

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ECONOMIA, ADMINISTRAÇÃO E
CONTABILIDADE
DEPARTAMENTO DE CONTABILIDADE E ATUÁRIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CONTROLADORIA E
CONTABILIDADE

RODRIGO SOARES LOPES

**Aplicação de Estratégias de High Frequency
Trading no Mercado Brasileiro de Dólar Futuro**

São Paulo
2018

Prof. Dr. Marco Antonio Zago
Reitor da Universidade de São Paulo

Prof. Dr. Adalberto Américo Fischmann
Diretor da Faculdade de Economia, Administração e Contabilidade

Prof. Dr. Ariovaldo dos Santos
Chefe do Departamento de Contabilidade e Atuária

Prof. Dr. Lucas Ayres Barreira de Campos Barros
Coordenador do Programa de Pós-Graduação em Controladoria e Contabilidade

RODRIGO SOARES LOPES

**Aplicação de Estratégias de High Frequency Trading no Mercado
Brasileiro de Dólar Futuro**

Dissertação apresentada ao Departamento de Contabilidade e Atuária da Faculdade de Economia, Administração e Contabilidade da Universidade de São Paulo como requisito parcial para a obtenção do título de Mestre em Ciências.

Orientador: Prof. Dr. Francisco Henrique Figueiredo de Castro Junior

**São Paulo
2018**

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

FICHA CATALOGRÁFICA

Elaborada por Rafael Mielli Rodrigues – CRB-8/7286
Seção de Processamento Técnico do SBD/FEA/USP

Lopes, Rodrigo Soares

Aplicação de estratégias de high frequency trading no mercado brasileiro de dólar futuro / Rodrigo Soares Lopes. – São Paulo, 2018.
59 p.

Dissertação (Mestrado) – Universidade de São Paulo, 2018.
Orientador: Francisco Henrique Figueiredo de Castro Junior.

1. Econometria 2. Análise de séries temporais 3. Negociação em alta frequência I. Universidade de São Paulo. Faculdade de Economia, Administração e Contabilidade. II. Título.

CDD – 330.015195

À minha família, por tudo.

Agradecimentos

Agradeço a Lara, minha filha, por dar significado a minha vida.

Agradeço aos meus pais, José e Vania, e meu irmão Felipe, pelo incansável apoio ao longo de toda a minha vida e por sempre estarem ao meu lado nos momentos mais felizes e nos de maior necessidade.

A Cristina, Fabiana, Patrícia e Alessandro Falcoski e a Lucas, Gael e Clara Magalhães por me aceitarem em suas famílias.

Agradeço também ao Prof. Henrique Castro, por me aceitar como seu orientado, por me ensinar de forma presente e disponível, por me ouvir e me acalmar em muitos casos. Agradeço à Verônica Santana por toda a paciência e carinho para sanar minhas dúvidas.

Agradeço aos membros da banca de qualificação, Prof. Lucas Ayres, Prof. Mauri Oliveira e Prof. Emerson Marçal, pelas contribuições efetuadas ao meu trabalho de pesquisa.

Aos professores do departamento, em especial ao Prof. Bruno Salotti, Prof. Fernando Murcia, Prof. Nelson Carvalho, Prof. Arioaldo dos Santos e Prof^a Silvia Casa Nova, por todo o aprendizado obtido ao longo do curso.

Aos meus colegas de turma por toda amizade, apoio, discussões acadêmicas e aprendizado mútuo, em especial, Antônio Schumaher, Eduardo Bona, Fernando Trambacos, Marcelo Guzella, Marília Ohlson, Renan Barabanov, Robson Pereira e Taynah Ueda.

Ao meu chefe e amigo, Eduardo Farias, por me sugerir o programa e por me incentivar e apoiar continuamente. Aos meus amigos, porque sem eles sem dúvida não teria conseguido, Ana Beatriz, César Carrasco, César Catanoce, Mariana Konno, Mariane Ceron, Marília Marton, Pérola Vianna e Thiago Peixoto.

Por fim, agradeço à administração da B3 por ter permitido que eu me ausentasse periodicamente para realizar o curso e por disponibilizar as informações utilizadas neste trabalho.

Resumo

Lopes, R. S. (2018). Aplicação de Estratégias de *High Frequency Trading* no Mercado Brasileiro de Dólar Futuro (Dissertação de Mestrado, Universidade de São Paulo, São Paulo).

A pesquisa tem por finalidade avaliar dois modelos econométricos de mudanças de preços, que podem ser utilizados em estratégias de arbitragem estatística, o probit ordenado e o de decomposição, estimando seus parâmetros em quatro pregões de mini contratos de dólar futuro negociados na bolsa de valores brasileira. O estudo da negociação em alta frequência com a utilização de dados de transação a transação revela informações relativas à microestrutura de mercado que o ferramental mais tradicional não é capaz de desvendar. Uma das razões é que modelos tradicionais trabalham com variações de preço como variáveis contínuas, enquanto que ao considerar as variações de preço uma variável contínua e não uma variável discreta, como nos modelos aqui avaliados. Este trabalho acrescenta à literatura sobre microestrutura de mercado ao aplicar os modelos estudados em um ativo distinto daqueles avaliados nos *papers* originais, voltados ao exame do mercado de ações. Esta pesquisa concluiu que os modelos probit ordenado e de decomposição podem ser utilizados para previsão de mini contratos de dólar futuro e que o modelo de decomposição apresenta parâmetros mais significantes. Também concluiu-se que, no modelo probit ordenado, as variáveis de volume e *time duration* não se apresentaram relevantes na determinação do preço desse contrato e que a quantidade de defasagens utilizadas nos parâmetros estimados pode variar dentre os pregões.

Palavras-chave: Econometria. Análise de séries temporais. Negociação em alta frequência.

Abstract

Lopes, R. S. (2018). *Application of High Frequency Trading Strategies in the US Dollar Futures Brazilian Market* (Master's thesis, University of São Paulo, São Paulo).

The research aims to evaluate two econometric models of price change, which can be used in strategies of statistical arbitrage, the ordered probit model and the decomposition model, estimating its parameters in four trading sessions of mini US dollar futures contracts traded on the Brazilian Stock Exchange. The study of high frequency trading with the use of trade-by-trade price movements reveals information related to the market microstructure that the more traditional econometric tools are not able to solve when considering the price changes as a continuous variable and not a discrete one, like in the models evaluated here. This work adds to the literature on market microstructure by applying the models studied in an asset different from those evaluated in the original papers, aimed at examining the stock market. This research concluded that the ordered probit and decomposition models can be used to predict mini US dollar futures price changes and that the decomposition model presents more significant parameters. It was also concluded that, in the ordered probit model, the volume and time duration variables were not relevant in determining the price of this contract and that the number of lags used to estimate parameters can vary among the trading sessions.

Keywords: Econometrics. Time series analysis. High frequency trading.

Sumário

Lista de Tabelas	3
Lista de Figuras	5
1 Introdução	7
2 Revisão de Literatura	11
2.1 Qualidade do Mercado	11
2.2 Estratégias de Negociação	14
2.3 Características Empíricas dos Dados de Alta Frequência	19
3 Metodologia	23
3.1 Dados e Amostra	23
3.2 Modelos	24
3.2.1 Modelo Probit Ordenado	24
3.2.2 Modelo de Decomposição	26
4 Resultados	31
4.1 Estatísticas Descritivas	31
4.2 Estimação de Parâmetros	36
4.2.1 Modelo Probit Ordenado	36
4.2.2 Modelo de Decomposição	40
5 Conclusão	43
Referências	45

Lista de Tabelas

4.1	Distribuição das mudanças de preços	33
4.2	Estatísticas descritivas dos preços	35
4.3	Resultados da estimação dos parâmetros das variáveis explicativas do modelo probit ordenado	37
4.4	Resultados da estimação dos interceptos do modelo probit ordenado .	39
4.5	Resultados da estimação do modelo de decomposição	40

Lista de Figuras

4.1	Gráficos da variação de preços no pregão	32
4.2	Histogramas das variações de preços	32
4.3	Gráficos das quantidades de contratos negociados	33
4.4	Gráficos da variação de tempo entre os negócios	34
4.5	Gráficos das mudanças de preços	35
4.6	Gráfico das probabilidades de mudanças de preços	38

1 Introdução

O avanço tecnológico promoveu a utilização de sistemas informatizados para automação de estratégias de investimento a um novo patamar. Esse processo desencadeou o surgimento de uma nova modalidade de negociação conhecida como negociação em alta frequência (*High Frequency Trading*).

O conceito de negociação em alta frequência é usualmente associado à negociação com a utilização de algoritmos. No entanto, enquanto a negociação em alta frequência é um tipo de negociação com algoritmos, nem toda a forma de negociação com algoritmos pode ser descrita como de alta frequência. A negociação com algoritmos antecede a de alta frequência e tem sido amplamente utilizada como uma ferramenta para determinar alguns aspectos da execução dos negócios como tempo, quantidade, preço e plataforma de negociação (Gomber, Arndt, Lutat, & Uhle, 2011).

Assim como no caso da negociação em alta frequência, inexiste uma definição universalmente aceita para negociação com algoritmos. Esse é o termo genérico utilizado para todas as estratégias em que ofertas são submetidas às plataformas de negociação de acordo com um conjunto pré-programado de regras (algoritmos). Uma dentre as diversas definições é formulada por Cartea e Penalva (2012, p. 2):

“We prefer to use algorithmic trading (AT) as the generic term that refers to strategies that use computers to automate trading decisions, and restrict the term high frequency (HF) trading to refer to the subset of AT trading strategies that are characterized by their reliance on speed differences relative to other traders to make profits based on short-term predictions and also by the objective to hold essentially no asset inventories for more than a very short period of time¹. ”

¹Livre tradução: “Preferimos utilizar negociação com algoritmos (AT), como um termo genérico referente às estratégias que utilizam computadores para automatizar decisões de investimento, e restringir o termo negociação em alta frequência (HF) para se referir ao subconjunto de estratégias

SEC and CFTC (2010, p. 45) definem a negociação em alta frequência com base em suas características da seguinte forma:

“HFTs are proprietary trading firms that use high speed systems to monitor market data and submit large numbers of orders to the markets. HFTs utilize quantitative and algorithmic methodologies to maximize the speed of their market access and trading strategies². ”

Segundo Gomber et al. (2011), a negociação em alta frequência é uma forma de negociação automatizada baseada em algoritmos matemáticos, não sendo uma estratégia de negociação em si, mas um meio de aplicação de determinadas estratégias em plataformas de negociação. Como pré-requisito, a negociação em alta frequência conta com acesso ao mercado em alta velocidade, ou seja, baixas latências; o uso de serviços de acesso direto ao mercado, conhecidos como *co-location*; a manutenção de ativos em custódia por curtos períodos, reduzindo o risco global da carteira; e uma elevada taxa de ofertas submetidas por transações executadas, ou seja, uma elevada taxa de cancelamento.

Investidores que realizam negociação em alta frequência o fazem por meio da aplicação de uma estratégia ou um conjunto destas. Segundo Durbin (2010), investidores de alta frequência, ou HFTs, como comumente são chamados em referência aos *High Frequency Traders*, fazem uso de estratégias de formação de mercado, arbitragem e arbitragem estatística. Esta representando aquelas nas quais investidores mantêm posições não cobertas carregadas por um período de tempo, normalmente curto. Dentre as estratégias de arbitragem estatística, encontram-se aquelas que empregam modelos de previsão de mudanças de preços esperados, com base na análise de padrões passados, conforme Aldridge (2009).

A pesquisa tem por finalidade avaliar dois modelos econométricos de mudança de preços, que podem ser utilizados em estratégias de arbitragem estatística, o probit ordenado, formulado por Hausman, Lo, e MacKinlay (1992), e o de decomposição, proposto por Rydberg e Shephard (2003), estimando seus parâmetros em quatro pregões de mini contratos de dólar futuro negociados na bolsa de valores brasileira.

de AT caracterizadas por sua dependência em diferenças de velocidade em relação a outros investidores para lucrar com base em previsões de curto prazo e também por não manter um inventário de ativos por um período de tempo muito longo.”

