**DRAFT SOBRE MODELO DE REGRESSÃO EM DADO EM PAINEL**

**CLIENTE: Carlos Benassuly Maués Filho**

**AUTHOR: Mário Diego Rocha Valente**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Será apresentado uma revisão bibliográfica a cerda do Modelo Estatístico Multivariado chamado Regressão de Dado em Painel. Descrevendo a formulação matemática, abordando os principais modelos de Dados Empilhados (Pooled OLS), Modelos de Efeitos Fixos (FE) e Modelos de Efeitos Aleatórios (RE). Ao final é apresentando o script criado na linguagem de Programação R.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Inicialmente, aplicou-se a análise fatorial a partir da matriz de correlações, no caso a base de dados é composta por 12 variáveis originais.

**Tabela 1**. Variáveis Originais utilizadas para Realização da Análise Fatorial Exploratória.

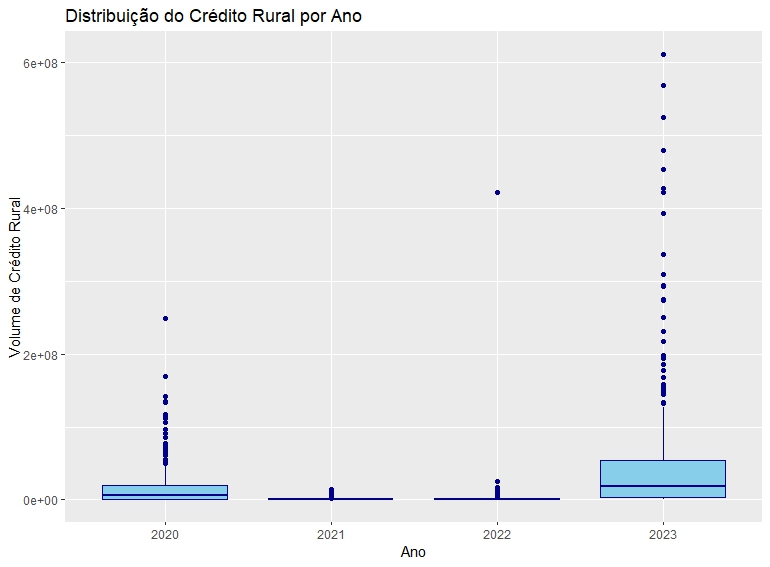
|  |  |
| --- | --- |
| **Itens** | **Variáveis** |
| Var1 | População |
| Var2 | Nº de Empregos Formais |
| Var3 | PIB\_(por mil) |
| Var4 | PIB Per Capita (R$ por pessoa) |
| Var5 | Nº de Agências Bancárias |
| Var6 | Inverso do Incremento do Desmatamento |
| Var7 | Nº de Estabelecimentos Agrícolas |
| Var8 | Rebanho Bovino (Cabeças) |
| Var9 | Área Colhida (LP)(hectare) |
| Var10 | Área Colhida (LT)(hectare) |
| Var11 | Educação Básica |
| Var12 | ICMS |

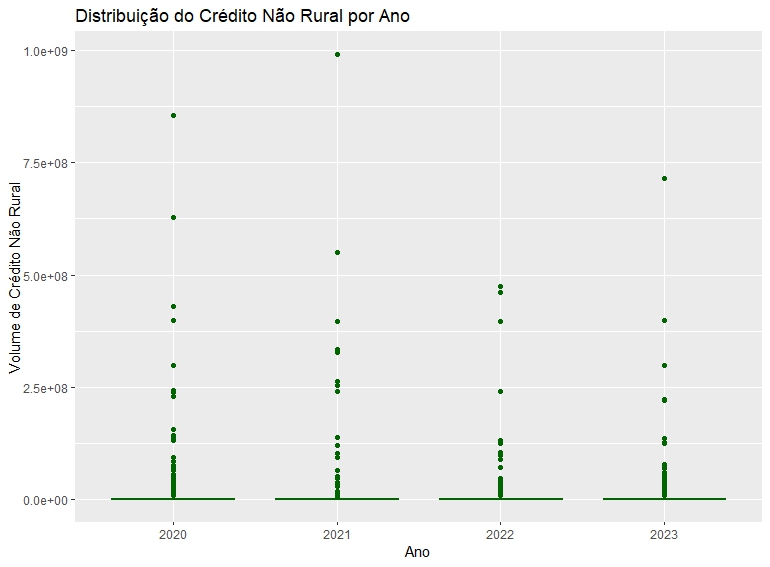
Aplicou-se a Analise Fatorial utilizando o método das componentes principais para extração dos fatores via rotação ortogonal do tipo varimax para obter melhores combinações e usando como critérios para a escolha do número de fatores a extrair, o Critério *Scree Test*, Critério da Raiz Latente e Critério da Percentagem da Variância Explicada. Assim, foi gerado um modelo com 3 fatores que explica 80,85% da estrutura de covariância inicial dos dados referentes ao Volume de Crédito Rural.

