

UNIVERSIDADE FEDERAL DE PERNAMBUCO  
CENTRO DE CIÊNCIAS SOCIAIS APLICADAS  
DEPARTAMENTO DE ECONOMIA

MODELOS DE SÉRIES TEMPORAIS PARA A PREVISÃO DA  
ARRECADAÇÃO TRIBUTÁRIA FEDERAL

MARCELO LETTIERI SIQUEIRA

RECIFE  
JUNHO/2002

UNIVERSIDADE FEDERAL DE PERNAMBUCO  
CENTRO DE CIÊNCIAS SOCIAIS APLICADAS  
DEPARTAMENTO DE ECONOMIA

MODELOS DE SÉRIES TEMPORAIS PARA A PREVISÃO DA  
ARRECADAÇÃO TRIBUTÁRIA FEDERAL

MARCELO LETTIERI SIQUEIRA

Dissertação apresentada em cumprimento às exigências para obtenção do grau de Mestre em Economia pela Universidade Federal de Pernambuco, PIMES - Pós-Graduação em Economia.

Orientador: Prof. Dr. Ricardo Chaves Lima.

RECIFE  
JUNHO/2002

# MODELOS DE SÉRIES TEMPORAIS PARA A PREVISÃO DA ARRECADAÇÃO TRIBUTÁRIA FEDERAL

MARCELO LETTIERI SIQUEIRA

**Aprovada com Distinção** em: 20.06.2002.

## **Banca Examinadora:**

Prof. Dr. Ricardo Chaves Lima

PIMES/UFPE

Prof. Dr. José Lamartine Távora Júnior

PIMES/UFPE

Prof. Dr. Luiz Rodrigues Kehrlé

UFPB

Prof. Dr. Ricardo Chaves Lima  
Orientador

**Para Janaina, pelo apoio  
incondicional a todas as  
minhas iniciativas.**

## AGRADECIMENTOS

Ao Programa de Pós-Graduação em Economia (PIMES) do Departamento de Economia da Universidade Federal de Pernambuco pela oportunidade de cursar o Mestrado.

Ao Professor Ricardo Chaves Lima pela orientação na elaboração deste trabalho e pelos conhecimentos repassados nas aulas de Econometria 2, principalmente em relação à utilização do Software RATS, numa sincronia perfeita entre a análise teórica e a implementação econométrica dos modelos.

Aos Professores José Lamartine Távora Júnior e Luiz Rodrigues Kehrlé pelas modificações sugeridas e pelo apoio na defesa desta dissertação.

A todos os companheiros de turma que de alguma forma contribuíram para a conclusão do curso.

Aos amigos: Herbertes Cordeiro, pela companhia nos estudos e pelo auxílio no tratamento estatístico dos modelos; Nelson Leitão, pela imensa ajuda com a coleta de dados; Giubran Zarur, principal responsável pela minha escolha pelo curso de economia; e Alberto Dahia, incentivador e colaborador na definição do tema desta dissertação.

A todos da Secretaria da Receita Federal que direta ou indiretamente contribuíram para a realização deste trabalho, sendo impossível tentar listar todos aqui sem correr o risco de, injustamente, esquecer alguém.

Às secretárias do PIMES e, em especial, à amiga Manuela que, mesmo não mais exercendo oficialmente funções no Departamento de Economia, contribuiu de forma decisiva para o cumprimento de todas as questões burocráticas relacionadas ao curso.

# SUMÁRIO

<b>AGRADECIMENTOS</b> .....	<b>II</b>
<b>SUMÁRIO</b> .....	<b>III</b>
<b>RESUMO</b> .....	<b>VI</b>
<b>ABSTRACT</b> .....	<b>VII</b>
1 - INTRODUÇÃO.....	1
2 - REVISÃO BIBLIOGRÁFICA .....	4
2.1 - <i>PREVISÃO DE RECEITAS PÚBLICAS</i> .....	4
2.1.1 - O Processo de Previsão .....	5
2.1.2 - Os Métodos de Previsão.....	6
2.1.2.1 - Métodos Qualitativos de Previsão .....	6
2.1.2.2 - Métodos Quantitativos de Previsão.....	8
2.2 – <i>MODELOS DE SÉRIES TEMPORAIS</i> .....	16
2.2.1 - Definições .....	17
2.2.2 – Processo Linear Geral .....	21
2.2.3 – Casos Particulares dos Modelos Lineares.....	23
2.2.3.1. Modelo Auto-Regressivo de Ordem $p$ : $AR(p)$ .....	23
2.2.3.2. Modelo de Médias Móveis de ordem $q$ : $MA(q)$ .....	28
2.2.3.3. Modelo Misto Auto-Regressivo e Médias Móveis de Ordem $(p,q)$ : $ARMA(p,q)$ .....	32
2.2.4 – Condições Formais para Estacionariedade e Invertibilidade .....	34
2.2.4.1. Estacionariedade .....	34
2.2.4.2. Invertibilidade .....	36
2.2.5 – Utilidade dos Modelos ARMA $(p,q)$ .....	38
2.2.6 – Estimação em Modelos Lineares de Séries Temporais.....	39
2.2.6.1 - Estimação da média do processo. ....	39
2.2.6.2 - Estimação da função de autocovariância e da função de autocorrelação. ....	40
2.2.6.3 - Estimação de parâmetros .....	41
2.2.7 – Construindo um Modelo ARIMA - Metodologia .....	44
2.2.7.1 - Identificação .....	45
2.2.7.2 – Estimação.....	48
2.2.7.3 – Verificação de Diagnóstico .....	50
2.2.8 – Modelando Séries Temporais Sazonais – O Modelo Sazonal Auto-Regressivo Integrado de Médias Móveis (SARIMA) .....	51
2.2.9 – Métodos de Comparação de Previsões .....	53
2.2.10 – Justificativas da escolha do Método de Box-Jenkins.....	54
2.3 – <i>MÉTODO DOS INDICADORES DA SRF</i> .....	56
3 – NATUREZA E TRATAMENTO DOS DADOS.....	59
3.1 - <i>PRINCIPAIS FATOS QUE AFETARAM A ARRECADAÇÃO NO PERÍODO 1990-2000</i> .....	60
3.2 - <i>ANÁLISE INDIVIDUAL DAS SÉRIES TRIBUTÁRIAS</i> .....	68

3.2.1 - Imposto sobre as Importações - II .....	69
3.2.2 - Imposto sobre Produtos Industrializados - IPI .....	71
3.2.3 - Imposto sobre a Renda (Total) - IR.....	75
3.2.4 - Outros Impostos - OI .....	77
3.2.5 - Receita de Impostos - RIMP .....	79
3.2.6 - Contribuição Provisória sobre Movimentação Financeira - CPMF.....	81
3.2.7 - Contribuição para Financiamento da Seguridade Social - COFINS e Contribuição para o PIS/PASEP .....	83
3.2.8 - Outras Contribuições - OCONTR.....	85
3.2.9 - Receita Total de Contribuições - RCONT .....	88
3.2.10 - Receita Administrada pela SRF - RECADM .....	90
<b>4 - RESULTADOS DA MODELAGEM BOX-JENKINS.....</b>	<b>95</b>
<b>4.1 - SÉRIE DO IMPOSTO DE IMPORTAÇÃO (II) .....</b>	<b>95</b>
4.1.1 - Identificação.....	95
4.1.2 - Estimção .....	99
4.1.3 - Checagem de Diagnóstico.....	100
4.1.4 - Previsão.....	101
<b>4.2 - SÉRIE DO IMPOSTO SOBRE PRODUTOS INDUSTRIALIZADOS (IPI).....</b>	<b>102</b>
4.2.1 - Identificação.....	102
4.2.2 - Estimção.....	107
4.2.3 - Checagem de Diagnóstico.....	107
4.2.4 - Previsão.....	108
<b>4.3 - SÉRIE DO IMPOSTO DE RENDA TOTAL (IR).....</b>	<b>110</b>
4.3.1 - Identificação.....	110
4.3.2 - Estimção.....	115
4.3.3 - Checagem de Diagnóstico.....	115
4.3.4 - Previsão.....	116
<b>4.4 - SÉRIE OUTROS IMPOSTOS (OI).....</b>	<b>118</b>
4.4.1 - Identificação.....	118
4.4.2 - Estimção.....	122
4.4.3 - Checagem de Diagnóstico.....	123
4.4.4 - Previsão.....	123
<b>4.5 - SÉRIE RECEITA TOTAL DE IMPOSTOS (RIMP) .....</b>	<b>125</b>
4.5.1 - Identificação.....	125
4.5.2 - Estimção.....	129
4.5.3 - Checagem de Diagnóstico.....	130
4.5.4 - Previsão.....	130
<b>4.6 - SÉRIE DA CPMF .....</b>	<b>133</b>
4.6.1 - Identificação.....	133
4.6.2 - Estimção.....	138
4.6.3 - Checagem de Diagnóstico.....	138
4.6.4 - Previsão.....	139
<b>4.7 - SÉRIE RECEITA DE OUTRAS CONTRIBUIÇÕES (OCONTR).....</b>	<b>141</b>
4.7.1 - Identificação.....	141
4.7.2 - Estimção.....	145

4.7.3 - Checagem de Diagnóstico.....	146
4.7.4 - Previsão.....	146
4.8 - <i>SÉRIE RECEITA TOTAL DE CONTRIBUIÇÕES (RCONT)</i> .....	148
4.8.1 - Identificação.....	148
4.8.2 - Estimção.....	152
4.8.3 - Checagem de Diagnóstico.....	153
4.8.4 - Previsão.....	153
4.9 - <i>SÉRIE DA RECEITA TOTAL ADMINISTRADA PELA SRF (RECADM)</i> .....	155
4.9.1 - Identificação.....	155
4.9.2 - Estimção.....	159
4.9.3 - Checagem de Diagnóstico.....	160
4.9.4 - Previsão.....	160
5 – DISCUSSÃO DOS RESULTADOS .....	162
5.1 – <i>SÉRIE DO IMPOSTO DE IMPORTAÇÃO – II</i> .....	162
5.2 - <i>SÉRIE DO IMPOSTO SOBRE PRODUTOS INDUSTRIALIZADOS - IPI</i> .....	164
5.3 - <i>SÉRIE DO IMPOSTO DE RENDA TOTAL - IR</i> .....	165
5.4 - <i>SÉRIE OUTROS IMPOSTOS - OI</i> .....	167
5.5 - <i>SÉRIE RECEITA TOTAL DE IMPOSTOS - RIMP</i> .....	169
5.6 - <i>SÉRIE DA CPMF</i> .....	171
5.7 - <i>SÉRIE RECEITA DE OUTRAS CONTRIBUIÇÕES - OCONTR</i> .....	174
5.8 - <i>SÉRIE RECEITA TOTAL DE CONTRIBUIÇÕES - RCONT</i> .....	175
5.9 - <i>SÉRIE RECEITA TOTAL ADMINISTRADA PELA SRF - RECADM</i> .....	177
6 – CONCLUSÃO E SUGESTÕES DE TRABALHOS FUTUROS .....	180
6.1 - <i>CONCLUSÕES GERAIS</i> .....	180
6.2 - <i>TRABALHOS FUTUROS</i> .....	182
REFERÊNCIAS BIBLIOGRÁFICAS.....	185
<b>APÊNDICE 1</b> .....	<b>190</b>
<b>APÊNDICE 2</b> .....	<b>197</b>

## RESUMO

A presente dissertação tem como principal objetivo apresentar uma metodologia alternativa para a previsão das receitas tributárias federais administradas pela Secretaria da Receita Federal.

Optou-se, então, pela utilização da metodologia de Box-Jenkins, especificamente o modelo Auto-Regressivo Integrado de Médias Móveis Sazonal – SARIMA.

O estudo será apresentado em 4 partes principais, assim desenvolvidas: revisão bibliográfica, natureza dos dados, apresentação e discussão dos resultados.

A revisão bibliográfica trará uma discussão geral sobre a previsão de receitas públicas e sobre o tratamento estatístico de séries temporais; apresentando, inclusive, a justificativa da opção pela metodologia de Box-Jenkins. Particularmente, em relação às séries temporais, serão apresentados os principais conceitos que norteiam o seu estudo, tais como: estacionariedade, estocasticidade, processos auto-regressivos – AR(p), de médias móveis – MA(q) e mistos (auto-regressivos e de médias móveis) – ARMA(p, q).

Na parte referente à natureza dos dados, serão apresentadas as séries tributárias objetos do presente estudo, os fatos que afetaram a arrecadação tributária federal no período, além do tratamento imposto aos dados em relação às mudanças estruturais observadas, à remoção de pontos discrepantes (“outliers”) e à correção com base em um índice geral de preços.

A parte referente à apresentação dos resultados trará os procedimentos utilizados na modelagem das séries tributárias, procedendo-se à identificação do modelo e sua respectiva estimação, à verificação de diagnóstico e, por fim, à previsão para cada um dos modelos selecionados.

Na parte final do trabalho, discutir-se-á os resultados obtidos, comparando-se as previsões obtidas pela metodologia de Box-Jenkins com aquelas originadas do método de indicadores utilizado pela Secretaria da Receita Federal, determinando-se quais seriam os melhores modelos para previsão dos valores futuros de cada uma das séries tributárias estudadas.

## ABSTRACT

This dissertation aims to advance an alternative methodology for forecast of the federal government receipts under management of Secretaria da Receita Federal (SRF).

We have chosen the Box-Jenkins Methodology, in special the Seasonal Autoregressive Integrated Moving Average – SARIMA.

The study consists of four main parts: literature review, type of data, presentation and analysis of results.

The literature review presents a discussion of aspects of revenue forecasting in the public sector and statistics properties of time series, and a justification for the methodology that was adopted in the study. In specific, with reference to time series, the main concepts that are used in this work are presented, such as: stationarity, expectations and stochastic processes, autoregressive – AR(p), moving average - MA(q) and autoregressive moving average – ARMA(p,q) processes.

As to data type, the study shows what series of taxes are used, the economics facts that have an influence on federal taxes collection in the period of analysis, and the modifications in data to remove outliers, to identify structural changes and to expurgate the inflation effects.

The presentation of results shows the procedures for modeling the taxes series, with the identification and estimation of the model, diagnostic's checks and, in the end, project revenue collection in future years.

In the final part of this work, an analysis of the results is conduct in order to compare the Box-Jenkins models with the model that is used for SRF, identifying the betters models for forecast of' the taxes revenue under analysis.

# 1 - INTRODUÇÃO

A Secretaria da Receita Federal (SRF), órgão específico singular, diretamente subordinado ao Ministro de Estado da Fazenda, tem como uma de suas finalidades, devidamente estabelecida em seu Regimento Interno (Portaria MF n.º 259, de 24 de agosto de 2001), a de realizar a previsão, o acompanhamento, a análise e o controle das receitas sob sua administração, bem assim coordenar e consolidar as previsões das demais receitas federais, para subsidiar a proposta orçamentária da União.

Especificamente no âmbito da SRF, a atividade de previsão concentra-se na estimação da arrecadação do exercício seguinte de todos os tributos por ela administrados<sup>1</sup>. Para executar essa tarefa, a Receita Federal optou por recorrer a um método denominado, institucionalmente, de método dos indicadores<sup>2</sup>.

Conforme já citado anteriormente, o presente trabalho tem por objetivo principal desenvolver um método de previsão de receitas tributárias federais baseado em modelos estatísticos e econométricos consistentes, como alternativa ao método ora utilizado pela SRF.

Para verificar a validade de se buscar alternativas de metodologias de previsão, analisou-se o poder preditivo do método de indicadores, comparando-o ao método de previsão que se utiliza de modelos ARIMA (metodologia Box-Jenkins). A análise detalhada dos procedimentos foi efetuada para as séries temporais da arrecadação individual e agregada dos diversos tributos do Sistema Tributário Brasileiro, para o período de janeiro de 1989 a dezembro de 2000, com os valores reais arrecadados nos meses do ano de 2001 servindo como parâmetros de comparação para as previsões geradas.

A presente dissertação apresenta quatro partes principais, além dessa introdução e da conclusão final, assim distribuídas: uma revisão bibliográfica que aborda os conceitos básicos de séries temporais e a atividade desenvolvida na área de previsão de receitas públicas; a apresentação da natureza e evolução temporal dos dados, com o respectivo tratamento de eventuais discrepâncias; a apresentação dos resultados obtidos na previsão dos valores futuros das séries estudadas, para o ano de 2001, pelos método dos indicadores e pelo método de Box-Jenkins; e a discussão completa dos resultados finais obtidos.

Em relação à organização da presente dissertação, as diversas partes acima referidas estão distribuídas por mais 5 capítulos, além do presente.

No capítulo 2, inicialmente é apresentada uma resenha literária concernente aos

---

<sup>1</sup> As receitas da contribuição previdenciária são administradas pelo Ministério da Previdência.

<sup>2</sup> A apresentação completa do método de indicadores utilizado pela SRF pode ser vista na seção 2.3 do Cap. 2.

estudos de previsão de receitas no âmbito das administrações públicas em geral. Nesta parte, discute-se os aspectos econômicos envolvidos na atividade de previsão de receitas governamentais, a natureza do processo de previsão e, por fim, os métodos de previsão recomendados pelos pesquisadores da área. Em seguida, procede-se à compilação dos principais conceitos sobre séries temporais, que envolvem, entre outras coisas, a caracterização da estacionariedade de uma série, a definição das funções de autocorrelação e de autocorrelação parcial, o desenvolvimento estatístico dos modelos auto-regressivos (AR), de médias móveis (MA) e mistos (ARMA), dos modelos integrados (ARIMA) e dos modelos sazonais (SARIMA). A análise se concentra na apresentação teórica da metodologia de Box-Jenkins, desenvolvendo as três etapas do ciclo iterativo que a compõe. Complementa a revisão da literatura a exposição dos métodos mais utilizados para a comparação dos resultados de previsão, os chamados índices de acurácia ou precisão e finaliza-se justificando a escolha do método.

No capítulo 3 é feita uma análise geral dos dados utilizados na presente dissertação, ressaltando-se os principais fatos que influenciaram a arrecadação de tributos no período considerado. Verifica-se a ocorrência de eventuais mudanças estruturais e procede-se à correção de pontos considerados discrepantes na série (“outliers”).

No capítulo 4 é efetuada a modelagem das séries individuais e agregadas corrigidas, gerando-se as previsões para cada uma delas, através da utilização da metodologia de Box-Jenkins. Todas as etapas dessa metodologia são explicadas individualmente para cada série, e os possíveis modelos são escolhidos por meio de um dos critérios de seleção de modelos apresentados no capítulo 2, particularmente o *BIC*. Os modelos escolhidos são testados pelos métodos de diagnósticos citados, rejeitando-se aqueles que se mostrarem inadequados em prol dos que lograrem resultados satisfatórios nos referidos testes.

No capítulo 5, os resultados obtidos no capítulo 4 são comparados entre si e com os apresentados pelo método dos indicadores da SRF, utilizando-se os valores da RQEQMP<sup>3</sup>, medida de acurácia escolhida para a confrontação dos modelos. Por fim, determina-se qual método possui melhores capacidades preditivas, dependendo do tipo de tributo e do horizonte de previsão considerados.

O 6º capítulo traz a conclusão do trabalho, aduzindo sobre a legitimidade do método de previsão empregado na dissertação, explicitando a necessidade de se promover pesquisas

---

<sup>3</sup> Trata-se da Raiz Quadrada do Erro Quadrado Médio de Previsão, a ser definida no capítulo 2.

adicionais sobre o tema e, por fim, recomendando a possibilidade de novos trabalhos.

## 2 - REVISÃO BIBLIOGRÁFICA

*Neste capítulo, apresenta-se uma resenha literária sobre a previsão de receitas públicas e sobre o tratamento estatístico de séries temporais, justificando-se a opção pelo método SARIMA. Serão apresentados, detalhadamente, os conceitos teóricos relevantes de séries temporais que subsidiarão a utilização do método de análise de séries temporais desenvolvido por Box & Jenkins (1976). Ao fim, apresenta-se detalhadamente o método de indicadores utilizado pela Secretaria da Receita Federal para a previsão da Arrecadação Tributária Federal.*

### 2.1 - PREVISÃO DE RECEITAS PÚBLICAS

A previsão de receitas envolve o uso de técnicas analíticas para projetar a quantidade de recursos financeiros disponíveis num determinado tempo futuro. No setor público, as receitas vêm dos impostos, das taxas, das contribuições sociais ou de transferências intergovernamentais, além daquelas relativas à atividade governamental empresarial. A previsão tenta identificar o relacionamento entre os fatores que afetam as receitas (alíquotas tributárias, variáveis macroeconômicas, volume de vendas a varejo, etc.) e a arrecadação governamental (tributos e outras receitas). A habilidade de projetar exatamente os recursos futuros é bastante crítica, na medida em que se objetiva, entre outras coisas, evitar desequilíbrios orçamentários ou uma excessiva arrecadação de tributos. Para o governo federal, mesmo pequenos erros na previsão de receitas podem resultar em problemas orçamentários sérios, principalmente se se tratarem de grandes déficits, objeto de preocupação constante das autoridades governamentais.

É importante observar que a previsão de receitas é fundamental para todos os níveis de governo (federal, estadual e municipal). Como os governos, em geral, continuam a deslocar o foco da tributação, da propriedade (mais estáveis) para as receitas baseadas no consumo (bastante instáveis), prevê-las será cada vez mais importante para a manutenção da atividade estatal.

As previsões das receitas podem ser aplicadas tanto à receita total agregada quanto às suas fontes individuais, tais como aquelas oriundas de impostos sobre o consumo (ICMS, IPI, IVA, etc.) ou as originadas de impostos sobre a propriedade (IPVA, IPTU, ITR, etc.), entre outras. Obviamente, não há nenhum método único para se projetar receitas. Assim sendo, os diferentes métodos pesquisados tendem a trabalhar melhor em função do tipo de receita que

se pretende prever. Da mesma forma, não há nenhuma estrutura padrão no tempo sobre a qual se pode fundar todas as previsões. Os governos podem ter como alvo a previsão do orçamento do ano seguinte, enquanto os gerentes de uma agência de energia elétrica, por exemplo, provavelmente estarão interessados num horizonte de tempo de vinte anos ou mais.

Finalmente, a previsão de receitas está intimamente ligada ao processo de política pública e, portanto, sujeita ao escrutínio e até mesmo à pressão política, que influenciarão decisivamente os resultados da arrecadação.

### **2.1.1 - O Processo de Previsão**

A política fiscal do governo é afetada pelo contexto em que ela é formada. Lida tanto com interesses políticos, quanto econômicos. Portanto, é essencial estabelecer as hipóteses e os procedimentos com os quais as partes interessadas concordam, assim como um mecanismo para avaliar a validade da receita que se prevê. Assim sendo, é fundamental que se estabeleça um processo bastante disciplinado.

Guajardo e Miranda (2000) sugerem um processo de sete estágios. Segundo eles, as seguintes etapas devem ser aplicadas a cada tipo de receita a ser prevista:

1º - A primeira etapa envolve a seleção de um período de tempo sobre o qual os dados serão examinados. O tamanho do período de tempo depende da disponibilidade e da qualidade dos dados, do tipo de receita a ser prevista, e do grau de exatidão desejado.

2º - Na segunda etapa, examina-se os dados a fim de determinar todos os padrões, taxas de variação, ou tendências que estejam evidentes. Aqui, os testes padrões podem sugerir que as taxas de variação sejam relativamente estáveis ou que variam, por exemplo, exponencialmente. Uma vez identificada qualquer tendência, o responsável pela previsão necessitará decidir em que grau a receita é previsível. Isto é feito examinando-se as características subjacentes da receita, tais como as estruturas de alíquotas usadas, as mudanças na demanda, ou as variações sazonais ou cíclicas.

3º - Os responsáveis pela previsão, em seguida, precisam compreender as hipóteses subjacentes associadas às fontes de receitas. Necessitam considerar a que grau a receita é afetada por circunstâncias econômicas, mudanças na demanda do cidadão e nas políticas do governo, entre outros efeitos. Estas suposições ajudam a determinar que método de previsão será mais adequado a cada caso.

4º - A etapa seguinte consiste em projetar os valores dos grupos de receitas para o horizonte de tempo desejado. O método selecionado para executar a projeção depende da

natureza e do tipo de receita. As fontes de receita com um grau elevado de incerteza, tais como receitas novas e de concessões ou vendas de patrimônio público (privatizações), podem empregar um método qualitativo de previsão, tal como a previsão de consenso, analisada mais adiante. As receitas que possuem baixo grau de incerteza, por sua vez, serão previstas, normalmente, através do emprego de um método quantitativo, baseados na análise de séries temporais ou na análise de regressão.

5º - Depois que as projeções forem feitas, as estimativas devem ser avaliadas para que se verifique a confiabilidade e validade dos procedimentos. Para avaliar a validade das estimativas, as hipóteses associadas às fontes de receitas devem ser reexaminadas. Se as suposições associadas ao ambiente político, econômico e administrativo existente se comprovarem coerentes, as projeções serão consideradas válidas. A confiabilidade é avaliada através de uma análise de sensibilidade, que envolve a verificação da variação dos parâmetros-chaves usados para criar as estimativas. Se isso resultar em grandes mudanças nas estimativas, supõe-se que a projeção tenha um baixo grau de confiabilidade.

6º - Na sexta etapa, os grupos de receitas são monitorados e comparados às estimativas. Tal monitoramento serve para avaliar a exatidão das projeções e determinar a probabilidade de ocorrência de quedas ou aumentos no orçamento.

7º - Finalmente, como as circunstâncias que afetam a geração de receitas governamentais mudam, a previsão precisará atualizar-se. As flutuações na arrecadação podem ser causadas por mudanças inesperadas nas condições econômicas e políticas, nos ajustes administrativos, ou nos padrões da demanda de consumidor.

### **2.1.2 - Os Métodos de Previsão**

Existe uma grande quantidade de técnicas de previsão disponíveis (Frank, 1993; Makridakis e Wheelwright, 1987, 1989; e Guajardo e Miranda, 2000). Elas variam de técnicas qualitativas relativamente informais a técnicas quantitativas altamente sofisticadas. Na previsão de receitas, mais sofisticado não significa, necessariamente, mais exato. Em geral, os responsáveis pela previsão usam uma variedade de técnicas, reconhecendo que algumas apresentam melhores resultados do que outras, dependendo da natureza da fonte de receita.

#### **2.1.2.1 - Métodos Qualitativos de Previsão**

Os métodos qualitativos de previsão se baseiam em conjeturas sobre a arrecadação futura de determinadas receitas. Estas técnicas são denominadas freqüentemente de aproximações conjeturais ou não-extrapolativas. Além de não apresentar uma formalidade

matemática mínima, tais técnicas freqüentemente não fornecem especificações rigorosas das hipóteses subjacentes.

Entre os métodos qualitativos de previsão, o mais geralmente usado é o da previsão conjetural. Esta técnica se baseia na utilização de um indivíduo ou um grupo pequeno de pessoas que fazem avaliações de circunstâncias futuras prováveis e analisam os efeitos destas sobre a receita a ser prevista. Embora seja *ad hoc*, a técnica pode produzir estimativas muito boas, especialmente quando pessoas experientes estão envolvidas no processo. O responsável pela previsão utilizará essa experiência em conjunto com a consideração de algumas tendências históricas, das circunstâncias econômicas atuais, e de outros fatores relevantes para a fonte de receita considerada.

As aproximações conjeturais tendem a trabalhar melhor, quando o ambiente econômico está mudando rapidamente. Quando as circunstâncias econômicas, políticas ou administrativas estão em fluxo intenso, métodos quantitativos não podem capturar informações importantes sobre os fatores que provavelmente estão alterando os padrões históricos.