²Livre tradução: “HFTs são empresas de investimento proprietário que usam sistemas de alta velocidade para monitorar dados de mercado e submeter grandes quantidades de ordens aos mercados. Os HFTs utilizam metodologias quantitativas e algorítmicas para maximizar a velocidade de seu acesso ao mercado e estratégias de negociação.”

O estudo da negociação em alta frequência com a utilização de dados de transação a transação revela informações relativas à microestrutura de mercado que o fissional mais tradicional não é capaz de desvendar. Uma das razões é que modelos tradicionais trabalham com variações de preço como variáveis contínuas, enquanto que ao considerar as variações de preço uma variável contínua e não uma variável discreta, como nos modelos aqui avaliados. Este trabalho acrescenta à literatura sobre microestrutura de mercado ao aplicar os modelos estudados em um ativo distinto daqueles avaliados nos *papers* originais, voltados ao exame do mercado de ações.

Esta pesquisa concluiu que os modelos probit ordenado e de decomposição podem ser utilizados para previsão de mini contratos de dólar futuro e que o modelo de decomposição apresenta parâmetros mais significantes. Também concluiu-se que as variáveis de volume e *time duration* não se apresentaram relevantes na determinação do preço desse contrato.

No capítulo 2, é realizada uma revisão da literatura relacionada à negociação em alta frequência, das estratégias de negociação utilizadas por esse tipo de investidor e das características desses dados. A amostra é apresentada no capítulo 3, assim como uma revisão da especificação dos modelos probit ordenado e de decomposição. Os dados são descritos no capítulo 4 com a apresentação detalhada de estatísticas descritivas, além de mostrar o resultado da estimativa dos modelos. As conclusões são apresentadas no capítulo 5.

2 Revisão de Literatura

2.1 Qualidade do Mercado

Uma parte da produção teórica relacionada à negociação em alta frequência diz respeito a análise de seus efeitos quanto à qualidade do mercado, em muito influenciada pelos eventos do *flash crash* de maio de 2010 (Kirilenko, Kyle, Samadi, & Tuzun, 2011), explicado em detalhes em SEC and CFTC (2010). Essas pesquisas representam contribuições empíricas na análise da relação entre a negociação de alta frequência e medidas de qualidade de mercado, notadamente volatilidade, liquidez e formação de preços.

Brogaard et al. (2010) estudam o impacto dos investidores de HFT no mercado de ações, avaliando suas estratégias, rentabilidade e relação com medidas de qualidade do mercado, incluindo liquidez, formação de preços e volatilidade. Brogaard et al. (2010) identificam que a negociação dos investidores de alta frequência contida em sua amostra representa um grande percentual do mercado e conclui que esses investidores proveêm liquidez, assim como a tomam, e que sua atividade tende a ser estável ao longo do tempo. Os investidores de alta frequência geralmente utilizam estratégias de reversão de preços, comprando após declínios de preços e vendendo após elevações de preços. São lucrativos, gerando cerca de US\$ 3 bilhões por ano em um volume de US\$ 30 trilhões de dólares negociados. E também conclui que investidores de HFT aumentam a liquidez do mercado, aumentam a eficiência dos mercados, melhorando o processo de formação de preços e tendem a diminuir a volatilidade.

Zhang (2010) examina o impacto do HFT no mercado de capitais americano, usando uma amostra de empresas do banco de dados CRSP e Thomson Reuters Institutional Holdings durante 1985-2009. Zhang (2010) estima que a negociação em alta frequência foi responsável por cerca de 78% do volume negociado em 2009, em relação

a quase zero em 1995, concluindo que o volume de negociação em alta frequência aumenta a volatilidade dos preços das ações. Especificamente, a volatilidade do preço das ações está positivamente correlacionada com a negociação em alta frequência após controlar a volatilidade dos fundamentos da empresa e outros determinantes de volatilidade exógena. A correlação positiva é mais forte entre as 3.000 maiores ações em capitalização de mercado e entre ações com altas participações institucionais. A correlação positiva também é mais forte durante períodos de alta incerteza no mercado. Além disso, conclui também que a negociação de alta frequência está negativamente relacionada à capacidade do mercado de incorporar informações sobre os fundamentos da empresa nos preços dos ativos. Os preços das ações tendem a reagir de forma exagerada às notícias quando o volume de negociação em alta frequência é elevado.

Hendershott, Jones, e Menkveld (2011) examinam o crescimento da negociação com algoritmos e as melhorias na liquidez ao longo de um período de 5 anos, a partir da constatação de que à medida que a negociação com algoritmos cresce, a liquidez melhora. No entanto, enquanto a negociação com algoritmos e a liquidez se movem na mesma direção, é possível que o relacionamento não seja causal. Para estabelecer a relação de causalidade, Hendershott et al. (2011) estudam um evento exógeno que aumenta a quantidade de negociação com algoritmos em algumas ações mas não em outras. Em particular, usam o início da cotação automatizada na NYSE como um instrumento para a negociação com algoritmos e concluem que a negociação com algoritmos, de fato, melhora a liquidez para ações com grande participação no mercado de capitais. Os *spreads* efetivos são estreitados com a cotação automatizada. Os *spreads* mais estreitos são o resultado de uma queda acentuada na seleção adversa, ou equivalentemente uma diminuição na formação de preços associada às transações. A negociação com algoritmos aumenta a formação de preços que ocorre sem negociação, o que implica que as cotações se tornam mais informativas.

Brogaard, Hendershott, e Riordan (2014) examinam o papel de investidores de alta frequência no processo de formação de preços usando um conjunto de dados da NASDAQ que identifica se as ordens tomadoras ou provedoras de liquidez de cada negócio são provenientes de investidores HFT, para uma amostra de ações em 2008 e 2009. Brogaard et al. (2014) mostram que esses investidores desempenham um papel benéfico na formação eficiente de preços ao negociar na direção oposta a erros de precificação transitórios, também referidos como volatilidade transitória ou ruído, e na mesma direção de movimentos de preços futuros eficientes e que isso é executado por meio de ordens tomadoras de liquidez. Em contraste, as ordens

provedoras de liquidez de investidores de alta frequência impõem custos de seleção adversa a outros investidores. Além disso, identificam que os investidores HFT fornecem liquidez em momentos de maior estresse no mercado, como dias de maior volatilidade e anúncios de notícias macroeconômicas. A direção da negociação de desses investidores prevê mudanças de preços em horizontes curtos, medidos em segundos, e está correlacionada com informações públicas, como anúncios de notícias macroeconômicas, movimentos de preços gerais do mercado e desequilíbrios nos livros de ordens.

Chaboud, Chiquoine, Hjalmarsson, e Vega (2014) estudam o impacto da negociação com algoritmos no mercado de câmbio usando um conjunto de dados que consiste na negociação minuto a minuto de setembro de 2003 a dezembro de 2007 obtidos do *Electronic Broking Services* (EBS) em três pares de moedas: o euro-dólar, o dólar-iene e o euro-iene, que identifica especificamente a atividade de negociação gerada por computadores. Chaboud et al. (2014) encontram evidências de que a negociação com algoritmos causa uma melhora em duas medidas de eficiência de preços neste mercado: a redução da frequência das oportunidades de arbitragem triangular entre os pares de moedas, o que demonstra uma melhor transmissão da informação aos preços, e a redução da autocorrelação de retornos de alta frequência, o que contribui para a diminuição temporária de desvios de preços de seus valores fundamentais, reduzindo portanto a volatilidade. Eles mostram que a redução das oportunidades de arbitragem está associada principalmente aos negócios realizados por algoritmos que tomam liquidez, enquanto a redução na autocorrelação dos retornos se deve mais à provisão de liquidez por algoritmos. Eles também acham evidências de que os negócios realizados por algoritmos e suas estratégias são mais correlacionados do que aqueles de investidores não-algorítmicos, potencialmente criando volatilidade excessiva. No entanto, a análise mostra que esse alto grau de correlação não parece causar uma degradação na qualidade do mercado.

Hruska e Linnertova (2015) analisam o impacto da negociação em alta frequência na liquidez das 22 ações mais negociadas na *London Stock Exchange* com observações minuto a minuto no período de 16 de setembro de 2014 a 30 de março de 2015. O modelo utilizado apresentou resultados mistos e foram influenciados pela medida de volatilidade utilizada. Embora os resultados mistos, o efeito da negociação em alta frequência na liquidez do mercado foi positivo. O motivo desses resultados mistos pode ser causado pelo uso de *proxies* para mensuração de liquidez devido à informação pública limitada sobre o mercado analisado.

Li, Cooper, e Van Vliet (2017), usando dados do livro de ofertas da NASDAQ, inves-

tigam como a negociação de alta frequência afeta a negociação de baixa frequência, categorizando ordens de ambos os tipos e examinando diversas medidas. Eles encontram evidências de que um aumento na negociação em alta frequência está associada a um aumento na frequência e no tamanho dos negócios das ordens limitadas de baixa frequência, o que implica que a atividade de HFT melhora a liquidez das ordens de baixa frequência. Eles também identificam que a negociação em alta frequência melhora a qualidade da execução das ordens, bem como a liquidez das ordens de baixa frequência. Um aumento na atividade de HFT reduz o tempo de espera para cancelamento ou execução, e melhora a probabilidade de execução das ordens limitadas de baixa frequência. Além disso, eles descobrem evidências de que o HFT contribui para a provisão de liquidez dos investidores de baixa frequência.

2.2 Estratégias de Negociação

Outra parte da produção acadêmica relacionada à negociação em alta frequência se dedica a estudar modelos econôméticos para implementar estratégias de negociação passíveis de aplicação em algoritmos. As estratégias de negociação em alta frequência podem ser definidas como um subconjunto das estratégias que utilizam algoritmos, que operando em alta frequência ou não, empregam diversos tipos de estratégias de negociação. Durbin (2010) agrupa as estratégias que utilizam negociação com algoritmos em três categorias: (i) formação de mercado, que se refere à estratégia de submeter simultaneamente uma ordem limitada de compra e uma de venda; (ii) arbitragem, uma estratégia sem riscos que explora erros de precificação de ativos; e (iii) arbitragem estatística, que envolve posições não cobertas carregadas por um determinado período de tempo, em antecipação a pequenas mudanças de preço intradiárias.

Autores como Johnson (2010) e Gomber et al. (2011) elencam uma quarta categoria de negociação por algoritmos, que não utilizaria alta frequência em si, com foco na otimização de execução de ordens, visando reduzir o impacto na formação de preços proveniente da execução de grandes ordens. Os algoritmos constantes dessa quarta categoria fatiam a ordem em diversas ordens menores executando-as ao longo do tempo (ou plataformas de negociação) de acordo com um modelo pré-estabelecido, chamado *benchmark*. Segundo Aldridge (2009), os modelos mais populares de *benchmark* são os *Time Weighted Average Price* (TWAP), que dividem uma ordem grande em fatias enviadas ao mercado em intervalos de tempo igualmente distribuídos, e os

Volume Weighted Average Price (VWAP), que tentam igualar ou superar o preço médio ponderado pelo volume ao longo de um período de tempo específico.

Uma série de modelos estatísticos para otimização de execução de ordens é apresentada em Cartea, Jaimungal, e Penalva (2015). Modelos específicos de TWAP e VWAP podem ser encontrados em Cartea e Jaimungal (2016), Guéant e Royer (2014) Konishi (2002) e Mitchell, Bialkowski, e Tompaidis (2013).

Estratégias de formação de mercado representam uma forma clássica de estratégia de negociação na qual técnicas de alta frequência podem ser empregadas. A formação de mercado é uma estratégia de provisão de liquidez em instrumentos listados. Na formação de mercado com a utilização de alta frequência, algoritmos calculam em alta velocidade os preços (compra e venda) em que um instrumento pode ser negociado. Modelos de formação de mercado são especificados em Avellaneda e Stoikov (2008), Carmona e Webster (2012), Cartea e Jaimungal (2015), Cartea, Jaimungal, e Ricci (2014), Cartea e Penalva (2012), Grossman (1976), Grossman e Miller (1988), Ho e Stoll (1981) e Kyle (1989).