Sendo que o primeiro fator explica **52,22%**, o seguindo fator explica **17,37%**, e o terceiro fator explica **11,26%**. Com isso, buscou-se identificar as variáveis que mais influenciam em cada fator, ou seja, as que possuem maiores cargas fatoriais.

Com os resultados obtidos, pode-se observar que o primeiro fator possui pesos mais altos nas variáveis: População, Nº de Empregos Formais, Nº de Agências Bancárias e Educação Básica. O segundo fator com as variáveis: Inverso do Incremento do Desmatamento, Rebanho Bovino e Nº de Estabelecimentos Agrícolas. Já o terceiro fator e composto por PIB e ICMS.

**Visualização Descritiva**





1. **Modelo de Regressão de Dados em Painel**

A análise de dados em painel, também conhecidos como dados longitudinais ou seccionais temporais, envolve observações de múltiplas unidades (como indivíduos, empresas ou em nosso casso “**Municípios**”) ao longo do tempo. A principal vantagem desse tipo de dado é que ele permite controlar tanto a heterogeneidade não observada entre unidades quanto os efeitos da evolução temporal, oferecendo inferências estatísticas mais robustas.

Formalmente, um painel de dados pode ser representado como:

Onde:

* é a variável dependente para a unidade i no tempo t;
* **xit ​** representa uma ou mais variáveis independentes;
* **α** é o intercepto;
* **β** são os coeficientes dos regressores;
* **uit​** é o erro aleatório.

No entanto, diferentes estratégias de modelagem são utilizadas dependendo de como tratamos os efeitos não observados uit​, que podem ser decompostos como:

**Uit = μi + λt + εi**

* μi​: efeito específico da unidade (ex.: município);
* λt ​: efeito específico do tempo (ex.: ano);
* εit​: erro idiossincrático.

A análise de dados em painel combina informações de corte transversal e séries temporais, permitindo explorar variações entre indivíduos ao longo do tempo. Este tutorial cobre os principais modelos para análise de dados em painel, suas especificidades, como escolher o modelo mais adequado, e técnicas para validar as estimativas.

Os modelos abordados incluem:

* **Modelo Pooled OLS (dados empilhados)**: Assume que não há diferenças entre indivíduos.
* **Modelo de Efeitos Fixos (FE)**: Captura variações intra-individuais, controlando para características invariantes.
* **Modelo de Efeitos Aleatórios (RE)**: Assume que as diferenças entre indivíduos são parte do termo de erro.

