

UNIVERSIDADE FEDERAL DE MINAS GERAIS
Instituto de Ciências Exatas
Programa de Pós-graduação em Estatística

Bernardo Medeiros de Freitas

**Estudo de caso da metodologia Dynamic Time Scan Forecasting para
previsão da série histórica de preços da ação preferencial da Petrobrás**

Belo Horizonte
2021

Bernardo Medeiros de Freitas

**Estudo de caso da metodologia Dynamic Time Scan Forecasting para
previsão da série histórica de preços da ação preferencial da Petrobrás**

Versão Inicial

Monografia de especialização apresentada ao Instituto de Ciências Exatas da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Especialista em Estatística.

Área de Ênfase: Estatística

Orientador: Marcelo Azevedo Costa

Belo Horizonte
2021

© 2021, Bernardo Medeiros de Freitas.
Todos os direitos reservados.

Freitas, Bernardo Medeiros de

F866e Estudo de caso da metodologia Dynamic Time Scan Forecasting para previsão da série histórica de preços da ação preferencial da Petrobrás [manuscrito] / Bernardo Medeiros de Freitas. Belo Horizonte - 2021.
47 f. : il. ; 29 cm corrigir

Orientador: Marcelo Azevedo Costa.

Monografia (especialização) - Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística.
Referências : f. 46-47

1. Estatística. 2. Análise de séries temporais.
3. Mercado de ações - Previsão. 4. Petrobrás - Estudo de caso. I. Costa, Marcelo Azevedo. II. Universidade Federal de Minas Gerais, Instituto de Ciências Exatas, Departamento de Estatística. III. Título.

CDU 519.2(043)

Ficha catalográfica elaborada pela bibliotecária Irénquer Vismeg
Lucas Cruz - CRB 6ª n° 819.



Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística
Programa de Pós-Graduação / Especialização
Av. Pres. Antônio Carlos, 6627 - Pampulha
31270-901 – Belo Horizonte – MG

E-mail: pgest@ufmg.br
Tel: 3409-5923 – FAX: 3409-5924

ATA DO 232º. TRABALHO DE FIM DE CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA DE BERNARDO MEDEIROS DE FREITAS.

Aos vinte e três dias do mês de agosto de 2021, às 14:00 horas, com utilização de recursos de videoconferência a distância, reuniram-se os professores abaixo relacionados, formando a Comissão Examinadora homologada pela Comissão do Curso de Especialização em Estatística, para julgar a apresentação do trabalho de fim de curso do aluno **Bernardo Medeiros de Freitas**, intitulado: “*Estudo de caso da metodologia Dynamic Time Scan Forecasting para previsão da série histórica de preços da ação preferencial da Petrobrás*”, como requisito para obtenção do Grau de Especialista em Estatística. Abrindo a sessão, o Presidente da Comissão, Professor Marcelo Azevedo Costa – Orientador, após dar conhecimento aos presentes do teor das normas regulamentares, passou a palavra ao candidato para apresentação de seu trabalho. Seguiu-se a arguição pelos examinadores com a respectiva defesa do candidato. Após a defesa, os membros da banca examinadora reuniram-se sem a presença do candidato e do público, para julgamento e expedição do resultado final. Foi atribuída a seguinte indicação: o candidato foi considerado Aprovado condicional às modificações sugeridas pela banca examinadora no prazo de 30 dias a partir da data de hoje por unanimidade. O resultado final foi comunicado publicamente ao candidato pelo Presidente da Comissão. Nada mais havendo a tratar, o Presidente encerrou a reunião e lavrou a presente Ata, que será assinada por todos os membros participantes da banca examinadora. Belo Horizonte, 23 de agosto de 2021.

Prof. Marcelo Azevedo Costa (Orientador)
Departamento de Engenharia da Computação / Escola de Engenharia / UFMG

Prof. Roberto da Costa Quinino
Departamento de Estatística / ICEX / UFMG

Prof. Anderson Laécio Galindo Trindade
Departamento de Engenharia da Computação / Escola de Engenharia / UFMG



Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística
Programa de Pós-Graduação / Especialização
Av. Pres. Antônio Carlos, 6627 - Pampulha
31270-901 – Belo Horizonte – MG

E-mail: pgest@ufmg.br

Tel: 3409-5923 – FAX: 3409-5924

DECLARAÇÃO DE CUMPRIMENTO DE REQUISITOS PARA CONCLUSÃO DO CURSO DE ESPECIALIZAÇÃO EM ESTATÍSTICA.

Declaro para os devidos fins que **Bernardo Medeiros de Freitas**, número de registro 2019705650, cumpriu todos os requisitos necessários para conclusão do curso de Especialização em Estatística, entregando a versão final do trabalho de conclusão de curso para seu orientador, o professor Marcelo Azevedo Costa, que aprovou a versão final. O trabalho foi apresentado no dia 23 de agosto de 2021 com o título “*Estudo de caso da metodologia Dynamic Time Scan Forecasting para previsão da série histórica de preços da ação preferencial da Petrobrás*”.

Belo Horizonte, 23 de novembro de 2021

Prof. Roberto da Costa Quinino
Coordenador da Comissão
do Curso de Especialização
em Estatística

Prof. Roberto da Costa Quinino
Coordenador do curso de
Especialização em Estatística
Departamento de Estatística / UFMG

À Thásia, por seu apoio incondicional, sempre.

AGRADECIMENTOS

Ao Professor Marcelo Azevedo Costa agradeço aceitar ser o meu orientador e sê-lo sempre com empenho e gentileza nas suas contribuições e ensinamentos. Ao Professor Anderson Trindade agradeço a indicação do curso de especialização em Estatística e por toda atenção que me foi dada mesmo no momento em que eu não estava vinculado as atividades da Universidade. Ao Professor Roberto Quinino agradeço pela coordenação realizada com primazia do curso de especialização em Estatística e por aceitar ser parte da banca examinadora desse trabalho.

A UFMG deixo meus agradecimentos pelo empenho em estar sempre entre as melhores universidades do país, mesmo nos cenários mais desafiadores, como têm sido os últimos anos.

Aos colegas de pós-graduação agradeço os momentos compartilhados, o apoio coletivo e os diálogos descontraídos. Em especial, não poderia deixar de agradecer ao Robert Bento, Vinícios Castro e Flora Guerra, que foram o maior apoio na superação dos desafios dessa trajetória, de vocês guardo para sempre a amizade.

Finalmente agradeço àquela que certamente foi a pessoa mais importante em toda minha trajetória acadêmica e profissional, que nos momentos mais difíceis sempre esteve presente, que me acolheu e cuidou sempre para que eu conseguisse alcançar o melhor de mim mesmo: minha mãe, Thásia de Medeiros.

RESUMO

Motivado pela capacidade do método Dynamic Time Scan Forecasting (DTSF) em identificar padrões através de funções de similaridade estatística para construir previsões em séries temporais, este trabalho se propôs a estudar sua eficiência em gerar previsões no mercado de ações, usando como estudo de caso a série histórica de preços das ações preferenciais da Petrobrás (PETR4).

Para as avaliações foram selecionados períodos de estabilidade das variações de preços, momentos de oscilações fortes entre altas e quedas de preços e momentos de eventos atípicos com forte impacto na volatilidade de preços do mercado e, em particular, das ações da Petrobrás.

A metodologia também foi comparada com metodologias já bem estabelecidas na análise de séries temporais, sendo estas a modelagem ARIMA, ETS e TBATS.

Palavras-chave: Previsão. Séries temporais. Mercado de ações. Petrobrás. ARIMA. ETS. TBATS. Dynamic Time Scan Forecasting (DTSF).

ABSTRACT

Motivated by the capacity of the Dynamic Time Scan Forecasting (DTSF) method in identifying patterns through statistical similarity functions in order to build forecasts in time series, this work proposes to study its efficiency in generating results in the stock market, using the series of prices of Petrobras preferred stock (PETR4) as a case study.

For the evaluations, periods of stability of price variations, moments of strong fluctuations between price increases and falls, and moments of atypical events with a strong impact on market price volatility and, in particular, on Petrobras shares, were selected.

The methodology was also compared with methodologies already quite common in the analysis of time series, these being the ARIMA, ETS and TBATS modeling.

Keywords: Forecasting. Time series. Stock market. Petrobrás. ARIMA. ETS. TBATS. Dynamic Time Scan Forecasting (DTSF).

LISTA DE FIGURAS

Figura 1 – Interface da aplicação Shiny desenvolvida	25
Figura 2 – Painel lateral	26
Figura 3 – Painel principal.....	28
Figura 4 – Janeiro (curva 18)	33
<i>Figura 5 – Fevereiro (curva 35)</i>	<i>33</i>
Figura 6 – Março (curvas 3, 4, 5 e 8)	33
<i>Figura 7 – Abril (curvas 1, 17 e 19)</i>	<i>33</i>
Figura 8 – Maio (curvas 7 e 14).....	34
<i>Figura 9 – 27/01/21 (curvas 1, 6, 23 e 45)</i>	<i>34</i>
<i>Figura 10 – 16/03/21 (curvas 32 e 37)</i>	<i>34</i>
<i>Figura 11 – 13/04/21 (curvas 12 e 39)</i>	<i>34</i>
<i>Figura 12 – 24/04/21 (curvas 26 e 36)</i>	<i>34</i>
Figura 13 – Janeiro (curvas 45 e 50).....	35
Figura 14 – Fevereiro (curvas 36 e 37)	35
Figura 15 – Março (curvas 9 e 15)	35
Figura 16 – Abril (curvas 11, 20 e 35)	35
Figura 17 – 26/02/20 (fora de controle).....	37
Figura 18 – 27/02/20 (curvas 6 e 32)	37
Figura 19 – 05/03/20 (curvas 17 e 42)	37
Figura 20 – 06/03/20 (curvas 27 e 44)	37
Figura 21 – 26/02/20 (fora de controle).....	38
Figura 22 – 27/02/20 (curvas 18, 27 e 49)	38
Figura 23 – 05/03/20 (curvas 23 e 25)	38
Figura 24 – 06/03/20 (fora de controle).....	38
Figura 25 – 19/02/21 (fora de controle).....	39
Figura 26 – 22/02/21 (curvas 15 e 30)	39
Figura 27 – 19/02/21 (fora de controle).....	40
Figura 28 – 22/02/21 (curva 17*).....	40
Figura 29 – DTSF (curva 18) – Janeiro	41
Figura 30 – ARIMA – Janeiro	41
Figura 31 – ETS – Janeiro	41
Figura 32 – TBATS – Janeiro	41
Figura 33 – DTSF (curva 35) – Fevereiro.....	41
Figura 34 – ARIMA – Fevereiro.....	41
Figura 35 – ETS – Fevereiro	41
Figura 36 – TBATS – Fevereiro.....	41
Figura 37 – DTSF (curva 45 e 50) – Janeiro	42
Figura 38 – ARIMA – Janeiro	42
Figura 39 – ETS – Janeiro	42
Figura 40 – TBATS – Janeiro	42
Figura 41 – DTSF (curvas 36 e 37) – Fevereiro	42
Figura 42 – ARIMA – Fevereiro.....	42

Figura 43 – ETS – Fevereiro	42
Figura 44 – TBATS – Fevereiro.....	42

LISTA DE TABELAS E QUADROS

Tabela 1 – Taxonomia dos modelos de alisamento exponencial	20
Tabela 2 – Relação da taxonomia com modelos conhecidos.....	21
Tabela 3 – Parâmetros de cada modelo de alisamento exponencial	22
Tabela 4 – Intervalo de confiança	32

SUMÁRIO

1	INTRODUÇÃO	13
2	METODOLOGIA	15
2.1	O Método <i>Dynamic Time Scan Forecasting</i>	15
2.2	Métodos Comparativos	17
2.2.1	A modelagem ARIMA	18
2.2.2	A modelagem de alisamento exponencial (ETS)	20
2.2.3	A modelagem TBATS	22
2.3	Painel de avaliação do DTSF	24
2.4	Objeto de estudo	30
3	RESULTADOS	31
3.1	Definição dos intervalos de confiança	31
3.2	Avaliação das projeções em momentos de estabilidade do mercado	32
3.2.1	Testes com a série original	32
3.2.2	Testes em dados com reversão de tendência	34
3.2.3	Testes com a série de retornos	35
3.3	Avaliação das projeções em momentos de instabilidade do mercado	36
3.3.1	Início dos casos de Covid-19 no Brasil e no mundo	36
3.3.2	Troca do presidente da Petrobrás	39
3.4	Comparação do método DTSF com o ARIMA, ETS e TBATS	40
4	CONCLUSÃO	44
5	REFERÊNCIAS	46

1 INTRODUÇÃO

As tentativas de prever valores futuros a fim de subsidiar decisões de compra e venda de ativos não são recentes. Técnicas como a análise de padrões de *candlesticks* podem ser rastreadas desde o século XVIII no mercado de arroz japonês [2]. Neste ínterim, diversas metodologias vêm sendo testadas e aplicadas a fim de produzir os melhores resultados possíveis em termos de lucratividade dos investimentos.

