



Universidade
Estadual de Londrina

CENTRO DE CIÊNCIAS EXATAS
DEPARTAMENTO DE MATEMÁTICA

Política Monetária Ótima no Regime de Metas para a Inflação

Geovani Nunes Grapiglia

LONDRINA - PR

2011

UNIVERSIDADE ESTADUAL DE LONDRINA
CENTRO DE CIÊNCIAS EXATAS
DEPARTAMENTO DE MATEMÁTICA

Política Monetária Ótima no Regime de Metas para a Inflação

Geovani Nunes Grapiglia

Dissertação de mestrado apresentada ao Departamento de Matemática da Universidade Estadual de Londrina, como requisito parcial para a obtenção do Título de MESTRE em Matemática Aplicada e Computacional.

Orientador: Prof. Dr. Naresh Kumar Sharma

Londrina - PR
Fevereiro - 2011

**Catálogo elaborado pela Divisão de Processos Técnicos da Biblioteca Central da
Universidade Estadual de Londrina.**

Dados Internacionais de Catalogação-na-Publicação (CIP)

G766p Grapiglia, Geovani Nunes.
Política monetária ótima no regime de metas para a inflação / Geovani
Nunes Grapiglia. – Londrina, 2011.
114 f. : il. + apêndice.

Orientador: Naresh Kumar Sharma.

Dissertação (Mestrado em Matemática Aplicada e Computacional) –
Universidade Estadual de Londrina, Centro de Ciências Exatas, Programa
de Pós-Graduação em Matemática Aplicada e Computacional, 2011.
Inclui bibliografia.

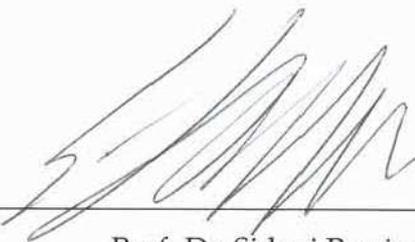
I. Matemática aplicada – Teses. 2. Política monetária.– Bancos Centrais
– Teses. 3. Matemática financeira – Inflação – Teses. I. Sharma, Naresh
Kumar. II. Universidade Estadual de Londrina. Centro de Ciências
Exatas. Programa de Pós-Graduação em Matemática Aplicada e Compu-
tacional. III. Título.

CDU 51-7:336.74

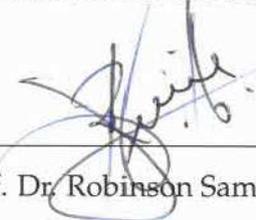
Política Monetária Ótima no Regime de Metas para a Inflação

Geovani Nunes Grapiglia

Dissertação de mestrado defendida em 28 de Fevereiro de 2011 e aprovada pela banca examinadora composta pelos seguintes professores doutores



Prof. Dr. Sidnei Pereira do Nascimento
Departamento de Economia
Universidade Estadual de Londrina



Prof. Dr. Robinson Samuel Viera Hoto
Departamento de Matemática
Universidade Estadual de Londrina



Prof. Dr. Naresh Kumar Sharma
Departamento de Matemática
Universidade Estadual de Londrina

*Dedico este trabalho ao meu pai, Alencar Antonio Grapiglia,
e à memória de minha amada mãe, Florisbela Nunes Grapiglia.*

Agradecimentos

Ao meu querido pai, Alencar Antonio Grapiglia, por todo amor e carinho.

Ao meu irmão e futuro biólogo, Eugenio O. N. Grapiglia, pela motivação.

A todos os meus familiares, especialmente ao tio Marcos Martini e à tia Islaine Aparecida Grapiglia, pelo apoio incondicional.

Aos amigos André Vicente, Daniela Grande e Poliane C. de Farias, por me estenderem a mão nos momentos de dificuldade.

Aos amigos Vinícius e Edilaine Peralta, João R. T. Garcia, Suellen R. Pardo, Camila Leão, Everton P. da Cruz, Cibele A. Ladeia, Rodrigo N. Monteiro, Jonatan M. da Silva, João A. W. Machado e Valdir Anhucchi, pelo prazer da companhia ao longo desses dois anos.

Ao Prof. Rogério Luís Rizzi e à Profa. Leila Deixum Franzini, por me incentivarem a prosseguir com meus estudos.

Ao Prof. Naresh Kumar Sharma, pela orientação neste trabalho e pelos conselhos para a vida.

À Profa. S. Shelly Sharma, por assistir meus seminários e contribuir com valiosas sugestões.

Aos professores do PGMAC, por ampliarem minha formação acadêmica.

À CAPES, pelo apoio financeiro.

Resumo

Este trabalho estima as preferências de política monetária do Banco Central do Brasil durante a Presidência de Henrique Meirelles. A metodologia empregada baseia-se na modelagem da política monetária como um problema do regulador linear-quadrático estocástico. Nessa abordagem, as preferências da autoridade monetária podem ser associadas aos pesos que definem a sua função perda. Assim, a estratégia de estimação consistiu em calibrar a função perda do Banco Central, escolhendo os pesos de modo a minimizar o desvio quadrático médio entre a política monetária real e a política monetária ótima simulada. Contudo, em vez de realizar essa escolha pelo procedimento usual de busca em malha, este trabalho propõe o uso de um algoritmo genético com codificação real. Os resultados sugerem que o Banco Central do Brasil tem adotado um regime de metas para a inflação flexível, no qual o maior peso é dado para a suavização da taxa de juros, enquanto um peso mínimo é atribuído ao hiato do produto.

Palavras Chave: Preferências Monetárias do Banco Central, Política Monetária Ótima, Regime de Metas para a Inflação, Problema do Regulador Linear-Quadrático.

Abstract

This work estimates the monetary policy preferences of the Central Bank of Brazil during Henrique Meirelles's mandate as chairman. The methodology is based on the modelling of monetary policy as a stochastic linear-quadratic regulator problem. In this framework, preferences of the monetary authority can be associated with the weights that define its loss function. Thus, the estimation strategy was to calibrate the loss function of the Central Bank by choosing the weights that minimize the mean-square deviation between the real monetary policy and the simulated optimal monetary policy. However, instead of performing this choice by the usual grid-search procedure, this work proposes the use of a genetic algorithm with real coding. The results suggest that the Central Bank of Brazil has adopted a flexible inflation targeting regime, in which greater weight is given to interest rate smoothing, while a minimal weight is assigned to the output gap.

Keywords: Monetary Preferences of the Central Bank, Optimal Monetary Policy, Inflation Targeting Regime, Linear-Quadratic Regulator Problem.

Sumário

1	Introdução	1
1.1	Organização da Dissertação	3
2	Modelo de Regressão Linear	5
2.1	Modelo Clássico de Regressão Linear	5
2.1.1	Hipóteses Clássicas	6
2.1.2	Estimador de Mínimos Quadrados Ordinários	7
2.1.3	Propriedades Estatísticas	9
2.1.4	Normalidade dos Erros e Teste de Hipóteses	14
2.2	Violação das Hipóteses Clássicas	19
2.2.1	Multicolinearidade	20
2.2.2	Autocorrelação e Heterocedasticidade	21
2.2.3	Não-normalidade dos Erros	25
2.3	Tópicos Especiais de Econometria	26
2.3.1	Regressão sobre Variáveis Binárias	26
2.3.2	Seleção de Regressores	27

2.3.3	Séries Temporais Não-Estacionárias	29
3	Programação Dinâmica em Tempo Discreto	34
3.1	Horizonte Finito	34
3.2	Horizonte Infinito	37
3.3	Problema do Regulador Linear-Quadrático	39
3.3.1	Caso Determinístico	40
3.3.2	Caso Determinístico Modificado	42
3.3.3	Caso Estocástico	44
3.3.4	Caso Estocástico Modificado	47
3.4	Equação de Riccati	49
4	Algoritmos Genéticos	52
4.1	Fundamentos Biológicos	52
4.1.1	Genética	52
4.1.2	Seleção Natural	53
4.2	Algoritmos Genéticos com Codificação Binária	53
4.2.1	Codificação	55
4.2.2	Seleção	56
4.2.3	Recombinação	58
4.2.4	Mutação	59
4.2.5	Critérios de Parada	60
4.3	Fundamento Teórico	60
4.3.1	A Teoria dos Esquemas	61

4.4	Algoritmos Genéticos com Codificação Real	64
4.4.1	Cruzamento Aritmético	65
4.4.2	Mutação Não-Uniforme	65
4.5	Vantagens e Limitações dos Algoritmos Genéticos	66
5	Problema de Condução da Política Monetária	67
5.1	Regime de Metas para a Inflação	67
5.1.1	Mecanismo de Trasmissão da Política Monetária	68
5.2	Modelo Macroeconômico Linear	72
5.2.1	Descrição das Variáveis	73
5.2.2	Análise das Séries Temporais	76
5.2.3	Estimação dos Coeficientes	77
5.3	Modelagem Matemática da Política Monetária	81
5.3.1	Solução via Programação Dinâmica	82
5.3.2	Regras Ótimas para a Política Brasileira	84
6	Preferências do Banco Central do Brasil	88
6.1	Revisão da Literatura	88
6.2	Estimação das Preferências do Banco Central do Brasil	91
6.2.1	Método de Solução Usual - Busca em Malha	92
6.2.2	Método de Solução Alternativo - Algoritmo Genético	92
6.2.3	Resultados	93
6.3	Política Adotada <i>versus</i> Política Ótima	95
6.4	Projeções Ótimas	96

6.4.1	Avaliação Preliminar	98
7	Conclusão	99
7.1	Sugestões para Trabalhos Futuros	100
7.1.1	Modelagem	100
7.1.2	Métodos	100
7.1.3	Questões	101
A	Distribuições de Probabilidade	102
1.1	Distribuição Normal Multivariada	102
1.2	Distribuições Correlatas	103
B	Filtro Hodrick-Prescott	104
2.0.1	Procedimento para Extração da Tendência	104
2.0.2	Estimação do PIB Potencial	106
	Referências Bibliográficas	107

Lista de Figuras

5.1	Síntese dos principais canais de transmissão da política monetária. . . .	71
5.2	Projeções para diferentes preferências de política monetária.	87
6.1	Taxa Selic adotada e taxa Selic ótima	96
6.2	Projeções Ótimas para a Taxa Selic, a Inflação e o Hiato do Produto. . . .	97
B.1	Produto Interno Bruto potencial obtido pelo Filtro HP	106

Lista de Tabelas

5.1	Teste de Raíz Unitária	76
5.2	Curva de Phillips	78
5.3	Curva IS	79
5.4	Curva da Expectativa de Inflação	80
5.5	Curva da Taxa de Câmbio	81
5.6	Preferências Simuladas	85
6.1	Pesos Calibrados pelo AG	94
6.2	Pesos Calibrados pelo AG ($\delta = 0,9991$)	94
6.3	Desempenho das projeções para a taxa Selic	98

Lista de Símbolos e Notações

- $M(m \times n)$: Conjunto das matrizes reais com m linhas e n colunas
- M' : Transposta da matriz M
- $tr(M)$: Traço da matriz M
- I_n : Matriz Identidade de ordem n
- $\tilde{\beta}$: Estimador linear qualquer
- $\hat{\beta}$: Estimador de Mínimos Quadrados Ordinários
- \hat{u} : Resíduo de Mínimos Quadrados Ordinários
- \hat{y} : Previsor de Mínimos Quadrados Ordinários
- R^2 : Coeficiente de Determinação
- $dim(x)$: Número de coordenadas do vetor x
- E_t : Esperança matemática condicionada à informação disponível no instante t
- p_c : Probabilidade de Cruzamento
- p_m : Probabilidade de Mutação
- T_{max} : Número máximo de gerações
- d : Parâmetro do Operador de Mutação Não-Uniforme
- $\delta(H)$: Comprimento do esquema H
- $o(H)$: Ordem do esquema H
- $m(H, t)$: Número de instâncias de H presentes na t -ésima geração
- $\hat{f}(H, t)$: Aptidão média das instâncias de H presentes na t -ésima geração
- $\bar{f}(t)$: Aptidão média da população na t -ésima geração.
- π_t : Taxa de inflação no período t
- π_t^e : Expectativa de inflação no período t

h_t : Hiato do produto no período t

c_t : Taxa de câmbio real no período t

i_t : Taxa nominal de juros no período t

$\Delta c_t = c_t - c_{t-1}$: Diferencial da taxa de câmbio real no período t

$\Delta i_t = i_t - i_{t-1}$: Diferencial da taxa nominal de juros no período t

$\varepsilon_{\pi,t}, \varepsilon_{h,t}, \varepsilon_{\pi^e,t}, \varepsilon_{c,t}$: Choques aleatórios ocorridos no período t

L_t : Perda do Banco Central no período t

λ_{π} : Peso atribuído ao desvio da inflação em relação à sua meta

λ_h : Peso atribuído ao hiato do produto

λ_i : Peso atribuído ao diferencial da taxa de juros

Lista de Abreviaturas e Siglas

ADF – “Augmented Dickey-Fuller”

AG – Algoritmo Genético

BLUE – “Best Linear Unbiased Estimator”

Copom – Comitê de Política Monetária

DF – “Dickey-Fuller”

DNA – “Deoxyribonucleic Acid”

DQM – Desvio Quadrático Médio

GMM – “Generalized Method of Moments”

HP – “Hodrick-Prescott”

IC – “Information Criteria”

IPCA – Índice Nacional de Preços ao Consumidor Amplo

MAE – “Mean Absolute Error”

MQO – Mínimos Quadrados Ordinários

MRE – “Mean Relative Error”

Selic – Sistema Especial de Liquidação e Custódia

SIC – “Schwarz Information Criteria”

SQE – Soma dos Quadrados Explicados

SQR – Soma dos Quadrados dos Resíduos

SQT – Soma dos Quadrados Totais

SOOS – “Simulated Out-of-Sample”

PMSE – “Prediction Mean Squared Error”

RMSE – “Root Mean Square Error”

Introdução

Em junho de 1999, o Brasil passou a adotar o regime de metas para a inflação como diretriz da sua política monetária. Nesse regime, o Banco Central anuncia uma meta numérica de médio prazo para a inflação e se compromete a atuar de modo que essa meta seja alcançada. Para isso, seu principal instrumento é a taxa Selic, cujo valor é definido em reuniões periódicas do Comitê de Política Monetária (Copom) [8].

Por conta da defasagem que existe entre as alterações no valor da taxa Selic e os seus efeitos sobre a economia, as decisões do Copom possuem um caráter antecipativo, tendo por base previsões quanto à evolução das variáveis macroeconômicas. De uma forma geral, a previsão persistente de uma inflação acima da meta indica a necessidade de elevação da taxa Selic, para que a pressão inflacionária seja contida. Por outro lado, a previsão de uma inflação abaixo da meta abre a possibilidade de redução da taxa Selic, a fim de evitar restrições desnecessárias sobre a atividade econômica [9, 13].

Embora as decisões de política monetária, no Brasil e no mundo, resultem de um processo complexo, envolvendo um amplo conjunto de informações e uma diversidade de avaliações subjetivas, percebeu-se que os movimentos do instrumento de política monetária poderiam ser descritos por equações lineares simples, as chamadas *regras de taxa de juros*. Partindo do influente trabalho de Taylor [69], numerosas especificações para essas regras tem sido estimadas com o intuito de modelar a reação da autoridade monetária frente ao desenvolvimento das variáveis macroeconômicas.

Seu uso inclui a caracterização do regime de política monetária [22, 50], bem como a avaliação de deslocamentos na postura do Banco Central [41].

Paralelamente a esses trabalhos, desenvolveu-se uma extensa literatura que analisa a política monetária sob a forma de um problema de otimização [3, 6, 51, 61, 67]. Nessa abordagem, supõe-se que o objetivo do Banco Central é minimizar uma função perda quadrática, tendo como restrição os vínculos que existem entre as variáveis macroeconômicas, os quais são representados por um modelo linear. Assim, a condução da política monetária é posta como um problema do regulador linear-quadrático estocástico, onde as preferências do Banco Central correspondem aos pesos que ele atribui a cada variável alvo na sua função perda. Como a solução desse tipo de problema resulta em uma regra linear, essa metodologia permite que as próprias regras de taxa de juros sejam encaradas como soluções de problemas de otimização específicos.

Sob esse enfoque, conforme apontado por Svensson [67], os parâmetros das regras de taxa de juros são determinados pelos coeficientes do modelo macroeconômico que restringe a otimização, junto com os pesos que definem a função perda do Banco Central. Como consequência, a estimação dessas regras não revela nada diretamente sobre as preferências que orientam as decisões de política monetária. Além disso, a verificação de quebra estrutural em seus parâmetros não pode ser atribuída unicamente à mudanças na postura da autoridade monetária, visto que ela também pode ocorrer por conta de alterações nos coeficientes do modelo macroeconômico que fundamenta a análise.

A partir dessas observações, a identificação das preferências do Banco Central surge como um ponto chave para uma compreensão mais profunda da política monetária. De acordo com Dennis [27], o conhecimento dessas preferências oferece as seguintes vantagens:

- (a) colabora para o entendimento das ações de política monetária;
- (b) permite avaliar se a política praticada é consistente com a política ótima associada a um determinado modelo; e
- (c) oferece um meio de comparar os objetivos da política monetária implementada em gestões e períodos diferentes.

Nesse sentido, existe um grande número de trabalhos que se dedicam à estimação dos pesos que a autoridade monetária atribui às variáveis alvo na sua função perda (por exemplo, [2, 21, 26, 29, 45]). Contudo, a maioria dessas pesquisas tem focado sobre as preferências do Federal Reserve¹ e dos bancos centrais de países europeus, especialmente aqueles que compõe a Área do Euro. No que se refere à política monetária brasileira, apenas Aragón e Portugal [6] apresentaram estimativas para as preferências do Banco Central do Brasil. Seguindo [19, 23, 64], esses autores empregaram uma estratégia de calibragem da função perda do Banco Central, na qual os pesos são escolhidos por um método de busca em malha, de modo que a solução do problema de otimização correspondente resulte no melhor ajuste entre a política monetária real e a política monetária ótima simulada.

Motivado pela escassez de estudos desse gênero sobre a política monetária brasileira, a presente dissertação tem como objetivo obter estimativas para as preferências do Banco Central do Brasil durante a presidência de Henrique Meirelles. Para isso, também utiliza-se a estratégia de calibragem da função perda, porém, como alternativa ao método de busca em malha, o presente trabalho propõe o uso de um algoritmo genético com codificação real.

1.1 Organização da Dissertação

O restante do texto está organizado da seguinte maneira. O Capítulo 2 apresenta noções básicas sobre o Modelo de Regressão Linear e o Método de Mínimos Quadrados Ordinários, com o propósito de fundamentar os procedimentos utilizados na estimação do modelo macroeconômico. No Capítulo 3 é abordado o Método de Programação Dinâmica, enfatizando o seu uso na resolução do problema do regulador linear-quadrático que descreve o processo decisório da política monetária. Em seguida, uma breve introdução sobre os algoritmos genéticos é feita no Capítulo 4, tendo em vista descrever os elementos do algoritmo utilizado para identificar as preferências do Banco Central. Finalmente, no Capítulo 5 é estimado um modelo macroeconômico linear, a

¹Banco Central dos Estados Unidos da América, referido pela sigla FED.

partir do qual a condução da política monetária é colocada como um problema do regulador linear-quadrático estocástico. Sob esse arcabouço, o Capítulo 6 apresenta estimativas para as preferências do Banco Central do Brasil, obtidas através de um algoritmo genético com codificação real. No Capítulo 7 são delineadas as conclusões e sugestões para trabalhos futuros. Complementando o trabalho, o Apêndice A enuncia algumas definições e resultados sobre as distribuições normal, qui-quadrado, F e t de Student, servindo como base para as demonstrações apresentadas no Capítulo 2. Por sua vez, o Apêndice B descreve o filtro Hodrick-Prescott aplicado para a estimação do hiato do produto. Ao final, apresenta-se a lista de referências da bibliografia que orientou esta dissertação.

Modelo de Regressão Linear

Baseado principalmente em Davidson [25], Gujarati [32] e Maddala [48], este capítulo expõe os fundamentos de Econometria necessários para a compreensão da modelagem matemática que é realizada no Capítulo 5.

2.1 Modelo Clássico de Regressão Linear

Considere as variáveis reais y, x_1, \dots, x_k , sobre as quais se possui uma amostra de n observações. Suponha que se deseje explicar o comportamento de y à partir das variáveis x_1, \dots, x_k . Formalmente, esse problema corresponde à busca por uma regra f , tal que:

$$y_t = f(x_{1t}, \dots, x_{kt}, u_t), \quad t = 1, \dots, n$$

onde

y_t é a t -ésima observação sobre a variável dependente;

x_{jt} é a t -ésima observação sobre a j -ésima variável explicativa (ou regressor); e

u_t é a t -ésima observação sobre um erro aleatório.

O modelo de regressão linear estabelece a relação entre as variáveis y e x_j , partindo do seguinte pressuposto:

$$y_t = \sum_{j=1}^k \beta_j x_{jt} + u_t \quad t = 1, \dots, n. \quad (2.1)$$

Nesse caso, o coeficiente β_j quantifica a influência que o regressor x_j exerce sobre y , enquanto os erros u_t representam o efeito aleatório de todas as variáveis que não são consideradas de forma explícita no modelo.

Em aplicações práticas, é comum ainda considerar um termo constante em (2.1), denominado *intercepto*. Por convenção, isso é feito assumindo-se $x_{1t} = 1$, para todo t , o que permite identificar o intercepto com o coeficiente β_1 :

$$y_t = \beta_1 + \sum_{j=2}^k \beta_j x_{jt} + u_t \quad t = 1, \dots, n.$$

2.1.1 Hipóteses Clássicas

A fim de explicitar as hipóteses clássicas que caracterizam o modelo de regressão linear, é conveniente resumir as equações (2.1) sob a forma de uma única equação matricial:

$$y = X\beta + u, \quad (2.2)$$

onde

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} \quad \text{e} \quad u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$

Com base nessa notação, as hipóteses clássicas do modelo de regressão linear podem ser enunciadas da seguinte maneira:

(H.1) A matriz de observações $X \in M(n \times k)$ tem posto completo: $\text{posto}(X) = k$;

(H.2) O vetor de erros aleatórios u possui esperança nula: $E(u) = 0$; e

(H.3) A matriz de variância-covariância de u é da forma $\text{var}(u) = \sigma^2 I_n$.

2.1.2 Estimador de Mínimos Quadrados Ordinários

Dados $y \in \mathbb{R}^n$ e $X \in M(n \times k)$, considere a função $S : \mathbb{R}^k \rightarrow \mathbb{R}$, dada por

$$S(b) = (y - Xb)'(y - Xb) = y'y - y'Xb - b'X'y + b'X'Xb.$$

Definição 2.1. *Seja a função $f : A \subset M(n \times k) \rightarrow \mathbb{R}^k$ tal que, para cada matriz de observações $X \in A$, o vetor $\hat{\beta} = f(X)$ é o ponto de mínimo global da função S . Nesse caso, diz-se que f é o estimador de Mínimos Quadrados Ordinários (MQO) para o vetor de coeficientes β em (2.2).*

Teorema 2.2 (Estimador de MQO). *Sob a hipótese (H.1), o estimador de MQO de β é dado pela regra*

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

Demonstração: Como $\text{posto}(X) = k$, segue-se que $X'X \in M(k \times k)$ é simétrica definida positiva¹. Logo, $X'X$ é invertível, donde conclui-se que o vetor $\hat{\beta}$ está bem definido.

Diferenciando S com respeito à b , e igualando o resultado a zero, o fato de $X'X$ ser invertível permite obter:

$$\begin{aligned} \frac{\partial S}{\partial b} &= 0 \\ \Leftrightarrow -2X'y + 2X'Xb &= 0 \\ \Leftrightarrow X'Xb &= X'y \\ \Leftrightarrow b &= (X'X)^{-1}X'y \\ \Leftrightarrow b &= \hat{\beta}, \end{aligned}$$

ou seja, $\hat{\beta}$ é o único ponto crítico de S .

Por outro lado, a matriz Hessiana de S em $\hat{\beta}$ é dada por:

$$H(\hat{\beta}) = 2X'X. \tag{2.3}$$

Como $X'X$ é simétrica definida positiva, tem-se que $\hat{\beta}$ é o mínimo local da função convexa S , sendo portanto um mínimo global. ■

¹Para maiores detalhes, ver na Seção 17 em [43] o tópico sobre *Matrizes de Gram*.

Corolário 2.3. Definindo o preditor $\hat{y} = X\hat{\beta}$ e resíduo $\hat{u} = y - X\hat{\beta}$ de Mínimos Quadrados Ordinários, verifica-se a seguinte igualdade:

$$y'y = \hat{y}'\hat{y} + \hat{u}'\hat{u}.$$

Demonstração: Pelo Teorema 2.2,

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1}X'y \\ \implies X'X\hat{\beta} &= X'y \\ \implies X'X\hat{\beta} &= X'(X\hat{\beta} + \hat{u}) \\ \implies X'X\hat{\beta} &= X'X\hat{\beta} + X'\hat{u} \\ \implies X'\hat{u} &= 0. \end{aligned}$$

Logo,

$$\begin{aligned} y'y &= (X\hat{\beta} + \hat{u})'(X\hat{\beta} + \hat{u}) \\ &= (\hat{\beta}'X' + \hat{u}')(\hat{u} + X\hat{\beta}) \\ &= \hat{\beta}'X'X\hat{\beta} + \hat{\beta}'X'\hat{u} + \hat{u}'X\hat{\beta} + \hat{u}'\hat{u} \\ &= \hat{y}'\hat{y} + \hat{u}'\hat{u} \end{aligned}$$

■

Ao longo deste trabalho, os escalares $y'y$, $\hat{y}'\hat{y}$ e $\hat{u}'\hat{u}$ serão referidos como Soma dos Quadrados Totais (SQT), Soma dos Quadrados Explicados (SQE) e Soma dos Quadrados dos Resíduos, respectivamente.

Definição 2.4. Com respeito ao modelo de regressão linear (2.2), o coeficiente de determinação R^2 é definido pela razão:

$$R^2 = \frac{SQE}{SQT}.$$

Observação 2.5. Tendo em vista o Corolário 2.3, R^2 pode ser posto na forma:

$$R^2 = \frac{\hat{y}'\hat{y}}{y'y} = \frac{y'y - \hat{u}'\hat{u}}{y'y} = 1 - \frac{\hat{u}'\hat{u}}{y'y} = 1 - \frac{SQR}{SQT}.$$

Note que, $0 \leq R^2 \leq 1$, sendo $R^2 = 1$ quando a regressão se ajusta perfeitamente aos dados, e $R^2 = 0$ quando a regressão não explica nada da variação de y . Portanto, o valor R^2 pode ser usado como uma medida do grau de ajuste do modelo em relação aos dados.

2.1.3 Propriedades Estatísticas

Teorema 2.6. *Sob a hipótese (H.2), o estimador $\hat{\beta}$ é não-tendencioso, isto é, $E(\hat{\beta}) = \beta$.*

Demonstração: Considerando $y = X\beta + u$ na expressão de $\hat{\beta}$, tem-se

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1}X'(X\beta + u) = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u \\ &= \beta + (X'X)^{-1}X'u.\end{aligned}\tag{2.4}$$

Então, aplicando o operador esperança sobre $\hat{\beta}$, o resultado enunciado segue pela hipótese (H.2):

$$E(\hat{\beta}) = \beta + E((X'X)^{-1}X'u) = \beta + (X'X)^{-1}X'E(u) = \beta,\tag{2.5}$$

■

Corolário 2.7. *Sob as hipóteses clássicas do modelo de regressão linear, a matriz de variância-covariância de $\hat{\beta}$ é dada por:*

$$\text{var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$$

Demonstração: Combinando o Teorema 2.6 e a equação (2.4), tem-se

$$\hat{\beta} - E(\hat{\beta}) = \hat{\beta} - \beta = (X'X)^{-1}X'u.\tag{2.6}$$

Por outro lado, note que

$$E(uu') = E[(u - 0)(u - 0)'] = E[(u - E(u))(u - E(u))'] = \text{var}(u) = \sigma^2 I_n,\tag{2.7}$$

onde a segunda e a última dessas igualdades decorrem das hipóteses (H.2) e (H.3), respectivamente. Como consequência das equações (2.6) e (2.7), obtém-se:

$$\begin{aligned}\text{var}(\hat{\beta}) &= E[(\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))'] \\ &= E\left[\left((X'X)^{-1}X'u\right)\left((X'X)^{-1}X'u\right)'\right] \\ &= E\left[(X'X)^{-1}X'uu'X(X'X)^{-1}\right] \\ &= (X'X)^{-1}X'E(uu')X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}.\end{aligned}$$

■

Na prática, a variância σ^2 também é desconhecida e portanto precisa ser estimada. A fim de derivar um estimador para esse termo, é útil considerar o seguinte resultado.

Lema 2.8. Se $M = I_n - X(X'X)^{-1}X'$ então as seguintes condições são satisfeitas:

- (i) M é uma matriz simétrica e idempotente;
- (ii) $(I_n - M)M = M(I_n - M) = 0$
- (iii) $\text{posto}(M) = \text{tr}(M) = n - k$; e
- (iv) $\hat{u} = Mu$.

Demonstração: A partir da definição de M decorre que

$$\begin{aligned} M' &= (I_n - X(X'X)^{-1}X')' \\ &= I_n' - (X')'(X'X)^{-1}X' \\ &= I_n - X(X'X)^{-1}X' = M, \quad \text{e} \end{aligned}$$

$$\begin{aligned} M^2 &= (I_n - X(X'X)^{-1}X')(I_n - X(X'X)^{-1}X') \\ &= I_n - X(X'X)^{-1}X' - X(X'X)^{-1}X' + X(X'X)^{-1}X'X(X'X)^{-1}X' \\ &= I_n - X(X'X)^{-1}X' - X(X'X)^{-1}X' + X(X'X)^{-1}X' \\ &= I_n - X(X'X)^{-1}X' = M, \end{aligned}$$

provando assim o item (i).

Por sua vez, (i) \Rightarrow (ii):

$$(I_n - M)M = M(I_n - M) = M - M^2 = M - M = 0.$$

Além disso, assumindo o item (i), o Teorema Espectral [43] e as propriedades do traço permitem mostrar que:

$$\text{posto}(M) = \text{tr}(M).$$

Portanto, o item (iii) é obtido pelo cálculo direto do traço de M :

$$\begin{aligned}
 \text{tr}(M) &= \text{tr}(I_n - X(X'X)^{-1}X') \\
 &= \text{tr}(I_n) - \text{tr}(X(X'X)^{-1}X') \\
 &= \text{tr}(I_n) - \text{tr}(X'X(X'X)^{-1}) \\
 &= \text{tr}(I_n) - \text{tr}(I_k) \\
 &= n - k,
 \end{aligned}$$

onde a segunda e a terceira igualdades acima resultam, respectivamente, da linearidade do traço, e da sua comutatividade com respeito ao produto de matrizes.

