

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE CIÊNCIAS ECONÔMICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

Juliano Marmitt

**Dados de Alta Frequência: Averiguando o Impacto de Microestrutura de
Mercado e Sazonalidade Intradiária na Detecção de Saltos e Estimação
da Variação Quadrática**

Porto Alegre

2012

Juliano Marmitt

Dados de Alta Frequência: Averiguando o Impacto de Microestrutura de Mercado e Sazonalidade Intradiária na Detecção de Saltos e Estimação da Variação Quadrática

Dissertação submetida ao programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como quesito parcial para obtenção do título de Mestre em Economia com ênfase em Economia Aplicada.

Orientador: Prof. Dr. Flávio Augusto Ziegelmann

Porto Alegre

2012

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)

Responsável: Biblioteca Gládis Wiebbelling do Amaral, Faculdade de Ciências Econômicas da UFRGS

M351d Marmitt, Juliano

Dados de alta frequência : averiguando o impacto de microestrutura de mercado e sazonalidade intradiária na detecção de saltos e estimação da variação quadrática / Juliano Marmitt. – Porto Alegre, 2012.
65 f. : il.

Orientador: Flávio Augusto Ziegelmann

Ênfase em Economia Aplicada.

Dissertação (Mestrado em Economia) – Universidade Federal do Rio Grande do Sul, Faculdade de Ciências Econômicas, Programa de Pós-Graduação em Economia, Porto Alegre, 2012.

1. Econometria. 2. Estimação. 3. Volatilidade. I. Ziegelmann, Flávio Augusto.
II. Universidade Federal do Rio Grande do Sul. Faculdade de Ciências Econômicas.
Programa de Pós-Graduação em Economia. III. Título.

CDU 330.115

Juliano Marmitt

Dados de Alta Frequência: Averiguando o Impacto de Microestrutura de Mercado e Sazonalidade Intradiária na Detecção de Saltos e Estimação da Variação Quadrática

Dissertação submetida ao programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como quesito parcial para obtenção do título de Mestre em Economia com ênfase em Economia Aplicada.

Aprovada em: Porto Alegre, 31 de agosto de 2012.

Prof. Dr. Flávio Augusto Ziegelmann – Orientador
UFRGS

Prof. Dr. Hudson da Silva Torrent
UFRGS

Prof. Dr. João Fróis Caldeira
UFRGS

Prof. Dr. Marcelo Perlin
UFRGS

RESUMO

Neste trabalho, visamos mostrar as características usuais dos dados de alta frequência, bem como utilizar modelagem não paramétrica para estimar a variância/volatilidade para esses dados. Após uma revisão sobre microestrutura de mercado, sazonalidade intradiária, variação quadrática e saltos, utilizamos os dados da PETR4 para estimar a variância realizada e variação bipotente. Determinadas essas séries, testamos se há saltos nas mesmas. Em seguida, analisamos o impacto que a microestrutura de mercado e a sazonalidade intradiária causam na detecção dos saltos. Concluimos que, enquanto a presença de microestrutura aponta para um número de saltos menor que o esperado, a sazonalidade intradiária aponta para o lado contrário, ou seja, ela causa um viés para detectar mais saltos, dada a estrutura típica da curva de volatilidade ao longo do dia em formato de J invertido, causando mais saltos incorretamente detectados no período mais volátil do dia (que corresponde a abertura da bolsa de valores).

Palavras-chave: Microestrutura de mercado. Sazonalidade intradiária. Variação quadrática. Variância realizada. Volatilidade realizada. Variação bipotente. Saltos. Dados de alta frequência.

ABSTRACT

In this work, we aim to show the usual characteristics of high-frequency data and the estimation of variance/volatility for this kind of data using nonparametric models. After reviewing concepts about market microstructure, intraday seasonality, quadratic variation and jumps, we use PETR4 data to estimate realized variance and bipower variation. With these series determined, we test for jumps. Then, we analyze the impact that market microstructure and intraday seasonality causes in jump detection. We conclude that while microstructure noise indicates fewer jumps than the ideal amount, intraday seasonality goes in the opposite direction, i.e., it detects more jumps than it should, since the typical inverted-J-shaped intraday volatility pattern tends to incorrectly detect more jumps at the most volatile period (which is when stock markets start negotiations).

Keywords: Market microstructure. Intraday seasonality. Quadratic variation. Realized variance. Realized volatility. Bipower variation. Jumps. High-frequency data.

LISTA DE FIGURAS

Figura 1: Função de autocorrelação das 10 primeiras defasagens de cada intervalo temporal para os retornos puros, juntamente com seus intervalos de confiança.	40
Figura 2: Retornos Realizados e Retornos Realizados Padronizados, respectivamente.	48
Figura 3: Variância Realizada, Variação Bipotente ($k = 2$), Função C e Volatilidade Realizada, respectivamente.	49
Figura 4: Magnitude dos saltos, Saltos Relativos, Estatística $Z_{SR,QP^{[3]}}$ ($\alpha = 0,1\%$ e $k = 2$) e Contribuição % dos saltos detectados sobre a VR, respectivamente.	50
Figura 5: Padrão Intradiário da volatilidade (em toda a amostra) e Padrão Semanal, respectivamente.	51
Figura 6: Padrão Intradiário Anual e Semanal (em série e empilhado).	53
Figura 7: Filtro VBP e Padrão Intradiário (filtrado e puro), respectivamente.	54
Figura 8: Função de Autocorrelação para retornos puros e filtrados (período de 5 dias), respectivamente.	55
Figura 9: Estimativas de densidade por Kernel para a estatística $Z_{SR,QP^{[3]}}$, VBP sem pulos (pura e filtrada) e VBP com dois pulos (pura e filtrada), respectivamente.	56

LISTA DE TABELAS

Tabela 1: Autocorrelação de primeira ordem e estatística Box-Pierce dos retornos intradiários para cada intervalo.....	37
Tabela 2: Número de saltos e sua respectiva proporção em relação ao total de dias, para cada nível de significância	45
Tabela 3: Estatísticas descritivas das séries da PETR4.....	47
Tabela 4: Estatísticas de Normalidade	56
Tabela 5: Saltos detectados para $Z_{SR,QP^{[3]}}$, retornos puros e filtrados	57
Tabela 6: Saltos detectados em cada ano para $Z_{SR,QP^{[3]}}$, $k = 2$, $\alpha = 0,1\%$, retornos puros e filtrados.....	57

SUMÁRIO

1 Introdução	9
2 Definições gerais de volatilidade, retornos, dados de alta frequência, microestrutura de mercado e sazonalidade intradiária	13
2.1 Volatilidade	13
2.2 Retornos e seus fatos estilizados	14
2.3 Dados de alta frequência e microestrutura de mercado.....	16
2.3.1 Irregularidade temporal e caráter discreto dos dados.....	18
2.3.2 <i>Spread</i> entre preços de compra e venda	19
2.4 Sazonalidade intradiária.....	21
3 Variação quadrática e variância realizada	23
3.1 Retornos intradiários	25
3.2 Modelos de difusão, semimartingales e variação quadrática.....	26
3.3 Variância Realizada e Volatilidade Realizada.....	29
3.4 Variação bipotente.....	30
4 Definições formais sobre <i>spread</i>, saltos e sazonalidade e análise dos dados	35
4.1 <i>Spread</i> entre preços de compra e venda	36
4.2 Saltos	38
4.3 Sazonalidade intradiária.....	48
5 Conclusões	58
Referências	61

1 INTRODUÇÃO

Na sociedade moderna, a quantidade de informação produzida diariamente é avassaladora. Não poderia ser diferente quando lidamos com finanças. Na medida em que o tempo passa, a profusão de informação vai moldando os preços dos ativos instantaneamente, caracterizando um comportamento bastante dinâmico. Dito isso, fazer previsões a respeito do comportamento dos preços dos ativos é tarefa complicada. Dada a quantidade de dados disponíveis, o que podemos fazer é analisá-los para inferir estimativas o mais precisas possíveis. Investidores que conseguem compreender esse dinamismo dos preços são os que possivelmente terão as melhores expectativas em relação aos preços futuros e quantidade de risco ao qual estão expostos. Podemos utilizar os mais variados tipos de informação para tentar determinar os preços dos ativos, e os investidores a usarão para obter resultados competitivos (TAYLOR, 2005). Modelar a dinâmica dos preços dos ativos e em particular da volatilidade usada em finanças é fundamental na precificação de ativos e opções, determinação e gerenciamento de riscos de mercado, seleção ótima de portfólio, *hedging*, etc., tópicos importantes no âmbito financeiro.

O resultado obtido através dessa modelagem auxilia nas decisões tomadas pelos agentes na hora de alocar os recursos em portfólios atrativos. Um dos modelos de volatilidade mais conhecidos e utilizados surgiu após o artigo de Engle (1982), onde o autor sugere a modelagem da inflação do Reino Unido por um modelo onde a variância condicional varia ao longo do tempo, ou seja, é heterocedástica, e não homocedástica como os modelos até então assim consideravam-na. Depois deste trabalho seminal que introduziu o modelo ARCH, várias outras generalizações foram surgindo, tentando explicar adicionalmente os fatos estilizados ligados aos retornos e a volatilidade. A generalização mais famosa foi a de Bollerslev (1986) com o modelo GARCH, que abriu caminho para várias outras surgirem em seguida. Ainda, outra abordagem para volatilidade vem através da modelagem via volatilidade estocástica. Os modelos univariados e suas generalizações foram por muito tempo dominantes na

literatura econométrica que trata de volatilidade. Porém, os dados de menor frequência, comumente utilizados para modelá-la, falham em capturar toda a movimentação que ocorre ao longo do dia no mercado de ações (ANDERSEN; BOLLERSLEV, 1998b).

Por consequência, surge a necessidade de explorar a verdadeira natureza da volatilidade, utilizando dados de maior frequência. Até pouco tempo atrás, a obtenção e armazenamento deste tipo de dado eram custosas, impedindo o seu uso de maneira mais eficaz. Como o processamento de computadores vem aumentando e o custo de armazenar dados vem diminuindo, este problema acaba se tornando menos relevante com o passar dos anos. A justificativa para tal abordagem é clara, dado um princípio estatístico básico: uma amostra que contenha mais observações (de maior frequência) traz mais informações do que uma amostra que contenha menos observações (de menor frequência). De outra maneira, podemos dizer que quanto mais observarmos o comportamento da volatilidade ao longo do tempo, mais saberemos sobre ela. Não só conseguimos obter informações relevantes da alta frequência como também da menor frequência da volatilidade. A utilização deste tipo de informação é relativamente recente, inclusive no Brasil. Com cada vez mais disponibilidade de dados intradiários de ativos financeiros, novos métodos de estimação de volatilidade destes retornos surgem na literatura para incorporar a estrutura peculiar destes dados – consequência da maior quantidade de informação contida neles.

Esse tipo de dado está disponível negociação a negociação, sendo que cada negociação que ocorre a qualquer instante no mercado de ações é computada. Isso gera inconveniências, pois os mecanismos de negociação que ocorrem nos mercados de ações não são perfeitos, ou seja, não há uma sincronia temporal perfeita, e as negociações vão ocorrendo em instantes diferentes. Todo esse excesso de informação em relação a dados diários de menor frequência acaba trazendo também outras complicações para a modelagem. Estas características reunidas são conhecidos na literatura como microestruturas de mercado (ou ruído de mercado), e podem ser

interpretados como resultado de fricções que vão ocorrendo a cada momento do tempo e sendo incorporadas aos preços dos ativos negociados. Essas microestruturas surgem em consequência de irregularidade temporal entre as transações, *spread* entre preços de compra e venda (*bid-ask spread*) e sazonalidade intradiária. A existência de tais complicações dificulta a modelagem direta, e por isso devemos recorrer a técnicas que consigam lidar com estes problemas. Assim, devemos tratar da microestrutura de mercado antes de tirar qualquer conclusão a respeito dos elementos correspondentes ao comportamento da volatilidade.

A literatura que lida com este tema utiliza muito a modelagem via volatilidade realizada, que consegue acomodar o problema de microestrutura de mercado sem descartar a maior carga de informação. Este tipo de modelagem para dados intradiários foi inicialmente sugerido por Andersen e Bollerslev (1998). Mais recentemente, novos modelos surgiram, relacionando a volatilidade realizada (ou, similarmente, variância realizada) com a variação quadrática (bastante usada em finanças para modelar precificação de derivativos), decompondo-a em um componente contínuo e um componente de saltos (que são observações maiores que o normal). O primeiro modelo desse tipo, chamado de modelo de variação bipotente, foi introduzido por Barndorff-Nielsen e Shephard (2004, 2006). Este modelo consegue trazer novas informações sobre a dinâmica dos dois componentes da volatilidade supracitados. Com isso, temos uma modelagem mais precisa e coerente na determinação da volatilidade e, por conseguinte, dos preços dos ativos.

Segundo Wright e Zhou (2007), a correta identificação dos saltos é uma questão muito importante no gerenciamento de risco e na precificação de contratos de derivativos. Dada a atual importância dos derivativos nos mercados financeiros, e sabendo quando e como os saltos ocorrem, os investidores se beneficiam muito, já que observações incomuns não previstas podem gerar grandes perdas. Quando o processo de movimentação de preços é contínuo, os investidores conseguem fazer *hedge* nos contratos de derivativos. Porém, na presença de saltos, o processo apresenta

descontinuidades, gerando um risco que não pode ser protegido por *hedging*. Se conseguirmos detectar os saltos corretamente, temos informações sobre quão frequentemente eles ocorrem e o quanto contribuem no processo de preços.

Buscamos assim extrair todo tipo de informação que os dados intradiários trazem consigo. Os dados escolhidos foram da Petrobrás (mais especificamente PETR4, ações preferenciais) de janeiro de 2007 a abril de 2011. A escolha deste ativo em particular se deve, primeiro, a maior facilidade de obtenção de dados de alta frequência de ações negociadas na BM&FBOVESPA e, segundo, por sua grande liquidez. Para reduzir o problema de microestrutura de mercado, utilizamos dados de alta frequência em um intervalo maior de tempo entre as observações, ou seja, os preços e retornos foram transformados para só mostrarem o último lance ocorrido a cada 5 minutos do pregão, dando assim uma sincronia temporal aos dados e menor ruído de microestrutura.

O capítulo dois do trabalho apresenta uma revisão sobre volatilidade e fatos estilizados dos retornos diários e intradiários, além de descrever os problemas de microestruturas e sazonalidade intradiária de forma detalhada. O capítulo três apresenta definições formais de vários termos utilizados na literatura de dados de alta frequência, bem como as abordagens não paramétricas por variância realizada e variação bipotente. O capítulo quatro apresenta definições formais sobre as microestruturas de mercado, saltos e sazonalidade intradiária, mostrando também os resultados obtidos após a análise dos dados e das estatísticas definidas no capítulo. Primeiro, testamos pela autocorrelação negativa de primeira ordem em intervalos de 1 minuto, 5 minutos, 30 minutos, 60 minutos e 24 horas, para verificar o comportamento do *spread* com o aumento do intervalo temporal; em seguida, utilizamos o teste de Barndorff-Nielsen e Shephard (2004, 2006) para detectar saltos diários na série de variância realizada; depois, após verificar o padrão intradiário nos dados, utilizamos um filtro para remover este padrão e testamos por saltos novamente. Por último, as considerações finais.

2 DEFINIÇÕES GERAIS DE VOLATILIDADE, RETORNOS, DADOS DE ALTA FREQUÊNCIA, MICROESTRUTURA DE MERCADO E SAZONALIDADE INTRADIÁRIA

2.1 Volatilidade

Os preços dos ativos tendem a mudar de acordo com as informações que chegam ao mercado - sendo que o fluxo de informações atualmente é enorme. Eles se movimentam de maneira mais acentuada em períodos de turbulência e ficam menos agitados quando há períodos de tranquilidade no mercado. A intensidade dessa mudança de preços é o que chamamos de volatilidade. Em outras palavras, a volatilidade é uma medida de variação nos preços em determinado período de tempo. A maneira mais comum a qual ela é referida é como sendo o desvio padrão dos retornos dos ativos. Sendo assim, parece evidente que esse conceito é de enorme interesse aos investidores, pois ao acompanhar a atividade da volatilidade estamos acompanhando a atividade na mudança dos preços dos ativos.

