

DETECÇÃO E ANÁLISE MULTI-SCALE DE EVENTOS ORIGINADOS POR
FATORES EXTERNOS DE INCERTEZA EM SÉRIES FINANCEIRAS

Diego Silva de Salles

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Ciência da Computação, Centro Federal de Educação Tecnológica Celso Suckow da Fonseca CEFET/RJ, como parte dos requisitos necessários à obtenção do grau de mestre.

Orientadores:
Eduardo Soares Ogasawara
Eduardo Bezerra da Silva

Rio de Janeiro,
Março de 2023

**Detecção e análise multi-scale de eventos originados por fatores
externos de incerteza em séries financeiras**

Dissertação de Mestrado em Ciência da Computação, Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, CEFET/RJ.

Diego Silva de Salles

Aprovada por:

Presidente, Professor. D.Sc. Eduardo Soares Ogasawara (CEFET-RJ)

D.Sc. Eduardo Bezerra da Silva (coorientador) (CEFET-RJ)

D.Sc. Rafaelli de Carvalho Coutinho (CEFET-RJ)

D.Sc. Carlos Eduardo Mello (UNIRIO)

Rio de Janeiro,
Março de 2023

DEDICATÓRIA

“E eis que vou hoje pelo caminho de toda a terra; e vós bem sabeis, com todo o vosso coração, e com toda a vossa alma, que nem uma só palavra falhou de todas as boas coisas que falou de vós o SENHOR vosso Deus; todas vos sobrevieram, nenhuma delas falhou.” Josué 23:14

AGRADECIMENTOS

A Deus primeiramente, por estar sempre a meu lado em minhas orações, e nos momentos mais difíceis me dar forças pra prosseguir. A Agnes, pela paciência e compreensão nos vários momentos em que, mesmo estando em casa, eu estava ausente. Pelo seu apoio e paciência, quando falava assuntos relacionados ao mestrado que ela sequer entendia, mas ouvia só pra me incentivar. A minha avó Neiva, que sempre me auxiliou em toda essa jornada, e mesmo não estando mais perto, sua generosidade permitiu que eu pudesse sonhar os meus sonhos. Aos meus pais que sempre me ensinaram o caminho da verdade e me aproximaram de Deus através dos ensinamentos passados ao longo dos anos. A meu orientador Dr Eduardo Ogasawara pelos ensinamentos e direcionamentos, e mesmo com as minhas limitações, sempre me ajudou e não desistiu de me auxiliar nessa jornada. Ao meu coorientador professor Dr. Eduardo Bezerra pelas orientações me guiando no melhor caminho a seguir. Serei sempre grato pelos conselhos, sugestões e paciência dos dois. Aos demais professores do PPCIC, e em especial a Rafaelli de Carvalho Coutinho que muito me auxiliou nos comentários da qualificação.

RESUMO

Detecção e análise multi-scale de eventos originados por fatores externos de incerteza em séries financeiras

Diego Silva de Salles

Orientadores:

Eduardo Soares Ogasawara

Eduardo Bezerra da Silva

Resumo da Dissertação submetida ao Programa de Pós-graduação em Ciência da Computação do Centro Federal de Educação Tecnológica Celso Suckow da Fonseca CEFET/RJ como parte dos requisitos necessários à obtenção do grau de mestre.

Existem diferentes fatores externos relatados na mídia que podem impactar uma série temporal financeira. Tais fatores, como transições de governo, crises econômicas ou escândalos de corrupção, podem estar relacionados a eventos que aumentam a incerteza na série temporal. Em particular, esses fatores externos podem aumentar o risco percebido em uma série temporal financeira por meio de eventos como anomalias ou pontos de mudança. Um estudo baseado nas diferentes características que compõem um evento pode determinar previsões, além de ajudar a minimizar o risco em investimentos. A influência desses fatores pode ter diferentes ciclos de flutuação, afetando uma série temporal ao longo de meses ou anos. Portanto, descobrir esses eventos na série temporal financeira é uma tarefa desafiadora. Essa dissertação apresenta o *Multi-Scale Event Detect (MSED)*, uma técnica para detecção de eventos em séries temporais não estacionárias e não lineares. Somado a isso, este trabalho faz um estudo associativo dos eventos encontrados pelos métodos de detecção nos componentes de *Intrinsic Mode Function (IMF)* com as marcações de eventos externos obtidos através da *Economic Policy Uncertainty (EPU)*. O objetivo é identificar que tipo de evento é refletido por um determinado fator de incerteza externo em uma série financeira, utilizando essa abordagem, é possível determinar a natureza mais predominante dos eventos com base nas variações de incerteza apresentadas na série de EPU. Essas informações permitem especificar um conjunto de séries temporais onde a influência da incerteza gera eventos aceitáveis para um determinado perfil de investimento, mitigando assim, o risco a que se pretende estar exposto.

Palavras-chave:

Economic Policy Uncertainty, multi-scale, Incerteza, Detecção de eventos

Rio de Janeiro,

Março de 2023

ABSTRACT

Detecção e análise multi-scale de eventos originados por fatores externos de incerteza em séries financeiras

Diego Silva de Salles

Advisors:

Eduardo Soares Ogasawara

Eduardo Bezerra da Silva

Abstract of dissertation submitted to Programa de Pós-graduação em Ciência da Computação - Centro Federal de Educação Tecnológica Celso Suckow da Fonseca CEFET/RJ as partial fulfillment of the requirements for the degree of master.

Different external factors reported in the media can impact a financial time series. Such factors, such as government transitions, economic crises, or corruption scandals, can be related to events that increase uncertainty in the time series. In particular, these external factors can increase the risk perceived in a financial time series through events such as anomalies or change points. A study based on the different characteristics that make up an event can determine predictions, in addition to helping to minimize the risk in investments. The influence of these factors can have different cycles of fluctuations, affecting a time series over months or years. Hence, discovering these events in the financial time series is a challenging task. This paper presents Multi-Scale Event Detect (MSED), a technique for detecting events in non-stationary and nonlinear time series. Added to this, this work makes an associative study of the events found by the detection methods in the components of the *Intrinsic Mode Function* (IMF) with the external factors labels obtained through the *Economic Policy Uncertainty* (EPU). The objective is to identify which type of event is reflected by a given external uncertainty factor in a financial series, using this approach it is possible to determine the most predominant nature of the events based on the uncertainty variations presented in the series of (EPU). This information makes it possible to determine a set of time series where the influence of uncertainty generates acceptable events for a given investment profile. Thus, mitigating the risk to which it is intended to be exposed.

Key-words:

Economic Policy Uncertainty, Multi-scaleUncertainty, Event detection

Rio de Janeiro,

Março de 2023

Sumário

I	Introdução	1
II	Referencial teórico	4
II.1	Séries temporais	4
II.2	Técnicas de decomposição de sinal <i>multi-scale</i>	6
II.2.1	Transformada de Fourier	6
II.2.2	Transformada Wavelet	7
II.2.3	<i>Empirical Mode Decomposition</i> (EMD)	8
II.2.4	<i>Ensemble Empirical Mode Decomposition</i> (EEMD)	11
II.2.5	<i>Complementary Ensemble Empirical Mode Decomposition</i> (CEEMD)	12
II.3	Filtro de tendência	14
II.3.1	<i>Energy Approach</i> (EA)	15
II.3.2	<i>Ratio Approach</i> (RA)	15
II.3.3	Componentes Zhang	17
II.4	Índice de Incerteza sobre a Política Econômica	17
II.5	Detecção de eventos	18
II.5.1	Pontos de mudança	18
II.5.2	Anomalias	20
II.5.3	Métricas de avaliação de resultados	22
III	Trabalhos Relacionados	25
IV	Metodologia	28
IV.1	Aquisição de dados	28
IV.2	Decomposição e filtro energético	28
IV.2.1	Componentes obtidos através do filtro de tendencia	29
IV.2.2	<i>Short Term Fluctuation</i> (STF)	29
IV.2.3	<i>Shock Significant Term</i> (SST)	31
IV.2.4	<i>Long Term</i> (LT)	31
IV.2.5	<i>Overall Adaptive Trend</i> (OAT)	32

IV.3 Detecção de eventos em componentes derivados do (MSED)	32
V Resultados	34
V.1 Séries de <i>data label</i>	34
V.1.1 <i>Annotated Approach</i> (AA)	35
V.1.2 <i>Envelop Approach</i> (EVA)	35
V.2 Taxas de câmbio para o real brasileiro	37
V.3 Índices do mercado de ações brasileiro	37
V.4 Índices do mercado de ações chinês	38
V.5 Índices do mercado de ações americano	39
V.6 Grupos de eventos gerados por componente Zhang	41
V.7 Discussões	41
VI Conclusão	43
A Eventos EPU	45
B Grupos de eventos por tipo e presença nos componentes Zhang	48
Referências	50

Lista de Figuras

II.1	Índice médio de ações S&P500	5
II.2	Desenvolvimento das técnicas de decomposição	6
II.3	Transformada de Fourier. Adaptada de [Huang et al., 1998]	7
II.4	Série histórica de fechamento Ibovespa decomposta utilizando (EMD)	9
II.5	Fluxograma do algoritmo <i>Empirical Mode Decomposition</i> (EMD), Adaptado Huang [2014]	10
II.6	Fluxograma do algoritmo (EEMD), Adaptado Zhang Xiaoyuan; Zhou [2013] e Zhao et al. [2011]	13
II.7	Fluxograma do algoritmo <i>Complementary Ensemble Empirical Mode Decomposition</i> (CEEMD), Adaptado Yeh et al. [2010]	14
II.8	<i>Ratio of Zero Crossing Numbers</i> RZCN	16
II.9	<i>Boxplot</i> , adaptado de Han et al. [2012]	23
IV.1	Série histórica de Fechamento Ibovespa	30
IV.2	Série histórica de Fechamento Ibovespa - Componente <i>Short Term Fluctuation</i> (STF)	30
IV.3	Série Histórica de Fechamento Ibovespa - Componente SST	31
IV.4	Série Histórica de Fechamento Ibovespa - Componente LT	32
IV.5	Série histórica de Fechamento Ibovespa - Componente OAT	33
IV.6	Multi-Scale Event Detect MSED	33
V.1	Annotated Approach - Série EPU do Brasil [ser, 2022]	35
V.2	Série EPU do Brasil com envelope superior e inferior. Fonte: Elaborada pelo autor	36
V.3	Série EPU do Brasil com definição de limites baseado no calculo de percentis. Fonte: Elaborada pelo autor	37
A.1	Índice de Incerteza da Política Econômica para a China	45
A.2	Índice de Incerteza da Política Econômica para o Brasil	46
A.3	Índice de Incerteza da Política Econômica para os Estados Unidos da América	47

Lista de Tabelas

IV.1 Séries estudadas	29
V.1 F_1 score para taxas de câmbio	38
V.2 F_1 score para índices médios no mercado de ações brasileiro	38
V.3 F_1 score para índices médios no mercado de ações chinês	39
V.4 F_1 score para índices médios no mercado de ações americano	39
V.5 Média dos valores de precisão e cobertura nos conjuntos de séries analisados	40
B.1 Grupos de eventos no conjunto de séries de taxa de câmbio	48
B.2 Grupos de eventos no conjunto de séries de índices médios no mercado de ações americano	48
B.3 Grupos de eventos no conjunto de séries de índices médios no mercado de ações brasileiro	49
B.4 Grupos de eventos no conjunto de séries de índices médios no mercado de ações americano chinês	49

Lista de Abreviações

AA	<i>Annotated Approach</i>	34, 35, 36, 37, 38, 39, 40, 41, 42
AD	<i>Anomaly Detect</i>	41
CEEMD	<i>Complementary Ensemble Empirical Mode Decomposition</i>	6, 12, 14, 28, 29, 31, 33
CHINA A50	<i>FTSE–Xinhua China A50 Index</i>	28
CSI 1000	<i>Compiled China Securities Index</i>	28
DJ COMPOSITE	<i>Dow Jones Composite Average</i>	28
DJI	<i>Down Jones Industrial Average</i>	28, 39
EA	<i>Energy Approach</i>	15, 16
EEMD	<i>Ensemble Empirical Mode Decomposition</i>	6, 11, 12, 17
EMD	<i>Empirical Mode Decomposition</i>	2, 6, 8, 9, 10, 11, 13, 17, 20, 22, 26, 33
EPU	<i>Economic Policy Uncertainty</i>	1, 3, 18, 34, 35, 36, 37, 41, 42, 43, 44
ERA	<i>Energy Rate Approach</i>	16, 29, 41
EUA	Estados Unidos Da América	34, 36, 41
EVA	<i>Envelop Approach</i>	34, 35, 36, 37, 38, 39, 40, 41
FBIAD	<i>Forward and Backward Inertial Anomaly</i>	20, 21, 36, 37, 38, 39, 40
GARCH	<i>Generalized Autoregressive Conditional Heteroscedasticity</i>	20, 21, 22, 36, 37, 38, 39, 40
HME	Hipótese Do Mercado Eficiente	1, 17
HSCE	<i>Hang Seng China Enterprises Index</i>	28
IBOVESPA	Índice Bovespa	28, 29
IBRA	Índice Brasil Amplo	28
IBRX100	Índice Brasil 100	28
IBRX50	Índice Brasil 50	28
ICSS	<i>Iterative Cumulative Sum of Squares</i>	26, 27, 33
IGC	Índice De Ações Com Governança Corporativa Diferenciada	28, 37
IMF	<i>Intrinsic Mode Function</i>	8, 9, 10, 11, 12, 13, 14, 15, 16, 22, 26, 28, 29, 30, 31, 33
LT	<i>Long Term</i>	17, 29, 31, 32, 33
LTCP	<i>Long Term Change Points</i>	41

MSED	<i>Multi-Scale Event Detect</i>	3, 32, 33, 36, 37, 38, 39, 40, 41, 43, 44
NASDAQ	<i>National Association of Securities Dealers Automated Quotations</i>	28, 39
OAT	<i>Overall Adaptive Trend</i>	29, 32
RA	<i>Ratio Approach</i>	15, 16
RZCN	<i>Ratio of Zero Crossing Numbers</i>	15, 16
SCP	<i>Seminal Change Point</i>	19, 33, 36, 37, 38
SFVE	<i>Short Fluctuation Variance Events</i>	41
SHANGHAI	<i>Shanghai Composite Index</i>	28
SMALL CAP 2000	<i>Small Cap Stock Market Index</i>	28
SP500	<i>Standard and Poor's 500</i>	28, 39
SSCP	<i>Shock Significant Change Points</i>	41
SST	<i>Shock Significant Term</i>	17, 29, 31, 33
SSVE	<i>Shock Significant Variance Event</i>	41
STF	<i>Short Term Fluctuation</i>	17, 22, 29, 30, 31, 32
SZSE COMPONENT	<i>Shenzhen Stock Exchange Component</i>	28
TF	<i>Transformada De Fourier</i>	6, 7
TW	<i>Transformada De Wavelet</i>	7, 8, 27
UIH	<i>Uncertain Information Hypothesis</i>	1, 17

Capítulo I Introdução

As séries temporais financeiras podem ser afetadas por razões complexas não facilmente observáveis [Zhang et al., 2008]. Por exemplo, as taxas de câmbio e os preços das ações se relacionam com as mudanças econômicas. Suas flutuações são acompanhadas de perto por formuladores de políticas e participantes do mercado financeiro [Malik, 2021]. A ocorrência de eventos externos como transições governamentais, crises econômicas e escândalos de corrupção podem ter diferentes formas de impacto na economia de um país. Esses eventos externos mostram a importância de ter um mecanismo que ajude a entender as séries temporais financeiras [Poon and Granger, 2003]. São considerados fatores extrínsecos os eventos externos que podem interferir nas séries temporais financeiras, como o processo de formação de preços [Wen et al., 2017].

Uma forma relevante de avaliar tais fatores extrínsecos é mapeando a incerteza político-econômica. O EPU é um índice criado a partir de notícias com termos de incertezas vinculadas em jornais de grande circulação [Samuelson, 2013]. O EPU visa captar a preocupação com a incerteza político-econômica destacada pela mídia [Samuelson, 2013]. Ao estudar a série temporal de EPU podemos entender a relação entre as variações de incerteza e os eventos detectados em uma série financeira, definindo assim, uma série temporal de *data labels* que representa os eventos externos refletidos através das variações de incerteza presentes na série de EPU. A série de *data label* criada, demonstra pontos na série temporal financeira onde é mais provável de encontrar eventos com os métodos de detecção utilizados.

Uma das teorias mais importantes em finanças é a Hipótese do Mercado Eficiente (HME). A HME afirma que os preços dos ativos resultam da informação disponível aos investidores. À medida que novas informações vão surgindo, os investidores tentam antecipar o movimento das ações, sendo o preço das ações uma consequência disso [Malkiel, 2005]. A incerteza é um conceito associado à incapacidade de prever um comportamento previamente conhecido à medida que o mesmo se repita ao longo do tempo [Samuelson, 2013]. A presença de incerteza está associada ao aumento do risco. Para os agentes financeiros, é importante ter confiança de que um evento conhecido do passado possa se perpetuar no futuro, e dessa forma, maximizar seus ganhos em investimentos futuros.

Além da HME, a *Uncertain Information Hypothesis* (UIH) afirma que os preços das ações se ajustam quase imediatamente à medida que novas informações chegam aos agentes financeiros. Mesmo que os investidores não tenham uma visão clara do impacto total desses eventos externos,

os preços das ações refletirão essas incertezas [Corrado and Jordan, 1997]. Esse comportamento faz com que os investidores tenham medo do aumento do risco inerente ao seu investimento, e acabam por negociar suas ações abaixo de um valor fundamental. A reação dos investidores é a aversão ao risco. Por isso, informações desfavoráveis tendem a apresentar uma reação contraproducente ao evento de incerteza. Por outro lado, informações positivas mostram uma lenta mudança nos preços dos ativos. A incerteza aumenta o risco dos ativos e diminui a expectativa de retorno. Se esse risco diminuir, os preços dos ativos tendem a ser reajustados.

Além disso, não só apenas eventos externos, como crises políticas, podem gerar mudanças no comportamento padrão das séries temporais financeiras, existem diversos outros tipos de eventos externos que podem impactar o processo de formação dos dados presentes em uma série. A presença desses eventos externos influencia o processo de formação de uma série temporal financeira, sendo que seus efeitos podem ser observados em período de semanas, meses ou anos [Aloui, 2007]. Por isso, há a necessidade de criar um mecanismo que possa detectar as informações geradas por estes eventos externos em diferentes intervalos de tempo [Wen et al., 2017].

A análise *multi-scale* é uma família de técnicas que permitem a decomposição de uma série temporal em componentes de diferentes ciclos, permitindo um maior entendimento dos eventos responsáveis pelo processo de formação de preços encontrados nas séries temporais financeiras [Leung and Zhao, 2021b]. Um algoritmo de decomposição eficaz pode converter uma série temporal original em componentes de diferentes ciclos [Wen et al., 2017; Huang et al., 1998]. Essa característica aumenta a capacidade de analisar informações locais da série temporal sob esses componentes [Leung and Zhao, 2021a]. A transformada de Fourier foi uma das primeiras técnicas de decomposição de sinal existentes [Jun et al., 2018]. No entanto, esta técnica tem algumas limitações. Seu uso é restrito a séries temporais lineares e estacionárias [Huang et al., 2003]. Uma das técnicas de decomposição mais conhecidas é a *Wavelet*, que tem várias vantagens ao lidar com multi-escala [Huang et al., 1998]. No entanto, este método pode gerar muitos espúrios harmônicos. Além disso, as escalas de decomposição não são ajustáveis, o que torna necessário definir a priori o tamanho da *wavelet* base. Mais recentemente, Huang et al. [1998] propuseram a *Empirical Mode Decomposition* EMD, que pode extrair funções intrínsecas derivadas de um sinal original conservando a escala de tempo e mantendo as características locais. O método EMD permite a análise de partes locais da série temporal enquanto fornece uma decomposição adaptativa de séries temporais não estacionárias [Salles et al., 2019].

Com base no que foi informado anteriormente, surge uma pergunta natural: técnicas de detecção de eventos podem ser aplicadas a componentes com diferentes flutuações, derivadas de uma decomposição de séries temporais financeiras, para mapear a incerteza de eventos externos?

