

rafa:: R Automatic Forecasting Algorithm

J. Renato Leripio
Codeplan/DF

9 de maio de 2019

Motivação

- 1 Geração de *benchmarks* (variáveis **com** referência).

¹Dentro da classe de modelos univariados utilizados.

Motivação

- ❶ Geração de *benchmarks* (variáveis **com** referência).
- ❷ Melhor previsão possível¹ (variáveis **sem** referência).

¹Dentro da classe de modelos univariados utilizados.

Objetivo

- Gerar previsões robustas:

Objetivo

- Gerar previsões robustas:
- ➊ Identificar o modelo com melhor desempenho preditivo.

Objetivo

- Gerar previsões robustas:
- 1. Identificar o modelo com melhor desempenho preditivo.
- 2. Calcular corretamente o intervalo de confiança.

O algoritmo: 5 etapas

- 1 A partir de uma série temporal qualquer, ajustar um grande conjunto de modelos.

O algoritmo: 5 etapas

- 1 A partir de uma série temporal qualquer, ajustar um grande conjunto de modelos.
- 2 Calcular estatísticas de acurácia para cada modelo.

O algoritmo: 5 etapas

- 1 A partir de uma série temporal qualquer, ajustar um grande conjunto de modelos.
- 2 Calcular estatísticas de acurácia para cada modelo.
- 3 Melhor modelo é selecionado.

O algoritmo: 5 etapas

- 1 A partir de uma série temporal qualquer, ajustar um grande conjunto de modelos.
- 2 Calcular estatísticas de acurácia para cada modelo.
- 3 Melhor modelo é selecionado.
- 4 Melhor modelo sofre processo adicional de refinamento.

O algoritmo: 5 etapas

- 1 A partir de uma série temporal qualquer, ajustar um grande conjunto de modelos.
- 2 Calcular estatísticas de acurácia para cada modelo.
- 3 Melhor modelo é selecionado.
- 4 Melhor modelo sofre processo adicional de refinamento.
- 5 São geradas as previsões (valores centrais/intervalo de confiança).

Etapa 1: Modelos disponíveis (pacote *forecast*)

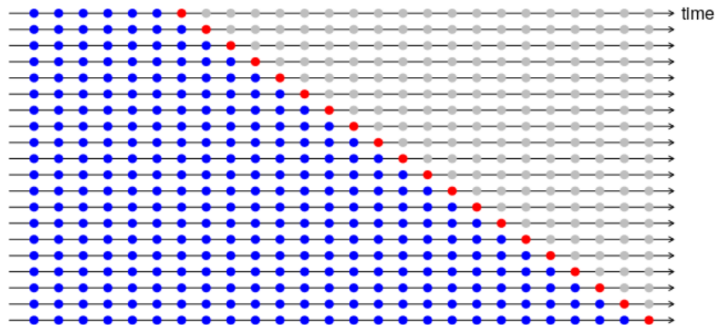
- 1 ARFIMA
- 2 ARIMA
- 3 ETS
- 4 HOLT
- 5 HOLT-WINTERS
- 6 MEANF
- 7 NNETAR
- 8 SES
- 9 SPLINEF
- 10 STRUCTS
- 11 TBATS
- 12 THETAF

Etapa 2: Computar os erros de previsão.

Forma convencional



Validação-cruzada



Fonte: Hyndman, R.J., & Athanasopoulos, G. (2018) Forecasting: principles and practice, 2nd edition, OTexts: Melbourne, Australia. [OTexts.com/fpp2](https://otexts.com/fpp2).

Etapa 3: Calcular acurácia dos modelos.

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (1)$$

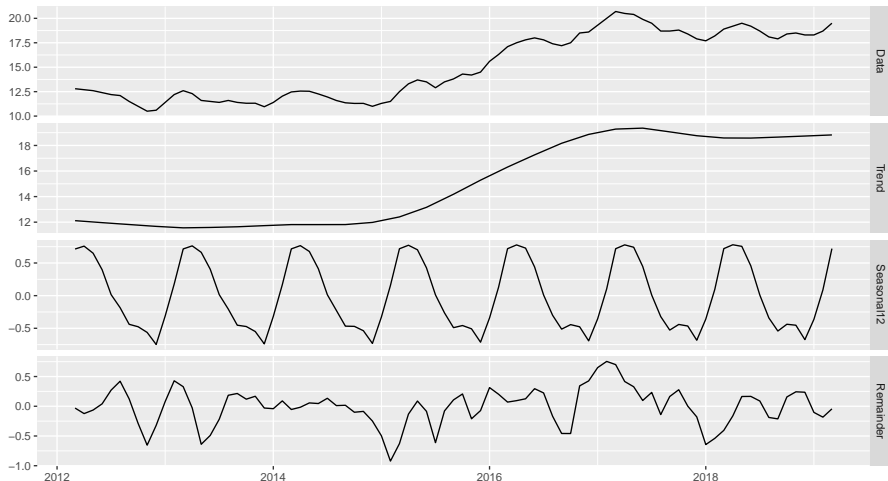
$$MAE = \frac{\left| \sum_{t=1}^T (y_t - \hat{y}_t) \right|}{n} \quad (2)$$

