

THIAGO KURIMORI GARCIA DA SILVEIRA

**MODELO DE PREVISÃO COM VOLATILIDADE
ESTOCÁSTICA**

Trabalho de Formatura apresentado à
Escola Politécnica da Universidade de
São Paulo para a obtenção do Diploma
de Engenheiro de Produção

SÃO PAULO
2008

THIAGO KURIMORI GARCIA DA SILVEIRA

**MODELO DE PREVISÃO COM VOLATILIDADE
ESTOCÁSTICA**

Trabalho de Formatura apresentado à
Escola Politécnica da Universidade de
São Paulo para a obtenção do Diploma
de Engenheiro de Produção

Orientadora:
Prof^a Dr^a Linda Lee Ho

SÃO PAULO
2008

FICHA CATALOGRÁFICA

Silveira, Thiago Kurimori Garcia da
Modelo de previsão com volatilidade estocástica, T. K. G.
Silveira – São Paulo, 2008.
107 p.

Trabalho de formatura – Escola Politécnica da
Universidade de São Paulo. Departamento de Engenharia de
Produção.

1.Modelos Não Lineares 2.Inferência Bayesiana 3.Métodos
MCMC I.Universidade de São Paulo. Escola Politécnica.
Departamento de Engenharia de Produção II.t.

DEDICATÓRIA

À minha família.

AGRADECIMENTOS

Aos meus pais Alberto e Lucia, meus irmãos Alberto e Luciana e minha namorada Helena pelo apoio, carinho e motivação.

À professora Linda Lee Ho, pelos conselhos, pela paciência e pela confiança depositada em mim.

RESUMO

A volatilidade é um tema de crescente importância no mercado financeiro, e diversos modelos têm sido criados e sugeridos para realizar a sua previsão. O presente trabalho apresenta a construção de um modelo de séries temporais para a previsão da volatilidade futura de uma série financeira. O modelo proposto incorpora componentes estocásticos tanto para a equação de retornos quanto para a equação das volatilidades condicionais, o que aumenta a complexidade do problema e impossibilita a estimação dos seus parâmetros via métodos de verossimilhança. Assim, são utilizadas técnicas de simulação MCMC, que surgem como alternativa para a resolução dos problemas de inferência Bayesiana em que não é possível derivar a função de verossimilhança analiticamente.

A capacidade preditiva do modelo foi comparada à de um modelo GARCH, que é amplamente utilizado para a previsão da volatilidade, e o desempenho dos modelos foram confrontados através de cálculos dos erros de previsão. O modelo de Volatilidade Estocástica proposto apresentou melhor desempenho na capacidade de previsão da volatilidade, mas também apresentou maior complexidade e maior demanda de capacidade de processamento computacional para seu ajustamento.

Palavras-chave: Modelos Não Lineares. Inferência Bayesiana. Métodos MCMC.

ABSTRACT

Volatility is an issue of growing importance in the financial market, and several models have been created and suggested to realize its forecast. This paper presents the construction of a time-series model to forecast the future volatility of a financial series. The proposed model incorporates stochastic components both to the equation of returns and for the equation of conditional volatilities, which increases the complexity of the problem and makes the estimation of its parameters impossible via the likelihood methods. Therefore, MCMC simulation techniques are used, which emerge as an alternative to solving the problems of Bayesian inference on which it is not possible to derive the likelihood function analytically.

The predictive ability of the model was compared to that of a GARCH model, which is widely used to predict the volatility, and the performance of the models were confronted by the calculation of forecasting errors. The proposed stochastic volatility model showed better performance in the prediction of volatility, but presented higher complexity and greater demand of computer processing capacity to realize its adjustment.

Keywords: Non-Linear Models. Bayesian inference. MCMC methods.

LISTA DE FIGURAS

Figura 2-1 - Volume global de contratos de futuros e de opções por categoria (Jan- Mai 2008).	28
Figura 2-2 – Resultado financeiro de uma opção de compra no vencimento.	31
Figura 3-1 - Trajetórias aleatórias de um processo estocástico.	41
Figura 3-2 – Fac de um ruído branco.	45
Figura 3-3 - Diagrama representativo de um filtro linear.	46
Figura 3-4 - Série de observações de um modelo AR(1).	48
Figura 3-5 - Fac amostral e teórica de um processo AR(1).	49
Figura 3-6 - Série de observações de um modelo MA(1).	50
Figura 3-7 - Fac amostral e teórica de um processo MA(1).	51
Figura 3-8 - Inversa FPA para uma distribuição normal padrão.	67
Figura 4-1 - Evolução das cotações diárias do IBOVESPA e os log retornos da série.	72
Figura 4-2 - Histograma dos log retornos do IBOVESPA.	73
Figura 4-3 - PPN dos log retornos do IBOVESPA.	73
Figura 4-4 - Fluxograma do processo de construção do modelo de volatilidade estocástica.	76
Figura 4-5 - Diagrama da simulação do Amostrador de Gibbs.	84
Figura 4-6 - Amostras realizadas dos parâmetros do modelo (parte 1/2).	86
Figura 4-7 - Amostras realizadas dos parâmetros do modelo (parte 2/2).	87
Figura 4-8 - Funções densidade de probabilidade das distribuições <i>a priori</i> e <i>a posteriori</i> dos parâmetros.	89
Figura 4-9 – Funções densidade de probabilidade dos parâmetros da volatilidade. ..	90
Figura 4-10 - Volatilidade condicional dos modelos ajustados.	91
Figura 4-11 - Resíduos dos modelos ajustados.	93
Figura 4-12 - Variância condicional prevista e os quadrados dos retornos observados.	96
Figura 4-13 - Resíduos EQMPV.	97

LISTA DE TABELAS

Tabela 2-1 - Direitos e obrigações de opções.....	30
Tabela 2-2 - Efeito do aumento das variáveis sobre o preço das opções.....	33
Tabela 4-1 - Estatísticas descritivas da série de log retornos do IBOVESPA.	74
Tabela 4-2 - Estimação do modelo GARCH(1,1) para a série de retornos do IBOVESPA.....	75
Tabela 4-3 - Teste de Ljung-Box paara os resíduos padronizados.....	92
Tabela 4-4 - Teste de Ljung-Box para os quadrados dos resíduos.....	92
Tabela 4-5 - Teste de multiplicadores de Lagrange.....	92
Tabela 4-6 - Quartis dos resíduos EMAP.	97
Tabela 4-7 - Comparação dos critérios de desempenho.	98

LISTA DE SIGLAS E ABREVIATURAS

AR	Autoregressivo
ARIMA	<i>Autoregressive Integrated Moving Average</i>
ARCH	<i>Autoregressive Conditional Heteroscedasticity</i>
BOVESPA	Bolsa de Valores de São Paulo
CDI	Certificado de Depósito Interfinanceiro
EQM	Erro Quadrado Médio
EQMPV	Erro Quadrado Médio de Previsão da Volatilidade
EMAP	Erro Médio Absoluto de Previsão
Fac	Função de autocorrelação
Facv	Função de autocovariância
FDP	Função de Densidade de Probabilidade
FPA	Função densidade de Probabilidade Acumulada
GARCH	<i>Generalized Autoregressive Conditional Heteroscedasticity</i>
i.i.d.	<i>Independent and identically distributed</i>
MA	<i>Moving Average</i>
MCMC	<i>Markov Chain Monte Carlo</i>
MVE	Modelo de Volatilidade Estocástica

SUMÁRIO

1. INTRODUÇÃO.....	21
1.1. O Estágio.....	22
1.2. Motivação	23
1.3. Organização do Trabalho.....	24
2. CONCEITOS	26
2.1. Ações	27
2.2. Derivativos.....	27
2.3. Opções	30
2.4. Modelo de Black & Scholes	33
2.5. Volatilidade.....	35
2.5.1 Volatilidade futura	36
2.5.2 Volatilidade histórica.....	36
2.5.3 Volatilidade implícita.....	37
3. FUNDAMENTAÇÃO TEÓRICA	39
3.1. Modelos de Séries Temporais	39
3.1.1 Processos estocásticos	40
3.1.2 Processos estacionários.....	42
3.1.3 Ruído branco.....	44
3.1.4 Função de autocovariância (facv)	44
3.1.5 Tipos de modelos.....	46
3.1.6 Modelos lineares.....	46
3.1.6.1 Modelos AR.....	48
3.1.6.2 Modelos MA	49
3.1.6.3 Modelos ARMA.....	51
3.1.6.4 Modelos ARIMA	52
3.1.7 Modelos não-lineares	52
3.1.7.1 Modelos ARCH.....	53
3.1.7.2 Modelos GARCH.....	54
3.1.7.3 Modelos de Volatilidade Estocástica.....	55

3.1.8	<i>Diagnósticos de Modelos</i>	55
3.1.8.1	Teste de Ljung-Box	55
3.1.8.2	Teste de multiplicadores de Lagrange.....	56
3.1.8.3	Análise da fac residual	57
3.2.	Inferência Bayesiana	58
3.2.1	<i>Distribuições a Priori Conjugadas</i>	59
3.2.2	<i>Simulação por Cadeias de Markov</i>	63
3.2.3	<i>Métodos baseados em simulação estocástica</i>	63
3.2.3.1	Amostrador de Gibbs	64
3.2.3.2	Griddy Gibbs	66
3.3.	O Modelo de Volatilidade Estocástica	69
4.	CONSTRUÇÃO DO MODELO	71
4.1.	Etapas da construção do modelo	75
4.1.1	<i>O Modelo Proposto</i>	77
4.1.2	<i>Escolha das distribuições a priori dos parâmetros</i>	78
4.1.3	<i>Estimação das posterioris condicionais</i>	79
4.1.4	<i>Simulação via Amostrador de Gibbs</i>	83
4.1.5	<i>Verificação da convergência do algoritmo</i>	85
4.1.6	<i>Estimativas dos parâmetros</i>	88
4.1.7	<i>Validação e comparação dos modelos ajustados</i>	90
5.	CONCLUSÕES	100
6.	REFERÊNCIAS BIBLIOGRÁFICAS	102
	ANEXO – CÓDIGO DO PROGRAMA ESCRITO EM MATLAB	104

1. INTRODUÇÃO

A volatilidade é um tema de crescente importância para todos os participantes econômicos que atuam direta ou indiretamente no mercado financeiro, pois é um dos principais critérios utilizados para análise de riscos, decisão de estratégias de *hedge*, oportunidades de especulação, otimização de *portfolios* de investimento, etc.

O mercado brasileiro de opções vem crescendo substancialmente nos últimos anos e, cada vez mais, novos participantes ingressam neste mercado. Os principais *players* são investidores que utilizam as opções como derivativos para se protegerem (os chamados *hedgers*); os especuladores, que tentam prever os movimentos de preços do mercado e lucrar com tais oscilações; e os arbitradores, que tentam aproveitar imperfeições de mercado para realizarem lucros com pouco ou sem riscos.

Uma das principais características de um mercado de opções é a quantidade de fatores que influem diretamente no apreamento desses derivativos. Uma dessas variáveis é a volatilidade, que representa a velocidade com que as cotações de um ativo variam com o tempo. O valor da volatilidade não pode ser determinado, pois ela não pode ser observada, e não há como se obter a real população dos retornos de um ativo qualquer. Assim, muitos esforços são realizados para obter a melhor estimativa possível do seu valor, e o sucesso financeiro de qualquer um dos *players* desse mercado está totalmente atrelado à eficiência na sua estimativa.

Atualmente, os modelos mais utilizados para a previsão da volatilidade são os modelos da família ARCH/GARCH, que utiliza a teoria de séries temporais para propor equações para a média e variância da variável de interesse, que usualmente é o retorno ou a variação de um instrumento financeiro. Assim, os modelos de volatilidade estocástica surgem como uma evolução direta aos modelos ARCH/GARCH, por permitirem maiores possibilidades de modelagem. O modelo de volatilidade estocástica foi apresentado por Jaquier, Polson e Rossi (1994), e não é tão popular provavelmente pela dificuldade de resolução do mesmo, que requer técnicas avançadas de simulação e demanda grande capacidade de processamento computacional.

O presente trabalho de formatura busca a resolução de um problema de previsão de uma variável de interesse: a volatilidade futura de um instrumento

financeiro. O objetivo é construir um modelo para o comportamento temporal de um instrumento financeiro, que permita inferir e prever os valores reais da volatilidade do mesmo. Assim, utiliza-se a inferência Bayesiana para a estimação dos parâmetros que, dado a impossibilidade de se obter analiticamente as distribuições necessárias no problema proposto, requer métodos de simulação de Cadeias de Markov (MCMC) para a obtenção de tais distribuições de probabilidade *a posteriori* dos parâmetros.

1.1. O Estágio

O desenvolvimento deste trabalho de formatura foi realizado na empresa Capitânia S.A., na área de gestão de ativos (*asset management*) através de fundos de investimentos multimercado, operando em ativos de renda variável na Bolsa de Valores de São Paulo (BOVESPA).

A Capitânia S.A. foi fundada em julho de 2003 pelas equipes de *trading*, estruturação e vendas de produtos da área de tesouraria do *Bank of America* no Brasil. Atualmente, a Capitânia S.A. é gestora de quatro fundos de investimento, sendo três da categoria multimercado e um fundo exclusivo de ações. Através destes quatro produtos, a empresa procura oferecer fundos de investimento com perfis diferenciados, utilizando em sua essência estratégias de renda variável, com o objetivo de gerar retornos consistentes, independentemente do cenário macroeconômico.

O estágio teve como principal objetivo a análise quantitativa de instrumentos financeiros, principalmente voltada a ações e opções de ações. Uma das principais estratégias de investimento da empresa é gerar ganhos financeiros através da especulação da volatilidade futura de uma ação. Assim, muitas pesquisas nesse sentido são realizadas, que têm em sua essência aplicações da teoria estatística para a realização de uma série de análises quantitativas. É nesse contexto que surge a proposta de realização do presente trabalho.

1.2. Motivação

Atualmente, há diversos modelos de previsão que incluem a importância da volatilidade nos seus parâmetros. Alguns deles consideram a volatilidade como constante ao longo do tempo (modelos ARIMA), enquanto outros a consideram como uma variável correlacionada a seus valores passados (modelos ARCH), e também a valores de outras variáveis explicativas (modelos GARCH). Tais classes de modelos buscam explicar os valores de uma variável em um dado instante a partir dos valores passados dessa mesma variável, ou seja, admite-se que os valores são autocorrelacionados. Caso exista essa explicação de valores futuros através de seus antecedentes, pode-se dizer que os valores da série são autocorrelacionados. A ideia principal da volatilidade estocástica é incrementar os modelos ARCH/GARCH para que este possa considerar a própria volatilidade como dependente não só de seus valores passados, mas também da dinâmica aleatória intrínseca do mercado financeiro.

Há alguns estudos que mostram a influência do volume de operações do mercado na volatilidade. O caótico fluxo de informações que atinge os diversos participantes do mercado produz um volume de operações aleatório. Tal comportamento deveria, se possível, ser incorporado a um modelo eficiente de previsão. Tauchen; Pitts (1983) e Gallant; Hsieh; Tauchen (1991) notaram que, se os fluxos de informações são autocorrelacionados, então um modelo que considera a variância condicional como variável no tempo e correlacionada com seus valores passados pode ser apropriado para estudar as séries financeiras. Com essa visão, Jaquier; Polson e Rossi (1994) propuseram uma metodologia que utiliza a análise Bayesiana para a construção do modelo de volatilidade estocástica. Tal análise busca a estimação dos reais parâmetros de um modelo a partir das informações que se dispõe para explicá-lo. Assim, dado um conjunto de informações disponíveis sobre uma série, a metodologia Bayesiana permite inferir sobre os reais valores dos parâmetros geradores dessa série analisada.

Contudo, para a incorporação do caráter aleatório da volatilidade, o modelo torna-se de difícil resolução pelos métodos clássicos – ou seja, que utilizam funções de verossimilhança. Tais funções podem ser obtidas analiticamente para os modelos

ARIMA, ARCH e GARCH, porém, para os modelos de volatilidade estocástica, esta obtenção não é possível. Por isso, como alternativa, são utilizadas técnicas Markov Chain Monte Carlo (MCMC) que, realizando simulações, constrói-se uma cadeia de Markov aproximadamente equivalente à distribuição conjunta *a posteriori* necessária.

Levando esses fatos em consideração, o objetivo deste trabalho é construir um modelo quantitativo que utilize um conjunto de informações disponíveis para a previsão da volatilidade, juntamente com a consideração de influências estocásticas no seu comportamento. Apesar de exigir técnicas de resolução mais trabalhosas e sofisticadas, o modelo será de grande valia caso tenha um desempenho marginalmente melhor que os modelos mais populares, visto que qualquer vantagem competitiva no mercado financeiro pode representar enormes benefícios para a empresa. O resultado poderá ser usado como auxílio direto no apreçamento de uma opção, e conseqüentemente servirá de apoio nas decisões de investimento do operador de opções.

1.3. Organização do Trabalho

O trabalho é organizado em seis partes básicas. Nesta seção, apresenta-se a forma como está organizado e descreve-se sucintamente cada um dos capítulos que o compõem:

- Capítulo 1 – São apresentados o problema a ser resolvido e os objetivos do trabalho, assim como uma breve introdução ao tema proposto. Apresenta-se também a descrição do estágio e da motivação para o projeto.
- Capítulo 2 – Conduz a parte conceitual necessária para a melhor compreensão e contextualização do tema proposto. São apresentados conceitos básicos do mercado financeiro, entre eles os instrumentos existentes, os principais participantes do mercado de capitais e um panorama da teoria de apreçamento de opções de Black & Scholes. Além disso, trata da variável objeto motivadora do presente trabalho, onde é apresentado o conceito de volatilidade no mercado financeiro: sua importância, suas

características, classificações, aplicações no mercado e peculiaridades no apreamento de opções.

- Capítulo 3 – Apresenta as fundamentações teóricas necessárias para o desenvolvimento do projeto. A teoria pesquisada para a resolução do problema envolve modelos de séries temporais, inferência Bayesiana e métodos de simulação MCMC.
- Capítulo 4 – Neste, detalha-se a construção do modelo de volatilidade estocástica, aplicado em uma série de dados históricos do IBOVESPA. O objetivo é propor um modelo que represente satisfatoriamente o comportamento dos retornos da série analisada e, principalmente, da sua volatilidade condicional. Assim, são aplicados métodos MCMC para que os parâmetros do modelo possam ser estimados via inferência Bayesiana. A resolução do modelo será dividida em: proposta do modelo, apresentação das distribuições *a priori* dos parâmetros, obtenção das distribuições *a posteriori* condicionais dos parâmetros, descrição dos algoritmos MCMC e resolução numérica do modelo. Este capítulo também apresenta a validação do modelo estatístico apresentado, verificando a coerência dos resultados obtidos e comparando seu desempenho com um modelo GARCH.
- Capítulo 5 – Conclui-se o trabalho, destacando os pontos mais relevantes do estudo realizado, bem como se analisa criticamente o projeto desenvolvido.

2. CONCEITOS

A definição de investir, segundo Bodie (2000), é comprometer o uso de recursos no presente com a expectativa de benefícios futuros. Atualmente, há inúmeros produtos financeiros que servem de meios para que essa alocação de recursos ocorra. Tais instrumentos podem ser classificados quanto à sua lucratividade, que pode ser fixa ou variável.

Ativos de renda fixa são investimentos que pagam, em períodos definidos, uma certa remuneração ao investidor. A rentabilidade pode ser determinada no momento da aplicação (títulos pré-fixados) ou no momento do resgate (títulos pós-fixados). Os títulos de renda fixa podem ser públicos ou privados, de acordo com a instituição encarregada de honrar o pagamento do investimento e a remuneração. Entre os ativos de renda fixa podem ser citados: a caderneta de poupança, o CDB (Certificado de Depósito Bancário) e debêntures.

Já os ativos de renda variável são títulos de propriedade que podem apresentar componentes de remuneração periódica, determinados contratualmente. O lucro nesse caso é apurado pela diferença entre o preço de compra mais os benefícios, e o preço de venda. Ações, cotas de fundos de investimento, moedas e commodities são exemplos dos ativos de renda variável mais conhecidos.

Os instrumentos financeiros mais importantes na realização e aplicação do presente trabalho são as ações, opções (uma classe de derivativos) e o CDI (Certificado de Depósito Interfinanceiro). A importância de tais ativos decorre da volatilidade, que interliga estes instrumentos. Em suma, a volatilidade de um ativo pode ser especulada através de opções que tenham este ativo como objeto, e o CDI serve como a taxa de retorno livre de risco, que corresponde ao custo de oportunidade de um investimento, e que é uma das entradas no apreamento de tais opções.

2.1. Ações

Segundo a Bolsa de Valores de São Paulo (BOVESPA), ações são títulos nominativos negociáveis que representam, para quem as possui, uma fração do capital social de uma empresa. Tais títulos conferem ao seu proprietário parcelas de participação no controle, nos bens e nos lucros da empresa. Existem basicamente dois tipos de ações:

- Ações ordinárias (ON): concedem, aos seus proprietários, poder de voto nas assembleias deliberativas da companhia.
- Ações preferenciais (PN): oferecem preferência na distribuição de resultados ou no reembolso do capital em caso de liquidação da companhia, mas não concedem poder de voto aos acionistas.

Uma empresa deve dividir os lucros com seus acionistas. Essa parcela direcionada aos detentores de ações é denominada de dividendo, e é distribuída na proporção da quantidade de ações que cada acionista possui, e apurada ao fim de cada exercício social. Assim, um dos principais motivos que atraem investidores a comprar ações é a expectativa de receber dividendos das empresas emissoras desses títulos. Em muitos casos, entretanto, é de interesse do proprietário de ações a busca por outros investimentos que resultem, por exemplo, em uma maior proteção aos seus investimentos, uma maior alavancagem do capital investido, dentre outros. Nestas e outras situações similares, os investidores podem utilizar os chamados derivativos.

2.2. Derivativos

Um derivativo é um título cujo valor depende dos valores de outras variáveis básicas que o referenciam. Um derivativo sobre uma ação, por exemplo, tem o seu valor atrelado à cotação desta ação. Os mercados de derivativos tiveram um ritmo de crescimento bastante acelerado nos últimos anos em todo o mundo. Além de instituições financeiras como bancos e investidores, o uso de derivativos foi

difundido para empresas de outros segmentos da economia (agrícola, energia, entre diversos outros).

As finalidades do investimento em derivativos são muitas, entre elas a proteção contra oscilações de mercado (*hedge*), ganhos financeiros através da especulação, transformação de um passivo de taxa flutuante por um de taxa fixa, entre outros. Existe uma variedade muito grande de contratos negociados, e a Figura 2-1 mostra os contratos de futuros e de opções mais negociados no período de janeiro a maio de 2008, classificados pelo tipo de ativo objeto do derivativo:

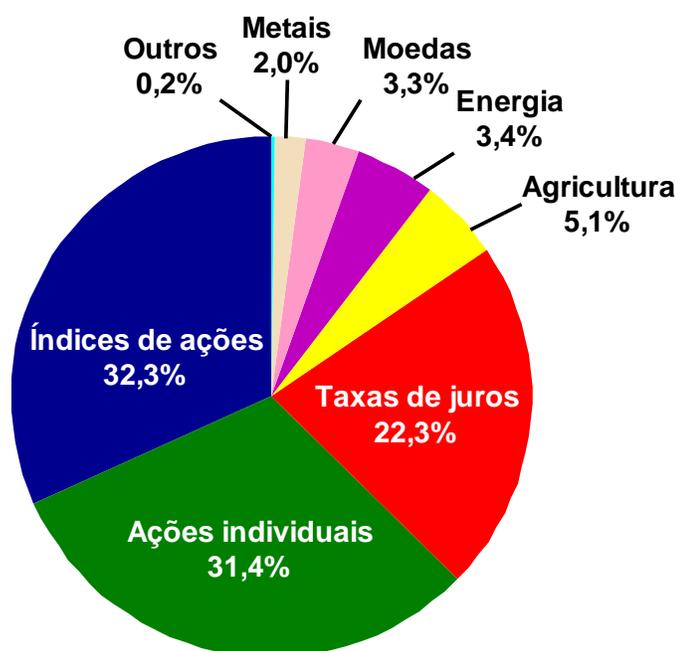


Figura 2-1 - Volume global de contratos de futuros e de opções por categoria (Jan-Mai 2008).
Fonte: FIA

O total de contratos negociados nesse período foi de aproximadamente 7 bilhões, considerando o mercado de balcão (isto é, negociações que não passam pelas Bolsas) e Bolsas de Valores de todo o mundo. A categoria “outros” envolve contratos baseados em carbono, índices de *commodities*, crédito, fertilizantes, construção civil, inflação, lenha e condições climáticas.