As razões para as mudanças de volatilidade não tem uma explicação específica, e podem depender de um ou mais fatores juntos. Atualmente, toda a informação divulgada que afete de algum modo os preços dos ativos é agregada instantaneamente aos preços. Quando há períodos caracterizados como crises, a perturbação é grande e é rapidamente transmitida ao mercado financeiro, que se agita (ANDERSEN; BOLLERSLEV; DIEBOLD; VEGA, 2003)¹. Variáveis como desemprego, inflação, crescimento econômico, expansão monetária, produção industrial, etc. podem catalisar a intensidade de variação da volatilidade (SCHWERT, 1989). A causa e consequência de cada uma delas separadamente (ou mais de uma conjuntamente) na determinação dos preços não são exatas, ou seja, não explicam a movimentação da volatilidade em sua totalidade. Apesar disso, as séries temporais de volatilidade podem ser estimadas

¹ O impacto de anúncios macroeconômicos é visto também em Andersen e Bollerslev (1998b). Mais sobre isso aparecerá no subcapítulo 2.3.

sem conhecermos a causa que determina seu comportamento. Contudo, o resultado nem sempre pode ser satisfatório aos olhos de quem tem interesse em estimações coerentes.

2.2 Retornos e seus fatos estilizados

Analisar os preços dos ativos diretamente é mais difícil do que analisar mudanças de preços. Assim, de acordo com Brooks (2008), é adequado trabalhar com variações nos preços de ativos – ou, em outras palavras, retornos dos ativos. Os retornos dos ativos são calculados da maneira tradicionalmente usada na literatura, através da transformação

$$y(t) = \ln \left(\frac{P(t)}{P(t-1)} \right),$$

onde $P(t)$ é o preço do ativo no tempo t . Eles não dependem de unidades monetárias como os preços, o que os tornam facilmente comparáveis, são ergódicos e geralmente estacionários (com média amostral próxima de zero). As regularidades estatísticas dos retornos são comuns a várias séries financeiras, gerando fatos estilizados, primeiramente enunciados nos trabalhos de Black (1976), Fama (1965) e Mandelbrot (1963). Os fatos estilizados podem ser encontrados em Taylor (2005) e são analisados de maneira mais detalhada em Cont (2001). Andersen e Bollerslev (1997) procuram esses fatos estilizados em retornos intradiários. Assim, ao estudar as séries de retornos, várias características comuns foram encontradas. Tais características, enumeradas abaixo, são:

(i) Autocorrelação entre retornos: Retornos geralmente apresentam pouca autocorrelação linear, sendo parecidos com um ruído branco. Por outro lado, os quadrados dos retornos (ou o valor absoluto dos retornos), apresentam autocorrelação significativa, sendo que essa autocorrelação tem um decaimento lento. Em relação aos

dados de alta frequência, estes geralmente apresentam autocorrelação devido aos efeitos causados pelas microestruturas de mercado.

(ii) Agrupamento de volatilidade (*clustering*): Os períodos de volatilidade geralmente aparecem em blocos, com períodos de alta e baixa volatilidade. Altos retornos absolutos tendem a seguir anteriores altos retornos absolutos e baixos retornos absolutos tendem a seguir anteriores baixos retornos absolutos. Ou seja, existe uma dependência temporal entre retornos absolutos (ou, similarmente, entre os quadrados dos retornos). Esses agrupamentos são recorrentes na economia, mas não seguem um padrão periódico específico. Eles podem ser interpretados como uma persistência na volatilidade.

(iii) Distribuições com caudas pesadas (Leptocurtose): Por inspeção gráfica de séries de retornos de vários ativos, podemos notar que geralmente as distribuições dos retornos não se assemelham à distribuição Gaussiana (Normal), e os testes clássicos levam à rejeição da hipótese de normalidade. Geralmente, essas distribuições têm caudas mais pesadas (leptocurticas), indicando que eventos mais extremos ocorrem com maior probabilidade, e há um pico agudo ao redor do valor zero. Andersen e Bollerslev (1997) relatam que quanto maior a frequência, maior a curtose.

(iv) Efeito alavancagem: É uma assimetria no impacto dos valores passados nos valores presentes. Períodos de calma geralmente apresentam baixa volatilidade, enquanto períodos turbulentos apresentam maior volatilidade. Logo, os retornos negativos (quedas nos preços dos ativos, causadas por alguma informação ruim que gera incerteza no mercado) resultam em uma maior volatilidade futura do que retornos positivos (ambos de mesmo módulo).

2.3 Dados de alta frequência e microestrutura de mercado

A maneira usual de medir volatilidade e gerenciar risco é posta em dúvida pelo atual sistema de negociações, altamente interligado. Nos últimos anos, a velocidade e quantidade de transações realizadas aumentaram drasticamente. Com isso, a modelagem de risco deve seguir o rápido passo do mercado para não ficar para trás. Qualquer agente que esteja ligado a esta atividade deve ampliar seu horizonte temporal de análise – do diário para o intradiário. Estes, quando envolvidos intensivamente com negociações ou na implementação de portfólios dinâmicos e estratégias de *hedging*, estão expostos a volatilidade de curto prazo, e consideram a informação intradiária como sendo essencial. Por outro lado, investidores mais passivos estão mais preocupados com dados de menor frequência. Com estes, podemos utilizar os modelos usuais de volatilidade. Quando usamos dados intradiários, surge a preocupação com microestrutura de mercado. Enquanto que dados de maior e menor frequência não são inteiramente independentes, as diferenças entre eles levam a diferentes interpretações e significados econômicos (ANDERSEN; BOLLERSLEV, 1998b).

Dados de menor frequência, que são registrados apenas com o último preço diário, não conseguem captar de maneira eficiente toda a movimentação que ocorre entre agentes no mercado financeiro. É notório que a quantidade de informação que o mercado recebe durante um dia de negociação não consegue ser descrita apenas por uma observação diária. Daí surge o termo (ultra) alta frequência em Engle (2000). Nesse caso, esses dados de maior frequência englobam todas as negociações realizadas no dia (uma a uma, a cada momento em que elas ocorrem), trazendo uma gama de informação muito maior do que a de dados de menor frequência. O maior número de observações ajuda a compreender melhor como os preços dos ativos reagem à maior quantidade de informação. Os dados de alta frequência disponíveis para consulta apresentam várias informações detalhadas - chegando à precisão dos segundos – de todas as negociações realizadas, cada dia. Como há muitas transações diárias, a

volatilidade dos ativos muda constantemente e a maior quantidade de informações geralmente melhora as previsões de volatilidade em relação aos dados de menor frequência.

Os dados de alta frequência são mais difíceis de serem obtidos do que os de menor frequência, e geralmente tem um custo associado com sua obtenção. Até duas décadas atrás, poucos estudos usando este tipo de dado eram feitos. Atualmente, a disponibilidade de dados de alta frequência é maior. Contudo, estes dados trazem problemas em sua estrutura. Dado o maior número de informações, os dados estão mais sujeitos a ruídos (TAYLOR, 2005). Ao verificar tais dados podemos notar que eles apresentam características únicas que dificultam sua modelagem, como: i) os retornos apresentam microestrutura de mercado, induzida pelos mecanismos comuns de transação que ocorrem nos mercados de ações (irregularidade temporal entre as negociações, caráter discreto dos dados, *spread* entre preços de compra e venda; ii) há um padrão intradiário comum, onde há maior volatilidade no início e término do pregão, induzindo uma espécie de sazonalidade intradiária na autocorrelação dos retornos absolutos (ou quadrado dos retornos) intradiários.

O estudo de microestruturas de mercado existe para tentar explicar o papel da informação e dos mecanismos de troca na determinação de preços. Se houvesse informação completa e perfeita, ela seria disseminada instantaneamente por todos os agentes econômicos, e os preços dos ativos seriam ajustados de acordo com sua preferência e informação disponível no momento. No entanto, na prática isso não ocorre, sendo que há agentes que possuem mais informação que os outros. Os mercados contam com muitos agentes realizando transações, sendo que estes agem de acordo com certas regras e conseguem verificar a ação de outros agentes. Além disso, as transações não ocorrem apenas em âmbito local, mas também internacional, aumentando ainda mais o número de transações e agentes envolvidos. Disso decorre um lapso temporal entre o recebimento de alguma informação e a real assimilação dela entre os agentes e o mercado, o que afeta os preços de maneira diferente em

várias partes do mundo. Logo, a assimetria de informação é um problema crônico e dela deriva a microestrutura de mercado² (ENGLE; RUSSEL, 2004).

2.3.1 Irregularidade temporal e caráter discreto dos dados

As negociações que ocorrem nos mercados não apresentam regularidade temporal, ou seja, há uma assincronia entre elas. Isso ocorre pois as negociações são livres para acontecerem a qualquer momento do dia em que os mercados estejam funcionando, e assim os agentes vão realizando-as no instante que preferirem, de acordo com suas preferências e informações disponíveis. Várias transações podem e vão ocorrendo ao mesmo tempo, no mesmo segundo, com preços diferentes. Ativos diferentes têm frequências de negociação diferentes; o mesmo vale para um ativo específico que geralmente apresenta muitas negociações na abertura e fechamento do mercado e menos negociações entre esses períodos. Além disso, todos os dados relativos a preços de ativos são discretos.

A maioria dos modelos econométricos utiliza séries de dados com regularidade temporal. No caso dos dados diários, o preço usado é o da última negociação de cada dia, dando uma falsa impressão que os retornos são igualmente distribuídos a cada 24 horas, o que pode gerar conclusões erradas sobre sua previsão diária. Por isso os dados intradiários são mais fidedignos e representam melhor as oscilações dos preços que ocorrem em um dia. Porém, o problema da irregularidade temporal traz complicações para a modelagem direta e os procedimentos usuais para análise de séries temporais não são mais adequados (TSAY, 2010). Ainda, muitas vezes os dados

² Ainda que não seja o escopo dessa dissertação, a informação tem papel fundamental na explicação das séries de mudanças de volatilidade. A assimetria de informação é um campo crescente de pesquisa na teoria econômica, e ficou notório após o trabalho de Akerlof (1970). Com a introdução do problema, alguns autores passaram a procurar as causas de várias crises que ocorreram na economia (trabalhos abrangentes nessa área são os de MISHKIN, 1990, 1992). Os autores concluem que a assimetria de informação é um grave problema, desencadeador de crises econômicas.

apresentam complicações como observações ausentes, devido a paralizações nas negociações (*circuit breaker*) ou feriados que adiam a abertura do pregão.

2.3.2 *Spread* entre preços de compra e venda

Chamado na literatura de *bid-ask spread*, é a diferença entre o preço de compra (*bid*) e venda (*ask*) que ocorre nas transações de ativos. Nos mercados de ativos, o papel do formador de mercado (*market maker*, ou provedor de liquidez) é fundamental: ele procura formar os preços dos ativos, fornecendo diversas ofertas de compra e venda que buscam trazer liquidez ao mercado, atendendo as demandas dos agentes por ativos. Dessa forma, o formador de mercado está disposto a comprar a um preço P_c e a vender a um preço P_v . O *spread* é então gerado por $P_v - P_c$ (TAYLOR, 2005). Segundo Engle e Russel (2004), é impossível distinguir os agentes que possuem informação daqueles que não a possuem, ou seja, quando os agentes negociam ativos no mercado, não há como saber se o negociador possui ou não informação. O formador de mercado naturalmente visa lucro, que provém do *bid-ask spread*. Assim, essa diferença entre os preços de compra e venda auferida pelo formador de mercado pode ser interpretada como uma compensação (ou corretagem, já que nesse caso ele age como um corretor) pelo risco que ele está sujeito ao negociar com um agente que possui mais informação que ele.

O *spread* pode ser constante (preços de compra e venda andam juntos) ou pode aumentar/diminuir (preços de compra e venda andam em direções opostas). As operações de maior risco (que envolvem ativos de menor liquidez ou expectativas heterogêneas em relação ao preço de um ativo – muita assimetria de informação) aumentam o *spread*, enquanto que operações menos arriscadas (que envolvem maior liquidez e/ou expectativas homogêneas em relação ao preço de um ativo) tendem a reduzir esse *spread*. Tsay (2010) afirma que esse *spread* tem consequências importantes nas séries de dados intradiários, pois introduz uma autocorrelação negativa na primeira defasagem nos preços dos ativos, chamada de *bid-ask bounce*.

Esse efeito pode ser explicado da seguinte maneira: como negociações de compra e venda acontecem o tempo todo, o preço efetivo do ativo tende a flutuar entre o preço de compra ou o preço de venda. Logo, o *bid-ask bounce* é muito mais evidente em séries de dados intradiários.

Entre os trabalhos pioneiros na área, Andersen e Bollerslev (1997) analisam essa autocorrelação negativa na primeira defasagem em um ano de retornos intradiários. As frequências analisadas vão de 5 minutos até 12 horas, sendo que para os períodos de 40 minutos em diante essa autocorrelação é insignificante. Para uma frequência de 5, 10, 15, 20 e 30 minutos da cotação do marco alemão sobre o dólar americano, os autores encontram autocorrelação por volta de -0,040, -0,070, -0,089, -0,082 e -0,043, respectivamente. Contudo, como por hipótese não há autocorrelação negativa além da primeira defasagem, os autores utilizam o modelo MA(1)-GARCH(1,1) para os dados, sendo que o componente de médias móveis de primeira ordem é incluído justamente para acomodar o *bid-ask bounce*. Engle e Russel (2004) chamam a atenção ao fato de que pode haver também uma autocorrelação positiva, que ocorre quando o formador de mercado divide ordens grandes em várias ordens menores na tentativa de conseguir, na média, preços melhores. Isso pode fazer com que as negociações induzam os preços a se movimentarem na mesma direção.

Para reduzir este problema, devemos recorrer a uma transformação nos dados que tenha periodicidade constante, como transações ocorrendo minuto a minuto, 5 a 5 minutos, etc. A tendência é de que, quanto maior for a frequência amostral escolhida, menor será a autocorrelação da primeira defasagem e, assim, menor o problema derivado da microestrutura de mercado.

2.4 Sazonalidade intradiária

Outra característica bastante comum dos ativos encontrados em séries de dados intradiários é a de que há um forte padrão intradiário (ou sazonalidade intradiária) da volatilidade, onde ela é maior na abertura e nos primeiros minutos do pregão, vai decaindo ao longo do dia e volta a crescer gradativamente até o fechamento do pregão, remetendo a um formato em “U” ou “J”. Este padrão surge da autocorrelação periódica dos retornos quadráticos ou absolutos (usados como *proxy* para volatilidade), e foi primeiramente observado por Wood, McInish e Ord (1985), onde os autores calcularam o desvio padrão dos retornos intradiários de ativos da bolsa de valores americana (NYSE) e verificaram que a volatilidade era maior no começo e no fim do pregão. Ainda, esse padrão pode ser encontrado no *spread* de compra e venda, no volume de negociações e nos retornos (ENGLE; RUSSEL, 2004).

O padrão intradiário presente nos dados reflete a atividade que ocorre durante o dia. Ele surge através de anúncios macroeconômicos, *lunch effects* (na hora do almoço a volatilidade é menor) e sobreposição de horários entre mercados de ações diferentes. Uma das possíveis explicações desse padrão ocorrer é que na abertura do pregão os participantes do mercado já analisaram toda a informação disponível anteriormente a abertura e já traçaram suas estratégias para o dia vindouro. Assim, quando o pregão começa, os agentes ajustam suas posições, dando ordens para criar novas posições ou limpar posições antigas. Logo, ocorrem mais transações na abertura, o que, por conseguinte, aumenta a volatilidade. Durante o dia, os agentes estão mais preocupados em ver o que acontece com suas posições e interpretam as informações que vão recebendo. Quando o dia de negociações vai chegando ao fim, os agentes fazem os ajustes finais, realizando mais negociações, o que aumenta a volatilidade novamente (SCHMID, 2009).

Mesmo outros mercados, que não possuem a mesma estrutura de abertura e fechamento dos mercados de ativos (como o mercado de câmbio), apresentam o

mesmo padrão intradiário. O mercado de câmbio funciona 24 horas por dia, sete dias por semana. Antes da introdução do euro, a volatilidade entre moedas europeias e o dólar americano era maior nos períodos em que o mercado americano e europeu estavam ativos ao mesmo tempo, ocorrendo na parte da manhã nos Estados Unidos e no final da tarde na Europa (ENGLE; RUSSEL, 2004). Andersen e Bollerslev (1997, 1998b) estudam este fenômeno no mercado de câmbio. No caso dos retornos de 5 minutos, há um total de 288 observações diárias em um dia de negociações no mercado internacional de câmbio. Os autores então verificam que a autocorrelação é 0,309 na primeira defasagem, declina para -0,033 na 144^a defasagem (que corresponde à metade do dia) e aumenta para aproximadamente 0,15 na 288^a defasagem (última observação do dia). O padrão então se repete para os dias consecutivos. Como ele introduz autocorrelação na volatilidade, precisa ser removido para que não haja problemas de especificação que geram análises estatísticas viesadas (ANDERSEN; BOLLERSLEV, 1997). Andersen e Bollerslev (1998b) confirmam o forte padrão da volatilidade, não só intradiário como interdiário também (como efeitos calendário – redução nas negociações por conta de feriados, por exemplo – ou dias da semana específicos que também apresentam um padrão – como o efeito sexta-feira).