As duas principais escolas que se desenvolveram ao longo deste tempo são conhecidas como as de análise fundamentalista e as de análise técnica (ou ainda, gráfica). Segundo John J. Murphy [3], enquanto a análise técnica se concentra no estudo dos movimentos de mercado, a análise fundamentalista foca nas forças econômicas de oferta e demanda e nas análises de todos os fatos relevantes que afetem os preços de mercado a fim de determinar o valor intrínseco daquele ativo. Essas metodologias coexistem há algumas décadas, no entanto foram desafiadas na década de 60 pela teoria do passeio aleatório, que ficou popularmente conhecida como hipótese dos mercados eficientes, que propõe que variações futuras nos preços de ações não podem ser previstas a partir de variações passadas. Apesar de alguns estudos terem mostrado a presença do 'passeio aleatório' nas movimentações de preços de ações, a maioria deles apontou para alguma componente passível de previsão [4].

Motivado por essa última constatação e pelo fato de consistentemente os analistas técnicos do mercado de ações usarem padrões de variações de preço para previsão de valores futuros, decidiu-se colocar a prova a metodologia do *Dynamic Time Scan Forecasting* (DTSF) como ferramenta preditiva de preços de ações. O DTSF consiste em uma metodologia que define uma janela de tamanho específico e fixo (w), na qual a janela de base comparativa é composta pelos últimos w valores, e estes são então comparados com todo o período anterior a janela base, em pelo menos o número de previsões que se deseja fazer. A comparação é feita por uma varredura de todas as janelas possíveis de w observações no período comparativo mencionado. Seleciona-se então os períodos mais similares à janela base e projeta-se os períodos futuros com base nos períodos subsequentes às janelas de maior semelhança estatística fazendo-lhes os ajustes a partir da função de similaridade.

Esse modelo é consistente com a lógica de analistas técnicos que defendem que prever movimentações futuras do mercado de ações consiste em reconhecer em extensas bases históricas de preços, padrões repetitivos de variações de preços que derivam dos comportamentos humanos diante de ditas variações [17].

2 METODOLOGIA

2.1 O MÉTODO *DYNAMIC TIME SCAN FORECASTING*

Neste trabalho será avaliada a eficiência do método Dynamic Time Scan Forecasting (DTSF) [1] em produzir previsões para o preço de ativos listados na Bolsa de Valores de São Paulo (Bovespa). Este é um método derivado da estatística de varredura (também conhecida como estatística de janela de varredura), que consiste em uma classe de estatísticas que fazem a varredura de uma sequência de dados a fim de encontrar comportamentos anômalos.

O método DTSF consiste então, basicamente, na varredura de uma série temporal usando uma janela de tamanho fixo. Sendo y_t uma série temporal de tamanho N , $t = 1, \dots, N$. Define-se com o vetor $y^{[w]}$ como as últimas w observações da série temporal:

$$y^{[w]} = [y_{N-w+1}, \dots, y_N]$$

O objetivo do DTSF é, então, identificar subconjuntos da série temporal que tenham forte correlação com o vetor $y^{[w]}$. Isso é feito através de uma varredura de subconjuntos de valores anteriores da série com o mesmo tamanho do vetor $y^{[w]}$. O conjunto de subconjuntos candidatos a melhor ajuste pode ser definido como:

$$x_t^{[w]} = [y_t, \dots, y_{t+w-1}]$$

Onde $t = 1, \dots, N - 2w + 1$. O limite superior deste intervalo garante que não haverá sobreposição do vetor $x_t^{[w]}$ com o vetor $y^{[w]}$.

O procedimento de varredura aplica uma função de similaridade $f(x_t^{[w]}; y^{[w]})$ para cada janela avaliada a fim de mapear observações anteriores (vista como variável preditora) em relação às últimas observações (vista como variável resposta). Um teste estatístico é calculado para medir a qualidade do ajuste da função de similaridade calculada para cada janela. Após varrer todos os subconjuntos da série temporal, as m melhores janelas, $[x_{t(1)}^{[w]}, \dots, x_{t(m)}^{[w]}]$, são selecionadas baseadas em sua

similaridade estatística em relação ao vetor $y^{[w]}$, onde o índice $t_{(i)}$ representa o momento de início da i -ésima melhor janela.

A proposta inicial para a função de similaridade é construir um modelo linear dos elementos do vetor $x_t^{[w]}$ em relação ao vetor $y^{[w]}$, considerando que se espera que os valores anteriores possam, em teoria, ser similares aos valores recentes, à exceção de sua escala e de seu deslocamento no eixo y .

Assumindo a função de similaridade como um modelo linear, pode-se estimar os parâmetros do modelo através da minimização da soma dos quadrados entre os elementos do vetor $y^{[w]}$ e a equação linear: $\beta_0^{[t]} + \beta_1^{[t]} \times x_t^{[w]}$. Além disso, a estatística de similaridade pode ser definida como o coeficiente de determinação da regressão linear (R^2):

$$R^2 = 1 - \frac{\sum_j (y_j^{[w]} - \hat{y}_j^{[w]})^2}{\sum_j (y_j^{[w]} - \bar{y}_j^{[w]})^2}$$

onde $y_j^{[w]}$ é o j -ésimo valor do vetor $y^{[w]}$ e $\hat{y}_j^{[w]}$ é o j -ésimo valor predito pela função linear. A estatística R^2 estará sempre no intervalo $[0,1]$, logo, se R^2 se aproxima de um, significa que vetor $x_t^{[w]}$ é bastante similar em relação ao vetor $y^{[w]}$ após feitos os ajustes de escala e deslocamento. Vale ressaltar que para avaliar a similaridade – ou seja, a qualidade do ajuste – outras estatísticas poderiam ser empregadas como o coeficiente de Pearson ou a função de verossimilhança.

O objetivo do método DTSF é construir uma previsão de k passos à frente da série temporal y_t , ou seja, $\hat{y}_N^{[k]} = [\hat{y}_{N+1}, \dots, \hat{y}_{N+k}]$. Para construir essas previsões, para cada uma das m melhores janelas $x_{t_{(i)}}^{[w]}$, usa-se sua respectiva função de similaridade aplicada aos respectivos k valores subsequentes de dita janela produzindo assim o vetor $\hat{y}_N^{[k]}$. Colocando em termos matemáticos, dada a i -ésima melhor janela, os valores preditos para cada tempo $N + l$, ($l = 1, \dots, k$), serão calculados pela Equação 1:

$$\hat{y}_{N+l,i} = g \left(y_{t_{(i)}+w-1+l}, x_{t_{(i)}}^{[w]}, \theta_i \right)$$

onde $i = 1, \dots, m$; $l = 1, \dots, k$; $g(\cdot)$ é a função de similaridade estimada usando a i -ésima melhor janela; $t_{(i)}$ é o índice temporal que representa o início da i -ésima melhor janela; $y_{t_{(i)}+w-1+l}$ é o l -ésimo valor observado subsequente à janela $x_{t_{(i)}}^{[w]}$ e θ_i é o vetor de coeficientes estimados da função de similaridade para dita janela.

Finalmente, pode-se combinar as previsões geradas pelas m melhores janelas para o período $N + l$ utilizando uma função de agregação. A média, caso escolhida para agregar as m previsões do período $N + l$, ($l = 1, \dots, k$), seria calculada pela Equação 2:

$$\hat{y}_{N+l} = \frac{1}{m} \sum_{i=1}^m \hat{y}_{N+l,i}$$

Por padrão, o método DTSF desenvolvido em R já gera um vetor da combinação das m previsões utilizando como função de agregação a mediana, a fim de minimizar os efeitos de previsões anômalas.

É importante notar que o modelo possui uma restrição intrínseca a respeito do tamanho de k (número de passos previstos adiante na série temporal). Este não deve ultrapassar o tamanho da janela de referência, pois isso garante que caso uma das m melhores janelas seja composta pelos valores imediatamente anteriores à janela de referência (vetor $y^{[w]}$), então os valores previstos serão uma função de tal vetor.

Nota-se que o modelo DTSF requer a especificação de três parâmetros: a extensão temporal da janela de varredura, w , a função de similaridade, $g = (x_t^{[w]}, y^{[w]})$, e finalmente o número de janelas melhor ajustadas encontradas na série temporal, m . A fim de otimizar a complexidade computacional do método, são oferecidas como alternativas de funções de similaridade: a linear, a quadrática, a cúbica e a de quarta potência. Para este trabalho apenas foi utilizada a função de similaridade linear.

2.2 MÉTODOS COMPARATIVOS

Para comparar a eficiência do DTSF foram construídas também previsões fazendo uso três outros métodos: o ARIMA, o Alisamento Exponencial e o TBATS. A seguir veremos o detalhamento de cada um deles.

2.2.1 A modelagem ARIMA

Modelos ARIMA são, em teoria, a classe de modelos mais geral para prever uma série temporal que possa ser levada à estacionaridade através de diferenciações, talvez em conjunto com transformações não lineares como, por exemplo, a logarítmica, caso necessário. A estacionaridade é um pré-requisito para a aplicação de processos autorregressivos e para processos de médias móveis. Ela pode ser entendida como uma série de dados que não apresenta tendência, cujas variações em torno de sua média têm amplitude constante e que se move de forma consistente, ou seja, seus padrões de curto prazo sempre se apresentam de formas estatisticamente equivalentes ao longo da série [5].

O acrônimo ARIMA significa autorregressivo integrado e de média móvel (Autoregressive Integrated Moving Average). As defasagens da série estacionarizada na equação de previsão são chamados de termos “autorregressivos”, já as defasagens dos erros de previsão são chamadas de termos de “médias móveis” e, finalmente, uma série temporal que requeira diferenciação para ser convertida em estacionária é dita ser uma versão integrada de uma série estacionária [5].

Os termos autoregressivos podem ser entendidos como um modelo de regressão no qual a variável de interesse é uma combinação linear de valores passados da variável (em lugar de uma variável preditora, um processo autoregressivo define termos futuros a partir de termos passados da série temporal) [9].

Assim, um modelo autoregressivo de ordem p pode ser escrito como:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t,$$

onde ε_t é um ruído branco. Este processo é similar a uma regressão múltipla, mas com valores defasados de y_t como preditores. Esse modelo é referido como AR(p), um modelo autoregressivo de ordem p [9].

Os termos de média móvel, por sua vez, podem ser entendidos, em lugar de uma regressão que usa valores passados da própria variável para prevê-la, como algo similar a uma regressão, mas que utiliza os erros de previsão passados como preditores.

Assim, um modelo de médias móveis de ordem q pode ser escrito como:

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q},$$

onde ε_t é um ruído branco. Esse modelo é referido como MA(q), um modelo de média móvel de ordem q . Naturalmente os valores de ε_t não são observados, então essa equação não é a de uma regressão no seu sentido mais comum [9].

Se combinarmos a diferenciação com os modelos autorregressivos e o de médias móveis, obtém-se um modelo ARIMA não sazonal. Este modelo pode ser escrito como:

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t,$$

onde y'_t é a série diferenciada d vezes e ε_t é um ruído branco [9]. Esse modelo é referido como ARIMA(p,d,q), onde:

- **p** é o número de termos autorregressivos,
- **d** é o número de diferenciações e
- **q** é o número de erros projetados defasados na equação de predição.

Uma componente sazonal pode ser incluída no modelo com essa mesma lógica, porém aplicando as defasagens como múltiplas do período sazonal em lugar de um período apenas. Esse modelo é referido como SARIMA(p,d,q)(P,D,Q) $_m$, onde os novos termos podem ser definidos como:

- **P** é o número de termos autorregressivos sazonais,
- **D** é o número de diferenciações sazonais,
- **Q** é o número de erros projetados defasados sazonais na equação de predição e
- **m** é o número de observações que conformam o período sazonal

A estipulação dos parâmetros (p , d , q , P , D , Q e m) é feita através de uma metodologia desenvolvida por Box e Jenkins [6], através da análise dos gráficos de Função de Autocorrelação e de Função de Autocorrelação Parcial, no entanto para este trabalho foi utilizada a função de definição automatizada do R para esses parâmetros (função `auto.arima()`).

