

**FUNDAÇÃO GETÚLIO VARGAS
ESCOLA DE ECONOMIA DE SÃO PAULO**

MARCELO TADEU MARCHI

**MICROESTRUTURA DE MERCADO : UMA ANÁLISE DOS DADOS DE ALTA
FREQUENCIA DO MINICONTRATO DE FUTURO DE ÍNDICE BOVESPA**

**SÃO PAULO
2011**

MARCELO TADEU MARCHI

**MICROESTRUTURA DE MERCADO : UMA ANÁLISE DOS DADOS DE ALTA
FREQUENCIA DO MINICONTRATO DE FUTURO DE ÍNDICE BOVESPA**

Dissertação apresentada à Escola de Economia da
Fundação Getúlio Vargas (FGV/EESP) como requisito
para obtenção do título de Mestre em Finanças e
Economia Empresarial.

Orientador: Prof. Dr. Afonso de Campos Pinto

SÃO PAULO

2011

Marchi, Marcelo Tadeu.

Microestrutura de mercado : uma análise dos dados de alta frequência do minicontrato de futuro de índice Bovespa / Marcelo Tadeu Marchi. - 2011.
67 f.

Orientador: Afonso de Campos Pinto

Dissertação (mestrado profissional) - Escola de Economia de São Paulo.

1. Mercado de capitais -- Análise. 2. Mercado de capitais -- Brasil. 3. Bolsa de Mercadorias & Futuros (Brasil) -- Indicadores. 4. Bolsa de Valores de São Paulo -- Índices. I. Pinto, Afonso de Campos. II. Dissertação (mestrado profissional) - Escola de Economia de São Paulo. III. Título.

CDU 336.76(81)

MARCELO TADEU MARCHI

**MICROESTRUTURA DE MERCADO : UMA ANÁLISE DOS DADOS DE ALTA
FREQUENCIA DO MINICONTRATO DE FUTURO DE ÍNDICE BOVESPA**

Dissertação apresentada à Escola de Economia da
Fundação Getúlio Vargas (FGV/EESP) como requisito
para obtenção do título de Mestre em Finanças e
Economia Empresarial.

Data de aprovação:

____/____/____

Banca Examinadora:

Prof. Dr. Afonso de Campos Pinto
(Orientador)
FGV-EAESP

Prof. Dr. Ricardo Rochman
FGV-EAESP

Prof. Dr. Flavio Almeida de Magalhaes Cipparrone
Escola Politécnica - USP

Agradecimentos

Aos professores e coordenador do MPFE-FGV, que me permitiram voltar a sentir o prazer de ser aprendiz.

Ao Prof. Dr. Afonso de Campos Pinto, orientador e, em certa medida, empreendedor deste trabalho.

Ao Prof. Dr. Emerson Fernandes Marçal, pela simpatia e atenção ao me receber para orientações na reta final.

Aos companheiros de sala, pelos sábados investidos, a minha admiração. Profissionais bem sucedidos que, humildemente, se dispuseram a aprender e trocar conhecimentos.

Ao prezado Alan de Genaro, da BM&FBOVESPA, o qual, pela gentileza, atenção e disponibilidade em compartilhar o seu objeto de estudo, reforçou e viabilizou minha escolha pelo tema.

Ao colega Jeferson Martin de Araújo, da BM&FBOVESPA, pelas incontáveis horas de empolgantes e bem fundamentadas discussões acerca dos impactos da latência na dinâmica do mercado.

Aos meus pais, os quais, sem o formalismo acadêmico, ensinaram-me as maiores lições. Ao meu querido irmão, pela inspiração de seguir adiante.

E por fim, à minha querida esposa Flavia, que me apoia incondicionalmente nesse desejo compartilhado de busca pelo espírito científico.

“A tarefa não é tanto ver aquilo que ninguém viu, mas pensar o que ninguém ainda pensou sobre aquilo que todo mundo vê”.

Arthur Schopenhauer

RESUMO

Este trabalho examina hipóteses de microestrutura de mercado nos dados de minicontratos futuros do Ibovespa divulgados em alta frequência pelo sinal de *market data* da BM&FBOVESPA. Para a análise, foi utilizado o modelo de duração condicional autoregressivo, o qual permite o tratamento estatístico de séries financeiras que possuem distribuição em espaços irregulares de tempo e padrões condicionais ao conjunto de informações disponíveis no mercado.

A contribuição deste trabalho reside na validação das premissas sobre microestrutura nos mercados de futuros de índice, que são diferentes dos mercados de ações, pois não existe a figura do formador de mercado, e a literatura global sugere que o efeito da assimetria de informação é insignificante na formação de preços devido à diversificação de ativos dos índices. Em contrapartida, o especulador é o provedor de liquidez neste mercado, e seu papel na intensidade de negociação, volume e *spread* é avaliado. Os resultados confirmam a relação negativa entre intensidade de negociação, volume e *spread* e corroboram a tese de que a liquidez de mercado pode ser entendida como o preço pela demanda por execução imediata de negócios.

ABSTRACT

This work examines the market microstructure hypotheses in the data for mini futures Ibovespa released in high frequency signal by the market data of the BOVESPA. For the analysis, the autoregressive conditional duration model, which allows the statistical treatment of financial series that have irregular distribution in space and time patterns conditional on the set of available market information.

The contribution of this work is the validation of assumptions about the microstructure of index futures markets, which are different from stock markets, where there is the market maker role and the literature suggests that the overall effect of information asymmetry is irrelevant in pricing discovery due to the diversification of asset indices. In futures index market the speculator is the liquidity provider and its role in the intensity of trading volume and spread is evaluated. The results confirm the negative relationship between trading intensity, volume and spread, and corroborates the theory of liquidity as the price for the demand for immediacy.

LISTA DE ILUSTRAÇÕES

Gráfico 1.1 - Participação nas operações em alta frequência	13
Gráfico 6.1 - Histograma e estatísticas da duração de preços	48
Gráfico 6.2 - Autocorrelação serial da duração de preços	49
Gráfico 6.3 - Efeito do horário na duração de preços.....	52
Gráfico 6.4 - Interpolação linear das durações de preços e variáveis explicativas...	53

LISTA DE TABELAS

Tabela 5.1 - Perfil dos dados do ativo WINZ08	40
Tabela 5.2 - Registro de ofertas e transações e topo do livro de ofertas	41
Tabela 5.3 - Duração entre ordens em milissegundos	42
Tabela 5.4 - Duração entre negócios em milissegundos.....	43
Tabela 6.1 - Estimação EACD para duração de preços.....	54
Tabela 6.2 – Modelos EACD(2,2): testes de agrupamento de preços	56

SUMÁRIO

1. Introdução	12
2. Revisão da literatura	16
3. Microestrutura do mercado e contratos de futuro de índice	22
4. Metodologia	26
4.1. Estudo de séries temporais de alta frequência	26
4.2. Modelos de Duração Condicional	29
4.3. Modelo de Duração Condicional Autoregressivo	33
4.4. Especificações do modelo de Duração Condicional Autoregressivo	34
4.5. Modelo Duração Condicional Autoregressivo Exponencial	34
5. Base de Dados	37
5.1. Minicontrato futuro do Ibovespa	37
5.2. Características dos dados	38
6. Análise Empírica	44
6.1. Considerações sobre a estimação	44
6.2. Duração de preços e volatilidade	46
6.3. Modelo de Duração Condicional Autoregressivo para o movimento de preços	47
6.4. Variáveis explicativas relacionadas ao processo de negociação	49
6.5. Sazonalidades intradiárias	51
6.6. Testes de hipóteses sobre microestrutura de mercado	54
7. Conclusão	58

1. Introdução

A Teoria da Microestrutura de Mercado, segundo O'Hara (1995), tem foco nos estudos de formação de preço, liquidez dos ativos, desenho de protocolos e regras de mercado, modelos de transparência das negociações e análise de risco. Isto é, cobre toda a estrutura institucional e, mais especificamente, a estrutura informacional na qual as transações financeiras são concluídas. Entretanto, os mercados de capitais ao redor do mundo tornaram-se, definitivamente, eletrônicos e as bolsas de valores cada vez mais automatizadas com sistemas de negociação que processam milhares de transações por segundo. A difusão em tempo real de informações detalhadas de todas as ofertas e transações efetuadas tem estimulado algumas inovações bastante relevantes na estrutura do mercado de capitais com um número cada vez maior de participantes enviando milhares de ordens automaticamente. Os participantes do mercado estão em busca de maior velocidade no acesso às praças de liquidez, ajustes cada vez mais eficientes das suas posições consolidadas e lucros de oportunidade. Tanto oportunidades de arbitragem, de especulação, quanto estratégias de investimento são mais efetivas quando um investidor identifica, age e recebe a confirmação da sua ação mais rapidamente que os demais. Tal necessidade de velocidade gera, por consequência, uma corrida tecnológica dos *traders* e bolsas ao redor do mundo em busca de um ambiente de negociação de baixa latência, isto é, um ambiente onde os eventos de negócio, as análises das informações, as ações dos investidores e as respostas das bolsas aconteçam dentro de poucos milissegundos. Neste ambiente veloz, a nova figura central do mercado passa a ser os *algorithmic traders*, ou *algo traders*.

Tal evolução, no Brasil, conta com os recentes investimentos em tecnologia da informação conduzidos pela BM&FBOVESPA em parceria com a *Chicago Mercantile Exchange Group Incorporation* (CME Group), com o objetivo de criar uma nova plataforma unificada de negociação que suporte mecanismos e instrumentos de negociação globais e que atenda um potencial aumento na demanda pelas operações em alta frequência (HFT - *High Frequency Trading*).

Segundo Martins (2010), as operações em alta frequência no mercado americano representam 70% do volume financeiro negociado, enquanto no Brasil, a estimativa é que apenas 6% dos negócios fechados têm como contraparte este tipo de participante, o que indica um grande espaço para crescimento deste tipo de operação. Em uma análise mais detalhada, segundo informações da BM&FBovespa (2011) apresentadas no Gráfico 1.1, o volume financeiro negociado em operações de alta frequência está essencialmente distribuído em 3 grupos de contratos, sendo os minicontratos os mais representativos com 54,7% do volume.



Gráfico 1.1 - Participação nas operações em alta frequência

É a partir desse contexto que surgem uma série de questões diretamente relacionadas com Teoria da Microestrutura de Mercado: os mercados de alta frequência são mais eficientes? Isto é, a divulgação de informações de negociações que ocorrem em milissegundos melhora a qualidade e liquidez do mercado? Investimentos na infraestrutura de tecnologia da informação das bolsas têm impacto positivo nas características do mercado? Como prever o comportamento do mercado no escopo de milissegundos? Quais as técnicas econométricas mais adequadas?

A recente literatura acadêmica avalia a estrutura institucional onde as demandas latentes dos investidores são convertidas em negócios a partir da disponibilidade de informações de transações em alta frequência, o que permite que muitas das teorias e hipóteses levantadas sobre a microestrutura do mercado possam ser testadas. Entretanto, para Pacurar (2006), a tarefa de estudar as séries financeiras que representam o comportamento dos investidores tem enfrentado um desafio especial, uma vez que, os dados destas transações apresentam intervalos de tempos pequenos e irregulares, enquanto os métodos econométricos tradicionais

são baseados em dados com intervalos de tempo iguais. O problema se torna ainda mais relevante se observarmos que a frequência de transações varia com o tempo. Eventos como a negociação de ativos, a qual apresenta baixa atividade, sendo repentinamente negociados com alta frequência, representam também um importante desafio de estudo, uma vez que, fixar intervalos de tempo e agrupar eventos podem tornar a base de dados de pesquisa ainda menos informativa.

Do ponto de vista econômico, o interesse de estudo reside não apenas na dinâmica apresentada pelas variáveis de interesse, mas também na relação de influência existente entre elas. Os modelos mais recentes da literatura de microestrutura de mercado argumentam que a dinâmica do tempo entre as transações transmite informação relevante ao mercado, e que também deve ser modelada para o estudo dos custos de transação, liquidez e análises dos impactos de diferentes mecanismos de formação de mercado.

A pesquisa em microestrutura de mercado tem suas próprias técnicas, as quais foram, em sua maioria, desenvolvidas na literatura com a aplicação e uso dos dados de alta frequência do mercado de ações da NYSE (*New York Stock Exchange*), seguindo as técnicas econométricas dos modelos autoregressivos de duração condicional (ACD - *Autoregressive Conditional Duration*), propostos por Engle e Russel (1998). Por isso, a maioria dos trabalhos internacionais tem examinado a microestrutura do mercado de ações, e alguns se concentram no mercado de câmbio.

Entretanto, este trabalho investiga a microestrutura de mercado de minicontratos de futuros de índice Ibovespa, em particular a relação intertemporal da dinâmica de preços, o volume negociado e o *spread* de compra e venda, utilizando-se dados de alta frequência de 2008. Segundo Tse (1999), o mercado de contratos futuros de índice tem um mecanismo de formação de preços e fluxo de ordens diferentes de outros mercados de valores mobiliários, pois não possui o papel do *market maker*, os custos de carregamento das posições são menores e a diversificação de ativos utilizados na formação do índice reduz os efeitos da informação privada dos participantes na formação de preço. Para o autor, os *traders* sempre preferem negociar mais ao mesmo preço, e isso induz à revisão do *spread* pelos participantes de mercado. Por isso, este trabalho pode gerar *insights*

importantes, pois corrobora a tese de imediatismo de microestrutura de mercado da literatura e identifica que o volume negociado e a intensidade de negociação são informativos. Embora o principal interesse deste trabalho seja a microestrutura dos minicontratos de futuro de índice, a análise é aplicável a outros mercados futuros e derivativos.

O restante do trabalho está organizado da seguinte forma: o capítulo 2 apresenta uma revisão bibliográfica sobre a microestrutura do mercado de capitais, desenvolvimento teórico e aplicações dos modelos autoregressivos de duração condicional, cujo objetivo é dar sustentação teórica à metodologia deste trabalho. O capítulo 3 descreve o contexto da Teoria de Microestrutura de Mercado aplicável ao mercado de contratos futuros de índice. O capítulo 4 formaliza a metodologia utilizada, com atenção especial às especificações que serão utilizadas nas estimações econométricas. O capítulo 5 relaciona os critérios de seleção dos dados e detalhes de elaboração da base de dados. O capítulo 6 apresenta a pesquisa empírica e resultados, e o capítulo 7 conclui e finaliza o presente trabalho.

2. Revisão da literatura

A Teoria de Microestrutura de mercado tem como tema o modelo institucional no qual as transações de ativos financeiros são realizadas. É o campo da economia e finanças que estuda o mecanismo pelo qual os preços são ajustados pelo mercado para refletir novas informações disponíveis.

