var_x1 var_x2, fe vce(cluster id)
```

```
estimates store FE_rob
```

```
quietly xtreg var_y var_x1 var_x2, re
```

```
estimates store RE
```

```
quietly xtreg var_y var_x1 var_x2, re vce(cluster id)
```

```
estimates store RE_rob
```

```
estimates table POLS_rob BE FE FE_rob RE RE_rob, b se stats(N r2 r2_o r2_b r2_w F chi2)
b(%7.5f)
```

Elaboração do Teste LM de Breusch-Pagan:

```
xttest0
```

Elaboração do Teste de Hausman para o Painel Curto:

```
hausman FE RE, sigmamore
```

Elaboração da Tabela de Coeficientes para cada País (Tabela 3):

```
preserve
statsby, by(pais) clear: xtreg var_y var_x1 var_x2, fe
list, clean
restore
```

Elaboração dos Modelos de Dados em Painel Longo:

- Efeitos Fixos:

```
xtreg var_y var_x1 var_x2, fe
```

- Efeitos Aleatórios:

```
xtreg var_y var_x1 var_x2, re
```

- Efeitos Fixos com Erros AR(1):

```
xtregar var_y var_x1 var_x2, fe
```

- Efeitos Aleatórios com Erros AR(1):

```
xtregar var_y var_x1 var_x2, re
```

- POLS com Erros AR(1) e correlação entre indivíduos:

```
xtpcse var_y var_x1 var_x2, corr(ar1)
```

- FGLS com Erros AR(1) e correlação entre indivíduos:

```
xtgls var_y var_x1 var_x2, corr(ar1) panels(correlated)
```

Elaboração da Tabela de Comparação dos Estimadores para os Modelos em Painel Longo (Tabela 5):

```
quietly xtreg var_y var_x1 var_x2, fe
estimates store FE
quietly xtreg var_y var_x1 var_x2, re
estimates store RE
quietly xtregar var_y var_x1 var_x2, fe
estimates store FEAR1
quietly xtregar var_y var_x1 var_x2, re
estimates store REAR1
quietly xtpcse var_y var_x1 var_x2, corr(ar1)
estimates store POLSAR1
quietly xtgls var_y var_x1 var_x2, corr(ar1) panels(correlated)
estimates store FGLSAR1
estimates table FE RE FEAR1 REAR1 POLSAR1 FGLSAR1, b se stats(N r2 r2_o r2_b r2_w
F chi2) b(%7.5f)
```

Elaboração do Teste de Hausman para o Painel Longo:

```
hausman FEAR1 REAR1
```