2.2.2 A modelagem de alisamento exponencial (ETS)

Segundo Hyndman e Athanasopoulos [9], o alisamento exponencial, proposto no final da década de 50 por Brown, 1959; Holt, 1957 e Winters, 1960, promoveu alguns dos mais bem-sucedidos métodos preditivos. Previsões utilizando esses métodos são médias ponderadas de observações anteriores com pesos que têm decaimento exponencial à medida que as observações se tornam mais antigas.

Os modelos ETS são uma família de modelos em que cada modelo consiste em uma equação de medida que descreve os dados observados e algumas equações de estado que descrevem como as componentes – ou estados – não observadas (nível, error, tendência e sazonalidade) variam ao longo do tempo [8, 9]. Assim, estes são referidos como modelos de espaço de estado.

Hyndman et al [10] propõem uma modificação da taxonomia proposta por Pegels [18] e ampliado por Gardner [19] para a classificação dos métodos de alisamento exponencial. Na tabela 1 podemos ver essa classificação:

Componente de Tendência	Componente Sazonal		
	N (nenhuma)	A (aditiva)	M (multiplicativa)
N (nenhuma)	NN	NA	NM
A (aditiva)	AN	AA	AM
M (multiplicativa)	MN	MA	MM
D (amortecida – “damped”)	DN	DA	DM

Tabela 1 – Taxonomia dos modelos de alisamento exponencial

Assim, podemos identificar nessa taxonomia vários dos métodos de alisamento exponencial mais conhecidos, como os seguintes exemplos:

Forma abreviada	Método
(N,N)	Alisamento Exponencial Simples
(A,N)	Método de Holt
(D,N)	Método de tendência amortecida aditiva
(A,A)	Método de Holt-Winters aditivo
(A,M)	Método de Holt-Winters multiplicativo
(D,M)	Método de Holt-Winters' amortecido

Tabela 2 – Relação da taxonomia com modelos conhecidos

Voltando a tabela 1, podemos escrever cada um dos 12 métodos de alisamento exponencial com o seguinte conjunto de equações:

$$\ell_t = \alpha P_t + (1 - \alpha)Q_t$$

$$b_t = \beta R_t + (\phi - \beta)b_{t-1}$$

$$s_t = \gamma T_t + (1 - \gamma)s_{t-m}$$

Onde ℓ_t corresponde ao nível da série no momento t , b_t corresponde à inclinação no momento t , s_t corresponde à componente sazonal da série no momento t e, finalmente, m corresponde ao número do período previsto. Os valores de P_t , Q_t , R_t e T_t variam de acordo com o método selecionado e α , β , γ e ϕ são constantes.

Componente Tendência	Componente Sazonal		
	N (nenhuma)	A (aditiva)	M (multiplicativa)
N (nenhuma)	$P_t = Y_t$	$P_t = Y_t - s_{t-m}$	$P_t = Y_t/s_{t-m}$
	$Q_t = \ell_{t-1}$	$Q_t = \ell_{t-1}$	$Q_t = \ell_{t-1}$
	$\phi = 1$	$\phi = 1$	$\phi = 1$
	$F_{t+h} = \ell_t$	$F_{t+h} = \ell_t + s_{t+h-m}$	$F_{t+h} = \ell_t s_{t+h-m}$
A	$P_t = Y_t$	$P_t = Y_t - s_{t-m}$	$P_t = Y_t/s_{t-m}$
	$Q_t = \ell_{t-1} + b_{t-1}$	$Q_t = \ell_{t-1} + b_{t-1}$	$Q_t = \ell_{t-1} + b_{t-1}$

(aditiva)	$R_t = \ell_t - \ell_{t-1}$ $\phi = 1$ $F_{t+h} = \ell_t + hb_t$	$R_t = \ell_t - \ell_{t-1}$ $T_t = Y_t - Q_t$ $\phi = 1$ $F_{t+h} = \ell_t + hb_t + s_{t+h-m}$	$R_t = \ell_t - \ell_{t-1}$ $T_t = Y_t/Q_t$ $\phi = 1$ $F_{t+h} = (\ell_t + hb_t)s_{t+h-m}$
M (multiplicativa)	$P_t = Y_t$ $Q_t = \ell_{t-1}b_{t-1}$ $R_t = \ell_t/\ell_{t-1}$ $\phi = 1$ $F_{t+h} = \ell_t b_t^h$	$P_t = Y_t - s_{t-m}$ $Q_t = \ell_{t-1}b_{t-1}$ $R_t = \ell_t/\ell_{t-1}$ $T_t = Y_t - Q_t$ $\phi = 1$ $F_{t+h} = \ell_t b_t^h + s_{t+h-m}$	$P_t = Y_t/s_{t-m}$ $Q_t = \ell_{t-1}b_{t-1}$ $R_t = \ell_t/\ell_{t-1}$ $T_t = Y_t/Q_t$ $\phi = 1$ $F_{t+h} = \ell_t b_t^h s_{t+h-m}$
D (amortecida – “damped”)	$P_t = Y_t$ $Q_t = \ell_{t-1} + b_{t-1}$ $R_t = \ell_t - \ell_{t-1}$ $\beta < \phi < 1$ $F_{t+h} = \ell_t + (1 + \phi + \dots + \phi^{h-1})b_t$	$P_t = Y_t - s_{t-m}$ $Q_t = \ell_{t-1} + b_{t-1}$ $R_t = \ell_t - \ell_{t-1}$ $T_t = Y_t - Q_t$ $\beta < \phi < 1$ $F_{t+h} = \ell_t + (1 + \phi + \dots + \phi^{h-1})b_t + s_{t+h-m}$	$P_t = Y_t/s_{t-m}$ $Q_t = \ell_{t-1} + b_{t-1}$ $R_t = \ell_t - \ell_{t-1}$ $T_t = Y_t/Q_t$ $\beta < \phi < 1$ $F_{t+h} = [\ell_t + (1 + \phi + \dots + \phi^{h-1})b_t]s_{t+h-m}$

Tabela 3 – Parâmetros de cada modelo de alisamento exponencial

Para cada um desses métodos descritos na tabela 1, ainda existem duas possibilidades de modelagem dos erros: aditiva ou multiplicativa [9].

Para o desenvolvimento deste trabalho, foi utilizado o método automatizado de definição destes parâmetros através da função `ets()` do R.

2.2.3 A modelagem TBATS

O modelo TBATS é um modelo totalmente automatizado que usa a combinação de termos de Fourier com alisamento exponencial e uma transformação de Box-Cox [9]. Ele foi desenvolvido para tentar contornar o fato de muitas séries temporais exibirem padrões de sazonalidades complexos, como por exemplo um padrão de sazonalidade não inteiro ou múltiplos padrões de sazonalidade simultaneamente [11].

TBATS é o acrônimo para os principais recursos do modelo, sendo estes:

- T: Sazonalidade trigonométrica
- B: Transformação de Box-Cox
- A: erros do ARIMA
- T: Tendência
- S: Componentes sazonais

Para contornar problemas de não-linearidade associados a modelos de estado de espaço para alisamento exponencial, a modelagem parte de realizar uma transformação de Box-Cox [12] na série temporal a fim de permitir certos tipos de não linearidade.

Estendendo o modelo de Holt Winters para incluir uma transformação de Box-Cox, erros do ARMA e T padrões sazonais, chega-se ao seguinte:

$$y_t^{(w)} = \begin{cases} \frac{y_t^w - 1}{w}; & w \neq 0 \\ \log y_t; & w = 0 \end{cases}$$

$$y_t^{(w)} = \ell_{t-1} + \phi b_{t-1} + \sum_{i=1}^T s_{t-m_i}^{(i)} + d_t$$

$$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha d_t$$

$$b_t = (1 - \phi)b + \phi b_{t-1} + \beta d_t$$

$$s_t^{(i)} = s_{t-m_i}^{(i)} + \gamma_i d_t$$

$$d_t = \sum_{i=1}^p \varphi_i d_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t$$

Onde m_1, \dots, m_T representam os períodos sazonais, ℓ_t é o nível local no período t , b é o tendência de longo prazo, b_t é a tendência de curto prazo no período t , $s_t^{(i)}$ representa a i -ésima componente sazonal no tempo t , d_t representa um processo ARMA(p,q) e ε_t é um ruído branco Gaussiano de média zero e variância constante σ^2 . Os parâmetros de alisamento são dados por α, β e γ_i para $i = 1, \dots, T$. O parâmetro de tendência amortecida (ϕ) é adotado, mas, conforme sugestão de Snyder [20], ele é suplementado com uma tendência de longo prazo (b). Essa alteração garante que as previsões de valores futuros da tendência de curto prazo (b_t) converjam para a tendência de longo prazo (b) ao invés de zero [11].

Essas modificações conformam o que pode ser identificado como BATS, que seria o acrônimo dos recursos chave dessa modelagem, conforme explicitado anteriormente (mas ainda sem a presença da sazonalidade trigonométrica). Os argumentos que suplementam essa modelagem são: $(\omega, \phi, p, q, m_1, m_2, \dots, m_T)$; estes representam, respectivamente, o parâmetro da transformação de Box-Cox, o parâmetro de amortecimento, os parâmetros ARMA (p e q) e os períodos sazonais (m_1, m_2, \dots, m_T) . Por exemplo, o modelo $BATS(1,1,0,0, m_1)$ representa o conhecido método de Holt-Winters aditivo de única sazonalidade [11].

Segundo De Livera et al [11], o modelo BATS é a generalização mais óbvia do tradicional modelo de inovação sazonal que permite múltiplos períodos sazonais. No entanto, ele não pode acomodar período de sazonalidade não inteiro e ele pode ter um número muito grande de estados. Na tentativa de se conseguir uma abordagem mais flexível e parcimoniosa, introduz-se a seguinte representação trigonométrica das componentes sazonais baseadas na série de Fourier.

$$s_t^{(i)} = \sum_{j=1}^{k_i} s_{j,t}^{(i)}$$

$$s_{j,t}^{(i)} = s_{j,t-1}^{(i)} \cos \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \sin \lambda_j^{(i)} + \gamma_1^{(i)} d_t$$

$$s_{j,t}^{*(i)} = -s_{j,t-1}^{(i)} \sin \lambda_j^{(i)} + s_{j,t-1}^{*(i)} \cos \lambda_j^{(i)} + \gamma_2^{(i)} d_t$$

Onde $\gamma_1^{(i)}$ e $\gamma_2^{(i)}$ são parâmetro de alisamento e $\lambda_j^{(i)} = 2\pi j/m_i$. O nível estocástico da i -ésima componente é descrita como $s_{j,t}^{(i)}$ e o crescimento no nível da i -ésima componente sazonal que é necessária para descrever a mudança na componente sazonal ao longo do tempo é descrita como $s_{j,t}^{*(i)}$. O número de harmônicos necessários para a i -ésima componente sazonal é representado por k_i .

2.3 PAINEL DE AVALIAÇÃO DO DTSF

Para construir as avaliações de eficiência da metodologia do método, foi desenvolvida uma aplicação em Shiny – pacote da linguagem R – no qual é possível desenvolver uma interface gráfica interativa.

