

**UNIVERSIDADE DE SÃO PAULO
ESCOLA POLITÉCNICA**

TIAGO CARVALHO DE MATOS MARQUES

MONITORAMENTO DE PERFIS APLICADO À ANÁLISE DE PORTFÓLIOS

São Paulo

2010

TIAGO CARVALHO DE MATOS MARQUES

MONITORAMENTO DE PERFIS APLICADO À ANÁLISE DE PORTFÓLIOS

**Trabalho de formatura apresentado à
Escola Politécnica da Universidade de
São Paulo para a obtenção do diploma
de Engenheiro de Produção.**

**São Paulo
2010**

TIAGO CARVALHO DE MATOS MARQUES

MONITORAMENTO DE PERFIS APLICADO À ANÁLISE DE PORTFÓLIOS

**Trabalho de formatura apresentado à
Escola Politécnica da Universidade de
São Paulo para a obtenção do diploma
de Engenheiro de Produção.**

Orientadora: Profa. Dra. Linda Lee Ho

São Paulo

2010

**AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTES
TRABALHOS, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO,
PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.**

Ficha Catalográfica

Marques, Tiago Carvalho de Matos
Monitoramento de perfis aplicado à análise de portfólios /
T.C.M. Marques. -- São Paulo, 2010.
119 p.

Trabalho de Formatura - Escola Politécnica da Universidade
de São Paulo. Departamento de Engenharia de Produção.

1. Controle estatístico de processo 2. Monitoramento de per-
fil linear 3. Beta de ações 4. Mercado (Modelo) 5. Análise de
regressão I. Universidade de São Paulo. Escola Politécnica.
Departamento de Engenharia de Produção II. t.

DEDICATÓRIA

À minha avó Prazeres.

AGRADECIMENTOS

À professora Lee Ho, que me direcionou no desenvolvimento do presente trabalho, mesmo com uma agenda de atividades repleta. Sua experiência, conhecimento e seu perfeccionismo foram cruciais, não apenas para este trabalho, mas para a minha formação como engenheiro.

Aos excelentes professores da Escola Politécnica, sem os quais eu jamais poderia absorver o conhecimento necessário para atingir meus resultados acadêmicos, me tornar um engenheiro e ser o que sou.

Aos funcionários da Escola Politécnica, Marlene, Dejanira e Patrícia, cuja extrema competência parece ser fruto do prazer que nutrem pelo trabalho que realizam. A dedicação dessas pessoas é transmitida aos alunos e refletida nos resultados obtidos por todos.

À École Centrale Paris e à França, pela oportunidade oferecida e pela forma como me acolheram durante os dois anos de estudos pelo duplo diploma, período que contribuiu decisivamente para minha formação acadêmica, profissional e humana, e que provocou uma mudança radical na minha relação com a sociedade e com a vida.

Aos meus pais, minha avó e meus irmãos que sempre me proveram em tudo o que precisei e com os quais eu sempre pude contar. À família Borges que, tão logo me conheceu, me integrou com extrema receptividade, ofereceu apoio e conforto e depositou em mim uma confiança desmedida.

Aos poucos que se mostraram dignos de receber minha confiança com tal intensidade que decidi permitir sua contribuição para o direcionamento da minha vida.

Aquela que esteve em meus pensamentos durante todos os dias dos últimos seis meses, fonte de inspiração em tudo o que fiz e com quem espero continuar a compartilhar momentos cujas lembranças me fazem sorrir nas mais ásperas aflições: Julia.

A todos aqueles que me fizeram perceber que lealdade, honestidade, humildade, amizade, altruísmo, polidez e perseverança são valores que levam ao sucesso e formam um verdadeiro homem. É pelos verdadeiros homens que faz sentido empregar todo o esforço que empreguei para preparar este trabalho.

EPIÍGRAFE

« Qui sait tout souffrir peut tout oser. »

Luc de Clapiers, marquis de Vauvenargues.

(Tradução livre: quem sabe suportar tudo pode ousar tudo.)

RESUMO

MARQUES, T. C. M. **Monitoramento de perfis aplicado à análise de portfólios.** 2010. 119 f. Trabalho de formatura (Graduação) – Escola Politécnica, Universidade de São Paulo, São Paulo, 2010.

A análise do risco de um portfólio de ativos é fundamental tanto para uma correta caracterização de produtos financeiros, quanto para seleção das técnicas mais adequadas para a neutralização do risco identificado. Desta forma, medir precisamente esse risco é de grande interesse para os envolvidos na negociação de tais ativos. Nesse contexto, o *Market Model* tem sido utilizado, visto ser uma representação simples e eficiente do risco a partir do coeficiente angular de um modelo de regressão que relaciona retornos do ativo analisado com retornos de um índice que reflita o desempenho da economia como um todo. No caso, foi considerado um índice da Bolsa de Valores. Para se caracterizar mais adequadamente o risco dos ativos e neutralizá-lo da forma desejada, é crucial uma correta estimação e monitoramento dos coeficientes desse modelo de regressão. O presente trabalho se propõe a aplicar ferramentas do controle estatístico de processos, particularmente gráficos de controle aplicados à fase I do monitoramento de perfis lineares, para estimar os parâmetros do *Market Model*. Tais ferramentas foram empregadas em análises de diversos setores industriais e períodos, de maneira a demonstrar a utilidade prática da metodologia proposta. Como principais resultados obtidos, ressaltam-se a aplicação detalhada da metodologia para ações de alguns setores industriais em diferentes períodos, evidenciando o ganho obtido na estimativa dos parâmetros, e análises gráficas. Tais aplicações e análises permitem monitorar a estabilidade de tais parâmetros ao longo de horizontes de tempo mais amplos.

Palavras-chave: controle estatístico de processo, monitoramento de perfil linear, beta de ações, *Market Model*, análise de regressão.

ABSTRACT

MARQUES, T. C. M. **Profile monitoring applied to portfolio analysis**. 2010. 119 f. Trabalho de formatura (Graduação) – Escola Politécnica, Universidade de São Paulo, São Paulo, 2010.

The risk analysis of an asset portfolio is essential not only to correctly characterize financial products, but also to select the most adequate risk neutralizing techniques. Therefore, portfolio stakeholders are highly interested in the definition of a procedure to precisely measure this risk. In this context, the Market Model has been used, as it provides a simple and efficient risk representation, by considering the slope coefficient of a regression model relating the asset returns with the returns of a market index. In the present work, a stock market index was used. To properly characterize the assets' risk and neutralize it as wanted, it is crucial to correctly estimate and monitor the regression model's coefficients. The purpose of this work is to apply statistical quality control tools, particularly control charts and phase I of linear profile monitoring, to estimate the Market Model parameters. These tools were used to develop analysis of a diverse set of industrial sectors and periods, attesting the proposed methodology's applicability. As main results, it can be pointed out the detailed application of the methodology for stocks of particular industrial sectors, in different periods of time, showing the benefits obtained by these estimations, and the development of graphical analysis. These analysis and applications allow the parameters' stability monitoring throughout longer time frames.

Keywords: statistical process control, linear profile monitoring, stocks *beta*, *Market Model*, linear regression.

LISTA DE ILUSTRAÇÕES

Figura 1 – Esquema representativo das técnicas usadas neste trabalho	18
Figura 2 – Esquema dos assuntos abordados no capítulo.....	26
Figura 3 – Benefícios da baixa correlação entre ativos	29
Figura 4 – Representação gráfica da <i>fronteira eficiente</i>	29
Figura 5 – O portfólio de mercado e a LMC	32
Figura 6 – Diversificação e os diferentes tipos de risco	33
Figura 7 – A linha do mercado de títulos	35
Figura 8 – A linha característica.....	37
Figura 9 – Interpretação gráfica da regressão simples	50
Figura 10 – Modelo de regressão com 3 variáveis	50
Figura 11 – Diagramas de Venn para diferentes níveis de colinearidade	53
Figura 12 – Representação de um gráfico de controle	60
Figura 13 – Resultados do teste de normalidade	77

LISTA DE GRÁFICOS

Gráfico 1 – Variação do preço de BBDC3 no período denominado 2010, evidenciando as amostras como sub-períodos	75
Gráfico 2 – Gráfico de controle para variância para BDC3 em período 2010	79
Gráfico 3 – Retas de regressão ajustadas para as amostras	82
Gráfico 4 – Retas ajustadas, após remoção de amostras <i>fora de controle</i>	84
Gráfico 5 – Market Neutral – influência do estimador de <i>beta</i>	86
Gráfico 6 – Comportamento do <i>beta</i> da categoria <i>Transporte Aéreo</i>	88
Gráfico 7 – Comportamento do <i>beta</i> da categoria <i>Água e saneamento</i>	89
Gráfico 8 – Comportamento do <i>beta</i> da categoria <i>Bancos</i>	89
Gráfico 9 – Comportamento do <i>beta</i> de ações <i>Blue Chips</i>	90
Gráfico 10 – Comportamento do <i>beta</i> de ações da categoria <i>Imóveis</i>	91
Gráfico 11 – Comportamento do <i>beta</i> de ações da categoria <i>Distribuição de Energia</i>	91
Gráfico 12 – Comportamento do <i>beta</i> da categoria <i>Rodoviário</i>	92
Gráfico 13 – Comportamento do <i>beta</i> da categoria <i>Siderurgia</i>	93
Gráfico 14 – Períodos tratados como amostras para análise mais ampla	95

LISTA DE TABELAS

Tabela 1 – Exemplo de matriz do modelo para 2 regressores e 4 amostras de 3 observações.	67
Tabela 2 – Ações selecionadas separadas nas categorias que a serem analisadas	73
Tabela 3 – Períodos definidos para coleta de dados	73
Tabela 4 – Exemplo de dados coletados.....	76
Tabela 5 – Teste de linearidade	77
Tabela 6 – Dados do gráfico de controle para variância de BBDC3 - 2010	78
Tabela 7 – Valores obtidos para teste F de detecção de amostras <i>fora de controle</i> de GOLL4 em 2009	80
Tabela 8 – Valores de <i>beta</i> antes e depois do tratamento das amostras	85
Tabela 9 – Resultados finais para VALE5	94
Tabela A. 1 – Resultados para o setor <i>Transporte aéreo</i>	104
Tabela A. 2 – Resultados para o setor <i>Água e saneamento</i>	104
Tabela A. 3 – Resultados para o setor <i>Bancos</i>	104
Tabela A. 4 – Resultados para o setor <i>Blue Chips</i>	105
Tabela A. 5 – Resultados para o setor <i>Imóveis</i>	106
Tabela A. 6 – Resultados para o setor <i>Distribuição de energia</i>	107
Tabela A. 7 – Resultados para o setor <i>Rodoviário</i>	108
Tabela A. 8 – Resultados para o setor <i>Siderurgia</i>	108

LISTA DE ABREVIATURAS E SIGLAS

CAPM: *capital asset pricing model*

CEP: controle estatístico de processos

LIC: limite inferior de controle

LMC: linha do mercado de capitais

LMT: linha do mercado de títulos

LSC: limite superior de controle

MMQG: método dos mínimos quadrados generalizado

MMQO: método dos mínimos quadrados ordinários

MMV: método da máxima verossimilhança

MRLC: modelo de regressão linear clássico

MSE: *mean of squared error*

SPC: *statistical process control*

SSE: *sum of squared error*

SSR: *sum of squares of the regression*

LISTA DE SÍMBOLOS

g : número de ativos num portfólio

m_0 : número de amostras coletadas

m : número de amostras restantes após aplicação do gráfico de controle para variância

m_{final} : número de amostras final usado para cálculo dos estimadores dos coeficientes

n : número de observações da amostra

N : número de observações total

p : número de variáveis explicativas (regressores) do modelo

u : erro estocástico ou *ruído branco*

SUMÁRIO

1.	INTRODUÇÃO.....	15
2.	MOTIVAÇÃO.....	17
3.	ESTRUTURA DO TRABALHO	19
4.	CONCEITOS FINANCEIROS UTILIZADOS.....	21
4.1.	AÇÕES	21
4.2.	PORTFÓLIO.....	21
4.3.	RETORNO.....	22
4.4.	AVERSÃO AO RISCO	23
4.5.	CONTRATO FUTURO DE IBOVESPA.....	23
5.	REVISÃO BIBLIOGRÁFICA	25
5.1.	MARKET MODEL	25
5.2.	ESTRATÉGIA MARKET NEUTRAL	41
5.3.	REGRESSÃO	43
5.3.1.	REGRESSÃO LINEAR MÚLTIPLA	44
5.3.2.	UTILIZAÇÃO DE VARIÁVEIS INDICADORAS NA REGRESSÃO..	55
5.3.3.	TESTES DE HIPÓTESES APLICADOS À REGRESSÃO	56
5.4.	MONITORAMENTO DE PERFIL LINEAR	58
6.	APLICAÇÃO	71
7.	CONCLUSÕES	97
	REFERÊNCIAS.....	99
	APÊNDICES	103

1. INTRODUÇÃO

Uma das atividades desempenhadas por instituições financeiras gestoras de fundos próprios ou de terceiros é a realização de análises envolvendo o retorno financeiro de seus produtos. Entre as motivações para tais análises estão a preocupação com a possibilidade de perda financeira (aversão ao risco) e medidas para evitá-la (estratégias de neutralização), o interesse em verificar se o desempenho observado corresponde ao proposto no momento da criação do produto e obrigações legais. Com relação a essas duas últimas motivações, observa-se a importância do conceito de risco, já que o nível de risco deve ser definido no momento da criação do fundo e comunicado aos investidores, sendo a instituição financeira legalmente obrigada a manter as características do produto de acordo com as especificações definidas.

A interpretação mais aceita para o risco é aquela que compara a variação dos retornos do ativo e a variação dos retornos médios da economia do país como um todo. O retorno de um índice de mercado, como o Índice Bovespa no Brasil, é usado como uma medida do nível de retorno global da economia do país. Outros indicadores, como o PIB, podem também servir ao mesmo fim.

O risco é medido pelo coeficiente angular de um modelo de regressão linear tendo como variável explicada o retorno do ativo e variável explicativa o retorno do índice de mercado. Esse modelo é conhecido como *Market Model*.

Produtos financeiros podem ser objeto de estudo do controle estatístico da qualidade. Neste trabalho será aplicada uma técnica de monitoramento de perfil linear para monitorar a estabilidade do risco (medido na forma de *beta* do *Market Model*) de algumas ações. Com base na medida confiável do *beta* no tempo, a instituição pode garantir a qualidade de seus produtos e tomar decisões de diversificação que permitam a melhoria dessa qualidade, fornecendo produtos com maior retorno para um mesmo nível de risco. Testes baseados na estatística F permitem a detecção de origens anormais de variabilidade e remoção de amostras *fora de controle*, com o objetivo de se obter uma melhor estimativa dos parâmetros do *Market Model*, gerando maior confiabilidade na caracterização dos produtos financeiros e melhores resultados de estratégias de neutralização do risco.

2. MOTIVAÇÃO

A motivação deste trabalho decorre do problema de controle e estimação do risco enfrentado por instituições financeiras, em especial no caso de ativos envolvendo ações. A medida do risco e seu monitoramento são de fundamental importância, tanto para a elaboração de novos produtos constituídos de portfólios de ações ou derivados destes, quanto para o controle de produtos já existentes.

O risco constitui uma característica básica de produtos financeiros e, portanto, noções de risco estão presentes em contratos de aquisição desses produtos. As instituições financeiras são legalmente obrigadas a declarar e seguir o nível de risco indicado nos contratos instituição-cliente. Dessa forma, vê-se que a importância do controle do risco vai além da proteção dos fundos próprios do banco, atingindo a esfera legal do ponto de vista dos direitos do consumidor (de produtos financeiros).

De posse de medidas de risco de diversos ativos é possível utilizar a *diversificação* como forma de criar produtos e controlar o risco de já existentes, questão também discutida neste trabalho.

A medida do risco de um produto financeiro, cuja intangibilidade é uma característica marcante, envolve ações de maior complexidade do que, por exemplo, a medida do diâmetro de uma circunferência produzida industrialmente, coletada com o intuito de verificar a conformidade com especificações de um projeto.

Assim, para avaliar o risco é necessário o uso de modelos, que envolvem hipóteses simplificadoras. De todo o modo, da mesma forma que se quer medir o diâmetro da circunferência por motivos de controle da qualidade, numa instituição financeira a medida e o controle do risco têm a mesma função: o controle de qualidade do produto financeiro em questão.

Este trabalho propõe uma solução para o problema da medida e controle do risco que envolve:

- a) Um modelo que propõe uma forma de mensurar o risco de um portfólio de ações;
- b) Uma metodologia que envolve controle estatístico de processos para o controle sistemático da medida de risco proposta por este modelo;

- c) A utilização de ferramentas estatísticas para aplicação do modelo e da metodologia de controle estatístico de processos.

A ferramenta do controle estatístico de processos utilizada é o monitoramento de perfis lineares. Encontram-se na literatura aplicações eficientes de tal ferramenta na indústria. De forma semelhante, através de tal ferramenta, podem-se obter resultados eficazes para a melhoria da caracterização do risco de ativos e para o desenvolvimento e aplicação de estratégias de neutralização de risco.

O esquema da Figura 1 ilustra o papel do monitoramento de perfis lineares no âmbito específico da indústria financeira.

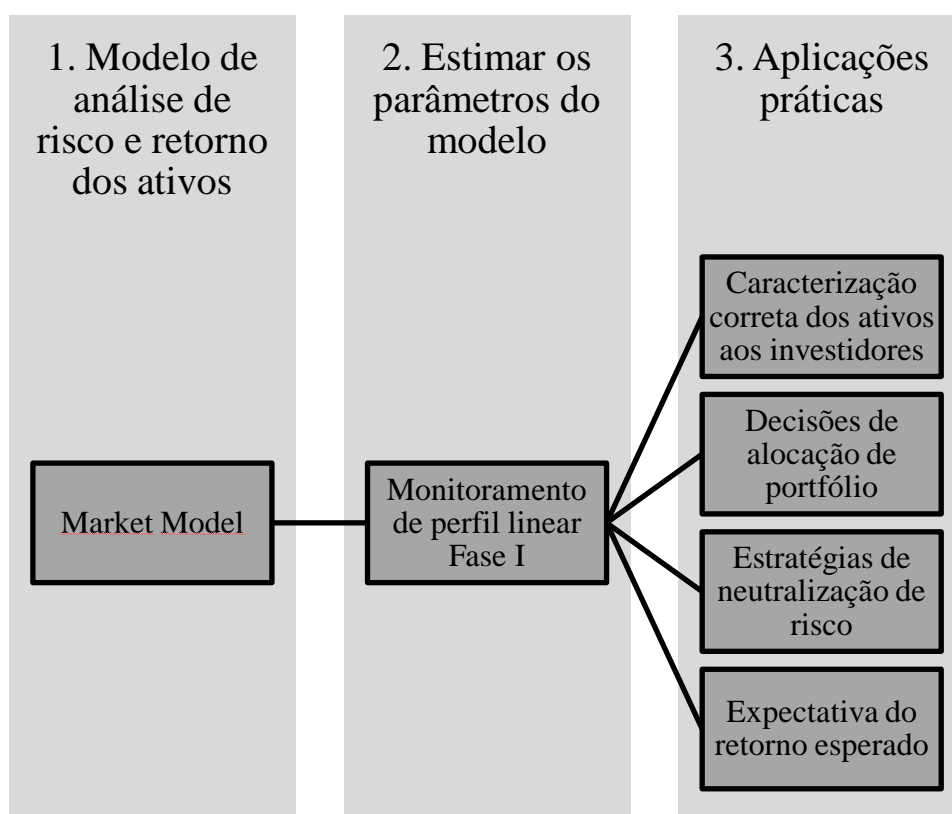


Figura 1 – Esquema representativo das técnicas usadas neste trabalho

3. ESTRUTURA DO TRABALHO

A estruturação do presente trabalho está apresentada a seguir.

O capítulo CONCEITOS FINANCEIROS UTILIZADOS apresenta e explica conceitos financeiros básicos empregados no decorrer do trabalho.

O capítulo REVISÃO BIBLIOGRÁFICA é dividido em quatro seções. A primeira explica o modelo *Market Model* como método de simplificação da análise de risco e retorno de ativos. Em seguida, a segunda parte apresenta a estratégia *Market Neutral* de neutralização de risco sistemático, que depende fortemente da precisão da estimativa dos parâmetros do *Market Model*. A terceira parte trata de conceitos relacionados ao modelo de regressão linear, de grande importância para a aplicação detalhada apresentada neste trabalho. A última parte trata da metodologia de monitoramento de perfil linear que permite, através da detecção de amostras *fora de controle*, uma melhor estimativa dos parâmetros do *Market Model*.

O capítulo APLICAÇÃO mostra em detalhes a aplicação numérica realizada neste trabalho para ações de determinados setores em diferentes períodos. Essa seção apresenta também uma análise dos setores com base nos resultados da aplicação. Em seguida, uma análise ampla é realizada através da consolidação das análises periódicas e da reaplicação da metodologia proposta, de forma a obter estimadores considerando um horizonte de tempo mais amplo.

O capítulo CONCLUSÕES discute a eficiência do método proposto com base nos resultados obtidos e propõe potenciais utilidades do método apresentado para a indústria.

4. CONCEITOS FINANCEIROS UTILIZADOS

Este capítulo tem como objetivo esclarecer os conceitos financeiros básicos que serão abordados neste trabalho. A definição apresentada cobre os aspectos úteis para o entendimento da metodologia abordada, não sendo o objetivo deste capítulo a definição de tais conceitos de forma completa e exaustiva.

4.1. AÇÕES

Segundo Mellagi Filho e Ishikawa (2000), ações são “títulos nominativos negociáveis que representam frações de capital social de uma empresa”. Ou seja, ações são frações do capital de uma sociedade anônima. Quem possui ações detém uma parte da empresa, e por isso recebe parte proporcional dos lucros. As ações podem ou não ser negociadas em Bolsas de Valores.

As ações são negociadas em quantidades pré-estabelecidas denominadas lotes ou em quantidades menores (fracionárias de lotes). O preço de uma ação é definido de acordo com a força de oferta e de demanda dos investidores.

As principais classes de ações são *ações preferenciais*, que dão ao seu proprietário o direito a voto em assembleias da empresa emissora da ação e cuja indicação na Bolsa de Valores é o sufixo PN ou PNA adjacente ao título da ação, e as *ações ordinárias*, que não dão direito a voto e cujo sufixo é ON.

4.2. PORTFÓLIO

Um portfólio é um conjunto de um ou mais ativos. No entanto, o portfólio é considerado também um ativo. Assim, um portfólio pode ser considerado um conjunto de um ou mais ativos “puros” ou “sub-portfólios”.

Segundo Elton e Gruber (1991), todo investidor aplica sua riqueza em um conjunto de ativos, o seu portfólio. A maneira como o investidor divide sua riqueza entre esses diversos ativos depende de vários fatores, incluindo planejamento deliberado, decisões ao acaso, escolhas pessoais, entre outros.

Este trabalho discutirá modelos que propõem técnicas para escolha de um portfólio ótimo com base em dois critérios apenas: o risco e o retorno. Cabe ao investidor, no entanto, fazer escolhas com base nesses dois critérios.

4.3. RETORNO

Retorno representa o percentual acrescido ao valor de um ativo ao fim de um período em relação ao valor do ativo no início do período, dado indicado em (1).

$$Retorno = \frac{P_{fim}}{P_{ini}} - 1 \quad (1)$$

P_{fim} = valor ou preço do ativo no fim do período

P_{ini} = valor ou preço do ativo no início do período

No caso de períodos sucessivos, é válida a expressão em (2).

$$Retorno = \frac{P_t}{P_{t-1}} - 1 \quad (2)$$

P_i = valor no início do período i

No caso de ações, no entanto, quando se fala em séries temporais históricas de retorno, alterações bruscas pontuais precisam ser removidas do valor das ações. Essas alterações são devidas à emissão de proventos aos acionistas. Assim, nesses casos considera-se a relação em (3).

$$Retorno = \frac{P_t + d_t}{P_{t-1}} - 1 \quad (3)$$

P_i = valor no início do período i

d_i = proventos emitidos aos acionistas no período i

4.4. AVERSÃO AO RISCO

Elton e Gruber (1997) descrevem que uma das principais questões com a qual o homem sempre se deparou é a de como alocar sua riqueza entre diferentes ativos. Um conceito importante para o entendimento de teorias que abordam essa questão, como a Moderna Teoria do Portfólio proposta por Markowitz (1952), é o conceito da aversão ao risco. Friedman e Savage (1948) e Markowitz (1952) introduziram tal conceito. A aversão ao risco refere-se à atitude frente ao risco de investidores possuindo as seguintes características:

- Preferência por ativos de baixo risco em detrimento a ativos de alto risco para um mesmo nível de retorno esperado;
- Preferência por um ativo com risco superior a outro somente se esse ativo oferecer maior retorno esperado.

Um investidor possuindo essas características é considerado avesso ao risco. Como propuseram Pratt (1964) e Arrow (1971), a aversão ao risco pode ser quantificada, podendo-se dizer que existem investidores mais avessos ao risco que outros, e riscos adicionais precisam ser recompensados com alguma vantagem (chamada de *utilidade* nesse contexto) adicional.

4.5. CONTRATO FUTURO DE IBOVESPA

Segundo o trabalho Futuro [...] (2006), o Contrato Futuro de Ibovespa tem como objeto de negociação o índice de ações da Bolsa de Valores de São Paulo, o Ibovespa. Sua cotação é dada em pontos desse índice, com o valor de cada ponto estabelecido pela BM&F Bovespa, empresa gestora da Bolsa de Valores. Operações com futuros se diferenciam das operações à vista por possuírem uma data predeterminada de liquidação. Essa liquidação financeira, através de operação de venda com data pré-fixada no momento da compra, evita a entrega física das ações que compõe o índice.

5. REVISÃO BIBLIOGRÁFICA

Este capítulo desenvolve os principais modelos, teorias e ferramentas que serão aplicados numericamente neste trabalho. Assim, o capítulo abordará o modelo usado no trabalho para a medida de risco (*Market Model*), a estratégia de neutralização de risco sistemático *Market Neutral*, uma breve descrição do modelo de regressão linear múltipla e testes de hipóteses para esse modelo e, por fim, as ferramentas de controle estatístico da qualidade que são o foco deste trabalho.

Será mostrada a importância do monitoramento e da estimativa dos parâmetros do *Market Model* para a estratégia *Market Neutral*. Será demonstrada também a eficiência das ferramentas do controle estatístico da qualidade para cumprir esse objetivo. O modelo de regressão linear múltipla e os testes de hipóteses são técnicas que farão parte da aplicação numérica elaborada neste trabalho, o que justifica uma breve revisão desses conceitos com foco nessa aplicação.

5.1. MARKET MODEL

Esta seção visa explicar o modelo *Market Model* e os princípios nos quais tal modelo se baseia. Nesse contexto, é fundamental entender a importância da determinação do retorno esperado de ativos. A abordagem que o *Market Model* propõe sobre esse tópico vem da teoria de portfólio e, portanto, será feita uma breve abordagem sobre essa teoria. Em seguida, como a noção de risco de *beta* advinda do CAPM é um dos alicerces do *Market Model*, a mesma será apresentada sob tal ótica. Por fim, com esses conceitos já detalhados, será apresentada a formulação do *Market Model*. O esquema da Figura 2 mostra a estrutura desta seção.

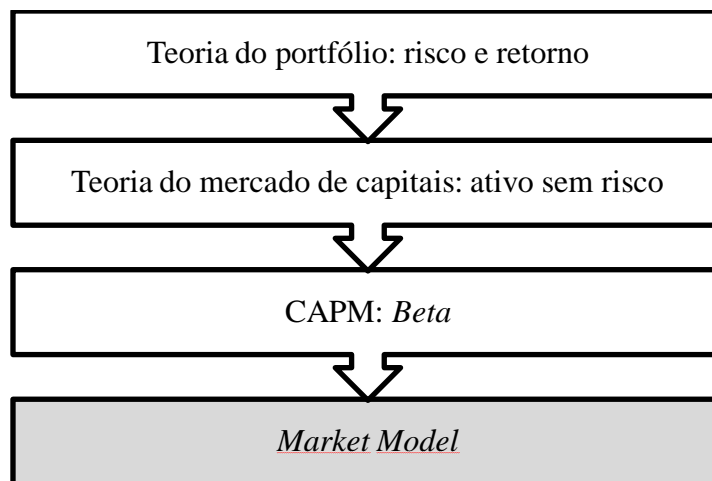


Figura 2 – Esquema dos assuntos abordados no capítulo

Markowitz (1952) propôs um modelo teórico para escolha de ativos visando atingir a melhor relação entre retorno e risco. Segundo Sharpe, Alexander et al. (1995), no que concerne o comportamento dos investidores, o modelo tem como principais premissas:

- Investidores maximizam sua utilidade esperada para um determinado horizonte de tempo.
- Investidores entendem risco como a variância dos retornos esperados de um ativo.
- Os únicos critérios levados em conta por investidores para fazer escolhas entre oportunidades de investimento são o risco e o retorno proporcionados por essas oportunidades.
- Investidores examinam oportunidades de investimento como uma distribuição de probabilidade de retornos esperados gerados pelos ativos num dado horizonte de tempo.
- Investidores são avessos ao risco. Dados dois investimentos de mesmo valor esperado de retorno, investidores optarão pelo de menor risco. Da mesma forma, dados dois investimentos de mesmo risco, eles optarão pelo de maior valor esperado de retorno, nunca estão satisfeitos com a riqueza já possuída.