Para remover a sazonalidade, devemos filtrar os dados intradiários antes de começar a estimação do modelo. Num dos primeiros trabalhos abordando sazonalidade intradiária, Andersen e Bollerslev (1997) utilizam o método de formas flexíveis de Fourier (FFF) para ajustar o padrão intradiário, usando diferentes frequências de funções seno e cosseno, dado o padrão consistente encontrado nos dados. Comparando os dados filtrados com os não filtrados, utilizando o mesmo modelo MA(1)-GARCH(1,1) os autores concluem que o parâmetros da volatilidade apresentam um comportamento mais coerente para os dados filtrados.

Porém, a filtragem da sazonalidade intradiária pode introduzir viés nos dados. Por exemplo, após a filtragem, a assimetria e a curtose diminuem e a autocorrelação negativa de primeira ordem aumenta. Seguindo esse raciocínio, Laakkonen (2007)

explica que a filtragem pode gerar problemas. Primeiro, nem toda a sazonalidade é removida dos dados, e isso pode afetar a correta estimação do modelo. Segundo, essa filtragem pode gerar novas informações na série de dados, como adição de novos saltos³ nos retornos ou remoção de informações importantes (algo relacionado a notícias ou eventos que ajudam a explicar o comportamento da volatilidade). Se estivermos interessados em estudar o impacto da informação nos dados intradiários, a filtragem pode atrapalhar nossos planos.

³ Saltos serão abordados no capítulo 4.

3 VARIAÇÃO QUADRÁTICA E VARIÂNCIA REALIZADA

A volatilidade é uma das principais variáveis financeiras, principalmente no mercado de derivativos, onde exerce papel fundamental na tomada de decisões. As empresas que lidam com derivativos e gerenciamento de risco procuram de toda forma determinar a volatilidade da maneira mais precisa possível, já que ela é o ponto chave nesses assuntos. Uma opção de um ativo qualquer é tão mais cara quanto maior for a volatilidade sobre o tempo de maturação do mesmo ativo. Como a volatilidade não é uma variável diretamente observável, como os retornos e preços de ativos, as tarefas que envolvem o uso de volatilidade na tomada de decisão se tornam mais complicadas. Para contornar o problema, os primeiros modelos de volatilidade usam os retornos quadráticos como medida de volatilidade.

Na ausência das microestruturas de mercado, como no caso dos dados diários de menor frequência, poderíamos modelar a volatilidade diretamente através de modelos convencionais como os modelos da família GARCH, largamente celebrados como modelos de resultados bastante precisos quando se usam dados diários. Como os desdobramentos resultantes das microestruturas de mercado discutidos no capítulo anterior complicam a modelagem direta, devemos encontrar modelos capazes de tratar de melhor maneira todos estes problemas encontrados nos dados de alta frequência. Assim podemos estimar a volatilidade via variância (ou volatilidade) realizada, que é baseada na teoria da variação quadrática, tida como variável muito importante em aplicações financeiras. Com a modelagem via volatilidade realizada, é também possível separar o componente contínuo do componente de saltos da volatilidade.

A volatilidade realizada é um método alternativo de mensurar a volatilidade dos retornos usando dados de alta frequência, introduzido no artigo de Andersen e Bollerslev (1998a) e passando depois a ser largamente utilizado na literatura de dados de alta frequência. Dada a estrutura um tanto complexa deste tipo de dado, modelos

paramétricos como o GARCH são difíceis de serem utilizados. Eles geralmente falham na descrição das características presentes nos dados intradiários (ANDERSEN; BOLLERSLEV; DIEBOLD; LABYS, 2003). Assim, é conveniente que seja usada uma aproximação não paramétrica, onde a volatilidade realizada é o método mais popular. A ideia básica deste método é simplesmente somar os retornos quadráticos intradiários em curtos espaços de tempo para gerar resultados mais eficazes em termos de estimação e previsão de volatilidade.

3.1 Retornos intradiários

Definamos o processo estocástico $y(t)$ como sendo o logaritmo do preço de um ativo qualquer no tempo contínuo t com $t \in [0, T]$. Para determinar os retornos discretos, supomos que Δ seja um intervalo fixo de tempo (como um dia ou um mês de negociação) e $\Delta > 0$. Seguindo Barndorff-Nielsen e Shepard (2004), quando fixamos Δ como sendo um intervalo fixo de 24 horas, o n -ésimo retorno de baixa frequência do intervalo de tempo $(n-1)\Delta$ a $n\Delta$ (que corresponde neste caso a um dia) é dado por

$$y_n = y(n\Delta) - y((n-1)\Delta), \quad \Delta = 1, 2, \dots . \quad [1]$$

Para cada n (ou seja, para cada dia diferente), observamos as mudanças em $y(t)$ em subintervalos de Δ (ou seja, subintervalos a cada 24 horas). Dividindo esse dia em m intervalos de tempo igualmente espaçados entre si, podemos definir os retornos de alta frequência como

$$y_{i,n} = y\left((n-1) + \frac{i}{m}\right)\Delta - y\left((n-1) + \frac{(i-1)}{m}\right)\Delta, \quad i = 1, 2, \dots, m. \quad [2]$$

onde m é o tamanho (ou frequência) amostral diário, $\frac{\Delta}{m}$ é a duração do subintervalo (5, 15, 30, etc. minutos) e $y_{i,n}$ é o i -ésimo Δ -retorno intradiário do n -ésimo dia. Nesse

caso, os retornos realizados diários são dados por $y_n = \sum_{i=1}^m y_{i,n}$. A frequência amostral diária vai depender da duração do subintervalo escolhido para os dados. Assim, num período de 7 horas de negociações, por exemplo, com um intervalo de 5 minutos, teríamos 12 observações por hora e 84 observações em um dia. No caso do mercado de câmbio, que funciona 24 horas por dia, para uma frequência de 5 minutos, temos 288 observações num período de um dia. Há, porém, um *trade-off* entre precisão e viés na escolha de $\frac{\Delta}{m}$, devido às microestruturas de mercado (XEKALAKI; DEGIANNAKIS, 2010). De um lado, na medida em que $\frac{\Delta}{m} \rightarrow 0$, maior a precisão da estimação; por outro lado, essa frequência amostral maior aumenta o ruído no estimador de volatilidade, dado que o ruído de microestrutura de mercado aumenta. Para reduzir o caráter ruidoso gerado pelas microestruturas, um intervalo de observação bastante utilizado na literatura é o de 5 minutos entre as negociações.

3.2 Modelos de difusão, semimartingales e variação quadrática

Para dados de menor frequência, em tempo discreto, os choques são geralmente descritos como um ruído branco. Em muitas aplicações financeiras que lidam com derivativos, é bastante comum assumir que os preços seguem uma trajetória em tempo contínuo. Assim, podemos descrever a trajetória de $y(t)$ por um modelo de difusão contínuo

$$dy(t) = \mu(t)dt + \sigma(t)dW(t), \quad t > 0, \quad [3]$$

onde $\mu(t)$ é um *drift* de trajetória contínua, $\sigma(t)$ é a volatilidade *spot* e $W(t)$ é o ruído Browniano (também conhecido como processo de Wiener), sendo que $\sigma(t)$ e $W(t)$ são independentes. O retorno de um dia do processo de difusão contínuo de preços é descrito por

$$y_{i,n} = \int_{(n-1)\Delta}^{n\Delta} \mu(s)ds + \int_{(n-1)\Delta}^{n\Delta} \sigma(s)dW(s). \quad [4]$$

De acordo com essa última equação, a volatilidade do processo de difusão contínuo no intervalo $[n\Delta - (n-1)\Delta]$ é relacionada à trajetória da volatilidade *spot* $\sigma(t)$. Então, condicional às trajetórias dos processos de *drift* e da volatilidade *spot*,

$$y_{i,n} \sim N\left(\int_{(n-1)\Delta}^{n\Delta} \mu(s)ds, VI_t\right), \quad [5]$$

onde VI equivale à Variância Integrada, que é definida (naquele intervalo) como

$$VI = \int_{(n-1)\Delta}^{n\Delta} \sigma^2(s)ds. \quad [6]$$

O processo $y(t)$ também pode ser representado de outra maneira similar. De acordo com Barndorff-Nielsen e Shephard (2002b, 2004), a teoria financeira moderna é bastante baseada no uso de semimartingales⁴ (\mathcal{SM}). Segundo Taylor (2005), um processo martingale é um processo estocástico onde o valor esperado de uma observação em um tempo t , condicionado a todas as observações anteriores, é igual a observação em $t-1$. Um processo \mathcal{SM} é, então, a soma de um processo martingale local e um processo de variação finita. O processo $y(t)$ é um \mathcal{SM} se ele pode ser decomposto da seguinte maneira

$$y(t) = \mu(t) + M(t), \quad [7]$$

onde $\mu(t)$ é um *drift* com variação local finita e $M(t)$ é um martingale local. Um dos aspectos mais importantes dos \mathcal{SM} é a Variação Quadrática (VQ) (BARNDORFF-NIELSEN; SHEPHARD, 2004). De uma maneira geral, a VQ pode ser estimada de maneira consistente pela soma dos quadrados dos retornos dispostos em intervalos de

⁴ Os semimartingales são brevemente apresentados aqui. A explicação destes modelos não é o escopo deste trabalho. Para maiores detalhes, consultar Barndorff-Nielsen e Shephard (2002b, 2004).

tempo muito pequenos. Seguindo Barndorff-Nielsen e Shephard, (2004, 2006), dado que $y(t)$ é um \mathcal{SM} temos que a VQ $[y](t)$ de $y(t)$ pode ser definida como

$$[y](t) = p \lim_{m \rightarrow \infty} \sum_{i=1}^m (y(t_i) - y(t_{i-1}))^2. \quad [8]$$

para todo intervalo de tempo $[0, T]$. Decompondo $[y](t)$, temos que

$$[y](t) = [y_c](t) + [y_d](t), \quad [9]$$

onde y_c e y_d são os componentes contínuo e descontínuo (saltos) da VQ, respectivamente. Se impusermos a restrição de que $\mu(t)$ é um processo de variação local finita com trajetória contínua, o processo se torna previsível, e assim podemos fazer a seguinte simplificação⁵,

$$[y](t) = [M](t). \quad [10]$$

É importante notar que toda a teoria da VQ desenvolvida em Barndorff-Nielsen e Shephard (2002a, 2002b, 2004, 2006) é baseada na hipótese de que $y(t)$ é um \mathcal{SM} . Se assim for, conseguimos estimar consistentemente a VQ.

A maioria dos modelos \mathcal{SM} que são contínuos pode ser representada por modelos de volatilidade estocástica (\mathcal{VE}). Para estender os resultados da VQ, e assumindo a restrição anterior de que $\mu(t)$ tem trajetória contínua como em [10], precisamos de mais uma restrição, onde $M(t)$ é um processo de \mathcal{VE} . Assim,

$$M(t) = \int_0^t \sigma(s) dW(s). \quad [11]$$

⁵ A prova e uma breve discussão desse resultado podem ser encontradas em Barndorff-Nielsen e Shephard (2002b, 2004).

3.3 Variância realizada e volatilidade realizada

O incremento na VQ do tempo $(n-1)\Delta$ ao tempo $n\Delta$ é definido como

$$VQ = \int_{(n-1)\Delta}^{n\Delta} \sigma^2(s) ds, \quad [12]$$

ou seja, a equação [12] mostra que a VQ no dia $n\Delta$ é igual a VI. Barndorff-Nielsen e Shephard (2002a, 2002b) mostram que a VQ pode ser bem aproximada pela soma dos quadrados dos retornos de alta frequência

$$VR = \sum_{i=1}^m y_{i,n}^2. \quad [13]$$

A equação [13] acima é definida como Variância Realizada (VR). Barndorff-Nielsen e Shephard estão entre os primeiros autores a estabelecer uma relação formal entre a teoria da VQ com a VR. Esta funciona como uma variável *proxy*, servindo muito bem na estimação daquela. Quando o tamanho amostral m torna-se grande o suficiente, para todos os \mathcal{SM} temos que

$$VR \xrightarrow{m \rightarrow \infty} \int_{(n-1)\Delta}^{n\Delta} \sigma^2(s) ds, \quad [14]$$

ou seja, a VR converge em probabilidade para a VQ. A volatilidade realizada nada mais é do que a raiz quadrada da VR, referindo-se então ao desvio-padrão da série de preços de ativos ao invés da variância. Em econometria, é comum se referir à volatilidade realizada como em [13]. Aqui se fará a distinção entre os termos por uma mera questão de nomenclatura financeira, apesar dos termos serem bastante similares. Assim, a volatilidade realizada é dada por

$$\sqrt{VR} = \sqrt{\sum_{i=1}^m y_{i,n}^2}. \quad [15]$$

3.4 Variação bipotente

Observações mostram que modelos de difusão contínuos falham ao tentar explicar algumas evidências empíricas dos retornos dos ativos, já que as séries de dados intradiários geralmente apresentam descontinuidades (saltos)⁶. Com isso, os processos puramente contínuos se tornam inadequados. Logo, o modelo de difusão contínuo é muito restrito, e com isso devemos considerar adicionar um componente descontínuo a ele, resultando em

$$dy(t) = \mu(t)dt + \sigma(t)dW(t) + \kappa(t)dq(t), \quad t > 0, \quad [16]$$

onde $dq(t)$ é um processo binário ($dq(t) = 0$ indica ausência de saltos e $dq(t) = 1$ indica a presença de um salto) e $\kappa(t)$ é o tamanho de um salto no retorno. É geralmente assumido que q é um processo de Poisson não correlacionado com $W(t)$ e direcionado pela intensidade $\lambda(t)$ do salto, ou seja, $[P(dq(t) = 1)] = \lambda(t)dt$.

Como vimos anteriormente, sabendo que $[y](t)$ é um \mathcal{SM} e assumindo que o *drift* tem trajetória contínua, para o caso do modelo de difusão contínuo temos que

$$[M](t) = \int_0^t \sigma(s)dW(s). \quad [17]$$

Nesta representação, $\sigma(s)$ pode conter saltos, agrupados com outros componentes de microestrutura de mercado. Na literatura que envolve a estimação não paramétrica de componentes contínuos e descontínuos da VQ, supõe-se que $[y](t)$ é um modelo \mathcal{SM} browniano como em [17] ou um modelo \mathcal{SM} browniano com saltos. Neste último caso,

⁶ Essas descontinuidades datam desde Merton (1976). Mais sobre saltos no capítulo 4.

$$[M](t) = \int_0^t \sigma(s) dW(s) + \sum_{i=1}^{N(t)} \kappa_i, \quad [18]$$

onde o último componente, dado pelo somatório de κ_i , é a parte descontínua que descreve os saltos, separadamente da volatilidade *spot*. Nesta parte, $N(t)$ é o número de saltos até o tempo t e κ_i determina a magnitude do salto. A VQ do modelo \mathcal{SM} browniano com saltos é dada por

$$VQ = [y](t) = [y_c](t) + [y_d](t) = \int_0^t \sigma^2(s) ds + \sum_{i=1}^{N(t)} \kappa_i^2, \quad [19]$$

Ou seja, a equação [19] mostra que a VQ é a soma da decomposição da VI juntamente com o quadrado do termo de saltos. Na ausência de saltos, ou seja, quando $N(t) = 0$ em [19] temos que $[y_c](t) = \int_0^t \sigma^2(s) ds$ e assim, na ausência de saltos, a VQ é exatamente igual à VI. Andersen, Bollerslev e Diebold (2007) mostram que o componente de saltos é muito mais imprevisível do que o componente contínuo da VQ. Na presença de saltos, podemos da mesma maneira utilizar a VR como *proxy* para a VQ. A estimação é consistente, ou seja,

$$VR \xrightarrow{m \rightarrow \infty} \int_{(n-1)\Delta}^{n\Delta} \sigma^2(s) ds + \sum_{i=N((n-1)\Delta)+1}^{N(n\Delta)} \kappa_i^2. \quad [20]$$

De acordo com [20], a VR (diária) converge para a VQ (diária) quanto maior for o tamanho amostral (intradiário). O problema em [20] é que a variação do preço é medida de maneira única, ou seja, a trajetória contínua não é diferenciada do componente de saltos e dessa maneira não conseguimos saber qual a variação de cada componente isolado. Em outras palavras, a VR apresenta uma medida consistente da variação total do processo de preços, incluindo o processo descontínuo de saltos. Como temos a VQ derivada de dois termos, o ideal é analisá-los separadamente, distinguindo corretamente um do outro. Para isso, Barndorff-Nielsen e Shephard

(2004) introduzem o método da variação bipotente (VBP). É um método não paramétrico bastante poderoso para estimação dos saltos nas séries de preços de ativos, sendo assintoticamente não-viesado e bastante útil na obtenção da parte contínua da VQ. Começando pelo conceito de variação potente realizada (VP), temos que

$$VP = \{y_{m,n}^{(a)}\} = \left\{ \frac{\Delta^{1-a/2}}{m} \right\} \sum_{i=1}^m |y_{i,n}|^a, \quad r > 0. \quad [21]$$

O termo $\frac{\Delta^{1-a/2}}{m}$ é uma normalização do modelo. Nesse caso, quando $a = 2$, a normalização é igual a 1 e temos a variância realizada como em [13]⁷. A variância pode ser estimada de forma consistente pela VBP, que é uma generalização da VP. A VBP é construída para ser robusta a saltos, pois sua forma funcional é dada pelo produto de dois retornos absolutos consecutivos ao invés do somatório dos quadrados dos retornos, como é o caso da VR⁸. A VBP realizada é definida como

$$VBP = \{y_{m,n}^{(a,b)}\} = \left\{ \frac{\Delta^{1-(a+b)/2}}{m} \right\} \sum_{i=1}^{m-1} |y_{i,n}|^a |y_{i+1,n}|^b, \quad a, b \geq 0, \quad [22]$$

onde a e b são as potências. Dada esta equação, Barndorff-Nielsen e Shephard (2004, 2006) chegam ao seguinte resultado geral:

$$\{y_{m,n}^{(a,b)}\} \xrightarrow{m \rightarrow \infty} \mu_a \cdot \mu_b \cdot \int_{(n-1)\Delta}^{n\Delta} \sigma^{a+b}(s) ds, \quad [23]$$

ou, de outra maneira,

⁷ A prova desse e dos outros resultados apresentados em seguida se encontra em Barndorff-Nielsen e Shephard (2004).