Finalmente, com base nas igualdades

$$\begin{aligned}
 \hat{u} &= y - X\hat{\beta} \\
 &= y - X \left[(X'X)^{-1}X'y \right] \\
 &= (I_n - X(X'X)^{-1}X')y \\
 &= My, \quad \text{e}
 \end{aligned} \tag{2.8}$$

$$\begin{aligned}
 MX &= (I_n - X(X'X)^{-1}X')X \\
 &= X - X(X'X)^{-1}X'X \\
 &= X - X = 0,
 \end{aligned} \tag{2.9}$$

o item (iv) segue de imediato:

$$\hat{u} = My = M(X\beta + u) = MX\beta + Mu = Mu.$$

■

Teorema 2.9. *Sob as hipóteses clássicas do modelo de regressão linear,*

$$\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{n - k}$$

é um estimador não-tendencioso de σ^2 .

Demonstração: Pelos itens (iv) e (i) do Lema 2.8,

$$\hat{u}'\hat{u} = u'M'Mu = u'M^2u = u'Mu. \tag{2.10}$$

Logo, considerando a propriedade $E(\text{tr}(A)) = \text{tr}(E(A))$, a equação (2.7) e o item (iii) do Lema 2.8, decorre que

$$\begin{aligned} E(\hat{u}'\hat{u}) &= E[\text{tr}(\hat{u}'\hat{u})] = E[\text{tr}(u'Mu)] \\ &= E[\text{tr}(Muu')] = \text{tr}(E(Muu')) \\ &= \text{tr}(ME(uu')) = \text{tr}(\sigma^2 M) \\ &= \sigma^2 \text{tr}(M) = \sigma^2(n-k), \end{aligned}$$

donde segue

$$E\left(\frac{\hat{u}'\hat{u}}{n-k}\right) = \frac{E(\hat{u}'\hat{u})}{n-k} = \sigma^2.$$

■

Concluindo esta subsecção, o Teorema de Gauss-Markov é demonstrado. Para tanto, é conveniente considerar a seguinte definição.

Definição 2.10. Com respeito ao modelo (2.2), diz-se que um estimador $\tilde{\beta}$ é linear quando ele é dado por $\tilde{\beta} = Cy$, onde $C \in M(k \times n)$.

Exemplo 2.11. O estimador de MQO $\hat{\beta}$ é um estimador linear, para o qual $C = (X'X)^{-1}X'$.

Teorema 2.12 (Gauss-Markov). Suponha que o modelo de regressão linear (2.2) satisfaz as hipóteses clássicas (H.1)-(H.3). Então, para qualquer estimador linear não-tendencioso $\tilde{\beta}$, a matriz

$$\text{var}(\tilde{\beta}) - \text{var}(\hat{\beta})$$

é simétrica semi-definida positiva.

Demonstração: Sendo $\tilde{\beta}$ um estimador linear qualquer, por definição

$$\tilde{\beta} = Cy.$$

Assim, aplicando o operador esperança sobre $\tilde{\beta}$, pela hipótese (H.2) verifica-se:

$$\begin{aligned} E(\tilde{\beta}) &= CE(y) = CE(X\beta + u) \\ &= CX\beta + CE(u) = CX\beta \end{aligned} \tag{2.11}$$

Como, além disso, $\tilde{\beta}$ é não-tendencioso, a equação (2.11) implica

$$CX = I_k. \quad (2.12)$$

Escrevendo $C = (X'X)^{-1}X' + D$, a equação acima, por sua vez, conduz ao seguinte resultado:

$$\begin{aligned} CX &= (X'X)^{-1}X'X + DX = I_k + DX \\ &\implies DX = 0, \end{aligned}$$

a partir do qual

$$\begin{aligned} CC' &= \left[(X'X)^{-1}X' + D \right] \left[X(X'X)^{-1} + D' \right] \\ &= (X'X)^{-1}X'X(X'X)^{-1} + DD' \\ &= (X'X)^{-1} + DD'. \end{aligned} \quad (2.13)$$

Com base em (2.7) e (2.13), a matriz de variância-covariância de $\tilde{\beta}$ pode então ser escrita como:

$$\begin{aligned} \text{var}(\tilde{\beta}) &= \text{var}(Cy) = \text{var}(CX\beta + Cu) \\ &= \text{var}(Cu) = E(Cuu'C') \\ &= CE(uu')C' = \sigma^2CC' \\ &= \sigma^2(X'X)^{-1} + \sigma^2DD' \\ &= \text{var}(\hat{\beta}) + \sigma^2DD', \end{aligned}$$

donde segue a igualdade

$$\text{var}(\tilde{\beta}) - \text{var}(\hat{\beta}) = \sigma^2DD'.$$

Isto conclui a demonstração, haja vista que DD' é simétrica semi-definida positiva, seja qual for a matriz D . ■

Corolário 2.13. *Sob as hipóteses clássicas, para todo estimador linear não-tendencioso $\tilde{\beta}$, o estimador de MQO $\hat{\beta}$ satisfaz a propriedade:*

$$\text{var}(\hat{\beta}_j) \leq \text{var}(\tilde{\beta}_j), \quad j = 1, \dots, k.$$

Nesse sentido, diz-se que $\hat{\beta}$ é o melhor estimador linear não-tendencioso para β (BLUE, do inglês, Best Linear Estimator Unbised).

Demonstração: Com efeito, assumindo e_j como o j -ésimo vetor da base canônica de \mathbb{R}^k , pelo Teorema 2.12 obtém-se

$$\begin{aligned} \text{var}(\tilde{\beta}_j) - \text{var}(\hat{\beta}_j) &= e_j' \text{var}(\tilde{\beta}) e_j - e_j' \text{var}(\hat{\beta}) e_j \\ &= e_j' (\text{var}(\tilde{\beta}) - \text{var}(\hat{\beta})) e_j \\ &= e_j' \sigma^2 D D' e_j \geq 0, \end{aligned}$$

ou seja

$$\text{var}(\hat{\beta}_j) \leq \text{var}(\tilde{\beta}_j).$$

■

2.1.4 Normalidade dos Erros e Teste de Hipóteses

Para aplicar o Método de Mínimos Quadrados Ordinários ao modelo clássico de regressão linear, não é necessário conhecer a distribuição de probabilidade dos erros u_t . Entretanto, se for do interesse não apenas obter $\hat{\beta}$, mas também usá-lo para testar hipóteses sobre o verdadeiro β , ou sobre o comportamento de y , então a distribuição dos erros u_t precisa ser especificada.

Tendo em vista derivar testes de hipóteses e procedimentos de previsão, daqui em diante será considerado o modelo clássico de regressão linear *normal*, o qual, além das hipóteses (H.1)-(H.3), também supõe que os erros u_t possuem distribuição normal, com média zero e variância constante σ^2 . Essa hipótese pode ser posta sob a forma vetorial:

$$(H.4) \quad u \sim N(0, \sigma^2 I_n).$$

Teste de Significância Global

Dado um vetor β_0 constante, considere a hipótese nula $H_0 : \beta = \beta_0$. Um procedimento estatístico para testar esse tipo de hipótese é oferecido pelo seguinte resultado.

Teorema 2.14. Seja $\hat{\beta} = (X'X)^{-1}X'y$ o estimador de MQO, e $\hat{\sigma}^2 = \left(\frac{\hat{u}'\hat{u}}{n-k}\right)$ o estimador para a variância dos erros u_t . Então,

$$\frac{(\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta)}{\hat{\sigma}^2} \frac{1}{k} \sim F(k, n - k).$$

Demonstração: Considerando a equação (2.6) e definindo $M = I_n - X(X'X)^{-1}X'$, obtém-se a igualdade

$$\begin{aligned} (\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta) &= u'X(X'X)^{-1}X'X(X'X)^{-1}X'u \\ &= u'X(X'X)^{-1}X'u \\ &= u'(I_n - M)u. \end{aligned} \tag{2.14}$$

Por outro lado, pelo item (i) do Lema 2.8, conclui-se que $(I_n - M)$ é simétrica e idempotente:

$$(I_n - M)' = I_n' - M' = I_n - M, \tag{2.15}$$

$$(I_n - M)^2 = I_n - 2M + M^2 = I_n - 2M + M = I_n - M. \tag{2.16}$$

donde segue, pelo item (iii) do mesmo Lema,

$$\text{posto}(I_n - M) = \text{tr}(I_n - M) = \text{tr}(I_n) - \text{tr}(M) = n - (n - k) = k. \tag{2.17}$$

A partir das equações (2.14)-(2.17), decorre do Teorema A.5 (ver Apêndice A)

$$\frac{(\hat{\beta} - \beta)'X'X(\hat{\beta} - \beta)}{\sigma^2} = \frac{u'(I_n - M)u}{\sigma^2} \sim \chi^2(k) \tag{2.18}$$

Além disso, por esse mesmo teorema, a equação (2.10) e os itens (i) e (iii) do Lema 2.8 implicam

$$\frac{\hat{u}'\hat{u}}{\sigma^2} = \frac{u'Mu}{\sigma^2} \sim \chi^2(n - k). \tag{2.19}$$

Finalmente, tendo em vista o item (ii) do 2.8, o Teorema A.8 garante que as variáveis aleatórias dadas em (2.18) e (2.19) são independentes. Em razão disso, aplicando a

Definição A.7 da distribuição F , chega-se ao resultado enunciado:

$$\begin{aligned}
 \frac{(\hat{\beta} - \beta)' X' X (\hat{\beta} - \beta) \frac{1}{k}}{\hat{\sigma}^2} &= \frac{(\hat{\beta} - \beta)' X' X (\hat{\beta} - \beta) / k}{\hat{u}' \hat{u} / (n - k)} \\
 &= \frac{[(\hat{\beta} - \beta)' X' X (\hat{\beta} - \beta) / k] / \sigma^2}{[(\hat{u}' \hat{u}) / (n - k)] / \sigma^2} \\
 &= \frac{[(\hat{\beta} - \beta)' X' X (\hat{\beta} - \beta) / \sigma^2] / k}{[(\hat{u}' \hat{u}) / \sigma^2] / (n - k)} \sim F(k, n - k).
 \end{aligned}$$

■

Observação 2.15. Com base no teorema anterior, sob a hipótese $H_0 : \beta = 0$, tem-se

$$\frac{\hat{\beta}' X' X \hat{\beta} \frac{1}{k}}{\hat{\sigma}^2} \sim F(k, n - k).$$

Portanto, para testar H_0 com respeito a um nível α , basta calcular o valor

$$F = \frac{\hat{\beta}' X' X \hat{\beta} \frac{1}{k}}{\hat{\sigma}^2}$$

e compará-lo com o valor crítico correspondente da distribuição $F_\alpha(k, n - k)$. Se $F > F_\alpha$, rejeita-se H_0 .

Teste de Significância Individual

Considere agora uma hipótese nula do tipo $H_0 : \beta_j = 0$. Nesse caso, um teste estatístico pode ser derivado a partir dos seguintes resultados.

Lema 2.16. Se $u \sim N(0, \sigma^2 I_n)$ então, para qualquer vetor $w \in \mathbb{R}^k$, verifica-se

$$w'(\hat{\beta} - \beta) \sim N(0, \sigma^2 w' (X' X)^{-1} w).$$

Demonstração: De fato, considerando a equação (2.6), obtém-se a igualdade

$$w'(\hat{\beta} - \beta) = w' (X' X)^{-1} X' u.$$

Logo, dada a normalidade de u , o Teorema A.3 permite concluir que $w'(\hat{\beta} - \beta)$ tem distribuição normal, com média zero e variância dada por:

$$\begin{aligned} w'(X'X)^{-1}X'var(u)(w'(X'X)^{-1}X')' &= w'(X'X)^{-1}X'(\sigma^2 I_n)X(X'X)^{-1}w \\ &= \sigma^2 w'(X'X)^{-1}X'X(X'X)^{-1}w \\ &= \sigma^2 w'(X'X)^{-1}w. \end{aligned}$$

■

Teorema 2.17. Para qualquer vetor $w \in \mathbb{R}^k$, ocorre o seguinte fato,

$$\frac{w'(\hat{\beta} - \beta)}{\hat{\sigma} \sqrt{w'(X'X)^{-1}w}} \sim t(n - k),$$

onde $\hat{\sigma}$ é a raiz quadrada do estimador de MQO para σ^2 .

Demonstração: Pelo Lema 2.16,

$$\frac{w'(\hat{\beta} - \beta)}{\sigma \sqrt{w'(X'X)^{-1}w}} \sim N(0, 1) \quad (2.20)$$

Além disso, substituindo (2.10) na definição de $\hat{\sigma}^2$, obtém-se a igualdade

$$\begin{aligned} \frac{\hat{\sigma}^2}{\sigma^2} &= \left(\frac{\hat{u}'\hat{u}}{n - k} \right) \frac{1}{\sigma^2} \\ &= \frac{(u'Mu)}{\sigma^2(n - k)} \end{aligned} \quad (2.21)$$

Por fim, com base na Definição A.6 da distribuição t de Student, os resultados (2.19), (2.20) e (2.21) implicam:

$$\begin{aligned} \frac{w'(\hat{\beta} - \beta)}{\hat{\sigma} \sqrt{w'(X'X)^{-1}w}} &= \frac{w'(\hat{\beta} - \beta)}{\sigma \sqrt{w'(X'X)^{-1}w}} \left(\frac{\sigma}{\hat{\sigma}} \right) \\ &= \frac{w'(\hat{\beta} - \beta) / \sigma \sqrt{w'(X'X)^{-1}w}}{\sqrt{\hat{\sigma}^2 / \sigma^2}} \\ &= \frac{w'(\hat{\beta} - \beta) / \sigma \sqrt{w'(X'X)^{-1}w}}{\sqrt{(u'Mu) / \sigma^2(n - k)}} \sim t(n - k) \end{aligned}$$

■

Observação 2.18. Com base no teorema anterior, sob a hipótese $H_0 : \beta = 0$, ocorre

$$\frac{\hat{\beta}_j}{\sqrt{\widehat{\text{var}}(\hat{\beta}_j)}} = \frac{e'_j \hat{\beta}}{\sqrt{e'_j (\text{var}(\hat{\beta})) e_j}} = \frac{e'_j (\hat{\beta} - 0)}{\hat{\sigma} \sqrt{e'_j (X'X)^{-1} e_j}} \sim t(n-k),$$

Portanto, para testar H_0 com respeito a um nível α , basta calcular o valor

$$t = \frac{\hat{\beta}_j}{\sqrt{\widehat{\text{var}}(\hat{\beta}_j)}},$$

e compará-lo com o valor crítico correspondente da distribuição $t(n-k)$. Se $t > t_\alpha(n-k)$, rejeita-se H_0 .

Previsões Pontuais e Intervalos de Confiança

Suponha a equação de regressão linear:

$$y_0 = x'_0 \beta + u_0.$$

Nesse caso, o Teorema 2.6 e a hipótese (H.2) implicam a seguinte igualdade:

$$E(x'_0 \hat{\beta}) = x'_0 E(\hat{\beta}) = x'_0 \beta = E(y_0).$$

Assim, conhecido o vetor x_0 , o número $\hat{y}_0 = x'_0 \hat{\beta}$ pode ser usado como uma previsão pontual de $E(y_0)$.

Por outro lado, assumindo a normalidade dos erros, segue pelo Teorema 2.17 que

$$\frac{\hat{y}_0 - E(y_0)}{\hat{\sigma} \sqrt{x'_0 (X'X)^{-1} x_0}} = \frac{x'_0 \hat{\beta} - x'_0 \beta}{\hat{\sigma} \sqrt{x'_0 (X'X)^{-1} x_0}} = \frac{x'_0 (\hat{\beta} - \beta)}{\hat{\sigma} \sqrt{x'_0 (X'X)^{-1} x_0}} \sim t(n-k)$$

Logo, fixado um nível de significância $0 < \alpha < 1$, o intervalo de confiança para a previsão de $E(y_0)$ é o seguinte:

$$\left[\hat{y}_0 - t_\alpha \hat{\sigma} \sqrt{x'_0 (X'X)^{-1} x_0}, \quad \hat{y}_0 + t_\alpha \hat{\sigma} \sqrt{x'_0 (X'X)^{-1} x_0} \right] \quad (2.22)$$

No entanto, existem situações nas quais o interesse está em obter intervalos de confiança para y_0 , ao invés de $E(y_0)$. Em ambos os casos, a previsão pontual é a mesma:

$\hat{y}_0 = x_0' \hat{\beta}$. A diferença está na amplitude do intervalo de confiança das previsões. Com efeito, definindo

$$e_0 = y_0 - \hat{y}_0 = x_0' \beta + u_0 - x_0' \hat{\beta} = u_0 - x_0' (\hat{\beta} - \beta),$$

segue que e_0 tem distribuição normal, com média zero e variância:

$$\begin{aligned} \text{var}(e_0) &= E \left[(u_0 - x_0' (\hat{\beta} - \beta))^2 \right] \\ &= E \left[u_0^2 - 2u_0 x_0' (\hat{\beta} - \beta) + (x_0' (\hat{\beta} - \beta))^2 \right] \\ &= \sigma^2 + \text{var}(x_0' (\hat{\beta} - \beta)) \\ &= \sigma^2 + \sigma^2 x_0' (X'X)^{-1} x_0 \\ &= \sigma^2 (1 + x_0' (X'X)^{-1} x_0). \end{aligned}$$

Como consequência,

$$\frac{\hat{y}_0 - y_0}{\sigma \sqrt{1 + x_0' (X'X)^{-1} x_0}} = \frac{e_0}{\sqrt{\text{var}(e_0)}} \sim N(0, I_n)$$

Assim, usando o mesmo argumento utilizado no Teorema 2.17, a substituição de σ por $\hat{\sigma}$ na expressão acima resulta em:

$$\frac{\hat{y}_0 - y_0}{\hat{\sigma} \sqrt{1 + x_0' (X'X)^{-1} x_0}} \sim t(n - k)$$

Portanto, fixado um nível de significância $0 < \alpha < 1$, o intervalo de confiança para a previsão de y_0 é o seguinte:

$$\left[\hat{y}_0 - t_\alpha \hat{\sigma} \sqrt{1 + x_0' (X'X)^{-1} x_0}, \quad \hat{y}_0 + t_\alpha \hat{\sigma} \sqrt{1 + x_0' (X'X)^{-1} x_0} \right] \quad (2.23)$$

Observação 2.19. Comparando (2.22) e (2.23), a única diferença é o aumento de uma unidade na raiz quadrada. Em razão disso, para um mesmo nível de significância α , o intervalo de confiança da previsão de y_0 é maior que o intervalo da previsão de $E(y_0)$.

2.2 Violação das Hipóteses Clássicas

Esta seção aborda, individualmente, os problemas que ocorrem quando as hipóteses (H.1), (H.3) e (H.4) são violadas. Em particular, são descritos alguns procedimentos estatísticos para testar a validade dessas hipóteses.

2.2.1 Multicolinearidade

O termo *multicolinearidade* refere-se à violação da hipótese (H.1), ou seja, diz respeito aos casos em que a matriz de observações X não tem posto completo. Considerando as colunas x_1, \dots, x_k de X , esse tipo de problema equivale à existência de escalares $\lambda_1, \dots, \lambda_k$, não todos nulos, tais que

$$\lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_k x_k = 0.$$

Como resultado da multicolinearidade, a matriz $X'X$ se torna singular, de modo que o sistema linear $(X'X)\hat{\beta} = X'y$ deixa de possuir solução única², culminando na indeterminação da estimativa $\hat{\beta}$ de mínimos quadrados ordinários.

Na prática, dificilmente se verifica a multicolinearidade. O que ocorre com maior frequência é a chamada *quase-multicolinearidade*, caracterizada pela existência de escalares $\lambda_1, \dots, \lambda_k$, não todos nulos, tais que

$$\lambda_1 x_1 + \lambda_2 x_2 + \dots + \lambda_k x_k + v = 0,$$

onde v é um erro aleatório. Neste caso, uma vez que as hipóteses clássicas sejam atendidas, o Corolário 2.13 continua válido, ou seja, o estimador de MQO $\hat{\beta}$ se mantém sendo BLUE. Todavia, isso não significa que a variância das suas coordenadas seja necessariamente pequena. O que se observa nos casos de quase-multicolinearidade é a ocorrência dos seguintes problemas:

- (a) Os estimadores $\hat{\beta}_i$ apresentam variância e covariância elevadas, comprometendo a precisão das suas estimativas;
- (b) Em decorrência de *a*), as estatísticas t e F tendem a ser menores, facilitando a aceitação da hipótese de insignificância dos coeficientes³;
- (c) Também em virtude de *a*), as previsões com o modelo de regressão linear se tornam mais imprecisas, dado o alargamento dos intervalos de confiança.

²Nesse caso, ou existem infinitas soluções, ou não existe solução.

³Ver a Seção 2.1.4.

Detecção da Quase-Multicolinearidade

Não existe um método único para medir a intensidade da quase-multicolinearidade⁴. Uma forma de quantificá-la é através do chamado *número de condição* (k). Ele é definido como a raiz quadrada da razão entre o maior autovalor λ_{max} e o menor autovalor λ_{min} da matriz $X'X$, em valores absolutos:

$$k = \sqrt{\frac{|\lambda_{max}|}{|\lambda_{min}|}}. \quad (2.24)$$

Obtido esse número, tem-se então a seguinte regra prática [32]:

- Se k está entre 10 e 30 então o nível de quase-multicolinearidade está entre moderado e forte;
- Se k exceder 30 então a multicolinearidade é grave.

2.2.2 Autocorrelação e Heterocedasticidade

Considere a situação em que a hipótese ($H.3$) não se verifica, isto é

$$var(u) = \Sigma \neq \sigma^2 I_n. \quad (2.25)$$

Pode-se distinguir duas causas possíveis para essa violação, a saber:

Autocorrelação: A matriz de variância-covariância Σ não é diagonal⁵;

Heterocedasticidade: Os elementos da diagonal principal de Σ não são todos iguais⁶.

Como o Teorema 2.6 depende apenas das hipóteses ($H.1$) e ($H.2$), em ambos os casos o estimador de MQO $\hat{\beta}$ continua sendo não-tendencioso. Porém, a sua matriz de

⁴Para uma discussão sobre vários desses métodos, ver [32, 48].

⁵Caso onde existem $t \neq s$ tais que $E(u_t, u_s) \neq 0$.

⁶Caso onde existem $t \neq s$ tais que a $var(u_t) \neq var(u_s)$.

variância-covariância passa a ser dada por:

$$\begin{aligned}
 \text{var}(\hat{\beta}) &= E [(\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))'] \\
 &= E \left[((X'X)^{-1}X'u)((X'X)^{-1}X'u)' \right] \\
 &= E \left[(X'X)^{-1}X'uu'X(X'X)^{-1} \right] \\
 &= (X'X)^{-1}X'E(uu')X(X'X)^{-1} \\
 &= (X'X)^{-1}X'\Sigma X(X'X)^{-1}.
 \end{aligned}$$

Isso significa que a fórmula convencional para a matriz de variância-covariância

$$\text{var}(\hat{\beta}) = \sigma^2(X'X)^{-1}$$

já não se aplica, o que compromete a validade dos procedimentos estatísticos que dela fazem uso, tais como os testes de significância.

A seguir, são descritos os testes estatísticos utilizados neste trabalho para investigar a ocorrência de autocorrelação e heterocedasticidade.

Teste Breusch-Godfrey para Autocorrelação

Seja o modelo de regressão linear dado na forma

$$y_t = x_t'\beta + u_t, \quad t = 1, \dots, n \quad (2.26)$$

onde os erros obedecem o seguinte esquema

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_p u_{t-p} + \epsilon_t, \quad \epsilon_t \sim IN(0, \sigma^2). \quad (2.27)$$

Nesse caso, a hipótese de ausência da autocorrelação pode ser posta como

$$H_0 : \rho_1 = \dots = \rho_p = 0.$$

O teste Breusch-Godfrey, também conhecido por teste LM, permite investigar se há ou não correlação entre os erros do modelo (2.26). Ele pode ser descrito pelos seguintes passos:

- 1) Estime β aplicando MQO;

2) Considerando os resíduos \hat{u}_t , estime o modelo de regressão auxiliar

$$\hat{u}_t = \sum_{i=1}^k \gamma_t x_{i,t} + \sum_{i=1}^p \rho_i \hat{u}_{t-i} + \eta_t. \quad (2.28)$$

Em seguida, obtenha o seu coeficiente de determinação R^2 e calcule $\theta = (n - p)R^2$;

3) Conforme mostrado em [30], sob a hipótese nula H_0 tem-se $\theta \sim \chi^2(p)$. Portanto, se o valor θ calculado exceder o valor crítico da distribuição $\chi^2(p)$ no nível de significância escolhido, a hipótese nula é rejeitada, ou seja, pode-se considerar que os erros apresentam autocorrelação.

Teste Breusch-Pagan para Heterocedasticidade

Com respeito ao modelo (2.26), suponha que os erros u_t são normalmente e independentemente distribuídos, com média zero e variância dada por

$$\sigma_t^2 = h(z_t' \alpha), \quad (2.29)$$

onde a função h possui derivadas de segunda ordem, $\alpha \in \mathbb{R}^p$ é um vetor de coeficientes não relacionados com aqueles que compõe β , e o primeiro elemento de z_t é igual à 1. Neste caso, a hipótese da ausência de heterocedasticidade, denominada *homocedasticidade*, pode ser escrita como:

$$H_0 : \alpha_2 = \dots = \alpha_p = 0,$$

uma vez que $\sigma_t^2 = h(z_t' \alpha) = h(\alpha_1) = \sigma^2$ é constante.

Utilizando a fórmula $\hat{\sigma}^2 = \frac{\sum \hat{u}_t^2}{n}$ para estimar a variância σ^2 dos erros⁷, pode-se estabelecer o seguinte resultado.

Teorema 2.20. *Considere os modelos (2.26) e (2.29) sob a hipótese nula $H_0 : \alpha_2 = \dots = \alpha_p$ de homocedasticidade dos erros. Então, definindo SQE como a soma dos quadrados explicados na regressão de $g_t = \frac{\hat{u}_t^2}{\hat{\sigma}^2}$ sobre z_t , tem-se $\frac{SQE}{2} \sim \chi^2(p - 1)$.*

⁷Essa fórmula corresponde ao estimador de Máxima Verossimilhança para σ^2 . Para maiores detalhes, ver Apêndice 4A em [32].

Demonstração: Ver Seção 2 em [17]. ■

Desse teorema decorre o teste Breusch-Pagan para a hipótese de homocedasticidade dos erros em um modelo de regressão linear. Esse teste pode ser descrito da seguinte maneira:

1) Estime (2.26) usando MQO, e obtenha os resíduos $\{\hat{u}_t\}_{t=1}^n$;

2) Calcule $\hat{\sigma}^2 = \left(\sum_{t=1}^n \hat{u}_t^2 \right) / n$;

3) Construa variáveis $g_t = \frac{\hat{u}_t^2}{\hat{\sigma}^2}$, para $t = 1, \dots, n$;

4) Estime por MQO os coeficientes da equação

$$g_t = \alpha_1 + \alpha_2 z_{2,t} + \dots + \alpha_p z_{p,t} + v_t, \quad t = 1, \dots, n, \quad (2.30)$$

onde v_t é o termo de erro;

5) Obtenha a soma dos quadrados explicados SQE de (2.30), e calcule $\lambda = \frac{1}{2} \text{SQE}$.

6) Se o valor λ calculado exceder o valor crítico da distribuição $\chi^2(p-1)$ no nível de significância escolhido, a hipótese H_0 é rejeitada, ou seja, pode-se assumir que os erros apresentam heterocedasticidade.

Observação 2.21. *Em amostras pequenas, o teste de Breusch-Pagan é sensível à hipótese de que os erros u_t tenham distribuição normal, exigindo que a hipótese de normalidade seja testada antes da sua aplicação.*

Teste de White para Heterocedasticidade

Um teste que não depende da hipótese de normalidade é o teste geral de White. Para ilustrar a idéia deste teste, considere o seguinte modelo de regressão linear:

$$y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + u_t. \quad (2.31)$$

O teste de White é feito da seguinte maneira:

1) A partir dos dados, estime a equação (2.31) e calcule os resíduos \hat{u}_t ;

2) Obtenha por MQO os coeficientes do modelo de regressão auxiliar

$$\hat{u}_t^2 = \alpha_0 + \alpha_1 x_{1,t} + \alpha_2 x_{2,t} + \alpha_3 x_{1,t}^2 + \alpha_4 x_{2,t}^2 + \alpha_5 x_{1,t} x_{2,t} + v_t,$$

e calcule o R^2 associado a essas estimativas;

3) Conforme mostrado em [71], sob a hipótese nula de homocedasticidade dos erros, tem-se $nR^2 \sim \chi^2(5)$, ou, no caso geral de um modelo com k regressores, $nR^2 \sim \chi^2(k(k+1)/2)$. Portanto, se o valor nR^2 calculado exceder o valor crítico da distribuição χ^2 no nível de significância escolhido, rejeita-se a hipótese de homocedasticidade.

Observação 2.22. *Se um modelo possuir muitos regressores, ao se introduzir todos eles na regressão auxiliar, seus termos elevados ao quadrado e seus produtos cruzados podem consumir rapidamente os graus de liberdade da regressão. Por isso, em alguns casos este teste é aplicado omitindo-se os termos de produto cruzado.*

2.2.3 Não-normalidade dos Erros

De acordo com a subseção 2.1.3, as propriedades estatísticas do estimador de Mínimos Quadrados Ordinários $\hat{\beta}$ não dependem da distribuição de probabilidade dos erros u_t . No entanto, a validade dos testes de hipóteses estabelecidos até aqui baseia-se na premissa (H.4) de que os erros u_t possuem distribuição normal. Assim, a aplicação desses testes exige que se aplique, antes, um teste estatístico para a hipótese:

$$H_0 : u_t \sim N(0, \sigma^2), \quad t = 1, \dots, n.$$

A seguir, faz-se um breve descrição de um desses testes.

Teste Shapiro-Wilk

O teste Shapiro-Wilk tem por base a estatística W e apresenta bom desempenho em amostras pequenas ($n < 20$). Dada uma amostra aleatória de tamanho n , composta pelas observações x_1, \dots, x_n , o procedimento de teste consiste nos seguintes passos:

1) Ordene as observações para obter uma amostra $y_1 \leq \dots \leq y_n$;

2) Calcule

$$S^2 = \sum_{t=1}^n (y_t - \bar{y})^2 = \sum_{t=1}^n (x_t - \bar{x})^2$$

onde $\bar{y} = \bar{x}$ denota a média aritmética da amostra considerada;

3) (a) Se n é par, considere $k = n/2$ e calcule:

$$b = \sum_{t=1}^k a_{n-t+1} (y_{n-t+1} - y_t), \quad (2.32)$$

onde os valores de a_{n-t+1} são dados na tabela 5 em [63].