Neste trabalho, buscamos identificar essa associação desenvolvendo um método chamado *Multi-*

*Scale Event Detect*MSED. Este método decompõe uma série temporal financeira em componentes de diferentes ciclos. Os componentes, uma vez construídos, são alvos de métodos de detecção de eventos que procuram anomalias e pontos de mudança nas séries temporais estudadas. Este trabalho tem duas contribuições principais:

1. Uma nova técnica de Detecção de Eventos MSED para aplicação em séries temporais financeiras;
2. Estudo associativo dos eventos encontrados pelos métodos de detecção com as variações de incerteza obtidas através do índice EPU.

Além da introdução, o texto desta dissertação é composto por cinco capítulos. No capítulo II é apresentada uma visão geral sobre séries temporais e os desafios decorrentes na análise de séries financeiras. Neste capítulo também é demonstrado a evolução das diferentes técnicas de decomposição de séries temporais, além de apresentar a importância do uso de filtro de tendência para análise *multi-scale* das séries. O conceito de incerteza também é apresentado no capítulo II em conjunto do processo de criação do índice EPU, e seu papel neste trabalho. O capítulo III apresenta um comparativo dos trabalhos relacionados com essa dissertação, além de explicar o método de pesquisa do tema apresentado. A apresentação da metodologia se encontra no capítulo IV, o capítulo V apresenta os resultados empíricos e o capítulo VI as considerações finais.

Capítulo II Referencial teórico

II.1 Séries temporais

Existem algumas abordagens na análise de dados onde os dados adjacentes têm uma alta correlação no tempo. Algumas soluções para esses casos podem limitar a aplicabilidade das técnicas estatísticas convencionais. Isso se deve ao fato de que os métodos mais tradicionais necessitam que os dados sejam independentes e identicamente distribuídos [Nava et al., 2018]. Nesses casos, existe uma forte dependência dos dados adjacentes entre o valor atual e as observações anteriores [Shumway and Stoffer, 2017]. A abordagem matemática amplamente utilizada para esses problemas é chamada de série temporal. Formalmente, uma série temporal $X(t)$ pode ser descrita como uma sequência de observações $\langle x_1, x_2, \dots, x_n \rangle$, tal que x_n é a observação mais recente e $|X(t)| = n$, que é o tamanho da série temporal [Shumway and Stoffer, 2017]. Essa abordagem se concentra em modelar algum valor futuro como uma função paramétrica dos valores atuais e passados [Nava et al., 2018].

As séries temporais podem ser estacionárias ou não estacionárias. Séries temporais estacionárias têm regularidade no comportamento ao longo do tempo [Shumway and Stoffer, 2017]. Uma série temporal estacionária tem média, variância e covariância constantes. Caso contrário, uma série temporal é definida como não estacionária quando viola as restrições sobre a existência de estacionaridade [Salles et al., 2019]. No entanto, é comum supor que a série temporal é estacionária ao analisá-la. Se isso não for feito, a análise da série temporal fica limitada [Nava et al., 2018]. No cenário econômico, isso tem um impacto relevante [Huang et al., 2003]. Suponha que a série temporal analisada não seja estacionária. Nesse caso, as análises podem negligenciar informações sobre mudanças reais no mercado financeiro perdendo a capacidade de olhar para as características locais presentes em uma série temporal, tornando imprecisa a análise das propriedades dos dados [Huang et al., 2003]. Além disso, isso é agravado porque as séries temporais do mercado financeiro são de natureza intrinsecamente não estacionária [Leung and Zhao, 2021b].

Séries temporais podem auxiliar análises científicas de diferentes formas, sendo utilizadas nas áreas mais diversas. Existem diferentes tipos de séries temporais que podem ser encontrados na economia. Observações de mercado e estudos empíricos demonstram que séries financeiras como índices de mercado, e preços de ativos, são frequentemente dirigidos por fatores *multi-scale* [Wen

et al., 2017] [Zhang et al., 2008]. Essas séries possuem variações com ciclos econômicos de longo prazo e flutuações mais rápidas de curto prazo. Tal comportamento indica que séries financeiras são potencialmente carregadas de informações em diferentes escalas de tempo [Leung and Zhao, 2021b] [Nava et al., 2018]. A representação gráfica de uma série temporal financeira pode ser visualizada na Figura II.1

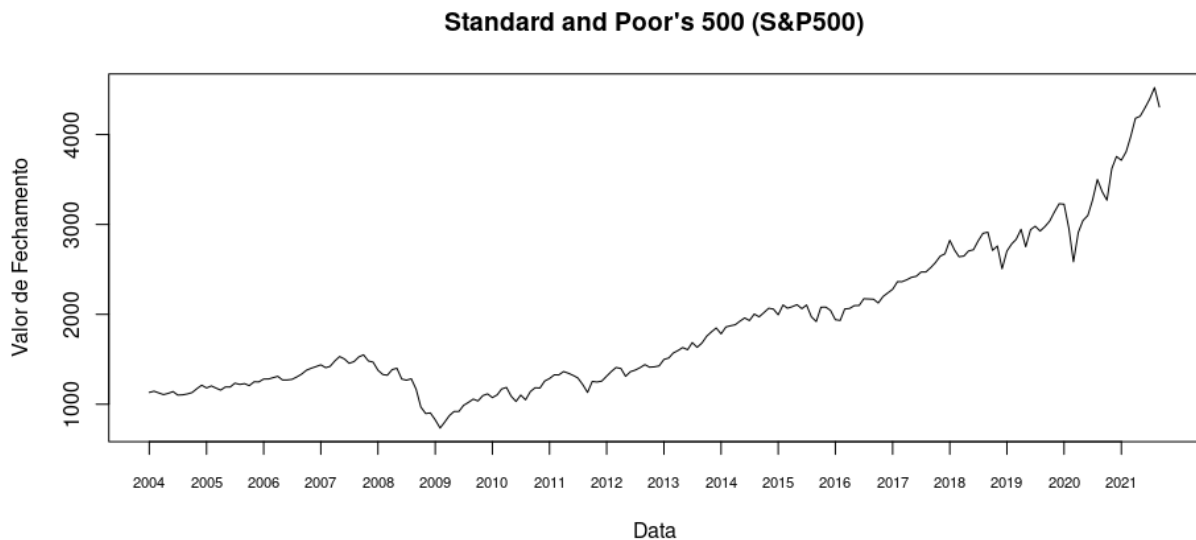


Figura II.1: Índice médio de ações S&P500

Componentes de uma série temporal

A decomposição é uma técnica importante na análise de séries temporais, permitindo a separação das séries temporais em um conjunto de componentes que não são facilmente observáveis. Esses componentes estão associados a diferentes variações temporais e são comumente classificados em três tipos de componentes: *(i)* tendência; *(ii)* sazonalidade e *(iii)* variações residuais [Hyndman and Athanasopoulos, 2018]. A Equação II.1 demonstra a fórmula geral da decomposição, onde: $S(t)$ representa o componente de sazonalidade, $T(t)$ o componente de tendência e $R(t)$ as variações residuais.

$$X(t) = S(t) + T(t) + R(t) \quad (\text{II.1})$$

A tendência é uma função monotônica ou uma função na qual pode haver no máximo um extremo dentro de um determinado período de dados [Moghtaderi et al., 2011]. A tendência é tomada como uma disposição do conjunto de dados da série temporal, que permanecerá em observações futuras [Flandrin et al., 2004a]. Muitas das dificuldades relacionadas à obtenção de valores de tendência surgem do uso em séries não estacionárias e não lineares. Existem vários processos para

determinar a tendência, os mais conhecidos são o uso de regressão linear determinando uma linha média dos dados, e o uso de média móvel em uma janela pré-determinada da série [Moghtaderi et al., 2010]. No entanto, Wu et al. [2007] relatam que esses algoritmos de remoção de tendência quando usados em séries não estacionárias e não lineares requerem parâmetros ou funções extrínsecas à série. Para auxiliar no processo de análise de tendências, uma abordagem para obter e remover uma tendência usando (EMD) foi proposta na Seção II.3

II.2 Técnicas de decomposição de sinal *multi-scale*

Na Figura II.2 é apresentada a evolução das técnicas de decomposição mais utilizadas em séries temporais. Nesta seção são abordados os principais métodos de decomposição, começando por Fourier [Huang et al., 1998]. Fourier tem sido uma das técnicas mais amplamente utilizadas para decomposição em séries temporais estacionárias, entretanto tem extrema dificuldade ao lidar com séries financeiras não estacionárias. Após Fourier, o trabalho apresenta Wavelet. Essa técnica consegue lidar com o problema da não estacionaridade, mas necessita de uma configuração prévia de parâmetros extrínsecos a série, mesmo com essa configuração extra é uma das técnicas mais utilizadas atualmente. Por fim, são apresentados EMD e os métodos derivados como: *Ensemble Empirical Mode Decomposition* (EEMD) e *Complementary Ensemble Empirical Mode Decomposition* CEEMD. Este conjunto de métodos consegue lidar com a não estacionaridade utilizando apenas parâmetros intrínsecos à série.



Figura II.2: Desenvolvimento das técnicas de decomposição

II.2.1 Transformada de Fourier

A Transformada de Fourier (TF) é uma das primeiras técnicas de decomposição de sinal existentes. Um de seus usos mais comuns é na análise de distribuição de domínio/frequência [Huang et al., 1998; Jun et al., 2018]. Esta técnica de decomposição tem sido útil para as mais variadas séries temporais, sendo aplicada nos mais diferentes domínios de conhecimento. No entanto, existem algumas limitações no uso geral de Fourier. A TF é restrita a séries temporais lineares, e os dados devem ser estacionários [Huang et al., 2003].

A TF cria uma representação baseada em frequência de uma série temporal em termos de

funções de base de Fourier. A TF de uma série temporal X , neste caso representada em função do tempo $X(t)$, é formulada na Equação II.2, onde $F(\xi)$ representa o espectro de Fourier, ξ é um componente de frequência e j é o número imaginário $j = \sqrt{-1}$ [Los, 2006].

$$F(\xi) = \int_{-\infty}^{+\infty} X(t)e^{-j\xi t} dt, \quad e^{-j\xi t} = \cos \xi t - j \sin \xi t \quad (\text{II.2})$$

Embora Fourier seja uma técnica amplamente utilizada, ela apresenta algumas limitações ao lidar com séries temporais financeiras. É importante notar que os componentes de Fourier são harmônicos em todo o espectro analisado da série temporal. Devido a isso, ao usar Fourier, é necessário um grande esforço para criar representatividade de componentes harmônicos para replicar as séries temporais não estacionárias [Huang et al., 1998, 1996]. Ao fazer isso, tenta se criar uma representação forçada da série em senos e cossenos, e quando a série não consegue ser representada por essas funções, a decomposição resultante da série conterà muitos espúrios harmônicos [Huang et al., 1996]. Este efeito faz com que a série original analisada perca o sentido local da informação, tendo pouca representatividade da informação em um domínio dos dados que se procura estudar [Wen et al., 2017]. A representação gráfica da transformada de Fourier pode ser visualizada na Figura II.3.

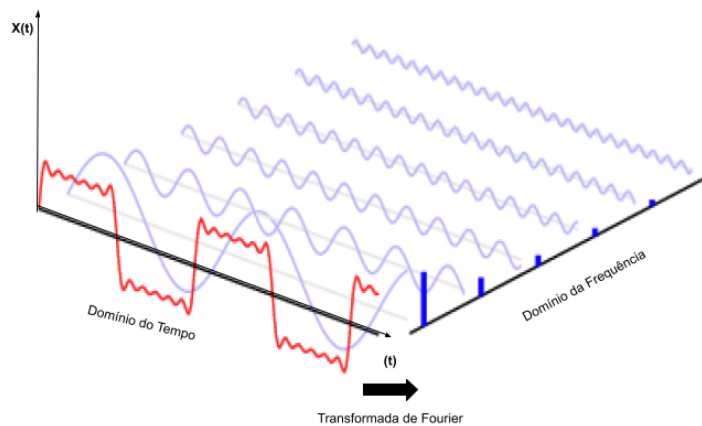


Figura II.3: Transformada de Fourier. Adaptada de [Huang et al., 1998]

II.2.2 Transformada Wavelet

A Transformada de Wavelet (TW) foi criada na década de 80 para análise de dados *multi-scale* [Mallat, 2008]. Uma das principais vantagens do wavelet sobre Fourier é sua capacidade de

capturar informações de frequência e localização [Jun et al., 2018]. A TW tem se apresentado ao longo dos anos como uma técnica capaz de analisar dados com mudanças graduais de frequência, sendo amplamente utilizada em séries temporais não estacionárias.

As *wavelets* são funções de base finita localizadas tanto no tempo quanto na frequência. A TW decompõe uma série temporal (sinal) correlacionando-a com uma família de *wavelets*, fornecendo uma representação tempo-frequência extremamente flexível [Los, 2006; Salles et al., 2019; Lahmiri, 2016]. A TW decompõe uma série temporal X , considerada como uma função do tempo $X(t)$, na série *wavelet* $\hat{X}(t)$ que pode ser vista na Equação II.3 [Joo and Kim, 2015; Salles et al., 2019]. Os componentes $\zeta(t)$ e o coeficiente b são responsáveis por modelar a tendência e a sazonalidade, representando parte da escala da série *wavelet*, enquanto os componentes $\psi(t)$ e seu coeficiente c correspondem ao ruído ou desvios aleatórios, na escala (nível de decomposição) l e posição k . Além disso, L é a definição de nível máximo da decomposição possível na série [Salles et al., 2019].

$$\hat{x}(t) = \sum_{k=1}^n b_{l,k} \zeta_{l,k}(t) + \sum_{l=1}^L \sum_{k=1}^n c_{l,k} \psi_{l,k}(t) \quad (\text{II.3})$$

Por mais que seu uso tenha aumentado nos últimos anos, um dos problemas comuns se encontra no vazamento de informações devido ao comprimento limitado do componente *wavelet* básico [Mallat, 2008; Huang et al., 1998]. Esta característica dificulta a definição do valor a ser distribuído quantitativamente na distribuição energia-frequência-tempo [Mallat, 2008; Huang et al., 1998]. Outra questão é que a interpretação *wavelet* não é intuitiva. Para definir uma mudança que ocorre localmente, deve-se procurar o resultado na faixa de alta frequência, quanto maior a frequência, maior é a *wavelet* básica. Se o evento ocorrer apenas na faixa de baixa frequência, ainda é necessário procurar os efeitos na faixa de alta frequência [Huang et al., 1998; Mallat, 2008]. Outra questão é a natureza não adaptativa da *wavelet*, uma vez que uma *wavelet* básica é selecionada, é necessário usá-la para todo o conjunto de dados [Bai et al., 2016]. Apesar de apresentar alguns problemas que precisam ser tratados ao se utilizar TW, essa técnica ainda é um dos métodos de análise de dados não estacionários mais utilizados.

II.2.3 *Empirical Mode Decomposition* (EMD)

A decomposição de uma série temporal pode determinar os componentes originais em diferentes ciclos como: tendência, variações cíclicas de longo prazo, sazonalidade e variáveis aleatórios [Shumway and Stoffer, 2017]. Huang et al. [1998] propuseram um método adaptativo para análise de séries temporais não estacionárias e não lineares o EMD. Entretanto, para fazer a decomposição dos dados obtendo componentes de oscilação IMFs, a série precisa obedecer os seguintes critérios: (i) A série deve ter pelo menos dois extremos, um máximo e um mínimo ; (ii) A escala de tempo característica

é definida pelo lapso de tempo entre os extremos; *(iii)* Se os dados são totalmente desprovidos de extremos, mas contêm apenas pontos de inflexão, eles podem ser diferenciados uma ou mais vezes para obter os extremos [Huang et al., 1998].

Intrinsic Mode Function - A IMF é um componente do processo de decomposição de modo empírico proposto por Huang et al. [1998]. A IMF representa o modo de oscilação de uma série temporal e possui as seguintes condições de existência: *(i)* Em todo o conjunto de dados o número de extremos e cruzamentos de zero deve ser igual ou diferir em pelo menos um; *(ii)* Em qualquer ponto da série temporal o valor médio do envelope definido pelos máximos e mínimos locais é igual a zero.

A Figura II.4 representa a série de histórica de fechamento da Ibovespa, com os componentes derivados da decomposição utilizando EMD.

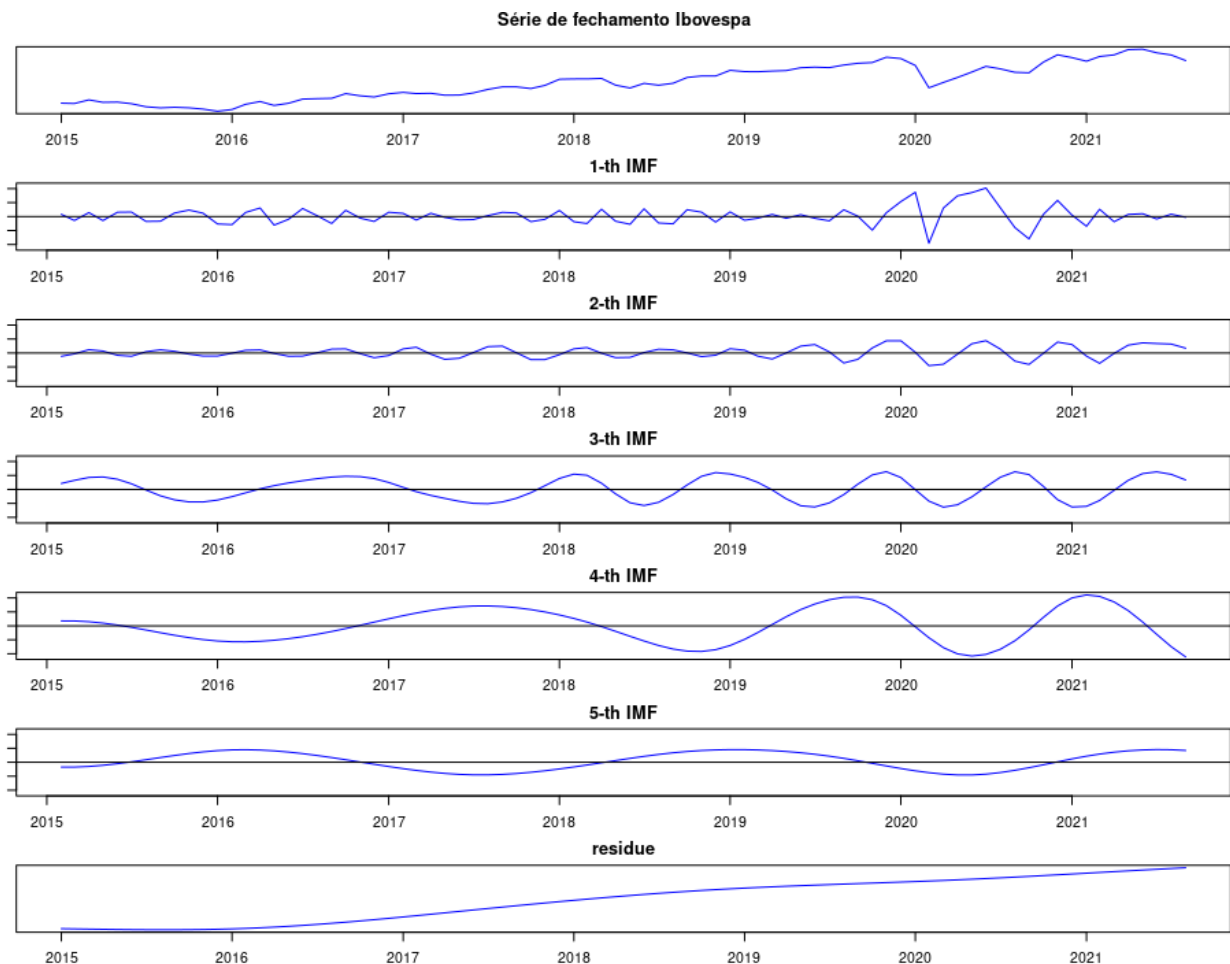


Figura II.4: Série histórica de fechamento Ibovespa decomposta utilizando (EMD)

Algoritmo EMD - Na Figura II.5 é demonstrado o diagrama de atividades do método EMD. O método EMD segue as seguintes etapas: *(i)* As variáveis são inicializadas $r(t) = X(t)$, $k = 0$;

(ii) São identificados todos os máximos e mínimos de séries temporais $h(t)$; (iii) São gerados os envelopes superior e inferior, $e_{max}(t)$; $e_{min}(t)$. (iv) Nesta etapa é feito o cálculo da média ponto a ponto dos envelopes superior e inferior: $m(t) = (e_{min}(t) + e_{max}(t))/2$; (v) Nesta etapa, a média é subtraída do sinal original $h(t) = h(t) - m(t)$; É necessário verificar se o novo valor de $h(t)$ é uma IMF (ver critérios de existência de uma IMF na Seção II.2.3); Se for um IMF, $IMF_{k+1}=h(t)$, onde k representa o número de componentes do IMF que podem ser encontrados em uma série temporal; (vi) Este IMF_k é subtraído da série temporal original $X(t)$, $X(t) = X(t) - IMF_k$. Este componente residual é tratado como novos dados $r(t) = X(t)$ e $h(t) = X(t)$, passando pelo processo descrito acima para calcular o próximo $IMF_{k+1}(t)$; (vii) Essas etapas são repetidas até que o componente residual final $h(t)$ não possa mais atender ao critério definido para que uma série seja decomposta utilizando o EMD, o valor final do resíduo encontrado é igual a $r(t) = h(t)$.