É amplamente difundido na literatura que o valor esperado de retorno R de um ativo é determinado pela relação expressa em (4).

$$E(R) = \sum_{i=1}^n P_i R_i = P_1 R_1 + P_2 R_2 + \cdots + P_n R_n \quad (4)$$

$E(R)$ = valor esperado do retorno do ativo R
 P_i = probabilidade de ocorrência da situação i
 R_i = retorno do ativo R caso a situação i ocorra

E para um portfólio $R_p = \sum w_i R_i$, seu retorno esperado e desvio padrão serão dados respectivamente por (5) e (6).

$$E(R_p) = \sum_{i=1}^n w_i E(R_i) \quad (5)$$

R_p = retorno do portfólio
 w_i = parcela do ativo i no portfólio
 R_i = retorno do ativo i

$$\begin{aligned} \sigma_p &= \sqrt{\sum_{i=1}^n w_i^2 \sigma_i^2 + \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n w_i w_j \text{Cov}_{i,j}} \\ &= \sqrt{\sum_{i=1}^n w_i^2 \sigma_i^2 + \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n w_i w_j \rho_{i,j} \sigma_i \sigma_j} \quad (6) \end{aligned}$$

σ_p = desvio padrão do portfólio
 σ_i = desvio padrão do ativo i
 $\text{Cov}_{1,2}$ = covariância entre os retornos dos ativos 1 e 2
 $\rho_{1,2}$ = coeficiente de correlação entre ativos 1 e 2

A expressão (6), do desvio padrão do portfólio, mostra que **o risco de um portfólio depende de 3 fatores: a participação de cada ativo no valor do portfólio, o risco (desvio padrão) de cada ativo e a correlação entre os ativos**. Como a variância é o quadrado do desvio padrão e a covariância é a correlação multiplicada pelos desvios padrão, pode-se de outro modo dizer que o risco do portfólio depende da variância dos ativos e da covariância entre eles, além da participação dos próprios ativos no portfólio.

As expressões (5) e (6) podem ser escritas na forma matricial, sendo respectivamente dadas por:

$$\begin{aligned}
 & E(R_p) = \mathbf{w}'\boldsymbol{\mu} \\
 & \sigma_R = \sqrt{\mathbf{w}'\boldsymbol{\Sigma}\mathbf{w}} \\
 & \text{Com} \\
 & \boldsymbol{\mu} = \begin{pmatrix} E(R_1) \\ E(R_2) \\ \vdots \\ E(R_g) \end{pmatrix} \\
 & \mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_g \end{pmatrix} \\
 & \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{1,1} & \dots & \sigma_{1,g} \\ \vdots & \ddots & \vdots \\ \sigma_{g,1} & \dots & \sigma_{g,g} \end{pmatrix}
 \end{aligned} \tag{7}$$

R_p = retorno do portfólio

$\boldsymbol{\mu}$ = matriz dos retornos esperados

R_i = retorno do ativo i

w = matriz das contribuições relativas dos ativos

w_i = contribuição relativa do ativo i

$\boldsymbol{\Sigma}$ = matriz de correlação

$\sigma_{i,i}$ = variância do ativo i

$\sigma_{i,j}$ = coeficiente de correlação entre ativos i e j

A matriz $\boldsymbol{\Sigma}$ em (7) é chamada **matriz de variância-covariância**.

É importante notar que **o máximo risco de um portfólio ocorre quando os ativos que o compõem são perfeitamente correlacionados** (coeficiente de correlação igual a 1). Se dois ou mais ativos são perfeitamente correlacionados não há benefício de diversificação em termos de redução de risco.

Segundo Ross, Westerfield et al. (2005), quanto menor a correlação entre ativos, maiores os benefícios possíveis da diversificação, ou seja, pode-se obter menor risco para um mesmo nível de retorno esperado ou maior retorno esperado para mesmo nível de risco, como mostrado na Figura 3.

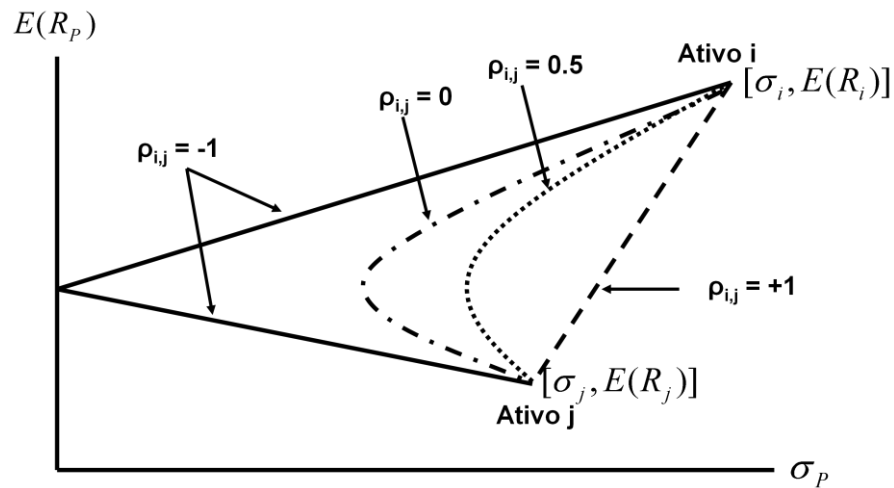


Figura 3 – Benefícios da baixa correlação entre ativos

Adaptado de Elton e Gruber (1991)

A teoria de portfólio de Markowitz determina que um portfólio que, dentre todos os portfólios possíveis, oferece o maior retorno para seu nível de risco é chamado de *portfólio eficiente*. O conjunto de todos os portfólios eficientes é chamado de *fronteira eficiente*. A Figura 4 mostra a representação gráfica da *fronteira eficiente* no plano de retorno esperado versus risco.

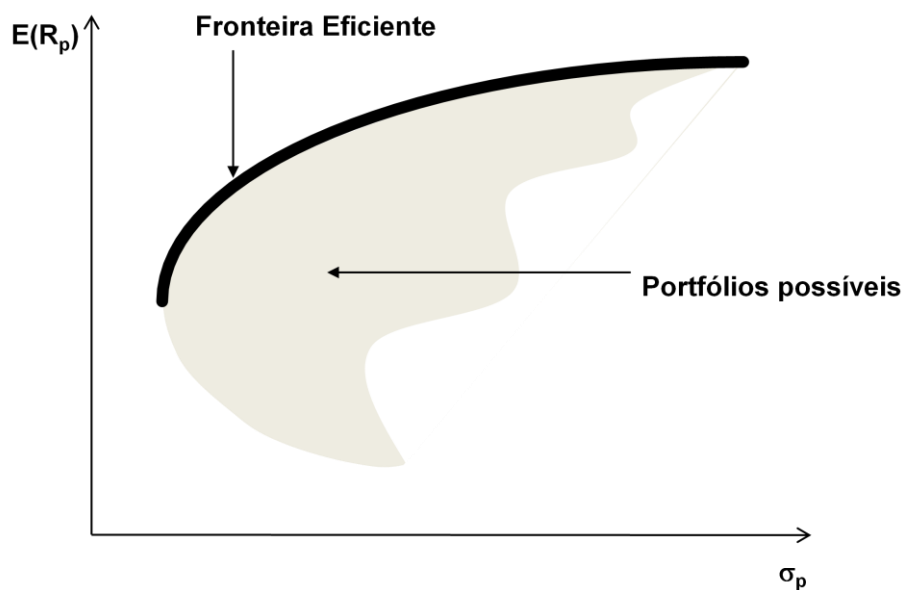


Figura 4 – Representação gráfica da fronteira eficiente

Adaptado de Bodie, Kane et al. (2008)

Dada a noção de retorno esperado de um ativo utilizada no *Market Model*, prossegue-se com outros elementos que permitirão a compreensão da noção de risco sistemático (*beta*) empregada na aplicação mostrada neste trabalho.

Sharpe (1964) e Lintner (1965) mostraram que combinações de um ativo sem risco (inexistente no modelo de portfólio) com qualquer portfólio da *fronteira eficiente*, quando representadas graficamente, estarão contidas numa linha reta entre um ponto representando o ativo sem risco e outro representando um portfólio denominado ***portfólio de mercado***, contido na *fronteira eficiente*. Essa teoria é chamada de ***teoria do mercado de capitais (capital market theory)***.

Segundo Reilly e Brown (2006) as oito hipóteses iniciais da teoria de mercado de capitais são:

- Os investidores do modelo de Markowitz apresentam comportamento conforme descrito no modelo de portfólio de Markowitz, ou seja, usam retorno esperado e variância para escolher ativos, e selecionam portfólios contidos na *fronteira eficiente*.
- Investidores podem emprestar e tomar emprestadas infinitas quantidades de capital pagando uma taxa de retorno sem risco.
- Investidores têm expectativas iguais quanto às distribuições de probabilidade dos ativos disponíveis, o que os leva a um único conjunto de portfólio ótimo.
- Todos os investidores têm um horizonte de tempo de um período.
- Ativos são infinitamente divisíveis.
- Não existem custos de transação.
- Não há inflação e taxas de juros são constantes.
- O mercado de capitais está em equilíbrio, ou seja, todos os investimentos estão corretamente avaliados de acordo com seu nível de risco.

A adição de um ativo sem risco à teoria de portfólio de Markowitz leva à teoria do mercado de capitais, mudando a *fronteira eficiente* de uma curva para uma linha reta chamada *linha do mercado de capitais* (LMC). Segundo Reilly e Brown (2006), esse ativo sem risco possui correlação nula com outros ativos, fornece uma taxa de retorno constante denominada *taxa sem risco* (TSR) e é representado no eixo vertical do gráfico retorno esperado versus risco (risco nulo). Além disso, por apresentar risco nulo, o desvio padrão de sua distribuição de retorno é nulo.

O retorno esperado de um portfólio formado de uma parcela investida num portfólio P e o restante no ativo sem risco tem seu retorno esperado dado por (8).

$$E(R_{pSR}) = w_{SR}TSR + (1 - w_{SR})E(R_p)$$

$$\begin{aligned} R_{pSR} &= \text{retorno do portfólio total} \\ w_{SR} &= \text{parcela relativa investida no ativo sem risco} \\ TSR &= \text{taxa de retorno sem risco} \\ R_p &= \text{retorno do portfólio envolvendo ativo com risco} \end{aligned} \quad (8)$$

Ao formar um portfólio a partir de um portfólio P e do ativo sem risco, o cálculo de seu desvio padrão (risco) é dado pela equação para o desvio padrão de um portfólio de dois ativos. Porém, como o desvio padrão do ativo sem risco é zero, assim como sua correlação com outros ativos, o desvio padrão resultante é diretamente proporcional à parcela investida no ativo com risco e é dado pela expressão (9).

$$\sigma_{pSR} = (1 - w_{SR})\sigma_p$$

$$\begin{aligned} w_{SR} &= \text{parcela relativa investida no ativo sem risco} \\ \sigma_{pSR} &= \text{desvio padrão do portfólio total} \\ \sigma_p &= \text{desvio padrão do portfólio com risco} \end{aligned} \quad (9)$$

Usando (9) em (8) e reagrupando, obtém-se a expressão (10).

$$E(R_{pSR}) = TSR + \sigma_{pSR} \left\{ \frac{[E(R_p) - TSR]}{\sigma_p} \right\}$$

$$\begin{aligned} R_{pSR} &= \text{retorno do portfólio total} \\ TSR &= \text{taxa de retorno sem risco} \\ R_p &= \text{retorno do portfólio envolvendo ativo com risco} \\ \sigma_{pSR} &= \text{desvio padrão do portfólio total} \\ \sigma_p &= \text{desvio padrão do portfólio com risco} \end{aligned} \quad (10)$$

Essa equação é conhecida como **linha do mercado de capitais (LMC)**. O retorno e o risco desse portfólio estão diretamente relacionados. A LMC representa todos os possíveis portfólios compostos por ativos com risco combinados com o ativo sem risco.

No gráfico retorno esperado versus risco, os portfólios representados pela linha que começa no ponto do ativo sem risco e que tange a *fronteira eficiente* são dominantes sobre os demais portfólios por oferecerem maior retorno esperado para um mesmo nível de risco. O portfólio representado por esse ponto de tangência será doravante denominado *portfólio M* ou **portfólio de mercado**, pois todos os investidores irão optar por investir neste único portfólio

(entre os portfólios constituídos unicamente por ativos com risco). A Figura 5 representa esses conceitos.

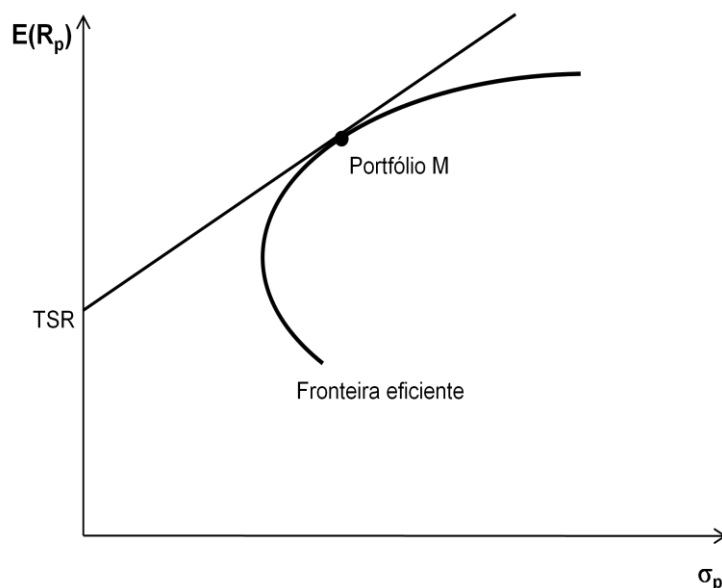


Figura 5 – O portfólio de mercado e a LMC
Adaptado de Elton e Gruber (1991)

Segundo Reilly e Brown (2006), como todos os investidores optarão pelo portfólio de mercado como ativo com risco, esse portfólio deve incluir todos os *ativos com risco* disponíveis no mercado. Se um ativo não está contido nesse portfólio, não haverá demanda por ele e, portanto, ele não terá valor algum. Como a oitava hipótese do modelo de mercado de capitais é a de mercado em equilíbrio, todos os ativos com risco devem estar inclusos no portfólio de mercado com parcelas proporcionais ao seu valor relativo ao valor do mercado como um todo.

Os ativos do mercado não são todos perfeitamente correlacionados. Assim, como mostrado anteriormente, o risco do portfólio de mercado é menor que a soma ponderada dos riscos dos ativos do mercado. Esse risco que desaparece no processo de diversificação é chamado de **risco não sistemático**, também chamado de *risco único* ou *risco diversificável*. O risco remanescente é chamado de **risco sistemático**, também chamado de *risco de mercado*. Este risco pode ser representado pelo *beta*, conceito chave deste trabalho. Ele não pode ser removido por diversificação, já que é atrelado a variáveis macroeconômicas e não depende de especificidades do ativo, como resultados financeiros e operacionais no caso de uma empresa,

mas sim da sua atividade. Tal risco está relacionado com o desempenho geral da macroeconomia, afetando da mesma maneira todas as empresas de um setor, já que age sobre o mercado no qual essas empresas estão inseridas.

O risco total de um portfólio qualquer (de um ou mais ativos) é formado por risco sistemático e risco não sistemático. A Figura 6 ilustra o papel da diversificação para redução do risco total.

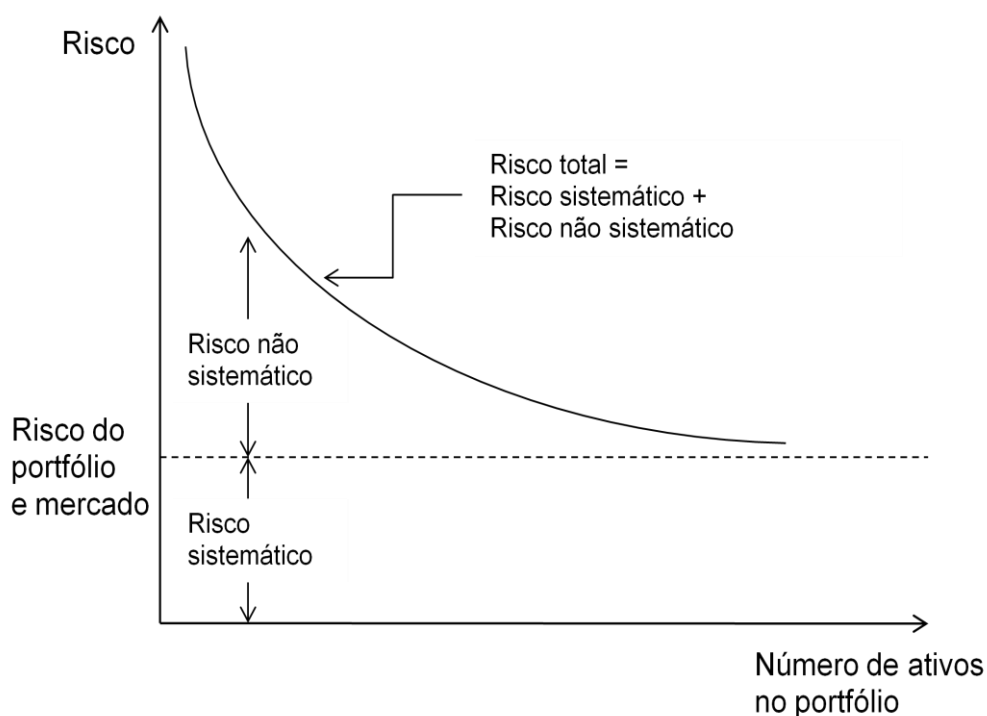


Figura 6 – Diversificação e os diferentes tipos de risco

Adaptado de Bodie, Kane et al. (2008)

Como, segundo a teoria de mercado de capitais, os investidores não têm custo para diversificação, os valores de equilíbrio dos ativos dependem apenas do risco não sistemático. Ou seja, o valor esperado do retorno de um ativo é relacionado apenas com seu risco sistemático e não com seu risco total.

A covariância entre o retorno de um ativo e o retorno do portfólio de mercado é uma medida para o risco sistemático desse ativo. Quanto maior a covariância, menor a parcela do seu risco que será eliminada através da diversificação (já que o portfólio de mercado é constituído por todos os ativos sem risco).

A partir do modelo de mercado de capitais, com foco nessa nova medida de risco, os autores Sharpe (1964), Lintner (1965) e Mossin (1966) desenvolveram paralelamente um modelo que descreve as taxas de retorno esperadas para ativos com risco numa situação de equilíbrio de mercado (ativos corretamente precificados).

Segundo Securato (1996), o modelo que relaciona retorno esperado com risco sistemático (*beta do CAPM*), é denominado **modelo de precificação de ativos financeiros**, nomenclatura proposta por Securato (1996), ou **CAPM**, da sigla em inglês *Capital Asset Pricing Model*. Segundo Gitman (2004), o modelo CAPM é descrito pela equação (12), e sua representação gráfica, ilustrada na Figura 7, é chamada de **linha do mercado de títulos (LMT)**, denominação proposta por Gitman (2004). A equação (12) é derivada de (11).

$$E(R_i) = TSR + Cov_{i,M} \left\{ \frac{[E(R_M) - TSR]}{\sigma_M^2} \right\}$$

$$R_i = \text{Retorno do portfólio } i \quad (11)$$

$$R_M = \text{Retorno do portfólio de mercado}$$

$$TSR = \text{Taxa de retorno sem risco}$$

$$\sigma_M = \text{desvio padrão do portfólio de mercado}$$

$$Cov_{i,M} = \text{covariância entre ativo } i \text{ e mercado}$$

$$E(R_i) = TSR + \beta_i [E(R_M) - TSR]$$

Com

$$\beta_i = \frac{Cov_{i,M}}{\sigma_M^2} \quad (12)$$

$$R_i = \text{Retorno do portfólio ou ativo } i$$

$$R_M = \text{Retorno do portfólio de mercado}$$

$$TSR = \text{Taxa de retorno sem risco}$$

$$\beta_i = \text{Beta do portfólio ou ativo } i$$

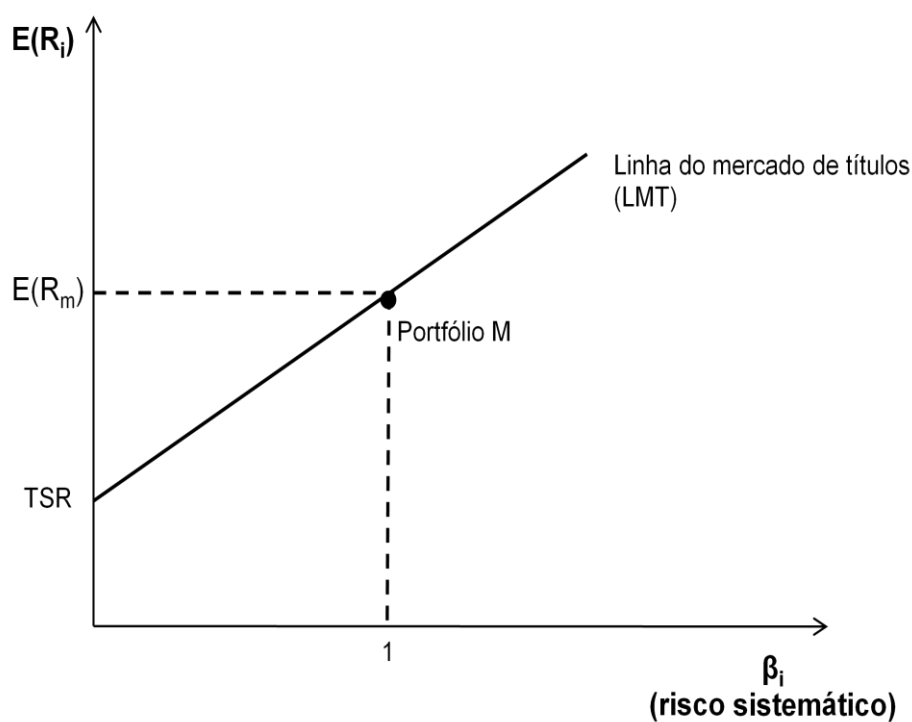


Figura 7 – A linha do mercado de títulos
Adaptado de Bodie, Kane et al. (2008)

O termo $\frac{Cov_{i,M}}{\sigma_M^2}$ é chamado de **beta** do modelo do modelo CAPM e constitui uma *medida padronizada de risco sistemático*.

Segundo Sharpe (1990), o CAPM é um modelo de equilíbrio. Em suma, o modelo atesta que, no equilíbrio, o retorno esperado de um ativo é a taxa sem risco mais um prêmio de risco ajustado pelo *beta*:

$$\beta_i \cdot [E(R_M) - TSR].$$

O CAPM estima o retorno esperado de um ativo a partir de três parâmetros: o retorno esperado do portfólio de mercado, o coeficiente *beta* do ativo e a taxa sem risco.

Quanto à diferença entre as linhas, vale ressaltar que a LMC tem o risco total como variável explicativa. Graficamente, a LMC tem o risco total no eixo das abscissas. Dessa maneira, *somente portfólios eficientes estarão representados na LMC*. Já a LMT tem o *beta*, risco sistemático, no eixo das abscissas. Assim, *todos os ativos corretamente avaliados (no equilíbrio) estarão representados na LMT*.

Segundo Sharpe (1990), as implicações-chave do CAPM são três:

- O portfólio de mercado é eficiente (maior retorno esperado para seu nível de risco);
- Todos os portfólios são equivalentes a um investimento no portfólio de mercado mais, possivelmente, o empréstimo ou tomada de empréstimo à taxa sem risco;
- Há relação linear entre retorno esperado e *beta*.

Definidos os conceitos necessários ao **Market Model** o modelo é agora apresentado em sua forma final. Para chegar-se ao portfólio ótimo, o modelo de Markowitz e o CAPM não podem ser aplicados na prática por dois motivos.

Primeiro, segundo Bodie, Kane et al. (2008) e Jones (1998), o modelo CAPM, assim como a própria teoria do portfólio trata de *expectativa* de retornos (evento *ex ante*). Além disso, o modelo CAPM refere-se à situação de *equilíbrio*. Por tratar apenas da situação de equilíbrio e de expectativas (esperança) dos retornos e não de valores efetivos, o modelo CAPM não se refere a análises posteriores ao evento (*ex post*) para se chegar ao conjunto de portfólios ótimos da *fronteira eficiente*. Sharpe (1990, p. 320) afirma “O CAPM não faz hipóteses sobre o processo de geração de retorno dos ativos, ele trata apenas do estado de equilíbrio. Assim, os resultados do CAPM são consistentes com qualquer processo que descreva a geração desses retornos.”.

Alguns modelos, partindo do modelo de Markowitz baseado em esperança e variância como critérios de seleção de portfólios, descrevem a relação entre a geração de retornos dos ativos e tratam do problema de seleção do portfólio ótimo.

Segundo Elton e Gruber (1991), o modelo de Markowitz determina que, para realizar a análise de um portfólio de *g* ativos e, posteriormente, definir a *fronteira eficiente*, são

necessários, conforme (7), os seguintes dados de entrada: g valores esperados, g desvios-padrão e $g(g-1)/2$ covariâncias pareadas (entre dois ativos), totalizando $g(g+3)/2$ parâmetros.

Alternativamente, Sharpe (1963) propôs um modelo simplificado no qual todos os ativos dependem de um único fator indicador do desempenho do mercado. Esse modelo é conhecido por *single-index model*. Há grande confusão na literatura quanto à nomenclatura dos modelos. A nomenclatura usada neste trabalho é a mais amplamente reconhecida e usada em trabalhos de grande reconhecimento internacional.

Posteriormente, Fama (1976) propôs outro modelo com hipóteses menos restritivas, conhecido hoje como *Market Model*.

Os modelos mencionados consistem em relações lineares entre o retorno em excesso (retorno menos a taxa de retorno sem risco) de um ativo e o retorno em excesso de um parâmetro de mercado. Como afirma Bodie, Kane et al. (2008), na prática, são usados retornos totais ao invés de retornos em excesso quando se trata de retornos diários. Segundo Reilly e Brown (2006) e Jones (1998), a reta obtida por regressão linear entre esses retornos usando dados *ex post* é chamada de *linha característica* (*characteristic line*). Ela é mostrada na Figura 8.

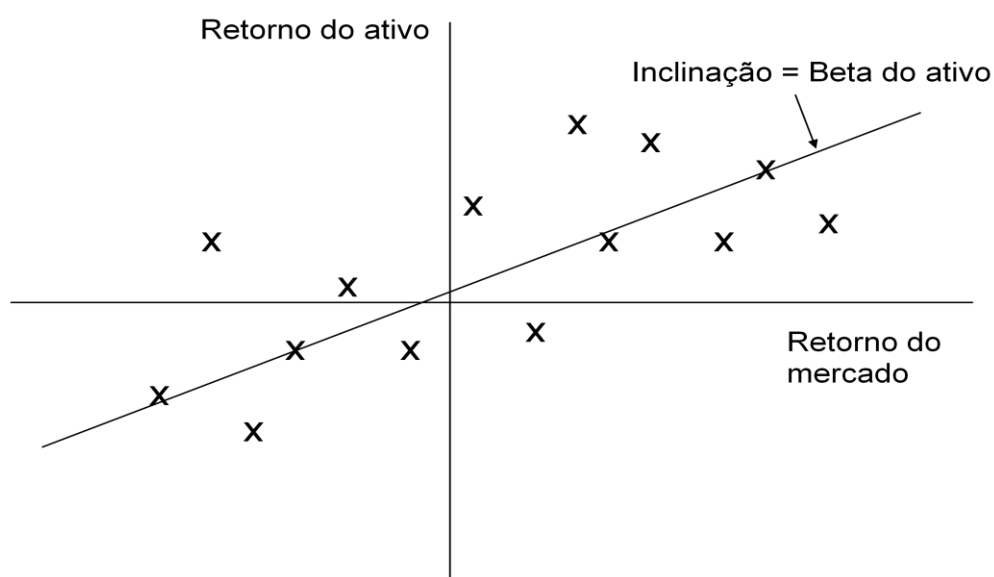


Figura 8 – A linha característica

Segundo Elton e Gruber (1991) e Hallerbach (1994), o *Market Model* é definido pela seguinte equação (13), cujas premissas são mostradas em (14). Neste trabalho será usado o *Market Model* com retornos totais ao invés de em excesso. Como mostrado por Bodie, Kane et al. (2008), a diferença nos cálculos não é significativa e tal adaptação representa grande simplificação prática.

$$r_i = \alpha_i + \beta_i r_m + u_i$$

$$\begin{aligned} \alpha_i &= \text{intercepto para o ativo } i \\ \beta_i &= \text{coeficiente angular (beta) do ativo } i \\ r_i &= \text{retorno do ativo } i \\ r_m &= \text{retorno do portfólio de mercado} \\ u_i &= \text{ruído branco relativo ao ativo } i \end{aligned} \quad (13)$$

Com as seguintes premissas:

$$\begin{aligned} E(u_i) &= 0 \text{ (por construção)} \\ u_i \text{ e } r_i &\text{ independentes (por hipótese)} \\ u_i &= \text{ruído branco relativo ao ativo } i \\ Cov &= \text{covariância} \\ \sigma_{i,j} &= \text{coeficiente de correlação entre ativos } i \text{ e } j \end{aligned} \quad (14)$$

O *beta* do *Market Model* é calculado de maneira análoga ao *beta* do CAPM, conforme equação (12). Nesse modelo, a variância de um ativo *i* e a covariância entre dois ativos *i* e *j* são dadas respectivamente por:

$$\begin{aligned} \sigma_i^2 &= \sigma_{i,i} = Cov(\alpha_i + \beta_i r_m + u_i, \alpha_i + \beta_i r_m + u_i) \\ &= Cov(\beta_i r_m, \beta_i r_m) + \sigma_{u_i}^2 \\ &= \beta_i^2 \sigma_M^2 + \sigma_{u_i}^2 \end{aligned}$$

e

$$\begin{aligned} \sigma_{i,j} &= Cov(\alpha_i + \beta_i r_m + u_i, \alpha_j + \beta_j r_m + u_j) \\ &= Cov(\beta_i r_m, \beta_j r_m) + \sigma_{u_i, u_j} \\ &= \beta_i \beta_j \sigma_M^2 + \sigma_{u_i, u_j} \end{aligned} \quad (15)$$

$$\begin{aligned} \alpha_i &= \text{intercepto para ativo } i \\ \beta_i &= \text{coeficiente angular (beta) do ativo } i \\ r_i &= \text{retorno em excesso do ativo } i \\ r_m &= \text{retorno em excesso do portfólio de mercado} \\ u_i &= \text{ruído branco relativo ao ativo } i \\ Cov &= \text{covariância} \\ \sigma_{i,j} &= \text{coeficiente de correlação entre ativos } i \text{ e } j \\ \sigma^2 &= \text{variância} \end{aligned}$$

Assim, como visto em (15), a variância de um ativo nesse modelo é composta por dois fatores, sendo: ***Variância total = variância relacionada ao mercado (sistemática, relacionada ao beta) + variância não relacionada ao mercado (não sistemática).***

Retomando as expressões (5), (6) e (7) que tratam de um portfólio de g ativos, e considerando que no modelo *Market Model*, o desvio-padrão do portfólio é dado por um termo de variância relacionado ao mercado ($\beta_i^2 \cdot \sigma_{r_m}^2$) e um termo de variância não relacionado ao mercado ($\sigma_{u_i}^2$), chega-se à expressão em (16).

$$\sigma_p = \sqrt{\sum_{i=1}^g w_i^2 \beta_i^2 \sigma_M^2 + \sum_{i=1}^g \sum_{\substack{j=1 \\ i \neq j}}^g w_i w_j \beta_i \beta_j \sigma_M^2 + \sum_{i=1}^g w_i^2 \sigma_{u_i}^2 + \sum_{i=1}^g \sum_{\substack{j=1 \\ i \neq j}}^g w_i w_j \sigma_{u_i u_j}} \quad (16)$$

β_i = coeficiente angular (beta) do ativo i
 w_i = parcela relativa do ativo i no portfólio
 σ_{ij} = coeficiente de correlação entre ativos i e j
 σ^2 = variância

A expressão (16), utilizando-se notação matricial, é dada por (17).