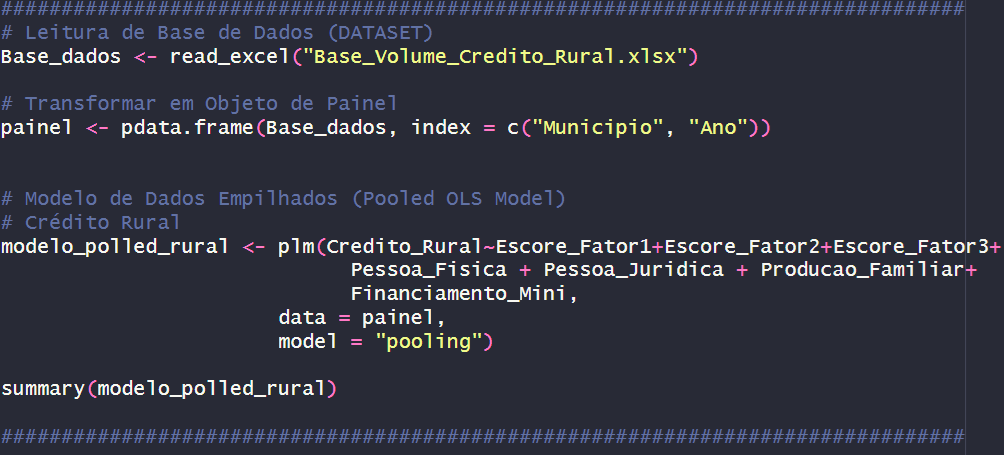
1. **Estimação dos Modelos**
   1. **Modelo de Dados Empilhados (Pooled OLS)**

Este modelo trata de “empilhar” todas as observações da base de dados, ignorando a estrutura de dados em painel. Desta forma, todas as observações são tratadas como não correlacionadas para os indivíduos, com erros homoscedásticos para com os indivíduos.

Trata-se, portanto, da forma mais simplista e ingênua pois desconsidera as dimensões de tempo e espaço combinados, ao mesmo tempo que estima a regressão pelo método dos Mínimos Quadrados Ordinários (MQO) (Gujarati e Porter 2011).

Para executar este modelo de regressão é necessário utilizar a função chamada plm do pacote ***plm*** (Croissant e Millo, 2008), juntamente com a variáveis dependente (no caso o Volume de Crédito Rural) e independentes, indicando a base de dados (*data*) e o tipo do modelo de Regressão (“*pooling*”).

**Figura 1**. Código criado na linguagem de Programação R4.4 para realizar o modelo de dados empilhados (Pooled Model).

****

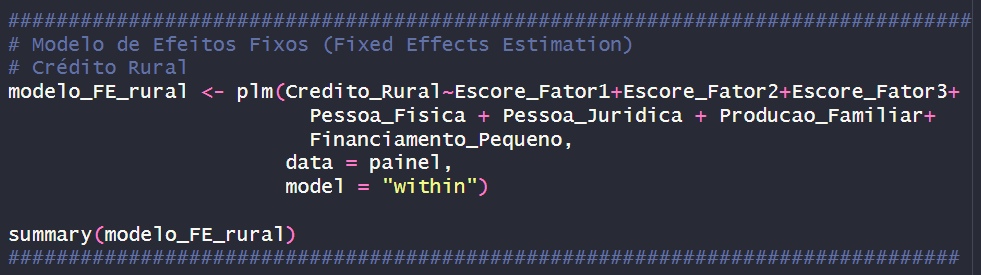
**Tabela 1.** Modelo de Regressão em Pooled para dos referentes ao Volume de Crédito Rural.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Coeficientes** | **Estimate** | **Std.Error** | **T-value** | **P-value** |
| (Intercept) | 1.8527e+07 | 3.8484e+06 | 4.8143 | 2.06e-06 \*\*\* |
| Escore\_Fator1 | -4.7168e+01 | 2.5476e+01 | -1.8515 | 0.064799 . |
| Escore\_Fator2 | 6.6936e+01 | 2.0702e+01 | 3.2333 | 0.001319 \*\* |
| Escore\_Fator3 | 1.8573e+03 | 1.3273e+01 | 139.9332 | < 2.2e-16 \*\*\* |
| Pessoa\_Fisica | 1.4694e+05 | 1.0322e+05 | 1.4236 | 0.155312 |
| Pessoa\_Jurídica | 1.1701e+05 | 7.4257e+04 | 1.5758 | 0.115815 |
| Produção\_Familiar | -1.7253e+05 | 1.2343e+05 | -1.3977 | 0.162921 |
| Financiamento\_Mini | -1.4408e+0 | 1.1644e+05 | -1.2374 | 0.216628 |
| **Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1** | | | | |
| **Total Sum of Squares**: 5.9969e+19 | | | | |
| **Residual Sum of Squares**: 1.249e+18 | | | | |
| **R-Squared**: 0.97917 | | | | |
| **Adj. R-Squared**: 0.97883 | | | | |
| **F-statistic**: 2841.05 on 7 and 423 DF, p-value: < 2.22e-16 | | | | |