⁸ O trabalho de Andersen, Bollerslev e Dobrev (2007) expõe trabalhos de vários outros autores que veem os modelos com saltos como uma alternativa bastante realista.

$$\mu_a^{-1} \cdot \mu_b^{-1} \cdot \{y_{m,n}^{(a,b)}\} \xrightarrow{m \rightarrow \infty} \int_{(n-1)\Delta}^{n\Delta} \sigma^{a+b}(s) ds, \text{ onde } \mu_a = 2^{a/2} \frac{\Gamma\left(\frac{1}{2}(a+1)\right)}{\Gamma\left(\frac{1}{2}\right)}, \quad [24]$$

para $a > 0$, sendo que Γ representa a função de distribuição Gama. Usando as equações obtidas acima, os autores chegam aos seguintes resultados para $a, b > 0$

$$\mu_a^{-1} \cdot \mu_b^{-1} \cdot \{y_{m,n}^{(a,b)}\} \xrightarrow{m \rightarrow \infty} \begin{cases} \int_{(n-1)\Delta}^{n\Delta} \sigma^{a+b}(s) ds, & \max(a,b) < 2 \\ x(t), & \max(a,b) = 2 \\ \infty, & \max(a,b) > 2 \end{cases} \quad [25]$$

onde $x(t)$ é um processo estocástico qualquer. Os autores atentam para o fato de somente o primeiro resultado de [25] ser usado na prática. Nesse caso, o impacto de saltos contíguos tende a zero quando $\frac{\Delta}{m} \rightarrow 0$ (ou, de maneira similar, $m \rightarrow \infty$) e $\max(a,b) < 2$. Disso decorre um caso especial quando $a \in (0,2)$. Assim, o limite de probabilidade da VBP não é afetado por saltos, ou seja,

$$\mu_a^{-1} \cdot \mu_{2-a}^{-1} \cdot \{y_{m,n}^{(a,2-a)}\} \xrightarrow{m \rightarrow \infty} \int_{(n-1)\Delta}^{n\Delta} \sigma^2(s) ds, \quad [26]$$

enquanto que

$$\{y_{m,n}^{(2)}\} - \left(\mu_a^{-1} \cdot \mu_{2-a}^{-1} \cdot \{y_{m,n}^{(a,2-a)}\} \right) \xrightarrow{m \rightarrow \infty} \sum_{i=N((n-1)\Delta)+1}^{N(n\Delta)} \kappa_i^2. \quad [27]$$

Logo, com o uso dos resultados em [26] e [27] é possível identificar separadamente a variação na trajetória de preços e dos saltos, não importando o tamanho dos saltos. Ainda, dado o que é exposto em, [6], [12] e [22], fica claro porque o interesse dos autores se volta para o caso em que $\max(a,b) < 2$, pois desse modo teremos uma estimação consistente da parte contínua da VQ. Assim, usando o caso especial de [26] e [27] quando $a = b = 1$, Barndorff-Nielsen e Shephard (2004, 2006) mostram que a VBP realizada é dada por

$$VBP = \mu_1^{-2} \cdot \{y_{m,n}^{(1,1)}\} = \mu_1^{-2} \cdot \sum_{i=1}^{m-1} |y_{i,n}||y_{i+1,n}|, \quad [28]$$

onde $\mu_1 = 2^{1/2} \frac{\Gamma(\frac{1}{2}(1+1))}{\Gamma(\frac{1}{2})} = \sqrt{2} \frac{\Gamma(1)}{\Gamma(\frac{1}{2})} = \frac{\sqrt{2}}{\sqrt{\pi}} \approx 0.79788$. Logo, $\mu_1^{-2} \cdot \{y_{m,n}^{(1,1)}\} = [y_c](t)$ converge para a parte contínua da VQ diária, mesmo com a presença de saltos,

$$VBP \xrightarrow{m \rightarrow \infty} \int_{(n-1)\Delta}^{n\Delta} \sigma^2(s) ds. \quad [29]$$

Dessa maneira, a VBP realizada consegue estimar os componentes contínuos e descontínuos separada e consistentemente. Temos assim um método simples para fazer inferência sobre a VQ. Uma última consideração a respeito de [29] é que calculamos a VBP realizada com uma pequena modificação para compensar a perda de uma observação e mantê-la comparável a VR:

$$VBP = \mu_1^{-2} \cdot \{y_{m,n}^{(1,1)}\} = \mu_1^{-2} \cdot \frac{m}{m-1} \cdot \sum_{i=1}^{m-1} |y_{i,n}||y_{i+1,n}|. \quad [30]$$

4 DEFINIÇÕES FORMAIS SOBRE *SPREAD*, SALTOS E SAZONALIDADE E ANÁLISE DOS DADOS

A análise dos modelos será baseada na série de dados intradiários dos preços das ações da Petrobrás (mais especificamente, PETR4, que correspondem a ações preferenciais), obtidos através da Economatica. A escolha deste ativo se deve à liquidez do mesmo, sendo um dos mais negociados na Bolsa de Valores, Mercadorias & Futuros de São Paulo (BM&FBOVESPA). A amostra vai de janeiro de 2007 até abril de 2011, contendo 1067 dias de negociação. Neste trabalho, dada a escolha da janela entre as observações como sendo de 5 minutos (veremos a justificativa a seguir), algumas observações em múltiplos de 5 que não tinham registro de negociação foram preenchidas com observações de minutos imediatamente anteriores ou posteriores.

Como vários dias apresentam diversas observações ausentes (quartas-feiras de cinzas e dias com paralização de uma ou mais horas seguidas nas negociações), optou-se por remover esses dias ao invés de preencher as observações ausentes, para evitar qualquer tipo de viés na VR e na VBP e, conseqüentemente, saltos espúrios. Assim, temos 1054 dias efetivos de negociação. Não foram incluídos os preços negociados no After-Market e, com isso, a matriz de preços contém 88.536 observações de preços da PETR4, com 84 observações diárias ($m = 84$). Ao converter para retornos, perdemos 1054 observações, dado que os retornos entre a última observação de um dia T qualquer e a primeira observação do dia $T + 1$ não são computados para evitar retornos *overnight*. As negociações geralmente começam as 10:00 e terminam as 17:00, com exceção do horário de verão. Nesse caso, quando as negociações começavam mais tarde devido ao horário de verão, optamos por trazer as observações para o horário normal, o que não modifica os resultados. Geralmente, na literatura, se ignora o período de pré-abertura do pregão removendo os primeiros 5 minutos de observações pós-abertura. Nesse trabalho, optamos por não remover os primeiros 5 minutos pós-abertura, já que os resultados praticamente não mudaram. Como um pouco mais de

20% das observações referentes às 10:00 estavam faltando, elas foram preenchidas por média aritmética simples entre os dois primeiros retornos do dia⁹. Por conveniência, utilizaremos agora $y_n = r_n$ para retornos diários e $y_{i,n} = r_{i,n}$ para retornos intradiários. Abaixo discutimos a evidência empírica dos componentes da microestrutura de mercado vistos no capítulo 2.

4.1 *Spread* entre preços de compra e venda

Como vimos no capítulo 3, a consistência da VR e da VBP depende do tamanho da amostra dos dados de alta frequência, ou seja, quando a frequência tende ao infinito, a combinação dessas medidas converge em probabilidade para a VQ. Porém, quanto maior a frequência dos dados que tivermos (minuto a minuto ou menos que isso), mais os dados estarão contaminados pelo ruído de microestrutura de mercado e, conseqüentemente, os resultados gerados pelo modelo serão viesados (BANDI; RUSSEL, 2004; ZHANG; MYKLAND; AÏT-SAHALIA, 2005). Em outras palavras, quanto maior for a frequência, maior será a consistência dos resultados, mas também maior a acumulação de microestrutura de mercado, o que levará a um maior viés. Seguindo os trabalhos supracitados, assumimos que

$$\tilde{y}_{i,n} = y_{i,n} + u_{i,n} \quad [31]$$

Onde $\tilde{y}_{i,n}$ é o i -ésimo log-preço observado no dia n , $y_{i,n}$ é o i -ésimo log-preço de equilíbrio (preço obtido na ausência de microestrutura de mercado) no dia n e $u_{i,n}$ é o ruído de microestruturas, por hipótese i.i.d. Para os retornos observados, temos que

⁹ Cappa e Valls Pereira (2010) utilizam a mesma metodologia em relação aos retornos *overnight* e a média aritmética simples. Os autores também utilizam a média simples para preencher os retornos em dias com muitas observações ausentes, que optamos por remover nesse trabalho. Schmid (2009) sugere o mesmo procedimento de média aritmética simples (interpolação linear) para contornar o problema das observações ausentes.

$$\tilde{r}_{i,n} = (y_{i,n} - y_{i-1,n}) + (u_{i,n} - u_{i-1,n}) = r_{i,n} + \varepsilon_{i,n}, \quad [32]$$

onde $\varepsilon_{i,n} = u_{i,n} - u_{i-1,n}$ é um processo de médias móveis de primeira ordem. Essa autocorrelação gerada pelo *spread* que causa o viés na VR.

Para reduzir a microestrutura, utilizaremos uma frequência menor. Assim, vários autores (ANDERSEN; BOLLERSLEV, 1997, 1998A, 1998B; ANDERSEN; BOLLERSLEV; DIEBOLD, 2007; ANDERSEN; BOLLERSLEV; DIEBOLD; LABYS, 2001, 2003; dentre outros citados nas referências), utilizam uma janela temporal de 5 a 30 minutos para os retornos, o que gera um bom equilíbrio entre viés e precisão, pois a redução não tão severa da frequência reduz o viés dos *spreads* sem excluir muita informação relevante, necessária para a maior consistência e precisão das estimativas. Podemos ver na Tabela 1 que para retornos de 1 minuto ainda há muito ruído causado pelas microestruturas de mercado, dado que a autocorrelação negativa na primeira defasagem é a maior dentre as frequências apresentadas. Ainda há autocorrelação negativa na primeira defasagem nos retornos de 5 minutos, mas ela é menor do que nos retorno de 1 minuto, o que indica que o ruído causado pelas microestruturas caiu com a diminuição da frequência.

Tabela 1: Autocorrelação de primeira ordem e estatística Box-Pierce dos retornos intradiários para cada intervalo.

	1 minuto	5 minutos	30 minutos	60 minutos	24 horas
$\rho_1(r_{i,n})$	-0,0398	-0,0284	-0,0063	-0,0170	0,0369
$Q_1(r_{i,n})$	519,828 (0,0000)	94,7975 (0,0000)	0,4242 (0,5148)	2,2423 (0,1342)	1,4570 (0,2274)
$Q_{10}(r_{i,n})$	61,691 (0,0000)	23,711 (0,0084)	14,266 (0,1611)	14,389 (0,1559)	11,405 (0,3268)

Fonte: elaboração própria. Abaixo das estatísticas de Box-Pierce se encontram seus respectivos p-valor.

Como dito anteriormente, é uma escolha entre viés e precisão. A partir dos 30 minutos em diante, a autocorrelação negativa já é irrisória, tanto que a estatística de Box-Pierce não rejeita a hipótese nula de a autocorrelação ser zero, com nível de significância de 10%. Neste trabalho, trataremos do problema de seleção entre precisão e viés aumentando a frequência para retornos observados de 5 a 5 minutos. Essa é uma solução provisória para lidar com a autocorrelação. Veremos mais adiante uma medida adicional para lidar com o ruído de microestruturas de mercado.

A Figura 1 mostra a FAC das 10 primeiras defasagens dos retornos de vários intervalos de escolha entre as observações. Como podemos ver pelo gráfico, de acordo com o intervalo de confiança de 95%, as autocorrelações são significativas em várias defasagens além da primeira para os intervalos de 1 e 5 minutos, o que contraria a hipótese de Andersen e Bollerslev (1997) discutida no capítulo 2 de que não há autocorrelação positiva além da primeira defasagem. O *bid-ask bounce* não aparece nos retornos para 30 e 60 minutos, mas há autocorrelação significativa em outras defasagens.

4.2 Saltos

Modelos de difusão simples, como em [3], são bastante convenientes para análise, pois a hipótese de que os preços seguem uma trajetória contínua simplifica a teoria de precificação de derivativos¹⁰. Porém, é bastante improvável que haja esse tipo de trajetória contínua em dados observados empiricamente. Supondo que essa desconfiança seja verdadeira, há então suspeitas de descontinuidades nas séries intradiárias de preços de ativos, chamados saltos. Os saltos são definidos como observações incomumente grandes que não fazem parte do processo difusivo contínuo. As explicações para essas descontinuidades aparecem em Merton (1976) pela primeira

¹⁰ Andersen, Bollerslev e Diebold (2007) demonstram que as medidas usuais para determinar a volatilidade de uma série de preços são em sua maioria baseadas na hipótese de que os preços seguem uma trajetória contínua.

vez. O autor sugere que os saltos em um ativo derivam da chegada de novas informações no mercado, referentes principalmente a indústria ou a firma específica do ativo. Desde então, outros autores, da mesma maneira, passaram a ligar a ocorrência de saltos à disponibilidade de informações que chegam aos mercados¹¹. Nesse caso, um mercado eficiente incorpora as novas informações imediatamente, causando as descontinuidades no processo até então contínuo.

