

Ricardo Semião e Castro

Efeito Agregado das Vacinas do COVID-19 no Brasil, uma metodologia SVAR

São Paulo, Brasil

2022

Ricardo Semião e Castro

Efeito Agregado das Vacinas do COVID-19 no Brasil, uma metodologia SVAR

Revisão da dissertação para obtenção do grau de
bacharelado apresentada à Escola de Economia
de São Paulo. Área de concentração: epidemio-
logia

Fundação Getúlio Vargas

Escola de Economia de São Paulo

Orientador: Marcel Bertini Ribeiro

São Paulo, Brasil

2022

Ricardo Semião e Castro

Efeito Agregado das Vacinas do COVID-19 no Brasil, uma metodologia SVAR/ Ricardo Semião e Castro. – São Paulo, Brasil, 2022-32p. : il. (algumas color.) ; 30 cm.

Orientador: Marcel Bertini Ribeiro

Tese (Graduação), revisão – Fundação Getúlio Vargas
Escola de Economia de São Paulo, 2022.

1. Coronavírus. 2. Vacinação. 3. Efeito causal. I. Marcel Bertini Ribeiro. II. Fundação Getúlio Vargas. III. Escola de Economia de São Paulo. IV. Efeito Agregado das Vacinas do COVID-19 no Brasil, uma metodologia SVAR

Ricardo Semião e Castro

Efeito Agregado das Vacinas do COVID-19 no Brasil, uma metodologia SVAR

Revisão da dissertação para obtenção do grau de
bacharelado apresentada à Escola de Economia
de São Paulo. Área de concentração: epidemio-
logia

Trabalho aprovado. São Paulo, Brasil, 22/09/2022:

Marcel Bertini Ribeiro
Orientador

São Paulo, Brasil

2022

Dedico esse trabalho à todas as famílias que perderam entes queridos para o coronavirus

Agradecimentos

Gostaria de agradecer o professor Marcel Bertini Ribeiro pelo incrível esforço feito em me orientar e contribuir para com o presente trabalho. Também sou grato pelos meus pais, que sempre incentivaram a busca de conhecimento. Por fim, gostaria de agradecer todos os professores, funcionários, e colegas da Escola de Economia de São Paulo, que contribuíram direto ou indiretamente para minha formação acadêmica. Um desses colegas, Juan Bonilla Villegas, merece um agradecimento especial, pois sua parceria sempre me encorajou a ir além nos estudos.

Resumo

Encontrar evidências para a efetividade das vacinas do COVID-19 é muito importante para guiar a ação de *policymakers*, ainda mais dado o quão recente a pandemia é. Esta tese contribui com essa literatura utilizando uma *Structural Vector Auto-Regression* para modelar a relação entre as variáveis, motivada e identificada por um modelo epidemiológico SIR. A captura das dinâmicas temporais entre as variáveis, o relaxamento de hipóteses de taxas de infecção, mortalidade, e vacinação constantes no tempo, e o potencial de gerar diferentes tipos de resultados, são os benefícios principais desse método. Foi descoberto que resposta de mortes à um choque de 1000 casos foi 60% (ou 10,24, em termos absolutos) menor depois da inclusão das vacinas. Também foi feita uma análise de contrafactuais, indicando que o número de vidas salvas vai de um *lower bound* de 16 mil a um *upper bound* de 279 mil, com indícios de maior proximidade do segundo.

Palavras-chave: Coronavírus. Vacinação. Efeito Causal. SVAR.

Abstract

Finding evidence for the effectiveness of COVID vaccines is very important to guide the action of policymakers, especially considering how recent the pandemic is. This thesis contributes to literature by using an Structural Vector Auto-Regression to model the relationship between the variables, motivated by and identified by a SIR epidemiological model. The capture of temporal dynamics between variables, the relaxation of assumptions on infection, mortality, and vaccination rates constant over time, as well as the potential of generating different types of results are this method's main benefits. It was discovered that the response of deaths to a shock of 1000 cases was 60% (or 10, 24, in absolute terms) less than after the inclusion of vaccines. There was also an analysis of counterfactuals, indicating that the number of lives saved goes from *lower bound* of 16 thousand to a *upper bound* of 279 thousand, with clues to a proximity to the latter.

Keywords: Coronavirus. Vaccination. Causal Effect. SVAR

Sumário

1	INTRODUÇÃO	9
2	REVISÃO DE LITERATURA	10
2.1	Abordagens	10
2.2	Resultados Existentes	11
3	MODELO TEÓRICO	12
3.1	O Modelo SIRDV	12
3.2	Relação Com o Modelo SVAR	13
4	DADOS	16
4.1	Fontes	16
4.2	Limitações	16
4.3	Estatísticas Descritivas	17
5	MODELO ECONOMETRICO	19
6	RESULTADOS E DISCUSSÃO	21
6.1	Resposta de Mortes à Casos Pré e Pós Vacinação	21
6.2	Análise de Contrafactual	23
7	CONCLUSÃO	25
	Referências	27
APÊNDICE A	IRF'S INDIVIDUAIS	29
APÊNDICE B	DIAGNÓSTICOS DO VAR	30

1 Introdução

A pandemia do coronavírus (COVID-19) causou mais de 38 milhões casos e 707 mil mortes no Brasil¹, portanto, estudar a eficácia das intervenções disponíveis é de suma importância.

A vacinação já se provou indiscutivelmente efetiva no controle de uma série de outras doenças, e existem muitos artigos com abordagens quase-experimentais atestando a diminuição da intensidade da pandemia trazida pelas vacinas para COVID-19. Ao mesmo tempo, ainda existe espaço para novas evidências acerca da eficiência da vacinação.

A presente tese inova ao propor um método de estimação com base em um *Structural Vector Auto-Regression* (SVAR) que, no melhor de meu saber, ainda não foi utilizado na literatura. A defesa desse método vem com base no celebrado modelo epidemiológico SIR, que explicita a natureza interconectada e temporalmente dinâmica das variáveis de interesse, algo que um SVAR consegue modelar muito bem.

Existem diferentes medidas para avaliar eficácia, algumas na linha de efeito marginal da vacina, outras na linha do efeito total da política de vacinação. Como mostrarei, o SVAR tem o potencial para criar estimativas de ambos através das funções de resposta ao impulso (IRF's) e da criação de contrafactuais.

Adicionalmente, o SVAR também permite relaxar hipóteses que outros métodos de estimação costumam fazer, ao ter as taxas de infecciosidade, mortalidade, e vacinação como dependentes da intensidade da pandemia no passado.

A principal limitação desse tipo de modelo é a identificação dos choques de cada variável. A teoria por trás dos modelos SIR motivou a realização de uma decomposição de Cholesky, na ordem de casos-vacinas-mortes, gerando uma identificação consistente do choque de casos. Por mais que os outros dois choques não tenham uma interpretação tão clara, a metodologia descrita nos próximos parágrafos não é afetada tão intensamente por esse problema.

O principal resultado deste trabalho vem da análise da variação no modelo antes e depois da introdução da política de vacinas, que gerou insumos para entender a mortalidade de um choque de casos pré e pós vacinas. Antes, um choque de 1.000 casos gerava aproximadamente

¹Dados do Ministério da Saúde (<https://covid.saude.gov.br>), de dezembro/2023

16, 91 casos ao longo das semanas subsequentes, e depois, apenas 6, 67, uma redução de 60%, atestando a importância da vacinação no Brasil. Esses resultados são consistentes com a análise exploratória das variáveis, que indica uma grande redução nas correlações cruzadas entre casos e mortes pós introdução das vacinas.

Outra abordagem foi utilizar o SVAR para criar contrafactuais onde nunca houve a política de vacinação. Foram testados vários métodos diferentes, o que permitiu a estimação de um *lower bound* de 16 mil vidas salvas, contra um *upper bound* de 279 mil. Essa banda é bastante larga, mas o resultado do parágrafo anterior, e resultados de outros trabalhos, indicam que o resultado está longe dos limites, com maior proximidade do *upper bound*.

A organização do restante da tese segue: a seção 2 descreve os estudos sobre vacinas do COVID-19 já feitos; a seção 3 apresenta o modelo SIR, motivando o SVAR; os dados são descritos na seção 4; a seção 5 traz o modelo econométrico estimado; 6 mostra os resultados; e por fim, concluo na seção 7.

2 Revisão de Literatura

2.1 Abordagens

Halloran et al., 1999 separa a literatura sobre efetividade de vacinas na busca de quatro efeitos:

- Efeito direto: $E[S_i | V_i]$;
- Efeito indireto: $E[S_i | \neg V_i, V_p]$;
- Efeito total: $E[S_p | V_p, V_i] - E[S_p | \neg V_p, \neg V_i]$;
- Efeito geral: $E[S_p | V_p] - E[S_p | \neg V_p]$.

Onde S_x é o estado de saúde individual ($x = i$) ou populacional ($x = p$), e V_x é o indicativo de indivíduo ou população vacinada. Os três primeiros efeitos são à nível individual, mas nesta tese busco o efeito geral da vacina, agregando tanto os efeitos diretos, quanto os indiretos, à nível populacional.

Os estudos sobre o efeitos individuais passam por estudos laboratoriais, de caso-controle, de coorte, e quase experimentais como escore de propensão. Já a literatura sobre o efeito agregado é mais diversa, podendo ser dividida em: (i) métodos que constroem um contrafactual com base em modelos teóricos e/ou preditivos; (ii) métodos quase-experimentais como séries temporais interrompidas, diferenças em diferenças, e desenhos de regressão descontínua; (iii) métodos de regressão com variáveis contínuas, como regressões *cross-section*, em painel, ou com variáveis

instrumentais.

