

Predição da Inadimplência

aplicação com séries temporais

Bruno Faria

Universidade Federal do Paraná
Departamento de Estatística

27 de novembro de 2024

Sumário

1. Introdução

2. Material e Método

3. Resultado e Discussão

4. Considerações Finais





Introdução

Introdução

sobre a inadimplência

No Mundo

- Inadimplência, descumprimento de uma obrigação financeira após prazo de vencimento;
- Em 1974, criação do Comitê de Regulação e Supervisão Bancária, o *Comitê de Basileia*;
- Em 1988 criação do *Basileia I*; em 2004, o *Basileia II*; e 2010, o *Basileia III*;
- Em geral, a gestão do risco analisa perdas e vulnerabilidades em empresas e instituições;

Introdução

sobre a inadimplência

No Mundo

- Inadimplência, descumprimento de uma obrigação financeira após prazo de vencimento;
- Em 1974, criação do Comitê de Regulação e Supervisão Bancária, o *Comitê de Basiléia*;
- Em 1988 criação do *Basiléia I*; em 2004, o *Basiléia II*; e 2010, o *Basiléia III*;
- Em geral, a gestão do risco analisa perdas e vulnerabilidades em empresas e instituições;

No Brasil

- Em 1994, implementação do *Basiléia I*, resolução CMN 2099/84;
- Em 2013, implementação do *Basiléia III*, resolução CMN 4192/13 e 4193/13;

Introdução

sobre a inadimplência

No Mundo

- Inadimplência, descumprimento de uma obrigação financeira após prazo de vencimento;
- Em 1974, criação do Comitê de Regulação e Supervisão Bancária, o *Comitê de Basiléia*;
- Em 1988 criação do *Basiléia I*; em 2004, o *Basiléia II*; e 2010, o *Basiléia III*;
- Em geral, a gestão do risco analisa perdas e vulnerabilidades em empresas e instituições;

No Brasil

- Em 1994, implementação do *Basiléia I*, resolução CMN 2099/84;
- Em 2013, implementação do *Basiléia III*, resolução CMN 4192/13 e 4193/13;

No Estudo

- Objetivo é ajustar um modelo para a inadimplência, no Brasil;
- O ajuste é com foco em modelos de séries temporais.

Introdução

sobre a inadimplência - série temporal

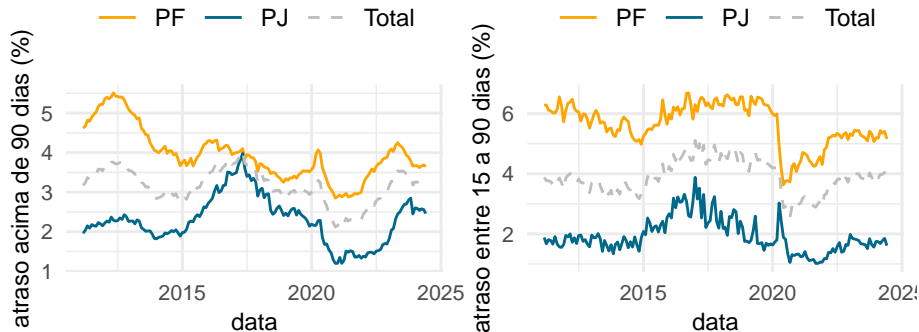


Figura 1: percentual da carteira de crédito com atraso acima de 90 dias e entre 15 e 90 dias após o vencimento - total, PF e PJ, 2011-2024

Introdução

sobre a inadimplência - segmentação do modelo

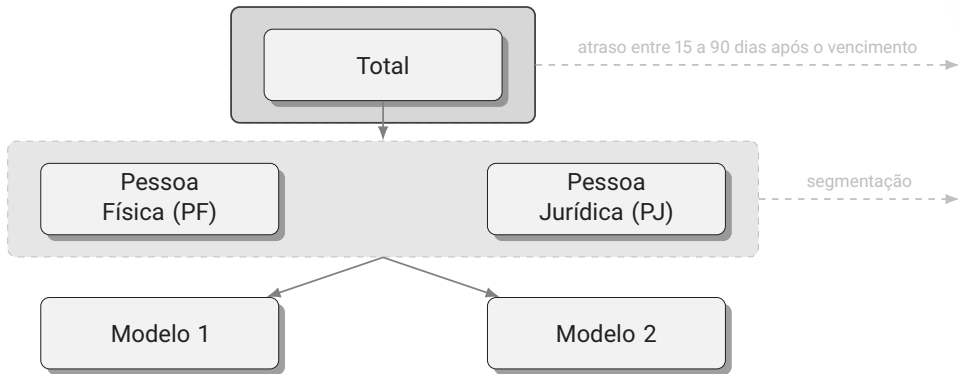


Figura 2: segmentação dos modelos de inadimplência com séries temporais



Material e Método

Material e Método

Material

- Conjunto de dados DELAY, 1944-2024;
- Dados coletados do Banco Central do Brasil (BACEN);
- Linguagem R (4.3.1) e Látex (3.14). Software Excel (2010) e VSCode (1.92.2).