As pesquisas de O'Hara (1995) inauguraram o referencial teórico para estudos de microestrutura. Os primeiros modelos foram voltados para o problema do estoque do formador de mercado (*Market maker*), o qual é sempre confrontado com a sucessão de eventos de compras e vendas, e, com o objetivo de maximizar ganhos e manter a solvência, define o preço de negociação. Em outra abordagem, o formador de mercado é visto como qualquer outro participante do mercado que deseja manter uma carteira eficiente no sentido de média-variância e, para compensar o risco de carregar uma carteira não eficiente, define cotações diferentes de compra e venda e, com isso, o *spread*. As conclusões mais importantes de Easley e O'Hara (1992) são:

- A variável tempo não é mais exógena;
- A sequência de preços pode ser informativa;
- O volume negociado revela informações importantes ao formador de mercado. O valor da informação sobre o volume negociado é importante, uma vez que o *trader* informado deseja sempre negociar mais ativos ao mesmo preço, o que implica que maiores volumes negociados exigem uma revisão do *spread* por parte do formador de mercado;

Glosten e Milgrom (1985) introduziram um grande avanço na pesquisa de microestrutura, pois elaboraram modelos baseados em informação. Nesses modelos, os participantes do mercado são diferenciados pela assimetria de informação entre eles: os *traders* informados e *traders* desinformados. Estes últimos não possuem qualquer tipo informação relevante em relação ao valor do ativo negociado, e entram no mercado por razões de liquidez. Já os *traders* informados, ao contrário, sempre sabem o verdadeiro valor do ativo e, portanto, eles sempre

vendem, se sabem de más notícias, e compram se as notícias forem boas em relação ao ativo. Nestes modelos, assume-se que o formador de mercado é um *trader* desinformado, de tal forma que ele sempre perde o negócio contra o *trader* informado. Para compensar essa perda, ele define a sua oferta de tal forma que obtenha ganhos dos *traders* desinformados. É nesse processo que o *spread* positivo de compra e venda surge.

Até o trabalho de Easley e O'Hara (1992), o papel da variável tempo ficou negligenciado; porém, os autores argumentam que a duração entre dois *negócios* transmite informações relevantes ao mercado: a longa duração entre *negócios* (baixa intensidade de negociação) significa que nenhuma informação nova está influenciando o comportamento do mercado e que, portanto, a probabilidade de se negociar com um *trader* informado é baixa. Neste contexto, o formador de mercado reduz o seu *spread* exigido para negociar. Isto implica uma relação inversa entre o *spread* de compra e venda (*bid-ask spread*) e a duração (tempo entre *negócios*).

A microestrutura de mercado vem desenvolvendo estudos para a mensuração dos custos de execução, liquidez, formação de preço e comparação de diferentes mecanismos de transparência de mercado.

Como alternativa às técnicas que utilizam tempo com intervalos idênticos, Engle e Russel (1997, 1998) propuseram os modelos ACD para modelar o tempo decorrido – duração – entre dois eventos sucessivos no mercado, tal como a compra ou venda de um ativo, condicional a variáveis observáveis de microestrutura do mercado.

Conforme publicado por Goodhart e O'Hara (1997), o tempo entre eventos intradiários no mercado desempenha um papel fundamental no entendimento sobre o processamento de informações públicas e privadas nos mercados financeiros. Em finanças, os modelos de Diamond e Verrecchia (1987) e de Easley e O'Hara (1992) forneceram uma justificativa teórica para o desenvolvimento de modelos de análise de séries temporais para avaliação da intensidade de negócios e da chegada de ordens. Para Pacurar (2006), os modelos autoregressivo de duração condicional constituem ferramentas econométricas eficientes para a análise de intervalos de tempo entre eventos, pois podem ser modificadas para análises e hipóteses sobre a

microestrutura de mercado, e têm sido utilizadas exclusivamente na análise de dados financeiros de alta frequência.

Os trabalhos inaugurais de Engle e Russell (1998) testaram as hipóteses de microestrutura do mercado em ativos negociados no mercado de ações da NYSE (*New York Stock Exchange*), e os testes de hipóteses de microestrutura para ações têm uma importante sustentação no papel do formador de mercado (*Market maker*). No Brasil, a negociação de derivativos não é, exceto para o mercado de *Brazilian Reais* e *Euros* (BRL / EUR), suportada por este tipo de participante.

Entretanto, os trabalhos de Holder, Qi e Sinha (2004), Eom e Hahn (2005) e, especialmente, Taylor (2004) abriram espaço para os testes de microestrutura com dados de alta frequência nos mercados futuros e de índice.

Taylor (2004) corroborou os resultados já observados em outros estudos de microestrutura de mercado, para os quais o *spread* de compra e venda (*bid/ask*) e o volume transacionado têm um impacto significativo na intensidade de negociação subsequente. Além disso, Taylor (2004) identificou que a ocorrência de diferenças entre o preço de mercado e o preço teórico de contratos futuros (erros de precificação) induzem a alterações na intensidade de transação de períodos subsequentes.

Bagehot (1971) apresentou uma teoria para explicar que a formação de preços é um processo dependente do equilíbrio informacional entre os agentes econômicos. Um resumo de elementos importantes da microestrutura de mercado pode ser encontrado em O'Hara (1995), uma pesquisa detalhada da literatura teórica sobre a microestrutura de mercado, fundamentando economicamente a importância do fluxo de informações na dinâmica de preço. Dentro da perspectiva lançada por O'Hara (1995), está a ideia de que a eficiência informacional do mercado é o objetivo central do desenho de um mecanismo de negociação, isto é, quão bem e quão rápido o mercado agrega as informações disponíveis no preço dos ativos.

Com efeito, O'Hara (1995) apresentou o desenvolvimento teórico de modelos de negociação baseados em informação, dividindo os participantes de mercado em duas classes: os *traders* informados e os não informados. Do ponto de vista da racionalidade dos agentes econômicos, os agentes informados vão ao mercado no

momento que suas informações privadas apresentam maior valor. Por isso, podemos identificar *clusters* de transações – períodos de maiores taxas de chegada de ordens. Em seu estudo, O'Hara (1995) afirmou que o *framework* Walrasiano de formação de preços não captura o custo do imediatismo presente nos mercados de capitais, isto é, não trata do prêmio que um participante do mercado deseja oferecer para induzir sua contraparte a negociar imediatamente.

A análise das microestruturas de mercado por meio das transações de alta frequência é marcada pelo fato de que elas ocorrem em intervalos irregulares de tempo, e os dados de interesse são, muitas vezes, multivariados. Motivados por essas considerações, Engle e Russell (1998) apresentaram o ACD, cujo objetivo é a modelagem de tempos entre os eventos. Desde sua concepção, o modelo ACD e suas extensões tornaram-se uma ferramenta importante na modelagem de eventos em que o tempo é irregularmente espaçado, abrindo a porta para a evolução teórica e empírica. A informação sobre o intervalo de tempo entre as transações efetivamente realizadas incorpora a ideia de frequência à modelagem, uma vez que os intervalos de tempo entre os negócios são modelados como variáveis aleatórias que seguem um processo estocástico pontual.

Engle e Russell (1997) aplicaram o modelo de duração condicional para a previsão da frequência das alterações de preços nas cotações do mercado de moeda que chegam ao terminal Reuters. Cada duração dos preços neste modelo pode ser interpretada como o tempo em que os *Market makers* respondem a novos eventos de informação no mercado.

Na formulação de Engle e Russel (1998), os modelos de duração condicional estabelecem uma relação de dependência entre a variável de interesse e suas realizações passadas. Se outras variáveis estocásticas explicativas que expressam a influência do ambiente econômico externo forem incorporadas aos intervalos de tempo entre as transações financeiras, o processo estocástico será chamado “processo pontual marcado”.

Engle e Russell (2002) evoluíram o modelo original com o objetivo de analisar processos multivariados, aplicando o modelo para prever a duração entre uma transação e uma nova cotação subsequente. O desafio dos autores foi modelar a

distribuição da duração entre cotações quando uma nova informação chega – um novo negócio - dentro da duração de interesse.

Bauwens e Giot (2000) desenvolveram uma versão logarítmica dos modelos de duração condicional, o log-ACD, a qual evita a necessidade de adoção de restrições de não negatividade da variável de interesse da formulação inicial de Engle e Russel, e aplicaram a sua formulação em três ativos altamente negociados na *New York Exchange* (NYSE), BOEING, DISNEY e IBM, encontrando evidências empíricas favoráveis aos modelos de informação de Easley e O'Hara (1992).

Com o objetivo de permitir que a duração esperada dependa não linearmente de informações passadas, Zhang, Russell e Tsay (2001) elaboraram uma extensão do modelo, chamada *Threshold Autoregressive Conditional Duration* (TACD), o que lhes permitiu uma identificação de quebras estruturais em períodos de transação de ativos da NYSE. Todos os pontos de quebra identificados foram satisfatoriamente explicados por eventos econômicos.

Fernandes e Grammig (2005) desenvolveram uma família de modelos autoregressivos de duração condicional utilizando a transformação Box-Cox com o parâmetro de duração condicional. Os autores definiram condições para a existência e possibilidade de curvas de choques assimétricos, também utilizando como fonte de informação dados de transação da NYSE, com especial atenção à duração de preços da IBM.

Engle e Lunde (2003) formularam um modelo bivariado de análise conjunta das negociações e cotações de uma série de ativos negociados na NYSE. A contribuição principal do trabalho foi a identificação de que as variáveis presentes no fluxo de informação, como altas taxas de negociação, grandes volumes e *spreads* maiores, têm grande capacidade preditiva para a velocidade de revisão de preços.

Manganelli (2005) apresentou uma evolução na utilização de modelos de duração condicional ao criar um *framework* analítico para a duração de preços, volume e retorno, ao incorporar os efeitos de causalidade entre estas variáveis, com importantes resultados na identificação de agrupamentos de volumes de negociação, presença de *traders* informados e convergência ao preço de equilíbrio.

Raposo (2006) apresentou uma extensa contribuição no estudo e aplicação das técnicas e modelos estendidos de duração condicional em conjunto com modelos *probit* e *logit* para a previsão de preços da IBM, também de um ativo do mercado americano.

Bortoluzzo et al. (2008) generalizaram o modelo de duração condicional autoregressivo para o caso de parâmetros do modelo variando no tempo por meio de ondaletas, permitindo a modelagem de processos não estacionários sem a transformação preliminar dos dados. Além do tratamento estatístico, o trabalho de Bortoluzzo et al. (2008) é uma das raras referências que aplicaram o modelo de duração condicional em uma série de ativos negociados no mercado brasileiro.

Como indicado neste capítulo, Engle e Russell (1998) introduziram os modelos ACD para lidar com os espaçamentos irregulares de dados no tempo e, conseqüentemente, testar algumas hipóteses sobre o mercado e a influência da microestrutura. Enquanto a maioria dos estudos concentra-se em ativos da NYSE ou em mercados de moeda, o presente trabalho analisa os dados de alta frequência do mercado brasileiro de derivativos, a fim de verificar se a relação intertemporal da duração entre *tradings*, o volume negociado e o *spread* entre os preços de compra e venda dos minicontratos de índice futuro do Ibovespa corroboram as teses de microestrutura de mercado.

3. Microestrutura do mercado e contratos de futuro de índice

A Teoria de Microestrutura de Mercado analisa os mecanismos e regras específicos de negociação com o objetivo de modelar como eles influenciam na formação de preços e desenvolvimento do mercado. Os principais modelos utilizados nessas análises são controle de inventário, inicialmente estudado por Stoll, Amihud e Mendelson (1980), modelos de informação, especialmente de seleção adversa, proposto por Copeland e Galai (1983), e modelos de especialistas e formadores de mercado, foco dos estudos de Brock e Kleidon (1992). Muitos trabalhos na literatura têm como alvo os mercados de ações e opções com base nesses modelos, mas poucos vinham estudando a microestrutura dos mercados futuros, devido a uma limitação do volume de transações cotações, fato já não mais relevante em mercados de alta frequência, conforme Tse (1999).

Entretanto, os mecanismos de formação de mercado e capacidade informacional do fluxo de ordens são diferentes nos mercados futuros se comparado com o mercado de ações. A primeira diferença é que não há o papel de formadores de mercado na negociação de contratos futuros. Todos os lances e ofertas são revelados a todos os participantes do pregão pelo sistema de leilão contínuo de negociação. Os mercados futuros contam também com a presença de especuladores, ou *traders* de curto prazo, que o negociam por interesses próprios e atuam como formadores de mercado "voluntários", já que eles proveem liquidez para as ordens neste mercado. Ao contrário do papel dos especialistas na New York Stock Exchange (NYSE), ou os *dealers* da Nasdaq, ou formadores de mercado da BM&FBovespa, os especuladores não têm obrigação de fazer ofertas continuamente, ou mesmo de permanecer em qualquer posição particular dentro de sua estratégia de negociação.

A segunda diferença, apontada por Manaster e Mann (1996), é que os custos de manutenção das posições nos mercados futuros são menores, pois os *traders* entendem que esse mercado é de fácil controle do inventário. Para os autores, o *trader* médio de contratos futuros da S&P 500 futuro pode reduzir seu estoque de ativos pela metade em um único negócio, enquanto um especialista leva uma semana para fazê-lo no mercado de ativos. Além disso, o especulador detém posições em aberto para um curto período de tempo (poucos minutos) e negocia em

pequenas quantidades (poucos contratos) em uma atividade bastante intensa de negociação.

A terceira diferença, sugerida por Subrahmanyam (1991), é que os preços dos contratos de futuros de índice não são afetados por informações privadas sobre empresas com ações negociadas, já que a diversificação de ativos na composição da carteira de índice reduz os efeitos de informações privadas e favorece a negociação de tais contratos. Portanto, o problema da seleção adversa induzida por informações assimétrica não é relevante nesse mercado. Por outro lado, é necessário notar que a ausência de informações assimétricas não implica que o fluxo de ordens no mercado futuro não carregue nenhuma informação. Miller (1990) apontou que os preços futuros são influenciados por informações mais abrangentes sobre o mercado, tais como notícias macroeconômicas; no entanto, a conjectura de Miller não pode ser facilmente testada nos mercados futuros de índice com base em informação de tempo real, já que a maioria das notícias macroeconômicas é relatada antes dos mercados serem abertos.

Para Tse (1999), tais características do mercado futuro de índice evidenciam que o modelo de especialistas, de controle de inventário, e de seleção adversa não são úteis para descrever a microestrutura dos mercados futuros, em particular para descrever os padrões intradiários de *spread* de compra e venda, da volatilidade, do volume negociado e do número de negócios. No entanto, os períodos de alta volatilidade e alto volume de negociação são oportunidades de lucro para os *traders* e, portanto, estes podem estar dispostos a assumir posições em *spreads* menores durante os períodos de abertura e de encerramento do pregão. Seguindo a abordagem de Gerety e Mulherin (1992), o agrupamento de volume e número de negócios ao abrir e ao fechar o mercado é uma evidência do desejo dos investidores de negociar o risco de manter posição durante a noite, isto é, os *spreads* de fechamento diminuem (aumentam) porque os *traders* não estão dispostos a manter os inventários e assumir os riscos de notícias desfavoráveis durante a noite, pois, para os autores, no mercado americano, as notícias macroeconômicas podem alterar os preços futuros rapidamente de forma significativa.