Os grupos (i) e (ii) são bons para estimar efeitos agregados, por gerarem um contrafactual simples de interpretar. O primeiro, têm o ponto negativo de não ser muito “agnóstico”, ao assumir um modelo populacional muitas vezes bastante específico. Já o segundo, não analisa a dimensão contínua do tratamento das vacinas. O grupo (iii) não traz a robustez de uma análise quase-experimental, mas consegue trazer resultados com base na intensidade da intervenção das vacinas.

Com base nesses prós e contras, argumento que existe uma janela para a aplicação de um modelo *Structural Vector Auto Regressive*. O SVAR consegue utilizar as variáveis contínuas, gera vários tipos de resultado diferentes, e dá atenção especial para a relação inter-temporal das variáveis, algo que argumentarei ser muito importante na seção 3.

2.2 Resultados Existentes

Pelo meu conhecimento, o uso de SVAR’s nessa literatura é no mínimo incomum, e não existe nenhum trabalho aplicando esse método em dados da pandemia do COVID-19. Similarmente, os estudos utilizando vacinas como uma variável contínua também são escassos. A seguir apresento os estudos mais comparáveis com a presente tese.

Barro (2022) mede o efeito da taxa de vacinação na taxa de mortes e casos, com base em dados a nível individual dos EUA. O estudo realiza regressões *cross-section* em modelos de *seemingly unrelated regressions* e variáveis instrumentais, para lidar com a possibilidade de causalidade reversa. A amostra é dividida entre quatro períodos diferentes, cada um com seu conjunto de regressões. Os resultados indicam que a vacina não parece fazer efeito no início da pandemia. Além disso, um aumento de um ponto percentual na taxa de vacinação diminui a taxa de mortalidade em 0,01 pontos percentuais, no período de pico, sendo um pouco menor antes e depois. O efeito na taxa de casos é parecido para os dois períodos medianos, indicando uma diminuição próxima de 0,45 pontos percentuais.

Chen et al. (2022) calcula o efeito de vacinas por cem habitantes na taxa de crescimento de casos e hospitalizações, com dados diários de estados americanos, utilizando regressões em painel com efeitos fixos de tempo e estado. O resultado principal é uma redução de 0,7 pontos percentuais (com erro padrão de 0,2 p.p.) na taxa de crescimento, para um aumento de uma vacina por 100 habitantes.

Ferreira et al. (2022) utiliza econometria Bayesiana para criar um contrafactual brasileiro onde não houveram vacinas. O efeito encontrado é um *lower-bound* porque somente leva em

conta o efeito direto da vacina, sem incluir o efeito de imunidade de rebanho. De acordo com as estimativas, a política de imunização durante janeiro a setembro de 2021, salvou diretamente 75 mil vidas.

3 Modelo Teórico

3.1 O Modelo SIRDV

Para motivar a utilização de um SVAR, apresento o modelo SIRDV (uma extensão do SIR), celebrado na literatura por ajudar a explicar o funcionamento de pandemias. Esse modelo expõe a relação inter-temporal das variáveis, o que faz do SVAR um modelo útil para o objetivo desse trabalho.

O modelo SIRDV é um sistema de equações que explica quantas pessoas estão em diferentes compartimentos da sociedade, nomeadamente: Suscetíveis $S(t)$, Infectados $I(t)$, Recuperados $R(t)$, Mortos $D(t)$, e Vacinados $V(t)$.

No dia t , o número de novos infectados é proporcional ao número de infectados na população em $t-1 - I(t-1)$, ao número de pessoas suscetíveis $- S(t-1)$, e à taxa de infecciosidade do vírus² $-\beta(d(t))$. Também em t , uma proporção v dos suscetíveis são vacinados. Ambos os grupos deixam de fazer parte dos suscetíveis a partir de $t+1$. Por fim, dentre os infectados, uma fração τ se recupera da doença, mas fração η falece, gerando o seguinte sistema:

$$\begin{aligned} S(t) &= S(t-1) - vS(t-1) - \beta(d(t-1))S(t-1)I(t-1) \\ I(t) &= I(t-1) + \beta(d(t-1))S(t-1)I(t-1) - \tau I(t-1) - \eta I(t-1) \\ R(t) &= R(t-1) + \tau I(t-1) \\ D(t) &= D(t-1) + \eta I(t-1) \\ V(t) &= V(t-1) + vS(t-1) \end{aligned}$$

O formato de $X(t) = X(t-1) + \dots$ já indica similaridades com um vetor auto-regressivo. Para mais detalhes, o efeito da vacina é decomposto na figura 1 e explicado a seguir.

A vacinação de um indivíduo em $t-1$ diminui $S(t)$ em 1. Essa diminuição é carregada para $S(t+1)$ em diante. Menos pessoas suscetíveis em $t-1$ implica em menos pessoas infectadas

²Que por sua vez é inversamente proporcional ao grau de distanciamento social $d(t)$ sendo empregado

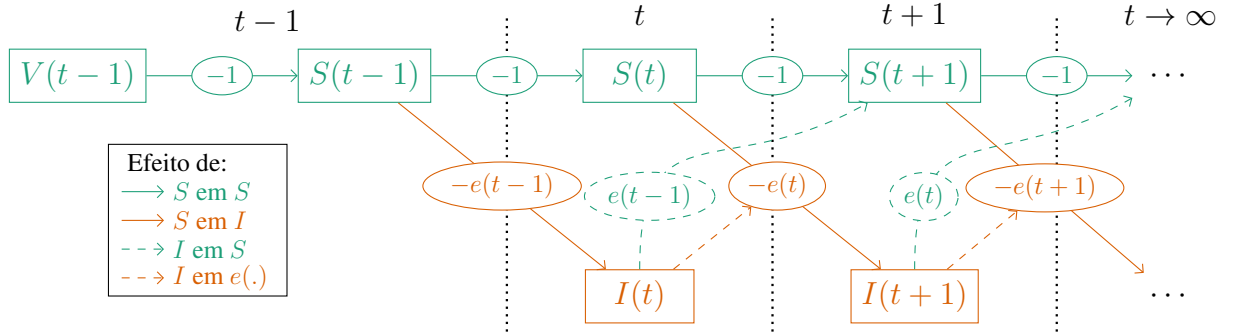
no período seguinte, através da derivada:

$$e(t) := \frac{\partial I(t+1)}{\partial S(t)} = \beta(d(t))I(t)$$

Avançando para o próximo período, a diminuição em $S(t)$ novamente gerará uma diminuição no número de infectados do próximo período, com valor de $e(t)$.

Adicionalmente, como o valor de $I(t)$ foi reduzido pela vacina em $e(t-1)$, a diminuição de casos seguinte ($-e(t)$) sofre dois efeitos: aumenta (em módulo) pois menos casos acumulados implica em menos novas infecções; diminui (em módulo) pois, por hipótese, menos casos acumulados implica em uma população menos preocupada e com menor nível de distanciamento $d(t)$, logo com mais infecções. A diminuição dos infectados também gera um efeito terciário de mais pessoas suscetíveis, que eventualmente poderão ser infectadas no futuro.

Figura 1 – Efeito da vacina – modelo SIRV



3.2 Relação Com o Modelo SVAR

Fica claro que a relação entre as variáveis não é trivial, e tem uma natureza muito interconectada, inclusive com *lag's* das outras variáveis. Um modelo SVAR consegue capturar essa dinâmica, e de uma maneira muito menos restritiva, permitindo relações diretas das outras variáveis $R(t)$, $D(t)$, e $V(t)$ nas outras equações. Por exemplo, a hipótese de que pessoas que se infectaram nunca poderão se infectar novamente é relaxada, bem como a hipótese que a taxa de vacinação é dada exógenamente e não dependa da intensidade da pandemia.

Vale notar, porém, que o termo de novos infectados no modelo SIRDV é uma interação não-linear das variáveis, o que não será modelado no SVAR. O resultado do modelo será, então, uma aproximação linear do efeito da vacina, que pode na realidade depender do nível das outras variáveis, inclusive da própria vacinação.

Essa é a realidade de muitos dos métodos de estimação com base em regressão já estabe-

lecidos na análise empírica de pandemias, mas uma possível expansão do presente trabalho é explorar esses efeitos não-lineares.

3.2.1 Efeitos a Estimar

O modelo SIRDV também ajuda a esclarecer as diferentes maneiras de se estimar o efeito da vacina:

- O efeito marginal de $V(j)$ em $[X(i)]_{i=j+1}^{\infty}$, sendo X casos ou mortes realizados;
- $[X'(i) - X(i)]_{i=1}^{\infty}$, onde $X'(i)$ é o contrafactual em que $[V(i)]_{i=1}^{\infty} = (0, 0, \dots)$;
- Seja f a função do efeito causal que $I(i)$ gera em $D(i)$: $D_t = f([I(i)]_{i=1}^{t-1})$, podemos comparar f do factual com f' do contrafactual.

O SVAR permite, uma abordagem diferente para cada efeito acima. O primeiro efeito pode ser analisado pela IRF de vacinas em casos/mortes, providenciado que o choque de vacinas esteja bem identificado, e que o SVAR não sofra com variáveis omitidas. O segundo efeito pode ser medido através da criação de um contrafactual, algo que será feito com diferentes métodos. O terceiro efeito pode ser estimado comparando as IRF's de um SVAR estimado pré vacinas, e de outro estimado pós vacinas.

A última abordagem, além de ser interessante por si só, tem uma certa robustez. Como estamos olhando para a variação do modelo, mesmo que ele apresente alguma falha, a mudança proporcional não deve ser tão diferente da mudança proporcional do modelo correto, por mais que as IRF's de um, tomadas individualmente, sejam diferentes das IRF's do outro.