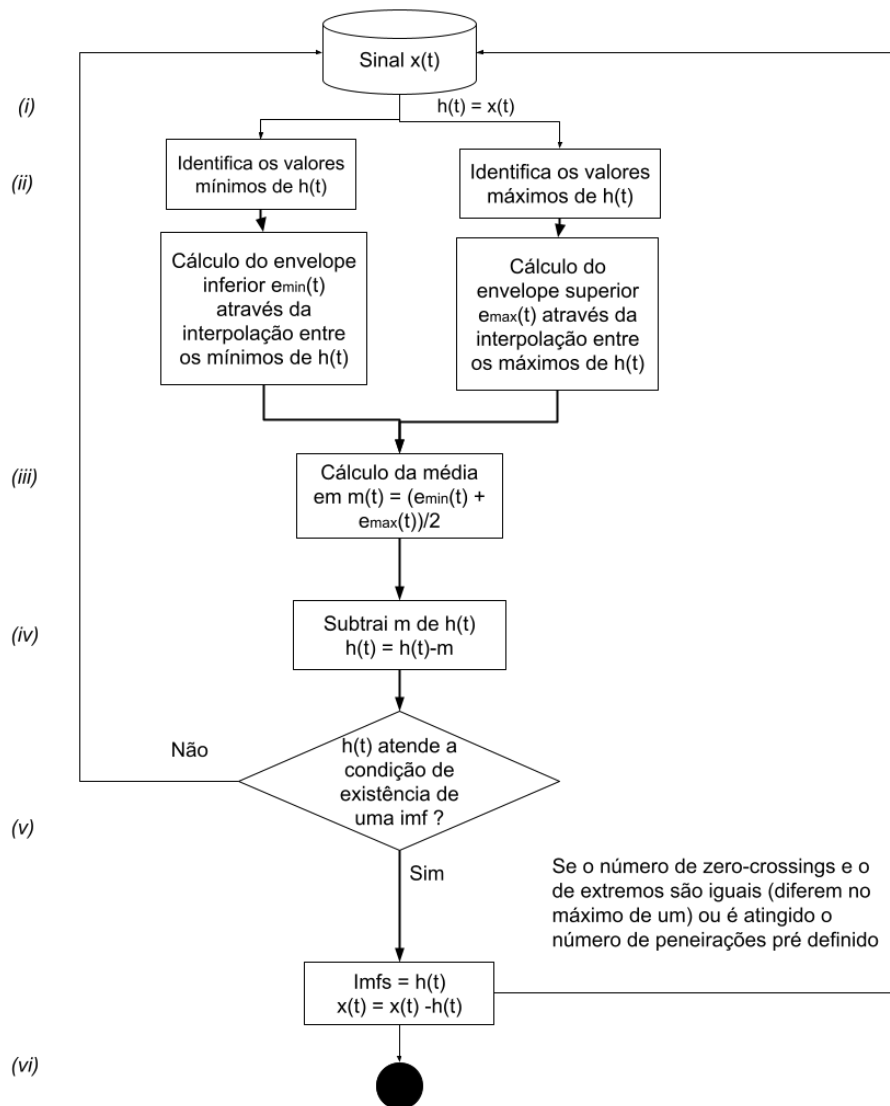


Figura II.5: Fluxograma do algoritmo EMD, Adaptado Huang [2014]

A Equação II.8 demonstra a função EMD: onde n é o enésimo componente do IMF, $c(t)$ representa cada componente de IMF, $r(t)$ é o resíduo e j o índice do componente do IMF.

$$X(t) = \sum_{j=1}^n c_j(t) + r(t) \quad (\text{II.4})$$

II.2.4 *Ensemble Empirical Mode Decomposition (EEMD)*

O algoritmo EMD tem uma desvantagem devido ao problema de *mode mixing* de frequência. O problema de *mode mixing* é definido por sinais com escalas semelhantes que residem em diferentes componentes de IMF [Wu and Huang, 2009]. No trabalho de Huang et al. [2003] é demonstrado o problema de *mode mixing* como um problema de sinal intermitente, essa intermitência pode não apenas causar sérios *aliasing* na distribuição tempo-frequência, mas tirar a clareza de significado estatístico individual das IMFs [Wu and Huang, 2009]. O Wu and Huang [2009] propuseram um método eficaz de *noise-assisted*, o EEMD. Ele reduz significativamente a chance do aparecimento de *mode mixing*, preservando a propriedade de ser um *adaptive dyadic filter* para qualquer conjunto de dados [Flandrin et al., 2004b; Wu and Huang, 2009]. O EEMD é uma versão aperfeiçoada do método EMD, sendo uma ferramenta mais aprimorada que o método EMD para o uso em processamento de sinais não lineares e não estacionários. O princípio do EEMD) é simples: o ruído branco adicionado preenche todo o espaço tempo-frequência uniformemente, facilitando uma separação natural das escalas de frequência, o que reduz a ocorrência do *mode mixing* [Flandrin et al., 2004b; Wu and Huang, 2009; Zhang Xiaoyuan; Zhou, 2013; Zhao et al., 2011].

Algoritmo EEMD - Na figura II.6 é demonstrado o diagrama de atividades do método EEMD. O método EEMD segue as seguintes etapas:

1. É definido o número do *ensemble* M ;
2. É estabelecido a amplitude do ruído branco gerado, onde $i = 1$;
3. O ruído branco com amplitude i é criado, $n_i(t)$;
4. Nessa etapa o ruído branco gerado $n_i(t)$ é somado ao sinal original $X(t)$, gerando um novo sinal:

$$X_i(t) = X(t) + n_i(t) \quad (\text{II.5})$$

Onde $n_i(t)$ representa a i -ésima série de ruído branco, e $X_i(t)$ representa o ruído adicionado à i -ésima iteração, quanto o $i = 1, 2, \dots, M$;

5. O algoritmo EMD é usado para decompor o sinal recém-gerado $X_i(t)$ em IMFs;

$$X_i(t) = \sum_{j=1}^n c_{i,j}(t) + r_{i,S}(t) \quad (\text{II.6})$$

Onde S é o número de IMFs, $r_{i,S}(t)$ é o resíduo final, e $c_{i,S}(t)$ representa os componentes de IMFs ($c_{i,1}, c_{i,2}, \dots, c_{i,S}$) que inclui diferentes bandas de frequência;

6. As etapas três e quatro são repetidas M vezes com uma série de ruído branco diferente a cada iteração. Obtendo assim, um *ensemble* de IMFs. Onde $c_S(t)$ é a s -ésima IMF decomposta pela (EMD), onde $i = 1, 2, \dots, M$, e $s = 1, 2, \dots, S$.

$$[c_{1,S}(t), c_{1,S}(t), \dots, c_{M,S}(t)] \quad (\text{II.7})$$

7. As médias dos *ensembles* das IMFs correspondentes da decomposição (EMD) são calculadas

$$c_S(t) = \frac{1}{M} \sum_{i=1}^M c_{i,S}(t) \quad (\text{II.8})$$

II.2.5 Complementary Ensemble Empirical Mode Decomposition (CEEMD)

Se os componentes produzidos usando EEMD contêm ruído residual, muitos testes são necessários para eliminar esse excesso de ruído, retardando o processo e aumentando o custo computacional [Yeh et al., 2010]. O EEMD resolve o problema de *model mixing* adicionando ruído excedente nos componentes IMF. Para remover esse excesso de ruído é necessário um processo de filtragem, o que aumenta o custo computacional do método [Wen et al., 2017]. Yeh et al. [2010] criaram uma nova abordagem chamada CEEMD, este método mantém a capacidade de decomposição do método EEMD enquanto compensa o uso de ruído branco adicional. O CEEMD pode eliminar o ruído residual usando pares de ruído branco (positivo e negativo), usando menos testes de ruído do que o método EEMD [Chen et al., 2018] [Yeh et al., 2010].

No CEEMD o ruído branco é próximo de zero, ao contrário de EEMD que retém muito ruído adicional. Devido a isso, foi utilizado o CEEMD como técnica de decomposição para as séries temporais utilizadas neste trabalho.

Algoritmo CEEMD - Na Figura II.7 é demonstrado o diagrama de atividades do método CEEMD. O método CEEMD segue as seguintes etapas:

1. O ruído branco com amplitude i é criado em pares, Onde $n_i^+(t)$ representa o ruído positivo, $n_i^-(t)$ o negativo e $x(t)$ é o sinal original;

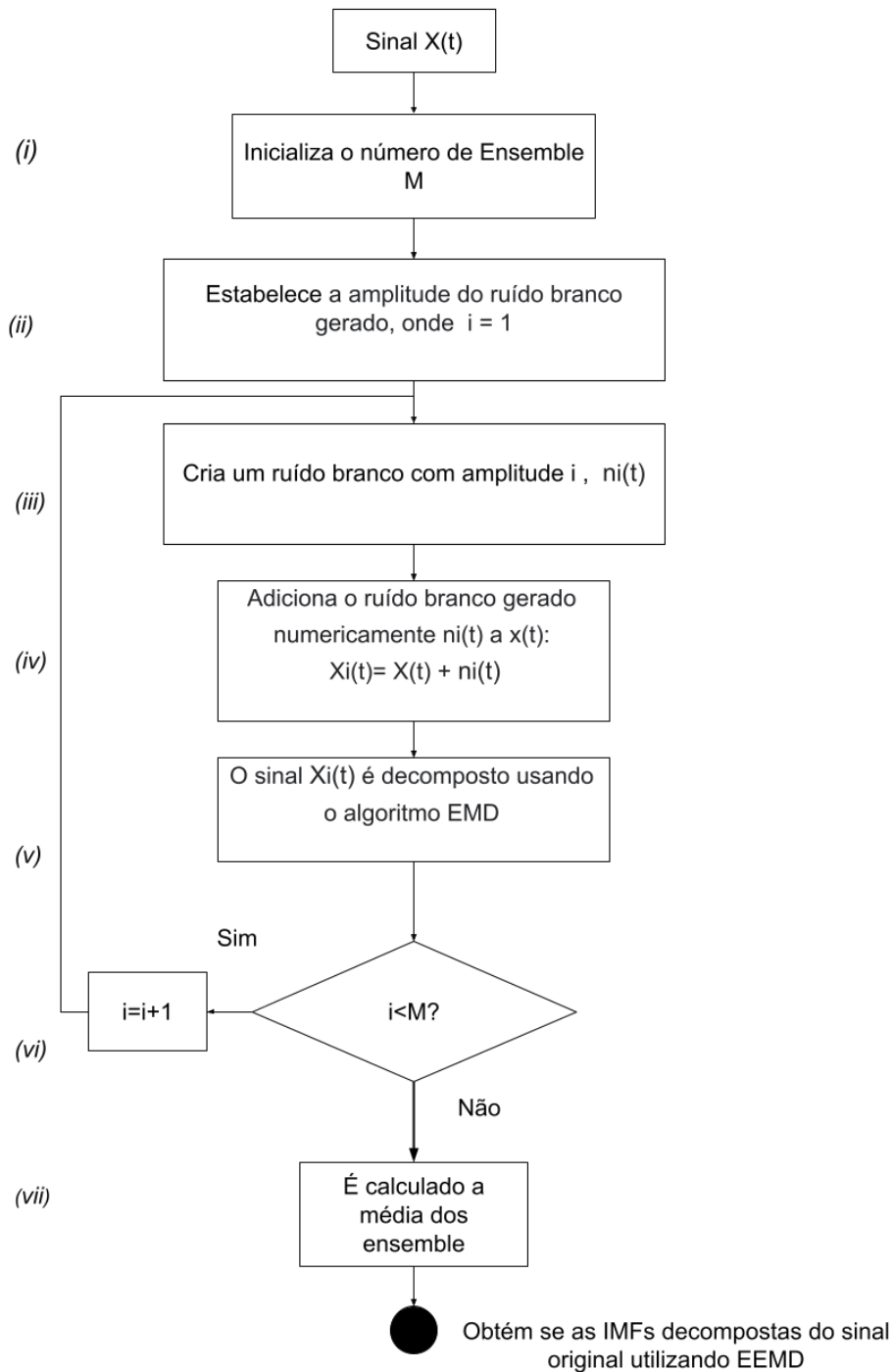


Figura II.6: Fluxograma do algoritmo (EEMD), Adaptado Zhang Xiaoyuan; Zhou [2013] e Zhao et al. [2011]

2. O ruído branco é adicionado em pares a partir do sinal original: $m_i^+(t) = x(t) + n_i^+(t)$; $m_i^-(t) = x(t) + n_i^-(t)$;
3. Decompõe $m_i^+(t)$ e $m_i^-(t)$ usando o método EMD obtendo dois grupos de IMF, onde IMF_{max} representa a decomposição em $m_i^+(t)$ e IMF_{min} a representação em $m_i^-(t)$

4. (iii) O resultado final é o *ensemble* do IMF positivo e negativo do componente IMF_m .

$$[IMF_m = \frac{1}{2}(IMF_{min} + IMF_{max})] \quad (II.9)$$

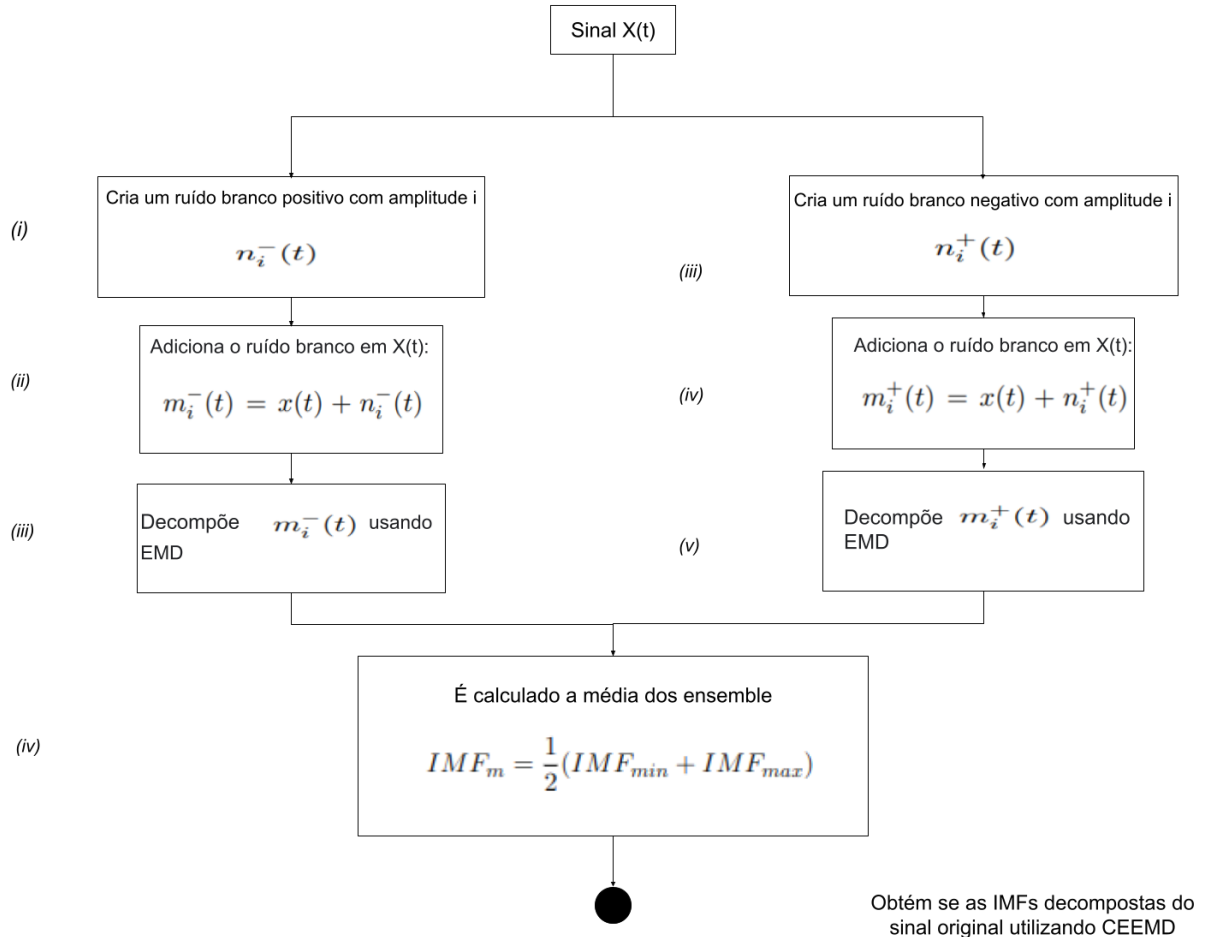


Figura II.7: Fluxograma do algoritmo CEEMD, Adaptado Yeh et al. [2010]

II.3 Filtro de tendência

O componente residual resultante da utilização do método CEEMD possui grande parte das informações de tendência presentes em uma série temporal, mas a presença de tendência não é uma condição de exclusividade deste componente. As IMFs com baixa flutuação em seus dados também podem possuir tendências de médio prazo. Devido a esta característica, para algumas análises é necessário criar um mecanismo que possa segregar os componentes do IMFs que possuem tendência de médio prazo dos que possuem informações de flutuação de curto prazo. Com esse objetivo Moghtaderi et al. [2011] propôs um filtro de tendência que utiliza duas abordagens: a energética e a de taxa de cruzamento de zeros.

II.3.1 *Energy Approach (EA)*

Sendo $X(t)_{t>0}$ uma determinada série temporal, pode se definir a quantidade de energia da IMF como G^i , assim:

1. M representa o conjunto de componentes de IMF presentes em uma série temporal.

$$M = \sum_{j=1}^n c_j(t) \quad (\text{II.10})$$

2. G^i representa o i -ésimo termo de um componente G , j é cada valor contido em um componente IMF e M^i é o i -ésimo IMF resultante da decomposição. Então, o valor absoluto da soma dos dados de um componente do IMF é elevado à segunda potência, obtendo-se assim o G^i de M^i . Onde i^* é o menor valor onde $i \geq 2$ e $G^i > G^{i-1}$. Essa abordagem é chamada de abordagem energética.

$$G^i = \sum_{j=0}^n |M_j^i|^2, 1 \leq i \leq i_m \quad (\text{II.11})$$

A limitação da técnica *Energy Approach (EA)* é que ela usa apenas um G^i para identificar o i^* usado no filtro de tendência. O problema de usar a EA é que como o filtro de tendência é baseado em apenas uma única série temporal, G^i pode obter valores menores que contêm informações não associadas à tendência.

$$T(t) = \sum_{j=i^*}^n c_j(t) + r(t), i \geq 2 \quad (\text{II.12})$$

II.3.2 *Ratio Approach (RA)*

Uma segunda abordagem a ser considerada para obter o componente de tendência é a abordagem de taxa de cruzamento de zeros *Ratio Approach (RA)*. Esta abordagem baseia-se no cálculo de cruzamentos zeros nos componentes de IMF, observa-se que o número de cruzamentos zero de um i -ésimo IMF é igual a Z^i . Então o i -ésimo *Ratio of Zero Crossing Numbers (RZCN)* é definido:

$$R^{(i+1)} = \frac{Z^{(i)}}{Z^{(i+1)}} \quad (\text{II.13})$$

O valor resultante de $R^{(i+1)}$ deve ser maior que dois. No artigo de Moghtaderi et al. [2010], 13 séries temporais de banda larga sem oscilações de baixa frequência foram testadas para identificar o limiar do componente de tendência i^* . O exemplo escolhido neste trabalho é baseado no uso

descrito no trabalho de Moghtaderi et al. [2010]. No entanto, isso não impede que os mesmos princípios sejam usados para obter i^* com a técnica RA em qualquer outra série temporal.