Contudo, os principais modelos da literatura utilizados para estudar a microestrutura do mercado de contratos futuros de índice são consistentes com a posição tomada por *traders* ativos, particularmente especuladores, que abastecem os mercados de futuros com liquidez, conforme sugerido por Grossman e Miller

(1988). Os autores defenderam que os investidores, em geral, preferem negociar imediatamente, e por isso, modelaram a ideia de liquidez como o preço que investidores estão dispostos a pagar pela necessidade de executar um negócio imediatamente ao invés de aguardar uma nova oportunidade mais tarde.

Em mercados competitivos, o imediatismo de execução é possibilitado por *traders* que enviam ofertas de compra e venda de forma contínua. Para Grossman e Miller (1986), a alta demanda pelo imediatismo nos mercados futuros está relacionada com as duas funções básicas destes mercados, que são a formação de preços e *hedging* de riscos: quanto mais voláteis são os preços dos contratos futuros, maior o risco de não negociar o ativo *spot* imediatamente e, portanto, maior a importância das funções do mercado futuro.

Nas praças de liquidez, onde as ofertas são transmitidas por sinais eletrônicos, cada *trader* age como um leiloeiro, e este pregão permite a competição aberta entre eles. Entretanto, os especuladores têm um horizonte de manutenção das suas posições menores (minutos ou segundos), enquanto que os investidores mantêm posições de mais longo prazo, como um dia ou mais. Neste sentido, os especuladores agem como *market makers* ou *dealers* e fornecem liquidez continuamente, enviando ofertas a mercado contra as ordens que podem ser executadas, e são identificados informalmente por outros *traders* pela simples observação de seu comportamento de negociação. Os especuladores não têm a obrigação de oferecer liquidez, nem mesmo permanecer presentes em qualquer posição ou praça de negociação em particular. Kuserk e Locke (1993) documentaram, em seus estudos, que uma estratégia bem sucedida de especulação envolve negociação frequente de pequenos volumes, e que a rentabilidade do especulador vem da diferença do *spread* de compra e venda e da sua previsão do movimento dos preços intradiários.

Para Grossman e Miller (1986), o modelo que melhor explica os mercados futuros é aquele em que o negócio ocorre em dois períodos e o ativo é liquidado no terceiro período. O que motiva a negociação é sempre o rebalanceamento da carteira do investidor induzida por choques exógenos de liquidez. A assincronia dos negócios cria a demanda por imediatismo no Período 1, com o desequilíbrio do portfólio no período 2. Assim, a mudança do preço entre os períodos 1 e 2 pode ser considerado o retorno, r , para um especulador que está disposto, por um dado preço, em manter portfólio não balanceado. O retorno esperado $E[r]$ é dado por

$$E[r] = (\text{posição do especulador}) \cdot (\text{coeficiente de aversão ao risco}) \cdot (\text{variância de } r) \quad (4.1.1)$$

Como o retorno dos especuladores aumenta com a variância do retorno, períodos de maior volatilidade irão atrair mais especuladores. Entretanto, eles exigem um retorno maior para compensar o maior risco. Grossman e Miller (1990) argumentaram que, em mercados competitivos, o risco suportado por qualquer formador de mercado, ou especulador, pode ser reduzido através da diversificação do risco por participação de muitos formadores de mercado. Em outras palavras, quanto maior o número de *traders*, menores são os custos para cada *trader* assumir os riscos de variação temporária de preços. Além disso, quanto maior o inventário, maior o retorno esperado entre os períodos para compensar o formador de mercado pelo risco.

Segundo os autores, a discussão acima sugere que a volatilidade e o volume devem ser inversamente relacionados com o *spread* de compra e venda. Especificamente, os especuladores estão ansiosos para negociar nos períodos mais voláteis, como os intervalos de abertura e encerramento do pregão, durante os quais os investidores têm maior necessidade de liquidez e, por isso, os especuladores estão dispostos a enviar ofertas de compra e venda para atrair fluxos de ordens, e, com isso, aumentar a intensidade de negociação.

4. Metodologia

Com a crescente automatização dos mercados financeiros, encontram-se disponíveis bases de dados intradiários que registram todas as transações de um dia de pregão junto às variáveis de interesse, o que vem impulsionando o desenvolvimento de uma nova área de pesquisa financeira conhecida como finanças em alta frequência, a qual utiliza técnicas de finanças, computação, e econometria de séries de tempo para a compreensão mais profunda da atividade do mercado de capitais.

Para lidar com essa nova realidade de dados e viabilizar os testes de microestrutura de mercado, os modelos de duração condicional buscam identificar a dinâmica que envolve o processo de convergência entre os preços de compra e venda, e o procedimento adotado toma por base a sequência de negócios fechados, ou ordens, e a informação relevante passa a ser o intervalo de tempo entre as transações efetivadas.

Neste capítulo, será desenvolvida uma revisão sobre os fundamentos teóricos e metodológicos da aplicação do modelo autoregressivo de duração condicional para os testes de microestrutura do mercado de minicontratos de futuro de índice.

4.1. Estudo de séries temporais de alta frequência

A especificação básica dos modelos de duração condicional estuda uma relação de dependência defasada entre a variável de interesse e suas realizações passadas, sendo que as demais variáveis exercem influência na dinâmica. Aos intervalos de tempo entre as transações financeiras, podem associar-se variáveis como volume, *spread* de compra e venda, e preço, adicionando um caráter explicativo à dinâmica temporal. Um fenômeno bem conhecido na literatura diz respeito ao comportamento sazonal intradiário, o qual cria agrupamentos de eventos (*clusters*), devendo ser tratados em qualquer estudo dessas séries.

Uma vez que a quantidade negociada e o preço em um dado período de tempo são variáveis-chave, o tratamento da dinâmica das negociações no tempo torna-se fundamento metodológico. Como já referido, os dados de transações financeiras

chegam inerentemente em intervalos de tempo irregulares, entretanto, as técnicas econométricas são baseadas, tradicionalmente, em estudos macroeconômicos, na análise em tempo fixo de intervalo.

Quando se trata de dados de ativos negociados em alta frequência, as operações ocorrem em intervalos de poucos segundos, o que induz o pesquisador a selecionar intervalos curtos para análise. Todavia, se o intervalo de tempo escolhido for muito curto, haverá muitos intervalos sem nenhuma informação nova e a heterocedasticidade será introduzida nas estatísticas. Em contrapartida, se um intervalo de tempo muito longo for escolhido, as características de microestrutura dos dados serão perdidas. O problema se torna ainda mais complicado quando se percebe que a taxa de chegada de dados e o tipo de transação podem variar ao longo do dia, em particular na abertura e fechamento do pregão.

Um fenômeno típico e intrigante é o caso de transações que geralmente são pouco frequentes, mas que repentinamente podem exibir uma atividade muito elevada. Nestes casos, a escolha de um intervalo fixo para a análise dos dados é equivocada, pois pode deixar o investigador com uma amostra sem informação suficiente.

Os intervalos irregulares de tempo, que são tratados na estatística como processos pontuais, possuem uma literatura bastante ampla e interdisciplinar de aplicações e desenvolvimento teórico. O exemplo talvez mais conhecido seja a análise da dinâmica de terremotos utilizando processos pontuais.

Mais formalmente, considere um processo estocástico que corresponda à sequência de instantes crescentes de tempo $\{t_0, t_1, \dots, t_n\}$. Para cada instante de tempo, associa-se uma função de contagem $N(t)$, a qual representa o número de eventos associados ao instante de tempo t . Caso existam informações adicionais associadas aos diferentes instantes de tempo, tais como preço ou volume, este processo será denominado “*marked point process*”. No estudo das séries financeiras associadas aos instantes de tempo de chegada estão informações como preço, volume, oferta de compra, oferta de venda, *spread*, profundidade do livro de ofertas, entre outras.

Um determinado processo pontual definido no espaço $[t_0, \infty)$ tem sua evolução sem efeitos passados se, para qualquer $t > t_0$, as realizações associadas aos instantes de tempo $[t, \infty)$ não apresentarem estrutura de dependência temporal com os eventos do espaço $[t_0, t)$. Por outro lado, um processo de contagem será dito condicionalmente ordenável no instante $t > t_0$ se, para um intervalo de tempo suficientemente curto, e de maneira condicional a um dado evento P , a probabilidade de dois ou mais eventos ocorrerem for pequena se comparada à probabilidade de ocorrência de um único evento.

Como já descrito, os processos pontuais são ferramentas estatísticas convenientes para a estrutura das séries financeiras pela capacidade de avaliar intervalos irregulares de tempo e de associação com vetores M-dimensionais de variáveis, *marks*, os quais representam informações associadas ao tempo do evento.

Contudo, dependendo da questão econômica em mãos, tanto o tempo de ocorrência de um evento (*arrival time*), como os *marks* são variáveis de interesse. Nesse caso, as questões econômicas podem ser elaboradas em um *framework* de expectativa condicional de valores futuros, por meio de um processo de filtração dos tempos de chegada de informação, definidos como $\hat{t}_i = \{t_i, t_{i-1}, \dots, t_0\}$, e dos vetores M-dimensional de marcadores (*marks*) no instante de tempo do $i^{\text{ésimo}}$ evento, definido $\hat{y}_i = \{y_i, y_{i-1}, \dots, y_0\}$. A estrutura de probabilidade para a dinâmica de um processo pontual estacionário e marcado pode ser completamente caracterizado e expresso como a distribuição conjunta de marcadores e instantes de tempo dos eventos dados pela filtração dos marcadores e instantes de chegada passados:

$$f(t_{N(t)+1}, y_{N(t)+1} | \hat{t}_{N(t)}, \hat{y}_{N(t)}) \quad (4.1.1)$$

Todavia, para responder perguntas econômicas de interesse, como quando o próximo evento vai ocorrer, ou qual o valor do marcador associado no próximo evento ou após um intervalo de tempo, ou ainda quanto tempo deve-se aguardar para se observar novamente o mesmo tipo de evento, algumas considerações devem ser feitas. Questões formuladas que não se concentram na relação contemporânea entre os marcadores e o instante de tempo têm tratamento facilitado pela restrição do foco na distribuição de probabilidade marginal para a condição em que os marcadores e tempo de chegada são fracamente exógenos. Se o tempo de

espera até o próximo evento, independente do valor dos marcadores é a variável de interesse, então a distribuição acumulada é dada pela seguinte equação, podendo ser analisada como:

$$f_t(t_{N(t)+1}|\hat{t}_{N(t)}, \hat{y}_{N(t)}) = \int_0^{+\infty} f(y, t_{N(t)+1}|\hat{t}_{N(t)}, \hat{y}_{N(t)})dy \quad (4.1.2)$$

Em outras palavras, este é um processo pontual que depende apenas das observações defasadas das chegadas e marcadores. Entretanto, se as hipóteses econômicas envolvem a dinâmica de marcadores, como o *spread* ou preço, o valor do próximo marcador no tempo é a variável de interesse independente de quando o evento ocorra. Neste caso, pode-se utilizar a formulação:

$$f_t(t_{N(t)+1}|\hat{t}_{N(t)}, \hat{y}_{N(t)}) = \int_{-\infty}^{+\infty} f(t, y_{N(t)+1}|\hat{t}_{N(t)}, \hat{y}_{N(t)})dy \quad (4.1.3)$$

Aqui, o conjunto de informações disponível é atualizado a cada novo instante de tempo de chegada de evento, além de outras formulações, como a previsão dos marcadores para um futuro determinado, Δt , ou mesmo para saber quando um determinado evento vai ocorrer.

4.2. Modelos de Duração Condicional

Para lidar com a dinâmica de chegada de informações no mercado financeiro, um novo modelo para processos pontuais dependentes foi formulado por Engle e Russell (1998) por meio de uma função de intensidade condicional, a qual pode ser parametrizada em termos de eventos passados de uma maneira particularmente adequada para o processo de transações.

A aplicação mais direta desta família de modelos é medir e prever a intensidade de chegadas de ordens, consistindo, essencialmente, na quantidade instantânea de transações. A formulação básica do modelo parametriza a intensidade condicional em função do tempo entre eventos passados; entretanto

numerosas extensões naturais incluem outros efeitos, como características associadas a operações passadas, ou qualquer outra influência externa.

A descrição completa deste tipo de processo deve ser formulada em termos da função de intensidade condicional a toda informação passada disponível, a qual deverá conter, pelo menos, os instantes de tempo e a contagem a esta associada. Dessa forma, o processo de intensidade condicional pode ser definido por:

$$\lambda(t|N(t), t_1, \dots, t_{N(t)}) = \lim_{\Delta t \rightarrow 0} \frac{P(N(t + \Delta t) > N(t) | N(t), t_1, \dots, t_{N(t)})}{\Delta t} \quad (4.2.1)$$

O mais simples exemplo de processo pontual que evolui com dependência temporal é o processo homogêneo de primeira ordem, em que $\lambda(t|N(t), t_1, \dots, t_{N(t)}) = \lambda(t|N(t), t_{N(t)})$ e a duração entre eventos $x_i = t_i - t_{i-1}$ formam uma sequência de variáveis aleatórias independentes. Todavia, talvez o processo pontual mais conhecido seja o processo de Poisson homogêneo, em que a probabilidade de um evento ocorrer em um dado instante é constante, isto é, $\lambda(t) = \lambda$, sendo a “taxa de chegada” uma função determinística do tempo.

Entretanto, no estudo das séries financeiras podemos observar períodos de explosões de atividades (*bursts*) seguidos de períodos de calma¹. Conforme Pacurar (2006), tal comportamento, chamado de *clustering*, pode ser considerado como um fato estilizado em estudos de séries temporais e estatisticamente comprovado pela análise de autocorrelação serial nas séries de intervalos entre dois eventos. No estudo de modelos de duração condicional, a função de intensidade é definida como uma função de t através de vários eventos, diferente da literatura macroeconômica, a qual tende a concentrar a análise em largas bases de informação com o mesmo intervalo de tempo.

Os processos de intensidade condicional são também comumente referenciados na literatura econométrica como “*hazard function*” (ou taxa de falha, taxa de risco, taxa de mortalidade, em literaturas correlatas), uma vez que representa a razão da função densidade de probabilidade $P(x)$ da variável de interesse pela sua função de confiabilidade (*survival function*) $S(x)$:

¹ A falta de previsibilidade deste tipo de comportamento é um grande incômodo para os profissionais envolvidos com o desempenho de sistemas de negociação eletrônica.

$$h(x) = \frac{P(x)}{S(x)} \quad (4.2.2)$$

Conforme discutida por Snyder e Miller (1975), a função de intensidade condicional das durações, ou *conditional survival function*, e a duração, ou “tempo de espera”, entre eventos descrevem completamente um processo estocástico pontual condicionalmente ordenável.