Material e Método

Material

- Conjunto de dados DELAY, 1944-2024;
- Dados coletados do Banco Central do Brasil (BACEN);
- Linguagem R (4.3.1) e Látex (3.14). Software Excel (2010) e VSCode (1.92.2).

Método

- Estrutura do dado com padronização das datas;
- Seleção de variáveis por teste *de estresse*;
- Validação cruzada considerando *pandemia*;
- Modelagem do dado para regressão com erros ARIMA.

Material e Método

padronização da data

Tabela 1: estrutura das datas para criar o conjunto de dados DELAY, 1944-2024

data	min	max	duplicate	data	min	max	duplicate
df_icc	2013-01-01	2024-07-01	FALSE	df_filter_cc	2013-01-01	2024-06-01	FALSE
df_unemployment	2012-03-01	2024-06-01	FALSE	df_filter_currency_dedup	2013-01-01	2024-06-01	FALSE
df_cc	2011-03-01	2024-07-01	FALSE	df_filter_delay_15_90	2013-01-01	2024-06-01	FALSE
df_delay_15_90	2011-03-01	2024-06-01	FALSE	df_filter_family_debt	2013-01-01	2024-06-01	FALSE
df_spread	2011-03-01	2024-07-01	FALSE	df_filter_ibc_br_full	2013-01-01	2024-06-01	FALSE
df_family_debt	2005-01-01	2024-06-01	FALSE	df_filter_icc	2013-01-01	2024-06-01	FALSE
df_ibc_br_full	2003-01-01	2024-06-01	FALSE	df_filter_igp_di	2013-01-01	2024-06-01	FALSE
df_public_debt	2001-12-01	2024-06-01	FALSE	df_filter_inpc	2013-01-01	2024-06-01	FALSE
df_selic_copom	1999-03-01	2024-08-01	TRUE	df_filter_ipca	2013-01-01	2024-06-01	FALSE
df_selic_copom_dedup	1999-03-01	2024-08-01	FALSE	df_filter_ipca12m	2013-01-01	2024-06-01	FALSE
df_tjlp	1994-12-01	2024-08-01	FALSE	df_filter_public_debt	2013-01-01	2024-06-01	FALSE
df_currency	1984-11-01	2024-08-01	TRUE	df_filter_selic_copom_dedup	2013-01-01	2024-06-01	FALSE
df_currency_dedup	1984-11-01	2024-08-01	FALSE	df_filter_spread	2013-01-01	2024-06-01	FALSE
df_ipca12m	1980-12-01	2024-07-01	FALSE	df_filter_tjlp	2013-01-01	2024-06-01	FALSE
df_ipca	1980-01-01	2024-07-01	FALSE	df_filter_unemployment	2013-01-01	2024-06-01	FALSE
df_inpc	1979-04-01	2024-07-01	FALSE				
df_igp_di	1944-02-01	2024-07-01	FALSE				