Uma variação da aproximação conjetural é a chamada previsão de consenso. Aqui, os peritos, familiarizados com os fatores que afetam um tipo particular de receita, encontram-se para discutir algumas circunstâncias gerais, a fim de definir o que provavelmente acontecerá à arrecadação de determinada receita. Por exemplo, os administradores públicos municipais podem encontrar-se com as pessoas familiarizadas com o mercado local de imóveis para coletar informações sobre os movimentos naquele mercado, de forma a subsidiar a previsão dos impactos sobre o Imposto sobre Propriedade Territorial e Urbana (IPTU); os economistas públicos podem monitorar as condições nacionais, estaduais ou municipais para ajustar a previsão de determinada receita às mudanças econômicas ocorridas; e representantes de instituições financeiras locais podem trazer uma previsão de consenso sobre o movimento das aplicações futuras no mercado financeiro, para auxiliar na definição de estratégias de arrecadação do Imposto sobre Operações Financeiras (IOF). A previsão de consenso tende a trabalhar melhor quando há pouca informação histórica para subsidiar a elaboração de um método quantitativo de previsão.

As aproximações conjeturais de previsão têm certamente seu lugar entre os métodos de previsão. Com alguma extensão, uma perspectiva conjetural pode complementar outras técnicas de previsão, mesmo aquelas baseadas em métodos quantitativos extremamente rigorosos. Como se pode suspeitar, entretanto, as aproximações conjeturais podem estar

sujeitas à polarização e a outras fontes de erro. Guajardo e Miranda (2000) fornecem a seguinte lista das principais deficiências dos métodos qualitativos de previsão:

1ª - ancoragem de eventos – os métodos conjecturais permitem que os eventos recentes influenciem percepções sobre os eventos futuros. Por exemplo, uma cidade que hospeda uma convenção recente pode influenciar as percepções sobre a taxa de ocupação futura dos quartos.

2ª - disponibilidade da informação – esses métodos costumam sobrevalorizar o uso da informação prontamente disponível, em detrimento daquelas que dependem de maiores análises.

3ª - correlação falsa - os responsáveis pela previsão podem incorporar informações sobre fatores que supostamente influenciariam as receitas, mas que efetivamente não tem qualquer relação com elas.

4ª - inconsistência nos métodos e nas conjecturas - os responsáveis pela previsão podem vir a utilizar estratégias diferentes, ao longo do tempo, na elaboração de suas conjecturas, tornando-as menos confiáveis.

5ª - percepções seletivas – os métodos qualitativos podem fornecer informações importantes sobre relacionamentos causais que se opõem ao ponto de vista do responsável pela previsão, sendo, então, por ele desconsideradas.

6ª - mais desejo do que realidade – normalmente, tais métodos possibilitam a inclusão de pesos impróprios para o que os responsáveis pela previsão e as autoridades governamentais gostariam de ver acontecer.

7ª - pensamento em grupo – isso acontece quando a dinâmica de formar um consenso tende a conduzir os indivíduos a reforçar outros pontos de vistas, em detrimento de julgamentos independentes.

8ª - pressão política – ocorre quando os responsáveis pela previsão ajustam as estimativas para atender aos imperativos da restrição orçamentária ou de orçamentos equilibrados.

#### 2.1.2.2 - Métodos Quantitativos de Previsão

Métodos quantitativos se baseiam em dados numéricos relevantes para a fonte de receita analisada. Tais métodos procuram, também, explicitar as hipóteses e os procedimentos usados para gerar as previsões, além de atribuírem, geralmente, uma margem de erro às previsões, fornecendo uma indicação do grau de incerteza associado às estimativas.

Há dois tipos gerais de métodos quantitativos de previsão. O primeiro trata da abordagem baseada na análise de série temporais, que consiste em um grande número de técnicas que se utilizam de tendências passadas para projetar as receitas futuras. A segunda abordagem geral, apesar de incorporar ainda dados de séries temporais, constrói modelos causais que usam variáveis explicativas que supostamente influenciam o nível de um certo tipo de receita.

Em geral, métodos quantitativos fazem um trabalho melhor do que os métodos qualitativos na previsão das receitas futuras (Cirincione, 1999; Makridakis e Wheelwright, 1989). Métodos quantitativos mais simples também executam previsões tão boas quanto os métodos mais complexos (Makridakis, 1984). Finalmente, a abordagem de séries temporais normalmente apresenta melhor performance que a baseada na modelagem causal, ao menos no curto prazo, dado a incerteza associada à captura de todos os fatores econômicos relevantes que influenciam a geração da receita (Frank, 1993).

#### a. Abordagem de Séries Temporais

As abordagens através de séries temporais são o "feijão-com-arroz" da modelagem de previsão de receitas. Foram intensivamente usadas no setor. Hoje, existem softwares que aplicam automaticamente as técnicas apropriadas de previsão, dadas as características dos dados incorporados. A hipótese principal da técnica de séries temporais é a de que os padrões associados aos valores passados de uma série de dados podem ser usados para projetar os valores futuros.

Utilizando técnicas de séries temporais, Frank (1993) identificou diversos conceitos essenciais que necessitam ser levados em consideração antes da seleção da técnica apropriada. O primeiro é o que se refere a tendência dos dados, no sentido mais amplo da palavra. Aqui, a questão fundamental é sobre que tamanho uma série de dados deve possuir para que a técnica possa identificar todo o padrão inerente aos dados. Não há nenhum guia definitivo a respeito do número de observações requeridas para a construção de uma série de dados adequada. Geralmente, os dados devem cobrir um período de pelo menos alguns anos e, dependendo da técnica usada, devem incluir tanto quanto 50 ou mais observações.

Outro conceito essencial é aquele referente ao padrão cíclico de uma série temporal, que se refere à extensão em que a fonte de receitas é influenciada por ciclos de negócio gerais. Outra vez, como os governos têm-se movido no sentido de se afastar do imposto sobre a propriedade, relativamente estável e previsível, para taxas sobre consumo e impostos sobre as

vendas, mais instáveis, a necessidade de se levar em consideração os efeitos dos ciclos de negócio torna-se relativamente mais importante.

Da mesma forma, a sazonalidade é outro fenômeno cíclico que precisa ser levado em consideração. Sua ocorrência é bastante típica, quando as observações são feitas em períodos mensais ou trimestrais. Conforme se verá, os modelos matemáticos empregados podem ser ajustados para incorporar a sazonalidade.

#### b. Descrição de Modelos de Previsão de Séries Temporais

Há um grande número de abordagens de séries temporais que são usadas na previsão. Cirincione (1999) discute um grande número de questões sobre o seu uso e fornece um sumário interessante sobre uma variedade de técnicas em um apêndice de seu artigo. Entre as abordagens apresentadas na descrição técnica lá apresentada, podemos citar as seguintes:

##### 1. O modelo ingênuo de previsão

O modelo ingênuo de previsão supõe simplesmente que a receita disponível no tempo “t” é a mesma que estava disponível no tempo “t - 1”. Tal modelo é conhecido, também, como a abordagem do caminho aleatório (“random walk”) e é matematicamente formulado como:

$$F_t = A_{t-1} \quad (2.1)$$

onde  $F_t$  é a previsão no tempo t, e  $A_{t-1}$  é o valor real no tempo t - 1.

Uma variação desta abordagem envolve calcular a média de dois períodos prévios para gerar a estimativa do próximo período. A previsão ingênua é normalmente usada quando a série de dados possui um comportamento altamente imprevisível.

##### 2. modelos de médias móveis

Os modelos de médias móveis são provavelmente a abordagem de séries temporais mais geralmente usada na área de previsão de receitas governamentais. Como o próprio nome diz, o valor futuro a ser previsto é baseado na média de n períodos precedentes. É uma média móvel, porque os pontos de dados mais antigos vão sendo deixados de fora à medida que novos dados são adicionados. A expressão matemática para uma média móvel é dada por:

$$F_t = \frac{\sum_{i=1}^n A_{t-i}}{n} \quad (2.2)$$

onde  $F_t$  é a previsão no tempo  $t$ ,  $A_{t-i}$  é o valor real no tempo  $t - i$  e  $n$  é o número de períodos de tempo considerados no cálculo da média.

O tamanho do período de tempo que se deseja incluir na média depende do grau de variação apresentado pela série sob estudo. Assim, se aparece um grau elevado de aleatoriedade nos dados, um período mais longo será utilizado. Similarmente, se ciclos e sazonalidade estão presentes nos dados, períodos de tempo ainda mais longos serão requeridos.

Uma grande quantidade de tentativas e erros serão necessárias até que se encontre o modelo mais apropriado, embora softwares recentes possam muito rapidamente identificar o período de tempo ideal, produzindo um erro de previsão mínimo. Enquanto técnicas mais complexas de séries temporais podem apresentar resultados melhores do que a de médias móveis, esta, faz um trabalho razoavelmente bom e é usada freqüentemente como um “benchmark” na comparação com outros métodos.

### 3. Modelos de Alisamento Exponencial

O modelo de alisamento exponencial simples é uma das técnicas de previsão mais usadas no setor privado. O modelo nada mais é do que uma média móvel das previsões, corrigida pelos erros observados em previsões precedentes. Neste primeiro modelo de alisamento analisado, supõe-se que não há tendência ou qualquer padrão sazonal. Matematicamente, tem-se:

$$F_t = F_{t-1} + a.(A_{t-i} - F_{t-1}) \quad (2.3)$$

onde  $F_t$  é a previsão no tempo  $t$ ,  $A_{t-i}$  é o valor real no tempo  $t - i$  e o parâmetro  $a$  é o coeficiente de alisamento, que possui um valor estimado que varia entre zero e um. Tal modelo é denominado de alisamento exponencial porque o valor de  $a$  tende a afetar exponencialmente os valores passados. Quando  $a$  se aproxima de um, a previsão assemelha-se a uma média móvel de curto prazo, e quando se torna mais próximo de zero tende a se assemelhar a médias móveis de longo prazo. Não obstante os valores assumidos para  $a$ , modelos de alisamento exponencial tendem a dar a valores mais recentes pesos implícitos mais elevados. Aqui, outra vez,  $a$  é estimado usando método de tentativa e erro, de forma a assegurar um modelo mais apropriado. Entretanto, softwares recentes podem rapidamente encontrar o modelo que minimiza o erro de previsão.

#### 4. O modelo de Holt

O único parâmetro apresentado no modelo de alisamento acima,  $\alpha$ , pode ser adaptado para levar em conta as tendências que podem estar presentes nos dados. A forma apresentada aqui é chamada de modelo de Holt. Além do parâmetro de alisamento estimado no modelo de alisamento exponencial, um parâmetro que representa a tendência é acrescentado.

Seguindo a exposição encontrada em Cirincione (1999), a previsão no tempo  $t$  para  $k$  períodos futuros é igual ao nível da série em  $t$  mais o produto de  $k$  pela tendência no tempo  $t$ . O nível da série é estimado como uma função do valor real da série no tempo  $t$ , do nível da série no tempo precedente e da tendência estimada no tempo precedente. Neste caso, verifica-se que o parâmetro é, também, um coeficiente de alisamento. A tendência no tempo  $t$  é estimada como uma função do valor “alisado” da mudança no nível entre os dois períodos de tempo e a tendência estimada para o período de tempo precedente. Os valores para os parâmetros de alisamento,  $a$  e  $\beta$ , estão entre zero e um. Matematicamente, tem-se:

$$\begin{aligned} F_{t+k} &= S_t + kT \\ S_t &= aA_t + (1 - a)(S_{t-1} + T_{t-1}) \\ T_t &= b(S_t - S_{t-1}) + (1 - b)T_{t-1} \end{aligned} \quad (2.4)$$

onde  $F_{t+k}$  é a previsão para  $k$  períodos de tempo no futuro,  $A_t$  é o valor real no tempo  $t$ ,  $S_t$  é o nível da série no tempo  $t$ ,  $T_t$  é a tendência no tempo  $t$ , e  $a$  e  $b$  são os parâmetros de alisamento.

#### 5. Alisamento Exponencial com Deterioração na Tendência

Embora o modelo de Holt leve em consideração a tendência, que pode ser inerente à série dos dados, não é muito realístico supor que tal tendência permaneça presente em toda a série, indefinidamente. Isto significa que tal procedimento pode fazer com que o modelo exceda às estimativas quando o horizonte de previsão for muito longo. Uma variação conhecida como Alisamento Exponencial com Deterioração na Tendência tem o efeito de variar a tendência, à medida que o tempo flui normalmente. Tal modelo inclui um terceiro parâmetro,  $f$ , com um valor entre zero e um, que especifica uma taxa de deterioração na tendência. Matematicamente, o modelo é assim representado:

$$\begin{aligned}
F_{t+k} &= S_t + \sum_{i=1}^k f^i T_t \\
S_t &= aA_t + (1-a)(S_{t-1} + fT_{t-1}) \\
T_t &= b(S_t - S_{t-1}) + (1-b)fT_{t-1}
\end{aligned} \tag{2.5}$$

onde  $F_{t+k}$  é a previsão para  $k$  períodos de tempo no futuro,  $A_t$  é o valor real no tempo  $t$ ,  $S_t$  é o nível da série no tempo  $t$ ,  $T_t$  é a tendência no tempo  $t$ , e  $a$ ,  $b$  e  $f$  são os parâmetros de alisamento.

#### 6. Alisamento linear sazonal de Holt-Winter

Este modelo adapta o método de Holt para incluir um componente sazonal, em adição ao coeficiente de alisamento e ao parâmetro de tendência. A primeira variante do modelo é aditiva. Supõe-se, inicialmente, que a sazonalidade é constante ao longo da série que está sendo prevista. A representação matemática do modelo é, portanto, a seguinte:

$$\begin{aligned}
F_{t+k} &= S_t + kT + I_{t-p+k} \\
S_t &= S_{t-1} + T_{t-1} + a(A_t - S_{t-1} - T_{t-1} - I_{t-s}) \\
T_t &= T_{t-1} + ab(A_{t-1} - S_{t-1} - T_{t-1} - I_{t-s}) \\
I_t &= I_{t-s} + d(1-a)(A_t - S_{t-1} - T_{t-1} - I_{t-s})
\end{aligned} \tag{2.6}$$

onde  $F_{t+k}$  é a previsão para  $k$  períodos de tempo no futuro,  $A_t$  é o valor real no tempo  $t$ ,  $S_t$  é o nível da série no tempo  $t$ ,  $T_t$  é a tendência no tempo  $t$ ,  $I_t$  é o índice sazonal no tempo  $t$ ,  $s$  é o contador do índice sazonal e  $a$ ,  $b$  e  $d$  são os parâmetros de alisamento.

A variante multiplicativa deste modelo supõe que a sazonalidade está mudando ao longo da série. Assim, matematicamente, tem-se:

$$\begin{aligned}
F_{t+k} &= (S_t + kT)I_{t-p+k} \\
S_t &= S_{t-1} + T_{t-1} + a[A_t - (S_{t-1} + T_{t-1})I_{t-s}] / I_{t-s} \\
T_t &= T_{t-1} + ab[A_{t-1} - (S_{t-1} + T_{t-1})I_{t-s}] / I_{t-s} \\
I_t &= I_{t-s} + d(1-a)[A_t - (S_{t-1} + T_{t-1})I_{t-s}] / S_t
\end{aligned} \tag{2.7}$$

Ao incorporar a sazonalidade, aumenta-se, naturalmente, a necessidade de inclusão de mais dados. Geralmente, três a quatro anos de dados mensais são suficientes. O modelo é

também bastante complexo, necessitando estimar os parâmetros de alisamento, de tendência e sazonal, simultaneamente.

### *7. Os modelos ARIMA de Box-Jenkins*

ARIMA é um acrônimo para o modelo auto-regressivo integrado de média móvel. Auto-regressivo e média móvel se referem a dois dos componentes do modelo, enquanto integrado se refere ao processo de representar os cálculos em uma métrica que possa ser corretamente interpretada<sup>4</sup> (Frank, 1993).

A modelagem ARIMA se compõe, basicamente, de três estágios (Pankratz, 1983). No estágio de identificação do modelo, o responsável pela previsão deve decidir se a série de tempo é auto-regressiva, de médias móveis, ou mista. Isto é feito, geralmente, de forma visual, inspecionando diagramas específicos obtidos dos dados ou empregando várias técnicas estatísticas. No segundo e terceiro estágios, estimação do modelo e checagem de diagnóstico, o responsável pela previsão verifica se a identificação do modelo original está correta. Isto requer que se sujeite o modelo a uma variedade de técnicas de diagnóstico. Se o modelo mostrar-se adequado, o responsável pela previsão prossegue então para a previsão.

Uma das principais vantagens de se utilizar a abordagem ARIMA é que o método pode incluir intervalos de confiança em torno das previsões. Isto serve como uma outra verificação da validade do modelo.

A fim de se obter melhores resultados na utilização da metodologia de Box-Jenkins (ARIMA), três hipóteses básicas precisam ser observadas.

A primeira é relativa ao tamanho inicial da amostra geralmente aceita, que deve ser de, no mínimo, 50 observações (Box & Jenkins, 1976). Este tende a ser um obstáculo significativo para muitos dados que somente são coletados anualmente.

A segunda suposição é a de que a série de dados seja estacionária, isto é, que a série varie em torno de uma média constante e com uma variância constante<sup>5</sup>. Rodar uma regressão com duas variáveis não-estacionárias pode causar resultados espúrios (Granger & Newbold, 1974). Se os dados forem não-estacionários, a série de dados deve ser diferenciada ou deve-se proceder à adição de uma tendência no tempo. Se os dados forem de tendência não-estacionária somente, então, adicionar ao modelo uma tendência linear do tempo tornará a

---

<sup>4</sup> Trata-se de uma colocação mais ampla da necessidade de se trabalhar com séries estacionárias. Estatisticamente, um processo integrado de ordem  $n$  precisa ser diferenciado  $n$  vezes para se tornar estacionário.

<sup>5</sup> A definição de estacionariedade encontra-se formalizada na seção 2.2 deste capítulo.

série estacionária. Os dados com tendência não-estacionária têm uma média e uma variância que mudam com o tempo, mas por uma quantidade constante. Se os dados forem não-estacionários em primeira diferença, uma única diferenciação tornará a série estacionária. Diferenciar uma única vez envolve subtrair a observação no tempo  $t$  da observação no tempo  $t - 1$  para todas as observações. Se os dados requerem este tipo de tratamento, isso deve ficar evidente no estágio da identificação.

A terceira hipótese para os modelos ARIMA é a de que a série seja homoscedástica, isto é, tenha uma variância constante ao longo do tempo. Se a amplitude da variação em torno da média aumentar com o tempo, mesmo após a diferenciação, a série é considerada heteroscedástica. A solução para este problema pode ser simples ou complexa e envolve medidas de ajustes tais como: usar o logaritmo natural dos dados, usar raízes quadradas ou cúbicas, truncar a série de dados, entre outros.

O primeiro componente do processo ARIMA é o auto-regressivo. Tal componente prevê os valores futuros com base em uma combinação linear dos valores passados. Sua definição teórica é apresentada na próxima seção deste capítulo.

O segundo componente, de médias móveis, fornece as estimativas das previsões com base nos erros presente e passados de previsão. O tratamento estatístico deste componente será apresentado, detalhadamente, na próxima seção.

Estes dois componentes formam, juntos, os modelos auto-regressivo de médias móveis (ARMA). Se uma série possuir tendência ou diferença estacionária, o modelo se transforma no modelo ARIMA de Box-Jenkins (Box & Jenkins, 1976). Uma descrição detalhada da metodologia de Box-Jenkins é apresentada na seção 2.2 deste capítulo.

### c. Os modelos causais

Modelos causais de previsão tendem a ser, geralmente, mais complexos, exigindo uma grande quantidade de dados e requerendo um grau elevado de habilidade estatística. Estas abordagens tendem a funcionar melhor para as receitas que são influenciadas pesadamente por fatores econômicos. Assim, os dados externos que representam indicadores relevantes de desempenho econômico são usados para prever o nível de receita esperado. Algumas informações econômicas normalmente incorporadas nestes modelos inclui a população, a renda e a informação de preço (Wong, 1995).

A complexidade de modelos causais é bastante variada. O tipo mais simples seria um modelo de regressão linear que pudesse tentar projetar a receita em função do tempo, por exemplo. Por outro lado, modelos de regressão múltipla incorporariam todo o número de

variáveis explicativas relevantes, incluindo variáveis importantes de política tributária, assim como variáveis dummies binárias. Para ilustrar, seguindo Cirincione (1999), quatro modelos comuns de regressão que empregam os mínimos quadrados ordinários podem ser mostrados:

$$\begin{aligned}
 F_t &= a + b_1T_t \\
 F_t &= a + b_1T_t + b_2T_t^2 \\
 F_t &= a + b_1T_t + g_1D_1 + g_2D_2 + \dots + g_sD_s \\
 F_t &= a + b_1T_t + b_2T_t^2 + g_1D_1 + g_2D_2 + \dots + g_sD_s
 \end{aligned}
 \tag{2.8}$$

onde  $F_t$  é o valor previsto no tempo  $t$ ,  $T_t$  é valor do tempo no tempo  $t$ ,  $T_t^2$  é o quadrado do valor do tempo no tempo  $t$ ,  $b_1$  é o parâmetro de tendência linear associado ao tempo,  $b_2$  é o parâmetro de tendência quadrática associado ao tempo,  $D$  é uma variável dummy binária do modelo para cada período de tempo  $s$ , e  $g_s$  é o parâmetro associado a cada período  $s$ . Os valores estimados das variáveis dummies do modelo revelam o nível médio da variável dependente durante o período de tempo considerado. Testar a igualdade dos coeficientes dessas variáveis entre os modelos pode revelar se há alguma diferença significativa no nível médio da variável dependente entre os períodos considerados.

As previsões econométricas são estruturalmente similares às equações de regressão, mas podem incluir estimativas através de sistemas de múltiplas equações. Assim, os eventos e os relacionamentos complexos podem ser modelados de forma que a saída de uma equação alimenta uma outra equação, devendo ser resolvidas simultaneamente. Os tipos de receita para as quais as previsões econométricas são mais úteis incluem aquelas cujo comportamento econômico é diretamente relacionado a variáveis perfeitamente identificáveis.

## 2.2 – MODELOS DE SÉRIES TEMPORAIS

Uma série temporal caracteriza-se como um conjunto de observações que representa uma variável observada ao longo do tempo. Quando as observações são obtidas continuamente, isto é, a todo instante ao longo do tempo, diz-se que a série temporal é contínua, podendo-se representá-la como  $X(t)$ . De outra forma, uma série temporal discreta é aquela em que as observações são tomadas em um conjunto discreto, ou seja, em intervalos fixos de tempo, cuja representação é dada por  $X_t$ . (Melo, 2001).

Uma série temporal observada  $\{X_t\}$  pode ser considerada como uma realização de algum processo estocástico ou randômico (Enders, 1995). A análise de tal processo objetiva:

- descrever o comportamento da série;
- investigar o mecanismo gerador da série temporal;
- fazer previsões de valores futuros da série.

Para tanto, é necessário descobrir um modelo probabilístico hipotético que represente os dados (Brockwell & Davis, 1996) e que possua propriedades similares àquelas do mecanismo gerador do processo estocástico. Assim, define-se que cada observação  $x_t$  é uma realização de uma variável aleatória  $X_t$ . Desta forma, uma série temporal  $\{x_t\}$  é a realização de uma família de variáveis aleatórias  $\{X_t\}$ . De outra maneira, um modelo de série temporal para dados observados  $\{x_t\}$  é uma especificação das distribuições de uma seqüência de variáveis aleatórias  $\{X_t\}$  da qual  $\{x_t\}$  é denominada uma realização (Brockwell & Davis, 1996).

### 2.2.1 - Definições

#### Definição 2.2.1.1

Seja  $T$  um conjunto arbitrário. Um processo estocástico é uma família  $X = \{X(t, w), t \in T, w \in \Omega\}$ , tal que para cada  $t \in T$ ,  $X(t, w)$  é uma variável aleatória definida num espaço de probabilidade  $(\Omega, A, P)$ . Se  $T = Z = \{0, \pm 1, \pm 2, \dots\}$ , tem-se um processo de tempo discreto; se  $T = R = (-\infty, +\infty)$ , tem-se um processo de tempo contínuo.

#### Definição 2.2.1.2

Seja  $X = \{X(t, w), t \in T, w \in \Omega\}$  um processo estocástico, definido num espaço de probabilidade  $(\Omega, A, P)$ . Para cada  $w$  fixo  $X(t, w)$  é uma série temporal, ou seja, é uma particular realização de um processo estocástico.

Daqui por diante denotaremos um processo estocástico mais resumidamente por  $X = \{X_t, t \in T\}$ .

Para cada  $t \in T$ ,  $X_t$  é uma variável aleatória e, portanto, possui um conjunto de valores possíveis. Para o caso de um processo  $X_t$  assumindo valores discretos, suas propriedades poderão ser obtidas a partir de uma função de probabilidade discreta,  $P_t(x_i)$ . Entretanto, na situação mais comum,  $X_t$  pode ser considerada uma variável aleatória contínua com densidade de probabilidade  $f_t(x)$ .

Neste caso, teremos:

$$E[X_t] = \int_{-\infty}^{+\infty} x \cdot f_t(x) dx = m_t(x) \quad (2.9)$$

e

$$Var[X_t] = E[\{X_t - m_t\}^2] = \int_{-\infty}^{+\infty} (x - m_t)^2 \cdot f_t(x) dx = s_t^2. \quad (2.10)$$

**Definição 2.2.1.3**

Um processo estocástico  $X = \{X_t, t \in T\}$  é chamado estritamente estacionário (ou fortemente estacionário) se a distribuição conjunta de  $(X_{t_1}, X_{t_2}, \dots, X_{t_k})$  é idêntica à distribuição conjunta de  $(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_k+h})$ , para quaisquer  $t_1, \dots, t_k, h$ , de  $T$ .

Segue-se, então, que os momentos de ordem superior, tais como,

$$E[X_t^l \cdot X_{t-p}^m \cdot X_{t-p}^0] \quad (2.11)$$

são funções somente de  $p$  e  $q$  e não de  $t$ .

Um processo  $X = \{X_t, t \in T\}$  é dito "estacionário até ordem  $m$ " se os momentos de  $X_{t_1+k}$  e  $X_{t_1}$  forem idênticos até ordem  $m$ . Da mesma maneira, as distribuições conjuntas de  $\{X_{t_1}, X_{t_2}\}$  e  $\{X_{t_1+k}, X_{t_2+k}\}$  não devem ser necessariamente idênticas, mas devem possuir momentos iguais até ordem  $m$ .

Estas idéias levam à próxima definição, para um processo estacionário de segunda ordem.

**Definição 2.2.1.4**

Um processo estocástico  $X = \{X_t, t \in T\}$  diz-se estacionário de 2ª ordem (ou fracamente estacionário) se:

- i)  $E[X_t] = \mu_t = \text{constante}, \forall t \in T$ ;
- ii)  $E[X_t^2] < \infty$ , também constante,  $\forall t \in T$ ;
- iii)  $\text{Cov}[X_t, X_s] = C(t, s) = C(t - s)$  é função apenas de  $t - s$ .

A partir de agora, denominaremos este tipo de processo simplesmente de processo estacionário, uma vez que, na maioria das situações, ele se aproxima bastante dos processos estacionários de ordem superior.

Considerando  $\{X_t\}$  um processo estacionário até segunda ordem, existem duas funções de importância fundamental, a função de autocovariância e a função de autocorrelação, as quais são definidas a seguir.