Tomando p_i como uma família de funções de densidade de probabilidade condicional para o instante de tempo t_i , o valor da função log-verossimilhança, em termos da densidade condicional ou intensidade, é representada como:

$$L = \sum_{i=1}^{N(T)} \log p_i(t_i | t_0, \dots, t_{i-1}) \quad (4.2.3)$$

$$L = \sum_{i=1}^{N(T)} \log \lambda(t_i | i-1, t_0, \dots, t_{i-1}) - \int_{t_0}^{N(T)} (u | N(u), t_0, \dots, t_{i-1}) du \quad (4.2.4)$$

A equação (4.2.2) corresponde à formulação geral da função de intensidade de um processo estocástico, no qual a evolução observada no passado influencia a estrutura de probabilidade dos eventos futuros. Tal formulação foi originalmente proposta por Hawkes (1972), a partir da qual inúmeras parametrizações vêm sendo propostas na literatura.

A classificação a respeito da memória do modelo pode ser definida em função da dependência que a intensidade tem do número de eventos, mas não do instante de tempo. Nesse caso, o processo estocástico é chamado de “*pure birth process*”.

Rubin (1972) introduziu os chamados “processos de memória finita” (*m-memory process*), cuja definição advém do fato de que apenas os “m” instantes de tempo mais recentes, correspondentes a realizações de determinado evento, se mostram presentes na função de intensidade condicional.

No que diz respeito aos processos de memória “longa”, há diversas formas de se parametrizar a função de intensidade condicional, das quais podemos destacar:

- Modelo de sequência de eventos que trata do intervalo de tempo verificado:

$$\lambda(t|N(t), t_1, \dots, t_{N(t)}) = w + \sum_{i=1}^{N(t)} \pi_i(t_{N(t)+1-i} - t_{N(t)-i}) \quad (4.2.5)$$

Esta formulação foi introduzida por Wold (1948) e, posteriormente, por Cox (1955). Wold propôs um modelo para intervalos correlacionados utilizando uma estrutura autoregressiva similar aos modelos Autoregressivos com Médias Móveis (*Autoregressive Moving Average* - ARMA) tradicionais. Tais modelos foram reformulados subsequentemente por Gaver e Lewis (1980), Lawrence e Lewis (1980) e Jacobs e Lewis (1977) como modelo exponencial autoregressivo de médias móveis – EARMA(p, q). Nesta formulação, assume-se que a duração condicional seja exponencialmente distribuída, sendo que a média do processo segue um processo ARMA. A formulação de um processo aditivo para o ruído com estas propriedades se mostra bastante complexa e os procedimentos de estimação (máxima verossimilhança) virtualmente intratáveis.

Algumas variações podem ser encontradas na literatura, como é o caso das derivações de Cox (1955) com os “modelos duplamente estocásticos”, nos quais o pressuposto fundamental diz respeito à existência de um processo independente que coordena o comportamento da taxa de chegada da informação.

Cox (1972) também contribuiu com a primeira proposta de análise que inclui durações defasadas como variáveis explicativas, sendo, posteriormente, generalizada como *proportional hazards framework*:

$$\lambda(t|z_{N(t)}, \dots, z_1) = \lambda(t) \exp\{\beta' z_{N(t)}\} \quad (4.2.6)$$

para a qual z_i é um vetor de variáveis explicativas associado ao instante de tempo i de chegada da informação. Uma das alternativas de modelagem já propostas por Cox (1972) foi a da inclusão de durações defasadas como variável explicativa, o que foi seguido pelas evoluções de Lancaster (1979) e Heckman

(1981).

4.3. Modelo de Duração Condicional Autoregressivo

Elaborado por Engle e Russell (1998), especificamente para a análise de dados de alta frequência, o modelo ACD é convenientemente especificado em termos da densidade condicional das durações entre eventos. Tomando $x_i = t_i - t_{i-1}$ como o intervalo de tempo entre duas transações financeiras (duração), os autores modelaram a densidade de probabilidade de x_i condicional às observações passadas, obtendo-as de maneira direta:

Seja ψ_i a esperança da i -ésima duração, teremos:

$$E(x_i | x_{i-1}, \dots, x_1) = \psi_i(x_{i-1}, \dots, x_1; \theta) \equiv \psi_i \quad (4.3.1)$$

Assim, a classe de modelos ACD consistirá na parametrização da equação (4.3.1) e no pressuposto de que a variável de interesse irá evoluir segundo o seguinte processo estocástico:

$$x_i = \psi_i \cdot \varepsilon_i \quad (4.3.2)$$

em que, $\{\varepsilon_i\} \sim \text{i.i.d.}$

Assim, podem-se propor inúmeras especificações diferentes para os modelos de duração condicional, mediante a adoção de especificações diferentes da duração esperada (ψ_i) e da distribuição dos resíduos (ε_i).

Com efeito, para derivarmos um modelo geral para a densidade condicional, tomando p_0 como a função densidade de probabilidade dos resíduos e a função de confiabilidade (“*survival function*”), podemos definir a taxa de intensidade condicional base (*baseline hazard*):

$$\lambda_0 = \frac{p_0(t)}{S_0(t)} \quad (4.3.3)$$

a partir da qual podemos estender a intensidade condicional do modelo ACD como:

$$\lambda(t|N(t), t_1, \dots, t_{N(t)}) = \lambda_0 \left(\frac{t - t_{N(t)}}{\psi_{N(t)+1}} \right) \cdot \frac{1}{\psi_{N(t)+1}} \quad (4.3.4)$$

Desta forma, os eventos passados influenciam a intensidade condicional por efeitos multiplicativos. Para as situações em que o investigador não está interessado em modelar o tempo entre as operações, mas sim em estudar os eventos associados e que influenciem os tempos de chegada, o modelo ACD também fornece uma estrutura para testar se a função intensidade é influenciada por variáveis observadas de mercado, os marcadores.

4.4. Especificações do modelo de Duração Condicional Autoregressivo

A principal motivação para os modelos ACD é a observação empírica de que o envio de ordens e negócios fechados ocorre em *clusters*, isto é, apresentam comportamentos de operações acumuladas no tempo, com períodos de grande intensidade e períodos de baixa intensidade. Este efeito foi identificado nos dados da NYSE por Engle e Russel (1998), e motivou os autores a tratarem estatisticamente as séries de tempo de durações com uma especificação inspirada nos modelos ARCH. Entretanto, a diferença entre os modelos reside no fato de que as durações, para os modelos ACD, não podem conter valores negativos, o que representa uma restrição acerca da distribuição de probabilidades requerida para a análise.

4.5. Modelo Duração Condicional Autoregressivo Exponencial

Na especificação mais simples, o modelo ACD assume que as durações são condicionalmente exponenciais, de forma que a taxa de intensidade condicional base é unitária e a intensidade condicional pode ser especificada como:

$$\lambda(t|N(t), t_1, \dots, t_{N(t)}) = \frac{1}{\psi_{N(t)+1}} \quad (4.5.1)$$

A partir daí, as variações das especificações dos modelos ACD estão baseadas em dois componentes:

$$\frac{x_i}{\psi_i} = \varepsilon_i \sim i.i.d \quad (4.5.2)$$

em que ε_i é uma distribuição geral no espaço $(0; \infty)$ e $E(x_i) = 1$. Além disso,

$$\psi_n = \omega_0 + \sum_{m=1}^p \alpha_m x_{n-m} + \sum_{m=1}^q \beta_m \psi_{n-m} \quad (4.5.3)$$

em que ψ_n é linearmente dependente das observações passadas da duração intensidade, com defasagens² p e q , sendo chamado modelo ACD(p, q).

A especificação do modelo ACD permite o cálculo dos momentos estatísticos, qualquer que seja a intensidade condicional base. Pela definição (4.3.1), vimos que a média condicional de x_i é igual à duração condicional ψ_i , mas a média não condicional de x_i é definida como:

$$E(x_i) = \mu = \frac{\alpha_0}{(1 - \sum_{m=1}^p \alpha_m + \sum_{m=1}^q \beta_m)} \quad (4.5.4)$$

De (4.3.2) podemos deduzir que a variância condicional de x_i é definida por:

$$Var(x_i|x_{i-1}, \dots, x_1) = \psi_i^2 Var(\varepsilon_i) \quad (4.5.5)$$

Assumindo a distribuição exponencial para os resíduos e defasagens de um período, a especificação mais comum na literatura é a família EACD, representada como:

$$\psi_i = \omega + \alpha x_{i-1} + \beta \psi_{i-1}, \text{ para qualquer } i \text{ e } \alpha, \beta > 0, \omega > 0. \quad (4.5.6)$$

² Nos modelos ACD (memória “ilimitada”), ACD (p, q), não existe um limite finito para a dependência entre eventos subsequentes, portanto, as defasagens “ m ” e “ q ” podem assumir qualquer valor positivo. Os modelos podem ser restritos ao ACD (p -memory) que apenas as “ p ” durações mais recentes da duração condicional são consideradas, e $q=0$.

Um recurso muito útil do modelo ACD é que o modelo autoregressivo também pode ser formulado como um processo ARMA (p,q) para as durações. Tomando $\eta_i = x_i - \psi_i$ como uma diferença martingal por construção, podemos definir o processo autoregressivo ARMA $(\max(p,q),q)$ de estimação com inovações não-Gaussianas:

$$x_i = w + \sum_{j=1}^{\max(p,q)} (\alpha_j + \beta_j)x_{i-j} - \sum_{j=1}^q \beta_j \eta_{i-j} + \eta_i \quad (4.5.7)$$

Importante notar que a persistência dessa formulação EARMA³ pode ser medida pela expressão $\alpha(1) + \beta(1) < 1$ para a maioria dos casos.

Como fica evidente, o modelo ACD e o modelo GARCH de Bollerslev (1986) compartilham várias características comuns. As duas famílias de modelos contam com a mesma premissa econômica de que os eventos nos mercados financeiros têm agregações (*clustering*) correlacionadas com notícias externas.

A modelagem autoregressiva da equação (4.5.7) permite a captura da duração observada no agrupamento de dados de alta frequência, isto é, pequenas (grandes) durações sendo seguido por outras pequenas (grandes) durações, de uma forma muito semelhante ao tratamento que o modelo GARCH dá para o agrupamento de volatilidade.

Portanto, da mesma forma que um modelo GARCH (1,1) com distribuição exponencial é frequentemente suficiente para remover a dependência temporal dos retornos, um modelo ACD de baixa ordem também é frequentemente bem sucedido na remoção da dependência temporal na duração. Variações dos modelos ACD são apresentadas no Anexo – Extensões do modelo ACD.

³ Tal formulação pode ser modelada como EARMA (*Exponential ARMA*) se as durações condicionais forem exponenciais.

5. Base de Dados

Conforme detalhado no capítulo anterior, os modelos ACD analisam a relação defasada entre a variável de interesse e as variáveis explicativas. Dessa forma, a pesquisa empírica exige uma base de dados adequada à aplicação dos modelos. Neste capítulo, será apresentado o ativo selecionado para o estudo e os critérios dessa seleção, bem como uma breve descrição do procedimento de construção da base de dados.

5.1. Minicontrato futuro do Ibovespa

Os modelos de duração condicional no mercado de derivativos ainda são raros na literatura, sendo mais frequentes nos ativos de câmbio. Entretanto, o presente trabalho optou por seguir Taylor (2004), que estudou a dinâmica do índice futuro FTSE 100 listado na *London Stock Exchange*, sendo que o ativo selecionado foi de minicontrato futuro do Ibovespa, negociado no segmento BM&F da BM&FBOVESPA. No período de análise, os minicontratos tiveram um volume médio de 360.000 negócios mensais. Embora o volume financeiro médio de R\$ 8 milhões seja bem inferior aos R\$ 1.4 bi de contratos de derivativos de taxas de juro, dos R\$ 800 milhões de dólares, e do R\$ 1.2 bi de taxas de câmbio, o número de negócios efetivados é maior que estes contratos, compondo uma característica fundamental para a análise da duração entre negócios efetivados, que é a intensidade de transações.

O minicontrato futuro do Ibovespa tem como ativo de referência o Índice *spot* da Bolsa de Valores de São Paulo (Índice Bovespa), sendo negociados com vencimentos nos meses pares, e cada contrato possuindo o tamanho de R\$0,20 do índice normal. Se o índice Bovespa está cotado a 50.000 pontos, pelo mini-índice um contrato equivale a $50.000 \times 0,20 = 10.000$. O vencimento do contrato futuro do Ibovespa ocorre na quarta-feira mais próxima do dia 15 dos meses pares.

A oscilação diária máxima permitida do índice é de 15 por cento sobre o valor do terceiro vencimento dos contratos futuros em aberto, calculados sobre o preço de ajuste do pregão anterior. Os dois primeiros vencimentos dos contratos futuros abertos à negociação não estão sujeitos a limites de oscilação. Para a liquidação do contrato futuro do Ibovespa na data de vencimento, as posições em aberto são liquidadas financeiramente no último pregão, mediante o registro de operação de natureza inversa à da posição (compra ou venda), na mesma quantidade de contratos, sendo utilizada a média do valor do Ibovespa à vista desta data, observada nas últimas duas horas e trinta minutos de negócios da sessão do pregão de viva voz. A esta média multiplica-se o valor (em reais) de cada ponto do índice estabelecido pela BM&F, sendo os resultados financeiros da liquidação movimentados no dia útil subsequente à data de vencimento.

As posições em aberto, ao final de cada pregão, são ajustadas com base no preço de ajuste do dia, e o preço de ajuste é determinado pela média ponderada das cotações dos negócios realizados nos últimos 30 minutos do pregão com movimentação financeira em $D + 1$ (dia útil seguinte). A margem de garantia do contrato consiste em um valor fixo por contrato, estabelecido com base na volatilidade (oscilação) do índice de referência.

5.2. Características dos dados

A construção da base de dados para o trabalho aqui desenvolvido representou um desafio particular. Os dados de alta frequência são registrados pelos sistemas de negociação e divulgados para o mercado no sinal de *Market data*, em cada transação, isto é, *tick-by-tick*, na ordem de chegada.

A utilização do método de Engle e Russell (1998) impõe a compilação de dados de alta frequência sem o agrupamento por unidade de tempo, isto é, sua aplicação depende de um conjunto detalhado de informações sobre todas as transações. A grande maioria dos trabalhos internacionais utiliza a base de dados

Negócios, Orders Reports and Quotes (TORQ) compilada por Joel Hasbrouck e NYSE com dados de Novembro de 1990 a Janeiro de 1991.

Por sua vez, os trabalhos brasileiros de microestrutura do mercado de câmbio utilizaram bancos de dados com agregações não convenientes para o estudo de modelos de duração condicional. Colussi (2008) utilizou as transações de moeda fornecidas pelo Sisbacen do Banco Central, que agrega os dados diariamente. Laurini et al. (2007) utilizaram o FAFX data do sistema *Bloomberg Data License*, que agrega os dados por segundos, todavia apresenta apenas as cotações indicativas de compra e venda, mas não as quantidades e o valor efetivamente transacionados.

Para viabilizar este trabalho e garantir dados corretos para a aplicação dos modelos aqui apresentados, optou-se por construir a própria base de dados a partir de informações do sinal incremental de *Market data*⁴ fornecido pela BM&FBOVESPA em formato *Financial Information eXchange protocol* (FIX).