Os tipos de derivativos mais negociados no mercado brasileiro atualmente são:

- **Contratos a termo:** são acordos de compra e venda de um determinado ativo para liquidação em uma data futura, estabelecida por um preço entre as partes, que é denominado de *preço de entrega*. Sua negociação ocorre principalmente em mercado de balcão.
- **Contratos futuros:** assim como os contratos a termo, as partes envolvidas são obrigadas a comprar ou vender, a um preço estabelecido por elas, determinada quantidade de um ativo, em certa data futura. A diferença é que os contratos são negociados em Bolsa e assim fica a cargo da Bolsa cumprir o contrato. Esses contratos são padronizados em relação à quantidade e qualidade do ativo, formas de liquidação, garantias, prazos de entrega, dentre outros. Além disso, é possível a liquidação do contrato antes do prazo de vencimento.
- **Swaps:** são acordos privados entre duas instituições para a troca futura de fluxos de caixa, realizada de acordo com uma fórmula preestabelecida. Assim, os pagamentos futuros dependem de seus respectivos indexadores, que podem ser flutuantes ou não. Como ambos os pagamentos são efetuados em uma mesma data, a liquidação do contrato geralmente ocorre pela diferença entre os valores contratuais corrigidos pelos seus respectivos indexadores. São negociados principalmente em mercado de balcão.
- **Opções:** são contratos que dão à parte compradora o direito (e não a obrigação) de comprar ou vender o ativo objeto em certa data a um preço preestabelecido. As opções são negociadas tanto em mercado de balcão quanto na Bolsa. Este tipo de derivativo, dada a sua relevância para o problema proposto, será apresentado com mais detalhes na Seção 2.3.

2.3. Opções

Segundo Costa (1998), opção é um “direito negociável de compra ou venda de um ativo a um preço futuro predeterminado”. O titular, agente econômico que compra a opção a um determinado preço na data inicial, adquire a escolha de exercer ou não o seu direito de compra ou venda. A contraparte, denominado de lançador da opção, adquire o dever de aceitar a decisão do titular, e assim torna-se obrigado a comprar ou vender o ativo objeto no preço determinado para o titular, em caso do exercício de direito.

Há inúmeros tipos de contratos de opções que apresentam uma série de características peculiares. As opções mais negociadas são as chamadas opções “*vanilla*”, que serão o único tipo de opção considerado no decorrer deste trabalho. Para maiores informações sobre os outros tipos de opções existentes (asiáticas, flexíveis, exóticas, etc) ver Hull (1997) e Costa (1998).

Há dois tipos de direitos que uma opção apresenta: direito de compra (*call*) e direito de venda (*put*), cujos direitos e obrigações são apresentados na Tabela 2-1:

Parte	Opção de compra (Call)	Opção de venda (Put)
Lançador	Obrigação de compra	Obrigação de venda
Titular	Direito de compra	Direito de venda

Tabela 2-1 - Direitos e obrigações de opções.

Além disso, as opções podem ser classificadas de acordo com a possibilidade de exercício antes ou somente na data de exercício preestabelecida. São ditas *opções americanas* aquelas que podem ser exercidas a qualquer instante até a data de vencimento, enquanto as *opções européias* podem ser exercidas somente na data de vencimento. No Brasil, a maioria das opções negociadas é do tipo europeu.

O resultado financeiro de uma operação de compra de uma opção *call* européia na data de vencimento é apresentado na Figura 2-2:

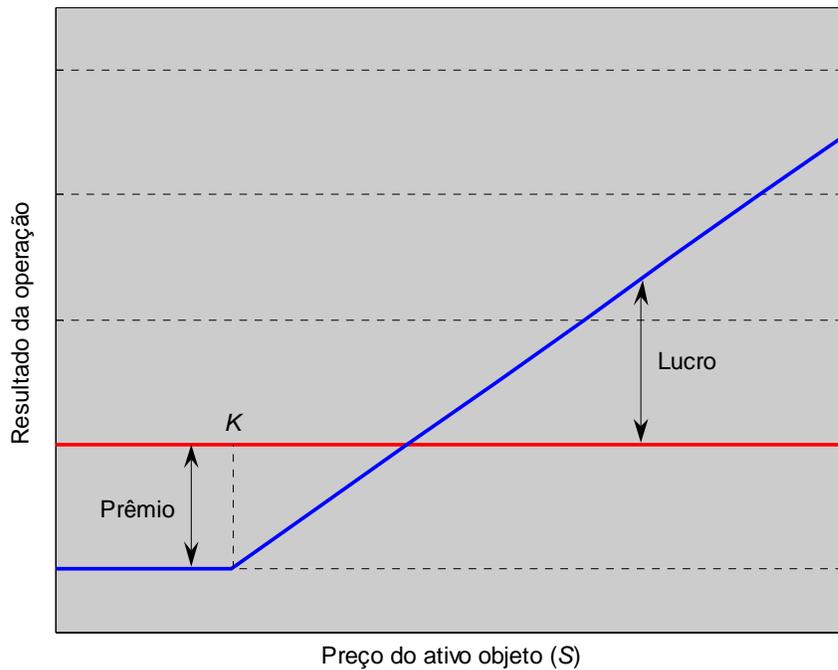


Figura 2-2 – Resultado financeiro de uma opção de compra no vencimento.

Na Figura 2-2, K corresponde ao preço de exercício da opção de compra e o prêmio é o preço pago pela opção na data negociada. Pode-se notar que se o preço do ativo objeto for maior que o preço de exercício, é compensador para o titular exercer o direito de compra, pois deste modo estaria comprando o ativo objeto por um preço menor do que o preço negociado nesta data. O valor recebido pelo titular no vencimento da opção é denominado *payoff*, e pode ser descrito como:

$$\max(0, S - K), \text{ para uma opção de compra}$$

$$\max(0, K - S), \text{ para uma opção de venda.}$$

onde S é a cotação à vista do ativo objeto (*spot*) e K é o preço de exercício da opção (*strike*). O *payoff* também é denominado de valor intrínseco da opção.

As opções podem ser negociadas em todo o período desde o seu lançamento até a data de vencimento, e neste período o seu valor não pode mais ser definido

apenas pelo valor intrínseco. Segundo Hull (1997) existem basicamente cinco variáveis que influem no valor de uma opção:

S – preço à vista do ativo objeto;

K – preço de exercício da opção;

r – taxa de juros livre de risco;

t – tempo até o vencimento;

σ – volatilidade do ativo objeto.

O preço à vista do ativo objeto e o preço de exercício podem ser analisados juntos através do valor intrínseco da opção. No caso de uma opção de compra, esta terá valor maior quanto maior for o preço do ativo e quanto menor for o seu *strike*. Já uma opção de venda terá valor maior quanto menor for o preço do ativo e maior for o seu *strike*.

A volatilidade de um ativo pode ser entendida como uma medida de incerteza sobre os seus preços futuros. Quanto maior a volatilidade, maior é a possibilidade de haver oscilações bruscas de preços, tanto para cima quanto para baixo. Assim, para uma opção, tanto de compra quanto de venda, a volatilidade aumenta a possibilidade de um resultado positivo, dado que as perdas são limitadas.

O tempo até o vencimento de uma opção é medido em dias úteis e influi no preço de uma opção, pois quanto maior o prazo até o vencimento, maiores as possibilidades para a trajetória de preços do ativo objeto.

A taxa de juros livre de risco afeta o preço das opções na medida em que quanto maior o seu valor, maior será a expectativa de retorno dos preços dos ativos. Além disso, por tratar de fluxos de caixas futuros, o valor futuro dos desembolsos e receitas diminuirão. Na prática, a taxa de juros livre de risco utilizada no Brasil é o CDI.

A Tabela 2-2 apresenta um panorama dos efeitos do aumento de cada variável sobre o preço da opção, enquanto as demais se mantêm constantes:

Variável	Opção de compra (Call)	Opção de venda (Put)
Preço do ativo	+	-
Preço de exercício	-	+
Prazo até o vencimento	+	+
Volatilidade	+	+
Taxa livre de risco	+	-

Tabela 2-2 - Efeito do aumento das variáveis sobre o preço das opções.

Fonte: Adaptado de Hull (1997)

A teoria de apreçamento de opções é extensa e bastante desenvolvida, no entanto o estudo publicado mais importante sobre o assunto é a derivação da equação de Black & Scholes (1973).

2.4. Modelo de Black & Scholes

A difusão e popularidade das opções no mercado financeiro devem-se, em grande parte, ao modelo proposto por Black e Scholes (1973) para o apreçamento de opções. Tal método foi apresentado na década de 70 e é amplamente utilizado no mercado financeiro, devido a sua fácil implementação, ao seu rigor teórico e à vasta aplicação prática no mercado de capitais. O modelo proposto por Black e Scholes apresenta uma fórmula cujo resultado é o valor justo de uma opção, respeitando determinadas hipóteses.

Dada a relevância prática ao presente trabalho, o modelo de Black e Scholes será apresentado sucintamente para que fique clara a principal implementação prática deste trabalho. Para maiores detalhes sobre o modelo de Black e Scholes, ver Hull (1997).

Considerando:

C : Preço da opção de compra

P : Preço da opção de venda

S : Preço do ativo objeto no instante t

K : Preço de exercício

r : Taxa de juros livre de risco, composta continuamente

T : Data de vencimento

$T - t$: Número de dias úteis até o vencimento

σ : Volatilidade do ativo objeto

$\Phi(d)$: Função de distribuição de probabilidade normal padrão acumulada

Black e Scholes (1973) desenvolveram um modelo que determina o preço de uma opção de compra do tipo européia, por meio da expressão:

$$C = S\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2) \quad (\text{Eq. 2.1})$$

Para uma opção de venda, a expressão torna-se:

$$P = Ke^{-r(T-t)}\Phi(-d_2) - S\Phi(-d_1) \quad (\text{Eq. 2.2})$$

onde:

$$d_1 = \frac{\ln(S/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}} \quad (\text{Eq. 2.3})$$

$$d_2 = \frac{\ln(S/K) + (r - \sigma^2/2)T}{\sigma\sqrt{T}} = d_1 - \sigma\sqrt{T}. \quad (\text{Eq. 2.4})$$

É importante notar que, dentre todas as variáveis utilizadas no apreamento de uma opção, a única que não é conhecida é a volatilidade do ativo objeto. Assim, este parâmetro se torna o mais relevante na negociação de opções, e muitos esforços são realizados a fim de estimá-lo do modo mais eficaz possível. Na prática, há muitos investidores que realizam as chamadas operações de volatilidade, que

consiste na especulação e arbitragem não do preço futuro de um ativo, mas da sua volatilidade futura. Desse modo, o investidor que conseguir estimar melhor o valor justo da volatilidade poderá operá-la de modo a obter lucros financeiros acima do seu custo de oportunidade.

Juntamente com a derivação do preço de uma opção, Black e Scholes (1973) apresentaram uma série de variáveis que permitem a realização de uma análise de sensibilidade do preço da opção em função de suas variáveis, que no jargão do mercado financeiro recebem o nome de *gargas*. Elas são extremamente úteis para analisar uma carteira de investimentos que contenha posições em opções. Cada uma mede uma dimensão do risco da posição em opção. Por não serem cruciais para o desenvolvimento deste trabalho, estas não serão detalhadas. Como referência, ver Hull (1997) e Costa (1998).

2.5. Volatilidade

O mercado financeiro é caracterizado pelo dinamismo dos preços dos ativos negociados nas Bolsas de Valores. Tais mudanças podem ser mais ou menos drásticas. A cotação da ação de uma empresa de tecnologia, por exemplo, exhibe um comportamento muito mais dinâmico do que a taxa de juros livre de risco. Nesse sentido, a volatilidade surge como uma medida da velocidade, da incerteza das movimentações de preços. Mercados que se movem lentamente são ditos mercados de baixa volatilidade. Em contrapartida, mercados altamente dinâmicos são ditos mercados de alta volatilidade.

Mais especificamente, a volatilidade pode ser entendida como uma medida de dispersão do preço ao redor de seu valor esperado em um determinado período de tempo. Assim sendo, é de extrema importância que os participantes do mercado consigam estimar a volatilidade para diferentes prazos – sejam eles diários, semanais, anuais, ou outros – para que o apreçamento de uma opção possa ser realizado de maneira acurada. A volatilidade, por não poder ser observada, deve ser devidamente medida e estimada para que o investidor possa tomar posições coerentes.

A volatilidade pode ser apresentada de diversas maneiras, e é importante saber distinguir qual é a abordagem em questão quando se trata deste tema. A seguir, serão descritos os diferentes tratamentos existentes em relação à volatilidade.

2.5.1 Volatilidade futura

A volatilidade futura é o que qualquer investidor gostaria de saber, pois corresponde à volatilidade que melhor descreve a distribuição futura dos preços de um determinado instrumento. Caso pudesse ser determinada, o investidor conheceria a real função de distribuição de probabilidade dos preços.

2.5.2 Volatilidade histórica

Uma das maneiras de se estimar a volatilidade real de um ativo é analisar a volatilidade histórica do mesmo. A volatilidade histórica, também chamada de volatilidade realizada, pode ser estimada para quaisquer intervalos de tempo, desde que se consiga obter os respectivos dados necessários. Usualmente, calcula-se a volatilidade histórica considerando observações diárias do preço de fechamento do ativo, calculando-se os retornos como a variação de tais observações.

A seguir, apresenta-se a definição da volatilidade histórica segundo Hull (1997). Sejam:

$n + 1$: Número de observações;

S_i : Preço do ativo ao final do i -ésimo intervalo ($i = 0, 1, \dots, n$);

u_i : Retorno continuamente capitalizado (não-anualizado), no i -ésimo intervalo;

τ : Intervalo de tempo em anos.

Pela definição de retorno continuamente capitalizado, temos:

$$S_i = S_{i-1} e^{u_i} \quad (\text{Eq. 2.5})$$

$$u_i = \ln \left(\frac{S_i}{S_{i-1}} \right) \quad (\text{Eq. 2.6})$$

A estimativa usual, s , do desvio padrão dos valores de u_i é dada por:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (u_i - \bar{u})^2} \quad (\text{Eq. 2.7})$$

ou

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n u_i^2 - \frac{1}{n(n-1)} \left(\sum_{i=1}^n u_i \right)^2} \quad (\text{Eq. 2.8})$$

onde \bar{u} é a média dos retornos u_i .

Desta maneira, utilizando a estimativa para o desvio padrão através da Equação 2.8, é possível estimar a volatilidade histórica para o período τ , designado por $\hat{\sigma}$:

$$\hat{\sigma} = \frac{s}{\sqrt{\tau}} \quad (\text{Eq. 2.9})$$

Deve-se frisar que a estimativa da volatilidade para o período considerado refere-se a dados passados, e que usar tal estimativa para apreçar opções equivale a presumir que o passado irá se repetir no futuro. Tal suposição requer cautela, pois o mercado financeiro é altamente complexo e dinâmico. Todavia, a volatilidade histórica é uma ferramenta bastante útil para que o comportamento do preço do ativo seja estudado, principalmente em horizontes de longo prazo.

2.5.3 Volatilidade implícita

O único parâmetro da equação de Black & Scholes que não pode ser diretamente observado é a volatilidade do preço do ativo. Por isso, a volatilidade é o principal fator de incerteza no preço de uma opção. Na prática, os operadores usualmente trabalham com a chamada volatilidade implícita, ou seja, a volatilidade associada ao preço da opção negociada no mercado.

A volatilidade implícita é usada para monitorar a opinião do mercado acerca da volatilidade futura de determinado ativo. Na prática é comum observar valores diferentes da volatilidade implícita de opções de um mesmo ativo objeto. Tal fato decorre de diversos aspectos, entre eles o prazo de vencimento e o preço de exercício da opção. Assim, não é usual negociar opções pelo seu preço, mas sim em termos de

sua volatilidade implícita. Um investidor em opções geralmente realiza suas decisões em relação ao valor da volatilidade implícita que este está disposto a receber ou pagar por uma opção, pois seu preço, por si só, não “carrega” as informações mais relevantes para as decisões de investimento. Neste sentido, o investidor especula a volatilidade implícita da opção.

3. FUNDAMENTAÇÃO TEÓRICA

O modelo proposto para a resolução do problema requer conhecimentos teóricos sobre estatística, mais especificamente a inferência Bayesiana, que não são abordados no curso de Engenharia de Produção. Porém, os conceitos de estatística desenvolvidos no curso apresentam uma base necessária para que o aluno tenha a competência necessária para se aprofundar em áreas de maior interesse. Assim, intensivas pesquisas foram realizadas sobre o conceito de séries temporais, volatilidade estocástica, teoria de Bayes e simulações MCMC. A argumentação teórica utilizada na resolução do problema é trabalhada nesta seção, e um resumo dos tópicos que serão abordados é apresentado a seguir:

- Modelos de séries temporais: contempla a teoria de séries temporais, apresentando os conceitos de processos estacionários, ruído branco, função de autocovariância, além de apresentar alguns dos principais modelos existentes e testes estatísticos de validação dos mesmos.
- Inferência Bayesiana: apresenta a teoria de inferência Bayesiana, juntamente com conceitos de distribuições conjugadas, simulações MCMC e os principais algoritmos desta categoria.

3.1. Modelos de Séries Temporais

Segundo Morettin; Tolo (2006), uma série temporal é qualquer conjunto de observações ordenadas no tempo. O objetivo da análise de séries temporais é construir modelos para as séries, com propósitos determinados, como:

- Investigar o mecanismo gerador da série temporal;
- Fazer previsões de valores futuros da série;
- Descrever apenas o comportamento da série, a partir da construção de gráfico, verificação da existência de tendências, ciclos, variações sazonais;
- Procurar periodicidades relevantes nos dados.

Em todos os casos, modelos probabilísticos são construídos. Caso seja possível, os modelos devem ser simples e parcimoniosos, ou seja, com um número enxuto de parâmetros envolvidos.

Os modelos utilizados para descrever séries temporais são processos estocásticos, isto é, processos descritos por leis probabilísticas.

Para descrever o comportamento de uma série, é possível considerar um número muito grande de modelos. A construção de um modelo adequado depende de vários fatores, tais como o comportamento do fenômeno ou o conhecimento prévio que temos de sua natureza e do objetivo da análise. Na prática, outras questões são relevantes, como a existência de métodos apropriados de estimação, disponibilidade de *softwares* adequados e capacidade de processamento de dados.

3.1.1 Processos estocásticos

Seja T um conjunto arbitrário representante da dimensão temporal. Um processo estocástico é uma família $Z = \{Z(t), t \in T\}$, tal que, para cada $t \in T$, $Z(t)$ é uma variável aleatória.

Nestas condições, um processo estocástico é a família de variáveis aleatórias, que se supõe definidas num mesmo espaço de probabilidades. O conjunto T é normalmente representado pelo tempo, e tomado como o conjunto dos inteiros $Z = \{0, \pm 1, \pm 2, \dots\}$ ou o conjunto dos reais \mathfrak{R} . Também, para cada $t \in T$, $Z(t)$ será uma variável aleatória real, definida sobre Ω , que é a dimensão dos parâmetros geradores do processo.

Logo, na realidade $Z(t)$ é uma função de dois argumentos, $Z(t, \omega)$, $t \in T$, $\omega \in \Omega$. A Figura 3-1 ilustra esta interpretação de um processo estocástico:

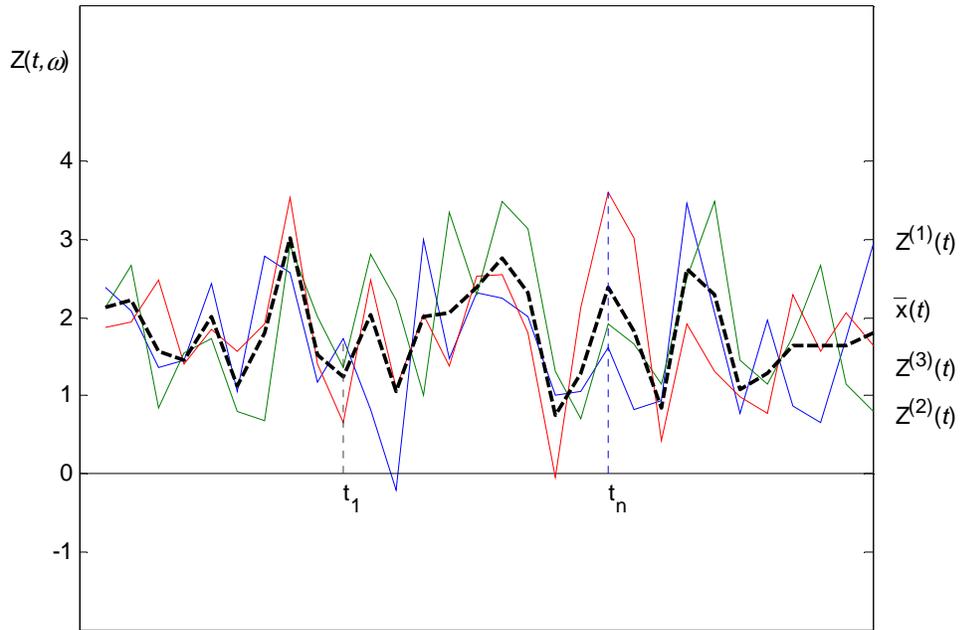


Figura 3-1 - Trajetórias aleatórias de um processo estocástico.

Para cada $\omega \in \Omega$ fixado, obtém-se uma função de t , ou seja, uma realização ou trajetória do processo, ou ainda, uma série temporal. No exemplo ilustrado, as trajetórias de $Z(t)$ foram geradas a partir de uma distribuição de probabilidade normal com parâmetro $\omega = (\mu, \sigma)$, para valores $\mu = 2$ e $\sigma = 1$. Portanto, nesse caso o parâmetro ω consiste em um vetor com os valores para a média e desvio padrão da série. A trajetória de $\bar{x}(t)$ denota a média amostral para as trajetórias de $Z(t)$ em cada instante t , isto é:

$$\bar{x}(t) = \frac{Z^{(1)}(t) + Z^{(2)}(t) + Z^{(3)}(t)}{3}. \quad (\text{Eq. 3.1})$$

O objetivo da inferência estatística trata-se de, a partir das observações disponíveis das trajetórias de uma variável, inferir os valores dos parâmetros geradores do processo. No exemplo descrito, o objetivo seria encontrar os valores de $\omega = (\mu, \sigma)$, isto é, a média e o desvio-padrão incondicional da distribuição normal geradora da série temporal.

A teoria de inferência será mais bem detalhada posteriormente, na Seção 3.2, mas a diferença fundamental entre a inferência estatística e a Bayesiana é que a última utiliza as chamadas distribuições de probabilidade *a priori*. Em suma, tais distribuições buscam implementar informações subjetivas sobre o processo em questão, através de funções de probabilidades. A inferência Bayesiana utiliza as distribuições *a priori*, e pondera tais informações com os dados observados amostralmente, na busca dos valores “justos” dos parâmetros.

Outra consideração importante é que na prática não se conhece, na maioria dos casos, a verdadeira função de densidade de probabilidade (fdp) da variável de interesse. Assim, primeiramente um modelo adequado deve ser proposto para que as séries sejam geradas a partir de equações especificadas. Há vários modelos de séries temporais existentes, e os principais serão apresentados na Seção 3.1.4.

3.1.2 Processos estacionários

A utilização de modelos de séries temporais frequentemente requer a introdução de suposições simplificadoras. Assim, algumas classes de processos estocásticos com podem ser definidas de acordo com suas características, segundo Morettin; Tolo (2006):

- a) Processos estacionários ou não-estacionários, de acordo com a independência ou não relativamente à origem dos tempos;
- b) Processos normais (Gaussianos) ou não-normais, de acordo com as funções de densidade de probabilidade que os caracterizam;
- c) Processos Markovianos ou não-Markovianos, de acordo com a independência dos valores do processo, em dado instante, de seus valores em instantes precedentes, ou seja, um processo Markoviano representa um processo no qual a série temporal não é autocorrelacionada. O conceito de autocorrelação será apresentado posteriormente.