Dynamic Time Scan Forecasting

Digite o código do papel a ser avaliado:

Intervalo Analisado:
 to

Trabalhar com retornos

Número de melhores curvas a projetar:
 curvas

Tamanho da janela de avaliação:
 dias

Ordem do polinômio da regressão:

Tamanho do período projetado:

Número de dias a mostrar no gráfico (antes da projeção):

Curvas projetadas a serem mostradas:

Mostrar combinação das curvas selecionadas
 Ocultar curvas selecionadas
 Mostrar combinação de todas as melhores curvas
 Mostrar previsão do modelo ARIMA
 Mostrar previsão do modelo de Ajustamento Exponencial
 Mostrar previsão do modelo TBATS

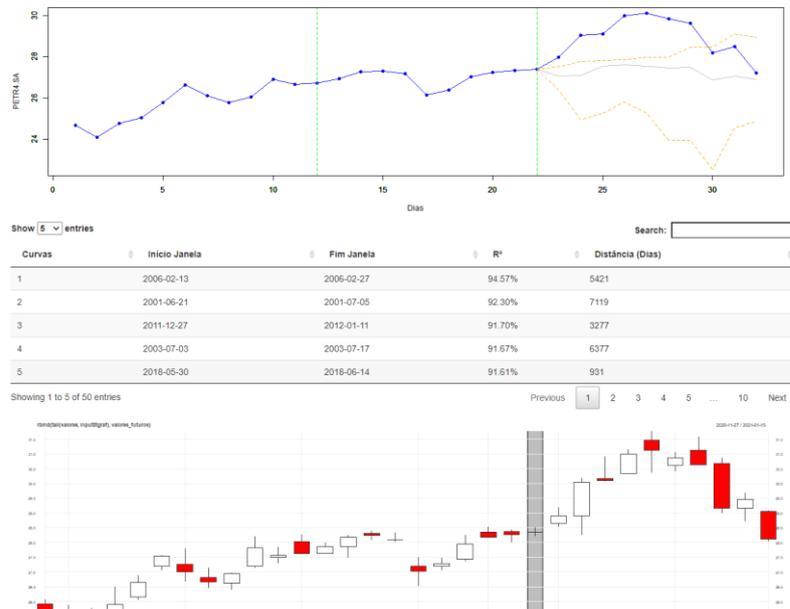


Figura 1 – Interface da aplicação Shiny desenvolvida

Conforme pode ser observado na Figura 1 a interface é composta por um painel lateral de entrada de dados e um painel principal com a saída corresponde ao especificado no painel lateral.

Dynamic Time Scan Forecasting

Digite o código do papel a ser avaliado:

Intervalo Analisado:

 to

Trabalhar com retornos

Número de melhores curvas a projetar:

Tamanho da janela de avaliação:

Ordem do polinômio da regressão:

Tamanho do período projetado:

Número de dias a mostrar no gráfico (antes da projeção):

Curvas projetadas a serem mostradas:

Mostrar combinação das curvas selecionadas

Ocultar curvas selecionadas

Mostrar combinação de todas as melhores curvas

Mostrar previsão do modelo ARIMA

Mostrar previsão do modelo de Alisamento Exponencial

Mostrar previsão do modelo TBATS

Figura 2 – Painel lateral

Detalhando o painel lateral, conforme mostrado na Figura 2, podemos definir os seus campos e as suas atribuições da seguinte forma:

1. **“Digite o código do papel a ser avaliado”**: neste campo deverá ser digitado o código da ação conforme listado no site *Yahoo Finance* para que a base histórica da mesma seja carregada do site através da utilização do pacote “QuantMod” do R.
2. **“Intervalo Analisado”**: nesse campo são especificadas as datas de corte de início e fim do período a ser analisado.

3. **"Trabalhar com retornos"**: essa caixa, quando marcada, faz com que a série seja convertida na série de retornos percentuais.
4. **"Número de melhores curvas a projetar"**: essa lista de valores (5, 10, 25 e 50) contém o número de melhores projeções (m) que o modelo irá gerar. Esse valor influenciará nos limites inferiores e superiores do intervalo de confiança.
5. **"Tamanho da janela de avaliação"**: essa lista de valores (11, 20 e 30) contém os possíveis tamanhos de janela de avaliação (w). Foi escolhido o valor onze como a menor opção, pois foi o menor valor de amostra com o qual o modelo não precisou forçar uma solução aproximada para a equação de regressão.
6. **"Ordem do polinômio da regressão"**: régua de valores entre 1 e 4 que define a ordem do polinômio de regressão a ser utilizado pelo modelo DTSP. Neste trabalho foi utilizado apenas a regressão linear (polinômio de ordem 1).
7. **"Tamanho do período projetado"**: régua de valores entre 1 e 15 que define o número de períodos a serem projetados pelo modelo.
8. **"Número de dias a mostrar no gráfico (antes da projeção)"**: é uma entrada com o intuito de auxiliar na visualização dos gráficos. É uma régua de valores entre 10 e 125 que irá definir quantos períodos serão mostrados nos gráficos antes do período projetado.
9. **"Curvas projetadas a serem mostradas"**: é uma lista de seleção múltipla na qual o usuário pode selecionar quais curvas projetadas deseja visualizar/combina.
10. **Sequência de caixas:**
 - a. **"Mostrar combinação das curvas selecionadas"**: Exibe uma curva vermelha no gráfico de linhas, que é resultado da combinação, através de mediana, das curvas selecionadas em "9";
 - b. **"Ocultar curvas selecionadas"**: Essa opção oculta as curvas selecionadas em "9", com o intuito de facilitar a visualização de curvas combinadas;

- c. **"Mostrar combinação de todas as melhores curvas"**: Exibe no gráfico de linhas a combinação via mediana de todas as m melhores curvas de acordo com o número selecionado em "4";
- d. **"Mostrar previsão do modelo ARIMA"**: Exibe no gráfico de linhas uma curva pontilhada com as previsões geradas pelo método ARIMA;
- e. **"Mostrar previsão do modelo de Alisamento Exponencial"**: Exibe no gráfico de linhas uma curva pontilhada com as previsões geradas pelo método ETS;
- f. **"Mostrar previsão do modelo TBATS"**: Exibe no gráfico de linhas uma curva pontilhada com as previsões geradas pelo método TBATS;

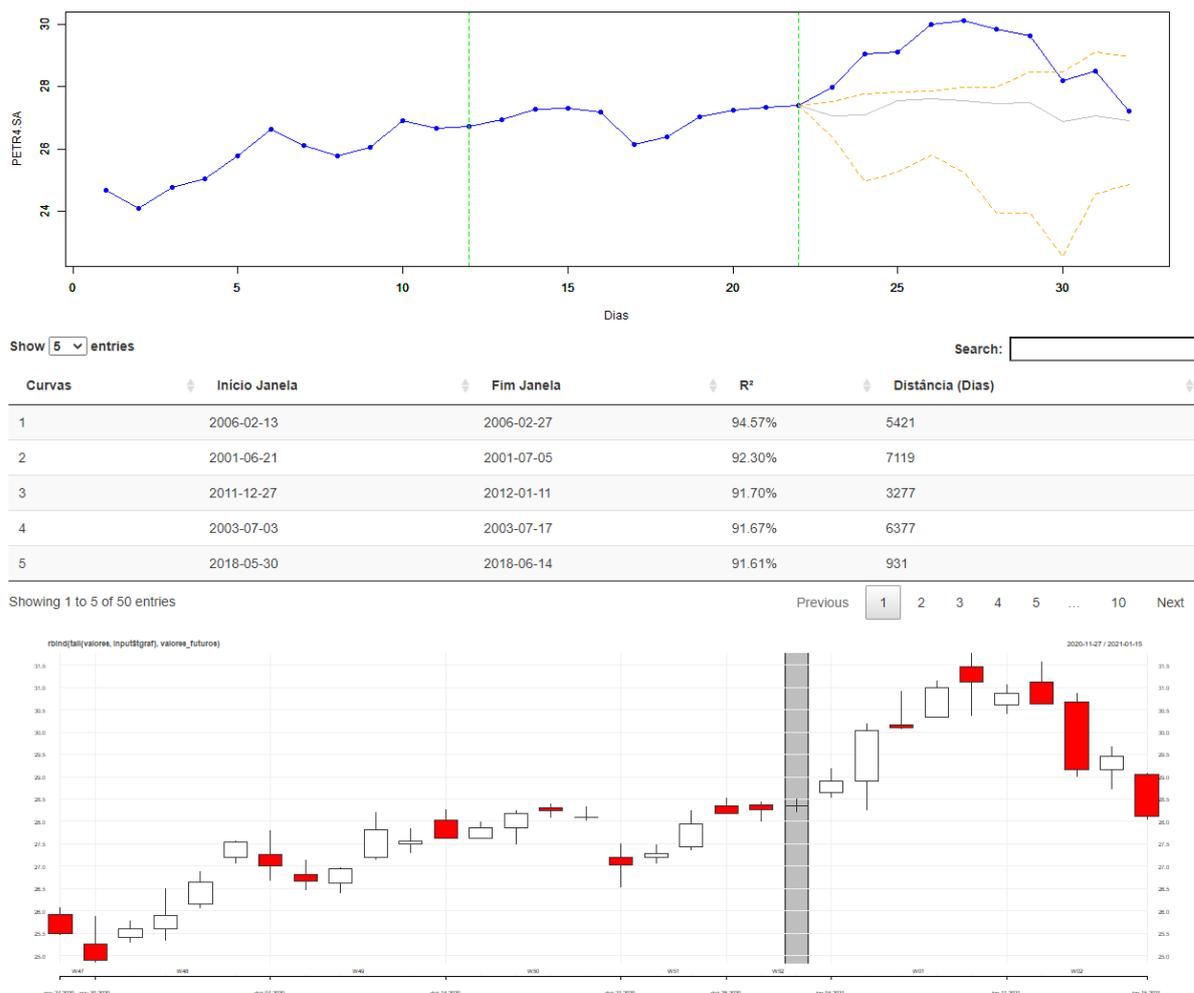


Figura 3 – Painel principal

Já no painel principal, conforme mostrado na Figura 3, podemos perceber a presença de três objetos:

1. O gráfico de linha;
2. A tabela resumo das melhores curvas projetadas e
3. O gráfico de *candlesticks*.

O gráfico de linhas está estruturado da seguinte forma: as linhas verticais pontilhadas e verdes marcam o início e o fim da janela de avaliação (w), isso quer dizer que a linha vertical mais à direita marca o último dado inserido no modelo. A linha azul é composta pelos dados reais da série analisada (em valores absolutos ou em retornos, a depender da configuração feita no painel lateral). Caso existam dados reais após o período avaliado, os mesmos serão mostrados em continuidade com a mesma cor. As linhas cinzas são formadas pelas diferentes projeções geradas pelo modelo e selecionadas no painel lateral (exceto se forem ocultadas). As linhas pontilhadas laranjas marcam os máximos e mínimos das curvas projetadas de acordo com a quantidade de curvas a serem geradas selecionada no painel lateral. As combinações de curvas projetadas e as curvas geradas pelos modelos comparativos serão exibidas conforme explicado no detalhamento do painel lateral.

A tabela resumo faz o detalhamento e ordenação das m melhores curvas geradas pelo modelo DTFSF em termos dos seguintes aspectos:

1. Ordem em termos de melhor ajuste de R^2 ;
2. Data de início da janela selecionada;
3. Data do fim da janela selecionada;
4. O valor do R^2 ajustado para a regressão dos dados da janela em questão em relação a janela de referência;
5. A distância em dias do final da janela selecionada em relação ao início do período projetado.

Finalmente o gráfico de *candlesticks* foi incluído no painel, pois é uma das visualizações mais comuns no mercado financeiro e, conforme mostrado em estudo anterior deste autor [7], existem certos padrões que podem funcionar como bons alarmes de reversão de tendência, em especial para reversão para queda. No entanto, para este trabalho não foi considerado o efeito de nenhum desses padrões.

2.4 OBJETO DE ESTUDO

Para essa investigação foram selecionados os preços históricos das ações preferenciais da Petrobrás (Petróleo Brasileiro S.A. – código PETR4). Essa escolha foi motivada em função do fato de essa ação estar sempre entre as mais negociadas na Bolsa de Valores brasileira (B3) e também pelo fato de ser uma ação com histórico de volatilidade sensível a diversas questões políticas e econômicas do Brasil e do mundo.

Os dados foram obtidos da página do *Yahoo Finance* através do uso do pacote “QuantMod” do R. Foram coletadas as informações entre janeiro de 2000 a maio de 2021, totalizando 5.132 observações diárias. Dentre as colunas de dados disponibilizadas, foi escolhida a coluna chamada “Ajustada”, pois traz os preços de fechamento de pregão diário com os ajustes de distribuição de dividendos, *splits*, *implits*, entre outros.

3 RESULTADOS

Para avaliar os resultados obtidos com o método uma bateria de testes foi realizada. Em primeiro lugar, dado que o método não produz um intervalo de confiança analítico, avaliou-se empiricamente quais eram os intervalos de confiança gerados pelos intervalos de máximos e mínimos do conjunto de projeções. Na sequência, realizou-se uma avaliação do método em termos de capacidade de prever valores ou tendências futuras em momentos de relativa estabilidade do mercado. Depois foi feita uma análise da capacidade do método em lidar com momentos de aumento de volatilidade do mercado. Por fim, analisou-se o método DTSF em comparação com os demais métodos preditivos.