**Definição 2.2.1.5**

Definimos a função de autocovariância de um processo  $\{X_t\}$  estacionário da média  $\mu$  através de expressão:

$$\text{Cov}[X_t, X_{t+\tau}] = E[\{X_t - m\}\{X_{t+\tau} - m\}]. \quad (2.12)$$

Para um processo estacionário, temos que a função de autocovariância não depende do instante  $t$ , mas apenas do valor da defasagem  $\tau$ . Assim podemos escrever  $\text{Cov}[X_t, X_{t+\tau}] = R(\tau)$ , onde para cada valor de  $\tau$ ,  $R(\tau)$  mede o grau de relação linear entre pares de valores do processo, separados por um intervalo de comprimento  $\tau$ , denominado comumente de defasagem ("lag").

Se agora escrevermos, para  $\tau$ ,  $\rho(\tau) = R(\tau)/R(0)$ , então  $\rho(\tau)$  é chamada função de autocorrelação de  $\{X_t\}$ .

Desde que  $R(0)$  pode ser escrita como:

$$R(0) = E[\{X_t - m\}^2] = \text{Var}[X_t] = s^2, \forall t \quad (2.13)$$

$\rho(\tau)$  pode ser expresso por:

$$r(\tau) = \frac{\text{Cov}[X_t, X_{t-\tau}]}{\text{Var}[X_t]} \quad (2.14)$$

Então, para cada valor de  $\tau$ ,  $\rho(\tau)$  representa o coeficiente de correlação entre pares de  $\{X_t\}$ , separados por um intervalo de comprimento  $\tau$ .

As funções de autocovariância e de autocorrelação fornecem uma medida útil do grau de dependência entre os valores de uma série temporal em diferentes períodos. As autocorrelações medem ainda o tamanho e a força da "memória" do processo. Por essa razão, essas funções são estimadas a partir dos valores realizados da série temporal estacionária  $\{X_t\}$ . (Melo, 2001).

2.2.1.1 Propriedades gerais de  $R(t)$  e  $\rho(\tau)$ 

Todas as funções de autocovariância possuem as seguintes propriedades:

- (1)  $R(0) = \sigma^2$ ,
- (2)  $|R(\tau)| \leq R(0), \forall \tau$ .

- (3) Quando o processo  $\{X_t\}$  assume valores reais, temos:  $R(-\tau) = R(\tau)$ ,  $\forall \tau$ , i.e.,  $R(\tau)$  é uma função par.

Da mesma maneira, todas as funções de autocorrelação possuem as seguintes propriedades:

- (1)  $\rho(0) = 1$ ,  
 (2)  $|\rho(\tau)| \leq 1$ ,  $\forall \tau$ ,  
 (3)  $\rho(-\tau) = \rho(\tau)$ ,  $\forall \tau$ , quando  $X_t$  assume valores reais.

**Definição 2.2.1.6 - RUÍDO BRANCO DE SEGUNDA ORDEM (RBSO)**

Uma série  $\{\eta_t; t = 0, \pm 1, \pm 2, \dots\}$ , tal que

$$E[h_t] = 0, E[h_t^2] = s_h^2, E[h_t h_s] = 0, \forall t \neq s \quad (2.15)$$

é denominada ruído branco.

A função de autocovariância é, portanto,

$$R(s) = Cov[h_t, h_{t+s}] = \begin{cases} 0, & s \neq 0 \\ s^2, & s = 0 \end{cases} \quad (2.16)$$

E a função de autocorrelação é dada por:

$$r(s) = \begin{cases} 0, & s \neq 0 \\ 1, & s = 0 \end{cases} \quad (2.17)$$

Assim, não existe relacionamento linear entre  $\eta_t$  e os valores prévios da série.

**Definição 2.2.1.7 RUÍDO BRANCO ESTRITO.**

Uma seqüência  $\{e_t, t = 0, \pm 1, \pm 2, \dots\}$  de variáveis aleatórias independentes e identicamente distribuídas (não necessariamente, mas, usualmente com distribuição normal de média zero e variância constante  $\sigma^2$ ) é chamada ruído branco estrito.

Claramente, um ruído branco estrito é um RBSO, mas não necessariamente o contrário. A partir de agora, denominaremos este tipo de processo simplesmente de ruído branco e usaremos a notação  $e_t$ .

### 2.2.2 – Processo Linear Geral

A forma mais simples de uma série temporal é aquela gerada pelo processo de ruído branco estrito. Porém, uma série temporal geral,  $\{X_t\}$ , tem uma estrutura mais complicada. A função média de  $\{X_t\}$ , particularmente  $\mu_t = E[X_t]$ , experimenta mudanças com o tempo, na presença de tendências e componentes sazonais. Essencialmente, a variação em torno da função média descrita pela série  $\{X_t^* = X_t - \mu\}$  será autocorrelacionada, isto é, em geral, haverá correlação entre os valores de  $X_t^*$  nos diferentes pontos do tempo.

Quando estudamos uma série temporal geral, tentamos construir um modelo que descreva a característica da função média e a dependência ou correlação em diferentes pontos do tempo. Supondo, por simplicidade, que foram removidas todas as tendências e componentes sazonais, podemos inferir, sem perda de generalidade, que a função média de  $\{X_t\}$  é igual a zero,  $\forall t \in T$ . O próximo passo é explicar as propriedades de autocorrelação da série a fim de poder encontrar algum tipo de relacionamento entre os valores passados, presentes e futuros que reduza a série a um ruído branco. Assim, um modelo para  $\{X_t, t = 0, \pm 1, \pm 2, \dots\}$ , de um modo geral, é descrito pela equação da forma

$$h(\dots, X_{t-2}, X_{t-2}, X_{t-2}, X_{t-2}, X_{t-2}, \dots) = e_t \quad (2.18)$$

onde  $e_t$ , é um processo ruído branco escrito com média zero e  $h(\cdot)$  é uma função conhecida.

A classe dos modelos lineares gerais corresponde ao caso em que  $h(\cdot)$  é uma função linear de,  $\dots X_{t-1}, X_t, X_{t+1}, \dots$ . Em tal situação, a equação (2.18) pode ser representada como:

$$\sum_{j=-\infty}^{\infty} a_j X_{t-j} = e_t \quad (2.19)$$

onde  $a_j$  é uma seqüência de constantes. Na prática, geralmente, supomos que  $X_t$  depende somente de seus valores passados, isto é,  $a_j = 0, \forall j < 0$ , desta maneira (2.19) pode ser representada na forma unilateral

$$\sum_{j=0}^{\infty} a_j X_{t-j} = e_t \quad (2.20)$$

Introduzindo o operador translação para o passado (também denominado operador de defasagem, denotado por  $BX_t = X_{t-1}, \dots, B^j X_t = X_{t-j}$ , etc., (2.20) pode ser, agora, expressa na forma.

$$a(B)X_t = e_t, \quad (2.21)$$

onde

$$a(z) = \sum_{j=0}^{\infty} a_j z^j. \quad (2.22)$$

A equação (2.21) pode ser resolvida, em alguns casos, para expressar  $X_t$  como uma função linear de valores passados e presentes de  $e_t$ , isto é, o processo linear (2.21) é invertível se  $a(B)$  convergir, para  $|B| \leq 1$ . Desta forma, invertendo (2.21), formalmente, podemos escrever

$$X_t = a^{-1}(B).e_t. \quad (2.23)$$

Se  $a^{-1}(Z)$  é expresso como uma série convergente para  $|Z| < 1$  (isto é, se  $a(Z)$  não possui zeros dentro do círculo unitário) podemos escrever que :

$$a^{-1}(Z) = b_0 + b_1 Z + b_2 Z^2 + \dots = \sum_{j=0}^{\infty} b_j Z^j. \quad (2.24)$$

Agora, escrevemos  $X_t$  explicitamente como

$$X_t = \sum_{j=0}^{\infty} b_j e_{t-j} \quad (2.25)$$

ou

$$X_t = b(Z).e_t, \quad (2.26)$$

onde

$$b(Z) = \sum_{j=0}^{\infty} b_j Z^j \quad (2.27)$$

teremos, certamente,  $b(Z) = a^{-1}(Z)$  ou  $b(Z).a(Z) = 1$ .

A equação (2.25) proporciona uma formulação alternativa para o modelo linear geral, na qual  $X_t$  é expressa como uma combinação linear (em geral infinita) dos valores passados e presentes de um processo ruído branco.

Algumas vezes é vantajoso interpretar (2.26) como a série temporal gerada através de um filtro linear (ou sistema linear), cuja entrada é a série ruído branco. Com esta interpretação, a função  $b(Z)$  é chamada de *função de transferência* do modelo.

Observa-se que o modelo linear (2.25) (ou (2.20)) é completamente determinado pela forma de  $b(Z)$ .

### 2.2.3 – Casos Particulares dos Modelos Lineares

#### 2.2.3.1. Modelo Auto-Regressivo de Ordem $p$ : AR( $p$ )

É o caso especial do modelo linear geral (2.20), onde  $a(Z)$  é um polinômio de grau finito, isto é

$$\Phi(Z) = 1 + \phi_1 Z + \dots + \phi_p Z^p. \quad (2.28)$$

Então, (2.20) se reduz à forma do modelo AR( $p$ ):

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + e_t, \quad (2.29)$$

onde  $e_t$  é ruído branco de 2ª ordem com média zero.

Este processo será estacionário se as raízes de  $\Phi(z) = 1 + \phi_1 z + \dots + \phi_p z^p = 0$  estiverem fora do círculo unitário  $|z| \leq 1$ . O processo é sempre invertível.

O modelo auto-regressivo de ordem 1, AR(1) é definido como

$$X_t = f_1 X_{t-1} + e_t, \quad (2.30)$$

onde

$$\varepsilon_t \sim \text{WN}(0, \sigma^2),$$

$f_1$  : parâmetro auto-regressivo,

e a notação é  $X_t \sim \text{AR}(1)$ .

Assumindo que o processo apresenta infinitos valores passados e que  $|f_1| < 1$ , então  $m_t = 0$ , para todo  $t$  (Granger & Newbold, 1986). Tal condição garante que o processo descrito por (2.30) é uma solução estacionária para  $\{X_t\}$ . Assim a raiz da solução da equação (2.30), que é  $f_1$ , deve estar fora do círculo unitário (no plano complexo) para que a condição de estacionariedade seja satisfeita, i.e.,  $|f_1| < 1$  (Granger & Newbold, 1986).

O modelo descrito em (2.30) pode ser generalizado para representar processos estacionários com médias diferentes de zero como

$$X_t = c + f_1 X_{t-1} + e_t, \quad (2.31)$$

onde  $c$  é um parâmetro incluído para que  $X_t$  tenha média diferente de zero e a notação desse processo é  $X_t \sim \text{AR}(1)$ . Assim, o processo descrito em (2.31) pode ser considerado como tendo média zero se for subtraída de cada observação uma média fixa.

A média de um processo AR(1) é obtida tomando os valores esperados dos dois lados da equação (2.31), assim

$$\begin{aligned} E(X_t) &= E(c + f_1 X_{t-1} + e_t) \Rightarrow E(X_t) = E(c) + f_1 E(X_{t-1}) \\ &\quad \therefore \\ m &= c + f_1 m \Rightarrow \\ m &= \frac{c}{1 - f_1} \end{aligned} \quad (2.32)$$

A variância é calculada utilizando o resultado (2.32) em (2.31); assim,

$$X_t = m(1 - f_1) + f_1 X_{t-1} + e_t \quad (2.33)$$

ou

$$X_t - m = f_1 (X_{t-1} - m) + e_t. \quad (2.34)$$

Elevando ambos os lados da equação acima ao quadrado e tomando os valores esperados,

$$E(X_t - m)^2 = f_1^2 E(X_{t-1} - m)^2 + 2f_1 E[(X_{t-1} - m) - e_t] + E(e_t^2). \quad (2.35)$$

O termo do meio é igual a zero (Hamilton, 1994) e assumindo estacionariedade tem-se que

$$E(X_t - m)^2 = E(X_{t-1} - m)^2 = g_0. \quad (2.36)$$

Assim, a equação (2.35) pode ser expressa como

$$g_0 = f_1^2 g_0 + 0 + s^2. \quad (2.37)$$

Então, a variância é dada por

$$g_0 = \frac{\sigma^2}{(1-f_1^2)} \quad (2.38)$$

A função de autocovariância é calculada multiplicando os dois lados da equação (2.30) e tomando os valores esperados; assim,

$$\begin{aligned} g_x(h) &= \text{cov}(X_t, X_{t-h}) \\ &= \text{cov}(f_1 X_{t-1}, X_{t-h}) + \text{cov}(e_t, X_{t-h}) \\ &= f_1 g_x(h-1) + 0 = \mathbf{L} = f_1^h g_x(0). \end{aligned} \quad (2.39)$$

A função de autocorrelação (ACF) para um processo AR(1) é dada por

$$r_x(h) = \frac{g_x(h)}{g_x(0)} = f_1^{|h|}, \quad h = 0, \pm 1, \mathbf{L} \quad (2.40)$$

Assim, o comportamento da ACF segue um decaimento geométrico para zero (Hamilton, 1994). E como a função de autocorrelação assume o mesmo sinal que  $f_1$  o decaimento será direto para valores positivos de  $f_1$  e seguirá um caminho oscilatório ao redor de zero para valores negativos de  $f_1$ . Ambos os decaimentos serão lentos se o valor de  $f_1$  estiver próximo da fronteira de não-estacionariedade, *i.e.*, +1 e -1 (Mills, 1990).

Além disso, o comportamento da série  $\{X_t\}$  está relacionado também com os valores de  $f$ . Para valores positivos de  $f$ , a série será mais suave que um processo do tipo *ruído branco* ( $f=0$ ) (Granger & Newbold, 1986), com valores adjacentes positivamente correlacionados e tendências de “baixa frequência” (Mills, 1990). Por outro lado, para valores negativos de  $f$ , a série será menos suave que o processo tipo *ruído branco* (Granger & Newbold, 1986), com valores adjacentes negativamente correlacionados e com oscilações rápidas e repentinas (Mills, 1990)

O processo auto-regressivo pode ser generalizado para uma ordem  $p>1$ , AR( $p$ ), como sendo

$$X_t = c + f_1 X_{t-1} + \mathbf{L} + f_p X_{t-p} + e_t = \sum_{j=1}^p f_j X_{t-j} + e_t, \quad (2.41)$$

onde

$$\varepsilon_t \sim \text{WN}(0, \sigma^2),$$

$f_1, \dots, f_p$ : parâmetros auto-regressivos,

com a notação de  $X_t \sim \text{AR}(p)$ . Alternativamente,

$$X_t = c + f_1 B X_t + \mathbf{L} + f_p B^p X_t + e_t \quad (2.42)$$

ou

$$f(B)X_t = c + e_t,$$

onde

$$f(B) = (1 - f_1 B - \mathbf{L} - f_p B^p).$$

Assim, para que  $\{X_t\}$  seja estacionário, as raízes do polinômio

$$1 - f_1 z - f_2 z^2 - \mathbf{L} - f_p z^p = 0$$

devem estar fora do círculo unitário (no plano complexo).

Assumindo que a condição de estacionariedade está satisfeita, a média do processo  $\text{AR}(p)$  é obtida tomando os valores esperados em (2.41),

$$m = c + f_1 m + f_2 m + \dots + f_p m,$$

ou ainda,

$$m = \frac{c}{(1 - f_1 - f_2 - \mathbf{L} - f_p)}$$

As autocovariâncias são calculadas multiplicando-se ambos os lados de (2.41) por  $(X_{t-j} - m)$  e tomando os valores esperados,

$$g_h = \begin{cases} f_1 g_{h-1} + \mathbf{L} + f_p g_{h-p} & , h = 1, 2, \mathbf{L} \\ f_1 g_1 + \mathbf{L} + f_p g_p + s^2 & , h = 0. \end{cases} \quad (2.43)$$

A autocorrelação é obtida dividindo-se a equação (2.43) por  $\gamma_0$ ,

$$r_h = f_1 r_{h-1} + \mathbf{L} + f_p r_{h-p} \quad , h = 1, 2, \mathbf{L} \quad (2.44)$$

As  $p$  equações obtidas de (2.44) são denominadas de equações de *Yule-Walker* e podem ser escritas na forma matricial como

$$\boldsymbol{\rho} = \mathbf{P}\boldsymbol{\varphi}, \quad (2.45)$$

onde

$$\boldsymbol{\rho} = (r_1, r_2, \mathbf{L}, r_p)' , \boldsymbol{\varphi} = (f_1, f_2, \mathbf{L}, f_p)'$$

e

$$\mathbf{P} = \begin{bmatrix} 1 & r_1 & r_2 & \mathbf{K} & r_{p-1} \\ r_1 & 1 & r_1 & \mathbf{K} & r_{p-2} \\ \mathbf{M} & \mathbf{M} & \mathbf{M} & & \\ r_{p-1} & r_{p-2} & r_{p-3} & \mathbf{K} & 1 \end{bmatrix}.$$

Então,

$$\boldsymbol{\varphi} = \mathbf{P}^{-1}\boldsymbol{\rho}, \quad (2.46)$$

de forma que os parâmetros auto-regressivos podem ser expressos como uma função das  $p$  autocorrelações (Mills, 1990).

O comportamento da função de autocorrelação de um processo auto-regressivo é uma mistura de decaimento exponencial e/ou decaimento senoidal. Se as raízes da equação auto-regressiva forem reais, então as autocorrelações decairão exponencialmente. Caso as raízes sejam complexas, o decaimento será na forma senoidal (Granger & Newbold, 1986).

Para um processo  $AR(p)$  não há correlação entre  $X_t$  e  $X_{t-k}$  para  $k > p$  (Mills, 1990). Então, todos os valores de  $\phi_{kk}$  para  $k > p$  são zero e a PACF para um processo  $AR(p)$  puro apresenta um “corte” para zero para defasagens maiores que  $p$  (Enders, 1995).

Assim, pode-se resumir que um processo  $AR(p)$  é descrito por:

- possuir uma função de autocorrelação, ACF, que é uma combinação de decaimentos exponenciais e senoidais e tamanho infinito; e
- possuir uma função de autocorrelação, PACF, que é zero para defasagens maiores que  $p$ .

Em um processo  $AR(1)$ ,  $X_t$  e  $X_{t-2}$  são correlacionados, mesmo que  $X_{t-2}$  não apareça diretamente no modelo. O valor da correlação entre  $X_t$  e  $X_{t-2}$  (i.e.,  $r_2$ ) é igual à correlação entre

$X_t$  e  $X_{t-1}$  ( $r_1$ ) multiplicada pela correlação entre  $X_{t-1}$  e  $X_{t-2}$  ( $r_1$ ), de forma que  $r_2 = r_1^2$ . Assim, toda essa correlação “indireta” está presente na ACF de qualquer processo auto-regressivo (Enders, 1995).

Dessa maneira, define-se a Função de Autocorrelação Parcial (PACF) como a seqüência de correlações entre  $(X_t$  e  $X_{t-1})$ ,  $(X_t$  e  $X_{t-2})$ ,  $(X_t$  e  $X_{t-3})$  e assim por diante, desde que os efeitos de defasagens anteriores sobre  $X_t$  permaneçam constantes (Hill, Griffiths & Judge, 1999). A PACF é calculada como o valor do coeficiente  $f_{kk}$  na equação

$$X_t = f_{k1} X_{t-1} + f_{k2} X_{t-2} + \mathbf{L} + f_{kk} X_{t-k} + e_t \tag{2.47}$$

O coeficiente  $f_{kk}$  é obtido das equações de *Yule-Walker* aplicadas a (2.47). Tais equações são dadas por (2.44) e, substituindo  $p = k$  e  $\phi_i = \phi_{ii}$ , tem-se (Mills, 1990)

$$f_{kk} = \frac{\begin{vmatrix} 1 & r_1 & \mathbf{L} & r_{k-2} & r_1 \\ r_1 & 1 & \mathbf{L} & r_{k-3} & r_2 \\ \mathbf{M} & \mathbf{M} & \mathbf{L} & \mathbf{M} & \mathbf{M} \\ r_{k-1} & r_{k-2} & \mathbf{L} & r_1 & r_k \end{vmatrix}}{\begin{vmatrix} 1 & r_1 & \mathbf{L} & r_{k-2} & r_{k-1} \\ r_1 & 1 & \mathbf{L} & r_{k-3} & r_{k-2} \\ \mathbf{M} & \mathbf{M} & \mathbf{L} & \mathbf{M} & \mathbf{M} \\ r_{k-1} & r_{k-2} & \mathbf{L} & r_1 & 1 \end{vmatrix}} \tag{2.48}$$

Assim, para um processo  $AR(p)$  não há correlação entre  $X_t$  e  $X_{t-k}$  para  $k > p$  (Mills, 1990). Então, todos os valores de  $\phi_{kk}$  para  $k > p$  são zero e a PACF para um processo  $AR(p)$  puro apresenta um “corte” para zero para defasagens maiores que  $p$  (Enders, 1995).

### 2.2.3.2. Modelo de Médias Móveis de ordem $q$ : $MA(q)$

É o caso especial de (2.25) em que  $b(Z)$  é um polinômio de grau finito, isto é,

$$\Theta(Z) = 1 + \theta_1 Z + \dots + \theta_q Z^q \tag{2.49}$$

Então, (2.25) se reduz à forma do modelo  $MA(q)$ .

$$X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \tag{2.50}$$

A condição de invertibilidade para um processo MA( $q$ ) é que as raízes da equação:

$$\Theta(Z) = 1 + \theta_1 Z + \dots + \theta_q Z^q = 0 \quad (2.51)$$

estejam fora do círculo unitário. O processo é sempre estacionário.

Incluindo a média  $m$  ao processo e considerando apenas 1 termo de média móvel temos:

$$X_t = m + e_t + q_1 e_{t-1} \quad (2.52)$$

onde  $m$  e  $q$  são valores constantes. Tal processo é dito ser de médias móveis de ordem 1, MA(1), com notação  $X_t \sim \text{MA}(1)$ .

Tomando o valor esperado de  $X_t$ , tem-se que a média do processo MA(1) é dada por

$$E(X_t) = E(m + e_t + q_1 e_{t-1}) = m + E(e_t) + q_1 E(e_{t-1}) = m \quad (2.53)$$

Assim, o termo  $m$  em (1.31) é na realidade a média do processo (Hamilton, 1994).

A variância de  $X_t$  é

$$\begin{aligned} E(X_t - m)^2 &= E(e_t + q_1 e_{t-1})^2 \\ &= E(e_t^2 + 2q_1 e_t e_{t-1} + q_1^2 e_{t-1}^2) \\ &= s^2 + 0 + q_1^2 s^2 \\ &= (1 + q_1^2) s^2. \end{aligned} \quad (2.54)$$

A primeira autocovariância é

$$\begin{aligned} g_1 &= E(X_t - m)(X_{t-1} - m) = E(e_t + q_1 e_{t-1})(e_{t-1} + q_1 e_{t-2}) \\ &= E(e_t e_{t-1} + q_1 e_{t-1}^2 + q_1 e_t e_{t-2} + q_1^2 e_{t-1} e_{t-2}) \\ &= 0 + q_1 s^2 + 0 + 0 \\ &= q_1 s^2. \end{aligned} \quad (2.55)$$

As autocovariâncias de ordem superior são

$$\begin{aligned} g_j &= E(X_t - m)(X_{t-j} - m) = E(e_t + q_1 e_{t-1})(e_{t-j} + q_1 e_{t-j-1}) \\ &= E(e_t e_{t-j} + q_1 e_{t-1} e_{t-j} + q_1 e_t e_{t-j-1} + q_1^2 e_{t-1} e_{t-j-1}) \\ &= 0 + 0 + 0 + 0 \\ &= 0, \quad j > 1. \end{aligned} \quad (2.56)$$

Assim, a primeira autocorrelação é expressa dividindo-se (2.55) por (2.54)

$$r_1 = \frac{q_1 s^2}{(1 + q_1^2) s^2} = \frac{q_1}{(1 + q_1^2)}. \quad (2.57)$$

Utilizando o resultado de (2.56), conclui-se que as autocorrelações de ordem superior são todas zero (Hamilton, 1994). Isso implica que, apesar de as observações separadas por um período sejam correlacionadas, as observações separadas por mais de um período são não-correlacionadas, de forma que a “memória” do processo é de somente um período (Mills, 1990). E como a média e as autocovariâncias não são funções do tempo, um processo MA(1) é estacionário, independentemente do valor de  $q$  (Hamilton, 1994).

De acordo com (2.57), o valor da autocorrelação  $r_1$  segue o sinal de  $q_1$ . Assim, para valores positivos de  $q_1$  as autocorrelações serão positivas e valores grandes (pequenos) de  $X_t$  são tipicamente seguidos de um valor acima (abaixo) da média para  $X_{t+1}$ . Contrariamente, para valores negativos de  $q_1$  as autocorrelações serão negativas e valores grandes de  $X_t$  são seguidos por um valor pequeno de  $X_{t+1}$  (Cribari-Neto, 2000).

Em (2.57), como  $q_1$  deve ser um número real, é necessário que  $|r_1| \leq 0,5$ . Dessa forma, para qualquer valor de  $r_1$ , entre -0,5 e 0,5, existem dois valores correspondentes para  $q_1$ , pois o valor de  $r_1$  não se altera quando se substitui  $q$  por  $1/q$  (Hamilton, 1994).

Generalizando o processo descrito em (2.52), o modelo de médias móveis de ordem  $q$ , MA( $q$ ), é caracterizado por

$$X_t = m + e_t + q_1 e_{t-1} + q_2 e_{t-2} + \dots + q_q e_{t-q} = m + \sum_{j=0}^q q_j e_{t-j}, \quad q_0 \equiv 1, \quad (2.58)$$

onde  $\varepsilon_t \sim \text{WN}(0, \sigma^2)$  e  $(q_1, q_2, \dots, q_q)$  são números reais e a notação é  $X_t \sim \text{MA}(q)$ . Alternativamente,

$$X_t = q(B)e_t, \quad q(B) = 1 + q_1 B + \dots + q_q B^q \quad (2.59)$$

Um processo MA( $q$ ) é dito ser *invertível* se as raízes de

$$1 + q_1 z + q_2 z^2 + \dots + q_q z^q = 0$$

se encontrarem fora do círculo unitário.

A média de (2.58) é dada por

$$E(X_t) = E(m + e_t + q_1 e_{t-1} + L + q_q e_{t-q}) = m + E(e_t) + q_1 E(e_{t-1}) + L + q_q E(e_{t-q}) = m.$$

A variância de um processo MA( $q$ ) é

$$\begin{aligned} E(X_t - m)^2 &= E(e_t + q_1 e_{t-1} + L + q_q e_{t-q})^2 \\ &= s^2 + q_1^2 s^2 + L + q_q^2 s^2 \\ &= (1 + q_1^2 + L + q_q^2) s^2, \end{aligned}$$

desde que os  $\varepsilon$ 's sejam não-correlacionados. As autocovariâncias de ordem superior são

$$g_j = E(X_t - m)(X_{t-j} - m) = E(e_t + q_1 e_{t-1} + L + q_q e_{t-q})(e_{t-j} + q_1 e_{t-j-1} + L + q_q e_{t-j-q})$$

e, como os termos envolvendo produtos de  $\varepsilon$ 's em diferentes instantes de tempo têm valor esperado zero, para  $j > q$ ,  $g_j = 0$ , seguindo que

$$\begin{aligned} g_j &= (q_1 + q_{j+1} q_1 + L + q_q q_{q-j}) s^2, \quad j=1, 2, \dots, q, \\ g_j &= 0, \quad j > q. \end{aligned}$$

Dessa feita, para um processo MA( $q$ ) a ACF apresenta um “corte” para zero para defasagens maiores que  $q$  (Mills, 1990). Assim, o máximo valor da primeira autocorrelação de um processo MA( $q$ ) é dado por (Granger & Newbold, 1986)

$$r_1(\max) = \cos\left[\frac{p}{q+2}\right]. \quad (2.60)$$

A função de autocorrelação parcial de um processo MA( $q$ ) possui ordem infinita (Mills, 1990). Assim, se as raízes de  $q(B)$  em (2.59) forem reais, a PACF apresentará decaimento exponencial. Caso as raízes de  $q(B)$  sejam complexas, a PACF apresentará decaimento senoidal ao longo de zero (Mills, 1990).