$$DIR = \begin{cases} 1 & \text{se } \text{sgn}(\hat{y}_t - y_{t-1}) = \text{sgn}(y_t - y_{t-1}) \\ -1 & \text{se } \text{sgn}(\hat{y}_t - y_{t-1}) \neq \text{sgn}(y_t - y_{t-1}) \end{cases} \quad (3)$$

Etapa 4: Refinamento das previsões (bagging).

Passo 1: Série é decomposta e parte “aleatória” sofre processo de reamostragem.

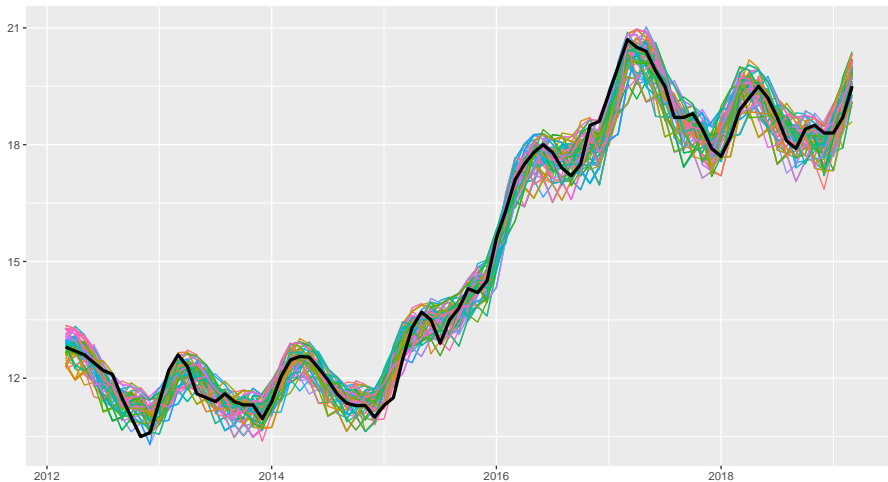
Decomposição STL da série.



Fonte: Elaboração própria com dados da PED/DF

Passo 2: Variações da parte aleatória são reintroduzidas à série.

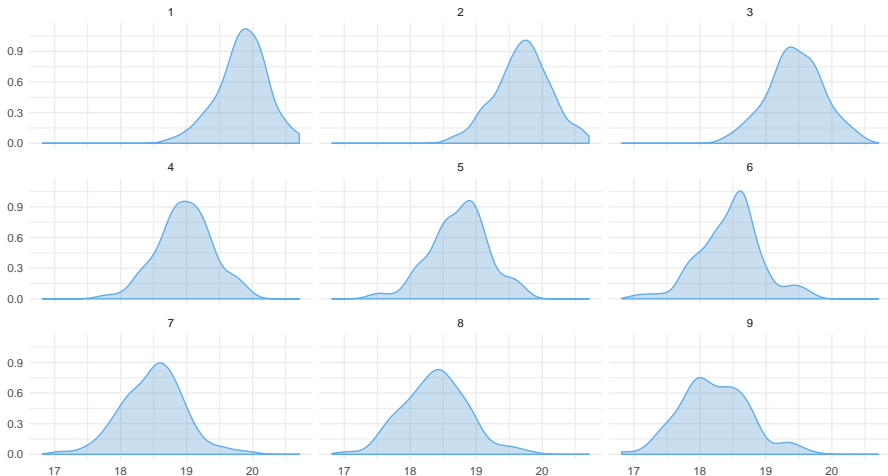
Séries calculadas por bootstrap



Fonte: Elaboração própria com dados da PED/DF

Projeções são computadas

Densidade das projeções para cada horizonte.



