Descrição das variáveis de Estudo

Tratamento para variáveis ...

Proxy das variáveis

Categoria de Notas Explicativas

**Descrição da Variável Dependente**

Por meio das hipóteses de pesquisa e do referencial teórico estudado, foi definida a variável dependente bem como as variáveis de independentes com a finalidade de viabilizar a questão de pesquisa.

Somatório das **Notas Explicativas** / **Qtt** de notas explicativas encontradas naquele respectivo ano.

**Descrição da Variáveis Independente**

Somatório do valor da **LegCPC** / **Qtt** de explicativas encontradas naquele respectivo ano

Quadro com as principais variáveis utilizada

A pesquisa utilizou a regressão de dados em painel, com uma amostra de 40 empresas, envolvendo 320 observações (painel não-balanceado). A unidade básica de estudo é representada por empresas, observadas em diferentes instantes do tempo (de 2010 a 2018). O objetivo da utilização da regressão em painel (ou dados longitudinais) é verificar a relação ente legibilidade da norma contábil e a legibilidade da nota explicativa das empresas, mas levando-se em consideração o tempo e as características individuais das mesmas.

Como iremos inserir no modelo o setores de referencias..

Objetiva-se também verificar há possibilidade de existirem diferenças entre os setores, como por exemplo o Y, que possuem regulação além da própria CVM, realizou-se as análises dos modelos anteriores com a inserção de *dummies* para o controle por setor, a fim de eliminar possíveis impactos decorrentes dos mesmos.

Assim, os modelos de dados em painel diferem dos modelos com dados temporais e *cross section* dado o caráter duplo que atribui a cada variável. De acordo Hill, Judge e Griffiths (2010) o modelo geral para os dados em painel é representado por:

(5)

Com: i =1,..., N os indivíduos (N indivíduos, países, regiões, empresas, setores); t=1,...,T os períodos de tempo que está sendo analisado (T períodos); β0 = parâmetro de intercepto; βk = coeficiente angular correspondente à k-ésima variável explicativa do modelo.

Se para cada indivíduo *i* dispõe-se do mesmo número de dados temporais, o painel chama-se *balanceado* (ou equilibrado). Se o número de dados temporais não é o mesmo para todos os indivíduos, o painel denomina-se de *não-balanceado*.

O teste de Hausman (1978) foi utilizado para decidir qual dos modelos é o mais apropriado: o modelo de efeitos aleatórios (Ho) ou o modelo de efeitos fixos (HA). O teste apresenta-se da seguinte forma:

H0: Cov (ai, Xit) = 0 (efeitos aleatórios)

HA: Cov (ai, Xit) ≠ 0 (efeitos fixos)

Sob a hipótese nula, os estimadores do modelo com efeitos aleatórios são consistentes e eficientes. Sob a hipótese alternativa, os estimadores MQG com efeitos aleatórios (e MQO) são não consistentes, mas os estimadores com efeitos fixos são. Esta é uma das vantagens dos modelos com efeitos fixos, uma vez que permite a endogeneidade dos regressores.

Desta forma, existem alguns modelos diferentes que podem ser utilizados para dados em painel. Assim, para a escolha dos modelos em painel, nesse artigo, por efeito fixo, aleatório ou *Pooled* foi aplicado os testes de Breusch-Pagan, Chow e Hausman. Para o primeiro teste rejeita-se a menos de 1% a hipótese nula. Portanto, o modelo estimado por efeitos aleatórios mostra-se mais adequado do que o modelo *pooled* (*pooled cross-section*).

Posteriormente, foi aplicado o teste de Chow. Rejeita-se a menos de 1% a hipótese nula. Portanto, o modelo estimado por efeitos fixos mostra-se mais adequado do que o modelo *pooled*. Após Teste de Breusch-Pagan e Chow, descarta-se o modelo *pooled*. Por último, demonstra que o efeito fixo foi a melhor opção, comparado o aleatório. Foi utilizado o software Stata 13 para a realização das análises empíricas. Todas as estimações foram realizadas, utilizando-se o comando *robust* para correção de qualquer tipo de heterocedasticidade.

