

Os pacotes: Diférenças entre Causal Impact

André F. B. Menezes | Junho, 2020

Os artigos

4 *Int. J. Mathematical Modelling and Numerical Optimisation*, Vol. 5, Nos. 1/2, 2014

Predicting the present with Bayesian structural time series

Steven L. Scott* and Hal R. Varian

Google, Inc.,
1600 Amphitheatre Parkway,
Mountain View, CA 94043, USA
E-mail: stevescott@google.com
E-mail: hal@google.com

*Corresponding author

Abstract: This article describes a system for short term forecasting based on an ensemble prediction that averages over different combinations of predictors. The system combines a structural time series model for the target series with a regression component capturing the contributions of contemporaneous search query data. A spike-and-slab prior on the regression coefficients induces sparsity, dramatically reducing the size of the regression problem. Our system averages over potential contributions from a very large set of models and gives easily digested reports of which coefficients are likely to be important. We illustrate with applications to initial claims for unemployment benefits and to retail sales. Although our exposition focuses on using search engine data to forecast economic time series, the underlying statistical methods can be applied to more general short term forecasting with large numbers of contemporaneous predictors.

The Annals of Applied Statistics
2015, Vol. 9, No. 1, 247–274
DOI: 10.1214/14-AOS788
© Institute of Mathematical Statistics, 2015

INFERRING CAUSAL IMPACT USING BAYESIAN STRUCTURAL TIME-SERIES MODELS

BY KAY H. BRODERSEN, FABIAN GALLUSSER, JIM KOEHLER,
NICOLAS REMY AND STEVEN L. SCOTT

Google, Inc.

An important problem in econometrics and marketing is to infer the causal impact that a designed market intervention has exerted on an outcome metric over time. This paper proposes to infer causal impact on the basis of a diffusion-regression state-space model that predicts the counterfactual market response in a synthetic control that would have occurred had no intervention taken place. In contrast to classical difference-in-differences schemes, state-space models make it possible to (i) infer the temporal evolution of attributable impact, (ii) incorporate empirical priors on the parameters in a fully Bayesian treatment, and (iii) flexibly accommodate multiple sources of variation, including local trends, seasonality and the time-varying influence of contemporaneous covariates. Using a Markov chain Monte Carlo algorithm for posterior inference, we illustrate the statistical properties of our approach on simulated data. We then demonstrate its practical utility by estimating the causal effect of an online advertising campaign on search-related site visits. We discuss the strengths and limitations of state-space models in enabling causal attribution in those settings where a randomised experiment is unavailable. The CausalImpact R package provides an implementation of our approach.

Organização

- Contextualização
- Modelos e Inferência
- Inferência Causal
- Aplicações
 - Campanha publicitária
 - SRAG

Contextualização

Soluções Computacionais

- JSS Volume 41 Statistical Software for State Space Methods (**Commandeur, Koopman, and Ooms, 2011**).
- **BATS**: escrito na década de 90 (**Pole, West, and Harrison, 1994**).
- **R**:
 - **d1m**: inferência frequentista e Bayesiana (Gibbs e ARMS).
 - **bsts**: inferência Bayesiana via MCMC, com implementações em C++.
 - **KFAS**: algoritmos de Kalman (*filter, smooth*) para modelos da família exponencial.
 - **bssm**: inferência Bayesiana para modelos não-lineares e não-gaussianos.
- **Ox**: SsfPack escrito em C por Siem Jan Koopman.
- **SAS**:
 - **SAS/ETS**: *procedures* SSM, STATESPACE e UCM. Implementações baseadas no livro de **Harvey (1989)**.
 - **SAS/IML**: *subroutines* KALDFS e KALDFF. Implementações baseadas em **Jong (1991)**.

Modelos

Modelos de espaço de estados

- Os modelos definidos em **West and Harrison (1997)** são também conhecidos como modelos de espaço de estados.
- Seguindo **Harvey (1989)** e **Durbin and Koopman (2001)**:

- Equação de observação:

$$y_t = Z_t \alpha_t + \epsilon_t, \quad \epsilon_t \sim N(0, H_t)$$

- Equação de transição:

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t \quad \eta_t \sim N(0, Q_t)$$

- Informação inicial:

$$\alpha_1 \sim N(a_1, P_1)$$

sendo que a_1 e P_1 são conhecidos.

- R_t em geral é a matriz identidade.
- **Durbin and Koopman (2001)** discutem abordagens para inicializar os algoritmos de Kalman (*filter* e *smooth*) quando a_1 e P_1 são desconhecidos.

Notação

Durbin e Koopman (2001)	West e Harrison (1997)
α_t	θ_t
Z_t	F_t^\top
ϵ_t	v_t
H_t	V_t
T_t	G_{t+1}
$R_t \eta_t$	w_{t+1}
R_t, Q_t	$W_{t+1} (= R_t Q_t R_t^\top)$
a_t	a_t
P_t	R_t
$a_{t t}$	m_t
$P_{t t}$	C_t

bsts

- Introduzido por **Steven L. Scott** quando trabalhava no Google.

bsts: Bayesian Structural Time Series

Time series regression using dynamic linear models fit using MCMC. See Scott and Varian (2014) <[doi:10.1504/IJMMNO.2014.059942](https://doi.org/10.1504/IJMMNO.2014.059942)>, among many other sources.

