

Modelos Econométricos para Dados em Painel: Aspectos Teóricos e Exemplos de Aplicação à Pesquisa em Contabilidade e Finanças

Autores:

PATRICIA CRISTINA DUARTE

(UNIVERSIDADE FEDERAL DE MINAS GERAIS)

WAGNER MOURA LAMOUNIER

(PROGRAMA DE MESTRADO EM CIÊNCIAS CONTÁBEIS DO CEPCON)

RENATA TUROLA TAKAMATSU

(UNIVERSIDADE FEDERAL DE MINAS GERAIS)

Resumo

A demanda crescente por informações pelos agentes econômicos tais como investidores, credores ou mesmo pelo próprio governo, evidenciou as técnicas estatísticas e econométricas como mecanismos eficientes para a extração de informações úteis a tomadas de decisões. Contudo, observa-se que a possibilidade de ganho oferecida pelo emprego de modelos estatísticos na análise financeira e patrimonial das organizações não é plenamente aproveitada pelos agentes do mercado de capitais. Nesse contexto, a metodologia de Dados em Painel que consiste na observação de n entidades para dois ou mais períodos de tempo, mostra-se de grande valor. Essa pode ser considerada como uma combinação entre dados de corte transversal e de séries temporais permitindo o controle da heterogeneidade presente nos indivíduos. Ainda são poucas as pesquisas contábeis envolvendo tal metodologia, principalmente em mercados financeiros menos desenvolvidos. Detecta-se uma ausência de textos em português que auxiliem o emprego das vantagens oferecidas pelos Dados em Painel. Por meio da apresentação de seus procedimentos e de algumas das principais pesquisas realizadas espera-se despertar o interesse de pesquisadores bem como, auxiliar na realização de futuras pesquisas empíricas.

Palavras-chave: Análise Financeira; Estatística; Contabilidade; Finanças.

1. Introdução

As demonstrações financeiras constituem-se em um dos principais instrumentos elaborados pela Contabilidade auxiliando a comunicação entre os usuários internos e externos a empresa. A partir de sua análise, dados são transformados em informações úteis aos agentes interessados fornecendo subsídios à tomada de decisões acerca de investimentos, aplicações e decisões relacionadas ao crédito.

Saporito (2005) afirma que, quando esgotados o rol de informações disponibilizadas aos usuários, o enriquecimento e ampliação do processo de análise somente pode evoluir por meio de alternativas de análise diferentes daquelas tradicionalmente difundidas nos meios acadêmicos e utilizadas por profissionais.

Nesse contexto, a utilização de técnicas e modelos estatísticos apresenta-se como uma eficiente maneira de extrair informações mais profundas e precisas dos dados inseridos nas demonstrações financeiras. A partir do uso desses modelos, pode-se propiciar um aumento da

compreensão dos agentes de mercado acerca dos riscos e especificidades de um empreendimento.

Dentre as técnicas estatísticas que vem ganhando destaque na literatura estrangeira está inserida a metodologia dos modelos para Dados em Painel, que consistem de observações de n entidades para dois ou mais períodos de tempo. Esse tipo de dados combina características de séries temporais com dados em corte transversal e são amplamente utilizados em estudos econométricos e nas ciências sociais aplicadas.

Todavia, o que se observa atualmente é uma escassez de trabalhos contábeis, principalmente na literatura nacional, empregando Dados em Painel que forneçam aos pesquisadores e investidores ferramentas necessárias a um completo aproveitamento das vantagens oferecidas pela metodologia. Desse modo, as possibilidades de ganho advindas da utilização das informações providas pelas demonstrações podem não estar sendo plenamente aproveitadas pelos agentes econômicos e pesquisadores envolvidos nessa área.

2. Objetivos

O objetivo geral do presente trabalho é aprofundar a discussão sobre o uso dos modelos sobre Dados em Painel para a análise das demonstrações financeiras das empresas, a fim de que as informações presentes nestas sejam melhor aproveitadas. Busca-se reunir, de forma integrada, um conjunto de referências e de resultados úteis para que futuros pesquisadores possam desenvolver trabalhos utilizando a metodologia.

3. Procedimentos Metodológicos

Segundo Oliveira (2001, p.119), a pesquisa bibliográfica visa conhecer as diferentes formas de contribuição científicas que foram realizadas sobre um determinado assunto ou fenômeno. Assim, a pesquisa caracterizou-se como sendo de natureza bibliográfica, contemplando livros, textos, dissertações, teses e artigos científicos. O presente artigo busca condensar os principais aspectos teóricos sobre o tema, bem como ilustrar a aplicação de modelos de Dados em painel para a análise de fenômenos contábeis / financeiros.

Adicionalmente, quanto ao objetivo da pesquisa, pode-se dizer que é uma pesquisa exploratória, ao buscar conhecer o assunto com mais profundidade. O trabalho foi desenvolvido no sentido de proporcionar uma visão geral acerca do tema, ainda pouco explorado na literatura contábil nacional.

4. Resultados e Discussão

Segundo Hsiao (1986), os modelos para dados em painel oferecem uma série de vantagens em relação aos modelos de corte transversal ou aos de séries temporais sendo que delas se refere ao fato de que esses modelos controlam a heterogeneidade presente nos indivíduos. Como exemplo, tem-se um estudo feito por Hajivassiliou (1987) apud Baltagi (1996) em vários países sobre o problema do pagamento de dívidas externas. É fato que cada país possui características específicas que influenciam a maneira como eles lidam com os empréstimos contraídos. Tais características estão ligadas à história colonial do país, ao regime político adotado, às instituições financeiras existentes, entre outros. Estas características afetam a variável que se deseja explicar, porém não são possíveis de serem mensuradas. A omissão dessas variáveis no modelo gera resultados viesados. Entretanto, o uso dos dados em painel permite controlar os efeitos das variáveis não observadas. Nesse caso, os modelos de dados em painel controlam as diferenças existentes entre os países.

Outra vantagem, conforme Hsiao (1986), é que os dados em painel permitem o uso de mais observações, aumentando o número de graus de liberdade e diminuindo a colinearidade entre as variáveis explicativas. Sabe-se que quando existe multicolinearidade torna-se difícil estabelecer se um regressor individual influencia a variável resposta. Uma vez eliminado esse problema, pode-se obter uma melhora na qualidade da estimação dos parâmetros.