O sinal de *Market data* contém informações detalhadas sobre cada ordem e transação efetuada, incluindo tipo (compra, venda ou negócio), preço, quantidade e o horário com precisão de milissegundos em que foi informada ao mercado. A ordem no livro de ofertas também é relatada, e inclui os preços de compra e venda, o número de contratos exigidos ou oferecidos.

O ativo utilizado para este estudo é o contrato WINZ08 (minicontrato de futuro do Ibovespa), com vencimento em 17/12/2008. O contrato WINZ08 é negociado na BM&FBOVESPA no sistema eletrônico de derivativos, o GTS, o qual opera com a prioridade preço e tempo de chegada da ordem. De acordo com as regras de negociação, os menores preços de compra prevalecem sobre os preços mais elevados, e o preço de venda mais elevado tem precedência sobre os preços inferiores. Se houver mais de uma oferta pelo mesmo preço, as ofertas mais antigas prevalecem sobre as mais novas. Todas as ofertas têm seus horários de chegadas das ordens registradas com precisão de milissegundos, o que permite que a determinação dos critérios de precedência seja bastante sensível.

⁴ A especificação do sinal incremental de *Market data* pode ser encontrado em <http://www.bmfbovespa.com.br/pt-br/servicos/servicos-de-informacao/sinal-de-informacoes-e-vendedores/informacoes-tecnicas.aspx?Idioma=pt-br>

Um aspecto importante em relação ao mecanismo de negociação adotado pela BM&FBOVESPA e seus impactos no mercado microestrutura é o fato de que todos os negócios nos mercados de derivativos são realizados com uma estrutura de livro aberto. Um segundo aspecto da negociação é o fato de que, em sua maioria, as ordens são do tipo limitada. Vale ressaltar que, nos mercados de derivativos, não existe o papel do *market maker*⁵.

A base de dados do presente trabalho foi construída considerando-se o período de amostragem entre 16 de outubro de 2008 a 16 de dezembro de 2008. As datas foram selecionadas por estarem após o vencimento dos contratos futuros WINV08 e anterior ao vencimento do próprio ativo WINZ08. A seleção foi feita com o objetivo de se concentrar a análise no período de negociação do contrato sem concorrência do mesmo produto com outro vencimento, e a exclusão do dia do vencimento do próprio contrato e o horário de negociação foi das 09h00 às 18h00 no horário local (12h00 – 20h00 em UTC). Não houve exclusão de datas dos 43 dias de negociação do período.

Após a consolidação de dados, temos o seguinte perfil da base de ordens e transações:

Tabela 5.1 - Perfil dos dados do ativo WINZ08

	Outubro	Novembro	Dezembro
Total de Negócios (negócios)	196.184	300.809	174.602
Média de Ordens (novas ou alterações) por dia	45.389	43.440	43.595
Média de Negócios por dia	16.300	16.800	15.870
Duração média entre negócios (milissegundos)	1.789	1.805	1.796
Duração média de negócios com alteração de valor (milissegundos)	4.337	4.045	4.157
Duração Mínima/Máxima	0/329.359	0/164.766	0/153.750

⁵ Exceção ao contrato BRL / EUR.

Com esse perfil de dados, e a partir do sinal incremental de *Market data*, foi possível construir uma base de dados⁶ representando a dinâmica de ordens e livro de ofertas com o seguinte formato:

Tabela 5.2 - Registro de ofertas e transações e topo do livro de ofertas

Sinal incremental (ofertas e negócios)							Topo do livro de ofertas			
Data	Horário	Tipo	Ativo	Qtd	Valor	Pos	Qte bid	Valor bid	Qtd ask	Valor ask
20081103	12:02:06.217	ask	WINZ08	7	38400	1	5	38750	7	38400
20081103	12:02:09.717	bid	WINZ08	8	38800	1	8	38800	7	38400
20081103	12:02:12.139	bid	WINZ08	5	38450	1	5	38450	7	38400
20081103	12:02:12.890	ask	WINZ08	1	38450	1	5	38800	1	38450
20081103	12:02:12.890	trade	WINZ08	1	38450	2	4	38450	7	38400
20081103	12:02:13.540	ask	WINZ08	5	38450	1	4	38450	5	38450
20081103	12:02:13.540	trade	WINZ08	10	38450	3	8	38800	1	38450

A primeira seção da Tabela 5.2 apresenta, de forma resumida, os eventos que estão sendo informados ao mercado pelo sinal de difusão da BM&FBOVESPA:

- Data: data do sinal incremental;
- Horário: *time stamp* do sinal no momento da saída do sistema de negociação;
- Tipo: informação se o sinal representa uma oferta de compra (*bid*), venda (*ask*) ou um negócio fechado;
- Ativo: objeto de negociação;
- Qtd: quantidade oferecida ou negociada;
- Valor: valor oferecido ou negociado;
- Pos: posição em que a ordem de compra ou venda entra no livro de ofertas;

A segunda seção da Tabela 5.2 informa a situação do topo do livro de ofertas após a atualização com a ordem correspondente. É importante notar que a dinâmica do livro de ofertas varia com a ordem que é informada ao mercado. Um detalhe operacional importante é que essa dinâmica do livro de ofertas é montada pela

⁶ Alguns detalhes do padrão do protocolo FIX e método de processamento utilizado podem ser conferidos no Apêndice 1. O horário utilizado na base de dados está em formato UTC (*Coordinated Universal Time*), portanto, 3 horas à frente do horário local na data estudada.

BM&FBOVESPA no sistema de negociação *Global Trading System* (GTS) e replicada pelos participantes de mercado, os quais utilizam sistemas próprios ou comerciais para a negociação. A BM&FBOVESPA informa apenas o sinal incremental.

Com a construção dessa base de dados, uma série de constatações ficou consistente com os trabalhos de Engle e Russell (1998). Como é possível observar, o processo de formação de preços é bivariado com a duração entre negócios e a duração entre ordens, sendo as variáveis aleatórias de interesse. Para efeito de aplicação, neste trabalho, consideramos apenas a duração entre *negócios*.

Para seguir as técnicas econométricas de Engle e Russell (1998), uma transformação na base de dados ainda se faz necessária. Como o objetivo dos modelos de duração condicional é modelar o tempo entre eventos, a duração entre um evento e outro precisa ser calculada. Seguindo a literatura, o horário da transação foi convertido em um número inteiro que representa o tempo decorrido, em milissegundos, desde a 00h do mesmo dia. No exemplo da Tabela 5.3, o horário 12:02:06.217 foi convertido no número 43326217, o que viabiliza a medida da duração entre eventos ser definida como $x_i = t_i - t_{i-1}$, em que t_i é o horário convertido do i -th evento e t_{i-1} é o horário convertido do evento imediatamente anterior. Para as primeiras transações e cotações do dia, o t_{i-1} é definido como o horário de abertura do pregão.

Com essa transformação, foi possível construir bases de dados com a duração entre cada evento por tipo.

Tabela 5.3 - Duração entre ordens em milissegundos

Tabela de duração entre ordens							
Data	Horário	Duração	Tipo	Ativo	Qty	Valor	Pos
20081103	43335736	6217	ask	WINZ08	7	38400	1
20081103	43350238	3500	bid	WINZ08	8	38800	1
20081103	43353894	2422	bid	WINZ08	5	38450	1
20081103	43354441	751	trade	WINZ08	1	38450	

Entretanto, para atender o objetivo de analisar a dinâmica da duração entre eventos de preços e seus marcadores, isto é, medir o intervalo de tempo decorrido

entre um negócio executado e outro, especialmente aqueles que representaram alteração no preço, e sua relação com variáveis como *spread*, volume e intensidade, é necessária uma base de dados que inclua apenas os eventos do tipo negócio e incluir-se a informação do topo do livro de ofertas, conforme a Tabela 5.4, a seguir:

Tabela 5.4 - Duração entre negócios em milissegundos

Tabela de duração entre trades								
Data	Horário	Duração	Tipo	Ativo	Qtd	Valor	Valor <i>bid</i>	Valor Ask
20081103	43368144	12890	trade	WINZ08	1	38450	37300	38450
20081103	43372019	650	trade	WINZ08	10	38460	37460	38450

Com a construção dessa base de dados, constatou-se uma série de consistências com os trabalhos de Engle e Russell (1998). Como é possível observar, o processo de formação de preços é bivariado com a duração entre *negócios* e duração entre ordens, sendo as variáveis aleatórias de interesse. Para efeito de aplicação, neste trabalho, consideramos apenas a duração entre *negócios* que representam alteração no preço.

6. Análise Empírica

Procuramos demonstrar que os parâmetros utilizados nos modelos podem ser facilmente deduzidos a partir de dados disponíveis do mercado, e que os testes de microestrutura de mercado são viáveis de serem executados. Neste capítulo, apresentamos os resultados da análise empírica da duração de preços do ativo minicontrato do índice BOVESPA, utilizando o modelo EACD por estimação Quase Máxima Verossimilhança (*Quasi-Maximum Likelihood QML*).

6.1. Considerações sobre a estimação

Do ponto de vista de propriedades estatísticas, a distribuição de ε_i deve ser compatível com uma distribuição no espaço $(0; \infty)$ para os modelos ACD e, portanto, qualquer distribuição positiva para as durações⁷ pode ser especificada, conforme Lancaster (1979).

A escolha natural, e muito conveniente, conforme já comentado, é a distribuição exponencial. Engle e Russell (1998) utilizaram a distribuição exponencial (ou seja, o parâmetro de forma é igual a um), pois tal distribuição é a que fornece estimadores QML para os parâmetros de ACD, o que facilita muito a estimação⁸. A forma da função de quase verossimilhança é definida como:

$$L(\theta) = - \sum_{i=1}^{N(T)} \left[\frac{x_i}{\psi_i} + \log(\psi_i) \right] \quad (6.1.1)$$

A extensão das propriedades de estimação por QML para o modelo GARCH (1,1) é formalizada para o modelo EACD (1,1) por Engle e Russell (1998, corolário, p.1135). Sob as condições definidas pelo seu teorema, estimativas consistentes e

⁷ A literatura apresenta a extensão Log-ACD, a qual flexibiliza essa restrição;

⁸ A literatura oferece inúmeras variações dos modelos ACD que apresentam vantagens sob a especificação EACD. Entretanto, a estimação requer a construção de rotinas para o cálculo da convergência da função de verossimilhança, por isso ficaram fora do escopo do presente trabalho.

assintoticamente normais θ são obtidos através da maximização da função de quase verossimilhança dada por (6.1.2), mesmo que a distribuição de ε_i , $p(\varepsilon_i, \theta)$ não seja exponencial. Os erros-padrão devem ser ajustados como definido por Bollerslev e Wooldridge (1992). O corolário de Engle e Russell (1998) também estabelece que as estimativas dos parâmetros ACD por QML podem ser obtidas utilizando-se o *software* padrão GARCH, em especial tomando a variável dependente igual a $\sqrt{x_i}$ e impondo-se a restrição de média condicional igual à zero. No entanto, um pressuposto fundamental para a obtenção de estimativas consistentes do modelo ACD por QML é que a esperança condicional da duração, ψ_i , seja especificada corretamente.

Uma importante restrição do presente trabalho é que a estimação por QML baseia-se no corolário derivado para o modelo EACD (1,1) linear, conforme Pacurar (2006), e não pode ser diretamente estendido para modelos ACD (m, q) mais gerais. A estimativa QML produz estimativas consistentes e os procedimentos de inferência, neste caso, estão disponíveis em pacotes de *softwares* estatísticos de mercado, mas com um custo de eficiência. Para estimativas mais robustas, dependendo da aplicação, a utilização de funções de máxima verossimilhança (ML) pode fazer-se necessária.

Outra limitação do presente trabalho é a escolha da distribuição da distribuição dos resíduos, que impactam diretamente nos resultados da função de intensidade condicional. A especificação exponencial implica em uma função de intensidade condicional plana, o que é uma condição bastante restritiva⁹ para várias aplicações empíricas em finanças, conforme Engle e Russell (1998) e Dufour e Engle (2000), entre outros.

Por outro lado, a utilização dos modelos GARCH como base apresenta a vantagem de que variáveis explicativas com valores reais (e potencialmente não negativas) podem ser introduzidas nos modelos sem complicações no processo de estimação. Em uma especificação mais geral, teremos:

⁹ Para maior flexibilidade, Engle e Russell (1998) utilizaram a distribuição padronizada Weibull com parâmetro de forma γ e parâmetro de escala igual a um. Importante notar que a distribuição se torna exponencial quando $\gamma = 1$.

$$\psi_n = \alpha_0 + \sum_{m=1}^p \alpha_m x_{n-m} + \sum_{m=1}^q \beta_m \psi_{n-m} + \theta' X_{n-1} \quad (6.1.3)$$

na qual X_{n-1} é um vetor de variáveis explicativas, e os parâmetros são restritos por α_m e β_m inteiros, $\alpha_m \geq 0$, $\beta_m \geq 0$, $\sum_{m=1}^p \alpha_m + \beta_m < 1$.

Como procedimento metodológico do presente trabalho, as especificações dos modelos ACD consideradas seguiram o estudo de Engle e Russell (1998), o qual utilizou os modelos EACD, de Bauwens e Giot (2000) e de Taylor (2004), os quais também utilizaram os modelos EACD como base de comparação das extensões. Nos três estudos, assim como no presente trabalho, o objetivo foi avaliar se as relações inversas entre a intensidade de negócios e volume, e o volume e o *spread* são identificadas, sendo que estimações utilizando outras especificações do modelo ficam como sugestões de trabalhos futuros.

6.2. Duração de preços e volatilidade

Uma importante contribuição do trabalho de Engle e Russel (1998) é a formalização da relação entre a duração de preços e as medidas de volatilidade de preços. Conforme indicam os autores, a partir da equação (4.2.1), a função da intensidade de preços pode caracterizar a probabilidade de um evento de preço ocorrer no instante seguinte, subordinada à história dos eventos passados, ou, mais precisamente, pode caracterizar a probabilidade de um preço se alterar em uma dada quantidade C . Então, se p_i é o preço associado à transação no instante de tempo t_i , a série de durações de preços é gerada retendo todos os i pontos do processo de pontual inicial, $i > 1$, tal que $|p_i - p_{i'}| > C$, para o qual $i' < i$, é o índice do ponto mais recentemente selecionado.