Logo, os saltos implicam em dificuldades para os problemas de precificação de derivativos. Surge assim a necessidade de detectarmos e estudarmos os saltos, dada a sua importância na explicação das variações nos preços. Nessa dissertação, iremos verificar a presença de saltos na série de dados intradiários da PETR4. Vimos anteriormente que a estimação da VQ por VBP é robusta na presença de saltos, pois ela consegue separar as trajetórias e com isso temos uma estimação mais precisa da VQ. Utilizaremos então o teste proposto por Barndorff-Nielsen e Shephard (2004, 2006), que é um teste bastante utilizado na literatura. Segundo estes autores, a diferença entre a VR e a VBP realizada dá uma estimação não paramétrica consistente do componente de saltos. O resultado deriva diretamente de [27],

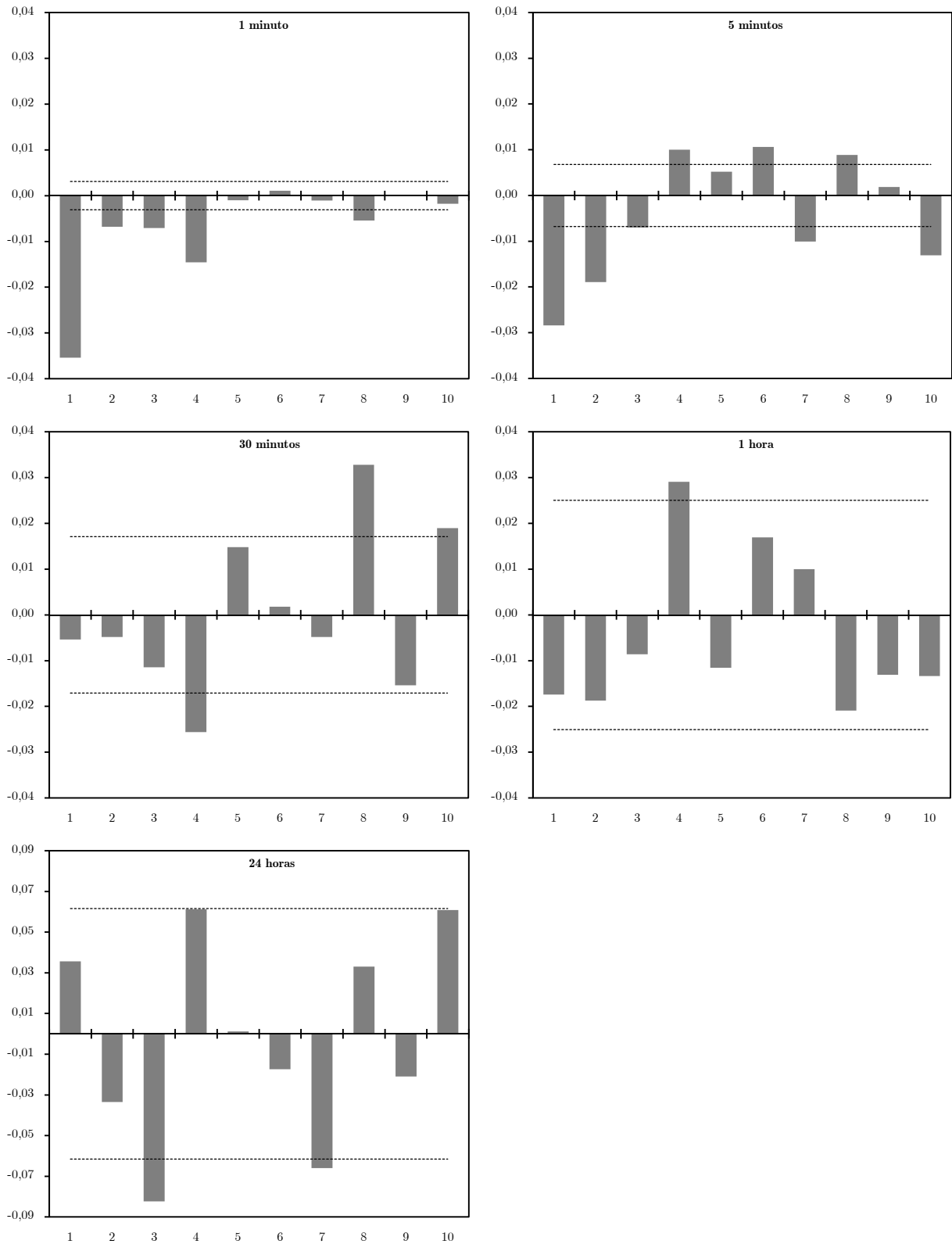
$$\{y_{m,n}^{(2)}\} - \left(\mu_1^{-2} \cdot \{y_{m,n}^{(1,1)}\}\right) = \sum_{i=1}^m r_{i,n}^2 - \left(\mu_1^{-2} \cdot \sum_{i=1}^{m-1} |r_{i,n}| |r_{i+1,n}|\right) \xrightarrow{m \rightarrow \infty} \sum_{i=N(\Delta(n-1))+1}^{N(\Delta n)} \kappa_i^2. \quad [33]$$

Com isso, temos a base para formular um teste. Necessitamos de uma estatística para testar a significância desse salto, que faz uso do desvio padrão condicional, dado pela Quarticidade Integrada (QI), definida como

$$QI = \int_0^t \sigma^4(s) ds. \quad [34]$$

¹¹ Christensen, Oomen e Podolskij (2011), ao estudar dados de altíssima frequência (na casa dos milissegundos), concluem que a influência causada pelos saltos na trajetória dos preços é irrisória, ou seja, o impacto dos saltos acaba praticamente desaparecendo quando usamos um intervalo de observação muito pequeno.

Figura 1: Função de autocorrelação das 10 primeiras defasagens de cada intervalo temporal para os retornos puros, juntamente com seus intervalos de confiança.



Fonte: Elaboração própria.

Para estimar a QI, Barndorff-Nielsen e Shephard (2004) utilizam a Quarticidade Tetra-potente ($QP^{[4]}$),

$$QP^{[4]} = m \cdot \mu_1^{-4} \cdot \frac{m}{m-3} \cdot \sum_{i=1}^{m-3} |r_{i,n}| |r_{i+1,n}| |r_{i+2,n}| |r_{i+3,n}|, \quad [35]$$

enquanto que Andersen, Bollerslev e Diebold (2007) utilizam a Quarticidade Tri-potente ($QP^{[3]}$),

$$QP^{[3]} = m \cdot \mu_{4/3}^{-3} \cdot \frac{m}{m-2} \cdot \sum_{i=1}^{m-2} |r_{i,n}|^{\frac{4}{3}} |r_{i+1,n}|^{\frac{4}{3}} |r_{i+2,n}|^{\frac{4}{3}} \quad [36]$$

Dadas essas estimativas para a VR, VBP e QI, podemos montar um teste formal. Andersen, Bollerslev e Diebold (2007), baseando-se nos resultados encontrados por Barndorff-Nielsen e Shephard (2006), sugeriram a seguinte estatística

$$Z_{QP^{[3]}} = \frac{VR - VBP}{\sqrt{(\mu_1^{-4} + 2\mu_1^{-2}) \cdot \frac{1}{m} \cdot QP^{[3]}}}. \quad [37]$$

Seguindo Andersen, Bollerslev, Diebold e Labys (2001, 2003), Andersen, Bollerslev e Diebold (2007) utilizam uma modificação de [37], utilizando os logaritmos das estimativas realizadas. Essa alteração visa melhorar a performance do teste quando utilizamos uma amostra finita,

$$Z_{\log QP^{[3]}} = \frac{\log(VR) - \log(VBP)}{\sqrt{(\mu_1^{-4} + 2\mu_1^{-2}) \cdot \frac{1}{m} \cdot \left(\frac{QP^{[3]}}{VBP^2}\right)}}. \quad [38]$$

Ainda, outra modificação que aparece em Barndorff-Nielsen e Shephard (2005) utiliza um ajuste de máximo da estatística [38],

$$Z_{\max \log QP^{[3]}} = \frac{\log(VR) - \log(VBP)}{\sqrt{\left(\mu_1^{-4} + 2\mu_1^{-2}\right) \cdot \frac{1}{m} \cdot \max\left(1, \frac{QP^{[3]}}{VBP^2}\right)}}. \quad [39]$$

De maneira análoga a [37], [38] e [39], podemos utilizar a $QP^{[4]}$ no lugar da $QP^{[3]}$ como estimativa para a QI. Huang e Tauchen (2005) sugerem o uso dos saltos relativos (SR), que informam a proporção dos saltos sobre a VR,

$$SR = \frac{VR - VBP}{VR}. \quad [40]$$

Após várias simulações de Monte Carlo, usando as estatísticas descritas em [37], [38] e [39] – também com o uso da $QP^{[4]}$ e dos saltos relativos – Huang e Tauchen (2005) verificam que o uso dos SR no numerador da estatística Z é considerado melhor, já que as outras estatísticas tendem a rejeitar demais a hipótese nula de ausência de saltos. Além disso, outro resultado das simulações afirma que o uso da $QP^{[3]}$ como estimador da QI é melhor. Dessa forma, após testarem todas estatísticas, os autores concluem, de forma similar a Barndorff-Nielsen e Shephard (2006)¹², que

$$Z_{SR, QP^{[3]}} = \frac{SR}{\sqrt{\left(\mu_1^{-4} + 2\mu_1^{-2}\right) \cdot \frac{1}{m} \cdot \max\left(1, \frac{QP^{[3]}}{VBP^2}\right)}} \rightarrow N(0,1). \quad [41]$$

Assim, utilizaremos a estatística sugerida por Huang e Tauchen (2005)¹³, já que ela é a melhor estatística para amostras finitas dentre as apresentadas acima. Calculada a estatística Z , testamos cada dia a hipótese nula de ausência de saltos usando o teste- z usual. Se o valor de Z for maior que o valor crítico do teste- z no nível de significância

¹² Barndorff-Nielsen e Shephard (2006) usam uma medida parecida, mas ela é negativa, ou seja, $-SR$. O resultado assintótico de [39] e [40] é, então, idêntico.

¹³ Andersen, Bollerslev e Diebold (2007), Bollerslev, Law e Tauchen (2008), Giot, Laurent e Petitjean (2010), Ishida e Watanabe (2009) e Rognlie (2010) também fazem o uso dos saltos relativos.

escolhido, rejeita-se a hipótese nula e, por consequência, concluímos que deve haver um salto. Formalmente,

$$S = I[Z > \Phi^{-1}(\alpha)] \cdot (VR - VBP), \quad [42]$$

onde $I[\bullet]$ é uma função indicadora¹⁴ e $\Phi^{-1}(\alpha)$ é o valo crítico ao nível de significância α da função distribuição Normal acumulada Φ . Rejeitada a hipótese nula, o valor do salto deve ser eliminado da série da VR. Com isso, o estimador da VQ é dado pela função

$$C = I[Z \leq \Phi^{-1}(\alpha)] \cdot VR + I[Z > \Phi^{-1}(\alpha)] \cdot VBP. \quad [43]$$

Logo, em um determinado dia t , quando não se rejeita a hipótese nula de ausência de saltos, a função C retorna a VR; caso contrário, na presença de saltos, a função C retorna a VBP.

Contudo, identificar esses saltos não é uma tarefa fácil quando utilizamos amostras finitas. Ao detectarmos uma observação anormal no processo difusivo contínuo de preços, há a dúvida se ela é resultado do próprio processo difusivo (devido a uma observação anômala gerada pelo ruído Browniano) ou se é um salto. Dito de outra forma, não sabemos se o salto é espúrio ou não. Segundo Bollerslev, Law e Tauchen (2008), quando fazemos os testes em vários dias seguidos, ocorre o problema de testes múltiplos, onde a expectativa de que ocorram saltos espúrios equivale ao nível de significância do teste individual multiplicado pelo número de dias da amostra. Com isso, se utilizarmos um nível de significância de 5%, há expectativa de que haja 52 saltos espúrios na nossa amostra de 1054 dias, enquanto que para um nível de significância de 1% a expectativa seria de 10 saltos. A proposta dos autores é

¹⁴ A função indicadora retorna $I[\bullet] = \begin{cases} 1 & \text{se a condição } [\bullet] \text{ é atendida} \\ 0 & \text{se a condição } [\bullet] \text{ não é atendida} \end{cases}$

então reduzir o nível de significância do teste para 0,1%, havendo dessa maneira a expectativa de somente 1 salto espúrio¹⁵ neste trabalho. Aqui há um *trade-off* entre poder do teste e quantidade de saltos espúrios.

Como vimos na Figura 1, os retornos intradiários observados de 5 a 5 minutos possuem autocorrelação negativa não só na primeira defasagem, mas também em defasagens posteriores. Para amenizar essa autocorrelação, Andersen, Bollerslev e Diebold (2007) sugerem utilizar retornos não contíguos no cálculo da VBP e da QP^[3], pulando um retorno. Os autores ainda sugerem que se há autocorrelação além da primeira defasagem, podemos pular mais de um retorno. Assim, para a comparação das estatísticas, utilizaremos

$$VBP = \mu_1^{-2} \cdot \frac{m}{m - (1 + k)} \cdot \sum_{i=1}^{m-(1+k)} |y_{i,n}| |y_{i+(1+k),n}|$$

e [44]

$$QP^{[3]} = m \cdot \mu_{4/3}^{-3} \cdot \frac{m}{m - 2(1 + k)} \cdot \sum_{i=1}^{m-2(1+k)} |r_{i,n}|^{\frac{1}{3}} |r_{i+(1+k),n}|^{\frac{1}{3}} |r_{i+2(1+k),n}|^{\frac{1}{3}},$$

onde k é o número de retornos a serem pulados. Caso $k = 0$, temos as fórmulas originais como em [30] e [36].

A Tabela 2 mostra o resultado para os três níveis de significância mais utilizados na literatura sobre dados intradiários. Comparando a estatística $Z_{SR,QP^{[3]}}$ com a estatística $Z_{\max \log QP^{[3]}}$, que usam o mesmo denominador, encontramos uma diferença considerável. Quando a VBP é calculada em sua forma sequencial normal (com $k = 0$), o número de saltos é bem menor em relação ao cálculo da VBP com

¹⁵ Giot, Laurent e Petitjean (2010) e Ishida e Watanabe (2009) também utilizam o nível de significância de 0,1%.

intervalos entre a multiplicação dos retornos. Entre $k = 1$ e $k = 2$ a diferença não é tão grande quanto de $k = 0$ para $k = 1$. Esse é um resultado conhecido e discutido em Huang e Tauchen (2005), onde os autores mostram que no caso de $k = 0$ as estatísticas Z são viesadas devido à autocorrelação dos retornos, fazendo com que menos saltos sejam detectados. O viés aparece dos dois lados: no numerador e no denominador. No numerador, pela VBP e, no denominador, pela $QP^{[3]}$, ambos contaminados pelo ruído de microestruturas. Os dois efeitos juntos causam viés na estatística Z contra a rejeição da hipótese nula. A escolha sugerida de k para este trabalho, baseada na autocorrelação além da primeira ordem, é de dois pulos entre os retornos, para garantir a eliminação do ruído de microestruturas. Huang e Tauchen (2005) também utilizam $k = 0, 1, 2$, sendo que eles atentam para o fato de que $k > 2$ leva ao problema da rejeição excessiva da hipótese nula. Um dos problemas de utilizar $k \neq 0$ é que isso piora os resultados assintóticos das estatísticas, como veremos adiante.

Tabela 2: Número de saltos e sua respectiva proporção em relação ao total de dias, para cada nível de significância.

	$\Phi^{-1}(0,95)$	$\Phi^{-1}(0,99)$	$\Phi^{-1}(0,999)$	
Saltos espúrios esperados	52 (5%)	10 (1%)	1 (0,1%)	
$k = 0$	$Z_{\max \log QP^{[3]}}$	177 (16,79%)	86 (8,15%)	36 (3,41%)
	$Z_{SR, QP^{[3]}}$	153 (14,51%)	62 (5,88%)	19 (1,80%)
$k = 1$	$Z_{\max \log QP^{[3]}}$	280 (26,56%)	148 (14,04%)	72 (6,83%)
	$Z_{SR, QP^{[3]}}$	240 (22,77%)	106 (10,05%)	31 (2,94%)
$k = 2$	$Z_{\max \log QP^{[3]}}$	314 (29,79%)	184 (17,45%)	102 (9,67%)
	$Z_{SR, QP^{[3]}}$	289 (27,41%)	139 (13,18%)	65 (6,16%)

Fonte: Elaboração própria.

A Tabela 3 mostra várias estatísticas para algumas medidas estudadas neste trabalho. Para as séries puras, as estatísticas de normalidade são bastante altas, diminuindo consideravelmente quando usamos a transformação logarítmica (ou seja, a

transformação logarítmica ajuda na distribuição das séries, fazendo com que elas fiquem mais próximas da Normal). Neste caso, a estatística Jarque-Bera é mais próxima da distribuição Normal para o log da VBP em relação ao log da VR. A série da magnitude dos saltos é a que apresenta maior assimetria e curtose entre as séries. Em relação a função C , seu log apresenta a distribuição mais próxima da Normal entre as séries de volatilidade. No caso dos retornos, vemos que há uma assimetria negativa e caudas pesadas, o que corrobora com os fatos estilizados vistos no capítulo 2. Quando os retornos são padronizados pela volatilidade realizada, a hipótese de normalidade não é rejeitada a 5% de significância. Os resultados destas transformações são similares aos encontradas em Andersen, Bollerslev, Diebold e Labys (2001).

