

ECONOMETRIA III

Trabalho para Conclusão da Disciplina de Econometria III

Professor André Luiz Squarize Chagas
FEA-USP - EAE 0327

Beatriz Moreira Petito N^o USP 10334954
Enrico Peceguini Ruggieri N^o USP 4649262
Fernanda Pucinelli de Souza Queiroz N^o USP 8592131
Gabriel Nascimento Pereira N^o USP 11840190
Gian Lucca du Mont Santoro N^o USP 8574882
João Victor Oliveira Balbino Silva N^o USP 11766972

Faculdade de Economia, Administração, Contabilidade e Atuária
UNIVERSIDADE DE SÃO PAULO

Resumo

Este trabalho se propõe a realizar uma análise de séries temporais associados ao desemprego no Brasil, destacando a relevância dessas variáveis para a avaliação e formulação de políticas econômicas. Valendo-se da metodologia estudada na disciplina de Econometria de Séries Temporais, são propostas aplicações de modelos econométricos univariados, como AR, MA, ARMA, ARIMA, SARIMA, e multivariados, tais como VAR e VECM, para previsão da população ocupada. Utilizando dados provenientes do Ipeadata, a pesquisa investiga mudanças dinâmicas que ocorreram tanto no período pré-pandemia quanto no pós-pandemia.

A abordagem metodológica utilizada se propõe a oferecer visão mais profunda sobre as complexidades do desemprego, contribuindo para o debate nesse assunto de notável importância social e econômica.

Palavras-chave: [Econometria de séries temporais, Emprego no Brasil, Variáveis macroeconômicas, Pandemia de COVID-19, Modelos SARIMA e VAR.]

Abstract

This study aims to conduct a time series analysis associated with unemployment in Brazil, emphasizing the significance of these variables in the assessment and formulation of economic policies. Leveraging the methodology studied in the Time Series Econometrics course, applications of univariate econometric models such as AR, MA, ARMA, ARIMA, SARIMA, and multivariate models like VAR and VECM are proposed for forecasting the employed population. Using data from Ipeadata, the research investigates dynamic changes that occurred both in the pre-pandemic and post-pandemic periods.

The methodological approach employed seeks to provide a deeper insight into the complexities of unemployment, contributing to the discourse on this subject of notable social and economic importance.

Keywords: [Time Series Econometrics, Employment in Brazil, Macroeconomic Variables, COVID-19 Pandemic, SARIMA and VAR Models.]

Sumário

1	Introdução	4
2	Revisão da Literatura	5
3	Dinâmicas da taxa de ocupação	7
3.1	Metodologia	7
3.1.1	Os Dados	8
3.2	Identificação do modelo	11
3.2.1	Forecasting	15
4	Tudo é endógeno: modelos multivariados	17
4.1	Metodologia	17
4.1.1	Caracterização dos dados	17
4.2	Resultados	20
4.2.1	Estimações e diagnósticos	20
4.2.2	Simulações: Choques da população ocupada na inflação, na expectativa de inflação, na taxa de cambio e no PIB	21
4.2.3	Backtesting e resultados dentro e fora da amostra	34
5	Conclusão	35
A	Resultados dos testes de raiz unitária	36
B	Resíduos	36
B.1	Resíduos dos modelos estimados	36
B.2	Histogramas dos resíduos	39

1 Introdução

O desemprego, uma métrica central em estudos econômicos e sociais, é definido como a condição de pessoas aptas ao trabalho que não estão empregadas, mas estão disponíveis e em busca ativa de emprego (IBGE, s.d.). Esta variável reflete o estado do mercado de trabalho e impacta, entre outros dados econômicos, a inflação - conforme demonstrado pela Curva de Phillips (A. W. Phillips, 1958). Assim, a modelagem e formulação de um modelo preditivo para o desemprego torna-se fundamental para a formulação de políticas econômicas e para a compreensão da dinâmica de crescimento de um país.

Taxas elevadas de desemprego estão frequentemente associadas ao aumento da pobreza e da instabilidade política e social, o que reforça a necessidade de políticas públicas eficazes e bem fundamentadas (Barroso et al., 2022). Para além dos impactos sociais e econômicos, o desemprego também é apontado pela literatura como detrimental para saúde mental e física, gerando consequências deletérias bastante profundas em diversos aspectos do tecido social (Linn, Sandifer e Stein, 1985).

Entender, portanto, as dinâmicas do desemprego é essencial para o desenvolvimento de políticas macro e microeconômicas efetivas. Nesse contexto, os modelos de séries temporais emergem como ferramentas valiosas, permitindo a identificação de tendências de longo prazo, análise de ciclos econômicos, volatilidade e sazonalidade. Estas técnicas são vitais para a projeção de cenários futuros e para análises conjuntas com outras variáveis econômicas (Stock e Watson, 2001).

Este artigo tem como objetivo analisar e sugerir modelos econométricos univariados e multivariados para a previsão da população ocupada brasileira. Com isso, pretende-se eventualmente verificar mudanças de dinâmicas ocorridas entre os anos pré-pandemia (2012-2019) e pós-pandemia e, eventualmente, compreender também o impacto deste choque na economia brasileira.

Inicialmente, abordaremos modelos univariados como AR, MA, ARMA, ARIMA e SARIMA, empregando testes estatísticos robustos para identificar raízes unitárias em séries temporais (Dickey e Fuller, 1979; P. C. B. Phillips e Perron, 1988).

A pesquisa começa, então, com a coleta e processamento de dados de séries temporais do Ipeadata, focando na série de 'pessoas ocupadas'. Nesta primeira parte do estudo aplicam-se métodos de decomposição de séries para identificar componentes de

tendência, sazonalidade e irregularidades. Testes de raiz unitária (ADF) e técnicas de dessazonalização e diferenciação em nível são empregados para verificar a estacionariedade das séries. Modelos SARIMA são desenvolvidos para analisar o comportamento da série temporal.

Na segunda parte, abordaremos modelos multivariados (VAR e VECM), que são utilizados para examinar as dinâmicas e interdependências entre as variáveis econômicas selecionadas e outras variáveis econômicas, como inflação e PIB (Toda e Yamamoto, 1995; Pesaran, Shin e Smith, 2001). Critérios como AIC e BIC orientam a escolha das defasagens e a adequação do modelo. Testes de causalidade de Granger e funções de resposta ao impulso são aplicados para explorar relações de causa e efeito entre as variáveis. Além disso, a decomposição de erro de previsão quantifica a contribuição de cada variável na previsão da outra.

Essa abordagem multivariada é crucial para compreender a complexidade do desemprego no Brasil. A integração desses modelos não apenas oferece uma visão acadêmica detalhada, mas também traz implicações práticas para a formulação de políticas públicas e planejamento econômico. Ao compreender as tendências e ciclos do desemprego através de uma lente econométrica robusta, podemos desenvolver estratégias mais eficientes para melhorar a saúde econômica e o bem-estar social (Arendt e Jakubik, 2021).

2 Revisão da Literatura

A econometria de séries temporais desempenha um papel crucial na análise do desemprego, fornecendo ferramentas para entender as dinâmicas econômicas subjacentes e para formular políticas eficazes. Um dos modelos mais significativos neste domínio é o modelo VAR, que permite examinar as interações entre múltiplas variáveis econômicas. O trabalho de Cavalcanti e Silva, 2010 é um exemplo notável, analisando a relação entre dívida pública, política fiscal e nível de atividade no Brasil. Este estudo resalta a importância da modelagem econométrica no entendimento das complexidades macroeconômicas e suas implicações para o desemprego.

Em relação à inflação e desemprego, a Curva de Phillips continua sendo um conceito central. Mazali e Divino, 2010 oferecem uma perspectiva aprofundada sobre a rigidez salarial real e a nova Curva de Phillips no Brasil, evidenciando as nuances da política

monetária e do mercado de trabalho e seu impacto na inflação e no desemprego. Essa relação é essencial para compreender os mecanismos econômicos que influenciam as taxas de desemprego.

A causalidade em econometria, explorada por Granger, 1969, é um elemento fundamental para identificar relações significativas entre variáveis econômicas. A metodologia de Granger permite determinar se uma variável pode ser considerada um preditor confiável para outra, um aspecto crucial na análise do desemprego e de suas variáveis correlatas.

Lee e Strazicich, 2003 contribuem significativamente para a análise de séries temporais com seu teste de raiz unitária, que é vital para a modelagem adequada de séries temporais econômicas, especialmente no contexto das taxas de desemprego. A identificação correta de séries estacionárias é fundamental para garantir a validade e precisão dos modelos econométricos.

A análise do desemprego em regiões metropolitanas do Brasil por Cruz, Anjos e Sáfadi, 2014 oferece uma visão detalhada das tendências regionais e urbanas do desemprego. Este estudo é particularmente relevante para políticas públicas direcionadas, uma vez que o desemprego pode variar significativamente entre diferentes áreas geográficas.

Bacciotti e Marçal, 2020 examinam a taxa de desemprego no Brasil em um período de quatro décadas, demonstrando a utilidade das séries temporais na compreensão das tendências de longo prazo do desemprego. Esta perspectiva histórica é crucial para entender como as políticas econômicas e as mudanças no mercado de trabalho afetam o desemprego ao longo do tempo.

A pesquisa de Barroso et al., 2022 é um exemplo de como a ciência de dados e os algoritmos de séries temporais podem ser aplicados para analisar o desemprego no contexto da pandemia de COVID-19. Este estudo destaca a importância de abordagens metodológicas inovadoras para entender o impacto de choques externos no mercado de trabalho.

Por fim, Casagrande et al., 1809 discutem métodos de previsão para a taxa de desemprego, destacando a importância de técnicas avançadas de séries temporais para antecipar mudanças no mercado de trabalho. A capacidade de prever as taxas de desemprego é fundamental para o planejamento econômico e a formulação de políticas proativas.

3 Dinâmicas da taxa de ocupação

3.1 Metodologia

Como discutido na introdução, o presente trabalho versa sobre uma das variáveis medidas pelo IBGE, o Instituto Brasileiro de Geografia e Estatística, para acompanhamento do mercado de trabalho brasileiro. A série de dados que será objeto de estudo principal da primeira parte deste trabalho será a de "Pessoas Ocupadas", coletada mensalmente através da PNAD continua, ela mede o número absoluto de indivíduos com alguma ocupação no período em que a pesquisa foi realizada.

Em contraste, outras métricas relacionadas ao mercado de trabalho divulgadas pelo IBGE são:

- **PIA:** Explora e analisa informações referentes às características estruturais fundamentais do setor empresarial na atividade industrial do país. O foco da pesquisa é em empresas industriais legalmente estabelecidas, considerando aquelas cuja principal fonte de receita derive da prática da atividade industrial.
- **Pessoas na Força de Trabalho:** São consideradas como parte da força de trabalho tanto as pessoas ocupadas quanto as desocupadas que tenham tomado ações efetivas para encontrar emprego.
- **Força de Trabalho Potencial:** Caracterizada como o grupo composto por pessoas com 14 anos de idade ou mais que não estavam atualmente empregadas nem desocupadas, mas que apresentavam o potencial de ingressar na força de trabalho. Esse grupo abrange os chamados "desalentados", que são indivíduos que, embora disponíveis para trabalhar, desistiram de procurar emprego devido a desafios ou condições desfavoráveis no mercado de trabalho. Essa categoria reflete a parcela da população que pode ser mobilizada para participar ativamente da força de trabalho, caso as condições melhorem.

Para a modelagem, análise e previsão utilizaremos a metodologia Box-Jenkins, a qual permite identificar o modelo adequado a série de dados através da função de autocorrelação desta. Também se pretende realizar a especificação do Modelo Autorregressivo Integrado de Médias Móveis com Sazonalidade (SARIMA), que leva em consideração tanto um componente de integração nas séries bem como a possível presença de sazonalidade, como pode ser visto:

$$\begin{aligned}
& \text{SARIMA}(p, d, q)(P, D, Q)_s = \\
& = (1 - \sum_{i=1}^p \phi_i L^i)(1 - \sum_{i=1}^P \Phi_i L^{is})(1 - L)^d(1 - L^s)^D y_t \\
& = (1 + \sum_{i=1}^q \theta_i L^i)(1 + \sum_{i=1}^Q \Theta_i L^{is}) \varepsilon_t
\end{aligned} \tag{1}$$

onde p, q, P, Q são as ordens das partes AR e MA (não sazonais e sazonais),

d, D são as ordens de diferenciação (não sazonal e sazonal),

s é o período de sazonalidade,

ϕ_i, Φ_i são os coeficientes para as partes AR (não sazonal e sazonal),

θ_i, Θ_i são os coeficientes para as partes MA (não sazonal e sazonal),

L é o operador de defasagem,

y_t é a série temporal,

ε_t é o termo de erro.