Fonte: Elaboração própria com dados gerados pelo modelo.

Sintaxe: `rafa::auto_forecast`

rafa :: **auto_forecast**(*data*, *h*, *h_cv*, *window*, *acc*, *n*, *level*, *exclude*, *test*)

data = série temporal univariada

h = horizonte para previsões (12)

h_cv = horizonte para validação-cruzada (1)

window = tamanho da janela para validação-cruzada (NULL)

acc = medida de acurácia ("MAE")

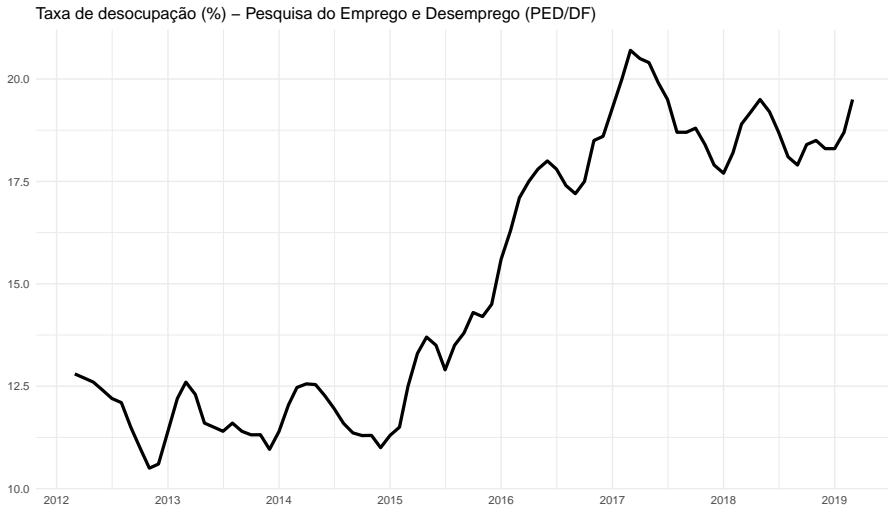
n = número de simulações por *bootstrap* (100)

level = nível de significância para os intervalos de confiança (0.05)

exclude = modelos a serem excluídos do processo (NULL)

test = data ou número da primeira observação da amostra de teste (NULL)

Exemplo: Taxa de desemprego.



Fonte: Elaboração própria com dados da PED/DF.

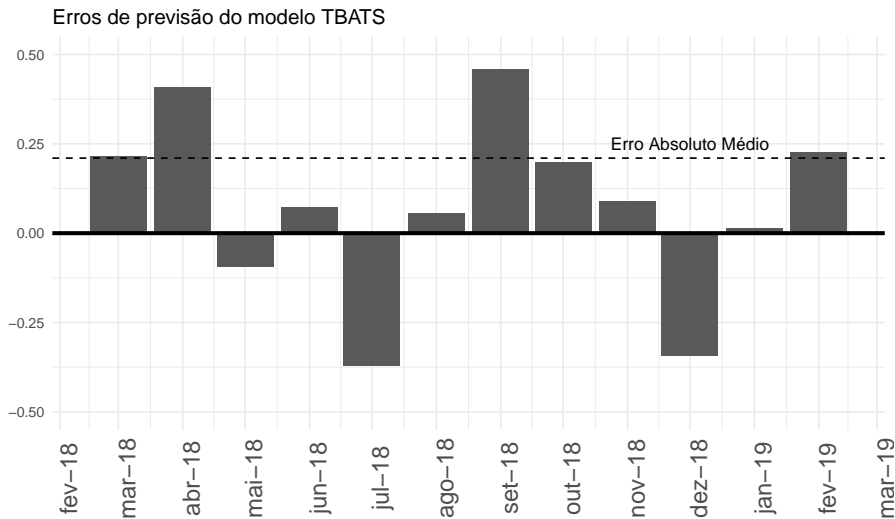
Performance dos modelos

Model	MAE	RMSE	Wrong dir	Right dir
tbats	0.21	0.26	3	9
ets	0.26	0.32	3	9
arfima	0.28	0.31	3	9
hw	0.31	0.37	3	9
holt	0.32	0.38	3	9
StrucTS	0.34	0.39	2	10
nnetar	0.35	0.42	5	7
ses	0.35	0.41	5	7
thetaf	0.35	0.41	4	8
splinef	0.37	0.42	2	10
meanf	3.73	3.79	7	5

Previsões para os próximos 12 meses

h	Point	Lower_0.95	Higher_0.95
1	19.81	18.84	20.54
2	19.69	18.78	20.44
3	19.48	18.41	20.31
4	19.01	17.99	19.85
5	18.76	17.68	19.56
6	18.45	17.47	19.29
7	18.48	17.40	19.37
8	18.39	17.27	19.31
9	18.23	17.19	19.34
10	18.65	17.37	19.76
11	19.24	17.89	20.34
12	19.84	18.47	20.91

Inspecionando os erros



Fonte: Elaboração própria com base nos resultados do modelo.