Se os valores dos parâmetros ( $q_1, q_2, \dots, q_q$ ) forem positivos o decaimento até zero será oscilatório e caso os parâmetros sejam negativos o decaimento será direto (Enders, 1995). Um processo MA( $q$ ) é estacionário, pois a média e as autocovariâncias são independentes de  $t$  para quaisquer valores de ( $q_1, q_2, \dots, q_q$ ) (Hamilton, 1994).

Pode-se resumir que um processo  $MA(q)$  é descrito por:

- possuir uma ACF que é zero para defasagens maiores que  $q$ ; e
- possuir uma PACF que é uma combinação de decaimentos exponenciais e senoidais e tamanho infinito.

### 2.2.3.3. Modelo Misto Auto-Regressivo e Médias Móveis de Ordem $(p,q)$ : $ARMA(p,q)$

Seja  $b(Z)$  uma função racional da forma

$$b(Z) = \frac{1 + q_1 Z + \dots + q_q Z^q}{1 + f_1 Z + \dots + f_p Z^p} \quad (2.61)$$

Diz-se que (2.25) é um modelo  $ARMA(p,q)$  se for escrito na forma

$$X_t + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}. \quad (2.62)$$

Este modelo pode ser expresso na forma de operador de defasagem como

$$\Phi(B) X_t = \Theta(B) \varepsilon_t \quad (2.63)$$

onde  $\Phi(B)$  e  $\Theta(B)$  estão definidas em (2.28) e (2.49), respectivamente.

O processo é estacionário se as raízes de  $\Phi(B) = 0$  caírem todas do círculo unitário e será invertível se isso ocorrer para todas as raízes de  $\Theta(B) = 0$ .

Assim, se  $q(z)$  for igual a 1, a série temporal  $\{X_t\}$  é dita ser um processo auto-regressivo de ordem  $p$  (ou  $AR(p)$ ), e se  $f(z)$  for igual a 1, a série temporal  $\{X_t\}$  é dita ser um processo de média móvel de ordem  $q$  (ou  $MA(q)$ ) (Brockwell & Davis, 1996). A notação desse processo é  $X_t \sim ARMA(p,q)$ .

Uma solução estacionária da série temporal  $\{X_t\}$  comandada pelo processo (2.58) existe (e é única) se e somente se (Brockwell & Davis, 1996)

$$f(z) = 1 - f_1 z - \dots - f_p z^p \neq 0, \quad \text{para todo } |z| = 1.$$

Assim, a estacionariedade de um processo  $ARMA$  depende inteiramente dos parâmetros auto-regressivos  $(f_1, \dots, f_p)$  e não dos parâmetros de média móvel  $(q_1, \dots, q_q)$ .

Uma série temporal  $\{X_t\}$  é dita ser *causal* se

$$f(z) = 1 - f_1 z - \mathbf{L} - f_p z^p \neq 0 \quad \text{para todo } |z| \leq 1$$

isto é, se as raízes do polinômio

$$1 - f_1 z - \mathbf{L} - f_p z^p = 0$$

se encontrarem fora do círculo unitário (Cribari-Neto, 2000). A condição de causalidade leva à condição de estacionariedade (Granger & Newbold, 1986).

Uma série temporal  $\{X_t\}$  é dita ser *invertível* se

$$q(z) = 1 + q_1 z + \mathbf{L} + q_q z^q \neq 0 \quad \text{para todo } |z| \leq 1$$

isto é, se as raízes do polinômio

$$1 + q_1 z + \mathbf{L} + q_q z^q = 0$$

se encontrarem fora do círculo unitário (Cribari-Neto, 2000). Assim, se um processo ARMA for estacionário ou causal esse processo pode ser aproximado por um processo MA de ordem elevada. E caso o processo ARMA seja invertível, esse processo pode ser aproximado por um processo AR de ordem elevada (Granger & Newbold, 1986).

Para os modelos ARMA, as funções de autocorrelação (ACF) e autocorrelação parciais (PACF) decaem até o infinito em vez de apresentarem um “corte” em alguma defasagem particular, como ocorre com os processos AR e MA puros (Mills, 1990). Assim, para um processo ARMA( $p, q$ ), a ACF decairá, seja diretamente ou oscilando ao redor de zero, a partir da defasagem  $q$ . Por sua vez, a PACF decairá, seja diretamente ou oscilando ao redor de zero, a partir da defasagem  $p$  (Enders, 1995).

O modelo mais geral, ARIMA ( $p, d, q$ ), que foi estudando em detalhes em Box & Jenkins (1976), corresponde ao caso onde o operador  $\Phi(B)$  em (2.63) contém um fator da forma  $(1 - B)^d$ . Um modelo ARIMA ( $p, d, q$ ) corresponde a uma série que pode ser modelada como um processo ARMA ( $p, q$ ) estacionário, desde que venha a sofrer diferenciações até a ordem “ $d$ ”.

Se  $d$  for um número inteiro não-negativo, então uma série temporal  $\{X_t\}$  é dita ser um processo ARIMA( $p, d, q$ ) ou um processo integrado de ordem  $d$  se

$$Y_t = (1 - B)^d X_t = \Delta^d X_t \quad (2.64)$$

for um processo ARMA( $p, q$ ) causal (Brockwell & Davis, 1996), *i.e.*,  $Y_t \sim \text{ARMA}(p, q)$ . Assim, um modelo ARMA( $p, q$ ) é um modelo ARIMA( $p, 0, q$ ). Alternativamente,  $\{X_t\}$  deve satisfazer

$$f(B)[(1 - B)^d X_t - m] = q(B)e_t,$$

onde  $e_t \sim WN(0, \sigma^2)$ ,  $f(B)$  e  $q(B)$  são polinômios de ordem  $p$  e  $q$ , respectivamente,  $f(B)$  é um operador estacionário,  $m$  é a média de  $D^d X_t$  e  $d$  é a ordem de diferenciação. Como já foi dito anteriormente, a ordem de diferenciação será 0 ou 1 para a maioria dos processos e raramente  $d = 2$  (Granger & Newbold, 1986). A notação de um processo integrado de ordem  $d$  é  $X_t \sim \text{ARIMA}(p, d, q)$ .

## 2.2.4 – Condições Formais para Estacionariedade e Invertibilidade

### 2.2.4.1. Estacionariedade

A equação característica de um modelo ARIMA pode ser dada em termos do operador auto-regressivo AR:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) = 0, \quad (2.65)$$

onde  $B$  é operador de defasagem ("backshift") definido anteriormente, o qual, neste caso, pode ser tratado como sendo uma variável algébrica qualquer. Para que o modelo seja estacionário devemos ter as raízes desta equação característica situadas fora do círculo unitário (no plano complexo). Apesar desta condição para estacionariedade ser bastante clara, a restrição provocada por esta condição nos valores dos coeficientes não é tão fácil de ser obtida na prática.

Para um modelo AR(1) é fácil mostrar que a condição de estacionariedade implica em  $|\phi_1| < 1$ . Na prática, o valor real de  $\phi_1$  não é conhecido, devendo-se tomar o seu valor estimado. Para que um modelo AR(2) ou ARMA(2,  $q$ ) satisfaça a condição de estacionariedade deve-se ter:

$$\begin{aligned} |\phi_2| &< 1 \\ \phi_2 + \phi_1 &< 1 \end{aligned}$$

$$\phi_2 - \phi_1 < 1$$

Todas as três condições acima devem ser satisfeitas para que um modelo AR(2) ou ARMA (2, q) seja estacionário.

As condições de estacionariedade se torna mais complicada quando  $p > 2$ . Para  $p = 3$  e  $q = 4$  existem soluções gerais para as raízes da equação característica, porém são bastante complicadas. Não existe solução geral para polinômios característicos de ordem igual ou superior a 5. Neste caso, o conjunto de valores dos coeficientes  $\phi$  que satisfazem a condição de estacionariedade pode ser encontrado apenas numericamente.

Para  $p > 2$  pode ser checada a seguinte condição necessária (mas não suficiente) para a estacionariedade de um modelo:

$$\sum_{j=1}^p f_j < 1.$$

Modelos ARIMA também podem ser expressos na forma multiplicativa, a fim de facilitar a análise das condições de estacionariedade e invertibilidade. Podemos expressar a equação característica na forma multiplicativa conforme mostrado abaixo:

$$(1 - \lambda_1 B)(1 - \lambda_2 B) \dots (1 - \lambda_p B) = 0, \quad (2.66)$$

onde o conjunto de condições de estacionariedade se torna:

$$|\lambda_i| < 1, \text{ para todo } i.$$

- **Razão para a estacionariedade**

Para se fazer alguma inferência sobre as observações de uma série temporal deve-se ter a condição de estacionariedade satisfeita, sob pena de não poderem ser obtidas estimativas úteis para os parâmetros do modelo. Por exemplo, para um modelo cuja média é variável no tempo, torna-se impossível de se fazer uma estimativa à respeito da média num certo instante, pois, para cada instante, dispomos de apenas uma observação.

A situação se torna ainda pior quando se trata de um processo onde a variância também não é constante ao longo do tempo. Neste caso, deve-se estimar  $2n$  parâmetros ( $n$  médias e  $n$  variâncias) com apenas  $n$  observações.

Pode-se, ainda, mostrar que um modelo não estacionário fará previsões cujas variâncias crescem ilimitadamente, o que é um resultado indesejável.

• **Checando a estacionariedade na prática.**

Para uma certa série temporal, existem três maneiras de determinar se a condição de estacionariedade está satisfeita:

1.º - Exame visual da série, para notar se a média e/ ou a variância parecem estar variando com o tempo.

2.º - Exame da função de autocorrelação (*ACF*), para detectar se as autocorrelações estão decaindo rapidamente para zero. Caso as autocorrelações estejam decrescendo vagarosamente com o tempo deve-se suspeitar de uma média não-estacionária e considerar uma possível diferenciação nos dados.

3.º - Exames das estimativas dos parâmetros auto-regressivos para saber se eles satisfazem as condições mencionadas anteriormente.

Para decidir a respeito da estacionariedade de um modelo deve-se proceder ao exame visual da estimativa da função de autocorrelação e ao exame das restrições para as estimativas dos coeficientes dos termos auto-regressivos.

Para modelos com  $p > 2$ , como temos expressões mais complicadas para as restrições nos valores dos parâmetros, deve-se preferir o exame da função de autocorrelação estimada e o gráfico da série original. Deve-se verificar, também, se a condição necessária para a estacionariedade foi satisfeita, ou seja, se:

$$\sum_{j=1}^p f_j < 1.$$

#### 2.2.4.2. Invertibilidade

Os requisitos matemáticos para a invertibilidade são os mesmos que os para a estacionariedade sendo que no lugar da equação característica de um operador auto-regressivo (AR), utiliza-se a equação característica de um operador de médias móveis (MA):

$$(1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) = 0, \quad (2.67)$$

Para haver invertibilidade, é necessário que todas as raízes da equação anterior localizem-se fora do círculo unitário (no plano complexo). Para a determinação das restrições

dos parâmetros, as dificuldades são as mesmas que foram discutidas em relação à estacionariedade, sendo que agora utilizamos os parâmetros de médias móveis,  $\theta$ . De uma maneira análoga ao que foi visto no caso da estacionariedade, o operador de médias móveis também pode ser expresso na forma multiplicativa:

$$(1 - \alpha_1 B)(1 - \alpha_2 B) \dots (1 - \alpha_q B) = 0, \quad (2.68)$$

onde o conjunto de condições de invertibilidade se torna:

$$|\alpha_i| < 1, \text{ para todo } i.$$

- **Razão para a invertibilidade**

Um modelo ARIMA não invertível significa que, para efeito de previsão de observações futuras, os pesos dados às observações passadas não diminuem com a antigüidade das observações. No entanto, o bom senso indica que as observações mais recentes devem receber pesos maiores. A invertibilidade garante justamente que isto aconteça.

É fácil de entender a lógica da condição de invertibilidade quando se trata, por exemplo, de um modelo MA(1). Um modelo MA(1) pode ser escrito como um processo AR de ordem infinita, isto é,

$$(1 + \phi_1 B + \phi_1^2 B^2 + \dots) X_t = e_t \quad (2.69)$$

ou

$$X_t = e_t - \theta_1 X_{t-1} - \theta_1^2 X_{t-2} - \dots \quad (2.70)$$

Os coeficientes  $\theta$  possuem pesos relacionados com a defasagem da respectiva observação. Se a condição de invertibilidade não for satisfeita, então, os pesos se tornam maiores à medida que se retrocede no tempo. Por exemplo, para  $\theta_1 = 2$ , os pesos  $(\theta_1^k)$  assumem os seguintes valores:

k	$\theta_1^k$
1	2
2	4
3	8
4	16
...	...

Considerando um caso onde a condição de invertibilidade é satisfeita, ou seja,  $\theta_1 = 0,8$ , teremos os pesos das observações diminuindo à medida que retrocedemos no tempo, conforme mostra a tabela abaixo:

k	$\theta_1^k$
1	0,8
2	0,64
3	0,51
4	0,41
...	...

O que foi mostrado nesses exemplos também é válido para qualquer processo de médias móveis. Para isso, deve-se escrever o modelo na forma de um modelo auto-regressivo de ordem infinita. Assim, pode-se verificar que os coeficientes das observações passadas não diminuirão à medida que retrocedemos no tempo, a não ser que a condição de invertibilidade seja satisfeita.

### 2.2.5 – Utilidade dos Modelos ARMA (p,q)

Os modelos lineares para séries temporais, introduzidos por Box & Jenkins (1976), têm uma importância fundamental na econometria, uma vez que possibilitaram as primeiras análises e previsões no domínio do tempo. As funções do autocovariância, autocorrelação e autocorrelação parcial são facilmente implementáveis computacionalmente e contribuíram bastante nas pesquisas em séries temporais.

Pode-se fazer uma analogia dos modelos lineares para séries temporais com modelos lineares de regressão. Ambos assumiram a forte hipótese da linearidade na estrutura dos dados, hipótese esta que na maioria dos casos práticos não é satisfeita, mas que permite um modelamento muitas vezes satisfatório.

Recentemente, modelos mais gerais têm sido propostos, os quais esperam melhor explicar o comportamento de muitas séries conhecidas, assim como permitir que melhores previsões sejam feitas.

No trabalho de Cordeiro (1998) são abordados alguns destes modelos, chamados não lineares de séries temporais, os quais têm sido extensivamente estudados nos últimos anos.

### 2.2.6 – Estimação em Modelos Lineares de Séries Temporais

Nos subitens anteriores foram discutidas as principais propriedades de um processo estacionário, onde foram introduzidas as funções de autocovariância e autocorrelação de uma série temporal. Neste subitem, veremos o método de estimação desta função e a estimação dos parâmetros em modelo ARMA(p, q).

#### 2.2.6.1 - Estimação da média do processo.

Seja  $\{X_1, X_2, \dots, X_N\}$  um conjunto de observações de um processo  $\{X_t\}$  estacionário, onde

$$\begin{aligned} E[X_t] &= m, \quad \forall t \\ \text{Var}[X_t] &= s^2, \quad \forall t \\ \text{Cov}[X_t, X_{t+\tau}] &= R(\tau) = s^2 r(\tau). \end{aligned}$$

Nosso principal interesse consiste em estimar as funções de autocovariância,  $R(\tau)$ , e autocorrelação,  $\rho(\tau)$ , para o processo.

Para isso devemos começar por estimar a média das observações através da expressão:

$$\bar{X} = \frac{1}{N} \sum_{t=1}^N X_t$$

o qual é estimador não viesado para  $\mu$ .

Além disso,

$$\text{Var}[\bar{X}] = \frac{1}{N^2} \sum_{s=1}^N \sum_{t=1}^N \text{Cov}[X_s, X_t] = \frac{s^2}{N^2} \sum_{s=1}^N \sum_{t=1}^N r(t-s) = \frac{s^2}{N^2} \sum_{t=-(N-1)}^{N-1} \left(1 - \frac{|t|}{N}\right) r(t).$$

Para o caso de termos

$$\lim_{N \rightarrow \infty} \left\{ \sum_{t=-(N-1)}^{N-1} \left(1 - \frac{|t|}{N}\right) r(t) \right\}$$

finito, teremos  $\text{Var}[X] \rightarrow 0$  quando  $N \rightarrow \infty$ , e por ser não viesado então  $\bar{X}$  será, também, um estimador consistente para  $\mu$ .

### 2.2.6.2 - Estimação da função de autocovariância e da função de autocorrelação.

Um estimador assintoticamente não viesado para a função de autocovariância pode ser dado pela expressão:

$$\hat{R}^*(t) = \frac{1}{N - |t|} \sum_{i=1}^{N-|t|} (X_i - \bar{X})(X_{i+|t|} - \bar{X}), \quad t = 0, \pm 1, \pm 2, \dots, \pm(N-1). \quad (2.71)$$

Alguns autores, notadamente Parzen (1961) e Schaerf (1964), sugerem um estimador alternativo,

$$\hat{R}(t) = \frac{1}{N} \sum_{i=1}^{N-|t|} (X_i - \bar{X})(X_{i+|t|} - \bar{X}), \quad t = 0, \pm 1, \pm 2, \dots, \pm(N-1). \quad (2.72)$$

Este estimador possui um viés maior do que o estimador  $\hat{R}^*$ , mas em geral apresenta um menor erro médio quadrático e, por isso, é muitas vezes preferido ao estimador  $\hat{R}^*$ . Além disso, o estimador de  $R(\tau)$  compartilha com o próprio  $R(\tau)$  a propriedade de ser uma função semi-definida positiva.

Uma vez que a função de autocorrelação  $\rho(\tau)$  está relacionada à função de autovariância  $R(\tau)$  pela expressão

$$\rho(\tau) = R(\tau) / R(0),$$

uma estimativa natural de  $\rho(\tau)$  é dada por :

$$\hat{r}(t) = \hat{R}(t) / \hat{R}(0), \quad t = 0, \pm 1, \dots, \pm(N-1). \quad (2.73)$$

Os gráficos dos estimadores de  $\rho(\tau)$  e  $R(\tau)$  em função de  $\tau$  ( $\tau = 0, \pm 1, \dots, \pm(n-1)$ ) são chamadas de função de autocovariância a função de autocorrelação amostrais, respectivamente.

As funções de autocovariância e autocorrelação amostrais podem ser calculadas para qualquer conjunto de dados e não estão restritas a observações de séries temporais estacionárias (Brockwell & Davis, 1996). Para dados contendo tendência, a ACF exibirá um decaimento lento na medida que  $t$  aumenta. Para dados com um componente periódico determinístico, como sazonalidade, a ACF exibirá um comportamento similar ao período.

Assim, o correlograma pode ser utilizado como um indicador de não-estacionariedade da série temporal (Brockwell & Davis, 1996).

### 2.2.6.3 - Estimação de parâmetros

Para um modelo ARMA ( $p, q$ ), que pode ser na forma :

$$X_t = f_1 X_{t-1} + f_2 X_{t-2} + \dots + f_p X_{t-p} + e_t - q_1 e_{t-1} - \dots - q_p e_{t-p}, \quad (2.74)$$

o problema de estimação consiste em obter, a partir de um conjunto de observações, as estimativas para os parâmetros.

$$\Phi = (f_1, f_2, \dots, f_p)', \quad \mu = E[X_t], \quad \Theta = (q_1, q_2, \dots, q_p)'$$

e

$$S_e^2 = E(e_t^2)$$

onde  $\{e_t\}$  é um processo ruído branco com média zero e variância igual a  $\sigma_e^2$ .

Discutiremos a seguir dois métodos que podem ser utilizados para estimação dos parâmetros de um modelo ARMA.

#### • Método dos momentos.

Este método consiste em substituir os momentos amostrais, como por exemplo, a média amostral,  $X_t$ , a variância amostral,  $R(0)$  e a função de autocorrelação amostral,  $\rho(t)$ , em seus respectivos valores teóricos e resolver as equações resultantes.

Por exemplo, para um processo AR ( $p$ )

$$X_t = f_1 X_{t-1} + f_2 X_{t-2} + \dots + f_p X_{t-p} + a_t, \quad (2.75)$$

a média  $\mu = E(X_t)$  será estimado por  $X_t$ .

Para estimar  $\phi$  usa-se o fato de que

$$r(t) = f_1 r(t-1) + f_2 r(t-2) + \dots + f_p r(t-p) \text{ (Wei, p.44)}$$

para  $t > 1$ , o que leva ao seguinte sistema de equações de Yule- Walker :

$$\begin{aligned} r_1 &= f_1 + f_2 r_1 + f_3 r_2 + \dots + f_p r_{p-1} \\ r_2 &= f_1 r_1 + f_2 + f_3 r_1 + \dots + f_p r_{p-2} \end{aligned} \tag{2.77}$$

**M**

$$r_p = f_1 r_{p-1} + f_2 r_{p-2} + f_3 r_{p-3} + \dots + f_p$$

onde  $r_t = r(t)$  é valor da autocorrelação de lag  $t$ .

Substituindo  $r_k$  por sua estimativa podemos obter as estimativas para  $(f_1, f_2, \dots, f_p)$  resolvendo o sistema de equação lineares mostrado acima.

Assim temos,

$$\begin{bmatrix} \hat{f}_1 \\ \hat{f}_2 \\ \mathbf{M} \\ \hat{f}_p \end{bmatrix} = \begin{bmatrix} 1 & \hat{r}_1 & \hat{r}_2 & \mathbf{L} & \hat{r}_{p-2} & \hat{r}_{p-1} \\ \hat{r}_1 & 1 & \hat{r}_1 & \mathbf{L} & \hat{r}_{p-3} & \hat{r}_{p-2} \\ \mathbf{M} & & & & & \\ \hat{r}_{p-1} & \hat{r}_{p-2} & \hat{r}_{p-3} & \mathbf{L} & \hat{r}_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{r}_1 \\ \hat{r}_2 \\ \mathbf{M} \\ \hat{r}_p \end{bmatrix} \tag{2.78}$$

Estas estimativas são chamadas de estimativas de Yule-Walker. Após obtermos as estimativas para  $(f_1, f_2, \dots, f_p)$ , utilizamos o resultado

$$\begin{aligned} R(0) &= E[X_t X_t] = E[X_t (f_1 X_{t-1} + f_2 X_{t-2} + \dots + f_p X_{t-p} + e_t)] \\ &= f_1 R(1) + f_2 R(2) + \dots + f_p R(p) + \sigma_e^2 \end{aligned} \tag{2.79}$$

e então, o estimador por momentos para  $\sigma_e^2$  segue de

$$\hat{s}_e^2 = \hat{R}(0)(1 - \hat{f}_1 \hat{r}_1 - \hat{f}_2 \hat{r}_2 - \mathbf{L} - \hat{f}_p \hat{r}_p) \tag{2.80}$$

- **Método da máxima verossimilhança**

Para um modelo estacionário ARMA (p, q) mostrado em (2.74) a função densidade de probabilidade conjunta de  $e = (e_1, e_2, \dots, e_n)'$  é dada por:

$$P(e / \Phi, m, \Theta, s_e^2) = (2ps_e^2)^{-n/2} \exp\left[-\frac{1}{2s_e^2} \sum_{t=1}^n e_t^2\right]. \quad (2.81)$$

Rescrevendo (2.74) como

$$e_t = q_1 e_{t-1} + \dots + q_p e_{t-p} + X_t - f_1 X_{t-1} - f_2 X_{t-2} - \dots - f_p X_{t-p} \quad (2.82)$$

podemos escrever a função de verossimilhança em função dos parâmetros  $\Phi$ ,  $\mu$ ,  $\Theta$ , e  $\sigma_e^2$ .

Considerando  $X = \{X_1, X_2, \dots, X_n\}'$  e assumindo as condições iniciais.

$$X^* = \{X_{1-p}, \dots, X_{-1}, X_0\}' \text{ e } e^* = (e_{1-q}, \dots, e_{-1}, e_0)',$$

então, a log-verossimilhança condicional é dada por:

$$\log L^*(\Phi, m, \Theta, s_e^2) = -\frac{n}{2} \log 2ps_e^2 - \frac{S^*(\Phi, m, \Theta)}{2s_e^2} \quad (2.83)$$

onde

$$S^*(\Phi, m, \Theta) = \sum_{t=1}^n e_t^2(\Phi, m, \Theta / X^*, e^*, X) \quad (2.84)$$

é função soma dos quadrados condicionais.

Os valores de  $\mu$  e  $\Theta$  que maximizam a equação (2.83) são chamados de estimativas de máxima verossimilhança condicional.

Desde que  $\log L^*(\Phi, \mu, \Theta, \sigma_e^2)$  depende dos dados apenas através de  $S^*(\Phi, \mu, \Theta)$ , estas estimativas são as mesmas estimativas obtidas a partir da minimização da função soma dos quadrados condicionais,  $S^*(\Phi, \mu, \Theta)$ , a qual não contém  $\sigma_e^2$ .

### 2.2.7 – Construindo um Modelo ARIMA - Metodologia

Dada uma série temporal não-sazonal não-estacionária  $\{X_t\}$ , considere que ela possa ser representada por um modelo da classe ARIMA( $p,d,q$ ),

$$f(B)(1-B)^d X_t = q(B)e_t, \quad (2.85)$$

onde

$$f(B) = (1 - f_1 B - \mathbf{L} - f_p B^p),$$

$$q(B) = (1 + q_1 B + \mathbf{L} + q_q B^q).$$

O objetivo da metodologia de Box-Jenkins é encontrar um modelo estocástico linear da classe ARIMA que possa ter gerado  $\{X_t\}$  e que esse modelo possa ser utilizado para fornecer previsões de valores futuros da série (Granger & Newbold, 1986). Caso a série temporal  $\{X_t\}$  apresente sazonalidade,  $\{X_t\}$  pode ser representada por um modelo da classe SARIMA( $p,d,q$ ) ( $P,D,Q$ ), conforme a equação (2.100).

A estratégia de modelagem, tanto para modelos sazonais quanto para não-sazonais, é baseada em um ciclo de três etapas iterativas:

1. identificação do modelo;
2. estimação do modelo; e
3. verificação de diagnóstico.

A etapa de identificação consiste em selecionar valores para  $p$ ,  $d$ ,  $q$  e  $P$ ,  $D$ ,  $Q$  (no caso de modelos sazonais). Essa etapa envolve subjetividade e julgamento pessoal. Na etapa de estimação, os coeficientes identificados na etapa anterior são estimados usando técnicas estatísticas. A última etapa indica se o modelo identificado e estimado descreve adequadamente o comportamento dos dados da série  $\{X_t\}$ . Caso o modelo não seja adequado, o ciclo deve começar novamente (Cribari-Neto, 2000).

Um conceito importante nessa metodologia é o princípio da parcimônia (Enders, 1995). Tal princípio sugere que modelos mais simples, com poucos parâmetros, produzem melhores previsões que modelos superparametrizados. Um modelo parcimonioso ajusta bem os dados sem incorporar coeficientes inúteis. O objetivo é se aproximar do processo gerador original dos dados e não descrevê-lo exatamente (Enders, 1995).

