



Gustavo Santos Raposo

**Análise de Dados de Alta Frequência e do
Processo de Formação de Preços: O Modelo
Multivariado Exponencial – EMACM**

Tese de Doutorado

Tese apresentada ao Programa de Pós-Graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica como parte dos requisitos parciais para obtenção do título de Doutor em Engenharia Elétrica.

Orientador: Prof. Álvaro de Lima Veiga Filho

Rio de Janeiro
Março de 2006



Gustavo Santos Raposo

**Análise de Dados de Alta Frequência e do
Processo de Formação de Preços: O Modelo
Multivariado Exponencial – EMACM**

Tese apresentada como requisito parcial para obtenção do grau de Doutor pelo Programa de Pós-Graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica do Centro Técnico Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Dr. Álvaro de Lima Veiga Filho
Orientador

Departamento de Engenharia Elétrica - PUC-Rio

Dr. Cristiano Augusto Coelho Fernandes
Departamento de Engenharia Elétrica - PUC-Rio

Dr. Marcelo Cunha Medeiros
Departamento de Economia – PUC-Rio

Dr. Antonio Marcos Duarte Júnior
IBMEC

Dr. Caio Ibsen Rodrigues de Almeida
IBMEC

Dr. Gyorgy Varga
QUANTUM/FGV

Prof. José Eugenio Leal
Coordenador Setorial do Centro
Técnico Científico - PUC-Rio

Rio de Janeiro, 06 de março de 2006

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Gustavo Santos Raposo

Doutorado em Engenharia Elétrica (2006) PUC/Rio, Mestrado em Engenharia de Produção (2004) PUC/Rio e Graduação em Engenharia Civil (1998) UFRJ.

Ficha Catalográfica

Raposo , Gustavo Santos

Análise de dados de alta frequência e do processo de formação de preços: modelo multivariado exponencial - Emacm / Gustavo Santos Raposo ; orientador: Álvaro de Lima Veiga Filho. - Rio de Janeiro : PUC, Departamento de Engenharia Elétrica, 2006.

131 f. ; 30 cm

Dissertação (mestrado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica.

Inclui referências bibliográficas.

1. Engenharia elétrica – Teses. 2. Dados de alta frequência. 3. GARCH. 4. Modelo condicional multivariado. 5. Séries temporais não-lineares. 6. Modelo Ordered. I. Veiga Filho, Álvaro de Lima. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. III. Título.

CDD: 621.3

A meus pais, esposa, irmãos, amigos,
colegas de profissão e colaboradores.

Agradecimentos

- Ao Professor Álvaro Veiga, orientador da tese de doutorado, pelo apoio e confiança depositados.
- Ao Professor Cristiano Fernandes, pelo apoio, orientação e estímulo.
- À Algorithmics do Brasil, pelo apoio concedido no desenvolvimento da pesquisa.
- Aos meus pais, pelo incentivo.
- À minha esposa Gisele Salomão Raposo, por todo o carinho e compreensão.
- Ao CNPq, pela ajuda financeira recebida durante o curso.

Resumo

Raposo, Gustavo Santos; Veiga Filho, Álvaro de Lima (Orientador). **Análise de Dados de Alta Freqüência e do Processo de Formação de Preços: O Modelo Multivariado Exponencial – EMACM**. Rio de Janeiro, 2006. 131p. Tese de Doutorado - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

A modelagem de dados que qualificam as transações de ativos financeiros, tais como, preço, *spread* de compra e venda, volume e duração, vem despertando o interesse de pesquisadores na área de finanças, levando a um aumento crescente do número de publicações referentes ao tema. As primeiras propostas se limitaram aos modelos de duração. Mais tarde, o impacto da duração sobre a volatilidade instantânea foi analisado. Recentemente, Manganelli (2002) incluiu dados referentes aos volumes transacionados dentro de um modelo vetorial. Neste estudo, nós estendemos o trabalho de Manganelli através da inclusão do *spread* de compra e venda num modelo vetorial autoregressivo, onde as médias condicionais do *spread*, volume, duração e volatilidade instantânea são descritas a partir de uma formulação exponencial chamada *Exponential Multivariate Autoregressive Conditional Model* (EMACM). Nesta nova proposta, não se fazem necessárias a adoção de quaisquer restrições nos parâmetros do modelo, o que facilita o procedimento de estimação por máxima verossimilhança e permite a utilização de testes de Razão de Verossimilhança na especificação da forma funcional do modelo (estrutura de interdependência). Em paralelo, a questão de antecipar movimentos nos preços de ativos financeiros é analisada mediante a utilização de um procedimento integrado, no qual, além da modelagem de dados financeiros de alta freqüência, faz-se uso de um modelo *probit ordenado* contemporâneo. O EMACM é empregado com o objetivo de capturar a dinâmica associada às variáveis e sua função de previsão é utilizada como *proxy* para a informação contemporânea necessária ao modelo de previsão de preços proposto.