Próximos passos

- 1 Estratégias rectify (Taieb e Hyndman, 2012): “o melhor dos dois mundos”.

Próximos passos

- ❶ Estratégia rectify (Taieb e Hyndman, 2012): “o melhor dos dois mundos”.
- Previsões recursivas: viesada, menor variância.

Próximos passos

- ❶ Estratégia rectify (Taieb e Hyndman, 2012): “o melhor dos dois mundos”.
 - Previsões recursivas: viesada, menor variância.
 - Previsões diretas: não-viesada, maior variância.

Próximos passos

- ❶ Estratégia rectify (Taieb e Hyndman, 2012): “o melhor dos dois mundos”.
 - Previsões recursivas: viesada, menor variância.
 - Previsões diretas: não-viesada, maior variância.
 - Solução?

Próximos passos

- ❶ Estratégia rectify (Taieb e Hyndman, 2012): “o melhor dos dois mundos”.
 - Previsões recursivas: viesada, menor variância.
 - Previsões diretas: não-viesada, maior variância.
 - Solução?
 - Previsões recursivas da série + diretas do erro de previsão

Próximos passos

- ❶ Estratégia rectify (Taieb e Hyndman, 2012): “o melhor dos dois mundos”.
 - Previsões recursivas: viesada, menor variância.
 - Previsões diretas: não-viesada, maior variância.
 - Solução?
 - Previsões recursivas da série + diretas do erro de previsão
 - Resultados parecem promissores.

Próximos passos

- ❶ Estratégia rectify (Taieb e Hyndman, 2012): “o melhor dos dois mundos”.
 - Previsões recursivas: viesada, menor variância.
 - Previsões diretas: não-viesada, maior variância.
 - Solução?
 - Previsões recursivas da série + diretas do erro de previsão
 - Resultados parecem promissores.
 - Mas altamente sensíveis ao método empregado.

Próximos passos

- ❶ Estratégia rectify (Taieb e Hyndman, 2012): “o melhor dos dois mundos”.
 - Previsões recursivas: viesada, menor variância.
 - Previsões diretas: não-viesada, maior variância.
 - Solução?
 - Previsões recursivas da série + diretas do erro de previsão
 - Resultados parecem promissores.
 - Mas altamente sensíveis ao método empregado.
- ❷ Covariáveis

Próximos passos

- ❶ Estratégia rectify (Taieb e Hyndman, 2012): “o melhor dos dois mundos”.
 - Previsões recursivas: viesada, menor variância.
 - Previsões diretas: não-viesada, maior variância.
 - Solução?
 - Previsões recursivas da série + diretas do erro de previsão
 - Resultados parecem promissores.
 - Mas altamente sensíveis ao método empregado.
- ❷ Covariáveis
- ❸ Deep learning (LSTM)

Próximos passos

- ❶ Estratégia rectify (Taieb e Hyndman, 2012): “o melhor dos dois mundos”.
 - Previsões recursivas: viesada, menor variância.
 - Previsões diretas: não-viesada, maior variância.
 - Solução?
 - Previsões recursivas da série + diretas do erro de previsão
 - Resultados parecem promissores.
 - Mas altamente sensíveis ao método empregado.
- ❷ Covariáveis
- ❸ Deep learning (LSTM)
 - **Sugestões?**

Repositório: <https://github.com/leripio/rafa>

Versão de teste:

```
devtools::install_github("leripio/rafa")
```

```
library(rafa)
```

Agradecimento: Esta ferramenta vem sendo desenvolvida na Codeplan <www.codeplan.df.gov.br> como um esforço para obter previsões confiáveis e tempestivas de variáveis econômicas.

Disclaimer: Deve-se ter em mente que esta ferramenta serve apenas para guiar análises econômicas e não deve ser considerada como uma ferramenta oficial da Codeplan ou do Governo do Distrito Federal.

Contato: joao.gomes@codeplan.df.gov.br