**Análise descritiva das variáveis utilizadas no Modelo**

Nesse sentido, o próximo passo consiste em analisar os resultados das principais medidas de análise em estatística descritiva a partir da analise de tendência central, dispersão e amplitude que são dados relevantes para compreender o comportamento das variáveis do modelo. Na tabela **XI** retrata as estatísticas descritivas para as variáveis dependente, independente e de controle. De modo geral, nota-se variabilidade entre as métricas estudadas, além disso, ressalta-se que nessa primeira análise dos dados não foram feitos tratamento para os outliers. A variável

**Tabela XI:** Estatística descritiva das variáveis utilizadas na pesquisa

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variável** | **Média** | **Desvio Padrão** | **Mínimo** | **Máximo** |
| LegNEMedio | 28,754 | 3,892 | 15,750 | 38,53333 |
| LegCPCMedio | 45,901 | 1,188 | 43,308 | 49,26667 |
| RevCPC | 0,375 | 0,485 | 0,000 | 1 |
| TAM\_log | 16,850 | 1,244 | 13,598 | 20,618 |
| COMPLEX | 3,444 | 1,579 | 1,000 | 7 |
| CAPIT | 18,625 | 9,180 | 0,000 | 31 |
| GC | 0,54375 | 0,4988623 | 0 | 1 |
| AUDIT | 0,938 | 0,242 | 0,000 | 1 |
| EXT | 86,328 | 35,544 | 20,000 | 231 |
| ADR | 0,375 | 0,485 | 0,000 | 1 |

Fonte: Elaborado pelo autor

Primeiramente fazer a análise não considerando os dados com transformação.

Tabela 2 – Análise descritiva das variáveis da pesquisa

Fonte: dados da pesquisa

|  |
| --- |
|  |

**Teste de verificação de adequação para aplicação das técnicas**

Com a finalidade de minimizar os problemas que por ventura existirem na aplicação do modelo em painel, foram realizados testes para especificação para as variáveis e modelos de regressão em Painel.

**Teste de Normalidade**

Primeiramente foi verificado se os dados seguiam distribuição normal. Portanto, através do teste Shapiro-Wilk as variáveis dependente e independente foram submetidas à estatística W.

**Tabela XII:** Teste de Normalidade (Shapiro-Wilk)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variável** | **W** | **V** | **Z** | **Prob>z** |
| LegNEMedio | 0,990 | 2,240 | 1,899 | 0,029 |
| LegCPCMedio | 0,990 | 2,148 | 1,800 | 0,036 |
| TAM | 0,378 | 140,331 | 11,641 | 0,000 |
| COMPLEX | 0,982 | 4,101 | 3,323 | 0,000 |
| CAPIT | 0,921 | 17,894 | 6,792 | 0,000 |
| EXT | 0,949 | 11,421 | 5,735 | 0,000 |

Fonte: Elaborado pelo autor

Segundo os resultados apresentados na tabela XII, rejeitamos a hipótese nula de dados normalmente distribuídos ao nível de significância de 5%. Os valores de grande de V´ indicam a não normalidade dos dados.

**Teste de Assimetria e Curtose**

Para testar formalmente se os resíduos seguem distribuição normal foi utilizado o teste de assimetria e curtose dos resíduos. Assim, com base no “valor p”, análise conjunta, foi possível rejeitar a hipótese nulo de que os dados possuem distribuição normal. No teste, foi utilizado o parâmetro “noadjust” e também

**Tabela XIII:** Teste de Assimetria e Curtose

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variáveis** | **Pr(Skewness)** | **Pr(Kurtosis)** | **chi2(2)** | **Prob>chi2** |
| LegNEMedio | 0,148 | 0,120 | 4,520 | 0,105 |
| LegCPCMedio | 0,906 | 0,297 | 1,100 | 0,576 |
| TAM | 0,339 | 0,000 | 13,930 | 0,001 |
| COMPLEX | 0,633 | 0,000 | 35,240 | 0,000 |
| EXT | 0,103 | 0,000 | 22,630 | 0,000 |

Fonte: Elaborado pelo autor

Com a finalidade de corrigir *outlier* e foram

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **W** | **V** | **Z** | **Prob>z** |
| LegNEMedio | 0,990 | 2,240 | 1,899 | 0,029 |
| LegCPCMedio | 0,995 | 1,047 | 0,109 | 0,457 |
| \*wlTAM | 0,987 | 2,918 | 2,521 | 0,006 |
| COMPLEX | 0,982 | 4,101 | 3,323 | 0,000 |
| CAPIT | 0,921 | 17,894 | 6,792 | 0,000 |
| \*WsqEXT | 0,989 | 2,377 | 2,039 | 0,021 |

Fonte: Elaborado pelo autor

Fazer os teste para inserir nessa seção.