## 2.2.7.1 - Identificação

Essa etapa é considerada a mais difícil e delicada, e não há consenso sobre qual a melhor estratégia a ser seguida (Granger & Newbold, 1986). Dentre as várias estratégias existentes, duas se destacam: a análise das funções de autocorrelação e autocorrelação parcial amostrais, e o uso de um critério de seleção de modelos.

Para utilizar a primeira estratégia é necessário reconhecer modelos AR, MA e ARMA por meio das características da ACF e da PACF. Para modelos AR( $p$ ), a ACF é uma combinação de decaimentos exponenciais e senoidais e a PACF apresenta um corte nas defasagens maiores que  $p$ . Para modelos MA( $q$ ), a ACF apresenta um corte nas defasagens maiores que  $q$  e a PACF é uma combinação de decaimentos exponenciais e senoidais. A Tabela 2.1 resume as propriedades da ACF e da PACF para diversos modelos ARIMA (Mills, 1990).

Para modelos sazonais, o comportamento da ACF e da PACF deve ser analisado também próximo da defasagem sazonal, por exemplo, defasagem 12 para dados mensais e defasagem 4 para dados trimestrais. A Tabela 2.2 resume as propriedades da ACF e da PACF para modelos ARIMA (Bowerman & O'Connell, 1987).

Além de identificar os valores para  $p$  e  $q$  (e os valores de  $P$  e  $Q$  para modelos SARIMA), o grau de diferenciação da série (valores  $d$  e  $D$ ) precisa ser conhecido. Para tanto, utiliza-se também a inspeção da ACF e da PACF amostrais ou um teste formal, tal qual o teste de raiz unitária.

Para um modelo não-sazonal, um comportamento suave persistente nas autocorrelações amostrais em defasagens altas indica não-estacionariedade, *i.e.*, necessidade de diferenciação. Assim, deve-se diferenciar a série para sucessivos valores positivos de  $d$  e examinar o correlograma de  $\{D^d X_t\}$  (Cribari-Neto, 2000).

O correlograma da série diferenciada  $d$  vezes que apresentar um decaimento rápido nas defasagens altas mostra-se estacionário e o valor de  $d$  correspondente indica o que será utilizado para estimar o modelo ARIMA. O mesmo procedimento vale para modelos sazonais, uma vez que valores elevados de autocorrelação nas defasagens sazonais indicam também não-estacionariedade.

Tabela 2.1 – Propriedades da ACF e da PACF para vários modelos ARIMA

Modelo	ACF	PACF
$(1,d,0)$	Decaimento exponencial ou oscilatório	$\phi_{kk}=0$ para $k > 1$
$(2,d,0)$	Decaimento exponencial ou senoidal	$\phi_{kk}=0$ para $k > 2$
$(p,d,0)$	Decaimento exponencial e/ou senoidal	$\phi_{kk}=0$ para $k > p$
$(0,d,1)$	$\rho_k = 0$ para $k > 1$	Dominado por decaimento exponencial
$(0,d,2)$	$\rho_k = 0$ para $k > 2$	Dominado por decaimento exponencial ou senoidal
$(0,d,q)$	$\rho_k = 0$ para $k > q$	Dominado pela combinação linear de decaimento exponencial e/ou senoidal
$(1,d,1)$	Decaimento exponencial a partir da defasagem 1	Dominado por decaimento exponencial a partir da defasagem 1
$(p,d,q)$	Decaimento exponencial e/ou senoidal depois da defasagem $q-p$	Dominado por decaimento exponencial ou senoidal depois da defasagem $q-p$

O teste de raiz unitária pode ser utilizado para descobrir se as raízes do polinômio de um modelo ARMA estão perto ou no círculo unitário (Brockwell & Davis, 1996). Uma raiz perto de 1 para o polinômio auto-regressivo sugere que os dados devem ser diferenciados, enquanto que para o polinômio de médias móveis uma raiz perto de 1 indica que os dados foram superdiferenciados (Brockwell & Davis, 1996).

A Segunda estratégia para identificar os valores de  $p$ ,  $d$ ,  $q$  é utilizar um critério de informação que selecione os modelos por meio de um conjunto de “regras” (Mills, 1990). Os critérios de seleção para modelos ARIMA mais utilizados são o *AIC* (*Akaike information criterion*), o *AICC* (*Akaike information criterion corrected*) e o *BIC* (*Bayesian information criterion*). Esses critérios incorporam um termo de penalidade para o aumento do número de parâmetros ( $p$  e  $q$ ) no modelo, de forma que modelos mais “parcimoniosos”, ou seja, com o menor número de parâmetros, sejam escolhidos.

Tabela 2 2 – Propriedades da ACF e da PACF para modelos SARIMA

Modelo	ACF	PACF
$(P,D,0)$	Decaimento	Picos nas defasagens $s, 2s, \dots Ps$ e corte após $Ps$
$(0,D,Q)$	Picos nas defasagens $s, 2s, \dots Qs$ e corte após $Qs$	Decaimento
$(P,D,0)$ ou $(0,D,Q)$	Picos nas defasagens $s, 2s, \dots Qs$ e corte após $Qs$	Picos nas defasagens $s, 2s, \dots Ps$ e corte após $Ps$
$(P,D,0)$ e $(0,D,Q)$	Decaimento rápido na defasagem sazonal	Decaimento rápido na defasagem sazonal
Nenhum operador sazonal	Valores pequenos em todas as defasagens sazonais (não há picos)	Valores pequenos em todas as defasagens sazonais (não há picos)

As equações para esses critérios, sendo  $T$  o número de observações, são (Cribari-Neto, 2000)

$$\begin{aligned}
 AIC &= -2\log \hat{L} + 2(p + q), \\
 AICC &= -2\log \hat{L} + \frac{2(p + q)T}{T - p - q - 1}, \\
 BIC &= -2\log \hat{L} + (p + q)\log T,
 \end{aligned}
 \tag{2.86}$$

onde  $\hat{L}$  representa a verossimilhança maximizada.

O critério  $AIC$  superestima assintoticamente a ordem verdadeira do modelo (Granger & Newbold, 1986) apresentando tendência a escolher modelos superparametrizados (Cribari-Neto, 2000). O  $AICC$  é uma versão corrigida do  $AIC$  que incorpora uma correção de viés para amostras finitas, possuindo uma penalidade mais forte para modelos de ordem elevada (Brockwell & Davis, 1996). O  $BIC$  é um critério consistente, de forma que ele fornece estimativas de  $p$  e  $q$  que convergem em probabilidade para os valores verdadeiros à medida que  $T$  tende a infinito (Brockwell & Davis, 1996). Já os critérios  $AIC$  e  $AICC$  não são consistentes.

Na prática, a seleção de modelos é feita calculando o valor do critério (o  $BIC$ , por

exemplo) para todos os modelos ARIMA associados aos valores de  $p$ ,  $d$  e  $q$  de forma que  $p, q = 0, 1, 2, 3, 4, 5$  e  $d = 0, 1$ . Assim, seleciona-se o modelo que apresenta o menor valor do BIC e modelos alternativos cuja diferença para o valor mínimo do BIC seja inferior a 2 (Brockwell & Davis, 1996). Para a modelagem SARIMA, a quantidade de modelos investigados é maior, pois além dos valores de  $p$ ,  $d$  e  $q$ , deve-se incluir ainda os valores para  $P, Q = 0, 1, 2$  e  $D = 0, 1$ . Porém, os modelos são selecionados pelos mesmos critérios que os utilizados para os modelos ARIMA.

Além dessa abordagem, os modelos alternativos podem ser obtidos calculando-se um valor  $\mathfrak{R}$  que é expresso por

$$\mathfrak{R} = \exp[-1/2T\{BIC(p_1, q_1) - BIC(p, q)\}], \quad (2.87)$$

onde  $BIC(p_1, q_1)$  é o valor do menor BIC e  $BIC(p, q)$  é valor BIC para um modelo particular. Assim, para valores de  $\mathfrak{R}$  menores do que 10, não há garantias suficientes para descartar o modelo particular (Poskitt & Tremayne, 1987).

#### 2.2.7.2 – Estimação

Assumindo que um modelo ARIMA da forma

$$(1 - f_1 B - \mathbf{L} - f_p B^p)(1 - B)^d X_t = (1 + q_1 B + \mathbf{L} + q_q B^q) e_t \quad (2.88)$$

seja escolhido conforme a etapa anterior, o objetivo agora é estimar, utilizando o método de máxima verossimilhança (ML), os parâmetros  $f = (f_1, \dots, f_p)'$ ,  $q = (q_1, \dots, q_q)'$  e  $S^2$ , a variância de  $e_t$ . Especificando  $x = (x_1, x_2, \dots, x_n)'$  como o vetor das observações originais (transformadas ou não) e especificando  $w = (w_1, w_2, \dots, w_T)'$  como o vetor das  $T = n - d$  diferenças estacionárias, de forma que  $w_t = D^d x_t$ , o modelo ARIMA( $p, d, q$ ) da equação (2.88) pode ser escrito como (Mills, 1990)

$$e_t = q_0 + q_1 e_{t-1} + \mathbf{K} + q_q e_{t-q} + w_t - f_1 w_{t-1} - \mathbf{K} - f_p w_{t-p}. \quad (2.89)$$

Considerando que os  $e_t$ 's são independentes e são originados de uma distribuição normal com média zero e variância  $S^2$ , a função de densidade de probabilidade conjunta de  $e = (e_1, e_2, \dots, e_T)'$  é dada por

$$p(\mathbf{e} | \mathbf{f}, \mathbf{q}, \mathbf{s}^2) = (2p\mathbf{s}^2)^{-T/2} \exp\left[-\frac{1}{2\mathbf{s}^2} \sum_{t=1}^T \mathbf{e}_t^2\right]. \quad (2.90)$$

A função de densidade de probabilidade conjunta de  $w$  (ou a função de verossimilhança dos parâmetros  $(\mathbf{f}, \mathbf{q}, \mathbf{s}^2)$ ) pode ser escrita como

$$L(\mathbf{f}, \mathbf{q}, \mathbf{s}^2 | \mathbf{w}) = g_1(\mathbf{f}, \mathbf{q}, \mathbf{s}^2) \exp\left[-\frac{1}{2\mathbf{s}^2} S(\mathbf{f}, \mathbf{q})\right], \quad (2.91)$$

onde  $g_1$  é uma função dos parâmetros  $(\mathbf{f}, \mathbf{q}, \mathbf{s}^2)$ . Assim, a estimativa da máxima verossimilhança (*ML*) dos parâmetros  $(\mathbf{f}, \mathbf{q}, \mathbf{s}^2)$  é obtida pela maximização da função de verossimilhança dada pela equação (2.90) (Mills, 1990).

A estimação da *ML* é difícil e geralmente requer muito tempo de processamento computacional. Desta forma, existem alternativas que aproximam a função de máxima verossimilhança ao ignorar a função  $g_1(\mathbf{f}, \mathbf{q}, \mathbf{s}^2)$ . Duas dessas alternativas são o *MQE* (mínimos quadrados exatos) e o *MQC* (mínimos quadrados condicional). Contudo, alguns estudos têm sugerido (Ansley & Newbold, 1980) que o método de máxima verossimilhança é superior aos demais.

Para a modelagem SARIMA o procedimento é idêntico ao mostrado para a modelagem ARIMA, com a superioridade da estimação por máxima verossimilhança sendo ainda mais pronunciada para modelos sazonais (Ansley & Newbold, 1980).

Contudo, apesar de a estimação pontual da função de máxima verossimilhança funcionar bem, a estimação intervalar ou as estimativas das variâncias não funcionam a contento (Ansley & Newbold, 1980). Isso ocorre porque:

- a distribuição dos estimadores em amostras de tamanho pequeno pode estar bem distante da distribuição assintótica normal (que é utilizada para a construção de intervalos de confiança);
- quando os parâmetros AR e/ou MA estão próximos do limite da superfície paramétrica, as variâncias assintóticas se tornam pequenas e qualquer viés nas estimativas afeta largamente a cobertura intervalar.

Desta maneira, os testes de significância para os valores estimados dos parâmetros  $\mathbf{f} = (f_1, \dots, f_p)'$ ,  $\mathbf{q} = (q_1, \dots, q_q)'$  em muitos casos podem ser viesados e, por isso, devem ser tomados com cautela.

Os testes de significância são calculados com base na hipótese nula que os parâmetros estimados são iguais a zero. Caso a estatística  $t$  de teste calculada como

$$t = \frac{|b|}{\sqrt{\text{var}(b)}}, \quad (2.92)$$

onde  $b$  corresponde ao valor estimados do parâmetro, for maior que o valor de 1,96, que corresponde ao valor do nível assintótico a 95% para a distribuição normal, então rejeita-se a hipótese nula que os parâmetros são nulos, ou seja, aceitam-se os valores estimados.

### 2.2.7.3 – Verificação de Diagnóstico

A correta especificação de um modelo ARIMA ou SARIMA é verificada no termo  $e_t$ , pois ele deve constituir um processo *ruído branco* (Granger & Newbold, 1986). Assim, a verificação da adequabilidade do modelo é efetuada nas autocorrelações amostrais dos erros ( $e_t$ ), as quais seguem assintoticamente uma distribuição normal, com média zero e desvio padrão  $n^{-1/2}$ , se forem provenientes de um *ruído branco*. Como os erros verdadeiros ( $e_t$ ) não são conhecidos, a inferência baseia-se nas estimativas dos erros, os resíduos  $\hat{e}_t$ .

Dessa forma, se o modelo estiver corretamente especificado, os resíduos não devem apresentar correlação serial, pois toda a dinâmica dos dados já foi capturada pelo modelo (Cribari-Neto, 2000). A autocorrelação amostral dos resíduos de ordem  $j$  é calculada como (Granger & Newbold, 1986)

$$r_j(\hat{e}) = \frac{\sum_{t=j+1}^T \hat{e}_t \hat{e}_{t-j}}{\sum_{t=1}^T \hat{e}_t^2}. \quad (2.93)$$

Então, os valores das autocorrelações residuais devem estar contidos no intervalo de confiança assintótico de 95% que é (Cribari-Neto, 2000)

$$\left[ -\frac{2}{\sqrt{T}}, \frac{2}{\sqrt{T}} \right], \quad (2.94)$$

onde  $T$  indica o número de observações da série.

Em adição ao exame das autocorrelações individuais dos resíduos um teste conjunto das primeiras  $m$  autocorrelações pode ser utilizado, que é conhecido por teste Ljung-Box. Tal teste compara o valor de

$$Q = T(T + 2) \sum_{j=1}^m (T - j)^{-1} r_j^2(\hat{e}_j) \quad (2.95)$$

com valores tabulados da distribuição do  $c^2$  (Qui-quadrado) com  $(m - p - q)$  graus de liberdade e com a rejeição da hipótese nula (de que o modelo é adequado) para valores de  $Q$  maiores que o valor crítico assintótico (Granger & Newbold, 1986). O valor de  $m$  deve ser pelo menos igual a  $\sqrt{T}$  (Cribari-Neto, 2000). Para reforçar a verificação de diagnóstico podem ser traçados os gráficos dos resíduos padronizados  $(a_t/\sigma_a)$ , de tal forma que picos inferiores a 3 ( $a_t < 3\sigma_a$ ) sugerem que os choques devem ser normalmente distribuídos e, portanto, o modelo se apresenta adequado.

### 2.2.8 – Modelando Séries Temporais Sazonais – O Modelo Sazonal Auto-Regressivo Integrado de Médias Móveis (SARIMA)

Suponha uma série temporal sazonal não-estacionária  $\{X_t\}$  observada  $s$  períodos por ano, de maneira que  $s = 4$  para séries trimestrais e  $s = 12$  para séries mensais. Uma forma de remover a sazonalidade da série e transformá-la em uma série estacionária  $\{Z_t\}$ , para que um modelo ARIMA possa ser empregado, é efetuar uma diferenciação sazonal, nos moldes da diferenciação vista anteriormente. Assim,

$$X_t - X_{t-s} = (1 - B^s)X_t = Z_t. \quad (2.96)$$

Contudo, em muitos casos é necessário modelar  $Z_t$  de acordo com o seu padrão sazonal, então

$$\Phi(B^s)(1 - B^s)^D X_t = \Theta(B^s)Z_t, \quad (2.97)$$

onde

$$\begin{aligned} \Phi(B^s) &= (1 - \Phi_{1s}B^s - \mathbf{L} - \Phi_{Ps}B^{Ps}), \\ \Theta(B^s) &= (1 + \Theta_{1s}B^s + \mathbf{L} + \Theta_{Qs}B^{Qs}). \end{aligned} \quad (2.98)$$

Pela equação (2.97), nota-se que o padrão sazonal é aleatório entre os ciclos  $s$  (Brockwell & Davis, 1996).

Se a sazonalidade da série  $Z_t$  tiver sido filtrada, um modelo ARIMA( $p,d,q$ ) regular pode representar  $Z_t$ , assim

$$f(B)(1-B)^d X_t = q(B)e_t, \quad (2.99)$$

onde

$$\begin{aligned} f(B) &= (1 - f_1 B - \mathbf{L} - f_p B^p), \\ q(B) &= (1 + q_1 B + \mathbf{L} + q_q B^q), \\ &\text{e } e_t \sim WN(0, \sigma^2). \end{aligned}$$

Combinando (2.97) e (2.99), chega-se a classe de modelos *sazonais multiplicativos* ARIMA( $p,d,q$ ) $\times$ ( $P,D,Q$ ) ou SARIMA,

$$f(B)\Phi(B^s)(1-B)^d(1-B^s)^D X_t = q(B)\Theta(B^s)e_t, \quad (2.100)$$

onde  $e_t \sim WN(0, \sigma^2)$ . Nessa classe de modelos permite-se tanto a diferenciação regular quanto a diferenciação sazonal. Nota-se que a série diferenciada pode ser representada usando tanto componentes auto-regressivos e de médias móveis regulares quanto sazonais. Em geral, o valor para  $D$  é raramente maior que um e os valores de  $P$  e  $Q$  não ultrapassam 2 (Brockwell & Davis, 1996).

O processo representado por (2.100) é *causal* se e somente se  $f(B) \neq 0$  e  $F(B) \neq 0$ , para  $|z| \leq 1$ , ou seja, se as raízes do polinômios auto-regressivos se encontrarem fora do círculo unitário (Brockwell & Davis, 1996).

Um modelo particular dessa classe de modelos é o chamado modelo “*airline*”, um modelo SARIMA( $0,1,1$ )( $0,1,1$ ). Esse modelo tem-se mostrado útil para especificar séries temporais sazonais (Cribari-Neto, 2000) e possui a seguinte forma

$$(1-B)(1-B^s)X_t = (1+q_1B)(1+\Theta_1B^s)e_t. \quad (2.101)$$

As autocorrelações desse modelo são (Cribari-Neto, 2000)

$$\begin{aligned}
r_1 &= \frac{q_1}{(1+q_1^2)}, \\
r_{s-1} = r_{s+1} &= \frac{q_1 \Theta_1}{(1+q_1^2)(1+\Theta_1^2)}, \\
r_s &= \frac{\Theta_1}{(1+\Theta_1^2)}, \\
r_k &= 0, \quad k = 2, \mathbf{L}, s+2, s+3, \mathbf{L}
\end{aligned} \tag{2.102}$$

Assim, o correlograma assume uma forma de identificação particular, com a primeira autocorrelação elevada, as seguintes todas zero e as autocorrelações próximas do termo sazonal com valores significativos também.

Deve-se acrescentar que as modelagens ARIMA/SARIMA permitem a inclusão de variáveis exógenas ao modelo. Dessa maneira, pode-se combinar um modelo de regressão com um modelo ARMA, por exemplo, o que resultaria em

$$X_t = a + f_1 X_{t-1} + \mathbf{L} + f_p X_{t-p} + b_0 X_t + b_1 X_{t-1} + \mathbf{L} + b_r X_{t-r} + e_t + q_1 e_{t-1} + \mathbf{L} + q_q e_{t-q},$$

onde  $a, b_i, f_i, q_i$  são parâmetros fixos e  $e_t$  é ruído branco. Tal classe de modelos é denominada ARMAX. Nota-se que caso os valores para  $b$  sejam iguais a zero, obtém-se o modelo ARMA( $p, q$ ) como caso particular.

### 2.2.9 – Métodos de Comparação de Previsões

Um dos métodos de escolha do melhor mecanismo de previsão é a comparação dos valores previstos ( $X_t$ ) com os valores observados da série ( $X_t$ ), o que caracteriza a acurácia ou a capacidade preditiva do mecanismo utilizado. Os três métodos mais populares de medição da acurácia utilizam os resíduos em seus cálculos (Kvanli, *et al.*, 1996). Esses métodos são o desvio absoluto médio (*DAM*), o erro quadrático médio (*EQM*) e o erro percentual absoluto médio (*EPAM*). Pankratz (1991) sugere a raiz quadrada do erro quadrado médio de (*RQEQMP*). Assim, os resíduos são definidos como

$$e_t = X_t - \hat{X}_t. \tag{2.103}$$

O desvio absoluto médio (*DAM*) é definido como a média dos valores absolutos de cada resíduo e é representado por

$$DAM = \frac{\sum |e_t|}{n}, \quad (2.104)$$

onde  $n$  é o número de valores previstos obtidos dos dados passados.

O erro quadrático médio (*EQM*) é a média dos valores quadráticos de cada resíduo, assim

$$EQM = \frac{\sum e_t^2}{n}. \quad (2.105)$$

O erro percentual absoluto médio (*EPAM*) considera o erro relativo de cada previsão. O erro relativo em cada período  $t$  é definido como  $e_t/X_t$ . Então,

$$EPAM = \frac{\sum \left| \frac{e_t}{X_t} \right|}{n}. \quad (2.106)$$

A raiz quadrada do erro médio de previsão (*RQEQMP*) nada mais é que a raiz quadrada da equação (2.105). Assim:

$$RQEQMP = \sqrt{\frac{\sum e_t^2}{n}}. \quad (2.107)$$

Não há consenso entre os estatísticos sobre qual método é preferível. Assim, se erros elevados de previsão são inaceitáveis, então o uso do EQM ou RQEQMP faz-se necessário. Entretanto, se é possível ignorar alguns erros elevados, o DAM funciona melhor. E o EPAM é utilizado para comparar a acurácia (precisão) de duas séries temporais diferentes (Kvanli, *et al.*, 1996). Dessa forma, o RQEQMP será utilizado como critério de acurácia para as comparações dos métodos de previsão apresentados neste trabalho e deles com o método dos indicadores.

### 2.2.10 – Justificativas da escolha do Método de Box-Jenkins

Em linhas gerais, há quatro abordagens de previsão baseada em dados de séries temporais (Gujarati, 2000): (1) modelos de regressão de equação única; (2) modelo de regressão de equações simultâneas; (3) modelos auto-regressivos de média móvel (ARMA); e (4) modelos de auto-regressão vetorial (VAR).

Diversos livros recentes ainda idolatram os modelos de equações simultâneas (Pyndick & Rubinfeld, 1988). No seu auge, durante as décadas de 1960 e 1970, modelos refinados da economia americana baseados em equações simultâneas dominaram a previsão econômica. Mas, nos últimos tempos, o glamour de tais previsões tem diminuído, graças aos choques de petróleo de 1973 e 1979 e à crítica de Lucas. O ataque desta crítica está no fato de que os parâmetros estimados a partir de um modelo econométrico dependem da política dominante no momento em que o modelo foi estimado e mudarão se houver uma alteração de política. Em suma, os parâmetros estimados não são invariáveis na presença de mudanças na política. Um modelo econométrico estimado com dados do passado terão pouco valor de previsão no novo regime.

A publicação por G.E.P. Box e G.M Jenkins de *Time Series Analysis: Forecasting and Control* (obra citada) abriu o caminho para uma nova geração de ferramentas de previsão, conhecida popularmente como metodologia de Box-Jenkins, mais referenciada como metodologia ARIMA. A ênfase desses novos métodos de previsão não está em construir modelos de equação única ou de equações simultâneas, mas em analisar as propriedades probabilísticas (ou estocásticas) de séries temporais econômicas em si mesmas, de acordo com a filosofia de “deixar que os dados falem por si mesmo”. Ao contrário dos modelos de regressão, nos quais  $Y_t$  é explicado por  $K$  regressores  $X_1, X_2, X_3, \dots, X_k$ , nos modelos de série temporal do tipo Box-Jenkins,  $Y_t$  pode ser explicado por valores passados (ou defasados) do próprio  $Y$  e dos termos de erro estocástico. Por esta razão, os modelos ARIMA são chamados às vezes de modelos ateóricos, pois não podem ser derivados de nenhuma teoria econômica que são freqüentemente a base de modelos de equação simultâneas.

Para o caso concreto que ora enfrentamos, qual seja, a previsão de receitas tributárias federais, optou-se por utilizar a Metodologia de Box- Jenkins pelos motivos que se passa a elencar:

1.º - quando as circunstâncias econômicas, políticas ou administrativas estão em fluxo intenso, modelos causais (equação única ou sistema de equações simultâneas) não conseguem capturar informações importantes sobre os fatores que provavelmente estão alterando os padrões históricos da série. Esse é particularmente o caso do Brasil, onde variações bruscas no cenário político-econômico tendem a inviabilizar a construção de modelos estruturais eficientes para a previsão econômica. Para contornar esse problema, utilizar-se de modelos de séries únicas, como na abordagem de Box-Jenkins, pode trazer resultados bastante promissores.

2.º - a abordagem de séries temporais normalmente apresenta melhor performance que a baseada na modelagem causal, ao menos no curto prazo, dado a incerteza associada a captura de todos os fatores econômicos relevantes que influenciam a geração da receita (Frank, 1993).

3.º - modelos causais de previsão tendem a ser, geralmente, mais complexos, exigindo uma grande quantidade de dados e requerendo um grau elevado de habilidade estatística. A modelagem introduzida por Box-Jenkins, além de trazer um tratamento estatístico extremamente robusto, pode ser implementada facilmente através de softwares comerciais de fácil manipulação, onde os fundamentos estatísticos estão inseridos no código do programa, sem a necessidade de grandes habilidades estatísticas por parte responsável pela previsão, a não ser àquelas necessárias à análise dos resultados.

4.º - Estudos anteriores (Melo, 2001; Pankratz, 1991; Frank, 1993 e outros) têm mostrado que os modelos ARIMA produzem bons resultados na previsão de receitas públicas, principalmente no curto prazo (até 6 meses), mostrando-se superior aos outros modelos normalmente aplicados na previsão de dados do setor privado, tais como alisamento exponencial e modelos de regressão única.

### 2.3 – MÉTODO DOS INDICADORES DA SRF

A Secretaria da Receita Federal, no que se refere à previsão das receitas tributárias federais, utiliza um método denominado Método dos Indicadores.

Tal método se baseia na aplicação de índices específicos sobre a arrecadação efetuada em determinado período para a previsão da receita do período seguinte. É elaborada por tributo e, visando a extração da sazonalidade dos dados, é feita em relação ao efetivo mês de arrecadação, para o mês correspondente do exercício financeiro seguinte.

Exemplificando: a previsão da receita do Imposto de Importação para março de 2001 é feita com base na arrecadação deste mesmo imposto em março de 2000.

Matematicamente, o método dos indicadores é assim definido:

$$A_t = A_{t-1}(1 + \Delta P)(1 + \Delta Q)(1 + \Delta L)(1 + \Delta U) \quad 2.108$$

Onde  $A_t$  é Arrecadação Prevista para o Período;

$A_{t-1}$  é a Arrecadação do Período Anterior;

$(1 + \Delta P)$  é o Efeito-preço;

$(1 + \Delta Q)$  é o Efeito-quantidade;

$(1 + \Delta L)$  é o Efeito-legislação; e

$(1 + \Delta U)$  é o Efeito-residual;

Os efeitos acima relacionados são definidos individualizadamente para cada tipo de tributo, levando em consideração as suas características econômicas específicas e a base de incidência a que se reporta.