3.1 DEFINIÇÃO DOS INTERVALOS DE CONFIANÇA

A fim de obter-se melhor entendimento do percentual de dados reais que não excedem os limites dos valores previstos pelo conjunto de previsões geradas a depender do número definido no painel lateral, foi realizado um teste utilizando-se a série de dados até a observação de número 4000 e gerando-se 10 previsões para cada dia a partir dessa observação e avançando uma observação a cada passo. A cada iteração eram estabelecidos os máximos e mínimos do conjunto de previsões e contados quantos valores reais ficavam fora dos intervalos estabelecidos.

O teste foi feito para todos os valores de números de previsões disponíveis (5, 10, 25 e 50) e para todos os tamanhos de janela de avaliação disponíveis (11, 20 e 30). Após a realização de ditos testes chegou-se ao seguinte resultado:

Número de melhores previsões	Tamanho da Janela de Avaliação		
	11 dias	20 dias	30 dias
5 previsões	67,85%	63,97%	58,56%
10 previsões	81,48%	77,80%	73,89%
25 previsões	90,30%	87,71%	85,00%
50 previsões	94,73%	92,09%	89,45%

Tabela 4 – Intervalo de confiança

Como pode ser observado, o período de 11 dias para a janela de avaliação produziu resultados sistematicamente melhores do ponto de vista de abranger o máximo de observações reais em relação aos demais tamanhos. Em relação ao número de previsões, os limites fornecidos pelas 50 melhores projeções em média ficaram próximos de fornecer um intervalo de 95% de confiança.

Tendo estes resultados em mente, os testes daqui em diante serão realizados com estes parâmetros (50 melhores projeções geradas pela janela de avaliação de 11 dias).

3.2 AVALIAÇÃO DAS PROJEÇÕES EM MOMENTOS DE ESTABILIDADE DO MERCADO

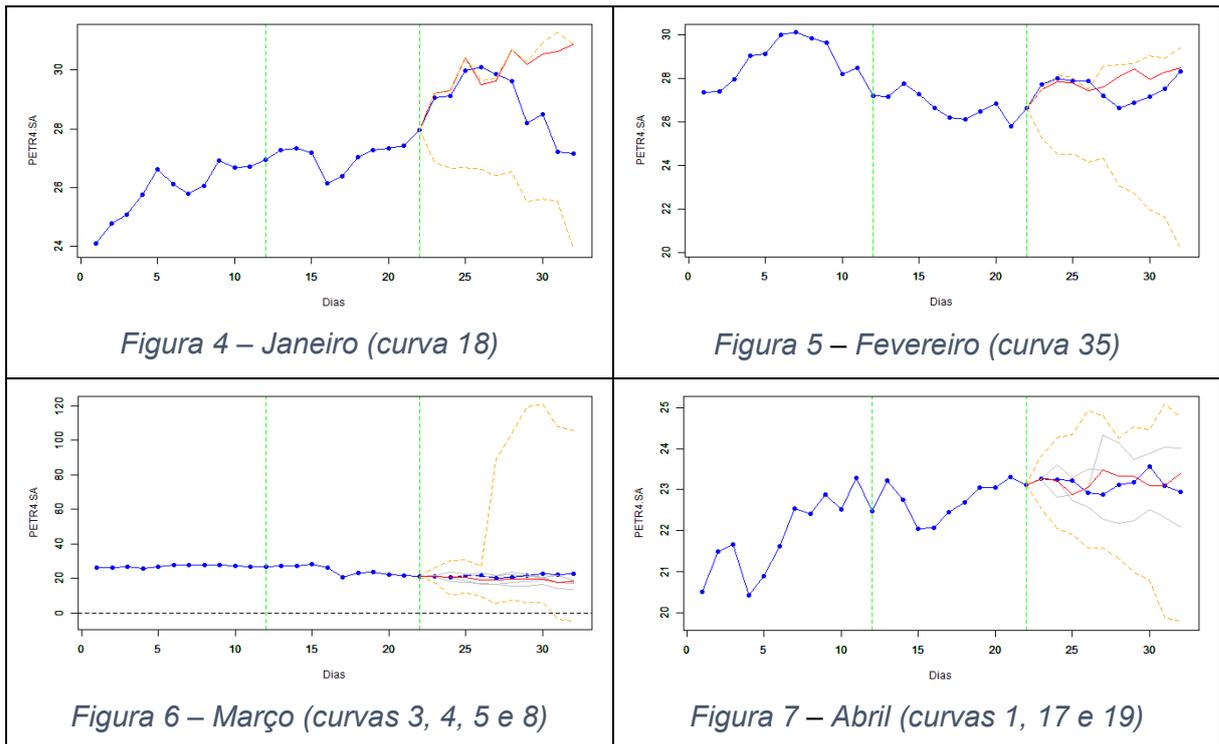
Para avaliar a qualidade das previsões geradas pelo método DTSTF em momentos de relativa estabilidade do mercado financeiro, foram selecionados como pontos de análise os primeiros dias de cada mês de 2021. Foram incluídos os meses de janeiro a maio.

Como foi adotado como ótimo o parâmetro de 50 melhores projeções, definiu-se que seria escolhida a curva que melhor previsse o primeiro ponto real como critério de seleção da projeção com valores mais prováveis de acerto. No caso de existirem duas ou mais curvas que fossem muito precisas em prever o primeiro ponto, a previsão seria gerada através da combinação dessas curvas.

Os testes então foram divididos entre testes feitos com a série original e com a série de retornos

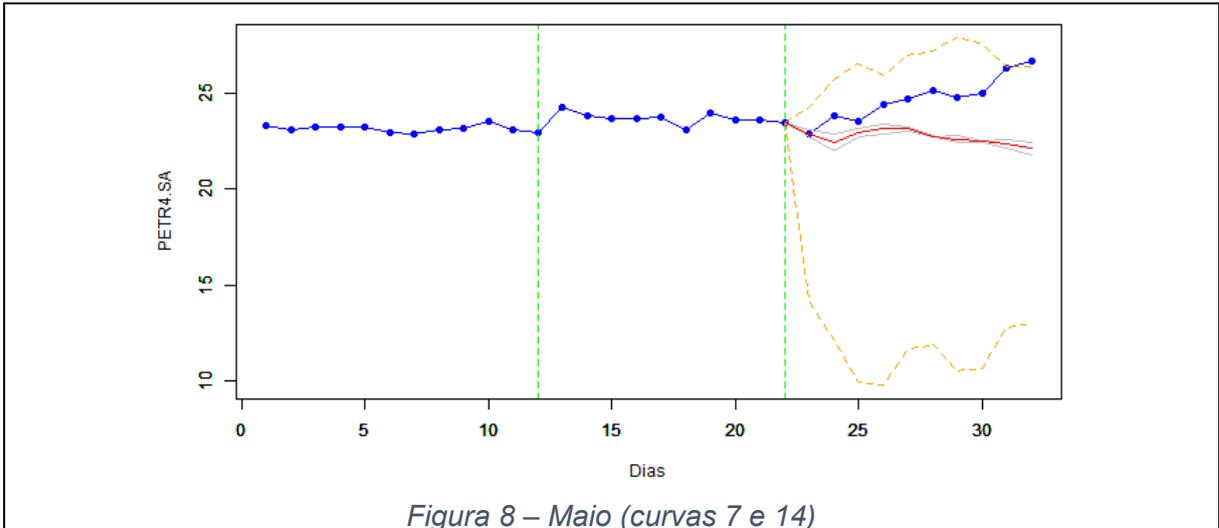
3.2.1 Testes com a série original

Conforme explicado anteriormente, foram realizadas as previsões considerando a data corte como o primeiro dia útil dos meses avaliados. Para os meses de janeiro a abril, foram obtidos os seguintes resultados:



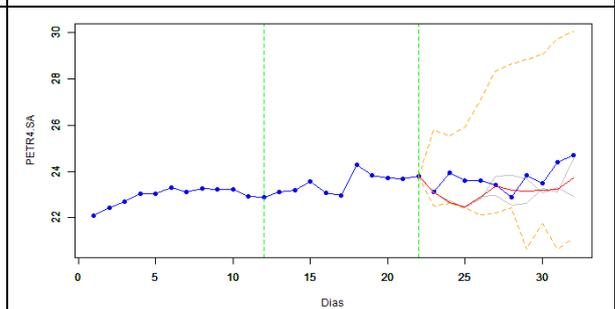
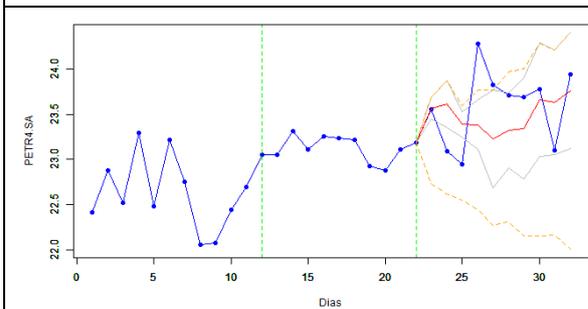
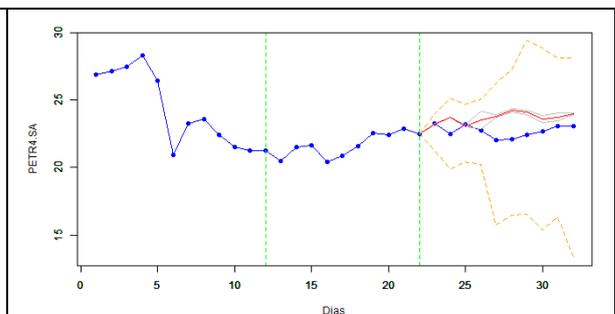
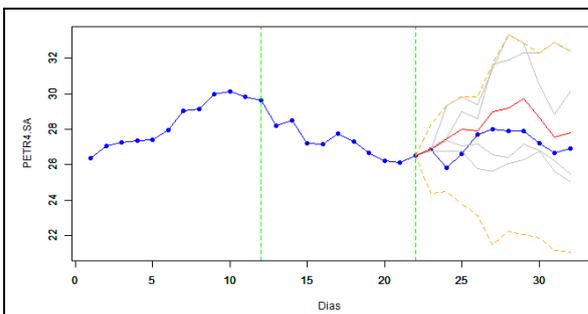
Nestes quatro gráficos é possível perceber uma boa aderência do modelo em termos de previsão. Considerando que a primeira previsão foi usada como critério de seleção da(s) curva(s), avaliando os períodos subsequentes é possível notar uma boa aderência nos primeiros 3 a 4 valores projetados. Algo interessante é que em dois casos – fevereiro e abril – houve uma certa divergência entre as previsões e dados reais a partir do quinto valor previsto, mas houve nova convergência a partir do nono e décimo valor.

O gráfico das previsões geradas para os primeiros dias de maio, por sua vez, pode ser visto na figura 8. Neste caso, podemos perceber que houve um descolamento da previsão em relação aos valores reais. Algo diferente desse extrato de dados em relação aos extratos anteriores é que a curva real sofre uma queda no primeiro ponto a ser previsto para logo em seguida voltar a subir com certa força. Isso suscita a dúvida sobre a capacidade preditiva do método em relação a momentos de fortes oscilações alternadas. Isso será investigado a seguir.



3.2.2 Testes em dados com reversão de tendência

Conforme identificado nos primeiros dias de maio, o modelo DTSF pareceu ter dificuldade de prever reversões de tendência no período imediatamente após o período de seleção de curvas. Para poder investigar essa questão com maior profundidade, foram identificados mais quatro momentos nos quais os dados reais apresentavam esse comportamento de alternância do sentido de oscilação.



Nas figuras de 9 a 12 podemos ver os gráficos das projeções realizadas para estes quatro momentos. De fato, conforme observado nos primeiros dias de maio, o modelo não foi capaz de prever a reversão logo após o período de seleção de curvas. No entanto, diferentemente do que aconteceu no primeiro exemplo, nestes quatro casos a previsão conseguiu nos períodos posteriores aproximar-se novamente dos dados reais.