(b) Se n é ímpar, considere $k = (n - 1)/2$ e calcule b utilizando (2.32);

4) Calcule $W = b^2/S^2$;

5) Na tabela 6 em [63], valores críticos da distribuição de W são fornecidos para vários níveis de significância. Se o valor W calculado for inferior ao valor crítico no nível de significância escolhido, então rejeita-se a hipótese de que a amostra x_1, \dots, x_n provém de uma distribuição normal.

Observação 2.23. No caso do modelo de regressão linear, o teste é aplicado sobre a amostra de resíduos $\hat{u}_1, \dots, \hat{u}_n$.

2.3 Tópicos Especiais de Econometria

Nesta seção são tratados três tópicos de econometria, a saber: Regressões sobre Variáveis Binárias, Seleção de Regressores e Séries Temporais Não-Estacionárias. O objetivo é descrever alguns conceitos e procedimentos empregados no Capítulo 5.

2.3.1 Regressão sobre Variáveis Binárias

Existem circunstâncias em que a variável dependente y também é influenciada por variáveis de natureza qualitativa (política econômica, sexo, religião, etc.). Como essas

variáveis geralmente indicam a presença ou ausência de um certo “atributo” (política A ou política B, homem ou mulher, católico ou não-católico, etc.), é possível representá-las por meio de *variáveis binárias*, caracterizadas por assumirem os valores “1” ou “0”, conforme o atributo se verifique ou não.

Para efeito de ilustração, considere dois grupos de observações descritos, respectivamente, pelos modelos:

$$y_t = \alpha_1 + \sum_{j=1}^{k-1} \beta_j x_{jt} + u_t \quad t = 1, \dots, n_1 \quad (2.33)$$

$$y_t = \alpha_2 + \sum_{j=1}^{k-1} \beta_j x_{jt} + u_t \quad t = n_1 + 1, \dots, n \quad (2.34)$$

Esses modelos podem ser combinadas em uma única equação:

$$y_t = \alpha_1 + (\alpha_2 - \alpha_1)D + \sum_{j=1}^{k-1} \beta_j x_{jt} + u_t \quad t = 1, \dots, n$$

onde,

$$D = 0, \quad \text{para } t = 1, \dots, n_1$$

$$D = 1, \quad \text{para } t = n_1 + 1, \dots, n$$

Nesse caso, D é uma variável binária, e o seu coeficiente mede a diferença entre os dois termos de intercepto, α_1 e α_2 .

2.3.2 Seleção de Regressores

Na prática, são raras as situações em que se possui de antemão um conhecimento exato sobre os regressores do modelo. Geralmente, existe um grande número de variáveis explicativas em potencial, e tem-se então o problema de escolher quais delas incluir no modelo. Este problema é referido na literatura como o *problema de seleção de regressores* [48].

Suponha que o objetivo seja obter um modelo de regressão linear que gere previsões para uma variável escalar y , sobre a qual estão disponíveis dados do período

$t = 1, \dots, n$. Sem perda de generalidade, considere apenas dois conjuntos de regressores, cujos dados no instante t são agrupados nos vetores $x_t \in \mathbb{R}^l$ e $z_t \in \mathbb{R}^k$, respectivamente. Dessa forma, obtém-se os seguintes modelos:

$$\text{Modelo 1: } y_t = \alpha' x_t + u_t$$

$$\text{Modelo 2: } y_t = \beta' z_t + v_t$$

Considere então o problema de se escolher entre os modelos 1 e 2 aquele que resulte na previsão mais acurada para o instante $n + 1$, em termos do erro quadrático médio de previsão (PMSE, do inglês, *Prediction Mean Squared Error*). Essa seleção pode ser feita pelo procedimento de simulação fora da amostra (SOOS, do inglês *Simulated Out-Of-Sample*), ou então aplicando um critério de informação (IC, do inglês, *Information Criteria*) [38].

No procedimento SOOS, cada modelo é estimado por MQO sobre as s primeiras observações, e procede-se a avaliação dos erros sobre as $(n - s)$ observações restantes:

$$PMSE_1 = \left(\sum_{t=s+1}^n \hat{u}_t^2 \right) / (n - s)$$

$$PMSE_2 = \left(\sum_{t=s+1}^n \hat{v}_t^2 \right) / (n - s)$$

onde $\hat{u}_t = y_t - \hat{\alpha}' x_t$ e $\hat{v}_t = y_t - \hat{\beta}' z_t$. Feito isso, escolhe-se o modelo que apresentar o menor PMSE, o qual então é reestimado sobre toda a amostra $t = 1, \dots, n$, e usado para prever y_{n+1} .

Por outro lado, a seleção por IC é implementada da seguinte maneira. Cada modelo é ajustado por MQO sobre a amostra completa de n observações. Então, os modelos são avaliados pelo seguinte critério:

$$IC_1 = \ln \left(\sum_{t=1}^n \hat{u}_t^2 / n \right) + \dim(\alpha) c_n \quad (2.35)$$

$$IC_2 = \ln \left(\sum_{t=1}^n \hat{v}_t^2 / n \right) + \dim(\beta) c_n \quad (2.36)$$

onde $\dim(x)$ denota o número de coordenadas do vetor x . Escolhe-se o modelo que apresentar o menor valor IC, sendo ele utilizado para prever y_{n+1} .

A forma de c_n depende do critério escolhido. Os Critérios de Schwarz (SIC) e de Akaike (AIC) estão entre os mais utilizados, sendo definidos por $c_n = \ln(n)/n$ e $c_n = 1/n$, respectivamente.

2.3.3 Séries Temporais Não-Estacionárias

Se uma variável aleatória X é indexada pelo tempo, denotado por t , o conjunto de observações $\{X_t, t \in \mathbb{T}\}$ é referido como *série temporal*, sendo \mathbb{T} um conjunto de índices (por exemplo, $\mathbb{T} = \mathbb{Z}$).

Na modelagem de um número finito de variáveis aleatórias, a matriz de variância-covariância é usualmente calculada para resumir a relação de dependência entre essas variáveis. No caso de uma série temporal, é necessário modelar a dependência de um número infinito de variáveis, o que é feito através da função de autocovariância.

Definição 2.24. *Dada uma série temporal $\{X_t, t \in \mathbb{T}\}$ com $\text{Var}(X_t) < \infty$, define-se a sua função de autocovariância como:*

$$\gamma_X(s, t) = \text{Cov}(X_s, X_t) = E[(X_s - E(X_s))(X_t - E(X_t))]. \quad (2.37)$$

A partir dessa função, pode-se definir a estacionariedade de uma série temporal.

Definição 2.25. *Diz-se que a série temporal $\{X_t, t \in \mathbb{Z}\}$ é estacionária se as seguintes condições são satisfeitas:*

- i) $E(X_t) = \mu$, para todo $t \in \mathbb{Z}$;
- ii) $\text{var}(X_t) = E(X_t - \mu)^2 = \sigma^2$, para todo $t \in \mathbb{Z}$;
- iii) $\gamma_X(s, t) = \gamma_X(s + h, t + h)$ para quaisquer $s, t, h \in \mathbb{Z}$.

Em outras palavras, para ser estacionária, uma série temporal precisa ter média e variância constante, enquanto a covariância entre dois períodos, s e t , deve depender apenas da distância $(t - s)$ que os separa. No caso em que a série temporal não satisfaz a definição anterior, ela é dita *não-estacionária*.

Teorema 2.26. Seja $\{y_t, t \in \mathbb{Z}\}$ uma série temporal definida por

$$y_t = \mu + \rho y_{t-1} + u_t, \quad (2.38)$$

onde $u_t \sim IN(0, \sigma^2)$. Se $|\rho| \geq 1$ então $\{y_t\}$ é uma série temporal não-estacionária⁸.

Demonstração: Supondo y_0 conhecido, pode-se obter por indução

$$\begin{aligned} y_1 &= \mu + \rho y_0 + u_1 \\ y_2 &= \mu + \rho y_1 + u_2 = \mu + \rho\mu + \rho^2 y_0 + \rho u_1 + u_2 \\ y_3 &= \mu + \rho y_2 + u_3 = \mu + \rho\mu + \rho^2\mu + \rho^3 y_0 + \rho^2 u_1 + \rho u_2 + u_3 \\ &\vdots \\ y_n &= \mu(1 + \rho + \rho^2 + \dots + \rho^{n-1}) + \rho^n y_0 + \sum_{t=1}^n \rho^{n-t} u_t \end{aligned}$$

Logo,

$$\begin{aligned} E(y_n) &= \mu(1 + \rho + \dots + \rho^{n-1}) + \rho^n y_0 + \sum_{t=1}^n \rho^{n-t} E(u_t) \\ &= \mu(1 + \rho + \dots + \rho^{n-1}) + \rho^n y_0 \quad e \end{aligned}$$

$$var(y_n) = \sum_{t=1}^n var(\rho^{n-t} u_t) = \sum_{t=1}^n \rho^{2(n-t)} \sigma^2 = \sigma^2 \left(\sum_{t=1}^n \rho^{2(n-t)} \right).$$

Assim, se $|\rho| \geq 1$ então ambas as sequências, $\{E(y_n)\}$ e $\{var(y_n)\}$, são divergentes, o que constitui uma violação das condições (i) e (ii) na Definição 2.25. Portanto, a série dada em (2.38) é não-estacionária⁹. ■

Regressões Espúrias

Regressões lineares envolvendo dados de séries temporais podem fornecer resultados duvidosos em determinadas situações, indicando uma forte relação entre variáveis que

⁸De fato, para todo $\rho \neq 0$ essa série é não-estacionária, porém, para a análise que segue é conveniente considerar apenas esse caso particular.

⁹Quando $\mu = 0$ e $\rho = 1$, a série temporal descrita em (2.38) é chamada de *passeio aleatório* [48].

na verdade se comportam de maneira independente. Esse fenômeno, conhecido como *regressão espúria*, caracteriza-se pela obtenção de um alto coeficiente de determinação R^2 e elevada significância estatística dos coeficientes, embora não exista, de fato, uma correlação entre as variáveis consideradas.

Conforme mostrado por Phillips [58], a regressão espúria pode ocorrer se as séries temporais sob análise forem não-estacionárias. Nesse caso, os testes de significância convencionais se tornam viesados, levando o pesquisador a rejeitar mais facilmente a hipótese nula $H_0 : \beta_j = 0$ (ou $H_0 : \beta = 0$).

Diante dessas considerações, torna-se interessante testar se as séries consideradas no modelo são estacionárias ou não. A seguir, um teste estatístico para esse propósito é apresentado.

Teste Dickey-Fuller

Considere o modelo autorregressivo:

$$y_t = \mu + \rho y_{t-1} + u_t, \quad t = 1, 2, \dots, n, \dots \quad (2.39)$$

onde $\mu, \rho \in \mathbb{R}$ e os erros u_t são independentes e normalmente distribuídos, com média zero e variância constante σ^2 . Observando o argumento utilizado para provar o Teorema 2.26, pode-se distinguir três situações:

$|\rho| < 1$: $\{y_t\}$ converge para uma série estacionária quando $t \rightarrow \infty$;

$|\rho| = 1$: $\{y_t\}$ é não-estacionária, com $\text{var}(y_t) = t\sigma^2$;

$|\rho| > 1$: $\{y_t\}$ é não-estacionária, com a sua variância crescendo de forma exponencial em relação a t .

A hipótese $H_0 : \rho = 1$ é de interesse em aplicações, pois nesse caso torna-se apropriado diferenciar a série temporal $\{y_t\}$, como mostra o teorema a seguir.

Teorema 2.27. *Seja $\{y_t\}$ uma série temporal dada por $y_t = \mu + y_{t-1} + u_t$, onde os erros u_t são independentemente e igualmente distribuídos, com média zero e variância σ^2 . Neste caso, a série temporal definida por $\Delta y_t = y_t - y_{t-1}$ é uma série estacionária.*

Demonstração: Com efeito, para ver isto basta notar que:

$$\begin{aligned}y_t &= \mu + y_{t-1} + u_t \\ \Rightarrow \Delta y_t &= y_t - y_{t-1} = \mu + u_t\end{aligned}$$

Como consequência,

$$E(\Delta y_t) = \mu, \quad \text{e} \quad \text{var}(\Delta y_t) = \text{var}(u_t) = \sigma^2.$$

Além disso, como os erros são independentes, tem-se

$$\begin{aligned}\gamma_{\Delta y}(s, t) &= E[(\Delta y_s - \mu)(\Delta y_t - \mu)] \\ &= E[(\mu + u_s - \mu)(\mu + u_t - \mu)] \\ &= E[(u_s)(u_t)] \\ &= 0,\end{aligned}$$

■

Nesse contexto, o teste Dickey-Fuller é um teste para a hipótese nula $H_0 : \rho = 1$. Os modelos considerados são os seguintes:

1. $y_t = \rho y_{t-1} + u_t, \quad t = 1, 2, \dots, n, \dots$
2. $y_t = \mu + \rho y_{t-1} + u_t, \quad t = 1, 2, \dots, n, \dots$
3. $y_t = \mu + \beta t + \rho y_{t-1} + u_t, \quad t = 1, 2, \dots, n, \dots$

A diferença entre o modelo 3 e os demais reside na inclusão da constante (intercepto) e do termo de tendência. Nesse caso, se a hipótese nula for verificada, deve-se subtrair a tendência de y_t (isto é, βt) e depois diferenciá-la, a fim de obter uma série estacionária.

O procedimento de teste ocorre de acordo com as seguintes etapas:

- 1) Especifique um modelo para a série (1,2 ou 3) e estime seus coeficientes aplicando MQO;
- 2) Designando por $\hat{\rho}$ e $\hat{\sigma}$, respectivamente, a estimativa de MQO de ρ e o erro padrão correspondente, considere a estatística $t = \frac{\hat{\rho} - 1}{\hat{\sigma}}$;

3) De acordo com Dickey e Fuller [28], o valor t assim obtido provém de uma distribuição de probabilidade diferente da t de Student. Sob a hipótese nula H_0 , os valores críticos dessa outra distribuição foram tabulados por esses autores com base em simulações de Monte Carlo. Tais valores dependem do tamanho da amostra e do modelo considerado, sendo t denotado por τ , τ_μ , e τ_τ para os modelos 1, 2 e 3, respectivamente. Se o valor absoluto da estatística calculada exceder o valor crítico informado pela tabela de Dickey e Fuller no nível de significância escolhido, então rejeita-se a hipótese nula.

Teste Dickey-Fuller Aumentado

Se nos modelos 1, 2 e 3, os erros u_t apresentam autocorrelação, então o teste de Dickey-Fuller é aplicado considerando a seguinte especificação:

$$y_t = \beta_1 + \beta_2 t + \rho y_{t-1} + \sum_{i=1}^p \alpha_i y_{t-1-i} + \epsilon_t. \quad (2.40)$$

O número p de termos defasadas é muitas vezes determinado de forma empírica: a idéia é incluir termos suficientes de modo que os erros em (2.40) não apresentem autocorrelação. Quando o teste de Dickey-Fuller é aplicado considerando um modelo do tipo (2.40), ele é referido como teste Dickey-Fuller Aumentado (ADF, do inglês, *Augmented Dickey-Fuller*). A estatística do teste ADF tem a mesma distribuição assintótica que a estatística DF, o que permite que os mesmos valores críticos sejam usados em ambos os testes.

Programação Dinâmica em Tempo Discreto

Este capítulo, fortemente baseado em Bellman [15], Bertsekas [16], Ljungqvist e Sargent [46] e Stokey *et al.* [65], tem como foco principal a aplicação do método de programação dinâmica ao problema do regulador linear-quadrático. Sua finalidade é fornecer o subsídio teórico suficiente para a solução do problema de condução da política monetária, o qual é objeto de estudo do Capítulo 5.

3.1 Horizonte Finito

Para a análise subsequente, convém primeiro considerar o seguinte problema de controle ótimo:

$$\min_{u_0, \dots, u_{N-1}} E_0 \left(\sum_{t=0}^{N-1} \delta^t c(x_t, u_t) \right) \quad (3.1)$$

$$\text{sujeito a } x_{t+1} = g(x_t, u_t, \epsilon_t) \quad (3.2)$$

onde $x_t \in X \subset \mathbb{R}^m$ é a variável de estado, $u_t \in A \subset \mathbb{R}^n$ é a variável de controle, $\delta \in (0, 1)$ é taxa de desconto intertemporal, $c : X \times A \rightarrow \mathbb{R}$ é a função perda, ϵ_t é uma variável aleatória que toma valores em \mathbb{R}^m , e E_0 é a esperança condicionada pela informação disponível no período $t = 0$. Além disso, suponha que $u_t = \mu_t(x_t)$, para todo $t = 0, \dots, N - 1$, onde μ_t é referida como uma *regra de decisão*.

Definição 3.1. Uma política de tamanho N é uma sequência de regras de decisão $\pi_N = (\mu_0, \dots, \mu_{N-1})$, onde $\mu_t : X \rightarrow \mathbb{R}^n$, para todo $t = 0, \dots, N-1$. Nesse caso, o conjunto das políticas viáveis de tamanho N é definido por $\Pi_N = \{\pi_N = (\mu_0, \dots, \mu_{N-1}), \mu_t(x) \in A, \forall x, t\}$.

Definição 3.2. Supondo que faltem T períodos para o fim do horizonte de programação, e que nesse estágio o estado do sistema seja x_0 , define-se o valor de uma política viável π_T como a soma:

$$W_T(x_0, \pi_T) = E_0 \left(\sum_{t=0}^{T-1} \delta^t c(x_t, u_t) \right)$$

A partir dessa notação, o problema (3.1)-(3.2) pode ser posto na forma:

$$\min_{\pi_N \in \Pi_N} W_N(x_0, \pi_N) \quad (3.3)$$

$$\text{sujeito a } x_{t+1} = g(x_t, \mu_t(x_t), \epsilon_t) \quad (3.4)$$

Definição 3.3. Se o problema (3.3)-(3.4) admite uma política ótima $\pi_N^* = (\mu_0^*, \dots, \mu_{N-1}^*)$ para todo o estado inicial x_0 , então a sua função valor ótimo $V_N : X \rightarrow \mathbb{R}$ é definida por

$$V_N(x) = W_N(x, \pi_N^*).$$

Teorema 3.4 (Equação de Bellman). Supondo que a função valor ótimo esteja bem definida para $j = 1, \dots, N$, então para todos esses índices verifica-se a seguinte igualdade:

$$V_j(x) = \min_u \{c(x, u) + \delta E [V_{j-1}(g(x, u, \epsilon))]\} \quad (3.5)$$

Demonstração: Partindo da Definição 3.3, e considerando a lei das expectativas iteradas¹, obtém-se:

$$\begin{aligned} V_N(x_0) &= W_N(x_0, \pi_N^*) = \min_{\pi_N \in \Pi_N} W_N(x_0, \pi_N) = \min_{u_0, \dots, u_{N-1}} E_0 \left(\sum_{t=0}^{N-1} \delta^t c(x_t, u_t) \right) \\ &= \min_{u_0, \dots, u_{N-1}} E_0 \left[c(x_0, u_0) + E_1 \left(\sum_{t=1}^{N-1} \delta^t c(x_t, u_t) \right) \right] \\ &= \min_{u_0} E_0 \left[c(x_0, u_0) + \min_{u_1, \dots, u_{N-1}} E_1 \left(\sum_{t=1}^{N-1} \delta^t c(x_t, u_t) \right) \right]. \\ &= \min_{u_0} E_0 \left\{ c(x_0, u_0) + \delta \left[\min_{u_1, \dots, u_{N-1}} E_1 \left(\sum_{t=1}^{N-1} \delta^{t-1} c(x_t, u_t) \right) \right] \right\}, \end{aligned} \quad (3.6)$$

¹Segundo essa lei, $E_0(\cdot) = E_0 [E_1(\cdot)]$.

onde a penúltima igualdade ocorre devido a independência de $c(x_0, u_0)$ em relação às decisões tomadas em $t \geq 1$.

Fazendo a mudança de variáveis

$$a_t = u_{t+1} \quad \text{e} \quad y_t = x_{t+1} \quad \text{para} \quad t = 0, \dots, N-2,$$

nota-se que

$$\begin{aligned} \min_{u_1, \dots, u_{N-1}} E_1 \left(\sum_{t=1}^{N-1} \delta^{t-1} c(x_t, u_t) \right) &= \min_{a_0, \dots, a_{N-2}} E_0 \left(\sum_{t=0}^{N-2} \delta^t c(y_t, a_t) \right) \\ &= \min_{\pi_{N-1} \in \Pi_{N-1}} W_{N-1}(y_0, \pi_{N-1}) \\ &= V_{N-1}(y_0). \\ &= V_{N-1}(x_1) \end{aligned} \tag{3.7}$$

Assim, substituindo (3.7) em (3.6), e considerando a *lei de movimento* (3.2), segue a igualdade:

$$\begin{aligned} V_N(x_0) &= \min_{u_0} E_0 \left\{ c(x_0, u_0) + \delta \left[\min_{u_1, \dots, u_{N-1}} E_1 \left(\sum_{t=1}^{N-1} \delta^{t-1} c(x_t, u_t) \right) \right] \right\} \\ &= \min_{u_0} E_0 [c(x_0, u_0) + \delta V_{N-1}(x_1)] \\ &= \min_{u_0} E_0 [c(x_0, u_0) + \delta V_{N-1}(g(x_0, u_0, \epsilon_0))] \\ &= \min_{u_0} \{c(x_0, u_0) + \delta E_0 [V_{N-1}(g(x_0, u_0, \epsilon_0))]\}. \end{aligned} \tag{3.8}$$

Finalmente, retirando o subscrito de x , u e ϵ , e escrevendo a equação (3.8) restando j períodos para o fim do horizonte de programação, obtém-se o resultado enunciado:

$$V_j(x) = \min_u \{c(x, u) + \delta E [V_{j-1}(g(x, u, \epsilon))]\}.$$

■

Da equação de Bellman decorre o resultado fundamental da Programação Dinâmica, conhecido como **Princípio da Otimalidade** [15]:

“An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decisions.”

Baseado no Princípio de Otimalidade, o *algoritmo de programação dinâmica* oferece um meio de se obter a política ótima do problema (3.3)-(3.4). O procedimento começa pela construção da função valor ótimo para o último período de programação: $V_1(x) = \min_u c(x, u)$. Como o ponto de mínimo u^* depende de x , segue-se que $u^* = \eta_1^*(x)$. Então, resolvendo iterativamente a equação de Bellman (3.5), são obtidas as funções V_j e as regras de decisão η_j^* , para $j = 2, \dots, N$. Ao final, fazendo $\mu_{N-j}^* = \eta_j^*$, tem-se a política ótima $\pi_N^* = (\mu_0^*, \dots, \mu_{N-1}^*)$. Esse algoritmo pode ser expresso da seguinte maneira:

Algoritmo 3.1: Programação Dinâmica Discreta com Horizonte Finito

Entradas: N, δ, x_0, c (função perda), g (lei de movimento)

Saída : Política ótima $(\mu_0^*, \dots, \mu_{N-1}^*)$

Inicialização:

Encontre $u^* = \eta_1^*(x)$ que minimiza o valor de $c(x, u)$;

$V_1(x) := c(x, \eta_1^*(x))$;

$\mu_{N-1}^* := \eta_1^*$;

Para $j := 2, \dots, N$ **faça**

Encontre $u^* = \eta_j^*(x)$ que minimiza o valor de $c(x, u) + \delta E [V_{j-1}(g(x, u, \epsilon))]$;
$V_j(x) := c(x, \eta_j^*(x)) + \delta E [V_{j-1}(g(x, \eta_j^*(x), \epsilon))]$;
$\mu_{N-j}^* := \eta_j^*$;

Fim

3.2 Horizonte Infinito

Considere agora o seguinte problema de controle ótimo:

$$\min_{\{u_t\}_t^\infty} E_0 \left(\sum_{t=0}^{N-1} \delta^t c(x_t, u_t) \right) \quad (3.9)$$

$$\text{sujeito a } x_{t+1} = g(x_t, u_t, \epsilon_t) \quad (3.10)$$

Como esse problema é o mesmo em cada período, admitindo que ele possua uma solução, a sua função valor ótimo V e a regra de decisão $\mu = \mu(x)$ também serão as mesmas, independente do período. Assim, a equação de Bellman apresenta-se como a

equação funcional:

$$V(x) = \min_u \{c(x, u) + \delta E [V(g(x, u, \epsilon))]\}. \quad (3.11)$$

Observe que o algoritmo 3.1 não pode ser aplicado neste caso, haja vista a inexistência de um período final por onde iniciar as iterações. Todavia, considerando o problema com horizonte finito (3.1)-(3.2), é natural conjecturar que a sua função valor ótimo V_N converge para a solução V de (3.11) quando $N \rightarrow \infty$. A fim de investigar essa proposição, considere a equação de Bellman para horizonte finito:

$$V_j(x) = \min_u \{c(x, u) + \delta E [V_{j-1}(g(x, u, \epsilon))]\}. \quad (3.12)$$

Em relação ao conjunto $\mathbb{B} = \{\rho : X \rightarrow \mathbb{R}; \rho \text{ é contínua e limitada}\}$, é interessante analisar o operador $T : \mathbb{B} \rightarrow \mathbb{B}$ definido por:

$$(T\rho)(x) = \min_u \{c(x, u) + \delta E [\rho(g(x, u, \epsilon))]\}. \quad (3.13)$$

Neste caso, a equação funcional de Bellman (3.11) toma a forma

$$V = TV, \quad (3.14)$$

de modo que a sua solução pode ser vista como um ponto fixo do operador T .

Lema 3.5. *Com respeito à norma do sup, o operador T definido por (3.13) é uma δ -contração.*

Demonstração: Dado $x \in X$, seja \bar{u} um ponto de mínimo na definição de T . Então,

$$\begin{aligned} (T\rho)(x) &= c(x, \bar{u}) + \delta E [\rho(g(x, \bar{u}, \epsilon))] \\ &= c(x, \bar{u}) + \delta E [\gamma(g(x, \bar{u}, \epsilon))] + \delta E [\rho(g(x, \bar{u}, \epsilon)) - \gamma(g(x, \bar{u}, \epsilon))] \\ &\leq c(x, \bar{u}) + \delta E [\gamma(g(x, \bar{u}, \epsilon))] + \delta \|\rho - \gamma\| \\ &= \min_u \{c(x, u) + \delta E [\gamma(g(x, u, \epsilon))]\} + \delta \|\rho - \gamma\| \\ &= (T\gamma)(x) + \delta \|\rho - \gamma\| \end{aligned}$$

donde segue

$$(T\rho)(x) - (T\gamma)(x) \leq \delta \|\rho - \gamma\|.$$

De forma análoga, substituindo ρ por γ no argumento anterior, obtém-se

$$(T\gamma)(x) - (T\rho)(x) \leq \delta \|\gamma - \rho\|.$$

Logo,

$$\begin{aligned}
 & -\delta \|\rho - \gamma\| \leq (T\gamma)(x) - (T\rho)(x) \leq \delta \|\gamma - \rho\| \\
 \implies & \quad | (T\rho)(x) - (T\gamma)(x) | \leq \delta \|\rho - \gamma\| \\
 \implies & \quad \|T\rho - T\gamma\| \leq \delta \|\rho - \gamma\|
 \end{aligned}$$

provando assim o resultado enunciado. ■

Teorema 3.6. *Existe uma única função $V \in \mathbb{B}$ que satisfaz a equação funcional de Bellman (3.11). Além disso, a sequência de funções V_j definida em (3.12), com $V_0 = 0$, converge para V quando $j \rightarrow \infty$.*

Demonstração: Considerando o Lema 3.5, e o fato de \mathbb{B} , munido da norma do sup, ser um espaço métrico completo, decorre do Teorema do Ponto Fixo de Banach² que o operador T possui um único ponto fixo V , o que prova a existência e unicidade da solução de (3.11). Além disso, por esse mesmo teorema, partindo da função contínua limitada $V_0 = 0$, a sequência $V_{j+1} = TV_j = \min_u \{c(x, u) + \delta E [V_j(g(x, u, \epsilon))]\}$ converge para V quanto $j \rightarrow \infty$. ■

3.3 Problema do Regulador Linear-Quadrático

Esta seção aborda a classe de problemas de controle discretos caracterizados por uma função perda quadrática e uma lei de movimento linear. Essa especificação corresponde ao chamado *problema do regulador linear-quadrático* (LQR, do inglês, *Linear-Quadratic Regulator*). Mais especificamente, aplicando o método de Programação Dinâmica, são obtidas as soluções de várias versões desse problema. As demonstrações aqui apresentadas seguem o procedimento indicado em [46], e justificam-se pelo fato de não terem sido encontradas de forma detalhada nos livros-texto e artigos consultados.

²Ver Teorema 5.1-2 em [42].

3.3.1 Caso Determinístico

O problema do regulador linear-quadrático determinístico com fator de desconto consiste em:

$$\min_{\{u_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \delta^t (x_t' R x_t + u_t' Q u_t), \quad (3.15)$$

$$\text{sujeito a } x_{t+1} = A x_t + B u_t, \quad \text{com } x_0 \text{ dado}, \quad (3.16)$$

onde:

$x_t \in \mathbb{R}^m$ é o vetor de variáveis de estado no período t ;

$u_t \in \mathbb{R}^n$ é o vetor de variáveis de controle no período t ;

$R \in M(m \times m)$ é uma matriz simétrica semi-definida positiva;

$Q \in M(n \times n)$ é uma matriz simétrica definida positiva;

$A \in M(m \times m)$ é a matriz que atua sobre o vetor de estados na lei de movimento;

$B \in M(m \times n)$ é a matriz que atua sobre o vetor de controle na lei de movimento.

Seja V a função valor ótimo associada a este problema. Conforme exposto na seção anterior, a equação de Bellman correspondente é a seguinte³:

$$V(x_t) = \min_{u_t} \{x_t' R x_t + u_t' Q u_t + \delta V(x_{t+1})\} \quad (3.17)$$

Pelo Teorema 3.6, V pode ser obtida como $\lim_{j \rightarrow \infty} V_j(x_t) = V(x_t)$, onde as funções V_j são definidas por:

$$V_{j+1}(x_t) = \min_{u_t} \{x_t' R x_t + u_t' Q u_t + \delta V_j(x_{t+1})\}. \quad (3.18)$$

O seguinte teorema especifica o termo geral de $\{V_j\}$ de uma forma conveniente, permitindo que se determine a função valor ótimo V .