“Intuitivamente, um processo Z é estacionário se ele se desenvolve no tempo de modo que a escolha de uma origem dos tempos não é importante. Em outras palavras, as características de $Z(t + \tau)$, para todo τ , são as mesmas de $Z(t)$. As

medidas das vibrações de um avião em regime estável de vôo horizontal, durante seu cruzeiro, constituem um exemplo de um processo estacionário” (MORETTIN; TOLOI, 2006, p.23).

Mais especificamente, uma série estacionária é aquela que se desenvolve no tempo aleatoriamente ao redor de uma média constante, refletindo uma espécie de equilíbrio estável. Na prática, a maioria das séries apresenta algum tipo de não-estacionariedade. As séries econômicas e financeiras, por exemplo, em geral apresentam tendências. Outros casos, como o crescimento populacional, apresentem uma forma de não-estacionariedade explosiva, e são mais difíceis de modelar.

Como a maioria dos procedimentos de análise estatística supõe a condição de estacionariedade, muitas vezes é necessário realizar algum tipo de transformação dos dados originais da série. O método mais comum consiste em tomar diferenças sucessivas da série original, até se obter uma série estacionária. Em muitos casos práticos, a tomada de diferenças de primeira e segunda ordem são bastante eficazes em tornar a série estacionária, o que facilita a utilização de modelos. A primeira diferença de $Z(t)$ é definida por:

$$\Delta Z(t) = Z(t) - Z(t-1) \quad (\text{Eq. 3.2})$$

De modo geral, a n -ésima diferença de $Z(t)$ é

$$\Delta^n Z(t) = \Delta[\Delta^{n-1} Z(t)]. \quad (\text{Eq. 3.3})$$

Um processo estacionário pode ser estritamente ou fracamente estacionário. As definições de tais processos são apresentadas a seguir:

Um processo estocástico $Z = \{Z(t), t \in T\}$ diz-se *estritamente estacionário* se todas as distribuições finito-dimensionais permanecem as mesmas sob translações no tempo, ou seja,

$$F(z_1, \dots, z_n; t_1 + \tau, \dots, t_n + \tau) = F(z_1, \dots, z_n; t_1, \dots, t_n), \quad (\text{Eq. 3.4})$$

para quaisquer t_1, \dots, t_n, τ de T .

Isto significa, em particular, que todas as distribuições unidimensionais são invariantes ao longo do tempo, logo a média e a variância da série são constantes, isto é:

$$\mu(t) = \mu, \quad V(t) = \sigma^2 \quad (\text{Eq. 3.5})$$

para todo $t \in T$.

3.1.3 Ruído branco

Ruído branco trata-se de uma sequência de valores aleatórios $\{\varepsilon_t\}$ que é amplamente utilizada nos modelos de séries temporais. Tais elementos são caracterizados por terem média zero e variância σ^2 ,

$$E(\varepsilon_t) = 0 \quad (\text{Eq. 3.6})$$

$$E(\varepsilon_t^2) = \sigma^2, \quad (\text{Eq. 3.7})$$

e terem todos os valores não-correlacionados no tempo:

$$E(\varepsilon_t \varepsilon_\tau) = 0 \text{ para } t \neq \tau. \quad (\text{Eq. 3.8})$$

Assim, um processo que satisfaz tais condições é chamado de *ruído branco*. Se além dessas características o processo seguir uma distribuição,

$$\varepsilon \sim N(0, \sigma^2), \quad (\text{Eq. 3.9})$$

então o processo é dito um *ruído branco gaussiano*.

3.1.4 Função de autocovariância (facv)

Uma importante análise na modelagem de séries temporais é observar a covariância entre os seus valores em diferentes instantes de tempo. Segundo Morettin; Toloi (2006), para um processo X estacionário de segunda ordem discreto qualquer, sua função de autocovariância (facv) é definida por:

$$\gamma(t_1, t_2) = E\{X(t_1)X(t_2)\} - E\{X(t_1)\}E\{X(t_2)\}. \quad (\text{Eq. 3.10})$$

Em particular, se $t_1 = t_2 = t$,

$$\gamma(t, t) = \text{Var}\{X(t)\} = E\{X^2(t)\} - E^2\{X(t)\}, \quad (\text{Eq. 3.11})$$

que é a (função) variância do processo X .

A facv também pode ser representada na forma γ_τ , onde $\gamma_\tau = E\{X_t X_{t+\tau}\}$.

Além da *facv*, outra função que é bastante utilizada na análise de séries temporais é a função de autocorrelação, definida por:

$$\rho_\tau = \frac{\gamma_\tau}{\gamma_0}, \tau \in Z. \quad (\text{Eq. 3.12})$$

O valor de ρ_τ é definido no intervalo $[-1,1]$, onde 1 indica perfeita correlação, -1 indica perfeita correlação negativa e zero indica correlação nula.

Para processos puramente aleatórios, como um ruído branco, os seus valores são independentes no tempo, ou seja, são não-correlacionados. Tal comportamento pode ser visualizado através da função de autocorrelação (*fac*) do processo:

Considere um processo puramente aleatório tal que:

$$\varepsilon_t \sim i.i.d.(0, \sigma_\varepsilon^2).$$

A *fac* desse ruído branco é ilustrada na Figura 3-2:

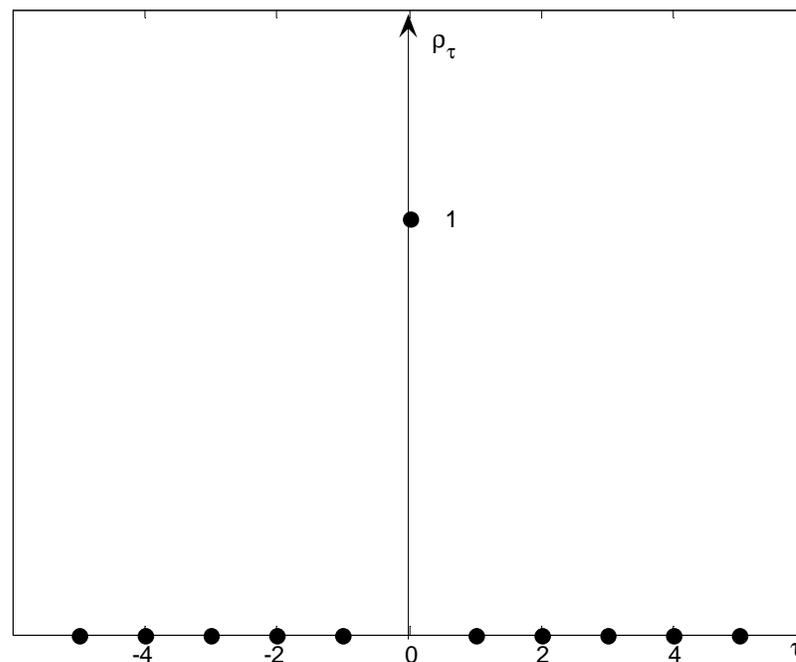


Figura 3-2 – Fac de um ruído branco.
Fonte: Morettin; Tolo (2006).

3.1.5 Tipos de modelos

Os modelos de séries temporais existentes podem ser classificados em duas classes:

1. Modelos lineares, que podem ser estacionários ou não-estacionários;
2. Modelos não-lineares, nos quais a série explicada apresenta uma combinação não-linear de parâmetros.

Os tipos de modelo que serão apresentados nesta seção são os modelos lineares estacionários (AR, MA e ARMA), modelos lineares não-estacionários (ARIMA) e modelos não-lineares (ARCH, GARCH e Volatilidade Estocástica). Como referência aos outros tipos de modelos existentes, ver Morettin; Tolo (2006).

A teoria de séries temporais apresenta alguns operadores para facilitar a manipulação dos modelos estudados. Os operadores que serão utilizados são:

- a) Operador translação para o passado:

$$BZ_t = Z_{t-1}, B^m Z_t = Z_{t-m}; \quad (\text{Eq. 3.13})$$

- b) Operador diferença, já definido anteriormente:

$$\Delta Z_t = Z_t - Z_{t-1} = (1 - B)Z_t. \quad (\text{Eq. 3.14})$$

3.1.6 Modelos lineares

Os modelos lineares estacionários podem ser representados como filtros lineares, como na Figura 3-3:

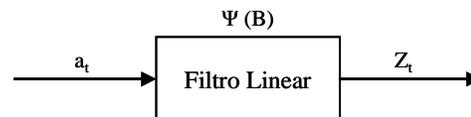


Figura 3-3 - Diagrama representativo de um filtro linear.

Este filtro pode ser representado na forma:

$$Z_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots = \mu + \psi(B)a_t, \quad (\text{Eq. 3.15})$$

em que

$$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots, \quad (\text{Eq. 3.16})$$

é denominada *função de transferência* do filtro, μ é um parâmetro determinando o nível da série e B é o operador translação para o passado.

Chamando $\tilde{Z}_t = Z_t - \mu$, temos que

$$\tilde{Z}_t = \psi(B)a_t. \quad (\text{Eq. 3.17})$$

Se a seqüência de pesos $\{\psi_j, j \geq 1\}$ for finita ou infinita e convergente, o filtro é dito estável (somável) e Z_t é estacionária. Neste caso, μ é a média do processo.

Uma forma alternativa de representar o processo \tilde{Z}_t é ponderando os valores passados $\tilde{Z}_{t-1}, \tilde{Z}_{t-2}, \dots$ mais um ruído branco a_t :

$$\tilde{Z}_t = \pi_1 \tilde{Z}_{t-1} + \pi_2 \tilde{Z}_{t-2} + \dots + a_t = \sum_{j=1}^{\infty} \pi_j \tilde{Z}_{t-j} + a_t. \quad (\text{Eq. 3.18})$$

Segue-se que

$$\left(1 - \sum_{j=1}^{\infty} \pi_j B^j \right) \tilde{Z}_t = a_t$$

ou

$$\pi(B)\tilde{Z}_t = a_t, \quad (\text{Eq. 3.19})$$

onde $\pi(B)$ é o operador

$$\pi(B) = 1 - \pi_1 B - \pi_2 B^2 - \dots$$

que é equivalente à

$$\pi(B)\psi(B)a_t = a_t,$$

de modo que

$$\pi(B) = \psi^{-1}(B). \quad (\text{Eq. 3.20})$$

Esta relação pode ser usada para obter os pesos π_j em função dos pesos ψ_j e vice-versa.

3.1.6.1 Modelos AR

Considerando o filtro linear apresentado com $\pi_j = 0, j > p$, obtém-se um *modelo auto-regressivo de ordem p*, denotado por AR(p):

$$\tilde{Z}_t = \sum_{j=1}^p \pi_j \tilde{Z}_{t-j} + a_t. \quad (\text{Eq. 3.21})$$

Assim, um modelo AR tenta prever o valor do instante t considerando uma combinação linear de seus valores anteriores.

O caso mais simples é o modelo AR(1):

$$\tilde{Z}_t = \pi_1 \tilde{Z}_{t-1} + a_t, \quad (\text{Eq. 3.22})$$

de maneira que \tilde{Z}_t depende apenas de \tilde{Z}_{t-1} e do ruído branco no instante t . A Figura 3-4 apresenta um exemplo de uma série de 50 observações geradas a partir de um processo AR(1):

$$Z_t = 0,8Z_{t-1} + a_t,$$

onde $a_t \sim N(0, 1)$.

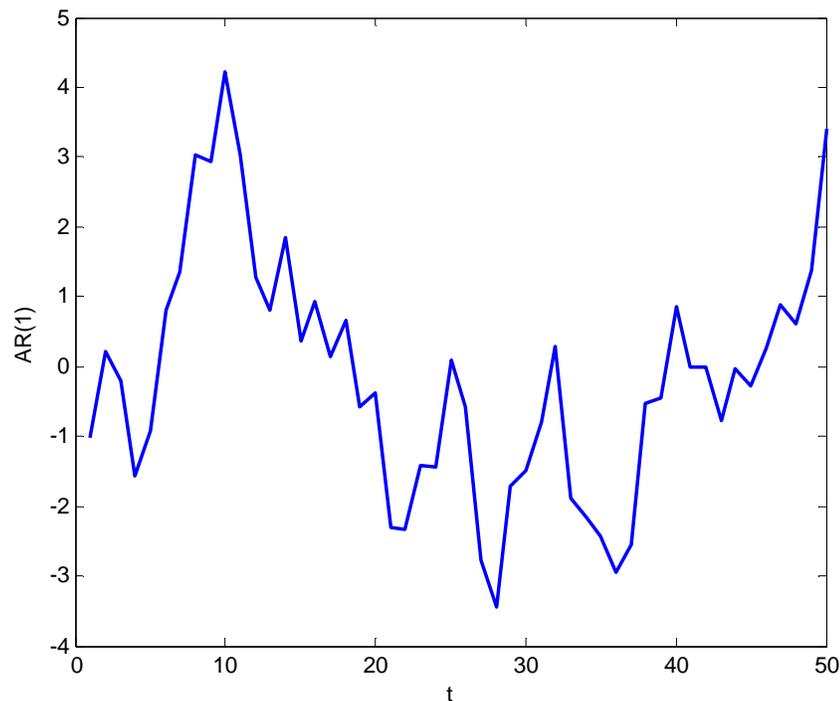


Figura 3-4 - Série de observações de um modelo AR(1).

No processo apresentado, temos que $\pi_1 = 0,8$, assim a função de autocorrelação do processo decai exponencialmente. A Figura 3-5 apresenta a fac teórica e amostral para o processo descrito, calculada com as observações do mesmo.

Deve-se ressaltar que os parâmetros de um modelo devem satisfazer algumas restrições para a modelagem adequada. No caso de um processo AR(1) qualquer, o parâmetro π_1 da equação de auto-regressão deve satisfazer $|\pi_1| \leq 1$ para que o processo seja estacionário e não exploda com a evolução do tempo, já que, caso contrário, o processo apresentaria um crescimento exponencial.

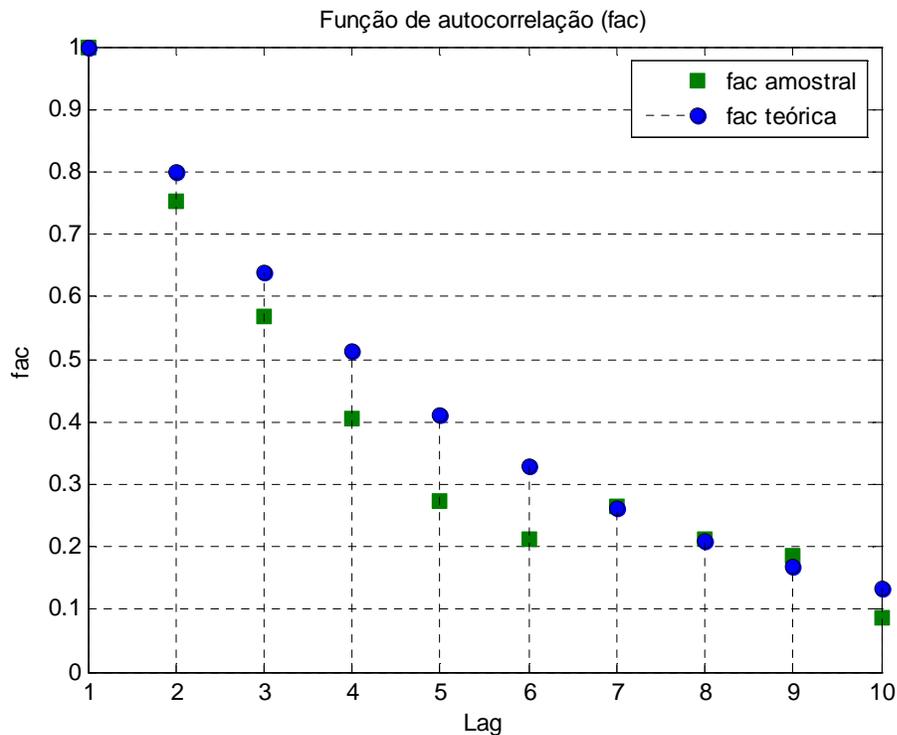


Figura 3-5 - Fac amostral e teórica de um processo AR(1).

3.1.6.2 Modelos MA

Considerando o filtro linear $Z_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots = \mu + \psi(B)a_t$, e supondo que $\psi_j = 0, j > q$, obtém-se um processo de médias móveis de ordem q , denotado por MA(p):

$$\tilde{Z}_t = \sum_{j=1}^p \psi_j a_{t-j} + a_t, \quad (\text{Eq. 3.23})$$

de maneira que \tilde{Z}_t depende apenas dos valores dos $(t-p)$ ruídos brancos passados e do ruído branco no instante t . A Figura 3-6 apresenta um exemplo de uma série de 50 observações para um processo MA(1):

$$\tilde{Z}_t = a_t - 0,8a_{t-1}.$$

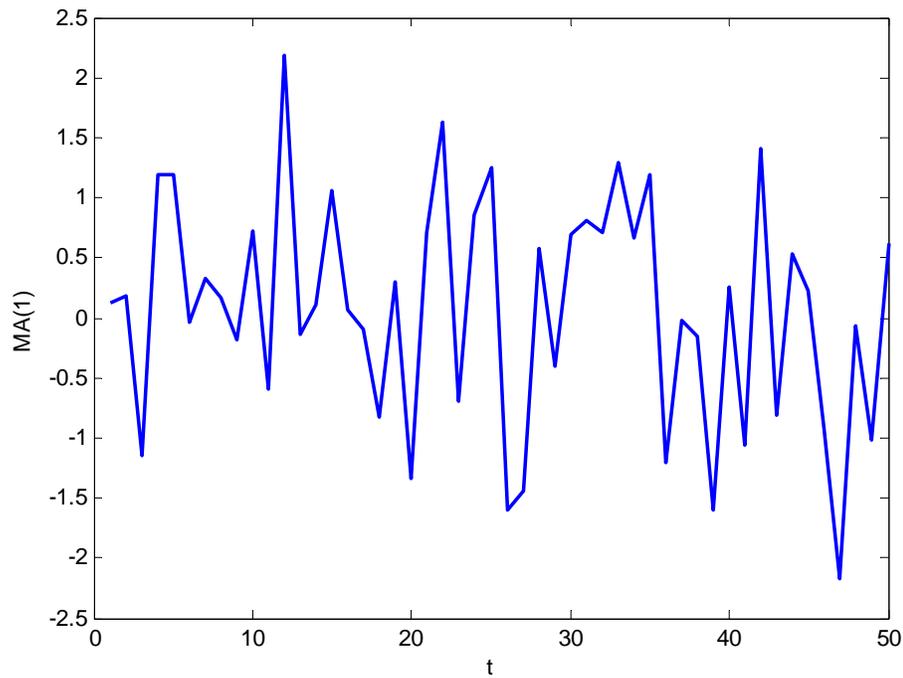


Figura 3-6 - Série de observações de um modelo MA(1).

E as funções de autocorrelação teóricas e amostrais do processo são apresentadas na Figura 3-7:

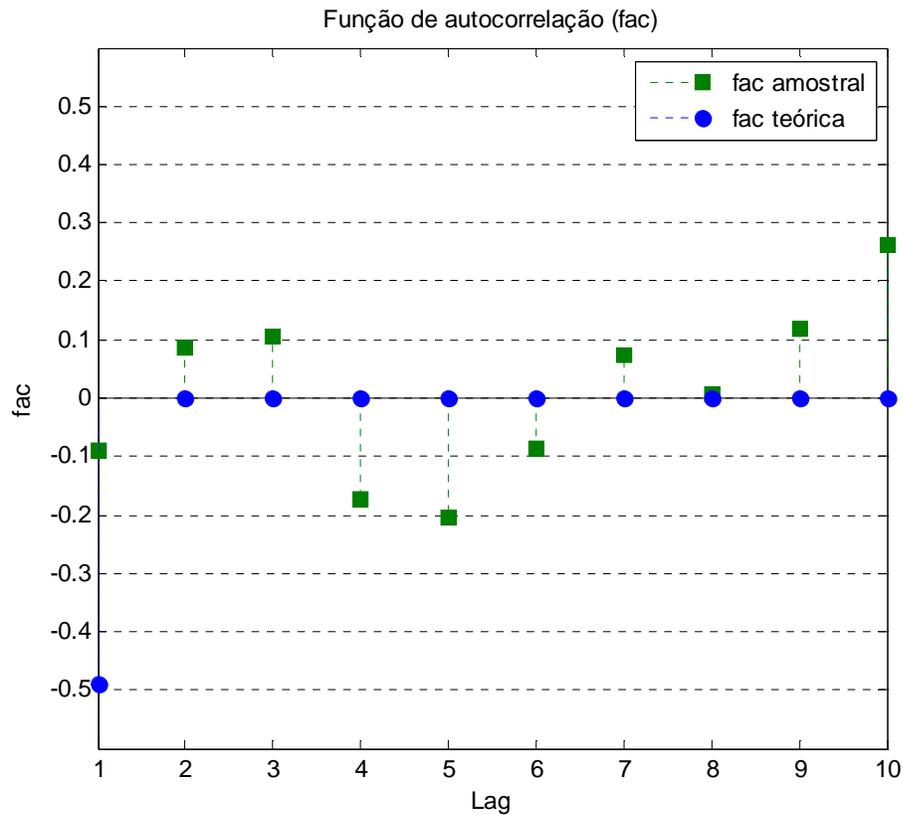


Figura 3-7 - Fac amostral e teórica de um processo MA(1).

Neste caso, pode-se notar que a fac teórica de um processo MA(q) é igual a zero para defasagens (“lags”) maiores do que q , ao contrário do que acontece para um processo AR, onde há um decaimento exponencial. Isto equivale dizer que um processo MA(q) possui memória apenas até o elemento $t - q$.

3.1.6.3 Modelos ARMA

Os modelos auto-regressivos e de médias móveis podem ser úteis para a descrição parcimoniosa de uma série, isto é, a construção de um modelo com um número de parâmetros não muito grande. Um modelo genérico ARMA(p, q) é construído a partir de uma parcela auto-regressiva e uma parcela de médias móveis:

$$\tilde{Z}_t = \sum_{j=1}^p \pi_j \tilde{Z}_{t-j} + \sum_{j=1}^q \psi_j a_{t-j} + a_t. \quad (\text{Eq. 3.24})$$

3.1.6.4 Modelos ARIMA

Os modelos auto-regressivos integrados de médias móveis (ARIMA) são freqüentemente utilizados para descrever o comportamento de séries econômicas e sociais, onde os erros (resíduos) observados são autocorrelacionados, e portanto influenciam a evolução do processo.

Os modelos ARIMA são bastante utilizados para descrever séries Z_t tais que, tomando-se um número finito de diferenças, d , tornam-se estacionárias (séries *não-estacionárias homogêneas*).

Se $W_t = \Delta^d Z_t$ for estacionária, podemos representar W_t por um modelo ARMA(p, q). Neste caso, trata-se de um processo ARIMA(p, d, q), que supõe que a d -ésima diferença da série Z_t pode ser representada por um modelo ARMA. Como a série W_t é gerada a partir de diferenças de Z_t , então Z_t é uma integral de W_t , e daí provém o nome de modelo auto-regressivo, integrado, de médias móveis.

3.1.7 Modelos não-lineares

Os modelos abordados até agora são freqüentemente utilizados para descrever séries estacionárias ou não-estacionárias homogêneas que apresentam variância constante. Para descrever o comportamento de séries que apresentam variância condicional que evolui com o tempo, os modelos do tipo ARIMA não são adequados.

Há uma variedade muito grande de modelos não-lineares que tratam a variância como uma variável no tempo. Exemplos mais conhecidos são os modelos ARCH (*autoregressive conditional heterocedasticity*), os modelos GARCH (*generalized autoregressive heterocedasticity*), e os modelos de volatilidade estocástica (*stochastic volatility*). O objetivo comum dessas classes é modelar a variância condicional dos retornos, ou seja, a sua volatilidade. Embora não possa ser medida diretamente, a volatilidade tem um comportamento com algumas características observáveis (Peña et. al., 2001):

- a) A volatilidade aparece em grupos, de maior ou menor variabilidade;

- b) A volatilidade evolui continuamente no tempo, e pode ser considerada estacionária;
- c) Ela reage de modo diferente a valores positivos ou negativos da série.