Com

$$\sigma_p = \sqrt{\boldsymbol{\beta}' \mathbf{w}' \boldsymbol{\Sigma} \mathbf{w} \boldsymbol{\beta} + \mathbf{w}' \mathbf{Z} \mathbf{w}}$$

$$\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{1,1} & \dots & \sigma_{1,n} \\ \vdots & \ddots & \vdots \\ \sigma_{n,1} & \dots & \sigma_{n,n} \end{pmatrix} \quad (17)$$

$$\mathbf{Z} = \begin{pmatrix} \sigma_{u_{1,1}} & \dots & \sigma_{u_{1,n}} \\ \vdots & \ddots & \vdots \\ \sigma_{u_{n,1}} & \dots & \sigma_{u_{n,n}} \end{pmatrix}$$

σ_p = desvio padrão do portfólio

β_i = coeficiente angular (beta) do ativo i

w_i = parcela relativa do ativo i no portfólio

$\boldsymbol{\Sigma}$ = matriz de correlação

\mathbf{Z} = matriz de perturbações

$\sigma_{i,i}$ = variância do termo i

$\sigma_{i,j}$ = coeficiente de correlação entre termos i e j

Neste trabalho será usado o *Market Model* "modificado", considerando o retorno total e não o retorno em excesso em relação à taxa sem risco. Como comentado por Bodie, Kane et al. (2008), essa prática é comum quando são usados preços diários, como no caso deste trabalho, já que nesse caso o retorno da taxa sem risco representa em média menos de 1% da variação de uma ação, sendo imperceptível sua influência nos cálculos.

Segundo Bodie, Kane et al. (2008), a relação da linha característica dos modelos *Market Model* e *single-index model* com o CAPM é que, na situação de equilíbrio do CAPM, os termos α tanto do *Market Model* quanto do *single-index model* são iguais a zero. A demonstração pode ser feita pelas equações (12) e (13), aplicando-se o valor esperado aos dois lados dessa segunda equação. Ainda segundo Bodie, Kane et al. (2008), de maneira análoga, pode-se verificar que, no equilíbrio, as linhas características dos modelos *Market Model* e *single-index model* determinam que, se o retorno total do fator de mercado é a taxa

sem risco (retorno em excesso nulo), o retorno esperado para os ativos também é a taxa sem risco (retorno em excesso nulo).

5.2. ESTRATÉGIA MARKET NEUTRAL

Esta seção mostra uma estratégia de proteção ao risco muito utilizada na área financeira, fortemente apoiada nos estimadores dos parâmetros do *Market Model*. A melhoria na estimação desses parâmetros e seu monitoramento, assuntos objeto deste trabalho, são de fundamental importância para aplicação de tal estratégia.

De acordo com o *Market Model*, assume-se uma relação linear entre o retorno de um ativo e um fator de mercado, como visto na equação (13). O fator de mercado pode ser o retorno de um índice da Bolsa de Valores, que reflete a média dos retornos ponderados de um conjunto de ações representativas da economia, como o Ibovespa da Bolsa de Valores Bovespa. Para o retorno r_i de um ativo i , considerando-se o retorno do índice de mercado representado por r_m , tem-se a relação dada por (18).

$$E(r_i|r_m) = \alpha_i + \beta_i r_m \quad (18)$$

Um estimador do valor esperado de r_i é mostrado em (19).

$$\hat{r}_i = \hat{\alpha}_i + \hat{\beta}_i r_m \quad (19)$$

Vale lembrar que o *beta* do *Market Model* é calculado de maneira análoga ao *beta* do CAPM, conforme equação (12). Pode-se usar essa abordagem para um portfólio de g ativos. Dado o peso W_i do ativo i no portfólio, o retorno r_p esperado para o portfólio é:

$$r_p = \sum_{i=1}^g W_i r_i$$

Com

$$\sum_{i=1}^g W_i = 1 \quad (20)$$

Por (20) e (13), tem-se a expressão (21).

$$r_p = \sum_{i=1}^g W_i \alpha_i + \sum_{i=1}^g W_i \beta_i r_m + \sum_{i=1}^g W_i \epsilon_i \quad (21)$$

Se um montante C for investido no portfólio P , o montante final \hat{R}_i esperado para um ativo i nesse investimento é dado por (22).

$$\hat{R}_i = CW_i \hat{\alpha}_i + CW_i \hat{\beta}_i r_m \quad (22)$$

Segundo o trabalho Futuro [...] (2006), para eliminar o *risco sistemático* relacionado ao ativo i desse portfólio, contratos futuros do índice no valor de $CW_i \beta_i$ precisam ser vendidos. Essa operação de neutralização de risco sistemático de uma carteira de ações mediante a venda de contratos futuros de índice é chamada de *hedge de venda* ou *short hedge*.

Seja r_f o retorno de cada contrato futuro, o montante final \hat{R} da estratégia é dado por:

$$\hat{R} = CW_i \hat{\alpha}_i + CW_i \hat{\beta}_i r_m - CW_i \hat{\beta}_i r_f \quad (23)$$

Em (23), se $r_f = r_m$, o que é esperado ao se usar o índice da Bolsa de Valores como portfólio indicador do desempenho da economia, então o risco sistemático é eliminado e tem-se a relação (24).

$$\hat{R} = CW_i \hat{\alpha}_i \quad (24)$$

Pode se prosseguir de maneira análoga para a neutralização do risco sistemático do portfólio P .

Veja que a quantidade a ser vendida depende do montante investido (C), do peso do ativo no portfólio (W_i) e do β . Destas variáveis, somente o β do ativo não depende de decisão do investidor, as demais podendo ser controladas por ele. Isso mostra a importância do monitoramento do *beta* e da precisão de sua avaliação para a neutralização do risco.

Uma estratégia comum de neutralização no caso de apenas uma ação é, portanto, a de estabelecer uma posição comprada no ativo i e uma posição vendida em contratos futuros do Ibovespa, na proporção equivalente ao β estimado. O retorno de tal estratégia é:

$$r_i - \hat{r}_i = \alpha_i - \hat{\alpha}_i + \beta_i r_m - \hat{\beta}_i r_m = (\alpha_i - \hat{\alpha}_i) + (\beta_i - \hat{\beta}_i) r_m$$

$$\begin{aligned} \beta_i &= \text{beta "verdadeiro" do ativo } i \\ \hat{\beta}_i &= \text{estimador do beta do ativo } i \\ (\alpha_i - \hat{\alpha}_i) &= \text{componente não dependente do mercado} \end{aligned} \quad (25)$$

Ou seja, tenta-se neutralizar o termo relacionado ao mercado, dado pelo seu β real (que nunca se consegue obter) multiplicado pelo retorno do índice de mercado, através da venda de contratos futuros na medida mais próxima possível desse valor, que é o β estimado multiplicado pelo índice. Quanto melhor se estima o β , melhores os resultados da estratégia, ou seja, mais próximo de zero fica o componente de retorno dependente do mercado da expressão (25), dado por $(\beta_i - \hat{\beta}_i) r_m$. Aproximar esse termo de zero é o objetivo da estratégia. Este resultado será usado mais adiante para testar a eficácia da metodologia no sentido de cumprir os objetivos da estratégia.

5.3. REGRESSÃO

Esta seção aborda alguns tópicos sobre o modelo de regressão linear que serão usados em diversas fases deste trabalho. Seu entendimento é importante para a correta leitura da aplicação que será detalhada. Serão discutidas brevemente algumas das bases do modelo de regressão linear múltipla, a escolha do método dos mínimos quadrados para estimativa dos parâmetros da regressão e suas premissas, algumas estatísticas importantes para esse método, e testes de hipótese aplicados à regressão que serão empregados neste trabalho.

O capítulo dará ênfase à notação matricial dos cálculos relacionados à regressão e testes de hipóteses. Esta notação é particularmente útil para cálculos computacionais, como no caso do programa desenvolvido neste trabalho para aplicação numérica dos modelos.

5.3.1. REGRESSÃO LINEAR MÚLTIPLA

Segundo Gujarati (2006), o termo *regressão* ou *modelo de regressão* significa a obtenção de uma função de esperança condicional, ou seja, uma função que retorna o valor de uma variável dita *explicada* com base nos valores de outras variáveis, ditas *explicativas*. Essa função, chamada de *função de regressão*, é indicada pela relação (26).

$$E(Y|X_1X_2 \cdots X_p) = f(X_1, X_2, \cdots, X_p) \quad (26)$$

$Y = \text{variável explicada}$
 $X_i = \text{variáveis explicativas (regressores)}$

O termo *regressão linear* refere-se a um modelo de regressão em que há *linearidade nos parâmetros*, havendo ou não *linearidade nas variáveis*. Por linearidade nos parâmetros entende-se que os coeficientes que multiplicam as variáveis explicativas na função de regressão têm expoente 1 e não são correlacionados. Para exemplificar, um modelo de regressão no qual a esperança da variável explicada é proporcional ao quadrado de uma das variáveis explicativas não é linear na variável. Porém, caso o coeficiente que multiplica o quadrado da variável explicativa tenha expoente 1, assim como os coeficientes de todas as variáveis explicativas, esse modelo é linear nos parâmetros e portanto pode ser chamado de *modelo de regressão linear*.

$$E(Y|X_aX_bX_c) = f(X_a, X_b, X_c) = \beta_0 + \beta_1^i X_a^\gamma + \beta_2^j X_b^\delta + \beta_3^k X_c^\varepsilon \quad (27)$$

$Y = \text{variável explicada}$
 $X_i = \text{variáveis explicativas (regressores)}$
 $\beta_0 = \text{intercepto}$
 $\beta_i = \text{coeficientes angulares (ou coeficientes parciais)}$

No modelo de regressão de três variáveis, indicado por (27), para que ele seja considerado um modelo de regressão linear, os coeficientes não podem ser correlacionados e os expoentes i , j e k deverão ser iguais a 1 (linear nos parâmetros). Os expoentes em letras gregas não necessariamente precisam ser iguais a um (não é necessária linearidade nas variáveis).

O termo *regressão linear múltipla* refere-se a um modelo de regressão linear nos parâmetros e que utiliza mais de uma variável explicativa.

A partir de valores das variáveis explicadas e explicativas obtidas de uma amostra da população total de observações, podem-se determinar os coeficientes da função de regressão. Os métodos mais usados são o *método dos mínimos quadrados ordinários* (MMQO) e o *método da máxima verossimilhança* (MMV). Neste trabalho será usado o MMQO por ser mais intuitivamente convincente e matematicamente mais simples que o MMV.

Estimados os parâmetros a partir dos dados da população, a função de regressão retorna o valor esperado para o valor da variável explicada condicionado aos valores das variáveis explicativas. Os valores observados para a variável explicada, no entanto, diferem do valor explicado. Essa diferença é chamada de *erro estocástico* ou *elemento aleatório* ou *não sistemático* do modelo de regressão, indicado por u (esse termo não tem relação com termo *não-sistemático* que é usado no modelo CAPM). A expressão (28) ilustra esse fato para a observação i da variável explicada Y para um modelo de regressão linear múltipla que também é linear nas variáveis.

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi} + u_i$$

$Y = \text{variável explicada}$
 $X_{ji} = \text{variáveis explicativas (regressores), } j=1,\dots,p$
 $\beta_0 = \text{intercepto}$
 $\beta_i = \text{coeficientes angulares (ou coeficientes parciais)}$
 $u_i = \text{erro estocástico}$

(28)

O modelo pode ser reescrito da seguinte forma:

$$Y_i = \beta_0 X_{0i} + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_p X_{pi} + u_i$$

$X_{0i} = 1 \text{ para todo } i$

(29)

Caso a obtenção dos estimadores dos parâmetros seja feita a partir de uma amostra da população, é possível calcular um estimador para o valor esperado da variável explicada condicionado aos valores das variáveis explicativas. Em (30) é mostrada a formulação desse modelo de regressão (modelo de regressão amostral) num caso análogo ao representado por (28) e (29). **A notação $\hat{}$ indica se tratar de estimadores dos parâmetros do modelo.**

$$\begin{aligned}
& \hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_p X_{pi} \\
\text{E} \quad & Y_i = \hat{\beta}_0 + \hat{\beta}_1 X_{1i} + \hat{\beta}_2 X_{2i} + \cdots + \hat{\beta}_p X_{pi} + \hat{u}_i \\
\text{Portanto} \quad & \hat{e}_i = Y_i - \hat{Y}_i
\end{aligned} \tag{30}$$

$Y = \text{variável explicada}$

$X_{ji} = \text{variáveis explicativas (regressores)}, j=1, \dots, p$

$\hat{\beta}_0 = \text{intercepto}$

$\hat{\beta}_i = \text{coeficientes angulares (ou coeficientes parciais)}$

$\hat{u}_i = \text{erro estocástico}$

Os termos $\hat{\beta}_j, j > 1$ em (30) são denominados *coeficientes parciais*. Cada um deles mede a variação no valor médio da variável explicada por unidade de variação da variável explicativa X_i pela qual estão multiplicados, *se fixadas as demais variáveis explicativas $X_j, j \neq i$* .

O objetivo do MMQO é estimar os parâmetros para a função de regressão de forma que a função minimize a soma dos quadrados das diferenças entre o valor esperado da variável observada previsto através da função de regressão e o valor efetivamente observado da variável explicada, para cada um dos valores das variáveis explicativas da amostra. Dito de outra maneira, para N observações, o objetivo é minimizar a soma dos $\hat{e}_i^2, i = 1 \dots N$.

É comum o uso da notação matricial ao se tratar de regressão linear múltipla. A equação de regressão linear múltipla em notação matricial é a mostrada em (31).

$$Y = X\hat{\beta} + \hat{\varepsilon}$$

Onde:

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$X = \begin{pmatrix} 1 & X_{11} & \dots & X_{p1} \\ 1 & X_{12} & \dots & X_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & \dots & X_{pn} \end{pmatrix}$$

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{pmatrix} \quad (31)$$

$$\hat{\varepsilon} = \begin{pmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_n \end{pmatrix}$$

Y = vetor com observações da variável explicada

X = matriz das observações dos regressores

$\hat{\beta}$ = matriz dos estimadores dos coeficientes

$\hat{\varepsilon}$ = matriz dos erros estocásticos das observações

y_i = valor da *i*-ésima observação da variável explicada

X_{Ni} = valor da *i*-ésima observação do regressor *N*

$\hat{\beta}_N$ = coeficiente parcial estimado do regressor *N*

\hat{u}_i = erro estocástico da *i*-ésima observação

Para uma amostra *j* tem-se a relação (32).

$$\hat{Y}_j = X_j \hat{\beta}_j \quad (32)$$

Como mostrado por Neter, Wasserman et al. (1990), os coeficientes que minimizam a soma dos quadrados dos desvios são calculados pela expressão (33).

$$\hat{\beta} = (X'X)^{-1}X'Y$$

(33)

$Y = \text{vetor com observações da variável explicada}$
 $X = \text{matriz das observações dos regressores}$
 $\hat{\beta} = \text{vetor dos coeficientes}$

Também por Neter, Wasserman et al. (1990), para estimar a variância dos coeficientes do modelo, calcula-se a matriz de variância-covariância Σ da regressão, dada por (34).

$$s^2(\beta) = \Sigma = (X'X)^{-1}s^2(\epsilon) = \begin{pmatrix} s_0^2 & \dots & s_{0,p}^2 \\ \vdots & \ddots & \vdots \\ s_{p,0}^2 & \dots & s_p^2 \end{pmatrix}$$

Onde:

$$s^2(\epsilon) = \frac{1}{n-p-1} \hat{\epsilon}'\hat{\epsilon} = \frac{(Y - X\hat{\beta})'(Y - X\hat{\beta})}{n-p-1} = MSE \quad (34)$$

$\Sigma = \text{matriz de variância-covariância}$

$n = \text{número de observações}$

$p = \text{número de regressores}$

$s_i^2 = \text{estimador da variância do coeficiente } i$

$s_{i,j}^2 = \text{estimador da covariância entre coeficientes } i \text{ e } j$

A matriz de variância-covariância é simétrica, e no caso de apenas uma variável explicativa, tem-se o resultado mostrado em (35).

$$\Sigma = (X'X)^{-1}MSE$$

$$= \begin{pmatrix} MSE \left(\frac{1}{n} + \frac{\bar{X}^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right) & \frac{-MSE\bar{X}}{\sum_{i=1}^n (X_i - \bar{X})^2} \\ \frac{-MSE\bar{X}}{\sum_{i=1}^n (X_i - \bar{X})^2} & \frac{MSE}{\sum_{i=1}^n (X_i - \bar{X})^2} \end{pmatrix} \quad (35)$$

Em (35), os elementos da diagonal principal são os estimadores da variância dos estimadores dos coeficientes do modelo. Para um modelo de regressão com um único regressor, o elemento (2,2) da matriz é o estimador da variância do coeficiente de índice 1 do modelo, que geralmente é o coeficiente angular.

Algumas medidas, mostradas em (36), são úteis para a determinação de indicadores da qualidade da regressão, ou seja, para medir a parcela da variabilidade dos dados que é explicada pelo modelo de regressão.

Segundo Neter, Wasserman et al. (1990).

$$SSE = Y'Y - \hat{\beta}'X'Y \quad (36)$$

e

$$SSR = \hat{\beta}'X'Y - \frac{1}{n}Y'UU'Y$$

$U = [1, 1, \dots, 1]'$, vetor unitário com n linhas e 1 coluna

n = número de observações

SSR = soma dos quadrados da regressão

Y = vetor com observações da variável explicada

X = matriz das observações dos regressores

SSE = soma dos quadrados dos erros

Y = vetor com observações da variável explicada

X = matriz das observações dos regressores

O termo SSE (*sum of squared error*) ou *soma dos quadrados dos resíduos* é a soma dos quadrados dos desvios dos valores da variável explicada na amostra em relação ao valor esperado pelo modelo para cada observação, condicionado aos valores das variáveis explicativas para cada observação da amostra. O termo SSR (*sum of squares of the regression*) ou *soma dos quadrados explicados pela regressão* é a soma dos quadrados dos desvios do valor esperado para as observações da variável explicada previsto pelo modelo, condicionado aos valores das observações das variáveis explicativas, em relação à média amostral da variável explicada. A Figura 9 mostra graficamente a interpretação dos resíduos (ou erros) na regressão.

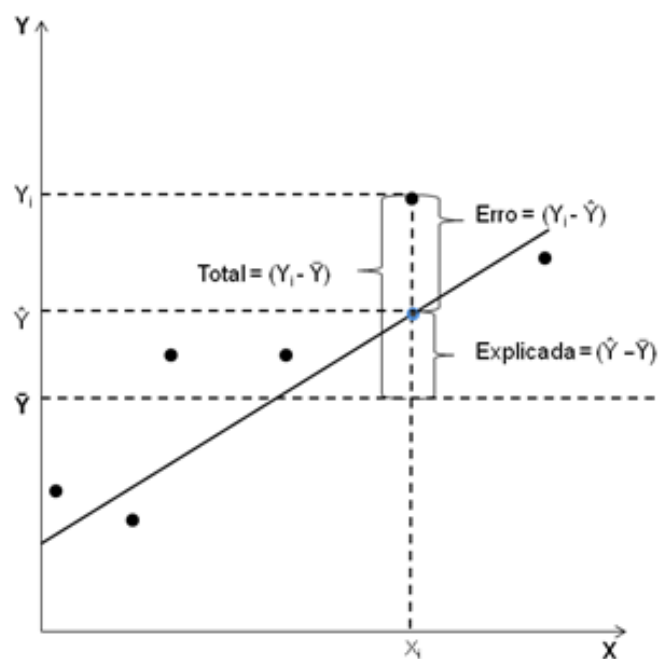


Figura 9 – Interpretação gráfica da regressão simples

A Figura 9 fornece uma interpretação da regressão no caso da regressão simples, onde há apenas uma variável explicativa. No caso de duas variáveis explicativas, o caso é mais bem ilustrado pela Figura 10.

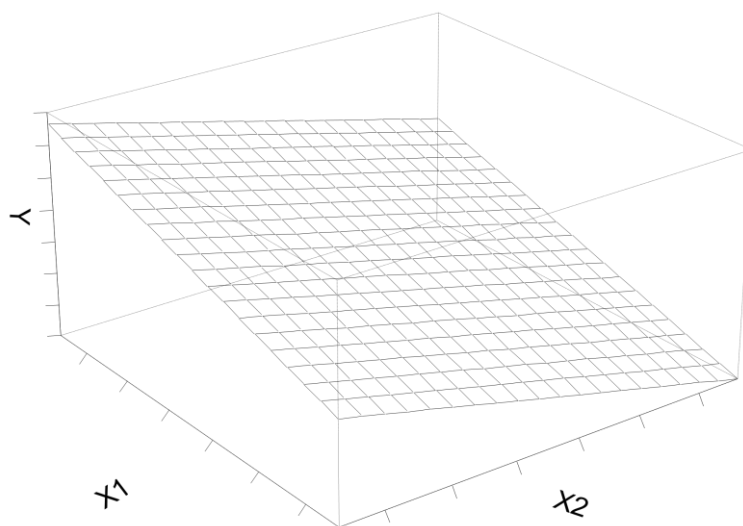


Figura 10 – Modelo de regressão com 3 variáveis

Outra medida importante para cálculo de estatísticas úteis para testes de hipóteses ligados à regressão é o MSE, representado em (37).

$$MSE = \frac{SSE}{n - p - 1}$$

$$\begin{aligned} MSE &= \text{soma dos quadrados médios} \\ SSE &= \text{soma dos quadrados dos erros} \\ n &= \text{número de observações} \\ p &= \text{número de regressores} \end{aligned} \quad (37)$$

O MSE (*mean squared error*) é um **estimador da variância**.

Pode-se sempre utilizar o MMQO para estimar os parâmetros da regressão linear. Porém, para se fazer inferências, é necessário assumir a distribuição normal para os resíduos. Quando se trata do modelo de regressão linear múltipla subentende-se que se fala do **modelo de regressão linear clássico, gaussiano ou padrão (MRLC)**, proposto por Gauss e De Morgan¹ apud Gujarati (2006, p. 53). Esse modelo pressupõe onze premissas, discutidas a seguir.

1. **O modelo é linear nos parâmetros.**
2. **Os valores das variáveis explicativas, ou regressores, são fixados em amostras repetidas.**
3. **Dados os valores das variáveis explicativas, o valor médio do termo de erro é zero.**

Caso essa premissa seja relaxada, não é possível estimar o coeficiente de intercepto. Porém, caso o valor médio do termo seja constante, os coeficientes parciais (das variáveis explicativas não constantes) não serão afetados.

4. **Dados os valores das variáveis explicativas, a variância do termo de erro é constante ou homocedástica.**

A ausência de *homocedasticidade*, ou a *heterocedasticidade*, é a dependência do termo quadrático de erro (real e não o estimado) em relação às variáveis explicativas. Na presença de heterocedasticidade, não se pode aplicar testes F

¹ GAUSS, C. F.; DE MORGAN, A. *Theoria combinationis observationum erroribus minimis obnoxiae*. Gottingae: Apud Henricum Dieterich, 1823. 58 p.

como testes de hipótese para o modelo, pois tais testes são baseados na hipótese de homocedasticidade. Entre os métodos de identificação da heterocedasticidade estão métodos informais, como a análise de gráficos nos quais se representa o erro quadrático amostral (estimador do erro quadrático) em função dos valores de Y previstos pelo modelo ou de valores de uma variável explicativa X . Segundo Gujarati (2006), há também testes formais como os de Park, de Glejser, de Goldfeld-Quandt, de Breusch-Pagan-Godfrey, de White, de Koenker-Basset e a análise do coeficiente de Spearman. Como métodos corretivos estão o MMQG, mínimos quadrados ponderados e transformações aplicadas ao MMQO que mantém as propriedades desejáveis de seus estimadores no caso de heterocedasticidade e a aplicação de erros-padrão dos testes de White.

5. Dados os valores das variáveis explicativas, não há autocorrelação entre os termos de erro.

Se esta hipótese não for verificada, os estimadores MMQO, embora lineares, não-tendenciosos e assintoticamente distribuídos normalmente, não são eficientes e, portanto, não se pode utilizar testes F como testes de hipótese. Segundo Gujarati (2006), a detecção da auto-correlação entre os termos de erro pode ser feita através de métodos gráficos, do *teste das carreiras*, do *teste d* de Durbin-Watson e do teste de Breusch-Godfrey. Como medidas corretivas podem-se citar o método dos mínimos quadrados generalizado, MMQG, transformações e o método de White (essas duas últimas medidas podem ser aplicadas de forma análoga à sugerida para o problema de heterocedasticidade).

6. Se as variáveis explicativas forem estocásticas, a correlação entre os termos de erro e as variáveis explicativas é nula.

7. O número de observações deve ser maior que o número de variáveis explicativas.

8. Deve haver variância não nula nos valores de cada variável explicativa em cada amostra.

9. A relação entre variável explicada e variáveis explicativas pressuposta pelo modelo é correta.

Segundo Gujarati (2006), testes de detecção desse tipo de erro são exame de resíduos, *estatística d* de Durbin-Watson, teste RESET de Ramsey e o teste do multiplicador de Lagrange. Se a relação pressuposta pelo modelo entre as variáveis explicada e explicativas é linear (variáveis todas com expoente 1), essa hipótese implica que esta relação deve ser realmente linear. Para verificar tal hipótese, testes de linearidade podem ser aplicados, como os indicados em Neter, Wasserman et al. (1990) e Burn e Ryan Jr (1983).

10. Não há relação linear exata (ou multicolinearidade) entre os regressores.

A multicolinearidade é observada para os dados da amostra, podendo ou não ser verdadeira para a população como um todo. A multicolinearidade perfeita torna impossível o cálculo dos estimadores MMQO. A Figura 11 ilustra os diferentes níveis de colinearidade.