Conforme Gomber et al. (2011), provedores de liquidez com negociação em alta frequência têm duas fontes básicas de receita: (i) ganham o *spread* entre ofertas de compra e venda; e (ii) recebem devoluções (rebates) ou reduções de taxas das plataformas de negociação como incentivo à provisão de liquidez.

Outro tipo de estratégia refere-se à arbitragem, cuja oportunidade frequentemente existe por períodos breves. Segundo Durbin (2010), uma vez que computadores são capazes de varrer os mercados para identificar tais possibilidades de curta duração, a arbitragem tornou-se uma importante estratégia aplicada por investidores que negociam em alta frequência. Esses investidores conduzem arbitragem da mesma forma que os seus homólogos tradicionais, utilizando tecnologia para lucrar com pequenas discrepâncias, de curta duração, entre ativos. As estratégias de arbitragem não estão limitadas aos investidores de alta frequência e são conduzidas pelos participantes não automatizados do mercado também. Uma vez que arbitradores reagem às ineficiências do mercado, eles são principalmente tomadores de liquidez e contribuem para a adequada formação de preços nos mercados em que atuam.

A arbitragem estatística envolve posições não cobertas carregadas por um determinado período de tempo, muitas vezes curto, em antecipação a pequenas mudanças de preços intradiárias, conforme Aldridge (2009). Em contraste com outras estratégias de alta frequência esses participantes não estão interessados em fornecer liquidez,

como formadores de mercado, nem explorar ineficiências de mercado, como arbitradores clássicos, eles costumam negociar tomando liquidez e visam lucrar com movimentos e tendências de mercado. Baseados em padrões passados, eles estimam mudanças nos preços esperados identificadas com base em sinais como movimentação dos preços de mercado, *spreads* e volume.

A estratégia de *pairs trading*, um tipo de arbitragem estatística, é realizada com o objetivo de ser neutra em relação ao mercado. Nesse tipo de estratégia, os arbitradores mantêm-se comprados em determinado instrumento enquanto se mantêm vendidos em outro. Uma vez que esses instrumentos estão estreitamente correlacionados (como por exemplo, ações de companhias do mesmo setor), ganhos e perdas, devido ao movimento geral do mercado, irão se compensar. Contudo, a fim de obter ganhos a partir dessa estratégia, arbitradores vendem um instrumento que consideram ter um valor intrínseco relativamente baixo e, simultaneamente, compram um instrumento, que reage de forma muito semelhante (idealmente idêntica) a alterações no ambiente de mercado, e que consideram ter um valor intrínseco relativamente mais elevado. Se as respectivas avaliações destes instrumentos se normalizarem na direção esperada, o arbitrador liquida sua posição neutra com relação ao mercado. Os ganhos provenientes da aplicação dessa estratégia resultam da diferença entre a avaliação individual dos ativos, no momento em que a posição é aberta, e os preços normalizados desses ativos, no momento em que a posição é encerrada. Modelos econométricos de estratégias de pairs trading são especificados em Elliott, Hoek, e Malcolm (2005), Leung e Li (2015), Leung e Ludkovski (2011) e Tourin e Yan (2013).

Outra forma de arbitragem estatística refere-se à construção de modelos de análise de mudanças de preços. Campbell, Lo, e MacKinlay (1997) descrevem o modelo de arredondamento proposto em Ball (1988), Cho e Frees (1988), Gottlieb e Kallay (1985) e Harris (1991), no qual fazem uma aproximação do preço discreto P_t^0 , observável, a partir do processo de preço “verdadeiro” de estado contínuo, não observável, P_t por meio do arredondamento de P_t em múltiplos de d , representando o incremento mínimo de variação de preços, ou seja, o *tick*. Campbell et al. (1997) também descrevem um segundo modelo de aproximação de preços discretos, suavemente diferente mas relacionado ao modelo anterior, proposto por Cho e Frees (1988) e Marsh e Rosenfeld (1986), conhecido como modelo de barreira, no qual P_t também é um processo de tempo contínuo e os negócios são observados sempre que P_t atinge um determinado nível ou barreira, estabelecida em múltiplos do *tick*.

Hausman et al. (1992) formulam o modelo probit ordenado levando em consideração

o fato de que mudanças de preço ocorrem seguindo um modelo de variável aleatória discreta e não contínua como nos modelos de arredondamento e de barreira, sendo o modelo probit ordenado considerado uma generalização destes sob determinadas restrições. Importante ressaltar que no período em que escreveram seu trabalho o tamanho do *tick* nas bolsas de valores norte-americanas era de 1/8 de dólar, o que tornava sua proposição mais evidente. Além disso o intervalo de execução dos negócios ocorre em espaços de tempo irregulares, dificultando a utilização de modelos para variáveis aleatórias contínuas. Adicionalmente, sua proposta de um modelo probit ordenado permite a consideração dos efeitos de outras variáveis econômicas nas mudanças de preços como volume, valores históricos defasados de preços e o intervalo entre negócios executados.

Hausman et al. (1992) utilizam dados de 100 ações americanas negociadas em 1988 para analisar três questões de pesquisa: primeiro, como a distribuição condicional de mudanças de preços é afetada pela sequência de negócios. Segundo, as mudanças de preços são afetadas pelo tamanho dos negócios, e caso afirmativo, qual o impacto sobre os preços. Terceiro, averiguar se é relevante considerar a mudança de preços como uma variável discreta, e ainda, se uma regressão linear simples pode ser utilizada para modelar a distribuição condicional das mudanças de preços. Para responder a estas perguntas Hausman et al. (1992) utilizam estimativas de máxima verossimilhança em três aplicações: (i) testando a dependência do fluxo de ordens; (ii) medindo o impacto nos preços; e (iii) comparando o probit ordenado com a regressão linear simples. O artigo conclui que a sequência de negócios afeta a distribuição condicional das mudanças de preços; o tamanho dos negócios é um importante fator na distribuição condicional das mudanças de preços e que o impacto dos negócios sobre os preços depende da sequência passada das mudanças de preço e do fluxo de ordens; e considerar a mudança de preços como uma variável aleatória discreta é importante no sentido de que uma regressão linear simples não consegue capturar todas as características das mudanças de preço.

Assim como Hausman et al. (1992), Russell e Engle (1998) também propõem um modelo de análise que considera a variação de preços como uma variável aleatória discreta. A pesquisa trata os tempos de chegada como um processo de ponto marcado¹, considerando a modelagem conjunta dos tempos de chegada e das mudanças de preços como uma função de variáveis predeterminadas ou fracamente exógenas.

¹O processo de ponto é a representação matemática de um conjunto de pontos localizados aleatoriamente em um plano matemático subjacente, como por exemplo os tempos de chegada das mudanças de preços. Um processo de ponto marcado é um processo de ponto que possui uma variável aleatória auxiliar (chamada de marca) associada a cada ponto de evento, como a modelagem conjunta dos tempos de chegada e das mudanças de preços. Vide Monteiro (2008)

Os autores propõem a decomposição da distribuição conjunta de mudanças de preços e tempos de chegada no produto da distribuição condicional das mudanças de preços e da distribuição marginal dos tempos de chegada. Russell e Engle (1998) desenvolvem o modelo de Duração Condicional Autorregressiva (DCA) para modelar a distribuição marginal dos tempos de chegada e sugerem a utilização de uma nova classe de modelo de séries temporais multinomial para analisar as mudanças de preços. Esse modelo é chamado de *Autoregressive Conditional Multinomial* (ACM) porque depende da distribuição histórica dos dados, bem como das realizações passadas.

Já o sistema proposto inicialmente em Rydberg e Shephard (2003) introduz um modelo de decomposição de mudanças de preços constituindo uma base econométrica para estudos empíricos em microestrutura de mercado com a utilização de séries temporais. O modelo proposto decompõe a variação de preços em três componentes: atividade, medindo se o preço mudou; direção, medindo se a variação foi positiva ou negativa; e tamanho, capturando a magnitude da mudança de preços; afirmado que essa seria uma estrutura econométrica mais simples para a análise de mudanças sequenciais de preços.

A pesquisa utiliza dados de negociação de IBM em 1995, para estimar o modelo, considerando uma utilização crucial da estrutura proposta a produção de previsões dos movimentos dos preços dos ativos em múltiplos passos. Isso pode ser expresso de duas maneiras básicas: (i) previsões dos movimentos dos preços ($p + 1$) períodos futuros²; e (ii) previsões dos níveis de preço dos ativos em ($p + 1$) períodos à frente.

Rydberg e Shephard (2003) também realizam uma comparação do modelo proposto com aqueles formulados em Hausman et al. (1992) e Russell e Engle (1998) concluindo que, no geral, suas análises estão bastante relacionadas e que os objetivos das pesquisas seriam os mesmos, embora as ferramentas utilizadas sejam muito diferentes. As principais vantagens do modelo proposto em Rydberg e Shephard (2003), segundo os autores, seriam: parcimônia, fácil interpretação através da decomposição e opções para extensões.

McCulloch e Tsay (2001) aplicam um modelo reduzido de Rydberg e Shephard (2003) ao considerar somente os negócios que possuem mudanças de preço, elencando uma variável adicional para denotar os negócios sem mudanças de preço com o objetivo de simplificar a análise. Os autores consideram que modelos de *threshold duration*

para uma definição matemática de processos de ponto marcado e Russell (1999) para a utilização de processos de ponto marcado.

²p é utilizado para denotar período.

se adequam melhor aos dados se comparados aos modelos DCA, utilizando então esse modelo para a estimação dos tempos entre negócios. Ao combinar ambos os modelos, a pesquisa faz uso de modelos hierárquicos, que consistem no uso de seis modelos condicionais simples para lidar com a estrutura dinâmica da negociação em um pregão versus a estrutura do período de negociação completo, utilizando também dados de IBM para tal propósito.

Em Kent (2015) o modelo de decomposição de preços proposto por Rydberg e Shephard (2003) é estendido com a definição da variável de magnitude de variação de preços seguindo uma adaptação do ACM proposto em Russell e Engle (1998), no lugar do modelo linear binomial negativo generalizado utilizado no artigo original. O objetivo dos autores era o de construir um novo modelo, denominado de ACM-Decomposição, para movimentos de preços de alta frequência que utilize a robustez do modelo ACM enquanto captura a informação de microestrutura adicional obtida por meio do modelo de decomposição considerando a atividade de preços e a direção.

Considerando a literatura apresentada serão utilizados como base para esta pesquisa os modelos originalmente propostos por Hausman et al. (1992) e Rydberg e Shephard (2003) para demonstrar sua aplicação no mercado de dólar futuro brasileiro.

2.3 Características Empíricas dos Dados de Alta Frequência

Os dados em alta frequência, que registram variáveis como preço e volume de cada negócio realizado em um determinado período de tempo, como um pregão, por exemplo, apresentam características distintas dos dados agrupados medidos diariamente ou com alguma outra forma de agregação, como minuto a minuto ou a cada cinco minutos. Essas características são ressaltadas no trabalho de Hausman et al. (1992) e descritas de forma empírica em Campbell et al. (1997).

Tsay (2005, p. 282) define de forma mais precisa os dados em alta frequência da seguinte forma:

“Let t_i be the calendar time, measured in seconds from midnight, at which the i th trade of an asset takes place. Associated with the transaction are several variables such as the trading price, the transaction volume, the prevailing bid and ask quotes, and so on. The collection of

t_i and the associated measurements are referred to as the *transactions data*³.”

Tsay (2005) resume quatro importantes características dos dados em alta frequência que não existem quando as observações são agrupadas em um período de tempo.