Além disso, dados em painel são capazes de identificar e mensurar efeitos que não são possíveis de serem detectados por meio da análise de dados em corte transversal ou de séries temporais isoladamente. Um exemplo é dado por Bratsberg et al. (2002) em um estudo sobre o impacto da naturalização nos salários dos imigrantes. Através da análise dos dados em corte transversal pode-se verificar que os imigrantes naturalizados ganham salários maiores e ocupam cargos melhores nas empresas em que trabalham. Porém, não foi possível determinar se as vantagens adquiridas pelo imigrante no mercado de trabalho foram consequência da naturalização ou da produtividade dos imigrantes que se naturalizaram. Esse efeito só pode ser identificado com o uso dos dados em painel, uma vez que esses modelos permitem controlar as características individuais dos imigrantes, como a idade e o tempo desde que imigrou. Entretanto, os dados em painel possuem algumas limitações. Conforme Hsiao (1986), como as variáveis são analisadas no tempo, os dados em painel exigem um grande número de observações e, portanto, são mais difíceis de serem implementados.

Quando se trata de uma pesquisa populacional, por exemplo, pode-se defrontar com a ausência de dados. Tal problema ocorre devido a diversas causas. Um entrevistado pode responder à pesquisa em um ano e não responder em outro. Também pode haver contagem incompleta da população de interesse, ocasionando a falta de dados da parte da população que não foi incluída na pesquisa. Além disso, o entrevistador pode cometer equívocos e não registrar a resposta de um entrevistado.

Geralmente os dados em painel cobrem um período de tempo pequeno, devido ao alto custo para a obtenção de novas informações, ou à indisponibilidade das informações no passado. Como os parâmetros estimados são assintoticamente consistentes é desejável que o número de observações seja grande. Neste sentido, quando o período de tempo coberto é pequeno, a propriedade de consistência só será satisfeita se o número de indivíduos for grande. A seguir serão apresentadas as principais variações do modelo de dados em painel, e discutidas as suas características principais.

4.1. O Modelo Geral para Dados em Painel

O modelo geral para os dados em painel é representado por:

$$y_{it} = \beta_{0it} + \beta_{1it}x_{1it} + \dots + \beta_{kit}x_{kit} + e_{it} \quad (1)$$

Nessa notação, o subscrito i denota os diferentes indivíduos e o subscrito t denota o período de tempo que está sendo analisado. β_0 refere-se ao parâmetro de intercepto e β_k ao coeficiente angular correspondente à k -ésima variável explicativa do modelo.

A forma matricial para o i -ésimo indivíduo será dada por:

$$\mathbf{y}_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix} \quad \mathbf{X}_i = \begin{bmatrix} x_{1i1} & x_{2i1} & \cdots & x_{ki1} \\ x_{1i2} & x_{2i2} & \cdots & x_{ki2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1iT} & x_{2iT} & \cdots & x_{kiT} \end{bmatrix} \quad \boldsymbol{\beta}_i = \begin{bmatrix} \beta_{0i1} & \beta_{1i1} & \beta_{2i1} & \cdots & \beta_{ki1} \\ \beta_{0i2} & \beta_{1i2} & \beta_{2i2} & \cdots & \beta_{ki2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \beta_{0iT} & \beta_{1iT} & \beta_{2iT} & \cdots & \beta_{kiT} \end{bmatrix} \quad \mathbf{e}_i = \begin{bmatrix} e_{i1} \\ e_{i2} \\ \vdots \\ e_{iT} \end{bmatrix}$$

Em que y_i e e_i são vetores de dimensão (T x 1) e contém, respectivamente, as T variáveis dependentes e os T erros. X_i é uma matriz de dimensão (K x T) com as variáveis explicativas do modelo. Assim, o elemento x_{kit} refere-se à k-ésima variável explicativa para o indivíduo i no instante de tempo t . Finalmente, β_i é a matriz dos parâmetros a serem estimados.

Neste modelo geral, o intercepto e os parâmetros resposta são diferentes para cada indivíduo e para cada período de tempo. Existindo, assim, mais parâmetros desconhecidos do que observações, não sendo possível, neste caso, estimar os seus parâmetros.

Nesse sentido, é necessário especificar suposições acerca do modelo geral a fim de torná-lo operacional. Entre os modelos que combinam dados de séries temporais e dados em corte transversal, três são os mais utilizados. São eles: *Modelo de Regressões Aparentemente Não-Relacionadas (Seemingly Unrelated Regressions - SUR)*, *Modelo de Efeitos Fixos* e *Modelo de Efeitos Aleatórios*.

4.2. Modelo de Regressões Aparentemente Não-Relacionadas (SUR)

A especificação feita para o modelo geral, que resulta no modelo de regressão SUR, segundo Hill, Griffiths e Judge (1999), é:

$$\beta_{0it} = \beta_{0i} \quad \beta_{1it} = \beta_{1i} \quad \dots \quad \beta_{kit} = \beta_{ki}$$

Deste modo, o modelo SUR supõe que o intercepto e os parâmetros de resposta diferem entre os indivíduos, mas são constantes no tempo. O modelo resultante é dado por:

$$y_{it} = \beta_{0i} + \beta_{1i}x_{1it} + \dots + \beta_{ki}x_{kit} + e_{it} \quad (2)$$

Ou na notação matricial, para o i-ésimo indivíduo:

$$\begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \beta_{0i} + \begin{bmatrix} x_{1i1} & x_{2i1} & \dots & x_{ki1} \\ x_{1i2} & x_{2i2} & \dots & x_{ki2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1iT} & x_{2iT} & \dots & x_{kiT} \end{bmatrix} \begin{bmatrix} \beta_{1i} \\ \beta_{2i} \\ \vdots \\ \beta_{ki} \end{bmatrix} + \begin{bmatrix} e_{i1} \\ e_{i2} \\ \vdots \\ e_{iT} \end{bmatrix}$$

Segundo Hill, Griffiths e Judge (1999), há outras duas suposições acerca do modelo SUR, apresentadas a seguir:

$$I. \quad Var(e_{it}) = \sigma_i^2 \quad e \quad Var(e_{jt}) = \sigma_j^2 \quad , \quad \text{sendo que } \sigma_i^2 \neq \sigma_j^2 \quad (3)$$

$$II. \quad Cov(e_{it}, e_{jt}) = \sigma_{ij} \neq 0 \quad (4)$$

A primeira suposição é de que a variância do erro é constante, mas varia de uma equação para outra. Sendo assim, existe heterocedasticidade entre as diferentes unidades observadas. A segunda suposição indica que existe correlação entre os erros das diferentes equações para o mesmo período de tempo, isto é, existe correlação contemporânea.