O período estudado das séries em questão compreende Março de 2012 até Junho de 2020. Dessa forma, como veremos adiante, será necessário tomar em consideração os efeitos da pandemia de Covid-19 na série observada.

3.1.1 Os Dados

Uma análise gráfica inicial apresentou a série temporal descrita na figura 1.

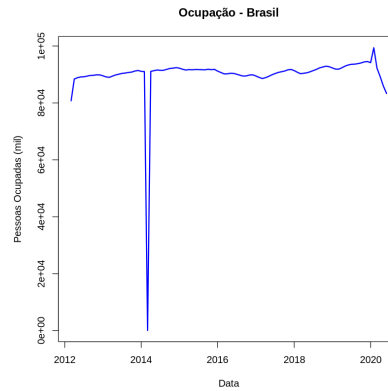


Figura 1: Pessoas Ocupadas

Como primeiro passo, investigamos a presença de possíveis *outliers* que pudessem prejudicar a estimação do modelo. Para isso, utilizamos a função `tsoutliers`, nativa da linguagem de programação *R*, cujo método é baseado em detectar observações que fogem da *trend* estimada para a série. Observações do tipo foram detectadas e confirmadas através de um boxplot (figura 2).

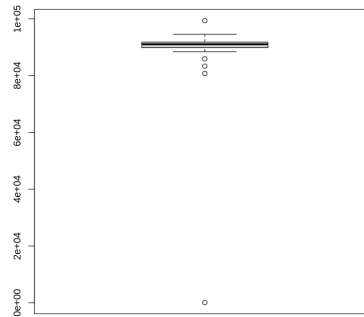


Figura 2: Boxplot para detecção dos *outliers*

Em seguida, através da função `tsclean()`, retiramos as observações problemáticas da nossa série, obtendo resultado descrito na figura 3.

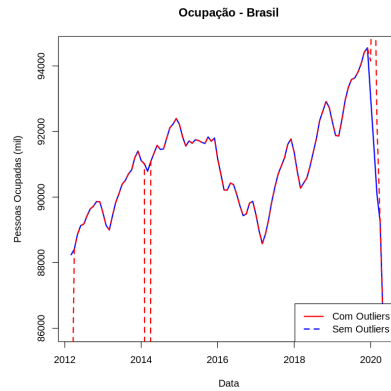


Figura 3: Gráfico com *outliers* detectados por função do R

Além da presença de *outliers*, também investigamos a presença de sazonalidade nos dados. Estando diretamente ligada ao mercado de trabalho, o número de pessoas ocupadas deve sofrer com a sazonalidade de setores do comércio, como o varejo, afetados pelas festas e remunerações adicionais de fim de ano, ou de turismo, sujeitos a temporadas em que há maior movimentação nas cidades turísticas.

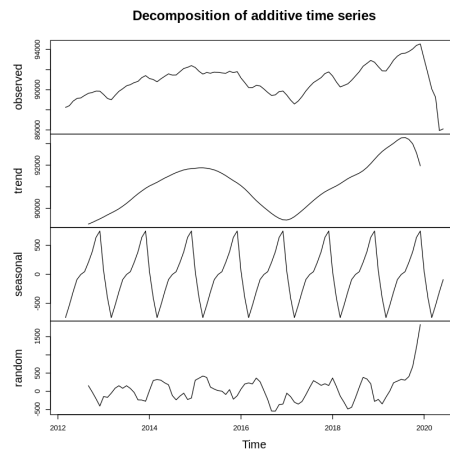


Figura 4: Decomposição tendência e sazonalidade da série sem *outliers*

Como pode ser visto na figura 4, o número de Pessoas Ocupadas possui uma clara sazonalidade com um aumento relevante do número de pessoas empregadas nos últimos meses de cada ano. Na parte da tendência, podemos identificar a crise brasileira de 2014-15 com a diminuição da parte *trend* dessa série nesse período. Também há o efeito do período mais agudo da pandemia que pode ser percebido no final da série decomposta,

que gera uma grande distorção nos níveis observado, de tendência e aleatório.

3.2 Identificação do modelo

Tratados os dados, podemos prosseguir com a metodologia de Box-Jenkins. Esse procedimento foi proposta inicialmente por George Box, Gwilym Jenkins e Gregory Reinsel, em um artigo de 1970 com o título *Time Series Analysis: Forecasting and Control*. A metodologia consiste na identificação, estimação e verificação das séries analisadas.

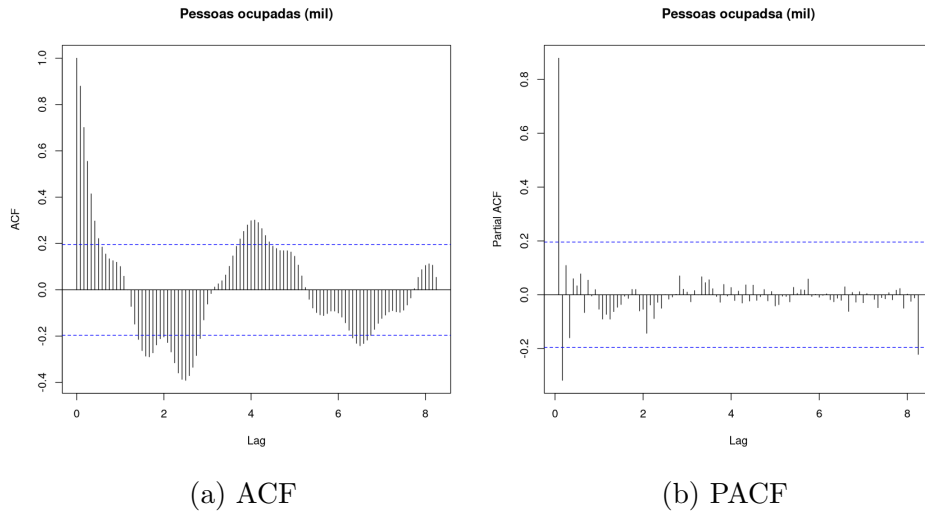


Figura 5: ACF e PACF

Em uma primeira análise das funções de autocorrelação e de autocorrelação parcial (figuras 5a e 5b), dado o decaimento após o primeiro *lag*, e seu comportamento oscilante, é possível inferir que o modelo seja um $ARMA(1, 1)$. No entanto, precisamos nos assegurar de que a série é estacionária. De acordo com Bueno, 2012, um processo estocástico ou série temporal é definido como (fracamente) estacionário se cumprir três requisitos:

- i. $E|y_t|^2 < \infty$
- ii. $E|y_t| = \mu$, para todo $t \in Z$
- iii. $E(y_t - \mu)(y_{t-j} - \mu) = \gamma_j$

O procedimento mais utilizado para testar a presença de estacionaridade se dá pelos testes de raiz unitária, em especial o teste de Dickey-Fuller e Dickey-Fuller aumentado, que testam a hipótese nula H_0 de que a série possui raiz unitária, e com a hipótese alternativa de estacionaridade. Utilizando a série original de dados, apenas desconsiderando os *outliers* identificados via *tsclean()*, não foi possível rejeitar a hipótese nula (estatística teste -1.62 vs. valor crítico menor ou igual à -3.15 para 90% de grau de confiança).

Através da diferenciação em nível, que envolve subtrair cada observação pelo valor da observação anterior, podemos verificar se a série temporal possui uma tendência ou padrão de crescimento ao longo do tempo. Ao testar para raiz unitária, a série em primeira diferença possibilita a rejeição da hipótese nula de não estacionaridade com 95% de grau de confiança (estatística teste -3.45 vs. valor crítico menor ou igual à -3.45 para 95% de grau de confiança), indicando que o modelo seja um ARIMA(1, 1, 3) (figuras 6a e 6b).

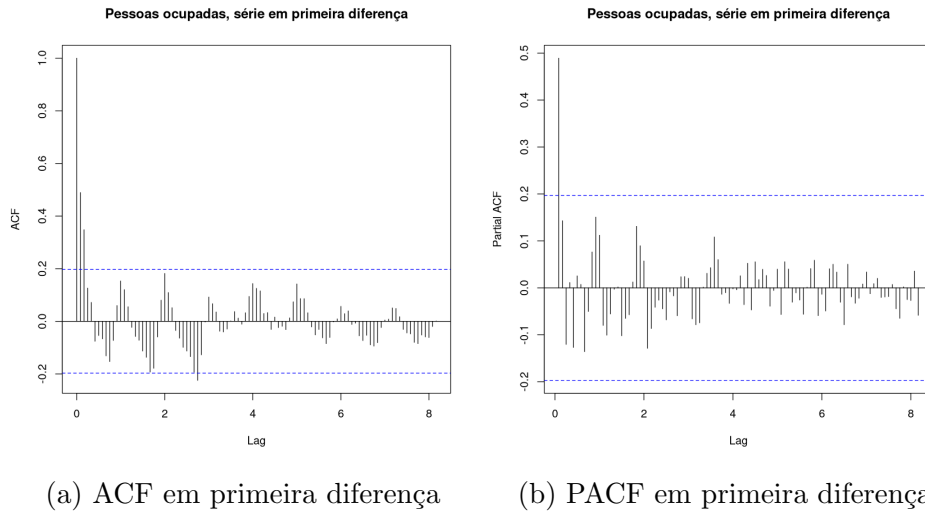


Figura 6: ACF e PACF em primeira diferença

Já a diferenciação sazonal, em contrapartida, envolve subtrair cada observação pelo valor da observação no mesmo período na temporada anterior. A diferenciação sazonal pode ser adequada quando existem padrões sazonais claros nos dados - como é esperado para o caso do desemprego. Na série dessazonalizada, entretanto, só foi possível obter a rejeição da hipótese nula no teste ADF a partir da segunda diferença.

Através da função `auto.arima()` aplicada à série em primeiras diferenças, o modelo

SARIMA(2, 0, 2)(1, 0, 1)[12] foi estimado. Nota-se que ele não possui primeira diferença (já que os dados alimentados na função já estavam em primeira diferença), uma bateria de testes foi realizada com base nas estimações anteriores para que mais dois modelos fossem escolhidos através do critério de BIC, resultando nos três finalistas:

- SARIMA(2, 0, 2)(1, 0, 1)[12] – (1505.636)
- SARIMA(1, 0, 0)(1, 0, 0)[12] – (1506.851)
- SARIMA(2, 0, 2)(2, 0, 1)[12] – (1509.606)

Quando partimos para análise gráfica dos resíduos dos modelos, todos os três parecem ter variância homocedástica, serem normalmente distribuídos e não possuírem autocorrelação (figuras 7, 8, 9 e 10).