As Figuras 2 e 3 exibem o agrupamento de volatilidade característico de séries financeiras, presentes na série de retornos realizados (retornos diários, obtidos pela soma dos retornos intradiários). Podemos ver que a volatilidade é maior em períodos de maior turbulência econômica – como no caso da crise do *subprime* em 2008. Uma exceção a isso aparece na Figura 2 no caso dos retornos realizados padronizados, onde este é muito parecido com um ruído branco.

Outro ponto a ser notado na Figura 3 é que a primeira grande variação nos retornos em 2007, por exemplo, não é detectada como um salto na função C , mesmo com essa mesma variação na VBP sendo menor; ou seja, nem toda observação elevada é um salto, podendo então ser resultado do movimento Browniano do processo de difusão contínuo. A Figura 4 mostra o seguinte: a magnitude dos saltos ($VR - VBP$); os saltos relativos $\left(\frac{VR-VBP}{VR}\right)$; a estatística $Z_{SR,QP^{[3]}}$ com $\alpha = 0,1\%$; e a contribuição percentual do salto sobre a VR $\left(100 \cdot \left(\frac{VR-VBP}{VR}\right)\right)$. Dadas as estatísticas dos SR na Tabela 3, vemos que eles correspondem em média a 8,68% da VR. Para verificar se há de fato a piora na distribuição assintótica da estatística para $k \neq 0$, vemos que a estatística $Z_{SR,QP^{[3]}}$ para $k = 2$ apresenta uma estatística

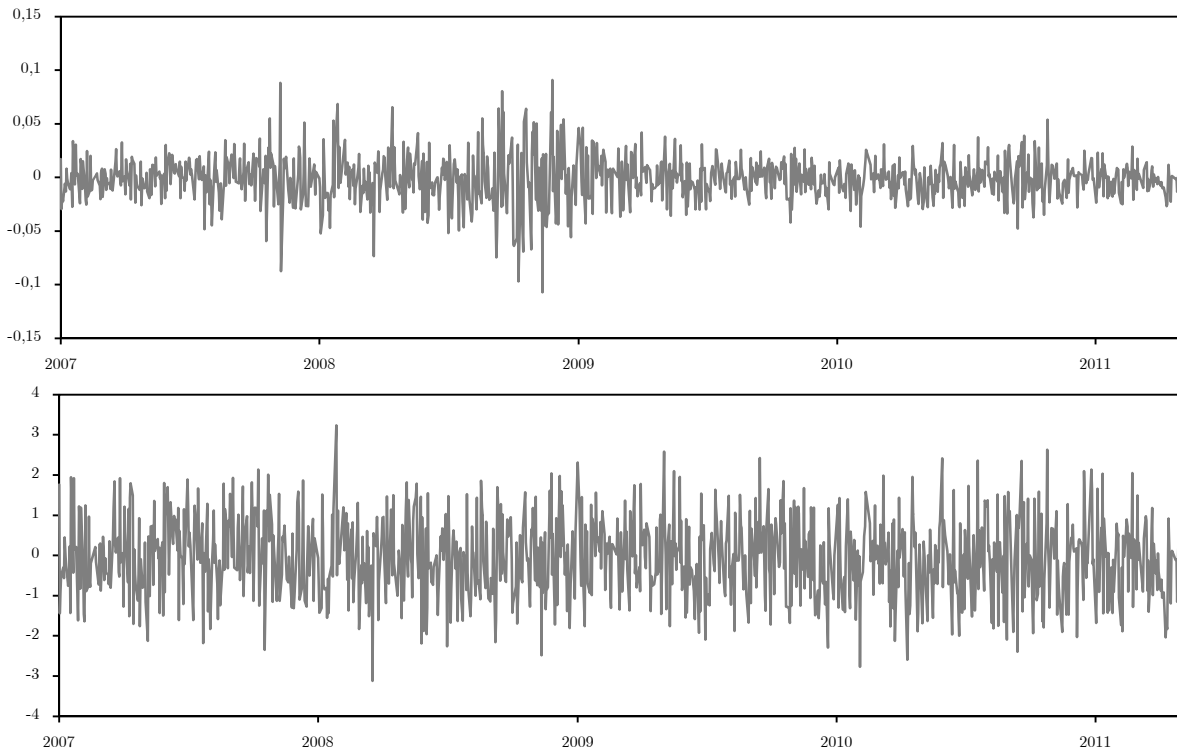
Tabela 3: Estatísticas descritivas das séries da PETR4.

Série	Média	DP	Mínimo	Máximo	Assimetria	Curtose	JB
VR	4,6694 (x10 ⁻⁴)	7,4515 (x10 ⁻⁴)	0,4209 (x10 ⁻⁴)	109 (x10 ⁻⁴)	6,5817	63,9180	18,7 (x10 ⁴)
VBP (k = 2)	4,3053 (x10 ⁻⁴)	7,0821 (x10 ⁻⁴)	0,3000 (x10 ⁻⁴)	107,90 (x10 ⁻⁴)	6,7226	66,912	20,4 (x10 ⁴)
$\sqrt{\text{VR}}$	1,8955 (x10 ⁻²)	1,0382 (x10 ⁻²)	0,6487 (x10 ⁻²)	10,4400 (x10 ⁻²)	2,7919	12,0910	7789
VR - VBP	36,4 (x10 ⁻⁴)	1,2325 (x10 ⁻⁴)	5,7959 (x10 ⁻⁴)	18,35 (x10 ⁻⁴)	7,8564	92,816	38,9 (x10 ⁴)
SR	0,0868	0,12503	-0,2340	0,6786	0,5187	0,7583	72
C	4,5590 (x10 ⁻⁴)	7,3557 (x10 ⁻⁴)	0,4114 (x10 ⁻⁴)	109 (x10 ⁻⁴)	6,7706	67,335	20 (x10 ⁴)
r_n	-0,1374 (x10 ⁻²)	2,0672 (x10 ⁻²)	-10,7090 (x10 ⁻²)	9,0742 (x10 ⁻²)	-0,0505	2,5993	170
log VR	-8,1397	0,8596	-10,076	-4,5190	0,7952	0,9003	146
log VBP	-8,2409	0,87815	-10,414	-4,5291	0,7707	0,8843	138
log $\sqrt{\text{VR}}$	-4,0698	0,4298	-5,0379	-2,2595	0,7952	0,9003	146
log $\left(\frac{\text{VR}}{\text{VBP}}\right)$	0,10125	0,14852	-0,2048	1,1363	1,2458	3,9088	943
log C	-8,1676	0,8655	-10,098	-4,5190	0,75799	0,89019	135
$\frac{r_n}{\sqrt{\text{VR}}}$	-0,0600	0,9839	-3,1160	3,2329	0,0902	-0,0322	5,97*

Fonte: Elaboração própria. DP e JB se referem ao Desvio-Padrão e a estatística Jarque-Bera, respectivamente. Todas as estatísticas de JB rejeitam a hipótese de normalidade, com exceção da estatística JB para os retornos padronizados $\frac{r_n}{\sqrt{\text{VR}}}$, que aceita a hipótese de normalidade para 5% de significância.

Jarque-Bera de 78, enquanto que para $k = 0$ ela é cerca de 48. De qualquer forma, ambas rejeitam a hipótese nula de normalidade a 10% de significância.

Figura 2: Retornos Realizados e Retornos Realizados Padronizados, respectivamente.

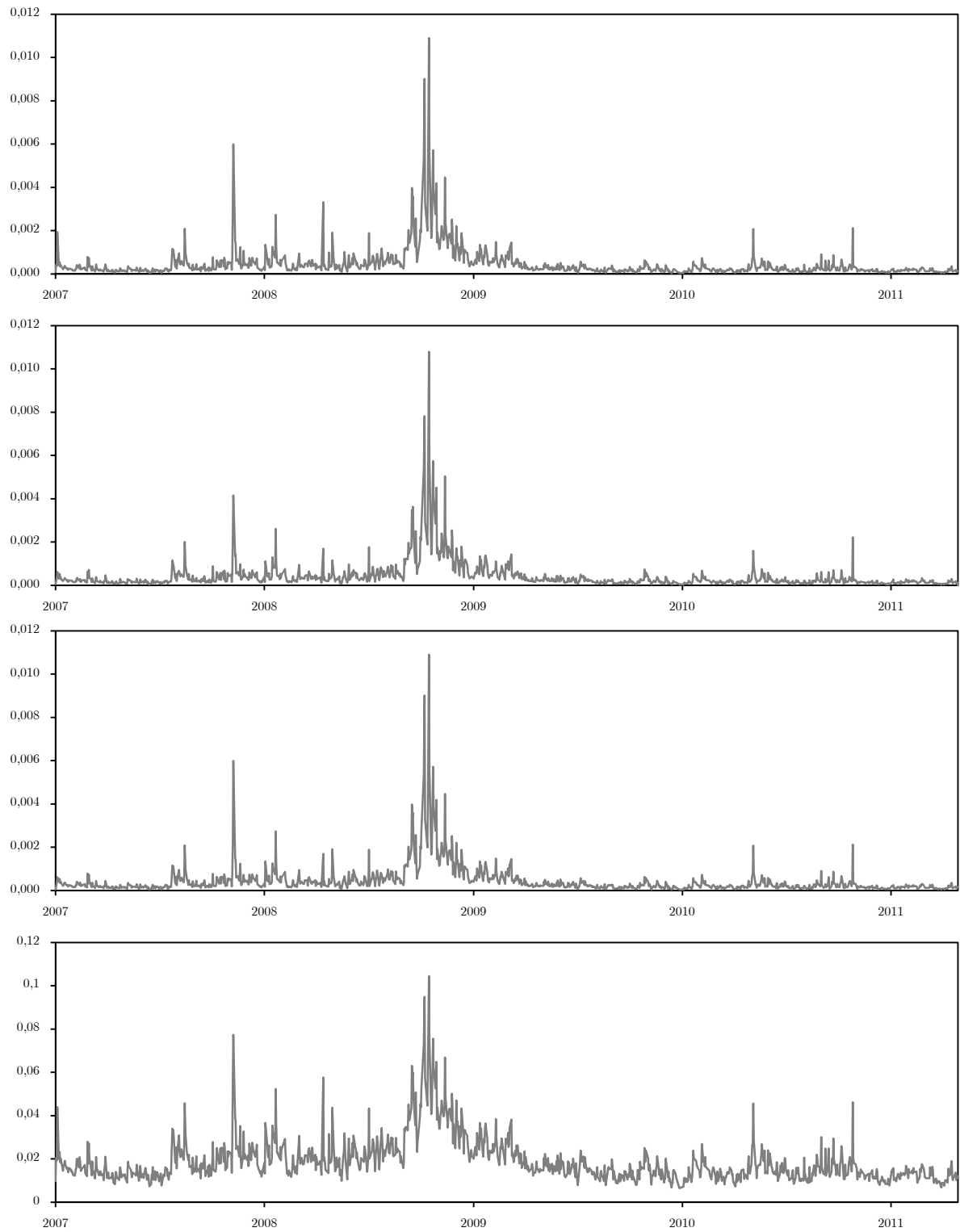


Fonte: Elaboração própria.

4.3 Sazonalidade intradiária

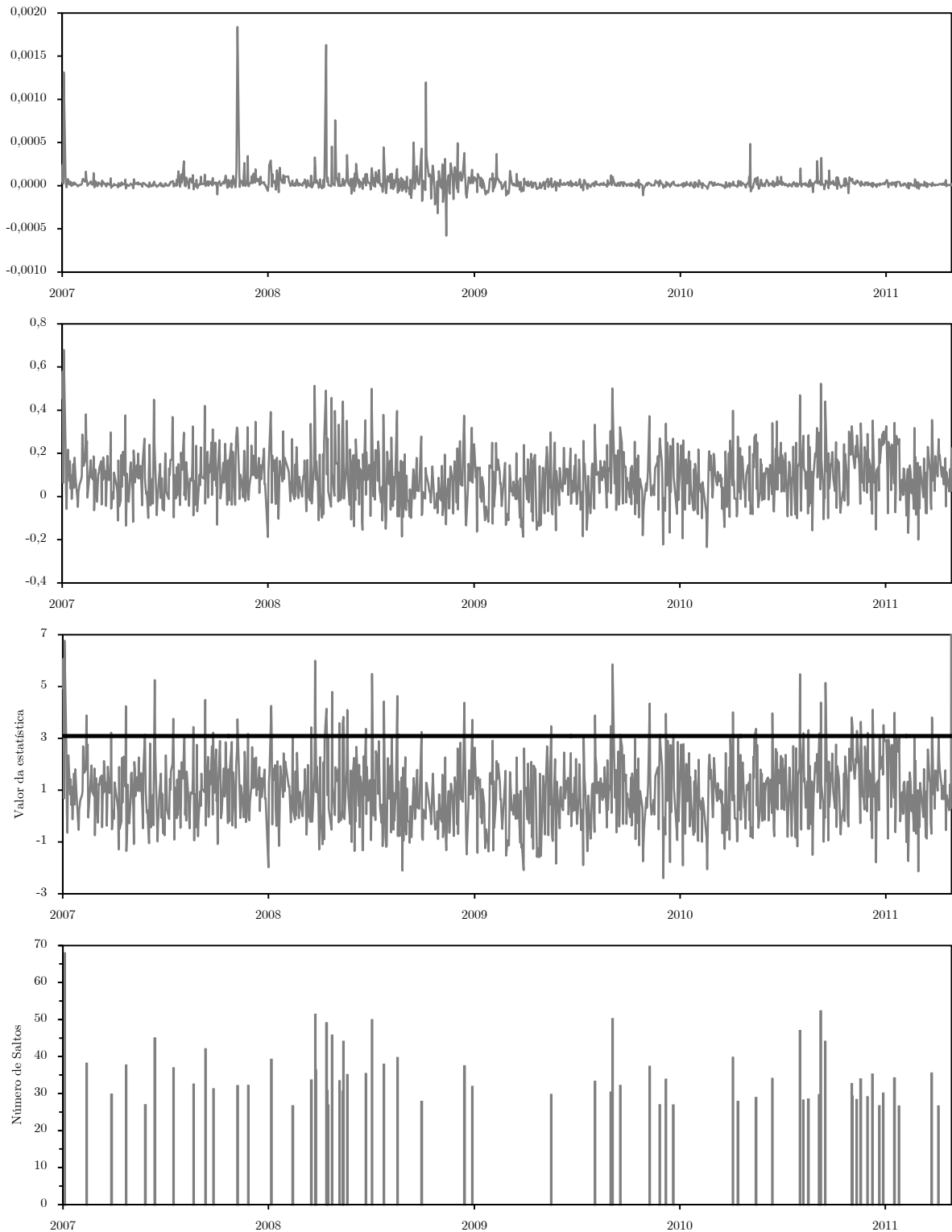
Esse padrão intradiário é derivado das grandes variações na volatilidade ao longo do dia, o que acarreta em viés na estimativa dos saltos, podendo gerar saltos espúrios. Assim, em períodos de alta volatilidade há tendência de detectar mais saltos do que o normal, enquanto que em períodos de baixa volatilidade há tendência de detectar menos saltos. Logo, é interessante verificar o quão importante essa sazonalidade intradiária interfere na estimação dos saltos. Para isso, utilizamos o valor absoluto dos retornos como *proxy* para volatilidade e tiramos a média de cada intervalo de 5 minutos contidos na amostra (10:05 até 16:55) para verificar a presença de tal padrão. A Figura 5 mostra os resultados obtidos ao longo de toda a amostra.

Figura 3: Variância Realizada, Variação Bipotente ($k = 2$), Função C e Volatilidade Realizada, respectivamente.