1. Intervalos de tempo espaçados desigualmente: transações tais como a negociação de ações não ocorrem em intervalos de tempo igualmente espaçados. Como tal, os preços de negócios de um ativo observados não constituem uma série temporal igualmente espaçada. A variação do tempo ou *time duration* entre os negócios torna-se importante e pode conter informações úteis sobre a microestrutura do mercado (por exemplo, intensidade de negociação).
2. Preços discretos: a mudança de preços de um ativo de uma transação para a próxima ocorre apenas em múltiplos de tamanho de *tick size*. Na NYSE, o *tick size* era um oitavo de dólar até 24 de junho de 1997 e era de dezesseis avos de dólar até 29 de janeiro de 2001. Portanto, os preços são variáveis discretas em dados de alta frequência. Apesar de os mercados de ações americano utilizar um sistema decimal atualmente, as mudanças de preços em negócios consecutivos tendem a ocorrer em múltiplos de um centavo e podem ser tratadas aproximadamente como uma variável discreta.
3. Existência de um padrão diário: nas condições normais de negociação, a atividade de negociação pode exibir padrões periódicos. Por exemplo, na NYSE, as transações são mais volumosas no início e no fechamento dos pregões e menos numerosas durante o horário de almoço, resultando em um volume de negociação em “forma de U”. Consequentemente, as variações de tempo entre transações também exibem um padrão cíclico diário.
4. Múltiplas transações dentro de um único segundo: é possível que várias transações, mesmo com preços diferentes, aconteçam ao mesmo tempo. Isto ocorre em parte devido ao fato de que o tempo é medido em segundos, o que pode ser uma escala de tempo muito longa em períodos de elevado volume de negociação.

³Livre tradução: “Considere-se t_i o tempo do calendário, medido em segundos a partir da meia-noite, no qual ocorre a i -ésima transação de um ativo. Associados à transação, existem diversas variáveis, como o preço da transação, o volume da transação, *spread* existente, e assim por diante. O conjunto de t_i e as variáveis associadas são referidas como os dados das transações”.

Essas características serão avaliadas utilizando os dados em alta frequência empregados neste trabalho nos capítulos mais adiante. A seguir são apresentados os dados utilizados, a especificação dos modelos e os resultados obtidos em sua aplicação.

3 Metodologia

A pesquisa utiliza metodologia quantitativa com análise de dados em archival, por meio da aplicação dos modelos de arbitragem estatística especificados a seguir. Com base nesses modelos são realizadas simulações de previsões de mudanças de preços, por meio da programação de algoritmos em software estatístico de análise de dados, utilizando informações de alta frequência de mini contratos de dólar futuro negociados na B3.

3.1 Dados e Amostra

Os dados utilizados nesta pesquisa foram obtidos em um ftp disponibilizado pela B3¹. Os dados ali disponibilizados abrangem um período de aproximadamente dois anos móveis, ou seja, na data na qual os dados foram adquiridos encontravam-se informações referentes ao período compreendido entre 3 de julho de 2015 e 22 de dezembro de 2017.

O ftp disponibiliza arquivos referentes aos negócios intradiários fechados e ofertas de compra e de venda referentes aos segmentos de negociação BM&F e Bovespa, ou seja, derivativos futuros, mercado à vista de renda variável e opções sobre ações, assim como os *layouts* dos respectivos arquivos.

Com o objetivo de delimitar o escopo de aplicação dos modelos aqui apresentados, que a princípio poderiam ser utilizados em qualquer classe de ativos, a pesquisa foi desenvolvida utilizando dados de mini contratos futuros de dólar, negociados em 06/04/2016 e 05/04, 06/10 e 01/11/2017. A escolha desse ativo para aplicação do modelo deveu-se a sua elevada liquidez, comparada aos demais, e consequente disponibilidade de dados em alta frequência. Essas datas foram selecionadas aleatoriamente com o objetivo de avaliar o comportamento dos modelos em diferentes

¹ftp é disponibilizado pela B3 no endereço [ftp://ftp.bmf.com.br/MarketData/](http://ftp.bmf.com.br/MarketData/)

cenários de tendência de preço, assim como observar a variação dos parâmetros estimados ao longo de diferentes períodos de tempo, aproximadamente um ano, seis meses e um mês.

Os mini contratos futuros de dólar são cotados em Reais por US\$ 1.000,00, possuem código de negociação WDO e tamanho mínimo de contrato de US\$ 10.000,00. A variação mínima de apreçoação, ou o *tick size* do contrato é de R\$ 0,50. O algoritmo desenvolvido para execução dos modelos econôméticos seleciona automaticamente os vencimentos mais líquidos desses contratos (historicamente a liquidez se concentra no vencimento de mais curto prazo), sendo este equivalente ao primeiro dia útil do mês subsequente a data de negociação.

3.2 Modelos

Este trabalho se apoia na utilização de dois modelos estatísticos para previsão de mudanças de preços utilizando dados de alta frequência com o emprego de estratégias de arbitragem estatística. O primeiro, Modelo Probit Ordenado, proposto por Hausman et al. (1992), e o segundo, Modelo de Decomposição, proposto por Rydberg e Shephard (2003), com uma versão modificada por McCulloch e Tsay (2001).

3.2.1 Modelo Probit Ordenado

Considerando que y_i representa as mudanças de preço não observáveis do ativo estudado $y_i = P_{ti} - P_{ti-1}$, tal que P_t representa o preço do ativo, t corresponde ao tempo contado a partir da meia noite, medido em segundos, e i denota a ordem da transação em cada segundo de negociação do ativo. O Modelo Probit Ordenado assume que y_i é uma variável contínua aleatória que segue o modelo:

$$y_i = \mathbf{x}_i\beta + \epsilon_i, \quad (3.1)$$

tal que \mathbf{x}_i representa um vetor linear de variáveis explicativas com dimensão p disponível em t_{i-1} , β é um vetor de parâmetros com dimensão $p \times 1$, $\mathbf{E}[\epsilon_i|\mathbf{x}_i] = 0$, $\mathbf{V}[\epsilon_i|\mathbf{x}_i] = \sigma_i^2$ e $\mathbf{C}[\epsilon_i|\epsilon_j] = 0$ para $i \neq j$.

Assumimos que a variância condicional σ_i^2 seja uma função positiva da variável explicativa \mathbf{w}_i , tal que, $\sigma_i^2 = g(\mathbf{w}_i)$, tal que $g(\cdot)$ é uma função positiva. Para dados

de transações financeiras \mathbf{w}_i pode conter o intervalo de tempo $t_i - t_{i-1}$ e variáveis heterocedásticas condicionais. Usualmente, também assumimos que a distribuição de ϵ_i dado \mathbf{x}_i e \mathbf{w}_i é Gaussiana.

As classes, que demonstram os grupos de possíveis valores que os preços podem assumir, são representadas por $\{s_1, \dots, s_k\}$. O modelo então postula a relação entre y_i e y_i^* como sendo:

$$y_i = s_j \quad \text{se} \quad \alpha_{j-1} < y_i^* \leq \alpha_j, \quad j = 1, \dots, k,$$

tal que α_j são números reais que satisfazem $-\infty = \alpha_0 < \alpha_1 < \dots < \alpha_{k-1} < \alpha_k = \infty$.

Sob a premissa de uma distribuição condicional Gaussiana, tem-se

$$\begin{aligned} P(y_i = s_j | \mathbf{x}_i, \mathbf{w}_i) &= P(\alpha_{j-1} < \mathbf{x}_i \beta + \varepsilon_i \leq \alpha_j | \mathbf{x}_i, \mathbf{w}_i) \\ &= \begin{cases} P(\mathbf{x}_i \beta + \varepsilon_i \leq \alpha_1 | \mathbf{x}_i, \mathbf{w}_i) & \text{se } j = 1, \\ P(\alpha_{j-1} < \mathbf{x}_i \beta + \varepsilon_i \leq \alpha_j | \mathbf{x}_i, \mathbf{w}_i) & \text{se } j = 2, \dots, k-1, \\ P(\alpha_{k-1} < \mathbf{x}_i \beta + \varepsilon_i | \mathbf{x}_i, \mathbf{w}_i) & \text{se } j = k, \end{cases} \\ &= \begin{cases} \Phi\left[\frac{\alpha_1 - \mathbf{x}_i \beta}{\sigma_i(\mathbf{w}_i)}\right] & \text{se } j = 1, \\ \Phi\left[\frac{\alpha_j - \mathbf{x}_i \beta}{\sigma_i(\mathbf{w}_i)}\right] - \Phi\left[\frac{\alpha_{j-1} - \mathbf{x}_i \beta}{\sigma_i(\mathbf{w}_i)}\right] & \text{se } j = 2, \dots, k-1, \\ 1 - \Phi\left[\frac{\alpha_{k-1} - \mathbf{x}_i \beta}{\sigma_i(\mathbf{w}_i)}\right] & \text{se } j = k, \end{cases} \end{aligned}$$

tal que $\Phi(x)$ é a função de distribuição acumulada da variável aleatória normal padrão avaliada em x , e $\sigma_i(\mathbf{w}_i)$ é uma função positiva de \mathbf{w}_i .

Conforme descrito anteriormente, o Modelo Probit Ordenado é utilizado para estimar uma variável aleatória contínua não observável. Os valores observados, que possuem um ordenamento baseado na execução dos negócios, podem ser interpretados como categorias ou classes representando o processo de formação de preços do ativo. O modelo gera, portanto, como resultado a probabilidade de mudanças de preço representadas por cada uma das categorias definidas.

Este modelo será aplicado considerando-se sete categorias para variação de preços, com foco na dependência dinâmica das mudanças de preço intradiárias. Portanto, foram definidas variáveis defasadas indicativas de mudanças de preço:

$$y_{m,j} = \begin{cases} 1 & \text{se } y_{i-m} = s_j, \\ 0 & \text{caso contrário,} \end{cases}$$

tal que s_j denota a j -ésima categoria de mudança de preço e y_{i-m} é a $(i-m)$ -ésima mudança de preço no momento t_{i-m} , tal que $j = 2, \dots, 7$ e $m = 1, 2, 3$, ou seja, classificamos as mudanças de preço para os três negócios consecutivos previamente executados. Também foram consideradas no modelo as mudanças de preço observadas y_{i-m} para $m = 1, 2, 3$. Consequentemente, o modelo utilizado é:

$$\mathbf{x}_i\beta = \sum_{m=1}^3 \beta_m y_{i-m} + \sum_{j=2}^7 \gamma_{1,j} y_{1,j} + \sum_{j=2}^7 \gamma_{2,j} y_{2,j} + \sum_{j=2}^7 \gamma_{3,j} y_{3,j}. \quad (3.2)$$

Este modelo será utilizado para estimar previsões probabilísticas para cada categoria de mudança de preço na transação seguinte.

3.2.2 Modelo de Decomposição

O outro modelo de análise de mudança de preço utilizado é o Modelo de Decomposição que desmembra a variação de preço em três componentes, sendo eles (i) um indicador da mudança de preço; (ii) a direção da variação de preço caso tenha ocorrido uma mudança; e (iii) o tamanho da variação de preço caso tenha ocorrido uma mudança.

Matematicamente a mudança de preço da i -ésima transação pode ser escrita da seguinte forma:

$$y_i = P_{t_i} - P_{t_{i-1}} = A_i D_i S_i, \quad (3.3)$$

tal que A_i representa uma variável binária definida como:

$$A_i = \begin{cases} 0 & \text{se existir uma mudança de preço na } i\text{-ésima transação,} \\ 1 & \text{se o preço permanecer constante na } i\text{-ésima transação.} \end{cases}$$

D_i é uma variável binária que denota a direção da variação de preço caso esta ocorra

$$D_i | (A_i = 1) = \begin{cases} 1 & \text{se o preço subir na } i\text{-ésima transação,} \\ -1 & \text{se o preço cair na } i\text{-ésima transação,} \end{cases}$$

tal que $D_i | (A_i = 1)$ significa que D_i é definido sob a condição $A_i = 1$, e S_i representa o tamanho da variação de preço.