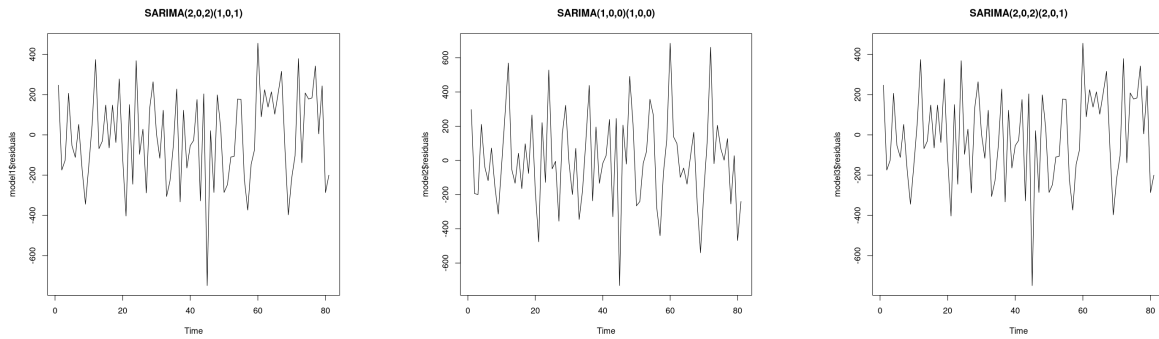


Figura 7: Resíduos dos modelos

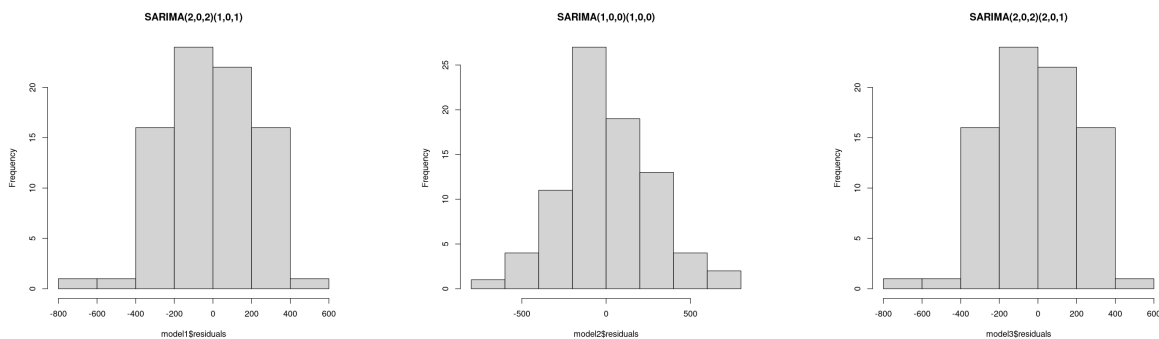


Figura 8: Distribuição dos resíduos

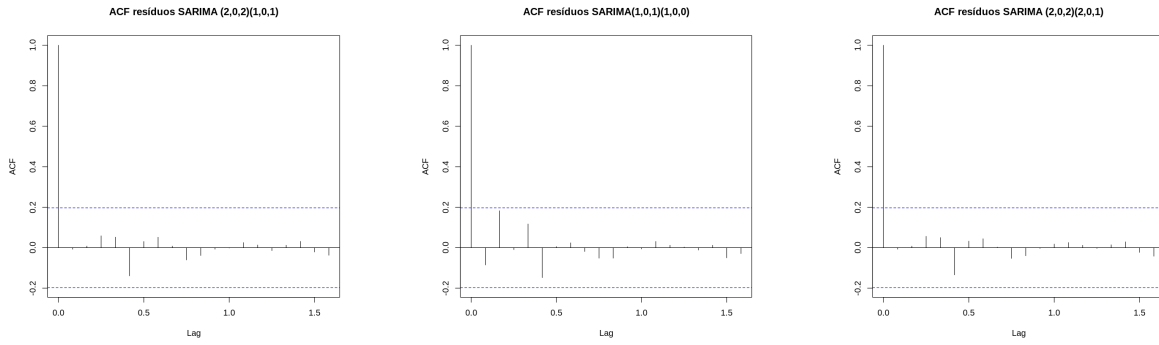


Figura 9: ACFs dos Resíduos dos modelos

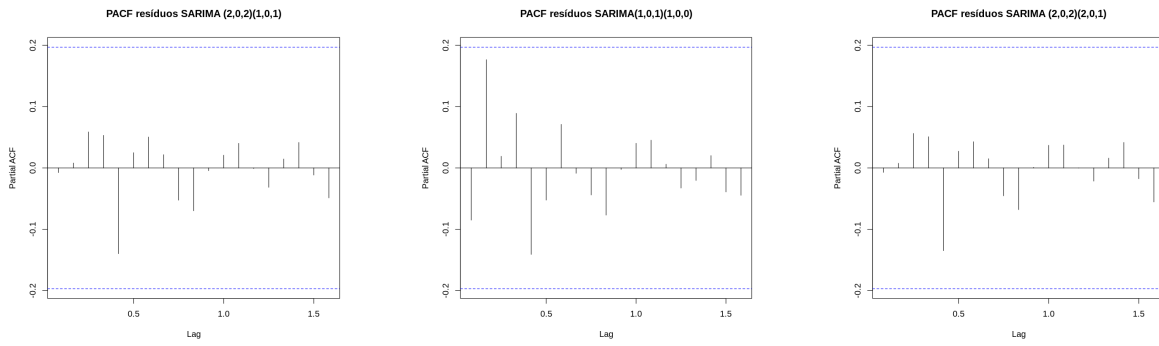


Figura 10: PACFs dos Resíduos dos modelos

No entanto, apenas a análise gráfica não é o suficiente para tais conclusões. Três testes foram usados para confirmar as hipóteses levantadas:

- Ljung-box: verifica a presença de autocorrelação nos resíduos, com H_0 : não há correlação;
- Jarque-Bera: verifica se os resíduos seguem uma distribuição normal, em que H_0 : skewness, kurtosis = 0;
- ARCH-LM: verifica se há heterocedasticidade nos resíduos, em que H_0 : não há um efeito ARCH-LM

	Ljung-box						Jarque-Bera		ARCH-LM	
	4		12		24					
	χ^2	p-valor	χ^2	p-valor	χ^2	p-valor	χ^2	p-valor	χ^2	p-valor
SARIMA(2, 0, 2)(1, 0, 1)[12]	0.66	0.96	3.69	0.99	7.05	1.00	1544.50	0.00	73.32	0.00
SARIMA(1, 0, 0)(1, 0, 0)[12]	5.62	0.23	8.65	0.73	11.96	0.98	2264.00	0.00	78.76	0.00
SARIMA(2, 0, 2)(2, 0, 1)[12]	0.60	0.96	3.37	0.99	6.95	1.00	1566.50	0.00	73.11	0.00

Tabela 1: Resumo das estatísticas

Através da tabela, é possível observar que, enquanto a hipótese nula não é rejeitada para nenhum modelo no teste Ljung-box, i.e., os resíduos não são autocorrelacionados, ela é para os testes de Jarque-Bera e ARCH-LM, indicando que os resíduos não estão normalmente distribuídos, e são heterocedásticos. Apesar dos resultados não serem animadores, é preciso considerar que as observações abarcam o período pandêmico, em que as dinâmicas do mercado de trabalho como conhecíamos foram severamente prejudicadas. Dito isto, a análise será continuada com os modelos previamente estimados.

3.2.1 Forecasting

Quando partirmos para o *forecasting* dos três modelos univariados sugeridos, para 24 meses com intervalo de confiança de 95%, obtemos resultados discrepantes entre os três, como era de se esperar (figura 11).

Realizamos então um teste de acurácia, que mede o quanto os modelos se adequam as observações. Dentre os três modelos, o ARIMA(2, 0, 2)(2, 0, 1)[12] indicou o menor RMSE (387), seguido de forma bem próxima pelo modelo ARIMA(2, 0, 2)(1, 0, 1)[12] (sugerido pela função *auto.arima()*) com RMSE de 389 e, por último, com menor acurácia e RMSE maior de 441 foi o modelo ARIMA(1, 0, 0)(1, 0, 0)[12]. É importante notar que o valor do RMSE usa o valor das observações como base, por isso são tão altos.

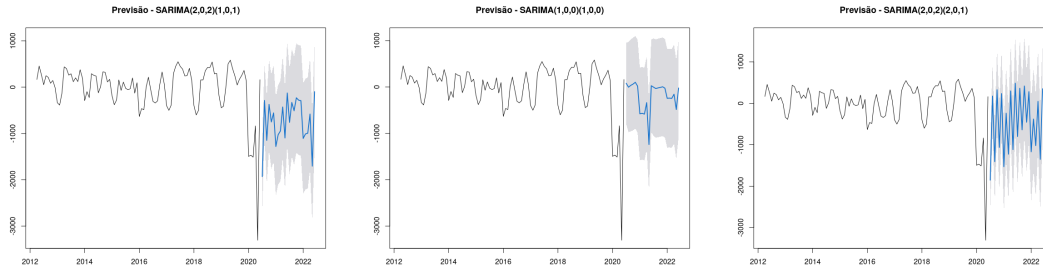


Figura 11: Previsão

Quando realizamos o teste de acurácia fora da amostra, tomando o período Mar/12 - Dez/18 como o período de treino, e o restante (Jan/19 - Jun/20), obtemos RMSEs muito mais elevados do que os anteriores, como pode ser comparado em: 2.

	Dentro da amostra	Fora da amostra
SARIMA(2, 0, 2)(1, 0, 1)	389	837
SARIMA(1, 0, 0)(1, 0, 0)	441	819
SARIMA(2, 0, 2)(2, 0, 1)	387	822

Tabela 2: Teste de acurácia

A causa destes resultados é a mesma da heterocedasticidade nos resíduos: o COVID-19. Através dessa comparação conseguimos observar o quão forte foi o choque causado pela pandemia, o qual praticamente inutilizou os modelos que eram usados até então, dificultando projeções e o próprio entendimento das dinâmicas do mercado de trabalho. E, por óbvio, apesar da força do choque, a acurácia dos modelos estimados com todas as observações é maior por levarem-no em conta a priori.

Através dos valores estimados com o modelo SARIMA(1, 0, 0)(1, 0, 0), da série original e das observações sobre o valor do salário mínimo real, conseguimos ter uma ideia da perda causada pela pandemia devido à queda no número de ocupações: cerca de R\$28,594,135.09.

4 Tudo é endógeno: modelos multivariados

4.1 Metodologia

A análise dos modelos multivariados será conduzida por meio de um Modelo Autorregressivo Vetorial (VAR), uma abordagem metodológica consolidada que incorpora informações passadas e contemporâneas para analisar variáveis, tratando-as todas como endógenas. Esta metodologia é fundamentada no trabalho seminal de Sims, 1980.

De maneira geral, o modelo autorregressivo de ordem p para n variáveis endógenas X_t pode ser expresso como segue (Equação 1):

$$Y_t = \sum_{i=1}^p \phi_i Y_{t-i} + \Gamma_t + v, \quad t = 1, 2, 3, \dots, p \quad (2)$$

Nesta equação, v é um vetor de ruído branco de erros de previsão, ϕ_i é a matriz de coeficientes, e Γ_t contém os termos determinísticos, como constante, tendência, dummies de sazonalidade e intervenção. O vetor de dados Y_t incorpora as variáveis contemporâneas do modelo.

Além disso, é possível representar um VAR estável na forma de um vetor de médias móveis (Vector Moving Average - VMA):

$$Y = \prod_t + \sum_{i=0}^{\infty} \Lambda^i v_{t-i} \quad (3)$$

Onde L é o operador de defasagem, então $\prod_t = (I_n - \sum_{i=1}^p \phi_i L^i)^{-1} \Gamma_t$ e $\sum_{i=0}^{\infty} \Lambda^i = (I_n - \sum_{i=1}^p \phi_i L^i)^{-1}$ são derivados da condição de estabilidade, com as raízes das equações situando-se fora do círculo unitário do domínio complexo. A análise da dinâmica entre as variáveis endógenas do vetor Y_t é realizada por meio de Funções de Resposta a Impulsos (FRI), partindo da representação VMA.

4.1.1 Caracterização dos dados

Para a realização da estimação do modelo VAR, iniciou-se com a verificação da sazonalidade das séries utilizando o pacote ‘seas’ do software R, e o subsequente remanejamento do componente sazonal de todas as séries analisadas.

Com o intuito de proporcionar uma análise mais robusta dos resultados, optou-se por modelar a variável de pessoas ocupadas no Brasil de duas maneiras distintas. Tradicionalmente, a série de ocupação é analisada em termos de taxa, representando o total de pessoas ocupadas sobre a população total na força de trabalho. Neste contexto, o presente trabalho busca abordar duas análises para explorar os resultados com maior robustez, considerando tanto a série em nível com o total da população ocupada quanto a série em taxa de pessoas empregadas.

Posteriormente à remoção da sazonalidade e à escolha da modelagem da variável de interesse, procedeu-se à verificação da estacionaridade das séries por meio dos testes de *Augmented Dickey-Fuller* (ADF). As estatísticas de teste, número de lags e os tipos de testes (com tendência, drift ou sem nenhuma das duas) são apresentados no Anexo 1 do trabalho.

Quanto à ordem de integração da série da taxa de pessoas ocupadas, destaca-se o estudo de Mazali e Divino, 2010, que, após uma bateria de testes de raiz unitária, concluiu que as séries brasileiras de inflação e desemprego são estacionárias. Portanto, descartou-se a possibilidade de cointegração dessas séries. Como a taxa de desemprego é parte da taxa de pessoas empregadas, argumentou-se que também é possível que a taxa de pessoas empregadas seja estacionária. Além disso, reproduziram-se os testes de Lee e Strazicich, 2003 para o conjunto informacional deste estudo, indicando a possibilidade de um breakpoint em janeiro de 2016 e dezembro de 2018. O resultado de 2016 coincide com o período de recessão econômica no Brasil, refletindo um aumento significativo na taxa de desemprego e uma queda expressiva na taxa de pessoas ocupadas¹. Já o segundo período está associado à recuperação econômica, com a taxa de pessoas ocupadas apresentando aumento. Portanto, ao adicionar essas duas quebras nos testes, concluiu-se que a série da taxa de pessoas empregadas é estacionária no período em consideração, destacando-se o teste t, que obteve um valor de -5, sendo significativo a 10%.