3.2.3 Testes com a série de retornos

Para as previsões de retornos a curva de combinação das previsões apenas foi utilizada quando as duas curvas mais próximas do primeiro valor real ainda estavam relativamente distantes, assim a reta de combinação conseguiria se aproximar do primeiro ponto. Não se usou mais de duas curvas combinadas, pois observou-se que isso produziria um efeito de regressão para a média.

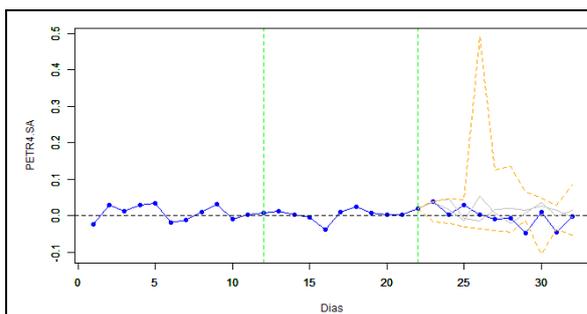


Figura 13 – Janeiro (curvas 45 e 50)

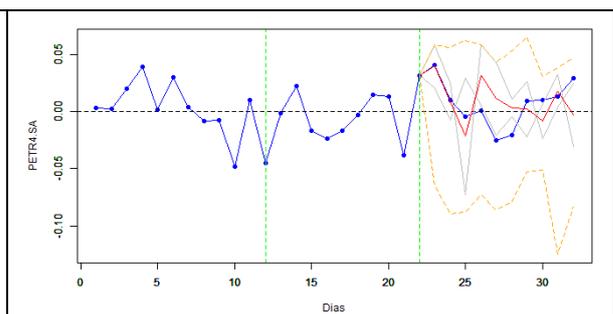


Figura 14 – Fevereiro (curvas 36 e 37)

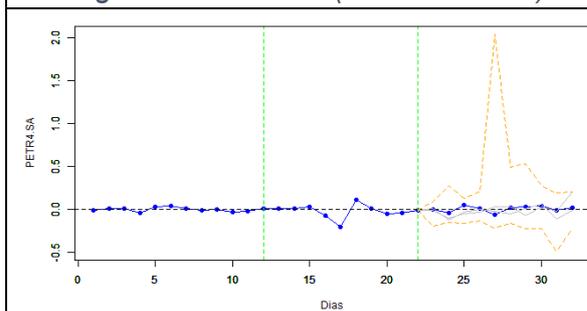


Figura 15 – Março (curvas 9 e 15)

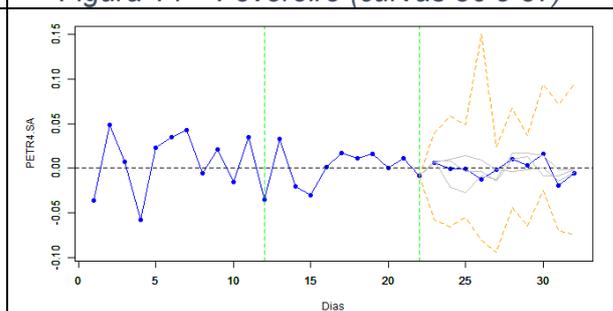


Figura 16 – Abril (curvas 11, 20 e 35)

Observando os gráficos, para estes casos avaliados, diferentemente do que se observou com os valores originais, o único período previsto consistentemente com alguma precisão é o período subsequente ao primeiro dado real (usado para selecionar as curvas utilizadas). No entanto, nestes períodos de estabilidade, a metodologia gerou previsões aderentes à volatilidade dos dados reais.

3.3 AVALIAÇÃO DAS PROJEÇÕES EM MOMENTOS DE INSTABILIDADE DO MERCADO

Com o intuito de observar a capacidade do modelo de se ajustar e gerar previsões aderentes em momentos de instabilidade do mercado, foram selecionados dois episódios recentes que promoveram grande impacto nos preços de negociação das ações da Petrobrás. Foram eles:

- Início dos temores do impacto do que viria a se tornar a pandemia de Covid-19 no mundo e primeiro caso confirmado da doença no Brasil, ocorrida na última semana de fevereiro de 2020.
- Anúncio do governo da troca do presidente da Petrobrás no dia 19 de fevereiro de 2020.

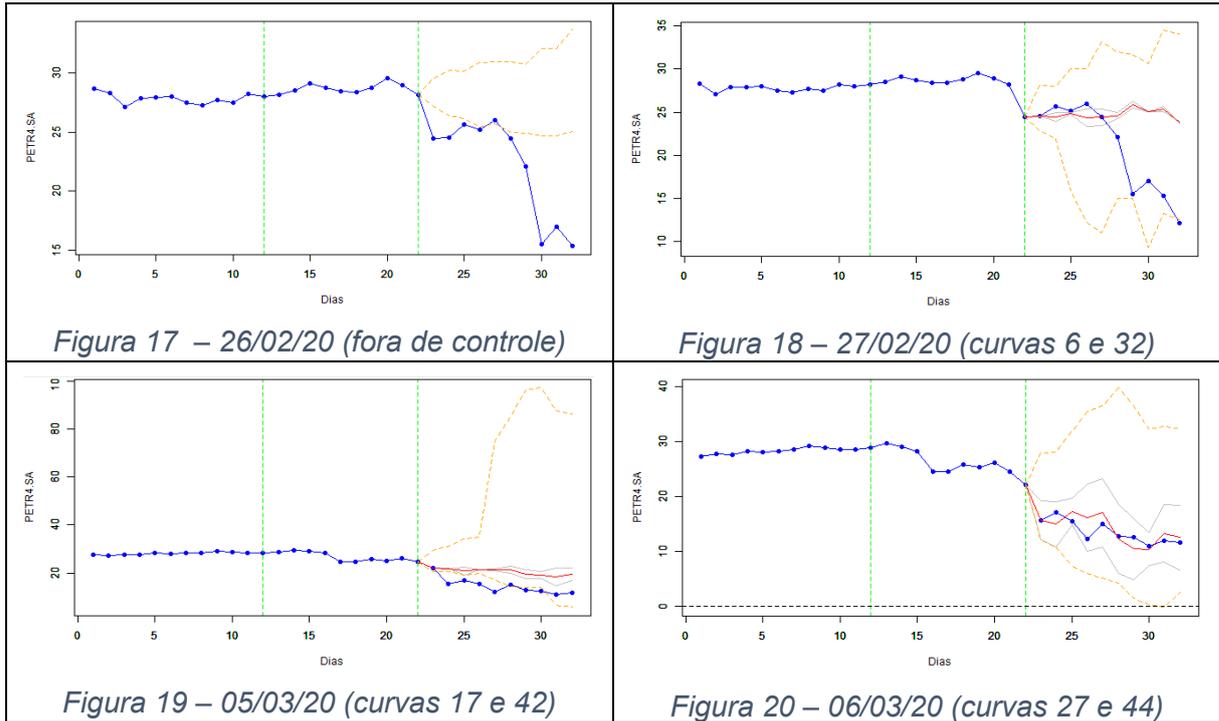
Estes são episódios recentes da história e que tiveram impacto relevante sobre o preço das ações da Petrobrás em particular e no mercado como um todo no caso da pandemia.

3.3.1 Início dos casos de Covid-19 no Brasil e no mundo

Apesar de os primeiros casos da doença que iria se tornar um caso de pandemia terem sido registrados no dia 8 de dezembro de 2019 em Wuhan na China [13], foi apenas na segunda quinzena (em especial na última semana) de fevereiro de 2020 que as tensões sobre os possíveis impactos do Covid-19 começaram a refletir nas oscilações dos mercados financeiros mundiais.

A partir do dia 27 de fevereiro de 2020 os preços da ação preferencial da Petrobrás iniciaram uma forte tendência de baixa, saindo do patamar dos R\$30 por ação na semana anterior para chegarem a ser negociados na casa dos R\$10 nos dias 18 e 19 de março.

Fazendo um acompanhamento evolutivo nos dias críticos dessa oscilação, chegamos nas seguintes previsões:



No dia 27 de fevereiro, utilizando-se a metodologia de ajustar a melhor curva para prever este dado com a janela de avaliação se encerrando no dia anterior (figura 17), percebemos que o dado real está fora dos limites de previsão do modelo. Isso é um indicativo que os dados reais estão fora dos limites de controle do modelo. No caso do DTSF infere-se que nada parecido foi detectado como análogo a esse período no passado que pudesse sinalizar o abrupto aumento de volatilidade que estava por vir.

Seguindo então para o dia 28 (figura 18), vemos que uma vez passada a primeira grande queda do mercado há uma estabilidade temporária dos preços, na qual o modelo foi capaz de gerar previsões aderentes pelos próximos quatro períodos subsequentes, de forma similar ao que foi observado nos períodos de estabilidades analisados na seção 3.2.1.

Avançando para o primeiro período sem aderência à previsão do dia 28 de fevereiro, vemos que no dia 6 de março (figura 19), apesar de termos curvas aderentes ao dado real deste dia, as previsões não foram aderentes dali em diante, devido à forte baixa no dia seguinte.

Finalmente, no dia 7 de março (figura 20), vemos que, devido à volatilidade presente na janela de avaliação, o modelo já é capaz de antever a volatilidade que

viria a seguir e, fazendo uso da metodologia já descrita e ajustando as curvas que melhor preveem o primeiro período vemos que foi obtida uma sequência de previsões muito aderentes aos dados reais em todo o horizonte de previsão.

Essas boas previsões para os dados não transformados em cenários de alta volatilidade induz naturalmente a dúvida sobre a capacidade do método em realizar previsões igualmente boas com os dados diferenciados (retornos). Essas previsões podem ser vistas a seguir:

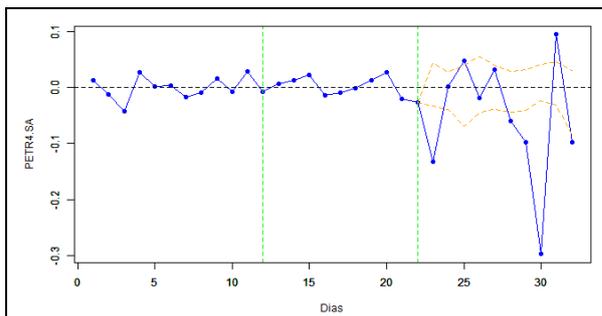


Figura 21 – 26/02/20 (fora de controle)

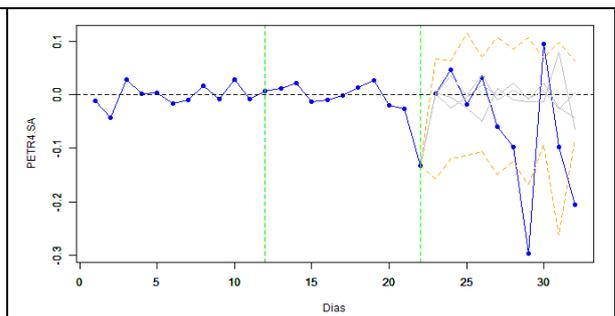


Figura 22 – 27/02/20 (curvas 18, 27 e 49)

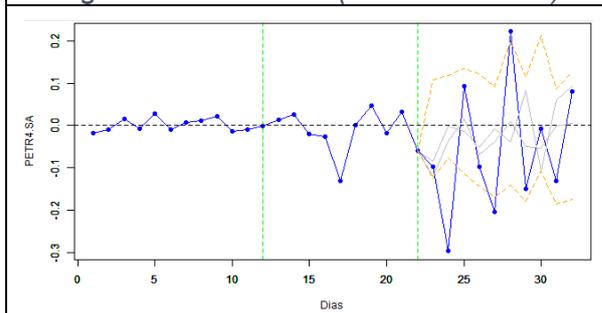


Figura 23 – 05/03/20 (curvas 23 e 25)

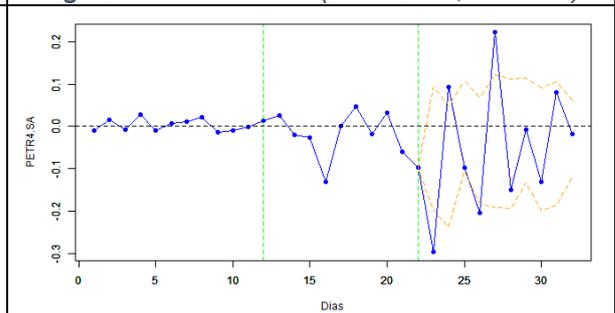


Figura 24 – 06/03/20 (fora de controle)