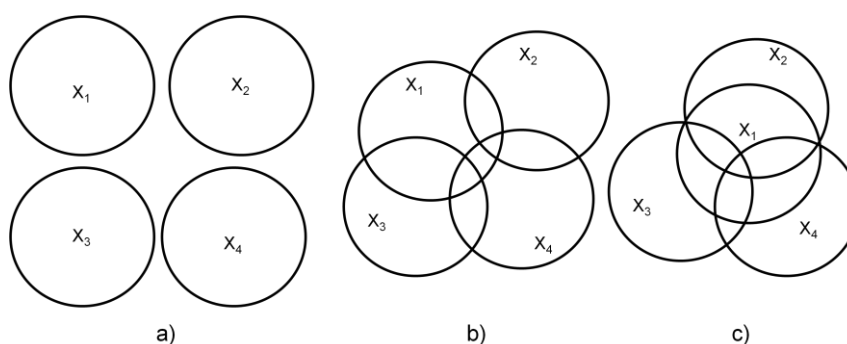


Figura 11 – Diagramas de Venn para diferentes níveis de colinearidade
Adaptado de Gujarati (2006)

- a) Ausência de colinearidade
- b) Presença de colinearidade
- c) Colinearidade perfeita, a variável X_1 é combinação linear das demais

Alta colinearidade pode dificultar o cálculo dos estimadores de MMQO, pois causa impossibilidade de inversão da matriz $\mathbf{X}'\mathbf{X}$, operação necessária para o cálculo dos coeficientes, como mostrado pela equação (33). Fatores que podem ser indicativos de alta multicolinearidade são coeficiente de determinação múltipla muito alto, altas correlações entre pares de regressores, exame das correlações parciais de Farrar e Glauber (1964), teste das regressões auxiliares de Klein (1962), análise por autovalores e do índice condicional e análise da tolerância e do valor de inflação da variância.

Como medidas corretivas, pode-se remover variáveis do modelo, adicionar dados (já que a multicolinearidade é referente às observações da amostra, podendo ser atenuada através de novos dados não pertencentes à mesma), uso de *análise fatorial* e *análise por componentes principais*, além de técnicas como a *regressão de cumeeira* (*ridge regression*).

11. O termo de erro estocástico se distribui normalmente.

Essa premissa garante a hipótese de que os estimadores do MMQO seguem uma distribuição normal e permite a utilização de testes F para verificação de hipóteses estatísticas relacionadas aos estimadores, qualquer que seja o tamanho da amostra. No entanto, pelo teorema central do limite, pode-se supor normalidade dos estimadores *assintoticamente*, isto é, em grandes amostras.

Alguns testes para validar essa premissa são o teste de normalidade *Anderson-Darling*, o teste *Jarque-Bera* e o teste de Shapiro-Wilk. Outros testes são sugeridos por Kutner, Nachtsheim et al. (2003) e Ryan (1997).

Em engenharia, um termo de erro apresentando as características descritas pelas hipóteses 3, 4 e 5 é chamado de *ruído branco*.

5.3.2. UTILIZAÇÃO DE VARIÁVEIS INDICADORAS NA REGRESSÃO

Esta seção aborda uma técnica aplicada ao modelo de regressão que será de importância crucial à metodologia aplicada neste trabalho.

Num modelo de regressão muitas vezes é desejável identificar, dentro de uma mesma amostra, subgrupos de observações que possuem alguma característica em comum. As variáveis indicadoras são introduzidas ao modelo com esse propósito.

Variáveis indicadoras assumem valores limitados e devem ser mutuamente exclusivas. No caso de variáveis indicadoras que assumem apenas valores 0 ou 1, dá-se o nome de variáveis binárias (*dummy variables*). Essas variáveis constituem um artifício para classificar dados em categorias mutuamente exclusivas como, por exemplo, masculino e feminino.

Variáveis binárias podem assumir valores diferentes de 0 e 1. O par (0, 1) pode ser transformado linearmente através de $Z=a+bD$ ($b \neq 0$), a e b constantes e $D = 0$ ou 1 . Assim, tem-se o par (a , $a+b$). No caso específico em que $a = 0$ e $b = 1$, o par é (0, 1). Portanto, as variáveis binárias ou qualitativas não têm uma escala natural de medição e são descritas como variáveis com escalas nominais.

No caso de um modelo de regressão linear, **ao se utilizar variáveis binárias para classificar observações, deve-se empregar uma variável a menos do que o número de classificações, para que não ocorra multicolinearidade entre as variáveis explicativas.** Chama-se de multicolinearidade entre variáveis explicativas o fato de uma variável ser combinação linear exata de outras. Neste caso, o poder explicativo dessa variável é nulo: sua capacidade explicativa já é completamente suprida por essas outras variáveis das quais ela é combinação linear.

Uma das aplicações das variáveis indicadoras é a de testar hipóteses entre amostras. Gujarati (1970) demonstra essa utilização para comparar coeficientes de regressão entre diversas amostras, aplicação muito semelhante à que será feita neste trabalho.

5.3.3. TESTES DE HIPÓTESES APLICADOS À REGRESSÃO

Esta seção aborda testes de hipóteses aplicados à regressão particularmente importantes a este trabalho, o teste F completo e o teste F parcial.

Segundo Gujarati (2006), pode-se aplicar testes de hipóteses aos modelos de regressão com diversos propósitos:

- Teste de significância do modelo de regressão múltipla estimado, ou seja, verificação se todos os coeficientes parciais são simultaneamente iguais a zero;
- Teste da contribuição marginal de variáveis explicativas;
- Testes de hipóteses relativos a um coeficiente parcial. Neste caso, a intenção é testar a hipótese $H_0: \beta_i = 0$;
- Teste de igualdade entre dois coeficientes parciais;
- Teste se os coeficientes parciais satisfazem certas restrições.

Este trabalho focará particularmente nos dois primeiros testes. Para realizar esses testes, são utilizados testes F que serão discutidos mais adiante.

Para testar a significância geral de uma regressão linear múltipla, segundo Gujarati (2006), é utilizado um teste F ou **teste F completo**. Indiretamente o que se quer é testar a hipótese nula:

$$H_0: \beta_0 = \beta_1 = \dots = \beta_p = 0$$

$$H_1: \text{pelo menos algum } \beta_i \neq 0$$

Para testá-la utiliza-se a seguinte estatística em (38).

$$F = \frac{\frac{SSR}{p}}{\frac{SSE}{n - p - 1}} \quad (38)$$

F = valor do teste F

SSR = soma de quadrados da regressão

SSE = soma de quadrados dos erros

p = número de regressores

n = número de observações

Segundo Levine (2008), essa estatística segue uma distribuição F com p e $n - p - 1$ graus de liberdade. Se $F > F_{\alpha}(p, n - p - 1)$ a hipótese H_0 deve ser rejeitada. $F_{\alpha}(p, n - p - 1)$ é o valor crítico da distribuição F no nível de significância α com p graus de liberdade para o numerador e $n - p - 1$ graus de liberdade para o denominador.

Segundo Albright, Winston et al. (2004) e Levine (2008), quando se quer comparar dois modelos, um chamado de *completo* e outro chamado de *reduzido* (as variáveis explicativas do modelo reduzido são um subconjunto do conjunto de variáveis explicativas do modelo completo) respectivamente representados em (39) e (40), a hipótese a ser feita é a de que os coeficientes das variáveis que estão no modelo completo e não estão no reduzido são todos iguais a zero neste modelo.

$$Y_i = \beta_0 X_{0i} + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_j X_{ji} + \dots + \beta_p X_{pi} + u_i \quad (39)$$

$$Y_i = \beta_0 X_{0i} + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_j X_{ji} + u_i \quad (40)$$

Ou seja, equivale a testar a hipótese:

$$H_0: \beta_{j+1} = \beta_{j+2} = \dots = \beta_p = 0$$

Neste caso, emprega-se o **teste F parcial**, cuja estatística é dada por (41).

$$F = \frac{\left[\frac{(SSE_{reduzido} - SSE_{completo})}{p - j} \right]}{MSE_{completo}}$$

$$F = \text{valor do teste F parcial} \quad (41)$$

$SSE_i = SSE$ do modelo i

$MSE_i = MSE$ do modelo i

$p =$ número de regressores do modelo completo

$j =$ número de regressores do modelo reduzido

Segundo Neter, Wasserman et al. (1990), essa estatística segue uma distribuição F com $p - j$ e $n - p - 1$ graus de liberdade. Se $F > F_{\alpha}(p - j, n - p - 1)$ a hipótese H_0 deve ser rejeitada. $F_{\alpha}(p - j, n - p - 1)$ é o valor crítico da distribuição F no nível de significância α com $p - j$ graus de liberdade para o numerador e $n - p - 1$ graus de liberdade para o denominador.

5.4. MONITORAMENTO DE PERFIL LINEAR

O monitoramento de perfil linear é uma das ferramentas utilizadas no controle estatístico de processos. Este capítulo começa com uma breve revisão sobre o controle estatístico de processos, seguida de uma explicação sobre gráficos de controle e de noções sobre as fases I e II do monitoramento de perfil linear, finalizando com uma explicação detalhada sobre o método de fase I de monitoramento de perfil linear, aplicado neste trabalho.

Segundo Albright, Winston et al. (2004), uma das áreas na qual a estatística teve maior impacto no mundo dos negócios é a área da qualidade. A história da produção industrial vivenciou grandes progressos em termos de melhoria da qualidade. Grande parte desse progresso deve-se ao uso de métodos quantitativos. Alguns desses métodos quantitativos são conhecidos como *controle estatístico do processo (CEP) ou statistical process control (SPC)*.

Duas das principais motivações do CEP são:

- Fazer o trabalho na primeira tentativa;
- Reduzir variância ou variabilidade.

Segundo Albright, Winston et al. (2004), no passado qualidade era sinônimo de inspeção no final do processo, enviando os produtos defeituosos para reajustes ao invés de entregá-los aos clientes. No entanto, o processo de inspeção é custoso, já que demanda tempo e trabalho. Os especialistas em qualidade observaram que seria vantajoso detectar problemas em etapas antecedentes do processo produtivo, poupando tempo e recursos e dando foco às causas dos problemas. Esse procedimento permitiria atingir a primeira motivação.

A segunda motivação é a de reduzir a variabilidade. Segundo Montgomery (2005), esse é o objetivo primário da engenharia da qualidade. Para que esse objetivo seja alcançado, é preciso um projeto que permita medir a variabilidade e dar aos operadores meios de reduzi-la.

O controle estatístico de processos foi e é muito discutido no contexto da indústria manufatureira, apesar de também ter sido muito empregado na indústria de serviços como o setor bancário e hoteleiro.

W. Edwards Deming foi um dos maiores responsáveis pela recente ênfase em qualidade, aplicando teorias estatísticas para redução da variabilidade de processos. Deming propôs 14 pontos, parcialmente quantitativos, para melhoria da qualidade e produtividade.

O controle estatístico de processo pode ser aplicado a qualquer processo. Suas sete principais ferramentas são:

- Apresentação em histogramas ou ramo-e-folhas;
- Folha de controle;
- Gráfico de Pareto;
- Diagrama causa-e-efeito;
- Diagrama de concentração de defeito;
- Diagrama de dispersão;
- Gráfico de controle.

O monitoramento de perfil linear pode ser interpretado como uma aplicação de um gráfico de controle.

Segundo Montgomery (2005), a variabilidade do processo de produção possui *causas aleatórias* e *causas atribuíveis*. Causas aleatórias (ou causas comuns) são naturais do processo, o ruído branco, que sempre existirá no processo por melhor que seja mantido. Causas atribuíveis (ou causas especiais) são provocadas por irregularidades no processo. Um processo ou característica que opera ou é decorrência da presença de causas atribuíveis é chamado de *fora de controle*. Um processo ou característica que não envolve causas atribuíveis é chamado de *estável* ou *sob controle*.

Gráficos de controle são uma das sete ferramentas do controle estatístico de processos. A ideia fundamental é a de um gráfico no qual estão marcadas 3 linhas: uma linha chamada *linha central*, que indica a média de uma estatística de interesse a ser monitorada, e duas linhas que, quando somente causas aleatórias estão agindo sobre o processo, são chamadas *limite inferior de controle (LIC)* e *limite superior de controle (LSC)*, respectivamente. Nesse gráfico são marcados os valores que a estatística assume no tempo ou em amostras sucessivas.

Há uma relação entre gráficos de controle e teste de hipóteses. Em geral, pode-se dizer que uma estatística monitorada marcada fora dos limites de controle significa que se deve rejeitar alguma hipótese de estabilidade do processo, ou seja, deve-se rejeitar a hipótese de que o processo encontra-se *sob controle*. Esse tipo de hipótese envolve a definição de um

nível de significância a ser determinado pelo criador do gráfico e, portanto, os limites superior e inferior dependem desse nível de significância. Montgomery (2005) indica que há pelo menos cinco motivos para a grande popularidade dos gráficos de controle, quais sejam:

1. Constituem uma técnica comprovada para a melhoria da produtividade;
2. São eficazes na prevenção de defeitos;
3. Evitam o ajuste desnecessário do processo;
4. Fornecem informações para diagnóstico;
5. Fornecem informações sobre a capacidade.

A Figura 12 apresenta um exemplo de gráfico de controle.

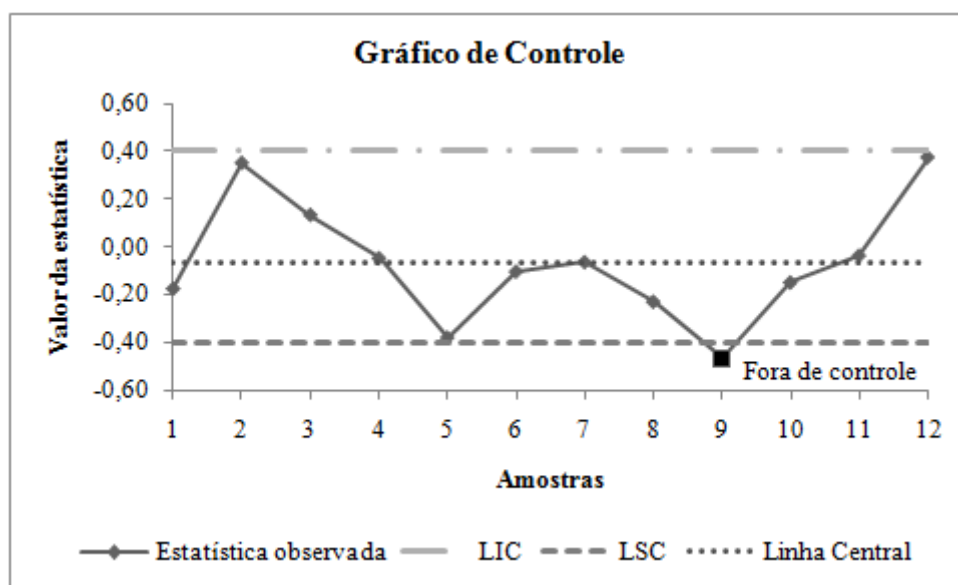


Figura 12 – Representação de um gráfico de controle

Dado que a ideia de monitoramento de perfil linear foi introduzida no âmbito do controle estatístico de processos, sua definição é agora apresentada.

É comum que uma característica relacionada à qualidade de um produto ou processo seja investigada através da aplicação do Controle Estatístico de Processo. Em muitos casos a qualidade de um processo ou produto é mais bem representada e sumariada através da realização da medição sequencial da característica, seja em espaços de tempo regulares ou através da análise de observações ou amostras consecutivas, seguida da sua representação gráfica por uma linha ou curva. Essa curva ou linha, que representa uma resposta funcional de

uma variável de qualidade, em função de uma ou mais variáveis independentes, é chamada de *perfil* na literatura. Caso a relação funcional entre a variável de qualidade representada no perfil e suas variáveis explicativas seja linear como, por exemplo, no caso em que a relação possa ser representada por um modelo de regressão linear, o perfil nesse caso é chamado de *perfil linear*.

Segundo Woodall (2007), *monitoramento de perfil* é o nome dado à técnica que consiste na análise de um perfil que representa uma característica referente à qualidade de um produto ou processo. Muitas vezes são usados gráficos de controle nesse processo, porém outras técnicas são possíveis.

O *monitoramento de perfil linear* é dividido em 2 fases: fase I e fase II.

Na fase I, o objetivo é de entender a variabilidade do processo, eliminar amostras cuja variabilidade pode ser relacionada a causas atribuíveis e estabelecer as condições para as quais o processo pode ser considerado sob-controle, ou seja, as causas agentes do processo são somente aleatórias. Com base no entendimento do processo obtido na fase I, cabe à fase II analisar amostras coletadas sequencialmente e detectar amostras *fora de controle*. O objetivo da fase II é detectar mudanças de estado de estável para *fora de controle* o mais rápido possível, evitando falsos alarmes (indicação mudança de estado quando, de fato, esta não ocorreu).

Neste trabalho será abordada a **fase I do monitoramento de perfil linear**.

A metodologia é composta por oito etapas, listadas a seguir e discutidas em detalhe em seguida:

1. Coleta de m_0 amostras, cada amostra j contendo n_j observações;
2. Aplicação do gráfico de controle proposto por Kim, Mahmoud et al. (2003) para monitoramento da variância das amostras;
3. Remoção das amostras detectadas como *fora de controle* pelo gráfico de controle para variância;
4. Agrupamento das amostras restantes em uma única amostra;
5. Aplicação do teste F global sugerido por Mahmoud (2008);
6. Aplicação do teste F para a inclinação (caso o teste F global seja significativo);
7. Aplicação do teste F para o intercepto (caso o teste F para a inclinação seja significativamente grande);

8. Identificação e tratamento (remoção) das amostras *fora de controle*.

Encontra-se na literatura uma série de abordagens para a fase I do monitoramento de perfil linear, como o gráfico de controle T^2 de Kang e Albin (2000), o método baseado em variáveis indicadoras de Mahmoud e Woodall (2004), o gráfico de controle T^2 baseado em diferenças sucessivas da matriz de estimadores de variância-covariância de Williams, Woodall et al. (2007) e o método de “ponto de mudança” proposto por Mahmoud, Parker et al. (2006). Esses métodos, salvo o último mencionado, não se aplicam a casos em que as amostras possuem conjuntos de diferentes valores para as variáveis explicativas. Além disso, quanto maior o número de parâmetros monitorados, menor o poder estatístico do método.

Mahmoud (2008) sugere um novo método que possibilita aplicação com diferentes conjuntos de valores para os regressores, caso para o qual não se encontra simulações na literatura. Esse método envolve o monitoramento de três parâmetros, independentemente do número de regressores presente no modelo.

Conforme Mahmoud (2008), a ideia principal dessa abordagem reside no fato de que, caso os parâmetros da fase I sejam estáveis, então o intercepto e a inclinação também serão estáveis (sob-controle) e, caso contrário, é esperado que ao menos um desses parâmetros seja *fora de controle*. Para monitorar esses dois parâmetros, esse modelo propõe um teste F como descrito por Mahmoud e Woodall (2004). Para monitorar a variância, é sugerido um gráfico de controle como proposto por Kim, Mahmoud et al. (2003).

Como mostrado por Mahmoud e Woodall (2004), para qualquer abordagem estatística da fase I do monitoramento de perfil linear as premissas de linearidade e normalidade são muito importantes para a coerência dos resultados. Assim, são recomendados testes de normalidade e linearidade antes da aplicação dos métodos. Alguns testes são propostos na seção sobre regressão deste trabalho.

O método que será aplicado no presente trabalho foi adaptado do proposto por Kim, Mahmoud et al. (2003). Este método faz uso de gráficos de controle de uma variável. Tal método apresenta, segundo Mahmoud (2008), sensibilidade muito boa a mudanças repentinas na variância quando comparado a outros métodos conhecidos na literatura.

Os limites de controle são aplicados para o MSE de cada amostra. Tais limites variam de acordo com a amostra e são mostrados na equação (42).

$$\begin{aligned}
 LIC &= \frac{m_0 F_{n_j-p-1, \sum_{i \neq j}^{m_0} (n_j-p-1), \alpha/2} MSE}{m_0 - 1 + F_{n_j-p-1, \sum_{i \neq j}^{m_0} (n_j-p-1), \alpha/2}} \\
 LSC &= \frac{m_0 F_{n_j-p-1, \sum_{i \neq j}^{m_0} (n_j-p-1), 1-\alpha/2} MSE}{m_0 - 1 + F_{n_j-p-1, \sum_{i \neq j}^{m_0} (n_j-p-1), 1-\alpha/2}} \\
 \text{Com} \quad MSE &= \frac{\sum_{j=1}^{m_0} MSE_j}{m_0}
 \end{aligned} \tag{42}$$

m_0 = número inicial de amostras
 n_j = número de observações da amostra j
 p = número de regressores
 α = nível de significância

A definição do nível de significância é arbitrária. Tendo como foco o nível de significância total de 5%, envolvendo tanto o gráfico de controle da variância como o de comparação entre as linhas de regressão, neste trabalho será usado o nível de significância (probabilidade de alarme falso) para o gráfico de controle da variância dado por (43).

$$1 - \left((1 - \alpha_{total})^{\frac{1}{p+2}} \right)^{\frac{1}{m_0}} \tag{43}$$

α_{total} = nível de significância total (variância e teste F)
 P = número de regressores
 m_0 = número de amostras inicial

As amostras cujo MSE está fora dos limites de controle (amostras *fora de controle*) devem então ser removidas (etapa 3). O número restante de amostras é dado pela expressão (44). As m amostras restantes devem ser unidas numa só amostra (etapa 4) contento todas as observações. O número total de observações da nova amostra agrupada final é dado por (45).

$$m = \sum_{j=1}^{m_0} \mathbb{I}(LIC < MSE_j < LSC) \tag{44}$$

m_0 = número de amostras inicial
 m = número de amostras final

$$N = \sum_{j=1}^{m_0} n_j \cdot \mathbb{I}(LIC < MSE_j < LSC) \quad (45)$$

n_j = número de observações da amostra j

m_0 = número de amostras inicial

N = número de observações a serem consideradas

O próximo passo (etapa 5) é a aplicação do método de monitoramento de perfil linear fase I sugerido por Mahmoud (2008). Através desse método é testada a hipótese de que os coeficientes de regressão das diferentes amostras (agora unidas numa só amostra) podem ser considerados iguais (mesmo intercepto e mesma inclinação) e, portanto, pode-se considerar um só modelo de regressão para todas as amostras. Caso, inicialmente, essa hipótese seja rejeitada, tenta-se identificar o coeficiente problemático (intercepto ou inclinação). São também removidas as amostras identificadas como *fora de controle*, amostras estas consideradas como causadoras da rejeição da hipótese de igualdade dos coeficientes. Deve-se utilizar $m-1$ variáveis indicadoras (vetores). Usando m variáveis ter-se-ia uma situação de multicolinearidade, conforme explicado na seção sobre variáveis indicadoras aplicadas à regressão. Assim, tem-se as variáveis indicadoras Z_j , $j=1,2,\dots,m'$ com $m'=m-1$, definidas como em (46).

$$Z_{ji} = \begin{cases} 1, & \text{se observação } i \text{ é originária da amostra } j \\ 0, & \text{caso contrário} \end{cases} \quad (46)$$

Z_j = variável indicadora j

Z_{ji} = linha (observação) i da variável Z_j

A seguir, considerando as variáveis descritas por (47), ajusta-se o modelo de regressão dado por (48).

$$\mathbf{X} = \begin{pmatrix} 1 & x_{1,1} & x_{2,1} & \dots & x_{p,1} \\ 1 & x_{1,2} & x_{2,2} & \dots & x_{p,2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1,N} & x_{2,N} & \dots & x_{p,N} \end{pmatrix}$$

$$\mathbf{Z} = \begin{pmatrix} Z1_1 & Z2_1 & \dots & Zm'_1 \\ Z1_2 & Z2_2 & \dots & Zm'_2 \\ \vdots & \vdots & \ddots & \vdots \\ Z1_N & Z2_N & \dots & Zm'_N \end{pmatrix} \quad (47)$$

\mathbf{y} = vetor com observações da variável explicada

\mathbf{X} = matriz do intercepto e regressores

N = número de observações

p = número de regressores

m' = número de amostras - 1 = $m - 1$

$\boldsymbol{\varepsilon}$ = vetor dos resíduos

\mathbf{Z} = matriz das variáveis indicadoras

$$y_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \boldsymbol{\beta}_{z0} + x_{1,i} \mathbf{Z}_i \boldsymbol{\beta}_{z1} + x_{2,i} \mathbf{Z}_i \boldsymbol{\beta}_{z2} + \cdots + x_{p,i} \mathbf{Z}_i \boldsymbol{\beta}_{zp} + \varepsilon_i \quad (48)$$

Para todas as observações $i, i = 1 \dots N$

Onde:

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}$$

$$\boldsymbol{\beta}_{zk} = \begin{pmatrix} \beta_{zk,1} \\ \beta_{zk,2} \\ \vdots \\ \beta_{zk,m'} \end{pmatrix}$$

y_i = valor da variável explicada para observação i

\mathbf{X}_i = vetor contendo linha i da matriz \mathbf{X} de regressores

$\boldsymbol{\beta}$ = vetor dos coeficientes parciais dos regressores

$\boldsymbol{\beta}_{zk}$ = vetor de coeficientes parciais auxiliares

\mathbf{Z}_i = vetor contendo linha i da matriz \mathbf{Z}

$x_{k,i}$ = valor do regressor k para observação i

ε_i = resíduo para observação i

p = número de regressores

m' = número de amostras - 1 = $m - 1$

A Tabela 1 mostra um exemplo da disposição dos dados para o modelo explicado anteriormente aplicado a uma situação hipotética com 2 regressores e 4 amostras de 3 observações cada.

Tabela 1 – Exemplo de matriz do modelo para 2 regressores e 4 amostras de 3 observações

Observação	Amostra de origem					Indicadoras			Auxiliares					
		y	1	X1	X2	Z1	Z2	Z3	X1Z1	X1Z2	X1Z3	X2Z1	X2Z2	X2Z3
1	1	y1	1	X1,1	X2,1	1	0	0	X1,1	0	0	X2,1	0	0
2	1	y2	1	X1,2	X2,2	1	0	0	X1,2	0	0	X2,2	0	0
3	1	y3	1	X1,3	X2,3	1	0	0	X1,3	0	0	X2,3	0	0
4	2	y4	1	X1,4	X2,4	0	1	0	0	X1,4	0	0	X2,4	0
5	2	y5	1	X1,5	X2,5	0	1	0	0	X1,5	0	0	X2,5	0
6	2	y6	1	X1,6	X2,6	0	1	0	0	X1,6	0	0	X2,6	0
7	3	y7	1	X1,7	X2,7	0	0	1	0	0	X1,7	0	0	X2,7
8	3	y8	1	X1,8	X2,8	0	0	1	0	0	X1,8	0	0	X2,8
9	3	y9	1	X1,9	X2,9	0	0	1	0	0	X1,9	0	0	X2,9
10	4	y10	1	X1,10	X2,10	0	0	0	0	0	0	0	0	0
11	4	y11	1	X1,11	X2,11	0	0	0	0	0	0	0	0	0
12	4	y12	1	X1,12	X2,12	0	0	0	0	0	0	0	0	0

A igualdade entre as retas de regressão das m amostras, cuja hipótese nula está mostrada em (49) tem como hipótese alternativa H_1 : ao menos um dos coeficientes parciais das variáveis indicadoras ou auxiliares é diferente de zero.

$$H_0: \beta_{z0,1} = \beta_{z0,2} = \dots = \beta_{z0,m'} = \beta_{z1,1} = \beta_{z1,2} = \dots = \beta_{z1,m'} = \beta_{zp,1} = \beta_{zp,2} = \dots = \beta_{zp,m'} = 0 \quad (49)$$

Sob a hipótese nula, o modelo em (48) pode ser reduzido a (50).

$$y = X\beta + \varepsilon \quad (50)$$

O modelo mostrado em (48) será doravante chamado de **modelo completo** e o modelo mostrado em (50) será chamado de **modelo reduzido**.

O modelo reduzido representa um total de p variáveis explicativas. O modelo completo apresenta um total de $(m-1)(p+1)+p$ variáveis explicativas, quais sejam: as p variáveis do modelo reduzido acrescidas de $(m-1)$ variáveis indicadoras e de $p(m-1)$ variáveis auxiliares. A diferença entre o número de variáveis indicadoras dos dois modelos é, portanto, $(p+1)(m-1)$. Assim, empregando (41), tem-se o teste F para a hipótese H_0 dado por (51).

$$F_{global} = \frac{\left[\frac{(SSE_{reduzido} - SSE_{completo})}{(p+1)(m-1)} \right]}{MSE_{completo}} \quad (51)$$

$MSE_{completo} = MSE \text{ do modelo completo}$
 $SSE_{reduzido} = SSE \text{ do modelo reduzido}$
 $SSE_{completo} = SSE \text{ do modelo completo}$

Essa estatística segue uma distribuição F com $(p+1)(m-1)$ e $N-m(p+1)$ graus de liberdade sob a hipótese nula. Caso o valor dessa estatística F seja significantemente grande, tem-se um sinal *fora de controle* e diz-se que o processo está *fora de controle*.

O nível de significância para o teste F global deve ser definido de forma que, juntamente com o gráfico de controle para variância, totaliza um nível de significância α_{total} desejado. A equação (43) define a probabilidade de alarme falso para o gráfico de controle da variância. Assim, de forma a compor $\alpha_{total} = 5\%$, a probabilidade de alarme falso para o teste F global deve ser como dado por (52).

$$1 - (1 - \alpha_{total})^{\frac{p+1}{p+2}} \quad (52)$$

$\alpha_{total} = \text{nível de significância total (variância e teste F)}$
 $P = \text{número de regressores}$

Para identificar a origem da diferença entre as retas de regressão: inclinação (etapa 6) ou intercepto (etapa 7), pode-se testar as seguintes hipóteses nulas:

H_{01} : Os coeficientes de regressão parcial $\beta_1, \beta_2, \dots, \beta_p$ são iguais para todas as amostras, ou seja, as retas de regressão são paralelas para todas as amostras. Assume-se que a variância é estável, o intercepto pode ser estável ou não.