3.1.7.1 Modelos ARCH

Os modelos ARCH (auto-regressivos com heteroscedasticidade condicional), apresentados por Engle (1982), surgiram com o objetivo de estimar a variância da inflação. Nestes modelos, o retorno X_t é considerado não-correlacionado serialmente, mas a volatilidade depende dos retornos passados, por meio de uma função quadrática. Um modelo ARCH(r) é definido por:

$$X_t = \sqrt{h_t} a_t, \quad (\text{Eq. 3.25})$$

$$h_t = \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_r X_{t-r}^2, \quad (\text{Eq. 3.26})$$

onde a_t i.i.d. $(0, 1)$, $\alpha_0 > 0, \alpha_i \geq 0, i > 0$.

Os coeficientes α_i devem satisfazer certas condições, dependendo do tipo de imposição que o processo X_t deve seguir (uma imposição necessária é que a volatilidade h_t seja estacionária e positiva).

Este modelo é caracterizado por apresentar caudas longas, isto é, caso X_t siga um modelo ARCH, as caudas serão mais pesadas do que as da normal (isto é, curtose de X_t é maior do que a de um processo normal). Dado que os retornos das séries financeiras apresentam este tipo de característica, os modelos ARCH apresentam uma propriedade vantajosa. Outra vantagem do modelo é que a estimação dos seus parâmetros pode ser realizada facilmente pelo método de máxima verossimilhança condicional (que será apresentado posteriormente). Uma desvantagem do modelo é que este não assume que retornos positivos ou negativos impactam de forma diferente na volatilidade, o que é observado na prática.

3.1.7.2 Modelos GARCH

Como o próprio nome diz, o modelo GARCH é uma generalização dos modelos ARCH (“*generalized* ARCH”), sugerida por Bollerslev (1986). Do mesmo modo que um modelo ARMA pode ser mais parcimonioso do que um modelo AR ou MA puro, um modelo GARCH pode ser usado para descrever a volatilidade com menos parâmetros do que um modelo ARCH.

Um modelo GARCH(r,s) é definido por:

$$X_t = \sqrt{h_t} a_t, \quad (\text{Eq. 3.27})$$

$$h_t = \alpha_0 + \sum_{i=1}^r \alpha_i X_{t-i}^2 + \sum_{j=1}^s \beta_j h_{t-j}, \quad (\text{Eq. 3.28})$$

onde a_t i.i.d. (0,1), $\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, $\sum_{i=1}^q (\alpha_i + \beta_i) < 1$, $q = \max(r, s)$.

Como no caso de um modelo ARCH, usualmente supõe-se que os a_t são normais ou seguem uma distribuição t de Student com ν graus de liberdade.

É possível derivar uma equação para a variância não-condicional do modelo, expressa por:

$$E(X_t^2) = \frac{\alpha_0}{1 - \sum_{i=1}^q (\alpha_i + \beta_i)}. \quad (\text{Eq. 3.29})$$

Assim como os modelos ARCH, o modelo GARCH apresenta algumas características vantajosas, como: a possibilidade pela estimação dos parâmetros via o método de máxima verossimilhança, apresenta caudas pesadas, volatilidades altas são precedidas de retornos ou volatilidades grandes.

Em geral, existe uma alta persistência na volatilidade das séries de retornos, o que implica num alto valor de r para um modelo ARCH(r) e, portanto, a necessidade de estimação de um grande número de parâmetros. Assim, o modelo GARCH surge como uma tentativa de expressar de forma mais parcimoniosa a dependência temporal dos retornos da variância condicional. Deste modo, além de depender do quadrado dos retornos passados, como no modelo ARCH, a volatilidade também depende das próprias variâncias condicionais passadas.

Há uma literatura muito grande sobre extensões dos modelos ARCH-GARCH, como os modelos IGARCH, EGARCH, TARCH, QGARCH, entre outros. Como referência a estes e outros tipos de modelos não-lineares, ver Morettin; Tolo (2006), Nelson (1991) e Zakoian (1994).

3.1.7.3 Modelos de Volatilidade Estocástica

O modelo que será desenvolvido no presente trabalho trata-se de um problema univariado de volatilidade estocástica. Para a resolução do modelo e estimação dos seus parâmetros na sua aplicação, é necessária a utilização da teoria de inferência Bayesiana e de algoritmos de simulação via MCMC, que serão apresentados na Seção 3.2. Assim, a apresentação dos modelos de volatilidade estocástica será realizada posteriormente, na Seção 3.3.

3.1.8 Diagnósticos de Modelos

Após a etapa de estimação dos parâmetros de um modelo, é necessário verificar se ele representa adequadamente a série objeto de estudo. Esta etapa de validação do modelo pode ajudar também a sugerir modelos alternativos em caso de inadequação.

Há diversos testes de adequação de um modelo de séries temporais, sendo que a maioria baseia-se nas autocorrelações estimadas dos resíduos (isto é, a autocorrelação dos erros entre a série real e a série ajustada). Os testes que serão apresentados são os testes comumente usados para a verificação de adequação dos modelos ARCH-GARCH.

3.1.8.1 Teste de Ljung-Box

O teste da estatística Q , proposto por Box e Pierce (1970), é utilizado para a verificação da aleatoriedade em um conjunto de dados. Assim, a partir das

autocorrelações dos resíduos de um modelo, o teste verifica a hipótese nula da ausência de correlação serial dos dados.

Assim, se o modelo for adequado, a estatística

$$Q(K) = n(n+2) \sum_{j=1}^K \frac{\hat{r}_j^2}{(n-j)} \quad (\text{Eq. 3.30})$$

terá aproximadamente uma distribuição χ^2 com $K - p - q$ graus de liberdade, onde n é o tamanho da amostra, K é o número de “lags” autocorrelacionados incluídos no teste, e \hat{r}_j^2 é o quadrado da autocorrelação amostral no “lag” j .

As hipóteses do teste para um conjunto de dados são:

H_0 : os dados são aleatórios.

H_1 : os dados não são aleatórios.

Para um nível de significância α , a hipótese nula é rejeitada se:

$$Q(K) > \chi_{1-\alpha, K}^2$$

onde $\chi_{1-\alpha, K}^2$ representa o α -ésimo quantil de uma distribuição chi-quadrada com K graus de liberdade.

3.1.8.2 Teste de multiplicadores de Lagrange

O teste de multiplicadores de Lagrange (ML), proposto por Engle (1982), consiste em testar:

H_0 : não há heterocedasticidade condicional dos resíduos.

H_1 : há heterocedasticidade condicional dos resíduos.

Em termos matemáticos, o ML consiste em testar:

$H_0 : \alpha_i = 0$, para todo $i = 1, \dots, r$, na regressão:

$$X_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \dots + \alpha_r X_{t-r}^2 + u_t, \quad (\text{Eq. 3.31})$$

para $t = r+1, \dots, n$, onde X_t^2 é o quadrado do resíduo do modelo no instante t . A estatística do teste é:

$$T = NR^2 \sim \chi^2(r)$$

em que R^2 é o coeficiente de determinação da regressão.

O teste rejeita a hipótese nula de ausência de heterocedasticidade condicional dos resíduos do modelo caso o valor de T seja maior que a sua estatística no nível de significância escolhido.

3.1.8.3 Análise da fac residual

A análise da autocorrelação dos resíduos do modelo tentativo pode ser feita através da construção da função de autocorrelação. Através dela, pode-se verificar se há persistência de autocorrelação nos resíduos, o que indicaria uma inadequação do modelo em explicar esta correlação ainda existente. É recomendável realizar esta análise também para os quadrados dos resíduos padronizados, para a verificação de heterocedasticidade residual, o que pode ser realizado através do Teste de multiplicadores de Lagrange.

3.2. Inferência Bayesiana

Inferência é o processo de utilização da teoria estatística para analisar e descrever alguma população a partir de uma amostra. O objetivo principal é realizar afirmações sobre quantidades desconhecidas, buscando extrair informações do todo a partir de um pedaço.

Atualmente, há duas correntes de inferência estatística: a freqüentista e a Bayesiana. A principal diferença entre elas é que na inferência Bayesiana, informações prévias sobre a variável de interesse podem ser incorporadas às análises. As informações *a priori*, para serem devidamente incorporadas na análise, devem ser traduzidas em uma distribuição de probabilidade. Tal distribuição do parâmetro populacional é chamada de distribuição *a priori*. Caso seja interessante considerar informações subjetivas, estas devem ser devidamente traduzidas em funções de distribuição de probabilidade, para que possam ser incorporadas na análise. A distribuição de densidade *a priori*, portanto, denota o grau de conhecimento acumulado sobre o processo, antes da observação dos dados amostrais.

O objetivo da inferência estatística é usar informações para realizar inferências sobre parâmetros desconhecidos. A inferência Bayesiana é uma parte da estatística que além de utilizar as informações disponíveis (dados) para fazer afirmações, agrega as informações *a priori*. Assim, a inferência Bayesiana busca atribuir funções de probabilidades para uma variável de maneira condicional aos dados da amostra e às informações *a priori*.

As informações *a priori* devem ser explicadas através de uma distribuição de probabilidade adequada para que, em conjunto com os dados observados, se consiga obter uma distribuição de probabilidade *a posteriori*, que é a distribuição de probabilidade de parâmetros dado as informações *a priori*, os dados observados e os valores de outros parâmetros.

Seja θ o vetor de parâmetros desconhecidos de um modelo definido, e X o vetor de dados observados. As informações *a priori* dos parâmetros são especificadas através de uma função $f(\theta)$. Seja $f(X|\theta)$ a função de verossimilhança. Assim, usando a teoria de probabilidade condicional:

$$f(\boldsymbol{\theta} | \mathbf{X}) = \frac{f(\boldsymbol{\theta}, \mathbf{X})}{f(\mathbf{X})} = \frac{f(\mathbf{X} | \boldsymbol{\theta})f(\boldsymbol{\theta})}{f(\mathbf{X})} \quad (\text{Eq. 3.32})$$

onde $f(\mathbf{X})$, distribuição marginal dos dados, pode ser obtida por:

$$f(\mathbf{X}) = \int f(\mathbf{X}, \boldsymbol{\theta})d\boldsymbol{\theta} = \int f(\mathbf{X} | \boldsymbol{\theta})f(\boldsymbol{\theta})d\boldsymbol{\theta}. \quad (\text{Eq. 3.33})$$

A distribuição $f(\boldsymbol{\theta} | \mathbf{X})$ na Eq. 3.32 é chamada de distribuição *a posteriori* de $\boldsymbol{\theta}$. Em geral, podemos escrever:

$$f(\boldsymbol{\theta} | \mathbf{X}) \propto f(\mathbf{X} | \boldsymbol{\theta})f(\boldsymbol{\theta}). \quad (\text{Eq. 3.34})$$

Tal equação também pode ser expressa usando a função de verossimilhança. Dado \mathbf{X} , qualquer função proporcional a $f(\mathbf{X} | \boldsymbol{\theta})$ é chamada de função de verossimilhança, $\ell(\boldsymbol{\theta})$. A forma da curva *posteriori* é inteiramente determinada por $\ell(\boldsymbol{\theta})$ e pela *priori* no numerador, o que pode ser escrito como:

$$f(\boldsymbol{\theta} | \mathbf{X}) \propto \ell(\boldsymbol{\theta})f(\boldsymbol{\theta}). \quad (\text{Eq. 3.35})$$

Caso $\ell(\boldsymbol{\theta}) = f(\mathbf{X} | \boldsymbol{\theta})$, então a constante de proporcionalidade é a distribuição marginal dos dados, $f(\mathbf{X}) = \int f(\mathbf{X}, \boldsymbol{\theta})d\boldsymbol{\theta} = \int f(\mathbf{X} | \boldsymbol{\theta})f(\boldsymbol{\theta})d\boldsymbol{\theta}$.

Na maioria das aplicações, é de interesse obter a distribuição marginal de cada um dos elementos do vetor de parâmetros, isto é, a partir da distribuição *posteriori* conjunta, obter as distribuições *posteriores* marginais dos parâmetros. Assim, considerando um problema de inferência com um vetor de parâmetros $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)'$, por exemplo, a distribuição *posteriori* marginal do parâmetro θ_1 pode ser obtida a partir da *posteriori* conjunta $f(\boldsymbol{\theta} | \mathbf{X}) = f(\theta_1, \theta_2, \theta_3 | \mathbf{X})$ por:

$$f(\theta_1 | \mathbf{X}) = \iint f(\theta_1, \theta_2, \theta_3 | \mathbf{X})d\theta_2d\theta_3. \quad (\text{Eq. 3.36})$$

3.2.1 Distribuições *a Priori* Conjugadas

Obter a distribuição *a posteriori* nem sempre é trivial, mas há alguns casos em que a *priori* e a *posteriori* pertencem a uma mesma família de distribuições.

Nesse caso, a *priori* é chamada de distribuição conjugada. A existência de tal distribuição implica a possibilidade de se obter uma expressão analítica para a *posteriori*, o que facilita muito a condução da análise Bayesiana. Em modelos que apresentam uma distribuição *a priori* não-conjugada, outros métodos devem ser utilizados para o cálculo das distribuições *a posteriori*. O método mais utilizado para tais situações é o Markov Chain Monte Carlo (MCMC), que a utilização de procedimentos numéricos para o cálculo das integrais necessárias. Entre os algoritmos MCMC mais aplicados estão o Amostrador de Gibbs, Metropolis-Hastings e o Griddy Gibbs, que serão apresentados posteriormente.

É comum a utilização de distribuições conjugadas na inferência Bayesiana, principalmente quando não há muita certeza nas informações sobre os parâmetros do modelo. Além disso, em casos de incerteza, as *prioris* são caracterizadas por possuírem valores pouco informativos de seus hiperparâmetros, que são os parâmetros das distribuições *a priori*. Nesse sentido, os hiperparâmetros tratam-se de parâmetros de parâmetros. É importante observar nos resultados da inferência Bayesiana a dependência dos resultados obtidos em relação às suas *prioris*. De modo geral, a obtenção de distribuições *posterioris* bem definidas e diferentes das suas respectivas *prioris* são um bom indicador da coerência e convergência do algoritmo.

A seguir, são apresentadas algumas distribuições conjugadas bastante utilizadas na inferência Bayesiana através de resultados disponíveis na literatura, apresentados por Tsay (2002). Serão abordadas as distribuições que serão utilizadas posteriormente na resolução do problema proposto, Para mais informações sobre outras distribuições conjugadas, ver DeGroot (1970).

Resultado 1: Suponha que x_1, \dots, x_n formem uma amostra aleatória de uma distribuição normal com média μ , que é desconhecida, e variância σ^2 , conhecida. Suponha que a distribuição *priori* de μ é uma normal com média μ_0 e variância σ_0^2 . Então a distribuição *posteriori* de μ dado os valores observados e a *priori* é uma normal com média μ_* e variância σ_*^2 segue uma distribuição normal com:

$$\mu_* = \frac{\sigma^2 \mu_0 + n \sigma_0^2 \bar{x}}{\sigma^2 + n \sigma_0^2} \quad \text{e} \quad \sigma_*^2 = \frac{\sigma^2 \sigma_0^2}{\sigma^2 + n \sigma_0^2},$$

onde $\bar{x} = \sum_{i=1}^n x_i / n$ é a média amostral.

Na análise Bayesiana, geralmente é conveniente usar o parâmetro de precisão $\eta = 1/\sigma^2$. Sendo $\eta_0 = 1/\sigma_0^2$ o parâmetro de precisão da distribuição *priori*, e $\eta_* = 1/\sigma_*^2$, o resultado apresentado pode ser reescrito na forma

$$\eta_* = \eta_0 + n\eta \quad \text{e} \quad \mu_* = \frac{\eta_0}{\eta_*} \times \mu_0 + \frac{n\eta}{\eta_*} \times \bar{x}.$$

Ainda em relação ao resultado apresentado, as informações observáveis (isto é, o conjunto de dados) sobre μ estão contidas na média amostral \bar{x} , que é a estatística suficiente de μ . A precisão de \bar{x} é $n/\sigma^2 = n\eta$. Conseqüentemente, a expressão dos parâmetros da *posteriori* mostra que (a) a precisão da distribuição *posteriori* é a soma das precisões da *priori* e da amostra, e (b) a média da *posteriori* é uma média ponderada da média da *priori* e da média amostral, com peso proporcional à precisão. Além disso, o resultado mostra que a contribuição da distribuição *priori* para a análise diminui com o aumento do número de observações n .

O resultado obtido pode ser estendido para a distribuição normal multivariada. Neste caso, dada uma distribuição com um vetor de média μ e matriz de covariância conhecida Σ com uma *priori* normal multivariada com vetor de média μ_0 e matriz de covariância Σ_0 , então a distribuição *posteriori* de μ também é multivariada normal com vetor de média μ_* e matriz de covariância Σ_* , onde

$$\Sigma_*^{-1} = \Sigma_0^{-1} + n\Sigma^{-1} \quad \text{e} \quad \mu_* = \Sigma_* (\Sigma_0^{-1} \mu_0 + n\Sigma^{-1} \bar{x}).$$

onde $\bar{x} = \sum_{i=1}^n \mathbf{x}_i / n$ é o vetor das médias amostrais, que é distribuída sendo uma normal multivariada com média μ e matriz de covariância Σ/n . Pode notar-se que

$n\Sigma^{-1}$ é a matriz de precisão de \bar{x} e Σ_0^{-1} é a matriz de precisão da distribuição a *priori*.

Resultado 2: Suponha que x_1, \dots, x_n formem uma amostra aleatória de uma distribuição normal com média zero e variância σ^2 . Suponha também que a distribuição *priori* de σ^2 é uma inversa chi-quadrada com ν graus de liberdade, isto é

$$\frac{(\nu\lambda)}{\sigma^2} \sim \chi_\nu^2, \text{ onde } \lambda > 0.$$

Nessas condições, a distribuição *posteriori* de σ^2 também será uma inversa chi-quadrada com $\nu + n$ graus de liberdade

$$\frac{(\nu\lambda + \sum_{i=1}^n x_i^2)}{\sigma^2} \sim \chi_{\nu+n}^2.$$

A distribuição inversa qui-quadrada é bastante utilizada em modelos que apresentam variáveis aleatórias com variância condicional. Uma variável Y tem uma distribuição inversa qui-quadrada com ν graus de liberdade se $1/Y$ segue uma distribuição qui-quadrada com os mesmos graus de liberdade. A função densidade de probabilidade de Y é

$$f(y | \nu) = \frac{2^{-\nu/2}}{\Gamma(\nu/2)} y^{-(\nu/2+1)} e^{-1/(2y)}, \quad y > 0. \quad (\text{Eq. 3.37})$$

Para essa distribuição tem-se que $E(Y) = 1/(\nu - 2)$ se $\nu > 2$ e $\text{Var}(Y) = 2/[(\nu - 2)^2(\nu - 4)]$ se $\nu > 4$.

3.2.2 Simulação por Cadeias de Markov

Para contornar os casos em que uma forma analítica da distribuição *a posteriori* não pode ser obtida (ou for extremamente difícil de obter), recentemente foram usados métodos baseados em simulação chamados Markov Chain Monte Carlo.

Os métodos de inferência Baysiana baseados em simulação estocástica (MCMC) utilizam a teoria de cadeias de Markov e técnicas de simulação. O conceito de cadeia de Markov é atribuído ao matemático russo Andrei Andreivich Markov, que desenvolveu um modelo probabilístico onde sucessivos resultados dependiam somente dos resultados imediatamente predecessores.

Em relação à inferência Bayesiana, os algoritmos existentes utilizam a teoria de Markov, e suas convergências são provadas usando as propriedades desta teoria. Entretanto, para fins práticos, ou seja, de implementação dos algoritmos, o conhecimento sobre a teoria de Markov não é crucial. Assim, a teoria de cadeias de Markov não será apresentada neste presente trabalho, mas para maiores detalhes ver Feller (1968), Meyn e Tweedie (1993), Nummelin (1984) e Ross (1996).

Em suma, o principal objetivo dos algoritmos MCMC presentes pode ser apresentado como a seguir:

Dado um vetor de parâmetros θ e um conjunto de dados X , a simulação de Markov Chain busca uma distribuição de transição estacionária para os parâmetros, que será a distribuição desejada $f(\theta|X)$. A idéia principal no MCMC é criar um processo de Markov cuja distribuição de transição estacionária seja $f(\theta|X)$ e rodar uma simulação suficientemente grande para obter sua distribuição estacionária.

3.2.3 Métodos baseados em simulação estocástica

Recentemente vários algoritmos foram desenvolvidos a fim de contornar as dificuldades existentes em modelos não-conjugados e, aliados ao avanço da

capacidade computacional de processamento de dados, apresentam extrema importância prática para a inferência Bayesiana.

Os principais algoritmos MCMC são:

- Amostrador de Gibbs (Gibbs sampling);
- Metropolis-Hastings;
- Griddy Gibbs.

Os algoritmos que serão utilizados na resolução deste trabalho são o Amostrador de Gibbs e o Griddy Gibbs. Para maiores detalhes sobre os algoritmos Metropolis-Hastings, entre outros, ver Gamerman (2006).

3.2.3.1 Amostrador de Gibbs

Apresentado por Geman and Geman (1984) e Gelfand e Smith (1990) é um dos métodos mais populares de MCMC. O Amostrador de Gibbs é usado em situações em que a função de verossimilhança é difícil de se obter, mas as distribuições condicionais de seus parâmetros, dados os outros parâmetros, são disponíveis.

O Amostrador de Gibbs baseia-se no conceito de “*data augmentation*”, que consiste na adição de variáveis auxiliares para a resolução de um problema. Assim, o problema é decomposto em diversos problemas menores, menos complexos, e mais fáceis de serem resolvidos.

O algoritmo de Gibbs diminui a complexidade do problema, pois decompõe um problema de estimação de muitas dimensões em vários problemas de dimensões menores, via as distribuições condicionais dos parâmetros. Assim, no extremo, um problema de dimensão N poderia ser decomposto em N problemas univariados de distribuição condicional. Cabe-se ressaltar que nem sempre tal decomposição é eficiente. Se os parâmetros forem altamente correlacionados, é difícil tirar amostras da distribuição *a posteriori* conjunta dos parâmetros.

Dado o conjunto de parâmetros θ de um modelo M proposto. Considere X como o conjunto de todas as variáveis explicativas do modelo. O objetivo é estimar os parâmetros do modelo para que este possa ser usado para realizar inferências. Suponha que a função de verossimilhança seja difícil de obter, ou que sua resolução seja infactível. Assim, se as distribuições condicionais de um único parâmetro, dados os outros parâmetros, forem fáceis de obter, o Amostrador de Gibbs pode ser usado para a construção de uma cadeia de Markov que, ao tornar-se estacionária, corresponderá a distribuição conjunta dos parâmetros do modelo.

A seguir, um exemplo genérico do algoritmo é apresentado:

Considere um modelo com três parâmetros θ_1 , θ_2 e θ_3 . Denote a coleção de dados disponíveis por X , e o modelo proposto por M . O Amostrador de Gibbs pode ser utilizado para calcular a função de verossimilhança caso as seguintes distribuições condicionais sejam conhecidas:

$$f_1(\theta_1 | \theta_2, \theta_3, X, M); \quad f_2(\theta_2 | \theta_3, \theta_1, X, M); \quad f_3(\theta_3 | \theta_1, \theta_2, X, M),$$

onde $f_i(\theta_i | \theta_{i \neq j}, X, M)$ denota a distribuição condicional do parâmetro θ_i dados o conjunto de observações, o modelo e os outros parâmetros θ_j , $j \neq i$.