Em seguida, determinaram-se as variáveis selecionadas para a estimação do VAR, utilizando o teste de causalidade de Granger (Granger, 1969). Este teste é frequentemente empregado para avaliar se uma variável X causa causalmente uma variável Y , sendo a hipótese nula (H_0) que os valores passados de X não causam Y . A estatística

¹“Brasil enfrenta pior crise já registrada poucos anos após um boom econômico”. Disponível em: <https://g1.globo.com/economia/noticia/brasil-enfrenta-pior-crise-ja-registrada-poucos-anos-apos-um-boom-economico.ghtml>. Acesso em: 26/11/2023

de teste baseia-se na melhoria da previsão de Y ao incluir os valores passados de X no modelo. O número de defasagens escolhido foi de 2 para obter um modelo parcimonioso. Adicionalmente, estimou-se como Y a série da população ocupada na segunda diferença e a taxa da população ocupada.

Com base nos resultados apresentados na Tabela 1, optou-se por incluir no modelo as variáveis de focus e PIB. A fim de enriquecer a análise e alinhar-se com a literatura recente sobre a relação entre emprego e outras variáveis macroeconômicas, adicionaram-se ao modelo as variáveis de inflação brasileira (IPCA) e a taxa de câmbio efetiva real. Essa adição visa explorar possíveis relações entre essas variáveis e a taxa de pessoas empregadas, complementando a análise com informações relevantes.

Tabela 3: Teste de Causalidade de Granger

height	Modelo 1	Modelo 2
	valor-p	valor-p
Pessoas de 14 anos ou mais	0.05	0.136
Pessoas na força de trabalho	0.94	0.1361
Pessoas ocupadas	-	-
Pessoas desocupadas	0.48	0.2118
Salário mínimo	0.83	0.6448
IPCA	0.18	0.04813
IBC-BR	2.619e-11	7.775e-08
PMC	2.511e-07	0.0001892
Expectativa média de inflação	0.001663	0.0004446
PIB	3.786e-10	4.475e-10
Taxa de câmbio - efetiva real	0.2796	0.6037

[†] Fonte: Elaboração própria.

[†] Nota: O modelo 1 refere-se ao modelo que utiliza a taxa de pessoas ocupadas sobre o número de pessoas na força de trabalho, utilizando apenas o percentual em $I(0)$. O modelo 2 refere-se ao modelo que utiliza a taxa de pessoas ocupadas utilizando $I(2)$ para diferenciação.

Por fim, procedeu-se à determinação do número de defasagens do modelo VAR, inicialmente considerando 12 defasagens. Para o Modelo 1, os resultados, de acordo com o Critério de Informação de Schwarz, indicaram uma preferência por uma defasagem. Entretanto, também foi estimado um modelo com 2 defasagens para permitir a comparação entre os modelos e identificar o mais adequado para explicar e prever a variável de interesse. Dessa forma, no total, foram estimados 4 modelos VAR, que serão de-

signados como Modelo 1, correspondendo ao modelo que utiliza como y a variável do percentual de pessoas ocupadas em relação ao total de pessoas na força de trabalho, com estimativas para as ordens 1 e 2 do VAR. Além disso, o Modelo 2 foi formulado considerando a variável y como o total da população ocupada na segunda diferença logarítmica e estimando o VAR para as ordens 1 e 2.

Essa abordagem permitirá uma compreensão mais abrangente e robusta da dinâmica das variáveis em estudo, auxiliando na escolha do modelo mais apropriado para análise e previsão da variável de interesse.

4.2 Resultados

4.2.1 Estimações e diagnósticos

Optou-se por utilizar um VAR com uma defasagem, incorporando uma constante, tendência, *drift*, e *dummies* de sazonalidade. A introdução da tendência linear e do *drift* visa controlar a trajetória decididamente decrescente da taxa de empregados durante a maior parte do período amostral.

A Tabela 2 apresenta os valores de λ para o VAR com uma defasagem e para o VAR com duas defasagens, considerando os dois tipos de transformações utilizadas na população ocupada. Observa-se que a todos os valores de λ possui magnitude inferior a 1, sugerindo a estabilidade do modelo. Além disso, no caso do λ maior que 1, Cavalcanti e Silva, 2010 propõem duas hipóteses para a possível instabilidade paramétrica. Primeiramente, afirmam que há a possibilidade de esses testes rejeitarem excessivamente a hipótese de estabilidade estrutural devido às limitações no tamanho da amostra. Em segundo lugar, argumentam que, sob a estabilidade dos parâmetros do VAR ao longo da amostra, as estatísticas que testam a igualdade da matriz de covariância tendem a ser significativas quando grande parte da volatilidade das séries está concentrada em um subperíodo amostral.

A Tabela 3 exibe os resultados do teste de heterocedasticidade. Ao analisar as estatísticas apresentadas, notamos que os valores-p das séries são estatisticamente inferiores a 5%, permitindo-nos rejeitar a hipótese nula (H_0). Portanto, concluímos que os erros em nosso modelo exibem homocedasticidade.

A Tabela 4 apresenta os resultados dos testes de correlação serial no VAR com 1 defasagem para ambos os modelos. Observa-se que, para h horizontes, temos evidências

Tabela 4: Autovalores dos modelos estimados

	roots	$\lambda - 1$	$\lambda - 2$	$\lambda - 3$	$\lambda - 4$	$\lambda - 5$	$\lambda - 6$	$\lambda - 7$	$\lambda - 8$	$\lambda - 9$	$\lambda - 10$
Modelo 1	1 defasagem	0.9889207	0.3877972	0.3423452	0.2169025	0.0986560					
	2 defasagens	0.9252426	0.6957474	0.6820284	0.6820284	0.5018691	0.5018691	0.4021387	0.4021387	0.3411139	0.3096350
Modelo 2	1 defasagem	0.8212033	0.4601361	0.2289204	0.1893636	0.1893636					
	2 defasagens	0.8210067	0.7197950	0.6402341	0.6402341	0.5419287	0.5419287	0.3667353	0.3667353	0.2850103	0.2850103

[†] Fonte: Elaboração própria.

[†] Nota: O Modelo 1 refere-se ao modelo que utiliza a taxa de pessoas ocupadas sobre o número de pessoas na força de trabalho, considerando apenas o percentual em $I(0)$. O Modelo 2 refere-se ao modelo que utiliza a taxa de pessoas ocupadas, adotando $I(2)$ para diferenciação. λ representa o inverso das raízes características do polinômio.

Tabela 5: Teste de heretocedasticidade

	Modelo 1	Modelo 2
	valor-p	valor-p
1 defasagem	7.655e-08	4.288e-05
2 defasagem	2.108e-05	4.228e-05

que permitem rejeitar a hipótese nula, com 99% de nível de confiança, de que nossos erros não exibem autocorrelação serial. Além disso, no Anexo 2, encontram-se os gráficos dos resíduos das variáveis utilizadas nos quatro modelos.

Tabela 6: Testes de correlação serial

	Modelo 1	Modelo 2
	valor-p	valor-p
1 defasagem	1.39e-08	0.0001751
2 defasagem	6.141e-08	2.211e-05

A Tabela 5 apresenta os resultados dos testes de não normalidade do tipo Jarque-Bera aplicados ao VAR. Observa-se que sempre se aceita a normalidade dos resíduos. No Anexo 2, foram incluídos os gráficos dos erros ao longo do tempo.

4.2.2 Simulações: Choques da população ocupada na inflação, na expectativa de inflação, na taxa de cambio e no PIB

Esta subseção apresenta os resultados das simulações, iniciando com choques de 1 desvio-padrão nas variáveis endógenas. A abordagem adotada baseou-se na fatoração de Cholesky, que, na prática, implica uma estrutura recursiva (matriz B é triangular inferior) conhecida como Wold causal chain. Destaca-se que essa decomposição é única

Tabela 7: Testes de normalidade

	Modelo 1	Modelo 2
	valor-p	valor-p
1 defasagem	9.992e-16	4.585e-06
2 defasagem	2.2e-16	2.2e-16

apenas para um determinado ordenamento das variáveis no vetor Y_t .

Adotou-se o ordenamento mais plausível segundo a teoria, em que

$$Y = (\Delta\theta_t, U_t, \Delta E\pi_{t+1}, \pi_t, \text{PIB}) \quad (4)$$

e para o segundo modelo, apenas alterando o U_t para ΔU_t .

As simulações foram conduzidas considerando um horizonte de 24 meses à frente, período no qual a maior parte dos choques já se dissipou ou a série se estabilizou em outro valor, restando pouca dinâmica.

O Gráfico 1 ilustra os resultados a uma inovação na taxa de pessoas empregadas (de 1 desvio padrão = 0.53 p.p). Nota-se que, mesmo controlando por um termo de tendência, a taxa de pessoas empregadas permanece a mais persistente de todas as séries, com um choque médio perdurando por quase 2 anos.

As expectativas de inflação foram significativamente afetadas apenas nos primeiros horizontes, enquanto a inflação propriamente dita e as expectativas de inflação nos demais horizontes não apresentaram efeitos significativos após o impulso. Contrariamente à literatura da Curva de Phillips, o sinal positivo do choque sobre variáveis de consumo sugere que um aumento na taxa de pessoas ocupadas leva a um aumento na taxa de inflação, indicando um possível aumento na demanda de bens e serviços devido ao aumento da renda nacional.

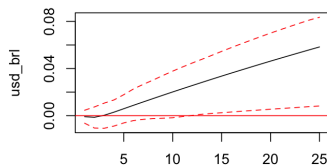
Os choques sobre as taxas de câmbio parecem persistir por um longo período, evidenciando um efeito positivo ao longo dos 2 anos de análise. Este resultado está em conformidade com a literatura que sugere um aumento na demanda de moeda devido ao aumento do consumo da população, levando a uma depreciação na taxa de câmbio interna.

Entretanto, não observamos o mesmo efeito sobre o PIB, onde os choques não são estatisticamente significativos. Pode-se argumentar que um aumento na taxa de pessoas

ocupadas não teria um efeito direto sobre o PIB, sendo este mais influenciado por outras variáveis de demanda. Considerando a fórmula de demanda agregada proposta por Keynes, o PIB seria afetado diretamente pelo consumo da população, investimentos privados, consumo do governo e exportações líquidas.

Figura 12: Respostas a um choque de 1 desvio-padrão na taxa de emprego, VAR(1)

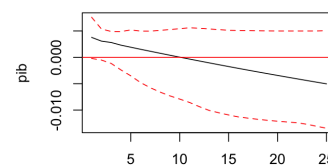
Orthogonal Impulse Response from tx_empreg (cumulative)



95 % Bootstrap CI, 100 runs

(a) IRF Câmbio

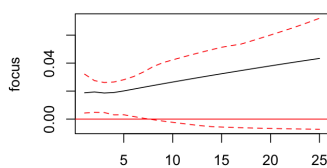
Orthogonal Impulse Response from tx_empreg (cumulative)



95 % Bootstrap CI, 100 runs

(b) IRF PIB

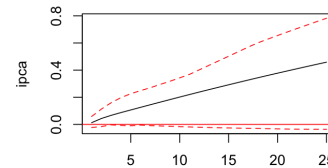
Orthogonal Impulse Response from tx_empreg (cumulative)



95 % Bootstrap CI, 100 runs

(c) IRFs Expectativa de Inflação

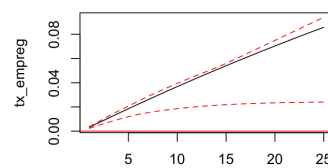
Orthogonal Impulse Response from tx_empreg (cumulative)



95 % Bootstrap CI, 100 runs

(d) IRFs IPCA

Orthogonal Impulse Response from tx_empreg (cumulative)



95 % Bootstrap CI, 100 runs

(e) IRFs Taxa de Emprego

Ao analisarmos a decomposição da variância na Tabela 6, é evidente que a taxa de ocupação apresenta a maior contribuição, o que é natural, uma vez que estamos medindo a variação em relação a ela mesma. Entretanto, à medida que avançamos no tempo, a taxa de ocupação continua sendo a principal fonte de variação, mas variáveis como o PIB e as expectativas de inflação começam a ganhar importância. Isso sugere, do ponto de vista das expectativas, que as percepções e previsões dos agentes econômicos desempenham um papel crucial na dinâmica do mercado de trabalho, possivelmente influenciando as decisões de investimento e contratação.