³Note que o operador esperança foi omitido da equação de Bellman, pois neste caso não há variáveis aleatórias envolvidas.

Teorema 3.7. Suponha $V_0 = 0$, e considere a sequência de matrizes simétricas,

$$P_{j+1} = R + \delta A' P_j A - \delta^2 A' P_j B (Q + \delta B' P_j B)^{-1} B' P_j A, \quad (3.19)$$

com $P_0 = 0$. Então, para todo $j \in \mathbb{N}$, tem-se

$$V_j(x_t) = x_t' P_j x_t. \quad (3.20)$$

Demonstração: A prova deste resultado é feita por indução. Primeiro, note que:

$$V_1(x_t) = \min_{u_t} \{x_t' R x_t + u_t' Q u_t\} \quad (3.21)$$

Diferenciando o lado direito de (3.21) e igualando o resultado à zero, pela simetria de Q obtém-se⁴:

$$\begin{aligned} & \frac{\partial x_t' R x_t}{\partial u_t} + \frac{\partial u_t' Q u_t}{\partial u_t} = 0 \\ \implies & (Q + Q') u_t = 0 \\ \implies & 2Q u_t = 0 \\ \implies & Q u_t = 0. \end{aligned} \quad (3.22)$$

A substituição de (3.22) em (3.21), resulta na seguinte igualdade:

$$V_1(x_t) = x_t' R x_t. \quad (3.23)$$

Como $P_0 = 0$, segue da equação (3.19) que $P_1 = R$. Logo, $V_1(x_t) = x_t' P_1 x_t$, ou seja, o resultado enunciado é verdadeiro para $j = 1$.

Agora, suponha que o resultado se verifique para $j = k$. Neste caso, tem-se o seguinte:

$$\begin{aligned} V_{k+1} &= \min_{u_t} \{x_t' R x_t + u_t' Q u_t + \delta V_k(x_{t+1})\} \\ &= \min_{u_t} \{x_t' R x_t + u_t' Q u_t + \delta (A x_t + B u_t)' P_k (A x_t + B u_t)\} \\ &= \min_{u_t} \{x_t' R x_t + u_t' Q u_t + \delta x_t' A' P_k A x_t + \delta x_t' A' P_k B u_t + \delta u_t' B' P_k A x_t + \delta u_t' B' P_k B u_t\} \\ &= \min_{u_t} \{x_t' R x_t + u_t' Q u_t + \delta x_t' A' P_k A x_t + 2\delta x_t' A' P_k B u_t + \delta u_t' B' P_k B u_t\}. \end{aligned} \quad (3.24)$$

⁴Daqui em diante, são utilizadas as seguintes regras para diferenciação de formas quadráticas:

$$\frac{\partial x' A x}{\partial x} = (A' + A)x, \quad \frac{\partial y' B z}{\partial y} = Bz \text{ e } \frac{\partial y' B z}{\partial z} = B'y.$$

Diferenciando o lado direito da igualdade (3.24) e igualando o resultado a zero, obtém-se:

$$\begin{aligned}
& \frac{\partial u_t' Q u_t}{\partial u_t} + 2\delta \frac{\partial x_t' A' P_k B u_t}{\partial u_t} + \delta \frac{\partial u_t' B' P_k B u_t}{\partial u_t} = 0 \\
\implies & 2Q u_t + 2\delta B' P_k A x_t + 2\delta B' P_k B u_t = 0 \\
\implies & Q u_t + \delta B' P_k A x_t + \delta B' P_k B u_t = 0 \\
\implies & u_t = -\delta(Q + \delta B' P_k B)^{-1} B' P_k A x_t
\end{aligned} \tag{3.25}$$

Então, substituindo (3.25) em (3.24), decorre que:

$$\begin{aligned}
V_{k+1} &= x_t' R x_t + \delta^2 x_t' A' P_k B (Q + \delta B' P_k B)^{-1} Q (Q + \delta B' P_k B)^{-1} B' P_k A x_t + \delta x_t' A' P_k A x_t \\
&\quad - 2\delta^2 x_t' A' P_k B (Q + \delta B' P_k B)^{-1} B' P_k A x_t \\
&\quad + \delta^2 x_t' A' P_k B (Q + \delta B' P_k B)^{-1} B' P_k B (Q + \delta B' P_k B)^{-1} B' P_k A x_t \\
&= x_t' \{ R + \delta^2 A' P_k B (Q + \delta B' P_k B)^{-1} Q (Q + \delta B' P_k B)^{-1} B' P_k A + \delta A' P_k A \\
&\quad - 2\delta^2 A' P_k B (Q + \delta B' P_k B)^{-1} B' P_k A \\
&\quad + \delta^2 A' P_k B (Q + \delta B' P_k B)^{-1} B' P_k B (Q + \delta B' P_k B)^{-1} B' P_k A \} x_t \\
&= x_t' \{ R + \delta A' P_k A + \delta^2 A' P_k B (Q + \delta B' P_k B)^{-1} B' P_k A \\
&\quad - 2\delta^2 A' P_k B (Q + \delta B' P_k B)^{-1} B' P_k A \} x_t \\
&= x_t' (R + \delta A' P_k A - \delta^2 A' P_k B (Q + B' P_k B)^{-1} B' P_k A) x_t \\
&= x_t' P_{k+1} x_t.
\end{aligned}$$

Portanto, o resultado enunciado é verdadeiro para todo $j \in \mathbb{N}$. ■

Observação 3.8. A equação (3.19) é denominada equação de diferenças matricial de Riccati. Se a sequência $\{P_j\}_{j=0}^{\infty}$ gerada por ela, com $P_0 = 0$, satisfaz $\lim_{j \rightarrow \infty} P_j = P$, então a função valor ótimo V que satisfaz a equação de Bellman (3.17) é dada por $V(x_t) = x_t' P x_t$. Assim, seguindo o mesmo argumento que conduziu à expressão (3.25), conclui-se que a regra de decisão ótima do problema do regulador linear-quadrático (3.15)-(3.16) é a seguinte:

$$u_t = -F x_t = -\delta(Q + \delta B' P B)^{-1} B' P A x_t, \tag{3.26}$$

3.3.2 Caso Determinístico Modificado

Considere a seguinte versão do problema do regulador linear-quadrático determinístico:

$$\min_{\{u_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \delta^t (x_t' R x_t + 2x_t' W u_t + u_t' Q u_t), \tag{3.27}$$

$$\text{sujeito a } x_{t+1} = Ax_t + Bu_t, \quad (3.28)$$

onde x_0 é dado e $0 < \delta < 1$. Conforme apontado por Pedersen [57], este problema pode ser facilmente transformado no problema anterior, eliminando-se o produto cruzado $2x'_t W u_t$ em (3.27). Para isso, considere a variável:

$$\hat{u}_t = u_t + Q^{-1}W'x_t. \quad (3.29)$$

Definindo $\bar{R} = R - WQ^{-1}W'$, segue que:

$$\begin{aligned} x'_t \bar{R} x_t + \hat{u}'_t Q \hat{u}_t &= x'_t (R - WQ^{-1}W') x_t + (u_t + Q^{-1}W'x_t)' Q (u_t + Q^{-1}W'x_t) \\ &= x'_t R x_t - x'_t W Q^{-1} W' x_t + u'_t Q u_t + u'_t Q Q^{-1} W' x_t \\ &\quad + x'_t W Q^{-1} Q u_t + x'_t W Q^{-1} Q Q^{-1} W' x_t \\ &= x'_t R x_t + 2x'_t W u_t + u'_t Q u_t. \end{aligned}$$

Por outro lado, considerando a lei de movimento (3.28) e definindo $\bar{A} = A - BQ^{-1}W'$, obtém-se:

$$\begin{aligned} \bar{A}x_t + B\hat{u}_t &= (A - BQ^{-1}W')x_t + B(u_t + Q^{-1}W'x_t) \\ &= Ax_t - BQ^{-1}W'x_t + Bu_t + BQ^{-1}W'x_t \\ &= Ax_t + Bu_t \\ &= x_{t+1}. \end{aligned}$$

Dessa forma, o problema do regulador (3.27)-(3.28) é reduzido a:

$$\min_{\{\hat{u}_t\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \{x'_t \bar{R} x_t + \hat{u}'_t Q \hat{u}_t\}, \quad (3.30)$$

$$\text{sujeito a } x_{t+1} = \bar{A}x_t + B\hat{u}_t. \quad (3.31)$$

Portanto, pela Observação 3.8, a regra de decisão ótima é dada por:

$$\hat{u}_t = -Fx_t = \delta(Q + \delta B'PB)^{-1}B'P\bar{A}x_t, \quad (3.32)$$

onde P é o limite das iterações sobre a equação de diferenças matricial de Riccati:

$$P_{j+1} = \bar{R} + \delta \bar{A}' P_j \bar{A} - \delta^2 \bar{A}' P_j B (Q + \delta B' P_j B)^{-1} B' P_j \bar{A}, \quad (3.33)$$

començando por $P_0 = 0$.

A fim de obter a regra ótima em termos de u_t , primeiro, note que:

$$\begin{aligned} \hat{u}_t &= -Fx_t \\ \implies u_t + Q^{-1}W'x_t &= -Fx_t \\ \implies u_t &= -(F + Q^{-1}W')x_t. \end{aligned} \quad (3.34)$$

onde

$$\begin{aligned}
& -(F + Q^{-1}W') \\
&= \delta(Q + \delta B'PB)^{-1}B'P\bar{A} + QW' \\
&= \delta(Q + \delta B'PB)^{-1}B'P(A - BQ^{-1}W') + Q^{-1}W' \\
&= \delta(Q + \delta B'PB)^{-1}B'PA - \delta(Q + \delta B'PB)^{-1}B'PBQ^{-1}W' + Q^{-1}W' \\
&= \delta(Q + \delta B'PB)^{-1}B'PA - (Q + \delta B'PB)^{-1}\delta B'PBQ^{-1}W' + Q^{-1}W' \quad (3.35)
\end{aligned}$$

Por outro lado,

$$\begin{aligned}
& -(Q + \delta B'PB)^{-1}\delta B'PBQ^{-1}W' + Q^{-1}W' \\
&= -(Q + \delta B'PB)^{-1}(QQ^{-1}W' + \delta B'PBQ^{-1}W' - QQ^{-1}W') + Q^{-1}W' \\
&= -(Q + \delta B'PB)^{-1}[(Q + \delta B'PB)Q^{-1}W' - W'] + Q^{-1}W' \\
&= -Q^{-1}W' + (Q + \delta B'PB)W' + Q^{-1}W' \\
&= (Q + \delta B'PB)W'. \quad (3.36)
\end{aligned}$$

Finalmente, substituindo (3.36) em (3.35), a equação (3.34) pode ser reescrita como:

$$\begin{aligned}
u_t &= - \left[\delta(Q + \delta B'PB)^{-1}B'PA + (Q + \delta B'PB)W' \right] x_t \\
&= - \left[(Q + \delta B'PB)^{-1}\delta B'PA + (Q + \delta B'PB)W' \right] x_t \\
&= -(Q + \delta B'PB)^{-1}(\delta B'PA + W')x_t.
\end{aligned}$$

3.3.3 Caso Estocástico

O problema do regulador linear-quadrático estocástico pode ser estabelecido da seguinte maneira:

$$\min_{\{u_t\}_{t=0}^{\infty}} E \left(\sum_{t=0}^{\infty} (x_t'Rx_t + u_t'Qu_t) \right), \quad (3.37)$$

$$\text{sujeito a } x_{t+1} = Ax_t + Bu_t + C\epsilon_t, \quad (3.38)$$

onde x_0 é dado e ϵ_t são variáveis aleatórias independentes, tomando valores em \mathbb{R}^m , com média zero e matriz de variância-covariância $E\epsilon_t\epsilon_t' = I_n$. As matrizes R, Q, A e B obedecem as mesmas hipóteses consideradas no caso determinístico. A equação de

Bellman para este problema é a seguinte:

$$V(x_t) = \min_{u_t} \{x_t' R x_t + u_t' Q u_t + \delta E [V(x_{t+1})]\} \quad (3.39)$$

Novamente, para se obter a função valor ótimo V , considera-se a sequência de funções $\{V_j\}$ definida por:

$$V_{j+1}(x_t) = \min_{u_t} \{x_t' R x_t + u_t' Q u_t + \delta E [V_j(x_{t+1})]\} \quad (3.40)$$

Nesse contexto, tem-se o seguinte teorema.

Teorema 3.9. *Suponha $V_0 = 0$, e considere as sequências:*

$$P_{j+1} = R + \delta A' P_j A - \delta^2 A' P_j B (Q + \delta B' P_j B)^{-1} B' P_j A, \quad \text{com } P_0 = 0 \quad (3.41)$$

$$\rho_{j+1} = \delta \rho_j + \delta \text{tr} (P_j C C'), \quad \text{com } \rho_0 = 0 \quad (3.42)$$

Além disso, suponha que as matrizes P_j sejam simétricas. Então, para todo $j \in \mathbb{N}$,

$$V_j(x_t) = x_t' P_j x_t + \rho_j \quad (3.43)$$

Demonstração: Também neste caso, a prova é feita por indução. Partindo da hipótese $V_0 = 0$, a equação (3.40) resulta em:

$$V_1(x_t) = \min_{u_t} \{x_t' R x_t + u_t' Q u_t\} \quad (3.44)$$

Diferenciando o lado direito de (3.44) e igualando o resultado à zero, pela simetria de Q obtém-se:

$$\begin{aligned} & \frac{\partial x_t' R x_t}{\partial u_t} + \frac{\partial u_t' Q u_t}{\partial u_t} = 0 \\ \implies & (Q + Q') u_t = 0 \\ \implies & 2Q u_t = 0 \\ \implies & Q u_t = 0. \end{aligned} \quad (3.45)$$

A substituição de (3.45) em (3.44), implica a seguinte igualdade:

$$V_1(x_t) = x_t' R x_t. \quad (3.46)$$

Como $P_0 = 0$, segue da equação (3.41) que $P_1 = R$. Por outro lado, considerando $\rho_0 = 0$ em (3.42), tem-se $\rho_1 = 0$. Portanto,

$$V_1(x_t) = x_t' P_1 x_t + \rho_1, \quad (3.47)$$

ou seja, o resultado enunciado é verdadeiro para $j = 1$.

Agora, suponha que o resultado seja verdadeiro para um natural $j = k$. Nesse caso,

$$\begin{aligned}
V_{k+1}(x_t) &= \min_{u_t} \{x_t' R x_t + u_t' Q u_t + \delta E [V_k(x_{t+1})]\} \\
&= \min_{u_t} \{x_t' R x_t + u_t' Q u_t + \delta E [x_{t+1}' P_k x_{t+1} + \rho_k]\} \\
&= \min_{u_t} \{x_t' R x_t + u_t' Q u_t + \delta E [x_{t+1}' P_k x_{t+1}] + \delta \rho_k\}. \tag{3.48}
\end{aligned}$$

Por outro lado,

$$\begin{aligned}
E [x_{t+1}' P_k x_{t+1}] &= E [(A x_t + B u_t + C \epsilon_t)' P_k (A x_t + B u_t + C \epsilon_t)] \\
&= E \{x_t' A' P_k A x_t + x_t' A' P_k B u_t + x_t' A' P_k C \epsilon_t \\
&\quad + u_t' B' P_k A x_t + u_t' B' P_k B u_t + u_t' B' P_k C \epsilon_t \\
&\quad + \epsilon_t' C' P_k A x_t + \epsilon_t' C' P_k B u_t + \epsilon_t' C' P_k C \epsilon_t\} \\
&= x_t' A' P_k A x_t + x_t' A' P_k B u_t + u_t' B' P_k A x_t + u_t' B' P_k B u_t \\
&\quad + E [\epsilon_t' C' P_k C \epsilon_t] \\
&= x_t' A' P_k A x_t + 2x_t' A' P_k B u_t + u_t' B' P_k B u_t + \text{tr} (P_k C C') \tag{3.49}
\end{aligned}$$

Substituindo (3.49) em (3.48), obtém-se:

$$\begin{aligned}
V_{k+1}(x_t) &= \min_{u_t} \{x_t' R x_t + u_t' Q u_t + \delta x_t' A' P_k A x_t + 2\delta x_t' A' P_k B u_t \\
&\quad + \delta u_t' B' P_k B u_t + \delta \text{tr} (P_k C C') + \delta \rho_k\} \tag{3.50}
\end{aligned}$$

Então, diferenciando o lado direito da igualdade (3.50) e igualando o resultado a zero, decorre que:

$$\begin{aligned}
&\frac{\partial u_t' Q u_t}{\partial u_t} + 2\delta \frac{\partial x_t' A' P_k B u_t}{\partial u_t} + \delta \frac{\partial u_t' B' P_k B u_t}{\partial u_t} = 0 \\
\implies &2Q u_t + 2\delta B' P_k A x_t + 2\delta B' P_k B u_t = 0 \\
\implies &Q u_t + \delta B' P_k A x_t + \delta B' P_k B u_t = 0 \\
\implies &u_t = -\delta (Q + \delta B' P_k B)^{-1} B' P_k A x_t \tag{3.51}
\end{aligned}$$

Por sua vez, a substituição de (3.51) na equação (3.50), culmina em:

$$\begin{aligned}
V_{k+1} &= x_t' R x_t + \delta^2 x_t' A' P_k B (Q + \delta B' P_k B)^{-1} Q (Q + \delta B' P_k B)^{-1} B' P_k A x_t + \delta x_t' A' P_k A x_t \\
&\quad - 2\delta^2 x_t' A' P_k B (Q + \delta B' P_k B)^{-1} B' P_k A x_t \\
&\quad + \delta^2 x_t' A' P_k B (Q + \delta B' P_k B)^{-1} B' P_k B (Q + \delta B' P_k B)^{-1} B' P_k A x_t + \delta \text{tr}(P_k C C') + \delta \rho_k \\
&= x_t' \{ R + \delta^2 A' P_k B (Q + \delta B' P_k B)^{-1} Q (Q + \delta B' P_k B)^{-1} B' P_k A + \delta A' P_k A \\
&\quad - 2\delta^2 A' P_k B (Q + \delta B' P_k B)^{-1} B' P_k A \\
&\quad + \delta^2 A' P_k B (Q + \delta B' P_k B)^{-1} B' P_k B (Q + \delta B' P_k B)^{-1} B' P_k A \} x_t + \delta \text{tr}(P_k C C') + \delta \rho_k \\
&= x_t' \{ R + \delta A' P_k A + \delta^2 A' P_k B (Q + \delta B' P_k B)^{-1} B' P_k A \\
&\quad - 2\delta^2 A' P_k B (Q + \delta B' P_k B)^{-1} B' P_k A \} x_t + \delta \text{tr}(P_k C C') + \delta \rho_k \\
&= x_t' (R + \delta A' P_k A - \delta^2 A' P_k B (Q + B' P_k B)^{-1} B' P_k A) x_t + \delta \rho_k + \delta \text{tr}(P_k C C') \\
&= x_t' P_{k+1} x_t + \rho_{k+1}.
\end{aligned}$$

Logo, o resultado enunciado é verdadeiro para todo $j \in \mathbb{N}$. ■

Observação 3.10. *Supondo que existam os limites $P = \lim_{j \rightarrow \infty} P_j$, com $P_0 = 0$, e $\rho = \lim_{j \rightarrow \infty} \rho_j$, com $\rho_0 = 0$, segue que $V(x_t) = \lim_{j \rightarrow \infty} V_j(x_t) = x_t' P x_t + \rho$. Com esta função valor ótimo, pelo mesmo procedimento que resultou em (3.51), conclui-se que a regra de decisão ótima para o problema (3.37)-(3.38) é dada por:*

$$u_t = -F x_t = -\delta(Q + \delta B' P B)^{-1} B' P A x_t \quad (3.52)$$

3.3.4 Caso Estocástico Modificado

Considere a seguinte versão do problema do regulador linear-quadrático estocástico:

$$\min_{\{u_t\}_{t=0}^{\infty}} E \left(\sum_{t=0}^{\infty} \delta^t (x_t' R x_t + 2x_t' W u_t + u_t' Q u_t) \right), \quad (3.53)$$

$$\text{sujeito a } x_{t+1} = A x_t + B u_t + C \epsilon_t, \quad (3.54)$$

onde são satisfeitas as mesmas condições que definem o problema anterior. Nesse caso, o produto cruzado $2x_t' W u_t$ pode ser eliminado utilizando-se variáveis iguais àquelas empregadas no caso determinístico modificado. Com efeito, definido

$$\hat{u}_t = u_t + Q^{-1} W' x_t, \text{ e}$$

$$\bar{R} = R - W Q^{-1} W',$$

segue-se que:

$$\begin{aligned}
x_t' \bar{R} x_t + \hat{u}_t' Q \hat{u}_t &= x_t' (R - WQ^{-1}W') x_t + (u_t + Q^{-1}W'x_t)' Q (u_t + Q^{-1}W'x_t) \\
&= x_t' R x_t - x_t' WQ^{-1}W'x_t + u_t' Q u_t + u_t' Q Q^{-1}W'x_t \\
&\quad + x_t' WQ^{-1}Q u_t + x_t' WQ^{-1}Q Q^{-1}W'x_t \\
&= x_t' R x_t + 2x_t' W u_t + u_t' Q u_t.
\end{aligned}$$

Por outro lado, considerando a lei de movimento (3.54) e definindo $\bar{A} = A - BQ^{-1}W'$, obtém-se:

$$\begin{aligned}
\bar{A}x_t + B\hat{u}_t + C\epsilon_t &= (A - BQ^{-1}W')x_t + B(u_t + Q^{-1}W'x_t) + C\epsilon_t \\
&= Ax_t - BQ^{-1}W'x_t + Bu_t + BQ^{-1}W'x_t + C\epsilon_t \\
&= Ax_t + Bu_t + C\epsilon \\
&= x_{t+1}.
\end{aligned}$$

Dessa forma, o problema do regulador (3.53)-(3.54) é reduzido ao seguinte problema:

$$\min_{\{\hat{u}_t\}_{t=0}^{\infty}} E \left(\sum_{t=0}^{\infty} \delta^t (x_t' \bar{R} x_t + \hat{u}_t' Q \hat{u}_t) \right), \quad (3.55)$$

$$\text{sujeito a } x_{t+1} = \bar{A}x_t + B\hat{u}_t + C\epsilon_t. \quad (3.56)$$

Como consequência, pela Observação 3.10, a regra de decisão ótima é dada por:

$$\hat{u}_t = -Fx_t = \delta(Q + \delta B'PB)^{-1} B'P\bar{A}x_t, \quad (3.57)$$

onde P o limite das iterações sobre a equação de diferenças matricial de Riccati:

$$P_{j+1} = \bar{R} + \delta \bar{A}' P_j \bar{A} - \delta^2 \bar{A}' P_j B (Q + \delta B' P_j B)^{-1} B' P_j \bar{A}, \quad (3.58)$$

começando por $P_0 = 0$.

Para obter a regra ótima em termos de u_t basta notar que:

$$\begin{aligned}
\hat{u}_t &= -Fx_t \\
\implies u_t + Q^{-1}W'x_t &= -Fx_t \\
\implies u_t &= -(F + Q^{-1}W')x_t.
\end{aligned}$$

Assim, pelo mesmo procedimento aplicado ao caso determinístico modificado, obtém-se:

$$u_t = -(Q + \delta B'PB)^{-1} (\delta B'PA + W')x_t.$$

Observação 3.11 (Princípio de Equivalência da Certeza). *Observe que em ambos os casos, determinístico e estocástico, a regra ótima é a mesma:*

(a) *Sem o produto cruzado $2x_t'Wu_t$, a regra é $u_t = -\delta(Q + \delta B'P_k B)^{-1}B'P_k Ax_t$;*

(b) *Com o produto cruzado $2x_t'Wu_t$, a regra é $u_t = -(Q + \delta B'PB)^{-1}(\delta B'PA + W')x_t$.*

Portanto, o problema do regulador linear-quadrático estocástico pode ser resolvido como se fosse o problema determinístico correspondente. Esse fato é conhecido como o Princípio da equivalência da Certeza.

3.4 Equação de Riccati

De acordo com o que foi exposto ao longo desta seção, pela aplicação do algoritmo de programação dinâmica, o problema do regulador linear-quadrático se reduz ao problema de se determinar o limite da sequência $\{P_j\}$, definida pela equação de diferenças matricial de Riccati:

$$P_{j+1} = R + \delta A'P_j A - \delta^2 A'P_j B(Q + \delta B'P_j B)^{-1}B'P_j A, \quad (3.59)$$

Com base em Bertsekas [16], esta subseção descreve o comportamento assintótico da sequência de matrizes $\{P_j\}$ gerada pela equação (3.59), fornecendo um algoritmo para calcular o seu limite. Para isto, no que segue, são introduzidas algumas noções essenciais da Teoria do Controle.

Definição 3.12. *Dadas as matrizes $A \in M(n \times n)$ e $B \in M(n \times m)$, diz-se que o par (A, B) é controlável se*

$$\left(B, AB, A^2B, \dots, A^{n-1}B \right) \in M(n \times n.m),$$

tem posto completo.

Observe que, se um par (A, B) é controlável então para qualquer estado inicial x_0 , existe uma sequência de controles u_0, \dots, u_{n-1} que conduz o sistema,

$$x_{t+1} = Ax_t + Bu_t \quad (3.60)$$

ao estado $x_n = 0$. De fato, a partir de (3.60), obtém-se de forma indutiva

$$\begin{aligned}
 x_1 &= Ax_0 + Bu_0 \\
 \implies x_2 &= A^2x_0 + Bu_1 + ABu_0 \\
 \implies x_3 &= A^3x_0 + Bu_2 + ABu_1 + A^2Bu_0 \\
 &\vdots \\
 \implies x_n &= A^n x_0 + Bu_{n-1} + ABu_{n-2} + \dots + A^{n-1}Bu_0
 \end{aligned}$$

ou equivalentemente,

$$x_n - A^n x_0 = (B, AB, \dots, A^{n-1}B) \begin{pmatrix} u_{n-1} \\ u_{n-2} \\ \vdots \\ u_0 \end{pmatrix}. \quad (3.61)$$

Assim, se (A, B) é controlável então a matriz $(B, AB, \dots, A^{n-1}B)$ tem posto completo, o que garante a existência e unicidade da solução do sistema linear (3.61), seja qual for o vetor $x_n - A^n x_0$. Em particular, pode-se obter uma solução $(u_0, u_1, \dots, u_{n-1})$ de modo que o lado esquerdo de (3.61) seja igual à $-A^n x_0$, o que corresponde ao caso $x_n = 0$.

Definição 3.13. Dadas as matrizes $A \in M(n \times n)$ e $C \in M(n \times m)$, diz-se que o par (A, C) é observável se o par correspondente (A', C') é controlável.

A noção de observabilidade implica que, dadas medidas z_0, \dots, z_{n-1} , tais que $z_t = Cx_t$, é possível inferir o estado inicial x_0 do sistema $x_{t+1} = Ax_t$, tendo em vista a relação,

$$\begin{pmatrix} z_{n-1} \\ \vdots \\ z_1 \\ z_0 \end{pmatrix} = \begin{pmatrix} CA^{n-1} \\ \vdots \\ CA \\ C \end{pmatrix} x_0.$$

Teorema 3.14. Sejam A uma matriz em $M(m \times m)$, B uma matriz em $M(m \times n)$, R uma matriz simétrica semi-definida positiva em $M(m \times m)$ e Q uma matriz simétrica definida positiva em $M(n \times n)$. Além disso, assuma que os pares (A, B) e (A, C) sejam, respectivamente, controlável e observável, onde $C'C = R$. Então, com respeito à equação de diferenças de Riccati

$$P_{j+1} = A'(P_j - P_j B(Q + B'P_j B)^{-1} B'P_j)A + R, \quad j = 0, 1, 2, \dots \quad (3.62)$$

existe uma matriz P simétrica definida positiva, tal que, para toda matriz inicial P_0 simétrica semi-definida positiva tem-se $\lim_{j \rightarrow \infty} P_j = P$.

Demonstração: Ver Proposição 4.1 em [16] ■

Voltando agora à equação (3.59), defina as matrizes $\tilde{Q} = Q/\delta$ e $\tilde{A} = \sqrt{\delta}A$. Desta forma, note que:

$$\begin{aligned}
 P_{j+1} &= R + \delta A' P_j A - \delta^2 A' P_j B (Q + \delta B' P_j B)^{-1} B' P_j A \\
 &= R + \delta \left(\frac{\tilde{A}}{\sqrt{\delta}} \right)' P_j \left(\frac{\tilde{A}}{\sqrt{\delta}} \right) - \delta^2 \left(\frac{\tilde{A}}{\sqrt{\delta}} \right)' P_j B (\delta \tilde{Q} + \delta B' P_j B)^{-1} B' P_j \left(\frac{\tilde{A}}{\sqrt{\delta}} \right) \\
 &= R + \tilde{A} P_j \tilde{A} - \delta \tilde{A} P_j B (\delta (\tilde{Q} + B' P_j B))^{-1} B' P_j \tilde{A} \\
 &= R + \tilde{A} P_j \tilde{A} - \tilde{A} P_j B (\tilde{Q} + B' P_j B)^{-1} B' P_j \tilde{A} \\
 &= \tilde{A}' (P_j - P_j B (\tilde{Q} + B' P_j B)^{-1} B' P_j) \tilde{A} + R,
 \end{aligned}$$

ou seja, pode-se reduzir (3.59) a uma equação do tipo (3.62). Além disso, a partir da definição $\tilde{A} = \sqrt{\delta}A$ tem-se que a controlabilidade de (A, B) implica a controlabilidade de (\tilde{A}, B) , enquanto a observabilidade de (A, C) implica a observabilidade de (\tilde{A}, C) . Portanto, a matriz P que define a função valor ótimo V pode ser obtida como o limite da sequência $\{P_j\}$ gerada pela equação (3.59), partindo de qualquer matriz P_0 simétrica semi-definida positiva. Esse fato motiva o seguinte algoritmo⁵:

Algoritmo 3.2: Método Iterativo para resolver a equação matricial de Riccati

Entradas: δ, A, B, R, Q e W (Parâmetros que definem a equação)

Saída : P (Solução da equação)

Inicialização: $erro := 1, P_0 := I_m, k := 1, tol := 10^{-9}, maxit := 5000$;

Enquanto $erro > tol$ e $k < maxit$ **faça**

$$\left| \begin{array}{l}
 P = R + \delta A' P_0 A - \delta^2 A' P_0 B (Q + \delta B' P_0 B)^{-1} B' P_0 A; \\
 erro := \max\{|(P)_{ij} - (P_0)_{ij}|; i, j = 1, \dots, m\}; \\
 P_0 = P; \\
 k := k + 1;
 \end{array} \right.$$

Fim

⁵O código MATLAB para esse algoritmo pode ser encontrado no Capítulo 9 em [57].