Version: 0.9.5
Depends: [BoomSpikeSlab](#) (≥ 1.2.3), [zoo](#) (≥ 1.8), [xts](#), [Boom](#) (≥ 0.9.6), R (≥ 3.4.0)
LinkingTo: [Boom](#) (≥ 0.9.6)
Suggests: [testthat](#)
Published: 2020-05-02
Author: Steven L. Scott
Maintainer: Steven L. Scott <steve.the.bayesian@gmail.com>
License: [LGPL-2.1](#) | file [LICENSE](#)
NeedsCompilation: yes
In views: [Bayesian](#), [TimeSeries](#)
CRAN checks: [bsts results](#)

Downloads:

Reference manual: [bsts.pdf](#)
Package source: [bsts_0.9.5.tar.gz](#)
Windows binaries: r-devel: [bsts_0.9.5.zip](#), r-release: [bsts_0.9.5.zip](#), r-oldrel: [bsts_0.9.5.zip](#)
macOS binaries: r-release: [bsts_0.9.5.tgz](#), r-oldrel: [bsts_0.9.5.tgz](#)
Old sources: [bsts archive](#)

Reverse dependencies:

Reverse depends: [CausalImpact](#)
Reverse imports: [cbar](#), [MarketMatching](#)

Linking:

Please use the canonical form <https://CRAN.R-project.org/package=bsts> to link to this page.

bsts

```
y <- my_data$ResponseVariable
ss <- AddLocalLinearTrend(
  state.specification = list(),
  y = y
)
ss <- AddSeasonal(
  state.specification = ss,
  y = y,
  nseasons = 7
)
model <- bst(
  formula = y ~ .,
  state.specification = ss,
  niter = 1000,
  data = my.data
)
plot(model)
predict(model, horizon = 7)
```

bsts

Função	Descrição
AddAr	AR(p)
AddLocalLevel	Local level model
AddLocalLinearTrend	Local linear trend
AddStudentLocalLinearTrend	Robust local linear trend
AddSemilocalLinearTrend	Semilocal linear trend
AddSeasonal	Factor seasonal
AddTrig	Trigonometric seasonal
AddRandomWalkHoliday	Random Walk Holiday
AddDynamicRegression	Dynamic regression

Componentes de tendência

- Local linear trend:

$$\begin{aligned}\mu_{t+1} &= \mu_t + \delta_t + \epsilon_t, & \epsilon_t &\sim N(0, \sigma_\mu^2) \\ \delta_{t+1} &= \delta_t + \eta_t, & \eta_t &\sim N(0, \sigma_\delta^2)\end{aligned}$$

- Semilocal linear trend:

$$\begin{aligned}\mu_{t+1} &= \mu_t + \delta_t + \epsilon_t, & \epsilon_t &\sim N(0, \sigma_\mu^2) \\ \delta_{t+1} &= D + \phi(\delta_t - D) + \eta_t, & \eta_t &\sim N(0, \sigma_\delta^2)\end{aligned}$$

em que D é um parâmetro inclinação de longo prazo.

Assumindo as seguintes prioridades:

- D : normal
- ϕ : normal truncada. Se for truncada em $(-1, 1)$ a inclinação exibe variação estacionária de curto prazo em torno da inclinação D .

Componentes de sazonalidade

- Factor seasonal:

$$\gamma_{t+1} = - \sum_{s=2}^S \gamma_{t-s} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\gamma^2)$$

em que S é o número de sazonalidades e γ_t representa a sazonalidade conjunta.

- Trigonometric Seasonal:

$$\begin{aligned} \gamma_{j,t+1} &= \gamma_{jt} \cos(\lambda_j) + \gamma_{jt}^* \sin(\lambda_j) + \omega_{jt}, \\ \gamma_{j,t+1}^* &= -\gamma_{jt} \sin(\lambda_j) + \gamma_{jt}^* \cos(\lambda_j) + \omega_{jt}^*, \quad j = 1, \dots, [S/2], \end{aligned}$$

em que $\lambda_j = \frac{2\pi j}{S}$, ω_{jt} e ω_{jt}^* são termos independentes com $N(0, \sigma_\omega^2)$.

Regressão dinâmica

- Seja $x_{j,t}$, $j = 1, \dots, p$ covariáveis no tempo t .

$$\mathbf{x}_t^\top \boldsymbol{\beta}_t = \sum_{j=1}^p x_{j,t} \beta_{j,t}$$

$$\beta_{j,t+1} = \beta_{j,t} + \epsilon_{j,t}$$

com $\epsilon_{j,t} \sim N(0, \sigma_{\beta_j}^2)$.

- Matricialmente têm-se:
 - $\mathbf{Z}_t = \mathbf{x}_t$.
 - $\boldsymbol{\alpha}_t = \boldsymbol{\beta}_t$.
 - $\mathbf{T}_t = \mathbf{I}_{p \times p}$.
 - $\mathbf{Q}_t = \text{diag}(\sigma_{\beta_j}^2)$.

Elicitação da priori

- Seja θ o conjunto de todos os parâmetros do modelo. Em geral, θ contém as variâncias.
- Seja $\alpha = (\alpha_1, \dots, \alpha_m)$ a sequência completa dos parâmetros de estado.
- No bsts a inferência Bayesiana é conduzida especificando uma priori para $p(\theta)$ e $p(\alpha_0 | \theta)$
- Então, amostra-se de $p(\alpha, \theta | \mathbf{y})$ via MCMC.
- Detalhes podem-se ser encontrados em [Durbin and Koopman \(2002\)](#).

Nível e sazonalidade

- bsts utiliza a seguinte priori:

$$\frac{1}{\sigma^2} \sim \text{Gamma} \left(\frac{v}{2}, \frac{s}{2} \right)$$

- Interpretações:
 - s/v é uma estimativa a priori de σ^2 .
 - v é um "peso" associada a estimativa a priori.
- Default:

$$\frac{1}{\sigma^2} \sim \text{Gamma} (10^{-2}, 10^{-2} s_y^2)$$

em que $s_y^2 = \frac{1}{n-1} \sum_t (y_t - \bar{y})^2$.

Regressão (*spike-and-slab prior*)

- Seja $\tau = (\tau_1, \dots, \tau_p)$, em que, $\tau_k = 1$ se $\beta_k \neq 0$ e $\tau_k = 0$ se $\beta_k = 0$.
- Seja β_τ os elementos diferentes de zero do vetor β .
- A priori *spike-and-slab* é definida por

$$p(\beta, \tau, \sigma_\epsilon^2) = p(\tau) p(\sigma_\epsilon^2 \mid \tau) p(\beta_\tau \mid \tau, \sigma_\epsilon^2)$$

- A distribuição marginal $p(\tau)$ é conhecida como "*spike*", e em geral supõe-se que:

$$p(\tau) = \prod_{j=1}^p \pi_j^{\tau_j} (1 - \pi_j)^{1-\tau_j}$$

- Valores de π_j podem ser especificados conforme o número esperado de preditores, M , i.e., $\pi_j = M/p$.