Ademais, o desempenho econômico, medido pelo PIB, também exerce influência, uma vez que um melhor desempenho econômico tende a resultar em maiores investimentos, conforme refletido pela fórmula da demanda agregada. Isso, por conseguinte, pode influenciar no aumento da contratação de pessoas. O Gráfico 2 exibe os resultados de um choque de um desvio padrão na taxa de pessoas ocupadas, representando a variação percentual das pessoas ocupadas sem normalização pela população na força de trabalho. Observa-se um choque positivo e significativo que perdura por todo o período analisado sobre a própria taxa de pessoas ocupadas.

Apesar de as demais variáveis apresentarem sinais esperados pela literatura, como choques positivos nas variáveis de consumo e na taxa de câmbio, elas não são estatisticamente significativas. Além disso, para a taxa de atividade e o PIB, que apresentou um sinal oposto ao Gráfico 1, não foi significativo em nenhum horizonte observado.

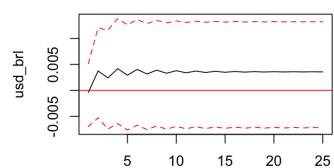
Tabela 8: Decomposição da variância – Taxa de empregados, VAR(1)

Tx. Empreg.	ipca	Expectativa de inflação	pib	Taxa de câmbio
100.00000	0.0000000	0.000000	0.000000	0.000000
81.93717	1.6021154	4.434940	10.730730	1.295043
79.86648	1.5910347	6.791249	9.570497	2.180738
79.00633	1.2536511	7.726490	9.491250	2.522279
78.71689	1.0146829	8.198230	9.316521	2.753674
78.52240	0.8519723	8.472074	9.237838	2.915711
78.38531	0.7375800	8.661034	9.176930	3.039149
78.27810	0.6531981	8.800879	9.134451	3.133375
78.19430	0.5885104	8.909049	9.101276	3.206864
78.12738	0.5373382	8.994885	9.075117	3.265275
78.07307	0.4958541	9.064530	9.053844	3.312703
78.02815	0.4615532	9.122108	9.036260	3.351932
77.99040	0.4327272	9.170489	9.021480	3.384906
77.95824	0.4081690	9.211705	9.008889	3.413002
77.93051	0.3870015	9.247229	8.998036	3.437220
77.90638	0.3685721	9.278159	8.988587	3.458306
77.88518	0.3523856	9.305324	8.980288	3.476826
77.86641	0.3380595	9.329367	8.972943	3.493217
77.84969	0.3252935	9.350792	8.966398	3.507823
77.83470	0.3138485	9.370000	8.960530	3.520918
77.82119	0.3035321	9.387313	8.955241	3.532722
77.80895	0.2941873	9.402996	8.950450	3.543414
77.79782	0.2856851	9.417265	8.946091	3.553142
77.78764	0.2779183	9.430300	8.942109	3.562028

Fonte: Elaboração própria.

Figura 13: Respostas a um choque de 1 desvio-padrão na Pop. Ocupada, VAR(1)

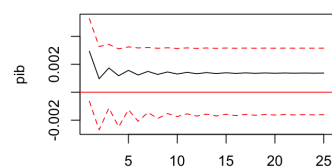
Orthogonal Impulse Response from empreg (cumulative)



95 % Bootstrap CI, 100 runs

(a) IRF Câmbio

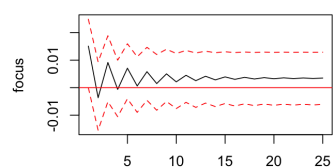
Orthogonal Impulse Response from empreg (cumulative)



95 % Bootstrap CI, 100 runs

(b) IRF PIB

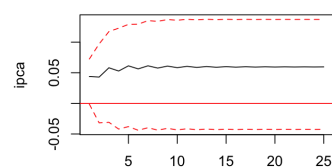
Orthogonal Impulse Response from empreg (cumulative)



95 % Bootstrap CI, 100 runs

(c) IRFs Expectativa de Inflação

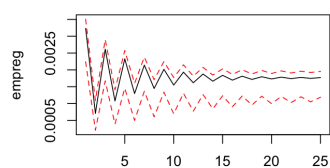
Orthogonal Impulse Response from empreg (cumulative)



95 % Bootstrap CI, 100 runs

(d) IRFs IPCA

Orthogonal Impulse Response from empreg (cumulative)



95 % Bootstrap CI, 100 runs

(e) IRFs Taxa de Emprego

Na decomposição da variância sobre a diferença percentual das pessoas ocupadas, os resultados são semelhantes à Tabela 7 em comparação com a Tabela 6, entretanto, a contribuição do PIB sobre a taxa de pessoas ocupadas torna-se mais significativa à medida que avançamos nos períodos analisados.

Tabela 9: Decomposição da variância – População ocupada, VAR(1)

TPop. Empreg.	ipca	Expectativa de inflação	pib	Taxa de câmbio
100.00000	0.000000	0.000000	0.00000	0.00000000
75.41311	1.633532	5.847635	17.07786	0.02785439
59.90645	2.839955	6.443940	30.78660	0.02305799
53.49583	2.628722	6.269692	37.57729	0.02846949
50.20927	2.633042	6.206620	40.91569	0.03538315
48.39613	2.596568	6.160703	42.80937	0.03722739
47.30885	2.585429	6.133247	43.93295	0.03952483
46.63581	2.574536	6.116071	44.63307	0.04051498
46.20447	2.568973	6.105044	45.08020	0.04131488
45.92413	2.564819	6.097870	45.37141	0.04177455
45.73932	2.562282	6.093142	45.56316	0.04210013
45.61669	2.560522	6.090003	45.69047	0.04230774
45.53485	2.559376	6.087909	45.77542	0.04244949
45.48005	2.558598	6.086506	45.83231	0.04254319
45.44327	2.558080	6.085565	45.87048	0.04260654
45.41854	2.557730	6.084932	45.89615	0.04264895
45.40190	2.557495	6.084507	45.91342	0.04267756
45.39070	2.557337	6.084220	45.92505	0.04269679
45.38315	2.557230	6.084027	45.93288	0.04270976
45.37806	2.557158	6.083897	45.93816	0.04271850
45.37463	2.557110	6.083809	45.94172	0.04272439
45.37232	2.557077	6.083750	45.94412	0.04272836
45.37076	2.557055	6.083710	45.94574	0.04273104
45.36971	2.557040	6.083683	45.94683	0.04273285

Fonte: Elaboração própria.

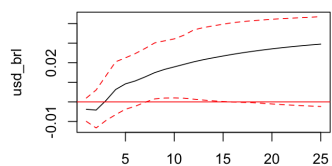
O Gráfico 3 apresenta as respostas a uma inovação na taxa de empregados para o modelo VAR(2). Observa-se um choque estatisticamente significativo que perdura por toda a análise. Diferentemente do modelo VAR(1) para a taxa de empregados, obtivemos resultados significativos que contradizem a análise sobre variáveis de consumo. No curto prazo, um choque na taxa de empregados gera um efeito positivo sobre as expectativas de inflação, devido à expectativa de que um aumento na taxa de empregados levará a um aumento do consumo devido ao aumento das pessoas empregadas. A longo

prazo, essa expectativa parece se realizar, pois nos horizontes mais longos, a inovação na taxa de empregados tem um efeito significativo sobre a taxa de inflação observada, indicando que o aumento das pessoas empregadas leva a um aumento da renda, gerando um aumento do consumo, conforme descrito na literatura sobre a relação entre emprego e inflação.

Ademais, também é possível observar um efeito significativo na taxa de câmbio apenas entre o oitavo e décimo quarto mês, indicando os efeitos mencionados da inovação sobre a taxa de câmbio. Novamente, não foram observados efeitos significativos sobre o indicador de atividade, o PIB, corroborando a análise anterior.

Figura 14: Respostas a um choque de 1 desvio-padrão na taxa de emprego, VAR(2)

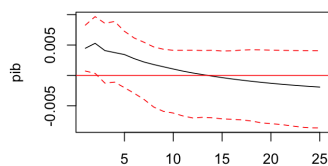
Orthogonal Impulse Response from tx_emploi (cumulative)



95 % Bootstrap CI, 100 runs

(a) IRF Câmbio

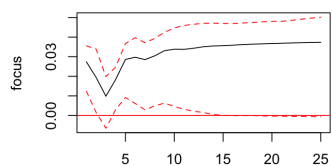
Orthogonal Impulse Response from tx_emploi (cumulative)



95 % Bootstrap CI, 100 runs

(b) IRF PIB

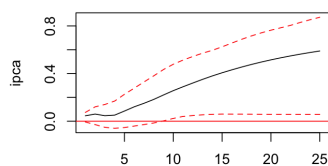
Orthogonal Impulse Response from tx_emploi (cumulative)



95 % Bootstrap CI, 100 runs

(c) IRFs Expectativa de Inflação

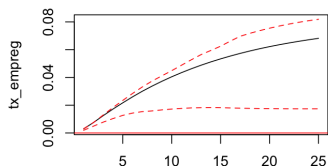
Orthogonal Impulse Response from tx_emploi (cumulative)



95 % Bootstrap CI, 100 runs

(d) IRFs IPCA

Orthogonal Impulse Response from tx_emploi (cumulative)

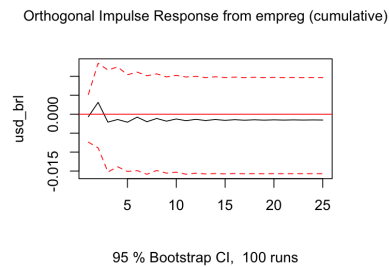


95 % Bootstrap CI, 100 runs

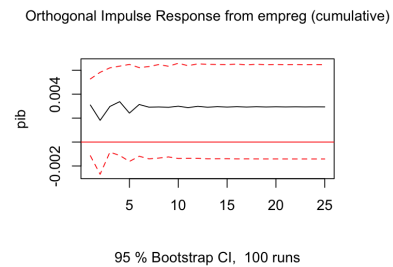
(e) IRFs Taxa de Emprego

O Gráfico 4 apresenta os resultados das Funções de Resposta ao Impulso (FRI) de um choque sobre a população ocupada para o modelo VAR(2). Mais uma vez, observa-se um choque positivo sobre a própria variável de população ocupada que persiste ao longo do tempo. Entretanto, para as demais variáveis, novamente encontramos resultados não significativos durante o período analisado.

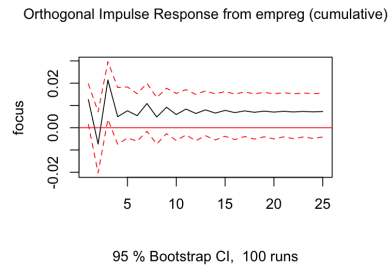
Figura 15: Respostas a um choque de 1 desvio-padrão Pop. Ocupada, VAR(2)



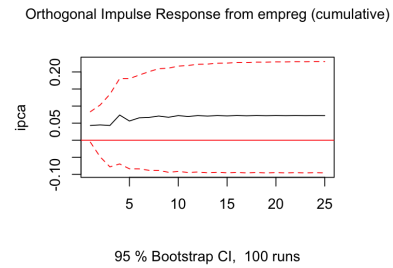
(a) IRF Câmbio



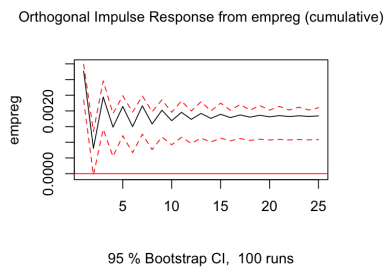
(b) IRF PIB



(c) IRFs Expectativa de Inflação



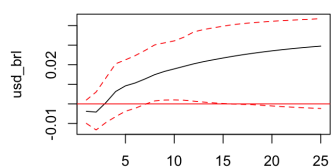
(d) IRFs IPCA



(e) IRFs Taxa de Emprego

Figura 16: Respostas a um choque de 1 desvio-padrão na taxa de emprego, VAR(2)

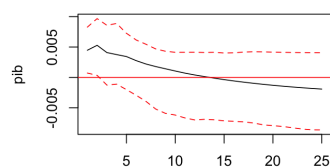
Orthogonal Impulse Response from tx_employment (cumulative)



95 % Bootstrap CI, 100 runs

(a) IRF Câmbio

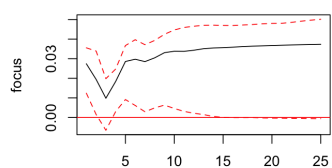
Orthogonal Impulse Response from tx_employment (cumulative)



95 % Bootstrap CI, 100 runs

(b) IRF PIB

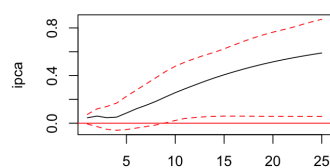
Orthogonal Impulse Response from tx_employment (cumulative)



95 % Bootstrap CI, 100 runs

(c) IRFs Expectativa de Inflação

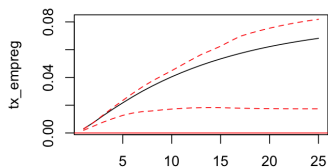
Orthogonal Impulse Response from tx_employment (cumulative)



95 % Bootstrap CI, 100 runs

(d) IRFs IPCA

Orthogonal Impulse Response from tx_employment (cumulative)



95 % Bootstrap CI, 100 runs

(e) IRFs Taxa de Emprego

As tabelas 8 e 9 apresentam as decomposições da variância para os demais modelos estimados.