“Tomar a IRF individualmente” é exatamente o que a primeira abordagem faz, ficando claro que ela não é tão robusta. Como nem a identificação do choque de vacinas, nem a inexistência de variáveis omitidas podem ser garantidos, essa abordagem foi deixada para o apêndice.

3.2.2 Distanciamento Social e Variáveis Omitidas

O modelo SIRDV também ajuda a entender seu papel do distanciamento social na estimação do SVAR, através da forma funcional de $\beta(d(t))$. Boa parte dos estudos em modelos SIR derivam um modelo de escolha individual do distanciamento ótimo, baseado na troca entre custo econômico de diminuir sua mobilidade, e o ganho em menor probabilidade de infecção. [Regula \(2010\)](#) utiliza um modelo – que foi replicado para a crise atual por [Cabrera et al. \(2021\)](#), [Mwalili et al. \(2020\)](#), e [Baker \(2020\)](#) – onde os indivíduos escolhem se distanciar mais quando pandemia está mais “intensa”, e o risco de infecção é maior, logo $d(t)$ seria proporcional ao número de casos diários. Similarmente, a população fica mais “confiante” após se vacinar, e relaxa o distanciamento.

Pode-se concluir que $d(t)$ deve ser proporcional ao número de mortes, e inversamente proporcional ao número de vacinas. Isso implica que, embora seja importante levar em conta o distanciamento na estimação empírica – e isso ficará claro nos dados –, parte da dinâmica já é controlada pelas próprias variáveis de casos, vacinas, e mortes.

Para que a análise “antes e depois” não contenha vieses, a única variação entre os períodos deveria ser a presença da vacina e como ela altera a relação casos & mortes. E essa hipótese não deve estar completamente satisfeita por alguns motivos:

- Como dito anteriormente, a escolha do distanciamento muda, embora essa dinâmica está parcialmente inclusa no modelo;
- Adicionalmente, o nível de imunização natural se altera com o passar do tempo, mas na data da primeira vacina, apenas 4,08% da população havia se recuperado da doença, que em comparação aos 85,05% vacinados, indica pouca relevância desse viés;
- Por fim, tanto as mutações do COVID-19, quanto as diferentes marcas de vacinas podem atrapalhar a análise, uma vez que a eficácia da vacinação varia muito a depender dessas características. Ainda assim, o efeito geral/médio deve-se manter pouco alterado³.

3.2.3 Identificação

Uma das principais limitações do SVAR é a necessidade de fazer uma hipótese apriorística de identificação sobre a relação contemporânea dos choques de cada variável. A literatura SIR apresenta argumentos para uma identificação por decomposição de Cholesky⁴ na ordem de casos-vacinas-mortes.

Mortes é a variável que mais responde às outras. Uma maior incidência de casos aumenta diretamente a probabilidade das mortes, inclusive através da pior provisão de serviços públicos de saúde. As evidências laboratoriais reforçam essa visão

Falta estudar ordem entre casos e vacinas. O modelo SIR é consistente com a interpretação de que a pandemia “surge” após um choque exógeno na variável de casos, motivando casos como a menos correlacionada contemporaneamente, que é a interpretação principal desta tese. Uma possível limitação da identificação vem do fato de que boa parte da provisão de vacinas depende de choques exógenos na criação e produção das vacinas, balanço do governo, e motivações políticas – como mostra a própria experiência brasileira da disputa política entre o governo federal e estaduais para vacinar a população. Ainda assim, parte da oferta, e principalmente da demanda por vacinas, parece responder contemporaneamente à intensidade da pandemia (pessoas indiferentes escolhem se vacinar após verem um número alto de casos e mortes).

³Adicionar dados relacionado a tais variáveis pode ser uma futura expansão para o modelo.

⁴Que consiste em assumir diferentes ordenamentos de relações contemporâneas entre as variáveis

A literatura de SVAR contém várias maneiras diferentes de explorar as relações teóricas entre as variáveis para criar hipóteses críveis, como identificações por efeitos de curto e longo prazos, e restrições de sinais. Não foi possível identificar nenhum paralelo entre a teoria epidemiológica e as possíveis identificações mais rebuscadas, abrindo espaço para futuras melhorias deste trabalho.

4 Dados

4.1 Fontes

O modelo SVAR somente necessita das variáveis de casos, mortes, e vacinas, uma vez que suscetíveis e recuperados são combinações lineares das três primeiras⁵. A fonte dessas variáveis foi repositório de dados da organização *Our World in Data* (Ritchie et al. (2020)) que apresenta dados diários. Para o Brasil, a organização captou os dados através do Ministério da Saúde, e da Universidade Federal de Viçosa.

A janela de dados vai de 02/03/2020 à 20/03/2023, totalizando 1113 dias. Os dados foram agregados em semanas (somando os valores de todos os dias da semana), por três motivos principais: (i) as limitações sobre o *timing* da captura dos dados, como será discutido mais a frente, afeta muito a utilidade de dados diários; (ii) a “periodicidade” do covid se assemelha à ciclos semanais, o período de maior chance de contágio seria aproximadamente 7 dias, e o período para eliminar o vírus do corpo pós sintomas seria aprox. 14 dias; e (iii) o VAR com dados semanais conseguiu ser muito mais parcimonioso, com bem menos lags, e conseguindo capturar a dinâmica das variáveis tão bem quanto o VAR diário.

Aplicando essa transformação, os 1113 dias se tornam 161 semanas. Como será descrito na seção do modelo econométrico (5), serão feitos modelos antes e depois da primeira aplicação de vacina no Brasil, que entrou na base de dados no dia 27/01/2021. Portanto, os modelos pré vacinas foram feitos com 48 observações, e os pós vacinas, com 113 observações.

4.2 Limitações

Em 4 observações as variáveis apresentavam valores negativos por causa de correções do ministério. Nesses casos, os valores foram aproximados usando a média das seis observações vizinhas, antes de agregar o dado semanalmente.

⁵Inclusive, sua inclusão geraria problemas de multicolinearidade

A qualidade das variáveis é uma limitação comum dos trabalhos sobre COVID, e este não é exceção. Boa parte das pessoas que pegam a doença não se testam, seja porque não apresentaram sintomas, porque não se preocuparam, ou porque não puderam arcar com o valor do teste, gerando muita subnotificação na variável. Além disso, idealmente queríamos que a variável de “novos casos” indicasse o dia da infecção, mas na realidade, um indivíduo infectado somente entra para a amostra após desenvolver os sintomas, se testar, e entrar para a base de dados do Ministério da Saúde. Vale notar que o SVAR ajuda a lidar com isso ao incluir vários *lags* das variáveis e gerar uma IRF, sem depender tanto da interpretação de um *lag* específico. Por fim, os testes do COVID costumam gerar falsos positivos e falsos negativos, também diminuindo a qualidade dessa variável. Além disso, os dados diários apresentam recorrentes correções nos finais/inícios de semanas, de modo que é necessário fazer uma suavização de média móvel de 7 dias. Por esse conjunto de problemas, a informação extra que dados diários trazem não é suficiente para justificar forte perda de performance gerada para a modelagem VAR.

O número de mortes sofre de todos esses problemas, mas em menor intensidade: a maior parte das mortes por COVID são (corretamente) identificadas e incluídas na base, com um *delay* menor que Casos. Essas limitações comprometem a interpretação estimação do SVAR, principalmente no que tange aos resultados referentes à casos, assim como indica [Escobar et al. \(2020\)](#), de modo que nossos resultados principais são aqueles relacionados à variável de mortes.

No que tange o distanciamento, os dados disponíveis, são de *surveys* de preenchimento voluntário e individual (sem um entrevistador), que podem apresentar dados viesados, não representativos, e disponíveis para uma janela bem pequena. Desse modo, nenhuma variável desse tema foi incluída. Vale lembrar que parte da dinâmica do distanciamento é explicada pelas próprias variáveis do sistema.

Uma possível expansão seria desagregar a variável de vacinas entre 1^{as}, 2^{as} e 3^{as} doses aplicadas, mas com a fraca especificação do choque, só estaríamos gerando problemas de dimensionalidade no VAR, sem trazer real benefício.

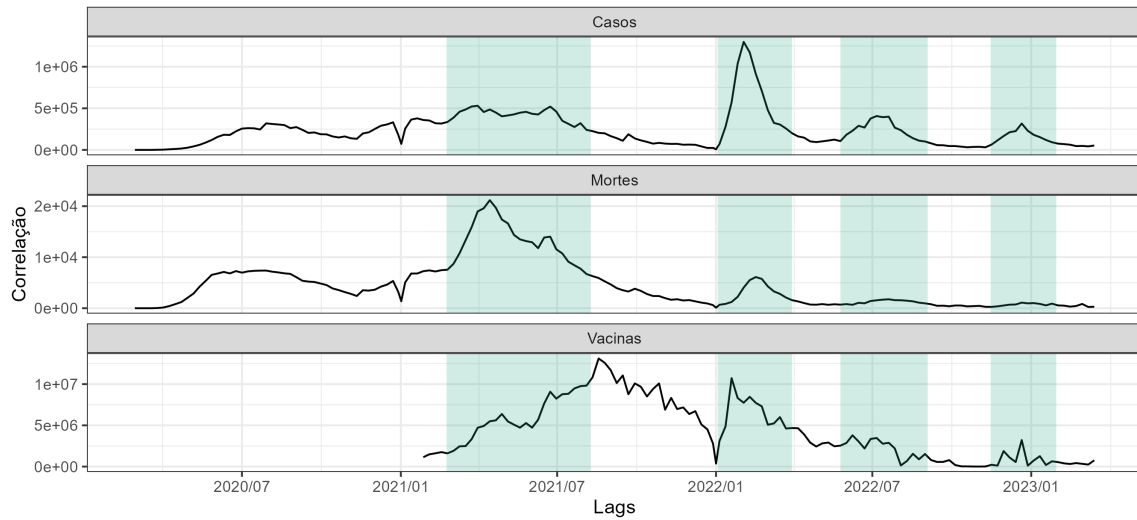
4.3 Estatísticas Descritivas

A figura 2 mostra como as variáveis de interesse progrediram no tempo. Foram marcadas algumas “ondas” de casos e mortes. É especialmente interessante notar a diferença na proporção casos/mortes entre a primeira e segunda onda.