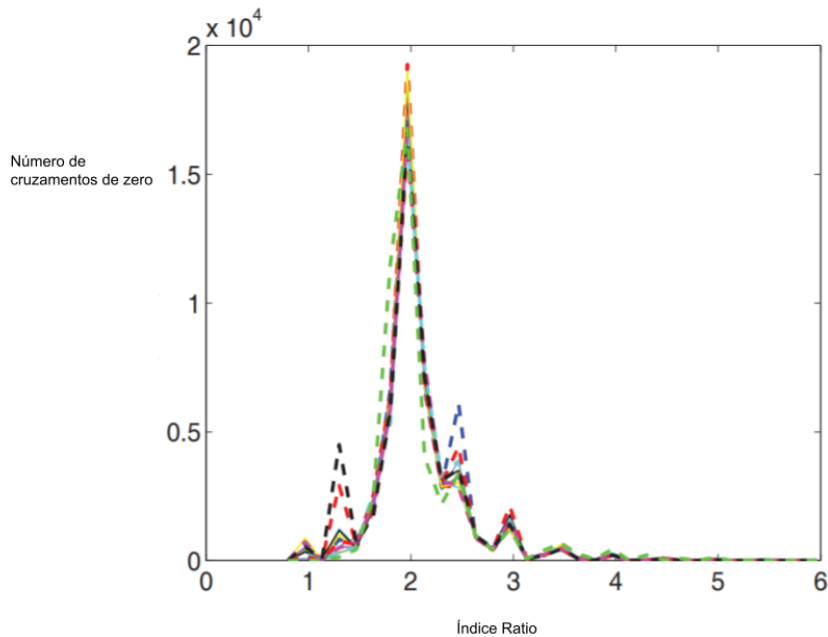


Figura II.8: *Ratio of Zero Crossing Numbers RZCN*

Para cada série temporal extraiu-se o IMF e calculou-se o RZCN. Em seguida, obteve-se o valor de R^i contido em cada série temporal utilizada no processo. Em seguida, foram calculados os limites superior e inferior de 95% e 5% para cada distribuição. Qualquer RZCN que difere em dois desvios padrão da média dos limites é significativamente diferente e representa o i^* dos componentes da amostra, qualquer índice *ratio* menor que dois é desconsiderado na análise. A linha preta na Figura II.8 indica a média dos valores de R^i , as estrelas representam os R^i s da série temporal e as linhas pontilhadas representam dois desvios padrão da média dos limites inferiores e superiores.

Como Moghtaderi et al. [2010] relata, o problema com esta abordagem é que os limites inferior e superior R^i são inteiramente baseados empiricamente. É possível que o menor $R^{(i+1)}$ resulte em uma detecção falsa, e conseqüentemente, um fator incorreto de i^* .

Combinando as abordagens A abordagem combinada *Energy Rate Approach* (ERA) pode reduzir as falhas de detecção no cálculo do i^* . Em EA o i^* é obtido através do menor valor de G^i sendo que $G^i > G^{i-1}$. Já o método RA, obtém todos os índices R^i , e após isso, utiliza o último R^i contido dentro do limite superior e inferior definidos. A partir desses dois resultados obtêm-se o i^* , sendo o menor valor comum de índice válido nas duas abordagens [Moghtaderi et al., 2011].

II.3.3 Componentes Zhang

A abordagem para obtenção de um filtro de tendência descrita na seção anterior permite separar a série em componentes de flutuação e tendência. Zhang et al. [2008] propuseram uma abordagem de análise de séries temporais utilizando EMD. Eles perceberam que a duração dos efeitos externos nas séries de preço de fechamento do óleo cru representava um problema de *multi-scale*. Wen et al. [2017] identificaram as mesmas características na série do ouro. Desta forma, o trabalho de Zhang et al. [2008], pode usar o EEMD para decompor a série e gerar três novos componentes: STF, *Shock Significant Term* (SST) e *Long Term* (LT). O objetivo é restringir as informações em escalas de tempo e fazer uma análise mais precisa do grau de impacto dos eventos externos sobre esses componentes.

II.4 Índice de Incerteza sobre a Política Econômica

Incerteza é um conceito associado a incapacidade de se prever um comportamento previamente conhecido a medida que o mesmo se repita no decorrer do tempo [Corrado and Jordan, 1997]. Por ser um conceito extenso não é facilmente observável, existindo uma grande dificuldade de mensurar seu impacto. Bernanke [1983] desenvolveu a teoria de escolhas irreversíveis, estabelecendo a relação de que empresas expostas a determinadas incertezas, tendem a adiar decisões de investimentos e de contratações, gerando inevitavelmente uma diminuição da atividade econômica num cenário maior.

Uma das teorias mais importantes em finanças HME afirma que o preço dos ativos são consequência das informações disponíveis para os investidores. À medida que novas informações surgem, os investidores tentaram antecipar o movimento das ações, sendo o preço das ações uma consequência disso [Bernanke, 1983]. Devido a esse comportamento, mudanças tornam-se difíceis de serem previstas, dificultando a capacidade de ganhos futuros.

Em complemento a HME foi desenvolvido a UIH. Jordan [1997] informa que na (UIH) os preços das ações são ajustados quase imediatamente de acordo com a chegada de novas informações, e mesmo que os investidores não possam ter a visão clara de todo impacto destes eventos, os preços das ações vão refletir essas incertezas, fazendo com que o investidor com medo do aumento do risco inerente no seu investimento, negocie suas ações abaixo de um determinado valor fundamental [Pástor Ľuboš; Veronesi, 2013]. Para os autores, a reação dos investidores é de aversão ao risco. Devido a isso, informações desfavoráveis tendem apresentar uma reação contraproducente ao evento de incerteza. Em contraponto, informações positivas apresentam uma mudança lenta no preço dos ativos. A incerteza aumenta o risco dos ativos e diminui a expectativa dos retornos esperados, uma vez que esse risco diminui, os preços dos ativos tendem a ser reajustados [Jordan, 1997].

Baseado nestas teorias, pode-se entender que a disseminação de notícias políticas e econômicas

com termos negativos, tendem a aumentar a incerteza em relação a qual política econômica será adotada por um determinado governo de um país. Conclui-se então, quanto maior seja a incerteza em relação a economia, maior é o potencial de perda e ganho nos retornos esperados por um investidor, sendo esta característica uma consequência do risco de investimento no mercado financeiro, a medida que a incerteza diminui, esse comportamento se dissipa [Samuelson, 2013; Golmohammadi and Zaiane, 2015]

Baker et al. [2016] desenvolveu um índice de incerteza da política econômica EPU. Este índice de incerteza foi criado a partir de informações presentes em diversos jornais de grande circulação. O índice captura termos de incerteza contidos nas notícias que são consumidas pelos agentes financeiros. As capturas dos termos buscam identificar tanto preocupações de curto, como de longo prazo que possam estar contidas nas publicações. Por fim, este trabalho utilizou as séries EPU de três países, Brasil, China e Estados Unidos da América [Baker et al., 2016].

II.5 Detecção de eventos

A literatura apresenta uma grande diversidade de tipos de eventos, estes comumente são classificados como anomalias ou pontos de mudança [Carreño et al., 2020]. Anomalias são observações de curto prazo que fogem de um comportamento padrão de um conjunto de dados [Braei and Wagner, 2020]. De outro modo, pontos de mudança representam uma alteração de estado de um processo que gera a série temporal [Aminikhanghahi and Cook, 2017]. A detecção de pontos de mudança tem como objetivo encontrar o instante (ou intervalo) específico no tempo, onde há uma inflexão em uma característica do domínio representado pela série temporal [Aminikhanghahi and Cook, 2017]. Apesar da diversidade de métodos, a escolha adequada de qual o melhor método a ser utilizado depende do tipo do evento que se busca observar em uma série. Métodos especializados em identificar um fenômeno, podem negligenciar outros com natureza distinta, e assim, detectar incidentes de forma imprecisa ou de maneira incorreta [Downey, 2008].

II.5.1 Pontos de mudança

O estudo de pontos de mudanças também é descrito como a capacidade de observar mudanças repentinas em dados quando uma propriedade da série muda [Downey, 2008]. Em resumo, os pontos de mudança buscam entender o comportamento de um domínio de dados quando transições entre estados acontecem mediante a um evento [Downey, 2008]. Aminikhanghahi and Cook [2017] descrevem que dada uma determinada série $X(t) = x_1, x_2, \dots, x_n$, um ponto de mudança pode ser definido como um evento, que pode ser identificado a partir de um teste de hipótese com duas alternativas. Na hipótese nula h_0 nenhuma mudança acontece, já na hipótese h_1 a mudança ocorre [Aminikhanghahi and Cook, 2017]. Algoritmos de detecção de pontos de mudança são tradicio-

nalmente classificados como *online* ou *offline*. Os algoritmos *offline* consideram todo conjunto de dados de uma só vez observando os dados passados para identificar onde a mudança ocorreu. Já os algoritmos *online* são executados simultaneamente aos processos que eles buscam monitorar, processando cada ponto de dados à medida que novos dados tornam-se disponíveis, o objetivo é detectar pontos de mudança o mais rápido possível [Lau and Yamamoto, 2010]. Neste trabalho foi utilizado somente algoritmos *offline*.

Iterative Cumulative Sum of Squares (ICSS)

Este método assume que uma série temporal tem uma variância estacionária em um segmento antes que um *breakpoint* ocorra, mudando o nível de volatilidade, o novo nível de variância permanece estacionária até que outra mudança ocorra. Para estimar o número e a localização dos pontos de interrupção, as somas cumulativas dos quadrados são usadas:

$$C_k = \sum_{t=1}^k \varepsilon_t^2, k = 1, 2, \dots, T \quad (\text{II.14})$$

Onde ε_t é uma série de variáveis aleatórias não correlacionadas, com média 0 e variância igual a σ_t^2 . As somas cumulativas centradas dos quadrados são definidas como:

$$D_k = \frac{C_k}{C_T} - \frac{k}{T}, k = 1, 2, \dots, T, D_0 = D_T = 0 \quad (\text{II.15})$$

O algoritmo procura pontos de mudança de variância em cada uma das seções criadas D_k até D_{kmax} . Ao repetir esse procedimento ao longo do tempo é possível determinar a mudança de variância na série, se este valor exceder um limite predeterminado, conclui-se que existe um ponto de mudança em torno deste ponto. Uma vez que novos pontos de interrupção não são mais encontrados, a busca é interrompida [Inclán and Tiao, 1994].

Seminal Change Point (SCP)

O *Seminal Change Point* (SCP) é um método de detecção de pontos de mudança proposto por Guralnik and Srivastava [1999] com abordagem univariada. Essa abordagem permite detectar pontos de mudanças pela detecção dos parâmetros do modelo que descrevem os dados subjacentes, para fazer isso ele usa uma combinação de pontos de mudança e técnicas de *model select*. O algoritmo segue dois passos: No primeiro passo, não é colocado nenhuma restrição na classe de funções que será ajustada às subsequências entre sucessivos pontos de mudança. Em segundo lugar, o número de pontos de mudança é não fixado a priori, sendo ajustados modelos aos segmentos de dados antes e depois do ponto, a existência de um ponto de mudança é determinado caso o erro de ajuste seja menor em comparação com o valor sem o ponto de mudança. Por não necessitar de

nenhum modelo anterior essa abordagem não pode ser considerada como bayesiana. O principal problema do ponto de mudança pode ser definido como encontrar um modelo m segmentado por parte: dado pela Equação II.16, onde $f_i(t, \mathbf{w}_i)$ é a função ajustada ao segmento i , \mathbf{w}_i é o vetor de parâmetros, θ_i são os pontos de mudança entre segmentos sucessivos e $e_i(t)$ são os termos de erro. Em cada segmento i o principal passa a encontrar $\hat{f}_i(t, \mathbf{w}_i)$ que minimiza o valor esperado da função de perda $E[\text{loss}(y(t), \hat{f}_i(t, \mathbf{w}_i))]$, onde x_t , $t = 1, 2, \dots, n$ é uma série temporal e t a variável de tempo. Por fim, para cada segmento, é examinado se ele pode ser dividido em dois novos segmentos.

$$\begin{aligned}
 x &= f_1(t, \mathbf{w}_1) + e_1(t), & 1 \leq t \leq \theta_1 \\
 &= f_2(t, \mathbf{w}_2) + e_2(t), & \theta_1 \leq t \leq \theta_2 \\
 &\dots \\
 &= f_k(t, \mathbf{w}_k) + e_k(t), & \theta_{k-1} \leq t \leq t(k)
 \end{aligned} \tag{II.16}$$

II.5.2 Anomalias

Detecção de anomalias é apresentada como um problema de identificar padrões nos dados fora de conformidade com a distribuição normal esperada [Han et al., 2012]. A análise deste fenômeno é de extrema relevância devido ao fato das anomalias demonstrarem características tão distintas do restante dos dados da série, que parece que não foram geradas pelo mesmo mecanismo [Han et al., 2012]. Por isso é importante entender quais são esses mecanismos que interferem no processo de criação desses dados anômalos presentes na série. Para entender e avaliar a presença desses mecanismos é necessário fazer diversas suposições sobre o restante dos dados da série e como estes dados anômalos detectados violam significativamente essas suposições.

Alguns processos de mineração de dados lidam com as anomalias, como ruídos e exceções [Aggarwal, 2013]. Estes processos entendem que tais eventos não necessitam ser analisados, removendo os da série. Contudo, esses eventos incomuns são de extrema importância para alguns modelos de análises, possuindo mais informações úteis do que os dados restantes [Carreño et al., 2020]. A grande maioria dos *datasets* consistem de dados normais, sendo as anomalias uma parte pequena do conjunto de dados [Golmohammadi and Zaiane, 2015; Chandola et al., 2009].

A literatura apresenta diversos métodos para a detecção de anomalias, neste trabalho foi utilizado três métodos que utilizam processos de decomposição para detecção de anomalias, *Forward and Backward Inertial Anomaly* (FBIAD), detecção de anomalias de volatilidade utilizando *Generalized Autoregressive Conditional Heteroscedasticity* (GARCH) e detecção de anomalias utilizando EMD.

Ao utilizar o processo de decomposição busca se efetuar a separação do componente de tendência e sazonalidade inerentes observados em uma série temporal. O processo de decomposição uma vez realizado, transforma a série em componentes com diferentes escalas de flutuação e um componente

com sinal monotônico que é a tendência. Quando se extrai da série original a tendência e a sazonalidade, obtêm-se um componente restante, e sobre este componente é realizada a busca por anomalias [Salles et al., 2019].

Forward and Backward Inertial Anomaly Detector (FBIAD)

O método FBIAD é uma variação da técnica conhecida como janelas deslizantes, foi criado para lidar com séries temporais heterocedásticas não estacionárias. A diferença entre os dois métodos é que o FBIAD transforma uma série temporal em uma sequência de dados na qual as propriedades estatísticas obtidas por um conjunto de amostras podem ser calculadas e consideradas no processo de normalização. Devido a esta característica, o método FBIAD consegue preservar as propriedades originais da série temporal dentro de cada janela de slide relativa, o que possibilita as janelas deslizantes geradas em FBIAD serem capazes de representar volatilidades diferentes [Lima et al., 2022].

O processo de detecção de eventos FBIAD compreende três fases: (i) transformação de séries temporais não estacionárias em uma sequência estacionária, que é feita através da criação de sequências de janelas deslizantes não sobrepostas para frente e para trás na série temporal; (ii) calcula se a diferenciação inercial para frente e para trás; (iii) registra se as anomalias para frente e para trás. O resultado desse processo consiste em dados que servirão de entrada para métodos de aprendizado de máquina [Lima et al., 2022].

A série temporal é explorada usando janelas deslizantes. Em cada slide, uma média móvel é calculada e todas as observações são diferenciadas da média móvel. Essa abordagem é feita tanto na série temporal quanto em seus valores inversos. A distribuição de todas as observações é calculada. As observações que estão fora da regra do intervalo interquartil são identificadas como anomalias [Lima et al., 2022; Ogasawara et al., 2010].

Volatility Anomaly

Em muitas séries temporais financeiras o nível de volatilidade varia ao longo do tempo, o que faz com que essas características não lineares presentes na série, não possam ser capturadas com os modelos lineares existentes [Malik, 2021]. Assim, surge a necessidade de um estudo que possa captar as mudanças presentes de volatilidade em uma série temporal. Neste contexto, existem diversos modelos econométricos que podem tratar a não linearidade existente nos dados, incluindo a volatilidade estocástica. Dentre os modelos econométricos que tratam da volatilidade, um dos que mais se destacam é o GARCH. O GARCH tem a capacidade de estimar a volatilidade nas séries temporais, baseando se nas observações passadas.

Sendo uma série yt segue um modelo GARCH(p,q) se a Equação II.17 é obedecida, onde μ_t é o

componente médio, σ_t é a variância condicional e R_t é um ruído branco Gaussian (sequencia i.i.d. de variáveis $N(0,1)$), de modo que a distribuição condicional de $\eta_t = y_t - \mu_t$, dado $\eta_{t-1}, \eta_{t-2}, \dots$ is $N(0, \sigma_t^2)$ [Chatzikonstanti, 2017].

$$x_t = \mu_t + \sigma_t R_t \quad (\text{II.17})$$

A Equação II.18 define a variância condicional em p períodos anteriores. Esse termo autor-regressivo p modela a variância condicional dos erros quadráticos, enquanto q modela a variação do processo [Vignoli, 2021]. A variável ϕ_j indica quanto a última observação têm de influência na variância condicional atual, enquanto que θ_j aponta quanto a volatilidade do período anterior deve influenciar a volatilidade atual. Quanto maior for o ϕ_j maior vai ser o impacto nos dados, e quanto maior for a duração, maior será o valor de θ_j .

$$\sigma_t^2 = \sigma^2 + \sum_{j=1}^p \phi_j \sigma_{t-j}^2 + \sum_{j=1}^q \theta_j \tilde{x}_{t-j}^2 \quad (\text{II.18})$$

A volatilidade dos mercados financeiros pode ser afetada por fatores externos atípicos, demonstrando o quão suscetíveis são séries financeiras a fatores extremos [Chatzikonstanti, 2017]. Ao utilizar a abordagem com GARCH é possível capturar essas variações extremas de volatilidade, toda observação é considerada anomalia se ultrapassar um limiar de valor interquartil pré definido.

Detecção de anomalias utilizando (EMD)

A técnica de detecção de anomalias usando EMD é obtida através da aplicação de um *boxplot* no componente STF. Após o processo de remoção de tendências, é possível separar as séries temporais em dois grupos de componentes de IMF, um com informações de alta flutuação e outro com informações de tendência. O componente de alta flutuação passa a ser o local onde será realizada a busca de anomalias, o processo de identificação de anomalias ocorre através do uso da técnica *boxplot*. Esta técnica baseia-se na definição do intervalo interquartil IQR que é o resultado da subtração ($Q3 - Q1$), sendo $Q3$ o quartil superior e $Q1$ o primeiro quartil de uma série temporal $X(t)$. Os dados de séries temporal que estiverem fora desse intervalo serão considerados anomalias [Han et al., 2012].

A Figura II.9 demonstra uma distribuição dos dados no formato de um *boxplot*, sendo as anomalias valores que ficam fora dos limites inferiores e superiores estabelecidos.

II.5.3 Métricas de avaliação de resultados

Nesta seção, são apresentadas as métricas que foram utilizadas para avaliar os resultados. A fim de obter uma melhor análise das técnicas de detecção estudadas, um conjunto de métricas

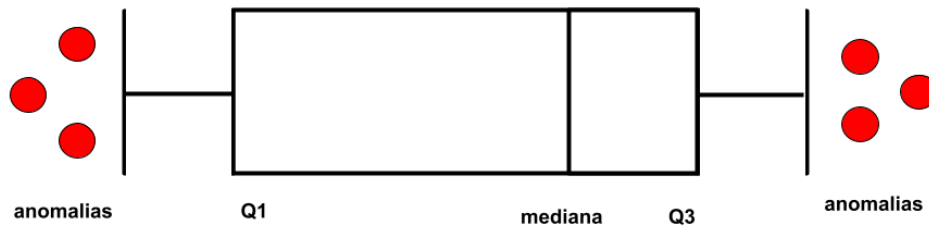


Figura II.9: *Boxplot*, adaptado de Han et al. [2012]

comumente usadas foi selecionado para comparar os resultados dos métodos de detecção de eventos de séries temporais [Aminikhanghahi and Cook, 2017]. Uma matriz de confusão tem a capacidade de relatar os erros e acertos obtidos através da aplicação de métodos de detecção. Para isso, compara os resultados da detecção com os valores de referência. A matriz de confusão tem quatro termos:

1. **VN**: Verdadeiro Negativo - ocorrências corretamente classificadas como negativas.
2. **FN**: Falso Negativo - ocorrências positivas incorretamente classificadas como negativas.
3. **FP**: Falso Positivo - ocorrências negativas classificadas incorretamente como positivas.
4. **VP**: Verdadeiro Positivo - ocorrências corretamente classificadas como positivas.