* 1. **Modelo de Efeitos Fixos (FE)**

O modelo de efeitos fixos é aquele utilizado para analisar o impacto de variáveis ao longo do tempo. Esse modelo considera os valores de interceptos de cada regressão que variam conforme os efeitos de cada companhia e os coeficientes angulares das variáveis independentes em cada equação que são os mesmos para cada Município do estudo. Desta forma, o intercepto da equação difere de Município para Município, mas os efeitos das variáveis independentes são os mesmo sobre a variável dependente.

**Figura 2**. Código criado na linguagem de Programação R4.4 para realizar o modelo de Efeito Fixos (Fixed Effects Model).

****

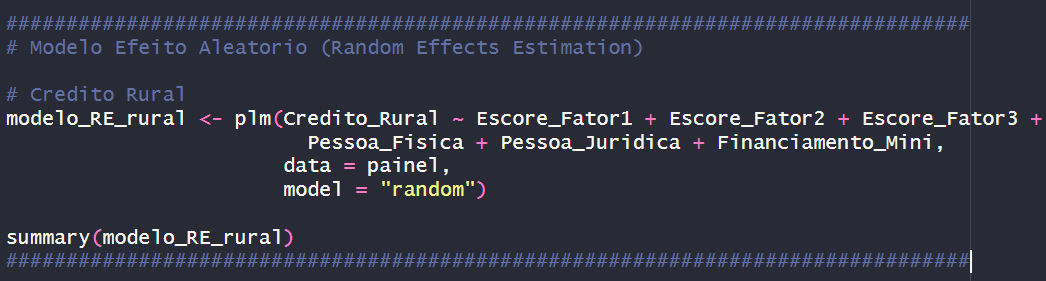
**Tabela 2.** Modelo de Regressão em Efeitos Fixos para dos referentes ao Volume de Crédito Rural.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Predictors** | **Estimate** | **Std.Error** | **T-value** | **P-value** |
| Escore\_Fator1 | -1.7920e+02 | 1.5094e+02 | -1.1872 | 0.2364348 |
| Escore\_Fator2 | -3.0559e+02 | 4.3974e+0 | -6.9494 | 4.226e-11 \*\*\* |
| Escore\_Fator3 | 1.8849e+03 | 6.6173e+00 | 284.8460 | < 2.2e-16 \*\*\* |
| Pessoa\_Fisica | 1.2129e+05 | 3.4449e+04 | 3.5209 | 0.0005241 \*\*\* |
| Pessoa\_Jurídica | 2.3927e+05 | 8.5012e+04 | 2.8145 | 0.0005241 \*\*\* |
| Produção\_Familiar | -9.3545e+04 | 1.0471e+05 | -0.8933 | 0.3726604 |
| Financiamento\_Pequeno | -2.6733e+05 | 9.9798e+04 | -2.6787 | 0.0079568 \*\* |
| **Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1** | | | | |
| **Total Sum of Squares**: 3.9349e+19 | | | | |
| **Residual Sum of Squares**: 1.0454e+17 | | | | |
| **R-Squared**: 0.99734 | | | | |
| **Adj. R-Squared**: 0.99471 | | | | |
| **F-statistic**: 11637.8 on 7 and 217 DF, p-value: < 2.22e-16 | | | | |

* 1. **Modelo de Efeitos Aleatórios (RE)**

O modelo de efeitos aleatórios, é aquele com efeitos aleatórios onde incluem variações dentro e entre as entidades. No modelo de regressão com efeitos aleatórios, os efeitos individuais dos Municípios são considerados variáveis aleatórias, ao contrário do modelo de efeitos fixos. Esse modelo estatístico apresenta alguma forma de variação aleatória.