Exemplificando: para o Imposto de Importação, os efeitos utilizados são os seguintes:

$\Delta P \rightarrow$  Variação cambial no período considerado;

$\Delta Q \rightarrow$  Variação do volume de importações;

$\Delta L \rightarrow$  Mudanças de alíquotas;

$\Delta U \rightarrow$  Efeito-residual  $\rightarrow$  outros fatores que possam influenciar a arrecadação, tais como crises externas, adoção de políticas protecionistas, etc.

Os órgãos de pesquisa de preços e acompanhamento da conjuntura econômica (IBGE, FGV, IPEA) são fontes importantes para se identificar quais são os índices de preço e quantidade que melhor se adequam aos vários tributos. Na Tabela 2.3 da página seguinte estão relacionados alguns tributos e seus principais indicadores de preço e quantidade.

A SRF mantém registro de séries históricas dos principais indicadores de preços e algumas séries de quantidade. As projeções dos parâmetros macroeconômicos (inflação, PIB, taxa de câmbio e taxa de juros) que influenciam os diversos indicadores são elaboradas pela Secretaria de Política Econômica do Ministério da Fazenda (SPE).

A Receita Federal faz a previsão da arrecadação do exercício seguinte (especificada mês a mês) em novembro do ano corrente e, em julho do ano seguinte faz uma correção da previsão.

Tabela 2 3 - Tributos e seus indicadores de preço e quantidade

TRIBUTOS/CONTRIBUIÇÃO	INDICADORES ESPECÍFICOS	
	PREÇO	QUANTIDADE
Imposto de Importação	Taxa de câmbio	Volume de importações tributadas, em dólar
IPI - Fumo	Índice de preços da indústria de fumo	Vendas de cigarro ao mercado interno
IPI - Bebidas	Índice de preços de bebidas	Volume de vendas de bebidas ao mercado interno
IPI - Automóveis	Índice de preços da indústria automobilística	Volume de vendas de automóveis ao mercado interno
Imposto de Renda Pessoa Física - IRPF	IGP - Índice Geral de Preços	Número de contribuintes
Imposto de Renda Pessoa Jurídica - IRPJ	IGP - Índice Geral de Preços	PIB
Imposto de Renda Retido na Fonte - IRRF - Trabalho	Variação nominal de salários	Nível de emprego
Imposto de Renda Retido na Fonte - IRRF - Capital	Taxa de juros	Volume em R\$ de aplicações financeiras
IOF - Imposto sobre operações financeiras	Variação nominal do volume de crédito e prêmios de seguro (em R\$)	
COFINS	IGP - Índice Geral de	PIB

### 3 – NATUREZA E TRATAMENTO DOS DADOS

Para a realização do presente trabalho, optou-se, inicialmente, por utilizar os dados das receitas federais administradas pela Secretaria da Receita Federal (SRF), relativos ao período de janeiro de 1989 a outubro de 2001; tendo sido previamente desconsiderados os dados de períodos anteriores, em virtude de a Constituição Federal de 1988 ter promovido uma grande mudança no Sistema Tributário Nacional, a partir de 1989, tornando tais observações estruturalmente não compatíveis com o período supracitado. Os valores observados foram corrigidos com base no IGP-DI – base dez/2000.

Assim, para o período em análise, tomou-se a decomposição das Receitas Administradas pela SRF em 15 subséries, cada uma representando um determinado imposto ou uma contribuição ou a soma de algumas categorias semelhantes de tributos, conforme apresentado na tabela 3.1, onde se informa a participação percentual de cada tributo na receita federal total administrada, representando esta cerca de 92% de toda a arrecadação federal.

Uma análise inicial da composição das receitas tributárias federais mostra que 52,44% de toda a arrecadação federal (exceto as contribuições para o INSS) são provenientes dos impostos, enquanto 46,79% provêm das contribuições. O Imposto sobre a Renda (total) é a maior fonte de receitas da União, representando mais de 33% da arrecadação administrada pela SRF (mais de 30% de toda a arrecadação da União). Do lado das contribuições, a Contribuição para Financiamento da Seguridade Social (COFINS) é a que mais contribui para a arrecadação total (aproximadamente 24%). Ressalta-se que não foi incluída na tabela a receita proveniente do Imposto de Exportação, por representar parcela insignificante da arrecadação total (menos de 0.1%). Outro imposto de arrecadação bastante inexpressiva, mas indicado na tabela 3.1, é o ITR, representando apenas 0,16% da receita total administrada.

Por fim, cabe aqui uma outra observação importante. A composição das receitas apresentada neste capítulo se refere ao ano de 2000, devendo-se, obviamente, alterar-se, se formos adotar um outro período de análise ou uma média de todo o período estudado, qual seja, jan/1989 a out/2001. Isto ocorre porque houve, no período em questão, diversas alterações na estrutura tributária brasileira, de tal forma que a participação de cada tributo na arrecadação total variou significativamente ao longo deste período. As principais alterações ocorridas e seus efeitos sobre a arrecadação serão aqui analisadas, primeiramente de forma global e em seguida individualmente para cada tributo.

Tabela 3 1 - Arrecadação total administrada pela SRF (2000) – em milhões de R\$

ORDEM	AGREGAÇÃO	VALOR	PARTICIPAÇÃO (%)
<b>1</b>	<b>IMPOSTO SOBRE IMPORTAÇÃO</b>	<b>8,510.20</b>	<b>5.12</b>
<b>2</b>	<b>IMPOSTO S/PRODUTOS INDUSTRIALIZADOS - TOTAL</b>	<b>18,838.90</b>	<b>11.34</b>
2.1	IPI - FUMO	1,997.90	1.20
2.2	IPI - BEBIDAS	1,942.00	1.17
2.3	IPI - AUTOMÓVEIS	2,374.70	1.43
2.4	IPI - OUTROS - TOTAL	5,061.20	3.05
2.5	IPI - VINC. ÀS IMPORTAÇÕES	7,463.00	4.49
<b>3</b>	<b>IMPOSTO SOBRE A RENDA - TOTAL</b>	<b>56,396.50</b>	<b>33.94</b>
3.1	IMPOSTO S/A RENDA-PESSOAS FÍSICAS	3,657.10	2.20
3.2	IMPOSTO S/A RENDA-PESSOAS JURÍDICAS	17,655.40	10.62
3.3	IMPOSTO S/ RENDA RETIDO NA FONTE	35,084.00	21.11
<b>4</b>	<b>IMPOSTO S/OPERAÇÕES FINANCEIRAS - TOTAL</b>	<b>3,126.60</b>	<b>1.88</b>
<b>5</b>	<b>IMPOSTO TERRITORIAL RURAL</b>	<b>267.20</b>	<b>0.16</b>
<b>6</b>	<b>RECEITA DE IMPOSTOS</b>	<b>87,139.40</b>	<b>52.44</b>
<b>7</b>	<b>CONT PROVISORIA S/MOVIIM FINANCEIRA - CPMF</b>	<b>14,544.70</b>	<b>8.75</b>
<b>8</b>	<b>CONTRIBUIÇÃO P/ FINANC. SEGURIDADE SOCIAL-COFINS</b>	<b>39,903.30</b>	<b>24.01</b>
<b>9</b>	<b>CONTRIBUIÇÃO PARA O PIS / PASEP</b>	<b>10,043.20</b>	<b>6.04</b>
<b>10</b>	<b>CONTRIBUIÇÃO SOCIAL SOBRE O LUCRO LÍQUIDO</b>	<b>9,278.10</b>	<b>5.58</b>
<b>11</b>	<b>CONTRIBUIÇÃO P/ PLANO DE SEGURIDADE SOCIAL SERVIDORES</b>	<b>3,626.70</b>	<b>2.18</b>
<b>12</b>	<b>CONTRIBUIÇÕES PARA O FUNDAF</b>	<b>372.40</b>	<b>0.22</b>
<b>13</b>	<b>RECEITA DE CONTRIBUIÇÕES</b>	<b>77,768,40</b>	<b>46.79</b>
<b>14</b>	<b>OUTRAS RECEITAS ADMINISTRADAS</b>	<b>1,275.20</b>	<b>0.77</b>
<b>15</b>	<b>RECEITA ADMINISTRADA PELA SRF</b>	<b>166,183.00</b>	<b>100.00</b>

### 3.1 - PRINCIPAIS FATOS QUE AFETARAM A ARRECADAÇÃO NO PERÍODO 1990-2000<sup>6</sup>

Em 1990, com a edição do Plano Collor I, foram estabelecidas, dentre outras, as seguintes medidas que contribuíram para o aumento de arrecadação:

- retenção de todos os ativos financeiros e tributação desses ativos através do IOF, cuja arrecadação passou de 0,15%, em 1989, para 1,30% do PIB, em 1990;

<sup>6</sup> Dados formatados a partir de informações disponíveis na página da Secretaria da Receita Federal na Internet ([www.receita.fazenda.gov.br](http://www.receita.fazenda.gov.br)).

- permissão para pagamento de tributos atrasados com os cruzados novos retidos, o que induziu um grande número de contribuintes a acertar as contas com o Fisco, aumentando a arrecadação de todos os tributos.

Em 1991, houve um processo de desindexação da economia, extinguindo-se, em fevereiro, a BTN-Fiscal, índice indexador dos tributos. Em seu lugar, foi criada a Taxa Referencial de Juros (TR), que funcionou, precariamente, como indexador tributário, porém, devido a questionamentos judiciais, foi abandonada pelo Governo Federal. Assim, os tributos ficaram, a partir de julho, sem qualquer indexação. Isto gerou uma queda significativa de arrecadação, dado que a inflação se acentuou naquele período. Por essa razão, as arrecadações do IRPJ, do IRPF e da CSLL caíram drasticamente, passando, em seu conjunto, de 2,41%, em 90, para 1,34% do PIB, em 1991.

Em 1992, houve a reindexação dos tributos, através da criação da Unidade Fiscal de Referência – UFIR. Adicionalmente, foi editada a Lei n.º 8.383/91 que estabeleceu, a partir de janeiro/92, o sistema de bases correntes para o Imposto de Renda das Pessoas Jurídicas – IRPJ. O sistema de apuração do Imposto passou de anual para mensal, fazendo com que, nesse ano, ocorressem arrecadações relativas ao ano-base 91 (cotas) e ao próprio ano-calendário 92. Isto elevou, significativamente, a arrecadação do IRPJ e da CSLL.

Em 1993, houve o reconhecimento da constitucionalidade da COFINS por parte do Supremo Tribunal Federal e uma elevação da alíquota da Contribuição Social sobre o Lucro Líquido de 15% para 23%, para as entidades financeiras. Adicionalmente, ocorreu uma série de pagamentos, pelas empresas, do imposto referente ao lucro inflacionário acumulado com alíquota reduzida de 5% (Lei 8.541/92). A Medida Provisória n.º 368/93, por sua vez, estabeleceu encurtamentos dos períodos de apuração e prazos de arrecadação para fatos geradores a partir de 1/11/93, resultando em ganho de arrecadação de cerca de 20 dias para o IPI.

Em 1994, teve início a cobrança do Imposto Provisório sobre Movimentação Financeira – IPMF (a partir de janeiro/94). Este imposto gerou uma arrecadação, no ano, de 4,98 bilhões, correspondendo a 7,74% do total arrecadado. Outras ocorrências influenciaram significativamente a arrecadação naquele ano, entre as quais se passa a relatar:

- a retomada do fluxo de pagamentos da COFINS por parte de grande número de contribuintes, em razão do julgamento da constitucionalidade dessa contribuição no final de 1993;

- a reestruturação das alíquotas, por faixas, do Imposto de Renda Retido na Fonte sobre o trabalho. A faixa com alíquota de 25% passou para 26,6% e foi criada uma nova faixa com alíquota de 35% (Lei n.º 8.848/94);
- a elevação da alíquota da Contribuição Social sobre o Lucro Líquido do setor financeiro de 23% para 30%, com reflexo na arrecadação a partir de julho/94. O aumento da alíquota foi decorrente da criação do Fundo Social de Emergência – FSE, por intermédio da Emenda Constitucional n.º 01 de 1994.

Nesse ano ocorre um dos fatos econômicos mais significativo da década, com forte influência na arrecadação tributária nacional: a implantação do Plano Real, a partir de julho/94, com redução acentuada da inflação. Este fato contribuiu para que as bases de cálculo dos principais tributos (IPI, IRPJ, COFINS e Contribuição Social sobre o Lucro Líquido) deixassem de sofrer os efeitos negativos, para arrecadação, decorrentes da corrosão inflacionária. Este efeito corrosivo, conhecido como "Efeito Tanzi", foi praticamente eliminado. Além disso, o início do Plano Real proporcionou um aquecimento acentuado do consumo, notadamente no segundo semestre de 1994, com repercussão positiva no resultado das empresas e, em consequência, na arrecadação.

Em 1995, o Governo Federal instituiu a limitação da compensação de prejuízo ao percentual de 30% do lucro obtido pelas empresas e elevou de 3,5% para 5% (regra geral) o percentual para estabelecimento da base de cálculo do lucro presumido/estimativa, fatos que tiveram repercussão positiva na arrecadação daquele ano. Adicionalmente, houve uma elevação de 62%, em relação a 1994, no volume das importações tributadas, principalmente de veículos, com acentuada repercussão na arrecadação do Imposto sobre Importações.

Em 1996, houve uma grande redução no volume das importações tributadas, principalmente de veículos, em razão das restrições impostas pelo aumento das alíquotas, retraindo o aumento observado no ano de 1995. Mais uma vez, fez-se uma reestruturação da tabela de retenção do Imposto de Renda na Fonte, ampliando-se a faixa de isenção, reduzindo-se alíquota de 26,6% para 25% e excluindo-se a alíquota de 35%. Tais fatos, em conjunto com a redução das alíquotas do IOF nas operações de crédito de 18% para 6% (pessoas físicas) e de 3% para 1,5% (pessoas jurídicas), afetaram negativamente a arrecadação daquele ano, que só não foi mais drástica porque ocorreu uma elevação de conversões de depósitos judiciais em renda da União (R\$ 2.782,6 milhões em 1996 contra R\$ 1.736,3 milhões em 1995).

Em 1997, um fato extremamente importante contribuiu, de forma bastante significativa, para o aumento da arrecadação federal. Trata-se da entrada em vigor, a partir de 23/01/97, da Contribuição Provisória sobre Movimentação Financeira – CPMF, cuja arrecadação, no ano de 1997 foi de R\$ 6,91 bilhões, correspondendo a 6,45% da arrecadação das receitas administradas pela SRF. A arrecadação do Imposto sobre Importações, por sua vez, sofreu um considerável acréscimo, em virtude da elevação do valor, em dólar, das importações tributadas, especialmente de veículos, com crescimento de 68% em relação ao ano anterior. Alguns outros fatos influenciaram de alguma forma a arrecadação daquele ano, dentre os quais destacamos:

- crescimento de 12% nas vendas de automóveis ao mercado interno;
- introdução da obrigatoriedade de retenção e recolhimento ao Tesouro, pelos órgãos públicos, a partir de 01/01/97, dos tributos oriundos do fornecimento de bens ou da prestação de serviços a esses órgãos (Lei n.º 9.430/96 – art. 64);
- elevação da alíquota do Imposto sobre Operações Financeiras – IOF sobre operações de crédito relativas às pessoas físicas de 6% para 15% a partir de 05/05/97 (Dec. n.º 2.219/97).

Em 1998, podem ser destacados os seguintes fatos relevantes que tiveram fortes influências sobre o comportamento da arrecadação:

- a ocorrência de arrecadação relativa à tributação dos saldos acumulados até 31/12/97 em fundos de investimento de renda fixa no valor de R\$ 1.763 milhões (janeiro a março); elevação da alíquota do imposto incidente sobre estas aplicações de renda fixa de 15% para 20% e mudança na sistemática de tributação relativamente às aplicações em fundos de renda fixa, passando de tributação no momento do resgate para tributação no vencimento da aplicação;
- pagamento referente a débitos em atraso, efetuado pelo setor público, no valor total de cerca de R\$ 1,77 bilhões;
- elevação das alíquotas do Imposto de Importação em três pontos percentuais, a partir de 13 de novembro de 1997 (Dec. n.º 2.376/97);
- elevação das alíquotas do IPI sobre bebidas em cerca de 10%, a partir de 21 de novembro de 1997 (AD n.º 74/97);

- elevação da alíquota do IRRF-Rendimentos do Trabalho de 25% para 27,5%, a partir de 1º de janeiro de 1998 (Lei n.º 9.532/97);
- redução de 16,2% no volume de vendas de cigarros no mercado interno;
- redução de 28,6% no volume de vendas de automóveis no mercado interno e redução de alíquotas a partir de agosto de 1998 (Dec. n.º 2.376/97);
- incremento no valor das compensações, no IPI, relativas ao crédito presumido na exportação pelo pagamento de PIS e COFINS;
- redução da alíquota do IOF, de 15% para 6%, nas operações de crédito das pessoas físicas, a partir de 10 de julho de 1998 (Portaria n.º 157 de 09/07/98); e
- redução do nível de atividade econômica.

Em 1999, arrecadação dos impostos e contribuições administrados pela Secretaria da Receita Federal e das demais receitas (taxas e contribuições controladas por outros órgãos, exclusive as contribuições previdenciárias) atingiu o valor de R\$ 15.021,5 milhões, no mês de dezembro de 1999, e de R\$ 151.516,6 milhões, no total.

A arrecadação da receita administrada pela SRF, no ano de 1999, apresentou crescimento nominal de 21,04% e real de 8,52%, em relação ao ano de 1998. No tocante à receita total, o crescimento nominal foi de 13,80% e o real de 2,06%.

Essa diferença de resultados entre a receita administrada (+8,52%) e a receita total (+2,06%) decorreu, basicamente, de arrecadações atípicas relativamente às “demais receitas”, especialmente outorga de serviços de telecomunicações, produto de depósitos abandonados, superávit financeiro, participações e dividendos, cujo montante, em 1999, importou em R\$ 5,2 bilhões, contra R\$ 11,9 bilhões, em 1998, o que concorreu para o decréscimo real de 47,51% nas “demais receitas”.

A arrecadação total das receitas federais, no mês de dezembro de 1999, apresentou crescimento nominal de 15,07% e real de 13,67% em relação ao mês anterior.

Esse desempenho foi decorrente, principalmente, de efeitos sazonais, especialmente a tributação e recolhimento do IRRF relativo ao 13º salário. Os demais fatores que também concorreram para o resultado alcançado foram os seguintes:

- recolhimento referente a cinco semanas de fatos geradores, neste mês, contra quatro semanas do mês de novembro, influenciando o desempenho dos tributos com apuração semanal, como: CPMF (+46,07%), IOF (+19,62%) e IRRF (+70,61%); e

- recolhimento, antecipado, relativo ao ajuste do IRPJ, no valor de R\$ 500 milhões.

Os decréscimos reais de 19,62% e 34,76% observados na arrecadação de COFINS e CSLL, respectivamente, em relação ao mês anterior, são decorrentes da arrecadação extra ocorrida em novembro, referente a conversão de depósitos judiciais no valor de R\$ 694 milhões (COFINS) e R\$ 312 milhões (CSLL).

Com relação à arrecadação da receita administrada total pela SRF, constatou-se um crescimento real de 8,58%, em razão, principalmente, dos seguintes fatores:

1. elevação da alíquota da **COFINS** de 2% para 3% e da obrigatoriedade do pagamento desta contribuição pelas entidades financeiras (Lei nº 9.718/98), a partir do mês de março de 1999; e
2. elevação da alíquota da CPMF de 0,20% para 0,38%, a partir de 17/06/99 (EC nº 21/99).

A seguir, estão relacionados outros tributos/contribuições com variações relevantes (positivas ou negativas) e os principais fatores que contribuíram para o desempenho da arrecadação do mês de dezembro de 1999 em relação a igual mês do ano anterior:

- **IPI-BEBIDAS (-33,58%)**: não pagamento de tributo, pela compensação do imposto devido com outros créditos, de terceiros, inclusive em virtude de decisão judicial;
- **IPI-AUTOMÓVEIS (+167,14%)**: esse desempenho foi decorrente, basicamente, da permissibilidade a partir de agosto/99 (Lei nº 9.826/99) da saída, do estabelecimento industrial, de insumos (peças e acessórios) destinados à montagem de veículos, com suspensão do IPI, ficando o recolhimento a cargo das montadoras. Esse fato vem repercutindo, positivamente, na arrecadação do IPI-Automóveis desde agosto/99 e concorreu para que a arrecadação, no ano de 1999, tenha sido praticamente igual, em termos reais, ao ano de 1998, a despeito de um decréscimo de 9,43% nas vendas, e redução das alíquotas para automóveis médios e populares, conforme Decreto nº 2.980/99 (acordo automotivo);
- **IRPJ (+40,30%)**: recolhimento, antecipado, relativo ao ajuste do IRPJ, no valor de R\$ 500 milhões;

- **IRRF-REMESSAS AO EXTERIOR (+81,37%)**: elevação no volume de remessas ao exterior e alteração na política cambial;
- **IOF (-9,75%)**: redução da alíquota do IOF, de 6% para 1,5%, nas operações de crédito das pessoas físicas (Port. nº 385, de 14/10/1999);
- **CSLL (-59,42%)**: recolhimento, em dezembro/98, de débitos em atraso no valor de R\$ 609 milhões;
- **CPSSS (+48,83%)**: tributação e recolhimento referente ao pagamento do 13º salário. No ano anterior essa tributação ainda não existia; e
- **OUTRAS RECEITAS ADMINISTRADAS (+114,56%)**: alocação das receitas oriundas de concursos e prognósticos neste item.

Em 2000, a arrecadação dos impostos e contribuições administrados pela Secretaria da Receita Federal e das demais receitas (taxas e contribuições controladas por outros órgãos, exclusive as contribuições previdenciárias) atingiu o valor de R\$ 17.022,1 milhões, no mês de dezembro e de R\$ 176.020,3 milhões no período de janeiro a dezembro.

A arrecadação das receitas administradas pela SRF, relativa ao mês de dezembro de 2000, apresentou crescimento nominal de 12,18% e real de 11,56% em relação ao mês anterior.

Este resultado decorreu, principalmente de efeitos sazonais, especialmente a tributação e recolhimento do IRRF-Rendimentos do Trabalho e CPSSS relativos ao 13º salário, o que justifica os crescimentos reais de 61,36% e 53,36%, respectivamente, na arrecadação desses tributos.

Outro fator relevante foi a arrecadação extra no valor de R\$ 531 milhões (IRPJ: R\$ 336 milhões e CSLL: R\$ 195 milhões) decorrente de pagamentos em atraso e liquidação de processos de parcelamento, contribuindo para os crescimentos reais de 35,58% (IRPJ) e 34,93% (CSLL).

A arrecadação relativa às receitas federais, administradas pela SRF, do mês de dezembro de 2000 apresentou crescimento nominal de 12,98% e real de 3,10% em relação a dezembro do ano anterior.

A seguir, estão enumerados os principais itens de receita que apresentaram variações reais significativas, quer positivas ou negativas, e as justificativas para estes comportamentos:

- **IPI-AUTOMÓVEIS (+45,17%):** elevação no volume de vendas de automóveis ao mercado interno em 55,75%;
- **IRRF-Rendimentos do Capital (-26,72%):** redução da taxa de juros sobre aplicações financeiras no mercado de capitais;
- **CPMF (-31,02%):** redução da alíquota de 0,38% para 0,30% a partir do mês de julho/00;
- **COFINS (+17,88%) e PIS/PASEP (+13,91%):** mudança na forma de tributação relativamente ao setor de combustíveis e derivados, passando a incidência dessas contribuições somente na saída da refinaria e a obrigatoriedade, pelas montadoras de veículos, de recolher na condição de contribuintes substitutos a COFINS e o PIS/PASEP devidos pelos comerciantes varejistas; e
- **CSLL (+65,46%):** arrecadação extra ocorrida em dezembro/00 no valor de R\$ 195 milhões referente a pagamentos de débitos em atraso e liquidação de processos de parcelamento

No ano de 2000, a arrecadação das receitas administradas pela SRF registrou crescimento nominal de 16,57% e real de 2,62% em relação ao ano de 1999.

Em 1999 ocorreram arrecadações extras em valores significativos, especialmente a partir do mês de fevereiro, contribuindo com uma participação importante no valor total arrecadado. Ao longo daquele ano foram efetuadas, também, alterações na legislação, como o aumento de 2% para 3% na alíquota da COFINS e inclusão das entidades financeiras em sua base de cálculo (março/99) e, ainda, reinício da cobrança da CPMF a partir do mês de julho de 1999.

As alterações na legislação, portanto, produziram efeitos parciais no ano de 1999, e integral, no ano de 2000. Além disso, ocorreram outros fatores que produziram efeitos pontuais na arrecadação de 1999.

As alterações introduzidas na legislação tributária, ainda em 1999, especialmente na COFINS e CPMF, em conjunto com a arrecadação de depósitos judiciais e administrativos, em 2000, produziram acréscimo de arrecadação de R\$ 33.259 milhões. O impacto deste acréscimo foi atenuado pela arrecadação, no valor de R\$ 28.113 milhões, proveniente, em grande parte, de fatores atípicos ocorridos em 1999.

A conjugação desses fatores e os relacionados abaixo, explicam o crescimento real de 2,62% na arrecadação das receitas administradas pela SRF, no ano de 2000, em relação ao ano de 1999.

A seguir estão relacionados outros itens de receita com variações relevantes e os principais fatores que contribuíram para este comportamento no ano de 2000 em relação ao ano anterior:

- **IMPORTAÇÃO-Petróleo: (-22,49%):** redução da alíquota de 9% para 6% a partir de janeiro de 2000;
- **IPI-AUTOMÓVEIS (+113,86%):** permissão, a partir de agosto/99, da saída de insumos (peças e acessórios) destinados à montagem de veículos com suspensão do IPI, ficando o recolhimento a cargo das montadoras (Lei n.º 9.826/99) e elevação do nível das vendas em 20,33%; e
- **IRRF-RENDIMENTO DO CAPITAL (-31,38%):** redução do rendimento das aplicações financeiras em consequência da queda nas taxas de juros e recolhimento de R\$ 2.275 milhões, em 1999, contra R\$ 474 milhões, em 2000, relativamente às operações de swap por conta da alteração da legislação e desvalorização cambial.

### 3.2 - ANÁLISE INDIVIDUAL DAS SÉRIES TRIBUTÁRIAS

Para todas as séries estudadas, verificou-se a necessidade de se promover uma transformação que estabilize uma possível variabilidade crescente, aplicando-se o logaritmo natural a todos os valores da série selecionada. Observou-se a ocorrência de "outliers", tratando de removê-los adequadamente.

A detecção e a remoção de "outliers" na modelagem de séries temporais é muito bem abordada nos trabalhos de Fox (1972), Hillmer, Bell & Tiao (1983), Tsay (1986) e Chang (1988). Outras discussões importantes sobre o tema podem ser encontradas nos livros de Box & Jenkins (1994), Mills (1990), Venables & Ripley (1999), Montgomery (1990), Pankratz (1991) e Wei (1990), entre outros. Adotaremos a abordagem de remoção destes "outliers" sugerida por Venables & Ripley (1999), que substitui os valores destoantes das séries pelas médias dos demais valores observados nos mesmos meses da ocorrência deles.

Adicionalmente, verificou-se, em cada série, se há evidências de ocorrência de alguma quebra estrutural, atentando para o fato de que a implantação do Plano Real em julho de 1994

tem demonstrado, em diversos trabalhos, que as variáveis econômicas apresentam uma mudança estrutural extremamente visível. Verificaremos se o mesmo se dá para as séries da Arrecadação Tributária.