Precisão - É definido como o número de verdadeiros positivos (VP) sobre a soma de verdadeiros positivos (VP) e o número de falsos positivos (FP). Essa métrica tem a capacidade de verificar entre as informações identificadas como positivas, qual estava realmente correta [Han et al., 2012].

$$precision = \frac{VP}{(VP + FP)} \quad (II.19)$$

Recall - É definido como o número de verdadeiros positivos (VP), sobre a soma do número de verdadeiros positivos (VP) e o número de falsos negativos (FN). Essa métrica analisa a quantidade de informações identificadas como positivas que estavam realmente corretas [Han et al., 2012].

$$Recall = \frac{VP}{(VP + FN)} \quad (II.20)$$

F1 - Consiste em uma média harmônica entre *precision* e *recall* [Han et al., 2012].

$$F1 = \frac{2 \times \textit{precision} \times \textit{recall}}{\textit{precision} + \textit{recall}} \quad (\text{II.21})$$

Capítulo III Trabalhos Relacionados

Neste capítulo será abordado a respeito do método de levantamento dos trabalhos relacionados para esta dissertação. Com objetivo de obter uma base para fundamentar os conceitos abordados neste trabalho, foi aplicado um filtro para busca de documentos que utilizassem as técnicas de decomposição *multi-scale* mais atuais. O foco é ter uma revisão que possa fornecer uma visão geral do estado da arte relacionado à temática abordada em relação à pesquisa, aumentando o entendimento a respeito dos trabalhos acadêmicos em temas similares, e contribuindo com a formulação de uma dissertação que auxilie na evolução da área de estudo escolhida.

A base de dados *Scopus* foi utilizada para a pesquisa dessa dissertação, foi executada duas buscas utilizando as seguintes strings: TITLE-ABS-KEY(“*Empirical Mode Decomposition*” AND “*multiscale*”) AND (LIMIT-TO(LANGUAGE, “*English*”)) e TITLE-ABS-KEY(“*Economic Policy Uncertainty*” AND “*multiscale*”) AND (LIMIT-TO(LANGUAGE, “*English*”)). A primeira string retornou 87 artigos e a segunda retornou 84 artigos, para os quais alguns critérios de inclusão e exclusão foram aplicados. Os critérios de inclusão são apresentados a seguir:

Código	Descrição
CI1	Trabalhos relacionados a ciência da computação e economias.
CI2	Trabalhos que envolvam detecção de eventos no cenário <i>multi-scale</i> .
CI3	Trabalhos que envolvam detecção de eventos com EMD ou variantes do método.
CI4	Trabalhos que utilizem métodos de detecção de pontos de mudança em séries temporais.
CI5	Trabalhos que utilizem métodos de detecção de anomalias em séries temporais.
CI6	Trabalhos que tratam de características <i>multi-scale</i> e EPU.

Além dos critérios de inclusão, foram definidos alguns critérios de exclusão para a filtragem de resultados. A Tabela a seguir apresenta os critérios de exclusão considerados. Na seleção dos artigos foram lidos os resumos dos 171 artigos resultantes da busca no *scopus*, selecionando os trabalhos mais coerentes com a pesquisa desta dissertação. No primeiro refinamento, cerca de 60% dos trabalhos foram eliminados.

Código	Descrição
CE1	O artigo é uma versão antiga de um outro já avaliado.
CE2	O trabalho não foi encontrado em nenhuma fonte de arquivos disponível.
CE3	Artigos fora dos critérios de inclusão.

A exclusão desses trabalhos foi baseada nas condições de exclusão demonstradas anteriormente. A próxima etapa foi ler a introdução e conclusão dos 49 artigos restantes e novamente separar aqueles mais relacionados a esta dissertação. Nesta segunda etapa restaram 28 artigos para leitura completa. Na etapa final foram consultados artigos considerados base para o tema dessa pesquisa, utilizando *snowballing* foram adicionados mais 10 novos arquivos. Consequentemente foi alcançado um total de 38 trabalhos para leitura completa e suporte à pesquisa. Na sequência serão apresentados os artigos de grande importância na elaboração desta dissertação.

Nos últimos anos tem tido um aumento de interesse no estudo de detecção de eventos em séries temporais. O uso de técnicas de detecção de eventos podem auxiliar na descoberta de informações inesperadas ou sequências de eventos anormais, esse cenário de pesquisa tem sido o foco em diversas áreas das ciências que lidam com séries temporais, fornecendo informações importantes para tomadas de decisões. No caso das séries financeiras, o estudo de eventos tem sido o foco de pesquisadores que procuram por eventos que podem gerar mudanças no comportamento padrão da série, podendo interferir no retorno esperado em determinados investimentos.

Alguns trabalhos escolhidos tiveram o foco na detecção de eventos sob diferentes componentes das séries temporais. O trabalho proposto por Zhang et al. [2008] apresentou um processo que permite construir componentes para análise *multi-scale* usando o EMD. O artigo demonstra que para obter esses componentes é necessário fazer a remoção do componente de tendência da série temporal. A abordagem proposta por Zhang et al. [2008] cria um filtro de tendência calculando a média da soma de cada componente de IMF. Depois disso, um *t-test* é usado para identificar qual IMF está associado ao componente de tendência.

Wen et al. [2017] propuseram uma análise do efeito de eventos externos sobre a volatilidade do preço do ouro. Neste trabalho é feito um estudo comparativo da correspondência de eventos como crises econômicas e guerras nas quebras estruturais presentes na série. Este trabalho se inspira nos conceitos de componentização *multi-scale* propostos por Zhang et al. [2008] e acrescenta um fluxo de análise para medir os efeitos destes eventos na mudança de volatilidade nos componentes com diferentes flutuações obtidos através da série do ouro. Utilizando *Iterative Cumulative Sum of Squares* (ICSS) para identificar pontos de mudança, ele aplica o *Chow Test* para verificar a presença ou não de uma quebra estrutural nos pontos de mudança encontrados. Embora o trabalho de Wen et al. [2017] se mostre uma ferramenta boa para análise de eventos externos refletidos em quebras estruturais encontradas na séries, o método implica de um conhecimento prévio dos eventos

históricos anotados para análise e é focado unicamente em eventos onde existe uma mudança de variância detectada.

Seguindo o conceito do estudo de fatores externos nas séries financeiras, o trabalho de Chatzikonstanti [2017] propõe um método para detecção de pontos de mudança utilizando uma variação do trabalho proposto por Inclán and Tiao [1994] ICSS. Além disso, ele apresenta um método baseado em TW para detecção de outliers. O resultado do trabalho demonstrou que as anomalias detectadas nas séries de retorno são associadas com eventos externos como crises financeiras.

Por fim, o trabalho de Moghtaderi et al. [2011] focado na obtenção de tendências em séries temporais [Flandrin et al., 2004a,b; Moghtaderi et al., 2010, 2011], trata-se de um extenso estudo com diferentes métodos de filtragem testados em diferentes séries. O trabalho de Moghtaderi et al. [2011] compara a abordagem proposta com outras técnicas de filtragem de tendências usadas em economia, como o filtro L1 e Hodrick–Prescott [Sakarya and de Jong, 2022; Yamada, 2018].

Capítulo IV Metodologia

A metodologia proposta nesta dissertação seguiu três etapas: (i) A primeira etapa apresenta o conjuntos de séries temporais econômico-financeiras utilizadas; (ii) A segunda etapa apresenta como é feita a aplicação do filtro de tendência na série temporal financeira para gerar os componentes *zhang* apresentados na Seção II.3.3. (iii) A última etapa demonstra a aplicação dos métodos de detecção de eventos nos componentes *zhang* gerados na etapa anterior.

IV.1 Aquisição de dados

O conjunto de séries temporais financeiras que analisamos neste trabalho foram todas coletadas do Yahoo! Finanças¹, no período de janeiro de 2004 a agosto de 2022 e se encontram na Tabela ??.

As séries temporais foram separadas em 4 grupos: índices do mercado de ações no Brasil, China, EUA e cotações de moedas frente ao real: **Índice de Ações no Brasil:** Índice Bovespa (Ibovespa), Índice Brasil 100 (Ibrx100), Índice Brasil 50 (Ibrx50), Índice Brasil Amplo (IBRA), Índice de Ações com Governança Corporativa Diferenciada (IGC). **Índice de ações na China:** *Compiled China Securities Index* (CSI 1000), *Shanghai Composite Index* (Shanghai), *Shenzhen Stock Exchange Component* (SZSE Component), *FTSE–Xinhua China A50 Index* (China A50) e *Hang Seng China Enterprises Index* (HSCE). **Índice de ações nos EUA:** *Down Jones Industrial Average* (DJI), *National Association of Securities Dealers Automated Quotations* (Nasdaq), *Standard and Poor’s 500* (SP500), *Dow Jones Composite Average* (DJ Composite) e *Small Cap Stock Market Index* (Small Cap 2000). **Taxas de câmbio:** A série temporal contida neste grupo são representados pelos valores das moedas estrangeiras convertidas em reais. As seguintes moedas foram convertidas: Euro, Dólar Americano, Yuanl Chinês, Peso Argentino e Peso Chileno. No total, 20 séries temporais com os valores de fechamento mensais foram utilizadas neste trabalho.

IV.2 Decomposição e filtro energético

A decomposição utilizando CEEMD permite obter diversos componentes que são séries derivadas do sinal original decomposto. As IMFs mais o resíduo são a resultante da decomposição de modos empírico. Cada uma das IMFs representa o modo de oscilação de uma série. E em cada

¹<https://finance.yahoo.com/>

Índice médio (Brasil)	Índice médio (China)	Índice médio (EUA)	Cotação de Moedas
Ibovespa	CSI 1000	DJI	EUR/BRL
Ibrx100	Shanghai	Nasdaq	USD/BRL
Ibrx50	SZSE Component	S&P 500	CNY/BRL
Ibra	China A50	DJ Composite	CLP/BRL
IGC	HSCE	Small cap 2000	ARS/BRL

Tabela IV.1: Séries estudadas

um desses componentes, podemos encontrar dados com diferentes ciclos de flutuações. Através desses componentes, podemos aprimorar o entendimento do comportamento local e geral da série. Contudo, as IMFs com menor oscilação podem ter informações de tendência de médio prazo. A presença desta tendência acaba dificultando a análise da flutuação de curto prazo. Para solucionar esse problema foi criado um filtro capaz de separar IMFs de tendência, de IMFs com flutuações de curto prazo.

O processo onde é feito a aplicação do filtro de tendência (usando ERA) pode determinar i^* , que é o ponto onde é possível dividir as IMFs em dois grupos. Através do uso deste filtro, obtém-se o i^* , que é o retorno do método ERA. Com essa variável, é possível encontrar o limiar onde se tem a IMF que separa informações de volatilidade e tendência, sendo a IMF limiar igual IMF_j e o índice $j = i^*$. Utilizando essa variável fornecida pelos métodos ERA pode-se somar as (IMFs) e criar os novos componentes.

IV.2.1 Componentes obtidos através do filtro de tendencia

Nesta parte do trabalho são apresentados cada um dos componentes que podem ser criados utilizando o filtro de tendência. A série de índice médio de ações da Ibovespa demonstrada na Figura IV.1, foi utilizada para apresentar os resultados da utilização do filtro ERA. Após a aplicação do método de decomposição CEEMD, quatro IMFs e um resíduo foram obtidos. Com esses 5 componentes em mãos, é possível criar quatro novos componentes utilizando o filtro de tendência ERA, a saber : STF, SST, LT e *Overall Adaptive Trend* (OAT). Cada um destes componentes será apresentado em detalhes a seguir.

IV.2.2 *Short Term Fluctuation (STF)*

O primeiro componente a ser apresentado é o STF, ele é gerado pela soma dos componentes de IMF de maior flutuação. Uma vez que a base de dados escolhida no experimento deste trabalho utiliza dados mensais, as oscilações presentes no componente STF correspondem a ciclos de menos de um ano. O processo de formação do STF se dá a partir de eventos externos com efeitos de pequena duração. Devido a curta duração dos efeitos, eles podem ser representados por eventos

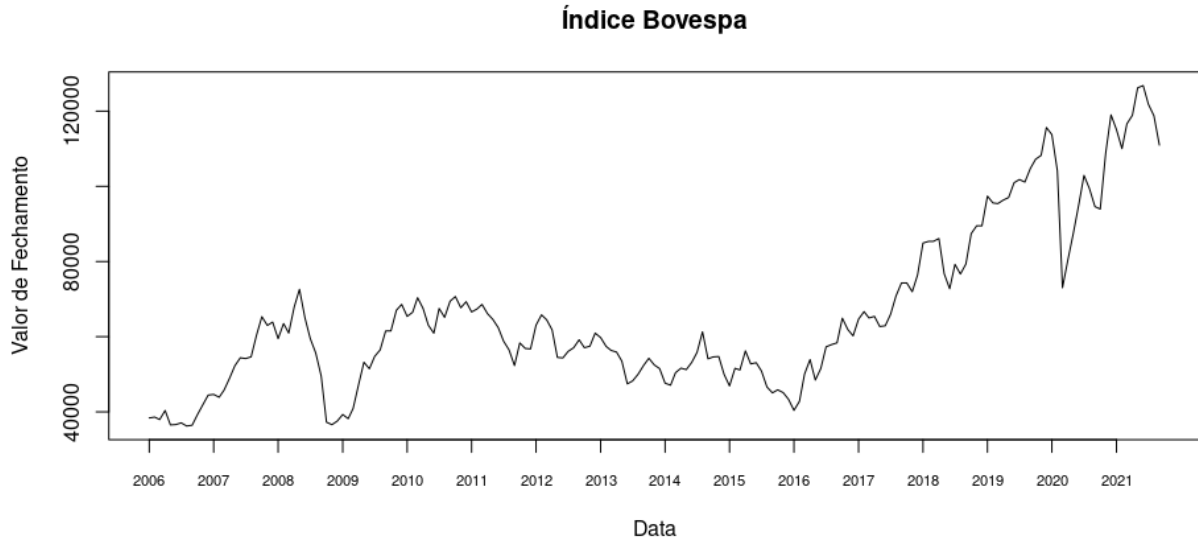


Figura IV.1: Série histórica de Fechamento Ibovespa

presentes no componente STF. Muitos desses eventos podem ser explicados como efeitos de desequilíbrio normal do mercado. Sendo categorizados como, eventos imprevisíveis com impacto de curto prazo no processo de formação de preços das séries financeiras.

$$STF = \sum_{j=1}^{i^*-1} c_j(t) \quad (IV.1)$$

A Figura IV.2 é mostrado o componente de STF. Essa série é construída a partir da obtenção do i^* , que no caso da Ibovespa especificamente é três, sendo $i^* = 3$. Todas as séries de IMFs com índice inferior a três são somadas ($IMF_1 + IMF_2$).

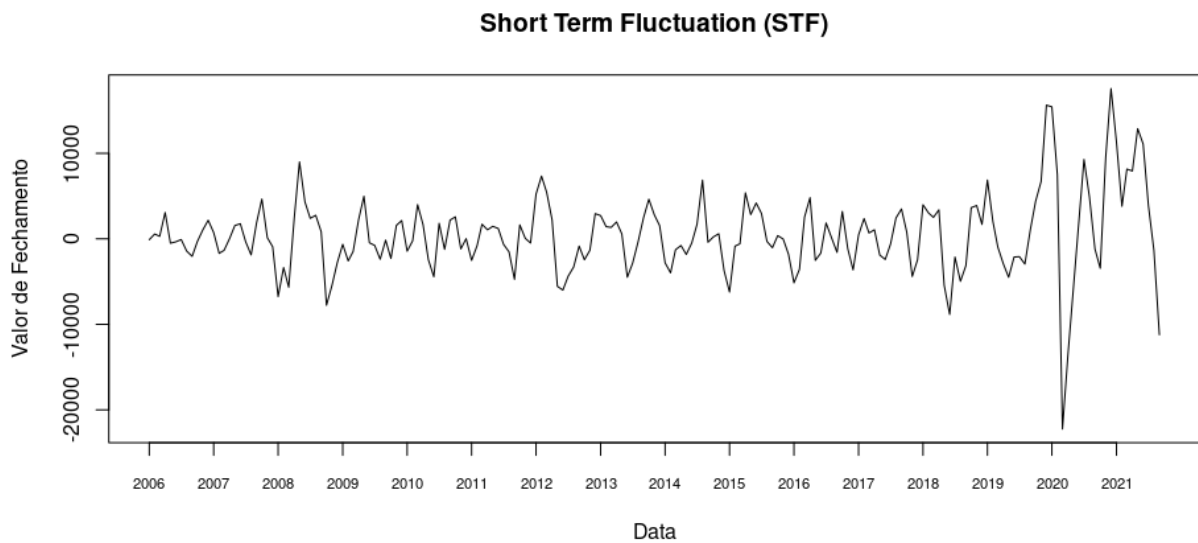


Figura IV.2: Série histórica de Fechamento Ibovespa - Componente STF

IV.2.3 Shock Significant Term (SST)

O componente SST é gerado pela soma de componentes de IMF com índices maiores ou iguais a i^* , como é descrito na equação IV.2. Ao contrário do STF, este componente tende a mudar mais lentamente, pois exclui os efeitos das atividades de alta frequência. Esta característica explica porque a curva gerada pelos eventos contidos neste componente tem uma aparência mais suavizada. As altas ou quedas acentuadas dos preços dos ativos são desencadeadas por eventos significativos cujo impacto pode durar por vários anos. Por mais que as flutuações normais contidas no SST possam ser negligenciadas em previsões ligadas ao estudo da tendência de longo prazo das séries financeiras, elas não podem ser ignoradas nas previsões de médio prazo.

$$SST = \sum_{j=i^*}^n c_j(t) \quad (IV.2)$$

Na Figura IV.3 demonstra a representação gráfica do componente SST. Esse componente é criado a partir i^* , onde essa variável representa os índice de IMFs maiores ou iguais a $i^* \geq 3$. O valor da nova série é resultado da soma de duas IMFs (IMF_3, IMF_4). Nesse componente é encontrado grande parte da informação de variação de preços mais significantes da série.

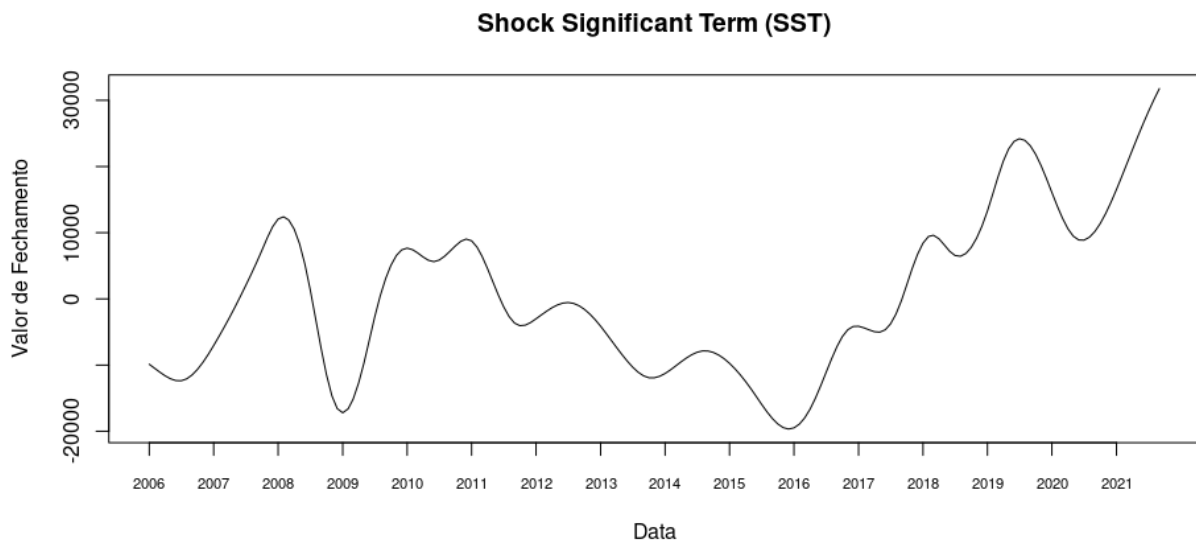


Figura IV.3: Série Histórica de Fechamento Ibovespa - Componente SST

IV.2.4 Long Term (LT)

O componente de LT é formado pelo resíduo proveniente da aplicação do método CEEMD em uma série temporal. Os dados presentes em uma série financeira são basicamente determinados pela tendência de longo prazo, o que faz com que os eventos encontrados neste componente determinem o padrão de crescimento/decrescimento que os dados vão seguir em LT. A Figura IV.4 demonstra

a representação gráfica do componente LT.