Tabela 10: Decomposição da variância – Taxa de empregados, VAR(2)

Tx. Empreg.	ipca	Expectativa de inflação	pib	Taxa de câmbio
100.00000	0.000000	0.000000	0.00000	0.00000000
75.41311	1.633532	5.847635	17.07786	0.02785439
59.90645	2.839955	6.443940	30.78660	0.02305799
53.49583	2.628722	6.269692	37.57729	0.02846949
50.20927	2.633042	6.206620	40.91569	0.03538315
48.39613	2.596568	6.160703	42.80937	0.03722739
47.30885	2.585429	6.133247	43.93295	0.03952483
46.63581	2.574536	6.116071	44.63307	0.04051498
46.20447	2.568973	6.105044	45.08020	0.04131488
45.92413	2.564819	6.097870	45.37141	0.04177455
45.73932	2.562282	6.093142	45.56316	0.04210013
45.61669	2.560522	6.090003	45.69047	0.04230774
45.53485	2.559376	6.087909	45.77542	0.04244949
45.48005	2.558598	6.086506	45.83231	0.04254319
45.44327	2.558080	6.085565	45.87048	0.04260654
45.41854	2.557730	6.084932	45.89615	0.04264895
45.40190	2.557495	6.084507	45.91342	0.04267756
45.39070	2.557337	6.084220	45.92505	0.04269679
45.38315	2.557230	6.084027	45.93288	0.04270976
45.37806	2.557158	6.083897	45.93816	0.04271850
45.37463	2.557110	6.083809	45.94172	0.04272439
45.37232	2.557077	6.083750	45.94412	0.04272836
45.37076	2.557055	6.083710	45.94574	0.04273104
45.36971	2.557040	6.083683	45.94683	0.04273285

Fonte: Elaboração própria.

Tabela 11: Decomposição da variância – População ocupada, VAR(2)

Pop. Empreg.	ipca	Expectativa de inflação	pib	Taxa de câmbio
100.00000	0.000000	0.000000	0.00000	0.000000000
74.42135	2.793161	5.172080	17.61199	0.001419546
55.10765	7.052450	6.425301	31.09474	0.319868155
47.36138	8.141595	5.632519	38.58976	0.274744193
45.22341	8.771750	5.268473	40.47891	0.257454457
45.05384	9.172086	5.153484	40.36898	0.251614867
45.22022	9.488889	5.077526	39.96564	0.247722194
45.26979	9.613657	5.045325	39.82637	0.244856244
45.09991	9.682082	5.017343	39.95641	0.244248635
44.93195	9.727744	4.989375	40.10808	0.242850512
44.84181	9.777933	4.968751	40.16987	0.241632122
44.80888	9.810041	4.956181	40.18397	0.240927057
44.79359	9.831701	4.948424	40.18577	0.240512793
44.78313	9.843962	4.943617	40.18904	0.240249812
44.77279	9.852642	4.940357	40.19415	0.240060181
44.76455	9.858131	4.938042	40.19936	0.239924926
44.75874	9.862205	4.936396	40.20283	0.239831736
44.75531	9.864884	4.935293	40.20474	0.239769465
44.75321	9.866780	4.934564	40.20572	0.239728222
44.75186	9.867985	4.934085	40.20637	0.239701018
44.75089	9.868814	4.933763	40.20685	0.239682739
44.75020	9.869353	4.933544	40.20723	0.239670290
44.74973	9.869730	4.933395	40.20749	0.239661789
44.74941	9.869980	4.933294	40.20766	0.239656055

Fonte: Elaboração própria.

4.2.3 Backtesting e resultados dentro e fora da amostra

A Tabela 9 apresenta os resultados dos modelos tanto dentro quanto fora da amostra. Observa-se que, com base nos resultados apresentados, o modelo que utiliza a taxa de empregados, normalizada pela população da força de trabalho, com um VAR(2) destacou-se como o mais eficaz dentro da amostra, enquanto o modelo que incorpora a diferença da população ocupada em um VAR(1) mostrou-se superior fora da amostra.

Além disso, em todos os modelos, observou-se uma significativa deterioração ao realizar estimativas fora da amostra. Tal fenômeno pode ser justificado pela ocorrência da pandemia em 2020, período no qual a taxa de empregados sofreu uma queda abrupta, inicialmente não explicada por fundamentos econômicos. Em virtude desse cenário, as modelagens econométricas foram adversamente afetadas, uma vez que o declínio não foi causado por um choque econômico convencional, mas sim por um impacto de saúde que reverberou na economia.

Atualmente, a literatura debate intensamente sobre a abordagem adequada para tratar esses modelos durante esse período. Em muitos casos, os coeficientes β que anteriormente eram eficazes durante períodos normais revelaram variações substanciais, distorcendo seus resultados. Algumas abordagens sugerem excluir esse período da amostra Lenza e Primiceri, 2020 , enquanto outras propõem uma modelagem mais complexa no VAR, permitindo que os erros possuam uma distribuição com caudas mais espessas do que a distribuição gaussiana Bobeica e Hartwig, 2021.

Dessa forma, considerando que é recente a ocorrência desse choque e que foi a primeira vez que um evento externo de tal magnitude impactou a economia, há uma discussão em ascensão sobre aprimoramentos nos modelos para melhorar a estimação desse período e, conseqüentemente, otimizar as previsões para esse contexto.

		Modelo 1		Modelo 2	
		Dentro da amostra	Fora da amostra	Dentro da amostra	Fora da amostra
1 defasagem	RMSE	0.0029	0.020	0.0029	0.012
	MAE	0.0022	0.013	0.0021	0.0055
2 defasagens	RMSE	0.0026	0.020	0.0026	0.872
	MAE	0.0019	0.013	0.0020	0.872

Fonte: Elaboração própria.

Nota: O modelo 1 refere-se ao modelo que utiliza a taxa de pessoas ocupadas sobre o número de pessoas na força de trabalho, utilizando apenas o percentual em I(0). O

modelo 2 refere-se ao modelo que utiliza a taxa de pessoas ocupadas utilizando $I(2)$ para diferenciação. O modelo dentro da amostra foi estimado de janeiro de 2012 até junho de 2020. Enquanto o modelo fora da amostra foi estimado até dezembro de 2018 e a previsão foi de janeiro de 2019 até junho de 2020.

5 Conclusão

O presente trabalho explorou uma vasta gama de metodologias, testes e procedimentos econométricos, com o objetivo de modelar e analisar variáveis relacionadas ao nível do emprego no Brasil.

Na parte I, valendo-nos de modelos univariados, pudemos comparar três modelos SARIMA. Notadamente, de acordo com as métricas da função *accuracy()*, obtivemos que o desempenho no *out-of-sample* de um modelo SARIMA(1,0,0)(1,0,0)[12] foi levemente superior ao modelo identificado pela função *auto.arima()*, que indicou um SARIMA(2,0,2)(1,0,1)[12] como mais adequado. Evidentemente que a métrica fora da amostra não deve ser a única avaliada, visto que está sujeita a vieses, mas o resultado obtido é interessante.

Na parte II a investigação se voltou aos modelos multivariados. Na análise do *backtest* fora da amostra, o modelo que obteve a melhor performance foi um VAR(1) adotando $I(2)$ para diferenciação. Esse também é um resultado interessante, visto que a primeira diferenciação, *i.e.* $I(1)$, costuma ser o procedimento mais frequentemente empregado em séries econômicas como as observadas.

Como discutido anteriormente, esse é um assunto destacadamente relevante para a avaliação de políticas públicas, bem como para a análise e previsão de variáveis de enorme importância para a sociedade como um todo. Para além do aspecto socioeconômico do desemprego, a taxa de desocupação é um dos indicadores mais relevantes para a condução da política monetária, estando fortemente associado a alguns núcleos específicos da inflação, como a dos serviços subjacentes e serviços intensivos em trabalho. Modelar e prever de forma efetiva esse dado, portanto, é um assunto de extrema importância, dado que contribui não só para uma compreensão mais aprofundada da dinâmica do mercado de trabalho, mas também para o embasamento de decisões cruciais no âmbito econômico e financeiro.

A Resultados dos testes de raiz unitária

Tabela 12: Testes de raiz unitária e estacionaridade

Variáveis	Escolha	N. Lags	tcode	Estatísticas do Teste		
				1	2	3
População de 14 anos ou mais	Trend	1	2	-7.8476	20.5401	30.7988
Pessoas na força de trabalho	Trend	3	2	-6.-216	12.1-75	18.1607
Pessoas ocupadas	Drift	3	6	-7.509	28.176	-
Pessoas desocupadas	Trend	8	6	-7.2671	17.62	26.4122
Salário mínimo	Drift	5	2	-5.82	16.9826	-
IPCA	Trend	8	1	-4.4553	6.6336	9.9351
Expectativas de Inflação	Trend	2	5	-9.9207	32.96	49.3656
PIB	Trend	2	5	-8.6508	25.03	37.53
IBC-BR	Drift	2	2	-8.5584	37.59	-
PMC	Drift	4	2	-7.3627	27.11	-
Taxa de câmbio efetiva real	Drift	5	5	-6.3721	20.3281	-
Taxa de emprego	Drift	3	1	0.9976	0.7835	-
Taxa desemprego	Trend	8	1	-2.2-85	1.9605	2.4516

Fonte: Elaboração própria.

Nota: Foram considerados como valores críticos as estatísticas ao nível de significância de 5%. Para as variáveis que foram escolhidas o tipo trend, foram usados os valores de referência: tau3:-3.45;phi2:4.88;phi3:6.49. Para as variáveis que foram escolhidas o tipo *drift*, foram usados os valores de referência: tau2:2.89;phi1:4.71.