Palavras-chave

Dados de alta freqüência, GARCH, Modelo Condicional Multivariado, Séries Temporais Não-lineares, Modelo Ordered Probit

Abstract

Raposo, Gustavo Santos; Veiga Filho, Álvaro de Lima (Advisor). **High Frequency Data and Price-Making Process Analysis: The Exponential Multivariate Autoregressive Conditional Model – EMACM**. Rio de Janeiro, 2006, 131p. Doctorate Thesis - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

The availability of high frequency financial transaction data – price, spread, volume and duration – has contributed to the growing number of scientific articles on this topic. The first proposals were limited to pure duration models. Later, the impact of duration over instantaneous volatility was analyzed. More recently, Manganelli (2002) included volume into a vector model. In this document, we extended his work by including the bid-ask spread into the analysis through a vector autoregressive model. The conditional means of spread, volume and duration along with the volatility of returns evolve through transaction events based on an exponential formulation we called Exponential Multivariate Autoregressive Conditional Model (EMACM). In our proposal, there are no constraints on the parameters of the VAR model. This facilitates the maximum likelihood estimation of the model and allows the use of simple likelihood ratio hypothesis tests to specify the model and obtain some clues about the interdependency structure of the variables. In parallel, the problem of stock price forecasting is faced through an integrated approach in which, besides the modeling of high frequency financial data, a contemporary ordered probit model is used. Here, EMACM captures the dynamic that high frequency variables present, and its forecasting function is taken as a proxy to the contemporaneous information necessary to the pricing model.

Keywords

High Frequency Data, GARCH, Autoregressive Conditional Multivariate Models, Nonlinear Time Series, Ordered Probit Model

Sumário

1. Relevância do Trabalho	12
1.1. Introdução	12
1.2. Estrutura do trabalho	16
2. Modelos de Volatilidade Condicional	18
2.1. Introdução	18
2.2. <i>Autoregressive Conditional Heteroskedasticity</i> (ARCH)	19
2.2.1. Representação alternativa	21
2.2.2. Estimacão dos parâmetros (máxima verossimilhança)	22
2.3. Outras formulações propostas	23
2.3.1. GARCH	23
2.3.2. IGARCH	26
2.3.3. ARCH-M (ARCH “na média”)	27
2.3.4. E-GARCH (<i>Exponential</i> GARCH)	28
2.3.4.1. Estimacão dos parâmetros: máxima verossimilhança	29
3. Teoria de Microestruturas de Mercado	30
3.1. Introdução	30
3.2. Dados de alta frequência	31
3.3. Modelagem de dados oriundos de transações financeiras	32
3.3.1. Modelo <i>probit</i> ordenado	33
3.3.1.1. Especificacão básica	33
3.3.1.2. Distribuicão condicional das variaçõs de preço	36
3.3.1.3. Processo de estimacão (máxima verossimilhança)	39
3.4. Modelo <i>logit</i> ordenado	40
4. Modelos de duracão condicional	43
4.1. Introdução	43
4.2. Histórico	44
4.3. Características dos dados	45
4.4. Processos de intensidade condicional	48
4.5. Principais modelos	52
4.5.1. Modelo ACD	52
4.5.2. Modelo EACD (<i>Exponential</i> ACD)	54
4.5.2.1. Características principais	54
4.5.3. Modelo WACD (Weibull ACD)	55
4.5.3.1. Distribuicão condicional da duracão: Weibull	55
4.5.3.2. Verossimilhança condicional	56
4.5.3.3. Observaçõs importantes	57
4.5.4. Modelo TACD	58
4.5.5. Modelo AACD (<i>Augmented</i> ACD – Transformacão Box-Cox)	62
4.5.6. Modelo conjunto para duracão, volume e volatilidade condicionais	65
4.5.6.1. Formulaçã de Manganeli	66