Material e Método

estrutura do dado

Tabela 2: estrutura da coluna e sua classe no conjunto de dados DELAY, 2013-2024

column	class	head	na	element	number	letter	accent	punctuation	space	empty
date	POSIXct, POSIXt	2013-01-01 2013~	no	yes	yes	no	no	yes	no	no
all_delay_15_90	numeric	3.7 3.76 3.81	no	yes	yes	no	no	yes	no	no
pj_delay_15_90	numeric	1.96 1.87 1.88	no	yes	yes	no	no	yes	no	no
pf_delay_15_90	numeric	5.75 5.97 6.11	no	yes	yes	no	no	yes	no	no
ipca	numeric	0.86 0.6 0.47	no	yes	yes	no	no	yes	no	no
ipca12m	numeric	6.15 6.31 6.59	no	yes	yes	no	no	yes	no	no
inpc	numeric	0.92 0.52 0.6	no	yes	yes	no	no	yes	no	no
igp_di	numeric	0.31 0.2 0.31	no	yes	yes	no	no	yes	no	no
max_selic_copom	numeric	7.25 7.25 7.25	no	yes	yes	no	no	yes	no	no
tjlp	numeric	5 5 5	no	yes	yes	no	no	yes	no	no
max_currency	numeric	2.0471 1.9893 2.~	no	yes	yes	no	no	yes	no	no
unemployment	numeric	7.3 7.8 8.1	no	yes	yes	no	no	yes	no	no
spread	numeric	14.47 14.59 14.2~	no	yes	yes	no	no	yes	no	no
cc	numeric	239.834 226.768 ~	no	yes	yes	no	no	yes	no	no
cc_i12	numeric	11.8811373125277~	no	yes	yes	no	no	yes	no	no
icc	numeric	19.94 19.98 19.7~	no	yes	yes	no	no	yes	no	no
family_debt	numeric	38.19 38.27 38.1	no	yes	yes	no	no	yes	no	no
public_debt	numeric	20.89 21.38 21.2~	no	yes	yes	no	no	yes	no	no
ibc_br_full	numeric	139.32 136.14 14~	no	yes	yes	no	no	yes	no	no
ibc_br_i12	numeric	4.48477576121193~	no	yes	yes	no	no	yes	no	no



Resultado e Discussão

Resultado e Discussão

exploração do dado

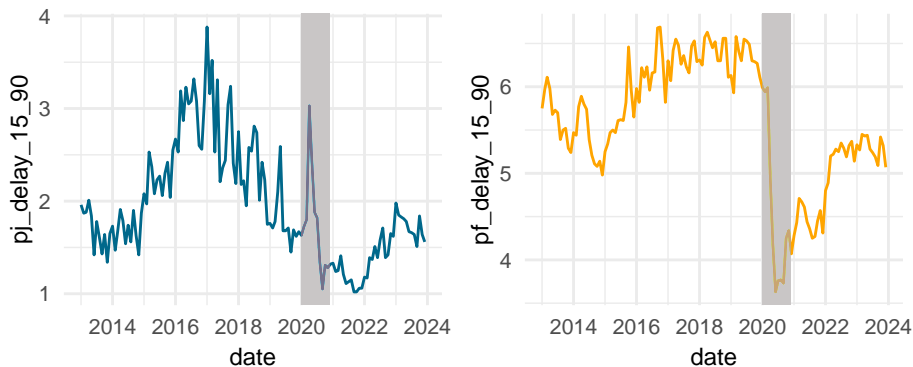


Figura 3: percentual da carteira de crédito com atraso entre 15 e 90 dias após o vencimento - PF e PJ, 2013-2024

Resultado e Discussão

exploração do dado

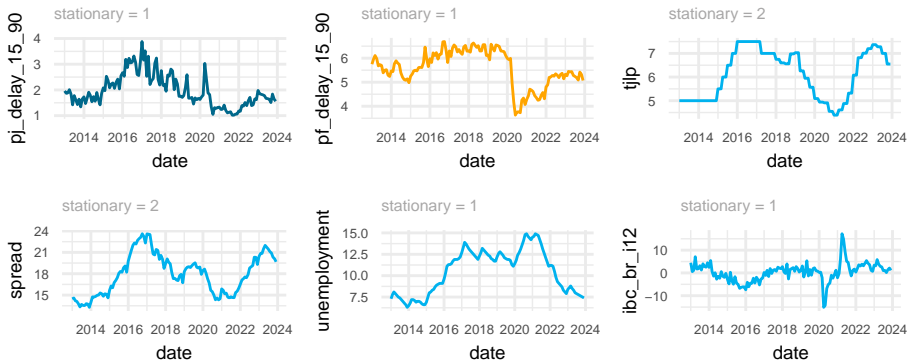


Figura 4: matriz de séries temporais da variável resposta e covariáveis selecionadas - dados de treino, 2013-2023