Algoritmos Genéticos

O presente capítulo oferece uma breve introdução aos Algoritmos Genéticos, tendo por base os trabalhos de Holland [35], Haupt [33], Alvarez [5] e Liu *et al.* [47]. O objetivo é descrever os elementos e o funcionamento do algoritmo utilizado neste trabalho para estimar as preferências monetárias do Banco Central do Brasil.

4.1 Fundamentos Biológicos

Algoritmos Genéticos (AGs) são heurísticas destinadas à resolução de problemas de otimização. Eles foram introduzidos de forma sistemática por Holland [35], como uma tentativa de imitar computacionalmente o processo de evolução natural das espécies. Em razão disso, esta seção aborda alguns termos e conceitos da Biologia, essenciais para o bom entendimento dos algoritmos genéticos.

4.1.1 Genética

Uma *célula* é um complexo de estruturas moleculares que compõe um todo capaz de produzir cópias de si mesmo. As informações que regem o seu funcionamento são armazenadas nos *cromossomos*, os quais são constituídos por moléculas de DNA (*Deoxyribonucleic Acid*). Por sua vez, o DNA consiste em uma sequência de *genes*, cada

um dos quais codificando uma certa característica da espécie. Para um indivíduo particular, a combinação completa dos seus genes é chamada de *genótipo*, enquanto as características observáveis associadas a esses genes são referidas como *fenótipo*.

A reprodução celular pode ocorrer de duas maneiras distintas, denominadas *mitose* e *meiose*. Na mitose, a mesma informação genética é copiada em uma nova célula. Esse processo é o responsável pelo crescimento de estruturas pluricelulares, tais como os órgãos. Por sua vez, a meiose promove a divisão da célula em *gametas*, cada um dos quais contendo a metade dos cromossomos da célula original. Nesse caso, dado um par de células-mãe, a partir da união dos seus gametas origina-se uma célula-filha, com a mesma quantidade de cromossomos das células genitoras [18].

4.1.2 Seleção Natural

Na natureza, os seres vivos estão em constante competição pelos recursos do meio em que vivem, tais como: água, comida, abrigo, luz, etc. Além disso, dentro de cada espécie, existe ainda a competição por parceiros para a reprodução. Como consequência disso, indivíduos com baixo desempenho tem menos chance de sobreviver, enquanto os indivíduos mais “aptos” produzem um número relativamente maior de descendentes. Este fato, conhecido como Princípio da Seleção Natural, foi estabelecido por Charles Darwin, e constitui a base para a Teoria da Evolução Natural.

4.2 Algoritmos Genéticos com Codificação Binária

Dada uma função $f : X \rightarrow \mathbb{R}$, um problema de otimização consiste na busca por um elemento $\bar{x} \in \Omega \subset X$, tal que $f(\bar{x}) \geq f(x)$ para todo $x \in X$. De forma concisa, é usual expressar esse tipo de problema da seguinte maneira:

$$\max_x f(x) \tag{4.1}$$

$$\text{sujeito a } x \in \Omega. \tag{4.2}$$

Nesse caso, diz-se que f é a função objetivo, enquanto $\Omega \subset X$ é o espaço de busca¹

Para descrever a aplicação dos algoritmos genéticos em problemas de otimização, primeiro firma-se a analogia entre os conceitos matemáticos e biológicos. Cada candidata a solução do problema (4.1)-(4.2) é considerada um indivíduo, sendo expressa por uma sequência de símbolos (genes), denominada cromossomo. A forma inicial de uma possível solução recebe a nome de fenótipo, enquanto a sua forma codificada em cromossomo caracteriza o seu genótipo.

Os algoritmos genéticos atuam sobre um conjunto de N indivíduos (população), e as iterações que os fazem progredir são chamadas de gerações. Em cada uma dessas gerações são executadas as operações genéticas referidas como seleção, recombinação e mutação. Estas operações fazem com que as soluções que compõe a população se aproximem, em média, do ponto de máximo da função objetivo f , de maneira similar ao processo de seleção natural, pelo qual as populações biológicas tendem a se tornar mais adaptadas ao meio em que vivem.

Inicialmente, os cromossomos são avaliados segundo a função objetivo f . A partir dessa avaliação, obtém-se a aptidão de cada indivíduo, com base na qual a seleção é realizada. Com o intuito de imitar a seleção natural, essa operação consiste na escolha dos indivíduos que irão permanecer na população, e tem por diretriz o seguinte critério: quanto maior a aptidão de um indivíduo, mais chance ele tem de ser copiado e assim se propagar para próxima geração. Feita a seleção, os indivíduos são codificados e ficam sujeitos à recombinação e à mutação, com probabilidades p_c e p_m respectivamente. Inspirada no processo de meiose celular, a recombinação organiza a população em pares, e promove a mistura de genes entre os seus cromossomos. Em seguida, a mutação promove a troca aleatória de genes em determinadas posições do cromossomo. Terminado este ciclo, os indivíduos são decodificados (isto é, retornam à forma original) e passam a compor a nova população, a qual será submetida aos mesmos procedimentos na geração seguinte. Este processo é repetido até que um certo critério de parada seja satisfeito, e pode ser sintetizado da seguinte maneira:

¹Minimizar uma função f é o mesmo que maximizar a função $-f$. Logo, todo o problema de otimização pode ser expresso no formato (4.1)-(4.2).

Algoritmo 4.1: Algoritmo Genético para solução de um problema de otimização

Entrada: p_c, p_m, N (Parâmetros que definem o AG)

Saída : M (Melhor candidata a solução obtida ao final do processo)

Inicialização: $k := 1$;

Escolha em Ω uma população inicial P_0 de tamanho N ;

Calcule a aptidão dos indivíduos em P_0 ;

$M :=$ indivíduo mais apto em P_0 ;

Enquanto o critério de parada não for satisfeito **repita**

 Selecione P a partir de P_0 ;

 Codifique P ;

 Aplique sobre P o operador recombinação com probabilidade p_c ;

 Aplique sobre P o operador mutação com probabilidade p_m ;

 Decodifique P e calcule a aptidão dos seus indivíduos;

$P_0 := P$;

$M :=$ indivíduo mais apto em P_0 ;

Fim

4.2.1 Codificação

O projeto de um algoritmo genético começa com a decisão sobre como codificar as soluções em cromossomos. Isso pode ser feito por meio de sequências binárias, vetores do espaço euclidiano, ou até mesmo listas compostas de números e símbolos. O tipo de codificação depende principalmente das características do problema a ser resolvido.

A codificação binária, como o nome sugere, consiste em representar as variáveis do espaço de busca Ω usando sequências de números binários. Por exemplo, no caso do vetor $(18, 3, 1, 4)$, convertendo cada uma das suas coordenadas em um número binário de tamanho cinco, obtém-se o seguinte cromossomo:

$$\left(10010:00011:00001:00100 \right)$$

Para efeitos didáticos, o restante desta seção descreve os procedimentos referentes aos AGs com codificação binária. Não obstante, a maioria deles pode ser utilizada em AGs com outras codificações.

4.2.2 Seleção

Após definir o tipo de codificação, o próximo passo é estabelecer um critério de seleção, ou seja, a forma de escolha dos indivíduos que irão permanecer na população para as fases seguintes de recombinação e mutação. Essa escolha se dá com base no valor da aptidão de cada indivíduo, o qual é calculada à partir da função objetivo do problema em análise. A proposta da seleção é fazer com que os indivíduos mais aptos tenham mais chance de permanecer na população, tal como preconiza a teoria de Darwin no caso dos seres vivos.

A intensidade com que os indivíduos mais aptos são favorecidos na seleção, denominada *pressão seletiva*, está diretamente relacionado com a taxa de convergência de um algoritmo genético. Uma pressão seletiva muito fraca pode implicar em um tempo desnecessariamente longo para que a solução ótima seja encontrada. Por outro lado, uma pressão muito forte, pode levar a uma perda precoce da diversidade da população, acarretando a convergência sobre um máximo local (solução sub-ótima).

Método da Roleta

Este é o procedimento de seleção tradicional dos algoritmos genéticos. Neste método, a probabilidade de seleção de cada indivíduo é dada por:

$$P_i = \frac{a_i}{\sum_{j=1}^N a_j}, \quad (4.3)$$

onde,

P_i é a probabilidade de seleção do i -ésimo indivíduo;

a_j é a aptidão do j -ésimo indivíduo.

N é o tamanho da população;

Portanto, quanto maior for a aptidão relativa de um indivíduo maior será a sua chance de deixar cópias. Como são produzidas N cópias a cada geração, o número de cópias esperadas para o i -ésimo indivíduo é igual ao produto NP_i .

O nome deste método deriva do fato dele poder ser entendido considerando-se uma roleta de cassino hipotética, na qual cada indivíduo corresponde à uma casa, cuja área é proporcional à sua probabilidade de seleção. Nesse caso, a roleta é “girada” N vezes, sendo selecionados os indivíduos que forem sorteados a cada rodada.

Seleção por Torneio

O Método da Roleta, embora simples de implementar, tem a desvantagem de impor uma forte pressão seletiva, o que pode induzir a convergência prematura do AG para uma solução sub-ótima. Uma forma de contornar esse problema é a Seleção por Torneio. Esse método executa a seleção realizando uma competição entre N' indivíduos da população escolhidos aleatoriamente. Define-se o *vencedor* como sendo o indivíduo que apresenta a maior aptidão no grupo sorteado. Assim, seleciona-se o vencedor de cada torneio, os quais são repetidos até que se chegue ao total de N indivíduos.

Elitismo

Tanto no Método da Roleta quanto na Seleção por Torneio, a probabilidade de que os melhores cromossomos sejam copiados é muito alta. Entretanto, ainda assim existe a chance deles serem descartados da população. Para evitar esse tipo de perda indesejável, o procedimento denominado *Elitismo* pode ser utilizado. Ele consiste em copiar diretamente os melhores indivíduos, sem submetê-los ao processo padrão de seleção. Uma vez garantida a permanência dos indivíduos mais aptos, o critério de seleção é aplicado normalmente sobre a população².

²Para evitar uma pressão seletiva excessiva, pode-se restringir a seleção aos indivíduos que não foram favorecidos pelo elitismo.

4.2.3 Recombinação

Feita a seleção, a população resultante é submetida à operação de recombinação. Essa operação começa pelo agrupamento arbitrário dos indivíduos em duplas. Para cada dupla, escolhe-se aleatoriamente um número s com distribuição uniforme no intervalo $[0, 1]$, o qual é comparado com a probabilidade de recombinação p_c pré-estabelecido. Se $s \leq p_c$, realiza-se uma mistura de genes entre os cromossomos, resultando assim em dois novos indivíduos. Caso contrário, os dois “novos” indivíduos são obtidos como cópias fiéis dos cromossomos originais.

Essa mistura de genes entre os cromossomos pode ser feita de várias formas. No que segue são discutidas algumas delas.

Recombinação de Um Ponto

Neste tipo de recombinação, escolhe-se aleatoriamente uma posição da sequência binária, referida como *ponto de clivagem*, e à partir dela os dígitos binários são trocados entre os cromossomos. Como exemplo, considerando os cromossomos x_1 e x_2 abaixo, e supondo que a terceira posição tenha sido escolhida como ponto de clivagem, realiza-se a seguinte transformação:

$$\begin{aligned} x_1 &= (010:101) \longrightarrow x'_1 = (010:011) \\ x_2 &= (100:011) \longrightarrow x'_2 = (100:101) \end{aligned}$$

Recombinação de Dois Pontos

Essa recombinação caracteriza-se pela escolha aleatória de dois pontos de clivagem, sendo trocado entre os cromossomos o conteúdo limitado por esses pontos. Por exemplo, com os mesmos cromossomos de antes, supondo que a segunda e a quarta

posições sejam sorteadas como pontos de clivagem, realiza-se a troca de genes:

$$\begin{aligned} x_1 &= (01:01:01) \longrightarrow x'_1 = (01:00:01) \\ x_2 &= (10:00:11) \longrightarrow x'_2 = (10:01:11) \end{aligned}$$

Recombinação Uniforme

Na recombinação uniforme, cada dupla de cromossomos é associada a uma sequência binária aleatória, de mesmo tamanho, referida como *padrão de recombinação* (PC). Então, dada uma dupla de cromossomos, percorre-se o padrão de corte correspondente e, toda vez que o valor “1” for obtido, são trocados entre os cromossomos os bits que encontram-se nessa mesma posição. No caso em que se obtém o valor “0”, nenhuma troca é realizada. Por exemplo:

$$PC = (0\boxed{1}00\boxed{1}0)$$

$$\begin{aligned} x_1 &= (0\boxed{1}01\boxed{0}1) \longrightarrow x'_1 = (0\boxed{0}01\boxed{1}1) \\ x_2 &= (1\boxed{0}00\boxed{1}1) \longrightarrow x'_2 = (1\boxed{1}00\boxed{0}1) \end{aligned}$$

4.2.4 Mutação

Uma vez concluída a recombinação, todos os cromossomos ficam sujeitos à operação de mutação. Esta operação é de grande importância, pois previne a perda irreversível de material genético. A sua aplicação ocorre da seguinte maneira. Para cada posição no cromossomo é sorteado um número s com distribuição uniforme no intervalo $[0, 1]$, o qual é comparado com a probabilidade de mutação p_m , inicialmente fixada. Se $s \leq p_m$, então o gene na posição considerada é trocado, caso contrário, ele se mantém inalterado. Em se tratando de cromossomos binários, as únicas trocas possíveis são de “1” para “0”, e vice-versa. Por exemplo, admitindo que a terceira coordenada do cromossomo abaixo foi sorteada para a mutação, tem-se:

$$x = (01\boxed{0}010) \longrightarrow x' = (01\boxed{1}010)$$

Observação 4.1. *Note-se que a mutação é realizada de forma independente em cada posição do cromossomo, ou seja, a sua ocorrência em uma dada posição não altera a probabilidade de mutação dos genes nas outras posições.*

4.2.5 Critérios de Parada

Conforme exposto no algoritmo 4.1, as gerações devem se repetir até que um certo critério de parada seja satisfeito. Neste momento, o algoritmo genético é interrompido, e o indivíduo mais apto é aceito como solução do problema de otimização. Segue abaixo uma breve descrição dos critérios de parada usualmente utilizados em aplicações práticas:

- (a) **Número máximo de Gerações:** Segundo este critério, o algoritmo genético é suspenso assim que um número máximo de gerações é alcançado;
- (b) **Tempo de Processamento:** Neste caso, o processo é interrompido após um certo tempo de processamento;
- (c) **Aptidão Constante:** Estabelece a parada do algoritmo genético quando o indivíduo mais apto da população não apresentar variação de aptidão por um certo número de gerações;
- (d) **Melhor indivíduo:** Dado um valor de aptidão pré-estabelecido, interrompe-se o algoritmo genético assim que a aptidão do melhor indivíduo ultrapassar esse valor.

4.3 Fundamento Teórico

O fundamento matemático que justifica o funcionamento dos algoritmos genéticos é fornecido pelo Teorema dos Esquemas, devido à Holland [35]. Esta seção apresenta uma demonstração desse teorema, bem como os conceitos subjacentes.

4.3.1 A Teoria dos Esquemas

Um *esquema* é uma sequência de símbolos que tomam valores no alfabeto $\{1, 0, *\}$, onde o asterisco $*$ representa ambos os valores, 1 e 0. Assim, um esquema $H = (1 * * 1)$, por exemplo, expressa o conjunto de todas as sequências de quatro bits que iniciam e terminam com 1. Cada sequência que satisfaz essa configuração (tais como (1001) e (1111)) é referida como uma *instância* de H .

Nesse contexto, os símbolos diferentes de $*$ recebem o nome de *bits definidores*. Define-se a *ordem* de um esquema H como a quantidade de seus bits definidores, indicada pela notação $o(H)$. Por outro lado, o comprimento de H é definido como a distância entre o primeiro e o último bit definidor desse esquema, sendo denotado por $\delta(H)$ ³. Portanto, para o esquema $H = (1 * * 1)$ tem-se $o(H) = 2$ e $\delta(H) = 3$.

A fim de estabelecer o Teorema dos Esquemas, considere um algoritmo genético com codificação binária, no qual emprega-se a seleção pelo método da roleta, a recombinação de um ponto, e a mutação usual, conforme a descrição feita na seção anterior. Tal especificação caracteriza o chamado *Algoritmo Genético Canônico*, por ter sido essa a primeira configuração de AG proposta por Holland. Para a análise subsequente, é útil ainda fixar a seguinte notação:

f é a função objetivo do problema, utilizada diretamente como função de aptidão;

$m(H, t)$ é o número de instâncias de H presentes t -ésima geração;

$\hat{f}(H, t)$ é a aptidão média das instâncias de H na t -ésima geração;

$\bar{f}(t)$ é a aptidão média da população na t -ésima geração;

$x \in H$ significa que “ x é uma instância do esquema H ”.

Lema 4.2. *Ignorando os efeitos da recombinação e da mutação, ao se aplicar o método da roleta a uma população de N cromossomos, tem-se*

$$E [m(H, t + 1)] = m(H, t) \left(\frac{\hat{f}(H, t)}{\bar{f}(t)} \right).$$

³Se n é o número de símbolos que separam os bits definidores mais afastados, então $\delta(H) = n + 1$.

Demonstração: Considerando a equação (4.3), o número esperado de cópias de um cromossomo x_i na geração t é dado por:

$$NP_i = N \left(\frac{f(x_i)}{\sum_{j=1}^N f(x_j)} \right) = f(x_i) / \left(\sum_{j=1}^N f(x_j) / N \right) = \frac{f(x_i)}{\bar{f}(t)}.$$

Assim, ignorando os efeitos da recombinação e da mutação

$$\begin{aligned} E [m(H, t + 1)] &= \sum_{x \in H} \frac{f(x)}{\bar{f}(t)} \\ &= \frac{m(H, t)}{m(H, t)} \sum_{x \in H} \frac{f(x)}{\bar{f}(t)} \\ &= \frac{m(H, t)}{\bar{f}(t)} \sum_{x \in H} \frac{f(x)}{m(H, t)} \\ &= m(H, t) \left(\frac{\hat{f}(H, t)}{\bar{f}(t)} \right), \end{aligned}$$

uma vez que, por definição, $\hat{f}(H, t) = \sum_{x \in H} \frac{f(x)}{m(H, t)}$, por definição. ■

Embora a recombinação e a mutação possam tanto destruir como criar instâncias de H , os próximos resultados levam em conta apenas os efeitos destrutivos dessas operações, ou seja, aqueles que implicam em uma diminuição do número de instâncias de H . Diz-se que o esquema H “sobreviveu” à recombinação ou à mutação, se ao menos um dos cromossomos resultantes dessas operações ainda permanecer como uma instância de H . Nesse sentido, são estabelecidos os seguintes Lemas.

Lema 4.3. *Considere o operador recombinação de um ponto, definido pela probabilidade p_c . Então, dado um esquema H de tamanho m , a probabilidade $S_c(H)$ dele sobreviver à recombinação satisfaz a seguinte desigualdade:*

$$S_c(H) \geq 1 - p_c \left(\frac{\delta(H)}{m - 1} \right).$$

Demonstração: Com efeito, se o ponto de clivagem ocorrer entre os bits definidores mais afastados de H então esse esquema não irá sobreviver à recombinação, a menos que os outros cromossomos envolvidos também sejam suas instâncias. Como a quantidade possível de pontos de clivagem é igual à $(m - 1)$, decorre que a probabilidade

de nenhuma instância de H ser propagada para a próxima geração é dada por:

$$p_c \left(\frac{\delta(H)}{m-1} \right) q,$$

onde q é a probabilidade das outras instâncias não pertencerem ao esquema H . Portanto, a probabilidade de H sobreviver à recombinação satisfaz a desigualdade:

$$\begin{aligned} S_c(H) &= 1 - p_c \left(\frac{\delta(H)}{m-1} \right) q \\ &\geq 1 - p_c \left(\frac{\delta(H)}{m-1} \right). \end{aligned}$$

■

Lema 4.4. *Suponha que a mutação seja aplicada com probabilidade p_m . Então a probabilidade de um esquema sobreviver à mutação satisfaz a desigualdade:*

$$S_m(H) \geq 1 - o(H)p_m.$$

Demonstração: De fato, a probabilidade de determinado bit em um cromossomo não ser alterado pela mutação é igual à $(1 - p_m)$. Logo, a chance de nenhum bit definidor de H ser modificado é $(1 - p_m)^{o(H)}$. Como $p_m \geq 0$ e $o(H) \in \mathbb{N}$, pela Desigualdade de Bernoulli⁴, a probabilidade de um esquema H sobreviver à mutação verifica a seguinte relação:

$$S_m(H) = (1 - p_m)^{o(H)} \geq 1 - o(H)p_m.$$

■

Teorema 4.5 (Teorema dos Esquemas). *Dada uma população de N sequências com m bits, seja H um esquema de ordem $o(H)$ e comprimento $\delta(H)$, com ao menos uma instância presente na t -ésima geração. Suponha que um algoritmo genético canônico é aplicado com probabilidades de recombinação e mutação dadas por p_c e p_m , respectivamente. Então, tem-se a seguinte desigualdade:*

$$E [m(H, t + 1)] \geq m(H, t) \left(\frac{\hat{f}(H, t)}{\bar{f}(t)} \right) \left[1 - p_c \left(\frac{\delta(H)}{m-1} \right) - o(H)p_m \right]. \quad (4.4)$$

⁴Ver Capítulo 2 em [44].

Demonstração: Considerando os Lemas (4.2)-(4.4), segue que:

$$\begin{aligned}
E [m(H, t + 1)] &= m(H, t) \left(\frac{\hat{f}(H, t)}{\bar{f}(t)} \right) S_c S_m \\
&\geq m(H, t) \left(\frac{\hat{f}(H, t)}{\bar{f}(t)} \right) \left[1 - p_c \left(\frac{\delta(H)}{m-1} \right) \right] [1 - o(H)p_m] \\
&= m(H, t) \left(\frac{\hat{f}(H, t)}{\bar{f}(t)} \right) \left[1 - o(H)p_m - p_c \left(\frac{\delta(H)}{m-1} \right) + p_c \left(\frac{\delta(H)o(H)}{m-1} \right) p_m \right] \\
&\geq m(H, t) \left(\frac{\hat{f}(H, t)}{\bar{f}(t)} \right) \left[1 - p_c \left(\frac{\delta(H)}{m-1} \right) - o(H)p_m \right]
\end{aligned}$$

■

Observação 4.6. *O Teorema dos Esquemas implica que esquemas curtos, de baixa ordem e com aptidão acima da média, tendem a aumentar o número de instâncias na população conforme avançam as gerações, justificando assim o funcionamento do algoritmo genético canônico.*

4.4 Algoritmos Genéticos com Codificação Real

Feitas essas considerações sobre os AGs binários, esta seção aborda os algoritmos genéticos com codificação real. Tais algoritmos apresentam melhores resultados que os binários no caso de problemas de otimização contínuos. Isto se dá pelos seguintes motivos:

- (a) Ao se representar as variáveis em ponto-flutuante, o algoritmo genético passa a ser limitado apenas pela precisão da máquina;
- (b) O armazenamento exigido pela codificação real é menor que o exigido pela codificação binária;
- (c) A codificação real implica em uma velocidade de processamento maior, em razão de não exigir a decodificação dos cromossomos para avaliar a função objetivo.

Um algoritmo genético com codificação real pode ser projetado usando-se os mesmos tipos de operações genéticas propostos para os algoritmos binários. Contudo,

o seu diferencial reside justamente na possibilidade de se aplicar novos métodos de recombinação e mutação. A seguir, são apresentados dois desses procedimentos, especificamente aqueles que caracterizam o AG aplicado nesta dissertação [47].

4.4.1 Cruzamento Aritmético

Este tipo de recombinação obtém novos indivíduos à partir de combinações lineares dos cromossomos de cada par. Se $x, y \in \mathbb{R}^m$ são dois cromossomos a serem recombina- dos, então os indivíduos resultantes são dados por:

$$x' = \alpha y + (1 - \alpha)x, \quad (4.5)$$

$$y' = \alpha x + (1 - \alpha)y, \quad (4.6)$$

onde α é um número aleatório com distribuição uniforme no intervalo $[0, 1]$.

4.4.2 Mutação Não-Uniforme

Na representação real, supondo que seja definido um número máximo de gerações T_{max} , o operador de mutação não-uniforme pode ser definido da seguinte maneira. Seja x_k a k -ésima coordenada de um vetor $x \in \mathbb{R}^m$. Suponha ainda que essa coordenada tenha sido selecionado para a mutação e esteja restrita ao intervalo $[a, b]$. Então, sorteando um número s no conjunto $\{0, 1\}$, o valor de x_k na t -ésima geração é alterado para x'_k , de acordo com as equações:

$$x'_k = x_k + (b - x_k) \left(1 - r \left(1 - \frac{t}{T_{max}} \right)^d \right), \quad \text{se } s = 0, \quad (4.7)$$

$$x'_k = x_k + (a - x_k) \left(1 - r \left(1 - \frac{t}{T_{max}} \right)^d \right), \quad \text{se } s = 1, \quad (4.8)$$

onde r é um número escolhido aleatoriamente, com distribuição uniforme no intervalo $[0, 1]$, e $d > 0$ é um parâmetro pré-estabelecido denominado *grau de não-uniformidade*.

Observação 4.7. Sem perda de generalidade, considere $s = 1$. Então, fixado r , para $t = 0$ e para t livre, tem-se respectivamente

$$\lim_{d \rightarrow \infty} x'_k = \lim_{d \rightarrow \infty} x_k + (a - x_k)(1 - r^d) = x_k + (a - x_k) = a$$

$$\lim_{t \rightarrow T_{max}} x'_k = x_k + (a - x_k)(1 - 1) = x_k$$

Assim, durante as primeiras gerações nota-se que a mutação não-uniforme induz sobre a coordenada uma forte perturbação, cuja intensidade é modulada pelo grau de não-uniformidade d . Por outro lado, as últimas gerações caracterizam-se por pequenas perturbações, realizando um ajuste mais fino em busca da solução ótima.

4.5 Vantagens e Limitações dos Algoritmos Genéticos

Comparando os algoritmos genéticos com os demais métodos de otimização, pode-se apontar duas principais vantagens:

- (a) Diferente dos métodos tradicionais de otimização, os algoritmos genéticos não impõe restrições matemáticas sobre a função objetivo, podendo ser aplicados a qualquer tipo de problema de otimização;
- (b) Ao invés de operar sobre apenas um ponto, os algoritmos genéticos atuam sempre sobre uma população de pontos, o que aumenta a chance de se encontrar o ponto de máximo, além de prevenir a convergência sobre pontos de máximo local (sub-ótimos).

Em contrapartida, entre as limitações dos algoritmos genéticos estão:

- (a) A dificuldade de se identificar a função de aptidão;
- (b) O problema de definir os vários parâmetros, tais como o tamanho da população, a forma de seleção, as probabilidades de recombinação e mutação, o critério de parada, etc;
- (c) A exigência de um grande número de avaliações da função objetivo, o que pode inviabilizar a sua aplicação, dependendo do problema em estudo.

Problema de Condução da Política Monetária

Neste capítulo, o processo decisório da política monetária é posto sob a forma de um problema do regulador linear-quadrático estocástico. Para isso, seguindo Rudebusch e Svensson [61], supõe-se que o objetivo do Banco Central é minimizar uma função perda quadrática, sujeito aos vínculos expressos por um modelo macroeconômico linear. Como ilustração dessa abordagem, a partir de um modelo estimado para a economia brasileira, realizam-se simulações de política monetária ótima considerando diferentes preferências para o Banco Central do Brasil.

5.1 Regime de Metas para a Inflação

Desde junho de 1999, o Brasil adota o regime de metas para a inflação como diretriz da sua política monetária. Nesse regime, o Banco Central anuncia ao público uma meta numérica de médio prazo para a inflação, e se compromete a atuar de forma que a inflação observada se encaminhe para essa meta. O principal instrumento de política monetária é a taxa básica de juros que equilibra o mercado de reservas bancárias, denominada taxa Selic [8].

Periodicamente, o Comitê de Política Monetária do Banco Central (Copom) se reúne para estabelecer a taxa Selic do período subsequente. Como existe uma defasagem entre as ações de política monetária e os seus efeitos sobre a economia [12], as decisões do Copom baseiam-se em previsões sobre o comportamento das variáveis macroeconômicas. Essas previsões, obtidas por meio de modelos econométricos, consideram dois cenários principais: o de referência e o de mercado. No cenário de referência, supõe-se que a taxa de juros se mantenha constante, no nível em que se encontra, ao longo de todo o horizonte de previsão. Por outro lado, o cenário de mercado admite que a taxa de juros varie de acordo com as expectativas do setor privado. Caso a inflação prevista seja maior (menor) que a meta pré-estabelecida, isso é interpretado pelo Copom como um indicativo de que a taxa Selic terá que ser elevada (reduzida) a um patamar acima (abaixo) do seu nível inicial, ou do nível esperado pelo mercado, a fim de trazer a inflação de volta para a sua meta [13].

Embora as deliberações do Banco Central resultem do voto majoritário entre os membros do Copom, é possível modelar matematicamente a política monetária, supondo que ela seja conduzida por um único agente, cujo objetivo é minimizar uma determinada função perda. Nessa abordagem, a minimização fica restrita às relações que existem entre as variáveis macroeconômicas, exigindo portanto um modelo matemático para representar os vínculos entre essas variáveis.

No que segue, são apresentados os principais canais pelos quais as ações de política monetária atuam sobre a economia. A partir desses fundamentos, um modelo matemático é estimado para representar o comportamento das variáveis macroeconômicas de interesse. Por fim, tendo esse modelo por base, o problema de otimização da autoridade monetária é formulado e resolvido, resultando em regras ótimas para a escolha da taxa Selic.

5.1.1 Mecanismo de Transmissão da Política Monetária

Em razão do seu caráter preventivo, a política monetária requer o conhecimento do mecanismo pelo qual as alterações na taxa Selic são transmitidas para a economia. De

acordo com [9], os principais canais de transmissão da política monetária são: a taxa de juros, a taxa de câmbio, o crédito e as expectativas.

Canal da Taxa de Juros

Ao mudar a taxa Selic, o Banco Central provoca a variação das demais taxas de juros da economia, haja vista a relação de dependência que existe entre elas. Considerando que no curto prazo os preços são rígidos, a ação do Banco Central induz a alteração das taxas reais de juros, o que tem um impacto direto sobre as decisões de investimento do setor privado. A variação no volume de investimentos, por sua vez, reflete-se sobre a demanda agregada, sendo assim transmitida para o nível de preços.