Regressão (*spike-and-slab prior*)

- Seja Ω^{-1} uma matriz simétrica e Ω_τ^{-1} denotando a matriz Ω^{-1} corespondente a $\tau_k = 1$.
- As distribuições prioris para $p(\sigma_\epsilon^2 \mid \tau)$ e $p(\beta_\tau \mid \tau, \sigma_\epsilon^2)$ são expressas via conjugação normal-inversa Gamma.

$$\beta_\tau \mid \sigma_\epsilon^2 \sim N\left(\mathbf{b}_\tau, \sigma_\epsilon^2 [\Omega_\tau^{-1}]^{-1}\right), \quad \frac{1}{\sigma_\epsilon^2} \mid \tau \sim \text{Gamma}\left(\frac{v_\epsilon}{2}, \frac{s_\epsilon}{2}\right)$$

- Especificações dos parâmetros da priori:
 - $s_\epsilon = v_\epsilon (1 - R^2) s_y^2$, sendo R^2 é uma espécie de coeficiente de determinação esperado da regressão, sendo que $R^2 = 0.5$ o *default* do bsts.
 - Para Ω^{-1} utiliza-se uma adaptação da **Zellner's g-prior**, ou seja,

$$\Omega^{-1} = \frac{g}{n} [wX^\top X + (1 - w) \text{diag}(X^\top X)]$$

em que g é interpretado como g observações valiosas de informação e w o peso dado a diagonal de $X^\top X$. Valores *default* são $g = 1$ e $w = 1/2$.

Inferência via MCMC

1. Simule do espaço de estado α a partir de $p(\alpha \mid \mathbf{y}, \theta, \beta, \sigma_\epsilon^2)$ utilizando algoritmo proposto por [Durbin and Koopman \(2002\)](#).
 2. Simule $\theta \sim p(\theta \mid \alpha, \mathbf{y}, \beta, \sigma_\epsilon^2)$.
 3. Simule β e σ_ϵ^2 de uma cadeia de Markov com distribuição estacionária $p(\beta, \sigma_\epsilon^2 \mid \alpha, \mathbf{y}, \theta)$.
- O `bsts` implementa as simulações em `C++`, proporcionando maior eficiência computacional ([Scott and Varian, 2014](#)).

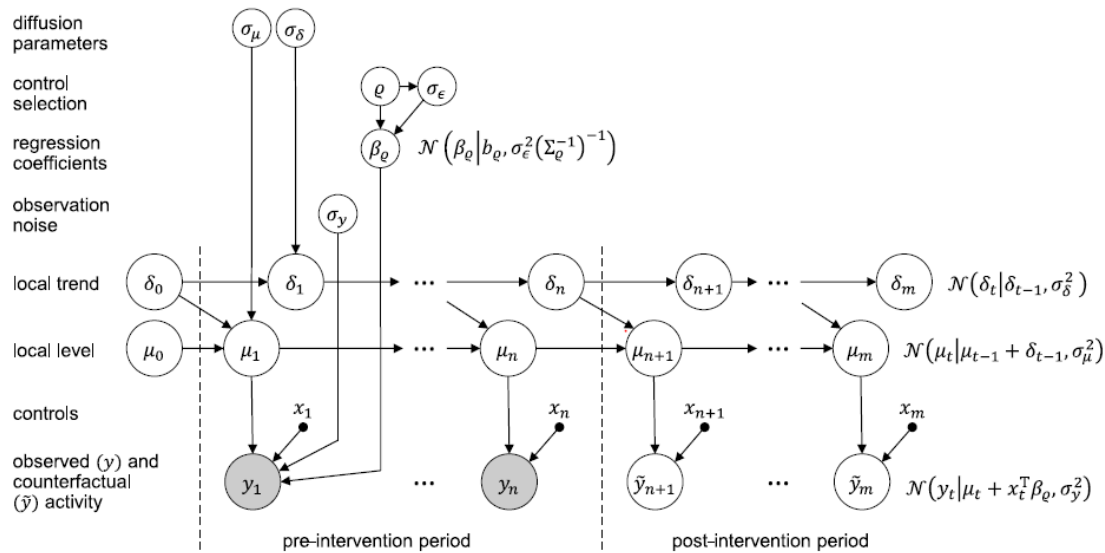
Inferência Causal

Inferência Causal

- O que aconteceria se não houvesse determinado evento?
- Como medir o efeito desse evento?
- **Randomized experiment:** caro, difícil e muitas vezes impraticável!
- Exemplos:
 - Qual a eficácia de determinada publicidade? (marketing).
 - Qual o impacto de determinada lei? (politica).
 - Qual a eficiência de uma intervenção pedagógica? (educação).
- **Problema:** estimar o efeito causal na ausência de um experimento randomizado.

Causal Impact

- Pacote desenvolvido por **Brodersen, Gallusser, Koehler, Remy, and Scott (2015)**, o carro-chefe é o pacote `bsts`.
- O que o pacote faz?
 - Estima o efeito causal de uma **intervenção** em uma série temporal utilizando modelos do `bsts`.
- Como funciona?



CausalImpact

CausalImpact: Inferring Causal Effects using Bayesian Structural Time-Series Models

Implements a Bayesian approach to causal impact estimation in time series, as described in Brodersen et al. (2015) <[doi:10.1214/14-AOAS788](https://doi.org/10.1214/14-AOAS788)>. See the package documentation on GitHub <<https://github.com/google/CausalImpact>> to get started.