Podemos ver que o *timing* das ondas não parece ser explicado apenas pelas três variáveis, e, de encontro com o descrito no modelo SIRDV, trabalhos como [Moraes \(2021\)](#) indicam que o relaxamento do distanciamento social é uma de suas causas. Como não temos dados de dis-

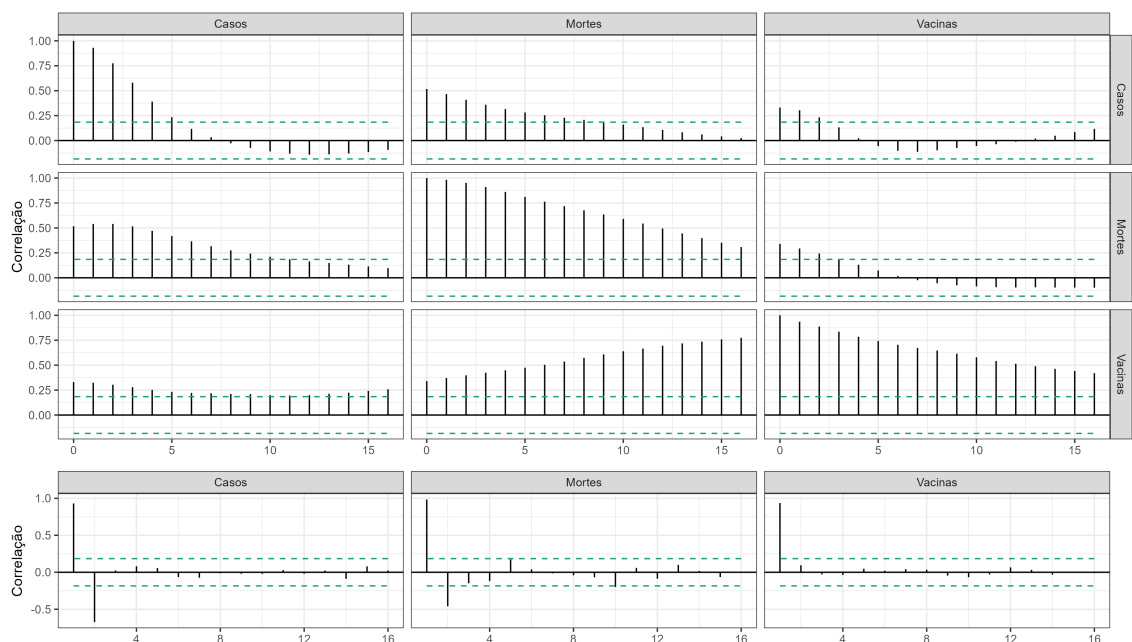
tanciamento, foram incluídas tendências determinísticas quadráticas no modelo SVAR, com o intuito de ajudar no controle desses *drivers* exógenos das ondas.

Figura 2 – Valores históricos das séries



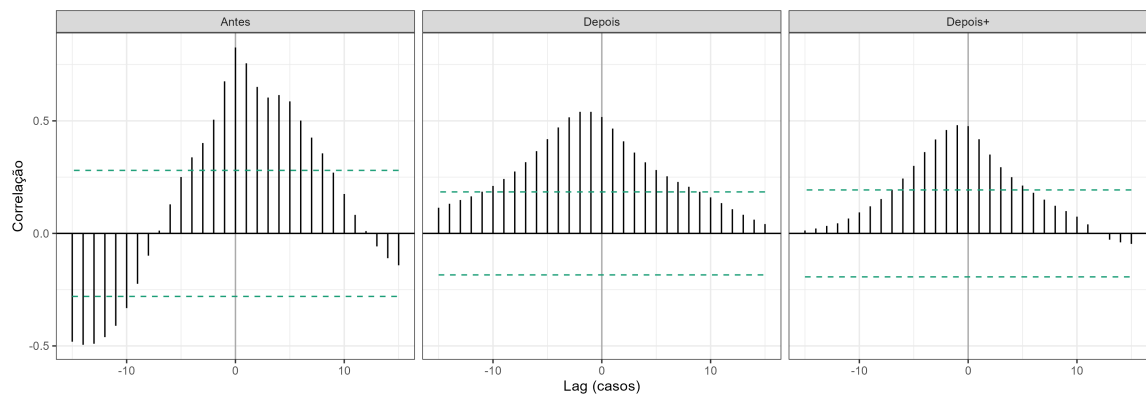
A figura 3 apresenta as autocorrelações (padrões, parciais, e cruzadas) das séries. As variáveis têm muitos *lags* significativos em suas ACF's, e apenas os primeiros em suas PACF's. Alguns fatos interessantes: casos tem a maior auto dependência; vacinas tem maior relação com mortes do que com casos – algo possivelmente relacionado com a “preferência” do governo, produtores, e “tomadores de vacinas”; as outras variáveis não apresentam correlação negativa com vacinas.

Figura 3 – Correlação cruzada e parcial das séries



Essa última constatação é contraintuitiva, e pode indicar como pode ser limitado medir o efeito direto da vacinação. Em oposição o efeito da vacina na relação das variáveis tem um resultado inicial interessante, exposto na figura 4. Uma análise das correlações cruzadas antes e depois da vacinação mostram uma maior relação entre *lags* de casos e mortes no período pré vacina, possivelmente indicando um aumento na resistência contra a doença na população pós vacina. Essa diferença é ainda maior no cenário da política de vacinação mais avançada como, por exemplo, 10 meses pós vacina.

Figura 4 – Correlação mortes-casos pré e pós vacinas



Por fim, no que tange a sazonalidade do COVID-19, a literatura é inconclusiva (Byun et al., 2021). Alguns estudos atestam relação com a temperatura ao longo do ano (Wiemken et al., 2023), porém, essa relação não pareceu estar tão significativa nos dados disponíveis, e gerou pouco efeito no modelo econométrico. Outras especificações de sazonalidade foram testadas, mas ao fim, nenhum tipo de correção pareceu necessária.

5 Modelo Econométrico

Para realizar as três abordagens diferentes citadas na seção do modelo teórico, foram estimados os modelos (1) a (3):

$$\begin{bmatrix} I_t \\ D_t \end{bmatrix} = \alpha_0^{(1)} + \sum_{j=1}^{p^{(1)}} \phi_j^{(1)} \begin{bmatrix} I_{t-j} \\ D_{t-j} \end{bmatrix} + \alpha_1^{(1)} \text{onda}_t + \alpha_2^{(1)} \text{onda}_t^2 + \epsilon_t^{(1)}, \quad t \in \{1, 2, \dots, t_v\} \quad (1)$$

$$\text{Modelo (1) estimado na amostra } t \in \{t_v, t_v + 1, \dots, t_f\} \quad (2)$$

$$\begin{bmatrix} I_t \\ V_t \\ D_t \end{bmatrix} = \alpha_0^{(3)} + \sum_{j=1}^{p^2} \phi_j^{(3)} \begin{bmatrix} I_{t-j} \\ V_{t-j} \\ D_{t-j} \end{bmatrix} + \alpha_1^{(3)} \text{onda}_t + \alpha_2^{(3)} \text{onda}_t^2 + \epsilon_t^{(3)}, \quad t \in \{t_v, t_v + 1, \dots, t_f\} \quad (3)$$

$$\text{Modelo (1) estimado na amostra } t \in \{t_v + g, t_v + g + 1, \dots, t_f\} \quad (4)$$

$$\text{Modelo (3) estimado na amostra } t \in \{t_v + g, t_v + g + 1, \dots, t_f\} \quad (5)$$

Onde t_v é a data da primeira vacina e t_f é a última data na base de dados. V_t é o número de vacinas aplicadas no período t , I_t é o número de casos registrados no período, e D_t é o número de mortes registrados no período. onda_t é um vetor com um índice de tempo para cada onda (como definidas pela figura 2).

O modelo (1) é o modelo pré vacinação, (3) o modelo pós vacinação, e (2) um modelo “intermediário”. Ele já apresenta a nova dinâmica entre casos e mortes pós vacinas, mas não leva em conta o efeito “indireto” dessas variáveis com vacinas, por isso, a princípio, deve apresentar os efeitos positivos da vacinação, mas em menor intensidade do que o segundo (3).

Os modelos (4) e (5) foram estimados considerando um “gap” na amostra, para melhor separá-los do período pré vacinação. O gap escolhido foi o menor que gerou uma diferença expressiva, $g = 10$ semanas.

A seleção do número de lags p para cada modelo, foi feita com base nos critérios de informação, e nos diagnósticos pós estimação. O modelo (1) teve $p = 1$, o resto $p = 2$. Todos os modelos foram estimados por MQO equação-a-equação. Os diagnósticos dos três modelos são similares, e os testes para o modelo (5) são apresentados no apêndice B.