**Figura 3**. Código criado na linguagem de Programação R4.4 para realizar o modelo de Efeito Aleatório (Randon Model).



**Tabela 3.** Modelo de Regressão em Efeitos Aleatórios para dados referentes ao Volume de Crédito Rural.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Coeficientes** | **Estimate** | **Std.Error** | **Z-value** | **P-value** |
| **(Intercept)** | 1.1929e+07 | 2.3082e+06 | 5.1683 | 2.363e-07 \*\*\* |
| Escore\_Fator1 | -2.3589e+01 | 1.9782e+01 | -1.1924 | 0.233104 |
| Escore\_Fator2 | 1.5936e+02 | 1.5780e+01 | -10.0992 | < 2.2e-16 \*\*\* |
| Escore\_Fator3 | 1.8728e+03 | 6.0527e+00 | 309.4055 | < 2.2e-16 \*\*\* |
| Pessoa\_Fisica | 2.4214e+05 | 3.8795e+04 | 6.2415 | 4.334e-10 \*\*\* |
| Pessoa\_Jurídica | 1.1255e+05 | 3.6976e+04 | 3.0438 | 0.002336 \*\* |
| Financiamento\_Mini | -9.1085e+04 | 4.1739e+04 | -2.1822 | 0.029093 \* |
| **Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1** | | | | |
| **Total Sum of Squares**: 4.6052e+19 | | | | |
| **Residual Sum of Squares**: 5.334e+17 | | | | |
| **R-Squared**: 0.98842 | | | | |
| **Adj. R-Squared**: 0.98836 | | | | |
| **Chisq:** 96210.4 on 6 DF, p-value: < 2.22e-16 | | | | |

**Comparação entre os Modelos**

Após a evidenciação dos modelos de regressão dos tipos agrupado (Pooled), de Efeitos Fixos e de Efeitos Aleatórios, é preciso efetuar os testes para definir qual é o melhor modelo e que por consequência deverá ser considerado.

**Modelo Pooled x Modelo de Efeitos Fixos**

O teste F (**F Test for individual Effects**) é em homenagem a Gregory Chow. Este teste estatístico é realizado para comprovação da “quebra” de uma tendência estável de uma série histórica e é amplamente utilizada em modelos econométricos.

Atualmente, é usado para verificar se há**efeitos individuais significativos** no modelo de dados em painel. Esse teste avalia se as diferenças entre os indivíduos (ou unidades) são estatisticamente significativas.

Os argumentos desta função são dois modelos: o primeiro sendo um modelo interno de efeitos fixos e o segundo um modelo de pool. Os efeitos testados são individuais, temporais ou de dois modos, dependendo dos efeitos introduzidos no modelo interno (EF).

**Teste de Hipóteses.**

**H0:** há igualdade nos interceptos e nas inclinações para topos os indivíduos, caracterizando um modelo (Pooled).

**H1**: O modelo de e Efeito Fixo é o mais indicado.

Para calcular o teste usa a função ***pFtest*** () do pacote plm na linguagem R.

**Tabela 4.** Teste F de Chow para dados referentes ao Volume de Crédito Rural.

|  |  |  |
| --- | --- | --- |
| **Coeficientes** | **F-Test value** | **P-value** |
| Escore\_Fator1 | 11,53 | 2.2e-16\*\*\* |
| Escore\_Fator2 |
| Escore\_Fator3 |
| Pessoa\_Fisica |
| Pessoa\_Jurídica |
| Financiamento\_Mini |
| **Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1** | | |
| **df1**: 206 | | |
| **df2:** 217 | | |

No teste acima, rejeita-se H0, ou seja, devemos considerar a presença de efeitos específicos. Há **evidências estatísticas fortes** de que os efeitos individuais são significativos. Isso significa que as diferenças entre os indivíduos (ou municípios) têm impacto na variável dependente (Crédito Rural) e, portanto, um modelo que considere esses efeitos (como o modelo de efeitos fixos) é mais apropriado do que um modelo Pooled OLS, que não leva em conta essas diferenças.