Em particular, uma elevação da taxa Selic implica uma alta da taxa real de juros, que aumenta o custo do capital, ocasionando a diminuição do investimento em capital fixo e estoques. Isto reduz a demanda do setor produtivo, forçando a queda dos preços, de acordo com a Lei da Oferta e da Procura [70]. De forma análoga, a elevação da taxa Selic também diminui o consumo de bens duráveis, levando a uma consequente redução dos seus preços.

Por outro lado, a diminuição da taxa Selic implica a redução da taxa real de juros, o que estimula investimento e consumo, provocando assim o aumento da demanda agregada. Isto induz a elevação dos preços por dois motivos. O primeiro é o aumento dos custos de produção, decorrentes da maior necessidade por trabalho, capital e matérias-primas. O segundo motivo refere-se à tendência das empresas monopolistas de explorar o aumento da demanda para elevar suas taxas de lucro.

Canal da Taxa de Câmbio

A transmissão das ações de política monetária através da taxa de câmbio funciona da seguinte maneira. Um aumento da taxa Selic torna os ativos internos relativamente mais rentáveis que os ativos externos. Nesse caso, haverá uma tendência entre os investidores externos para converter moedas estrangeiras em reais (R\$), aumentando a demanda pela moeda nacional e alterando com isto a sua relação de troca (moeda

estrangeira/real). Por um lado, essa valorização da taxa de câmbio acaba tornando os produtos nacionais mais caros ao consumidor estrangeiro, o que reduz a demanda mundial por exportações brasileiras, forçando assim a queda dos preços. Por outro, ela faz com que os produtos importados fiquem mais baratos ao consumidor brasileiro, o que tem um efeito deflacionário pelas seguintes razões:

1. Os bens domésticos passam a ser substituídos por bens importados, o que reduz a demanda interna;
2. A redução dos preços das *commodities* importadas diminui os custos de produção;
3. Dado que os bens importados também fazem parte da cesta que compõe os índices que medem a inflação, a queda dos seus preços contribui diretamente para a redução dos índices de inflação.

Canal das Expectativas

Alterando a taxa Selic, o Banco Central influencia as expectativas do setor privado quanto à evolução futura das variáveis macroeconômicas. De acordo com [11], pode-se citar três formas pelas quais a expectativa de inflação pode interferir sobre o nível dos preços:

1. Diretamente: pela sua incorporação aos preços de produtos e serviços;
2. Via Salários: ao ser levada em conta nas negociações trabalhistas;
3. Indiretamente: alterando a taxa real de juros *ex ante*, a qual tem grande importância para empresas e indivíduos na tomada de decisões sobre investimento e poupança.

Canal do Crédito

Reduzindo a taxa Selic, o Banco Central eleva o volume de crédito disponível na economia, permitindo que os bancos comerciais aumentem seus empréstimos. Isso estimula

tanto o investimento quanto o consumo, levando a um aumento da demanda agregada, o qual provoca uma alta dos preços.

Por outro lado, ao elevar a taxa Selic, o Banco Central restringe o acesso ao crédito, desestimulando o investimento privado. Como consequência, tem-se uma redução da demanda agregada, que contribui para frear a inflação.

Condução da Política Monetária

Para todos os canais de transmissão tratados acima, verifica-se que a elevação (redução) da taxa Selic induz a diminuição (aumento) da taxa de inflação. Essa relação inversa entre a taxa Selic e a inflação fornece a base para o procedimento de condução da política monetária, o qual consiste em: elevar (reduzir) a taxa Selic, quando a inflação prevista no horizonte considerado se mostrar acima (abaixo) da meta pré-estabelecida.

Encerrando esta seção, a Figura 5.1 resume os principais canais de transmissão da política monetária¹:

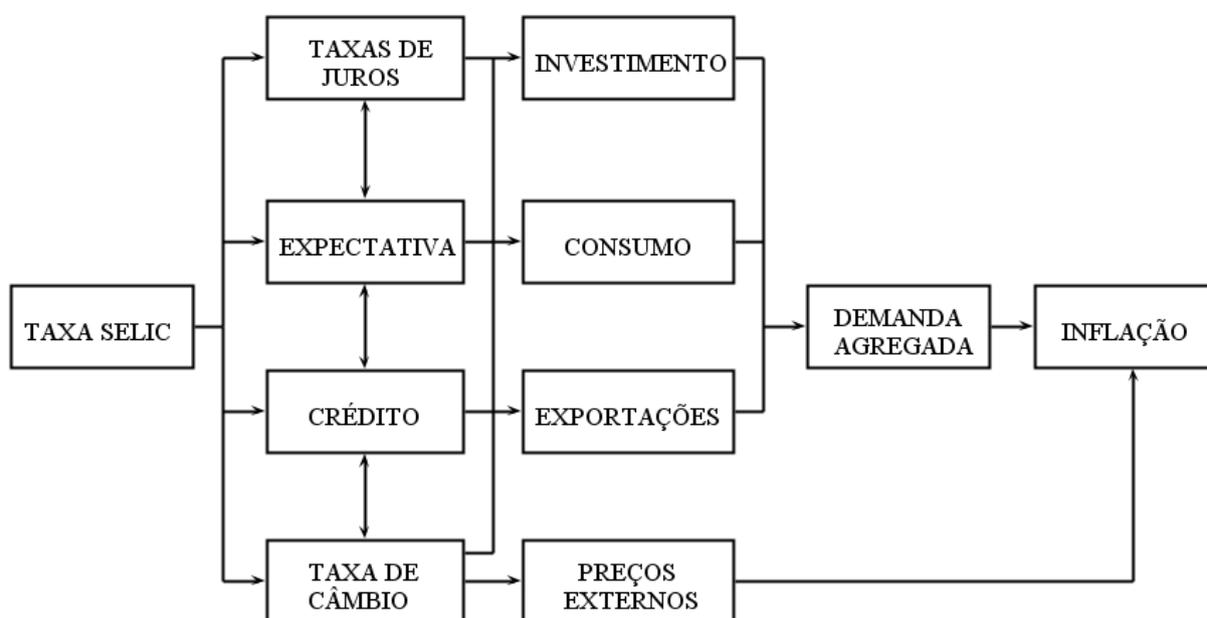


Figura 5.1: Síntese dos principais canais de transmissão da política monetária.

¹Esta figura é uma adaptação do diagrama apresentado em [9], página 91.

5.2 Modelo Macroeconômico Linear

Com o objetivo de representar o mecanismo de transmissão da política monetária, neste trabalho, considera-se um modelo macroeconômico linear inspirado naquele proposto por Rudebusch e Svensson [61]. As equações que o compõe são as seguintes:

$$\pi_{t+1} = \alpha_{\pi}(L)\pi_t + \alpha_{\pi^e}(L)\pi_t^e + \alpha_h(L)h_t + \alpha_c(L)\Delta c_t + \varepsilon_{\pi,t}, \quad (5.1)$$

$$h_{t+1} = \beta_h(L)h_t + \beta_r(L)(i_t - \pi_t^e) + \beta_c(L)c_t + \varepsilon_{h,t}, \quad (5.2)$$

$$\pi_{t+1}^e = \gamma_{\pi^e}(L)\pi_t^e + \gamma_{\pi}(L)\pi_t + \gamma_c(L)\Delta c_t + \gamma_i(L)\Delta i_t + \varepsilon_{\pi^e,t}, \quad (5.3)$$

$$c_{t+1} = \delta_c(L)c_t + \delta_{\pi}(L)\pi_t + \delta_{\pi^e}(L)\pi_t^e + \delta_i(L)\Delta i_t + \varepsilon_{c,t}, \quad (5.4)$$

onde:

$\alpha(L), \beta(L), etc.$, são polinômios no operador defasagem L , definido por $Lx_t = x_{t-1}$;

π_t corresponde a taxa de inflação no período t ;

π_t^e corresponde a expectativa de inflação que os agentes do mercado tem no período t ;

h_t corresponde ao hiato do produto no período t ;

c_t corresponde a taxa de câmbio real no período t , sendo $\Delta c_t = c_t - c_{t-1}$;

i_t corresponde a taxa nominal de juros no período t , com $\Delta i_t = i_t - i_{t-1}$;

$(i_t - \pi_t^e)$ corresponde a taxa real de juros *ex ante* no período t ;

$\varepsilon_{\pi,t}, \varepsilon_{h,t}, \varepsilon_{\pi^e,t}, \varepsilon_{c,t}$ são erros aleatórios, associados aos choques ocorridos no período t .

A equação (5.1), conhecida como *Curva de Phillips*, expressa a taxa de inflação em função das suas defasagens, da expectativa de inflação, do hiato do produto e do diferencial da taxa de câmbio real. Por sua vez, a equação (5.2) modela os canais de taxa de juros e de crédito, sendo chamada de *Curva IS*. Nela, o hiato do produto é explicado em função das suas defasagens, da taxa real de juros *ex ante* e da taxa de câmbio real. O canal das expectativas é representado pela equação (5.3), referida neste texto como *Curva da Expectativa de Inflação*. Ela supõe que a expectativa de inflação seja determinada pelas suas próprias defasagens, pela inflação realizada, pelo diferencial da taxa

de câmbio real e pelo diferencial da taxa de juros, sendo a escolha dessas variáveis explicativas baseada em [11]. Por fim, o canal da taxa de câmbio é representado pela equação (5.4), denominada aqui como *Curva da Taxa de Câmbio*. Essa equação explica o comportamento da taxa de câmbio real em função das suas defasagens, da taxa de inflação, da expectativa de inflação e do diferencial da taxa de juros.

5.2.1 Descrição das Variáveis

O modelo (5.1)-(5.4) possui quatro variáveis endógenas, quais sejam: a taxa de inflação, o hiato do produto, a expectativa de inflação e a taxa de câmbio real. A única variável exógena é a taxa nominal de juros, que no caso é a taxa Selic².

Para a estimação dos coeficientes das equações (5.1)-(5.4), utilizou-se dados trimestrais da economia brasileira, do terceiro trimestre de 2001 (2001/T3) ao segundo trimestre de 2010 (2010/T2). Esse dados foram obtidos nos sites do Banco Central do Brasil e do IPEADATA³. A seguir faz-se uma descrição detalhada das variáveis do modelo, e também das séries históricas utilizadas neste trabalho para representá-las.

Inflação

A inflação é definida como o aumento generalizado dos preços em uma economia. No Brasil ela é acompanhada por diversos índices de preços, sendo o Índice Nacional de Preços ao Consumidor Amplo (IPCA) a referência do Banco Central para efeitos de política monetária. De acordo com o Decreto número 3.088 [9], considera-se que a meta de inflação foi cumprida quando a taxa de variação acumulada do IPCA, relativa ao período de janeiro a dezembro do ano em questão, situar-se na faixa do seu respectivo intervalo de tolerância. Por isso, como medida da inflação, optou-se por utilizar a taxa de variação do IPCA acumulada em 12 meses⁴, sendo $\pi_t = \ln(1 + IPCA_t/100)$.

²Com respeito a um determinado modelo, uma variável é dita *endógena* quando a sua dinâmica é explicado por esse modelo. Por outro lado, uma variável é dita *exógena* quando a sua dinâmica *não* é explicada pelo modelo em questão.

³Respectivamente, <https://www3.bcb.gov.br/sgspub> e <http://www.ipeadata.gov.br>.

⁴Série 13522 do Sistema Gerenciador de Séries Temporais (SGS) do Banco Central do Brasil.

Hiato do Produto

O hiato do produto é um conceito amplamente usado para representar as flutuações da demanda agregada. Ele é definido como a diferença entre o produto realizado, e o nível máximo de produção que a economia é capaz de manter sem gerar pressões inflacionárias, conhecido como *produto potencial* [31].

No presente trabalho, como medida do produto potencial utiliza-se a tendência de longo prazo do PIB a preços de mercado dessazonalizado⁵, obtida através do filtro Hodrick-Prescott (ver detalhes no Apêndice B). Dessa forma, calculado o hiato, define-se $h_t = \ln(1 + HIATO_t)$.

Expectativa de Inflação

A expectativa de inflação é representada pela taxa de variação do IPCA esperada para os próximos 12 meses, pesquisada pelo Banco Central junto ao setor privado⁶, sendo $\pi_t^e = \ln(1 + IPCA_t^e/100)$.

Taxa de Câmbio real

Corresponde a média no trimestre do índice da taxa de câmbio efetiva real⁷, o qual representa a dinâmica do real (R\$) em relação às moedas estrangeiras. Define-se $c_t = \ln(CAMBIO_t)$.

Taxa nominal de juros

A taxa nominal de juros considerada é a taxa Selic fixada pelo Copom⁸, vigente ao fim do trimestre considerado, sendo $i_t = \ln(1 + SELIC_t/100)$.

⁵Série 1253 do Sistema Gerenciador de Séries Temporais (SGS) do Banco Central do Brasil.

⁶Disponível no site IPEADATA

⁷Série 11752 do Sistema Gerenciador de Séries Temporais (SGS) do Banco Central do Brasil.

⁸Série 432 do Sistema Gerenciador de Séries Temporais (SGS) do Banco Central do Brasil.

Variáveis Binárias

Seguindo [3, 6], para evitar que as estimativas dos coeficientes fossem viesados por dados discrepantes, foram adicionadas variáveis binárias às equações (5.1)-(5.4). A descrição dessas variáveis, bem como a sua justificativa são apresentada a seguir.

Na Curva de Phillips (5.1), a fim de capturar o forte aumento da inflação verificado durante o último trimestre de 2002 e o primeiro trimestre de 2003, são incluídas, respectivamente, as variáveis binárias $D_{\pi,1}$ e $D_{\pi,2}$.

Por sua vez, a Curva IS (5.2) é modificada com a inclusão de quatro variáveis binárias, denotadas por $D_{h,1}$, $D_{h,2}$, $D_{h,3}$ e $D_{h,4}$. As variáveis $D_{h,1}$ e $D_{h,2}$ buscam captar os efeitos negativos decorrentes das condições de crédito adversas verificadas durante o primeiro e o segundo trimestres de 2003, respectivamente. Por outro lado, as variáveis $D_{h,3}$ e $D_{h,4}$ referem-se aos efeitos da Crise Financeira Internacional observados durante o último trimestre de 2008 e o primeiro trimestre de 2009, respectivamente.

Com respeito à Curva da Expectativa de Inflação (5.3), consideram-se as variáveis $D_{\pi^e,1}$, $D_{\pi^e,2}$ e $D_{\pi^e,3}$. As duas primeiras correspondem ao forte aumento das expectativas de inflação do setor privado observado durante o terceiro e quarto trimestres de 2002, em decorrência da eleição presidencial. Já a variável $D_{\pi^e,3}$ busca captar a queda das expectativas de inflação no segundo trimestre de 2008, frente a Crise Financeira Internacional.

Por fim, na Curva da Taxa de Câmbio (5.4) são incluídas quatro variáveis binárias, referidas como $D_{c,1}$, $D_{c,2}$, $D_{c,3}$ e $D_{c,4}$. As três primeiras buscam capturar a desvalorização cambial ocorrida durante o segundo, terceiro e quarto trimestres de 2002, respectivamente, associadas à crise de confiança diante da alternância presidencial. Quanto à variável $D_{c,4}$, ela representa a forte desvalorização do real verificada no último trimestre de 2008 com a eclosão da Crise Financeira Internacional.

5.2.2 Análise das Séries Temporais

Como primeiro passo para a estimação dos coeficientes do modelo macroeconômico, o teste ADF de raiz unitária foi aplicado a fim de investigar a estacionariedade das variáveis consideradas. Foram utilizadas as especificações com e sem intercepto, resultando nas estatísticas τ e τ_μ , respectivamente⁹. Para a escolha do número de defasagens, aplicou-se o Critério de Informação de Schwarz¹⁰, admitindo um número máximo de nove defasagens. Os resultados são apresentados na Tabela 5.1.

Tabela 5.1: Teste de Raiz Unitária

Série	Estatística τ	Defasagens	Estatística τ_μ	Defasagens
h_t	-3,863*	1	-3,7964*	1
π_t	-1,336	1	-2,7851***	1
π_t^e	-1,778	6	-4,6088*	5
c_t	-2,179**	4	-0,2960*	4
Δc_t	-5,219*	1	-4,1705*	3
i_t	-3,105*	5	-1,9146	1
Δi_t	-3,761*	1	-5,7069*	4
$(i_t - \pi_t^e)$	-1,172	1	-2,0461	1

* Rejeita a hipótese de raiz unitária ao um nível de 1%.

** Rejeita a hipótese de raiz unitária ao um nível de 5%.

*** Rejeita a hipótese de raiz unitária ao um nível de 10%.

Com exceção da taxa real de juros ($i_t - \pi_t^e$), em pelo menos uma das especificações do teste, todas as variáveis se mostraram estacionárias ao nível de 10% de significância. No caso da taxa real de juros, optou-se por mantê-la em nível, tendo em vista o valor τ_μ correspondente, e o baixo poder do teste ADF.

⁹Ver a subseção 2.3.3

¹⁰Ver a subseção 2.3.2.

5.2.3 Estimação dos Coeficientes

Cada equação do modelo, separadamente, teve seus coeficientes estimados pelo Método dos Mínimos Quadrados Ordinários, utilizando-se dados trimestrais da economia brasileira, referentes ao período [2001/T3 - 2010/T2]. Como as equações (5.1)-(5.4) envolvem polinômios no operador defasagem L , foi preciso escolher quais defasagens incluir no modelo. Essa escolha deu-se com base no Critério de Informação de Schwarz, o qual foi aplicado de maneira individual sobre as equações do modelo. Mais especificamente, admitindo um número máximo de três defasagens para as variáveis, foram estimadas todas as combinações possíveis para cada equação. Então, dentre as especificações obtidas, escolheu-se aquela que resultou no menor valor do Critério de Informação de Schwarz (SIC). Os resultados obtidos são apresentados e discutidos a seguir.

Curva de Phillips

Incluindo as variáveis binárias, a especificação selecionada para a equação (5.1) foi a seguinte:

$$\pi_{t+1} = \alpha_1\pi_t + \alpha_2\pi_{t-1} + \alpha_3\pi_{t-1}^e + \alpha_4h_{t-1} + \alpha_5\Delta c_t + \alpha_6D_{\pi,1} + \alpha_7D_{\pi,2} + \varepsilon_{\pi,t} \quad (5.5)$$

A tabela 5.2 descreve os coeficientes estimados, o grau de ajuste obtido, e também os testes estatísticos aplicados ao modelo. Com respeito aos testes realizados, pode-se assumir que a Curva de Phillips (5.5) satisfaz as condições do Teorema de Gauss-Markov. De fato, mesmo a um nível de 10% de significância, o teste Shapiro-Wilk não rejeitou a hipótese de que os erros sejam provenientes de uma distribuição normal. Além disso, os testes Breusch-Godfrey e White não indicaram a presença de autocorrelação e heterocedasticidade, respectivamente. Por fim, note-se ainda que todos os coeficientes se mostraram estatisticamente significantes, ao menos a um nível de 10%.

Tabela 5.2: Curva de Phillips

Coeficiente	Valor Estimado	Desvio Padrão	Estatística t
α_1	0,7157	0,1425	5,0228
α_2	-0,2667	0,1048	-2,5567
α_3	0,6086	0,1057	5,7568
α_4	0,1538	0,0914	1,6824
α_5	-0,0368	0,0243	-1,5118
α_6	0,0628	0,0092	6,7963
α_7	0,0548	0,0103	5,3334
R^2	0,9657	SIC	-198,63
White	3,7117	Graus de Liberdade	4
Breusch-Godfrey	1,3502	Graus de Liberdade	4
Shapiro-Wilk	0,9707	P-valor	0,9999

Curva IS

Incluindo as variáveis binárias, a especificação selecionada para a equação (5.2) foi a seguinte:

$$h_{t+1} = \beta_1 h_t + \beta_2 (i_t - \pi_t^e) + \beta_3 c_t + \beta_4 D_{h,1} + \beta_5 D_{h,2} + \beta_6 D_{h,6} + \beta_7 D_{h,7} + \varepsilon_{h,t} \quad (5.6)$$

A tabela 5.3 descreve os coeficientes estimados, o grau de ajuste obtido, e também os testes estatísticos aplicados ao modelo. No caso da Curva IS (5.6), note-se que o teste Shapiro-Wilk rejeitou a hipótese de que os erros provêm de uma distribuição normal. Enquanto isso, os testes Breusch-Godfrey e White não indicaram a presença de autocorrelação e heterocedasticidade, respectivamente. Por fim, quanto à estatística t , todas as variáveis se mostraram significativas, ao menos a um nível de 10%. Contudo, é importante notar que esses testes de significância dependem da normalidade dos erros¹¹, de modo que os seus resultados devem ser encarados com cautela.

¹¹Ver a subseção 2.1.4.

Tabela 5.3: Curva IS

Coeficiente	Valor Estimado	Desvio Padrão	Estatística t
β_1	0,8273	0,0710	11,65
β_2	-0,1518	0,0394	-3,85
β_3	0,0038	0,0008	4,60
β_4	-0,0239	0,0063	-3,77
β_5	-0,0093	0,0064	-1,44
β_6	-0,0515	0,0068	-7,52
β_7	-0,0370	0,0063	-5,86
R^2	0,89	SIC	-215,08
White	4,4756	Graus de Liberdade	3
Breusch-Godfrey	8,1755	Graus de Liberdade	4
Shapiro-Wilk	0,9048	P-valor	0,014

Curva da Expectativa de Inflação

A especificação selecionada para a equação (5.3), incluindo as variáveis binárias, foi a seguinte:

$$\pi_{t+1}^e = \gamma_1 \pi_t^e + \gamma_2 \pi_{t-1}^e + \gamma_3 \pi_t + \gamma_4 \Delta c_t + \gamma_5 \Delta i_t + \gamma_6 D_{\pi^e,1} + \gamma_7 D_{\pi^e,2} + \gamma_8 D_{\pi^e,3} + \varepsilon_{\pi^e,t} \quad (5.7)$$

A tabela 5.4 descreve os coeficientes estimados, o grau de ajuste obtido, e também os testes estatísticos aplicados ao modelo. Pelos testes realizados, também pode-se admitir que a Curva da Expectativa de Inflação (5.7) satisfaz as condições do Teorema de Gauss-Markov. Com efeito, os testes Breusch-Godfrey e White não indicaram a presença de autocorrelação e heterocedasticidade, respectivamente. Além disso, o teste Shapiro-Wilk não rejeitou a hipótese de que os erros sejam provenientes de uma distribuição normal. Por fim, note-se que todos os coeficientes se mostraram estatisticamente significantes, ao menos a um nível de 10%.

Tabela 5.4: Curva da Expectativa de Inflação

Coefficiente	Valor Estimado	Desvio Padrão	Estatística t
γ_1	1,048	0,0788	13,30
γ_2	0,098	0,0888	11,10
γ_3	-0,1877	0,0539	-3,48
γ_4	-0,0237	0,0113	-2,09
γ_5	-0,222	0,0592	-3,75
γ_6	0,0243	0,0048	5,05
γ_7	0,0747	0,0054	13,78
γ_8	0,01347	0,0047	2,85
R^2	0,94	SIC	-220,50
White	7,015	Graus de Liberdade	4
Breusch-Godfrey	8,1394	Graus de Liberdade	4
Shapiro Wilk	0,9381	P-valor	0,1168

Curva da Taxa de Câmbio

Por fim, a especificação selecionada para a equação (5.4), incluindo as variáveis binárias, foi a seguinte:

$$c_{t+1} = \delta_1 c_t + \delta_2 \pi_t + \delta_3 \pi_t^e + \delta_4 \Delta i_t + \delta_5 D_{c,1} + \delta_6 D_{c,2} + \delta_7 D_{c,3} + \delta_8 D_{c,4} + \varepsilon_{c,t} \quad (5.8)$$

A tabela 5.5 apresenta de forma resumida os coeficientes estimados, o grau de ajuste obtido, e também os testes estatísticos aplicados ao modelo. Também neste caso, os testes Breusch-Godfrey e White não indicaram a presença de autocorrelação e heterocedasticidade. Além disso, mesmo a um nível de 10% de significância, o teste Shapiro-Wilk não rejeitou a hipótese de que os erros possuem uma distribuição normal. Quanto aos coeficientes, todos se mostraram estatisticamente significantes, ao menos a um nível de 10%.

Tabela 5.5: Curva da Taxa de Câmbio

Coeficiente	Valor Estimado	Desvio Padrão	Estatística t
δ_1	0,9820	0,0080	122,67
δ_2	-0,9397	0,4607	2,039
δ_3	1,9831	1,1058	1,793
δ_4	-1,7591	0,7415	-2,3724
δ_5	0,083	0,0412	2,0149
δ_6	0,2898	0,0428	6,7668
δ_7	0,1532	0,0408	3,7505
δ_8	0,2751	0,0421	6,5330
R^2	0,97	SIC	-80,14
White	5,9371	Graus de Liberdade	4
Breusch-Godfrey	6,1547	Graus de Liberdade	4
Shapiro Wilk	0,9390	P-valor	0,1128

5.3 Modelagem Matemática da Política Monetária

Conforme exposto na seção 5.1, o objetivo operacional da política monetária brasileira é manter a inflação estável em torno de uma meta numérica pré-estabelecida. Para isso, o Banco Central atua sobre a demanda agregada por meio da taxa Selic, cujo valor é definido periodicamente pelo Copom.

Com base em [8, 9, 11, 12, 13], é razoável supor que o processo decisório do Copom seja orientado pelos seguintes critérios:

1. Manter a inflação na meta, ou trazê-la de volta após a ocorrência de um desvio;
2. Promover a estabilidade econômica;
3. Quando ajustes sobre a taxa Selic se fizerem necessários, realizá-los de forma gradual e consistente com o padrão de resposta anteriormente apresentado.

O grau pelo qual esses critérios são satisfeitos pode ser expresso matematicamente através da função perda [4]:

$$L_t = \lambda_\pi(\pi_t - \pi^*)^2 + \lambda_h h_t^2 + \lambda_i(i_t - i_{t-1})^2. \quad (5.9)$$

O primeiro termo em (5.9) está associado ao Critério 1, segundo o qual, quanto maior for o desvio entre a inflação atual π_t e a sua meta π^* , maior será a perda associada. Por sua vez, o segundo termo corresponde ao Critério 2, e mostra que a perda do Banco Central aumenta com a flutuação da atividade econômica, expressa pelo hiato do produto. Finalmente, o terceiro termo representa o Critério 3, e estabelece que o valor atual da taxa Selic i_t não deve se desviar muito do seu valor no período anterior i_{t-1} , implicando uma suavização da taxa de juros.

A partir da função (5.9), a trajetória da taxa Selic que melhor satisfaz os critérios 1-3 corresponde àquela que minimiza o valor esperado das perdas atuais e futuras do Banco Central. Dessa forma, considerando os vínculos expressos pelo modelo macroeconômico (5.5)-(5.8), a condução da política monetária pode ser vista como o seguinte problema de otimização:

$$\min_{\{i_t\}_{t=0}^{\infty}} E_0 \left(\sum_{t=0}^{\infty} \delta^t L_t \right), \quad (5.10)$$

$$\text{sujeito a } (5.5) - (5.8). \quad (5.11)$$

Nesse contexto, as decisões do Banco Central são determinadas pelos valores que ele atribui aos pesos δ , λ_π , λ_h e λ_i , os quais são referidos como as *preferências de política monetária*.

5.3.1 Solução via Programação Dinâmica

Para resolver o problema de otimização (5.10)-(5.11), este trabalho segue a mesma metodologia empregada em [61]. O primeiro passo é escrever o modelo (5.5)-(5.8) em

notação matricial:

$$\begin{bmatrix} \pi_{t+1} \\ \pi_t \\ \pi_{t+1}^e \\ \pi_t^e \\ h_{t+1} \\ h_t \\ c_{t+1} \\ c_t \\ i_t \end{bmatrix} = \begin{bmatrix} \alpha_1 & \alpha_2 & 0 & \alpha_3 & 0 & \alpha_4 & \alpha_5 & -\alpha_5 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \gamma_3 & 0 & \gamma_1 & \gamma_2 & 0 & 0 & \gamma_4 & -\gamma_4 & -\gamma_5 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\beta_2 & 0 & \beta_1 & 0 & \beta_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ \delta_2 & 0 & \delta_3 & 0 & 0 & 0 & \delta_1 & 0 & -\delta_4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \pi_t \\ \pi_{t-1} \\ \pi_t^e \\ \pi_{t-1}^e \\ h_t \\ h_{t-1} \\ c_t \\ c_{t-1} \\ i_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \gamma_5 \\ 0 \\ -\beta_2 \\ 0 \\ \delta_4 \\ 0 \\ 1 \end{bmatrix} i_t + \begin{bmatrix} \varepsilon_{\pi,t} \\ 0 \\ \varepsilon_{\pi^e,t} \\ 0 \\ \varepsilon_{h,t} \\ 0 \\ \varepsilon_{c,t} \\ 0 \\ 0 \end{bmatrix}$$

Esta equação, por sua vez, pode ser sintetizada como:

$$X_{t+1} = AX_t + Bi_t + u_t \quad (5.12)$$

Por outro lado, supondo que a meta de inflação seja zero ($\pi^* = 0$)¹² e definindo o vetor

$Y_t = [\pi_t \quad h_t \quad (i_t - i_{t-1})]'$, tem-se as seguintes igualdades:

$$L_t = Y_t'KY_t \quad (5.13)$$

$$Y_t = C_x X_t + C_i i_t, \quad (5.14)$$

onde

$$K = \begin{bmatrix} \lambda_\pi & 0 & 0 \\ 0 & \lambda_h & 0 \\ 0 & 0 & \lambda_i \end{bmatrix}, \quad C_x = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \end{bmatrix} \quad \text{e} \quad C_i = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Logo, combinando as equações (5.13) e (5.14), obtém-se:

$$\begin{aligned} L_t &= Y_t'KY_t \\ &= (C_x X_t + C_i i_t)'K(C_x X_t + C_i i_t) \\ &= X_t' C_x' K C_x X_t + X_t' C_x' K C_i i_t + i_t C_i' K C_x X_t + i_t C_i' K C_i i_t \\ &= X_t' R X_t + i_t Q i_t + 2i_t W X_t \end{aligned} \quad (5.15)$$

¹²A hipótese de inflação zero justifica-se no presente trabalho principalmente pela sua conveniência analítica. Não obstante, dentro da literatura econômica existe um extenso debate sobre qual deve ser a meta de inflação a ser perseguida pela política monetária. Argumentos que defendem uma meta de inflação nula podem ser encontrados em Hoskins [36], enquanto argumentos favoráveis a uma meta de inflação baixa, mas positiva, podem ser vistos em Aiyagari [1].

onde

$$R = C'_x K C_x, \quad Q = C'_i K C_i \text{ e } W = C'_i K C_x. \quad (5.16)$$

Finalmente, substituindo (5.12) e (5.15) em (5.10) e (5.11), respectivamente, o problema de condução da política monetária assume a forma de um problema do regulador linear-quadrático estocástico modificado¹³:

$$\min_{\{i_t\}_{t=0}^{\infty}} E_0 \left(\sum_{t=0}^{\infty} \delta^t (X'_t R X_t + i_t Q i_t + 2i_t W X_t) \right), \quad (5.17)$$

$$\text{sujeito a } X_{t+1} = A X_t + B i_t. \quad (5.18)$$

De acordo com o estudo realizado na subseção 3.3.4, a solução desse problema é dada pela seguinte regra ótima:

$$i_t = -F X_t = -(Q + \delta B' P B)^{-1} (\delta B' P A + W') X_t, \quad (5.19)$$

onde P é a matriz simétrica definida-positiva que satisfaz a equação matricial algébrica de Riccati:

$$P = \bar{R} + \delta \bar{A}' P \bar{A} - \delta^2 \bar{A}' P B (Q + \delta B' P B)^{-1} B' P \bar{A} \quad (5.20)$$

Observação 5.1. Como R , Q e W dependem da matriz K , as equações (5.19) e (5.20) implicam que $F = F(\delta, \lambda_{\pi}, \lambda_h, \lambda_i)$. Em razão disso, regras ótimas distintas podem ser obtidas quando se consideram diferentes preferências para o Banco Central.