5. Modelo de Dados de Alta Frequência	72
5.1. Introdução	72
5.2. O EMACM	73
5.3. Ajuste sazonal (padrão intradiário)	78
5.4. Estimacão dos modelos	79
5.4.1. Programacão seqüencial quadrática	79
5.4.2. Método Simplex de Nelder-Mead	80
5.5. Capacidade de identificacão do modelo	82
5.6. Análise empírica	86
5.6.1. Base de dados	86
5.7. Testes empíricos	88
5.7.1. Principais resultados	90
6. Modelo de Previsão de Preços	95
6.1. Introdução	95
6.2. Modelo proposto	96
6.2.1. EMACM	96
6.2.2. O Modelo de Apreçamento (Modelo <i>Probit</i> Ordenado)	97
6.2.3. Testes Empíricos	98
7. Sistema Computacional (<i>Market Sensor</i>)	103
7.1. Introdução	103
7.2. Aspectos Tecnológicos	104
7.3. Aspectos Funcionais	106
8. Comentários finais e conclusões	109
8.1. Conclusão	109
8.2. Sugestão de Trabalhos Futuros	111
9. Referências Bibliográficas	113
10. Apêndice	116
10.1. Apêndice I	116
10.2. Apêndice II	119
10.3. Apêndice III	122
10.4. Apêndice IV	124
10.5. Apêndice V	125

Lista de Ilustrações

Figura 4.1. Série de duração – IBM	46
Figura 4.2. ACF e PACF (duração e duração com ajuste sazonal) – IBM	46
Figura 4.3. Resíduo modelo de duração – IBM	47
Figura 4.4. Prazo sazonal (duração) – IBM	47
Figura 4.5. Diferentes topologias AACD	64
Figura 5.1. Resposta da duração devido a impulso nas componentes	84
Figura 5.2. Resposta do volume devido a impulso nas componentes	85
Figura 5.3. Resposta do <i>spread</i> devido a impulso nas componentes	85
Figura 5.4. Resposta da volatilidade devido a impulso nas componentes	86
Figura 5.5. Padrão sazonal intradiário das componentes do sistema	89
Figura 5.6. ACF duração (resíduos x observações)	90
Figura 5.7. ACF volume (resíduos x observações)	91
Figura 5.8. ACF <i>spread</i> (resíduos x observações)	91
Figura 5.9. ACF volatilidade instantânea (resíduos x observações)	92
Figura 5.10. Previsto x realizado – variáveis financeiras de alta frequência	92
Figura 6.1. ACF da duração (resíduos x observações)	98
Figura 6.2. ACF do volume (resíduos x observações)	99
Figura 6.3. ACF do <i>spread</i> (resíduos x observações)	99
Figura 6.4. Número de acertos (direção e magnitude)	101
Figura 6.5. Número de acertos (direção)	101
Figura 6.6. Número total de acertos (direção)	102
Figura 7.1. Visão esquemática do funcionamento do sistema	104
Figure 7.2. Interface gráfica do sistema (<i>Market Sensor</i>)	105
Figura 1 (apêndice III): Análise descritiva dos dados de duração	122
Figura 2 (apêndice III): Análise descritiva dos dados de volume	122
Figura 3 (apêndice III): Análise descritiva dos dados de <i>spread</i>	123
Figura 4 (apêndice III): Análise descritiva dos dados de retornos	123

Lista de Tabelas

Tabela 1: Experimento de SMC (nível de significância: 90% bi-caudal)	83
Tabela 2: Resultados Ljung-Box – dependência linear	93
Tabela 3: Resultados do Teste de Razão de Verossimilhança	94
Tabela 4: Teste de Razão de Verossimilhança (relação de causalidade)	94
Tabela 5: Análise da dependência linear (Ljung-Box)	99
Tabela 6: Estimção (modelo completo)	119
Tabela 7: Estimção (modelo “livre de variação”)	120
Tabela 8: Estimção (modelo individual)	121
Tabela 9: Resultados da previsão de preços	124