A correlação contemporânea é caracterizada pela correlação entre os erros de duas equações no mesmo período de tempo e surge devido à omissão de variáveis. Uma vez que as variáveis explicativas de cada equação do modelo SUR são idênticas, as informações presentes nas variáveis que explicam a variável resposta, mas não entram no modelo, passam a fazer parte do erro de cada um dos indivíduos. Deste modo, o erro de um indivíduo será correlacionado com o erro de outro.

A estimação dos parâmetros do modelo SUR não pode ser feita através do método de mínimos quadrados ordinários (MQO), pois há correlação contemporânea, o que levaria a estimadores viesados. Nessa situação, o método que fornece o melhor estimador não viesado para o modelo é o de mínimos quadrados generalizados (MQG). Contudo, Griffiths et al. (1993) ponderam que para o cálculo do estimador de MQG de $\hat{\beta}$ é necessária a matriz de covariância dos erros das diferentes equações. Como essa matriz é desconhecida, torna-se necessário estimar os seus elementos. Uma vez obtida a matriz de covariância dos erros estimada, é possível calcular o estimador de mínimos quadrados generalizados estimado ($\hat{\beta}$). O estimador $\hat{\beta}$ é não viesado, e para amostras grandes possui distribuição normal com média β .

A estimação desse modelo produzirá um intercepto para cada indivíduo, resultando em n interceptos diferentes e k coeficientes angulares para cada indivíduo, totalizando $n \times k$ coeficientes.

Como visto, o modelo SUR pressupõe que existe correlação contemporânea e heterocedasticidade. Para verificar se estas suposições estão adequadas são necessários alguns testes. Primeiramente, é desejável testar se os termos estocásticos são realmente correlacionados. Griffiths, Hill e Judge (1993) sugerem o teste para correlação contemporânea dado sob as seguintes hipóteses:

$$H_0 : \sigma_{ij} = 0 \quad \forall \quad i \neq j$$

H_1 : pelo menos uma covariância é diferente de zero

Ainda segundo esses autores, a estatística de teste λ é definida como T vezes a soma de todas as correlações ao quadrado, em que T é o número de observações no tempo para cada indivíduo. Sob H_0 , a estatística de teste λ possui distribuição assintótica qui-quadrado χ^2 , com número de graus de liberdade igual ao número de correlações. Por exemplo, em um modelo que possui três indivíduos a estatística de teste será dada por:

$$\lambda = T(r_{21}^2 + r_{31}^2 + r_{32}^2) \quad (5)$$

Em que r_{ij}^2 é a correlação entre os resíduos ao quadrado e λ tem distribuição χ^2 com três graus de liberdade. Caso a hipótese nula seja rejeitada, conclui-se que há correlação contemporânea entre os erros, e desse modo, o modelo SUR é apropriado para os dados observados. Por outro lado, se a hipótese nula não for rejeitada, ou seja, se não existir evidências de que os erros são correlacionados, pode-se estimar as equações para cada indivíduo separadamente e, então, aplicar o método de Mínimos Quadrados Ordinários em cada equação.

O modelo SUR também pressupõe que as variâncias dos erros são diferentes. Uma maneira de verificar isso é através da análise do gráfico dos quadrados dos resíduos e_i^2 versus \hat{Y}_i . Conforme Matos (2000), além da análise gráfica, pode-se realizar testes sendo que os mais utilizados para verificar se existe o problema da heterocedasticidade são: Teste de Quandt-Goldfeld, Teste de Glejser, Teste de Park e Teste de Pesaran e Pesaran.

O modelo SUR apresenta uma grande vantagem ao considerar a correlação contemporânea entre os erros. Como foi visto, quando os erros das equações de cada indivíduo são correlacionados, as equações ficam relacionadas. E essa informação adicional, que é incluída no modelo SUR, aumenta a precisão da estimação dos parâmetros.

Entretanto, este modelo possui uma limitação. Quando o número de indivíduos é muito grande o número de parâmetros a ser estimado também será grande, o que pode levar a estimativas pouco fidedignas do modelo em questão.

Como uma aplicação da metodologia SUR à contabilidade pode-se citar o trabalho de Wells (2002) que utilizou o modelo SUR em um estudo sobre o gerenciamento dos resultados contábeis na época de mudança do CEO (Chief Executive Officer) de empresas australianas. Ele visava investigar se as empresas pioram os seus resultados contábeis antes da troca do CEO, a fim de apresentar resultados melhores no futuro.

As acumulações discricionárias foram consideradas proxy para o gerenciamento de resultados contábeis, isto é, elas determinaram se estava ou não havendo gerenciamento dos lucros ou prejuízos reportados pela empresa. O modelo estimou as acumulações não discricionárias. Para obter as acumulações discricionárias a partir daí, subtraiu-se das acumulações totais os valores obtidos estimados pela regressão. O modelo modificado de Jones utilizado por Wells (2002) foi:

$$EA_{i,t} = \alpha_1 \left(\frac{1}{A_{i,t-1}} \right) + \alpha_2 \left(\frac{\Delta REV_{i,t} - \Delta REC_{i,t}}{A_{i,t-1}} \right) + \alpha_3 \left(\frac{PPE_{i,t}}{A_{i,t-1}} \right) \quad (6)$$

Em que, EA denota as acumulações não discricionárias, ΔREV refere-se à variação das receitas e ΔREC à variação da conta contas a receber, PPE denota a conta Propriedades, Plantas e Equipamentos, e finalmente, $A_{i,t-1}$ denota o ativo total no período t-1.