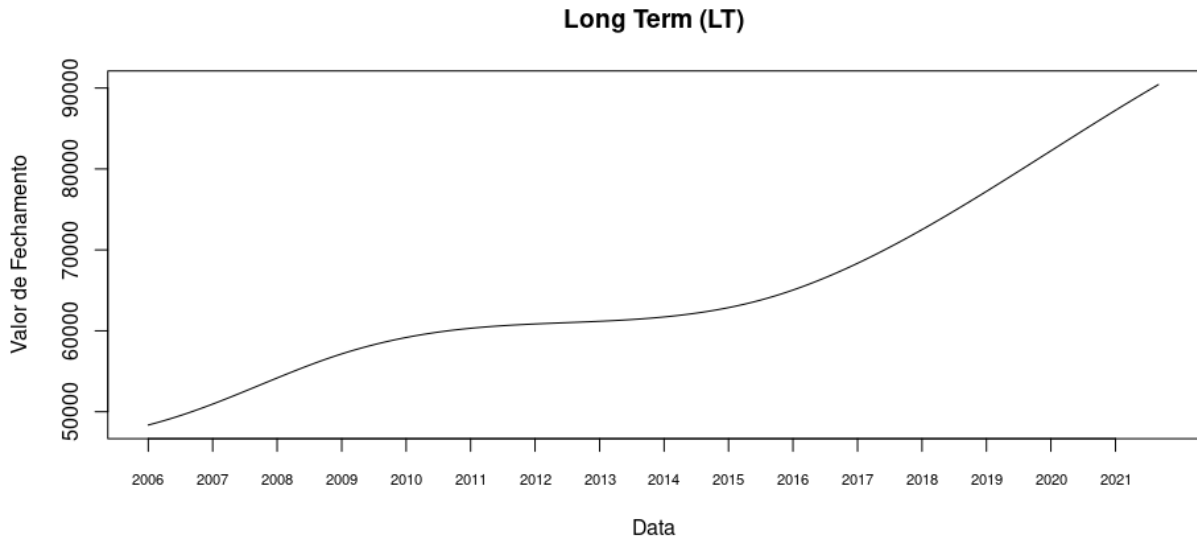


Figura IV.4: Série Histórica de Fechamento Ibovespa - Componente LT

IV.2.5 Overall Adaptive Trend (OAT)

A Figura IV.5 demonstra a representação gráfica do componente OAT. O OAT é o componente gerado a partir da série temporal financeira menos o componente STF. Para o conjunto de dados estudado, corresponde ao movimento da tendência anual da série. O valor da nova série é resultado da soma dos componentes (IMF_3 , IMF_4 , $residuo$), conforme descrito pela Equação IV.3. A Figura IV.5 demonstra a representação gráfica do componente OAT.

$$OAT = \sum_{j=i^*}^n c_j(t) + r(t) \quad (IV.3)$$

Ao utilizar MSED podemos entender em qual componente de *zhang* um tipo de evento possui maior presença. Os eventos presentes em componentes de alta flutuação podem não ser encontrados em componentes de tendência, anomalias por exemplo, podem ser mais fáceis de se observar em alguns componentes, enquanto que em outros componentes é mais fácil de encontrar a presença de mudanças de variância. A relação entre os componentes e os eventos-alvo, pode ajudar a entender como a série financeira analisada reflete os eventos externos aos quais está exposta.

IV.3 Detecção de eventos em componentes derivados do (MSED)

O método MSED usa diferentes métodos de detecção de eventos nos componentes obtidos após aplicação do filtro de tendência. O componente STF é representado por séries com alta flutuação onde podem ser encontrados fenômenos isolados de curto prazo e mudanças abruptas de variância.

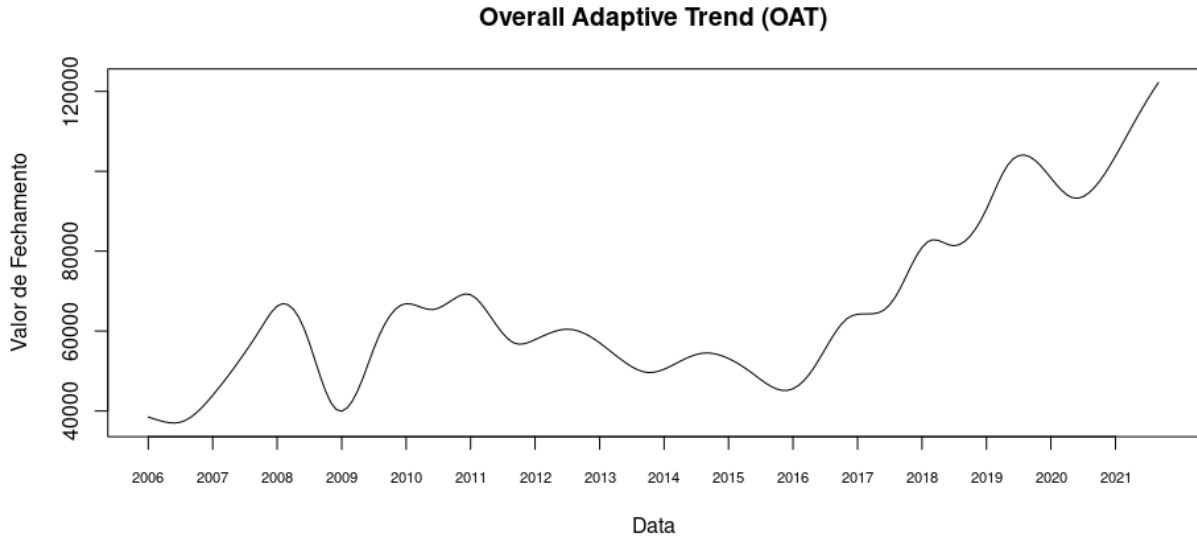


Figura IV.5: Série histórica de Fechamento Ibovespa - Componente OAT

Para este componente, foram escolhidos os métodos de ICSS e detecção de anomalias usando EMD. Na componente SST é possível observar eventos mais significativos cujo impacto pode durar vários anos. Neste componente, o MSED detecta os eventos utilizando SCP e ICSS. O componente LT é um bom lugar para entender as inflexões de direção dos preços das séries financeiras, então o MSED também usou o SCP neste componente.

A Figura IV.6 representa os passos usados pelo método MSED para detectar eventos na série temporal. O MSED tem quatro etapas. Primeiramente, as séries financeiras utilizadas são decompostas pelo CEEMD, e o resultado desse procedimento são os componentes de IMF + resíduo. Na próxima etapa, o filtro de tendência é aplicado aos componentes do IMF obtendo i^* , que é o resultado da utilização da técnica *ERA*. O i^* define o ponto de corte da tendência do filtro de tendência. Na etapa três, os componentes de alta flutuação, flutuação média e resíduo são criados a partir do resultado contido em i^* . Na última etapa, os métodos de detecção de eventos são aplicados a cada componente. Os resultados permitiram identificar a natureza mais predominante dos eventos em uma série temporal. Por meio dessas informações, o analista pode escolher a série temporal com características de incerteza aceitáveis para o perfil de investimento que melhor se ajusta ao risco ao qual pretende estar exposto.

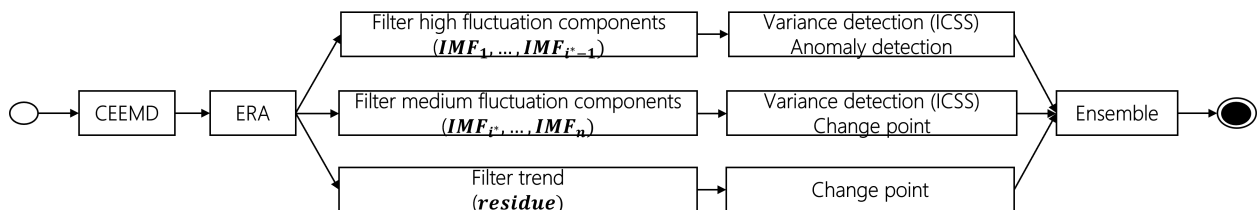


Figura IV.6: Multi-Scale Event Detect MSED

Capítulo V Resultados

Nesse capítulo são apresentados as métricas com os resultados dos experimentos visando a comparação do desempenho dos diferentes métodos de detecção de eventos utilizados. Este capítulo é organizada em cinco partes: (i) Na primeira etapa são descritas as abordagens usadas para construir as séries de *data label* utilizadas para criar a relação entre os eventos externos representados na séries de EPU pelas variações de termos de incerteza, e os eventos detectados nas séries temporais financeiras; (ii) Análise das métricas presentes no conjunto de dados de taxas de câmbio para o real brasileiro; (iii) Análise das métricas presentes do conjunto de dados índices do mercado de ações brasileiro; (iv) Análise das métricas presentes no conjunto de dados de índices do mercado de ações chinês; (v) Análise das séries presentes no conjunto de dados de índices do mercado de ações americano. Por fim é apresentada uma discussão sobre os principais resultados obtidos em cada uma das análises.

V.1 Séries de *data label*

A primeira tarefa a ser realizada uma vez selecionada as séries financeiras é identificar pontos na série financeira, onde os eventos externos contidos na série do EPU têm a maior probabilidade de ser identificados. Os *labels* correspondentes à instância de dados da série temporal podem identificar a presença ou ausência de um evento associado a uma determinada série temporal [Chandola et al., 2009]. Para gerar essa análise de correspondência, séries de *data labels* foram criados com base na séries de EPU do Brasil, China e Estados Unidos da América (EUA) [Baker et al., 2022]. A série de *data label* é uma série mensal que possui 0 ou 1 como dado. A presença de um evento em um determinado mês recebe o *label* 1, caso contrário receberá o *label* 0.

Neste trabalho os valores presentes na série de *data label* foram criados a partir de duas abordagens. A primeira abordagem é a *Annotated Approach* (AA) que considera todas as marcações presentes no índice AA, que é baseado no trabalho do Baker et al. [2016]. A segunda abordagem é a de *Envelop Approach* (EVA), essa abordagem considera um envelope superior e inferior sobre os máximos e mínimos locais presentes na série de EPU, e define um limiar de corte a partir do cálculo de percentis. Cada uma dessas abordagens é explicada em detalhes a seguir:

V.1.1 Annotated Approach (AA)

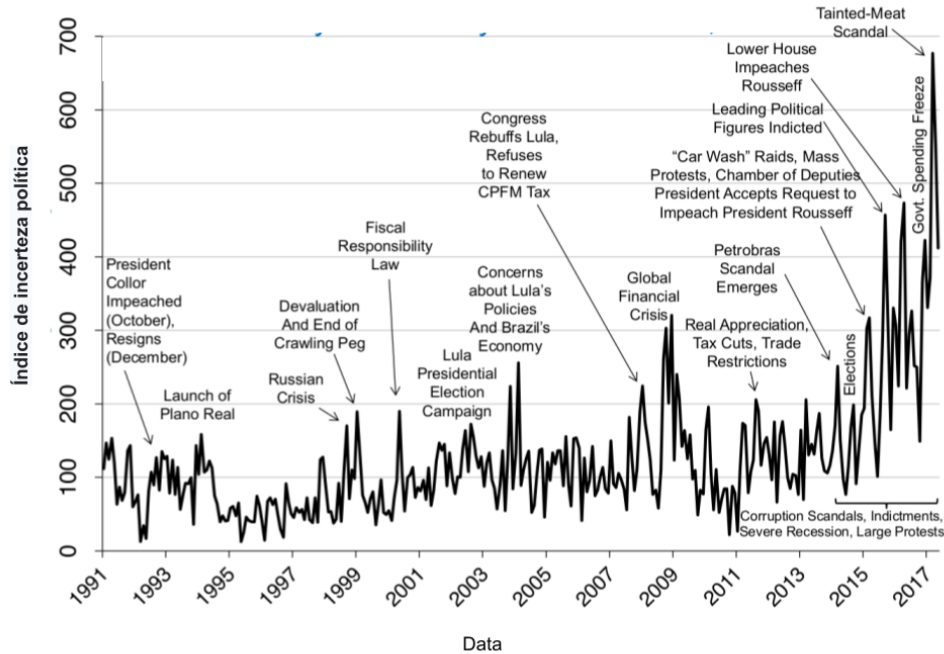


Figura V.1: Annotated Approach - Série EPU do Brasil [ser, 2022]

Esta abordagem considera apenas os eventos anotados em Baker et al. [2022]. O objetivo da abordagem AA é identificar os eventos na série temporal sob a ótica do aumento da incerteza presente na série temporal da EPU [Baker et al., 2016]. A figura V.1 demonstra pontos onde o aumento de incerteza recebe *labels* que representam eventos significativos marcados entre o período de janeiro de 1991 a maio de 2017.

Esta abordagem considera apenas os eventos anotados em Baker et al. [2022]. O objetivo ao utilizar a abordagem AA é identificar os eventos na série temporal sob a perspectiva do aumento da incerteza presente na série temporal de EPU [Baker et al., 2016]. A figura V.1 retrata os pontos onde o aumento da incerteza recebe *labels* que representam eventos significativos marcados entre o período de janeiro de 1991 a maio de 2017.

V.1.2 Envelop Approach (EVA)

Enquanto na abordagem AA são observados apenas os picos onde há aumento nos termos de incerteza, na segunda abordagem, é criado um envelope sobre os máximos e os mínimos locais presentes na série de EPU. Para criar a EVA é necessário primeiramente obter dois conjuntos de dados: um formado pelo envelope superior com os máximos locais, e outro com os mínimos locais. A partir da criação desses grupos é possível fazer o cálculo dos percentis, esses percentis determinam o ponto de corte que será utilizado para obtenção dos dados presentes na série de *data*

label. Essa série *data label* representa pontos onde há maior propensão de um evento ser detectado na série financeira analisada. Um fato importante nesta abordagem é que ela não apenas observa pontos onde há aumento da incerteza, mas pode identificar áreas onde a diminuição da incerteza também gera eventos. Ao definir os limites superior e inferior com base nos percentis, um novo campo é criado referente aos rótulos nos conjuntos de dados usados. Nesta abordagem, o intervalo é determinado pelos limites percentuais definidos a priori com um intervalo fixo. A Figura V.3 demonstra os limiares de percentis utilizados.

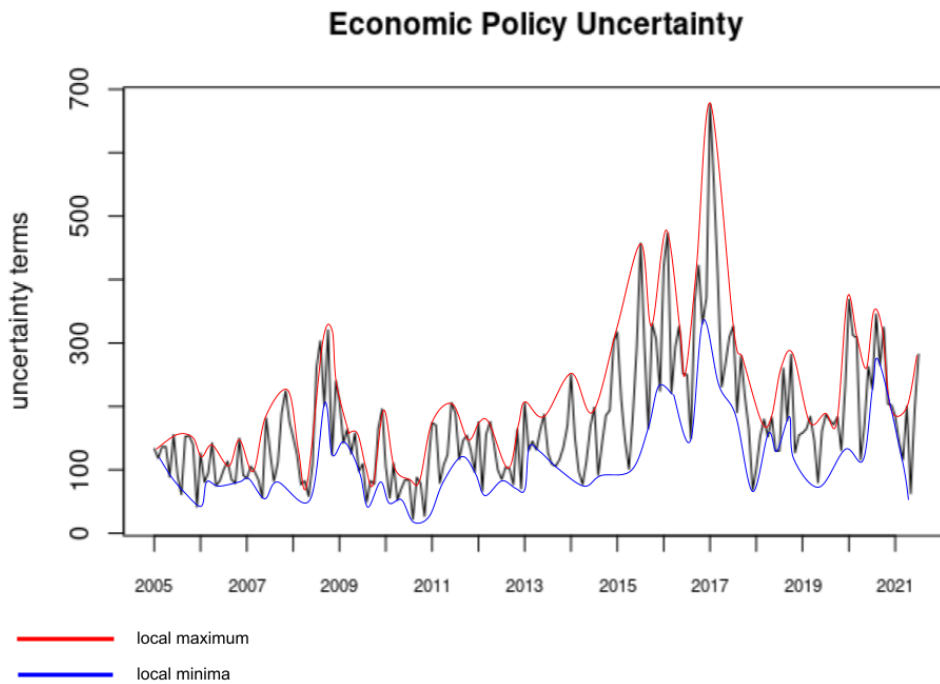


Figura V.2: Série EPU do Brasil com envelope superior e inferior. Fonte: Elaborada pelo autor

A Figura V.2 demonstra os máximos e mínimos locais marcados na série de EPU do Brasil. O mesmo procedimento é feito para as demais séries de EPU utilizadas. Dois conjuntos de dados são gerados a partir dos envelopes que representam os máximos e mínimos locais, e é sobre eles que os percentis são calculados.

Nesta sessão, os experimentos para métodos de detecção de eventos de comparação são organizados de acordo com quatro grupos de séries temporais financeiras: taxas de câmbio para o real brasileiro, índices médios de ações no Brasil, China e EUA. As tabelas representam os resultados dos métodos de detecção de eventos em cada conjunto de dados. As linhas correspondem à análise da série financeira e as colunas representam os métodos utilizados para detectar os eventos. Os métodos usados são FBIAD, GARCH, SCP e MSED. As tabelas são separadas de acordo com a abordagem usada para definir os rótulos de dados. Os resultados incluem a AA e a EVA. Os melhores resultados da aplicação dos métodos de detecção são sublinhados nas respectivas colunas. Por fim, a métrica utilizada neste trabalho é $F1$. A métrica $F1$ foi escolhida porque a série possui

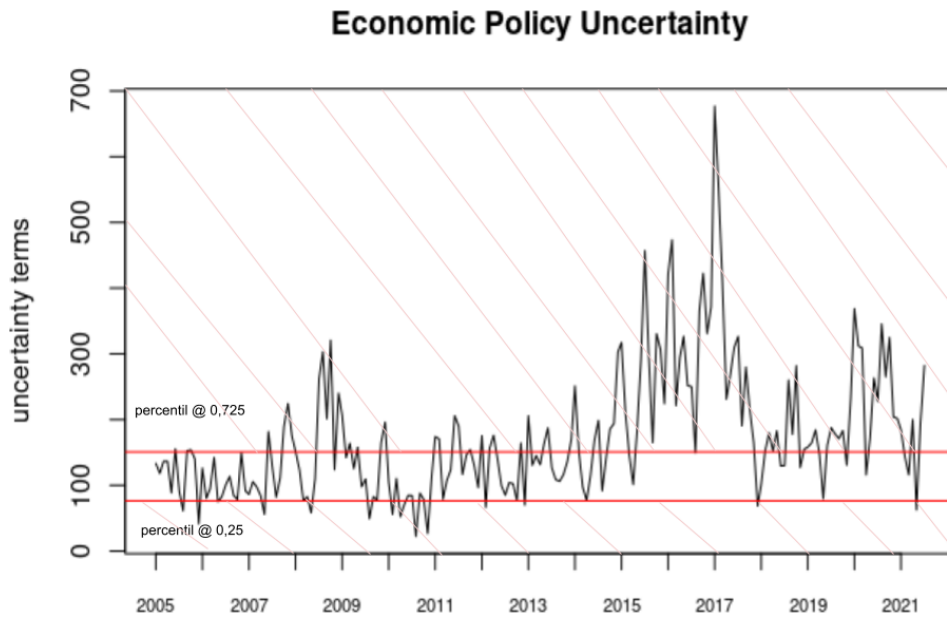


Figura V.3: Série EPU do Brasil com definição de limites baseado no calculo de percentis. Fonte: Elaborada pelo autor

classes desbalanceadas.

V.2 Taxas de câmbio para o real brasileiro

A Tabela V.1 apresenta a pontuação $F1$ dos métodos de detecção de eventos para cada série presente no conjunto de dados da taxa de câmbio para o real brasileiro. Na abordagem AA, a métrica resultante dos métodos de detecção de eventos apresentou que GARCH, FBIAD e SCP tiveram resultados inferiores ao MSED. De fato, o $F1$ score do MSED apresentou uma média com resultado três vezes superior aos demais métodos.

Ao analisar os resultados utilizando os métodos de detecção de eventos na EVA, verificou-se que as métricas tiveram uma melhora nos resultados obtidos em cerca de 12 % em relação a abordagem AA. A análise dos resultados apresentou que o aumento ou diminuição da incerteza neste conjunto de séries gerou mais eventos de pontos de mudança. Assim, os métodos que detectam esse tipo de evento obtiveram os melhores resultados.