Em seguida os efeitos contemporâneos entre as variáveis foram modelados utilizando a decomposição de Cholesky, estimando os modelos estruturais. Foi feita a decomposição na ordem casos-vacinas-mortes, gerando a seguinte matrizes de relações contemporâneas:

$$\begin{bmatrix} \varphi_{II}^1 & 0 & 0 \\ \varphi_{IV}^1 & \varphi_{VV}^1 & 0 \\ \varphi_{ID}^1 & \varphi_{VD}^1 & \varphi_{DD}^1 \end{bmatrix} \quad (6)$$

Todo o trabalho computacional foi realizado utilizando o *software* R com os pacotes *vars*,

svars, e *varutils*. O repositório de github com o código e outras informações pode ser encontrado [neste link](#).

6 Resultados e discussão

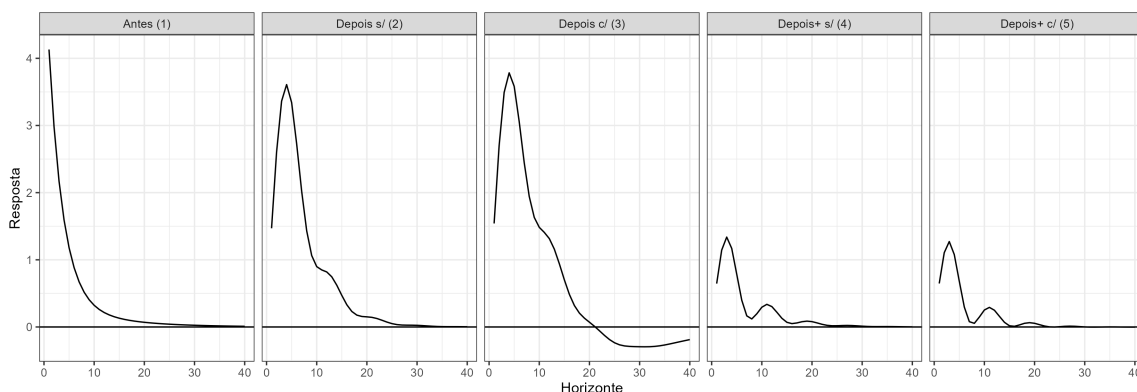
A seção 6.1 calcula as IRF's dos cinco modelos VAR, utilizando os modelos SVAR, mostrando como a inclusão de vacinas afeta as funções. A seção 6.2 constrói contrafactuais da não existência de vacinas com vários métodos diferentes, utilizando os modelos VAR, obtendo estimativas do efeito total da política de vacinação. O apêndice A mostra a análise não prioritária das IRF's do modelo (SVAR) (5).

Antes de apresentar os resultados principais, vale a pena mostrar como correlação entre $\hat{\varepsilon}_{Casos}^i$ e $\hat{\varepsilon}_{Mortes}^i$ ⁶ diminuiu após o início da vacinação. Nos três modelos VAR, respectivamente: $COR(\hat{\varepsilon}_{Mortes}^i, \hat{\varepsilon}_{Casos}^i) = 0,91, 0,46, 0,48, 0,30$, e $0,30$. Esse fato é uma evidência preliminar sobre como a vacina alterou a relação entre os choques das variáveis.

6.1 Resposta de Mortes à Casos Pré e Pós Vacinação

O resultado principal deste trabalho é mostrar a diferença do impacto de um choque de casos pré e pós vacinação, algo que fica claro na figura 5, que mostra a função de resposta (em mortes) ao impulso (de 1000 casos), para os cinco modelos⁷.

Figura 5 – IRF casos em mortes - modelos (1) a (5)



Antes da vacinação, um choque de 1000 casos gerava 4,13 mortes na primeira semana. Após a vacina, os modelos sem *gap* indicam um primeiro choque em torno de 1,5, e os com,

⁶ $\hat{\varepsilon}^i$ indica vetor de resíduos do modelo (i).

⁷Intervalos de confiança não foram calculados por imprecisões nos pacotes disponíveis no R.

um choque próximo de 0,64.

Só o pico não passa toda a mudança de dinâmica, a vacinação pode afetar também a trajetória das funções: existem os efeitos sobre a confiança e distanciamento; o tempo de cama antes do falecimento pode aumentar; além de possivelmente afetar a própria taxa de mortalidade.

Por esse motivo, vale também analisar as mudanças nas somas dos valores das funções (tabela 1). A soma da resposta à 1000 casos é de 16,91 mortes, reduzindo para 7,99 no modelo intermediário, e 6,67 no último modelo (ambos com *gap*), representando uma diminuição na resposta de 10,24 mortes (ou 40%).

Os modelos sem *gap* mostram um aumento grande nas respostas, totalizando, contraintuitivamente, um maior número de mortes. Esse resultado deve estar relacionado à má separação dos momentos pré e pós desses modelos, indicando a superioridade dos modelos com *gap*.

Tabela 1 – Variações na Soma das IRF's

Modelo	Estimativa	(1) - (i)	%
(1)	16,91	0	100%
(2)	27,81	< 0	< 0
(3)	26,91	< 0	< 0
(4)	7,99	8,92	53%
(5)	6,67	10,24	60%

É importante lembrar que as IRF's analisadas são aproximações lineares das respostas durante o período analisado, sem permitir variações nessas funções com base nos níveis das outras variáveis, algo que o modelo SIRDV assumia ser importante. Se, por exemplo, a resposta à vacinas cai muito conforme o número de imunizados sobe, essa metodologia deve apresentar um resultado muito superestimado.

Por último, note que estamos interpretando a IRF sem vacina como um contrafactual da “IRF no período pós vacinas, mas no mundo onde a vacinação não ocorreu”⁸. Com isso em mente, podemos extrapolar os resultados dessa seção, e multiplicar a IRF sem vacinas pelo total de casos⁹. Os resultados desse método são apresentados no fim da seção seguinte.

⁸A princípio essa interpretação faz sentido, como discutido no final da seção 3.2

⁹Dividido por mil, uma vez que essa era o tamanho do choque.

6.2 Análise de Contrafactual

Para ter uma medida do efeito total da política de vacinação, é possível construir um contrafactual onde nunca houve a vacinação, e comparar com os casos e mortes observados. Para isso, foram elencados alguns métodos diferentes: (i) utilizar o modelo (1) para “prever” o período pós vacinas; (ii) utilizar o modelo (5)¹⁰ para “prever” a série de casos e mortes (a partir do ponto pré vacinas), mas zerando o efeito das vacinas; (iii) usar o contrafactual de casos como “choque total” e multiplicar pela IRF sem vacinas da seção anterior. (i*), (ii*), e (iii*) envolvem as mesmas duas abordagens anteriores, mas em vez de calcular a variável de casos com o modelo, utilizar a variável de casos observada para calcular um contrafactual apenas para mortes.

Alguns comentários sobre cada abordagem:

- Os contrafactuais são diferentes:
 - (i*) a (iii*) estimam “qual seria o número de mortes no ‘mundo’ sem vacinas, mas onde os casos se mantiveram iguais?”;
 - (i) a (iii) estimam “qual seria o número de mortes no mundo sem vacinas, dado que a falta de vacinas também afeta o número de casos?”;
- (i) a (iii) são mais “completos” mas dependem da problemática interpretação da variável de casos. (i*) a (iii*) são mais “conservadores”, mas mais precisos.
- (i*) a (iii*) também são mais conservadores por já carregar a mudança na relação entre as variáveis em seus coeficientes;
- (i) e (i*) foram estimados com metade das observações de (ii) e (ii*). Adicionalmente, (i) não tem muita informação para “prever” trajetórias, e deve ser um método ainda mais errático;
- (iii) e (iii*) são métodos menos tradicionais, extrapolando os choques individuais ao nível agregado, com alto potencial de superestimação.
- Por fim, todos os modelos utilizaram as tendências determinísticas. Embora no contrafactual as ondas poderiam não ter ocorrido, os métodos não tem poder para determinar isso¹¹.

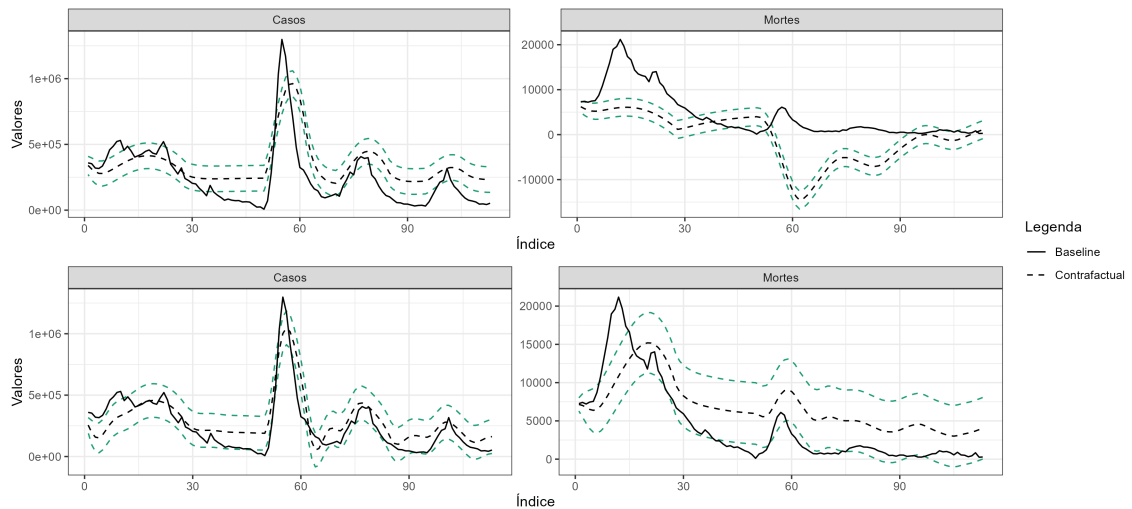
Em suma, estamos mais interessados no método (ii*) como o mais conservador, (ii) como uma opção menos conservadora, e (iii*) como uma abordagem adicional, potencialmente superestimada.