Os dados utilizados na regressão foram obtidos das 100 maiores empresas listadas na bolsa de valores australianas e compreenderam o período entre 30 de junho de 1984 e 30 de junho de 1994. Nesse período, ocorreram 77 trocas de CEO, porém, devido à falta de dados financeiros, apenas 65 foram utilizadas para a regressão.

A equação (6) foi estimada pelo modelo SUR, uma vez que existem variáveis não observadas e variáveis omitidas na regressão. Além disso, existia heterocedasticidade entre as diferentes empresas observadas e correlação contemporânea conforme o trabalho de Wells demonstrava.

Os resultados obtidos da regressão foram fracos, mostrando que não haviam evidências empíricas que confirmassem o gerenciamento dos resultados contábeis no ano em que ocorreu a mudança de CEO nas empresas da amostra e também no ano anterior à mudança. No entanto, o autor reconheceu que o modelo de Jones possui algumas limitações, por não considerar problemas de erros nas variáveis e simultaneidade.

4.3. Modelo de Efeitos Fixos

O modelo de efeitos fixos pretende controlar os efeitos das variáveis omitidas que variam entre indivíduos e permanecem constantes ao longo do tempo. Para isto, supõe que o intercepto varia de um indivíduo para o outro, mas é constante ao longo do tempo; ao passo que os parâmetros resposta são constantes para todos os indivíduos e em todos os períodos de tempo. De acordo com Hill, Griffiths e Judge (1999), as suposições do modelo são:

$$\beta_{0it} = \beta_{0i} \quad \beta_{1it} = \beta_1 \dots \beta_{kit} = \beta_k$$

O modelo de efeitos fixos será, portanto, dado por:

$$y_{it} = \alpha_i + \beta_1 x_{1it} + \dots + \beta_k x_{kit} + e_{it} \quad (7)$$

A forma matricial para o i -ésimo indivíduo, como sugerido por Griffiths, Hill e Judge (1993), será:

$$\begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \alpha_i + \begin{bmatrix} x_{1i1} & x_{2i1} & \cdots & x_{ki1} \\ x_{1i2} & x_{2i2} & \cdots & x_{ki2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1iT} & x_{2iT} & \cdots & x_{kiT} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} e_{i1} \\ e_{i2} \\ \vdots \\ e_{iT} \end{bmatrix}$$

Neste modelo α_i representa os interceptos a serem estimados, um para cada indivíduo. Como os parâmetros resposta não variam entre os indivíduos e nem ao longo do tempo, todas as diferenças de comportamento entre os indivíduos deverão ser captadas pelo intercepto. Desse modo, α_i pode ser interpretado como o efeito das variáveis omitidas no modelo.

Outra importante suposição do modelo de efeitos fixos é que o intercepto é um parâmetro fixo e desconhecido que capta as diferenças entre os indivíduos que estão na amostra. Assim, as inferências feitas acerca do modelo são somente sobre os indivíduos dos quais dispõe-se de dados.

Pode-se fazer uma especificação do modelo de efeitos fixos utilizando variáveis binárias para representar os interceptos específicos para cada indivíduo. Nesse caso, a equação geral será definida como:

$$y_{it} = \beta_0 + \beta_1 x_{1it} + \dots + \beta_k x_{kit} + \gamma_1 D_{1i} + \gamma_2 D_{2i} + \gamma_3 D_{3i} + \dots + \gamma_n D_{ni} + e_{it} \quad (8)$$

Em que D_{ni} representa uma variável binária para cada indivíduo e equivale a um quando $i = n$ e a zero caso contrário. No entanto, essa equação apresenta uma variável binária para cada indivíduo (n variáveis binárias), resultando no problema econométrico de multicolinearidade perfeita. Para eliminar-se a multicolinearidade deve-se omitir uma variável binária. Assim, o modelo proposto por Stock e Watson (2004) será escrito na forma:

$$y_{it} = \beta_0 + \beta_1 x_{1it} + \dots + \beta_k x_{kit} + \gamma_2 D_{2i} + \gamma_3 D_{3i} + \dots + \gamma_n D_{ni} + e_{it} \quad (9)$$

Neste modelo, $\beta_0, \beta_1, \dots, \beta_k, \gamma_2, \dots, \gamma_n$ são os parâmetros que deverão ser estimados. Quando $i = 1$, o intercepto é dado por $\beta_0 = \alpha_1$. Para $i \geq 2$, o intercepto é dado por $\beta_0 + \gamma_i$. O modelo de efeitos fixos é estimado usando mínimos quadrados ordinários. As estimativas obtidas são não tendenciosas e consistentes, uma vez que o modelo pressupõe que os erros possuem distribuição normal, variância constante e não são correlacionados.

De acordo com Wooldridge (2002), a estimação do modelo com variáveis binárias produz os mesmos resultados da estimação de efeitos fixos. Contudo, o uso de variáveis dummy no modelo não é prático. Isto ocorre porque, mesmo com uma quantidade pequena de indivíduos no modelo, o número de parâmetros a serem estimados aumenta muito. Essa regressão pode até mesmo ser impossível de ser implementada em alguns pacotes econométricos quando se tem um grande número de indivíduos.

Para testar se as suposições sobre o modelo estão adequadas é conveniente fazer um teste para verificar se os interceptos são diferentes entre os indivíduos. As hipóteses nula e alternativa sugeridas por Griffiths, Hill e Judge (1993) são:

$$H_0 : \beta_{01} = \beta_{02} = \dots = \beta_{0k}$$

$$H_1 : \text{os interceptos } \beta_{0i} \text{ não são todos iguais}$$

Estas hipóteses podem ser testadas através da estatística F. Rejeitando-se H_0 , conclui-se que os interceptos não são todos iguais, satisfazendo a suposição do modelo de n interceptos diferentes. Wooldridge (2002) observa que o estimador do intercepto do modelo, $\hat{\alpha}_i$, apesar de não viesado, é inconsistente quando o número de indivíduos observados é muito grande ($N \rightarrow \infty$). Isso ocorre porque cada observação em corte seccional adicionada aumenta um novo coeficiente α_i . Desse modo, os estimadores de α_i serão melhores quando o número de períodos de tempo observado for grande.