Considere $\theta_{2,0}$ e $\theta_{3,0}$ como dois valores arbitrários de θ_2 e θ_3 . O amostrador de Gibbs procede da seguinte maneira:

1. Tire uma amostra de $f_1(\theta_1 | \theta_{2,0}, \theta_{3,0}, X, M)$. Denote a amostra calculada por $\theta_{1,1}$.
2. Tire uma amostra de $f_2(\theta_2 | \theta_{3,0}, \theta_{1,1}, X, M)$. Denote a amostra calculada por $\theta_{2,1}$.
3. Tire uma amostra de $f_3(\theta_3 | \theta_{1,1}, \theta_{2,1}, X, M)$. Denote a amostra calculada por $\theta_{3,1}$.

A iteração é repetida atualizando-se os valores iniciais para $\theta_{1,1}$, $\theta_{2,1}$ e $\theta_{3,1}$. Assim, o ciclo repete-se por um número m de vezes até que se obtenha a seqüência de amostras aleatórias:

$$\theta_{1,1}, \theta_{2,1}, \theta_{3,1}, \dots, (\theta_{1,m}, \theta_{2,m}, \theta_{3,m}).$$

Se algumas condições forem respeitadas, pode-se mostrar que, para um número suficientemente grande de m , $(\theta_{1,m}, \theta_{2,m}, \theta_{3,m})$ é aproximadamente equivalente à uma amostra aleatória retirada da distribuição conjunta $f(\theta_1, \theta_2, \theta_3 | \mathbf{X}, M)$ dos três parâmetros. As condições necessárias para a convergência do algoritmo essencialmente requerem que os valores paramétricos iniciais $(\theta_{1,0}, \theta_{2,0}, \theta_{3,0})$ podem visitar todo o espaço paramétrico Θ .

Na prática, usa-se um número suficientemente grande n e descartam-se as primeiras m amostras aleatórias retiradas do Amostrador de Gibbs.

3.2.3.2 Griddy Gibbs

Em alguns problemas de inferência, a forma da distribuição condicional de todos os parâmetros (a chamada distribuição condicional “cheia”) não é conhecida, o que impossibilita a amostragem via os algoritmos tradicionais. Isto ocorre na maioria das aplicações financeiras onde há distribuições *a posteriori* não-lineares dos parâmetros (usual em modelos de volatilidade).

Ritter e Tanner (1992) propuseram um esquema de amostragem para casos que apresentam condicionais cheias difíceis. O método baseia-se em avaliar a condicional cheia em alguns pontos selecionados, como apresentado por Gamerman (2006), a seguir:

Considere θ_i um parâmetro escalar com distribuição condicional *posteriori* $f(\theta_i | \mathbf{X}, \theta_{-i})$, onde θ_{-i} é o vetor de parâmetros após remover θ_i . Por exemplo, se $\theta = (\theta_1, \theta_2, \theta_3)'$, então $\theta_{-1} = (\theta_2, \theta_3)'$. Assim, o Griddy Gibbs consiste em:

1. Selecionar um conjunto de pontos de um intervalo adequadamente escolhido de θ_i , como $\theta_{i1} \leq \theta_{i2} \leq \dots \leq \theta_{im}$. Avaliar a função densidade da condicional posteriori para obter $w_j = f(\theta_{ij} | \mathbf{X}, \theta_{-i})$ para $j = 1, \dots, m$.

2. Usar w_1, \dots, w_m para obter uma aproximação da função inversa de distribuição de probabilidade acumulada (FPA) de $f(\theta_{ij} | X, \theta_{-i})$.
3. Tirar uma amostra de uma variável uniformemente distribuída em (0,1) e transformar a observação da distribuição uniforme através da função inversa FPA aproximada, para obter uma amostra aleatória de θ_i .

Uma aproximação simples para a função inversa FPA é uma distribuição discreta para $\{\theta_{ij}\}_{j=1}^m$ com probabilidade $p(\theta_{ij}) = w_j / \sum_{v=1}^m w_v$.

A amostragem de um θ_i genérico é ilustrada na Figura 3-8 da função inversa FPA para uma distribuição normal:

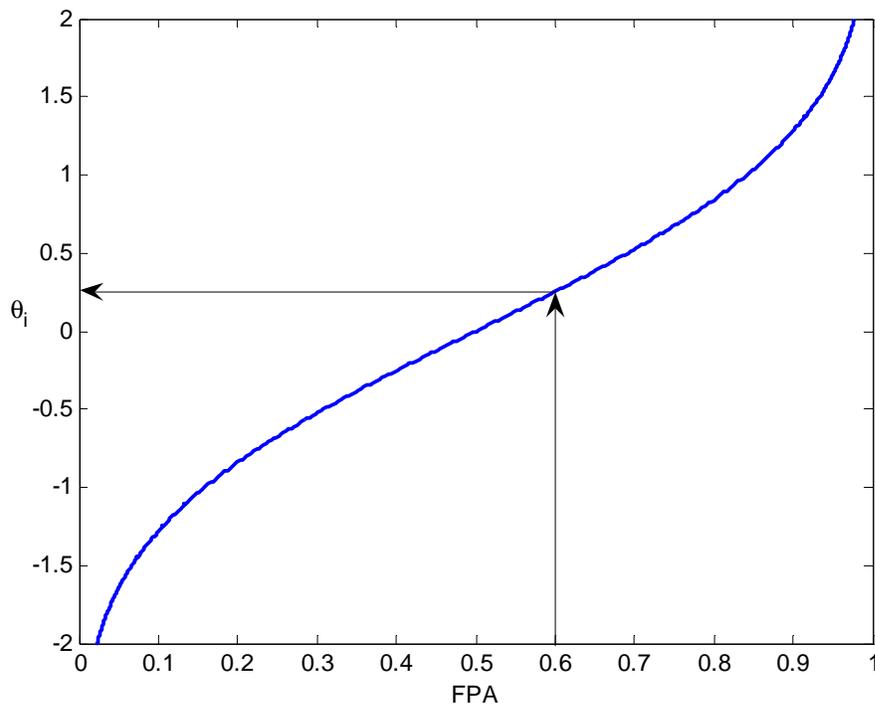


Figura 3-8 - Inversa FPA para uma distribuição normal padrão.

A Figura 3-8 ilustra uma inversa FPA para uma distribuição normal com média 0 e desvio-padrão 1. A amostragem de θ_i foi realizada com um valor da

variável uniforme de 0,60, que resultou em um valor de 0,2533 para θ_i . O cálculo de θ_i foi realizado usando a inversa FPA da curva normal:

$$x = F^{-1}(p | \mu, \sigma) = \{x : F(x | \mu, \sigma) = p\} \quad (\text{Eq. 3.38})$$

onde

$$p = F(x | \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \quad (\text{Eq. 3.39})$$

e, para o exemplo apresentado,

$$p = 0,60, \mu = 1 \text{ e } \sigma = 1.$$

A curva apresentada é contínua, porém na prática a obtenção do valor de θ_i pode ser realizada através de uma interpolação simples entre os pontos da FPA discreta aproximada.

Em aplicações práticas, a escolha do intervalo $[\theta_{i1}, \theta_{im}]$ deve ser feita cuidadosamente. Um procedimento simples para a validação do intervalo é observar o histograma das amostras de θ_i . Caso estas estejam muito concentradas nos extremos do intervalo, o intervalo deve ser expandido. Em contrapartida, caso os valores estejam muito distantes dos extremos, o intervalo deve ser diminuído, pois caso contrário o Griddy Gibbs torna-se ineficiente visto que a maioria dos w_j seriam zero.

Outra consideração relevante é que o Griddy Gibbs pode ser utilizado juntamente com o Amostrador de Gibbs para obter amostras de vários parâmetros de interesse.

3.3. O Modelo de Volatilidade Estocástica

O modelo que será desenvolvido no presente trabalho foi proposto por Jacquier et. al. (1994), e as equações para a média e para a volatilidade (h_t) de uma série de retornos r_t são:

$$r_t = \beta_0 + \beta_1 x_{1t} + \dots + \beta_p x_{pt} + a_t, \quad a_t = \sqrt{h_t} \varepsilon_t \quad (\text{Eq. 3.40})$$

$$\ln h_t = \alpha_0 + \alpha_1 \ln h_{t-1} + v_t \quad (\text{Eq. 3.41})$$

onde $\{x_{it} | i=1, \dots, p\}$ são variáveis explicativas conhecidas no tempo $t-1$, β_j são parâmetros, $\{\varepsilon_t\}$ são variáveis independentes e de idênticas distribuições de probabilidade (iid) com média 0 e variância 1, $\{v_t\}$ também é um ruído branco com média 0 e variância σ_v^2 , e $\{\varepsilon_t\}$ e $\{v_t\}$ são independentes. A transformação logarítmica é utilizada para assegurar que h_t seja positivo para qualquer t . Além disso, assume-se que $|\alpha_1| < 1$ para que $\ln h_t$ seja estacionário. Caso seja de interesse, um modelo AR de maior ordem pode ser utilizado para $\ln h_t$, o que equivale a utilizar diferenças $\ln h_{t-p}$, onde p refere-se à ordem do modelo AR.

Sejam:

$\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)'$ o vetor de coeficientes da equação da média;

$\boldsymbol{\omega} = (\alpha_0, \alpha_1, \sigma_v^2)'$ o vetor de parâmetros da equação da volatilidade;

$\mathbf{R} = (r_1, \dots, r_n)'$ a coleção dos retornos observáveis;

\mathbf{X} a coleção das variáveis explicativas do modelo;

$\mathbf{H} = (h_1, \dots, h_n)'$ o vetor de volatilidades não-observáveis.

Os parâmetros $\boldsymbol{\beta}$ e $\boldsymbol{\omega}$ são “tradicionais”, enquanto \mathbf{H} é um vetor de variáveis auxiliares. A estimação do modelo via o método de máxima verossimilhança seria complicada, pois a função de verossimilhança é uma mistura sobre a distribuição n-dimensional \mathbf{H} , dada por:

$$f(\mathbf{R} | \mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\omega}) = \int f(\mathbf{R} | \mathbf{X}, \boldsymbol{\beta}, \mathbf{H}) f(\mathbf{H} | \boldsymbol{\omega}) d\mathbf{H}$$

Contudo, sobre a ótica Bayesiana, o vetor de volatilidade \mathbf{H} consiste em parâmetros segmentados. Condicionando em \mathbf{H} , pode-se focar nas funções de distribuição de probabilidades $f(\mathbf{R}|\mathbf{H},\boldsymbol{\beta})$, $f(\mathbf{H}|\boldsymbol{\omega})$, e a distribuição priori $p(\boldsymbol{\beta},\boldsymbol{\omega})$. Assume-se que a distribuição a priori pode ser particionada em $p(\boldsymbol{\beta},\boldsymbol{\omega}) = p(\boldsymbol{\beta})p(\boldsymbol{\omega})$, ou seja, que as distribuições a priori das equações da média e da volatilidade são independentes. Assim, uma abordagem pelo Amostrador de Gibbs para estimar a volatilidade estocástica nas equações envolve tirar amostras das seguintes distribuições condicionais *a posteriori*:

$$f(\boldsymbol{\beta}|\mathbf{R},\mathbf{X},\mathbf{H},\boldsymbol{\omega}), f(\mathbf{H}|\mathbf{R},\mathbf{X},\boldsymbol{\beta},\boldsymbol{\omega}), f(\boldsymbol{\omega}|\mathbf{R},\mathbf{X},\boldsymbol{\beta},\mathbf{H}),$$

sendo que, na convergência do algoritmo, a cadeia estacionária resultante será equivalente a amostrar diretamente da distribuição conjunta *a posteriori* $f(\boldsymbol{\beta},\mathbf{H},\boldsymbol{\omega}|\mathbf{X})$.

4. CONSTRUÇÃO DO MODELO

O objetivo desta seção é desenvolver um modelo de séries temporais para o índice IBOVESPA que considere a volatilidade como uma variável que evolui no decorrer do tempo. Para tal, dois modelos serão realizados: GARCH e MVE (Modelo de Volatilidade Estocástica). O intuito é verificar se o modelo de volatilidade estocástica apresenta um desempenho melhor que o modelo GARCH em termos de descrição e previsão da série de retornos do IBOVESPA. Há vários tipos de modelos de volatilidade estocástica, e o que será desenvolvido nesta seção é a formulação proposta por Jacquier et. al. (1994).

A proposta por um modelo MVE surge como um aprimoramento dos modelos ARCH-GARCH, por considerar termos estocásticos tanto para a equação da média quanto da volatilidade dos retornos. Considerando um MVE simples genérico:

$$r_t = \beta_0 + a_t, \quad a_t = \sqrt{h_t} \varepsilon_t, \quad (\text{Eq. 4.1})$$

$$\ln h_t = \alpha_0 + \alpha_1 \ln h_{t-1} + \nu_t, \quad (\text{Eq. 4.2})$$

onde ε_t e ν_t são ruídos brancos independentes, tais que $\varepsilon_t \sim N(0,1)$ e $\nu_t \sim N(0, \sigma_\nu^2)$, r_t é o retorno do instante t , h_t é a volatilidade do instante t e $\beta_0, \alpha_0, \alpha_1, \sigma_\nu^2$ são os parâmetros do modelo.

Neste caso, pode-se notar que a volatilidade possui um termo estocástico ν_t , que não está presente nos modelos ARCH-GARCH. O fato de os modelos MVE apresentarem termos de erro tanto para a equação da média quanto da volatilidade surge como uma consideração natural de que ambas variáveis são processos aleatórios, ou seja, apesar de serem dependentes de outras variáveis (inclusive de seus próprios valores passados), sempre haverá uma componente de incerteza, proveniente da sua característica estocástica.

A série de retornos que será modelada é uma série de valores diários do IBOVESPA de 2-jan-2002 a 28-dez-2007, num total de 1484 observações. Na Figura 4-1 são ilustradas as cotações e os log retornos da série:

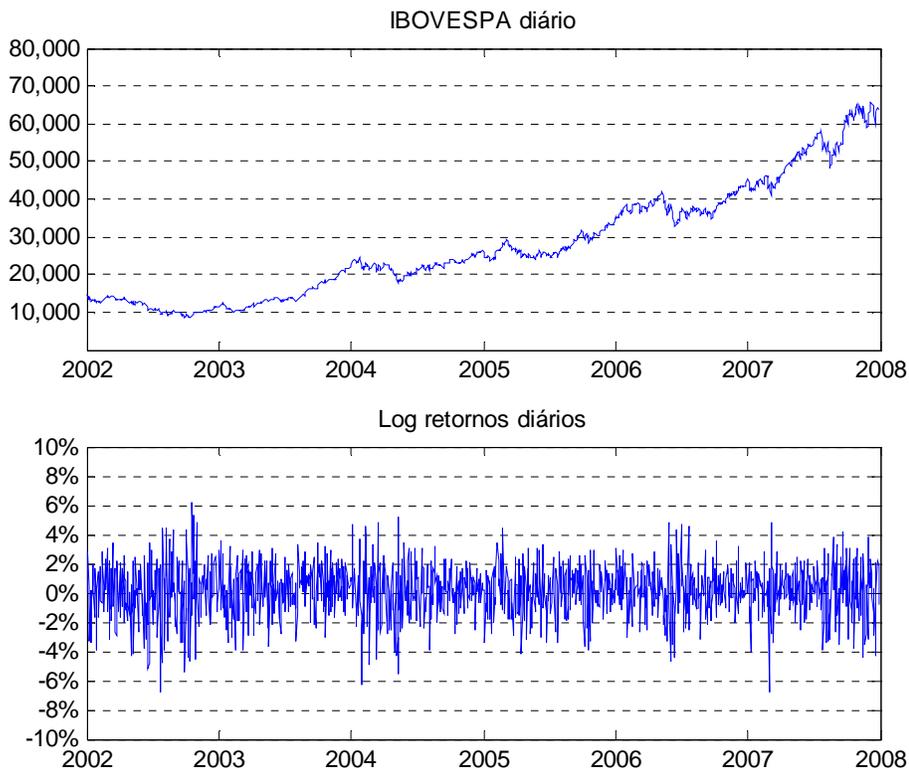


Figura 4-1 - Evolução das cotações diárias do IBOVESPA e os log retornos da série.

É interessante observar o histograma e o gráfico PPN (Normal Probability Plot) para visualizar algumas características da série em estudo na Figura 4-2 e Figura 4-3, respectivamente:

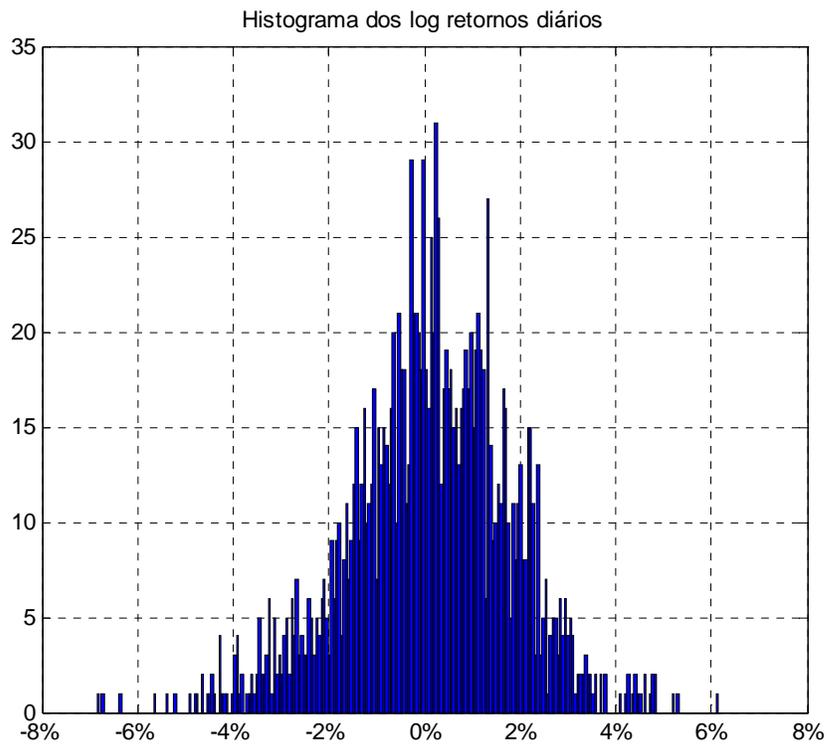


Figura 4-2 - Histograma dos log retornos do IBOVESPA.

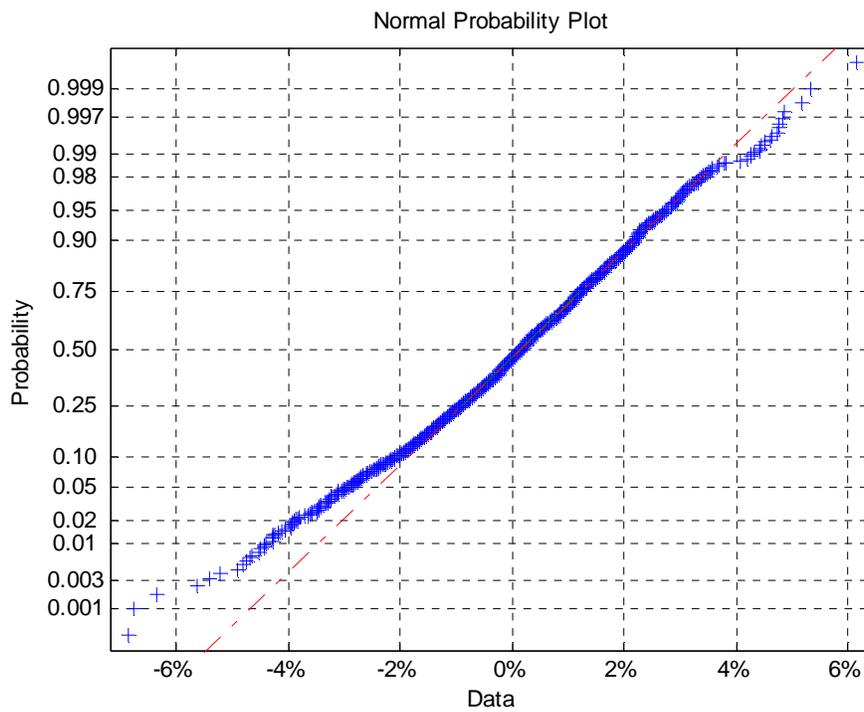


Figura 4-3 - PPN dos log retornos do IBOVESPA.

Os gráficos mostram a característica de caudas pesadas nos retornos, o que é chamado de excesso de curtose. Este fato pode ser claramente observado no gráfico PPN, onde há diversos pontos (principalmente em retornos negativos) fora da reta, indicativo da não-normalidade dos dados.

As principais estatísticas da série são apresentadas na Tabela 4-1:

Estatística	Valor
Média	0,1028
Mediana	0,1600
Desvio padrão	1,7156
Mínimo	-6,8565
Máximo	6,1503
Assimetria	-0,2920
Curtose	3,6792

Tabela 4-1 - Estatísticas descritivas da série de log retornos do IBOVESPA.

O primeiro passo para a resolução do problema é a proposição de um modelo GARCH para a série de retornos do IBOVESPA. O modelo que será utilizado é o GARCH(1,1) dado por:

$$r_t = \mu + a_t, \quad a_t = \sqrt{h_t} \varepsilon_t, \quad (\text{Eq. 4.3})$$

$$h_t = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 h_{t-1}. \quad (\text{Eq. 4.4})$$

Como visto anteriormente, os modelos GARCH podem ser estimados via o método de máxima verossimilhança condicional. Supondo normalidade dos ε_t , pode-se derivar a função de log-verossimilhança, condicionada às m primeiras observações, dada por:

$$\ell(\alpha, \beta | x_1, x_2, \dots, x_m) \propto -\frac{1}{2} \sum_{t=m+1}^n \ln(h_t) - \frac{1}{2} \sum_{t=m+1}^n \frac{x_t^2}{h_t}. \quad (\text{Eq. 4.5})$$

Bollerslev (1986) utiliza $h_j = \hat{\sigma}^2$, $j = 1, \dots, s$, onde $\hat{\sigma}^2 = \sum_{t=1}^n X_t^2 / n$, e n é o número de observações. Assim, as estimativas dos parâmetros são obtidas através de métodos de maximização da função $\ell(\alpha, \beta | x_1, \dots, x_m)$, tais como Newton-Raphson,

Scoring, Gauss-Newton, etc. Para a maximização da função e estimação dos parâmetros, utilizou-se o *software* matemático MATLAB.

A saída da estimação do modelo GARCH(1,1) e a maximização da função de verossimilhança condicional é apresentada na Tabela 4-2:

Parâmetro	Valor	Erro Padrão	<i>t-stat</i>
μ	0,14134	0,043872	3,2216
α_0	0,083756	0,031805	2,6334
α_1	0,92357	0,020105	45,9380
β_1	0,047729	0,011612	4,1103

Valor da log-verossimilhança: -2868,17

Tabela 4-2 - Estimação do modelo GARCH(1,1) para a série de retornos do IBOVESPA.

Assim, o modelo torna-se:

$$r_t = 0,14134 + a_t, \quad a_t = \sqrt{h_t} \varepsilon_t,$$

$$h_t = 0,083756 + 0,92357a_{t-1}^2 + 0,0477295h_{t-1}.$$

Para verificar a adequação ou não do modelo GARCH ajustado, os resíduos do modelo devem ser analisados e testes estatísticos devem ser feitos. A verificação do modelo ajustado será realizada juntamente com a verificação do modelo de volatilidade estocástica.

4.1. Etapas da construção do modelo

A proposta de um modelo de volatilidade estocástica para a série de retornos do IBOVESPA não foi feita ao acaso. Para a escolha de um modelo a ser ajustado, os principais fatores que foram considerados são:

1. Necessidade de considerar a variância da série como condicional e que evolui com o tempo;

2. Constatação (através da análise das estatísticas amostrais e histograma dos dados) da não-normalidade dos dados de retornos, evidenciado pela presença de caudas pesadas no histograma e pelo comportamento do PPN;
3. Observação, no gráfico dos retornos, que as variações da série aparecem em grupos de maior ou menor volatilidade (como é de senso comum na análise de séries financeiras);
4. Tentativa de construir um modelo que obtenha um desempenho melhor que os modelos ARCH-GARCH. O desempenho pode ser medido principalmente pela análise dos resíduos do modelo ajustado.