V.3 Índices do mercado de ações brasileiro

A Tabela V.2 mostra os resultados de $F1$ referentes ao conjunto de dados dos principais índices de ações do Brasil. O MSED apresentou um resultado geralmente melhor do que os outros métodos, exceto para a série IGC em AA. Os resultados também mostraram que os métodos com capacidade

Tabela V.1: F_1 score para taxas de câmbio

Labels: AA Series	F_1			
	FBIAD	MSED	GARCH	SCP
EUR/BRL	0.29	<u>0.57</u>	0.19	0.32
USD/BRL	0.14	<u>0.58</u>	0.03	0.44
CNY/BRL	0.18	<u>0.54</u>	0.16	0.15
CLP/BRL	0.35	<u>0.47</u>	0.28	0.26
ARS/BRL	0.14	<u>0.52</u>	0.09	0.27
Labels: EVA Series	F_1			
FBIAD	MSED	GARCH	SCP	
EUR/BRL	0.27	<u>0.60</u>	0.23	0.33
USD/BRL	0.10	<u>0.64</u>	0.02	0.42
CNY/BRL	0.18	<u>0.60</u>	0.18	0.21
CLP/BRL	0.33	<u>0.52</u>	0.29	0.31
ARS/BRL	0.13	<u>0.52</u>	0.20	0.30

de detectar pontos de mudança SCP tiveram melhor desempenho do que os outros métodos de detecção de anomalias FBIAD e GARCH.

Tabela V.2: F_1 score para índices médios no mercado de ações brasileiro

Labels: AA Series	F_1			
	FBIAD	MSED	GARCH	SCP
Ibovespa	0.22	<u>0.57</u>	0.23	<u>0.48</u>
Ibrx100	0.03	<u>0.47</u>	0.24	0.45
Ibrx50	0.05	<u>0.48</u>	0.21	0.46
Ibra	0.03	<u>0.58</u>	0.27	0.41
IGC	0.18	0.39	0.26	<u>0.54</u>
Labels: EVA Series	F_1			
FBIAD	MSED	GARCH	SCP	
Ibovespa	0.17	<u>0.60</u>	0.23	0.40
Ibrx100	0.02	<u>0.44</u>	0.22	0.34
Ibrx50	0.03	<u>0.52</u>	0.21	0.36
Ibra	0.02	<u>0.63</u>	0.26	0.36
IGC	0.11	<u>0.55</u>	0.25	0.36

V.4 Índices do mercado de ações chinês

As Tabelas V.3 mostraram resultados do $F1$ para os principais índices de ações chineses. O método GARCH foi comumente o melhor método em AA, enquanto MSED teve melhores resultados em EVA. Isso indica que as mudanças de volatilidade são mais relevantes no conjunto de dados chinês. Além disso, também pode indicar que alguns eventos relevantes não foram completamente anotados na abordagem AA. Ao analisar os resultados utilizando os métodos de detecção de eventos na EVA, verificou-se que a métrica de $F1$ obteve um valor médio 29% superior ao encontrado na abordagem AA utilizando o método de detecção MSED.

Tabela V.3: F_1 score para índices médios no mercado de ações chinês

Labels: AA		F_1			
Series	FBIAD	MSED	GARCH	SCP	
CSI 1000	0.21	0.20	0.35	<u>0.36</u>	
Shanghai	0.25	0.31	<u>0.35</u>	0.34	
SZSE Component	0.19	0.31	<u>0.34</u>	0.32	
China A50	0.07	<u>0.30</u>	0.15	0.11	
HSCE	0.22	0.35	<u>0.37</u>	0.29	
Labels: EVA		F_1			
Series	FBIAD	MSED	GARCH	SCP	
CSI 1000	0.22	<u>0.31</u>	0.22	0.20	
Shanghai	0.38	<u>0.42</u>	0.28	0.27	
SZSE Component	0.19	<u>0.50</u>	0.20	0.31	
China A50	0.04	<u>0.40</u>	0.12	0.06	
HSCE	0.29	<u>0.52</u>	0.24	0.31	

V.5 Índices do mercado de ações americano

A Tabela V.4 apresenta F_1 para os principais índices do mercado de ações americano. Os resultados mostraram maior predominância de anomalias nas séries DJI, Nasdaq e SP500. Tanto no AA quanto no EVA há predominância do MSED. Também é possível observar um equilíbrio entre eventos de volatilidade e pontos de mudança.

Tabela V.4: F_1 score para índices médios no mercado de ações americano

Labels: AA		F_1			
Series	FBIAD	MSED	GARCH	SCP	
DJI	0.33	<u>0.35</u>	0.30	0.28	
Nasdaq	0.31	<u>0.41</u>	0.39	0.29	
S&P 500	0.27	<u>0.38</u>	0.36	0.32	
DJ Composite	0.25	<u>0.40</u>	0.32	0.39	
Small Cap 2000	0.26	<u>0.39</u>	0.32	0.33	
Labels: EVA		F_1			
Series	FBIAD	MSED	GARCH	SCP	
DJI	0.33	<u>0.57</u>	0.30	0.27	
Nasdaq	0.34	<u>0.58</u>	0.34	0.30	
S&P 500	0.30	<u>0.47</u>	0.32	0.31	
DJ Composite	0.22	<u>0.47</u>	0.33	0.33	
Small Cap 2000	0.23	<u>0.56</u>	0.27	0.29	

Resumo de Precisão e Cobertura - A Tabela V.5 demonstra os resultados das métricas de precisão e cobertura para todos os conjuntos de dados. Para a série de taxas de câmbio, o método FBIAD foi geralmente melhor nos resultados de precisão na abordagem AA, mas em EVA, o GARCH foi um pouco superior. O MSED também demonstrou um resultado de cobertura superior aos demais métodos apresentados, sendo possível observar um maior equilíbrio entre a métrica de precisão e cobertura, e por esse motivo, obteve um resultado de $F1$ superior aos demais. Por fim,

os métodos FBIAD e GARCH obtiveram os melhores resultados de precisão, mas apresentaram um baixo resultado de cobertura, o que demonstra incapacidade de detecção de muitos dos eventos presentes nas séries financeiras.

Tabela V.5: Média dos valores de precisão e cobertura nos conjuntos de séries analisados

Labels: AA Series	Precision			
	FBIAD	MSED	GARCH	SCP
Exchange Rate	<u>0.81</u>	0.56	0.76	0.57
Brazilian Stock Indexes	<u>0.97</u>	0.58	0.64	0.68
Chinese Stock Indexes	0.29	0.23	<u>0.39</u>	0.32
US Stock Indexes	0.29	0.29	<u>0.34</u>	0.30
Labels: AA Series	Recall			
	FBIAD	MSED	GARCH	SCP
Exchange Rate	0.13	<u>0.55</u>	0.09	0.20
Brazilian Stock Indexes	0.05	<u>0.47</u>	0.15	0.34
Chinese Stock Indexes	0.16	<u>0.45</u>	0.31	0.27
US Stock Indexes	0.29	<u>0.69</u>	0.34	0.36
Labels: EVA Series	Precision			
	FBIAD	MSED	GARCH	SCP
Exchange Rate	0.83	0.64	<u>0.86</u>	0.72
Brazilian Stock Indexes	<u>0.99</u>	0.70	0.96	0.83
Chinese Stock Indexes	<u>0.90</u>	0.70	0.76	0.67
US Stock Indexes	0.87	0.87	<u>0.98</u>	0.79
Labels: EVA Series	Recall			
	FBIAD	MSED	GARCH	SCP
Exchange Rate	0.12	<u>0.56</u>	0.10	0.24
Brazilian Stock Indexes	0.02	<u>0.44</u>	0.14	0.23
Chinese Stock Indexes	0.22	<u>0.31</u>	0.12	0.14
US Stock Indexes	0.17	<u>0.42</u>	0.18	0.18

O método FBIAD obteve um valor de precisão dominante na abordagem AA e EVA, no conjunto de séries de índices do mercado de ações brasileiro. Entretanto, esse método teve uma baixa taxa de cobertura para o conjunto de séries estudado. O segundo melhor resultado de precisão foi o método GARCH em AA, e depois MSED em EVA. No entanto, por mais que o MSED tenha apresentado o terceiro maior resultado, ele possui uma cobertura muito superior aos demais métodos em AA e EVA. Os resultados obtidos demonstraram que esse conjunto de dados estão de acordo com a série de taxas de câmbio.

Nos conjuntos de séries de índices do mercado de ações chineses, o método GARCH superou outros métodos na métrica de precisão para AA, mas o FBIAD foi melhor em EVA. Além disso, há uma diferença significativa na precisão entre AA e EVA, o que pode indicar que os dados marcados com AA tenham a capacidade de capturar apenas alguns resultados relevantes. Finalmente, o MSED superou todos os métodos em termos de cobertura nas abordagens AA e EVA.

Finalmente, nos conjuntos de séries de índices do mercado de ações americano, o GARCH apresentou métrica de precisão com resultados melhores do que outros métodos para AA e EVA. Nova-

mente, há uma disparidade de magnitude entre AA e EVA. No entanto, MSED superou todos os métodos em termos de cobertura.

V.6 Grupos de eventos gerados por componente Zhang

Além da aplicação das técnicas de detecção nos grupos obtidos através do filtro energético, foi feita uma comparação entre os eventos de incertezas anotados EPU e os encontrados nos métodos de detecção. Para isso, foi desenvolvido um algoritmo que compara cada evento encontrado nos componentes derivados do filtro de tendência, e após isso faz a detecção dos eventos nos diferentes componentes gerados a partir do filtro ERA. Os eventos detectados podem ser divididos em dois tipos, anomalias e pontos de mudança. Uma vez que esses eventos são identificados nos componentes derivados do filtro ERA, podemos dividi-los segundo a presença de eventos com maior dominância. Separando-os não só pelo subtipo de evento, mas em questão da duração e impacto dos eventos relatados. Sendo uma importante ferramenta para análise da série temporal financeira. Grupos de eventos por tipo e presença nos componentes Zhang: (i) *Short Fluctuation Variance Events* (SFVE); (ii) *Shock Significant Change Points* (SSCP); (iii) *Shock Significant Variance Event* (SSVE); (iv) *Long Term Change Points* (LTCP); (v) *Anomaly Detect* (AD). Os resultados de cada grupo de eventos por conjunto de séries temporais se encontra no Apêndice B.

V.7 Discussões

Nessa sessão são feitas as discussões a respeito dos resultados obtidos com base nos objetivos a serem alcançados na pesquisa. Os resultados que obtivemos nos experimentos levantaram duas questões relevantes: (Q1) Quais tipos de eventos são gerados em períodos onde existe aumento ou ausência de incerteza em uma série financeira? (Q2) Quais conjuntos de séries mostraram a maior sensibilidade para influência de eventos externos presentes nos termos de incertezas contido na série de EPU?. Já no conjunto dos principais índices de ações da China e dos EUA, a presença de anomalias de volatilidade foi maior. O método MSED geralmente apresentou melhores resultados no EVA e AA na maioria dos conjuntos de dados, exceto para as séries financeiras chinesas no AA. Esses resultados indicam que a abrangência do método de detecção para encontrar diferentes eventos permite que os mesmos métodos capturem a presença ou ausência de incerteza gerada por fatores externos nas séries financeiras.

A respeito do primeiro questionamento, os métodos aplicados encontraram eventos de diferentes tipos, como: pontos de mudança e anomalias. Os eventos de pontos de mudança demonstraram uma predominância no conjunto de dados de taxas de câmbio e índices do mercado de ações brasileiro, já no conjunto de índices de ações da China e EUA a presença de anomalias de volatilidade foi maior. O método MSED geralmente apresentou melhores resultados no EVA e AA na maioria dos

conjuntos de dados, exceto para as séries financeiras chinesas no AA. Esses resultados demonstraram que métodos de detecção que possuem a capacidade de encontrar eventos distintos, capturam a presença ou ausência de incerteza gerada por eventos externos nas séries financeiras de forma mais assertiva.

Sobre o questionamento Q2, é possível verificar que as métricas demonstraram que o conjunto de séries de cotações de moedas apresentou maior sensibilidade a eventos externos. Os resultados da aplicação dos métodos de detecção de eventos mostraram que o conjunto de dados analisados possui a capacidade de encontrar diversos eventos nas séries analisadas, o que confirma a sensibilidade a eventos externos gerados a partir da variação da presença de incerteza contida na série de EPU.

Finalmente, é importante considerar qual abordagem é utilizada para definição da série de *data label*, dependendo da escolha determinados tipos de eventos podem ser ignorados, criando uma representação forçada da presença de determinados tipos eventos em detrimento de outros, gerando um ruído na análise. A melhor abordagem de série de *data label* é a que não observa somente o aumento da incerteza mas o período onde a ausência dos termos de incerteza. Através dos resultados, é possível observar que a ausência ou a presença de incerteza tem capacidade de gerar eventos nas séries correlacionadas com aquela incerteza. As séries financeiras onde existe pouca correlação com a série de incerteza analisada, tendem a apresentar resultados com valores inferiores nas métricas utilizadas.

Capítulo VI Conclusão

Esta dissertação abordou o impacto da incerteza gerada a partir de eventos externos como crises políticas e econômicas na criação de eventos em diferentes séries temporais financeiras. Os resultados obtidos demonstraram que as séries financeiras são impactadas pelas variações de incertezas presentes nas séries de EPU. Para alcançar esses resultados, foi necessário criar uma correlação entre as séries de EPU e as séries financeiras utilizadas neste trabalho. O propósito deste estudo é buscar séries financeiras onde existe uma maior probabilidade das variações de incertezas estarem refletidas nos eventos encontrados. Finalmente, o presente trabalho demonstrou que o método MSED apresentou um desempenho médio geral bem superior aos outros métodos de detecção, com a capacidade de fazer uma análise do grau de impacto destes eventos no processo de formação dos valores presentes na série.

O trabalho contribui também demonstrando a natureza intrínseca não estacionária e não linear das séries financeiras e como o método de detecção *multi-scale* pode auxiliar nessa análise. Os componentes gerados a partir do uso de MSED possuem diversos tipos de informações que auxiliam no entendimento do processo de formação de uma determinada série financeira. Entender em qual componente existe a maior presença de determinados eventos, auxilia no entendimento de qual é o comportamento das séries financeiras sobre a influência de variações de incertezas geradas por eventos externos, essa característica auxilia no entendimento da sensibilidade a incerteza presente na série financeira, de modo que uma maior presença de eventos em determinados componentes, pode aumentar o risco de investimentos de curto, médio e longo prazo.

A dissertação proposta não tem a capacidade de analisar o efeito de contágio de incertezas geradas por outros países que são parceiros comerciais do país de onde foi obtida a série de EPU. Uma das limitações deste trabalho é identificar o efeito contágio de incerteza de outros países nas séries financeiras estudadas, países que possuem uma forte dependência econômica tendem a ser impactados pela presença de incerteza política e econômica de outros países. Um exemplo claro disso é o da economia americana, onde incertezas políticas e econômicas oriundas dos EUA tendem a gerar reações na bolsa de valores de diversos países ao redor do mundo. A questão é que a incerteza observável através de eventos em séries financeiras, nem sempre vêm unicamente de uma fonte, incertezas geradas por grandes parceiros comerciais de um país, podem gerar eventos em séries financeiras de outro. Quanto maior é a capacidade de análise dessas influências externas,

maior será a capacidade da análise dos resultados dos métodos, o que permitirá que os resultados obtidos apresentem um entendimento mais assertivo do comportamento das séries financeiras.

A extensão dessa pesquisa pode auxiliar o entendimento de problemas que não foram explorados neste trabalho, e que podem ser foco de estudos futuros. Um dos limites desse trabalho é o entendimento do efeito contágio de incerteza gerada por parceiros comerciais de um determinado país, sendo uma boa abertura para estudos futuros, essa alteração implicaria na mudança da abordagem utilizada para definir a série de *data label*, sendo necessário que a nova abordagem criada possa englobar essas informações provenientes de outras séries de EPU.

O trabalho proposto demonstrou a capacidade do método MSED em detectar eventos em séries financeiras, apresentando também a correlação desses eventos com incertezas externas captadas nas séries de EPU. O método MSED provou ser um importante ferramental na análise de séries temporais financeiras. A análise dos componentes oriundos do processo de decomposição permitiu ampliar a visão mais apurada dos processos de formação dos valores presentes na série, o que permitiu uma maior compreensão da dimensão e o tempo de impacto de um determinado evento em uma série financeira.

Apêndice A Eventos EPU

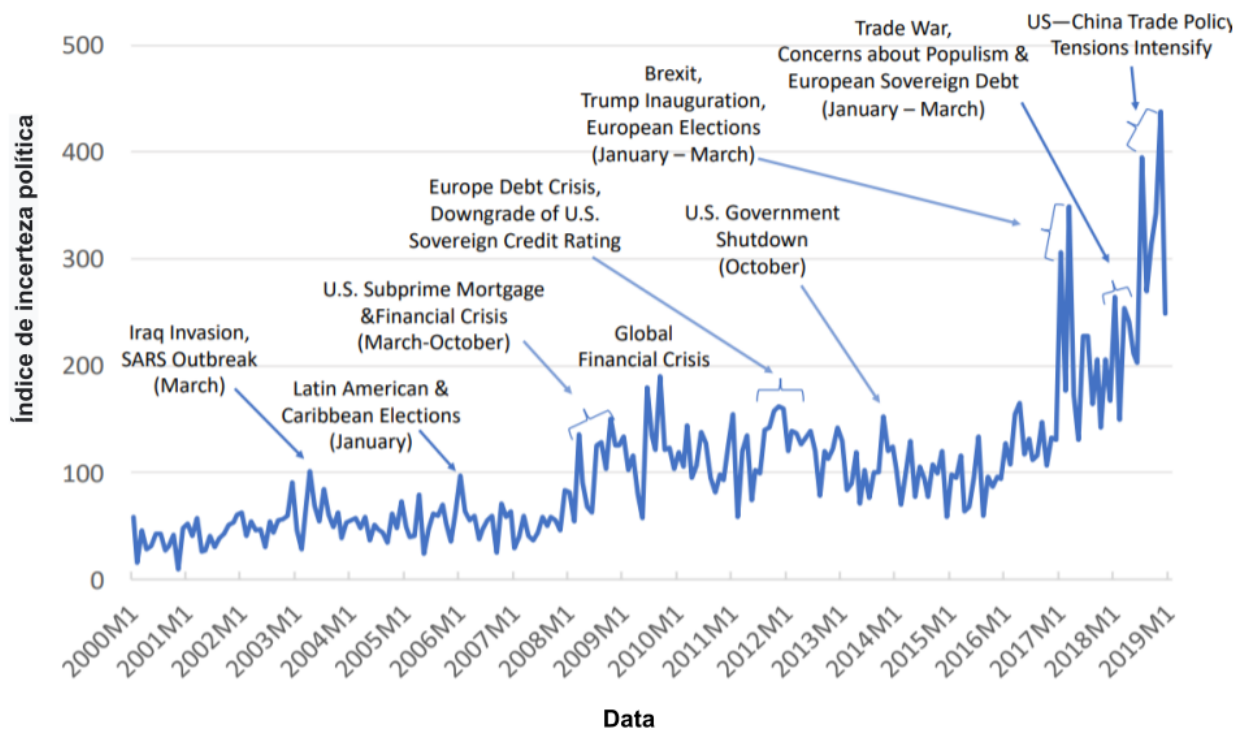


Figura A.1: Índice de Incerteza da Política Econômica para a China

Observação: O índice reflete contagens mensais em escala de artigos contendo pelo menos uma palavra-chave em cada uma das três categorias: (i) Economia (ii) Política Comercial e (iii) Incerteza. A série é normalizada para média 100 de janeiro de 2000 a dezembro de 2018 e com base em consultas executadas em 4 de janeiro de 2019 para *Renmin Daily* e *Guangming Daily*.

...