O modelo de efeitos fixos é a melhor opção para modelar os dados em painel, quando o intercepto, α_i , é correlacionado com as variáveis explicativas em qualquer período de tempo. Além disso, como o intercepto do modelo é tratado como um parâmetro fixo, também é desejável usar efeitos fixos quando as observações são obtidas de toda a população e o que se deseja fazer são inferências para os indivíduos dos quais dispõe-se de dados.

Como uma aplicação do modelo de efeitos fixos em finanças, pode-se citar Yermack (1996), que utilizou o modelo de efeitos fixos para testar a teoria de que há uma relação inversa entre o valor de mercado de uma empresa e o número de diretores que compõem o conselho de administração. Isto é, quanto menor o conselho de administração, maior a eficiência da empresa, e por consequência, maior o seu valor de mercado.

A regressão implementada por ele utilizou como variável dependente o valor Q de Tobin, que expressa uma aproximação do valor de mercado da empresa. Como uma das variáveis explicativas, utilizou-se o logaritmo do tamanho do conselho de administração. Outras variáveis explicativas foram adicionadas ao modelo para controlar variáveis que poderiam afetar o valor de mercado diretamente.

Para controlar a performance da empresa, utilizou-se as variáveis retorno sobre o ativo (ROA) do ano corrente, o ROA defasado em um ano e o ROA defasado em dois anos. O tamanho da empresa foi controlado com o logaritmo do capital total. O controle da composição do conselho de administração foi feito através da porcentagem de diretores de fora da empresa que participam do conselho.

Para estimar a regressão, foram utilizados dados de 452 empresas entre os anos de 1984 e 1991. As empresas da amostra selecionadas foram classificadas como uma das 500 maiores dos Estados Unidos pela revista Forbes por pelo menos quatro, dos oito anos da análise. Também foram incluídas variáveis dummy para cada ano.

O modelo de efeitos fixos foi utilizado porque existem variáveis não observadas que provavelmente afetam o valor de mercado das empresas. Nesse sentido, esse modelo controla as variáveis omitidas da regressão. Além disso, o modelo de efeitos fixos permite um único intercepto para cada empresa.

O coeficiente do logaritmo do tamanho do conselho de administração foi conforme o esperado, negativo, evidenciando uma relação inversa entre o valor de mercado da empresa e o tamanho do seu conselho de administração. Segundo Yermack (1996), a variável dependente, Tobin's Q, é reduzida em 23% quando o tamanho do conselho de administração dobra. Desse modo, ficou evidenciado pelo seu trabalho que o tamanho do conselho foi um determinante para a performance e a eficiência das empresas analisadas em seu estudo.

4.4. Modelo de Efeitos Aleatórios

O modelo de efeitos variáveis possui as mesmas suposições do modelo de efeitos fixos, isto é, o intercepto varia de um indivíduo para o outro, mas não ao longo do tempo, e os

parâmetros resposta são constantes para todos os indivíduos e em todos os períodos de tempo. A diferença entre os dois modelos refere-se ao tratamento do intercepto.

O modelo de efeitos fixos, como dito anteriormente, trata os interceptos como parâmetros fixos. Já o modelo de efeitos variáveis trata os interceptos como variáveis aleatórias. Isto é, este modelo considera que os indivíduos sobre os quais dispõe-se de dados são amostras aleatórias de uma população maior de indivíduos. Como sugerido por Hill, Griffiths e Judge (1993), os n interceptos serão modelados como:

$$\beta_{0i} = \bar{\beta}_0 + \alpha_i \quad i = 1, \dots, n \quad (10)$$

Pode-se notar que este intercepto é composto pelo intercepto do modelo de efeitos fixos, α_i , que capta as diferenças de comportamento dos indivíduos, e por um segundo componente, $\bar{\beta}_0$, que corresponde ao intercepto populacional.

O modelo geral de efeitos fixos é dado a seguir:

$$y_{it} = \bar{\beta}_0 + \beta_1 x_{1it} + \dots + \beta_k x_{kit} + v_{it} \quad (11)$$

Em que $v_{it} = e_{it} + \alpha_i$ representa o erro.

A forma matricial desse modelo, para o i -ésimo indivíduo, será dada por:

$$\begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \bar{\beta}_0 + \begin{bmatrix} x_{1i1} & x_{2i1} & \cdots & x_{ki1} \\ x_{1i2} & x_{2i2} & \cdots & x_{ki2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1iT} & x_{2iT} & \cdots & x_{kiT} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} e_{i1} \\ e_{i2} \\ \vdots \\ e_{iT} \end{bmatrix}$$

Hill, Griffiths e Judge (1999) apresentam as quatro propriedades do novo termo estocástico v_{it} :

$$\text{I.} \quad E(v_{it}) = 0 \quad (12)$$

$$\text{II.} \quad \text{var}(v_{it}) = \sigma_e^2 + \sigma_\alpha^2 \quad (13)$$

$$\text{III.} \quad \text{cov}(v_{it}, v_{is}) = \sigma_\alpha^2, \quad \forall t \neq s \quad (14)$$

$$\text{IV.} \quad \text{cov}(v_{it}, v_{jt}) = 0, \quad \forall i \neq j \quad (15)$$

As duas primeiras propriedades indicam que v_{it} possui média zero e variância constante, isto é, o erro é homocedástico. Da terceira propriedade tem-se que os erros do mesmo indivíduo em diferentes períodos de tempo são correlacionados, caracterizando a autocorrelação. Finalmente, a quarta propriedade mostra que os erros de diferentes indivíduos no mesmo instante de tempo não são correlacionados, isto é, não existe correlação contemporânea.

Como existe correlação entre os erros do mesmo indivíduo em períodos de tempo diferentes, o método de mínimos quadrados ordinários (MQO) não é o mais apropriado para estimar os coeficientes do modelo de efeitos aleatórios. Desse modo, o método que oferece os melhores estimadores é o de mínimos quadrados generalizados (MQG).