Para um melhor entendimento das etapas que devem ser realizadas, e da ordem em que ocorrerão, é apresentado um fluxograma do processo de construção do modelo, conforme a Figura 4-4:

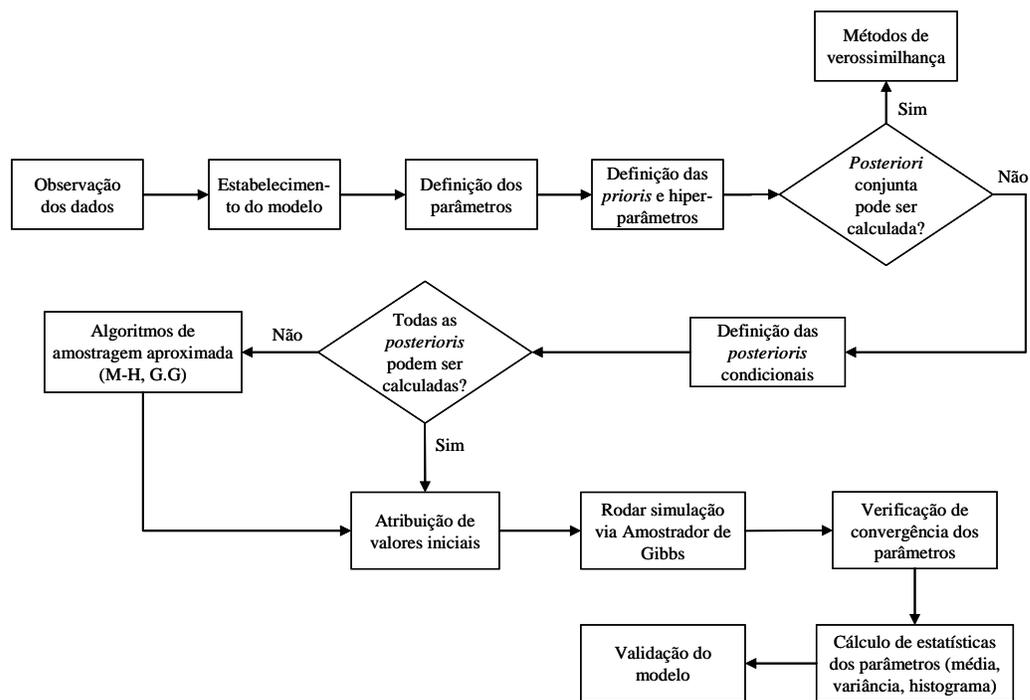


Figura 4-4 - Fluxograma do processo de construção do modelo de volatilidade estocástica.

4.1.1 O Modelo Proposto

O modelo proposto trata um conjunto de dados, \mathbf{R} , como um vetor gerado através de um modelo $p(\mathbf{R} | \mathbf{h})$, onde \mathbf{h} é um vetor de volatilidades. Cada ponto r_t tem uma variância h_t , que é dependente do tempo. As volatilidades \mathbf{h} não são observáveis, e assume-se que são geradas a partir de uma função de probabilidade $p(\mathbf{h} | \boldsymbol{\omega})$. As equações para a média e volatilidade foram apresentadas anteriormente (Eq. 4.1 e Eq. 4.2). Assim, assume-se que o logaritmo da variância condicional da série \mathbf{R} segue um processo AR(1).

Especificamente, o modelo de volatilidade estocástica que será ajustado para a série de retornos diários do IBOVESPA é similar ao proposto por Jacquier et. al. (1994), que utilizou a inferência Bayesiana para estimar os parâmetros do modelo através da simulação via o Amostrador de Gibbs das seguintes distribuições de probabilidades condicionais:

$$f(\boldsymbol{\beta} | \mathbf{R}, \mathbf{X}, \mathbf{H}, \boldsymbol{\omega}) \quad f(\mathbf{H} | \mathbf{R}, \mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\omega}) \quad f(\boldsymbol{\omega} | \mathbf{R}, \mathbf{X}, \boldsymbol{\beta}, \mathbf{H})$$

onde $\boldsymbol{\beta}$ é o vetor de parâmetros da equação da média, $\boldsymbol{\omega}$ é o vetor de parâmetros da equação da volatilidade e \mathbf{H} é um vetor de volatilidades que foi incorporado ao modelo como uma variável auxiliar, baseado no conceito de segmentação (“data augmentation”).

Cabe ressaltar que a simulação via métodos MCMC é necessária, pois a função de verossimilhança para o processo é definida como uma integral de dimensão n para a distribuição de \mathbf{H} :

$$f(\mathbf{R} | \mathbf{X}, \boldsymbol{\beta}, \boldsymbol{\omega}) = \int f(\mathbf{R} | \mathbf{X}, \boldsymbol{\beta}, \mathbf{H}) f(\mathbf{H} | \boldsymbol{\omega}) d\mathbf{H} ,$$

onde n é o número de dados da série.

Devido à complexidade da função de verossimilhança, não é possível derivar analiticamente a mesma e, portanto, métodos de máxima verossimilhança não podem ser aplicados. Assim, uma alternativa para realizar a estimação dos parâmetros do modelo é simular uma cadeia que, na sua convergência, será aproximada à função *posteriori* conjunta dos parâmetros $\boldsymbol{\beta}$, \mathbf{H} e $\boldsymbol{\omega}$.

4.1.2 Escolha das distribuições *a priori* dos parâmetros

Como visto anteriormente, a inferência Bayesiana permite a consideração de informações prévias de um processo, através das distribuições *a priori* dos parâmetros. Mais importante do que atribuir valores prováveis a um parâmetro, a distribuição *a priori* é responsável por, em muitos casos, moldar o comportamento da distribuição *a posteriori* (o que acontece nas distribuições *a priori* conjugadas, vistas anteriormente). Assim, é necessário escolher as distribuições *a priori* com bom senso, e sempre que possível considerar informações de pessoas mais experientes sobre o processo em questão.

Por não haver muito interesse em incluir conhecimentos específicos sobre os retornos do IBOVESPA e pelo desejo de minimizar os impactos das informações *a priori* na análise de modo a traduzir adequadamente a incerteza existente, foram escolhidas distribuições *prioris* não muito informativas, com grandes valores para as variâncias das distribuições, e com base nas distribuições comumente utilizadas na literatura.

É importante observar ao final do ajustamento do modelo a influência dos valores dos hiperparâmetros (isto é, os parâmetros dos parâmetros do modelo) sobre os resultados obtidos. Em geral, um simples confronto da função de distribuição de probabilidades *a posteriori* com a sua respectiva *priori* pode mostrar a grande ou pequena influência das informações prévias acrescentadas na análise.

As distribuições *a priori* que serão utilizadas no modelo de volatilidade estocástica proposto são:

$$\beta \sim N(\beta_0, A_0), \quad \alpha \sim N(\alpha_0, C_0), \quad \frac{m \times \lambda}{\sigma_v^2} \sim \chi_5^2$$

onde

- β segue uma distribuição normal com hiperparâmetros $\beta_0 = 0$ e $A_0 = 9$;
- α segue uma distribuição normal multivariada com vetor de média

$$\alpha_0 = \begin{bmatrix} 0,4 \\ 0,8 \end{bmatrix} \text{ e matriz de covariância } C_0 = \begin{bmatrix} 0,09 & 0 \\ 0 & 0,04 \end{bmatrix};$$

- σ_v^2 segue uma distribuição inversa chi-quadrada com $m = 5$ graus de liberdade e $\lambda = 0,2$.

4.1.3 Estimação das *posteriors* condicionais

1. Estimação da *posteriori* de β

Dado \mathbf{H} , a equação para a média é uma regressão linear não-homogênea.

Dividindo a equação por $\sqrt{h_t}$, a equação pode ser reescrita por

$$r_{0,t} = \mathbf{x}_{0,t} \boldsymbol{\beta} + \varepsilon_t, \quad t = 1, \dots, n \quad (\text{Eq. 4.6})$$

onde $r_{0,t} = r_t / \sqrt{h_t}$ e $\mathbf{x}_{0,t} = \mathbf{x}_t / \sqrt{h_t}$, com $\mathbf{x}_t = (1)'$, que é o vetor de variáveis explicativas da equação da média, que no modelo proposto representa apenas uma constante. Supondo que a distribuição *priori* de $\boldsymbol{\beta}$ seja uma normal multivariada com vetor de média $\boldsymbol{\beta}_0$ e matriz de covariância \mathbf{A}_0 , então a distribuição *posteriori* de $\boldsymbol{\beta}$ também será multivariada normal com média $\boldsymbol{\beta}_*$ e matriz de covariância \mathbf{A}_* de modo que (Resultado 1):

$$\mathbf{A}_*^{-1} = \sum_{t=1}^n \mathbf{x}_{0,t} \mathbf{x}'_{0,t} + \mathbf{A}_0^{-1}, \quad \boldsymbol{\beta}_* = \mathbf{A}_* \left(\sum_{t=1}^n \mathbf{x}_{0,t} r_{0,t} + \mathbf{A}_0^{-1} \boldsymbol{\beta}_0 \right), \quad (\text{Eq. 4.7})$$

onde entende-se que a soma começa com $p+1$ se r_{t-p} for o maior lag de retorno utilizado como variável explicativa da equação da média. No modelo proposto, não há termos auto-regressivos para a equação da média, apenas uma constante, de modo a simplificar o problema, e considerando que a volatilidade como a principal variável de interesse.

2. Estimação da *posteriori* de h_t

O vetor de volatilidades \mathbf{H} é amostrado elemento por elemento. A distribuição *posteriori* condicional necessária é $f(h_t | \mathbf{R}, \mathbf{X}, \mathbf{H}_{-t}, \boldsymbol{\beta}, \boldsymbol{\omega})$, sendo proporcional ao produto da distribuição normal de a_t e da distribuição log-normal da volatilidade h :

$$\begin{aligned}
f(h_t | \mathbf{R}, \mathbf{X}, \mathbf{H}_{-t}, \boldsymbol{\beta}, \boldsymbol{\omega}) & \\
& \propto f(a_t | h_t, r_t, \mathbf{x}_t, \boldsymbol{\beta}) f(h_t | h_{t-1}, \boldsymbol{\omega}) f(h_{t+1} | h_t, \boldsymbol{\omega}) \\
& \propto h_t^{-0.5} \exp\left[-(r_t - \mathbf{x}_t' \boldsymbol{\beta})^2 / (2h_t)\right] h_t^{-1} \exp\left[-(\ln h_t - \mu_t)^2 / (2\sigma^2)\right] \\
& \propto h_t^{-1.5} \exp\left[-(r_t - \mathbf{x}_t' \boldsymbol{\beta})^2 / (2h_t) - (\ln h_t - \mu_t)^2 / (2\sigma^2)\right], \quad (\text{Eq. 4.8})
\end{aligned}$$

onde $\mu_t = [\alpha_0(1 - \alpha_1) + \alpha_1(\ln h_{t+1} + \ln h_{t-1})] / (1 + \alpha_1^2)$ e $\sigma^2 = \sigma_v^2 / (1 + \alpha_1^2)$. Para a derivação da condicional *posteriori* de h_t , as seguintes propriedades foram utilizadas, assim como descrito por Jacquier et. al. (1994) e Tsay (2002):

1. $a_t | h_t \sim N(0, h_t)$;
2. $\ln h_t | \ln h_{t-1} \sim N(\alpha_0 + \alpha_1 \ln h_{t-1}, \sigma_v^2)$;
3. $\ln h_{t+1} | \ln h_t \sim N(\alpha_0 + \alpha_1 \ln h_t, \sigma_v^2)$;
4. $d \ln h_t = h_t^{-1} dh_t$, onde d denota diferenciação;
5. A igualdade

$$(x - a)^2 A + (x - b)^2 C = (x - c)^2 (A + C) + (a - b)^2 AC / (A + C)$$

onde $c = (Aa + Cb) / (A + C)$ com $A + C \neq 0$. Esta igualdade é uma versão escalar do Lema 1 de Box e Tiao (1973, pág. 418). Na presente aplicação, temos:

$$\begin{aligned}
A &= 1, \\
a &= \alpha_0 + \ln h_{t-1}, \\
C &= \alpha_1^2, \\
b &= (\ln h_{t+1} - \alpha_0) / \alpha_1.
\end{aligned}$$

O termo $(a - b)^2 AC / (A + C)$ é independente da variável h_t e, portanto, é integrado para fora da derivação da distribuição *posteriori* condicional.

Para a amostragem de h_t , Jacquier et. al. (1994) utilizou o algoritmo Metropolis-Hastings. O algoritmo que será utilizado neste trabalho será o Griddy Gibbs, assim como foi utilizado em Tsay (2002).

Para tirar uma amostra de h_t via Griddy Gibbs, primeiramente devem ser escolhidos os parâmetros do algoritmo: o intervalo de valores e o número de pontos de gradação (“grid points”). Assim, o algoritmo será realizado da seguinte maneira:

- Intervalo de dados: $]0; 1,5 s^2]$, onde s^2 é a variância amostral incondicional dos retornos;
- Número de pontos de graduação: $m = 500$.

O Griddy Gibbs procede da seguinte forma:

1. Avaliar a função de densidade de probabilidade da distribuição derivada, proporcional à $f(h_t | \mathbf{R}, \mathbf{X}, \mathbf{H}_{-t}, \boldsymbol{\beta}, \boldsymbol{\omega})$, para cada um dos pontos do “grid”, obtendo-se $w_j = f(h_{t_j} | \mathbf{R}, \mathbf{X}, \mathbf{H}_{-t}, \boldsymbol{\beta}, \boldsymbol{\omega})$ para cada $j = 1, \dots, m$.
2. Normalizar cada w_j , isto é, calcular $\frac{w_j}{\sum_{i=1}^m w_i}$, e utilizar estes pesos para obter uma aproximação da inversa da função de probabilidade acumulada (FPA) de $f(h_t | \mathbf{R}, \mathbf{X}, \mathbf{H}_{-t}, \boldsymbol{\beta}, \boldsymbol{\omega})$.
3. Tirar uma amostra de uma variável uniforme no intervalo $]0,1]$, e transformar a observação via a inversa FPA, para obter uma amostra aleatória de h_t . Isto é feito interpolando linearmente a variável aleatória uniforme amostrada na inversa FPA.

Como apresentado anteriormente, é necessário checar a adequação do intervalo escolhido para h_t . Isto será realizado através da análise de histogramas para os valores amostrados para alguns elementos h quaisquer.

Uma observação importante é que a expressão da distribuição *posteriori* condicional de h_t é válida para $1 < t < n$, onde n é o tamanho da amostra. Para os dois pontos extremos, h_1 e h_n , algumas modificações são necessárias. Uma alternativa plausível é empregar uma previsão de h_{n+1} e uma previsão retrógrada de h_0 , para poder continuar aplicando a fórmula normalmente. Como h_n é a variável de interesse, pode-se fazer uma previsão de $\ln h_{n+1}$ partindo de $n-1$ de 2 passos:

$$\hat{\phi}_{n-1}(2) = \alpha_0 + \alpha_1(\alpha_0 + \alpha_1 \ln h_{n-1}), \quad (\text{Eq. 4.9})$$

onde $\hat{\phi}_{n-1}$ é a previsão de $\ln h_{n-1}$.

A previsão retrógrada de h_0 é baseado na reversibilidade temporal do modelo

$$(\ln h_t - \eta) = \alpha_1 (\ln h_{t-1} - \eta) + v_t,$$

onde $\eta = \alpha_0 / (1 - \alpha_1)$ e $|\alpha_1| < 1$. O modelo da série reversa é

$$(\ln h_t - \eta) = \alpha_1 (\ln h_{t+1} - \eta) + v_t^*, \quad (\text{Eq. 4.10})$$

onde $\{v_t^*\}$ também é um ruído branco gaussiano com média zero e variância σ_v^2 .

Conseqüentemente, a previsão retrógrada de 2 passos de $\ln h_0$ em $t = 2$ é

$$\hat{\phi}_2(-2) = \alpha_1^2 (\ln h_2 - \eta). \quad (\text{Eq. 4.11})$$

Outra observação importante é que os valores iniciais de h_t para a simulação via Amostrador de Gibbs podem ser obtidos através de um ajustamento de um modelo ARCH-GARCH. Assim, os valores obtidos no modelo GARCH(1,1) ajustado serão utilizados no modelo de volatilidade estocástica como valores iniciais da simulação.

3. Estimação da *posteriori* de ω

Para realizar a amostragem de ω , este vetor é particionado em $\alpha = (\alpha_0, \alpha_1)'$ e σ_v^2 . Assim a *priori* de ω também deve ser particionada adequadamente. Assumindo que as *prioris* dos parâmetros particionados são independentes é possível escrever $p(\omega) = p(\alpha)p(\sigma_v^2)$. Assim, as distribuições *posteriores* condicionais podem ser obtidas:

- $f(\alpha | \mathbf{R}, \mathbf{X}, \mathbf{H}, \beta, \sigma_v^2) = f(\alpha | \mathbf{H}, \sigma_v^2)$: Dado \mathbf{H} , $\ln h_t$ segue um modelo AR(1). Assim, se a distribuição *priori* de α é multivariada normal com média α_0 e matriz de covariância \mathbf{C}_0 , então $f(\alpha | \mathbf{H}, \sigma_v^2)$ é multivariada normal com média α_* e matriz de covariância \mathbf{C}_* , onde

$$\mathbf{C}_*^{-1} = \frac{\sum_{t=2}^n \mathbf{z}_t \mathbf{z}_t'}{\sigma_v^2} + \mathbf{C}_0^{-1}, \quad \alpha_* = \mathbf{C}_* \left(\frac{\sum_{t=2}^n \mathbf{z}_t \ln h_t}{\sigma_v^2} + \mathbf{C}_0^{-1} \alpha_0 \right), \quad (\text{Eq. 4.12})$$

onde $\mathbf{z}_t = (1, \ln h_{t-1})'$.

- $f(\sigma_v^2 | \mathbf{R}, \mathbf{X}, \mathbf{H}, \boldsymbol{\beta}, \boldsymbol{\alpha}) = f(\sigma_v^2 | \mathbf{H}, \boldsymbol{\alpha})$: Dado \mathbf{H} e $\boldsymbol{\alpha}$, é possível calcular $v_t = \ln h_t - \alpha_0 - \alpha_1 \ln h_{t-1}$ para $t = 2, \dots, n$. Portanto, se a distribuição *priori* de σ_v^2 for $(m\lambda)/\sigma_v^2 \sim \chi_m^2$, então a distribuição *posteriori* condicional de σ_v^2 é uma inversa chi-quadrada com $m + n - 1$ graus de liberdade, isto é

$$\frac{m\lambda + \sum_{t=2}^n v_t^2}{\sigma_v^2} \sim \chi_{m+n-1}^2. \quad (\text{Eq. 4.13})$$

4.1.4 Simulação via Amostrador de Gibbs

Com as distribuições *posteriors* condicionais definidas, o próximo passo é a implementação de uma simulação via o Amostrador de Gibbs. O objetivo do algoritmo é, a partir das distribuições condicionais dos parâmetros, simular uma cadeia que, ao chegar na estacionariedade, será equivalente à distribuição *posteriori* conjunta dos parâmetros do modelo

$$f(\boldsymbol{\beta}, \mathbf{H}, \boldsymbol{\omega} | \mathbf{R}, \mathbf{X}).$$

Com a distribuição *posteriori* conjunta, é possível estimar os valores dos parâmetros que melhor explicam a série de retornos. Assim, ao término da simulação via o Amostrador de Gibbs, serão calculadas estatísticas descritivas dos parâmetros, bem como serão construídos gráficos auxiliares para uma série de análises, dentre elas a verificação da convergência do algoritmo e a influência das informações *a priori* nos resultados finais.

A Figura 4-5 representa a simulação do algoritmo de Gibbs para o problema proposto, apresentando os parâmetros a serem amostrados:

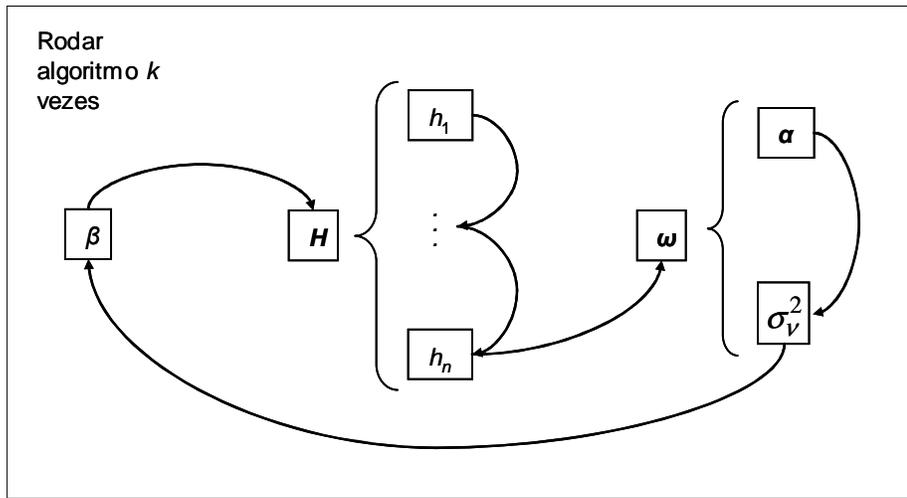


Figura 4-5 - Diagrama da simulação do Amostrador de Gibbs.

A Figura 4-5 mostra as etapas do algoritmo na amostragem dos parâmetros β , \mathbf{H} (que é segmentado em n elementos h_t , representantes da volatilidade no instante t) e ω (que também é particionado em dois parâmetros auxiliares α e σ_v^2). O objetivo do algoritmo é rodar a cadeia um número k de vezes suficientemente grande para que os valores amostrados, na convergência, sejam equivalentes à uma amostragem diretamente da distribuição *posteriori* conjunta dos parâmetros.

Para que a simulação se inicie é necessária a definição de valores iniciais para os parâmetros. Em geral, tais valores não influenciam nos resultados do algoritmo visto que, na convergência do mesmo, a dependência dos valores dos parâmetros em relação aos seus valores iniciais será desprezível.

Os valores iniciais que serão utilizados para a simulação são:

$$\sigma_v^2 = 0,5 \text{ e } \beta = 0,103 \text{ (média amostral dos retornos),}$$

e os valores iniciais para $\{h_t\}$ são os valores resultantes do modelo GARCH(1,1) ajustado anteriormente.

Para a simulação do problema de inferência o autor escreveu um programa computacional no *software* MATLAB (Anexo A). Os resultados obtidos na simulação serão analisados posteriormente, pois é necessário verificar se o número de iterações k é suficiente para que a cadeia simulada atinja a estacionariedade.

Assim, além das estimativas dos parâmetros, a programação gerou uma série de gráficos que serão utilizados para a verificação da convergência do algoritmo e para o diagnóstico e validação do modelo ajustado.