Tratamento de *outliers*

Após o tratamento

1. Teste de Normalidade Assimetria e Curtose

Com a finalidade de avaliar se a amostra coletada é oriunda de uma população a qual apresenta normalidade, Hair Jr. *et al.* (2009) e Kline (2011) recomendam que sejam analisadas as medidas de assimetria e curtose da amostra. Fávero *et al.* (2009, p. 58), explica que a assimetria corresponde ao “grau de desvio ou achatamento, da simetria de uma distribuição, ou seja, se uma distribuição for simétrica, os valores da média, da mediana e da moda serão iguais”. Em relação à curtose, ou achatamento, os autores definem que trata-se da “altura do ponto máximo da curva de distribuição” (FÁVERO *et al*, 2009, p. 60).

De acordo com Hair Jr. *et al.* (2009) e Kline (2011), a amostra segue uma distribuição normal quando os valores de assimetria são inferiores a 3 e os de curtose são menores do que 8. Os resultados das análises dos cálculos dos construtos usados nesta pesquisa indicam que seus valores de assimetria e curtose estão dentro dos limites apontados pelos autores, conforme se pode verificar Tabela 1. Foi necessário utilizar a transformação logarítmica na base 10 para a variável TAM .

1. Teste de Verificação de Multicolinearidade

Refere-se a existência de correlação alta entre duas ou mais variáveis independentes.  
Resultado do teste de Multicolinearidade com as variáveis explicativas do modelo.

Fonte: dados da pesquisa

Critério VIF < 10 e Vif médio < 10

Como observado não a problemas relacionados a multicolinearidade entre as variáveis, sendo assim, não houve indicativo para remoção de nenhuma das variáveis do modelo.

1. Teste de autocorrelação

Será utilizado o Teste de Wooldridge para testar se existe autocorreção ou correlação serial quando os erros ou perturbações da regressão são correlacionados ao longo do tempo violando a hipótese de que os erros são aleatórios ou não correlacionados. Sendo assim, as hipóteses deste teste são:

H0: Ausência de autocorrelação no modelo e;

H1: Presença de autocorrelação no modelo.

Quando foi detectado o problema de autocorrelação, no primeiro momento, utilizou-se a estratégia de correção serial AR(1), processo auto regressivo de primeira ordem, com o objetivo de tratar esse problema (WOOLDRIDGE, 2002).

**Tabela XV:** Teste Wooldridge para autocorrelação

|  |  |  |
| --- | --- | --- |
| **Equação** | **F** | **Prob > F** |
| **01** | F( 1, 39) = 11.112 | 0.0019 |

Fonte: Elaborado pelo autor

O teste de Wooldrige apresentou Prob > F 0.0000. O que faz com que a hipótese nula de ausência de autocorrelação seja rejeitada. Nesse caso, vamos utilizar o parâmetro robust para rodar os modelos.

**Tabela XVI:** Teste Wald para heterocedasticidade

|  |  |  |
| --- | --- | --- |
| **Equação** | **chi2** | **Prob > chi2** |
| **01** | chi2(40) = 2747.02 | 0.0000 |

Fonte: Elaborado pelo autor

Com relação ao teste de Wald retornou um Prob> chic2 =0.0000, desse modo, rejeitando a hipótese nula de homocedasticidade.