Para testar se o modelo de efeitos aleatórios é apropriado, Breusch e Pagan (1980) apud Greene (1997) desenvolveram um teste baseado no multiplicador de Lagrange, definindo-se as seguintes hipóteses nula e alternativa:

$$H_0 : \sigma_\alpha^2 = 0$$

$$H_1 : \sigma_\alpha^2 \neq 0$$

A estatística de teste foi definida como:

$$LM = \frac{nT}{2(T-1)} \left[\frac{\sum_{i=1}^n \left[\sum_{t=1}^T e_{it} \right]^2}{\sum_{i=1}^n \sum_{t=1}^T e_{it}^2} - 1 \right]^2 \quad (16)$$

Sob a hipótese nula, LM tem distribuição qui-quadrado com um grau de liberdade. Se a hipótese nula for aceita, o modelo de efeitos fixos é preferível ao modelo de efeitos aleatórios. Caso contrário, deve-se assumir que o modelo de efeitos aleatórios é melhor para os dados que estão sendo tratados.

Wooldridge (2002) defende que o principal determinante para decidir entre o modelo de efeitos fixos e o modelo de efeitos variáveis é o efeito não observado α_i . Em situações em que α_i não é correlacionado com todas as variáveis explicativas, o modelo de efeitos variáveis é o mais indicado. Caso contrário, se α_i for correlacionado com algumas variáveis explicativas, então o modelo de efeitos fixos deve ser utilizado. Nesse último caso, o modelo de efeitos variáveis gera estimadores inconsistentes.

Para checar se existe correlação entre α_i e as variáveis explicativas, Greene (1997) sugere o teste de Hausman e formula as hipóteses nula e alternativa:

$$H_0 : \alpha_i \text{ não é correlacionado com as variáveis explicativas}$$

$$H_1 : \alpha_i \text{ é correlacionado com as variáveis explicativas}$$

A estatística de teste pode ser encontrada em Greene (1997). Caso a hipótese nula seja aceita, não há evidências que α_i seja correlacionado com as variáveis explicativas. Sendo assim, o modelo de efeitos aleatórios deve ser utilizado. Por outro lado, se a hipótese alternativa for aceita, deve-se utilizar o modelo de efeitos fixos.

Como uma aplicação do *Modelo de Efeitos Aleatórios* pode-se citar o estudo feito por Hatzinikolaou et al. (2002) que buscou identificar se a incerteza acerca da inflação exerceria algum efeito sobre a estrutura de capital das empresas.

A equação utilizada é:

$$y_{it} = \beta_0 + \beta_1 a_{it} + \beta_2 r_t + \beta_3 u_t + v_i + w_{it} \quad (17)$$

A variável dependente, y_{it} , que expressa a estrutura do capital da firma no momento t era a razão entre as exigibilidades e o valor de mercado do patrimônio líquido da empresa. O ativo total foi representado por a_{it} . A variável r_t era a taxa de juros real esperada, e a variável u_t a incerteza da inflação. Como r_t e u_t são variáveis macroeconômicas, variam no tempo,

mas não entre as diferentes empresas observadas. Além disso, assume-se que os coeficientes angulares $\beta_1, \beta_2, \beta_3$ são os mesmos para todas as empresas em todos os períodos de tempo.

Os dados utilizados para se estimar a regressão foram obtidos para 30 empresas do ramo industrial listadas no índice Dow Jones entre os anos de 1978 e 1997. Para decidir entre utilizar o modelo de efeitos fixos e o modelo de efeitos variáveis, primeiro assumiu-se que a variável v_i era um parâmetro fixo, e assim, o parâmetro intercepto da equação (17) passou a ser $\beta_0 + v_i$. Desse modo, a equação (17) foi estimada através do método de mínimos quadrados ordinários, uma vez que esse é o melhor estimador para o modelo de efeitos fixos.

Uma vez realizada a regressão de efeitos fixos, assume-se que a variável v_i é um parâmetro aleatório. Para testar essa hipótese, utiliza-se o teste de Breusch e Pagan, sendo $H_0 : \sigma_v^2 = 0$. O resultado do teste não rejeitou H_0 , concluindo-se que não há evidências para supor que a variável v_i deve ser tratada como um parâmetro aleatório. Dessa forma, o modelo de efeitos fixos foi o mais apropriado para o problema em questão.

4.5. Modelos com Variáveis Instrumentais

A regressão de variáveis instrumentais consiste em um método que fornece estimadores consistentes quando o erro e uma variável regressora são correlacionados. Sabe-se que, sob esta condição, o método de mínimos quadrados ordinários produz estimadores viesados e inconsistentes.

A correlação entre o erro e a variável regressora é determinada por três motivos: omissão de variáveis relevantes para o modelo, erros nas variáveis e simultaneidade. Stock e Watson (2004) explicam os vieses que surgem devido a esses motivos. O *viés de omissão de variáveis* surge quando uma variável que determina a variável resposta é omitida da regressão e quando pelo menos uma das variáveis explicativas é correlacionada com a variável omitida. Neste caso, o estimador de MQO é inconsistente, mesmo para amostras grandes.

O *viés de erro* nas variáveis é originado quando uma variável explicativa é medida de forma imprecisa. Os erros de medida são determinados por vários fatores. Por exemplo, em uma pesquisa o entrevistado pode dar uma resposta equivocada, ou o entrevistador pode registrar as respostas de forma incorreta, e até mesmo digitar um dado errado no momento que estiver passando os resultados da pesquisa para o computador. Este viés resulta em MQO viesado e inconsistente. Finalmente, o *viés de simultaneidade* surge quando existe simultaneidade na relação causal entre a variável dependente e uma variável explicativa. Isto é, além da variável explicativa determinar a variável dependente, a variável dependente determina uma variável explicativa.

Desse modo, a regressão de variáveis instrumentais soluciona os vieses que surgem devido à correlação entre o erro e a variável explicativa. Este método se baseia na utilização de uma variável adicional, z , denominada variável instrumental que satisfaça duas condições: relevância e exogeneidade. Para uma *variável instrumental* ser relevante, ela deve ser correlacionada com a variável explicativa. Por outro lado, para satisfazer a condição de exogeneidade, o instrumento não pode ser correlacionado com o erro. Quando a variável instrumental satisfaz as duas condições, diz-se que é um instrumento válido.