A comprovação formal de quebras estruturais em séries temporais pode ser obtida através do Teste-F para mudança estrutural, na forma apresentada por Enders (1996). Como a mudança estrutural pode ser facilmente observada através dos gráficos das séries estudadas, não faremos para todas elas o teste estatístico supramencionado, mas a mudança estrutural ocorrida com a edição do Plano Real será comprovada estatisticamente para a série da Receita Total Administrada pela SRF, na subseção 3.2.9.

Em virtude da ocorrência da mudança estrutural observada nas séries estudadas, utilizaremos somente os dados referentes ao último período sem mudanças para a identificação e estimação do modelo, desde que se tenha um total de observações suficientes para a utilização da modelagem ARIMA, admitindo que haja sazonalidade na série. As previsões serão feitas para os meses de janeiro a dezembro de 2001, ou seja, doze passos à frente.

### **3.2.1 - Imposto sobre as Importações - II**

A série do Imposto sobre as Importações, II, está apresentada no gráfico da figura 3.1. Uma análise desse gráfico evidencia que há um pequeno aumento da variância à medida que o valor da série aumenta. A transformação da série através da aplicação do logaritmo natural a todos os seus valores mostra uma nova série, LII\*, que possui uma variância aproximadamente constante, à exceção de "outliers" que serão corrigidos oportunamente, conforme se observa na fig. 3.2.

Através da simples observação da série LII\*, constata-se que há dois pontos extremamente destoantes do conjunto restante das observações, indicando possíveis "outliers" que, se não removidos, podem comprometer a estimação do modelo e, conseqüentemente, sua capacidade preditiva.

Os "outliers" ali observados, maio e junho de 1995, podem ser explicados através do seguinte fato: houve, naquele ano, uma elevação de 62%, em relação a 1994, no volume das importações tributadas, principalmente de veículos, com acentuada repercussão na arrecadação do Imposto sobre Importações nos meses supracitados. Procedendo às devidas correções, a série corrigida, LII, apresenta-se grafada na fig. 3.3.

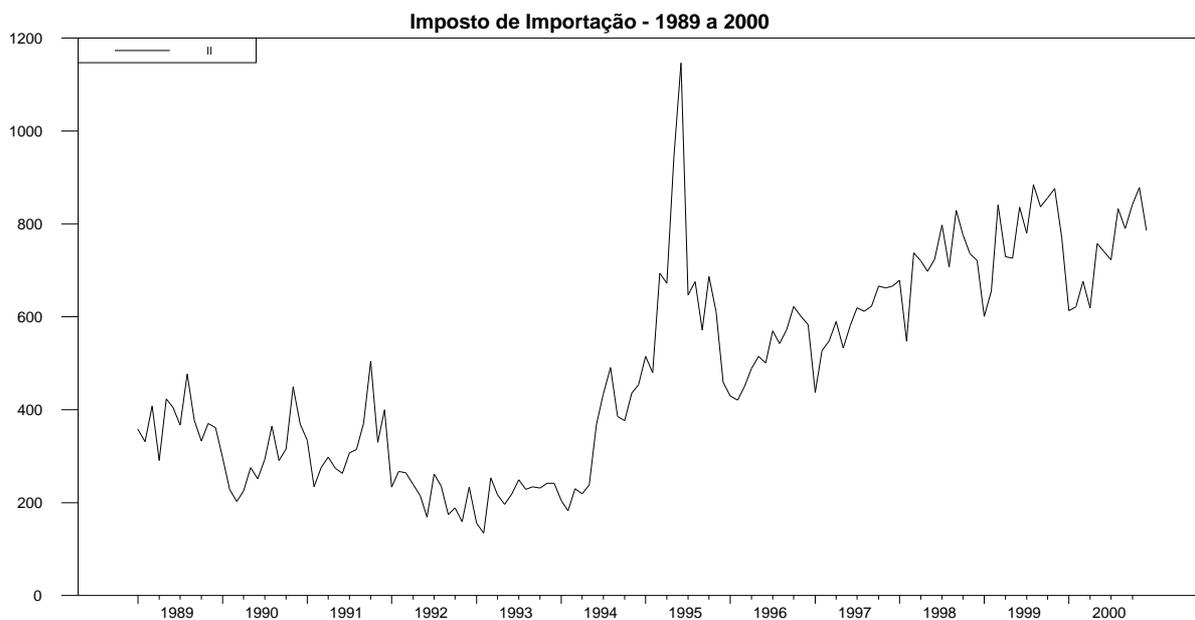


Figura 3.1 – Imposto sobre as Importações - em milhões de R\$ - II - jan/1989 a dez/2000

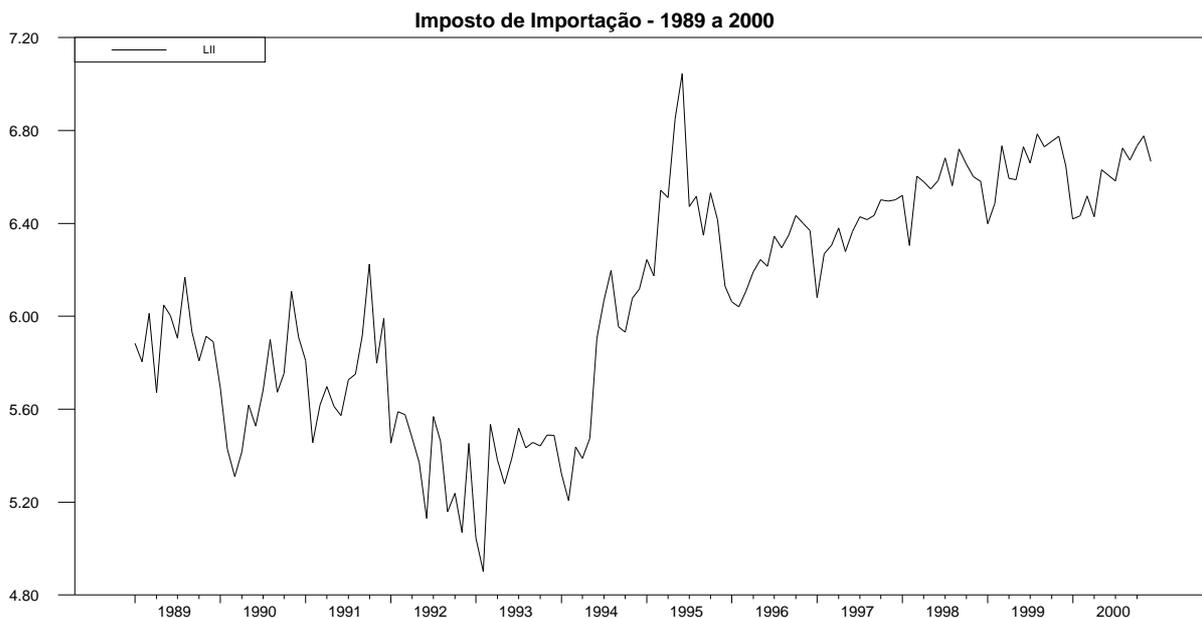


Figura 3.2 – Série Ln do Imposto sobre as Importações - LII\* - jan/1989 a dez/2000

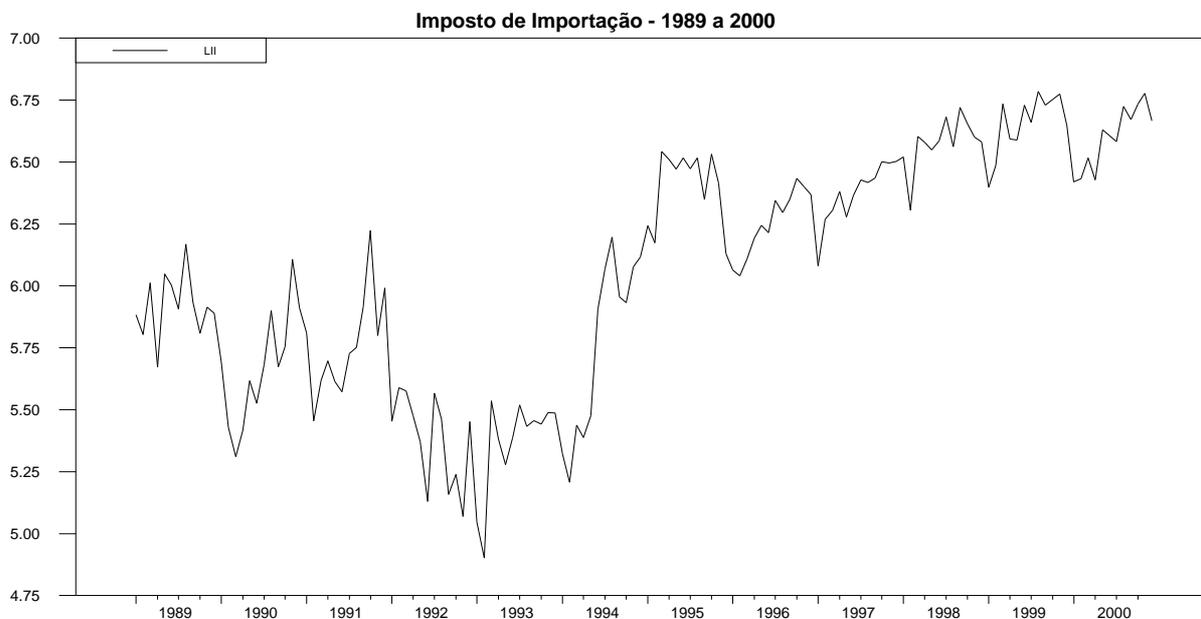


Figura 3.3 – Série ln do Imposto sobre as Importações Corrigida - LII - jan/1989 a dez/2000

Os dados apresentados na fig. 3.3 mostram nitidamente uma mudança estrutural em julho de 1994, mês de implantação do Plano Real. Algumas das explicações para este comportamento podem ser vistas na seção 3.1 deste capítulo, com a análise dos eventos importantes que afetaram a arrecadação do período em estudo.

Em virtude da ocorrência da mudança estrutural observada na série do Imposto sobre as Importações, utilizaremos somente os dados referentes ao segundo período (jul/94 a dez/2000) para a identificação e estimação do modelo, perfazendo um total de 78 observações, suficientes para a utilização da modelagem ARIMA, mesmo que haja sazonalidade na série. O gráfico da série LII para o período considerado está, agora, apresentado na figura 3.4.

### 3.2.2 - Imposto sobre Produtos Industrializados - IPI

A série do Imposto sobre Produtos Industrializados - IPI está apresentada no gráfico da figura 3.5. A transformação da série através da aplicação do logaritmo natural a todos os seus valores mostra uma nova série, LIPI\*, que possui uma variância aproximadamente constante, à exceção de "outliers" que serão corrigidos oportunamente, conforme se observa na fig. 3.6.

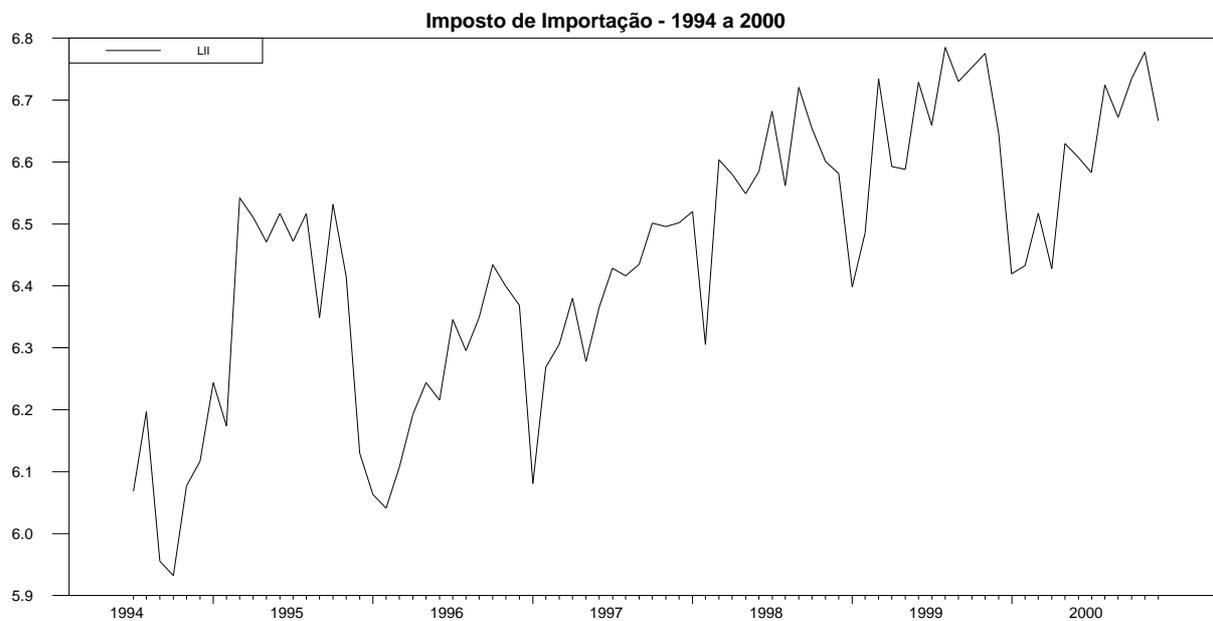


Figura 3.4 – Série In do Imposto sobre as Importação Corrigida - LII - jul/94 a dez/2000



Figura 3.5 – Imposto sobre Produtos Industrializados - IPI - jan/1989 a dez/2000



Figura 3.6 – Série In do Imposto sobre Produtos Industrializados - LIPI\* - jan/1989 a dez/2000

Através da simples observação da série LIPI\*, constata-se que há diversos pontos que destoam do conjunto restante das observações, indicando possíveis "outliers". Procedendo às devidas correções, a série corrigida, LII, apresenta-se grafada na fig. 3.7.

Os dados apresentados na fig. 3.7 mostram nitidamente duas mudanças estruturais, uma ocorrida em janeiro de 1992 e a outra, em julho de 1994, mês de implantação do Plano Real. Algumas das explicações para este comportamento pode ser vista na seção 3.1 deste capítulo.

Em virtude da ocorrência das mudanças estruturais observadas na série do Imposto sobre Produtos Industrializados, utilizaremos somente os dados referentes ao segundo período (jul/94 a dez/2000) para a identificação e estimação do modelo, perfazendo um total de 78 observações, suficientes para a utilização da modelagem ARIMA, mesmo que haja alguma sazonalidade na série. O gráfico da série LIPI para o período considerado está, agora, apresentado na figura 3.8.



Figura 3.7 – Série ln do Imposto sobre Produtos Industrializados Corrigida - LII - jan/1989 a dez/2000



Figura 3.8 – Série ln do Imposto sobre Produtos Industrializados Corrigida - LIPI - jul/94 a dez/2000

### 3.2.3 - Imposto sobre a Renda (Total) - IR

A série do Imposto sobre a Renda (Total) - IR está apresentada no gráfico da figura 3.9. O Imposto sobre a Renda (total) representa o somatório de três componentes: o Imposto sobre a Renda de Pessoas Físicas - IRPF, o Imposto sobre a Renda de Pessoas Jurídicas - IRPJ e o Imposto sobre a Renda Retido na Fonte - IRRF. A transformação da série através da aplicação do logaritmo natural a todos os seus valores mostra uma nova série, LIR\*, que possui uma variância aproximadamente constante, à exceção de "outliers" que serão corrigidos oportunamente, conforme se observa na fig. 3.10.

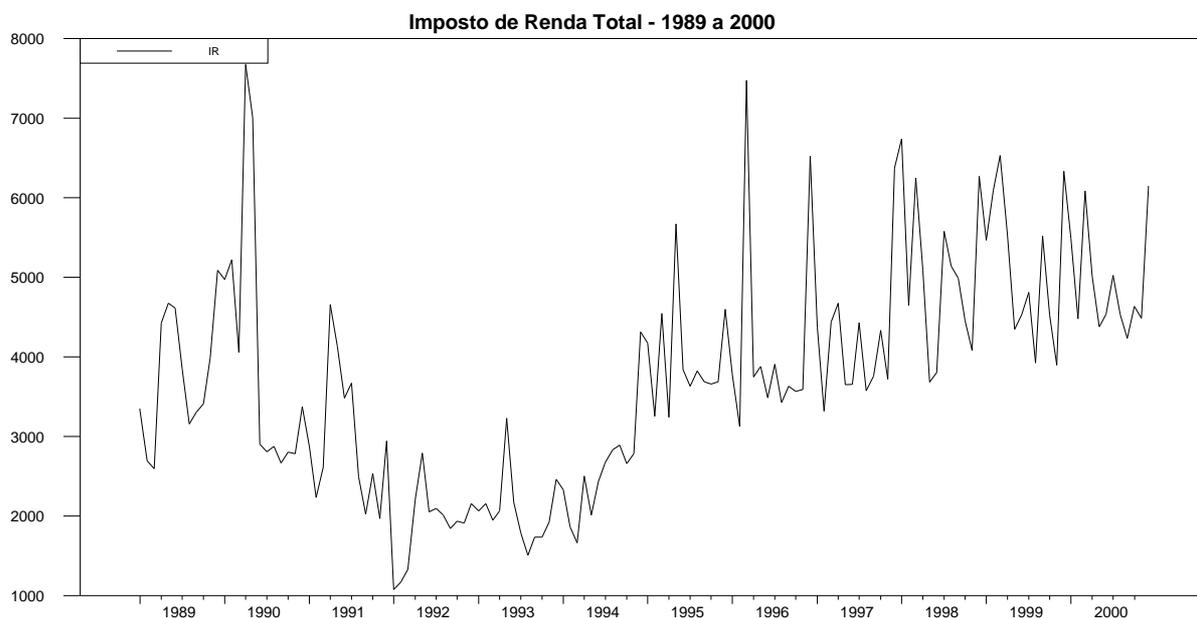


Figura 3.9 – Imposto sobre a Renda (Total) - IR - jan/1989 a dez/2000

Através da simples observação da série LIR\*, constata-se que há alguns poucos pontos que destoam do conjunto restante das observações, indicando possíveis "outliers". Procedendo às devidas correções, a série corrigida, LIR, apresenta-se grafada na fig. 3.11.

Os dados apresentados na fig. 3.11 mostram nitidamente uma mudança estrutural em julho de 1994, mês de implantação do Plano Real.

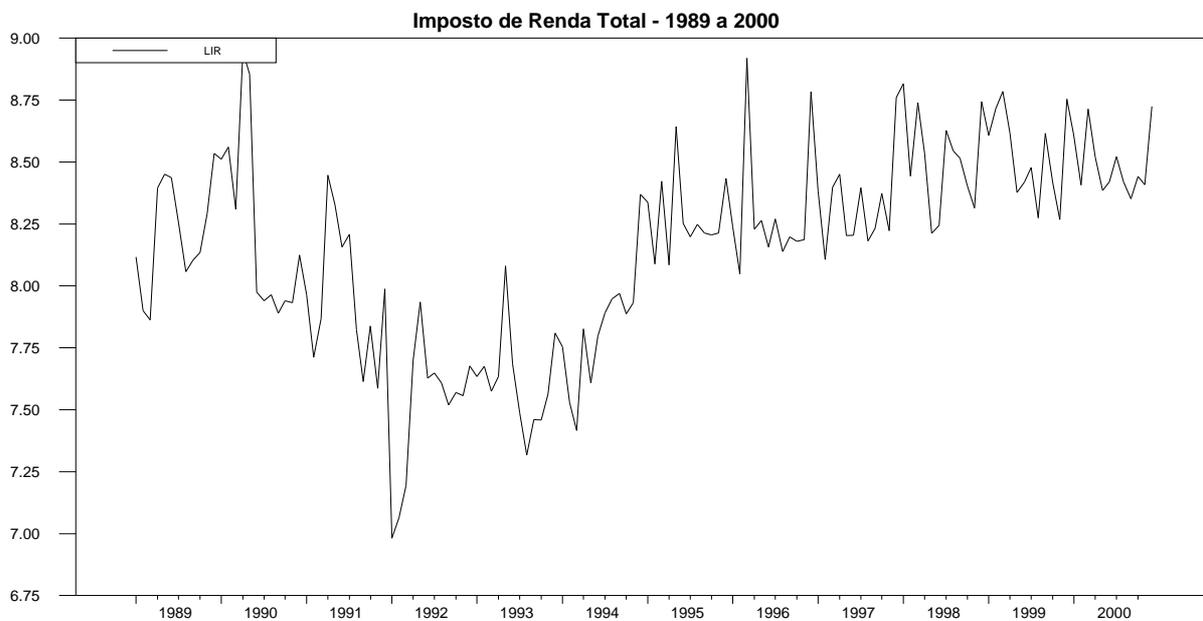


Figura 3.10 – Série ln do Imposto sobre a Renda (Total) - LIR\* - jan/1989 a dez/2000

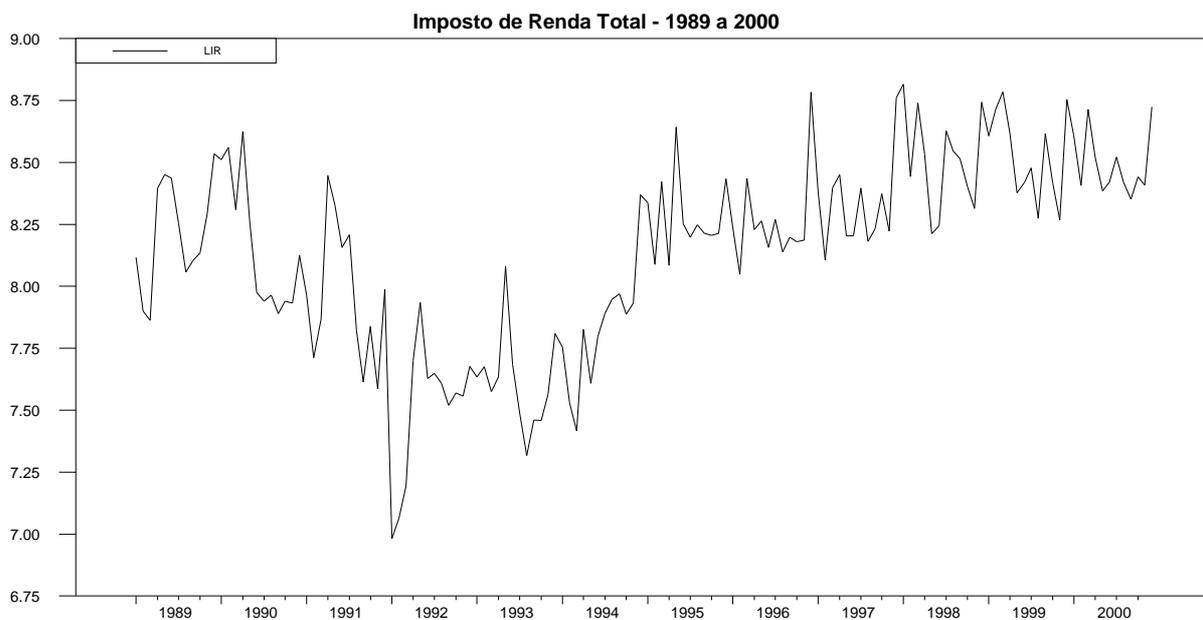


Figura 3.11 – Série ln do Imposto sobre a Renda (Total) Corrigida- LIR - jan/1989 a dez/2000

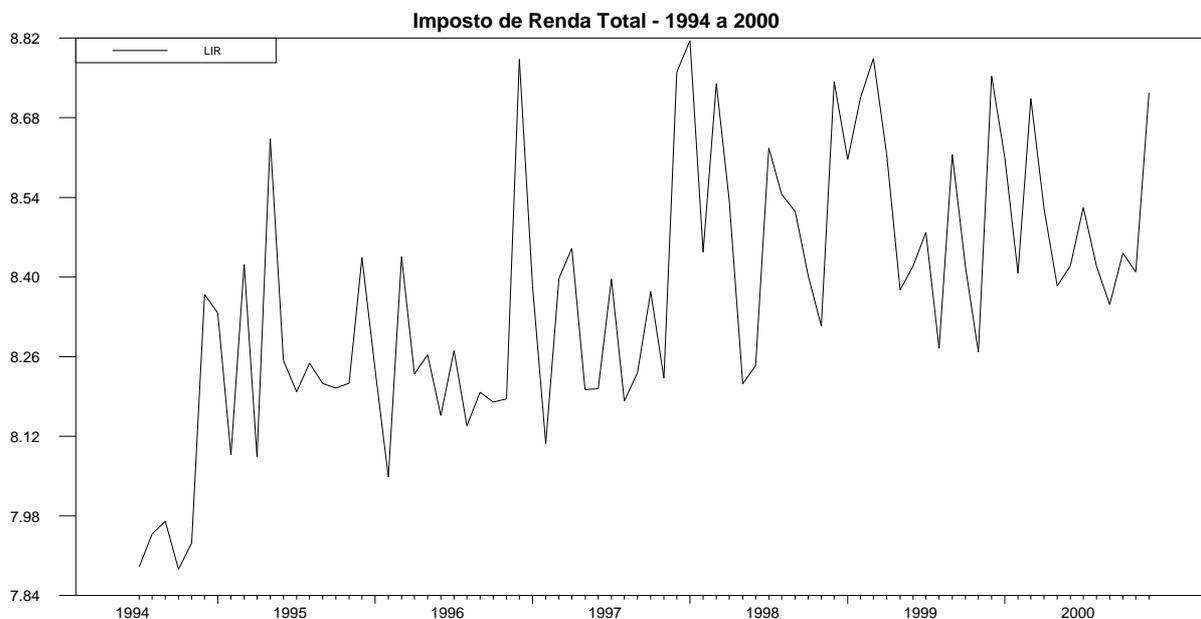


Figura 3.12 – Série In do Imposto sobre a Renda (Total) Corrigida - LIR - jul/94 a dez/2000

### 3.2.4 - Outros Impostos - OI

Os demais impostos federais (ITR, IOF, IE, IPMF - alguns meses), por representarem individualmente parcelas pouco significativas da arrecadação total, foram consolidados numa única série, aqui denominada de Outros Impostos - OI. Tal série está apresentada no gráfico da figura 3.13. Uma análise do gráfico desta série agregada evidencia que não há variabilidade significativa da variância à medida que o valor da série aumenta, indicando, a princípio, que não há necessidade de se promover qualquer transformação para estabilização da variância.

Através da simples observação da série OI, constata-se que há alguns poucos pontos que destoam do conjunto restante das observações, indicando possíveis "outliers" que, se não removidos, podem comprometer a estimação do modelo e, conseqüentemente, sua capacidade preditiva.