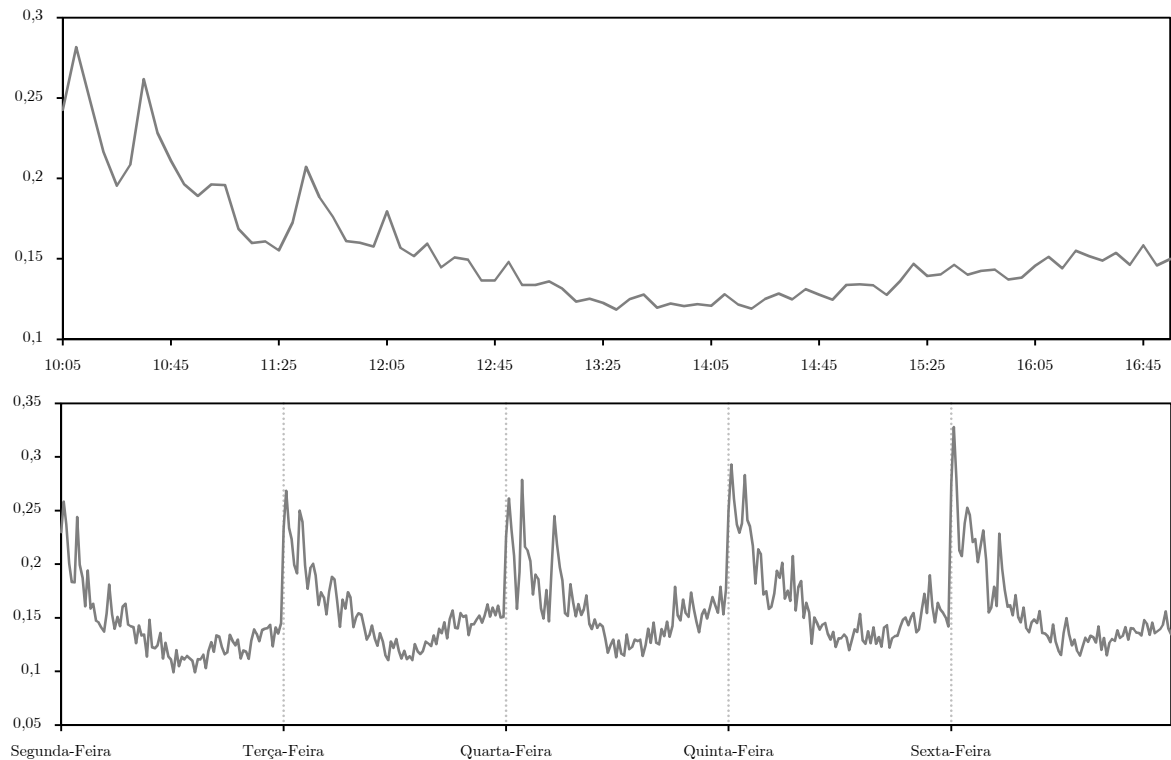
Fonte: Elaboração própria.

Figura 4: Magnitude dos saltos, Saltos Relativos, Estatística $Z_{SR,QP^{[3]}}$ ($\alpha = 0,1\%$ e $k = 2$) e Contribuição % dos saltos detectados sobre a VR, respectivamente.



Fonte: Elaboração própria. Nota: a linha horizontal para a estatística $Z_{SR,QP^{[3]}}$ refere-se ao valor crítico $\Phi^{-1}(0,999) = 3,09023$.

Figura 5: Padrão Intradiário da volatilidade (em toda a amostra) e Padrão Semanal, respectivamente.



Fonte: Elaboração própria.

Fica evidente na Figura 5 o padrão intradiário tipicamente observado em retornos intradiários, no formato J invertido¹⁶. A volatilidade ao final do dia não parece ser muito maior que a observada por volta da metade do dia. Ainda na Figura 5 vemos que na sexta-feira a volatilidade é a maior no começo do dia em comparação com outros dias da semana, possivelmente devido ao fato de que os agentes não querem manter posições durante os finais de semana. A Figura 6 desmembra a volatilidade vista em toda a amostra em períodos menores para observar se a tendência é isolada em alguma periodicidade específica ou não. Para cada ano e para cada dia da semana, o padrão parece se manter, o que é confirmado pelo gráfico de linhas empilhadas, que vai somando a contribuição de cada efeito. Os picos dos

¹⁶ Cappa e Valls Pereira (2010), utilizando uma série intradiária da PETR4, chegam a resultados similares em relação ao padrão intradiário da volatilidade.

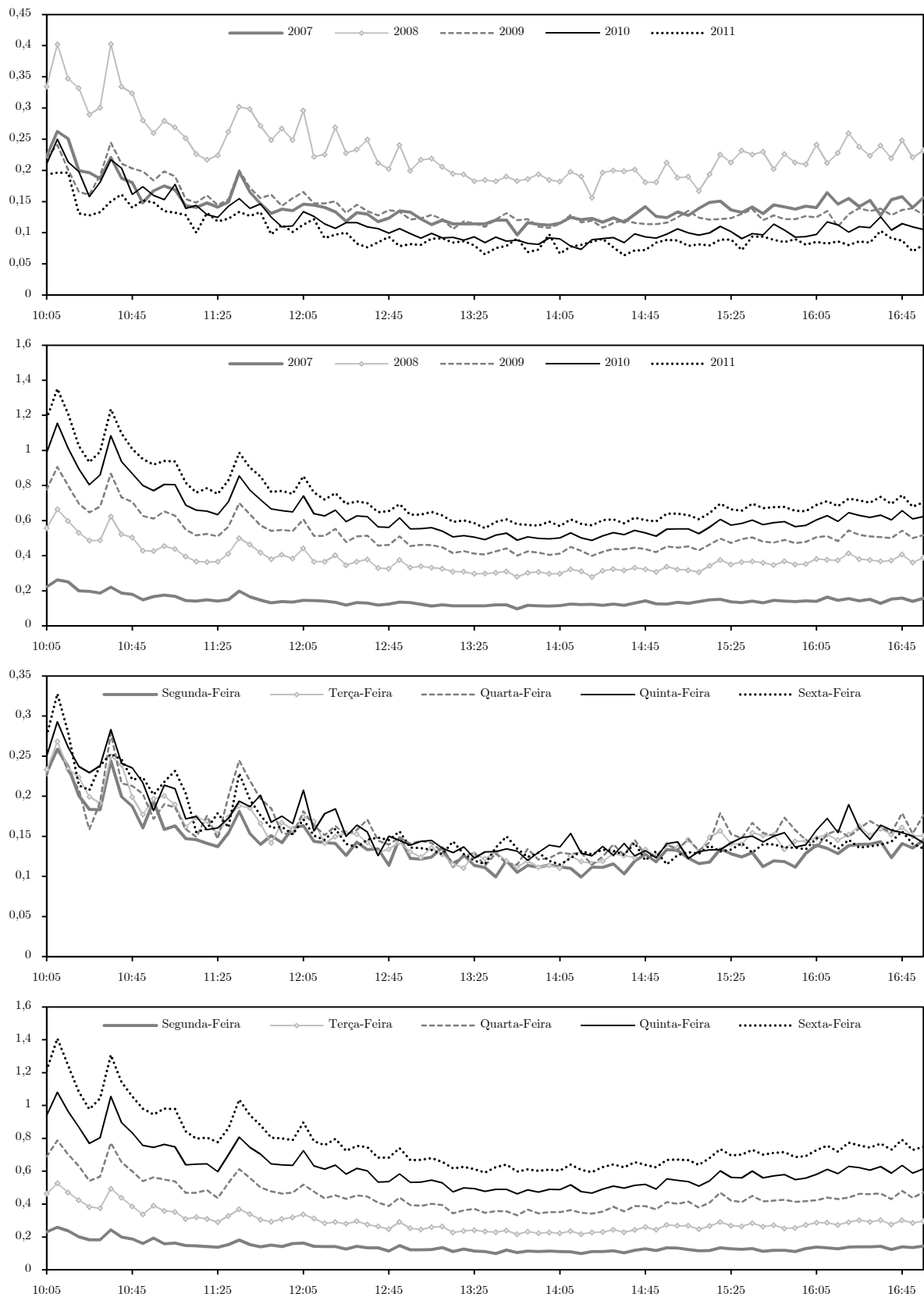
gráficos com linhas empilhadas ficam cada vez maiores no começo do dia, enquanto que no resto do dia isso não é evidente.

Olhando os gráficos em série da Figura 6, vemos que no caso da periodicidade anual, 2008 foi um ano bastante volátil devido à crise do *subprime*, ficando descolado dos outros anos, que se mantem mais coesos. O ano de 2011, apesar de contabilizar na amostra apenas quatro meses, já mostrou que tem um padrão similar aos dos outros anos completos. Em relação a periodicidade semanal, o padrão mantém-se parecido, mas individualmente a segunda-feira vem abaixo dos outros dias, enquanto que quinta e sexta-feira apresentam os maiores picos intradiários no começo do pregão, e quarta e quinta-feira no meio pro final do pregão¹⁷. Para remover a sazonalidade intradiária, devemos recorrer a alguma técnica. Segundo Rognlie (2010), o viés que ocorre nos testes de saltos vem dos diferentes padrões de volatilidade entre retornos consecutivos. Se conseguirmos manter um padrão estável para a volatilidade ao longo de toda a amostra, a estatística de teste será menos viesada. Para isso, Rognlie (2010) sugere padronizar os retornos pela raiz quadrada da VBP média local, que é uma medida robusta a saltos. Dessa maneira, temos

$$\bar{r}_{i,n} = \begin{cases} \frac{r_{1,n}}{\sqrt{\frac{1}{m} \sum_{n=1}^T |r_{1,n}| |r_{2,n}|}} & i = 1 \\ \frac{r_{i,n}}{\sqrt{\frac{1}{2m} \left(\sum_{n=1}^T |r_{i-1,n}| |r_{i,n}| + \sum_{n=1}^T |r_{i,n}| |r_{i+1,n}| \right)}} & 1 < i < c, \\ \frac{r_{c,n}}{\sqrt{\frac{1}{m} \sum_{n=1}^T |r_{c-1,n}| |r_{c,n}|}} & i = c \end{cases} \quad [45]$$

¹⁷ Este é outro fato que parece confirmar a eliminação de posições mantidas pelos agentes na sexta-feira, já que a volatilidade está entre as maiores no começo do dia e decai bastante após as 15:00, inclusive ficando próxima do padrão observado na segunda-feira. Esta peculiaridade é chamada de efeito sexta-feira.

Figura 6: Padrão Intradiário Anual e Semanal (em série e empilhado).

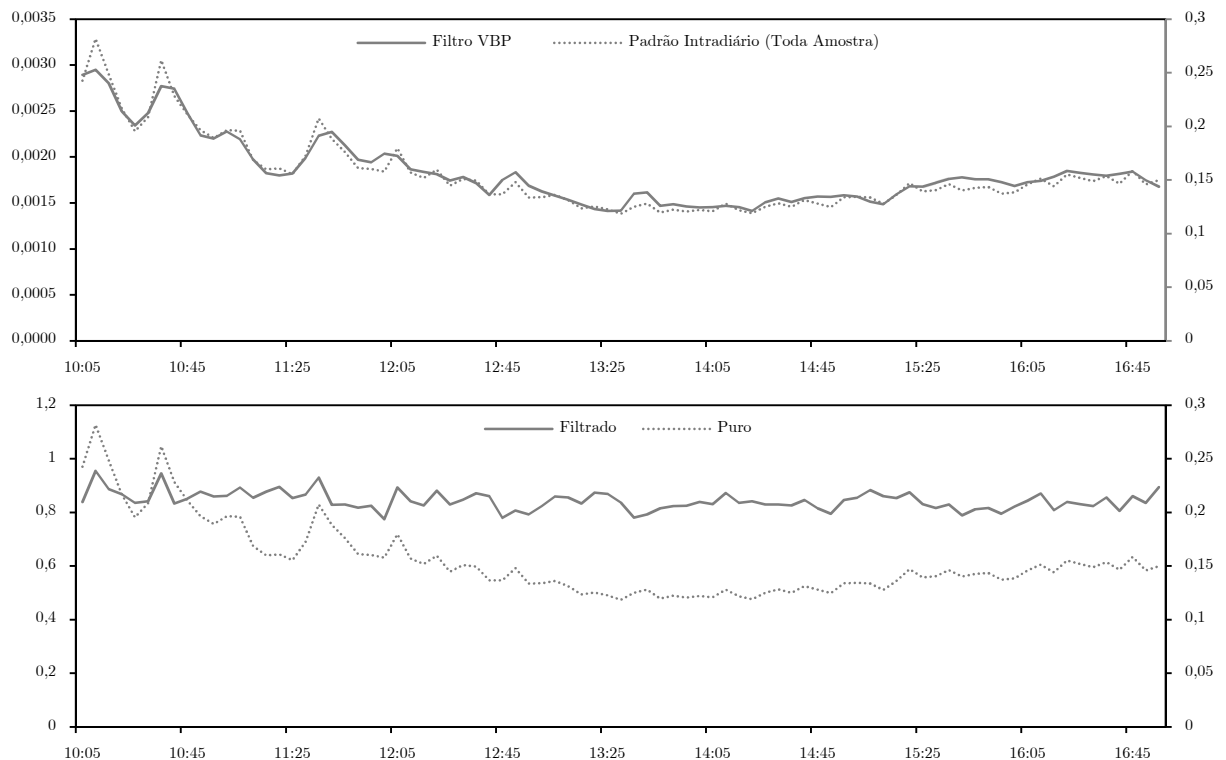


Fonte: Elaboração própria.

onde c é um índice para a última observação intradiária. Através de [45] obtemos um filtro sazonal composto de m valores (que serão usados como denominador). Depois, dividimos cada valor dos retornos puros pelo filtro no horário respectivo.

A Figura 7 compara o filtro e o padrão intradiário da Figura 5, bem como a série pura e filtrada da média dos retornos absolutos. O filtro de VBP mostra seu efeito contra saltos: onde o padrão intradiário apresenta picos mais elevados, o filtro mostra uma suavização, justamente para evitar saltos espúrios.

Figura 7: Filtro VBP e Padrão Intradiário (Filtrado e Puro), respectivamente.

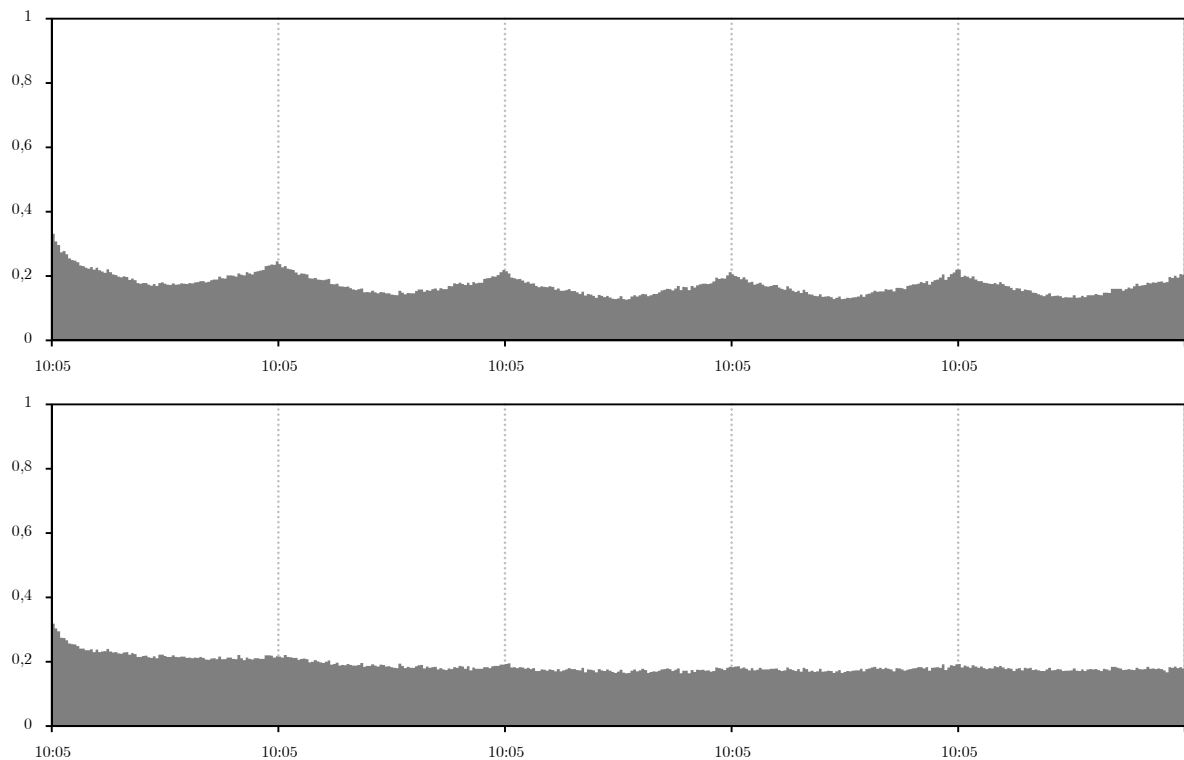


Fonte: Elaboração própria.

Ao dividirmos os retornos pelo filtro, vemos que a série fica ajustada (série “Filtrado”), não apresentando mais o padrão em J invertido (série “Puro”, mostrada para comparação). Desse modo, utilizando os retornos filtrados, esperamos que a estatística $Z_{SR,QP^{[3]}}$ seja menos viesada, já que agora a volatilidade apresenta um comportamento mais estável que anteriormente. A Figura 8 contém as FAC dos 415

(415 = 83.5 = 1 semana) primeiros retornos puros e filtrados¹⁸. Os retornos puros apresentam o formato em U, enquanto que os retornos filtrados estão livres desse padrão na autocorrelação.