4.1.5 Verificação da convergência do algoritmo

A seguir, é analisada a condição de convergência, isto é, se o número k de vezes que a simulação foi iterada é suficiente para atingir a condição de estacionariedade da cadeia. Isto pode ser observado através de um gráfico que mostre os valores amostrados para os parâmetros do modelo ao longo das simulações. A condição de convergência pode ser aceita caso a curva apresente uma característica estacionária (isto é, sem inclinações e tendências). A Figura 4-6 e a Figura 4-7 apresentam os gráficos das amostras tiradas dos parâmetros para cada ciclo simulado (devido ao grande número de parâmetros da volatilidade, visto que cada elemento h_t é amostrado individualmente, a convergência de apenas alguns deles será verificada):

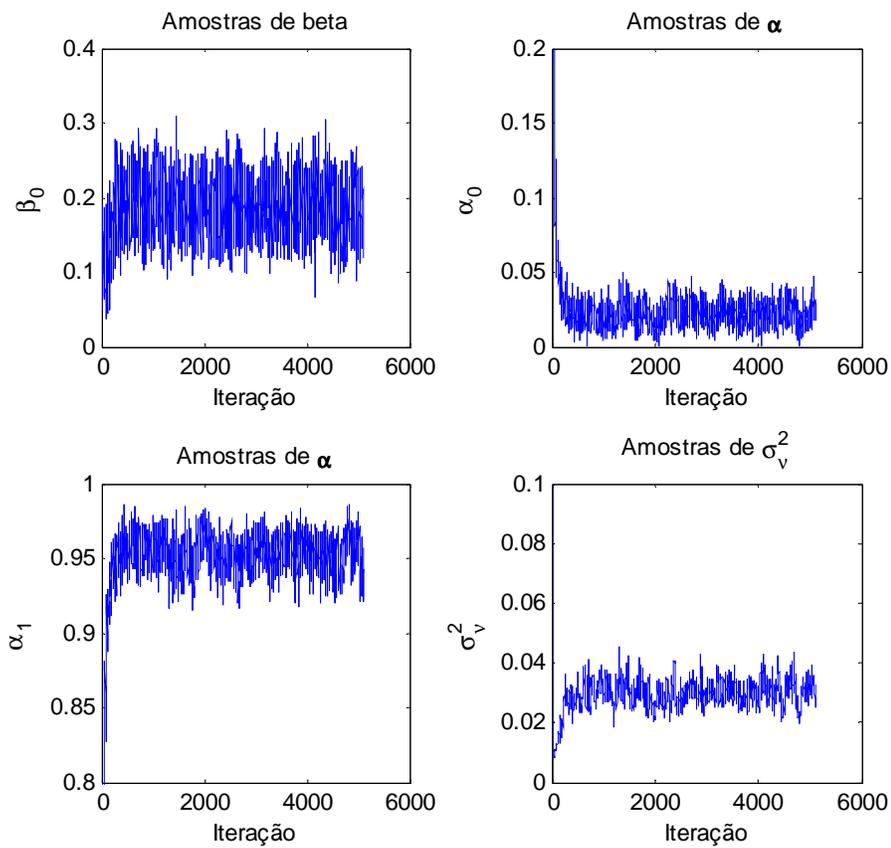


Figura 4-6 - Amostras realizadas dos parâmetros do modelo (parte 1/2).

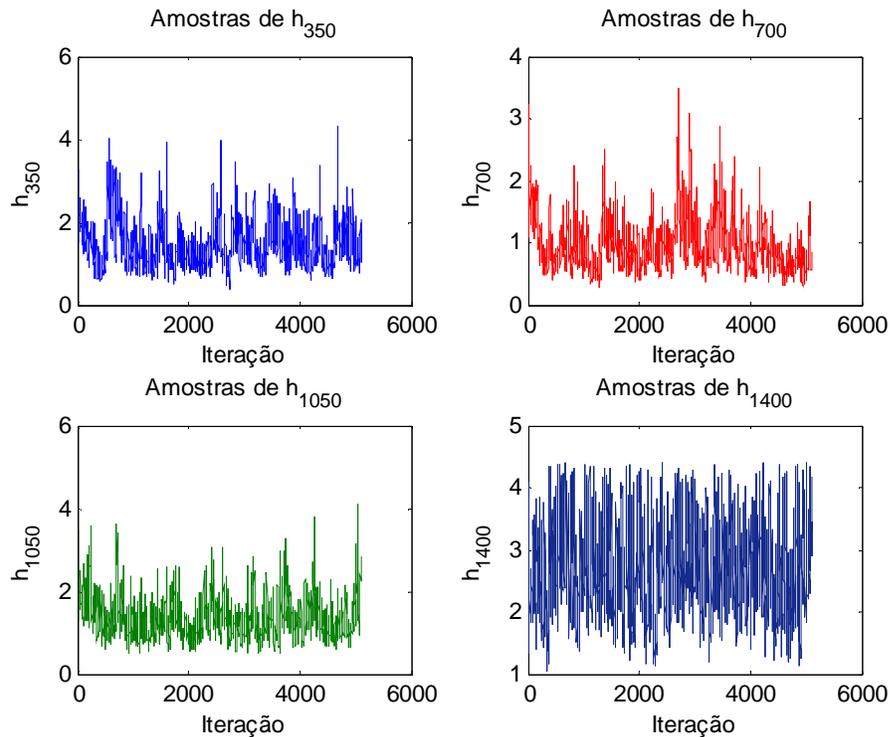


Figura 4-7 - Amostras realizadas dos parâmetros do modelo (parte 2/2).

Os gráficos mostram que, nas primeiras simulações, os valores dos parâmetros mudam drasticamente, o que evidencia a inacurácia dos valores iniciais utilizados. Este padrão de mudança brusca somente não se verifica na amostragem das volatilidades $\{h_t\}$, visto que os seus valores iniciais foram provenientes do modelo GARCH(1,1) ajustado anteriormente. Nos elementos de h escolhidos, observa-se que os valores iniciais são próximos aos valores obtidos na convergência do algoritmo.

A convergência da cadeia também é evidenciada nos gráficos apresentados, visto que as curvas da amostragem ao longo das iterações não apresentam tendências de inclinação ou sazonalidades. Assim, pode-se considerar válidos os resultados obtidos pela simulação via o Amostrador de Gibbs. O que deve ser analisado, agora, é se o modelo ajustado é eficiente na explicação da série temporal dos retornos e das suas volatilidades.

4.1.6 Estimativas dos parâmetros

A programação da inferência Bayesiana para a estimação dos parâmetros do modelo gerou os seguintes resultados:

Parâmetro	beta0	alpha0	alpha1	sigma_v2
Média	0.186689	0.022581	0.952755	0.029743
Desv. Pad.	0.033364	0.007943	0.011685	0.004318

Número de simulações: 5100

Burn-in sample: 100

Número de dados: 1483

Tempo decorrido para realizar GARCH: 5.56 seg.

Tempo decorrido para realizar simulação: 135 min.

Assim, é possível reescrever o modelo ajustado da seguinte forma:

$$r_t = 0,186689 + a_t, \quad a_t = \sqrt{h_t} \varepsilon_t,$$

$$\ln h_t = 0,022581 + 0,952755 \ln h_{t-1} + \sigma_v^2 \eta_t,$$

onde $\eta_t \sim N(0,1)$ é um ruído branco.

As estimativas dos parâmetros foram calculadas com base nas $k - b$ últimas amostras simuladas, onde b é o *burn-in sample*. Isto é realizado pois as primeiras amostras tiradas são fortemente dependentes dos valores iniciais, o que poderia afetar os valores das estimativas.

Antes de realizar o diagnóstico do modelo (isto é, verificar se a série de retornos pode ser efetivamente explicada) é sugerido analisar o grau de relevância das informações *a priori* adicionadas ao modelo. Uma grande influência das informações *a priori* pode ser indicativo de que os resultados são fortemente influenciados pelas informações anteriores à observação amostral, o que não é desejável no problema proposto pois, como dito anteriormente, não há informações precisas sobre o processo em questão. Em suma, é de interesse que os dados “falem por si só”, e que os resultados não tenham grande dependência das informações *a priori*.

Para analisar o grau de informação *a priori* nos resultados do modelo ajustado, é possível analisar as funções de densidade de probabilidade *a priori* e *a posteriori* dos parâmetros, que são apresentadas na Figura 4-8:

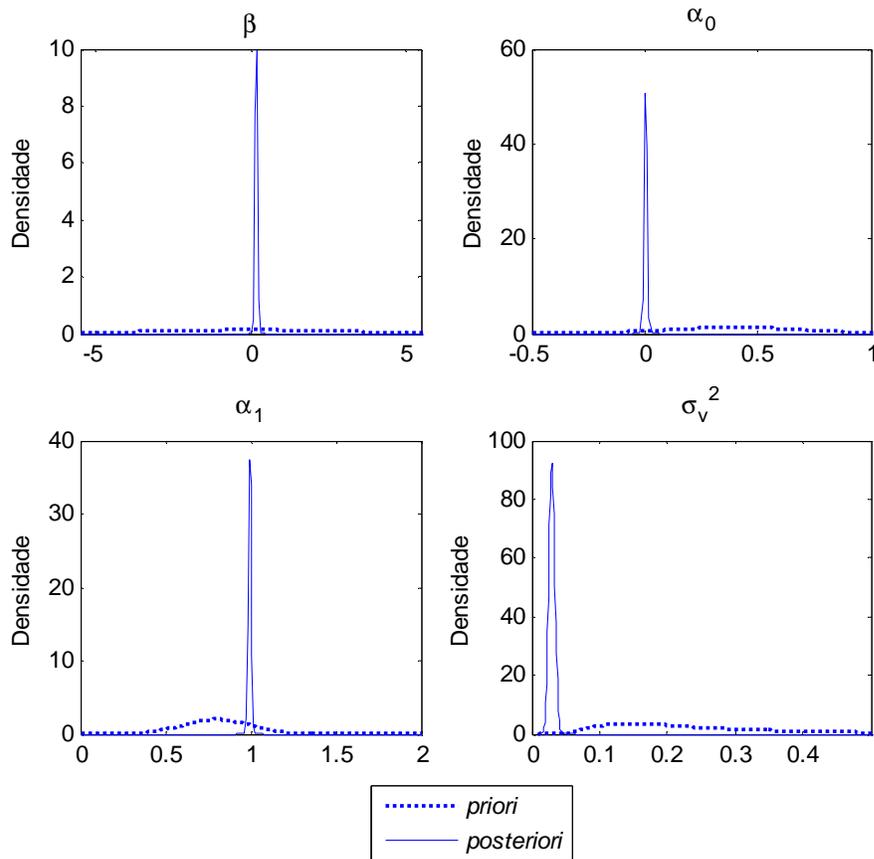


Figura 4-8 - Funções densidade de probabilidade das distribuições *a priori* e *a posteriori* dos parâmetros.

Os gráficos das funções densidade de probabilidade dos parâmetros mostram que as distribuições *a priori* não são informativas para o modelo ajustado. Isto pode ser observado pelo deslocamento das médias entre as curvas, além da notável diferença da amplitude das mesmas (isto é, as curvas apresentam valores de desvio padrão diferentes). Como dito, esta característica já era desejável, pois não há informações precisas em relação à série temporal de retornos do IBOVESPA.

A Figura 4-9 apresenta as funções densidade de probabilidade *a posteriori* dos parâmetros da volatilidade h_t , segundo à Eq. 4.8, para visualização. Do mesmo modo que a Figura 4-7, apenas alguns elementos são ilustrados:

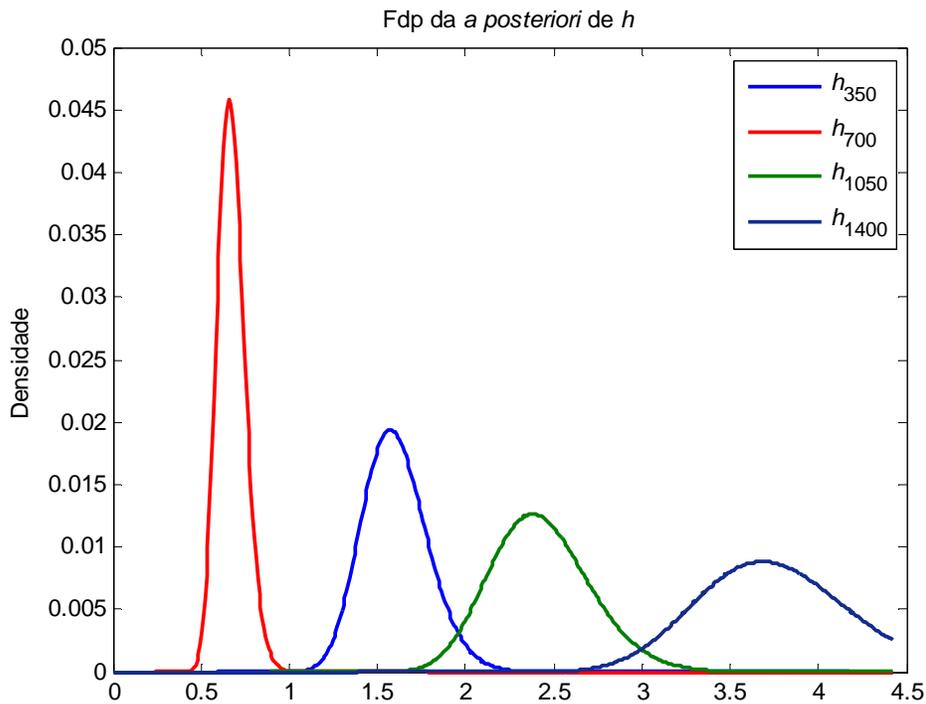


Figura 4-9 – Funções densidade de probabilidade dos parâmetros da volatilidade.

Cada curva apresentada na Figura 4-9 corresponde a uma distribuição de probabilidade da previsão da volatilidade, para cada instante t . Assim, a previsão da volatilidade no modelo MVE gera distribuições de valores, e não apenas valores pontuais de previsão.

4.1.7 Validação e comparação dos modelos ajustados

Os resultados dos modelos ajustados apresentam um comportamento similar, o que pode ser visto na Figura 4-10, que apresenta a volatilidade condicional dos modelos ajustados:

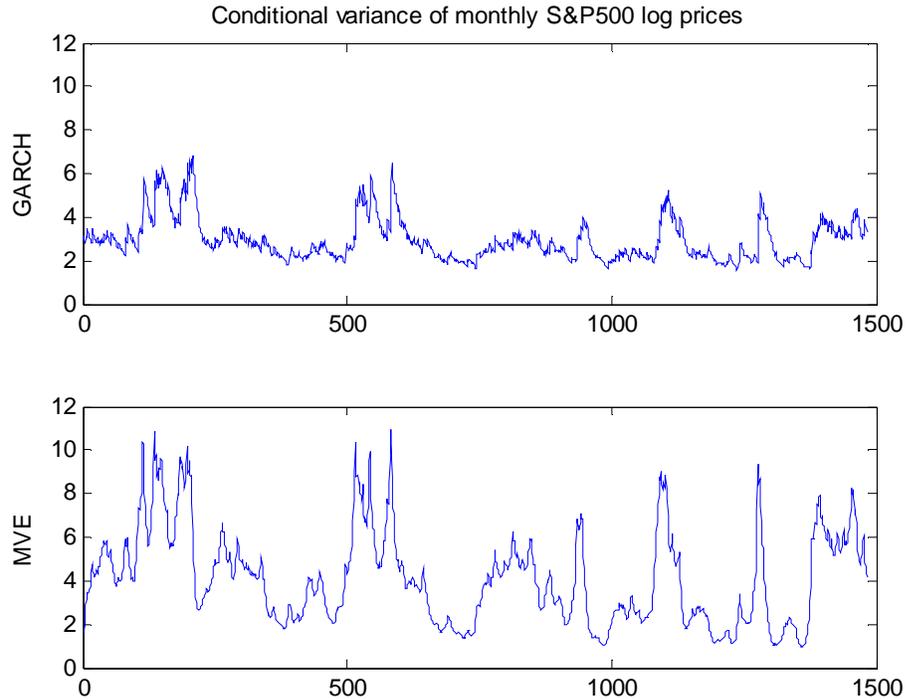


Figura 4-10 - Volatilidade condicional dos modelos ajustados.

Os dois modelos apresentam comportamentos similares, com os mesmos picos e vales de volatilidades. A diferença mais notável é que no MVE a curva apresenta maior amplitude e apresenta previsões maiores do que o GARCH.

Para a validação dos modelos ajustados e verificação de qual deles apresenta um melhor desempenho, dois testes estatísticos serão realizados: o teste de Ljung-Box e o teste de multiplicadores de Lagrange (ML).

O primeiro teste estatístico que será realizado é o teste de Ljung-Box, que testa a hipótese de ruído branco para os resíduos padronizados e para os quadrados dos resíduos padronizados, através de uma medição do grau de correlação dos resíduos para alguns *lags*. Os valores obtidos, com um valor de 5% de significância, são apresentados na Tabela 4-3 e Tabela 4-4:

Teste de Ljung-Box – resíduos padronizados							
GARCH(1,1)				MVE			
Lag (K)	Rejeição de H_0	p-valor	$Q(K)$	Lag (K)	Rejeição de H_0	p-valor	$Q(K)$
10	FALSO	0.5913	8.3851	10	FALSO	0.6554	7.7275
20	FALSO	0.7473	15.4972	20	FALSO	0.3240	22.308

Tabela 4-3 - Teste de Ljung-Box para os resíduos padronizados.

Teste de Ljung-Box – quadrados dos resíduos padronizados							
GARCH(1,1)				MVE			
Lag (K)	Rejeição de H_0	p-valor	$Q(K)$	Lag (K)	Rejeição de H_0	p-valor	$Q(K)$
10	FALSO	0.3883	10.6149	10	FALSO	0.8998	3.0366
20	FALSO	0.3915	21.0959	20	FALSO	0.0635	30.4099

Tabela 4-4 - Teste de Ljung-Box para os quadrados dos resíduos.

Pode-se observar que todos os testes não rejeitaram a hipótese nula de resíduos padronizados aleatórios e, portanto, confirmam o poder explicativo de ambos os modelos.

O segundo teste estatístico que será realizado é o teste ML, que verifica se há persistência de heterocedasticidade condicional na série ajustada. Usualmente, a estatística é realizada para alguns valores de *lag*. Os resultados dos testes são apresentados a seguir:

Teste de multiplicadores de Lagrange							
GARCH(1,1)				MVE			
Lag (K)	Rejeição de H_0	p-valor	T	Lag (K)	Rejeição de H_0	p-valor	T
1	FALSO	0.1927	1.697	1	FALSO	0.1414	2.163
3	VERDADEIRO	0	31.473	3	VERDADEIRO	0	31.4022

Tabela 4-5 - Teste de multiplicadores de Lagrange.

O teste mostra que não há efeitos ARCH remanescentes para o valor de *lag* 1, isto é, a hipótese de homocedasticidade nos elementos correspondentes de *lag* 1 não pode ser rejeitada. Assim, a heterocedasticidade dos resíduos para *lag* 1 foi explicada adequadamente em ambos os modelos. Em contrapartida, para um valor de *lag* 3, os testes ML mostram que ainda há resquícios de heterocedasticidade condicional nos

resíduos de ambos os modelos ajustados. Desse modo, o teste sugere que modelos de maior ordem podem ser testados para uma maior explicação dos resíduos.

Outro aspecto importante a ser considerado é que, como pode ser observado em Jacquier et. al. (1994), Morettin; Tolo (2006) e Tsay (2002), em geral, quanto maior a complexidade de um modelo, menor a capacidade de previsão do mesmo. Isto se justifica pois modelos com um grande número de parâmetros tendem a se ajustar melhor nos dados históricos, mas esse ajuste não contribui com uma melhor capacidade de previsão (o que é chamado de super-ajustamento).

Os resíduos dos modelos são apresentados na Figura 4-11:

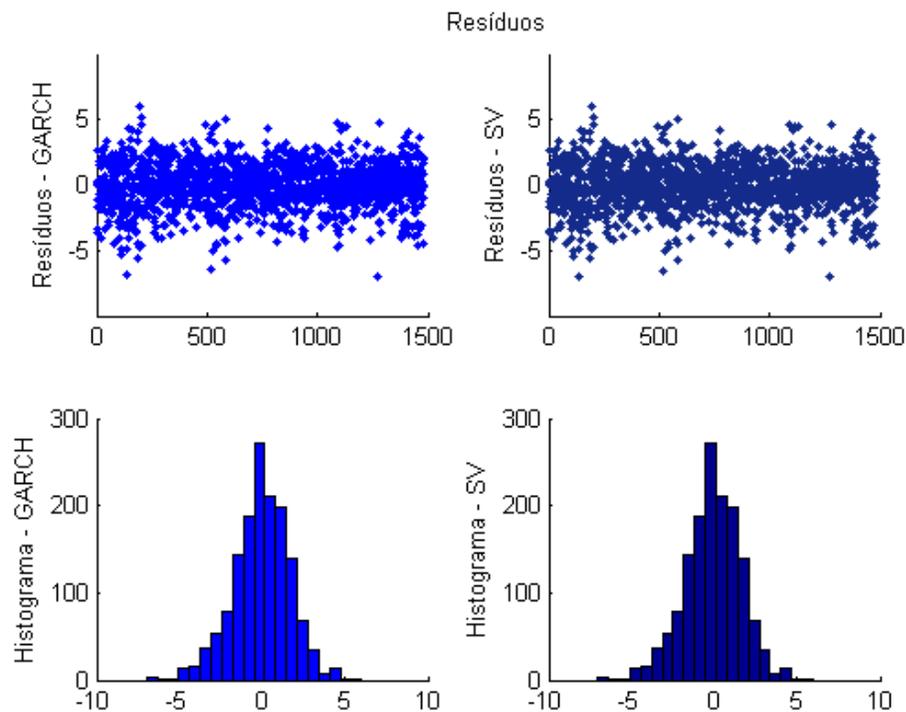


Figura 4-11 - Resíduos dos modelos ajustados.

A Figura 4-11 indica a similaridade entre os resíduos dos modelos, e não é possível distinguir visualmente diferenças entre eles. Isto se justifica pois ambos os modelos têm equações idênticas para as médias dos retornos (Eq. 4.1 e Eq. 4.3), e as diferenças decorrem somente dos pequenos desvios na estimação do parâmetro da equação da média (β_0 no modelo SV e μ no modelo GARCH).

Assim, além dos testes qualitativos realizados para a validação dos modelos, é fundamental avaliar quantitativamente a capacidade de previsão dos mesmos, e medir qual tem o melhor desempenho neste aspecto. Os testes de desempenho que serão realizados são:

1. Erro quadrático médio (EQM) da previsão dos retornos:

O primeiro teste de desempenho que será realizado é o erro quadrático médio (EQM), que é definido por:

$$EQM = \frac{\sum_{t=1}^n (r_t - \hat{r}_t)^2}{n}. \quad (\text{Eq. 4.14})$$

O erro quadrático médio é uma medida que verifica o erro médio de previsão do modelo ajustado, em comparação com os valores reais observados. Assim, o EQM serve como uma função penalizadora, e o quadrado do resíduo é utilizado para penalizar mais os maiores erros.

Deve-se destacar que o EQM não deve ser diferente entre os modelos, visto que ambos possuem as mesmas equações para a média dos retornos. Assim, quaisquer previsões dos retornos são equivalentes (visto que a esperança de cada retorno seria a mesma, ou seja, μ no modelo GARCH e β no modelo MVE).

Esta característica já era de interesse no problema proposto, visto que o objetivo é comparar o desempenho de previsão da volatilidade entre os modelos. Assim, ao igualar as equações para a média dos retornos, é possível concentrar as divergências entre os modelos apenas nas suas capacidades de explicação da volatilidade da série.

Portanto, o EQM será calculado apenas para fins de verificação da condição de igualdade dos parâmetros da média. Além disso, a verificação de erros quadráticos médios semelhantes atua como evidência de que o algoritmo MCMC simulado foi capaz de estimar o parâmetro β satisfatoriamente.