Define-se $S_i = 0$ caso não exista mudança de preço na i -ésima transação. Caso exista uma variação de preços S_i é positivo. Note que D_i não é necessário quando $A_i = 0$, e que existe uma ordem natural da decomposição. D_i é definido somente quando $A_i = 1$ e S_i somente é relevante quando $A_i = 1$ e D_i existe.

A especificação do modelo de decomposição faz uso do ordenamento das variáveis, podendo ser descrito da seguinte forma:

$$\begin{aligned} P(y_i|F_{i-1}) &= P(A_i D_i S_i | F_{i-1}) \\ &= P(S_i | D_i, A_i, F_{i-1}) P(D_i | A_i, F_{i-1}) P(A_i | F_{i-1}), \end{aligned}$$

sendo F_i o conjunto informacional disponível na i -ésima transação.

Como A_i é uma variável binária, é suficiente considerar a evolução da probabilidade $p_i = P(A_i = 1)$ ao longo do tempo. Assumimos que

$$\ln\left(\frac{p_i}{1-p_i}\right) = \mathbf{x}_i \beta \text{ ou } p_i = \frac{e^{\mathbf{x}_i \beta}}{1 + e^{\mathbf{x}_i \beta}}, \quad (3.4)$$

tal que \mathbf{x}_i é um vetor contendo os elementos de F_{i-1} e β é um vetor de parâmetros.

Condicionado a $A_i = 1$, D_i também é uma variável binária, com o seguinte modelo para sua variação $\delta_i = P(D_i = 1 | A_i = 1)$:

$$\ln\left(\frac{\delta_i}{1-\delta_i}\right) = \mathbf{z}_i \gamma \text{ ou } \delta_i = \frac{e^{\mathbf{z}_i \gamma}}{1 + e^{\mathbf{z}_i \gamma}}, \quad (3.5)$$

tal que \mathbf{z}_i é um vetor contendo os elementos de F_{i-1} e γ é um vetor de parâmetros.

Com o objetivo de permitir assimetria nas mudanças de preços positivas e negativas, assumimos que:

$$S_i | (D_i, A_i = 1) \sim 1 + \begin{cases} g(\lambda_{u,i}) & \text{se } D_i = 1, A_i = 1, \\ g(\lambda_{d,i}) & \text{se } D_i = -1, A_i = 1, \end{cases}$$

tal que $g(\lambda)$ com parâmetros λ , tal qual os parâmetros $\lambda_{j,i}$ variam no tempo conforme

$$\ln\left(\frac{\lambda_{j,i}}{1-\lambda_{j,i}}\right) = \mathbf{w}_i \theta_j \text{ ou } \lambda_{j,i} = \frac{e^{\mathbf{w}_i \theta_j}}{1 + e^{\mathbf{w}_i \theta_j}}, \quad j = u, d, \quad (3.6)$$

tal que \mathbf{w}_i é um vetor contendo os elementos de F_{i-1} e θ_j é um vetor de parâmetros. Nessa equação foi realizada uma transformação logística para assegurar que $\lambda_{j,i} \in$

$[0, 1]$.

A função distribuição de probabilidade da variável aleatória x , que segue uma distribuição geométrica $g(\lambda)$, é dada por:

$$p(x = m) = p(S_i = 1 + m) = \lambda(1 - \lambda)^m, \quad m = 0, 1, 2, \dots$$

Foi adicionado 1 na função distribuição de probabilidades geométrica para que a variação de preços, caso ocorra, seja de ao menos uma unidade.

As especificações descritas anteriormente classificam a i -ésima transação em uma das seguintes categorias:

1. Sem mudança de preço: $A_i = 0$ com probabilidade $(1 - p_i)$.
2. Aumento de preço: $A_i = 1$, $D_i = 1$ e a probabilidade associada é $p_i\delta_i$. O tamanho do aumento de preço é dado por $1 + g(\lambda_{u,i})$.
3. Redução de preço: $A_i = 1$, $D_i = -1$ e a probabilidade associada é $p_i(1 - \delta_i)$. O tamanho da redução de preço é dado por $1 + g(\lambda_{d,i})$.

Seja $I_i(j)$ para $j = 1, 2, 3$ a variável indicadora de cada uma das três categorias. Isto é, $I_i(j) = 1$ caso a j -ésima categoria ocorra e $I_i(j) = 0$ caso contrário. A função logarítmica apresentada anteriormente, se torna portanto

$$\begin{aligned} \ln[P(y_i|F_{i-1})] &= I_i(1)\ln[(1 - p_i)] \\ &\quad + I_i(2)[\ln(p_i) + \ln(\delta_i) + \ln(\lambda_{u,i}) + (S_i - 1)\ln(1 - \lambda_{u,i})] \\ &\quad + I_i(3)[\ln(p_i) + \ln(1 - \delta_i) + \ln(\lambda_{d,i}) + (S_i - 1)\ln(1 - \lambda_{d,i})]. \end{aligned}$$

O modelo de decomposição que será empregado possui o seguinte formato:

$$\ln\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 A_{i-1}, \quad (3.7)$$

$$\ln\left(\frac{\delta_i}{1 - \delta_i}\right) = \gamma_0 + \gamma_1 D_{i-1}, \quad (3.8)$$

$$\ln\left(\frac{\lambda_{u,i}}{1 - \lambda_{u,i}}\right) = \theta_{u,0} + \theta_{u,1} S_{i-1}, \quad (3.9)$$

$$\ln\left(\frac{\lambda_{d,i}}{1 - \lambda_{d,i}}\right) = \theta_{d,0} + \theta_{d,1} S_{i-1}, \quad (3.10)$$

tal que A_{i-1} representa o indicador da mudança de preço anterior; D_{i-1} o indicador da direção da variação de preço anterior; e S_{i-1} o tamanho da variação de preço anterior.

4 Resultados

4.1 Estatísticas Descritivas

Diferentemente dos trabalhos que utilizam dados de alta frequência de ações, que selecionam parte do pregão com o objetivo de excluir os efeitos dos leilões de abertura e de fechamento, este trabalho utilizou a integralidade dos dados dos dias de negociação selecionados, devido ao fato de inexistirem leilões para formação de preços de abertura e de fechamento no mercado de dólar futuro.

A Figura 4.1 mostra os preços dos negócios executados nos pregões avaliados. A negociação de 06/04/2016 mostra uma tendência de queda no valor dos mini contratos futuros de dólar, enquanto que 05/04/2017 apresenta uma tendência inicialmente neutra com uma elevação de preços ao final do pregão. A negociação de 06/10/2017 inicia-se com uma grande volatilidade nos preços, o que também fica claro com a análise da Figura 4.5, apresentando uma leve tendência de queda ao final do pregão. Já em 01/11/2017, o dia começa com uma forte tendência de alta, seguido por uma reversão de preços e uma subsequente tentativa de recuperação dos valores interrompida por uma abrupta queda no terceiro quarto do pregão.

O histograma disposto na Figura 4.2 e a distribuição apresentada na Tabela 4.1 mostram uma concentração de negócios sem variação de preços de 84,69 a 90,01% nos quatro pregões avaliados. As mudanças de preço com oscilação de até dois *ticks* foram de aproximadamente 15%, em 06/04/2016, e de 9,94 a 10,96% nos demais pregões. Já as variações com mais de dois *ticks* representam menos de 0,4% das mudanças de preço, demonstrando uma grande concentração destas em operações que não geraram alterações de preço ou produziram pequenas variações. Além disso, a distribuição de modificações de preço positivas e negativas aparenta ser aproximadamente simétrica.

A Figura 4.3 demonstra a quantidade de contratos negociados nos quatro pregões

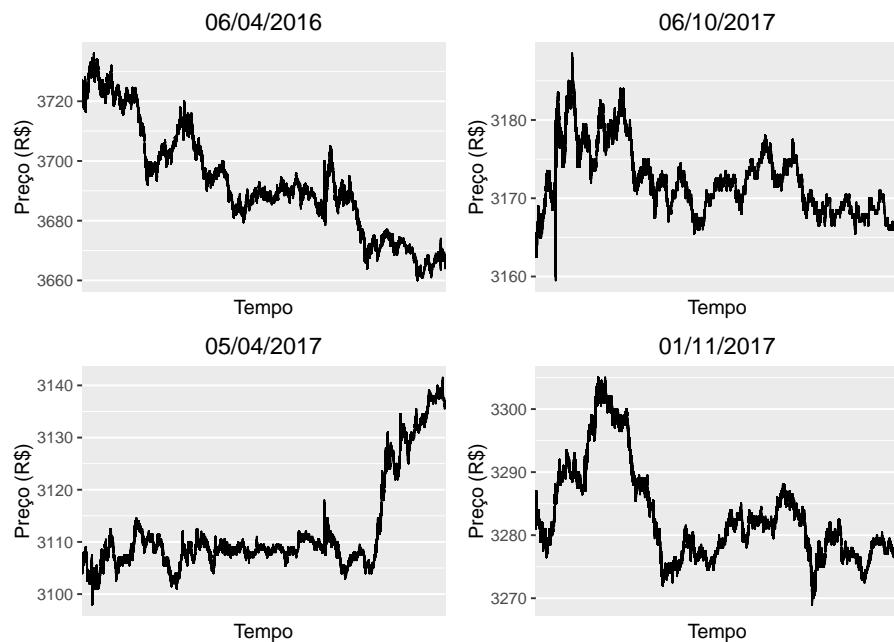


Figura 4.1
Gráficos da variação de preços no pregão

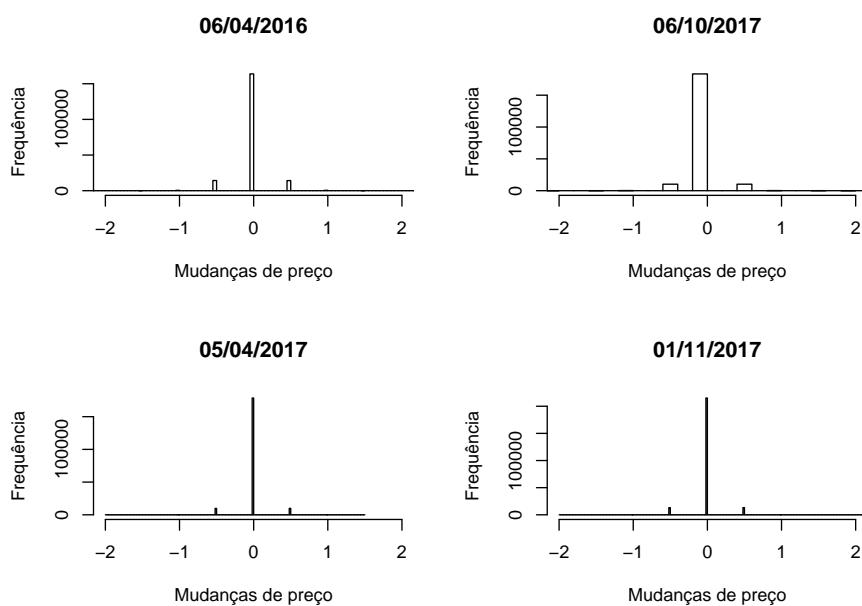


Figura 4.2
Histogramas das variações de preços

Tabela 4.1 Distribuição das mudanças de preços

Frequência relativa às mudanças de preços de mini contratos de dólar futuro negociados na bolsa de valores, mercadorias e futuros brasileira, em quantidade e percentual, nos pregões indicados.