Resultado e Discussão

exploração do dado

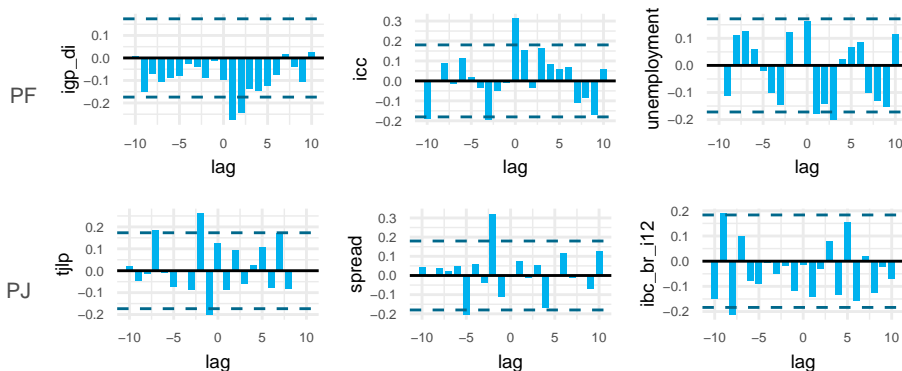


Figura 5: matriz de correlação cruzada com método de pré-branqueamento - variáveis respostas e covariáveis selecionadas, dados de treino, 2013-2023

Resultado e Discussão

modelagem do dado

Tabela 3: coeficientes estimados - PF, dados de treino, 2013-2023

model	parameter	estimate	std_error	t_value	Pr(> t)
model _{pf1}	ar1	0.8087	0.1029	7.8574	0.0000
	ma1	-0.9459	0.1552	-6.0946	0.0000
	sma1	-0.9459	0.1552	-6.0946	0.0000
	spread	0.1509	0.0475	3.1770	0.0019
	icc	0.1151	0.1038	1.1094	0.2695
	arimab::lagb(igp_di, 1)	-0.0421	0.0211	-1.9950	0.0483
	arimab::lagb(unemployment, 1)	-0.1452	0.0581	-2.5013	0.0137
	arimab::lagb(public_debt, -3)	0.0925	0.0206	4.4949	0.0000
model _{pf2}	ar1	-0.3559	0.4254	-0.8366	0.4045
	ma1	-0.5061	0.4075	-1.2417	0.2168
	ma2	-0.4479	0.3745	-1.1959	0.2341
	sma1	-0.9440	0.2992	-3.1554	0.0020
	arimab::lagb(unemployment, 1)	-0.2208	0.0915	-2.4121	0.0174
	arimab::lagb(inpc, 1)	-0.0141	0.0481	-0.2925	0.7704
	arimab::lagb(public_debt, -3)	0.0731	0.0252	2.9012	0.0044
	arimab::lagb(tjlp, 1)	0.0089	0.1055	0.0843	0.9330

Resultado e Discussão

modelagem do dado

Tabela 4: coeficientes estimados - PJ, dados de treino, 2013-2023

model	parameter	estimate	std_error	t_value	Pr(> t)
model _{pj1}	ar1	0.4698	0.1120	4.1936	0.0001
	ar2	0.0311	0.1043	0.2978	0.7664
	ma1	-1.9882	0.1779	-11.1757	0.0000
	ma2	0.9899	0.1777	5.5697	0.0000
	sma1	-0.7708	0.1373	-5.6141	0.0000
	arimab::lagb(spread, -2)	0.1541	0.0197	7.8235	0.0000
	arimab::lagb(cc_i12, 1)	-0.0078	0.0032	-2.4032	0.0179
	arimab::lagb(ibr_br_i12, -9)	0.0069	0.0088	0.7927	0.4296
model _{pj2}	ar1	0.4845	0.1083	4.4736	0.0000
	ar2	0.0289	0.0996	0.2896	0.7726
	ma1	-1.9934	0.1193	-16.7099	0.0000
	ma2	0.9956	0.1191	8.3622	0.0000
	sma1	-0.8143	0.1526	-5.3372	0.0000
	arimab::lagb(tjlp, -2)	0.1302	0.0834	1.5613	0.1212
	arimab::lagb(spread, -2)	0.1217	0.0327	3.7178	0.0003
	arimab::lagb(ibr_br_i12, -9)	0.0013	0.0094	0.1394	0.8894

Resultado e Discussão

performance do ajuste

Tabela 5: performance do ajuste - PF e PJ, dados de treino e teste, 2013-2024

model	data	n	rmse	mae
$model_{pf1}$	train	128	0.1941	0.1465
	test	5	0.1948	0.1699
$model_{pf2}$	train	128	0.1750	0.1203
	test	5	0.2439	0.2114
$model_{pj1}$	train	122	0.2761	0.2015
	test	5	0.4087	0.3965
$model_{pj2}$	train	123	0.2765	0.1892
	test	6	0.3371	0.3184