**Modelo Pooled x Modelo de Efeitos Aleatórios**

Para comparar os modelos foi realizado o **Teste de Lagrange Multiplier (LM)** ou **Teste de Breusch-Pagan**. Esse teste foi desenvolvido em 1979 por Trevor Breusch e Adrian Pagan, e é utilizado para testar a homoscedasticidade em um modelo de regressão linear, ele avalia a variância dos erros de uma regressão e se os valores dependem das variáveis independente.

Esse teste foi utilizado para decidir entre um Modelo de regressão OLS agrupada (Pooled) ou uma Regressão de Efeitos Aleatórios.

**Teste de Hipóteses.**

**H0:** Não há variância significativa nos efeitos aleatórios (modelo Pooled OLS é apropriado).

**H1**: Existe variância significativa nos efeitos aleatórios (modelo de efeitos aleatórios é mais apropriado).

Para calcular o teste usa a função ***plmtest*** () do pacote plm na linguagem R.

**Tabela 4.** Teste de Breusch**-Pagan** para dados referentes ao Volume de Crédito Rural.

|  |  |  |
| --- | --- | --- |
| **Coeficientes** | **Chisq-Test value** | **P-value** |
| Escore\_Fator1 | 482.49 | 2.32e-16\*\*\* |
| Escore\_Fator2 |
| Escore\_Fator3 |
| Pessoa\_Fisica |
| Pessoa\_Jurídica |
| Financiamento\_Mini |
| **Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1** | | |
| **df1**: 1 | | |

Isso significa que há evidências de que a variância dos efeitos aleatórios é significativa, o que justifica o uso do modelo de **efeitos aleatórios (RE)** ao invés do modelo Pooled OLS.

**Modelo de Efeitos Fixos x Modelo de Efeitos Aleatórios**

O **Teste de Hausman** é utilizado para decidir entre os modelos de **efeitos fixos (FE)** e **efeitos aleatórios (RE)** em uma análise de dados em painel. Ele verifica se as estimativas dos coeficientes do modelo de efeitos aleatórios são consistentes em relação ao modelo de efeitos fixos.

**Teste de Hipóteses.**

**H0:** O modelo de **efeitos aleatórios (RE)** é consistente e, portanto, apropriado.

**H1**: O modelo de **efeitos aleatórios (RE)**é inconsistente, indicando que o modelo de **efeitos fixos (FE)** é mais apropriado.

Para calcular o teste usa a função ***phtest*** () do pacote plm na linguagem R.

**Tabela 4.** Teste de Hausmanpara dados referentes ao Volume de Crédito Rural.

|  |  |  |
| --- | --- | --- |
| **Coeficientes** | **Chisq-Test value** | **P-value** |
| Escore\_Fator1 | 23,67 | 0,00025\*\*\* |
| Escore\_Fator2 |
| Escore\_Fator3 |
| Pessoa\_Fisica |
| Pessoa\_Jurídica |
| Financiamento\_Mini |
| **Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1** | | |
| **df1**: 5 | | |

O p-valor é muito baixo (menor que 0,05), levando a rejeição da hipótese nula. Isso significa que o modelo de **efeitos fixos** é consistente e apropriado para os dados

**Pressupostos do Modelo**

1. **Teste de Correlação Cross-Section ou Contemporânea**

Para investigar se existe a dependência cross-section, é aplicado o de Breusch-Pagan LM.

**Teste de Hipóteses.**

**H0:** Não tem correlação cross-section

**H1**: Tem Correlação

Para calcular o teste usa a função ***pcdtest*** (*modelo, test = c ("lm")*) do pacote plm na linguagem R.

**Tabela 4.** Teste de Breusch-Pagan LM

|  |  |  |
| --- | --- | --- |
| **Coeficientes** | **Chisq-Test value** | **P-value** |
| Escore\_Fator1 | 9754,9 | 2.2e-16\*\*\* |
| Escore\_Fator2 |
| Escore\_Fator3 |
| Pessoa\_Fisica |
| Pessoa\_Jurídica |
| Financiamento\_Mini |
| **Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1** | | |
| **df1**: 5439 | | |

O p-valor é muito baixo (menor que 0,05), levando a rejeição da hipótese nula. Isso significa que o modelo de **efeitos fixos** possui dependência cross-section. Assim, é preciso corrigir para erros robustos.