H_{02} : Os interceptos de todas as retas de regressão são iguais, assumindo que variância e coeficientes de inclinação são estáveis.

Sob a hipótese nula H_{01} : o modelo completo fica como descrito em (53).

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\beta}_{z0} \quad (53)$$

Este modelo será chamado de *modelo intermediário*. Este modelo apresenta $p+m$ variáveis explicativas, contra $(m-1)(p+1)+p$ do modelo completo. Assim, a estatística para teste da hipótese H_A é dada por (54).

$$F_{\text{inclinação}} = \frac{\left[\frac{(SSE_{\text{intermediário}} - SSE_{\text{completo}})}{p(m-1)} \right]}{MSE_{\text{completo}}} \quad (54)$$

$SSE_{\text{intermediário}} = SSE \text{ do modelo intermediário}$

O teste (54) segue uma distribuição F com $p(m-1)$ e $N-m(p-1)$ graus de liberdade sob a hipótese nula. Se o teste F referente à inclinação for significativamente grande, considera-se que as linhas não são paralelas e não se testa a hipótese H_B , já que os coeficientes parciais não são considerados estáveis. Caso contrário, usa-se a seguinte estatística para testar a hipótese H_B :

$$F_{\text{intercepto}} = \frac{\left[\frac{(SSE_{\text{reduzido}} - SSE_{\text{intermediário}})}{(m-1)} \right]}{MSE_{\text{intermediário}}} \quad (55)$$

$MSE_{\text{intermediário}} = MSE \text{ do modelo intermediário}$

O teste (55) segue uma estatística F com $(m-1)$ e $N-p-m$ graus de liberdade sob a hipótese nula. Caso este teste apresenta valor significativamente elevado, atribui-se ao intercepto a instabilidade.

O próximo passo (etapa 8) consiste em identificar as amostras *fora de controle*. Para tanto, o teste F global é decomposto em $m-1$ testes, cada um tendo como objetivo identificar se uma amostra está *fora de controle*. Uma das amostras deve ser usada como referência, por isso são feitos $m-1$ testes. Neste trabalho, a amostra de referência será a última amostra coletada.

Para testar se o perfil de uma amostra j é considerado *fora de controle*, consideram-se as observações de apenas duas amostras: a amostra j e a última amostra coletada. A seguir, unem-se as observações dessas duas amostras em uma única, usam-se as variáveis indicadoras propostas em (46) e aplica-se o teste F global cujo desenvolvimento foi mostrado em (48), (49), (50) e (51).

A aplicação do teste F global neste caso é semelhante ao feito anteriormente para todas as observações provenientes das m amostras. Porém, tem-se agora 2 amostras ($m=2$). Assim, o teste proposto em (51) pode ser simplificado para a relação mostrada em (56).

$$F_{amostra} = \frac{\left[\frac{(SSE_{reduzido} - SSE_{completo})}{(p + 1)} \right]}{MSE_{completo}} \quad (56)$$

A estatística (56) segue uma distribuição F com $(p+1)$ e $N'-2(p+1)$ graus de liberdade sob a hipótese nula (basta substituir $m=2$ nos graus de liberdade do teste global). N' é o número total de observações das duas amostras, ou seja, o número de observações da amostra de referência somado a número de observações da amostra j . Ainda, como agora o número de amostras é 2, tem-se apenas uma variável indicadora no modelo completo aplicado a essas duas amostras.

Se essa estatística F aplicada à amostra j for significativamente grande, considera-se que a amostra j é *fora de controle* e deve ser removida para permitir análises sobre o processo em estabilidade. Repete-se tal procedimento para $m-1$ amostras.

6. APLICAÇÃO

É comum a utilização da regressão simples como método prático para mensurar o *beta* de ações. Tal método considera alterações pontuais ou temporárias nos *betas* e resultam em medidas potencialmente incorretas. Tais alterações podem ser fruto de choques econômicos pontuais (como crises), fusões envolvendo a empresa da qual se quer medir o *beta* ou de outras empresas do mesmo setor, entrada de novos agentes no mercado, alterações temporárias na administração, boatos envolvendo a empresa, entre outros fatores.

A metodologia proposta neste trabalho sugere a monitoração do *beta* (do modelo *Market Model*) de forma a reduzir o impacto de suas reduções temporárias, resultando num estimador mais fidedigno do comportamento normal esperado para esse estimador – o que é de grande interesse de instituições financeiras.

Esta metodologia envolve monitoramento de perfil como técnica principal de estimação do *beta*. Como o objetivo é estimar o valor de uma variável importante para o controle da qualidade (o *beta*), o que se quer é aplicar a fase I do monitoramento de perfil. Será aplicado o método alternativo para fase I de monitoramento de perfil proposto por Mahmoud (2008). A razão para escolha desse método é sua capacidade de aplicação em casos nos quais os conjuntos de valores das variáveis explicativas não são os mesmos para diferentes amostras (embora ainda não sejam encontradas simulações na literatura). Este método é um dos poucos que permite aplicação em tal situação e sua aplicação é consideravelmente mais simples.

Será também aplicado o gráfico de controle para monitoramento da variância proposto por Kim, Mahmoud et al. (2003), um método simples e sugerido por Mahmoud (2008).

A metodologia empregada, cujas etapas foram discutidas no capítulo anterior deste trabalho, utilizará como dados de entrada amostras de cujas observações são retornos diários de algumas ações selecionadas e do índice de Bovespa. Os retornos serão calculados com base nos valores de fechamento de cada dia (ultimo valor negociado na bolsa de valores). O índice Bovespa será considerado o fator de mercado do modelo *Market Model*, usado para mensurar o *beta* de cada ação.

A aplicação foi feita através de um programa de computador, desenvolvido em linguagem Visual Basic; que faz os cálculos e realiza as iterações e exclusões necessárias. O

programa é mostrado no Apêndice B deste trabalho. As tabelas contemplando todos os resultados obtidos através da aplicação da metodologia sobre os dados coletados encontram-se no Apêndice A deste trabalho.

Para a coleta de dados e para as análises efetuadas, foram utilizados os softwares Microsoft Excel 2007, MatLab 2008b, SAS JMP 7, GraphPad Prism 5 e Bloomberg.

Os dados coletados referem-se ao preço de fechamento (último preço negociado no dia) ajustado para os proventos (conforme explicado na seção 4.3) para um conjunto de ações pré-selecionado. Para obtenção desses dados foi utilizado o sistema Bloomberg, o sistema mais utilizado mundialmente para obtenção de dados sobre mercado financeiro, através dos parâmetros *PDFE* e *Cash Div Adjustment Intraday*. Os retornos foram calculados conforme (3).

As ações selecionadas estão listadas na Tabela 2. Foram selecionados diversos setores industriais, bem como empresas com diferentes níveis de representatividade (capitalização de mercado). Foram definidas categorias compostas por companhias de setores específicos diferentes, mas dentro de um mesmo setor industrial. A exceção é a categoria *Blue chip* que considera apenas a representatividade da empresa para as negociações da Bolsa de Valores. As empresas dessa categoria são as duas maiores da Bovespa em capitalização de mercado (preço de suas ações multiplicado pela quantidade de ações no mercado) e também as de maior volume médio diário negociado nos últimos 12 meses.

Tabela 2 – Ações selecionadas separadas nas categorias que a serem analisadas

<i>Categoria</i>	<i>Setor específico</i>	<i>Nome</i>	<i>Código Bovespa</i>
<i>Água e Saneamento</i>	Água e Saneamento	Sabesp ON	SBSP3
<i>Transporte aéreo</i>	Transporte aéreo	GOL ON	GOLL4
<i>Financeiro</i>	Bancos	Bradesco ON	BBDC3
<i>Blue chip</i>	Exploração e Refino	Petrobrás PN	PETR4
	Minerais Metálicos	Vale PN	VALE5
<i>Imobiliário</i>	Construção Civil	Cyrela Realt ON	CYRE3
	Construção Civil	Gafisa ON	GFSA3
	Exploração de Imóveis	Iguatemi ON	IGTA3
<i>Distribuição de energia</i>	Energia Elétrica	Cemig ON	CMIG3
	Energia Elétrica	Eletrobras ON	ELET3
	Energia Elétrica	Cemig PN	CMIG4
	Energia Elétrica	Tran Paulist PN	TRPL4
	Gás	Comgas PNA	CGAS5
<i>Rodoviário</i>	Material Rodoviário	Randon Part PN	RAPT4
	Exploração de Rodovias	Ccr Rodovias ON	CCRO3
<i>Siderurgia</i>	Siderurgia	Cia Sid Nacional PN	CSNA3
	Siderurgia	Ferbasa PN	FESA4

Para cada ação, foram coletadas 12 amostras de 12 observações cada (retornos diários ajustados) para cada um dos 9 períodos mostrados na Tabela 3.

Tabela 3 – Períodos definidos para coleta de dados

<i>Nome dado ao período</i>	<i>Início</i>	<i>Fim</i>	<i>Número de dias úteis</i>
2003	03/01/2003	01/08/2003	144
2004	05/01/2004	30/07/2004	144
2005	03/01/2005	29/07/2005	144
2006	02/01/2006	31/07/2006	144
2007	02/01/2007	31/07/2007	144
2008a	02/01/2008	31/07/2008	144
2008b	01/08/2008	27/02/2009	144
2009	02/01/2009	31/07/2009	144
2010	14/09/2009	16/04/2010	144

A coluna *número de dias* indica o número de dias que apresentaram operações na bolsa de valores Bovespa entre as datas de início e fim, com essas datas inclusas.

Assim, para cada um dos períodos, analisados separadamente, foram obtidas 144 observações.

Há dois períodos, *2008a* e *2008b*, cujo início situa-se no ano de 2008. O motivo da seleção de dois períodos nesse ano é a possibilidade de análise de períodos imediatamente antes e após a grande crise econômica mundial iniciada em setembro de 2008.

Para cada ação, as 144 observações de cada um dos 9 períodos foram divididas em 12 amostras de 12 observações cada. Isso foi feito considerando que, para intervalos relativamente curtos (12 dias úteis, por exemplo), pode-se considerar que dados provindos do mercado financeiro fazem parte de uma mesma situação macroeconômica, podendo, portanto, serem considerados dados provenientes de uma mesma situação experimental. O Gráfico 1 ilustra a divisão feita para um período da ação BBDC3.

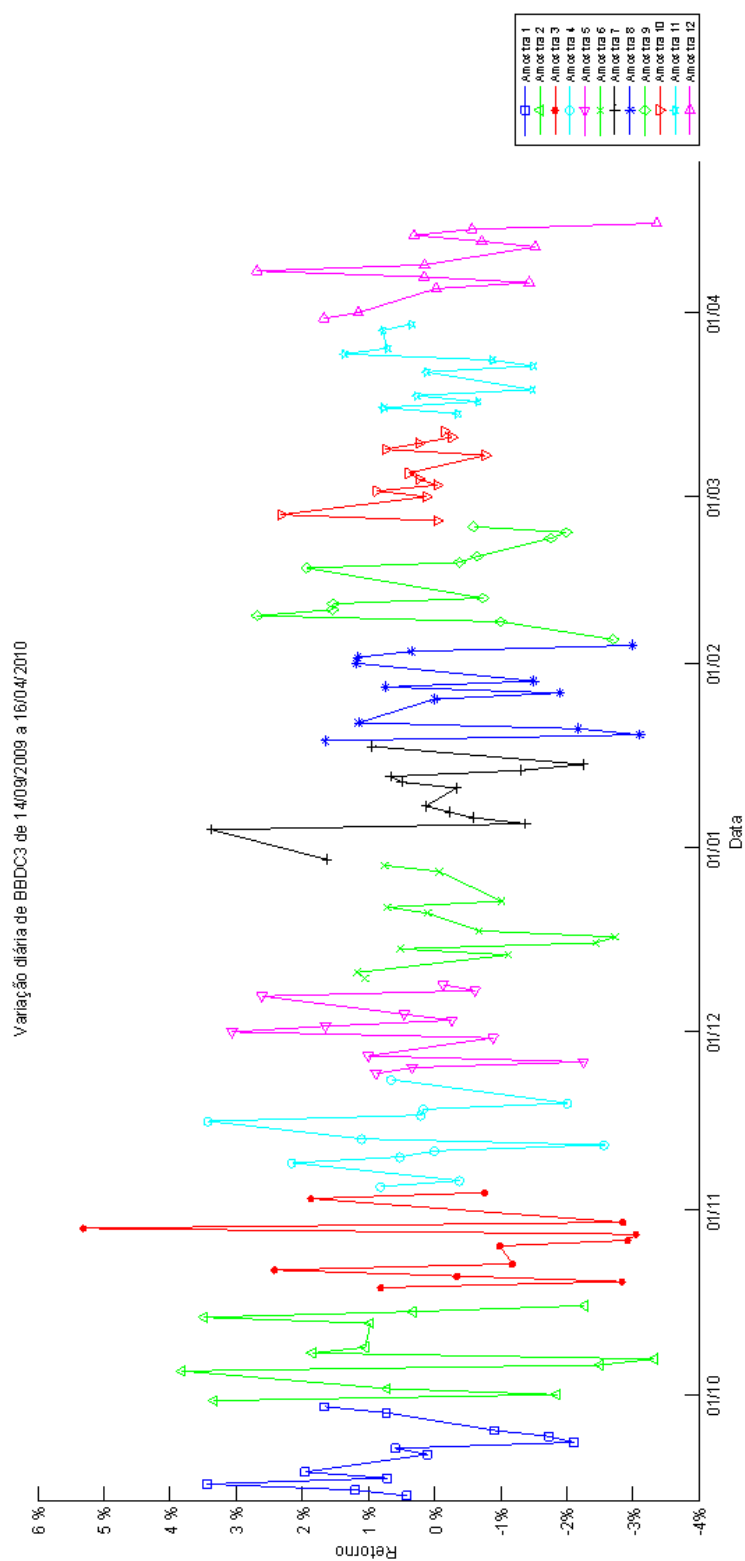


Gráfico 1 – Variação do preço de BBDC3 no período denominado 2010, evidenciando as amostras como sub-períodos

O código MatLab para geração desse gráfico encontra-se no Apêndice C.

A Tabela 4 ilustra os dados coletados para o índice Ibovespa e para as ações do banco Bradesco (código BBDC3 na Bolsa de Valores Bovespa) no período denominado **2010**, indicando a atribuição de cada amostra a uma observação.

Tabela 4 – Exemplo de dados coletados

	Amostra	Data	Índice Ibovespa		Ação BBDC3	
			Pontos	Variação	Preço da ação (R\$)	Variação
<i>Dia anterior</i>	-	11/09/2009	58366,38		R\$ 23,74	
<i>Início do período</i>	1	14/09/2009	58867,55	0,86%	R\$ 23,84	0,41%
	1	15/09/2009	59263,86	0,67%	R\$ 24,12	1,20%
	1	16/09/2009	60410,66	1,94%	R\$ 24,95	3,44%
	1	17/09/2009	60236,03	-0,29%	R\$ 25,13	0,71%
	1	18/09/2009	60703,01	0,78%	R\$ 25,62	1,95%
	1	21/09/2009	60928,02	0,37%	R\$ 25,65	0,11%
	1	22/09/2009	61493,39	0,93%	R\$ 25,80	0,59%
	1	23/09/2009	60496,19	-1,62%	R\$ 25,26	-2,11%
	1	24/09/2009	60046,28	-0,74%	R\$ 24,82	-1,73%
	1	25/09/2009	60355,73	0,52%	R\$ 24,60	-0,90%
	1	28/09/2009	61316,62	1,59%	R\$ 24,77	0,72%
	1	29/09/2009	61235,26	-0,13%	R\$ 25,18	1,66%
	2	30/09/2009	61517,89	0,46%	R\$ 26,02	3,33%
	2	01/10/2009	60459,33	-1,72%	R\$ 25,54	-1,85%

	11	29/03/2010	69939,12	1,83%	R\$ 25,66	0,79%
	11	30/03/2010	69959,58	0,03%	R\$ 25,75	0,35%
	12	31/03/2010	70371,54	0,59%	R\$ 26,18	1,67%
	12	01/04/2010	71136,34	1,09%	R\$ 26,48	1,15%
	12	05/04/2010	71289,68	0,22%	R\$ 26,47	-0,03%
	12	06/04/2010	71095,65	-0,27%	R\$ 26,09	-1,44%
	12	07/04/2010	70792,94	-0,43%	R\$ 26,13	0,15%
	12	08/04/2010	71784,78	1,40%	R\$ 26,83	2,68%
	12	09/04/2010	71417,27	-0,51%	R\$ 26,87	0,15%
	12	12/04/2010	70614,36	-1,12%	R\$ 26,46	-1,53%
	12	13/04/2010	70792,40	0,25%	R\$ 26,27	-0,72%
	12	14/04/2010	71034,85	0,34%	R\$ 26,35	0,30%
	12	15/04/2010	70524,35	-0,72%	R\$ 26,20	-0,57%
	12	16/04/2010	69421,35	-1,56%	R\$ 25,32	-3,36%

Como explicado no detalhamento da metodologia, a premissa de normalidade é de fundamental importância para aplicação desta metodologia. Um teste que pode ser aplicado para verificação de tal premissa é o teste Shapiro-Wilk. O software estatístico JMP, fabricado

pela SAS, permite a aplicação do teste sobre os resíduos. O resultado de tal teste aplicado aos dados do modelo de regressão dos retornos da ação BBDC3 no período 2010 explicados pelos retornos do Ibovespa nesse mesmo período é mostrado na Figura 13.

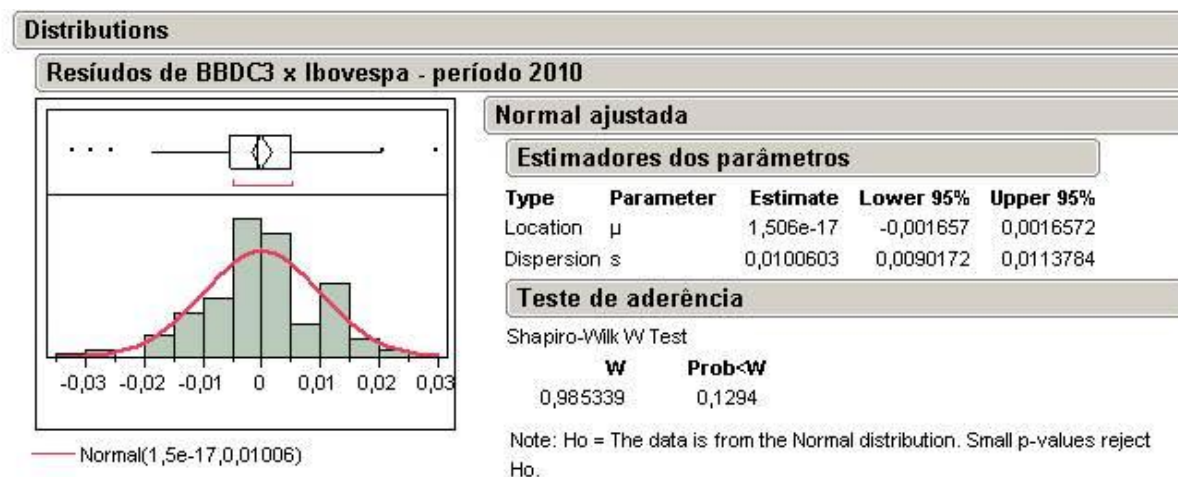


Figura 13 – Resultados do teste de normalidade
Imagem gerada pelo software SAS JMP 7.0

Como o valor P ($prob < W$) da estatística W calculada do teste de Shapiro-Wilk é superior ao nível de significância (definido arbitrariamente, em geral 5%), não se rejeita a hipótese H_0 : os dados seguem uma distribuição normal.

A validade da hipótese de linearidade é também crucial para aplicação desta metodologia. Para testar a hipótese da linearidade pode ser usado o software GraphPad Prism, que executa um *teste runs* para verificar a linearidade dos dados. Os resultados estão apresentados na Tabela 5.

Tabela 5 – Teste de linearidade

<i>Teste Runs</i>	
<i>Pontos acima da reta</i>	70
<i>Pontos abaixo da reta</i>	74
<i>Numero de runs</i>	80
<i>Valor P (teste runs)</i>	0,8972
<i>Desvio da linearidade</i>	Não significativo

Como o valor P não é significativamente pequeno, o teste não rejeita a hipótese de que os dados seguem uma relação linear.

Os testes de normalidade e linearidade, mostrados na Figura 13 e na Tabela 5, foram aplicados a todas as amostras coletadas, confirmando tais hipóteses. Tais resultados não foram aqui reproduzidos, pois sua apresentação não contribuiria para os objetivos deste trabalho. Assim, prossegue-se com a etapa seguinte que consiste na aplicação de um gráfico de controle para a variância.

Inicialmente, define-se o nível de significância do teste. De acordo com (43), considerando um nível de significância total de 5%, número de regressores $p = 1$, e número de amostras $m = 12$, a probabilidade de alarme falso para o gráfico de controle da variância é 0,001424.

Conforme expressão (42), para cálculo dos limites de controle, e de (36) e (37), para o cálculo dos estimadores de variância MSE de cada amostra, quer-se verificar a estabilidade da variância dos retornos de BBDC3 no período denominado 2010. Foram obtidos os resultados mostrados na Tabela 6 para os MSE das amostras. Ademais, para os demais parâmetros, foram obtidos os seguintes resultados: $MSE\ médio = 1,00E-04$, $LIC = 1,43E-05$ e $LSC = 2,82E-04$.

Tabela 6 – Dados do gráfico de controle para variância de BBDC3 - 2010

<i>Amostra</i>	<i>MSE</i>
1	1,21E-04
2	2,99E-04
3	1,12E-04
4	8,78E-05
5	6,99E-05
6	8,61E-05
7	9,31E-05
8	9,04E-05
9	5,14E-05
10	5,11E-05
11	7,61E-05
12	6,30E-05

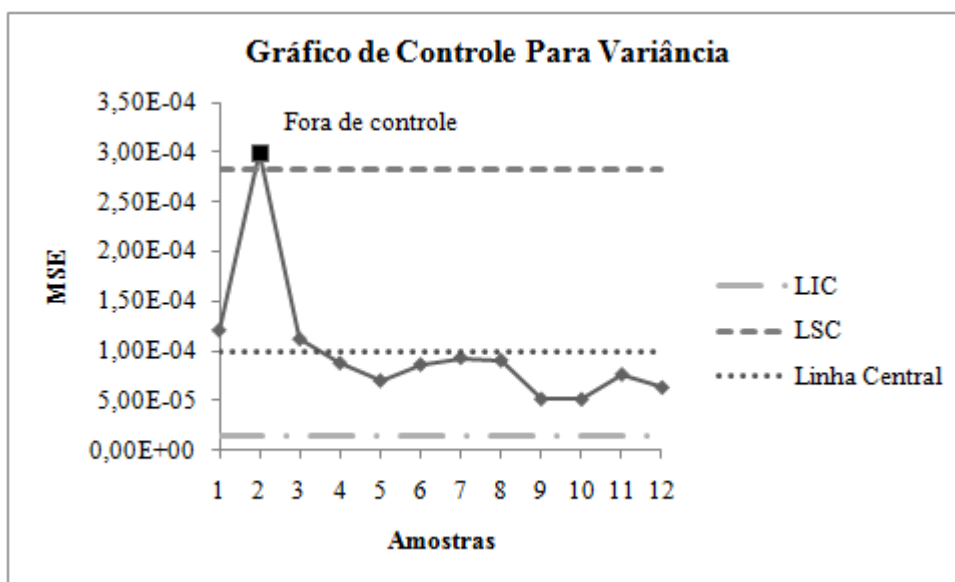


Gráfico 2 – Gráfico de controle para variância para BDC3 em período 2010

Como mostrado no Gráfico 2, para a ação BBDC3 no período 2010 a amostra 2 foi detectada como *fora de controle* e deve ser portanto excluída das análises posteriores. Veja que uma análise visual do Gráfico 1 permite concluir que a amostra 2 possui variância maior que as demais, o que faz sentido já que o MSE é um estimador da variância. Depois de excluída a amostra 2, que contém 12 observações, restam $N = 144 - 12 = 132$ observações.

Esse procedimento foi realizado para todas as ações em todos os períodos, visando a remoção de amostras *fora de controle* em relação à estabilidade da variância. Após essa etapa, o próximo passo é a execução do teste F Global, para a hipótese nula de que todas as retas de regressão nesse período são coincidentes. Novamente, inicia-se com a definição do nível de significância para o teste. Como no caso em questão tem-se $p = 1$, o uso da expressão (52) resulta no nível de significância para o teste F global de 0,033617.

Os resultados finais dessa análise para todas as ações e períodos coletados podem ser vistos no Apêndice A. Para o caso da ação Gol Linhas Aéreas (código GOLL4 na Bolsa de Valores Bovespa), período 2009, observou-se $SSE_{\text{reduzido}} = 0,1466$, referente ao *modelo reduzido* apresentado em (50), $SSE_{\text{completo}} = 0,1084$ e $MSE_{\text{completo}} = 0,0009033$, referentes ao *modelo completo* apresentado em (48). O resultado do teste F global dado por (51) é 1,9206, cujo valor P é 0,01377. Esse valor é inferior ao nível de significância do teste, que é de 0,033617, indicando situação *fora de controle*, ou seja, situação na qual *rejeita-se* a hipótese

de que as retas de regressões sejam coincidentes. Esta etapa foi executada para todos os períodos e ações, detectando quais apresentavam condição *fora de controle*.

Para os casos que apresentavam situação *fora de controle* para a hipótese de igualdade das retas de regressão das amostras segundo o teste F global, foi realizado o teste F visando identificar se a causa da variabilidade é a inclinação. No caso da ação GOLL4 para período 2009, o valor do teste F para a inclinação, dado pela relação (54), foi de 2,1033, indicando que a causa da divergência é a inclinação. Caso não fosse, prosseguir-se-ia com o teste F análogo para teste do intercepto, conforme detalhamento da metodologia no capítulo anterior.

Identificadas as causas da variabilidade, devem-se identificar quais são as amostras *fora de controle*. Conforme discutido em (56), para todas as amostras com exceção da última (amostra de referência), calcula-se a estatística F em (56). Dando prosseguimento à análise da ação GOLL4 no período 2009, obtiveram-se os resultados mostrados na Tabela 7.

Tabela 7 – Valores obtidos para teste F de detecção de amostras *fora de controle* de GOLL4 em 2009

<i>Amostra</i>	<i>F Global</i>	<i>Valor P</i>
1	0,6110	0,5526
2	1,1600	0,3342
3	1,3200	0,2884
4	0,8690	0,4345
5	4,4800	0,0247
6	3,2100	0,0618
7	1,5400	0,2388
8	0,4370	0,6517
9	1,0400	0,3730
10	3,1300	0,0656
11	1,0400	0,3722

As amostras identificadas como *fora de controle* são as amostras 5 e 6. É possível confirmar essa constatação através de uma análise do Gráfico 3(a), contendo as retas de regressão ajustadas aos dados das amostras: as retas das amostras 5 e 6 aparentam ser diferentes das demais. Além disso, a comparação desse gráfico com o Gráfico 3(b) e Gráfico 3(c) permite uma avaliação quantitativa a respeito do nível de ajuste do modelo. Nota-se que as retas do Gráfico 3(a) possuem inclinação muito mais variada que as dos outros dois gráficos, o Gráfico 3(c) possuindo retas com inclinações mais semelhantes entre si que os

outros três. O nível de semelhança entre as inclinações das retas constitui um indicador qualitativo do nível de ajuste esperado para o método.