Version: 1.2.4
Depends: [bsts](#) (≥ 0.9.0)
Imports: [assertthat](#) (≥ 0.2.0), [Boom](#), [dplyr](#), [ggplot2](#), [zoo](#)
Suggests: [knitr](#), [testthat](#)
Published: 2020-01-08
Author: Kay H. Brodersen, Alain Hauser
Maintainer: Alain Hauser <alhauser@google.com>
License: [Apache License 2.0](#) | file [LICENSE](#)
Copyright: Copyright (C) 2014-2019 Google, Inc.
URL: <https://github.com/google/CausalImpact/>
NeedsCompilation: no
Citation: [CausalImpact citation info](#)
Materials: [README](#)
CRAN checks: [CausalImpact results](#)

Downloads:

Reference manual: [CausalImpact.pdf](#)
Vignettes: [CausalImpact](#)
Package source: [CausalImpact 1.2.4.tar.gz](#)
Windows binaries: r-devel: [CausalImpact 1.2.4.zip](#), r-release: [CausalImpact 1.2.4.zip](#), r-oldrel: [CausalImpact 1.2.4.zip](#)
macOS binaries: r-release: [CausalImpact 1.2.4.tgz](#), r-oldrel: [CausalImpact 1.2.4.tgz](#)
Old sources: [CausalImpact archive](#)

Reverse dependencies:

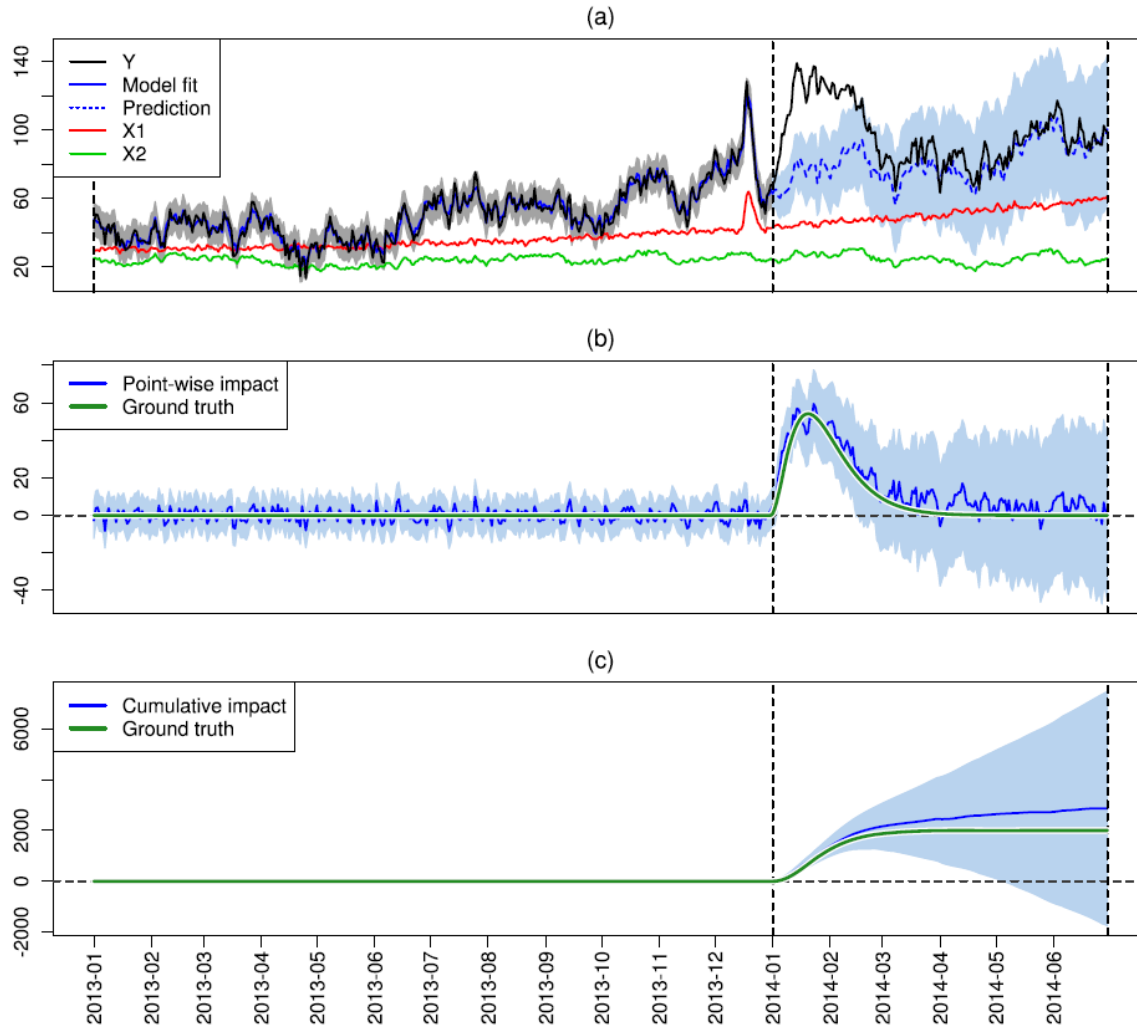
Reverse imports: [MarketMatching](#)

Linking:

Please use the canonical form <https://CRAN.R-project.org/package=CausalImpact> to link to this page.

```
pre.period <- c(1, 70)
post.period <- c(71, 100)
impact <- CausalImpact(data, pre.period, post.period)
summary(impact, "report")
plot(impact)
```