Figura 8: Função de Autocorrelação para retornos puros e filtrados (período de 5 dias), respectivamente.



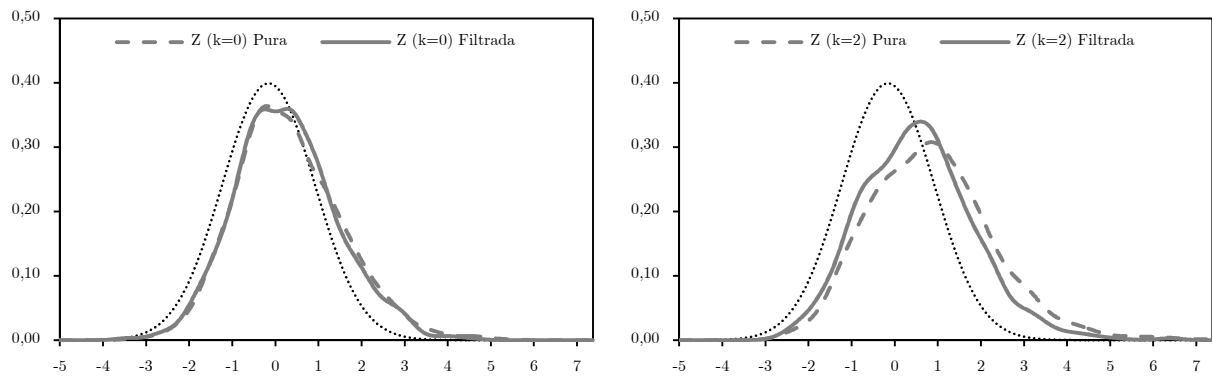
Fonte: Elaboração própria.

A Figura 9 apresenta as densidades estimadas por Kernel para a estatística $Z_{SR,QP^{[3]}}$ usando retornos puros e filtrados, tanto no caso onde não há pulos entre retornos quanto no caso onde há dois pulos. Em ambos os casos, a densidade dos retornos filtrados é mais próxima da Normal padrão do que a densidade dos retornos puros. Quando não há pulos no cálculo da VBP e $QP^{[3]}$, a distribuição é melhor assintoticamente do que quando há dois pulos, confirmando o que foi encontrado por Huang e Tauchen (2005). A Tabela 4 mostra as estatísticas Jarque-Bera para

¹⁸ O padrão se repete acima das 415 observações.

$Z_{SR,QP^{[3]}}$. O melhor resultado assintótico é no caso em que não há pulos entre os retornos e eles estão filtrados. Quando há dois pulos, o retorno filtrado também apresenta uma estatística de normalidade melhor.

Figura 9: Estimativas de densidade por Kernel para a estatística $Z_{SR,QP^{[3]}}$, VBP sem pulos (pura e filtrada) e VBP com dois pulos (pura e filtrada), respectivamente.



Fonte: Elaboração própria.

Tabela 4: Estatísticas de Normalidade

Jarque-Bera de $Z_{SR,QP^{[3]}}$.

	$k = 0$	$k = 2$
Retornos Puros	48 (0,00)	78 (0,00)
Retornos Filtrados	18 (0,00)	61 (0,00)

Fonte: Elaboração própria. Nota: Entre parênteses se encontra o p-valor da estatística.

Analisando os saltos, vemos que na Tabela 5 eles diminuíram em relação aos saltos em retornos puros para todos os níveis de significância, o que confirma a suspeita de que haja saltos espúrios advindos da sazonalidade. Assim, a sazonalidade intradiária é um fenômeno relevante no estudo dos saltos na série de VR. A Tabela 6 mostra a diferença entre os saltos detectados para retornos puros e filtrados ano a ano. Analisando os números brutos, 2010 é o ano que mais teve saltos espúrios (12), enquanto que em proporção o maior resultado fica com 2011 (75% dos saltos para

retornos filtrados não foram detectados em relação aos retornos puros), seguido por 2009 e 2010 (66,66%).

Tabela 5: Saltos detectados para $Z_{SR,QP^{[3]}}$,
retornos puros e filtrados.

		$\Phi^{-1}(0,95)$	$\Phi^{-1}(0,99)$	$\Phi^{-1}(0,999)$
$k = 0$	Puros	153 (14,51%)	62 (5,88%)	19 (1,80%)
	Filtrados	134 (12,71%)	52 (4,93%)	10 (0,94%)
$k = 2$	Puros	289 (27,41%)	139 (13,18%)	65 (6,16%)
	Filtrados	199 (18,88%)	80 (7,59%)	31 (2,94%)

Fonte: Elaboração própria. Entre parênteses está a proporção entre saltos detectados e o total de dias da amostra.

Tabela 6: Saltos detectados em cada ano para
 $Z_{SR,QP^{[3]}}$, $k = 2$, $\alpha = 0,1\%$, retornos puros e filtrados.

	2007	2008	2009	2010	2011	Total
Puros	14	20	9	18	4	65
Filtrados	8	13	3	6	1	31
Diferença	6 (42,85%)	7 (35%)	6 (66,66%)	12 (66,66%)	3 (75%)	34 (52,30%)

Fonte: Elaboração própria. Entre parênteses está a proporção entre a diferença de saltos e os saltos detectados para retornos puros.

5 CONCLUSÕES

Este trabalho procurou mostrar a importância dos saltos e do correto tratamento do ruído de microestrutura presente nos dados financeiros de alta frequência. Quando estimamos a variação quadrática pela conjugação da variância realizada e variação bipotente, estamos contando com a correta detecção de saltos pelo teste, para que a trajetória contínua estimada dos preços seja a mais precisa possível. Sendo assim, se há presença de microestrutura de mercado e sazonalidade intradiária, devemos corrigir estes fatores para evitar viés na correta detecção dos saltos.

Ao analisar os dados financeiros da PETR4, primeiro observamos se há ruído de microestrutura causado pelo *spread* de compra e venda, que por hipótese só introduz autocorrelação negativa na primeira defasagem. Para diferentes frequências, vimos que de fato, ao aumentarmos o intervalo entre as observações dos preços e retornos, a autocorrelação de primeira ordem decai, o que indica que a microestrutura diminui (inclusive fica nula para certos intervalos em diante) com intervalos cada vez maiores. Como há um *trade-off* entre viés e precisão, optamos pelo intervalo de 5 minutos entre as observações de retornos. Selecionado o intervalo, partimos para a estimação da variação quadrática via variância realizada (estima a trajetória contínua e descontínua) e variação bipotente (estima a trajetória contínua somente, sem os saltos). Tendo estas séries disponíveis, bem como a QI, podemos montar um teste de saltos. Como para o intervalo de 5 minutos ainda há microestrutura, inclusive além da primeira defasagem, optamos pela metodologia de Andersen, Bollerslev e Diebold (2007) para reduzir a microestrutura, pulando dois retornos entre o cálculo da variação bipotente. Ainda, seguindo Huang e Tauchen (2005), usamos a estatística de saltos relativos, também utilizada por outros autores na literatura de dados de alta frequência.

Saltos foram detectados e eles são, de fato, parte importante no processo de difusão de preços. Caso fosse utilizada a VR tão somente para estimar a VQ,

estariamos sempre adicionando este componente no processo contínuo, o que pode gerar interpretações erradas sobre a trajetória correta dos preços. Nesse caso, as informações capazes de remodelar as expectativas dos agentes em relação aos preços ao longo do dia, que não fazem parte da trajetória contínua, estariam sempre presentes, descrevendo esta trajetória de forma incorreta. De outra forma, um retorno fora do comum pode ser resultado da trajetória normal do ruído Browniano. Nesse caso, a correta detecção de saltos também é fundamental, para que essa observação não seja detectada como tal e permaneça inclusa na trajetória contínua. Vimos que a microestrutura de fato causa uma detecção menor de saltos, como visto na literatura. Corrigida a microestrutura, o número de saltos aumenta. Porém, a correta detecção de saltos ainda é uma questão aberta¹⁹. As diferentes abordagens mostradas anteriormente e testadas através de simulações na literatura mostram que esse fenômeno ainda vai ser amplamente estudado, dada a cada vez maior disponibilidade de dados intradiários e interesse acadêmico.

Vimos também que os saltos geralmente são atribuídos a certos eventos que modificam os preços. Nesse caso, os saltos são estocásticos. Porém, podem ocorrer saltos devido a sazonalidade intradiária dos dados – fato que foi confirmado no capítulo 4. Assim, retornos acima do esperado seriam resultado da maior volatilidade presente nos horários de abertura e encerramento do pregão. Esses saltos não podem ser detectados como saltos estocásticos como os relatados no primeiro caso. Assim, a remoção do padrão intradiário se faz fundamental. Ao remover a sazonalidade intradiária, esperamos que menos saltos sejam detectados, o que de fato ocorre na série de dados intradiários da PETR4.

Portanto, para uma estimação que contenha mais informações sobre o processo de movimentação dos preços, os dados intradiários se mostram úteis e viáveis. Contudo, devemos estar atentos ao fato de que há o problema de microestrutura de

¹⁹ Há outros testes para detecção de saltos, mas eles não serão abordados nesse trabalho, dado que o teste de Barndorff-Nielsen e Shephard (2006) é ainda muito utilizado na literatura. Para uma literatura a respeito da detecção de saltos intradiários, ver Lee e Mykland (2008).

mercado e sazonalidade intradiária que introduzem um viés na correta detecção dos saltos – alheios à trajetória contínua dos preços – que ocorrem ao longo do dia devido a fatores diversos. Tratados estes problemas, esperamos ter uma estimação bastante precisa da trajetória contínua da VQ.

REFERÊNCIAS

AKERLOF, G. A. The Market for ‘Lemons’: Quality Uncertainty and the Market Mechanism. **Quarterly Journal of Economics**, n. 84, p. 488-500, 1970.

ANDERSEN, T. G.; BOLLERSLEV, T. Intraday Periodicity and Volatility Persistence in Financial Markets. **Journal of Empirical Finance**, n. 4, p. 115-158, 1997.

_____. Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts. **International Economic Review**, n. 39, p. 885-905, 1998a.

_____. DM-Dollar Volatility: Intraday Activity Patterns, Macroeconomic Announcements and Longer Run Dependencies. **The Journal of Finance**, n. 53, p. 219-265, 1998b.

ANDERSEN, T. G.; BOLLERSLEV, T.; DIEBOLD, F. X. Roughing it Up: Including Jump Components in the Measurement, Modeling and Forecasting of Return Volatility. **The Review of Economics and Statistics**, Vol. 89, p. 701-720, 2007.

ANDERSEN, T. G.; BOLLERSLEV, T.; DIEBOLD, F. X.; LABYS, P. The Distribution of Exchange Rate Volatility. **Journal of the American Statistical Association**, n. 96, p. 42-55, 2001.

_____. Modeling and Forecasting Realized Volatility. **Econometrica**, n. 71, p. 529-626, 2003.

ANDERSEN, T. G.; BOLLERSLEV, T.; DIEBOLD, F. X.; VEJA, C. Micro Effects of Macro Announcements: Real-Time Price Discovery in Foreign Exchange. **American Economic Review**, n. 93, p. 38-62, 2003.

BACK, K. Asset Pricing for General Processes. **Journal of Mathematical Economics**, n. 20, p. 371-395, 1991.

BANDI, F. M.; RUSSELL, J. R. Microstructure Noise, Realized Volatility, And Optimal Sampling. **Econometric Society 2004 Latin American Meetings**, Econometric Society, n. 220, 2004.

BARNDORFF-NIELSEN, O. E.; SHEPHARD, N. Econometric Analysis of Realized Volatility and its Use in Estimating Stochastic Volatility Models. **Journal of the Royal Statistical Society**, n. 64, p. 253-280, 2002a.

_____. Estimating Quadratic Variation Using Realized Variance. **Journal of Applied Econometrics**, n. 17, p. 457-477, 2002b.

_____. Power and Bipower Variation with Stochastic Volatility and Jumps. **Journal of Financial Econometrics**, n. 2, p. 1-37, 2004.

_____. How Accurate Is The Asymptotic Approximation to the Distribution of Realized Volatility? **Cambridge University Press**, p. 306-331, 2005.

_____. Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation. **Journal of Financial Econometrics**, n. 4, p. 1-30, 2006.

BLACK, F. Studies of Stock Market Volatility Changes. **Proceedings of the American Statistical Association**, p. 177-181, 1986.

BOLLERSLEV, T. Generalized Autoregressive Conditional Heteroskedasticity. **Journal of Econometrics**, n. 31, p. 307-327, 1986.

BOLLERSLEV, T; LAW, T. H.; TAUCHEN, G. Risk, Jumps, and diversification. **Journal of Econometrics**, n. 144, p. 234-256, 2008.

CAPPA, L. e VALLS PEREIRA, P. L. Modelando a volatilidade dos retornos de Petrobrás usando dados de alta frequência. **Textos para discussão da escola de economia de São Paulo da Fundação Getúlio Vargas**, n. 258, 2010.

CHRISTENSEN, K.; OOMEN, R. C. A.; PODOLSKIJ, M. Fact or Friction: Jumps at Ultra High Frequency, **Working Paper Series**, 2011.

CONT, R. Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues. **Quantitative Finance**, n. 1, p. 223-236, 2001.

ENGLE, R. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. **Econometrica**, n. 50, p. 987-1007, 1982.

_____. The Econometrics of Ultra-High-Frequency Data. **Econometrica**, n. 68, p. 1-22, 2000.

ENGLE, R.; RUSSEL, J. R. Analysis of High Frequency Financial Data. **Handbook of Financial Econometrics**, 2004.

FAMA, E. The Behavior of Stock-market Prices. **The Journal of Business**, n. 38, p. 34-105, 1965.

GIOT, P.; LAURENT, S.; PETITJEAN, M. Trading Activity, Realized Volatility and Jumps. **Journal of Empirical Finance**, p. 168-175, 2010.

HUANG, X.; TAUCHEN, G. The Relative Contribution of Jumps to Total Price Variance. **Journal of Financial Econometrics**, p. 456-499, 2005.

ISHIDA, I.; WATANABE, T. Modeling and Forecasting the Volatility of the Nikkei 225 Realized Volatility Using the ARFIMA-GARCH Model. **Hi-stat Discussion Paper**, n. 32, 2009.

LAAKKONEN, H. Exchange Rate Volatility, Macro Announcements and the Choice of Intraday Seasonality Filtering Method. **Research discussion papers: Bank of Finland**, 2007.

LEE, S. S.; MYKLAND, P. A. Jumps in Financial Markets: A New Nonparametric Test and Jump Dynamics. **Review of Financial Studies**, n. 21, p. 2535-2563, 2008.

MANDELBROT, B. The Variation of Certain Speculative Prices. **The Journal of Business**, n. 36, p. 394-419, 1963.

MERTON, R. Option Pricing When Underlying Stock Returns are Discontinuous. **Journal of Financial Economics**, p. 125-144, 1976.

MISHKIN, F. S. Asymmetric Information and Financial Crisis. **National Bureau of Economic Research: Financial Markets and Financial Crisis**, p. 69-108, 1990.

_____. Anatomy of a Financial Crisis. **Journal of Evolutionary Economics**, p. 115-130, 1992.

ROGNLIE, M. **Spurious Jump Detection and Intraday Changes in Volatility**. Duke University, 2010.

SCHMID, F. High Frequency Financial Markets Data Cleaning and Stylized Facts in Financial Markets Data. **Seminar of Economic and Social Statistics**, 2009.

SCHWERT, G. W. Why Does Stock Market Volatility Change Over Time? **Journal of Finance**, Vol. 44, p. 1115-1153, 1989.

TAYLOR, S. J. **Asset Price Dynamics, Volatility and Prediction**. Princeton University Press, 2005.

TSAY, R. S. **Analysis of Financial Time Series**. 3. ed. John Wiley & Sons, 2010.

WOOD, R. A.; MCINISH, T. H.; ORD, J.K. An Investigation of Transaction Data for NYSE Stocks. **Journal of Finance**, Vol. 40, p. 723-739, 1985.

WRIGHT, Z.; ZHOU, H. Bond Risk Premia and Realized Jump Volatility. **Finance and Economics Discussion Series**, 2007.

XEKALAKI, E.; DEGIANNAKIS, S. **ARCH Models for Financial Applications**. 1. ed. John Wiley & Sons Ltd, 2010.

ZHANG, L.; MYKLAND, P. A.; AÏT-SAHALIA, Y. A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High Frequency Data. **Journal of The American Statistical Association**, n. 100, p. 1394-1411, 2005.