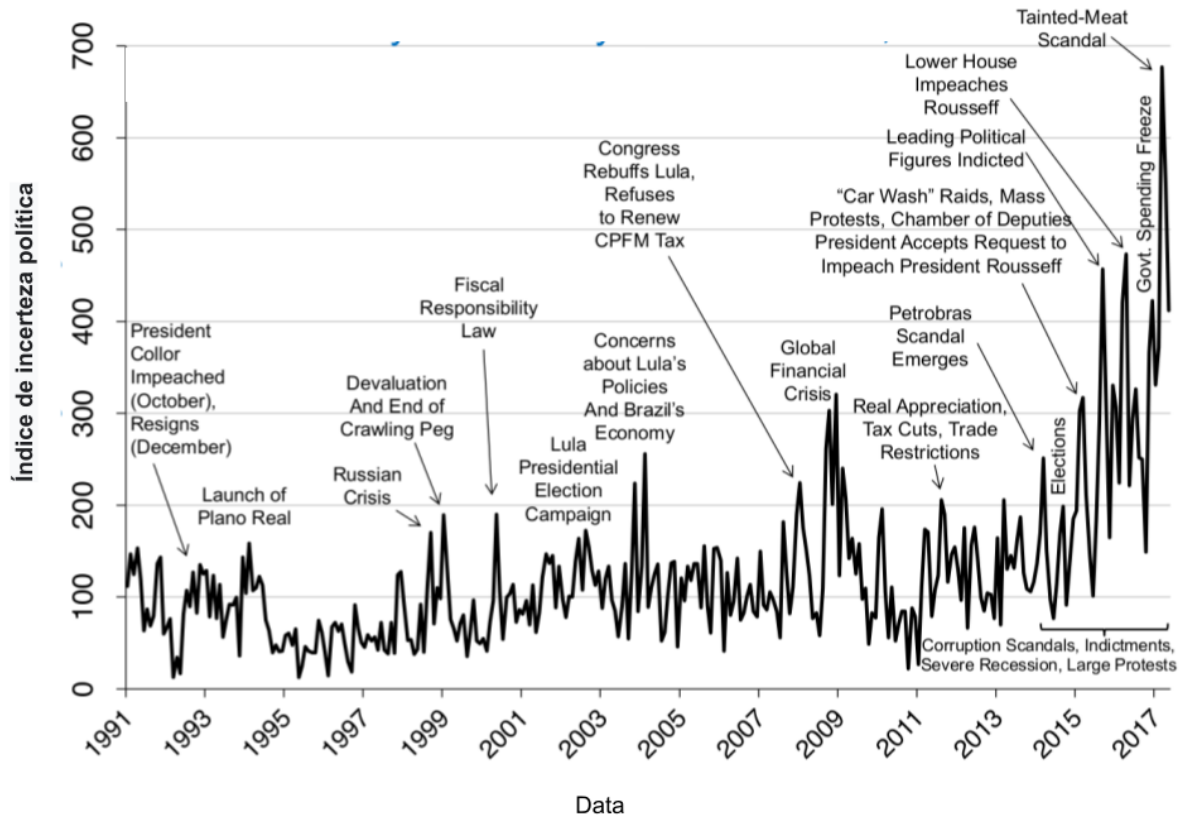
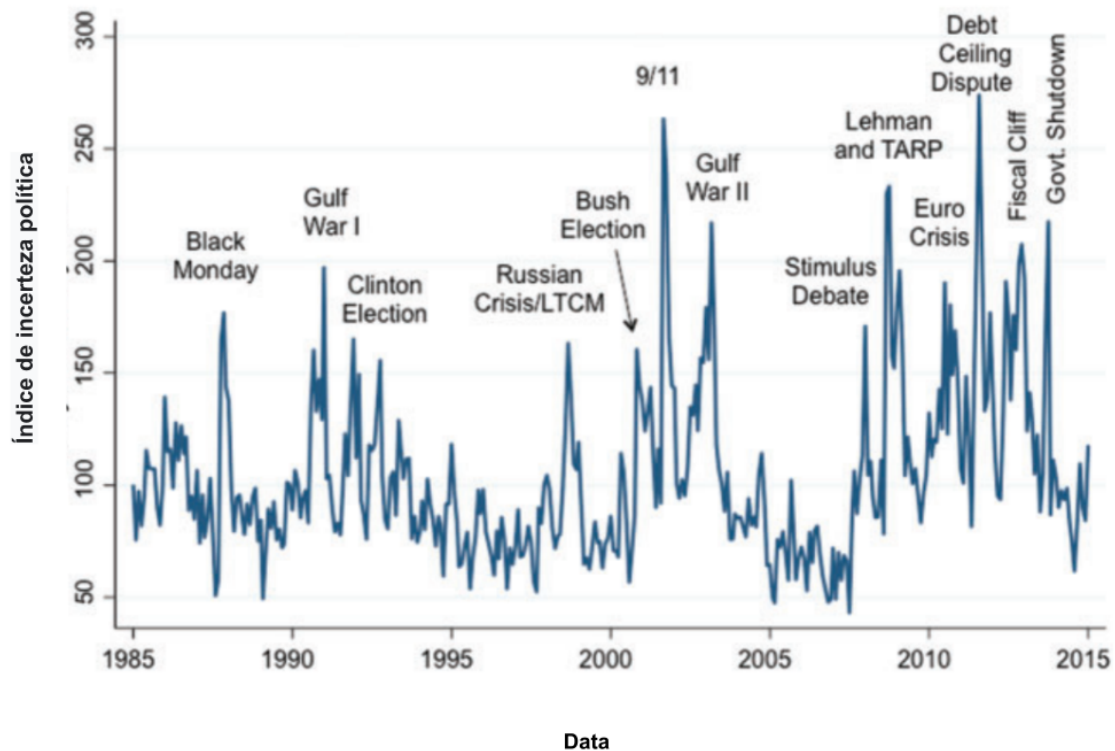


Figura A.2: Índice de Incerteza da Política Econômica para o Brasil

Observação: O índice reflete contagens mensais de artigos na Folha de São Paulo contendo “incerto” ou “incerteza”, “econômico” ou “economia” e um ou mais termos relevantes para políticas que incluem regulação, deficit, orçamentário, imposto, “banco central”, planalto, congresso, senado, legislação e tarifa. Normalizado para uma média de 100 de 1991 a 2011.



Observação: O índice reflete contagens mensais de 10 jornais de grande circulação nos EUA (USA Today, Miami Herald, Chicago Tribune, Washington Post, Los Angeles Times, Boston Globe, San Francisco Chronicle, Dallas Morning News, Houston Chronicle, e WSJ). Para construir o índice, foram realizadas pesquisas mensais em cada jornal com termos relacionados à incerteza econômica e política. Os termos procurados nos artigos dos jornais foram: 'uncertainty' ou 'uncertain', e os termos 'economic' ou 'economy' e um ou mais dos seguintes termos: 'congress', 'legislation', 'white house', 'regulation', 'federal reserve' ou 'déficit'.

Figura A.3: Índice de Incerteza da Política Econômica para os Estados Unidos da América

Apêndice B Grupos de eventos por tipo e presença nos componentes Zhang

Tabela B.1: Grupos de eventos no conjunto de séries de taxa de câmbio

Taxa de câmbio					
Labels: AA	Componentes Zhang				
Séries	SFVE	SSCP	SSVE	LTCP	AD
CNY/BRL	3	20	1	0	22
CLP/BRL	1	9	1	0	4
ARS/BRL	3	21	1	1	13
EUR/BRL	2	18	2	0	14
USD/BRL	0	23	2	0	25

Taxa de câmbio					
Labels: EVA	Componentes Zhang				
Séries	SFVE	SSCP	SSVE	LTCP	AD
CNY/BRL	4	21	2	0	31
CLP/BRL	2	17	3	0	4
ARS/BRL	4	19	2	0	14
EUR/BRL	1	26	3	0	20
USD/BRL	1	18	3	0	20

Tabela B.2: Grupos de eventos no conjunto de séries de índices médios no mercado de ações americano

Índice médio de ações EUA					
Labels: AA	Componentes Zhang				
Séries	SFVE	SSCP	SSVE	LTCP	AD
DJI	1	10	0	0	5
Nasdaq	1	17	0	0	10
S&P500	1	4	1	0	7
DJ Composite	1	8	1	0	6
Small Cap 2000	2	9	1	0	5

Índice médio de ações EUA					
Labels: EVA	Componentes Zhang				
Séries	SFVE	SSCP	SSVE	LTCP	AD
DJI	2	36	2	0	14
Nasdaq	2	39	1	0	30
S&P500	2	18	3	0	19
DJ Composite	1	29	2	0	16
Small Cap 2000	3	34	2	0	18

Tabela B.3: Grupos de eventos no conjunto de séries de índices médios no mercado de ações brasileiro

Índice médio de ações no Brasil					
Labels: AA	Componentes Zhang				
Séries	SFVE	SSCP	SSVE	LTCP	AD
Ibovespa	1	19	0	0	7
Ibrx100	1	15	2	0	13
Ibrx50	2	20	1	0	10
Ibra	2	23	1	0	16
IGC	1	7	2	2	14

Taxas de câmbio					
Labels: EVA	Componentes Zhang				
Séries	SFVE	SSCP	SSVE	LTCP	AD
Ibovespa	1	21	0	0	8
Ibrx100	1	12	2	0	13
Ibrx50	2	22	3	0	11
Ibra	2	25	1	0	15
IGC	1	6	2	4	17

Tabela B.4: Grupos de eventos no conjunto de séries de índices médios no mercado de ações americano chinês

Taxas de câmbio					
Labels: AA	Componentes Zhang				
Séries	SFVE	SSCP	SSVE	LTCP	AD
CSI 1000	1	1	1	0	4
Shanghai	1	7	2	0	6
SZSE Component	2	11	1	0	2
China A50	2	12	1	0	6
HSCE	1	12	1	0	3

Taxas de câmbio					
Labels: EVA	Componentes Zhang				
Séries	SFVE	SSCP	SSVE	LTCP	AD
CSI 1000	2	11	3	0	8
Shanghai	2	6	2	0	19
SZSE Component	3	22	2	0	14
China A50	2	27	4	0	19
HSCE	2	21	2	0	17

Referências

- Indice incerteza epu brasil. https://www.policyuncertainty.com/brazil_monthly.html, 2022. 35
- Aggarwal, C. C. *Outlier Analysis*. Springer Science & Business Media, 2013. 20
- Aloui, C. Price and volatility spillovers between exchange rates and stock indexes for the pre- and post-euro period. *Quantitative Finance*, 7(6):669–685, 2007. 2
- Aminikhanghahi, S. and Cook, D. A survey of methods for time series change point detection. *Knowledge and Information Systems*, 51(2):339–367, 2017. 18, 23
- Bai, Y., Li, Y., Wang, X., Xie, J., and Li, C. Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions. *Atmospheric Pollution Research*, 7(3):557–566, 2016. 8
- Baker, D., Bloom, N., and Davis, S. J. Brazil monthly index: Economic Policy Uncertainty Index. Technical report, https://www.policyuncertainty.com/brazil_monthly.html, 2022. 34, 35
- Baker, S., Bloom, N., and Davis, S. Measuring economic policy uncertainty. *Quarterly Journal of Economics*, 131(4):1593–1636, 2016. 18, 34, 35
- Bernanke, B. S. Irreversibility, uncertainty, and cyclical investment. *Quarterly Journal of Economics vol. 98 iss. 1*, 98, 1983. 17
- Braei, M. and Wagner, S. Anomaly Detection in Univariate Time-series: A Survey on the State-of-the-Art. Technical report, arXiv, 2020. 18
- Carreño, A., Inza, I., and Lozano, J. Analyzing rare event, anomaly, novelty and outlier detection terms under the supervised classification framework. *Artificial Intelligence Review*, 53(5):3575–3594, 2020. 18, 20
- Chandola, V., Banerjee, A., and Kumar, V. Anomaly detection: A survey. *ACM Computing Surveys*, 41(3), 2009. 20, 34
- Chatzikonstanti, V. Breaks and outliers when modelling the volatility of the U.S. stock market. *Applied Economics*, 49(46):4704–4717, 2017. 22, 27

- Chen, D., Lin, J., and Li, Y. Modified complementary ensemble empirical mode decomposition and intrinsic mode functions evaluation index for high-speed train gearbox fault diagnosis. *Journal of Sound and Vibration*, 424:192–207, 2018. 12
- Corrado, C. and Jordan, B. Risk aversion, uncertain information, and market efficiency reexamining the evidence. *Review of Quantitative Finance and Accounting*, 8(1):51–68, 1997. 2, 17
- Downey, A. B. A novel changepoint detection algorithm. Technical report, arXiv, 2008. 18
- Flandrin, P., Gonçalves, P., and Rilling, G. Detrending and denoising with empirical mode decompositions. In *2004 12th European Signal Processing Conference*, pages 1581–1584, 2004a. 5, 27
- Flandrin, P., Rilling, G., and Gonçalves, P. Empirical mode decomposition as a filter bank. *IEEE Signal Processing Letters*, 11(2 PART I):112–114, 2004b. 11, 27
- Golmohammadi, K. and Zaiane, O. Time series contextual anomaly detection for detecting market manipulation in stock market. In *Proceedings of the 2015 IEEE International Conference on Data Science and Advanced Analytics, DSAA 2015*, 2015. 18, 20
- Guralnik, V. and Srivastava, J. Event Detection from Time Series Data. In *Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '99*, pages 33–42, New York, NY, USA. ACM, 1999. 19
- Han, J., Kamber, M., and Pei, J. *Data Mining: Concepts and Techniques*, 2012. , 20, 22, 23, 24
- Huang, N., Long, S., and Shen, Z. The Mechanism for Frequency Downshift in Nonlinear Wave Evolution. *Advances in Applied Mechanics*, 32(C):59–117,117A,117B,117C, 1996. 7
- Huang, N., Shen, Z., Long, S., Wu, M., Snin, H., Zheng, Q., Yen, N.-C., Tung, C., and Liu, H. The empirical mode decomposition and the Hubert spectrum for nonlinear and non-stationary time series analysis. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 454(1971):903–995, 1998. , 2, 6, 7, 8, 9
- Huang, N., Wu, M.-L., Qu, W., Long, S., and Shen, S. Applications of Hilbert-Huang transform to non-stationary financial time series analysis. *Applied Stochastic Models in Business and Industry*, 19(3):245–268, 2003. 2, 4, 6, 11
- Huang, N. E. *Hilbert-Huang Transform and Its Applications*. World Scientific, 2014. , 10
- Hyndman, R. J. and Athanasopoulos, G. *Forecasting: principles and practice*. OTexts, 2018. 5
- Inclán, C. and Tiao, G. Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association*, 89(427):913–923, 1994. 19, 27

- Joo, T. and Kim, S. Time series forecasting based on wavelet filtering. *Expert Systems with Applications*, 42(8):3868–3874, 2015. 8
- Jordan, C. J. C. B. D. Risk aversion, uncertain information, and market efficiency. *Review of Quantitative Finance and Accounting vol. 8 iss. 1*, 8, 1997. 17
- Jun, W., Yuyan, L., Lingyu, T., and Peng, G. A new weighted CEEMDAN-based prediction model: An experimental investigation of decomposition and non-decomposition approaches. *Knowledge-Based Systems*, 160:188–199, 2018. 2, 6, 8
- Lahmiri, S. Interest rate next-day variation prediction based on hybrid feedforward neural network, particle swarm optimization, and multiresolution techniques. *Physica A: Statistical Mechanics and its Applications*, 444:388–396, 2016. 8
- Lau, H. and Yamamoto, S. Bayesian online changepoint detection to improve transparency in human-machine interaction systems. In *Proceedings of the IEEE Conference on Decision and Control*, pages 3572–3577, 2010. 19
- Leung, T. and Zhao, T. Financial Time Series Analysis and Forecasting with HHT Feature Generation and Machine Learning. Technical report, arXiv, 2021a. 2
- Leung, T. and Zhao, T. Financial time series analysis and forecasting with Hilbert–Huang transform feature generation and machine learning. *Applied Stochastic Models in Business and Industry*, 37(6):993–1016, 2021b. 2, 4, 5
- Lima, J., Salles, R., Porto, F., Coutinho, R., Alpis, P., Escobar, L., Pacitti, E., and Ogasawara, E. Forward and Backward Inertial Anomaly Detector: A Novel Time Series Event Detection Method. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2022. 21
- Los, C. *Financial Market Risk: Measurement and Analysis*. Routledge, London; New York, 1 edition, 2006. 7, 8
- Malik, F. Volatility spillover between exchange rate and stock returns under volatility shifts. *Quarterly Review of Economics and Finance*, 80:605–613, 2021. 1, 21
- Malkiel, B. Reflections on the efficient market hypothesis: 30 years later. *Financial Review*, 40(1):1–9, 2005. 1
- Mallat, S. *A Wavelet Tour of Signal Processing: The Sparse Way*. Academic Press, 2008. 7, 8

- Moghtaderi, A., Borgnat, P., and Flandrin, P. Trend filtering: Empirical mode decompositions versus l1 and hodrick-prescott. *Advances in Adaptive Data Analysis*, 3(1-2):41–61, 2011. 5, 14, 16, 27
- Moghtaderi, A., Flandrin, P., and Borgnat, P. Time-varying spectrum estimation of uniformly modulated processes by means of surrogate data and empirical mode decomposition. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pages 3678–3681, 2010. 6, 15, 16, 27
- Nava, N., Di Matteo, T., and Aste, T. Financial time series forecasting using empirical mode decomposition and support vector regression. *Risks*, 6(1), 2018. 4, 5
- Ogasawara, E., Martinez, L., De Oliveira, D., Zimbrão, G., Pappa, G., and Mattoso, M. Adaptive Normalization: A novel data normalization approach for non-stationary time series. In *Proceedings of the International Joint Conference on Neural Networks*, 2010. 21
- Pástor Ľuboš; Veronesi, P. Political uncertainty and risk premia. *Journal of Financial Economics vol. 110 iss. 3*, 110, 2013. 17
- Poon, S.-H. and Granger, C. Forecasting volatility in financial markets: A review. *Journal of Economic Literature*, 41(2):478–539, 2003. 1
- Sakarya, N. and de Jong, R. The spectral analysis of the Hodrick–Prescott filter. *Journal of Time Series Analysis*, 43(3):479–489, 2022. 27
- Salles, R., Belloze, K., Porto, F., Gonzalez, P., and Ogasawara, E. Nonstationary time series transformation methods: An experimental review. *Knowledge-Based Systems*, 164:274–291, 2019. 2, 4, 8, 21
- Samuelson, P. A. Proof that Properly Anticipated Prices Fluctuate Randomly. In *The World Scientific Handbook of Futures Markets*, volume Volume 5 of *World Scientific Handbook in Financial Economics Series*, pages 25–38. WORLD SCIENTIFIC, 2013. 1, 18
- Shumway, R. H. and Stoffer, D. S. *Time Series Analysis and Its Applications: With R Examples*. Springer, 2017. 4, 8
- Vignoli, L. E. G. Análise comparativa de métodos para detecção de eventos em séries temporais. Master’s thesis, Centro Federal de Educação Tecnológica Celso Suckow da Fonseca, CEFET/RJ, 2021. 22

- Wen, F., Yang, X., Gong, X., and Lai, K. Multi-Scale Volatility Feature Analysis and Prediction of Gold Price. *International Journal of Information Technology and Decision Making*, 16(1):205–223, 2017. 1, 2, 4, 7, 12, 17, 26
- Wu, Z. and Huang, N. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, 1(1):1–41, 2009. 11
- Wu, Z., Huang, N., Long, S., and Peng, C.-K. On the trend, detrending, and variability of nonlinear and nonstationary time series. *Proceedings of the National Academy of Sciences of the United States of America*, 104(38):14889–14894, 2007. 6
- Yamada, H. A New Method for Specifying the Tuning Parameter of l1 Trend Filtering. *Studies in Nonlinear Dynamics and Econometrics*, 22(4), 2018. 27
- Yeh, J.-R., Shieh, J.-S., and Huang, N. Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method. *Advances in Adaptive Data Analysis*, 2(2):135–156, 2010. , 12, 14
- Zhang, X., Lai, K., and Wang, S.-Y. A new approach for crude oil price analysis based on Empirical Mode Decomposition. *Energy Economics*, 30(3):905–918, 2008. 1, 5, 17, 26
- Zhang Xiaoyuan; Zhou, J. Multi-fault diagnosis for rolling element bearings based on ensemble empirical mode decomposition and optimized support vector machines. *Mechanical Systems and Signal Processing 2013-dec vol. 41 iss. 1-2*, 41, 2013. , 11, 13
- Zhao, F., Mei, X., Tao, T., Jiang, G., and Zhou, Y. Fault diagnosis of a machine tool rotary axis based on a motor current test and the ensemble empirical mode decomposition method. *ARCHIVE Proceedings of the Institution of Mechanical Engineers Part C Journal of Mechanical Engineering Science 1989-1996 (vols 2011-apr 28 vol. 225 iss. 5*, 225, 2011. , 11, 13