Através das variáveis instrumentais é possível isolar a parte da variável explicativa que não está correlacionada com o erro, como consequência, será possível obter estimadores

consistentes e não viesados dos parâmetros da regressão. O modelo geral da regressão de variáveis instrumentais é dado por:

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + \beta_{k+1} w_{1i} + \dots + \beta_{k+r} w_{ri} + e_i \quad (18)$$

Nesse modelo, $x_{1i} \dots x_{ki}$ representam as k variáveis explicativas endógenas, isto é, que são correlacionadas com o erro. E $w_{1i} \dots w_{ri}$ representam as r variáveis explicativas exógenas, ou seja, que não são correlacionadas com o erro e_i .

Os coeficientes do modelo são estimados por meio do método de mínimos quadrados em dois estágios (MQ2E). No primeiro estágio é feita a regressão de cada uma das variáveis que estão correlacionadas com o erro sobre as variáveis instrumentais e sobre as variáveis que não estão correlacionadas com o erro. Desse modo, as variáveis x 's serão as variáveis resposta e as variáveis z 's e w 's serão as variáveis explicativas. Supondo que o número de variáveis instrumentais é equivalente ao número de variáveis explicativas correlacionadas com o erro, então a regressão do primeiro estágio para a primeira variável independente, x_1 , é da forma:

$$x_{1i} = \pi_0 + \pi_1 z_{1i} + \dots + \pi_k z_{ki} + \pi_{k+1} w_{1i} + \dots + \pi_{k+r} w_{ri} + u_i \quad (19)$$

Essa regressão é realizada para cada uma das variáveis explicativas endógenas, ou seja, para todas as variáveis que são correlacionadas com o erro. Através dessas regressões, todas as variáveis explicativas endógenas são decompostas em duas partes, de modo que uma parte não esteja correlacionada com o erro. A parte que é correlacionada com o erro e_i é justamente o erro u_i .

Nesse estágio, utiliza-se o método de mínimos quadrados ordinários para estimar os coeficientes da equação (19). Uma vez calculados estes coeficientes, pode-se obter os valores previstos \hat{x} 's para cada uma das variáveis endógenas, utilizando-se a seguinte equação:

$$\hat{x}_{1i} = \hat{\pi}_0 + \hat{\pi}_1 z_{1i} + \dots + \hat{\pi}_k z_{ki} + \hat{\pi}_{k+1} w_{1i} + \dots + \hat{\pi}_{k+r} w_{ri} \quad (20)$$

Pode-se perceber que a equação (20) não possui o termo de erro u_i . Isso ocorre porque esta é a parte que está correlacionada com o erro e_i , e portanto deve ser desconsiderada da equação.

O segundo estágio realiza a regressão da equação do modelo geral de variáveis instrumentais. Porém, as variáveis que são correlacionadas com o erro são trocadas pelos seus valores previstos, que foram calculados no primeiro estágio. Realiza-se, assim, uma regressão de y_i sobre os k valores previstos das variáveis endógenas e sobre as r variáveis exógenas. A equação do segundo estágio será:

$$y_i = \beta_0 + \beta_1 \hat{x}_{1i} + \dots + \beta_k \hat{x}_{ki} + \beta_{k+1} w_{1i} + \dots + \beta_{k+r} w_{ri} + e_i \quad (21)$$

Os coeficientes da equação (21), $\beta_0, \beta_1, \dots, \beta_{k+r}$, também são estimados por meio do método de mínimos quadrados ordinários. Os estimadores de MQ2E são consistentes e possuem distribuição assintótica normal. Contudo, os estimadores serão inconsistentes e tendenciosos se os instrumentos utilizados na regressão não forem válidos. Desse modo, é necessário verificar as condições de relevância e de exogeneidade.

Um instrumento é considerado relevante se estiver correlacionado com a variável explicativa. Quanto mais relevante for um instrumento, maior será a correlação entre x e z , e maior será a variação em x explicada pelo instrumento. Em consequência disso, os

estimadores gerados pela regressão serão mais precisos. Por outro lado, a utilização de instrumentos pouco relevantes na regressão gera estimadores de MQ2E viesados. Dessa maneira, as inferências estatísticas baseadas na estatística t são pouco confiáveis, mesmo em amostras grandes. Por esse motivo, instrumentos que não têm forte correlação com a variável explicativa são denominados instrumentos fracos.

Segundo Stock e Watson (2004), quando há mais de uma variável explicativa endógena na regressão não há como realizar-se testes estatísticos para verificar se um instrumento é fraco ou não. Nesse caso, é necessário que o pesquisador tenha um grande conhecimento do problema com que está trabalhando e faça uma análise dos dados e das variáveis disponíveis para avaliar esta condição.

Além disso, instrumentos que não satisfizem a condição de exogeneidade, e desse modo, são correlacionados com o termo de erro geram estimadores de MQ2E inconsistentes. Como na caso da relevância, não existem testes estatísticos que permitam verificar a condição de exogeneidade quando o número de instrumentos é no mínimo igual ao número de variáveis explicativas. Portanto, mais uma vez torna-se necessária uma avaliação do problema empírico e dos dados utilizados pelo pesquisador para verificar se um instrumento é exógeno ou não.

Uma importante aplicação em contabilidade das variáveis instrumentais é dada por Kang e Sivaramakrishnan (1995) em estudos sobre gerenciamento dos resultados contábeis a questão mais importante é determinar como estimar as acumulações discricionárias, usadas como proxy para o gerenciamento. A idéia proposta por Kang e Sivaramakrishnan (1995) é realizar uma regressão para estimar as acumulações totais, sendo estas, a soma das acumulações discricionárias e das acumulações não discricionárias. Feita esta regressão, obtêm-se os coeficientes das variáveis utilizadas. A partir desses coeficientes, calcula-se a diferença entre o valor observado das acumulações totais e o valor estimado das acumulações. Essa diferença representa as acumulações discricionárias.