Para formalizar tal relação, Engle e Russell (1998) demonstraram que a volatilidade instantânea pode ser definida em função do preço P do ativo estudado como:

$$\sigma^2 = \lim_{\Delta t \rightarrow 0} E \left\{ \frac{1}{\Delta t} \left(\frac{P(t - \Delta t) - P(t)}{P(t)} \right)^2 \right\} \quad (6.2.1)$$

Com as substituições baseadas na equação (4.2.1) e tirando os limites, a expectativa da volatilidade condicional instantânea pode ser, então, definida como:

$$\sigma^2(t|t_{n(t)}, \dots, t_1) = \left(\frac{C}{P(t)} \right)^2 \lambda(t|t_{n(t)}, \dots, t_1) \quad (6.2.2)$$

em que $P(t)$ é o preço no instante t , após a alteração em relação ao último preço observado. Embora uma solução fechada para o cálculo da volatilidade instantânea condicional não exista na literatura, os dois dados da equação (6.2.2) podem ser calculados em cada ponto e também por simulação, uma vez que a função de intensidade condicional de preços é definida para todo $t > t_0$, havendo, pois, uma função contínua da volatilidade ao longo do dia de negociação. Consequentemente, tem-se um modelo EACD aplicado nas durações preço, $\lambda(t|t_{n(t)}, \dots, t_1) = 1/\psi_i$.

Tal relação tem importância fundamental nos resultados empíricos de testes de microestrutura de mercado, pois a intensidade de negociação é variável explicativa do modelo.

6.3. Modelo de Duração Condicional Autoregressivo para o movimento de preços

Os modelos ACD, conforme já mencionado, podem ser definidos por um subconjunto da chegada de transações com características específicas ou marcadores. Para se concentrar na análise do processo de formação de preços, este trabalho segue Engle e Russel (1998) e Bauwens e Giot (2000), e trata apenas dos pontos de *negócios* nos quais o preço foi alterado. Portanto, a duração de *negócios* é calculada não apenas entre chegadas de novas transações, mas entre transações em que ocorre um novo preço, um processo de seleção de dados que a literatura chama de *thinned point process*.

O modelo ACD será, portanto, aplicado à rapidez com que o preço está mudando. A função de intensidade condicional é, neste escopo, uma medida instantânea da probabilidade de um movimento de preço chamada de "intensidade de preço". Entretanto, várias especificações sobre o movimento de preços são possíveis. A literatura informa que o principal cuidado que se deve ter ao modelar o comportamento de preços é o efeito "*bid-ask bounce*", isto é, o salto repentino de ofertas de compras e vendas como um efeito transitório, explicado por O'Hara (1992) como sendo causado por *traders* impacientes.

Para construir uma medida que elimine o efeito de saltos bruscos no preço, Engle e Russel (1998) sugeriram a utilização do ponto médio entre a oferta de compra e venda, definido em t_i como

$$p_i = (bid_i + ask_i) * 0.5 \quad (6.3.1)$$

em que bid_i e ask_i são as ofertas correntes e p_i o preço de oferta médio. Após a construção da série de preços médios de ofertas, foram calculadas as durações entre transações cujo preço de oferta médio mudou $x_i = t_i - t_{i-1}$. Com essa seleção, o tamanho da amostra foi reduzido para 279.141 observações, ou cerca de 43% do tamanho original da amostra, cujas estatísticas estão apresentadas a seguir:

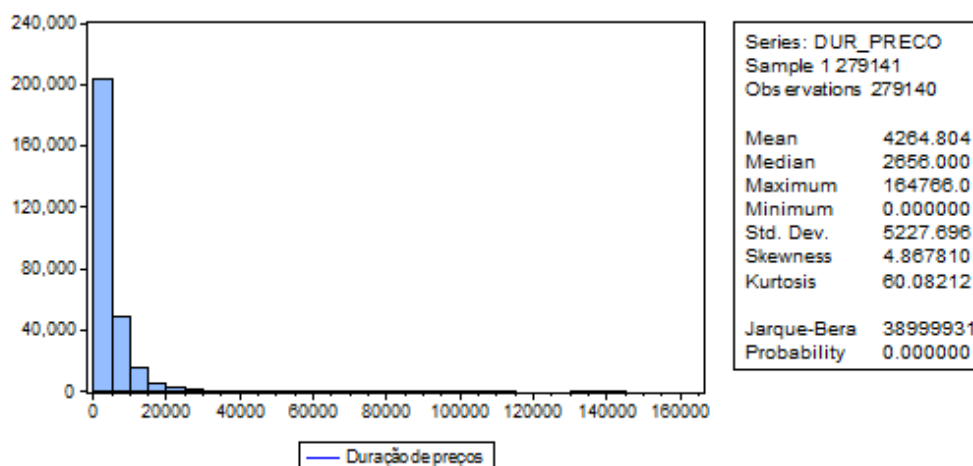


Gráfico 6.1 - Histograma e estatísticas da duração de preços

Da mesma forma que a distribuição dos tempos entre *negócios*, a duração de preços também é uma série que apresenta correlação serial, com todos os coeficientes de autocorrelação e autocorrelação parcial positivos, como demonstrado

no Gráfico 6.2, o que indica, segundo Engle e Russel (1998), o *clustering* de eventos de preços, isto é, períodos de grande variação da duração e períodos de pequena variação, característica bem documentada na literatura de modelos de duração condicional e também de modelos de volatilidade.

Duração de preços

Sample: 1 279141

Included observations: 279140

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
		1	0.180	0.180	9076.8	0.000
		2	0.165	0.137	16709.	0.000
		3	0.149	0.103	22876.	0.000
		4	0.142	0.087	28493.	0.000
		5	0.132	0.070	33329.	0.000
		6	0.126	0.061	37768.	0.000
		7	0.125	0.058	42152.	0.000
		8	0.122	0.052	46308.	0.000
		9	0.113	0.040	49863.	0.000
		10	0.113	0.040	53402.	0.000
		11	0.110	0.037	56803.	0.000
		12	0.109	0.036	60139.	0.000
		13	0.106	0.032	63294.	0.000
		14	0.109	0.035	66618.	0.000
		15	0.104	0.028	69647.	0.000

Gráfico 6.2 - Autocorrelação serial da duração de preços

6.4. Variáveis explicativas relacionadas ao processo de negociação

O modelo de Easley e O'Hara (1992) prevê que a intensidade de transações, o volume e o *spread* influenciam o processo de preços através de informação representada pelo *clustering* de transações. Segundo Engle e Russel (1998), tais hipóteses podem ser testadas estatisticamente, pela relação entre as variáveis explicativas e a duração de preços.

A especificação da duração condicional dada apenas pela equação (4.5.3) não é suficiente para que as hipóteses de microestrutura de mercado sejam testadas, pois é um modelo "autocontido", isto é, a duração depende apenas das durações passadas através de um processo autoregressivo. Variáveis explicativas relevantes devem ser incluídas nas especificações conforme informações disponíveis aos participantes no mercado em cada período x_i .

Este trabalho segue as investigações de Engle e Russell (1998), Taylor (2004) e Bauwens e Giot (2000), e concentra a análise nas três variáveis relacionadas por Easley e O'Hara (1992) como características do processo de negociação: a intensidade, o volume médio, e o *spread*. A escolha destas variáveis é também referenciada por Pacurar (2006) para os modelos de informação de microestrutura de mercado desenvolvidos pela literatura recentemente. Tais variáveis são definidas como:

- **Intensidade de negociação (λ_i):** durante cada duração de preços médios de ofertas, a intensidade de negociação é definida como o número de negócios registrados neste intervalo, dividido pelo comprimento da duração x_i , de forma que um grande número de *negócio*, ao longo de um curto período de estabilidade dos preços, gera uma intensidade alta de negociação. Para Easley e O'Hara (1992), um aumento na intensidade de *negócios* deveria levar a revisões mais frequentes das cotações, isto é, quanto maior a frequência, menores serão as durações.
- **Volume médio negociado:** o papel do volume negociado é também indicado por Easley e O'Hara (1992) como a variável que leva aos participantes do mercado um conteúdo informativo que não está contido no processo de preço. Para Easley e O'Hara (1992), o excesso de volume é indicativo de possível chegada de *traders* informados, o que também deve levar a revisões mais frequentes nas cotações e, portanto, também a uma relação negativa com as durações. No escopo deste trabalho, analisamos este efeito pelo volume médio negociado ao longo de uma duração, isto é, dividindo a quantidade negociada do ativo ao longo de um período de estabilidade de preços pela respectiva duração.
- **Spread:** os resultados de Easley e O'Hara (1992) indicam que um *spread* elevado é indicativo de *traders* informados, e que, portanto, também devem estar relacionados negativamente às longas durações. Para investigar este efeito, este trabalho utiliza o *spread* médio correspondente às operações realizadas durante cada duração x_i .

6.5. Sazonalidades intradiárias

Um fato bem documentado na literatura é que, ao longo do dia de negociação, os mercados financeiros são caracterizados por uma forte sazonalidade intradiária. Os principais estudos reportados por Pacurar (2006) demonstram a presença de sazonalidade por meio do agrupamento das durações em intervalos de tempo regulares (por exemplo, a cada 10 minutos).

Engle e Russell (1998) relataram maior atividade comercial (portanto, menor duração, em média) no início e final do pregão, e atividade mais lenta (portanto, maior duração) no meio do dia. Estes padrões estão associados a características do mercado de câmbio, aberturas de mercados relacionados, e aos hábitos dos *traders*, que são ativos na abertura do pregão para conseguir operações com alguma vantagem informativa, e no encerramento, quando buscam fechar suas posições. Esse comportamento é apresentado na literatura como o formato de “U invertido”.

A exemplo dos dados dos trabalhos de referência, é possível também identificar esse formato de U invertido nos dados de negociação do contrato de derivativo WINZ08, conforme Gráfico 6.3. No caso do derivativo em questão, um efeito especialmente interessante é observado: a maior intensidade de alterações nos preços ocorre imediatamente após o horário de abertura do pregão americano, às 11h30 (no horário de Brasília), e não na abertura da bolsa brasileira. Como também representado no Gráfico 6.3, o período anterior à abertura do mercado americano apresenta um comportamento muito variável em relação ao padrão das durações de preços, e é seguida pelo formato de “U invertido”.

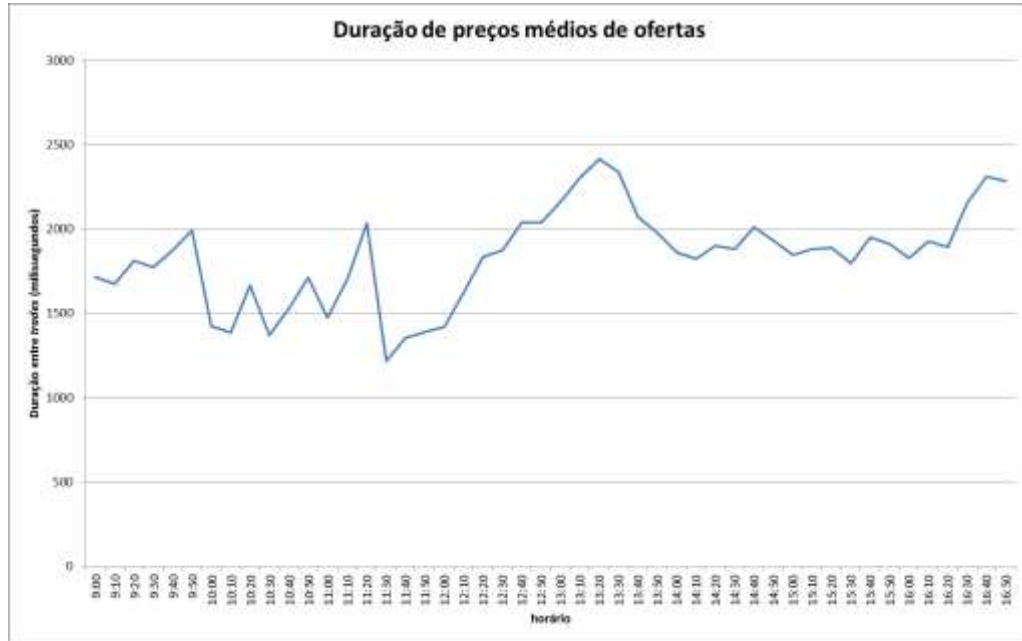


Gráfico 6.3 - Efeito do horário na duração de preços

Pacurar (2006) afirmou que ignorar esses padrões intradiários pode distorcer os resultados de estimação. Para controlar esse efeito, o procedimento mais utilizado na literatura correlata é o mesmo originalmente empregado por Engle e Russell (1998). Trata-se de decompor durações intradiárias em uma parte determinística com base no horário do dia, e uma parte estocástica, da seguinte forma:

$$\tilde{x}_i = \frac{x_i}{f(t_i)} \quad (6.5.1)$$

em que \tilde{x}_i denota a duração "ajustado pelo horário do dia" ¹⁰, e $f(t_i)$ representa o fator sazonal em t_i . A equação (4.3.1) para a duração condicional torna-se, então:

$$E(\tilde{x}_i | \tilde{x}_{i-1}, \dots, \tilde{x}_1) f(t_i) \quad (6.5.2)$$

No que diz respeito à especificação do fator sazonal $f(t_i)$, o presente trabalho segue a abordagem dominante na literatura e originalmente elaborada por Engle e

¹⁰ Para evitar problemas de convergência das funções de verossimilhança, Engle e Russell (1998) sugeriram que o processo seja feito em duas etapas: primeiro as durações são ajustadas e, em seguida, o modelo ACD é estimado com durações dessazonalizadas.