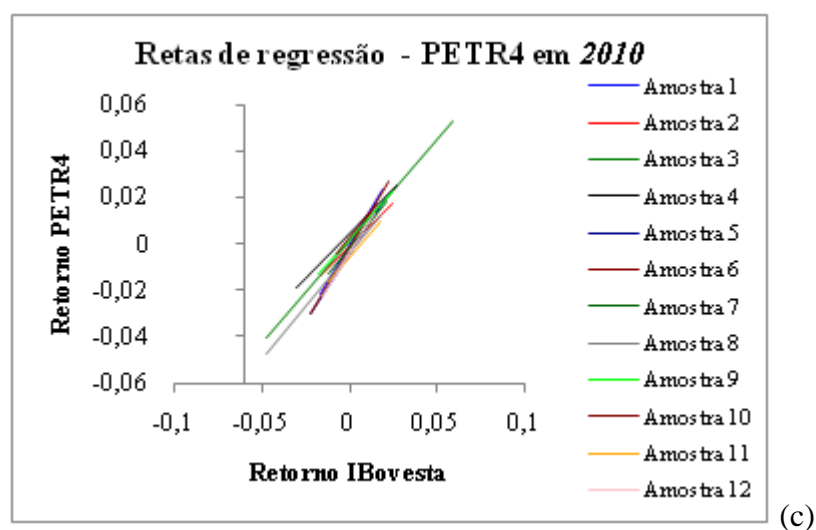
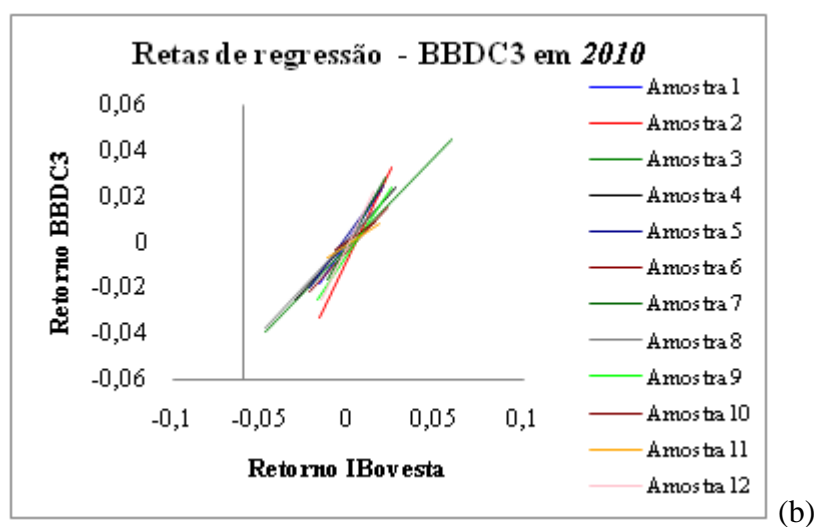
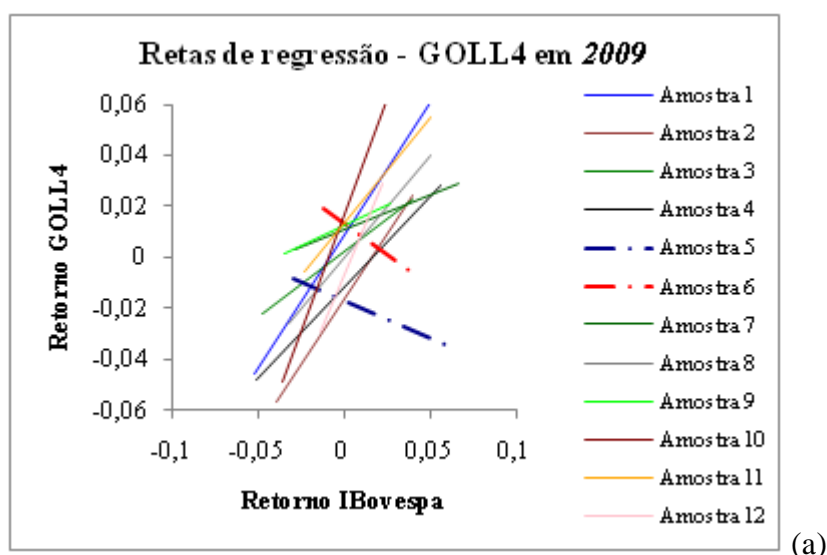


Gráfico 3 – Retas de regressão ajustadas para as amostras

Em suma, com referência à análise de GOLL4 em 2009, após remoção das amostras detectadas como *fora de controle* pelos testes F, o número de amostras final foi $m_{final} = 10$. Pode-se reaplicar o teste F global para verificar se a situação é estável. Procedendo dessa maneira, obtém-se $SSE_{reduzido} = 0,1169$, $SSE_{completo} = 0,09617$ e $MSE_{completo} = 0,0009617$, o resultado do teste F global é 1,1944, com *valor P* = 0,2803. Esse valor é superior ao nível de significância do teste, que é de 0,033617, o que indica situação *sob controle*.

Assim, depois de removidas as amostras instáveis tanto no que concerne o gráfico de controle para a variância quanto os testes F, a hipótese de que se pode considerar a mesma reta de regressão (com os mesmos coeficientes) para todas as amostras não é rejeitada. Ou seja, a hipótese de igualdade das retas de regressão não é rejeitada. O Gráfico 4 mostra as retas de regressão após exclusão das amostras *fora de controle*, evidenciando que uma reta de regressão única pode ser considerada para todas as amostras.

Dessa forma, uma única reta de regressão é ajustada considerando o conjunto de observações das amostras *sob controle*, o coeficiente angular resultante dessa reta sendo o *beta* final resultado do modelo. Tais retas ajustadas são mostradas no Gráfico 4(a), Gráfico 4(b) e Gráfico 4(c) para três os casos, GOLL4 2009, BBDC3 2010 e PETR4 2010, respectivamente.

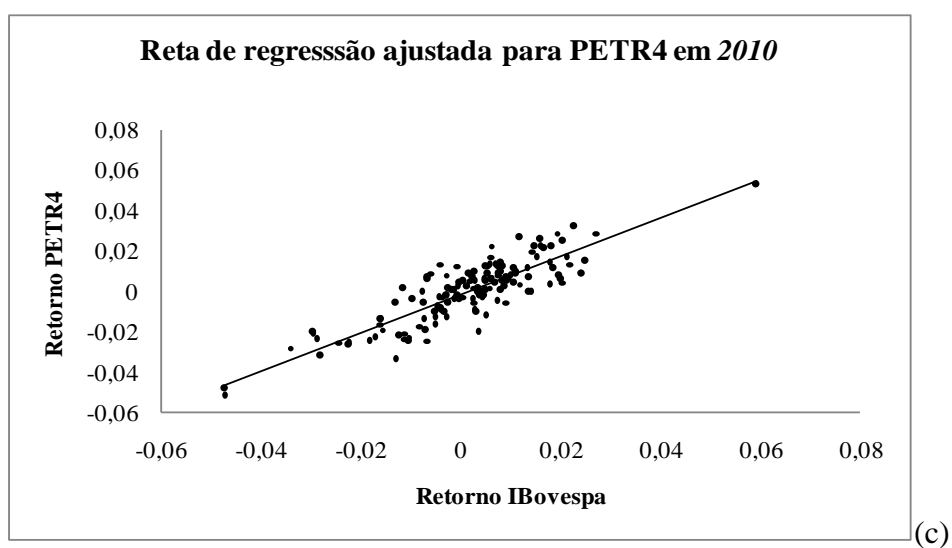
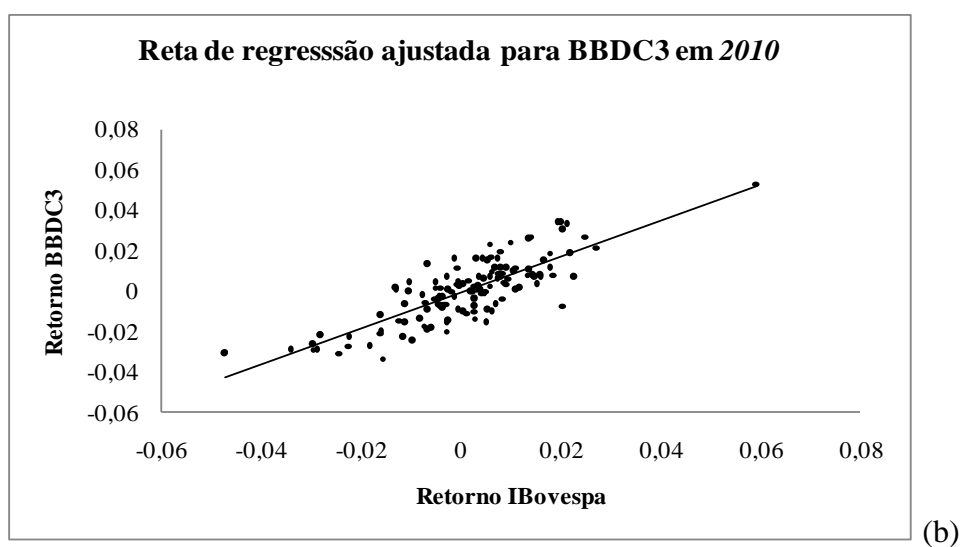
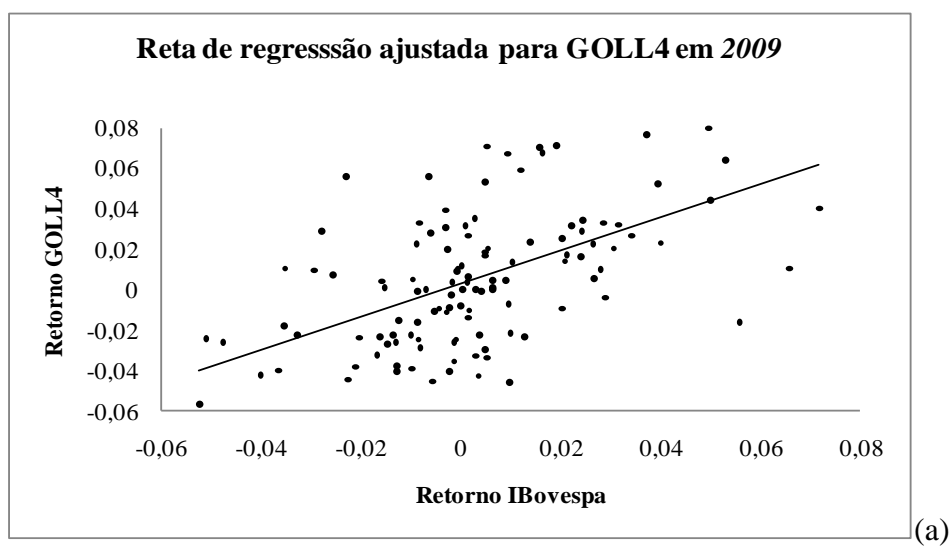


Gráfico 4 – Retas ajustadas, após remoção de amostras *fora de controle*

Podem-se comparar os estimadores da inclinação dos modelos de regressão (*betas*) e suas variâncias estimadas antes e depois da remoção das amostras *fora de controle*. A Tabela 8 mostra as variações obtidas. O estimador da variância do *beta* é calculado segundo as expressões (34) e (35).

Tabela 8 – Valores de *beta* antes e depois do tratamento das amostras

<i>Ação</i>		<i>GOLL4</i>	<i>BBDC3</i>	<i>PETR4</i>
<i>Período</i>		2009	2010	2010
<i>Antes</i>	β original	0,6613	0,9228	0,9501
	$Var(\beta)$ antes	1,40E-02	3,54E-03	2,24E-03
	<i>Remoções variância</i>	0	1	0
<i>Final</i>	<i>Remoções testes F</i>	2	0	0
	β final	0,8194	0,8883	0,9501
	$Var(\beta)$ final	1,57E-02	3,07E-03	2,24E-03
	<i>Variação β</i>	23,91%	-3,74%	0,00%
	<i>Variação $Var(\beta)$</i>	11,52%	-13,16%	0,00%

A exclusão de amostras detectadas como *fora de controle* pelo gráfico de controle aplicado para variância gera necessariamente uma redução no estimador da variância de *beta* (equações (34) e (35)). A remoção de amostras detectadas como *fora de controle* na fase I do monitoramento de perfil linear resulta em uma melhoria significativa na medida do *beta* (variação de 23,91% para o caso GOLL4 no período 2009). A ação PETR4 no período 2010 não apresentou ganho com a aplicação do método, já que os dados apresentavam suficiente estabilidade para este caso, não levando a nenhum sinal *fora de controle*.

A metodologia consiste, portanto, em obter um estimador do *beta* mais preciso que o obtido equivocadamente ao considerar todas as amostras iniciais. O *beta* final obtido após a remoção das amostras *fora de controle* é um indicador mais confiável da relação entre os retornos da ação e do índice, proporcionando uma maior precisão para aplicação de estratégias de neutralização. Além disso, como demonstrado pela teoria do *Market Model*, esse *beta* representa uma melhor medida do risco sistemático da empresa.

Para demonstrar o resultado do método, pode-se analisar o retorno acumulado da estratégia *Market Neutral* usando o *beta* obtido antes e depois da aplicação do método. O retorno da estratégia é dado por (25). Usando os dados da Tabela 8, pode-se escolher um valor “verdadeiro” de *beta* que esteja mais próximo do valor do *beta* final que do original. Assim,

para ilustração, usar-se-á como valor verdadeiro de β para GOLL4 no período 2009 e para BBDC3 no período 2010, o valor 0,85, que se encontra dentro do intervalo de confiança de 95% para os dados de β final da tabela. No Gráfico 5(a) e no Gráfico 5(b) são mostradas as curvas de retorno acumulado dadas pela estratégia definida pela expressão (25) aplicada aos ativos GOLL4 no período 2009 e BBDC3 no período 2010, respectivamente.

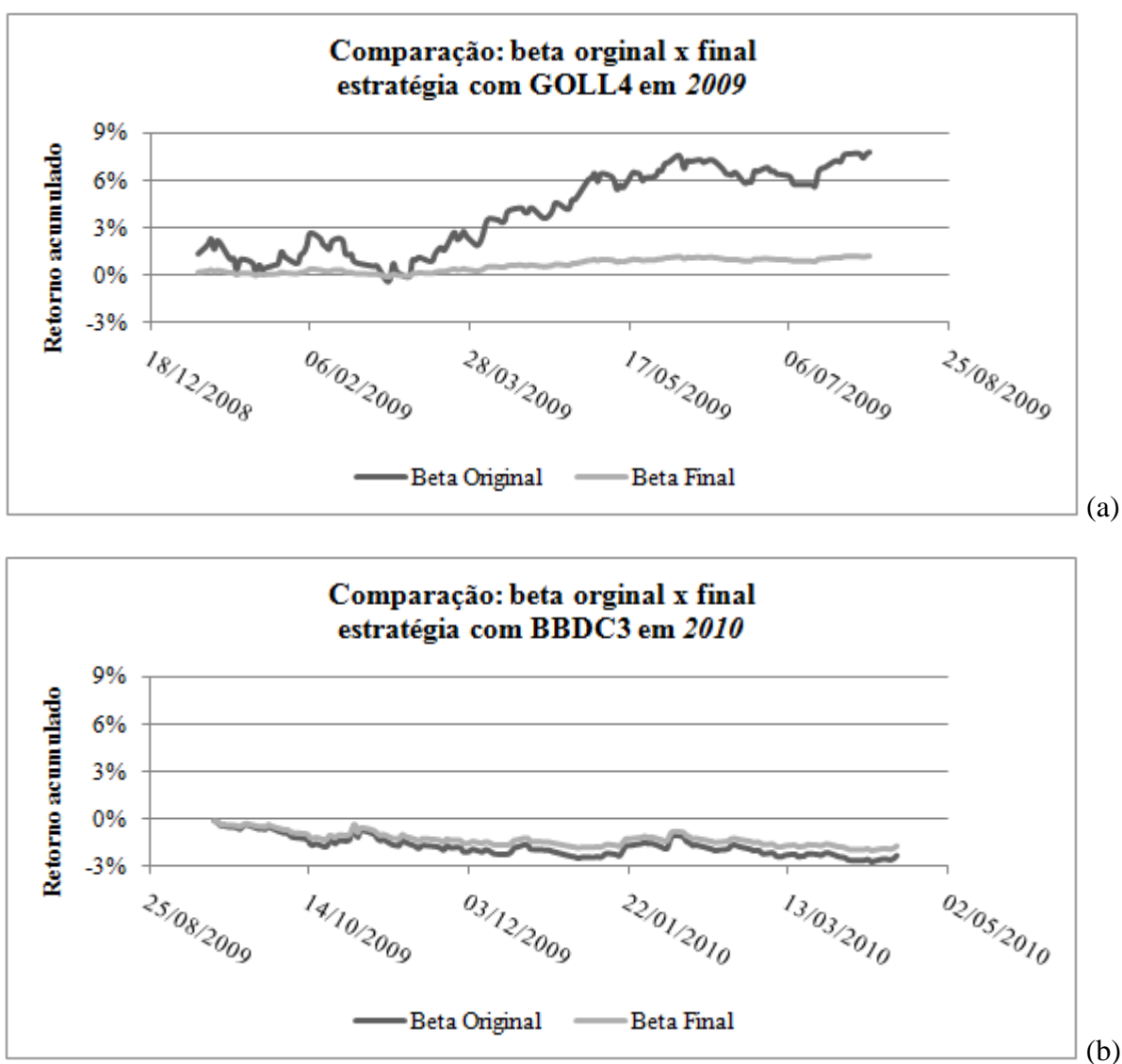


Gráfico 5 – Market Neutral – influência do estimador de β

Para o caso de GOLL4 2009, a diferença foi maior, o que está de acordo com as diferenças indicadas na Tabela 8. Ademais, fica clara a eficiência dos novos estimadores no sentido da proteção: redução de variabilidade e manutenção da proximidade do ponto de retorno neutro.

As etapas de aplicação do gráfico de controle para variância, remoção de amostras *fora de controle*, agrupamento das amostras em uma única amostra e aplicação do teste F global (seguido da aplicação dos dois outros testes F, caso o teste F global indicasse situação de instabilidade) e identificação e remoção das amostras *fora de controle* foram repetidas para as 17 ações selecionadas em cada um dos 9 períodos. Para algumas ações não existem dados disponíveis para aplicação da metodologia a todos os períodos, sendo, portanto, feita a aplicação para os períodos nos quais se tem todos os dados necessários. Caso todas as ações tivessem dados de preço de fechamento para todas as observações (dias) consideradas nos 9 períodos, poder-se-ia aplicar a metodologia $17 * 9 = 153$ vezes. Porém, algumas dessas ações começaram a ser negociadas na Bovespa em períodos posteriores ao primeiro período analisado ou apresentavam baixa liquidez nos primeiros períodos coletados, impedindo a coleta de uma relação completa de variações de preço diárias no período. Por esse motivo, os resultados são fruto de pouco menos de 153 repetições da metodologia apresentada anteriormente.

Para viabilizar a execução dessa análise foi desenvolvido um programa em linguagem Visual Basic que realiza todas as etapas do método apresentado, realizando os cálculos matriciais, testes de hipóteses e remoção de amostras *fora de controle*. Tal programa, bem como a relação completa dos resultados obtidos através do uso do mesmo para aplicação da metodologia aos dados coletados, encontra-se no Apêndice B.

A seguir é apresentada uma análise por setor baseada nesses resultados. O objetivo de tal análise não é estudar as causas mercadológicas ou econômicas envolvidas no comportamento do *beta* através dos períodos, mas ilustrar uma possível utilização do método no sentido de melhorar a qualidade de conclusões a respeito das inferências sobre o risco sistemático das ações. Para entendimento de tais análises é importante saber que o *beta* representa medida de risco sistemático (risco não diversificável), isto é, um *beta* intrínseco à atividade da empresa. Assim, é de se esperar que setores semelhantes apresentem *betas* de valor semelhante no decorrer do tempo. É também importante ressaltar que a medida do *beta* feita através do preço das ações na Bolsa de Valores recebe influência de atitudes puramente

especulativas de investidores, desvinculadas da atividade da empresa, mas que têm influência sobre o comportamento da medida de risco sistemático, feita através do *Market Model*, usando dados de preços de ações e índices da bolsa.

Transporte aéreo

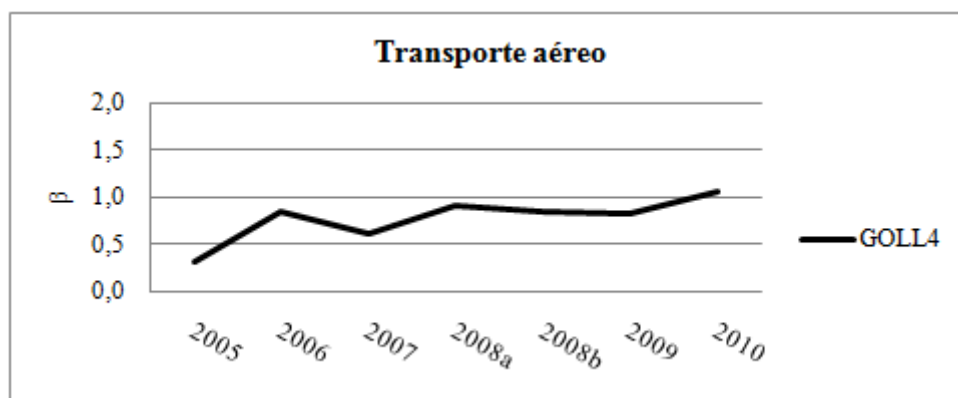


Gráfico 6 – Comportamento do *beta* da categoria *Transporte Aéreo*

Através do Gráfico 6 observa-se que a ação GOLL4, da empresa Gol Linhas Aéreas, apresenta *variabilidade decrescente* do *beta* do período 2004 a 2010. Uma provável hipótese que se pode levantar é a de que, a partir do início da negociação de uma ação na Bolsa de Valores, o *beta* de uma empresa leva alguns anos para assumir um valor estável, que reflita o consenso do mercado sobre o risco sistemático da ação. Tal efeito seria mais evidente para ações de empresas pertencentes a setores onde há poucos atuantes no mercado, como é o caso do setor de transporte aéreo, já que esse fato dificulta a tomada de outras ações como referência de risco.

Água e saneamento

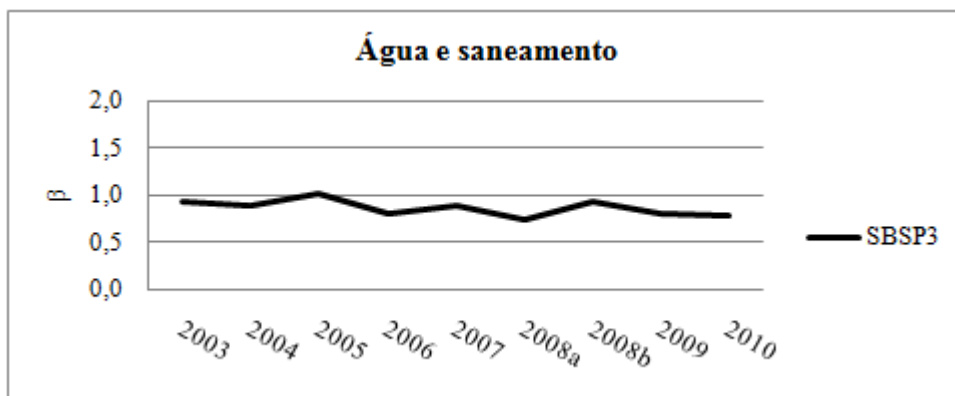


Gráfico 7 – Comportamento do *beta* da categoria *Água e saneamento*

O Gráfico 7 mostra que, no período analisado, o setor de distribuição de água e saneamento apresenta boa estabilidade quando comparado aos demais setores. Uma hipótese provável é a de que por se tratar de um serviço básico à população e um monopólio, esse setor tem risco sistemático muito bem assimilado pelo mercado e sua percepção pelos investidores sofre pouca influência de oscilações na economia geral do país.

Bancos

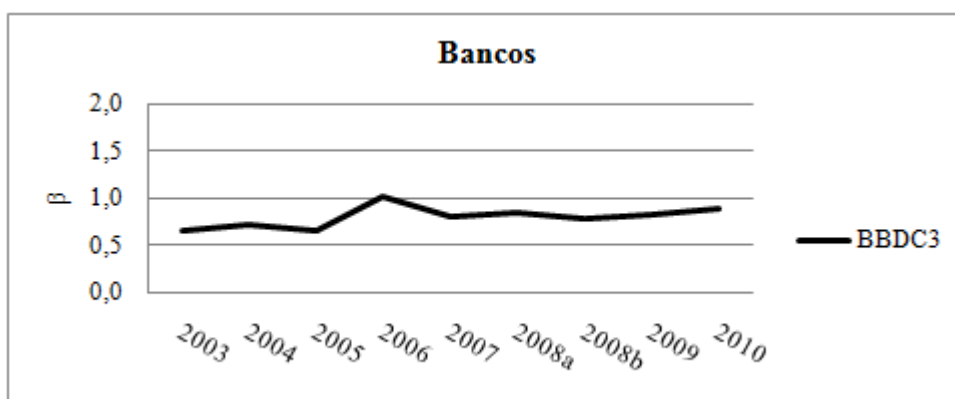


Gráfico 8 – Comportamento do *beta* da categoria *Bancos*

A análise do comportamento das ações do Banco Bradesco no decorrer dos períodos tratados revela variabilidade relativamente baixa, dado que entre 2008a e 2008b houve a grande crise econômica global (ver Gráfico 8). Essa baixa variabilidade do risco sistemático, revelando a não variação das percepções dos investidores no que refere a variabilidade dos

retornos da ação com relação aos retornos da economia como um todo, pode indicar uma possível proteção existente no sistema bancário brasileiro às variações da economia. Um estudo mais detalhado sobre este comportamento pode ser abordado em futuros trabalhos no ramo de ciências financeiras e econômicas.

Blue chips

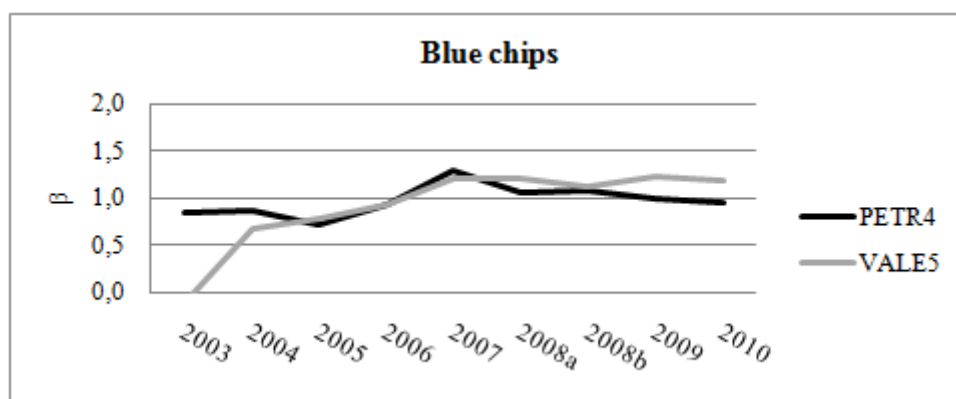


Gráfico 9 – Comportamento do *beta* de ações *Blue Chips*

Esta categoria não trata de um setor específico da economia, mas sim de ações de grande capitalização de mercado e cujos valores médios diários negociados são bastante significativos quando comparados à média das ações do mercado. Observa-se nesta categoria que, excluído o *beta* de VALE5 em 2003, cujo valor difere significativamente tanto dos valores de *beta* para essa ação nos demais períodos quanto do valor do *beta* para a ação PETR4 em períodos subsequentes (provavelmente devido a aquisições realizadas pela empresa no período), o comportamento da variação do *beta* para as duas ações é bastante similar. Uma hipótese para tal resultado é a de que os parâmetros de capitalização de mercado e de volume diário negociado, quando atingem certos limites, passam a representar um aspecto mais importante para ajustes referentes ao risco sistemático das empresas do que o fato de pertencerem a um setor ou a outro. Uma razão para isso seria a parcela significativa de especulação financeira sofrida por essas ações através dos grandes volumes diários negociados e do montante em capital aberto (ver Gráfico 9).

Imóveis

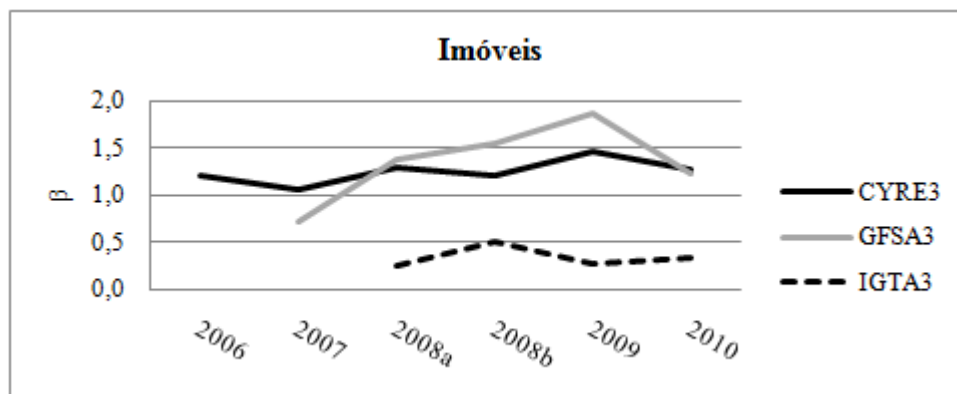


Gráfico 10 – Comportamento do *beta* de ações da categoria *Imóveis*

Neste caso fica evidente a diferença de percepção do risco sistemático entre dois segmentos específicos: o da construção imobiliária e o da exploração (comercialização de imóveis). A análise do Gráfico 10 leva à percepção de que o risco sistemático tem por vezes relação mais fraca com o setor industrial no sentido amplo no qual a empresa está inserida (neste caso *Imóveis*) que com o nicho industrial específico (no caso construção para CYRE3 e GFSA3 e comercialização para IGTA3). Pode haver, no entanto, casos em que o efeito do setor abrangente encobre possíveis variações causadas pelo nicho específico. As análises das duas categorias seguintes representam argumentos para essa hipótese.