5.3.2 Regras Ótimas para a Política Brasileira

Nesta subseção são apresentadas regras ótimas para a política monetária brasileira, simulando diferentes preferências para o Banco Central. Para obtê-las, a equação de Riccati (5.20) foi resolvida para matrizes K distintas aplicando-se o algoritmo 3.2.

¹³Esta especificação para o problema de otimização do Banco Central decorre da função perda quadrática e do modelo macroeconômico linear. Recentemente, duas abordagens tem sido propostas como alternativas a esse arcabouço linear-quadrático. Na primeira delas, a política monetária ótima é investigada supondo modelos não-lineares para a estrutura econômica (por exemplo, Nobay e Peel [53] e Orphanides e Wieland [54]). Já a outra abordagem, com o intuito de modelar a existência de preferências assimétricas, considera funções não-quadráticas para representar a perda do Banco Central (por exemplo, Cukierman [24] e Surico [66]).

Assumindo a taxa de desconto intertemporal $\delta = 0.98$, foram consideradas as preferências descritas na tabela 5.6 a seguir, tendo em vista comparar os Critérios 1, 2 e 3 de condução da política monetária. A partir das regras obtidas, realiza-se uma análise das suas projeções obtidas para o período [2010/T3 - 2011/T4].

Tabela 5.6: Preferências Simuladas

Pesos na Função Perda	λ_π	λ_h	λ_i
Critério 1	1	0	0
Critérios 1 e 2	1	0,4	0
Critérios 1,2 e 3	1	0,4	4

Critério 1

Para as preferências associadas ao Critério 1, obteve-se a seguinte regra ótima de política monetária:

$$i_t = -0,2058\pi_t - 0,2458\pi_{t-1} + 2,6401\pi_t^e + 0,6628\pi_{t-1}^e + 1,4878h_t + 0,1417h_{t-1} - 0,05182c_t - 0,05840c_{t-1} + 0,1315i_{t-1}. \quad (5.21)$$

Como neste caso o único compromisso do Banco Central é a manutenção da inflação em torno da sua meta, a regra (5.21) apresenta coeficientes relativamente altos para a expectativa de inflação e o hiato do produto, o que está de acordo com a teoria econômica. Conforme mostra a Figura 5.2 ao final deste capítulo, no quarto trimestre de 2011, a condução ótima da política monetária sob o Critério 1 reduziria a inflação a um valor próximo do limite inferior da sua faixa de tolerância¹⁴. Para isso, logo no terceiro trimestre de 2010, a taxa Selic seria elevada a um nível acima dos 20% ao ano, implicando uma diminuição substancial do hiato do produto.

¹⁴Conforme estabelecido pelo Banco Central na Resolução 3.748 de 2009, a meta de inflação para 2011 é um IPCA de 4,5% com tolerância de dois pontos percentuais para mais ou para menos, ficando portanto limitado ao intervalo 2,5-6,5 (%).

Critérios 1 e 2

Com respeito às preferências que representam os Critério 1 e 2, obteve-se a seguinte regra ótima para a política monetária:

$$i_t = -0,0151\pi_t - 0,2297\pi_{t-1} + 1,8695\pi_t^e + 0,5630\pi_{t-1}^e + 2,3202h_t \quad (5.22) \\ + 0,1325h_{t-1} - 0,0256c_t - 0,04101c_{t-1} + 0,0328i_{t-1}.$$

Neste caso, o Banco Central busca controlar a inflação e manter a atividade econômica estável no seu nível potencial, caracterizando o chamado regime de metas para inflação *flexível*. Em razão disso, o valor do coeficiente do hiato é quase o dobro do seu correspondente na regra (5.21). Com respeito a dinâmica gerada por (5.22), na Figura 5.2 observa-se uma grande semelhança com aquela referente ao Critério 1. Entretanto, por conta do peso λ_h , os Critérios 1 e 2, juntos, implicam uma elevação maior da taxa Selic, a qual chegaria a um valor próximo de 24,5% ao ano já no terceiro trimestre de 2010. Como consequência, a atividade econômica seria reduzida ao seu nível potencial, com o hiato do produto praticamente nulo.

Critérios 1, 2 e 3

Finalmente, considerando as preferências que representam, conjuntamente, os Critérios 1, 2 e 3, obteve-se a seguinte regra ótima para a política monetária:

$$i_t = -0,0983\pi_t - 0,0184\pi_{t-1} + 0,3420\pi_t^e - 0,0127\pi_{t-1}^e + 0,1258h_t \quad (5.23) \\ - 0,0106h_{t-1} - 0,0007c_t - 0,0045c_{t-1} + 0,8163i_{t-1}.$$

Neste cenário, também tem-se um regime de metas para inflação flexível, no qual o Banco Central busca controlar a inflação e estabilizar a economia por meio de uma trajetória suave para a taxa Selic. Em decorrência disso, o coeficiente sobre i_{t-1} mostra-se sensivelmente maior que os seus correspondentes nas regras anteriores. Além disso, na Figura 5.2 observa-se uma elevação gradual da taxa Selic, chegando a um nível próximo de 12,5% a.a., ao final de 2011. Por conta dessa política, o IPCA acumulado em 12 meses terminaria exatamente sobre a sua meta central de 4,5%, enquanto a atividade econômica se manteria estável, mas ainda acima do seu nível potencial.

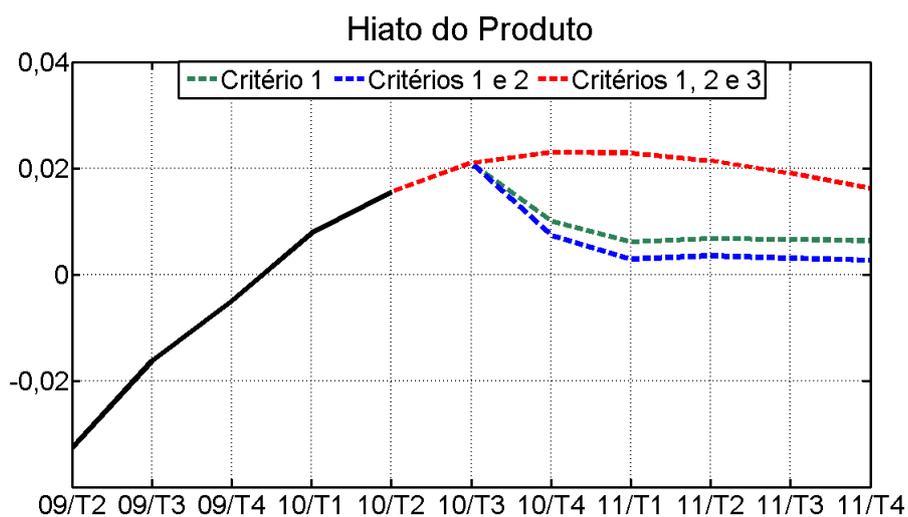
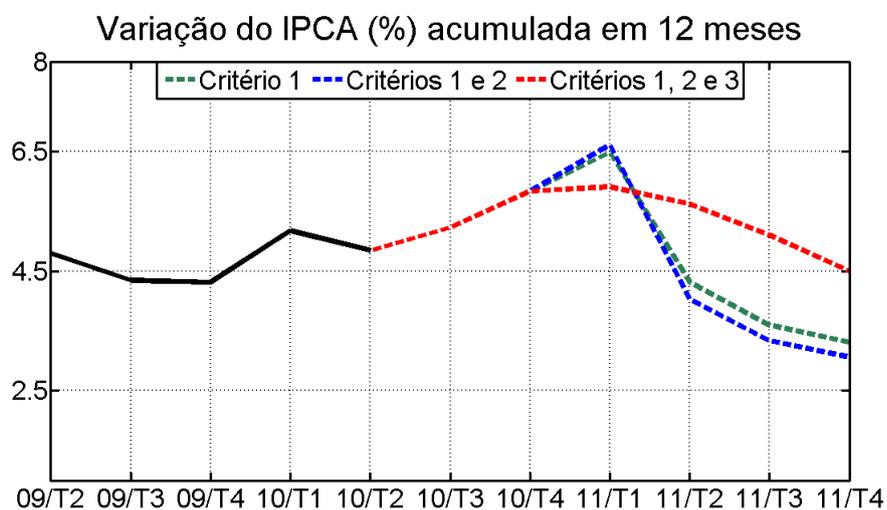
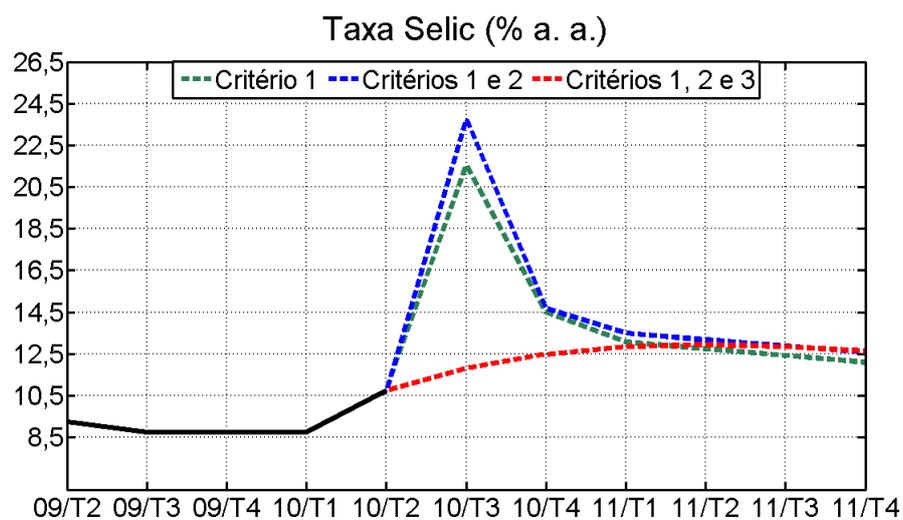


Figura 5.2: Projeções para diferentes preferências de política monetária.

Preferências do Banco Central do Brasil

Neste capítulo, a função perda do Banco Central do Brasil é calibrada utilizando-se um algoritmo genético com codificação real. Esse procedimento permite estimar as preferências que tem orientado o Banco Central na condução da política monetária.

6.1 Revisão da Literatura

A busca por um entendimento mais profundo da política monetária, em termos de metas e preferências, tem motivado vários trabalhos que se dedicam à estimação dos pesos que a autoridade monetária atribui às variáveis alvo na sua função perda.

Com respeito ao Banco Central dos EUA, o Federal Reserve (FED), Favero e Rovelli [29] e Ozlale [55] estimam as preferências monetárias em conjunto com os coeficientes do modelo macroeconômico que restringe a otimização. Para isso, esses autores consideram uma função perda com horizonte finito e utilizam os métodos de momentos generalizados (GMM) e máxima verossimilhança, respectivamente. Dennis [26, 27], por sua vez, supõe um horizonte de programação infinito e obtém estimativas para as preferências aplicando o método de máxima verossimilhança. De maneira diferente, Castelnuovo e Surico [19] e Söderström *et al.* [64] identificam as preferências do FED por um procedimento de calibragem da função perda, com o intuito de ajustar a política monetária ótima simulada com a política praticada no período sob análise. Em-

bora esses trabalhos considerem modelos, amostras e métodos de estimação diferentes, em relação ao período Volcker-Greenspan¹, todos eles sugerem que o FED deu um peso insignificante para o hiato do produto, enquanto atribuiu um peso considerável para a suavização da taxa de juros.

No que se refere às preferências monetárias em outros países, Cecchetti *et al.* [20] realizam um estudo sobre dez bancos centrais europeus, sete deles pertencentes à União Monetária Européia. Segundo esses autores, a maioria dos países analisados atribuiu um peso para a inflação significativamente maior que aquele dado ao hiato do produto. Por outro lado, usando um modelo VAR estrutural, Cecchetti e Ehrmann [21] estimam as preferências de política monetária de 23 países industrializados e em desenvolvimento. Em todos esses países, verificou-se um aumento da aversão à inflação durante os anos noventa, o qual se mostrou um pouco mais pronunciado nos países praticantes do regime de metas para a inflação. Lippi e Neri [45], estimam as preferências monetárias na Área do Euro pelo método de máxima verossimilhança, e concluem que a autoridade monetária européia tem atribuído um peso considerável sobre a suavização da taxa de juros, o qual se mostrou maior que o peso dado para a inflação e o hiato do produto. Aguiar e Martins [2] obtém estimativas para as preferências monetárias na Área do Euro durante o período 1995-2002, supondo um horizonte de programação de quatro trimestres e utilizando GMM para obter os parâmetros. De acordo com esse estudo, a função perda do Banco Central Europeu tem sido caracterizada por uma meta implícita de inflação em torno de 1,6%, com um peso significativo sobre a suavização da taxa de juros e nenhum peso sobre o hiato do produto. Ilbas [37], utilizando métodos bayesianos, também obtém estimativas para as preferências monetárias na Área do Euro. Seus resultados sugerem que a inflação é a maior preocupação da autoridade monetária européia, seguida pela suavização da taxa de juros e, por último, pelo hiato do produto.

Ampliando o estudo das preferências monetárias para além do eixo América do Norte e Europa, Colins e Siklos [23] estimam e comparam as preferências monetárias da Austrália, Canadá, Nova Zelândia e Estados Unidos, utilizando GMM e consideran-

¹Período de [1979 - 2006], no qual o FED esteve sob a presidência de Paul A. Volcker [1979 - 1987] e Alan Greenspan [1987 - 2006].

do o período a partir de 1988. Os resultados mostram que a política monetária do FED, sob a presidência de Alan Greenspan, pode ser associada a um regime de metas para a inflação ótimo, no qual há uma propensão significativa para a suavização da taxa de juros, com um pequeno peso dado ao hiato do produto. Para os outros países analisados, resultados similares foram obtidos, indicando que a política monetária nesses países tem sido muito parecida desde 1988. Tachibana [68] estima as preferências do Japão, Reino Unido e Estados Unidos, supondo um horizonte de programação de dois trimestres e utilizando o Método dos Mínimos Quadrados Ordinários. Nos três países analisados, após a primeira crise do petróleo em 1980, verifica-se um aumento do peso atribuído à inflação, principalmente no Reino Unido. Sánchez [62], obtém estimativas para o Banco Central da Coreia do Sul utilizando o método de máxima verossimilhança. Os resultados indicam que a autoridade monetária sul-coreana, embora tenha priorizado o controle da inflação, também atribuiu um peso considerável para a suavização da taxa de juros. Além disso, pesos negligenciáveis foram estimados para o hiato do produto e a taxa de câmbio, sugerindo que o controle dessas variáveis não constitui uma meta em si para a autoridade monetária sul-coreana.

Seguindo a metodologia proposta por Cecchetti e Ehrmann [21], Bandeón [7] estima as preferências do Banco Central de Reserva do Perú (BCRP) durante a década de noventa, obtendo um peso sobre a inflação bem maior que aquele atribuído ao hiato do produto. Rodriguez [60] também apresenta estimativas para as preferências do BCRP, supondo um horizonte de programação finito e utilizando GMM para obter os parâmetros. Embora seus resultados mostrem grande sensibilidade em relação ao método de cálculo do hiato do produto, observa-se uma redução significativa da meta implícita de inflação peruana ao longo do período 1979-2005. Mais recentemente, Pasca [56] identifica as preferências do BCRP durante o regime de metas para a inflação, utilizando um processo de calibragem da sua função perda. Os resultados sugerem que a autoridade monetária peruana tem priorizado o controle da inflação, com um peso considerável sobre a suavização da taxa de juros, acampanhada por um peso ínfimo para o hiato do produto.

Por fim, com respeito à política monetária brasileira, Aragón e Portugal [6] apresentam estimativas originais para as preferências do Banco Central do Brasil, utilizando

também um processo de calibragem da função perda. De acordo com esse estudo, a autoridade monetária brasileira tem praticado um regime de metas para a inflação flexível, onde grande peso é conferido ao controle da inflação, mas pouca importância é dada ao hiato do produto e à suavização da taxa de juros. Não obstante, o estudo indica que o Banco Central atribui um peso maior para a suavização da taxa de juros, em relação ao peso dada ao hiato do produto.

A seguir, procede-se a identificação das preferências do Banco Central do Brasil, a fim de confrontá-las com os resultados obtidos por Aragón e Portugal.

6.2 Estimação das Preferências do Banco Central do Brasil

Seguindo [6, 19, 23, 56, 64], a estratégia utilizada para estimar as preferências da autoridade monetária consiste em calibrar a função perda do Banco Central, de modo que a política ótima resultante apresente o melhor ajuste em relação à política monetária no período sob análise. Conforme salientam Castelnuovo e Surico [19], Aragón e Portugal [6] e Pasca [56], em relação aos métodos de máxima verossimilhança e de momentos generalizados, a vantagem do procedimento de calibragem reside na sua robustez à má-especificação dos erros no modelo macroeconômico que restringe a otimização.

Formalmente, a calibragem da função perda consiste na busca pelo vetor de parâmetros $(\delta, \lambda_\pi, \lambda_h, \lambda_i)$, cuja trajetória ótima da taxa de juros, dada por $\hat{i}_t = -FX_t$, apresenta o menor desvio quadrático médio em relação à trajetória efetivamente implementada pelo Banco Central. Dessa forma, as preferências monetárias são identificadas com a solução do seguinte problema de otimização:

$$\min_{(\delta, \lambda_\pi, \lambda_h, \lambda_i)} \left(\sum_{t=1}^T [i_t - \hat{i}_t(\delta, \lambda_\pi, \lambda_h, \lambda_i)]^2 \right) / T \quad (6.1)$$

$$\text{sujeito a } (\delta, \lambda_\pi, \lambda_h, \lambda_i) \in \Omega. \quad (6.2)$$

6.2.1 Método de Solução Usual - Busca em Malha

Devido o caráter peculiar do problema (6.1)-(6.2), os trabalhos anteriormente citados obtêm soluções sub-ótimas aplicando o método de busca em malha. Nessa abordagem, primeiro, o espaço Ω é discretizado, a fim de restringir a busca a um subconjunto finito $S \subset \Omega$. Em seguida, para todos os elementos $(\delta, \lambda_\pi, \lambda_h, \lambda_i) \in S$, o problema (5.17)-(5.18) correspondente é resolvido, resultando em regras ótimas do tipo $\hat{i}_t = -F(\delta, \lambda_\pi, \lambda_h, \lambda_i)X_t$. A partir dessas regras, substituindo os valores reais de X_t , calcula-se a trajetória ótima da taxa de juros associada a cada elemento de S . Ao final, dentre todos os vetores $(\delta, \lambda_\pi, \lambda_h, \lambda_i)$ considerados, as preferências monetárias são representadas por aquele que apresenta o menor desvio quadrático médio:

$$DQM(i_t, \hat{i}_t) = \left(\sum_{t=1}^T [i_t - \hat{i}_t(\delta, \lambda_\pi, \lambda_h, \lambda_i)]^2 \right) / T. \quad (6.3)$$

6.2.2 Método de Solução Alternativo - Algoritmo Genético

De acordo com a formalização do regulador linear-quadrático (5.17)-(5.18), essencialmente, a calibragem das preferências monetárias corresponde ao problema de se determinar matrizes de ponderação Q , R e W , tais que a solução ótima correspondente apresente características de desempenho desejadas, no caso, o melhor ajuste com a política monetária praticada. Sob essa perspectiva, esse tipo de problema tem sido amplamente estudado tendo como foco o desenvolvimento de sistemas de controle [14, 40]. Nesse escopo, a aplicação de algoritmos genéticos para a escolha das matrizes de ponderação tem sido proposta em diversos trabalhos, demonstrando um desempenho superior quando comparada às técnicas usualmente utilizadas [39, 49, 52, 59].

Reconhecendo essa similaridade que existe entre o projeto de sistemas de controle e a identificação das preferências monetárias, o presente trabalho propõe o uso de um algoritmo genético para resolver o problema de calibragem da função perda do Banco Central. Especificamente, o algoritmo genético utilizado possui a seguinte configuração²:

²Ver detalhes no Capítulo 4.

- **Codificação:** Emprega-se a codificação real pela facilidade de implementação, e também devido as vantagens discutidas na seção 4.4;
- **Espaço de Busca:** Respeitando a definição de δ , e buscando ampliar a busca tanto quanto possível, considera-se

$$\Omega = \{(\delta, \lambda_\pi, \lambda_h, \lambda_i) \in \mathbb{R}^4; 0 \leq \delta \leq 1, 0 \leq \lambda_\pi, \lambda_h, \lambda_i \leq 20\}; \quad (6.4)$$

- **População:** Composta por 40 elementos de Ω , é escolhida inicialmente de forma aleatória;
- **Aptidão:** Dado que o objetivo é resolver o problema (6.1)-(6.2), a aptidão de cada $(\delta, \lambda_\pi, \lambda_h, \lambda_i)$ em Ω é calculada como o inverso do desvio quadrático médio associado a esses pesos, isto é,

$$a(\delta, \lambda_\pi, \lambda_h, \lambda_i) = 1/D(\hat{i}_t, \hat{i}_t); \quad (6.5)$$

- **Seleção:** Realiza-se pelo Método da Roleta;
- **Recombinação:** Aplica-se o Cruzamento Aritmético, com $p_c = 0,6$;
- **Mutação:** Utiliza-se a Mutação Uniforme, com $p_m = 0,3$;
- **Critério de Parada:** Número Máximo de Iterações.

6.2.3 Resultados

Por se tratar de um procedimento estocástico, execuções repetidas de um algoritmo genético podem resultar em soluções subótimas distintas. Assim, o AG descrito acima foi executado seis vezes, sendo cada uma delas limitada por um número máximo de 500 iterações. Os resultados são reportadas na tabela 6.1. O menor desvio quadrático médio entre a trajetória ótima e a trajetória real da taxa Selic foi de 1,7764 pontos percentuais, e corresponde às preferências obtidas na quarta execução do algoritmo genético. Essas preferências diferem das demais, principalmente, no que se refere ao valor δ obtido (0,9991), o qual se aproxima bastante dos valores usualmente utilizados em trabalhos desse gênero [6, 51, 56]

Tabela 6.1: Pesos Calibrados pelo AG

Execução	δ	λ_π	λ_h	λ_i	DQM
1	0,1878	9,7353	16,6239	1,0172	1,9208
2	0,2175	8,8582	5,0123	0,5423	1,8357
3	0,2604	9,7968	11,5449	1,5277	1,8984
4	0,9991	11,3651	0,2032	19,2743	1,7764
5	0,1710	9,9038	8,1460	0,5727	1,8209
6	0,1793	9,3351	11,7320	0,8672	1,8478

Em razão dessa discrepância, e com o intuito de melhorar o ajuste obtido, o algoritmo genético foi aplicado novamente, restringindo a busca ao subconjunto:

$$S = \{(\delta, \lambda_\pi, \lambda_h, \lambda_i) \in \Omega; \delta = 0,9991\}. \quad (6.6)$$

Além disso, o tamanho da população foi reduzido para 30 elementos, enquanto o critério de parada foi limitado à 300 iterações. Os resultados desse experimento são apresentados na tabela 6.2.

Tabela 6.2: Pesos Calibrados pelo AG ($\delta = 0,9991$)

Execução	λ_π	λ_h	λ_i	DQM
1	7,0812	0,00152	13,6505	1,764855
2	5,8007	0,00016	10,7573	1,764299
3	7,8588	0,00120	14,3290	1,764304
4	7,7889	0,05534	14,1324	1,768254
5	7,4778	0,03353	13,4662	1,766994
6	6,0764	0,01478	11,5421	1,765792

As preferências que resultaram no melhor ajuste entre a política ótima simulada e a política monetária real foram aquelas obtidas na segunda execução do AG:

$$\delta = 0,9991, \lambda_\pi = 5,8007, \lambda_h = 0,00016 \quad \text{e} \quad \lambda_i = 10,7573. \quad (6.7)$$

Para efeitos de comparação, simulando as preferências estimadas por Aragón e Portugal³, obteve-se em um desvio quadrático médio significativamente maior (5,2411 pontos percentuais), o que pode ser atribuído às diferenças de modelagem, amostragem e método de estimação utilizados.

As preferências (6.7) sugerem que a autoridade monetária brasileira tem atribuído um peso considerável para a suavização da taxa de juros e um peso ínfimo para o hiato do produto, o que está de acordo com os resultados reportados pela literatura internacional com respeito a outros bancos centrais. No entanto, diferente do trabalho de Aragón e Portugal [6], o peso obtido para o controle da inflação é inferior ao peso dada para a suavização da taxa Selic. Isso vem de encontro com os resultados obtidos por Dennis [26] em relação ao FED, e por Lippi e Neri [45] para a Área do Euro, refletindo uma grande cautela do Banco Central do Brasil no que se refere à realização de alterações sobre a taxa Selic.

6.3 Política Adotada *versus* Política Ótima

A regra de política monetária ótima, associada às preferências (6.7), é dada por:

$$i_t = -0,2058\pi_t - 0,2458\pi_{t-1} + 2,6401\pi_t^e + 0,6628\pi_{t-1}^e + 1,4878h_t \quad (6.8) \\ + 0,1417h_{t-1} - 0,05182c_t - 0,05840c_{t-1} + 0,1315i_{t-1}.$$

A figura 6.1, a seguir, apresenta em um mesmo plano a trajetória real da taxa Selic e a trajetória ótima associada às preferências obtidas pelo algoritmo genético. Embora verifique-se um ajuste razoável entre a política ótima e a política adotada, algumas diferenças importantes podem ser salientadas. Segundo (6.8), durante o primeiro semestre de 2003, a decisão ótima de política monetária teria conduzido à valores da taxa Selic menores que aqueles praticados. Já no período seguinte [2003/T3 - 2004/T2], em consonância com a política ótima, o Banco Central promoveu uma redução substancial da taxa Selic. Todavia, observa-se que a taxa Selic manteve-se sempre abaixo da trajetória ótima postulada pela regra (6.8). A partir do segundo trimestre de 2004,

³ $\delta = 0,98, \lambda_\pi = 0,727, \lambda_h = 0,073$ e $\lambda_i = 0,2$.

o Banco Central interrompeu a trajetória decrescente da taxa Selic, dando início a um período de aperto monetário, o qual se estendeu até o terceiro trimestre de 2005. Nesse intervalo, nota-se uma elevação maior da taxa Selic, em relação àquela que seria realizada caso a política monetária fosse conduzida conforme a regra ótima. Finalmente, com respeito ao período restante [2005/T4 - 2010/T2], em geral, a trajetória da taxa Selic implementada pelo Banco Central fica sempre abaixo da trajetória ótima⁴.

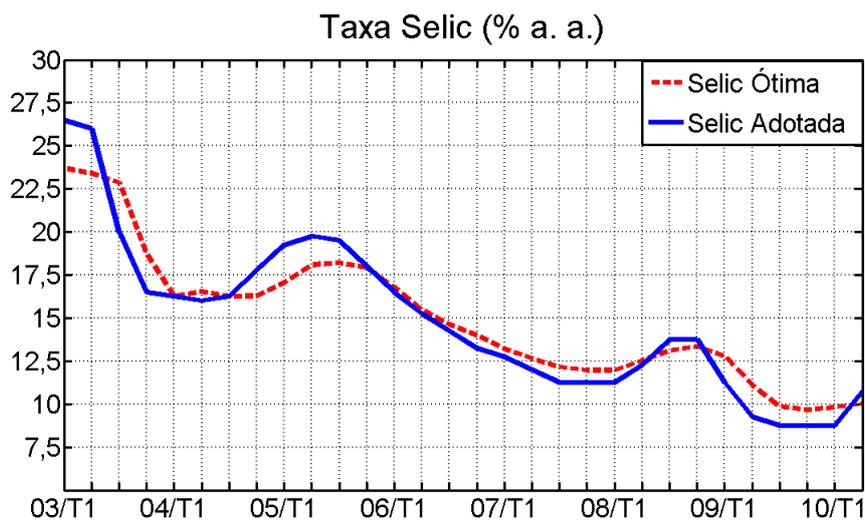


Figura 6.1: Taxa Selic adotada e taxa Selic ótima

6.4 Projeções Ótimas

Utilizando a regra (6.8), e supondo que o Banco Central mantenha esse padrão de reação, foram obtidas projeções para a taxa Selic, o IPCA acumulado em 12 meses e o hiato do produto, com respeito ao período [2010/T3 - 2011/T4]. Como essas projeções resultam da solução de um problema de otimização, elas são referidas como *projeções ótimas*.

⁴Qualitativamente, o comparativo entre a trajetória ótima da taxa Selic e a trajetória real adotada pelo Banco Central é idêntico àquela obtido por Aragón e Portugal [6], considerando [2003/T1 - 2007/T3].