Russell (1997, 1998) em que as durações x_i são regredidas contra nós de horários do dia igualmente espaçados de hora em hora do utilizando uma interpolação linear, sendo, então, a série \tilde{x}_i obtida. Este processo é executado em todas as variáveis explicativas, e o Gráfico 6.4 apresenta as funções interpoladas $f(t_i)$ utilizadas para o ajuste da duração de preços, intensidade, volume e *spread*:

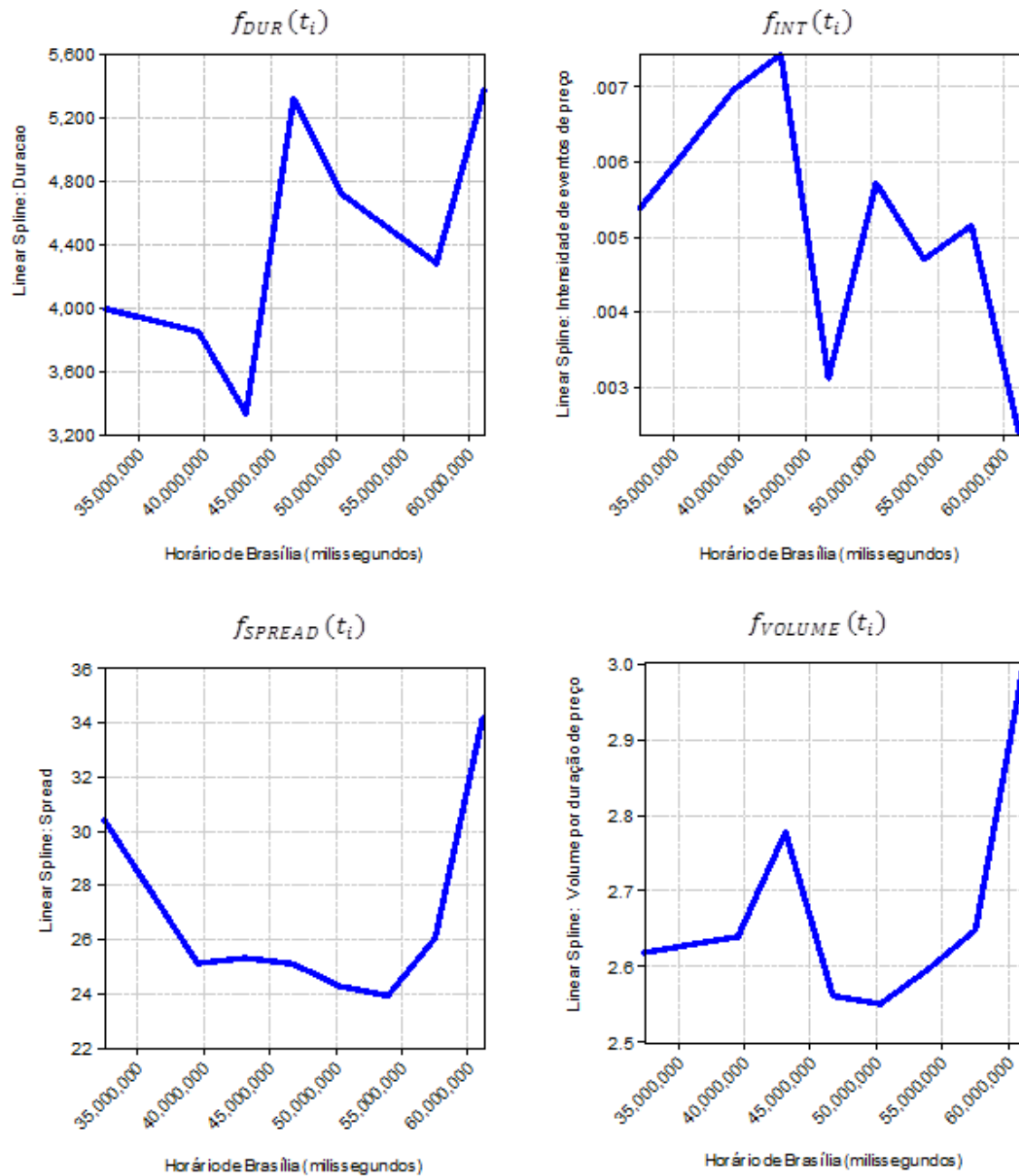


Gráfico 6.4 - Interpolação linear das durações de preços e variáveis explicativas

Os padrões das funções de interpolação diferem levemente dos da literatura, pois apresentam o efeito da alteração da duração após a abertura do mercado americano. Nenhuma observação foi retirada da amostra por questões de horários.

6.6. Testes de hipóteses sobre microestrutura de mercado

Após a criação das variáveis de controle e os ajustes dos dados que serão utilizados nos modelos, especialmente o controle da sazonalidade diária, a dinâmica de duração de preços do ativo WINZ08 pode ser analisada e, como primeiro passo, verificou-se a adequação do modelo EACD a esse objetivo.

As estimações das variações EACD (1,1) e EACD (2,2), com a $\sqrt{x_i}$ como variável dependente na equação da média, foram feitas utilizando-se a implementação ARCH do *software* EVIEWS, com estatísticas t^{11} para a distribuição de erros, e utilizando-se o algoritmo Bernt-Hall-Hall-Hausman, como apresentado na Tabela 6.1. O primeiro ponto a ser notado é que todos os fatores estocásticos α e β , para os dois modelos estimados, foram significativos estatisticamente, e a soma dos fatores $\alpha + \beta < 1$ de ambos os modelos demonstram persistência dos modelos e processo não explosivo.

Tabela 6.1 - Estimação EACD para duração de preços

Parâmetros	EACD(1,1)			EACD(2,2)		
	Estimativa	Estatística t	p -Valor	Estimativa	Estatística t	p -Valor
ω	0.002321	21.91052	0.0000	0.714721	10.95204	0.0000
α_1	0.025339	39.81100	0.0000	0.18945	8.278649	0.0000
α_2				-0.034142	-31.97193	0.0000
β_1	0.963488	986.7335	0.0000	0.720951	7.790505	0.0000
β_2				-0.062337	-3.448036	0.0006
	<i>R-squared</i>	0.059382		<i>R-squared</i>	-0.786700	
	<i>Adjusted R-squared</i>	0.059379		<i>Adjusted R-squared</i>	-0.786707	
	<i>Akaike info criterion</i>	1.198801		<i>Akaike info criterion</i>	2.053934	
	<i>Schwarz criterion</i>	1.199028		<i>Schwarz criterion</i>	2.054236	
LM Tests	<i>F-statistic</i>	1296.669		<i>F-statistic</i>	1296.669	
	<i>Prob. F(15,279108)</i>	0.0000		<i>Prob. F(15,279108)</i>	0.0000	
	<i>Obs*R-squared</i>	18183.97		<i>Obs*R-squared</i>	18183.97	
	<i>Prob. Chi-Square(15)</i>	0.0000		<i>Prob. Chi-Square(15)</i>	0.0000	

¹¹ Segundo Pacurar (2006), na literatura ARCH as distribuições normais são frequentemente rejeitadas em favor de distribuições leptocúrticas para a distribuição dos erros

Modelo	$\psi_n = \omega_0 + \sum_{m=1}^1 \alpha_m x_{n-m} + \sum_{m=1}^1 \beta_m \psi_{n-m}$	$\psi_n = \omega_0 + \sum_{m=1}^2 \alpha_m x_{n-m} + \sum_{m=1}^2 \beta_m \psi_{n-m}$
--------	---	---

O teste de autocorrelação serial pela estatística Ljung-Box demonstrou que os modelos EACD (1,1) e EACD (2,2) tratam bem da autocorrelação dos resíduos. De maneira geral, o modelo EACD (1,1) apresentou um melhor ajuste aos dados pelos critérios de informação. Entretanto, como é possível observar nos resultados dos testes dos Multiplicadores de Lagrange (*Lagrange Multiplier – LM*), os modelos ainda apresentam efeito ARCH (heterocedasticidade). Este resultado também está presente nos trabalhos correlatos, e indica uma má especificação da distribuição dos erros. Engle e Russell (1998) apontaram esse resultado como uma forte evidência contra a utilização do modelo EACD em favor de outras modelagens para a distribuição dos erros, como logarítmica, ou Burr e Weibull. Um ponto relevante é que o R^2 do modelo EACD (2,2) apresenta valor negativo. A literatura de econometria informa que $R^2 < 0$ indica que o modelo adiciona ruídos não explicados na modelagem da variância; entretanto, tal fato não representa um problema de especificação para Engle e Russel (1998).

Ainda que a presença de heterocedasticidade seja uma forte evidência para uma má especificação do modelo dos erros, Engle e Russell (1998), Taylor (2004) e Pacurar (2005) indicaram que o modelo EACD pode ser utilizado para os testes de hipóteses de microestrutura de mercado, e a Tabela 6.2 apresenta os resultados de tais testes. Seguindo a literatura, este trabalho utiliza a especificação EACD (2,2) para os testes de relação defasada entre as variáveis relacionadas com o processo de negociação. Duas versões de modelos foram construídas: o modelo 1 utiliza apenas a intensidade de negociação como variável de explicação; o modelo 2 utiliza as variáveis intensidade, volume e *spread*. É possível observar, na Tabela 6.2, que os parâmetros estocásticos (α e β) e os regressores (intensidade, volume e *spread*) do modelo possuem relevância estatística, já que se rejeitam facilmente a hipótese nula.

O sinal negativo das variáveis explicativas tem significado fundamental na análise do modelo 1, pois indica que a duração esperada de preços é mais curta, e

que, portanto, a volatilidade equivalente é maior, após períodos de altas taxas de transação. Este resultado é coerente com o modelo de Easley e O'Hara (1992).

O modelo 2 também apresenta evidências de alinhamento com a literatura. Glosten e Milgrom (1985) demonstraram que os modelos de informação assimétrica geralmente preveem que *spreads* tendem a se tornar maiores com o aumento da probabilidade de haver agentes informados na negociação, o que está relacionado com maior intensidade e volume de ativos negociados. Para testar tais hipóteses, o modelo 2 acrescenta o volume médio por transação observado em relação à duração de preço anterior e o *spread* de compra e venda no final do período passado, sendo cada variável também corrigida para sazonalidades diárias.

Tabela 6.2 – Modelos EACD(2,2): testes de agrupamento de preços

Parâmetros	Modelo 1			Modelo 2		
	Estimativa	Estatística t	p-Valor	Estimativa	Estatística t	p-Valor
ω	0.884352	1.682210	0.0925	0.945041	6.229677	0.0000
α_1	0.049098	7.095916	0.0000	0.083606	9.360651	0.0000
α_2	-0.019699	-1.724950	0.0779	0.006913	1.422697	0.0725
β_1	0.423996	1.988668	0.0346	0.455341	2.617506	0.0089
β_2	-0.015893	-1.659430	0.0733	0.015440	1.142425	0.0867
Intensidade (λ)	-0.000613	-29.51429	0.0000	-0.000399	-1.663878	0.0944
Volume				-0.042303	-9.550094	0.0000
Spread				-0.040981	-5.046226	0.0000
LM Tests	<i>R-squared</i>	-2.243711		<i>R-squared</i>	-2.799870	
	<i>Adjusted R-squared</i>	-2.243722		<i>Adjusted R-squared</i>	-2.799883	
	<i>Akaike info criterion</i>	2.714365		<i>Akaike info criterion</i>	2.866720	
	<i>Schwarz criterion</i>	2.714705		<i>Schwarz criterion</i>	2.867135	
	F-statistic	1105.160		F-statistic	783.4021	
	Prob. F(15,279108)	0.0000		Prob. F(15,279108)	0.0000	
	Obs*R-squared	15648.90		Obs*R-squared	11276.92	
	Prob. Chi-Square(15)	0.0000		Prob. Chi-Square(15)	0.0000	
Modelo	$\psi_n = \omega_0 + \sum_{m=1}^1 \alpha_1 x_{n-1} + \sum_{m=1}^1 \beta_1 \psi_{n-1} + \theta' X_{n-m}$			$\psi_n = \omega_0 + \sum_{m=1}^2 \alpha_2 x_{n-m} + \sum_{m=1}^2 \beta_2 \psi_{n-m} + \theta' X_{n-m}$		

O coeficiente de intensidade tem sua relevância estatística reduzida em relação ao modelo 1, mas ainda é significativo no nível de 10% de confiança, e ainda é negativo. O volume e o *spread* também são significativos pela estatística t, com *p*-valores que rejeitam a hipótese nula. Estes resultados confirmam ainda mais fortemente as hipóteses de Glosten e Milgrom (1985), bem como a conjectura de

Easley e O'Hara (1992), de que processos de negociação com *traders* informados resultam em volatilidades superiores. Os coeficientes de determinação R^2 ainda foram negativos para o modelo EACD(2,2), o que sugere o exercício de novas especificações, mas, segundo Pacurar (2006), não invalida o modelo que pode conter não-linearidades.

Considerando a possibilidade de que o agrupamento de negociação pode estar ocorrendo em momentos diferentes, por diferentes razões, Easley e O'Hara (1992) e Engle e Russel (1998) sugeriram que, em certos momentos, o agrupamento de transações ocorra devido à negociação baseada em informação, mas que pode haver momentos em que o agrupamento de transações com mudança de preços é devido à necessidade de liquidez. A análise da queda do coeficiente de intensidade no modelo 2, em relação ao modelo 1, como apresentado na Tabela 6.2, permitiu a inferência, baseado em Taylor (2004), Engle e Russell (1998) e Easley e O'Hara (1992), de que o preço tende a se mover menos rapidamente, ou quase não é afetado pelas taxas de intensidade de transações quando os *spreads* são menores, o que, em tese, ocorre quando os *traders* entram no mercado por liquidez. Isto sugere, que nos dados analisados, ocorreram agrupamentos tanto por liquidez, como por informação.

Engle e Russell (1998) sugeriram como procedimento a utilização de classes de *spread* como variável *dummy* para entender o agrupamento de preços por liquidez ou por informação. Se há um efeito heterogêneo em relação ao *spread*, seria correto se esperar que o coeficiente de intensidade de transações seja negativo quando o "*spread* é maior", refletindo movimentos de alterações de preços mais rapidamente quando a fração de *traders* informados for alta. O coeficiente de intensidade com "*spread* menor" deve ser insignificante de acordo com Admati e Pfleiderer (1998). Em um contexto mais geral, em que informações privadas nem sempre existem, podemos esperar que o coeficiente sobre "*spread* menor" seja positivo, refletindo um mercado muito líquido. Isto é, na ausência de *traders* informados, o mercado é provavelmente mais líquido quando a atividade de negociação é maior. O presente trabalho sugere que estudos futuros avaliem a relação de níveis de *spread* e a dinâmica de preço.

7. Conclusão

Este trabalho utilizou as técnicas de modelagem de séries temporais irregularmente espaçadas no tempo utilizando a classe de modelos autoregressivos de duração condicional para testar as hipóteses de microestrutura de mercado com dados de minicontratos de futuros de índice Bovespa. Os resultados da literatura foram observados neste trabalho o que indica que os modelos se adaptaram bem para dados financeiros de alta frequência de *Market data*. Os resultados aqui apresentados corroboraram a tese de imediatismo de microestrutura de mercado da literatura e identificaram que o volume negociado e o *spread* de compra e venda são informativos e negativamente relacionados com a intensidade de negociação.

As análises concentraram-se nas propriedades assintóticas da literatura ARCH/GARCH para obter estimações via QML, para analisar a dinâmica de duração de preços utilizando a versão exponencial do modelo ACD. Esta especificação se demonstrou adequada para reduzir a autocorrelação serial associada com a duração de preços, o que ficou comprovado pela redução do valor da estatística Ljung-Box para níveis marginalmente significantes. Entretanto, os testes para o excesso de dispersão sugerem que as versões exponenciais do modelo ACD podem não ser totalmente adequadas para a modelagem de duração de preços, fato também observado na literatura. Uma limitação deste trabalho reside no fato de que apenas a especificação exponencial foi utilizada, quando outras especificações, como *logarithmic ACD* (log-ACD) e *Weibull ACD* (WACD), são apresentadas pela literatura como alternativas adequadas para estimação de durações de preços.

Um modelo para o processo de preço foi estimado, selecionando-se uma sub amostra dos dados que carregam informações sobre o comportamento do mercado. O modelo ACD pode ser interpretado como um modelo para a volatilidade instantânea, e proporcionou uma imagem contínua da volatilidade ao longo do dia, recurso importante para gestores de risco que operam em alta frequência. O modelo de preços foi utilizado para testar as hipóteses sobre a origem do agrupamento de durações dos preços, e foi apresentada a evidência de que o agrupamento das

transações que geram alterações de preços ocorre tanto devido à concentração de *traders* informados, como de *traders* que operam por liquidez.