Distribuição de energia

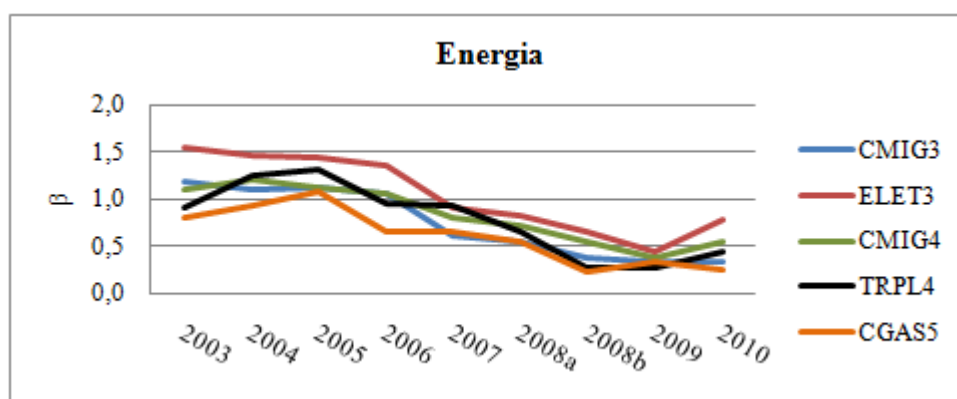


Gráfico 11 – Comportamento do *beta* de ações da categoria *Distribuição de Energia*

Conforme mostrado pelo Gráfico 11, esta categoria mostra forte sincronia no comportamento do *beta* entre as ações. Embora se trate de empresas de tamanhos diferentes, atuando em territórios variados ou mesmo possuindo focos energéticos distintos (CGAS5 é distribuidora de gás natural enquanto as demais estão no domínio de energia elétrica), a variação do risco sistemático das empresas ocorreu de maneira bastante similar no decorrer do período analisado. Foram analisadas ações diferentes pertencentes à CMIG, que, conforme esperado, apresentaram comportamento similar quando ao *beta*. O valor de *beta* relativamente baixo obtido para este setor em 2010 indica um setor *defensivo* nesse período, ou seja, um setor que apresenta menor variabilidade de retornos do que o mercado como um todo e, portando, pode ser visto pelo investidor como um alvo de investimentos visando à proteção contra crises macroeconômicas. Assim, sob certas condições de capitalização de mercado e volume diário negociado, pode-se confirmar a hipótese de que o risco sistemático de empresas de determinados setores pode ser bem estimado através da média das empresas do setor. Tal tópico é deixado para futuras análises.

Rodoviário

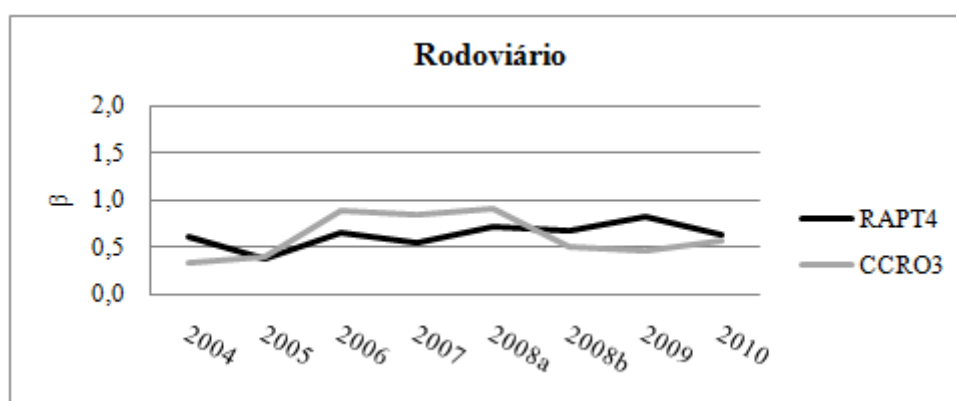


Gráfico 12 – Comportamento do *beta* da categoria Rodoviário

Nesta categoria, pelo Gráfico 12, novamente é observado o comportamento semelhante do risco sistemático para empresas pertencentes a um mesmo setor industrial abrangente, o setor rodoviário. Apesar da diferença das áreas de atuação entre Randon (RATP4), que comercializa peças para veículos de transporte rodoviário, e CCR Rodovias (CCRO3), que atua na área de concessão e manutenção de rodovias, os perfis dos gráficos das duas ações tem aparência bastante similar. Pode-se neste caso confirmar o uso do *beta* como

estimador do risco sistemático da empresa, que não depende de particularidades de cada empresa (não é diversificável), mas sim da atividade exercida pelo setor.

Siderurgia

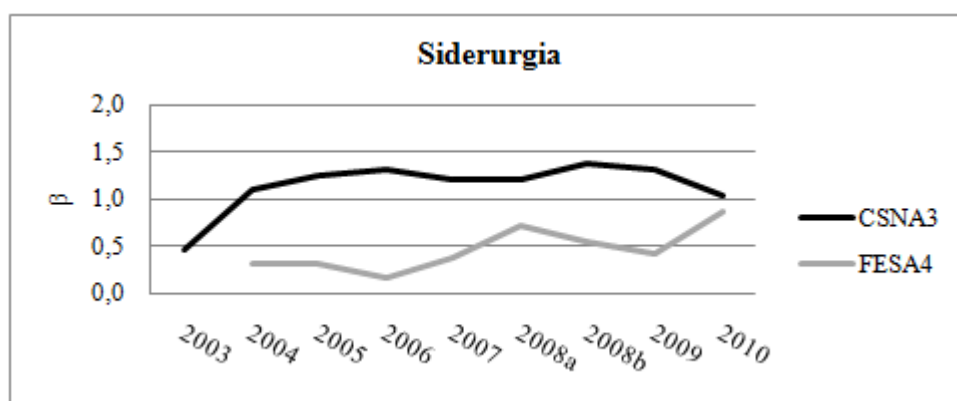


Gráfico 13 – Comportamento do *beta* da categoria Siderurgia

Nesta categoria a diferença dos perfis de *betas* no decorrer dos períodos é evidente entre as duas empresas, apesar de as duas possuírem atividades muito semelhantes (vide Gráfico 13). Uma hipótese provável para tal diferença é o contraste entre os tamanhos das empresas, medidos na forma de volumes médios diários negociados e capitalizações de mercado. Segundo dados da agência Reuters, enquanto CSNA3 tem, capitalização de mercado por volta de 40 bilhões de reais e volume diário médio por volta de 4 milhões e 300 mil ações, FESA4 tem capitalização de mercado de 954 milhões de reais e volume diário médio negociado de 156.830 ações. Esses valores referem-se a dados atuais, mas a diferença relativa manteve-se consideravelmente constante nos últimos anos.

Pode-se estimar que uma parcela do risco sistemático das duas empresas é devida às suas exposições ao mercado financeiro e aos especuladores. O fato de tais exposições serem tão distintas resulta em variabilidade e valores também diversos para a medida de *beta* do *Market Model* aplicado ao preço das ações na Bolsa de Valores como medida do risco sistemático. Ademais, o tamanho inferior da Febrasa (FESA4) pode dificultar a estimação de seus riscos por parte dos investidores, levando a uma flutuação no retorno de suas ações quando comparadas ao índice Bovespa e, portanto, a uma maior variabilidade do *beta*.

Outra aplicação possível da metodologia pode ser feita considerando simultaneamente todos os períodos analisados, após aplicação da metodologia individualmente aos mesmos da maneira detalhada neste trabalho. Tal composição consiste em considerar, para cada ação selecionada e cada um dos 9 períodos, o conjunto das amostras finais (as consideradas *sob controle*) como sendo uma única amostra. Assim, para cada ação ter-se-ia 9 amostras de entrada para aplicação da mesma metodologia, agora envolvendo o horizonte amplo dos 9 períodos, ou seja, $m_0 = 9$ amostras.

No Gráfico 14 está esquematizada, usando o exemplo da ação VALE5, a composição das amostras dos períodos para esta nova abordagem, no qual os períodos já tratados são considerados amostras únicas. Um detalhe interessante desse gráfico é o aumento visível da variabilidade dos retornos durante o período 2008b, auge da crise econômica mundial. Para o exemplo da ação VALE5, os parâmetros resultantes das análises individuais dos períodos, mostrados na Tabela 9, seriam utilizados.

Tabela 9 – Resultados finais para VALE5

<i>Período</i>	<i>Amostras removidas na análise do período</i>	<i>Total de observações restantes</i>
2003	0	144
2004	0	144
2005	3	108
2006	0	144
2007	0	144
2008a	0	144
2008b	0	144
2009	0	144
2010	0	144

A metodologia não se limita ao caso de amostras com mesmo número de observações. Assim, é possível reaplicar a metodologia considerando os conjuntos das observações restantes dos períodos como amostras com números de observações distintos.

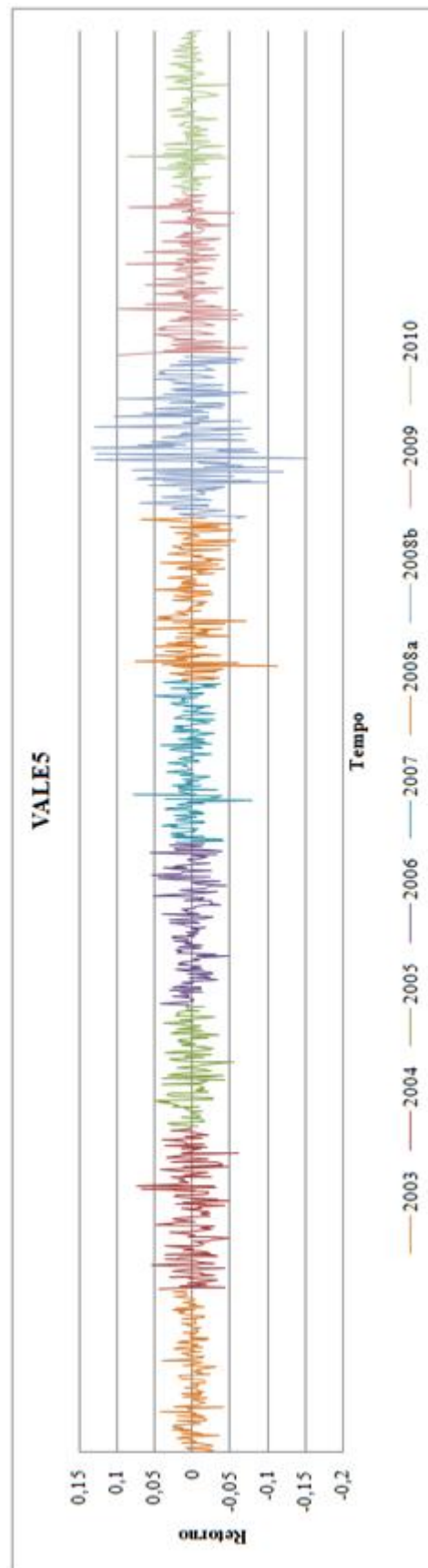


Gráfico 14 – Períodos tratados como amostras para análise mais ampla

7. CONCLUSÕES

O presente trabalho propôs-se a estudar o modelo *Market Model* e utilizar ferramentas de controle estatístico de processo para estimar seus parâmetros com precisão maior que a de métodos comumente usados com este intuito, como a regressão linear simples, no contexto de análise de risco de um portfólio de ativos. Ademais, a aplicação detalhada do método para ações de um conjunto variado de setores industriais e o posterior desenvolvimento de análises dos resultados obtidos permitiram a demonstração de aplicações práticas da metodologia apresentada.

O estudo possibilitou o detalhamento das diversas etapas constituintes do método utilizado, evidenciando os cálculos desenvolvidos, bem como as hipóteses consideradas. Foram expostas aplicações de testes de verificação para as premissas consideradas para os dados coletados, como o teste Shapiro-Wilk para a normalidade e o teste *Runs* para a linearidade. Ademais, foram exploradas ferramentas de controle estatístico de processo, como gráficos de controle e a fase I do monitoramento de perfis lineares.

Para possibilitar a aplicação do método a um grupo de ações selecionadas e considerando períodos representativos de diferentes situações econômicas, foram utilizados softwares como Microsoft Excel e MatLab e um programa de computador, desenvolvido em linguagem Visual Basic especificamente para a aplicação realizada. A eficiência do método foi comprovada através de uma análise gráfica envolvendo a redução do risco sistemático, aplicada aos resultados obtidos.

Foi indicada uma forma de aplicação que considera resultados dos períodos já analisados como dados de entrada para uma nova análise que abrange um horizonte de tempo mais amplo e tende a ressaltar aspectos menos voláteis e mais representativos do comportamento do risco sistemático da empresa.

As aplicações desenvolvidas no presente trabalho tomaram como dados de entradas preços de ações de empresas de diversos setores para vários períodos, disponíveis no mercado. Os resultados obtidos por meio de tais aplicações permitiram o desenvolvimento de análises de tais setores, nas quais foram sugeridas algumas hipóteses sobre o comportamento do *beta* calculado (representante do risco sistemático dos ativos analisados) com base em fatos observados na economia mundial. Apesar de não ser o foco do presente trabalho,

algumas observações interessantes puderam ser verificadas, como o comportamento semelhante para o *beta* de todas as ações de empresas do setor energético estudadas, condizente com o tipo de produto que fornecem, um serviço básico para a população.

Como sugestões de estudos futuros, pode-se mencionar:

- Aplicação do método para períodos mais extensos, de ordem superior a 10 anos, proporcionando uma análise ainda mais completa das empresas e setores analisados;
- Análise estruturada de setores industriais, verificando a correlação entre o comportamento dos *betas* calculados por meio do método apresentado e acontecimentos da economia mundial.
- Aplicação do método utilizado, tomando como parâmetro representante do retorno do mercado, outros índices que não o IBovespa, tais como PIB, índice de desemprego ou outros índices representativos do desempenho da economia.

REFERÊNCIAS

ALBRIGHT, S. C. et al. **Data analysis for managers with Microsoft Excel**. 2nd ed. Belmont: Thomson, 2004. 952 p.

ARROW, K. J. **Essays in the theory of risk-bearing**. Chicago: Markham Pub. Co., 1971. 278 p.

BODIE, Z. et al. **Investments**. 8th ed. Boston: McGraw-Hill, 2008. 990 p.

BURN, D.; RYAN JR, T. A diagnostic test for lack of fit in regression models. **ASA - Proceedings of the Statistical Computing**, p. 286-290, 1983.

ELTON, E.; GRUBER, M. Modern portfolio theory, 1950 to date. **Journal of Banking and Finance**, v. 21, n. 11/12, p. 1743–1760, 1997.

ELTON, E. J.; GRUBER, M. J. **Modern portfolio theory and investment analysis**. 4th ed. New York: John Wiley, 1991. 736 p.

FAMA, E. F. **Foundations of finance**. New York: Basic Books, 1976. 395 p.

FARRAR, D. E.; GLAUBER, R. R. **Multicollinearity in regression analysis**. Cambridge: MIT, 1964. 49 p.

FRIEDMAN, M.; SAVAGE, L. The utility analysis of choices involving risk. **The Journal of Political Economy**, v. 56, n. 4, p. 279-304, 1948.

FUTURO de índice Bovespa. São Paulo: Bolsa de Mercadorias & Futuros, 2006. 24 p. Disponível em: <http://www.bmfbovespa.com.br/pt-br/educacional/download/series-mercados_futuro-de-indice-bovespa.pdf>. Acesso em: 14 fev. 2010.

GITMAN, L. **Princípios de administração financeira**. 10. ed. São Paulo: Addison Wesley, 2004. 745 p.

GUJARATI, D. N. Use of dummy variables in testing for equality between sets of coefficients in linear regressions: a generalization. **American Statistician**, v. 24, n. 5, p. 18-22, 1970.

- GUJARATI, D. N. **Econometria básica**. 4. ed. Rio de Janeiro: Elsevier, 2006. 697 p.
- HALLERBACH, W. G. **Multi-attribute portfolio selection**. Rotterdam: Erasmus University Rotterdam, 1994. 347 p.
- JONES, C. P. **Investments analysis and management**. 6th ed. New York: John Wiley, 1998. 663 p.
- KANG, L.; ALBIN, S. On-line monitoring when the process yields a linear. **Journal of Quality Technology**, v. 32, n. 4, p. 418-426, 2000.
- KIM, K. et al. On the monitoring of linear profiles. **Journal of Quality Technology**, v. 35, n. 3, p. 317-328, 2003.
- KLEIN, L. R. **An introduction to econometrics**. Englewood Cliffs: Prentice-Hall, 1962. 280 p.
- KUTNER, M. H. et al. **Applied linear regression models**. 4th ed. Maidenhead: McGraw-Hill Education, 2003. 701 p.
- LEVINE, D. M. **Statistics for managers using Microsoft Excel**. 5th ed. Harlow: Prentice Hall, 2008. 858 p.
- LINTNER, J. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. **The Review of Economics and Statistics**, v. 47, n. 1, p. 13-37, 1965.
- MAHMOUD, M. Phase I analysis of multiple linear regression profiles. **Communications in Statistics: Simulation and Computation**, v. 37, n. 10, p. 2106-2130, 2008.
- MAHMOUD, M. et al. A change point method for linear profile data. **Quality and Reliability Engineering International**, v. 23, n. 2, p. 247-268, 2006.
- MAHMOUD, M.; WOODALL, W. Phase I analysis of linear profiles with calibration applications. **Technometrics**, v. 46, n. 4, p. 380-391, 2004.
- MARKOWITZ, H. Portfolio selection. **Journal of finance**, v. 7, n. 1, p. 77-91, 1952.

MELLAGI FILHO, A.; ISHIKAWA, S. **Mercado financeiro e de capitais**. Editora Atlas, 2000. 335 p.

MONTGOMERY, D. C. **Introduction to statistical quality control**. 5th ed. Hoboken: John Wiley, 2005. 759 p.

MOSSIN, J. Equilibrium in a capital asset market. **Econometrica: Journal of the Econometric Society**, v. 34, n. 4, p. 768-783, 1966.

NETER, J. et al. **Applied linear statistical models**. 3rd ed. Homewood: Irwin, 1990. 1181 p.

PRATT, J. Risk Aversion in the Small and in the Large. **Econometrica: Journal of the Econometric Society**, v. 32, n. 1, p. 122-136, 1964.

REILLY, F. K.; BROWN, K. C. **Investment analysis and portfolio management**. 8th ed. Mason: Thomson, 2006. 1174 p.

ROSS, S. A. et al. **Corporate finance**. 7th ed. Boston: McGraw-Hill, 2005. 942 p.

RYAN, T. P. **Modern regression methods**. New York: John Wiley, 1997. 515 p.

SECURATO, J. **Decisões financeiras em condições de risco**. São Paulo: Atlas, 1996. 250 p.

SHARPE, W. F. A simplified model for portfolio analysis. **Management science**, v. 9, n. 2, p. 277-293, 1963.

SHARPE, W. F. Capital asset prices: A theory of market equilibrium under conditions of risk. **Journal of finance**, v. 19, n. 3, p. 425-442, 1964.

SHARPE, W. F. Capital asset prices with and without negative holdings. In: **Economic Sciences: 1981-1990 (Nobel Lectures)**. Stanford: Nobel Foundation, 1990. p. 312-332.

SHARPE, W. F. et al. **Investments**. 5th ed. Englewood Cliffs: Prentice Hall, 1995. 1058 p.

WILLIAMS, J. et al. Statistical monitoring of nonlinear product and process quality profiles. **Quality and Reliability Engineering International**, v. 23, n. 8, p. 925-941, 2007.

WOODALL, W. Current research on profile monitoring. **Produção**, v. 17, p. 420-425, 2007.

APÊNDICES

APÊNDICE A – RESULTADOS DA APLICAÇÃO DA METODOLOGIA

Legenda

Ex. Var. : Número de amostras excluídas após aplicação do gráfico de controle para variância.

Ex. H_0 : Número de amostras excluídas após aplicação dos testes de hipóteses sobre a igualdade das retas de regressão.

N: número final de observações consideradas para obtenção dos coeficientes finais.

α final: Intercepto para o modelo de regressão final, após remoção das amostras fora de controle.

B final : Intercepto para o modelo de regressão final, após remoção das amostras fora de controle.

MSE final : MSE para o modelo final.

Os traços indicam dados que não puderam ser calculados devido à ausência de informações completas sobre preços da ação no período mencionado. Em geral isso ocorre devido à baixa liquidez da ação no período.

Os dados serão apresentados agrupados por categoria.

Transporte Aéreo

Tabela A. 1 – Resultados para o setor *Transporte aéreo*

<i>Nome</i>	<i>Código</i>	<i>Período</i>	<i>Ex. Var</i>	<i>Ex. H₀</i>	<i>N</i>	<i>α final</i>	<i>B</i>	<i>MSE final</i>
<i>GOL ON</i>	<i>GOLL4</i>	2003	-	-	-	-	-	-
		2004	-	-	-	-	-	-
		2005	0	0	144	-0,0009	0,3185	7,382E-04
		2006	1	0	132	-0,0013	0,8528	7,590E-04
		2007	1	0	132	-0,0036	0,6084	2,462E-04
		2008a	2	0	120	-0,0080	0,9134	8,700E-04
		2008b	3	0	108	0,0005	0,8374	2,552E-03
		2009	0	2	120	0,0028	0,8194	9,903E-04
		2010	0	0	144	0,0009	1,0477	5,072E-04

Água e saneamento

Tabela A. 2 – Resultados para o setor *Água e saneamento*

<i>Nome</i>	<i>Código</i>	<i>Período</i>	<i>Ex. Var</i>	<i>Ex. H₀</i>	<i>N</i>	<i>α final</i>	<i>B</i>	<i>MSE final</i>
<i>Sabesp ON</i>	<i>SBSP3</i>	2003	0	0	144	0,0002	0,9230	3,885E-04
		2004	0	0	144	-0,0008	0,8938	4,929E-04
		2005	0	0	144	-0,0003	1,0259	3,971E-04
		2006	1	0	132	0,0018	0,8086	3,179E-04
		2007	0	0	144	0,0004	0,8896	3,988E-04
		2008a	0	0	144	0,0001	0,7488	3,828E-04
		2008b	0	0	144	0,0002	0,9277	9,796E-04
		2009	0	0	144	-0,0008	0,8064	6,234E-04
		2010	0	0	144	-0,0005	0,7735	1,825E-04

Bancos

Tabela A. 3 – Resultados para o setor *Bancos*

<i>Nome</i>	<i>Código</i>	<i>Período</i>	<i>Ex. Var</i>	<i>Ex. H₀</i>	<i>N</i>	<i>α final</i>	<i>B final</i>	<i>MSE final</i>
<i>Bradesco ON</i>	<i>BBDC3</i>	2003	0	0	144	-0,0006	0,6653	2,333E-04
		2004	0	0	144	-0,0008	0,7238	2,154E-04
		2005	0	0	144	0,0023	0,6654	3,817E-04
		2006	0	1	132	0,0009	1,0242	2,788E-04
		2007	0	0	144	0,0003	0,8070	1,888E-04
		2008a	0	0	144	-0,0004	0,8386	2,284E-04
		2008b	2	0	120	-0,0005	0,7760	3,819E-04
		2009	0	0	144	-0,0006	0,8247	2,057E-04
		2010	1	0	132	-0,0005	0,8883	8,289E-05

Blue Chips**Tabela A. 4 – Resultados para o setor *Blue Chips***

<i>Nome</i>	<i>Código</i>	<i>Período</i>	<i>Ex.</i> <i>Var</i>	<i>Ex.</i> <i>H₀</i>	<i>N</i>	<i>α final</i>	<i>B final</i>	<i>MSE final</i>
<i>Petrobrás</i>	<i>PETR4</i>	2003	0	0	144	0,0005	0,8471	1,956E-04
		2004	0	0	144	0,0002	0,8594	1,967E-04
		2005	0	0	144	0,0011	0,7233	1,442E-04
		2006	0	0	144	0,0009	0,9401	1,556E-04
		2007	0	7	60	0,0002	1,2886	7,770E-05
		2008a	0	0	144	-0,0007	1,0618	3,061E-04
		2008b	0	0	144	0,0017	1,0836	4,428E-04
		2009	0	0	144	-0,0002	0,9918	9,124E-05
		2010	0	0	144	-0,0011	0,9501	6,460E-05
<i>Vale</i>	<i>VALE5</i>	2003	0	0	144	0,0002	-0,0989	2,204E-04
		2004	0	0	144	-0,0006	0,6752	3,867E-04
		2005	0	3	108	0,0000	0,7913	2,776E-04
		2006	0	0	144	-0,0003	0,9269	1,629E-04
		2007	0	0	144	0,0011	1,2019	9,953E-05
		2008a	0	0	144	-0,0007	1,2157	1,826E-04
		2008b	0	0	144	0,0009	1,1128	2,915E-04
		2009	0	0	144	-0,0009	1,2339	1,551E-04
		2010	0	0	144	0,0015	1,1753	7,008E-05

Imóveis**Tabela A. 5 – Resultados para o setor *Imóveis***

<i>Nome</i>	<i>Código</i>	<i>Período</i>	<i>Ex.</i> <i>Var</i>	<i>Ex.</i> <i>H₀</i>	<i>N</i>	<i>α final</i>	<i>B final</i>	<i>MSE final</i>
<i>Cyrela Realit ON</i>	<i>CYRE3</i>	2003	-	-	-	-	-	-
		2004	-	-	-	-	-	-
		2005	-	-	-	-	-	-
		2006	0	0	144	-0,0008	1,2132	1,035E-03
		2007	0	0	144	-0,0006	1,0622	5,153E-04
		2008a	0	0	144	0,0006	1,2982	6,913E-04
		2008b	1	0	132	-0,0017	1,2163	1,470E-03
		2009	0	0	144	0,0021	1,4625	1,102E-03
		2010	0	0	144	-0,0023	1,2731	3,149E-04
<i>Gafisa ON</i>	<i>GFSA3</i>	2003	-	-	-	-	-	-
		2004	-	-	-	-	-	-
		2005	-	-	-	-	-	-
		2006	-	-	-	-	-	-
		2007	0	0	144	-0,0012	0,7230	4,874E-04
		2008a	0	0	144	-0,0003	1,3695	6,884E-04
		2008b	0	0	144	-0,0005	1,5456	2,051E-03
		2009	0	7	60	0,0099	1,8585	1,158E-03
		2010	1	1	120	-0,0026	1,2262	2,756E-04
<i>Iguatemi ON</i>	<i>IGTA3</i>	2003	-	-	-	-	-	-
		2004	-	-	-	-	-	-
		2005	-	-	-	-	-	-
		2006	-	-	-	-	-	-
		2007	-	-	-	-	-	-
		2008a	2	0	120	-0,0032	0,2617	2,606E-04
		2008b	2	3	84	0,0066	0,4995	8,142E-04
		2009	0	0	144	0,0034	0,2712	3,393E-04
		2010	0	0	144	0,0015	0,3330	3,341E-04

Distribuição de energia**Tabela A. 6 – Resultados para o setor *Distribuição de energia***

<i>Nome</i>	<i>Código</i>	<i>Período</i>	<i>Ex. Var</i>	<i>Ex. H₀</i>	<i>N</i>	<i>α final</i>	<i>B final</i>	<i>MSE final</i>
<i>Cemig ON</i>	<i>CMIG3</i>	2003	0	0	144	-0,0017	1,1847	4,287E-04
		2004	0	0	144	0,0008	1,0997	4,599E-04
		2005	0	0	144	0,0009	1,1144	4,036E-04
		2006	0	0	144	-0,0005	1,0601	3,655E-04
		2007	3	0	108	-0,0004	0,6077	1,968E-04
		2008a	0	0	144	0,0006	0,5411	3,573E-04
		2008b	1	0	132	0,0006	0,3876	5,509E-04
		2009	0	0	144	0,0000	0,3275	2,848E-04
		2010	0	0	144	0,0002	0,3352	1,648E-04
<i>Eletrobras ON</i>	<i>ELET3</i>	2003	0	0	144	-0,0026	1,5372	4,961E-04
		2004	0	0	144	0,0000	1,4626	7,101E-04
		2005	0	0	144	-0,0007	1,4456	4,460E-04
		2006	0	0	144	0,0007	1,3628	5,278E-04
		2007	0	0	144	-0,0014	0,9192	1,675E-04
		2008a	1	0	132	0,0013	0,8204	4,709E-04
		2008b	1	0	132	0,0002	0,6485	6,456E-04
		2009	0	0	144	0,0000	0,4501	3,724E-04
		2010	4	0	96	-0,0011	0,7743	1,875E-04
<i>Cemig PN</i>	<i>CMIG4</i>	2003	1	0	132	-0,0002	1,0996	2,642E-04
		2004	0	0	144	0,0006	1,1996	4,160E-04
		2005	0	0	144	0,0015	1,1169	4,310E-04
		2006	0	0	144	-0,0007	1,0656	2,631E-04
		2007	0	0	144	-0,0004	0,7990	1,949E-04
		2008a	0	0	144	0,0019	0,7253	4,027E-04
		2008b	0	0	144	0,0010	0,5394	5,600E-04
		2009	0	0	144	-0,0001	0,3887	3,282E-04
		2010	0	0	144	-0,0001	0,5490	2,049E-04
<i>Tran Paulist PN</i>	<i>TRPL4</i>	2003	0	0	144	0,0008	0,9003	3,582E-04
		2004	0	0	144	0,0001	1,2398	5,376E-04
		2005	0	3	108	-0,0004	1,3152	6,492E-04
		2006	0	0	144	-0,0024	0,9452	7,104E-04
		2007	0	0	144	0,0005	0,9240	2,768E-04
		2008a	0	0	144	0,0034	0,6652	4,188E-04
		2008b	1	0	132	0,0000	0,2834	6,698E-04
		2009	1	0	132	0,0011	0,2690	2,393E-04
		2010	1	0	132	-0,0005	0,4489	1,800E-04
<i>Comgas PNA</i>	<i>CGAS5</i>	2003	0	0	144	0,0017	0,8049	5,017E-04
		2004	0	0	144	0,0010	0,9289	5,794E-04
		2005	0	0	144	-0,0004	1,0810	4,447E-04
		2006	0	0	144	0,0003	0,6648	4,174E-04
		2007	0	0	144	0,0007	0,6624	2,301E-04
		2008a	3	0	108	0,0027	0,5576	1,724E-04
		2008b	3	0	108	-0,0034	0,2227	3,671E-04
		2009	1	0	132	-0,0002	0,3381	1,870E-04
		2010	0	0	144	0,0003	0,2585	1,040E-04