Categoria	06/04/2016	05/04/2017	06/10/2017	01/11/2017
	(1)	(2)	(3)	(4)
< -2	9 (0,01)	1 (0,00)	206 (0,10)	1 (0,00)
[-2, -1)	553 (0,29)	93 (0,05)	291 (0,14)	86 (0,04)
[-1, 0)	14.300 (7,40)	9.781 (4,93)	10.204 (4,98)	13.208 (5,46)
0	163.759 (84,69)	178.487 (90,01)	183.406 (89,55)	215.145 (89,00)
(0, 1]	14.678 (7,59)	9.929 (5,01)	10.411 (5,08)	13.294 (5,50)
(1, 2]	53 (0,03)	3 (0,00)	112 (0,06)	1 (0,00)
> 2	7 (0,00)	0 (0,00)	171 (0,08)	2 (0,00)
Total	193.359	198.294	204.801	241.737

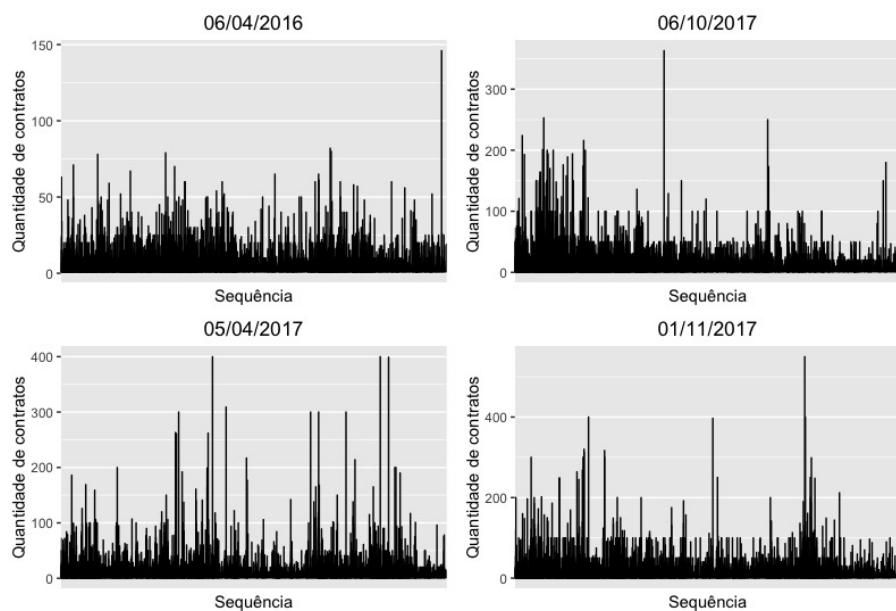


Figura 4.3
Gráficos das quantidades de contratos negociados

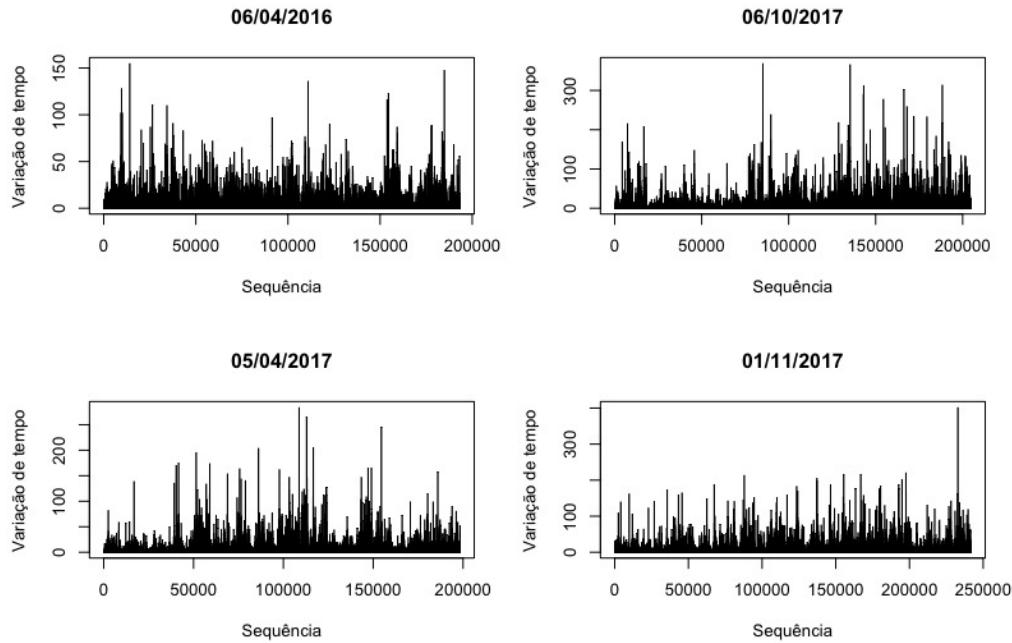


Figura 4.4
Gráficos da variação de tempo entre os negócios

avaliados, a qual não apresenta indícios de existência de um padrão diário evidenciada por uma concentração de volume no início e no fechamento dos pregões, com uma menor quantidade durante o horário de almoço, resultando em um volume de negociação em “forma de U”. A Figura 4.3 também mostra uma elevação dos picos de negociação em 2017 se comparado a 2016.

A Figura 4.4 apresenta gráficos da *time duration* dos pregões analisados, medido em segundos, e a Figura 4.5 exibe as variações de preços em negócios consecutivos. A Figura 4.4 evidencia um padrão de maior espaçamento dos últimos negócios do pregão, com exceção de 05/04/2017, onde esse espaçamento ocorre no meio do pregão, e a Figura 4.5 mostra que as variações de preço, quando ocorrerem, são em sua maioria de um *tick*, com destaque para os pregões de 06/04/2016 e 06/10/2017, com maior variação no final e no início do pregão, respectivamente.

A Tabela 4.2 exibe as estatísticas descritivas dos preços ao longo dos quatro pregões avaliados. A quantidade de observações, que representa o volume de negócios executados no pregão, varia de 193.360, em 06/04/2016, a 241.738, em 01/11/2017, representando uma variação superior a 43 mil negócios no período de aproximadamente um ano e meio. O pregão com maior amplitude de variação de preços e desvio padrão foi 06/04/2016, com 76 e 18, respectivamente. As médias representam as

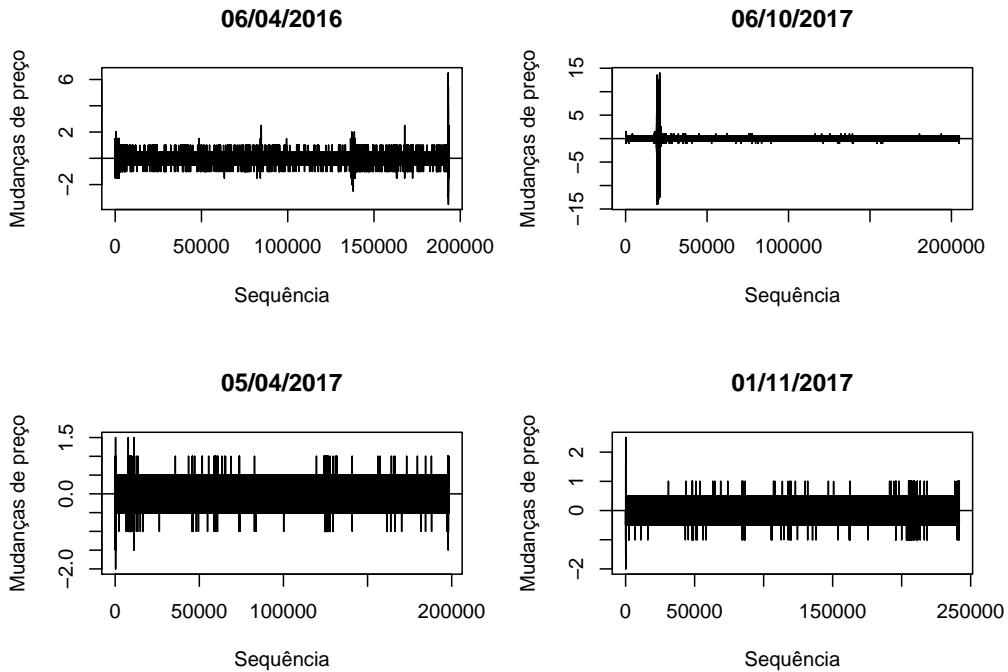


Figura 4.5
Gráficos das mudanças de preços

Tabela 4.2 Estatísticas descritivas dos preços

Estatísticas descritivas dos preços (R\$) de mini contratos de dólar futuro negociados na bolsa de valores, mercadorias e futuros brasileira, nos pregões indicados.

Estatística	06/04/2016	05/04/2017	06/10/2017	01/11/2017
	(1)	(2)	(3)	(4)
média	3.697,925	3.111,662	3.174,771	3.285,232
desvio	17,996	8,637	5,151	8,402
mediana	3.696,000	3.109,000	3.174,500	3.283,000
mínimo	3.660,000	3.098,000	3.159,500	3.269,000
máximo	3.736,000	3.141,500	3.188,500	3.305,000
amplitude	76,000	43,500	29,000	36,000
assimetria	0,046	1,496	0,084	0,535
curtose ^a	-0,828	1,342	-0,743	-0,705
erro padrão	0,041	0,019	0,011	0,017
observações	193.360	198.295	204.802	241.738

^a Nota: Refere-se ao excesso de curtose.

expectativas futuras, entendidas como o mês subsequente devido ao vencimento dos contratos analisados, da cotação do câmbio entre real e dólar, sendo de 3,70, em 06/04/2016, e de 3,28, em 01/11/2017. Quanto à assimetria, os preços de todos os pregões apresentam-se positivamente assimétricos, com destaque para 05/04/2017 e de 01/11/2017 que exibem formação mais assimétrica.

4.2 Estimação de Parâmetros

4.2.1 Modelo Probit Ordenado

A Tabela 4.3 apresenta as estimativas por máxima verossimilhança dos parâmetros das variáveis explicativas do modelo Probit ordenado para os quatro pregões analisados. As entradas em cada uma das colunas, correspondentes a cada um dos pregões, são as estimativas dos parâmetros para as mudanças de preço de mini contratos futuros de dólar, e entre parêntesis é apresentado o erro padrão correspondente ao parâmetro.

O modelo Probit ordenado foi aplicado preliminarmente levando em consideração o volume defasado, definido como $v_{i-m} = V_{i-m}/100$ para $m = 1, 2, 3$, tal que V representa a quantidade de contratos negociada ou o volume corrente. Também foi considerado inicialmente o *time duration* com três defasagens, determinado como $\Delta t_i^* = t_i - t_{i-1}$ entre o $(i-1)$ -ésimo e o i -ésimo negócio medido em milissegundos. Ambas as variáveis não apresentaram significância estatística a um nível de 5% e por isso foram desconsideradas no modelo final aqui apresentado.

As estimativas dos parâmetros β_1 e β_2 , referentes às categorias de mudanças de preço com uma e duas defasagens, respectivamente, apresentam sinal negativo e significante, indicado uma tendência a reversão de preços. O nível de significância dos parâmetros estimados mostra que existe uma dependência dinâmica na variação intradiária do preço. O modelo ajustado, portanto, pode ser usado para fornecer previsões de probabilidade para a próxima mudança de preços. Já o parâmetro β_3 apresentou elevado nível de significância no pregão de 06/10/2017.

Vale notar que os parâmetros γ não foram significantes em 05/04/2017, o que não pode ser explicado devido ao baixo número de mudanças de preço nas categorias mais extremas, 1 e 7, pois a negociação de 01/11/2017 apresenta distribuição semelhante. Uma possibilidade é que isso tenha ocorrido devido a inexistência de mudanças de

Tabela 4.3 Resultados da estimação dos parâmetros das variáveis explicativas do modelo probit ordenado

Estimativa por máxima verossimilhança dos parâmetros das variáveis explicativas do modelo probit ordenado aplicado em $\mathbf{x}_i\beta = \sum_{m=1}^3 \beta_m y_{i-m} + \sum_{j=2}^7 \gamma_{1,j} y_{1,j} + \sum_{j=2}^7 \gamma_{2,j} y_{2,j} + \sum_{j=2}^7 \gamma_{3,j} y_{3,j}$ para mudanças de preços intradiárias de mini contratos de dólar futuro negociados na bolsa de valores, mercadorias e futuros brasileira, nos pregões indicados.