Este trabalho abre espaço para pesquisas futuras em várias direções. Na frente de aplicações em finanças, é possível observar que, conforme literatura correlata, o processo de preços é essencialmente uma árvore binomial com espaçamentos aleatórios. As aplicações na precificação de derivativos e cálculo da estrutura a termo de taxa de juros são diretas. Outro campo fértil para aplicações em finanças é a gestão de riscos, uma vez que a volatilidade, aqui modelada como a função recíproca da duração, é componente fundamental dos modelos VaR (*Value at Risk*). Como exemplo de aplicação, Giot (2002) apresentou uma quantificação de risco intradiário, utilizando um *framework* de VaR paramétrico. As análises da variação temporal de liquidez, e a modelagem conjunta de mais de um ativo, em particular a dinâmica dos movimentos de preços de ativos subjacentes e seus derivativos, também representa um estoque de grande aplicação para as técnicas de modelagem da duração.

Uma segunda frente de aplicação e pesquisa futura reside na relação de durações de transações e preços com as variáveis explicativas para análise dos custos de transação. Um aspecto fundamental na tendência de automação das praças de liquidez é o efeito que a latência e o desempenho dos sistemas de negociação podem ter no processo de formação de preços. Portanto, estudos sobre a compreensão dos impactos dos sistemas eletrônicos nos mecanismos fundamentais de microestrutura de mercado brasileiro, a exemplo dos estudos da literatura sobre os investimentos em tecnologia da informação na NYSE, NASDAQ e Deutsche Börse, são uma fonte rica de pesquisa e de aplicação prática.

REFERÊNCIAS

- Admati, R., Pfleiderer, P. (1988). A Theory of Intraday Patterns: Volume and Price Variability. *The Review of Financial Studies*, 1, 3-40.
- Bagehot, W. (1971). The Only Game in Town. *Financial Analysts Journal*, 22, 12-14.
- Bauwens, L. and P. Giot (2000). The Logarithmic ACD Model: an Application to the Bid-Ask Quote Process of Three NYSE Stocks. *Annales d'Économie et de Statistique*, 60, 117-149.
- Bauwens, L. and P. Giot (2001). Econometric Modeling of Stock Market Intraday Activity. *Advanced Studies in Theoretical and Applied Econometrics*. Kluwer Academic Publishers: Dordrecht, 196 pages.
- Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*. P. 31, 307-327.
- Bollerslev, T. and J. Wooldridge (1992). Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances. *Econometric Reviews*, 11, 143-172.
- Bortoluzzo, A.B., Morettin, P.A., Toloi, C.M.C. Time-varying autoregressive conditional duration model. *Journal of Applied Statistics*. P. 847–864. 2008
- Colussi, P. B. (2009). O mercado de câmbio brasileiro pela ótica da microestrutura. FGV.
- Cox, D. R. (1955). Some Statistical Models Connected with Series of Events (with Discussion). *Journal of the Royal Statistical Society. Series B*, 17, 129-164.
- Cox, D. R. (1955). The Statistical Analysis of Dependencies in Point Processes. in *Symposium on Point Processes*, ed. by P. A. W. Lewis. New York: John Wiley, 55-66.

- Diamond, D.W. and R.E.Verrechia (1987). Constraints on Short-Selling and Asset Price Adjustments to Private Information. *Journal of Financial Economics*, 18, 277-311.
- Dufour, A. and R.F. Engle (2000). Time and the Impact of a Trade. *Journal of Finance*, 55, 2467-2498.
- Easley, D. and M. O'Hara (1992). Time and the Process of Security Price Adjustment. *Journal of Finance*, 47, 577-606.
- Easley, D. and M. O'Hara (1992). Time and the Process of Security Price Adjustment. *Journal of Finance*, 47, 577-606.
- Engle, R.F and Lunde, A. (2003). Negócios and Quotes: A Bivariate Point Process. *Journal of Financial Econometrics*.
- Engle, R.F. (2000). The Econometrics of Ultra-High Frequency Data. *Econometrica*, 68, 1-22.
- Engle, R.F. and J.R. Russell (1997). Forecasting the Frequency of Changes in Quoted Foreign Exchange Prices with Autoregressive Conditional Duration Model. *Journal of Empirical Finance*, 4, 187-212.
- Engle, R.F. and J.R. Russell (1998). Autoregressive Conditional Duration: a New Model for Irregularly Spaced Transaction Data. *Econometrica*, 66, 1127-1162
- Engle, R.F. and J.R. Russell (2002). Analysis of High Frequency Data,. *Handbook of Financial Econometrics*, ed. by Y. Ait-Sahalia and L. P. Hansen, Elsevier Science: North-Holland.
- Eom, K.S. and S.B. Hahn (2005). Traders' Strategic Behavior in an Index Options Market. *Journal of Futures Markets*, 25, 105-133.
- Fernandes, M. and J. Grammig (2005). Nonparametric Specification Tests for Conditional Duration Models. *Journal of Econometrics*, 127, 35-68.
- Gaver, D. P., e LEWIS P. A. W. (1980). First Order Autoregressive Gamma Sequences and Point Processes. *Advances in Applied Probability*, 12, 727-745.
- Gerety, M. S., & Mulherin, J. H. (1992). Trading Halts and Market Activity: An Analysis of Volume at the Open and the Close. *Journal of Finance*, 47, 1765–1784.

- Giot, P. (2002). Market Risk Models for Intraday Data. *Discussion Paper, CORE, Université Catholique de Louvain, forthcoming in the European Journal of Finance*.
- Glosten, L.R. and Milgrom P.R.(1985). Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders. *Journal of Financial Economics*, 14, 71-100.
- Goodhart, C. A. E., O'HARA, M. High frequency data in financial markets: issues and applications. *Journal of Empirical Finance*, 4, 73, 114, 1997.
- Grossman, S. J., & Miller, M. H. (1986). Economic Costs and Benefits of the Proposed One-Minute Time Bracketing Regulation. *The Journal of Futures Markets*, 6, 141–166.
- Hawkes, A. G. (1972). Spectra of Some Mutually Exciting Point Processes with Associated Variables. *Stochastic Point Processes* ed. by P. A. W. Lewis. New York: Wiley, 261-271.
- Heckman J. (1981). Heterogeneity and State Dependence Studies in Labor Markets. *National Bureau of Economic Research*, 91-139.
- Holder, M.E., Qi, M., Sinha A.K. (2004). The Impact of Time Duration between Negócios on the Price of Treasury Note Futures Contracts. *Journal of Futures Markets*, 24,965-980.
- INFORMAÇÕES financeiras BM&FBovespa Janeiro de 2011. *Relações com Investidores*. Disponível em <
http://ri.bmfbovespa.com.br/upload/portal_investidores/pt/informacoes_financeiras/apresentacoes_videoconferencias/BM&FBovespa%20-%20Janeiro_2011.pdf >. Acesso em 18 Fev. 2011.
- Jacobs, P. A., Lewis, P. A. W. (1977). A Mixed Autoregressive-Moving Average Exponential Sequence and Point Process (EARMA1, 1). *Advances in Applied Probability*, 9, 87-94.
- Lancaster, T. (1979). Econometric Methods for the Duration of Unemployment. *Econometrica*, 47, 939-956.
- Laurini, M.P., Furlani, L.G.C. e Portugal, M.S. (2008). Empirical market microstructure: An analysis of the BRL/US\$ exchange rate market. *Emerging Markets Review*, 9,247-265

- Lawrence, A. J., Lewis P. A. W. (1980). The Exponential Autoregressive-Moving Average ERMA(p, q) Process. *Journal of the Royal Statistical Society, B*, 42, 150-161.
- Manaster, S., & Mann, S. M. (1996). Life in the Pits: Competitive Market Making and Inventory Control. *Review of Financial Studies*, 9, 953–975.
- Manganelli, S. (2005). Duration, Volume and Volatility Impact of Negócios. *Journal of Financial Markets*, 8, 377-399.
- Martins, M. (2010). As negociações de alta frequência e o mercado brasileiro. *Valor Econômico*, São Paulo, 16 Set. 2010. Disponível em <<http://www.valoronline.com.br/impresso/investimentos/119/309420/as-negociacoes-de-alta-frequencia-e-o-mercado-brasileiro>>. Acesso em 18 de Fev. 2011.
- Miller, M. H. (1990). International Competitiveness of U.S. Futures Exchanges. *Journal of Financial Services Research*, 4, 387–408.
- O'Hara, M. (1995). Market Microstructure Theory. *Blackwell Publishers*. 304 pages.
- Pacurar, M. (2006). Autoregressive Conditional Duration (ACD) Models in Finance: A Survey of the Theoretical and Empirical Literature. *Journal of Economic Surveys*
- Rubin, I. (1972). Regular Point Processes and Their Detection. *IEEE Transactions on Information Theory*, ITT-18, 547-557.
- Subrahmanyam, A. (1991). A Theory of Trading in Stock Index Futures. *Review of Financial Studies*, 4, 17–52.
- Taylor, N. (2004). Trading Intensity, Volatility, and Arbitrage Activity. *Journal of Banking & Finance*, 28, 1137-1162.
- Tsay, R.S. (2002). Analysis of Financial Time Series. *Wiley: New York*, 472 pages.
- Veredas, D., Rodriguez-Poo, J., Espasa A. (2001). On the (Intradaily) Seasonality and Dynamics of a Financial Point Process: A Semiparametric Approach. *Working Paper 2001-19, CREST*.
- Wold, H. (1948). On Stationary Point Processes and Markov Chains. *Skandinavisk Aktuaetidskrist*, 31, 229-240.

Zhang, M. Y., Russell, J.R., Tsay, R.S. (2001). A Nonlinear Autoregressive Conditional Duration Model with Applications to Financial Transaction Data. *Journal of Econometrics*, 104, 179-207.

APÊNDICE

Rotina de interpolação linear para ajuste das sazonalidades diárias

'LINEAR SPLINE das durações de preços

```
!k1=36000000 '10h = 36000000
!k2=!k1+3600000 '11h = 39600000
!k3=!k2+3600000 '12h = 43200000
!k4=!k3+3600000 '13h = 46800000
!k5=!k4 + 3600000 '14h = 50400000
!k6=!k5+ 3600000 '15h = 54000000
!k7=!k6+ 3600000 '16h = 57600000
```

```
!nobs=279141
smpl 1 !nobs
genr t=time_br
genr t1=t-!k1
genr t2=(t-!k2)*(t>!k2)
genr t3=(t-!k3)*(t>!k3)
genr t4=(t-!k4)*(t>!k4)
genr t5=(t-!k5)*(t>!k5)
genr t6=(t-!k6)*(t>!k6)
genr t7=(t-!k7)*(t>!k7)
```

```
' ajuste da sazonalidade da duração de preços
equation spline.ls dur_preco c t1 t2 t3 t4 t5 t6 t7
spline.fit dur_precof
genr durs=dur_preco/dur_precof
```

```
' ajuste da sazonalidade dos spreads
equation spline.ls spread_avg c t1 t2 t3 t4 t5 t6 t7
spline.fit spread_avgf
genr spread_avgs= spread_avg/ spread_avgf
```

```
' ajuste da sazonalidade dos volumes de negócios
equation spline.ls volume_transacao c t1 t2 t3 t4 t5 t6 t7
spline.fit volume_transacaof
genr volume_transacaos= volume_transacao/ volume_transacaof
```

```
' ajuste da sazonalidade das intensidades
equation spline.ls intensidade_transacao c t1 t2 t3 t4 t5 t6 t7
spline.fit intensidade_transacaof
genr intensidade_transacaos= intensidade_transacao/ intensidade_transacaof
```

```
scat time_br dur_precof
scat time_br spread_avgf
scat time_br volume_transacaof
scat time_br intensidade_transacaof
```

ANEXOS

Extensões do modelo ACD

Na literatura é possível encontrar diversas extensões dos modelos ACD. Conforme já mencionado, muitas extensões do modelo são possíveis considerando-se diferentes distribuições de probabilidade para as variáveis de interesse e resíduos. Pacurar (2006) apresentou uma completa pesquisa das extensões e aplicações dos modelos ACD.

WACD – Weibull ACD

A generalização mais comum dos modelos ACD consiste na parametrização da distribuição condicional da variável de interesse, Weibull, sendo equivalente a assumir que x^γ é exponencial. Seguindo a especificação de Engle e Russell (1998), a função de taxa de risco (*hazard function*) é definida por:

$$h(x) = \kappa^\gamma x^{\gamma-1} \gamma \quad (6.6.1)$$

em que (κ, γ) são os parâmetros da distribuição e x a duração. A intensidade condicional, em termos de ψ_n , toma a seguinte forma:

$$\lambda(t|x_{N(t)}, \dots, x_1) = \gamma \left(\Gamma(1 + 1/\gamma) \psi_{N(t)+1}^{-1} \right)^\gamma (t - t_{N(t)+1})^{\gamma-1} \quad (6.6.2)$$

para a qual $\Gamma(\cdot)$ é a função gamma e γ é o parâmetro Weibull. Uma versão simples de um modelo ACD, utilizando-se a distribuição Weibull, é a WACD(1,1). Tomando ψ_i como o logaritmo da esperança condicional da duração x_i , podemos escrever $\psi_i = E(x_i|I_{i-1}) = \gamma + \delta \Gamma(1 + \frac{1}{\beta})$.

A função log-verossimilhança para a especificação Weibull-ACD¹² pode ser escrita como:

$$\sum_{i=1}^{N(T)} \ln \left(\frac{\gamma}{x_i} \right) + \gamma \ln \left(\frac{\Gamma(1 + 1/\gamma) x_i}{\psi_i} \right) - \left(\frac{\Gamma(1 + 1/\gamma) x_i}{\psi_i} \right)^\gamma \quad (6.6.3)$$

¹² A função log verossimilhança da especificação WACD pode ser convertida para exponencial se $\gamma = 1$

Log-ACD – Logarithmic ACD

A versão logarítmica dos modelos ACD pode ser elaborada pela alteração do processo de inovações (4.3.2) pela equação:

$$x_i = e^{\phi_i \cdot \varepsilon_i} \quad (6.6.4)$$

no qual os resíduos ε_i são i.i.d e seguem uma distribuição *Weibull* $(1, \gamma)$, e ϕ_i é a função proporcional ao logaritmo da esperança condicional da duração x_i . Tomando ψ_i como o logaritmo da esperança condicional da duração x_i , podemos escrever $\psi_i = \ln E(x_i | I_{i-1})$ e especificar o processo autoregressivo como:

$$\psi_i = \alpha_0 + \alpha \ln(x_{i-1}) + \beta \psi_{i-1} \quad (6.6.5)$$

A equação (6.6.5) deverá seguir a restrição de que $\ln|\alpha + \beta| < 1$.

A função log-verossimilhança para a especificação Log-ACD, em função das observações de x_i pode ser escrita como:

$$\sum_{i=1}^{N(T)} \ln(\gamma) - \ln(x_i) + \gamma \ln(\Gamma(1 + 1/\gamma)x_i) - \gamma \psi_i - \left(\frac{\Gamma(1 + 1/\gamma)x_i}{e^{\psi_i}} \right)^\gamma \quad (6.6.6)$$