Causal Impact

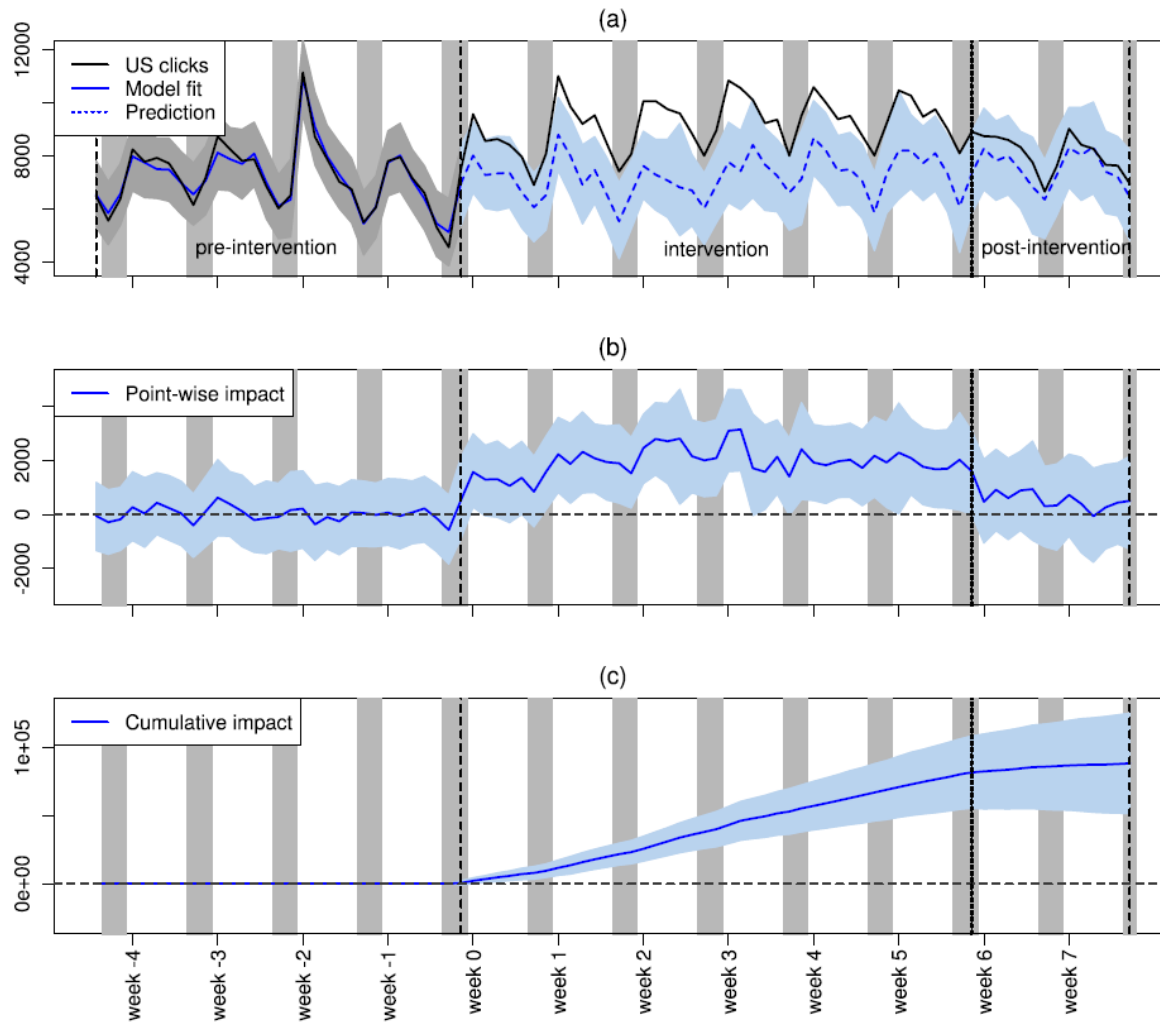


Aplicações

Campanha publicitária

- Realizada por anunciantes do Google no Estado Unidos.
- **Objetivo:** Inferir o efeito causal da campanha no número de vezes que um usuário foi direcionado ao site do anunciante na página de resultados de pesquisa do Google.
- Aleatoriamente 95 das 190 áreas de marketing receberam os anúncios.
- **Variável resposta:** número de cliques.
- Covariáveis: número de cliques das campanhas não afetadas pelos anúncios.
- O modelo:
 - Nível local: $s/v = 0.1\sigma_y$ e $v = 32$.
 - Regressão estática: $M = 3$, $R^2 = 0.8$ e $v_\epsilon = 50$.

Campanha publicitária

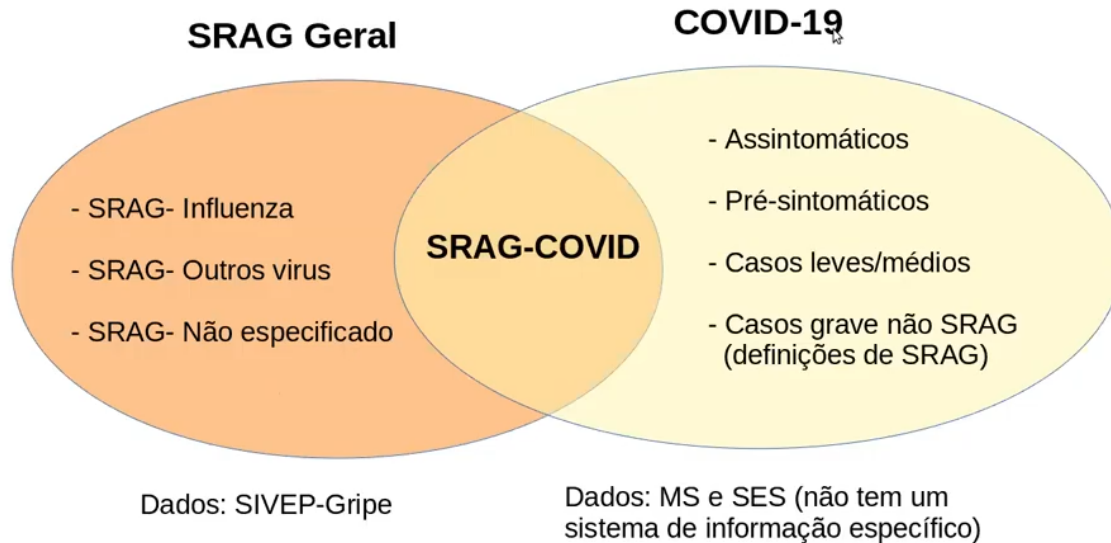


SRAG

- Síndrome Respiratória Aguda Grave (SRAG), no Brasil, se refere há uma síndrome (conjunto de sintomas).
- Indivíduo hospitalizado com:
 - Febre.
 - Tosse ou dor de garganta.
 - Dispneia ou saturação de $O_2 < 95\%$ ou desconforto respiratório.
- Indivíduo que evolui para óbito por SRAG independente de internação.
- Por lei, SRAG tem notificação obrigatória.

Dados

- Pesquisadores da Fiocruz desenvolveram o sistema **InfoGripe** que armazena dados de SRAG.
- Desde 21/03/2020 o SARS-COV-2 entrou na lista de vírus que são testados.