Os resultados dos métodos (i) e (ii) podem ser vistos na figura 6. A figura apresenta o intervalo de confiança de 80%. Como esperado, o método (i) é bastante errático, mas o outro segue as séries realizadas com muito mais precisão.

¹⁰Os resultados utilizando o modelo (3) são similares

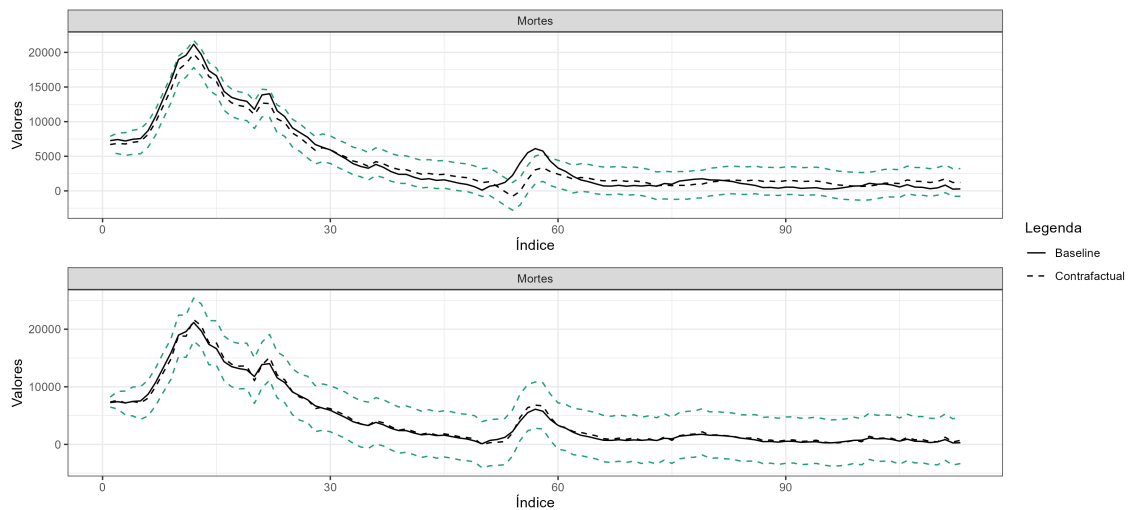
¹¹Adicionalmente, as ondas estão relacionadas ao surgimento de variantes, algo exógeno ao contrafactual.

Figura 6 – Contrafactuais (i) e (ii)



Os resultados dos métodos (i*) e (ii*) podem ser vistos na figura 7. Os contrafactuais são realmente pouco disruptivos.

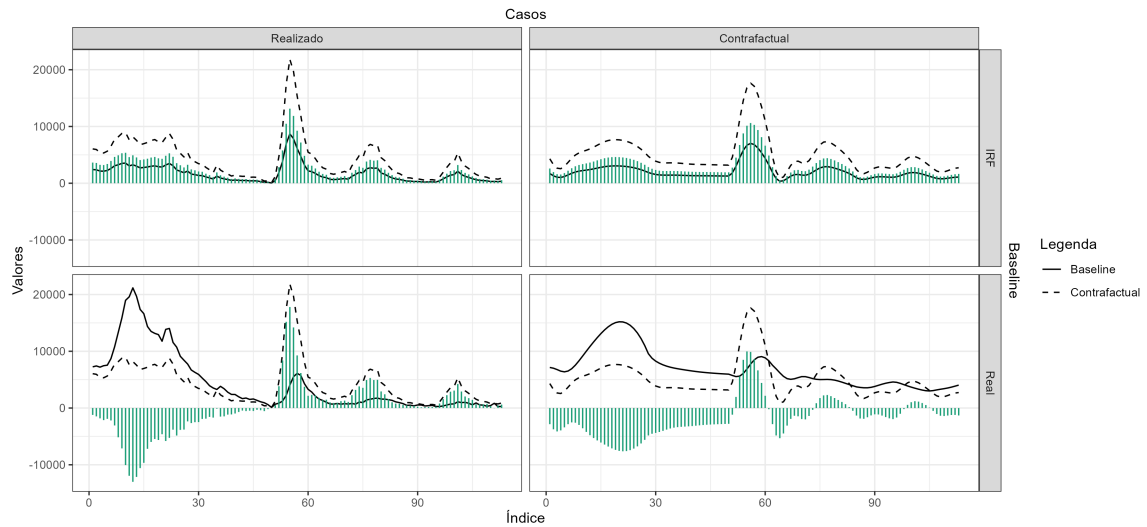
Figura 7 – Contrafactuais (i*) e (ii*)



A construção do contrafactual multiplicando a série de casos pela IRF sem vacinas é limitada. A IRF é uma média das respostas, e na realidade, mortes não respondem sempre da mesma forma. na figura 8, realmente vemos o descasamento das séries. Isso motiva a criação de outro *baseline*, mais comparável, multiplicando a série de casos pela IRF com vacinas (modelo (5)). Embora essa comparação não seja exatamente na mesma ordem das outras abordagens, efetivamente transformamos a disparidade das IRF's em uma medida de quantidade total¹².

¹²Os segmentos em verde são a diferença contrafactual – *baseline*

Figura 8 – Contrafactuais (iii) e (iii*)



A tabela 2 apresenta os modelos (ii), (ii*), (iii), e (iii*). Vemos que o contrafactual de casos realmente é limitado, apresentando uma banda de 30 milhões de casos, enfraquecendo a confiança nos métodos que o utilizam. A abordagem (ii*) apresenta um resultado bastante conservador, 16 mil “vidas salvas”¹³. Os outros métodos, por sua vez, realmente apresentaram números mais expressivos, tanto pelo uso do contrafactual de casos, quanto pela aproximação feita na IRF.

Tabela 2 – Análise de Contrafactual

Método	Série	Diferença	Int. inferior	Int. superior
(ii)	Casos	4.873.072	-10.424.667	20.170.811
	Mortes	279.473	-159.341	718.288
(ii*)	Mortes	16.024	-422.701	454.750
(iii)	Mortes	289.089	-	-
(iii*)	Mortes	338.402	-	-

7 Conclusão

Nesta tese, foram desenvolvidas diferentes abordagens para medir medido efeito das vacinas no número de casos mortes que o Brasil apresentou durante a pandemia do COVID-19.

¹³Um valor substancialmente menor do que aquele calculado por [Ferreira et al., 2022](#)

O uso de um modelo SVAR se mostrou atrativo. O modelo teórico SIRDV e os dados explicitaram a natureza de dinâmica temporal entre as variáveis; a teoria epidemiológica motivou a identificação estrutural; e foi possível obter diversas ordens de resultados.

A abordagem mais robusta comparou as IRF's de SVAR's estimados antes e depois do início da vacinação, controlando possíveis falhas na especificação do modelo ao analisar como a vacina o altera. Essa abordagem mostrou a mudança na soma da resposta em mortes, dado um choque de 1000 casos, é de 10, 24 em termos absolutos, e 60% em termos proporcionais, indicando o alto nível de eficiência da política de vacinação.

A outra abordagem foi a criação de um contrafactual para estimar o efeito agregado da vacinação. Foi identificado um *lower* e um *upper bound* de, respectivamente 16 mil e 279 mil vidas salvas até mar/2023. O valor real deve se encontrar longe dos *bounds* uma vez que os valores apresentam indícios de sub/superestimação (respectivamente). Também foi criado um contrafactual com base na redução da IRF comentada anteriormente, e este é mais consistente com uma redução da ordem do supracitado *upper bound*.

Esta tese apresentou mais uma metodologia para atestar a eficiência da vacinação, servindo como mais um insumo para a busca de um consenso na literatura e fim da desinformação, além de ser mais uma ferramenta para os *policymakers* definirem o gasto e urgência ótima para as políticas de vacinação.

O potencial da metodologia não foi explorado até o fim, existe espaço para a busca de novas técnicas de identificação, busca de dados de melhor qualidade, inclusão do distanciamento social e não linearidade nos modelos. Com isso, seria possível aprimorar as estimativas obtidas, e possibilitar analisar as IRF's de vacinas em mortes e casos, gerando um resultado de efeito marginal muito útil. Similarmente, existe muito espaço para aplicações dessa metodologia em outros contextos para evidenciar seu bom funcionamento.