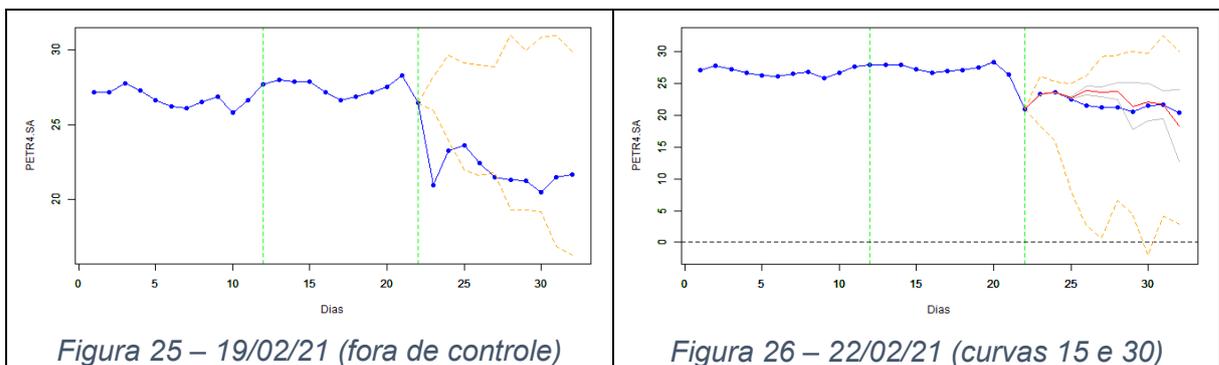
Analisando a sequência, vemos que, em congruência ao observado com os dados originais, no dia 27 de fevereiro o dado real está fora dos limites de previsão gerados pelas curvas do dia 26 (figura 21). No entanto, no dia 7 de março, diferentemente do que acontece com os dados originais, a série de retornos fica fora do limite de controle o que nos leva a concluir que as previsões geradas não devem ser consideradas confiáveis.

No geral, a série de retornos nos mostram o crescimento da volatilidade futura, no entanto muitos dados ficam fora dos limites de controle da previsão, o que sugere que trabalhar com os dados não diferenciados gera percepções mais assertivas sobre os períodos previstos do que trabalhar com os retornos.

3.3.2 Troca do presidente da Petrobrás

No dia 19 de fevereiro de 2021, numa sexta-feira à noite, o presidente Jair Bolsonaro anunciou Joaquim Silva e Luna para a presidência da Petrobrás no lugar de Roberto Castello Branco. Esta troca gerou muita preocupação no mercado em relação a um movimento intervencionista na estatal [14,15] e isso fez com que as ações preferenciais (PETR4) fechassem em queda de mais de 21% no pregão do dia 22 de fevereiro, a segunda-feira posterior ao anúncio [15].

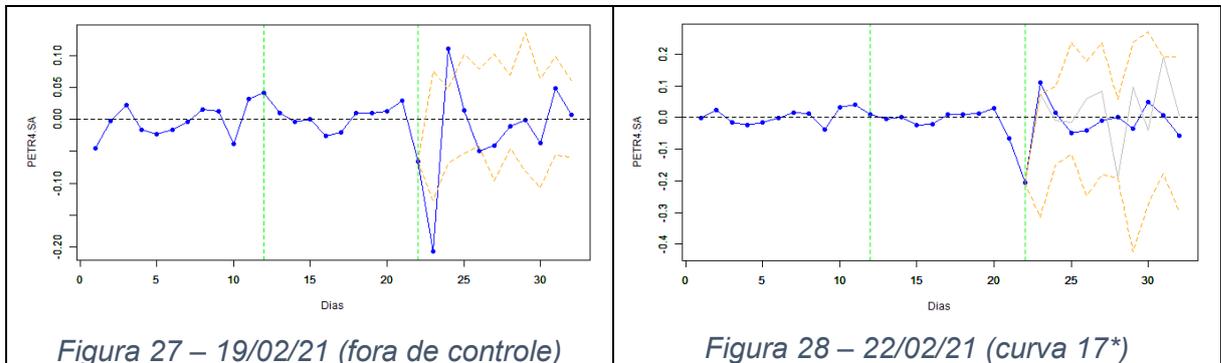
Similarmente ao que foi feito em 3.3.1, vamos fazer um acompanhamento evolutivo das oscilações ocasionadas por esse evento e as previsões geradas pelo DTFSF:



Partindo-se do dia 22 de fevereiro, podemos ver que o primeiro dado real após a data de corte para a previsão (19/fev/21) está fora dos limites de controle do DTFSF. Isso já era esperado, dado que a informação que causou o impacto foi divulgada fora do horário de pregão, assim os dados de preços até o dia 19 não continham nenhum (ou talvez, quase nenhum [16]) efeito de dita informação.

No dia 23 de fevereiro, por sua vez, utilizando-se as informações até o dia 22 e ajustando as melhores curvas de previsão ao dado real obtém-se uma curva bastante aderente aos dados reais ao longo de todo período projetado. Esse período, no entanto, foi bem mais estável, uma vez que o impacto inicial do anúncio de troca de presidente da Petrobrás não gerou maiores desdobramentos subsequentes e o modelo, conforme já observado anteriormente, tem boa capacidade preditiva em momentos de estabilidade.

Novamente, vamos analisar o comportamento do modelo diante da série de retornos dos dados:



Diferentemente do observado na série não diferenciada, a análise dos retornos nos apresenta o primeiro dado real fora do intervalo de controle nos dois casos. No entanto, para a análise do dia 23 (com curvas projetadas com dados até o dia 22) o dado real está fora, mas relativamente próximo do limite de controle, assim foi testada a curva mais próxima. Mas, apesar de alguma aderência ao primeiro dado real subsequente ao dado de escolha de curva, a previsão aponta para uma volatilidade maior que a real.

Finalmente, observando esses dois casos de instabilidade no mercado foi possível perceber neles, que o modelo, apesar de naturalmente não ser capaz de prever efeitos anômalos nos preços, mostra uma capacidade de reagir a esses eventos com bastante celeridade, algumas vezes até incorporando o aumento de volatilidade causado e com boa consistência, capturando tendências de curto prazo.

3.4 COMPARAÇÃO DO MÉTODO DTSF COM O ARIMA, ETS E TBATS

Para avaliar o método DTSF em comparação com os métodos preditivos mais comuns utilizados para a análise de séries temporais foram selecionados os métodos ARIMA, ETS e TBATS.

Para fazer essa comparação, foram utilizados os primeiros dias de cada um dos cinco primeiros meses de 2021, de forma similar ao realizado em 3.2.1 e 3.2.3. No entanto, dada a similaridade de resultados entre todos os meses, serão

apresentados os gráficos apenas de janeiro e fevereiro, podendo seus resultados serem generalizados para os demais meses.

Primeiramente avaliamos janeiro de 2021:

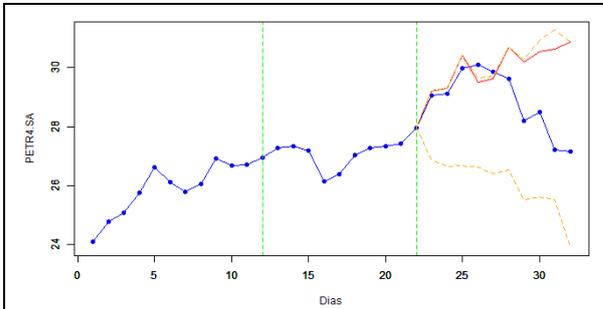


Figura 29 – DTSF (curva 18) – Janeiro

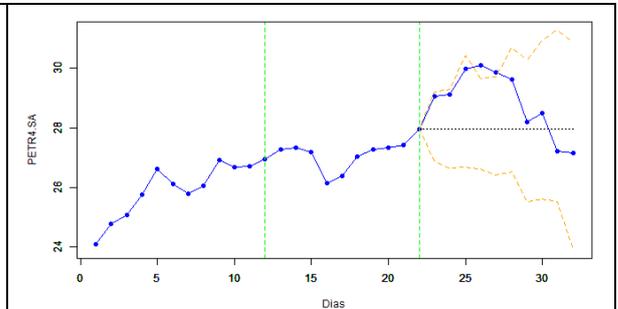


Figura 30 – ARIMA – Janeiro

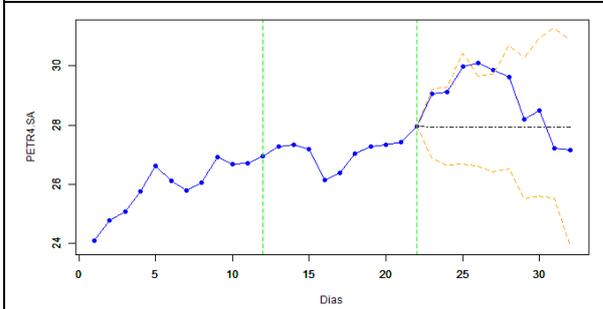


Figura 31 – ETS – Janeiro

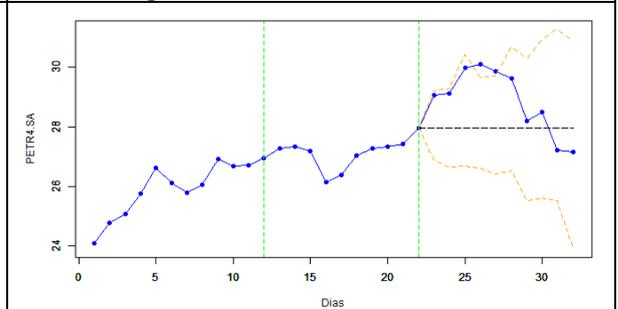


Figura 32 – TBATS – Janeiro

E na sequência, fevereiro:

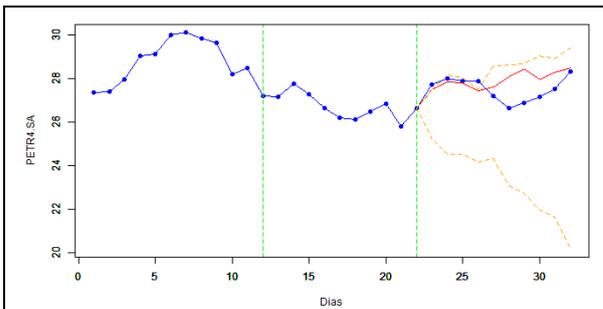


Figura 33 – DTSF (curva 35) – Fevereiro

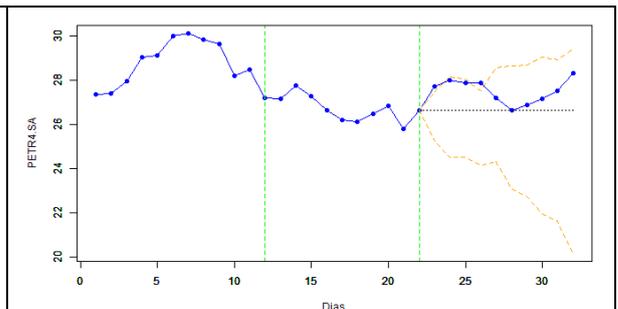


Figura 34 – ARIMA – Fevereiro

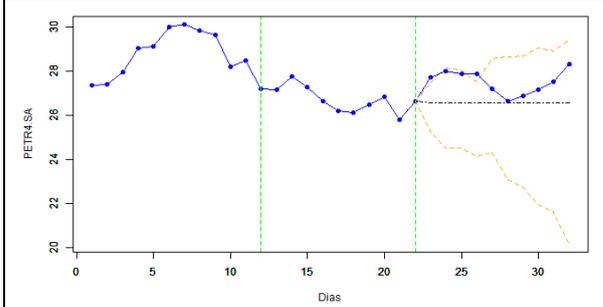


Figura 35 – ETS – Fevereiro

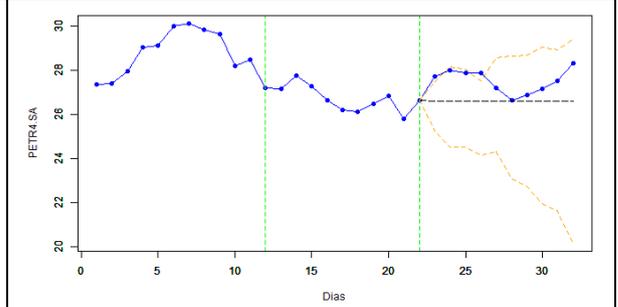


Figura 36 – TBATS – Fevereiro

Foi feita também a análise considerando a série de retornos. Assim temos para janeiro:

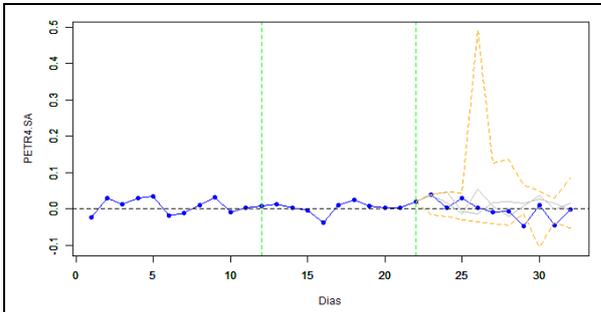


Figura 37 – DTSF (curva 45 e 50) – Janeiro

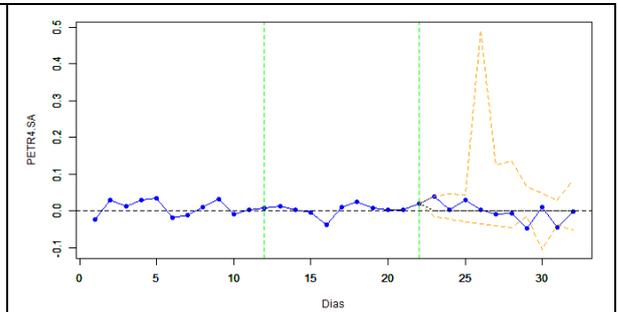


Figura 38 – ARIMA – Janeiro

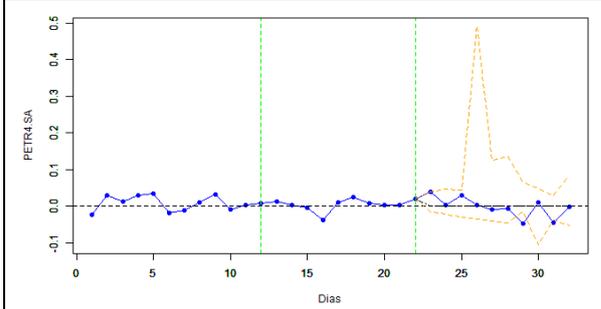


Figura 39 – ETS – Janeiro

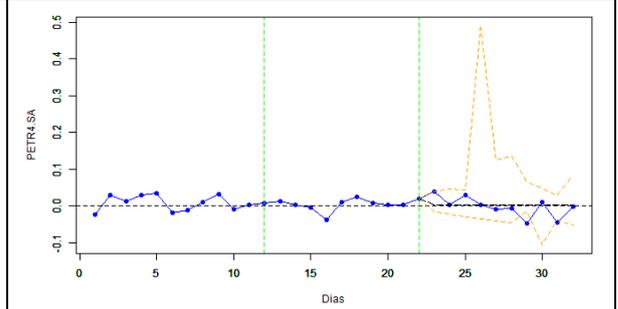


Figura 40 – TBATS – Janeiro

Fevereiro:

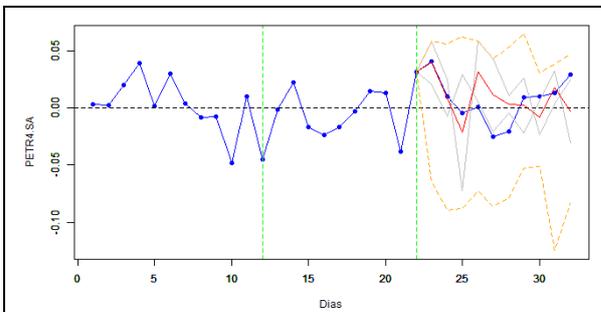


Figura 41 – DTSF (curvas 36 e 37) – Fevereiro

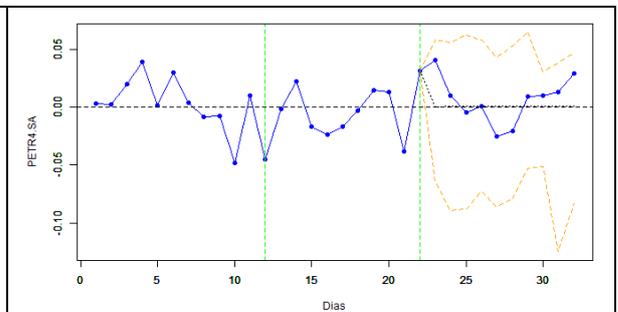


Figura 42 – ARIMA – Fevereiro

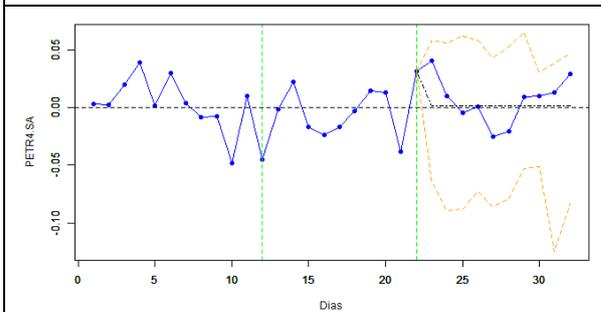


Figura 43 – ETS – Fevereiro

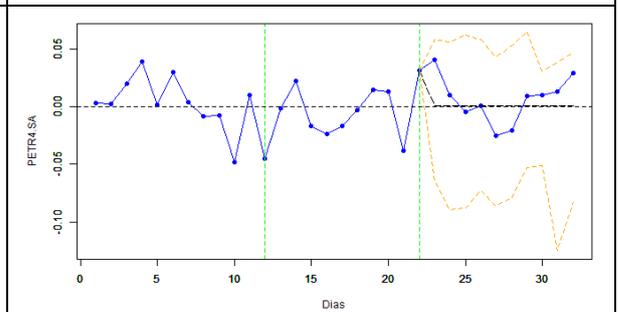


Figura 44 – TBATS – Fevereiro

Como é possível perceber, todos os métodos alternativos ao DTSTF não foram capazes de perceber informação na série histórica da PETR4, classificando-a assim como um movimento de caminho aleatório. Dessa forma a única previsão que esses métodos foram capazes de gerar foi a repetição do último dado real.

Com essa constatação em mente, pode-se dizer que o DTSTF é um método interessante a ser usado como alternativa aos avaliados no momento que estes se mostrarem ineficazes em detectar informação preditiva na série temporal em questão.

4 CONCLUSÃO

Este trabalho se propôs a investigar a possível eficiência do método *Dynamic Time Scan Forecasting* (DTSF) em realizar previsões de preços de ativos na bolsa de valores. Especificamente, foi avaliada sua eficiência em construir previsões para as ações preferenciais da Petrobrás. Inicialmente desenvolvido e testado para prever a velocidade média de ventos como ferramenta de planejamento na geração de energia eólica, a lógica do método de buscar padrões anteriores de alta similaridade ao padrão imediatamente anterior ao período que se deseja projetar instigou a curiosidade em prever movimentos do mercado financeiro, dado que é a mesma de muitas análises preditivas já utilizadas em dito mercado.

Ao longo de todos os testes realizados a conclusão mais forte que se pôde chegar foi a de que o método DTSF é uma ferramenta poderosa em prever movimentos de curto prazo (quatro a dez dias úteis) em momentos de relativa estabilidade com os dados não diferenciados da série histórica. Como já era de se esperar, o método é incapaz de prever movimentos atípicos do mercado, no entanto ele é capaz de se ajustar a esses impactos com rapidez e produzir, logo que concretizado o movimento atípico, previsões aderentes aos períodos subsequentes. Mesmo em movimentos de forte oscilação no sentido das variações o método ainda foi capaz de apontar uma tendência aderente com a realidade.

Em relação ao uso do método com a série de retornos percebeu-se que ela foi menos eficiente em produzir entendimentos mais profundos sobre a série analisada, ainda que tenha se mostrado capaz de refletir em algum grau as variações de volatilidade nos preços do ativo avaliado.

Finalmente, ao compará-lo com as metodologias mais tradicionais da estatística em termos de modelagem de séries temporais, foi possível notar que o DTSF se adequa com muito mais facilidade a fenômenos comuns da série histórica de preços de ativos financeiros – em particular as de dados diários – tais como a volatilidade e ausência de padrões claros de sazonalidade. Enquanto métodos tradicionais como o ARIMA, o alisamento exponencial e o TBATS perceberam a série avaliada como um caminho aleatório e foram incapazes de produzir previsões de movimentos de mercado, o método DTSF foi capaz de produzir várias previsões e, aliado a estratégia de seleção de curvas com o primeiro dado real, capaz de produzir

boa aderência à realidade, mesmo em momentos de certa instabilidade na movimentação de preços.

5 REFERÊNCIAS

- [1] COSTA, Marcelo Azevedo, et al. Dynamic Time Scan Forecasting for multi-step wind speed prediction. *Renewable Energy*, 2021.
- [2] NISON, Steve. *Japanese candlestick charting techniques: a contemporary guide to the ancient investment techniques of the Far East*. Penguin, 2001.
- [3] MURPHY, John J. *Technical analysis of the financial markets: A comprehensive guide to trading methods and applications*. Penguin, 1999.
- [4] MALLIKARJUNA, Mehari; RAO, R. Prabhakara. Evaluation of forecasting methods from selected stock market returns. *Financial Innovation*, 2019, 5.1: 1-16.
- [5] NAU, Robert. Introduction to ARIMA models. *Introduction to ARIMA Models, people.duke.edu/~rnau/Slides_on_ARIMA_models--Robert_Nau.pdf*, 2014.
- [6] BOX, George EP, et al. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.
- [7] COELHO, Bernardo, FREITAS, Bernardo *Estratégias de trade através do uso de padrões de candlesticks*, Trabalho de Conclusão de Curso (Graduação em Engenharia de Produção) - Universidade Federal de Minas Gerais; Orientador: Anderson Laécio Galindo Trindade, 2012.
- [8] Statsmodels. ETS models, 2021. Página inicial. Disponível em: <https://www.statsmodels.org/>. Acesso em: 21, jun. 2021.
- [9] HYNDMAN, Rob J.; ATHANASOPOULOS, George. *Forecasting: principles and practice*. OTexts, 2018.
- [10] HYNDMAN, Rob J., et al. A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of forecasting*, 2002, 18.3: 439-454.
- [11] DE LIVERA, Alysha M.; HYNDMAN, Rob J.; SNYDER, Ralph D. Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American statistical association*, 2011, 106.496: 1513-1527.
- [12] BOX, George EP; COX, David R. An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 1964, 26.2: 211-243.
- [13] Linha do tempo do Coronavírus no Brasil. Sanar Saúde, 11, mai. De 2021. Disponível em: <https://www.sanarmed.com/linha-do-tempo-do-coronavirus-no-brasil>. Acesso em 23, jun. 2021.
- [14] Bolsonaro troca presidente da Petrobras; repercussão. G1, 19, fev. de 2021. Disponível em: <https://g1.globo.com/economia/noticia/2021/02/19/bolsonaro-troca-presidente-da-petrobras-repercussao.ghtml>. Acesso em: 23, jun. 2021.
- [15] Ações da Petrobras despencam mais de 21% após anúncio de troca de comando. UOL, São Paulo, 22, fev. de 2021. Disponível em: <https://economia.uol.com.br/cotacoes/noticias/redacao/2021/02/22/acoes-petrobras-bolsa-de-valores.htm>. Acesso em: 23, jun. 2021.

[16] Petrobras: suspeita de insider trading antes de mudança na presidência. Exame, 02, mar. de 2021. Disponível em: <https://invest.exame.com/me/petrobras-suspeita-de-insider-trading-antes-de-mudanca-na-presidencia>. Acesso em: 23, jun. 2021.

[17] EDWARDS, Robert D.; MAGEE, John; BASSETTI, WH Charles. *Technical analysis of stock trends*. CRC press, 2018.

[18] PEGELS, C. Carl. Exponential forecasting: some new variations. *Management Science*, 1969, 311-315.

[19] GARDNER JR, Everette S. Exponential smoothing: The state of the art. *Journal of forecasting*, 1985, 4.1: 1-28.

[20] SNYDER, Ralph. Discussion| International Journal of Forecasting-Volume 22, Issue 4. *International Journal of Forecasting*, 2006, 22.4: 673-676.