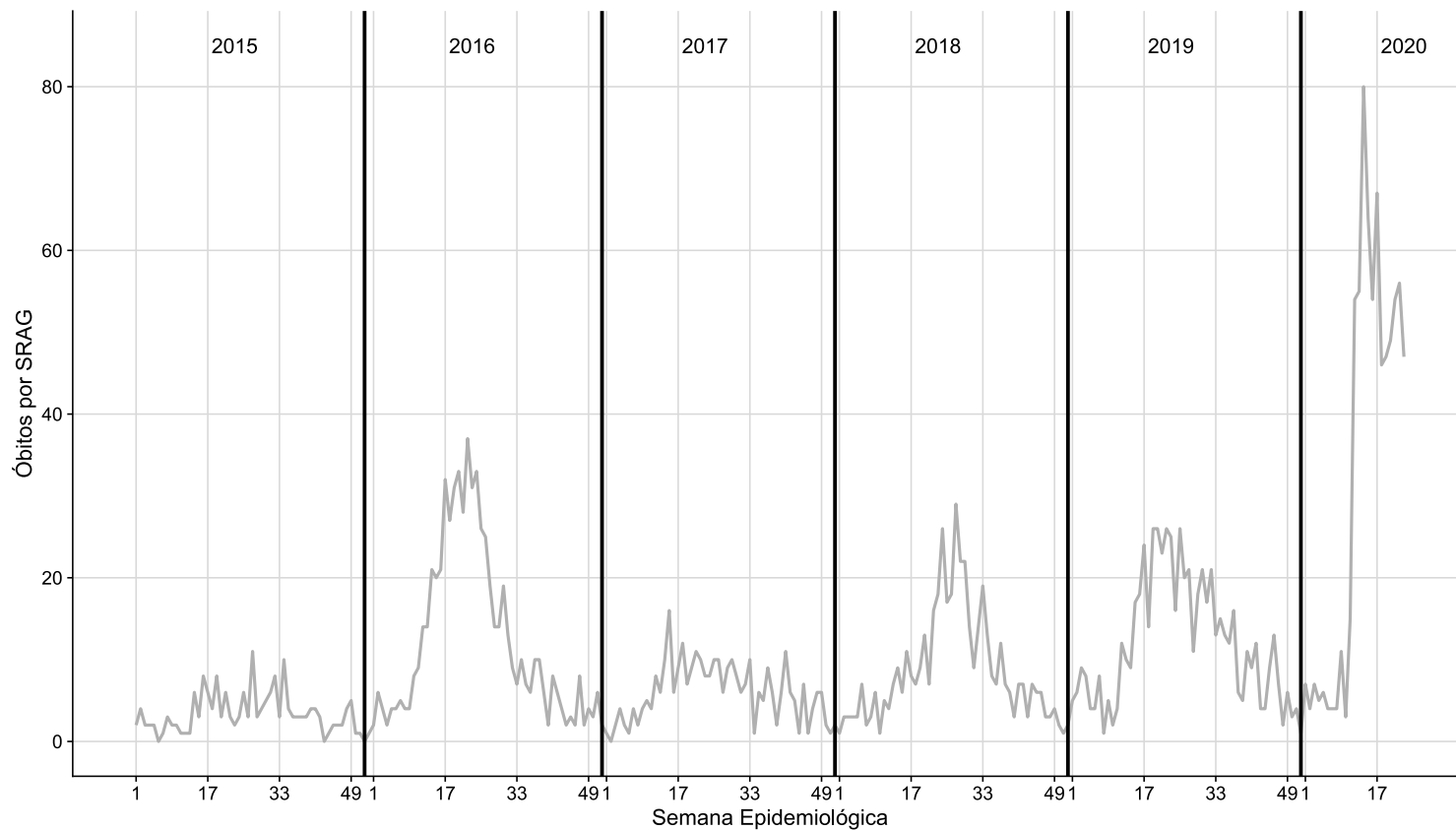
- **Objetivo:** inferir o efeito causal da pandemia do coronavírus no número de óbitos por SRAG no estado do Paraná.

Dados

- Consideramos os óbitos por SRAG no Paraná durante as semanas epidemiológicas dos anos de 2015 a 2020.

Ano	Semanas	Óbitos	Casos
2015	52	179	1950
2016	52	641	4915
2017	52	312	2978
2018	52	449	4659
2019	52	622	5581
2020	23	743	5254

Dados



Modelo

- O período de 2015 a 2019 foi utilizado como "pré-intervenção" no sentido de que não houve pandemia (mesmo sabendo que 2016 tivemos H1N1).
- O modelo considerado consiste em uma componente de tendência mais sazonalidade via Fourier.
- Seguindo a notação de **West and Harrison (1997)** temos:

- $\boldsymbol{\theta}_t = (\theta_{tT}, \theta_{tS_1}, \dots, \theta_{tS_{51}})$