Referências

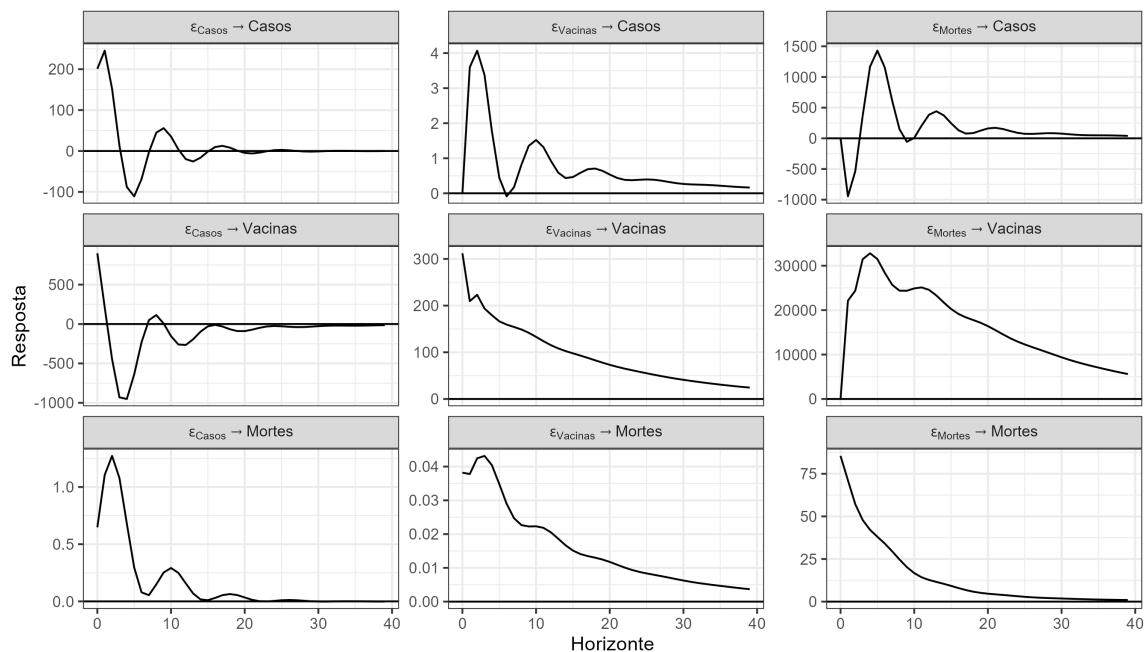
- Baker, R. (2020). Reactive Social distancing in a SIR model of epidemics such as COVID-19. <https://doi.org/10.48550/ARXIV.2003.08285>
- Barro, R. J. (2022). *Vaccination Rates and COVID Outcomes across U.S. States* (rel. técn.). National Bureau of Economic Research. <https://doi.org/10.3386/w29884>
- Byun, W. S., Heo, S. W., Jo, G., Kim, J. W., Kim, S., Lee, S., Park, H. E., & Baek, J.-H. (2021). Is coronavirus disease (COVID-19) seasonal? A critical analysis of empirical and epidemiological studies at global and local scales. *Environmental Research*, 196, 110972. <https://doi.org/https://doi.org/10.1016/j.envres.2021.110972>
- Cabrera, M., Córdova-Lepe, F., Gutiérrez-Jara, J. P., & Vogt-Geisse, K. (2021). An SIR-type epidemiological model that integrates social distancing as a dynamic law based on point prevalence and socio-behavioral factors. *Scientific Reports*.
- Chen, X., Huang, H., Ju, J., Sun, R., & Zhang, J. (2022). Impact of vaccination on the COVID-19 pandemic in U.S. states. *Scientific Reports*, 12(1), 1554. <https://doi.org/10.1038/s41598-022-05498-z>
- Escobar, L. E., Molina-Cruz, A., & Barillas-Mury, C. (2020). BCG vaccine protection from severe coronavirus disease 2019 (COVID-19). *Proceedings of the National Academy of Sciences*, 117(30), 17720–17726. <https://doi.org/10.1073/pnas.2008410117>
- Ferreira, L. S., Darcie Marquitti, F. M., da Silva, R. L. P., Borges, M. E., Costa Gomes, M. F. d., Cruz, O. G., Kraenkel, R. A., Coutinho, R. M., Prado, P. I., & Bastos, L. S. (2022). Estimating the impact of implementation and timing of COVID-19 vaccination programme in Brazil: a counterfactual analysis. *medRxiv*. <https://doi.org/10.1101/2021.12.24.21268384>
- Halloran, M. E., Longini, J., Ira M., & Struchiner, C. J. (1999). Design and Interpretation of Vaccine Field Studies. *Epidemiologic Reviews*, 21(1), 73–88. <https://doi.org/10.1093/oxfordjournals.epirev.a017990>
- Moraes, R. (2021). NT DINTE 31 - A Segunda Onda da Pandemia (Mas Não do Distanciamento Físico): Covid-19 E Políticas de Distanciamento Social dos Governos Estaduais no Brasil. *Notas Técnicas*, 1–18. <https://doi.org/10.38116/ntdinte31>
- Mwalili, S., Kimathi, M., Ojiambo, V., Gathungu, D., & Mbogo, R. (2020). SEIR model for COVID-19 dynamics incorporating the environment and social distancing. *BMC Research Notes*. <https://doi.org/10.1186/s13104-020-05192-1>
- Reluga, T. C. (2010). Game theory of social distancing in response to an epidemic. *PLoS computational biology*.

- Ritchie, H., Mathieu, E., Rod  s-Guirao, L., Appel, C., Giattino, C., Ortiz-Ospina, E., Hasell, J., Macdonald, B., Beltekian, D., & Roser, M. (2020). Coronavirus Pandemic (COVID-19). *Our World in Data*. <https://ourworldindata.org/coronavirus>
- Wiemken, T. L., Khan, F., Puzniak, L., Yang, W., Simmering, J., Polgreen, P., Nguyen, J. L., Jodar, L., & McLaughlin, J. M. (2023). Seasonal trends in COVID-19 cases, hospitalizations, and mortality in the United States and Europe. *Sci Rep*, 13(1), 3886.

A IRF's Individuais

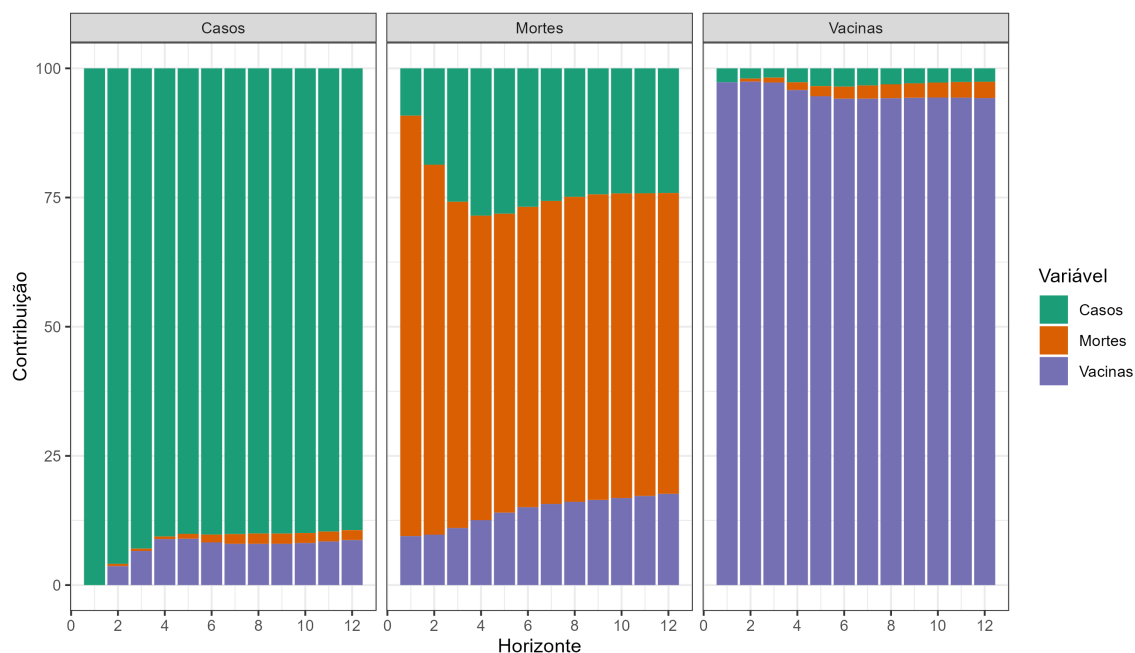
Como foi explicado, a identificação dos choques não é muito robusta, e existem variáveis omitidas no modelo, de modo que a análise das IRF's individualmente não é prioritária. De qualquer forma, pode-se vê-las na figura 9. De fato, contra o que as análises mais robustas indicam, o efeito da vacina em casos e mortes não se mostra negativo.

Figura 9 – IRF's do modelo (5)



Outro possível objeto de interesse são as *Forecast Error Variance Decomposition* (FEVD), que indicam a fração da contribuição de cada variável para a variância da previsão de cada equação. A figura 10 As vacinas parecem explicar uma maior porcentagem da variância nas mortes do que dos casos, embora não seja o *driver* majoritário em nenhum dos dois.