**Tabela XVII:** Matriz de Correlação das variáveis de pesquisa

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | LegNEMedio | cLegCPCMedio | RevCPC | wlTAM | COMPLEX | CAPIT | GC | AUDIT | WsqEXT | ADR |
| LegNEMedio | 1.00 |  |  |  |  |  |  |  |  |  |
| LegCPCMedio | -0.1904\* | 1.00 |  |  |  |  |  |  |  |  |
| RevCPC | -0.0178 | 0.0219 | 1.00 |  |  |  |  |  |  |  |
| wlTAM | -0.0445 | 0.0392 | 0.0495 | 1.00 |  |  |  |  |  |  |
| COMPLEX | -0.0799 | 0.1985\* | 0.0072 | 0.3868\* | 1.00 |  |  |  |  |  |
| CAPIT | 0.0466 | -0.1671\* | 0.0423 | 0.6452\* | 0.2023\* | 1.00 |  |  |  |  |
| GC | 0.0554 | 0.0223 | 0.0097 | -0.4765\* | -0.1441\* | -0.4715\* | 1.00 |  |  |  |
| AUDIT | -0.2250\* | 0.2523\* | -0.0667 | -0.1245\* | 0.2611\* | 0.0458 | 0.0486 | 1.00 |  |  |
| WsqEXT | -0.1962\* | 0.3196\* | -0.0127 | 0.4142\* | 0.3763\* | 0.1109\* | -0.2729\* | 0.1911\* | 1.00 |  |
| ADR | -0.1520\* | -0.0609 | 0.0000 | 0.6552\* | 0.3430\* | 0.5078\* | -0.2236\* | 0.0400 | 0.3964\* | 1.00 |

Fonte: Elaborado pelo autor

Estimações Econométricas

**Escolha do melhor modelo**

1. Testes para escolha entre modelos de regressão pool, efeito fixo ou efeito aleatório

Na literatura temos diferentes modelos que podem ser utilizados para dados em painel. Desse modo, faz se necessário a escolha dos modelos em painel, sendo estes, modelo por efeito fixo, aleatório ou Pooled. Para realizar a escolha do modelo foi aplicado o teste de Breusch-Pagan, Chow e Hausman. Para o primeiro teste foi rejeitado a hipótese nula, sendo assim, o modelo estiamdo por efeitos aleatórios mostra se maiis adqueando do que o modeo pooled.

O próximo teste foi de Chow, que verifica a adequação da relação entre as variáveis ao testar a igualdade dos regressões em duas regressões lineares assumindo uma variância comum. As hipóteses desse teste são: h0: Modelo restrito (pooled) e H1: Modelo irrestrito (dados em painel com efeitos fixos) ( Cameron; Trivedi, 2009). O teste retornou um Prob > chic2 =0.000, desse modo, rejeitou-se a hipótese nula, o que indica que modelo estimado por modelo fixo mostra se mais adequado do que o modelo *pooled*. Logo, após esses dois testes ( Breusch-Pagan e Chow) descarta-se o modelo *Polled*.

É essencial que se discuta a distinção entre os modelos de efeitos fixos e aleatório na análise de dados em painel. Sob a hipótese nula de que os estimadores são similares(efeitos aleatórios) ou divergem entre si (efeitos fixos) para casa indivíduo. Assim, foi realizado teste de Hausman no qual rejeita-se a hipótese nula (H0: efeito aleatório, H1: efeito fixo) ocorre a correção para chi2<0 (Prob>chi2 = 0.5075). Assim sendo, para esse estudo o modelo de efeito fixo mostra mais adequado.

Análise das regressões

Empregou-se a estatística VIF (*Variance Inflation Factor*) para avaliar possíveis problemas relacionados à multicolinearidade, para que nenhuma das variáveis independentes ou exploratórias expliquem a mesma função, ou seja, possuírem correlação entre si (GUJARATI; PORTER, 2011).

**Tabela XVIII:** Teste VIF (*Variance Inflation Factor*)

|  |  |  |
| --- | --- | --- |
| **Variável** | **VIF** | **1/VIF** |
| wlTAM | 3,29 | 0,304308 |
| CAPIT | 2,30 | 0,434184 |
| ADR | 2,07 | 0,481958 |
| WsqEXT | 1,67 | 0,598780 |
| GC | 1,50 | 0,666665 |
| COMPLEX | 1,40 | 0,714427 |
| AUDIT | 1,36 | 0,733890 |
| cLegCPCMedio | 1,30 | 0,769361 |
| RevCPC | 1,01 | 0,987258 |
| VIF médio | **1,77** |  |

Fonte: Elaborado pelo autor.

O teste VIF considera para ausência de multicolinearidade valores da estatística do teste inferiores a 10 (dez). Nesse sentido, observando os resultados da Tabela 10, nenhuma variável apresentou multicolinearidade.

Além disso, testou-se a hipótese nula de homocedasticidade dos resíduos. A estatística do teste retornou *p-valor* de 0,000; portanto, os resíduos apresentam heterocedasticidade. Assim, utilizou-se um modelo de erros padrão robustos para amenizar os resíduos em homocedásticos, isto é, distribuição de resíduos como uma variância constante.