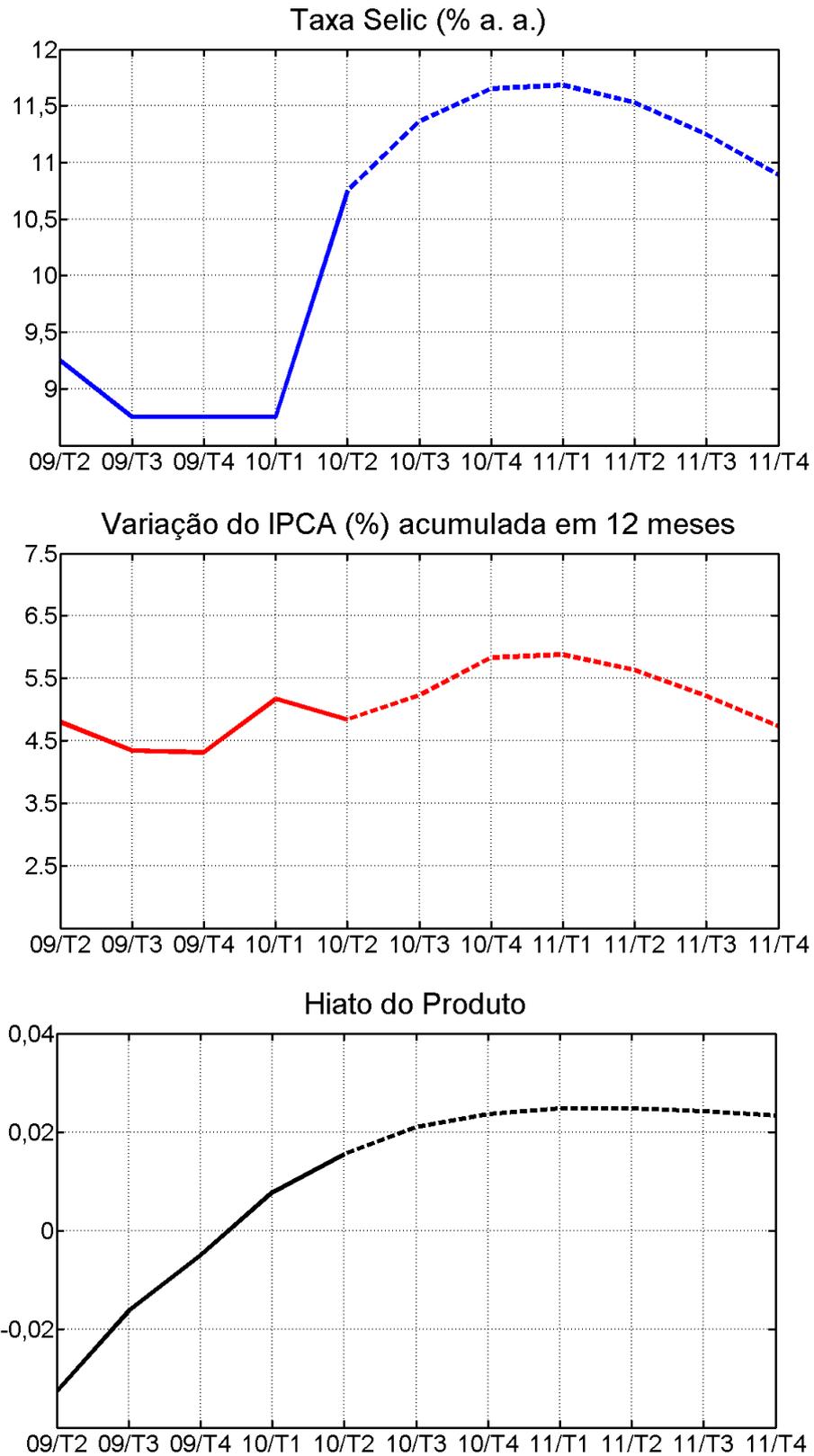


Figura 6.2: Projeções Ótimas para a Taxa Selic, a Inflação e o Hiato do Produto.

De acordo com os gráficos apresentados na Figura 6.2, sob as preferências (6.7), a política monetária ótima resultaria no retorno da inflação para a sua meta central de 4,5% ao final de 2011. Para isso, o aperto monetário iniciado no segundo trimestre de 2010 teria que ser mantido até o primeiro trimestre de 2011, com a elevação da taxa Selic a um patamar próximo de 11,75% ao ano. Como consequência disso, segundo as projeções, haveria uma estabilização do hiato do produto, ainda que a um nível acima do seu potencial.

6.4.1 Avaliação Preliminar

Considerando os dados disponíveis do período [2010/T3 - 2011/T1], a projeção ótima obtida para a taxa Selic mostrou-se mais acurada⁵ que a média das projeções de curto prazo realizadas pelo grupo TOP 5 do Relatório Focus⁶, e coletadas ao longo do mês de julho de 2010, como mostra a tabela 6.3.

Tabela 6.3: Desempenho das projeções para a taxa Selic

PROJEÇÃO	MRE (%)	RMSE (p.p)	MAE (p.p)
ÓTIMA	5,95	0,45	0,14
TOP 5	10,36	1,33	0,35

Embora a amostra seja pequena, esses resultados sugerem o grande potencial desse arcabouço teórico como ferramenta de previsão para a taxa Selic, e vem de encontro com a metodologia empregada pelo Norges Bank, o qual desde março de 2005 publica projeções ótimas para o seu próprio instrumento de política monetária [67].

⁵As medidas de erro utilizadas foram as seguintes: Erro Relativo Médio (MRE, do inglês, Mean Relative Error), Raíz Quadrada do Erro Quadrático Médio (RMSE, do inglês, Root Mean Square Error) e o Erro Absoluto Médio (MAE, do inglês, Mean Absolute Error).

⁶O TOP 5 consiste nas cinco melhores Instituições Financeiras participantes do Sistema de Expectativas de Mercado do Banco Central do Brasil no que se refere à acuidade das projeções.

Conclusão

Neste trabalho de mestrado foram identificadas as preferências que tem orientado o Banco Central do Brasil na condução da Política Monetária. Para isso, o processo decisório que envolve a escolha da taxa Selic foi modelado como um problema do regulador linear-quadrático estocástico, o que permitiu estabelecer uma correspondência entre as preferências do Banco Central e os pesos que definem a sua função perda.

Seguindo [6, 19, 23, 56], a estratégia utilizada consistiu em calibrar a função perda da autoridade monetária, de modo que a solução do seu problema de otimização resultasse no melhor ajuste entre a política real e a política ótima simulada. Entretanto, ao invés de escolher os pesos pelo método usual de busca em malha, o presente trabalho propôs o uso de um algoritmo genético com codificação real.

Com respeito à gestão de Henrique Meirelles, os resultados sugerem que o Banco Central do Brasil praticou um regime de metas para a inflação flexível. Em consonância com as preferências obtidas para outros bancos centrais [26, 27, 45], o peso estimado para o hiato do produto foi praticamente nulo, indicando que a sua presença na regra ótima deve-se à informação que ele contém sobre a inflação futura, e não que a sua estabilização seja uma meta em si. Todavia, divergindo do trabalho de Aragón e Portugal [6], o peso obtido para a suavização da taxa de juros se mostrou superior ao da inflação.

7.1 Sugestões para Trabalhos Futuros

O trabalho realizado nesta dissertação pode ser incrementado de várias maneiras. A fim de descrever algumas delas, é útil considerar o problema sob três aspectos: Modelagem, Métodos e Questões.

7.1.1 Modelagem

Com respeito à modelagem matemática da economia e da política monetária, as ações que podem ser adotadas incluem:

- (a) Acrescentar variáveis fiscais no modelo macroeconômico, tais como a dívida pública e o superávit primário, dada a relevância dessas variáveis no processo decisório do Copom [8];
- (b) Considerar a meta de inflação na função perda (em vez de $\pi^* = 0$), para que ela possa ser estimada junto com as preferências e comparada com a meta oficial anunciada pelo Banco Central;
- (c) Incluir na função perda um peso para o desvio da taxa Selic em relação ao valor preconizado por uma regra simples de taxa de juros (regra de Taylor), tal como faz o Norges Bank [4], a fim de tornar as projeções ótimas mais robustas frente à má-especificação do modelo.

7.1.2 Métodos

No que se refere aos métodos quantitativos, o presente trabalho pode ser melhorado pelas seguintes medidas:

- (a) Considerar outras formas de se estimar o hiato do produto, tais como o filtro Kalman e a função de produção, tendo em vista contornar o problema do filtro HP conhecido como “efeito de borda”[13];

- (b) Utilizar outros critérios e procedimentos para selecionar as defasagens no modelo, tais como o critério de Akaike ou a Simulação fora da Amostra (SOOS), a fim de obter previsões mais acuradas;
- (c) Estimar em conjunto todas as equações do modelo macroeconômico pelo Método dos Mínimos Quadrados Generalizados, frente uma eventual correlação contemporânea entre os erros nas diferentes equações;
- (d) Testar outras especificações para os operadores genéticos, com o intuito de melhorar o desempenho do AG;
- (e) Resolver a equação matricial algébrica de Riccati (5.20) utilizando o método de Newton, ou métodos baseados em subespaços invariantes [72], para que o cálculo das aptidões no algoritmo genético se dê de forma mais rápida, viabilizando assim experimentos maiores.

7.1.3 Questões

Por fim, no que tange aos temas de pesquisa, o instrumental teórico abordado nesta dissertação pode ser dirigido para as seguintes questões:

- (a) Analisar as preferências monetárias a partir de diferentes modelos, amostras e métodos, a fim de avaliar a robustez das estimativas obtidas;
- (b) Identificar as preferências monetárias durante o período em que Armínio Fraga esteve na Presidência do Banco Central do Brasil, com o intuito de testar se houve mudança significativa de conduta entre a sua gestão e a de Henrique Meirelles;
- (c) Acrescentar outras variáveis na função perda do Banco Central, tais como a dívida pública e a taxa de câmbio, tendo em vista investigar se o controle dessas variáveis constitui ou não uma meta implícita para a autoridade monetária brasileira.

Distribuições de Probabilidade

Servindo como base para as demonstrações realizadas ao longo do Capítulo 2, este apêndice enuncia algumas definições e teoremas referentes às distribuições normal, qui-quadrado, F e t de Student.

1.1 Distribuição Normal Multivariada

Definição A.1 (Distribuição Normal). Diz-se que um vetor aleatório $x = (x_1, \dots, x_n)$ tem distribuição normal multivariada se existirem $\mu \in \mathbb{R}^n$ e $\Sigma \in M(n \times n)$ simétrica definida positiva tais que a função densidade de probabilidade de x é dada por:

$$f(x) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det(\Sigma)}} \exp\left(-\frac{1}{2}(x - \mu)' \Sigma^{-1} (x - \mu)\right). \quad (\text{A.1})$$

Para indicar esse fato, escreve-se $x \sim N(\mu, \Sigma)$.

Observação A.2. Se μ e $\Sigma = \sigma^2$ são escalares, então a função dada em (A.1) se reduz à densidade da distribuição normal univariada

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right). \quad (\text{A.2})$$

Teorema A.3. Dado um vetor aleatório $x = (x_1, \dots, x_n)$, seja $y = Bx + d$, onde $d \in \mathbb{R}^m$ e $B \in M(m \times n)$. Nesse caso, se $x \sim N(\mu, \Sigma)$ então $y \sim N(B\mu + d, B\Sigma B')$

Demonstração: Ver Teorema C.1.1 em [25]. ■

1.2 Distribuições Correlatas

Definição A.4 (Distribuição Qui-quadrado). Diz-se que uma variável aleatória w possui distribuição qui-quadrado com n graus de liberdade, $w \sim \chi^2(n)$, se $w = z'z$ para $z = (z_1, \dots, z_n) \sim N(0, I_n)$.

Teorema A.5. Sejam $z = (z_1, \dots, z_n) \sim N(0, I_n)$ e $M \in M(n \times n)$ uma matriz simétrica idempotente com posto r . Então $z'Mz \sim \chi^2(r)$.

Demonstração: Ver Teorema C.3.3 em [25]. ■

Definição A.6 (Distribuição t de Student). Diz-se que uma variável aleatória y possui distribuição t de Student com n graus de liberdade, $y \sim t(n)$, se $y = \frac{z}{\sqrt{w/n}}$, sendo $z \sim N(0, I_n)$ e $w \sim \chi^2(n)$.

Definição A.7 (Distribuição F). Diz-se que uma variável aleatória y possui distribuição F se $y = \frac{w_1/n_1}{w_2/n_2}$, onde $w_1 \sim \chi^2(n_1)$ e $w_2 \sim \chi^2(n_2)$ são variáveis independentes. Nesse caso, emprega-se a notação $y \sim F(n_1, n_2)$.

As aplicações da distribuição F realizadas no Capítulo 1 tem como fundamento o seguinte teorema.

Teorema A.8. Sejam $z'Mz$ e $z'Qz$ formas quadráticas idempotentes sobre o vetor aleatório $z = (z_1, \dots, z_n) \sim N(0, I_n)$. Se $MQ = 0$ então essas variáveis aleatórias são independentes.

Demonstração: Ver Teorema C.4.1 em [25]. ■

Filtro Hodrick-Prescott

Teoricamente, o hiato do produto é a diferença entre o produto realizado e o produto potencial. Na prática, por se tratar de uma variável não-observável, o produto potencial costuma ser estimado como a tendência do produto realizado. Para extrair essa tendência, nesta dissertação, optou-se por utilizar o filtro Hodrick-Prescott (HP), tendo em vista o seu extenso uso em aplicações do gênero, inclusive pelo Banco Central do Brasil [10]. Este apêndice apresenta uma breve descrição sobre os principais aspectos desse filtro.

2.0.1 Procedimento para Extração da Tendência

Dada uma série de observações $\{y_t\}_{t=1}^n$ ordenadas no tempo, considere a seguinte decomposição:

$$y_t = g_t + c_t, \quad \text{para } t = 1, \dots, n, \quad (\text{B.1})$$

onde $\{g_t\}_{t=1}^n$ é a tendência da série e $\{c_t\}_t^n$ são as suas flutuações cíclicas.

Hodrick e Prescott [34] assumem como tendência a solução do seguinte problema de otimização irrestrita:

$$\min_{\{g_t\}_{t=1}^n} \left\{ \sum_{t=1}^n (y_t - g_t)^2 + \lambda \sum_{t=2}^{n-1} [(\tau_{t+1} - g_t) - (g_t - \tau_{t-1})]^2 \right\}. \quad (\text{B.2})$$

O primeiro termo em (B.2) é a soma dos quadrados dos desvios entre a série e a sua tendência, medindo o grau de ajuste entre elas. Por sua vez, o segundo termo é a soma

dos quadrados das diferenças entre componentes da própria tendência, expressando assim o seu grau de suavidade.

Para a análise que segue, é útil escrever o problema (B.2) em notação matricial:

$$\min_g f(g) = (y - g)'(y - g) + \lambda(Ag)'(Ag) \quad (\text{B.3})$$

onde

$$A = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & -2 & 1 & \dots & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \dots & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -2 & 1 \end{bmatrix}, \quad y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \cdot \\ \cdot \\ y_n \end{bmatrix}, \quad \text{e} \quad g = \begin{bmatrix} g_1 \\ g_2 \\ g_3 \\ \cdot \\ \cdot \\ g_n \end{bmatrix}$$

Nesses termos, observe que

$$\begin{aligned} f(g) &= (y - g)'(y - g) + \lambda(Ag)'(Ag) \\ &= y'y - y'g - g'y + g'g + \lambda g'A'Ag \\ &= y' - 2y'g + g'g + \lambda g'A'Ag \end{aligned}$$

Logo, diferenciando f com respeito à g e igualando o resultado a zero, obtém-se:

$$\begin{aligned} 0 &= -2Y + (I_n + I'_n)g + \lambda(A'A + (A'A)')g \\ \Rightarrow 0 &= -2Y + 2g + 2\lambda A'Ag \\ \Rightarrow y &= g + \lambda A'Ag \\ \Rightarrow y &= (I_n + \lambda A'A)g \\ \Rightarrow g &= (I_n + \lambda A'A)^{-1}y \end{aligned} \quad (\text{B.4})$$

Por outro lado, sendo

$$\frac{\partial f}{\partial g} = -2y + 2g + 2\lambda A'Ag$$

decorre que

$$\frac{\partial^2 f}{\partial g^2} = 2I_n + 2\lambda A'A = 2(I_n + \lambda A'A),$$

onde para todo $x \neq 0$ tem-se

$$x'2(I_n + \lambda A'A)x = 2(x'x + 2\lambda x'A'Ax) = 2(x'x + 2\lambda(Ax)'(Ax)) > 0, \text{ se } \lambda > 0.$$

Portanto, o vetor g especificado em (B.4) minimiza a função f , sendo assumido como a tendência de $\{y_t\}_{t=1}^n$. Como consequência imediata, pela equação (B.1) o componente cíclico dessa série é dado por:

$$c = y - g \quad (\text{B.5})$$

$$= y - (I_n + \lambda A'A)^{-1}y \quad (\text{B.6})$$

$$= (I_n - (I_n + \lambda A'A)^{-1})y \quad (\text{B.7})$$

2.0.2 Estimação do PIB Potencial

O hiato do produto utilizada neste trabalho foi obtido aplicando-se o filtro HP à série do logaritmo natural do PIB a preços de mercado dessazonalizado (ano-base=1995), referente ao período [2001/T3 - 2010/T2]. Seguindo [34], utilizou-se $\lambda = 1600$, por tratar-se de dados com periodicidade trimestral. A figura abaixo mostra a série do produto junto com o componente tendencial obtido pelo filtro:

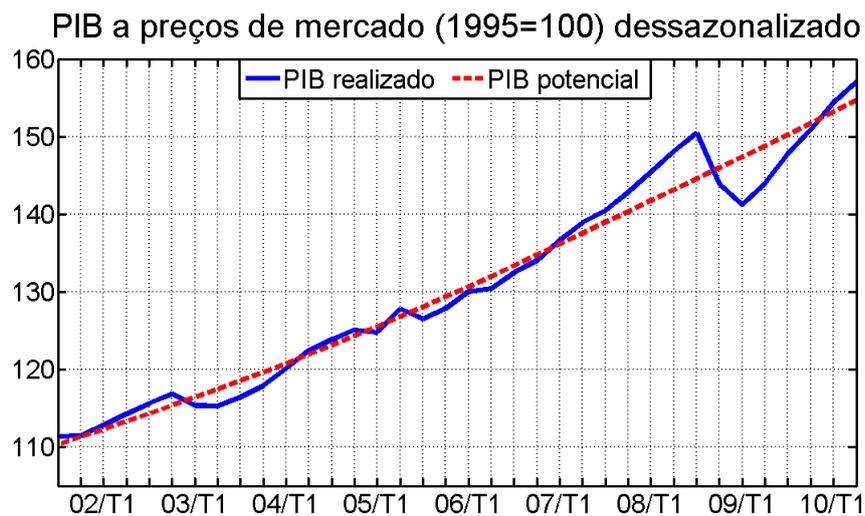


Figura B.1: Produto Interno Bruto potencial obtido pelo Filtro HP

Referências Bibliográficas

- [1] AIYAGARI, S. R. Response to a Defense of Zero Inflation. *Federal Reserve Bank of Minneapolis Quarterly Review*, 15, 2 (1991), 21-24.
- [2] AGUIAR, A. AND MARTINS, M. M. F. The Preferences of the Euro Area Monetary Policy-maker. *Journal of Common Market Studies* 43, 2 (2005), 221-250.
- [3] ANDRADE, J. P., AND DIVINO, J. A. C. A. Optimal Rules for Monetary Policy in Brazil. *Texto para Discussão n° 806*, IPEA, 2001.
- [4] ALSTADHEIM, R., *et al.* Monetary Policy Analysis in Practice. *Staff Memo n° 11*, Norges Bank, 2010.
- [5] ALVAREZ, V. H. *et al.* Parameter Estimation for VLE Calculation by Global Minimization: The Genetic Algorithm. *Brazilian Journal of Chemical Engineering* 25, 2 (2008), 409-418.
- [6] ARAGÓN, E. K. S. B., AND PORTUGAL, M. S. Central Bank Preferences and Monetary Rules under the Inflation Targeting Regime in Brazil. *Brazilian Review of Econometrics* 29, 1 (2009), 79-109.
- [7] BALDEÓN, W. B. Algunas lecciones y preferencias reveladas de la política monetaria del Perú en la última década. *Concurso de investigación para jóvenes economistas 2001-2002*, 2002.

- [8] BANCO CENTRAL DO BRASIL. *Regime de Metas para a Inflação no Brasil*. Disponível em: <http://www4.bcb.gov.br/pec/gci/port/focus/faq.asp>. Acesso em: 25 de janeiro de 2011.
- [9] BANCO CENTRAL DO BRASIL. *Relatório de Inflação, Volume 1, n°1*. Comitê de Política Monetária, 1999.
- [10] BANCO CENTRAL DO BRASIL. *Relatório de Inflação, Volume 1, n°2*. Comitê de Política Monetária, 1999.
- [11] BANCO CENTRAL DO BRASIL. *Relatório de Inflação, Volume 10, n°3*. Comitê de Política Monetária, 2008.
- [12] BANCO CENTRAL DO BRASIL. *Relatório de Inflação, Volume 11, n°2*. Comitê de Política Monetária, 2009.
- [13] BANCO CENTRAL DO BRASIL. *Relatório de Inflação, Volume 12, n°1*. Comitê de Política Monetária, 2010.
- [14] BARAKATI, S. M., APLEVICH, J. D., AND KAZERANI, M. Design of a State-Feedback Controller and Optimal Observer for a Wind Turbine System Including a Matrix Converter. In *40th North American Power Symposium, NAPS'08*, 2008.
- [15] BELLMAN, R. E. The Theory of Dynamic Programming. *Bulletin of the American Mathematical Society* 60, 6 (1954), 503-515.
- [16] BERTSEKAS, D. P. *Dynamic Programming and Optimal Control, Volume 1*. Athena Scientific, 1995.
- [17] BREUSCH, T. S., AND PAGAN, A. R. A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica* 47, 5 (1979), 1287-1294.
- [18] BROWN, T. A. *Genética: Um Enfoque Molecular*, 3 ed. Guanabara Koogan, 1999.
- [19] CASTELNUOVO, E., AND SURICO, P. What does Monetary Policy Reveal about Central Bank's Preferences? *Economic Notes by Banca Monte dei Paschi di Siena SpA* 39, 3 (2003), 335-359.

- [20] CECCHETTI, S. G., McCONNELL, M. M., AND QUIROS, G. P. Policymakers' Revealed Preferences and the Output-Inflation Variability Trade-Off: Implications for the European System of Central Banks. *Manchester School* 70, 4 (2002), 596-618.
- [21] CECCHETTI, S. G., AND EHRMANN, M. Does Inflation Targeting Increase Output Volatility? An International Comparison of Policymakers' Preferences and Outcomes. *NBER Working Paper n°7426*, National Bureau of Economic Research, 1999.
- [22] CLARIDA, R., GALÍ, J., AND GERTLER, M. Monetary policy rules in practice: Some international evidence. *European Economic Review* 41 (1998), 1033-1067.
- [23] COLLINS, S., AND SIKLOS, P. L. Optimal Monetary Policy Rules and Inflation Targets: Are Australia, Canada, and New Zealand Different from the U.S.? *Open Economies Review* 15, 4 (2004), 347-362.
- [24] CUKIERMAN, A., AND GERLACH, S. The Inflation Bias Revisited: Theory and Some International Evidence. *The Manchester School*, 71, 5 (2003), 541-565.
- [25] DAVIDSON, J. *Econometric Theory*. Blackwell Publishing, 2000.
- [26] DENNIS, R. Inferring policy objectives from economic outcomes. *Oxford Bulletin of Economics and Statistics* 66, Supplement (2004), 735-764.
- [27] DENNIS, R. The Policy Preferences of the US Federal Reserve. *Journal of Applied Econometrics* 21, 1 (2006), 55-77.
- [28] DICKEY, D. A., AND FULLER, W. A. Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association* 74, 366 (1979), 427-431.
- [29] FAVERO, C. A., AND ROVELLI, R. Macroeconomic stability and the preferences of the Fed. A formal analysis, 1961-98. *IGIER Working Papers n°200*, Innocenzo Gasparini Institute for Economic Research, 2001.
- [30] GODFREY, L. G. Testing Against General Autoregressive an Moving Average Error Models when the Regressors Include Lagged Dependent Variables. *Econometrica* 46, 6 (1978), 1293-1301.

- [31] GORDON, R. J. *Macroeconomia*, 7 ed. Bookman, 2000.
- [32] GUJARATI, D. *Econometria Básica*. Makron Books, 2005.
- [33] HAUPT, R. L., AND HAUPT, S. E. *Practical Genetic Algorithms*, 2 ed. Wiley-Interscience, 2004.
- [34] HODRICK, R. J., AND PRESCOTT, E. C. Post-War U.S. Business Cycles: An Empirical Investigation. *Discussion Papers n°451*, Northwestern University, Center for Mathematical Studies in Economics and Management Science, 1981.
- [35] HOLLAND, J. H. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence*. MIT Press, 1992.
- [36] HOSKINS, W. L. Defending Zero Inflation: All for Naught. *Federal Reserve Bank of Minneapolis Quarterly Review*, 15, 2 (1991), 16-20.
- [37] ILBAS, P. Estimation of monetary policy preferences in a forward-looking model: a Bayesian approach. *Working Paper Research n°129*, National Bank of Belgium, 2008.
- [38] INOUE, A., AND KILIAN, L. On the Selection of Forecasting Models. *Working Paper Series n°214*, European Central Bank, 2003.
- [39] JAZEARI, M., AND WANG, H. F. Multi-Mode Genetic Algorithm Based Linear Optimal Control Design for Power Systems. In *International Conference on Sustainable Power Generation and Supply, SUPERGEN'09*, 2009.
- [40] JOHNSON, M. A., AND GRIMBLE, M. J. Recent trends in linear optimal quadratic multivariable control system design. In *IEE Proceedings Part D Control Theory and Applications* 134, 1 (1987), 53-71.
- [41] JUDD, J. P., AND RUDEBUSCH, G. D. Taylor's Rule and the Fed: 1970 - 1997. *FRBSF Economic Review*, 3 (1998), 3-16.
- [42] KREYSZIG, E. *Introductory Functional Analysis with Application*. John Wiley and Sons, 1978.

- [43] LIMA, E. L. *Álgebra Linear*, 7 ed. IMPA, 2008.
- [44] LIMA, E. L. *Análise Real, Volume 1*. IMPA, 1989.
- [45] LIPPI, F., AND NERI, S. Information variables for monetary policy in an estimated structural model of the euro area. *Journal of Monetary Economics* 54, 4 (2007), 1256-1270.
- [46] LJUNGQVIST, L., AND SARGENT, T. *Recursive macroeconomic theory*, 2 ed. MIT Press, 2004.
- [47] LIU, Z., *et al.* Evolving Neural Network Using Real Coded Genetic Algorithm (GA) for Multispectral Image Classification. *Future Generation Computer Systems* 20, 7 (2004), 1119-1129.
- [48] MADDALA, G. S. *Introdução à Econometria*, 3 ed. LTC, 2003.
- [49] MANSOURI, N., AND KHALOOZADEH, H. The GA Approach to the Eigenvalue Placement of an LQ Optimal System. In *IEEE International Conference on Artificial Intelligence Systems, ICAIS'02*, 2002.
- [50] MOHANTY, M. S., AND KLAU, M. Monetary policy rules in emerging market economies: issues and evidence. *BIS Working Papers n°149*, Bank for International Settlements, 2004.
- [51] MORAIS, J. F. M., ANDRADE, J. P., AND SILVA, M. L. F. Silva. Política Monetária Ótima em um Contexto de Elevada Dívida Pública: O Caso Do Brasil. In *Anais do XXXII Encontro Nacional de Economia*, 2004.
- [52] NETO, J. V. F., *et al.* Modelos e Convergência de um Algoritmo Genético para Alocação de Auto-estrutura via RLQ. *IEEE Latin America Transactions* 6, 1 (2008).
- [53] NOBAY, A. R., AND PEEL, D. A. Optimal Monetary Policy with a nonlinear Phillips Curve. *Economic Letters*, 67, 2 (2000), 159-164.
- [54] ORPHANIDES, A., AND WIELAND, V. Inflation Zone Targeting. *Working Paper Series n°8*, European Central Bank, 1999.

- [55] OZLALE, U. Price Stability vs. output stability: tales of federal reserve administrations. *Journal of Economic Dynamics and Control* 27, 9 (2003), 1595-1610.
- [56] PASCA, N. M. C. Preferências do Banco Central de Reserva do Perú e Regras Monetárias Ótimas sob o Regime de Metas de Inflação. Dissertação de Mestrado, Universidade Federal do Rio Grande do Sul, 2010.
- [57] PEDERSEN, T. M. *Understanding Business Cycles with MATLAB applications*. Department of Economics, University of Copenhagen, 1999. Disponível em: <http://alumni.econ.queensu.ca/allenj/RBC/>. Acesso em: 04 de fevereiro de 2011.
- [58] PHILLIPS, P. C. B. Understanding Spurious Regressions in Econometrics. *Journal of Econometrics* 33, (1986), 311-340.
- [59] ROBAND, I., *et al.* Optimal feedback control design using genetic algorithm in multimachine power system. *International Journal of Electrical Power and Energy Systems* 23, 4 (2001), 263-271.
- [60] RODRÍGUEZ, G. Eficiencia de la política monetaria y la estabilidad de las preferencias del Banco Central. Evidência empírica para el Perú. *Revista Estudios Económicos*, 15 (2008), 9-20, Banco Central de Reserva del Perú.
- [61] RUDEBUSCH, G., AND SVENSSON, L. E. O. Policy rules for inflation targeting. *NBER Working Paper n°6512*, National Bureau of Economic Research, 1998.
- [62] SÁNCHEZ, M. What does South Korean inflation targeting target? *Journal of Asian Economics* 21, 6 (2010), 526-539.
- [63] SHAPIRO, S. S., AND WILK, M. B. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika* 52, 3-4 (1965), 591-611.
- [64] SÖDERSTRÖM, U., SÖDERLIND, P., AND VREDIN, A. Can a Calibrated New-Keynesian Model of Monetary Policy Fit the Facts? *Working Paper n°140*, Sveriges Riksbank, 2002.
- [65] STOKEY, N. L., R. E. LUCAS, AND PRESCOTT, E. C. Prescott (1989): *Recursive Methods in Economic Dynamics*. Harvard University Press, 1989.

- [66] SURICO, P. The Fed's monetary policy rule and U.S. inflation: The case of asymmetric preferences. *Journal of Economic Dynamic & Control*, 31, 1 (2007), 305-324.
- [67] SVENSSON, L.E.O. Optimal Inflation Targeting: Further Developments of Inflation Targeting. *Working Paper n°403* Central Bank of Chile, 2006.
- [68] TACHIBANA, M. Central Banks' preferences in Japan, the UK, and the US. *Japan and the World Economy*, 16 (2004), 81-93.
- [69] TAYLOR, J. B. Discretion versus police rules in practice. In *Carnegie-Rochester Conference Series on Public Policy*, 39 (1993), 195-214.
- [70] VASCONCELLOS, M. A. S., AND GARCIA, M. E. *Fundamentos de Economia*, 2 ed. Saraiva, 2005.
- [71] WHITE, H. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. *Econometrica* 48, 4 (1980), 817-838.
- [72] ZHOU, K., DOYLE, J. C., AND GLOVER, K. *Robust and Optimal Control*. Prentice Hall, 1995.