2. Comparação do Erro Quadrado Médio de Previsão da Volatilidade condicional com o quadrado do retorno observado (EQMPV):

Conforme proposto por Andersen e Bollerslev (1998), a comparação de desempenho de previsão da volatilidade entre dois modelos pode ser realizada pelo cálculo de uma função penalizadora da diferença entre as volatilidades previstas e os retornos quadrados observados (que é representante da volatilidade realizada e funciona como *benchmark* para comparação). Assim, o principal indicador de desempenho que será utilizado para a comparação entre os modelos consiste em calcular o erro quadrático médio de previsão da volatilidade (EQMPV), ou seja:

$$EQMPV = \frac{\sum_{t=1}^n (r_t^2 - h_t)^2}{n} \quad (\text{Eq. 4.15})$$

3. Comparação do Erro Médio Absoluto Padronizado da variância condicional prevista com o quadrado do retorno observado (EMAP):

A comparação de desempenho da volatilidade prevista pode ser realizada considerando os erros absolutos padronizados, que consideram os desvios do modelo relativos à grandeza dos retornos ao quadrado, através da expressão:

$$EMAP = \frac{\sum_{t=1}^n \frac{|(r_t^2 - h_t)|}{r_t^2}}{n} \quad (\text{Eq. 4.16})$$

A Figura 4-12 apresenta as volatilidades previstas nos dois modelos e os quadrados dos retornos, que são representantes da volatilidade diária realizada no período:

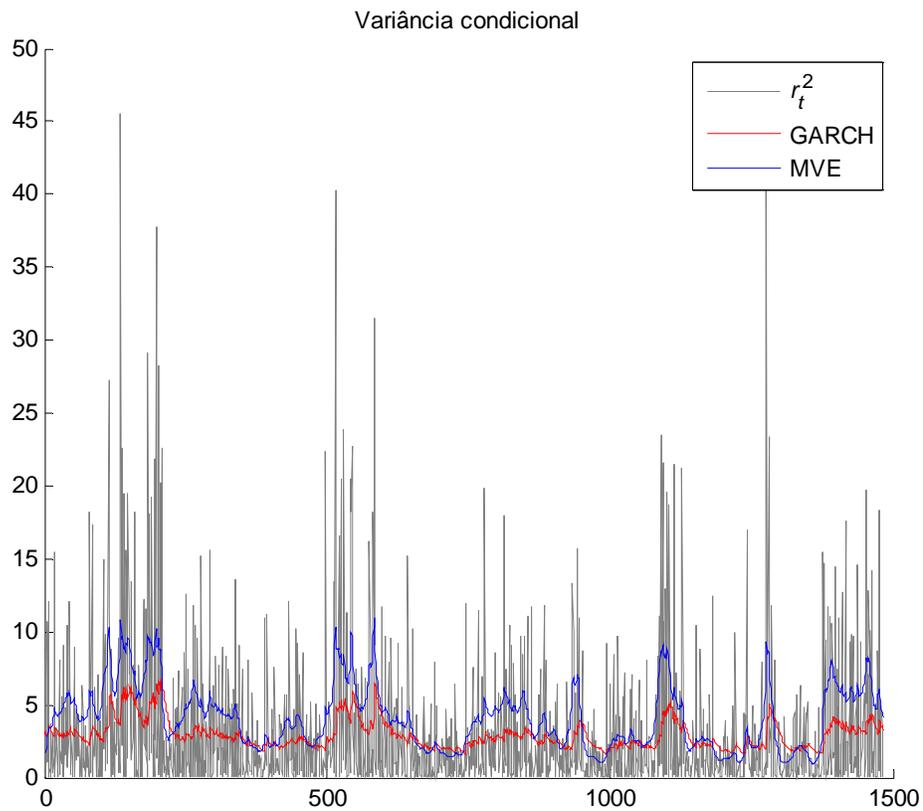


Figura 4-12 - Variância condicional prevista e os quadrados dos retornos observados.

A Figura 4-13 apresenta os resíduos EQMPV, referentes aos erros quadráticos da Figura 4-12:

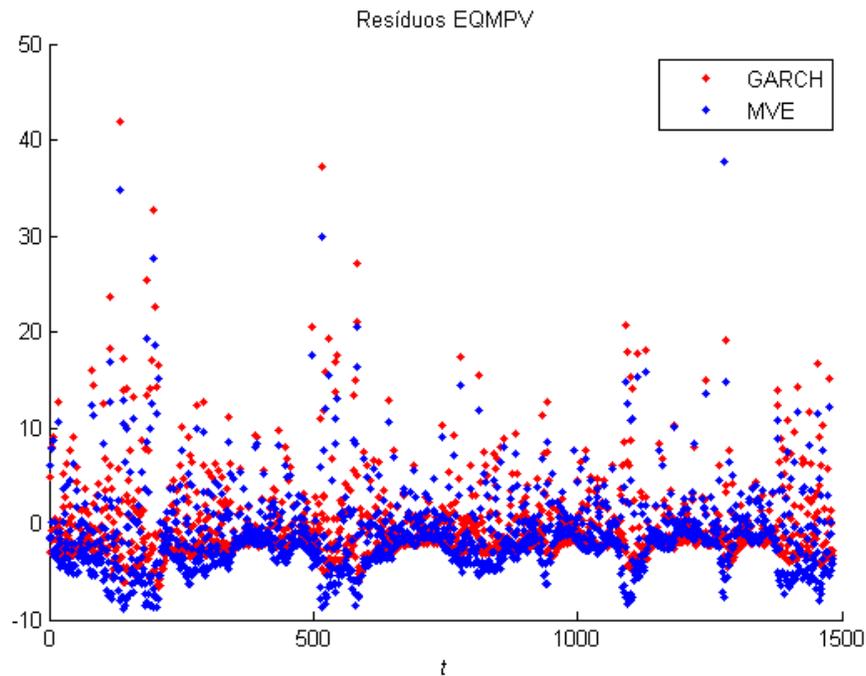


Figura 4-13 - Resíduos EQMPV.

A Figura 4-13 mostra que os resíduos de ambos os critérios apresentam alguns valores extremos. Os resíduos EQMPV apresentam valores elevados coincidentes com os picos de volatilidade, que podem ser observados na Figura 4-12 (os maiores picos ocorrem em $t_{1276} = 47,01$, $t_{135} = 45,59$ e $t_{515} = 40,24$).

A Tabela 4-6 apresenta os quartis dos resíduos EMAP:

Erro Médio Absoluto de Previsão (EMAP)		
Quartil (k)	GARCH	MVE
25%	0.5199	0.4765
50%	1.3590	0.8282
75%	10.9845	6.7466
100%	59092.58	35120.37

Tabela 4-6 - Quartis dos resíduos EMAP.

Os quartis apresentados na Tabela 4-6 denotam o valor dos resíduos no qual há k ocorrências menores ou iguais a este valor. Assim, para $k = 50\%$, por exemplo, pode-se afirmar que metade dos resíduos EMAP são menores ou iguais a 1,3590 para o GARCH e 0,8282 para o MVE. Cabe ressaltar que, no caso em que $k = 50\%$, o percentil é equivalente à mediana dos resíduos.

A análise da Tabela 4-6 mostra que alguns valores bastante elevados são observados, porém a ocorrência é pequena, para valores acima do quartil de 75%. A incidência de tais valores elevados ocorre nos casos em que o denominador do cálculo do EMAP (ou seja, os retornos ao quadrado r_t^2) apresentaram valores bem próximos de zero.

Cabe ressaltar que, na prática, trabalha-se com os erros absolutos de previsão da volatilidade, e não com os erros relativos. Isto ocorre pois a sensibilidade do valor de uma opção em relação às variações da volatilidade é aproximadamente linear, segundo o modelo de Black e Scholes (1973). Assim, o cálculo do EQMPV, proposto por Andersen e Bollerslev (1998), desconsidera os erros de previsão em termos relativos (isto é, um erro de 2% sobre um valor observado de 10% é equivalente a um erro de 2% sobre um valor observado de 60%, por exemplo).

Os valores dos critérios de comparação de desempenho são apresentados na Tabela 4-7, com os resultados obtidos para cada modelo:

Critério	GARCH	MVE
EQM	2,9447*	2,9503
EQMPV	21,7169	19,2190*
EMAP	661.323	390.523*

Tabela 4-7 - Comparação dos critérios de desempenho.

A Tabela 4-7 mostra qual modelo apresentou o melhor desempenho por critério, indicado pelo símbolo (*).

Com base nos valores da Tabela 4-7, pode-se observar que, pelo critério EQM, o modelo MVE obteve um desempenho 0,2% pior que o GARCH. Como

observado anteriormente, já era esperado que os modelos obtivessem EQM muito próximos, pois ambos possuem uma mesma equação para a média dos retornos. Este fato ficou evidenciado pela pequena diferença do EQM entre os dois modelos.

Já pelo critério EQMPV, observa-se que o modelo MVE apresentou um desempenho 13% melhor que o GARCH. Considerando os erros absolutos padronizados da volatilidade (EMAP), o modelo MVE também apresentou melhor desempenho, com um valor 41% menor que o modelo GARCH.

Com base nos resultados obtidos, pode-se afirmar que o modelo proposto tem maior capacidade de previsão da volatilidade da série do IBOVESPA, tomando-se como critério comparativo o erro quadrado médio de previsão da volatilidade em relação aos quadrados dos retornos observados.

5. CONCLUSÕES

Foi desenvolvido um modelo de previsão de uma série de retornos e das suas volatilidades condicionais (MVE), no qual ambas variáveis possuem componentes estocásticos em suas equações. Para tal, utilizou-se a teoria de inferência Bayesiana e algoritmos de simulação MCMC para obter os valores dos parâmetros do modelo.

O modelo proposto por Jacquier et. al. (1994) surgiu como uma evolução dos modelos ARCH-GARCH, por acrescentar um termo estocástico na equação da volatilidade. Esta característica, entretanto, aumenta a complexidade de resolução do modelo, pois a função de verossimilhança não pode ser derivada analiticamente, e portanto a estimação dos seus parâmetros não pode ser realizada pelos métodos tradicionais de otimização desta função.

Como alternativa de resolução dos problemas de inferência Bayesiana nos quais os métodos de máxima verossimilhança não podem ser aplicados, surgiu a utilização dos algoritmos de simulação via MCMC, que possibilitam a amostragem da distribuição *a posteriori* conjunta dos parâmetros, e que permitem a estimação dos parâmetros do modelo.

A eficiência do modelo proposto foi avaliada através do cálculo dos erros quadrados de previsão dos retornos e das volatilidades, em confronto com um modelo GARCH, e ambos foram ajustados à uma série de retornos diários do IBOVESPA. Como o principal objetivo é a previsão da volatilidade da série, adotou-se um modelo GARCH com uma mesma equação para a média dos retornos, para que o poder de previsão dos modelos pudesse ser devidamente comparado entre eles. Assim, ambos modelos apresentaram desempenhos similares na capacidade de previsão dos retornos, com o modelo MVE obtendo um desempenho de apenas 0,2% pior que o GARCH. Entretanto, o modelo MVE apresentou um desempenho 13% melhor do que o GARCH na capacidade de previsão da volatilidade da série pelo critério de erro quadrado médio de previsão, e 40% melhor pelo critério de erro médio absoluto padronizado.

Apesar da melhor capacidade de previsão da volatilidade pelo modelo MVE, alguns aspectos negativos em relação ao GARCH devem ser salientados. A principal

dificuldade de implementação do MVE decorre da estimação dos parâmetros do modelo, que requer métodos complexos de simulação que demandam grande capacidade de processamento computacional e tempo para sua resolução (o modelo GARCH foi ajustado em apenas 6 segundos, utilizando-se o pacote existente no *software* MATLAB, enquanto a estimação do modelo MVE demorou 135 minutos).

A utilização do modelo MVE deve ser realizada considerando-se a finalidade da previsão da volatilidade, e o nível de acurácia necessário. No contexto da empresa onde o trabalho foi realizado, a volatilidade é especulada no mercado financeiro, e a correta previsão da mesma proporciona ganhos financeiros diretos para a empresa. Assim, é fundamental que a sua estimação seja a melhor possível, e nesse caso o modelo MVE deve ser aplicado, em contrapartida ao GARCH. Porém, em outras aplicações o modelo pode tornar-se inconveniente, devido à sua complexidade e demora no ajustamento. Um exemplo é a utilização da volatilidade futura para avaliação de medidas de risco em carteiras de investimento, nas quais a sensibilidade dos seus resultados em relação à pequenas variações da volatilidade prevista pode ser pouco significativa.

Outro importante aspecto do trabalho desenvolvido é a possibilidade da estimação dos parâmetros de praticamente qualquer modelo de séries temporais via simulação MCMC. Por não haver necessidade de derivar funções de verossimilhança, os algoritmos de simulação possibilitam a estimação de modelos mais complexos, com maior número de parâmetros e variáveis explicativas.

O MVE construído trata-se de uma formulação simples e enxuta de um modelo de volatilidade estocástica. Uma recomendação para trabalhos futuros seria a inclusão de novas variáveis, como o volume de negociação do IBOVESPA, cotações de outros índices (como o S&P500, um dos principais índices de ações dos Estados Unidos), preços de *commodities* (como o petróleo e o níquel), entre outras.

6. REFERÊNCIAS BIBLIOGRÁFICAS

ANDERSEN, T. G.; BOLLERSLEV, T. **Answering the skeptics**. International Economic Review, 1998.

BLACK, F.; SCHOLES, M. **The Pricing of Options and Corporate Liabilities**. Journal of Political Economy, v.81, 1973.

BODIE, Z.; KANE, A.; MARCUS, A. **Fundamentos de Investimentos**. Porto Alegre: Bookman, 2000.

BOLLERSLEV, T. **Generalized Autoregressive Conditional Heteroskedasticity**. Journal of Econometrics, 1986.

BOX G. E. P.; JENKINS G. M. **Time Series Analysis: Forecasting and Control**. Holden-Day Inc., 1970.

BOX, G. E. P.; TIAO, G. C. **Bayesian inference in statistical analysis**. Addison-Wesley Publ., 1973.

COSTA, C. L. **Opções, Operando a Volatilidade**. São Paulo: Cultura, 1998.

DEGROOT, M. H. **Optimal Statistical Decisions**. New York: McGraw Hill, 1970.

ENGLE, R. F. **Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of UK Inflation**. Econometrica, 1982.

FELLER, W. **An introduction to probability theory and its applications**. Wiley, 1968.

GALLANT, A. R.; HSIEH, D.; TAUCHEN, G. **Estimation of stochastic volatility models with diagnostics**. Journal of Econometrics, 1991.

GAMERMAN, D. L.; FREITAS, H. **Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference**. 2. ed. Londres: Chapman & Hall/CRC, 2006.

GELFAND, A. E.; SMITH, A. F. **Sampling-based Approaches to Calculating Marginal Densities**. Journal of the American Statistical Association, 1990.

GEMAN, S.; GEMAN, D. **Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images**. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1984.

GUJARATI, D. N. **Econometria Básica**. Pearson Education do Brasil, 2001.

HAMILTON, J. D. **Time Series Analysis**. New Jersey: Princeton University Press, 1994.

- HULL, J. C. **Options, Futures & Other Derivatives**. New Jersey: Prentice Hall, 1997.
- JAQUIER, E.; POLSON, N. G.; ROSSI, P. E. **Bayesian Analysis of Stochastic Volatility Models**, 1994.
- MEYN, S. P.; TWEEDIE, R. I. **Markov chains and stochastic stability**. Springer, 1993.
- MORETTIN, P. A. TOLOI, C. M. **Análise de séries temporais**. 2. ed. São Paulo: Edgard Blücher, 2006.
- NELSON, D. B. **Conditional heteroskedasticity in asset returns: A new approach**. *Econometrica*, 1991.
- NUMMELIN, E. **General Irreducible Markov Chains and Non Negative Operators**. Cambridge University Press, 1984.
- PEÑA, D.; TIAO, G. C.; TSAY, R. T. **A Course in Time Series Analysis**. John Wiley & Sons Inc., 2001.
- RITTER, C.; TANNER, M. A. **Facilitating the Gibbs sampler: the Gibbs stopper and the griddy-Gibbs sampler**. *Journal of the American Statistical Association*, 1992.
- ROSS, S. **Stochastic Processes**. Wiley, 1996.
- ROSSI, P. E. **Bayesian Statistics and Marketing**. London: John Wiley & Sons Ltd, 2005.
- SANTOS, B.L. **Medidas de Risco em Carteiras de Ativos Financeiros**. Trabalho de Formatura para obtenção do título de Engenheiro de Produção, Escola Politécnica da Universidade de São Paulo, 2007.
- SARTORIS, A. **Estatística e introdução à economia**. São Paulo: Saraiva, 2003.
- SHEPHARD, N. **Stochastic volatility selected readings**. New York: Oxford University Press Inc., 2005.
- TAUCHEN, G.; PITTS, M. **The Price Variability-Volume Relationship on Speculative Markets**. *Econometrica*, Econometric Society, 1983.
- TSAY, R. S. **Analysis of financial time series**. New York: John Wiley & Sons Inc., 2002.
- ZAKOIAN, J. M. **Threshold heteroskedastic models**. *Journal of Economic Dynamics and Control*, 1994.

ANEXO – Código do programa escrito em MATLAB

```

%Apaga todas as variaveis e gráficos existentes
clear
close all

%Pega os dados da serie de retornos no arquivo "IBOV_d.m"
IBOV_d

t1 = clock;

%Ajusta um modelo GARCH(1,1)
Spec = garchset('P',1,'Q',1,'Display','off');
[Coeff, Errors, LLF, Innovations, Sigmas, Summary] = garchfit(Spec,r);

t2 = clock;

% Valores iniciais de {ht}: Sigmas
%Priori de mu (beta)
beta0 = 0;
sigma_beta0 = sqrt(9);
mu = 0.66;
grid_beta = (-15:0.05:15);
priori_beta = normpdf(grid_beta,beta0,sigma_beta0)';
%Priori de alpha
alpha0 = [0.4 0.8]';
C0 = [0.09 0;0 0.04];
grid_alpha = (-2:0.01:2);
priori_alpha(:,1) = normpdf(grid_alpha,alpha0(1),sqrt(C0(1,1)))';
priori_alpha(:,2) = normpdf(grid_alpha,alpha0(2),sqrt(C0(2,2)))';
%Priori de sigma_v^2
m = 5;
lambda = 0.2;
sigma_v = sqrt(0.5);
grid_sigma = linspace(.01,.99,1001);
priori_sigma = chi2pdf(1./grid_sigma,m)./grid_sigma.^2;

%Parametros do Amostrador de Gibbs
n_it = 5100; %numero de iteracoes
n_burn = 100; %burn-in sample

%Parametros do Griddy Gibbs
s = std(r);
n_gp = 500; %numero de grid points
h_lb = 0; %banda inferior
h_ub = 1.5*s^2; %banda superior
inter = (h_ub-h_lb)/n_gp;

```

```

h_grid = (h_lb+inter:inter:h_ub)';

% Valores iniciais (k=1)
beta_sim(1,1:n_it) = 0;
h_sim(1:n,1:n_it) = 0;
alpha_sim(1:2,1:n_it) = 0;
sigma_v2_sim(1,1:n_it) = 0;

beta_sim(1,1) = mu;
h_sim(:,1) = Sigmas.^2;
alpha_sim(:,1) = alpha0;
sigma_v2_sim(1,1) = sigma_v^2;
ln_Sigmas = log(Sigmas.^2);

t3 = clock;

% Inicia Amostrador de Gibbs
for k = 2:n_it

    x = 1;
    r0t = r./sqrt(h_sim(:,k-1));
    x0t = x./sqrt(h_sim(:,k-1));

    % Amostragem de beta
    sigma_beta_ast = inv(sum(x0t.*x0t) + inv(sigma_beta0^2));
    beta_ast = sigma_beta_ast*(sum(x0t.*r0t)+inv(sigma_beta0^2)*beta0);
    beta_sim(k) = mvnrnd(beta_ast,sigma_beta_ast);

    % Amostragem de h
    nep = alpha_sim(1,k-1)/(1-alpha_sim(2,k-1));
    h_0 = (alpha_sim(2,k-1)^2)*(ln_Sigmas(2,k-1)-nep);
    h_n1 = alpha_sim(1,k-1) + alpha_sim(2,k-1)*(alpha_sim(1,k-1) + alpha_sim(2,k-1)*ln_Sigmas(n-1,k-1));
    for i = 1:n
        if i == 1
            mu_t(i,1) = (alpha_sim(1,k-1)*(1-alpha_sim(2,k-1)) + alpha_sim(2,k-1)*(ln_Sigmas(i+1,k-1)+h_0))/(1+alpha_sim(2,k-1)^2);
        elseif i == n
            mu_t(i,1) = (alpha_sim(1,k-1)*(1-alpha_sim(2,k-1)) + alpha_sim(2,k-1)*(h_n1+log(h_sim(i-1,k))))/(1+alpha_sim(2,k-1)^2);
        else
            mu_t(i,1) = (alpha_sim(1,k-1)*(1-alpha_sim(2,k-1)) + alpha_sim(2,k-1)*(ln_Sigmas(i+1,k-1)+log(h_sim(i-1,k))))/(1+alpha_sim(2,k-1)^2);
        end

        sigma2 = sigma_v2_sim(k-1)/(1+alpha_sim(2,k-1));

```

```

    f(:,i) = (h_grid.^(-1.5)).*exp(-(r(i,1)-x'*beta_sim(k)).^2./(2*h_grid)-
(log(h_grid)-mu_t(i,1)).^2)/(2*sigma2));
    f_norm = sum(f(:,i));
    f(:,i) = f(:,i)/f_norm;
    cdf = cumsum(f(:,i));
    h_sim(i,k) = interp1q(cdf,h_grid,rand);
end
ln_Sigmas(:,k) = log(h_sim(:,k));

%Amostragem de w (particionado em alpha e sigma_v^2)
%alpha
z(1,2:n) = 1;
z(2,2:n) = ln_Sigmas(1:n-1,k);
C_ast = [0 0;0 0];
for t = 2:n
    C_ast = C_ast + (z(:,t)*z(:,t)');
end
C_ast = inv(C_ast/sigma_v2_sim(k-1) + inv(C0));
alpha_ast = [0 0]';
for t = 2:n
    alpha_ast = alpha_ast + (z(:,t)*ln_Sigmas(t,k)');
end
alpha_ast = C_ast*(alpha_ast/sigma_v2_sim(k-1) + inv(C0)*alpha0);
alpha_sim(:,k) = mvnrnd(alpha_ast',C_ast)';

%sigma_v^2
v(2:n,1) = ln_Sigmas(2:n,k) - alpha_sim(1,k) - alpha_sim(2,k)*ln_Sigmas(1:n-
1,k);
sigma_v2_sim(k) = (((5*0.2)+sum(v.*v))./chi2rnd(m+n-1));

if mod(k,50) ==0
    display(k)
end
end

t4 = clock;

output(1:n,1) = mean(h_sim(:,n_burn+1:n_it),2);

beta_stat = mean(beta_sim(n_burn+1:n_it));
alpha_stat = mean(alpha_sim(:,n_burn+1:n_it),2);
sigma_stat = mean(sigma_v2_sim(:,n_burn+1:n_it),2);

beta_stat2= std(beta_sim(n_burn+1:n_it));
alpha_stat2(1) = std(alpha_sim(1,n_burn+1:n_it));
alpha_stat2(2) = std(alpha_sim(2,n_burn+1:n_it));
sigma_stat2 = std(sigma_v2_sim(:,n_burn+1:n_it));

```

```

disp('Parameter  beta0  alpha 0  alpha 1  sigma_v2')
disp(sprintf('Mean      %f  %f  %f
%f',beta_stat,alpha_stat(1),alpha_stat(2),sigma_stat))
disp(sprintf('St. Error  %f  %f  %f
%f',beta_stat2,alpha_stat2(1),alpha_stat2(2),sigma_stat2))

%Posterioris
x = 1;
r0t = r./sqrt(output);
x0t = x./sqrt(output);
sigma_beta_ast = inv(sum(x0t.*x0t) + inv(sigma_beta0^2));
beta_ast = sigma_beta_ast*(sum(x0t.*r0t)+inv(sigma_beta0^2)*beta0);
posteriori_beta = normpdf(grid_beta,beta_ast,sqrt(sigma_beta_ast));

%alpha
z(1,2:n) = 1;
z(2,2:n) = log(output(1:n-1));
C_ast = [0 0;0 0];
for t = 2:n
    C_ast = C_ast + (z(:,t)*z(:,t));
end
C_ast = inv(C_ast/sigma_stat + inv(C0));
alpha_ast = [0 0]';
for t = 2:n
    alpha_ast = alpha_ast + (z(:,t)*log(output(t)));
end
alpha_ast = C_ast*(alpha_ast/sigma_stat + inv(C0)*alpha0);
posteriori_alpha(:,1) = normpdf(grid_alpha,alpha_ast(1),sqrt(C_ast(1,1)));
posteriori_alpha(:,2) = normpdf(grid_alpha,alpha_ast(2),sqrt(C_ast(2,2)));
%sigma
v(2:n,1) = log(output(2:n)) - alpha_stat(1) - alpha_stat(2)*log(output(1:n-1));
posteriori_sigma = chi2pdf(1./grid_sigma,m+n-
1).*(m*lambda+sum(v.*v))./grid_sigma.^2;

disp(sprintf('Número de simulações: %d',n_it))
disp(sprintf('Burn-in sample: %d',n_burn))
disp(sprintf('Número de dados: %d',n))
disp(sprintf('Tempo decorrido para realizar GARCH: %1.2f seg.',etime(t2,t1)))
disp(sprintf('Tempo decorrido para realizar simulação: %1.0f min.',etime(t4,t3)/60))

%Gera gráficos dos parâmetros
graficos

```