Assim, a equação para estimar as acumulações totais será:

$$AB_{i,t} = \Phi_0 + \Phi_1[\delta_{1,i}REV_{i,t}^*] + \Phi_2[\delta_{2,i}EXP_{i,t}^*] + \Phi_3[\delta_{3,i}GPPE_{i,t}^*] + \beta PART_{i,t} + v_{i,t} \quad (22)$$

Em que:

$$\delta_1 = \frac{AR^*_{t-1}}{REV^*_{t-1}} \quad \delta_2 = \frac{APB^*_{t-1}}{EXP^*_{t-1}} \quad \delta_3 = \frac{DEP^*_{t-1}}{GPPE^*_{t-1}} \quad (23)$$

Nessa equação, AR^* refere-se à conta Contas a Receber, REV^* refere-se à Receita, APB^* denota a soma das contas patrimoniais relacionadas às despesas, EXP^* refere-se às despesas, DEP^* denota as despesas de depreciação, $GPPE^*$ refere-se à conta Propriedades, Plantas e Equipamentos e $PART$ é a variável relacionada às acumulações discricionárias. O asterisco representa as contas *unmanaged*.

A equação utilizada para estimar as acumulações discricionárias será dada por:

$$DA_{i,t} = AB_{i,t} - \{ \Phi_0 + \Phi_1[\delta_{1,i}REV_{i,t}^*] + \Phi_2[\delta_{2,i}EXP_{i,t}^*] + \Phi_3[\delta_{3,i}GPPE_{i,t}^*] + \beta PART_{i,t} \} \quad (24)$$

No entanto, observa-se que existe o problema de erro nas variáveis porque algumas das variáveis contábeis utilizadas para estimar as acumulações não discricionárias podem ser alvos de gerenciamento, ou então, não observadas pelos pesquisadores. Também existe o problema de simultaneidade, pois as variáveis explicativas e as variáveis respostas são determinadas conjuntamente no processo de estimativa. Desse modo, a variável resposta também determinará a variável explicativa.

Para solucionar esses problemas são utilizadas variáveis instrumentais para estimar as acumulações totais. Os instrumentos sugeridos pelos autores são as variáveis δ_1, δ_2 e δ_3 defasadas de um ano. Desse modo, os instrumentos são correlacionados com as variáveis explicativas. Por exemplo, a razão entre contas a receber e receita de um ano é correlacionada com a do ano anterior. Por outro lado, os instrumentos não são contemporaneamente correlacionados com os erros das equações.

5. Conclusões

A presente pesquisa buscou ressaltar da utilização de métodos estatísticos para a análise das mais diversas instituições auxiliando no seu processo decisório. Por meio da apresentação dos principais aspectos metodológicos dos Dados em painel, esse artigo apresentou novas informações acerca do tema, para que futuras pesquisas empíricas possam ser realizadas.

A adoção dos Dados em painel oferece respaldos estatísticos e econométricos para a pesquisa científica transmitindo maior credibilidade aos possíveis trabalhos desenvolvidos, pelo seu maior rigor metodológico. Todavia, é necessário julgar ponderadamente cada situação uma vez que as regressões em dados em painel, apesar apresentam uma série de vantagens na pesquisa aplicada não são adequadas a qualquer situação.

Ressalta-se a necessidade de pesquisas em mercados emergentes, uma vez que a comparação do resultado desses estudos com os realizados em mercados desenvolvidos pode auxiliar o entendimento do comportamento das empresas, investidores e mercados. A utilização de técnicas estatísticas oferece informações adicionais ao mercado financeiro e a academia anteriormente não aproveitadas, auxiliando pesquisadores, gestores e investidores para controle dos riscos a que estão sujeitos e para seus processos de decisórios e em suas pesquisas.

6. Referências Bibliográficas

BALTAGI, Badi H. **Econometric analysis of panel data**. New York: John Wiley, 1996.

BRATSBURG, Brent; RAGAN, James F.; NASIR, Zafar M. The Effect of Naturalization on Wage Growth: A Panel Study of Young Male Immigrants. **Journal of Labor Economics**. Chicago, v.20, p.568-592, 2002.

GREENE, William H. **Econometric Analysis**. 2nd ed. New York: MacMillan, 1993.

GRIFFITHS, W. E., HILL, R. C. e JUDGE, G. G. **Learning and Practicing Econometrics**. New York: John Wiley & Sons Inc., 1993.

HATZINIKOLAOU, Dimitris; KATSIMBRIS, George M.; NOULAS, Athanasios G. Inflation uncertainty and capital structure: Evidence from a pooled sample of the Dow-Jones industrial firms. **International Review of Economics and Finance**. North Holland, p. 45-55, 2002.

HILL, R. C; GRIFFITHS, W. E. e JUDGE, G. G.. **Econometria**. São Paulo: Saraiva, 1999.

HSIAO, Cheng. **Analysis of panel data**. Cambridge: Cambridge University Press, 1986.

KANG, S. H.; SIVARAMAKRISHNAN, K. Issues in Testing Earnings Management and an Instrumental Variable Approach. **Journal of Accounting Research**. Rochester, 33 (2), p.353-367, 1995.

KMENTA, Jan. **Elementos de econometria**. 2. ed. São Paulo: Atlas, 1988.

MATOS, Orlando Carneiro de. **Econometria básica: teoria e aplicações**. 3 ed. rev. e ampl. São Paulo: Atlas, 2000.

OLIVEIRA, Silvio Luiz de. **Tratado de metodologia científica: projetos de pesquisas, TGI,TCC, monografias, dissertações e teses**. 3. ed; São Paulo: Pioneira Thomson Learning, 2001.

SAPORITO, Antonio. **Análise referencial: proposta de um instrumento facilitador da análise a longo prazo de demonstrações contábeis**. 191 p. Tese (Doutorado). Faculdade de Economia, Administração e Contabilidade, Universidade de São Paulo, 2005. Disponível em: < <http://www.teses.usp.br/>>. Acesso em: 7 de fevereiro de 2007.

STOCK, James H; WATSON, Mark W. **Econometria**. São Paulo: Pearson Addison Wesley, 2004.

WELLS, Peter. Earnings management surrounding CEO changes. **Accounting and Finance**. Australia, v.42, p.169-193.

WOOLDRIDGE, Jeffrey M. **Econometric analysis of cross section and panel data**. Cambridge, Mass.: MIT Press, 2002.

YERMACK, David. Higher evaluation of companies with a small board of directors. **Journal of Financial Economics**. Chicago, v.40, p.185-211, 1996.