A coluna "tcode" denota a seguinte transformação para os dados:

1. sem transformação
2. Δx_t
3. $\Delta^2 x_t$
4. $\ln(x_t)$
5. $\Delta \ln(x_t)$
6. $\Delta^2 \ln(x_t)$

B Resíduos

B.1 Resíduos dos modelos estimados

Figura 17: Gráficos dos resíduos para o modelo da taxa de empregados para o VAR(1)

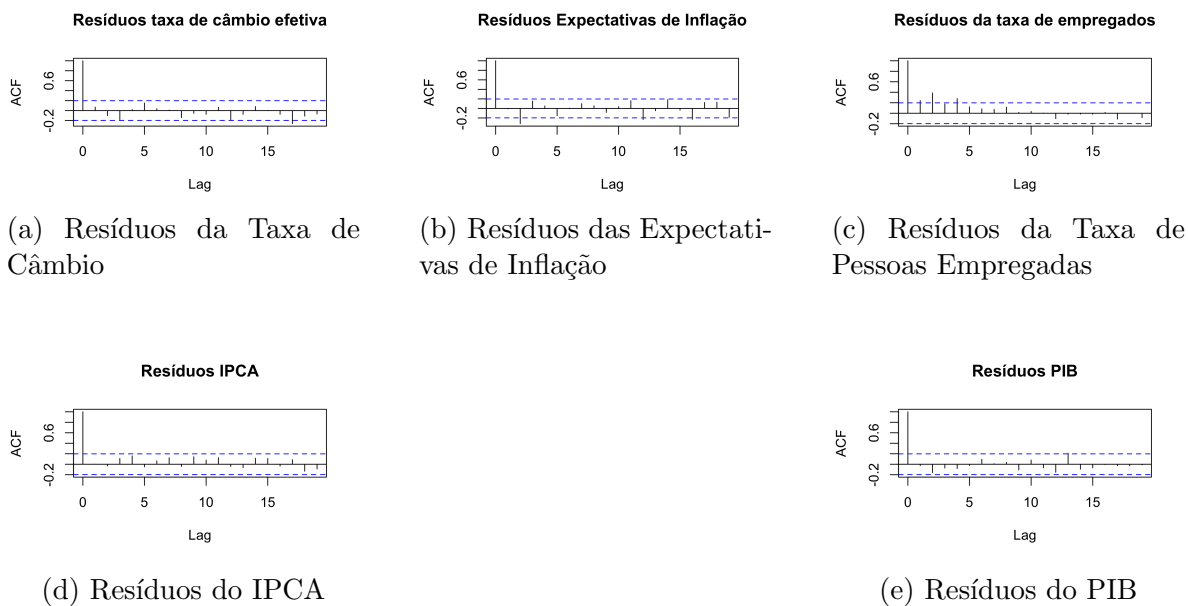


Figura 18: Gráficos dos resíduos para o modelo da diferença percentual do número de pessoas ocupadas, sem a normalização pela força de trabalho, para o VAR(1)

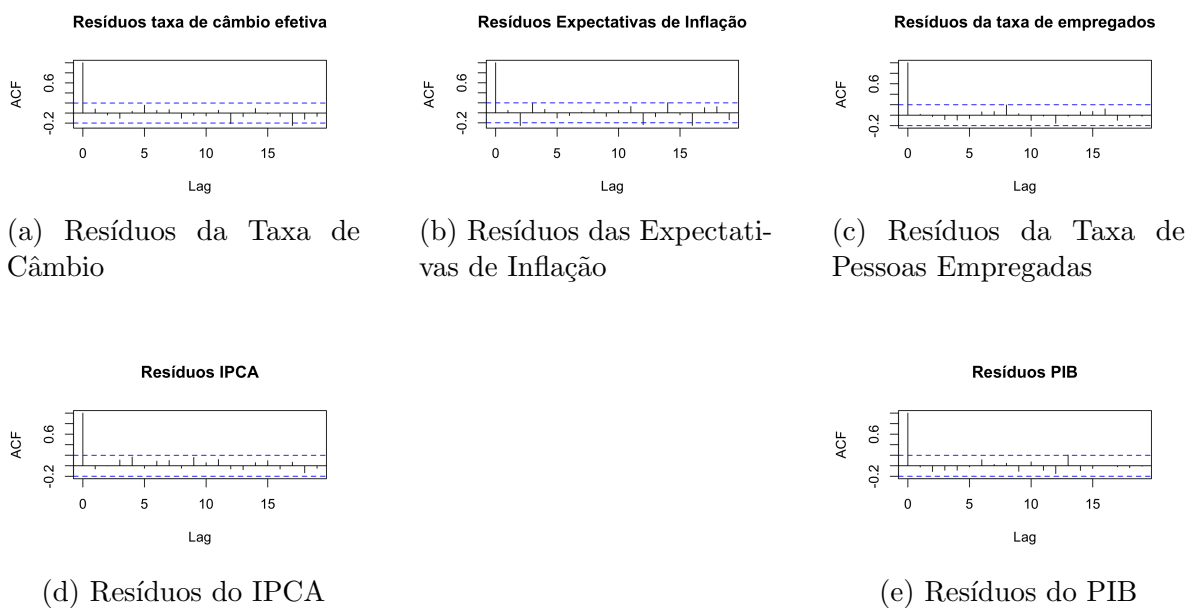


Figura 19: Gráficos dos resíduos para o modelo da taxa de empregados para o VAR(2)

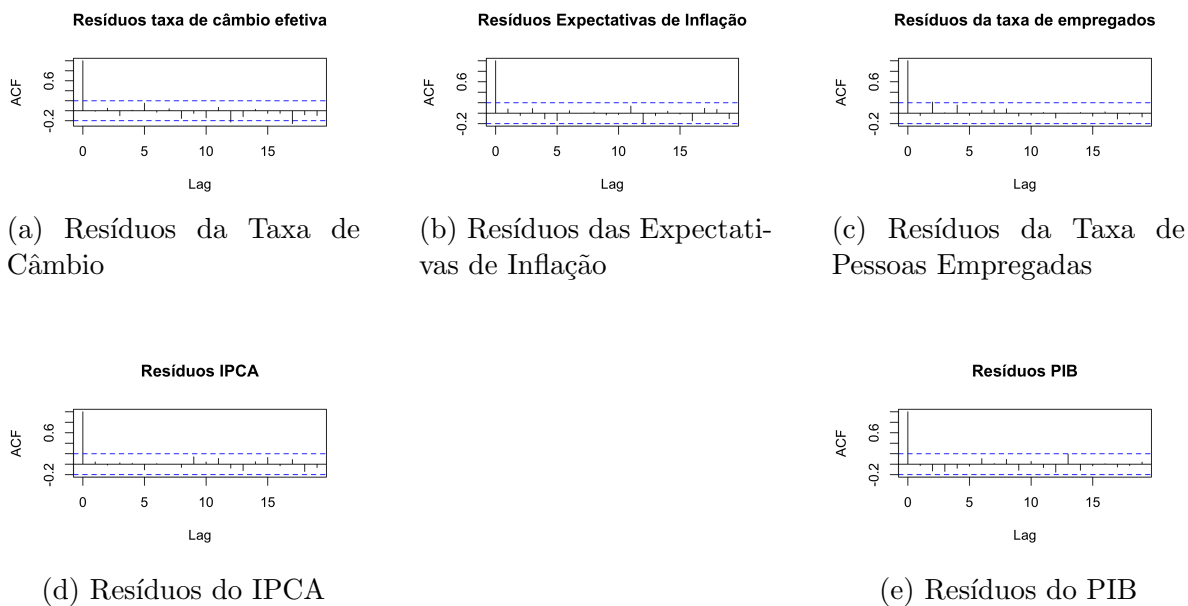
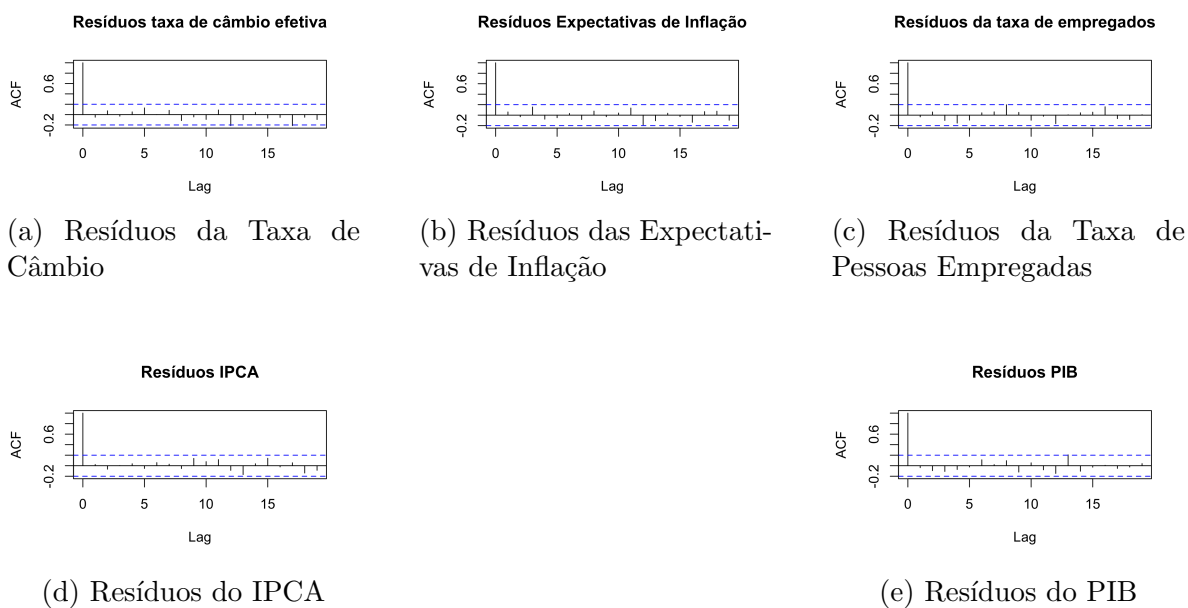
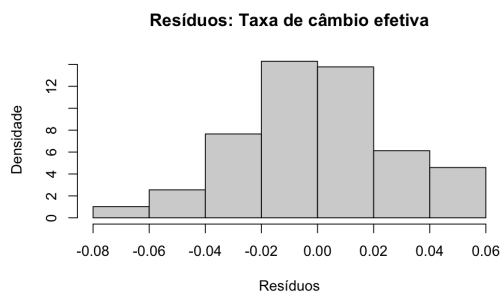


Figura 20: Gráficos dos resíduos para o modelo da diferença percentual do número de pessoas ocupadas, sem a normalização pela força de trabalho, para o VAR(2)

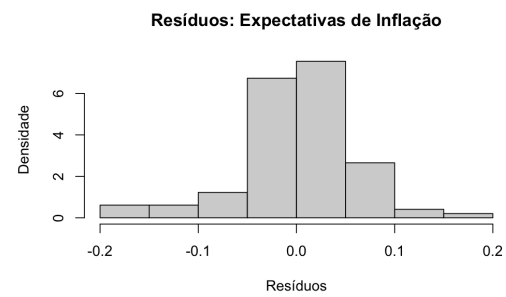


B.2 Histogramas dos resíduos

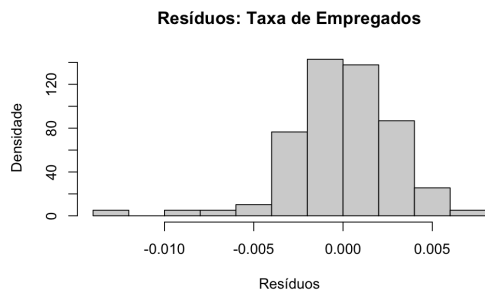
Figura 21: Histograma do modelo para a taxa de empregados para o VAR(1)



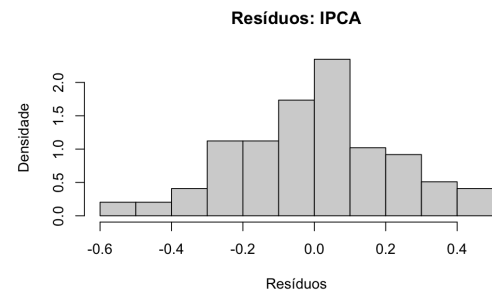
(a) Resíduos da Taxa de Câmbio



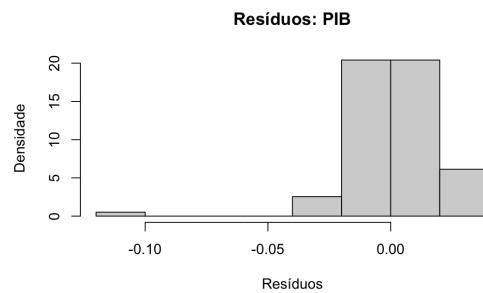
(b) Resíduos das Expectativas de Inflação