Parâmetro	06/04/2016 (1)	05/04/2017 (2)	06/10/2017 (3)	01/11/2017 (4)
β_1	-0,899*** (0,097)	-1,260*** (0,245)	-0,156*** (0,021)	-2,041*** (0,285)
β_2	-0,241** (0,107)	-1,296*** (0,264)	-0,073*** (0,022)	-0,997*** (0,316)
β_3	0,005 (0,106)	-0,514* (0,268)	-0,106*** (0,021)	-0,215 (0,338)
$\gamma_{1,2}$	1,876*** (0,457)	1,876 (1,357)	1,093*** (0,135)	9,135*** (1,114)
$\gamma_{1,3}$	2,047*** (0,474)	2,139 (1,379)	1,599*** (0,125)	9,630*** (1,151)
$\gamma_{1,4}$	1,463*** (0,496)	0,974 (1,416)	0,037 (0,132)	8,807*** (1,212)
$\gamma_{1,5}$	0,882* (0,522)	-0,163 (1,464)	-1,543*** (0,141)	8,010*** (1,286)
$\gamma_{1,6}$	1,327** (0,607)	-0,114 (1,719)	-0,065 (0,197)	10,994*** (2,371)
$\gamma_{1,7}$	4,544*** (0,873)		0,560** (0,247)	15,448*** (1,970)
$\gamma_{2,2}$	0,242 (0,462)	1,888 (1,579)	0,880*** (0,138)	7,512*** (1,291)
$\gamma_{2,3}$	0,098 (0,481)	2,016 (1,601)	1,393*** (0,130)	7,484*** (1,334)
$\gamma_{2,4}$	-0,292 (0,507)	1,648 (1,638)	0,489*** (0,137)	6,894*** (1,400)
$\gamma_{2,5}$	-0,681 (0,537)	1,267 (1,686)	-0,425*** (0,146)	6,278*** (1,480)
$\gamma_{2,6}$	-1,233* (0,630)	1,938 (2,009)	1,030*** (0,201)	3,657* (1,983)
$\gamma_{2,7}$	0,616 (0,944)		0,686*** (0,253)	7,363*** (2,050)
$\gamma_{3,2}$	0,070 (0,434)	1,235 (1,826)	1,156*** (0,138)	0,934 (1,662)
$\gamma_{3,3}$	0,132 (0,455)	0,887 (1,846)	1,565*** (0,130)	0,951 (1,700)
$\gamma_{3,4}$	-0,146 (0,483)	0,622 (1,879)	1,126*** (0,137)	0,454 (1,760)
$\gamma_{3,5}$	-0,425 (0,515)	0,368 (1,923)	0,709*** (0,146)	-0,020 (1,834)
$\gamma_{3,6}$	-1,057* (0,615)	1,867 (2,237)	1,535*** (0,200)	-2,625 (2,743)
$\gamma_{3,7}$	1,344 (0,928)		1,535*** (0,250)	-0,147 (2,458)
Observações	193.356	198.291	204.798	241.734

Nota:

*p<0,1; **p<0,05; ***p<0,01

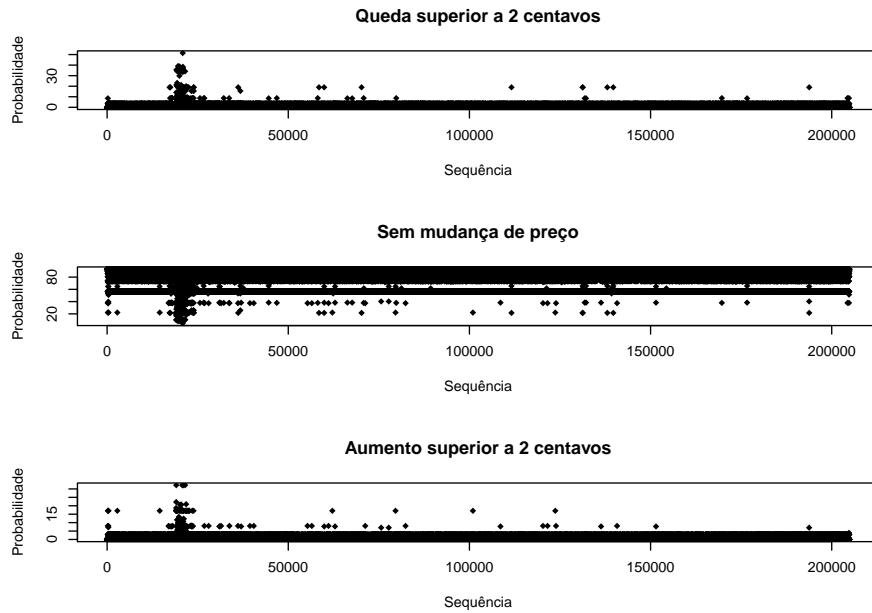


Figura 4.6
Gráfico das probabilidades de mudanças de preços

preço na categoria 7. Já em 06/04/2016, os parâmetros referentes à γ_2 e γ_3 não se apresentaram significantes em sua maioria, e em 01/11/2017 foi o parâmetro γ_3 que não demonstrou significância. Contudo, em 06/10/2017, exceto pelos parâmetros $\gamma_{1,4}$ e $\gamma_{1,6}$ que só se mostraram significantes nos pregões constantes das colunas 1 e 4, todos os parâmetros foram significantes. Essas constatações evidenciam que, para utilização prática desse modelo, ele precisaria ser recalculado com uma frequência determinada.

A Figura 4.6 apresenta as probabilidades de mudanças de preço nas categorias 1, 4 e 7, para o pregão de 06/10/2017, o qual apresentou o maior número de parâmetros significantes, calculadas com as estimativas do modelo probit ordenado. Como esperado, as probabilidades referentes a categoria 4, que representa a ausência de mudanças de preços, são elevadas, porém há casos nos quais as probabilidades de variação de preços superiores e inferiores a dois centavos são elevadas. Isso ocorre sobretudo no início do pregão devido a elevada volatilidade nesse período, conforme demonstra a figura .4.5.

A Tabela 4.4 exibe as estimativas por máxima verossimilhança dos parâmetros referentes aos interceptos do modelo Probit ordenado para os quatro pregões analisados. As entradas em cada uma das colunas, correspondentes a cada um dos pregões, são as estimativas dos limites de participação de cada uma das categorias de mudanças de

Tabela 4.4 Resultados da estimação dos interceptos do modelo probit ordenado

Estimativas por máxima verossimilhança dos interceptos do modelo probit ordenado aplicado em $\mathbf{x}_i\beta = \sum_{m=1}^3 \beta_m y_{i-m} + \sum_{j=2}^7 \gamma_{1,j} y_{1,j} + \sum_{j=2}^7 \gamma_{2,j} y_{2,j} + \sum_{j=2}^7 \gamma_{3,j} y_{3,j}$, para mudanças de preços intradiárias de mini contratos de dólar futuro negociados na bolsa de valores, mercadorias e futuros brasileira, nos pregões indicados.

Limites de Partição	06/04/2016	05/04/2017	06/10/2017	01/11/2017
	(1)	(2)	(3)	(4)
α_1	-3,232*** (0,870)	-1,411 (2,870)	-1,835*** (0,233)	10,888*** (2,553)
α_2	-1,977** (0,866)	-0,646 (2,870)	-1,562*** (0,233)	12,075*** (2,555)
α_3	-0,509 (0,866)	1,383 (2,870)	-0,168 (0,233)	14,319*** (2,556)
α_4	2,567** (0,866)	5,110* (2,870)	3,464*** (0,233)	17,993*** (2,556)
α_5	4,735*** (0,867)	8,267** (2,875)	5,041*** (0,234)	21,244*** (2,557)
α_6	5,285*** (0,869)	88,744*** (2,875)	5,186*** (0,234)	21,327 (2,558)

Nota:

*p<0,1; **p<0,05; ***p<0,01

preço de mini contratos futuros de dólar, e entre parêntesis é apresentado o erro padrão correspondente ao parâmetro.

Finalmente, analisamos os limites de participação calculados com o modelo probit ordenado, apresentados na Tabela 4.4, no pregão de 06/10/2017. Primeiro, porque as variáveis explicativas podem ter médias diferentes de zero, as estimativas dos limites de participação α_i não são simétricas em relação a 0. Em segundo lugar, $\hat{\alpha}_2 - \hat{\alpha}_1 = 0,273$ e $\hat{\alpha}_6 - \hat{\alpha}_5 = 0,145$. Os dois intervalos não têm aproximadamente o mesmo tamanho. Contudo, $\hat{\alpha}_3 - \hat{\alpha}_2 = 1,394$ possui valor próximo de $\hat{\alpha}_5 - \hat{\alpha}_4 = 1,577$. A análise dos limites de participação aponta que a distribuição dos preços parece ser simétrica mais ao centro, mas nem tanto nos extremos.

4.2.2 Modelo de Decomposição

Tabela 4.5 Resultados da estimação do modelo de decomposição

Estimativa dos parâmetros do modelo de decomposição aplicado em $\ln\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 A_{i-1}$, $\ln\left(\frac{\delta_i}{1-\delta_i}\right) = \gamma_0 + \gamma_1 D_{i-1}$, $\ln\left(\frac{\lambda_{u,i}}{1-\lambda_{u,i}}\right) = \theta_{u,0} + \theta_{u,1} S_{i-1}$, $\ln\left(\frac{\lambda_{d,i}}{1-\lambda_{d,i}}\right) = \theta_{d,0} + \theta_{d,1} S_{i-1}$, para mudanças de preços intradiárias de mini contratos de dólar futuro negociados na bolsa de valores, mercadorias e futuros brasileira, nos pregões indicados.

Parâmetro	06/04/2016	05/04/2017	06/10/2017	01/11/2017
	(1)	(2)	(3)	(4)
β_0	-1,927*** (0,007)	-2,552*** (0,009)	-2,531*** (0,009)	-2,449*** (0,008)
β_1	1,067*** (0,015)	1,927*** (0,017)	1,992*** (0,017)	1,860*** (0,015)
γ_0	-0,010 (0,013)	0,009 (0,017)	0,016 (0,016)	-0,007 (0,015)
γ_1	-2,768*** (0,045)	-4,397*** (0,110)	-2,994*** (0,053)	-4,615*** (0,104)
$\theta_{u,0}$	5,664*** (0,162)	9,865*** (1,353)	4,331*** (0,081)	8,874*** (0,849)
$\theta_{u,1}$	-0,671** (0,212)	-2,574** (1,227)	-1,416*** (0,053)	-1,701* (0,902)
$\theta_{d,0}$	3,563*** (0,058)	4,707*** (0,132)	3,610*** (0,061)	5,440*** (0,163)
$\theta_{d,1}$	-0,776*** (0,082)	-0,170 (0,211)	-1,217*** (0,048)	-0,909*** (0,214)

Nota:

*p<0,1; **p<0,05; ***p<0,01

A Tabela 4.5 expõe as estimativas por máxima verossimilhança do modelo de decomposição para os quatro pregões analisados. As entradas em cada uma das colunas, correspondentes a cada um dos pregões, são as estimativas de parâmetros para as mudanças de preço de mini contratos futuros de dólar, e entre parêntesis o erro padrão correspondente a esse parâmetro.

Os parâmetros estimados do modelo apresentam significância estatística em todos os pregões avaliados, com exceção do parâmetro referente ao intercepto do indicador referente ao sinal da variação de preços e, no pregão de 05/04/2017, o parâmetro relativo ao tamanho da variação de preços quando ocorrem quedas de preço. Este modelo aparenta resultados mais parcimoniosos aos comparados ao do modelo probit ordenado ao se levar em consideração o nível de significância nos diferentes pregões avaliados.