1. **Teste de Correlação Serial**

Para investigar se existe a dependência serial, é aplicado o de Breusch-Godfrey/Wooldrige.

**Teste de Hipóteses.**

**H0:** Não tem correlação serial nos erros idiossincráticos

**H1**: Tem Correlação serial nos erros idiossincráticos

Para calcular o teste usa a função ***pbgtest*** () do pacote plm na linguagem R.

**Tabela 4.** Teste de Breusch-Godfrey/Wooldrige

|  |  |  |
| --- | --- | --- |
| **Coeficientes** | **Chisq-Test value** | **P-value** |
| Escore\_Fator1 | 25,02 | 5,72e-07\*\*\* |
| Escore\_Fator2 |
| Escore\_Fator3 |
| Pessoa\_Fisica |
| Pessoa\_Jurídica |
| Financiamento\_Mini |
| **Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1** | | |
| **df1**: 1 | | |

O p-valor é muito baixo (menor que 0,05), levando a rejeição da hipótese nula. Isso significa que o modelo de **efeitos fixos** possui correlação serial nos erros idiossincráticos. Com isso, é preciso corrigir para erros robustos.

1. **Teste de Heterocedasticidade**

**Teste de Hipóteses.**

**H0:** homocedásticos

**H1**: heterocedasticos

Para calcular o teste usa a função ***bptest*** () do pacote lmtest na linguagem R.

**Tabela 4.** Teste de Breusch-Pagan

|  |  |  |
| --- | --- | --- |
| **Coeficientes** | **Chisq-Test value** | **P-value** |
| Escore\_Fator1 | 1809,3 | 2.2e-16\*\*\* |
| Escore\_Fator2 |
| Escore\_Fator3 |
| Pessoa\_Fisica |
| Pessoa\_Jurídica |
| Financiamento\_Mini |
| **Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1** | | |
| **df1**: 7 | | |

O p-valor é muito baixo (menor que 0,05), levando a rejeição da hipótese nula. Isso significa que o modelo de **efeitos fixos** possui erros heterocedasticos.

Nesse contexto, será aplicado a metodologia de Driscoll e Kraay (1998), que utiliza uma um estimador Robusto para a Matriz de Variância/Covariância, corrigindo o modelo escolhido nos testes anteriores

**Modelo Final**

**REFERENCIAS BIBLIOGRÁFICAS**

CROISSANT, Y; MILLO, G. “Econometria de dados em painel em R: o pacote plm.” Journal of Statistical Software , 27 (2), 1-43, 2018. doi: 10.18637/jss.v027.i02 (URL: <http://doi.org/10.18637/jss.v027.i02> ).

GREENE, William H. Análise econométrica. 7ª ed. Pearson Education, 2012.

TORRES-REYNA, Oscar. Introdução a Modelos de Efeitos Fixos/Aleatórios usando R. Princeton: Universidade de Princeton, 2010. Disponível em: <http://www.princeton.edu/~otorres/Panel101R.pdf> .

Breusch, Trevor S. 1978. «Testing for autocorrelation in dynamic linear models». Australian Economic Papers 17 (31): 334–55.

Breusch, Trevor S, e Adrian R Pagan. 1979. «A simple test for heteroscedasticity and random coefficient variation». Econometrica: Journal of the Econometric Society, 1287–94.

Breusch, Trevor Stanley, e Adrian Rodney Pagan. 1980. «The Lagrange multiplier test and its applications to model specification in econometrics». The Review of Economic Studies 47 (1): 239–53.

Gujarati, Damodar N., e Down C Porter. 2011. Econometria básica. 5a ed. New York: Mc Graw Hill. <https://doi.org/10.1126/science.1186874>.

Hausman, Jerry A. 1978. «Specification tests in econometrics». Econometrica: Journal of the econometric society, 1251–71.

Pesaran, M Hashem. 2015. «Testing weak cross-sectional dependence in large panels». Econometric Reviews 34 (6-10): 1089–1117.

Wooldridge, Jeffrey M. 2010. Econometric analysis of cross section and panel data. MIT Press.