(c) Resíduos da Taxa de Pessoas Empregadas

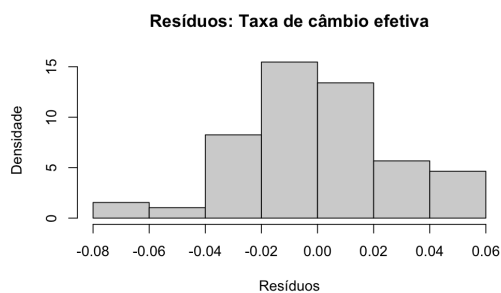


(d) Resíduos do IPCA

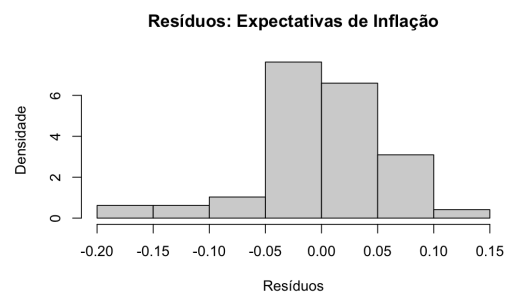


(e) Resíduos do PIB

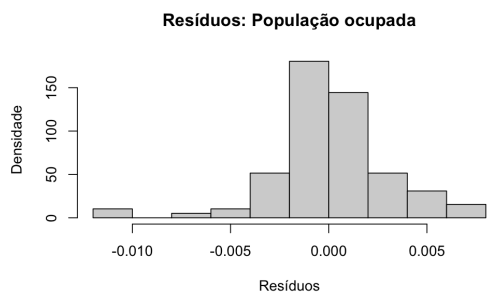
Figura 22: Histogramas para o modelo que utiliza a taxa de ocupados, sem a normalização pela força de trabalho, para o VAR(1)



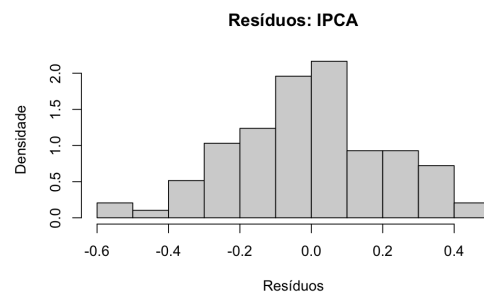
(a) Resíduos da Taxa de Câmbio



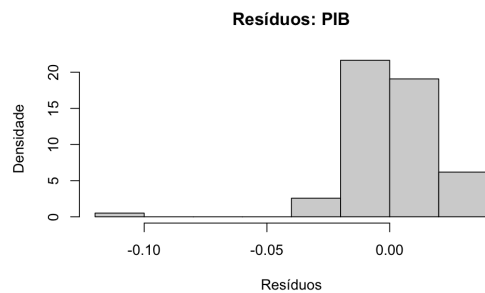
(b) Resíduos das Expectativas de Inflação



(c) Resíduos da Taxa de Pessoas Empregadas

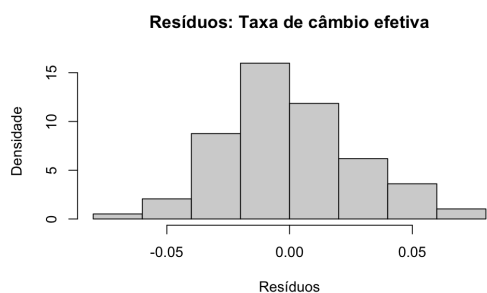


(d) Resíduos do IPCA

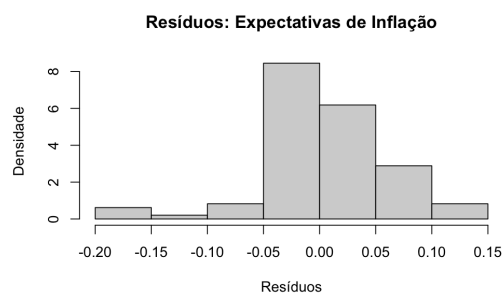


(e) Resíduos do PIB

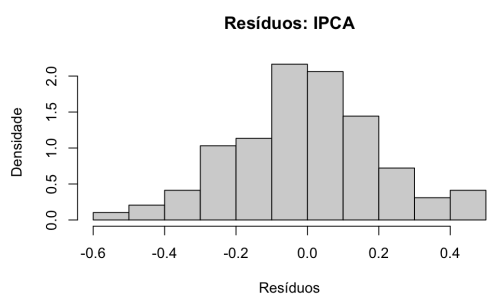
Figura 23: Histograma do modelo para a taxa de empregados para o VAR(2)



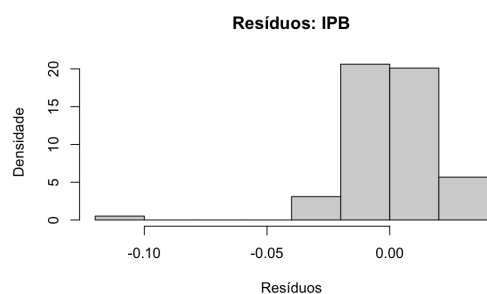
(a) Resíduos da Taxa de Câmbio



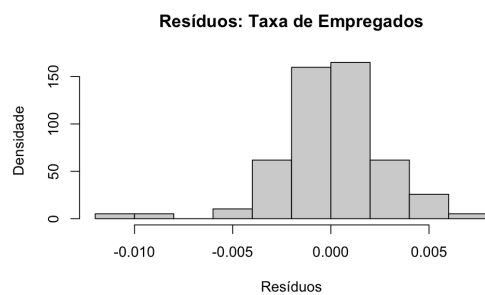
(b) Resíduos das Expectativas de Inflação



(c) Resíduos do IPCA

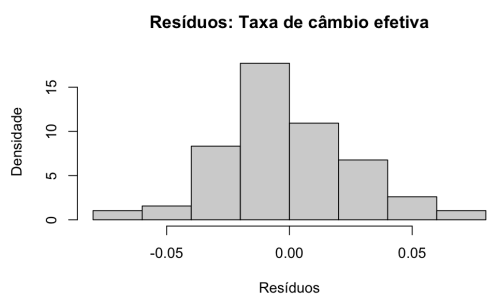


(d) Resíduos do PIB

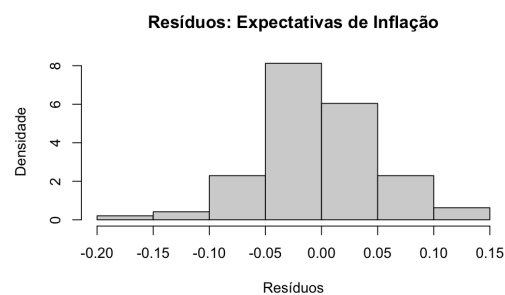


(e) Resíduos da Taxa de Empregados

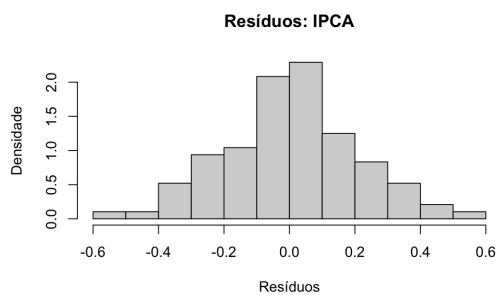
Figura 24: Histogramas para o modelo que utiliza a taxa de ocupados, sem a normalização pela força de trabalho, para o VAR(2)



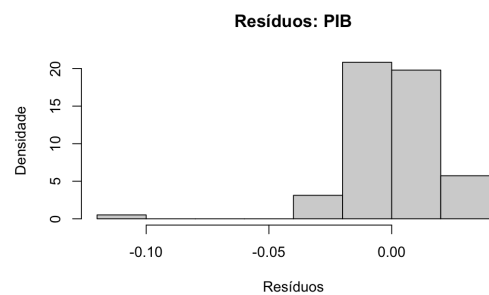
(a) Resíduos da Taxa de Câmbio



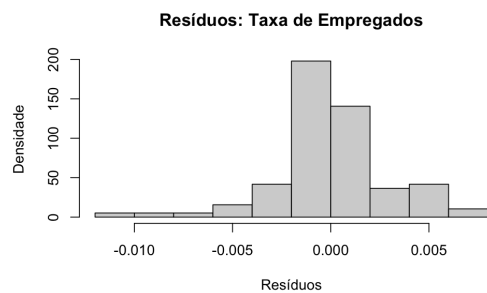
(b) Resíduos das Expectativas de Inflação



(c) Resíduos do IPCA



(d) Resíduos do PIB



(e) Resíduos da Taxa de Empregados

Referências

- Arendt, Lucas e Adam Jakubik (2021). “Labor Market Trends and Unemployment in Brazil: An Econometric Analysis”. Em: *International Journal of Economics and Financial Issues* 11.2, pp. 50–58.
- Bacciotti, Rafael e Emerson Fernandes Marçal (2020). “Taxa de Desemprego no Brasil em quatro décadas: retropolação da PNAD contínua de 1976 a 2016”. Em: *Estudos Econômicos (São Paulo)* 50, pp. 513–534.
- Barroso, Bianca A et al. (2022). “Ciência de dados aplicada à uma análise do desemprego no Brasil antes e depois da covid-19: uma abordagem por algoritmos de séries temporais”. Em.
- Bobeica, Elena e Benny Hartwig (2021). “The COVID-19 shock and challenges for time series models”. Em.
- Bueno, Rodrigo De Losso Da Silveira (2012). “Econometria de Séries Temporais”. Em.
- Casagrande, Dieison Lenon et al. (1809). “Métodos de previsão para a taxa de desemprego mensal: uma análise de séries temporais”. Em: *Revista Eletrônica de Economia da Universidade Estadual de Goiás UEG ISSN*, p. 970X.
- Cavalcanti, Marco A.F.H. e Napoleão L.C. Silva (out. de 2010). “Dívida pública, política fiscal e nível de atividade: uma abordagem VAR para o Brasil no período 1995-2008”. Em: *Economia Aplicada* 14.4, pp. 391–418. ISSN: 1413-8050. DOI: 10.1590/S1413-80502010000400007. URL: <https://doi.org/10.1590/S1413-80502010000400007>.
- Cruz, Érica Fernanda da, Tamara Aparecida Nogueira dos Anjos e Thelma Sáfadi (2014). “Análise da série temporal do desemprego em regiões metropolitanas do Brasil (pp. 104-108)”. Em: *Revista da Estatística da Universidade Federal de Ouro Preto* 3.3.
- Dickey, David A. e Wayne A. Fuller (1979). “Distribution of the Estimators for Autoregressive Time Series with a Unit Root”. Em: *Journal of the American Statistical Association* 74.366, pp. 427–431.
- Granger, C. W. J. (1969). “Investigating Causal Relations by Econometric Models and Cross-spectral Methods”. Em: *Econometrica* 37.3, pp. 424–438. ISSN: 00129682, 14680262. URL: <http://www.jstor.org/stable/1912791> (acesso em 26/11/2023).
- IBGE (s.d.). *Desemprego — IBGE — ibge.gov.br*. <https://www.ibge.gov.br/explica/desemprego.php>. [Accessed 28-11-2023].
- Lee, Junsoo e Mark Strazicich (2003). “Minimum Lagrange Multiplier Unit Root Test with Two Structural Breaks”. Em: *The Review of Economics and Statistics* 85.4,

- pp. 1082–1089. URL: <https://EconPapers.repec.org/RePEc:tpr:restat:v:85:y:2003:i:4:p:1082-1089>.
- Lenza, Michele e Giorgio E Primiceri (2020). *How to Estimate a VAR after March 2020*. Rel. técn. National Bureau of Economic Research.
- Linn, MW, R Sandifer e S Stein (1985). “Effects of unemployment on mental and physical health”. Em: *American Journal of Public Health* 75.5.
- Mazali, Antonio Alberto e José Angelo Divino (set. de 2010). *Real wage rigidity and the new Phillips curve: the Brazilian case*. DOI: 10.1590/S0034-71402010000300005. URL: <https://doi.org/10.1590/S0034-71402010000300005>.
- Pesaran, M. Hashem, Yongcheol Shin e Richard J. Smith (2001). “Bounds Testing Approaches to the Analysis of Level Relationships”. Em: *Journal of Applied Econometrics* 16.3, pp. 289–326.
- Phillips, Alban William (1958). “The Relation Between Unemployment and the Rate of Change of Money Wage Rates in the United Kingdom, 1861–1957”. Em: *Economica* 25.100.
- Phillips, Peter C. B. e Pierre Perron (1988). “Testing for a Unit Root in Time Series Regression”. Em: *Biometrika* 75.2, pp. 335–346.
- Sims, Christopher A. (1980). “Macroeconomics and Reality”. Em: *Econometrica* 48.1, pp. 1–48. ISSN: 00129682, 14680262. URL: <http://www.jstor.org/stable/1912017> (acesso em 26/11/2023).
- Stock, James H. e Mark W. Watson (2001). “Vector Autoregressions”. Em: *Journal of Economic Perspectives* 15.4, pp. 101–115.
- Toda, Hiro Y. e Taku Yamamoto (1995). “Statistical Inference in Vector Autoregressions with Possibly Integrated Processes”. Em: *Journal of Econometrics* 66.1-2, pp. 225–250.