Resultado e Discussão

performance do ajuste

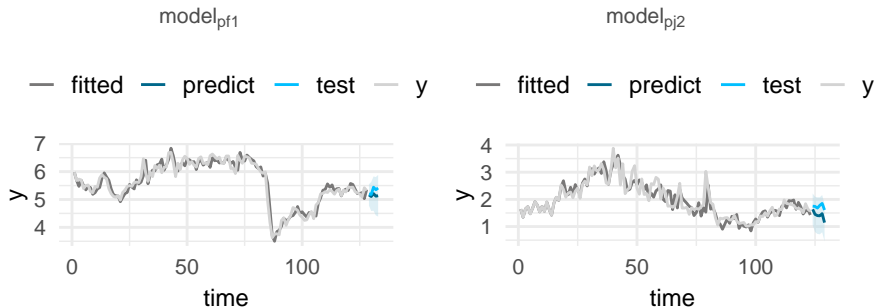


Figura 6: modelos ajustados selecionados - PF e PJ, dados de treino e teste, 2013-2024

Resultado e Discussão

Discussão

- Método *stepwise* para covariáveis;
- Problema com método MASE;
- Ajuste com modelo linear misto com efeitos fixos e aleatórios.

Resultado e Discussão

discussão

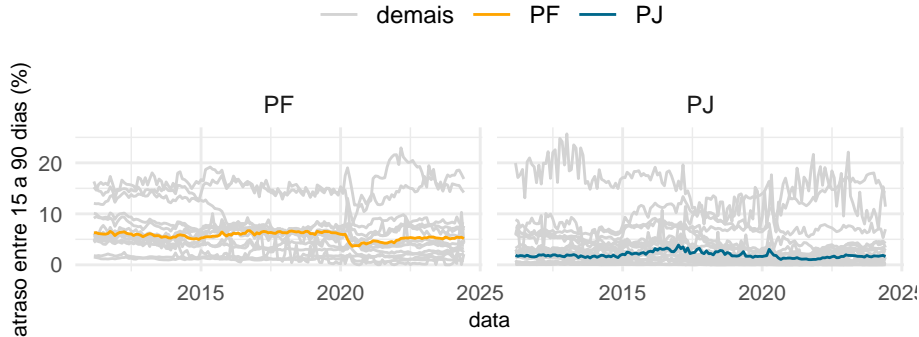


Figura 7: percentual da carteira de crédito livre por produto com atraso entre 15 e 90 dias após o vencimento - PF e PJ, 2011-2024



Considerações Finais

Considerações Finais

Sobre o projeto

- Objetivo é ajustar um modelo para a inadimplência, no Brasil;
- O ajuste é com foco em modelos de séries temporais.

Considerações Finais

Sobre o projeto

- Objetivo é ajustar um modelo para a inadimplência, no Brasil;
- O ajuste é com foco em modelos de séries temporais.

Sobre o resultado

- Segmentação de modelos: PF e PJ;
- Seleção de variáveis pela literatura *teste de estresse*;
- Seleção de covariáveis pelo método de pré-branqueamento;
- Ajuste dos modelos utilizando lags;
- Destaque para o modelo PF;

Considerações Finais

Sobre o projeto

- Objetivo é ajustar um modelo para a inadimplência, no Brasil;
- O ajuste é com foco em modelos de séries temporais.

Sobre o resultado

- Segmentação de modelos: PF e PJ;
- Seleção de variáveis pela literatura *teste de estresse*;
- Seleção de covariáveis pelo método de pré-branqueamento;
- Ajuste dos modelos utilizando lags;
- Destaque para o modelo PF;

Sobre o futuro

- Considerar uma estratégia de ajuste mais geral para obter um modelo razoável.



Apêndice

Apêndice

criação de pacote

O pacote ARIMAB é um *wrapper* da função ARIMA, do r-base, com o objetivo de facilitar o ajuste para modelo de regressão com erros arima. Em geral, o escopo do pacote é composto em:

1. Por definição, o modelo ajustado será

$$Z_t = X_t\beta + \epsilon + \eta_t$$

2. Implementação da fórmula para o modelo;
3. Implementação do método de pré-branqueamento (BOX, 2015);
4. Implementação do método de análise de resíduos da regressão linear com ACF e PACF (CRYER; CHAN, 2008);
5. Implementação de gráficos para exploração compacta de variáveis entre outras melhorias para o ajuste de séries temporais.



- Pacote: arimab
- Autor: Bruno Faria
- Versão: 0.0.0.9000

Obrigado!