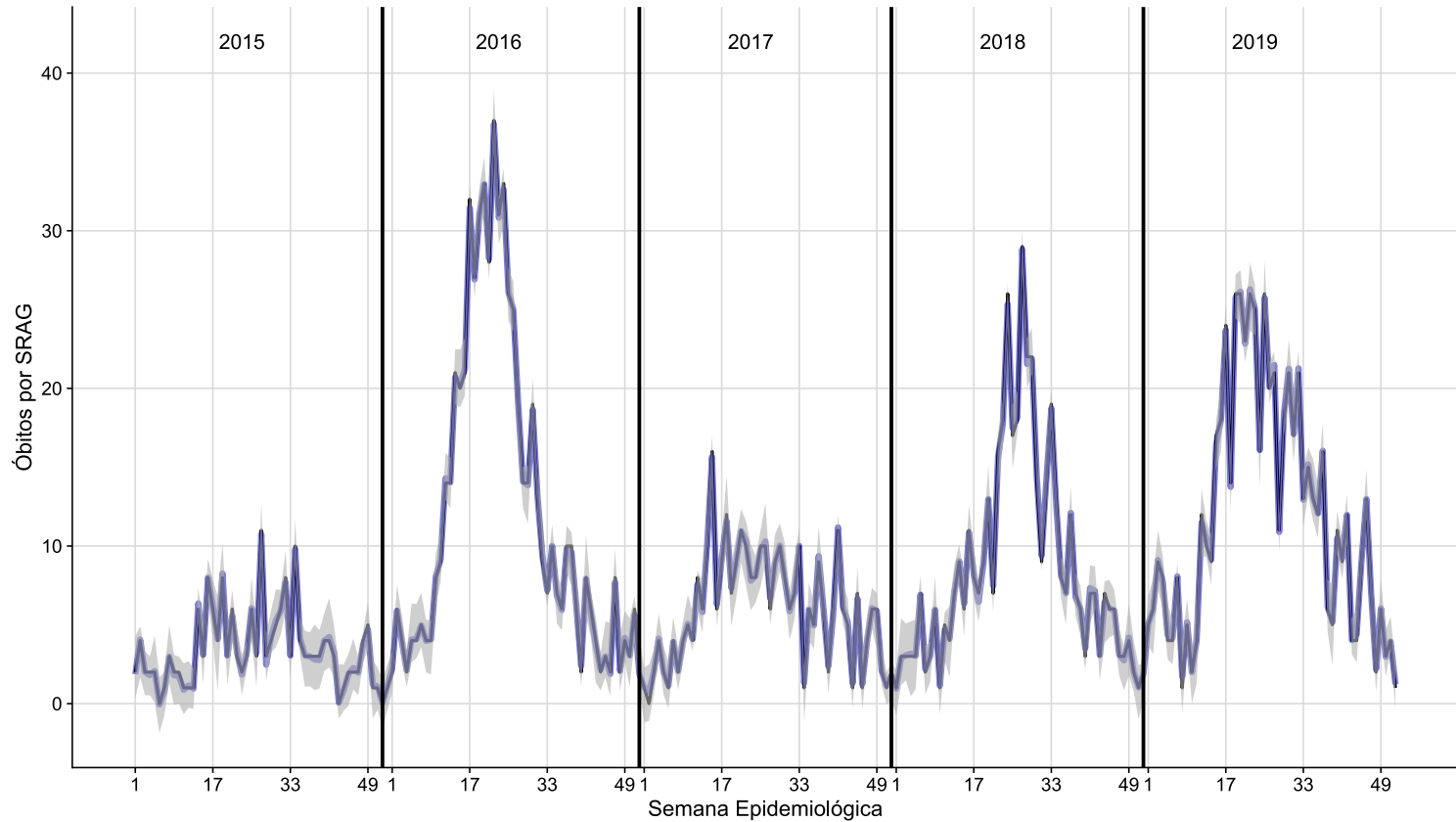
- $\mathbf{F}_t = (1, 1, 0, 1, 0, \dots, 1, 0)$.

- $\mathbf{G}_t = \text{block diag}(1, G_1, \dots, G_{51}, -1)$, em que

$$G_j = \begin{bmatrix} \cos \lambda_j & \sin \lambda_j \\ -\sin \lambda_j & \cos \lambda_j \end{bmatrix}, \quad \lambda_j = \frac{2\pi j}{s}, \quad j = 1, \dots, [s/2]$$

- Nesta análise $s = 52$, pois as observações são semanais.