Figura 10 – FEVD's do modelo (5)



B Diagnósticos do VAR

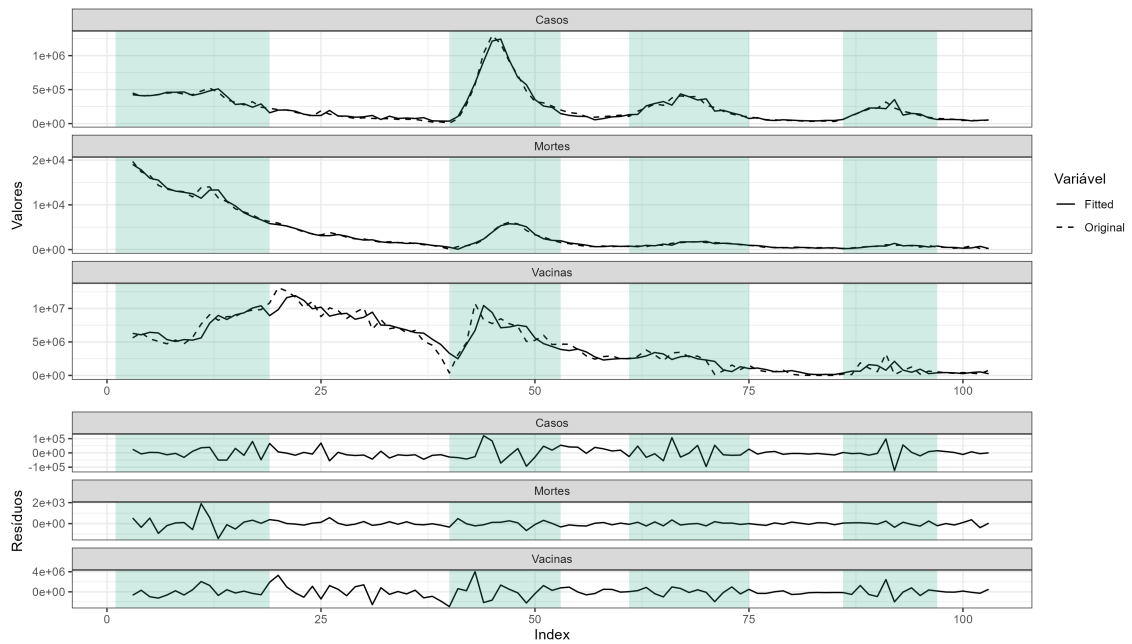
Antes de sua inclusão no VAR (5), foi realizado um teste de raiz unitária sobre as séries, apresentado na tabela 3. Podemos ver que apenas a variável de casos é estacionária. Uma boa especificação do modelo pareceu lidar bem com essa limitação.

Tabela 3 – Teste Dickey-Fuller Aumentado

Variável	Estatística	Lag	P-valor
Casos	-3.61	5	0.03
Vacinas	-2.97	4	0.17
Mortes	-2.71	5	0.28

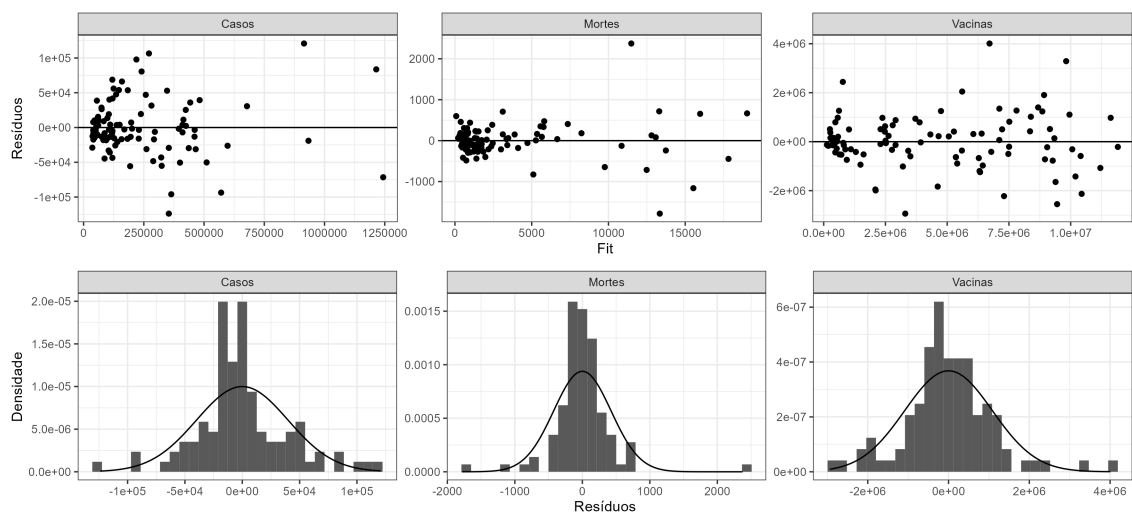
A comparação do fit do modelo VAR e os valores reais, bem como o resíduo resultante, está presente na figura 11. A distribuição dos resíduos parece relativamente ter uma volatilidade similar ao longo da reta de valores fitados, e de fato, o teste ARCH-LM não indicou a presença de erros ARCH no modelo – estatística do teste de 206, 61, com 180 graus de liberdade, gerando um p-valor $< 0,0848$. O R quadrado de cada equação é: Casos 0,960, Vacinas 0,918, Mortes 0,988

Figura 11 – Fit e resíduos



Podemos ver mais detalhes dos resíduos na figura 12, que mostra uma variância dos resíduos não tão dependente dos valores do modelo, mas uma distribuição não exatamente normal. Realmente, o teste Jarque-Bera multivariado indica erros não normais – estatística do teste de 42,113, com 6 graus de liberdade, gerando um p-valor de $1,747e-07$. Dito isso, esse resultado não compromete nenhuma hipótese deste trabalho.

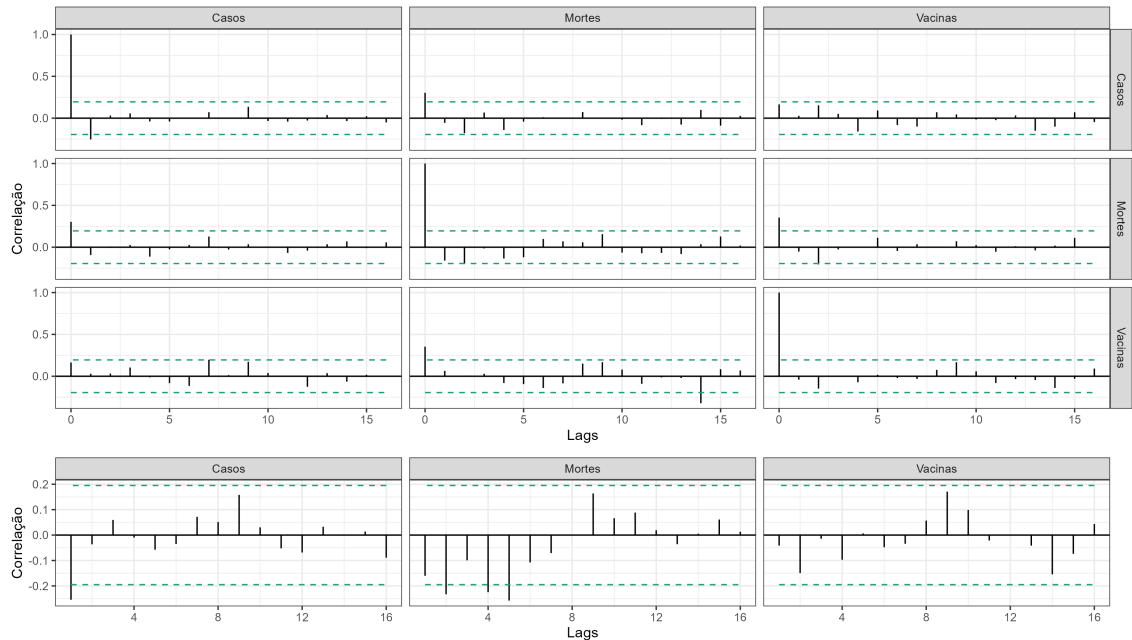
Figura 12 – Dispersão e distribuição dos resíduos



A figura 13 mostra as autocorrelações dos resíduos, indicando dependência temporal insignificante das e entre as variáveis. De fato, o modelo passou teste Portmanteau de autocorrelação

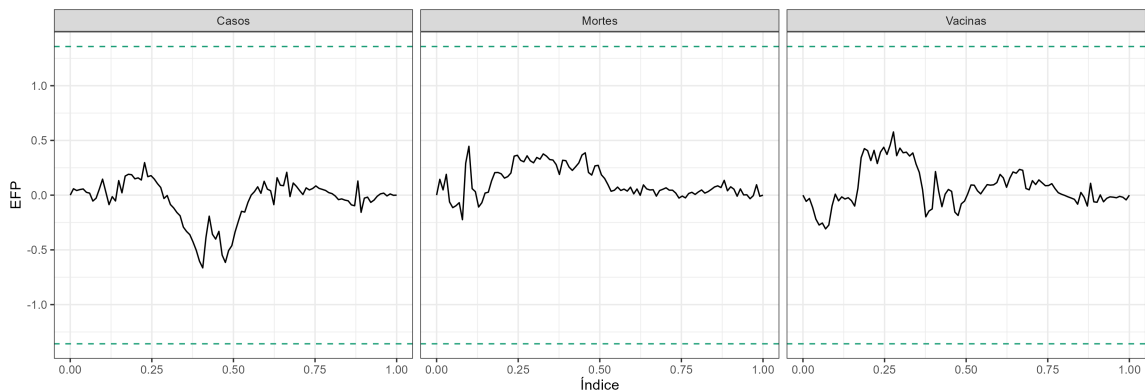
dos resíduos (em até 12 lags) – estatística do teste de 123, 54, com 126 graus de liberdade, gerando um p-valor = 0,3215. Similarmente, o teste Edgerton-Shukur também rejeitou a presença de autocorrelação.

Figura 13 – Correlação cruzada e parcial dos resíduos



A figura 14 mostra a estabilidade dos resíduos do modelo, mais especificamente, o resultado da função “stability” do pacote “vars” com o método “CUMSUM-OLS” no R. Ela indica que alterar a amostra que baseou a estimativa do modelo não altera drasticamente os valores dos coeficientes.

Figura 14 – Estabilidade dos resíduos



Vale atentar que as raízes do VAR são todas menores que 1: $0,909(< 1)$, $0,886(< 1)$, $0,886(< 1)$, $0,754(< 1)$, $0,754(< 1)$, $0,563(< 1)$, $0,563(< 1)$, $0,522(< 1)$, $0,522(< 1)$.