Rodoviário**Tabela A. 7 – Resultados para o setor Rodoviário**

<i>Nome</i>	<i>Código</i>	<i>Período</i>	<i>Ex. Var</i>	<i>Ex. H₀</i>	<i>N</i>	<i>α final</i>	<i>B final</i>	<i>MSE final</i>
<i>Randon Part PN</i>	<i>RAPT4</i>	2003	-	-	-	-	-	-
		2004	0	0	144	0,0042	0,6181	4,218E-04
		2005	0	0	144	-0,0010	0,3891	2,353E-04
		2006	0	0	144	-0,0010	0,6559	3,217E-04
		2007	0	0	144	0,0035	0,5549	3,245E-04
		2008a	2	0	120	-0,0008	0,7249	4,771E-04
		2008b	1	0	132	-0,0060	0,6700	7,433E-04
		2009	0	4	96	0,0062	0,8284	7,359E-04
		2010	0	0	144	0,0001	0,6298	2,572E-04
<i>Ccr Rodovias ON</i>	<i>CCRO3</i>	2003	-	-	-	-	-	-
		2004	1	0	132	0,0016	0,3442	6,093E-04
		2005	0	0	144	-0,0003	0,3975	3,496E-04
		2006	0	0	144	0,0005	0,8930	6,803E-04
		2007	2	0	120	0,0001	0,8502	2,366E-04
		2008a	0	0	144	0,0018	0,9182	5,424E-04
		2008b	1	0	132	-0,0001	0,5061	6,783E-04
		2009	0	0	144	0,0009	0,4598	5,412E-04
		2010	0	0	144	0,0013	0,5789	3,104E-04

Siderurgia**Tabela A. 8 – Resultados para o setor Siderurgia**

<i>Nome</i>	<i>Código</i>	<i>Período</i>	<i>Ex. Var</i>	<i>Ex. H₀</i>	<i>N</i>	<i>α final</i>	<i>B final</i>	<i>MSE final</i>
<i>Cia Sid Nacional</i>	<i>CSNA3</i>	2003	2	3	84	0,0076	0,4612	3,970E-04
		2004	0	0	144	0,0010	1,0919	4,925E-04
		2005	0	0	144	0,0005	1,2452	2,952E-04
		2006	0	0	144	0,0022	1,3085	2,679E-04
		2007	0	0	144	0,0023	1,2109	2,004E-04
		2008a	0	0	144	0,0021	1,2135	3,655E-04
		2008b	1	0	132	-0,0005	1,3682	5,248E-04
		2009	1	0	132	0,0012	1,3090	2,592E-04
		2010	0	0	144	0,0010	1,0425	1,504E-04
<i>Ferbasa</i>	<i>FESA4</i>	2003	-	-	-	-	-	-
		2004	1	0	132	0,0003	0,3083	1,591E-04
		2005	0	0	144	-0,0025	0,3189	2,579E-04
		2006	0	0	144	-0,0008	0,1711	1,614E-04
		2007	3	0	108	0,0019	0,3810	3,476E-04
		2008a	1	0	132	0,0058	0,7227	7,260E-04
		2008b	0	0	144	-0,0052	0,5586	9,117E-04
		2009	1	0	132	0,0021	0,4247	4,494E-04
		2010	0	0	144	0,0001	0,8595	4,017E-04

APÊNDICE B – CÓDIGO FONTE DO PROGRAMA DESENVOLVIDO EM LINGUAGEM VISUAL BASIC PARA MICROSOFT EXCEL

```

'Autor: Tiago C. M. Marques
'Data: Novembro 2009 a Maio 2010
'Algoritmo para aplicação de metodologia fase I de monitoramento de perfil linear
Option Explicit
Option Base 1
Private Enum Modelo 'Identificador do modelo
    modeloCompleto = 0
    modeloIntermediario = 1
    modeloReduzido = 2
End Enum
'Níveis de significância definidos arbitrariamente
Private Const FalseAlarmProbTotal As Double = 0.05
Private Const FalseAlarmProbAmostra As Double = 0.065
Function CalcN(ByVal Ns As Variant)
'Soma o número de observações das amostras
'Para retornar o número total de observações
Dim i As Integer
For i = 1 To UBound(Ns)
    CalcN = CalcN + Ns(i)
Next
End Function
Function Calcular(ByVal m As Integer, ByVal k As Integer, ByRef Ns As Variant, _
ByRef Xred As Variant, ByRef Y As Variant, _
ByRef IteracoesVar As Long, ByRef RemovidosVar As Long, _
ByRef IteracoesF As Long, ByRef RemovidosF As Long, _
ByRef MatrixIteracoesF As Variant) As Variant
'Função principal, realiza todas as etapas da metodologia

'Entrada:
'Número de amostras m
'Número de regressores k
'Vetor Ns, de m linhas, contendo o número de observações de cada amostra
'Matriz Xred, contendo os valores dos regressores para todas as observações ordenadas
'pelo número da amostra
'e um vetor unitário (intercepto) como primeira coluna
'Matriz Y contendo os valores da variável explicativa para todas as observações ordenadas
'por amostra

'Saída por referência:
'IteracoesVar, contendo o número de iterações feitas durante aplicação do gráfico de controle
' para variância
'RemovidosVar, com o número de amostras removidas após detecção pelo gráfico de controle
'para variância
'IteracoesF, com o número de iterações feitas durante aplicação do teste F global
'RemovidosF, com o número de amostras removidas após detecção via teste F
'MatrixIteracoesF, contendo dados sobre coeficientes e estatísticas F para cada iteração do teste F global

'Resposta da função:
'Matriz contendo intercepto, primeiro coeficiente angular e seu estimador de variância
'para 3 momentos:
'Início
'Após remoção das amostras detectadas como fora de controle pelo gráfico de controle da variância
'Fim do algoritmo

```

'OBS

'Usa funções do Microsoft Excel para executar operações matriciais

Dim res(3)

Dim Betas(3) As Double

Dim coefs() As Variant

Dim i As Long

Dim aux As Long

Dim Fs() As Double

Dim Xfull As Variant

Dim Xinter As Variant

Dim FalseAlarmProbVar As Double

Dim FalseAlarmProbF As Double

Dim FalseAlarmProbSample As Double

Dim MatrixF As Variant

'Monta matrizes dos demais modelos a partir da matriz do modelo reduzido

Matrizes Xred, Xfull, Xinter, m, Ns, k

'Calcular probabilidades de alarme falso

FalseAlarmProbVar = 1 - ((1 - FalseAlarmProbTotal) ^ (1 / (k + 2))) ^ (1 / m) '0.001424

FalseAlarmProbF = (1 - (1 - FalseAlarmProbTotal) ^ ((k + 1) / (k + 2))) '0.033617

FalseAlarmProbSample = FalseAlarmProbAmostra

'Calculo matricial do estimador da variância da inclinação

With Excel.WorksheetFunction

Betas(3) = .MInverse(.MMult(.Transpose(Xred), Xred))(2, 2) * _

CalcularSSE(modeloReduzido, Xred, Y, 0, Ns, k, coefs) / (CalcN(Ns) - k - 1)

End With

Betas(1) = coefs(1, 1)

Betas(2) = coefs(2, 1)

res(1) = Betas

'Executa iterações para gráfico de controle da variância

IteracoesVar = 1

Do Until Variancia(m, Xred, Y, Ns, k, FalseAlarmProbVar, Xfull, Xinter, RemovidosVar) = 0

IteracoesVar = IteracoesVar + 1

Loop

'Calculo matricial do estimador da variância da inclinação

With Excel.WorksheetFunction

Betas(3) = .MInverse(.MMult(.Transpose(Xred), Xred))(2, 2) * _

CalcularSSE(modeloReduzido, Xred, Y, 0, Ns, k, coefs) / (CalcN(Ns) - k - 1)

End With

Betas(1) = coefs(1, 1)

Betas(2) = coefs(2, 1)

res(2) = Betas

'Reseta matriz dos dados de iterações de F

ReDim MatrixIteracoesF(1) As Variant

'Executa iterações do monitoramento de perfil

```

IteracoesF = 1
Do Until Calculos(Xfull, Ns, k, Xred, Y, Xinter, m, FalseAlarmProbF, RemovidosF, _
MatrixF, FalseAlarmProbSample) = 0 Or m = 1
    IteracoesF = IteracoesF + 1
    MatrixIteracoesF(UBound(MatrixIteracoesF)) = MatrixF
    ReDim Preserve MatrixIteracoesF(UBound(MatrixIteracoesF) + 1) As Variant
Loop
MatrixIteracoesF(UBound(MatrixIteracoesF)) = MatrixF

'Calculo matricial do estimador da variância da inclinação
With Excel.WorksheetFunction
    Betas(3) = .MInverse(.MMult(.Transpose(Xred), Xred))(2, 2) * _
    CalcularSSE(modeloReduzido, Xred, Y, 0, Ns, k, coefs) / (CalcN(Ns) - k - 1)
End With
Betas(1) = coefs(1, 1)
Betas(2) = coefs(2, 1)
res(3) = Betas

    Calcular = res
End Function
Private Function CalcularSSE(ByVal iModelo As Modelo, ByVal Xtotal As Variant, _
ByVal Yori As Variant, _
ByVal Amostra As Integer, ByVal Ns As Variant, ByVal k As Integer, ByRef Coef As Variant, _
Optional ByVal SoAmostra As Boolean = False) As Double
'Função que executa os cálculos matriciais de SSE

'Entrada:
'iModelo, indicando se trata-se do modelo reduzido, intermediário ou completo
'Xtotal, matriz com intercepto e regressores do modelo
'Yori, matriz com os valores da variável explicada
'Amostra, indicando a amostra que deseja-se analisar. Informar 0 caso deseje analisar o modelo todo.
'Vetor Ns com o número de observações de cada amostra
'k o número de regressores
'SoAmostra, VERDADEIRO caso deseje somente realizar o SSE da regressão da amostra.
'FALSO caso deseje realizar testes F referentes à fase I do monitoramento de perfil linear

'Saída
'Coef, vetor contendo os coeficientes resultantes da regressão realizada para cálculo do SSE

    Dim SSE As Double
    Dim ychap As Double
    Dim ini As Integer
    Dim fim As Integer
    Dim aux As Integer
    Dim i As Integer
    Dim X As Variant
    Dim m As Long
    Dim Y As Variant

    ini = 1
    fim = UBound(Xtotal, 1)

    If Amostra Then
        'Análises referentes a uma amostra em particular

        '---Estrutura matriz de variáveis explicativas e explicadas de acordo com o modelo tratado
        m = UBound(Ns)
        If Not SoAmostra Then

```



```

ReDim Y(Ns(Amostra) + Ns(m), 1)
Select Case iModelo
Case modeloCompleto
    ReDim X(Ns(Amostra) + Ns(m), 1 + k + (2 - 1) * (k + 1))
Case modeloIntermediario
    ReDim X(Ns(Amostra) + Ns(m), 1 + k + (2 - 1) * 1)
Case modeloReduzido
    ReDim X(Ns(Amostra) + Ns(m), 1 + k)
End Select
Else
ReDim Y(Ns(Amostra), 1)
Select Case iModelo
Case modeloCompleto
    ReDim X(Ns(Amostra), 1 + k + (2 - 1) * (k + 1))
Case modeloIntermediario
    ReDim X(Ns(Amostra), 1 + k + (2 - 1) * 1)
Case modeloReduzido
    ReDim X(Ns(Amostra), 1 + k)
End Select
End If
If Not SoAmostra Then
    ini = (UBound(Xtotal) - Ns(m) + 1)
    fim = UBound(Xtotal)
    For i = ini To fim
        For aux = 1 To k + 1
            X(i - ini + 1 + Ns(Amostra), aux) = Xtotal(i, aux)
            Select Case iModelo
            Case modeloCompleto
                X(i - ini + 1 + Ns(Amostra), aux + (k + 1)) = 0
            Case modeloIntermediario
                If aux = 1 Then X(i - ini + 1 + Ns(Amostra), aux + (k + 1)) = 0
            End Select
        Next
        Y(i - ini + 1 + Ns(Amostra), 1) = Yori(i, 1)
    Next
End If

```

'Determina primeira e última obsevações da amostra a ser tratada
 ini = IniM(Ns, Amostra)
 fim = FimM(Ns, Amostra)

```

For i = ini To fim
    For aux = 1 To k + 1
        X(i - ini + 1, aux) = Xtotal(i, aux)
        Select Case iModelo
        Case modeloCompleto
            X(i - ini + 1, aux + (k + 1)) = Xtotal(i, aux)
        Case modeloIntermediario
            If aux = 1 Then X(i - ini + 1, aux + (k + 1)) = Xtotal(i, aux)
        End Select
    Next
    Y(i - ini + 1, 1) = Yori(i, 1)
Next
'---

```

'Calculo matricial dos coeficientes e do SSE

With Excel.WorksheetFunction

```

    Coef = .MMult(.MMult(.MInverse(.MMult(.Transpose(X), X)), .Transpose(X)), Y)

```

```

        SSE = .MMult(.Transpose(Y), Y)(1) - .MMult(.Transpose(Coef), .MMult(.Transpose(X), Y))(1)
    End With

Else

    X = Xtotal

    'Calculo matricial dos coeficientes e do SSE
    With Excel.WorksheetFunction
        Coef = .MMult(.MMult(.MInverse(.MMult(.Transpose(X), X)), .Transpose(X)), Yori)
        SSE = .MMult(.Transpose(Yori), Yori)(1) - .MMult(.Transpose(Coef), _
        .MMult(.Transpose(X), Yori))(1)
    End With
End If

CalculaSSE = SSE

End Function

Private Function Calculos(ByRef Xfull As Variant, ByRef Ns As Variant, ByVal k As Long, _
ByRef Xred As Variant, ByRef Y As Variant, ByRef Xinter As Variant, ByRef m As Integer, _
ByVal FalseAlarmProb As Double, ByRef RemovidosF As Long, _
ByVal FalseAlarmProbSample As Double, _
ByRef MatrixF As Variant) As Long
'Função que executa testes F referentes ao monitoramento de perfil linear fase I
'Calculos(Xfull, Ns, k, Xred, Y, Xinter, m, FalseAlarmProbF, RemovidosF, _
MatrixF, FalseAlarmProbSample)
'Entrada:
'Xfull, matriz contendo intercepto e valores dos regressores para as observações para modelo completo
'Xred, a matriz para o modelo reduzido
'Xinter, a matriz para o modelo intermediário
'Vetor Ns, contendo número de observações para cada amostra
'k, o número de regressores
'Y, os valores da variável explicativa para todas as observações
'm, o número de amostras
'FalseAlarmProb, a probabilidade de alarme falso para o teste F global
'FalseAlarmProbSample, a probabilidade de alarme falso para o teste de identificação de
'amostras fora de controle

Dim coefs As Variant
Dim SSEfull As Double
Dim SSEred As Double
Dim SSEinter As Double
Dim GlobalFTest As Double
Dim HaFTest As Double
Dim HbFTest As Double
Dim FValue As Variant
Dim MSEfull As Double
Dim MSEinter As Double
Dim ychap As Double
Dim SSEfulltemp As Double
Dim MSEfulltemp As Double
Dim mtemp As Integer
Dim i As Long
Dim j As Integer
Dim GlobalFTestPValue As Double
Dim HaFTestPValue As Double
Dim HbFTestPValue As Double
Dim Removidos As Long
Dim B0 As Double

```

Dim B1 As Double

Dim Removs() As Long

Dim linrem As Long

ReDim FValue(m - 1, 2) As Double

'Calculo dos SSE dos modelos, do intercepto e do primeiro coef. angular

SSEfull = CalcularSSE(modeloCompleto, Xfull, Y, 0, Ns, k, coefs)

SSEred = CalcularSSE(modeloReduzido, Xred, Y, 0, Ns, k, coefs)

B0 = coefs(1, 1)

B1 = coefs(2, 1)

With Excel.ThisWorkbook.Worksheets("Resultados")

For i = 1 To k + 1

.Cells(2, 3 + i) = coefs(i, 1)

Next

End With

SSEinter = CalcularSSE(modeloIntermediario, Xinter, Y, 0, Ns, k, coefs)

MSEfull = SSEfull / (CalcN(Ns) - (k + (k + 1) * (m - 1)) - 1)

MSEinter = SSEinter / (CalcN(Ns) - (k + (m - 1)) - 1)

'Cálculo da estatística F global e dos testes de hipóteses de igualdade de inclinação e intercepto

GlobalFTest = (SSEred - SSEfull) / ((k + 1) * (m - 1) * MSEfull) 'Teste global

HaFTest = (SSEinter - SSEfull) / (k * (m - 1) * MSEfull) 'Ha

HbFTest = (SSEred - SSEinter) / ((m - 1) * MSEinter) 'Hb

'Valores P para as estatísticas F consideradas

'p-value do F-global

GlobalFTestPValue = Excel.WorksheetFunction.FDist(GlobalFTest, (k + 1) * (m - 1), _
CalcN(Ns) - m * (k + 1))

'p-value do Ha

HaFTestPValue = Excel.WorksheetFunction.FDist(HaFTest, (m - 1) * k, CalcN(Ns) - m * (k + 1))

'p-value do Hb

HbFTestPValue = Excel.WorksheetFunction.FDist(HbFTest, (m - 1), CalcN(Ns) - m - k)

'Cálculo das estatísticas F para todas as amostras, bem como seus valores P

For i = 1 To m - 1

mtemp = 2

SSEfulltemp = CalcularSSE(modeloCompleto, Xfull, Y, i, Ns, k, coefs)

MSEfulltemp = SSEfulltemp / (Ns(i) + Ns(m) - (k + (k + 1) * (mtemp - 1)) - 1)

FValue(i, 1) = (CalcularSSE(modeloReduzido, Xred, Y, i, Ns, k, coefs) - SSEfulltemp) / _
((k + 1) * (mtemp - 1) * MSEfulltemp)

FValue(i, 2) = Excel.WorksheetFunction.FDist(FValue(i, 1), _
(mtemp - 1) * (k + 1), Ns(i) + Ns(m) - mtemp * (k + 1))

Next

'Reinicia matriz com dados das estatísticas F de cada amostra

ReDim MatrixF(9) As Variant

MatrixF(1) = GlobalFTest

MatrixF(2) = HaFTest

MatrixF(3) = HbFTest

MatrixF(4) = B0

MatrixF(5) = B1

MatrixF(6) = SSEred / (CalcN(Ns) - k - 1)

```

'Indicação de aceitação ou não da hipótese de igualdade de inclinação das retas de regressão
If GlobalFTestPValue < FalseAlarmProb And HaFTestPValue < FalseAlarmProb Then
    MatrixF(7) = "R"
Else
    MatrixF(7) = "A"
End If

```

```

MatrixF(8) = m
MatrixF(9) = FValue

```

```

If GlobalFTestPValue < FalseAlarmProb Then
    'Caso rejeite-se a hipótese de igualdade das retas, deve-se identificar as amostras fora de controle
    For i = 1 To m - 1
        If FValue(i, 2) < FalseAlarmProbSample Then
            'Amostra identificada, deve ser removida
            RemoverLinhas Xred, IniM(Ns, i) - linrem, FimM(Ns, i) - linrem
            RemoverLinhas Y, IniM(Ns, i) - linrem, FimM(Ns, i) - linrem
            linrem = linrem + Ns(i)
            Removidos = Removidos + 1
            If Removidos = 1 Then
                ReDim Removs(1) As Long
                Removs(1) = i
            Else
                ReDim Preserve Removs(UBound(Removs) + 1) As Long
                Removs(UBound(Removs)) = i
            End If
        End If
    End If
Next
End If

```

```

'Caso tenham ocorrido remoções, deve-se reajustar as matrizes para o conjunto de amostras restante
If Removidos > 0 Then
    For i = 1 To Removidos
        RemoverLinhas1Dim Ns, Removs(i) - (i - 1), Removs(i) - (i - 1)
    Next
End If
m = m - Removidos
Calculos = Removidos

```

```

RemovidosF = RemovidosF + Removidos
If m > 1 Then
    Matrizes Xred, Xfull, Xinter, m, Ns, k
End If

```

```

End Function
Function FimM(ByVal Ns As Variant, ByVal j As Long)
    'Identifica a última amostra referente à observação j
    Dim i As Integer
    For i = 1 To j
        FimM = FimM + Ns(i)
    Next
End Function
Function IniM(ByVal Ns As Variant, ByVal j As Long)
    'Identifica a primeira amostra referente à observação j
    Dim i As Integer
    For i = 1 To j - 1
        IniM = IniM + Ns(i)
    Next
IniM = IniM + 1

```

```

End Function
Private Function Matrizes(ByVal Xred As Variant, ByRef Xfull As Variant, ByRef Xinter As Variant, _
ByVal m As Integer, ByVal Ns As Variant, ByVal k As Integer)
'Função que, a partir da matriz do modelo reduzido,
'monda as variáveis indicadoras (dummy) e estrutura as matrizes dos demais modelos
    Dim auxX As Long
    Dim Z() As Byte
    Dim ZXs() As Variant
    Dim ZX() As Double
    Dim i As Long
    Dim aux As Long

    ReDim Z(CalcN(Ns), m)
    ReDim ZXs(k)
    For auxX = 1 To k
        ReDim ZX(CalcN(Ns), m - 1)
        For i = 1 To CalcN(Ns)
            For aux = 1 To m - 1
                Z(i, aux) = 0
                If RetM(Ns, i) = aux Then Z(i, aux) = 1
                ZX(i, aux) = Xred(i, auxX + 1) * Z(i, aux)
            Next
        Next
        ZXs(auxX) = ZX
    Next

    ReDim Xfull(CalcN(Ns), 1 + k + (m - 1) + k * (m - 1))
    ReDim Xinter(CalcN(Ns), 1 + k + (m - 1))

    For i = 1 To CalcN(Ns)
        Xfull(i, 1) = 1
        Xinter(i, 1) = 1
        For aux = 2 To 1 + k
            Xfull(i, aux) = Xred(i, aux)
            Xinter(i, aux) = Xred(i, aux)
        Next
        For aux = 1 + k + 1 To 1 + k + (m - 1)
            Xfull(i, aux) = Z(i, aux - (1 + k))
            Xinter(i, aux) = Z(i, aux - (1 + k))
        Next
        For aux = 1 + k + (m - 1) + 1 To 1 + k + (m - 1) + k * (m - 1)
            Xfull(i, aux) = ZXs(Int((aux - (1 + k + (m - 1) + 1)) / _
(m - 1) + 1))(i, (aux - (1 + k + (m - 1) + 1)) Mod (m - 1) + 1)
        Next
    Next
End Function
Private Function RemoverLinhas(ByRef Matrix As Variant, ByVal LinIni As Long, _
ByVal LinFin As Long)
'Remocao de linhas para matrizes com duas dimensões
    Dim MatrixAux() As Variant
    Dim i As Long
    Dim iaux As Long
    Dim j As Long
    ReDim MatrixAux(UBound(Matrix, 1) - (LinFin - LinIni + 1), UBound(Matrix, 2))
    For i = 1 To UBound(Matrix, 1)
        If i < LinIni Or i > LinFin Then
            iaux = iaux + 1
            For j = 1 To UBound(Matrix, 2)
                MatrixAux(iaux, j) = Matrix(i, j)
            Next
        End If
    Next
End Function

```

```

        Next
    End If
Next
Matrix = MatrixAux
End Function
Private Function RemoverLinhas1Dim(ByRef Matrix As Variant, ByVal LinIni As Long, _
ByVal LinFin As Long)
'Remocao de linhas para matrizes com uma dimensão
    Dim MatrixAux() As Variant
    Dim i As Long
    Dim iaux As Long
    Dim j As Long
    ReDim MatrixAux(UBound(Matrix) - (LinFin - LinIni + 1))
    For i = 1 To UBound(Matrix)
        If i < LinIni Or i > LinFin Then
            iaux = iaux + 1
            MatrixAux(iaux) = Matrix(i)
        End If
    Next
    Matrix = MatrixAux
End Function

Function RetM(ByVal Ns As Variant, ByVal Pos As Long)
'Dada uma amostra, esta função retorna a amostra da qual ela faz parte
    Dim i As Integer
    Dim aux As Long
    For i = 1 To UBound(Ns)
        aux = aux + Ns(i)
        If Pos <= aux Then
            RetM = i
            Exit Function
        End If
    Next
End Function

Private Function Variancia(ByRef m As Integer, ByRef Xred As Variant, ByRef Y As Variant, _
ByRef Ns As Variant, ByVal k As Long, ByVal FalseAlarmProb As Double, ByRef Xfull As Variant, _
ByRef Xinter As Variant, ByRef RemovidosVar As Long) As Long
'Aplica gráfico de controle para variância e remove amostras fora de controle

'Entrada:
'Número de amostras m
'Matriz do modelo reduzido Xred
'Matriz do modelo completo Xfull
'Matriz do modelo intermediário Xinter
'Vetor com valores das observações para variável explicada, Y
'Vetor com número de observações das amostras Ns
'número de regressores k
'Probabilidade de alarme falso FalseAlarmProb

'Saída:
'Número de amostras fora de controle removidas

    Dim MSE As Double
    Dim MSEs() As Double
    Dim LCL As Double
    Dim UCL As Double
    Dim F As Double
    ReDim MSEs(m) As Double
    Dim i As Long

```

```

Dim coefs As Variant
Dim linrem As Long
Dim remov() As Long

ReDim remov(1) As Long

'Cálculo dos MSE das amostras
For i = 1 To m
    MSEs(i) = CalcularSSE(modeloReduzido, Xred, Y, i, Ns, k, coefs, True) / (Ns(i) - k - 1)
    MSE = MSE + MSEs(i)
Next

'Cálculo do MSE médio
MSE = MSE / m

For i = 1 To m

    'Cálculo do LSC
    F = Excel.WorksheetFunction.FInv(FalseAlarmProb / 2, Ns(i) - k - 1, CalcN(Ns) - Ns(i) - _
    (m - 1) * (k + 1))
    UCL = m * F * MSE / (m - 1 + F)

    'Cálculo do LIC
    F = Excel.WorksheetFunction.FInv(1 - FalseAlarmProb / 2, Ns(i) - k - 1, CalcN(Ns) - Ns(i) - _
    (m - 1) * (k + 1))
    LCL = m * F * MSE / (m - 1 + F)

    'Identificação e remoção de amostras fora de controle
    If MSEs(i) < LCL Or MSEs(i) > UCL Then
        RemoverLinhas Xred, IniM(Ns, i) - linrem, FimM(Ns, i) - linrem
        RemoverLinhas Y, IniM(Ns, i) - linrem, FimM(Ns, i) - linrem
        linrem = linrem + Ns(i)
        If Not (UBound(remov) = 1 And remov(1) = 0) Then
            ReDim Preserve remov(UBound(remov) + 1)
        End If
        remov(UBound(remov)) = i
        Variancia = Variancia + 1
    End If
Next
m = m - Variancia
For i = 1 To UBound(remov)
    If Not remov(i) = 0 Then RemoverLinhas1Dim Ns, remov(i) - i + 1, remov(i) - i + 1
Next
Matrizes Xred, Xfull, Xinter, m, Ns, k
RemovidosVar = RemovidosVar + Variancia
End Function

```

APÊNDICE C – CÓDIGO FONTE DO PROGRAMA DESENVOLVIDO EM LINGUAGEM MATLAB

Este programa gera arquivos gráficos usando arquivos do tipo Microsoft Excel como fontes de dados. Ademais, este programa emprega um mecanismo para diferenciação entre as amostras mostradas no gráficos.

```
%Autor: Tiago C. M. Marques
%Data: Fevereiro à Abril de 2010
%Código para leitura de dados a partir de arquivo Microsoft Excel
%E geração de arquivo tipo gráfico contendo um gráfico com amostras
%diferenciadas
echo off;
cd('C:\TFTiago\');
clear;
dados = xlsread('DadosBBDC3-2010.xls','1');

%conversão de datas do arquivo Excel para o MatLab
matlabDates = x2mdate(dados(:,1));

[idx, m, n] = unique(dados(:,3));
for i=1:length(idx)
    l(i)=cellstr(['Amostra ' num2str(idx(i))]);
end
symb='s^o<x+*dvp>'; %símbolos usados nos pontos do gráfico
cor='bgrcmgkbggrcm';
h = figure('visible','off');
set(h,'paperpositionmode','auto');
set(h,'Position',[1, 600, 1280, 600]);
numdiv = 12;
set(gca,'Position',[0.05 0.08 0.93 0.88]);
axis([734030 734260 -0.1 0.1]);
set(gca,'XTick',-734030:30:734270);
for k=1:numdiv
    h2 = gscatter(matlabDates((k-1)*12+1:k*12,1),...
        dados((k-1)*12+1:k*12,2)*100,...
        dados((k-1)*12+1:k*12,3),cor(k),symb(k),[],[],...
        'Data','Retorno');
    set(h2,'LineStyle','-');
    hold on;
end
leg = legend(l(1:length(l)));
set(leg,'Location','SouthEastOutside');
set(leg,'FontSize',7);
title('Variação diária de BBDC3 de 14/09/2009 a 16/04/2010');
cd('C:\TFTiago\');
labelsy = get(gca, 'YTickLabel');
for i=1:size(labelsy,1)
    labelsy_modif(i,:)=[labelsy(i,:) '%'];
end
set(gca,'Yticklabel',labelsy_modif);
datetick('x',19,'keeplimits');
print(h, '-dmeta','-r300', 'testea');%geracao do arquivo
```