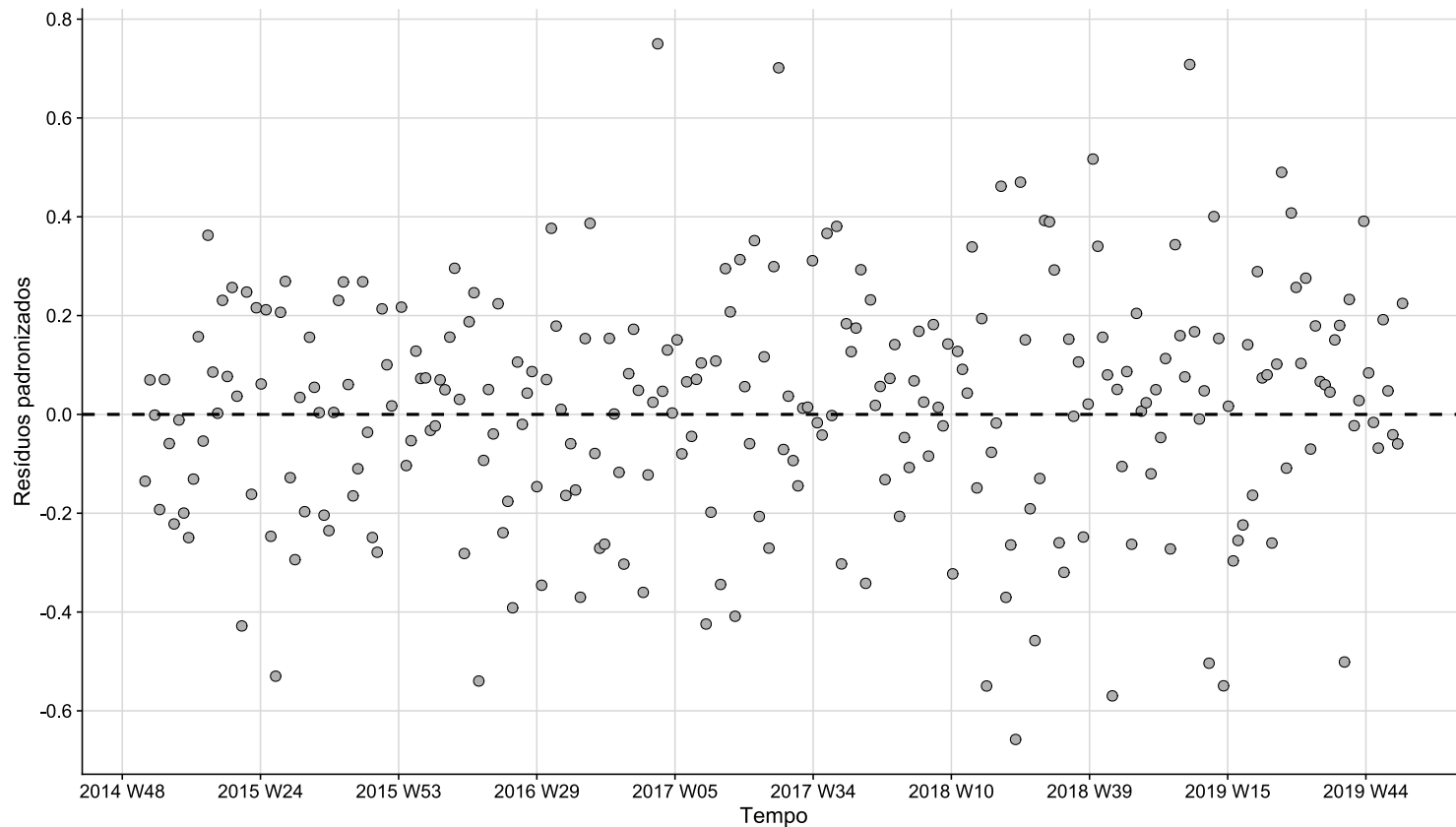
Ajuste



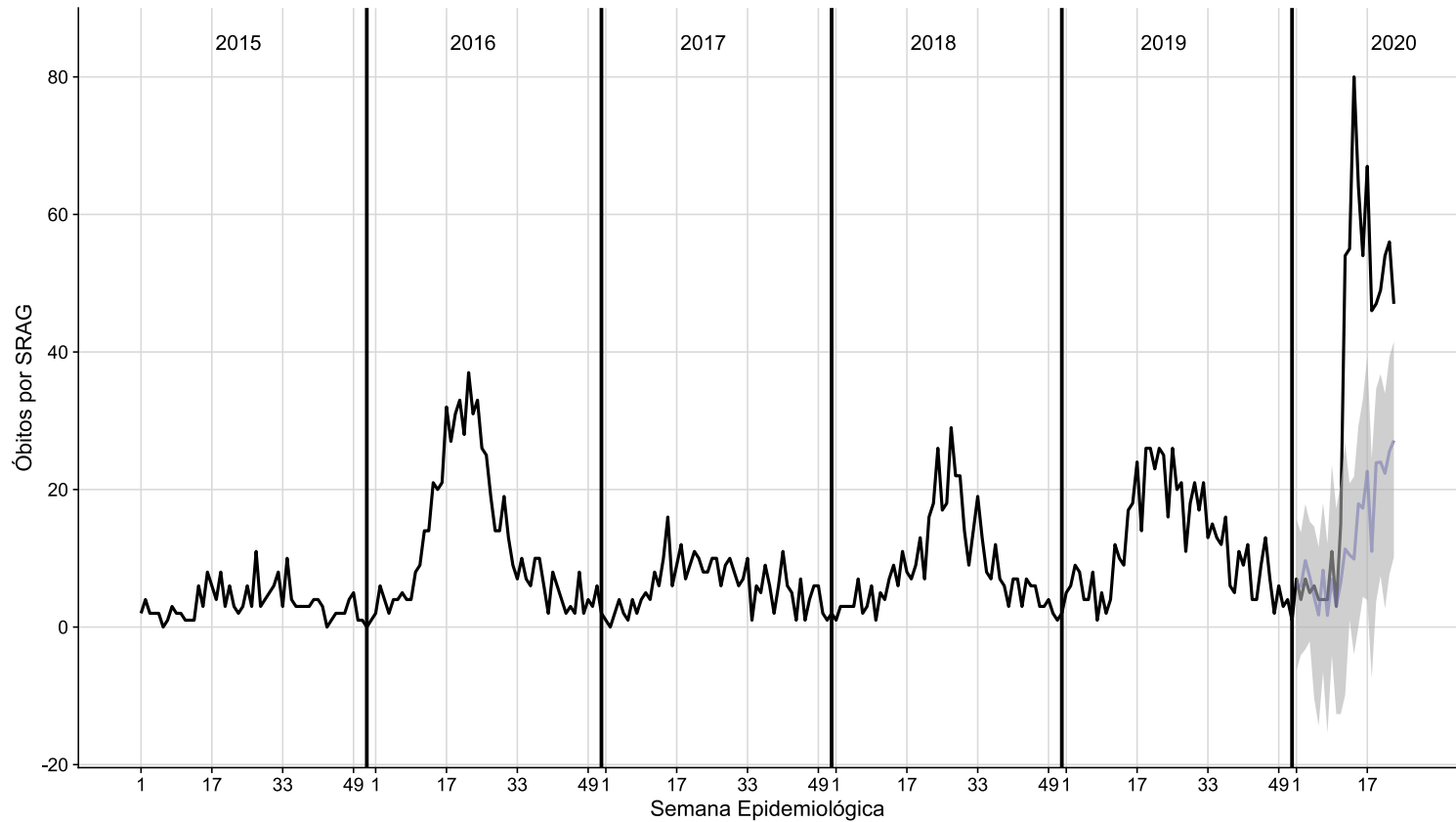
Ajuste - Medidas

Ano	RMSE	MDAE	MAE
2015	0.199	0.163	0.163
2016	0.203	0.137	0.162
2017	0.246	0.137	0.185
2018	0.264	0.152	0.205
2019	0.247	0.152	0.189
Todos	0.233	0.151	0.181

Ajuste - Resíduos



Previsões



Previsões

Semana	Óbitos	Predito	Efeito
10	3	3.2	-0.2
11	15	6.5	8.5
12	54	11.4	42.6
13	55	10.5	44.5
14	80	9.9	70.1
15	64	17.9	46.1
16	54	17.3	36.7
17	67	22.7	44.3
18	46	11.0	35.0
19	47	23.9	23.1
20	49	24.0	25.0
21	54	22.3	31.7
22	56	25.6	30.4

Referências

Brodersen, K. H, F. Gallusser, J. Koehler, et al. (2015). "Inferring causal impact using Bayesian structural time-series models". In: *Annals of Applied Statistics* 9.1, pp. 247-274.

Commandeur, J, S. J. Koopman, and M. Ooms (2011). "Statistical Software for State Space Methods". In: *Journal of Statistical Software, Articles* 41.1, pp. 1-18.

Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford University Press.

Durbin, J. and S. J. Koopman (2002). "A simple and efficient simulation smoother for state space time series analysis". In: *Biometrika* 89.3, pp. 603-616.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Referências

Jong, P. D. (1991). "The Diffuse Kalman Filter". In: *Annals of Statistics* 19.2, pp. 1073-1083.

Pole, A, M. West, and J. Harrison (1994). *Applied Bayesian Forecasting and Time Series Analysis*. Chapman & Hall.

Scott, S. L. and H. R. Varian (2014). "Predicting the present with Bayesian structural time series". In: *International Journal of Mathematical Modelling and Numerical Optimisation* 5.1/2, pp. 4-23.

West, M. and J. Harrison (1997). *Bayesian Forecasting and Dynamic Models*. Springer.

Obrigado!