Estimando as probabilidades de mudanças de preço utilizando o modelo de decomposição estimado para o pregão de 06/10/2017, que possui o maior número de parâmetros significantes e que foi analisado com mais detalhes na seção anterior, tem-se:

$$P(A_i = 1|A_{i-1} = 0) = 0,074,$$

$$P(A_i = 1|A_{i-1} = 1) = 0,368.$$

O resultado indica que a maioria das transações não apresenta mudanças de preço, conforme esperado. Quando nenhuma mudança de preço ocorreu no $(i - 1)$ -ésimo negócio, então apenas cerca de 7,4

A direção das mudanças de preço é ditada por:

$$P(D_i = 1|F_{i-1}, A_i) = \begin{cases} 0,504 & \text{se } D_{i-1} = 0, \quad A_{i-1} = 0, \\ 0,048 & \text{se } D_{i-1} = 1, \quad A_i = 1, \\ 0,953 & \text{se } D_{i-1} = -1, \quad A_i = 1. \end{cases}$$

Este resultado diz que (i) se nenhuma mudança de preço ocorreu no $(i - 1)$ -ésimo negócio, então as chances de aumento ou diminuição de preços no i -ésimo negócio são aproximadamente as mesmas; e (ii) as probabilidades de mudanças de preços

consecutivas são pequenas. A probabilidade de um aumento de preços no negócio i , dado que ocorreu uma mudança de preço no i -ésimo negócio, e houve um aumento de preço no $(i - 1)$ -ésimo negócio é de apenas 4,8%. No entanto, a probabilidade de um aumento de preços é cerca de 95,3%, uma vez que ocorre uma mudança de preço no negócio i e houve uma diminuição de preço no $(i - 1)$ -ésimo negócio. Consequentemente, esse resultado demonstra o efeito *bid-ask bounce* e suporta a ideia de reversões de preços na negociação de alta frequência.

Há fracas evidências sugerindo que grandes mudanças de preços têm maior probabilidade de serem seguidas por outra grande mudança de preços. Considere o tamanho de um aumento de preço. Temos:

$$S_i | (D_I = 1) \quad 1 + g(\lambda_{u,i}), \quad \lambda_{u,i} = 4,331 - 1,416S_{i-1}$$

Usando a função de probabilidades de uma distribuição geométrica, obtém-se que a probabilidade de um aumento de preços em 1 centavo é de 0,949 no i -ésimo negócio se a transação resulta em um aumento de preços e $S_{i-1} = 1$. A probabilidade reduz para 0,817 se $S_{i-1} = 2$ e a 0,521 se $S_{i-1} = 3$. Da mesma forma, a probabilidade de um preço aumentar em 2 centavos é 0,048 no negócio i se a transação resultar em um aumento de preço e $S_{i-1} = 1$. A probabilidade aumenta para 0,150 se $S_{i-1} = 2$ e para 0,250 se $S_{i-1} = 3$. Consequentemente, a probabilidade de um S_i grande é proporcional a S_{i-1} , dado que ocorreu um aumento de preços no i -ésimo negócio.

5 Conclusão

Utilizando dados de mini contratos futuros de dólar, em quatro pregões, disponibilizados pela B3 em seu banco de dados, identificou-se que os modelos probit ordenado e de decomposição podem ser utilizados para previsão da variação de preços desse ativo e que o modelo de decomposição apresenta resultados mais significantes, ao nível de 5%, se comparado ao probit ordenado, para o conjunto de pregões analisados. Também deduziu-se que as variáveis de volume e *time duration* não se apresentaram relevantes na determinação do preço desse contrato e que os modelos geram resultados consistentes com diferentes tendências de mercado.

Este trabalho contribuiu com a literatura da área de microestrutura de mercado ao realizar a aplicação de dois modelos estatísticos em um mercado e um ativo até então ainda não explorados. Os modelos foram criados para a avaliação de ações no mercado Norte-americano e ao aplicá-los no mercado brasileiro de mini contratos futuros de dólar pôde-se avaliar algumas importantes distinções entre os mercados, como a ausência de padrões de negociação, como o volume de negociação em “forma de U”, no mercado brasileiro, assim como similaridades tais como a tendência à reversão de preços em ambos os mercados e a concentração de negócios sem mudanças de preço, embora o percentual desse tipo de transação ainda seja maior no mercado brasileiro.

A pesquisa aqui realizada ainda pode ser estendida com a aplicação desses modelos em outras classes de ativos e utilizando diferentes regressores em seu emprego, ou na comparação das previsões desses modelos com os movimentos reais dos ativos para se obter uma janela ótima de dados utilizados na estimação dos parâmetros e também identificar qual seria o modelo mais assertivo do ponto de vista preditivo. A pesquisa poderia ainda ser ampliada com a utilização, e subsequente comparação, de outros modelos como o *Autoregressive Conditional Multinomial* (ACM), proposto por Russell e Engle (1998), ou mesmo com a proposição de novos modelos, como realizado em Kent (2015) com o modelo denominado de ACM-Decomposição.

Referências

- Aldridge, I. (2009). *High-frequency trading: a practical guide to algorithmic strategies and trading systems*. Wiley.
- Avellaneda, M. & Stoikov, S. (2008). High-frequency trading in a limit order book. *Quantitative Finance*, 8(3), 217–224.
- Ball, C. A. (1988). Estimation bias induced by discrete security prices. *The Journal of Finance*, 43(4), 841–865.
- Brogaard, J. et al. (2010). High frequency trading and its impact on market quality. *Northwestern University Kellogg School of Management Working Paper*, 66.
- Brogaard, J., Hendershott, T., & Riordan, R. (2014). High-frequency trading and price discovery. *The Review of Financial Studies*, 27(8), 2267–2306.
- Campbell, J. Y., Lo, A. W.-C., & MacKinlay, A. C. (1997). *The econometrics of financial markets*. Princeton University Press.
- Carmona, R. & Webster, K. (2012). High frequency market making. *arXiv preprint arXiv:1210.5781*. Recuperado de <https://arxiv.org/abs/1210.5781>
- Cartea, Á. & Jaimungal, S. (2015). Risk metrics and fine tuning of high-frequency trading strategies. *Mathematical Finance*, 25(3), 576–611.
- Cartea, Á. & Jaimungal, S. (2016). A closed-form execution strategy to target volume weighted average price. *SIAM Journal on Financial Mathematics*, 7(1), 760–785. Recuperado de <https://doi.org/10.1137/16M1058406>
- Cartea, Á., Jaimungal, S., & Penalva, J. (2015). *Algorithmic and high-frequency trading*. Cambridge University Press.
- Cartea, Á., Jaimungal, S., & Ricci, J. (2014). Buy low, sell high: A high frequency trading perspective. *SIAM Journal on Financial Mathematics*, 5(1), 415–444.
- Cartea, Á. & Penalva, J. (2012). Where is the value in high frequency trading? *The Quarterly Journal of Finance*, 2(03), 1250014.
- Chaboud, A. P., Chiquoine, B., Hjalmarsson, E., & Vega, C. (2014). Rise of the machines: Algorithmic trading in the foreign exchange market. *The Journal of Finance*, 69(5), 2045–2084.

- Cho, D. C. & Frees, E. W. (1988). Estimating the volatility of discrete stock prices. *The Journal of Finance*, 43(2), 451–466.
- Durbin, M. (2010). *All about high-frequency trading*. McGraw Hill Professional.
- Elliott, R. J., Hoek, J. V. D., & Malcolm, W. P. (2005). Pairs Trading. *Quantitative Finance*, 5(3), 271–276.
- Gomber, P., Arndt, B., Lutat, M., & Uhle, T. (2011). High-Frequency Trading. Goethe University. Deutsche Börse Group.
- Gottlieb, G. & Kalay, A. (1985). Implications of the discreteness of observed stock prices. *The Journal of Finance*, 40(1), 135–153.
- Grossman, S. (1976). On the efficiency of competitive stock markets where trades have diverse information. *The Journal of Finance*, 31(2), 573–585.
- Grossman, S. & Miller, M. H. (1988). Liquidity and market structure. *The Journal of Finance*, 43(3), 617–633.
- Guéant, O. & Royer, G. (2014). VWAP execution and guaranteed VWAP. *SIAM Journal on Financial Mathematics*, 5(1), 445–471.
- Harris, L. (1991). Stock price clustering and discreteness. *The Review of Financial Studies*, 4(3), 389–415.
- Hausman, J. A., Lo, A. W., & MacKinlay, A. C. (1992). An ordered probit analysis of transaction stock prices. *Journal of Financial Economics*, 31(3), 319–379.
- Hendershott, T., Jones, C. M., & Menkveld, A. J. (2011). Does algorithmic trading improve liquidity? *The Journal of Finance*, 66(1), 1–33.
- Ho, T. & Stoll, H. R. (1981). Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial Economics*, 9(1), 47–73.
- Hruska, J. & Linnertova, D. (2015). Liquidity of the European Stock Markets Under the Influence of HFT. *Procedia Economics and Finance*, 26, 375–381.
- Johnson, B. (2010). *Algorithmic Trading & DMA: An introduction to direct access trading strategies*. 4Myeloma Press.
- Kent, A. (2015). High Frequency Decomposition and Trade Arrivals. *Undergraduate Honors Theses*.
- Kirilenko, A., Kyle, A. S., Samadi, M., & Tuzun, T. (2011). The flash crash: The impact of high frequency trading on an electronic market. Available at SSRN, 1686004. Recuperado de <https://ssrn.com/abstract=1686004>
- Konishi, H. (2002). Optimal slice of a VWAP trade. *Journal of Financial Markets*, 5(2), 197–221.
- Kyle, A. S. (1989). Informed speculation with imperfect competition. *The Review of Economic Studies*, 56(3), 317–355.

- Leung, T. & Li, X. (2015). Optimal mean reversion trading with transaction costs and stop-loss exit. *International Journal of Theoretical and Applied Finance*, 18(03), 1550020.
- Leung, T. & Ludkovski, M. (2011). Optimal timing to purchase options. *SIAM Journal on Financial Mathematics*, 2(1), 768–793.
- Li, K., Cooper, R., & Van Vliet, B. (2017). How Does High-Frequency Trading Affect Low-Frequency Trading? *Journal of Behavioral Finance*, 1–14.
- Marsh, T. A. & Rosenfeld, E. R. (1986). Non-trading, market making, and estimates of stock price volatility. *Journal of Financial Economics*, 15(3), 359–372.
- McCulloch, R. E. & Tsay, R. S. (2001). Nonlinearity in high-frequency financial data and hierarchical models. *Studies in Nonlinear Dynamics & Econometrics*, 5(1).
- Mitchell, D., Bialkowski, J. P., & Tompaidis, S. (2013). Optimal VWAP Tracking. Available at SSRN, 2333916. Recuperado de <https://ssrn.com/abstract=2333916>
- Monteiro, A. (2008). *The dynamics of cooperative credit risk. An intensity-based econometric*. Rozenberg Publishers.
- Russell, J. R. (1999). Econometric modeling of multivariate irregularly-spaced high-frequency data. Available at SSRN, 2333916. Recuperado de <http://faculty.chicagobooth.edu/jeffrey.russell/research/multi.pdf>
- Russell, J. R. & Engle, R. F. (1998). Econometric analysis of discrete-valued irregularly-spaced financial transactions data using a new autoregressive conditional multinomial model. Available at SSRN, 106528. Recuperado de <https://ssrn.com/abstract=106528>
- Rydberg, T. H. & Shephard, N. (2003). Dynamics of trade-by-trade price movements: decomposition and models. *Journal of Financial Econometrics*, 1(1), 2–25.
- SEC and CFTC. (2010). Findings regarding the market events of May 6, 2010. *Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues*. Recuperado de <https://www.sec.gov/news/studies/2010/marketevents-report.pdf>
- Tourin, A. & Yan, R. (2013). Dynamic pairs trading using the stochastic control approach. *Journal of Economic Dynamics and Control*, 37(10), 1972–1981.
- Tsay, R. S. (2005). *Analysis of financial time series*. Wiley.
- Zhang, F. (2010). High-frequency trading, stock volatility, and price discovery. Available at SSRN, 1691679. Recuperado de <https://ssrn.com/abstract=1691679>