Adotadas as verificações e correções descritas anteriormente, estimou-se o modelo de regressão utilizando o modelo fixo (*robust*), para análise da hipótese principal de que a legibilidade da norma contábil afeta a legibilidade das notas explicativas emitidas pelas companhias abertas brasileiras.

**Tabela XIV:** Modelo de Regressão fixo - *Robust*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Variáveis** | **Beta** | **Erro Padrão Robusto** | **Estat. t** | **p-valor** | **Intervalo de Confiança 95%** | |
| LegCPCMedio | 0,502 | 0,294 | 1,71 | 0,095 | -0,092 | 1,097 |
| RevCPC | -0,255 | 0,303 | -0,84 | 0,405 | -0,866 | 0,357 |
| wlTAM | 0,310 | 0,952 | 0,33 | 0,746 | -1,615 | 2,234 |
| COMPLEX | 0,174 | 0,342 | 0,51 | 0,614 | -0,518 | 0,866 |
| CAPIT | 0,136 | 0,119 | 1,14 | 0,259 | -0,104 | 0,377 |
| GC | -1,241951 | 1,019 | -1,22 | 0,23 | -3,302 | 0,818 |
| AUDIT | -0,038 | 0,601 | -0,06 | 0,951 | -1,253 | 1,178 |
| WsqEXT | -0,222 | 0,182 | -1,22 | 0,23 | -0,591 | 0,146 |
| \_cons | 0,167 | 18,966 | 0,01 | 0,993 | -38,196 | 38,529 |

Notas: \*\*\*Significante ao nível **de 0,01.**

\*\*\* p<0.001, \*\*p<0.05.

Todas as estimativas foram calcular por efeitos fixos tendo em vista o teste Hausman.

R² within= 0,089, R² between= 0,002, R² overall= 0,0887, F= 18,13

Fonte: Elaborado pelo autor.

A tabela XIV apresenta os resultados para o modelo estimado em que parte da variabilidade, estatisticamente significativa, do índice de legibilidade das notas explicativas pode ser justificada pelas variáveis independentes utilizadas na regressão. Três variáveis se mostraram importantes para o modelo, já que os valores críticos da estatística T foram estatisticamente significativos ao nível de 1%, sendo elas: Legibilidade da norma contábil (LegCPC), Tamanho da empresa dado pelo ativo total (Tam) e a *dummy* para auditoria (Big4).

A principal variável explicativa *LegCPC* que mensurou a facilidade de leitura da norma contábil registrou relação positiva com o índice de legibilidade das notas explicativas, confirmando a hipótese central da tese e indicando que para cada unidade aumentada no índice Flesch de um pronunciamento contábil, em média, consegue-se um ganho de 0,339 na legibilidade da nota explicativa respectiva. Esse resultado é de suma importância para o Comitê de Pronunciamentos Contábeis, sendo esse o órgão responsável pela convergência das normas internacionais, ao passo que tornar os pronunciamentos técnicos mais legíveis resulta em maior compreensão da norma pelo elaborador das demonstrações financeiras, e por fim, implica em notas explicativas menos complexas. Dessa forma, apoiando-se nas dimensões técnica, semântica e pragmática da Teoria da Comunicação de Shannon (1948), o objetivo principal da contabilidade que é de gerar informações úteis aos usuários, tende a ser atingido com mais eficiência. Ademais, sistemas de evidenciação que possuem alto padrão de qualidade fornecem aos investidores confiabilidade e credibilidade nas demonstrações financeiras (LEVITT, 1998).

A tabela 12 representa os sinais para as variáveis de controle encontrados na literatura (esperado) e os sinais informados pelo modelo econométrico aqui testado (observado).

**Referencias**

FÁVERO, Luiz Paulo et al. **Análise de Dados:**Modelagem multivariada para tomada de decisões. Rio de Janeiro: Elsevier, 2009.

HAIR JR.; Joseph F.; BLACK, William C.; BABIN, Barry J.; ANDERSON, Rolph E.; TATHAM, Ronald L. **Análise Multivariada de Dados**. 6. ed. Porto Alegre: Bookman, 2009.

KLINE, Rex B. **Principles and Practice of Structural Equation Modeling**. 3. ed. New York: The Guilford Press, 2011.