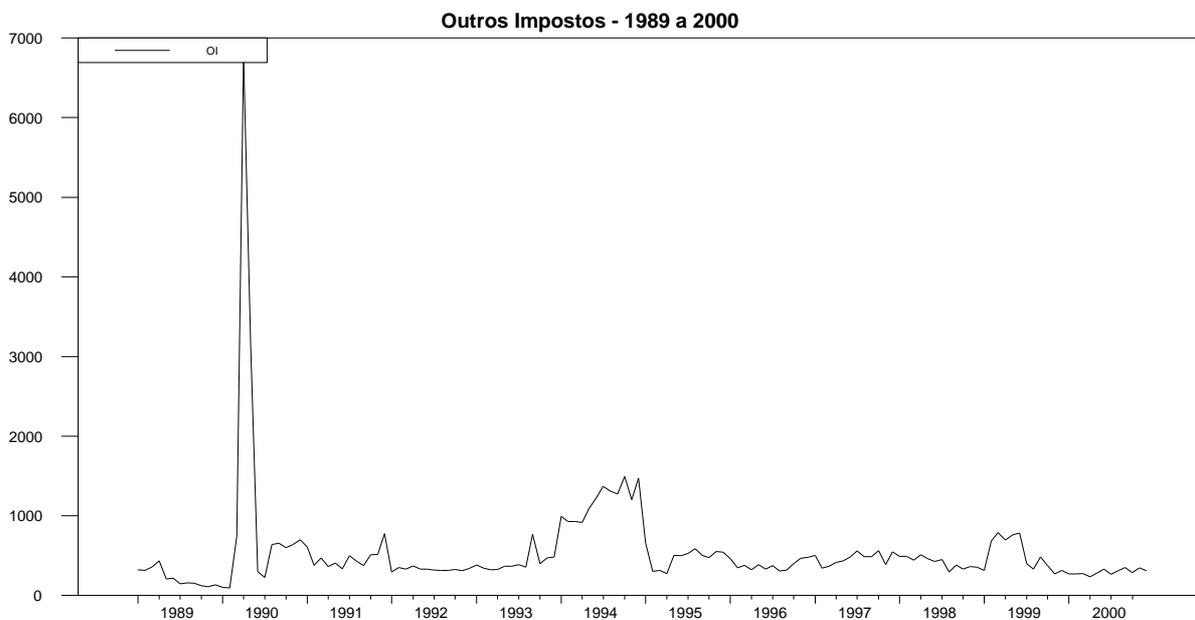


Figura 3.13 – Outros Impostos - OI - jan/1989 a dez/2000

Os dados apresentados na fig. 3.14 mostram nitidamente uma mudança estrutural em 1994, devido ao retorno da cobrança do IPMF entre os meses de janeiro de 1994 e janeiro de 1995.

Em virtude da ocorrência da mudança estrutural observada na série aqui estudada, utilizaremos somente os dados referentes ao período fev/95 a dez/2000 para a identificação e estimação do modelo, perfazendo um total de 71 observações, suficientes para a utilização da modelagem ARIMA, mesmo que haja alguma sazonalidade na série. O gráfico da série OI para o período considerado está, agora, apresentado na figura 3.15.

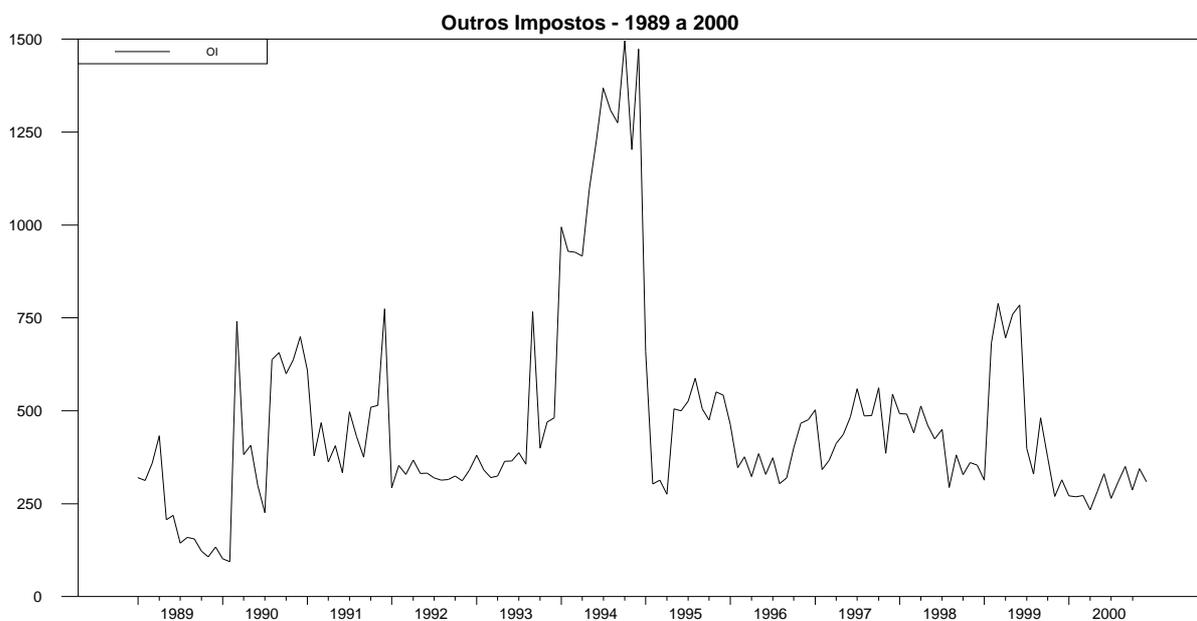


Figura 3.14 – Série Outros Impostos Corrigida- OI\* - jan/1989 a dez/2000

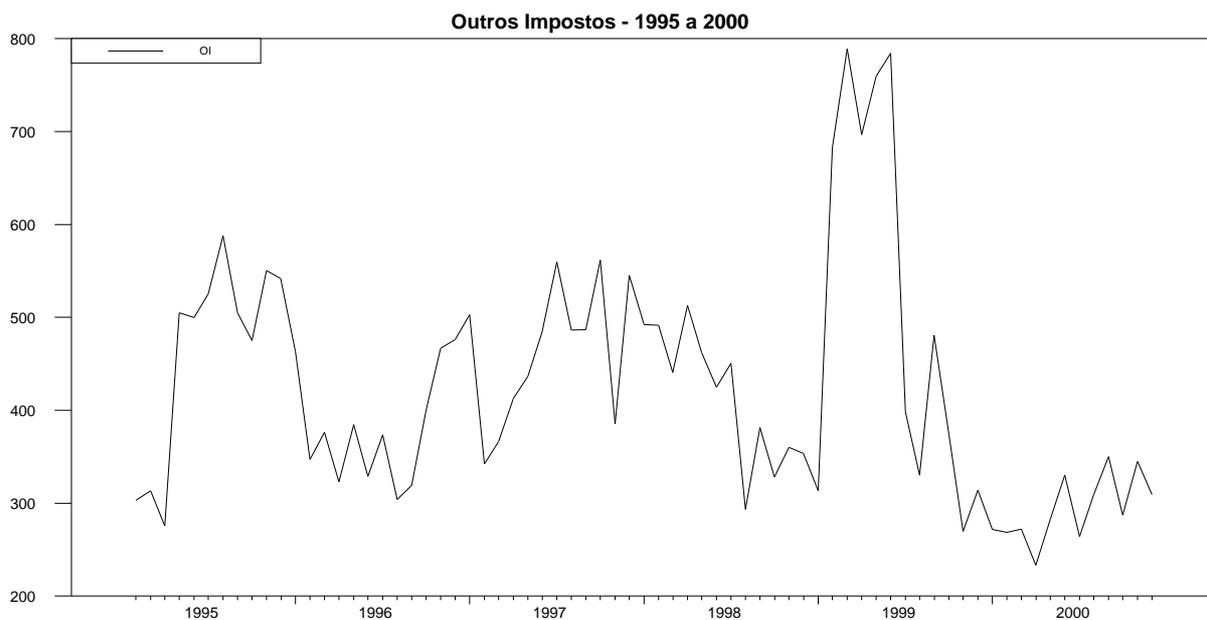


Figura 3.15 – Série Outros Impostos Corrigida - OI - fev/95 a dez/2000

### 3.2.5 - Receita de Impostos - RIMP

Esta série representa o somatório de todas as séries referentes aos impostos federais, ou seja, o somatório das quatro séries estudadas até o presente momento. Tal série está

apresentada no gráfico da figura 3.16. O gráfico da série logaritmizada, LRIMP\*, está apresentado na figura 3.17.

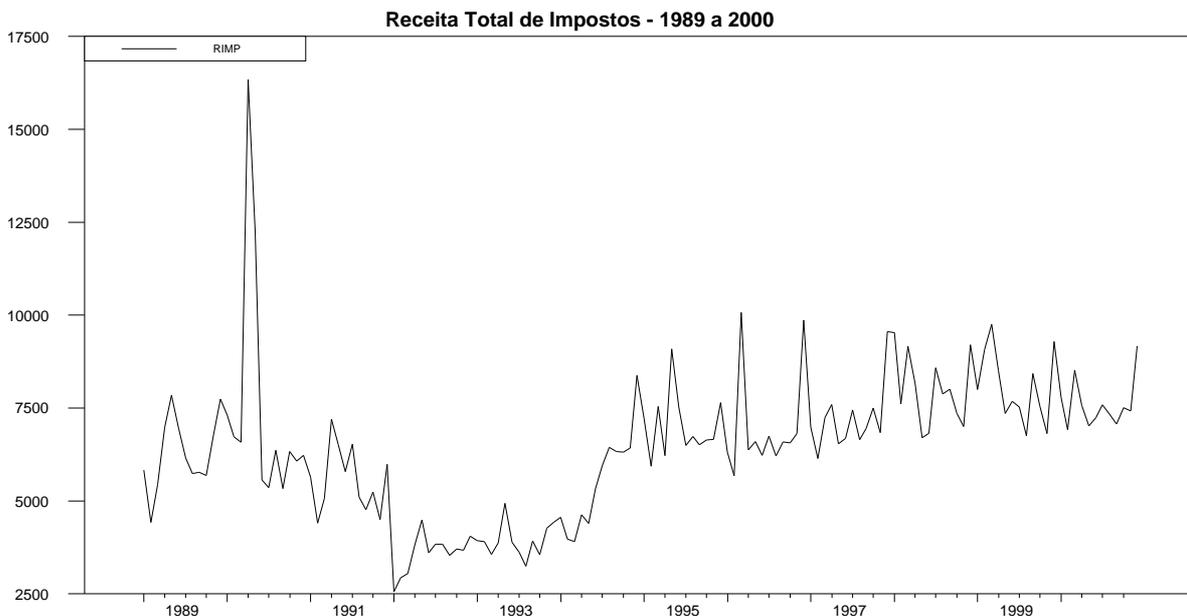


Figura 3.16 – Receita de Impostos - RIMP - jan/1989 a dez/2000

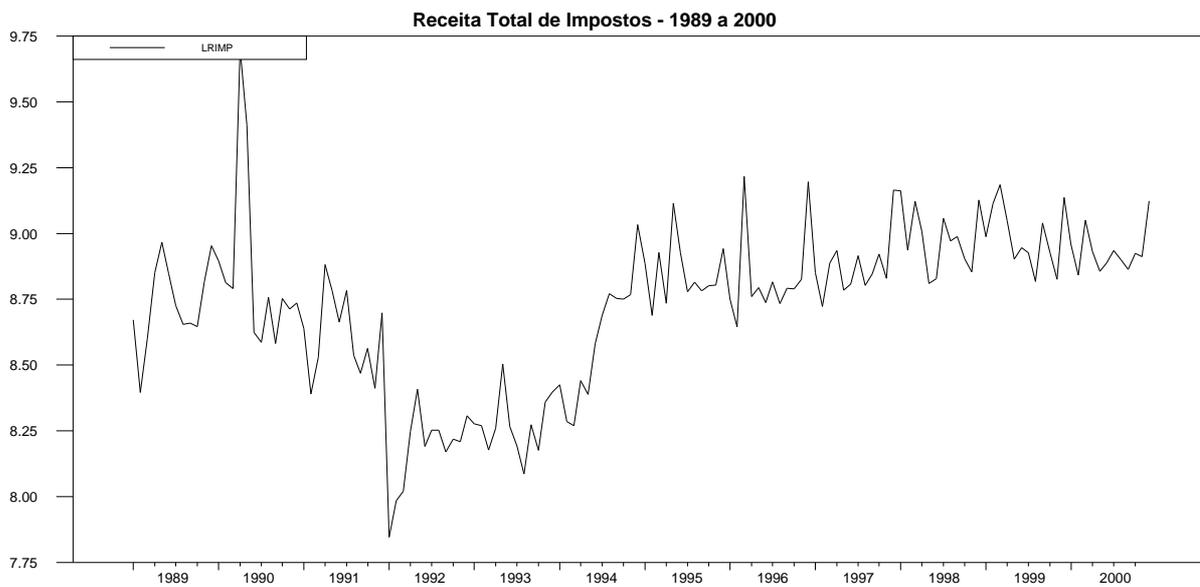


Figura 3.17 – ln da Receita de Impostos - LRIMP\* - jan/1989 a dez/2000

Através da simples observação da série LRIMP\*, constata-se que há alguns poucos pontos que destoam do conjunto restante das observações, indicando possíveis "outliers".

Procedendo às devidas correções, a série corrigida, LRIMP, apresenta-se assim grafada (fig. 3.18).

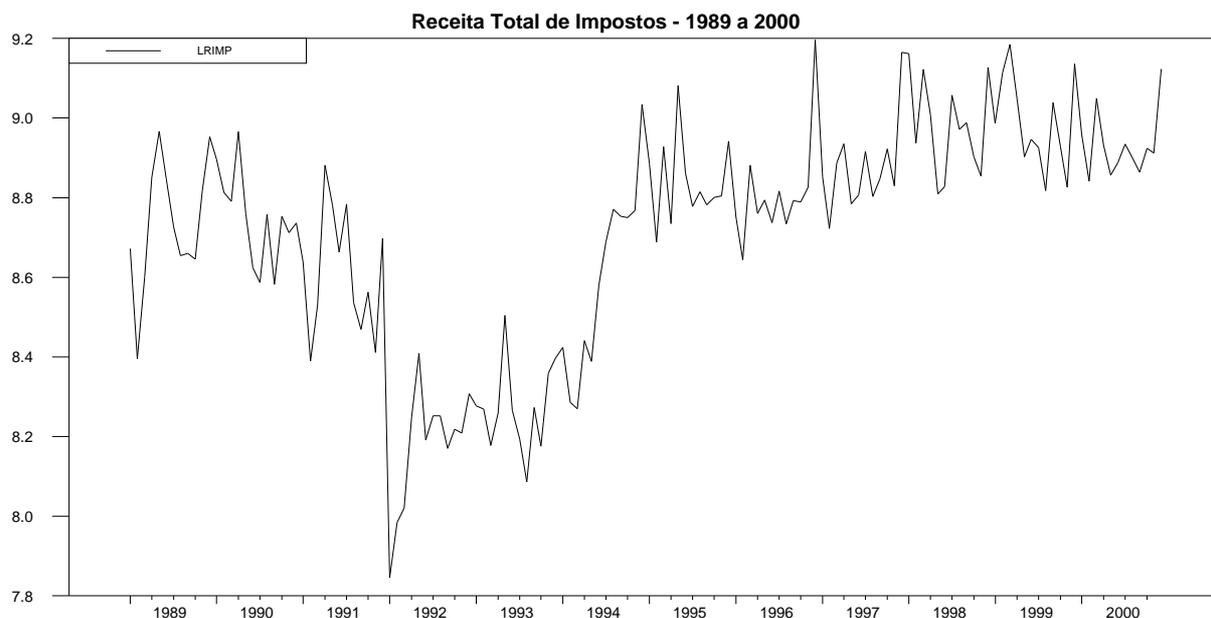


Figura 3.18 – Série ln da Receita de Impostos Corrigida- LRIMP - jan/1989 a dez/2000

Os dados apresentados na fig. 3.18 mostram nitidamente uma mudança estrutural em julho de 1994, mês de implantação do Plano Real. Em virtude da ocorrência da mudança estrutural observada na série em estudo, utilizaremos somente os dados referentes ao segundo período (jul/94 a dez/2000) para a identificação e estimação do modelo, perfazendo um total de 78 observações, suficientes para a utilização da modelagem ARIMA, mesmo que haja alguma sazonalidade na série. O gráfico da série LRIMP para o período considerado está, agora, apresentado na figura 3.19.

### 3.2.6 - Contribuição Provisória sobre Movimentação Financeira - CPMF

Como a CPMF só foi criada em 23/01/1997, apresentaremos o gráfico da série a partir do mês da sua criação. Tal série está apresentada, portanto, no gráfico da figura 3.20. O gráfico da série logaritmizada, LCPMF\*, está apresentado na figura 3.21.

Através da simples observação da série LCPMF\*, constata-se que há quatro pontos que destoam do conjunto restante das observações, indicando possíveis "outliers". Tais "outliers" se referem ao período entre a extinção e a nova entrada em vigor da CPMF, qual seja, fevereiro a maio de 1999.

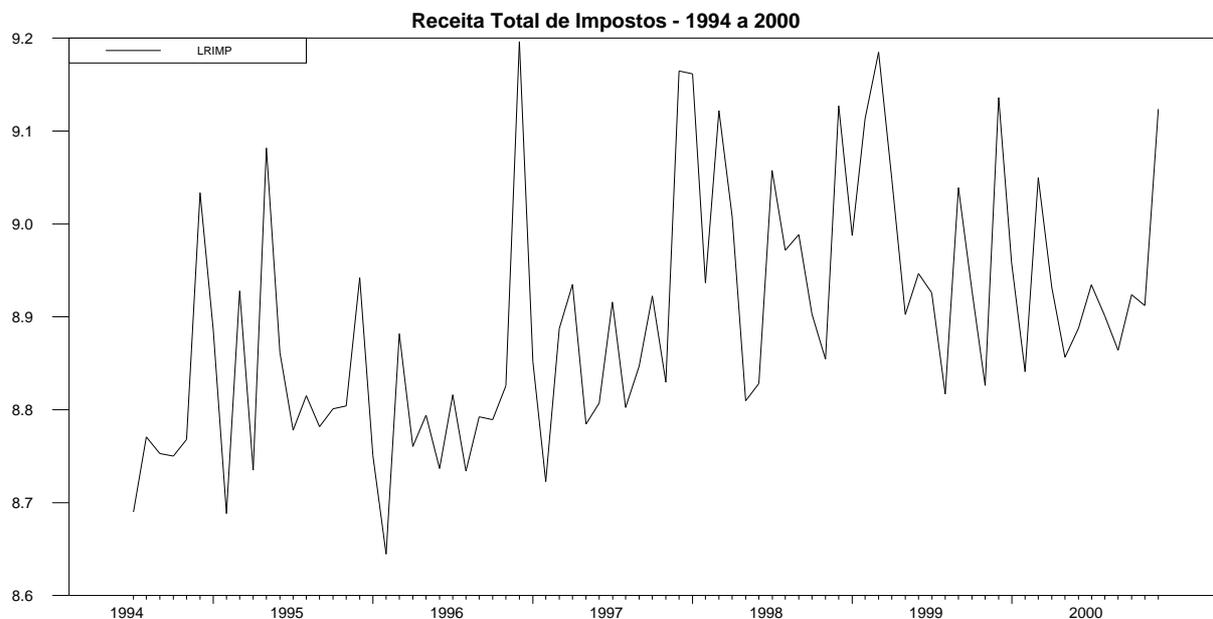


Figura 3.19 – Série In da Receita de Impostos Corrigida- LRIMP - jul/1994 a dez/2000

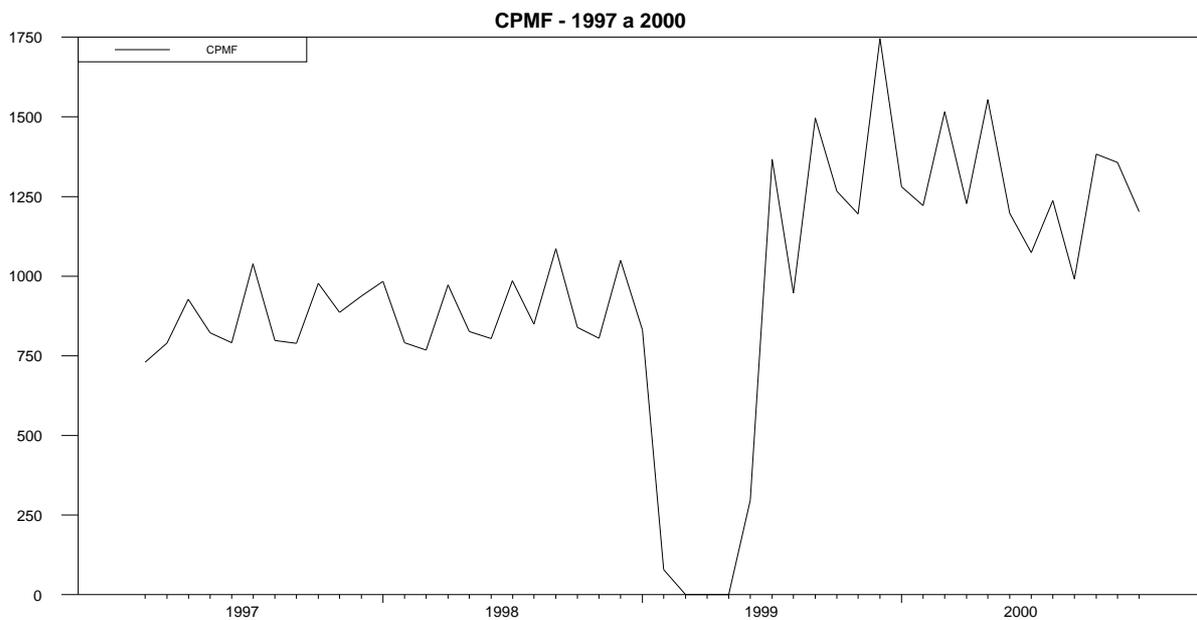


Figura 3.20 – Contribuição Provisória sobre Movimentação Financeira - CPMF - fev/1997 a dez/2000

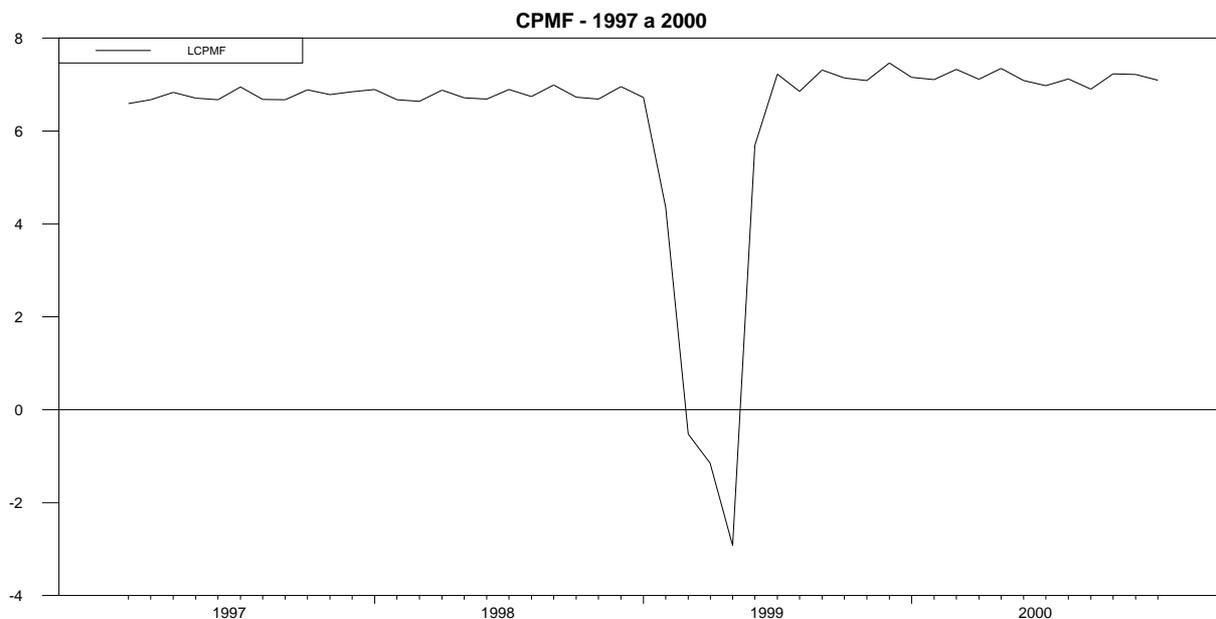


Figura 3.21 – ln da Contribuição Provisória sobre Movimentação Financeira - LCPMF\* -fev/1997 a dez/2000

Os dados apresentados na fig. 3.22 não mostram nitidamente qualquer mudança estrutural relevante. Como o período analisado já é no Pós-Real e não houve nenhuma mudança estrutural na economia brasileira neste período, utilizaremos somente os dados referentes aos meses de fev/1997 a dez/2000 para a identificação e estimação do modelo, perfazendo um total de 47 observações, consideradas insuficientes para a utilização da modelagem ARIMA, quando há alguma sazonalidade na série. Apesar de se ter poucas observações para a série da CPMF, ajustaremos o modelo de Box-Jenkins para estes dados e compararemos as previsões calculadas com aquelas feitas pela SRF.

### 3.2.7 - Contribuição para Financiamento da Seguridade Social - COFINS e Contribuição para o PIS/PASEP

Como as contribuições sociais COFINS e PIS/PASEP possuem, regra geral, a mesma base de cálculo (faturamento bruto das empresas), suas séries apresentaram o mesmo comportamento temporal, podendo ser agregadas no estudo de previsão, sem perda de generalidade. Constrói-se, portanto, a série agregada COFPIS, que apresenta a soma dessas duas contribuições e está apresentada no gráfico da figura 3.23. A transformação da série através da aplicação do logaritmo natural a todos os seus valores mostra uma nova série, LCOFPIS\*, que possui uma variância aproximadamente constante, à exceção de "outliers" que serão corrigidos oportunamente, conforme se observa na fig. 3.24.

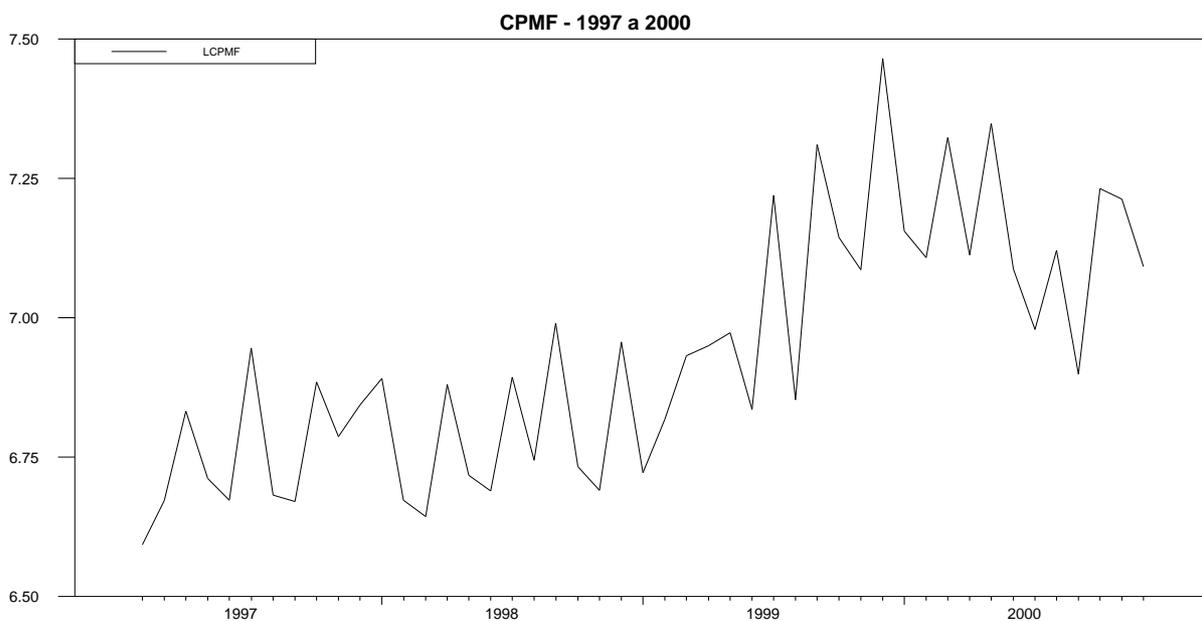


Figura 3.22 – Série ln da CPMF - LCPMF -fev/1997 a dez/2000

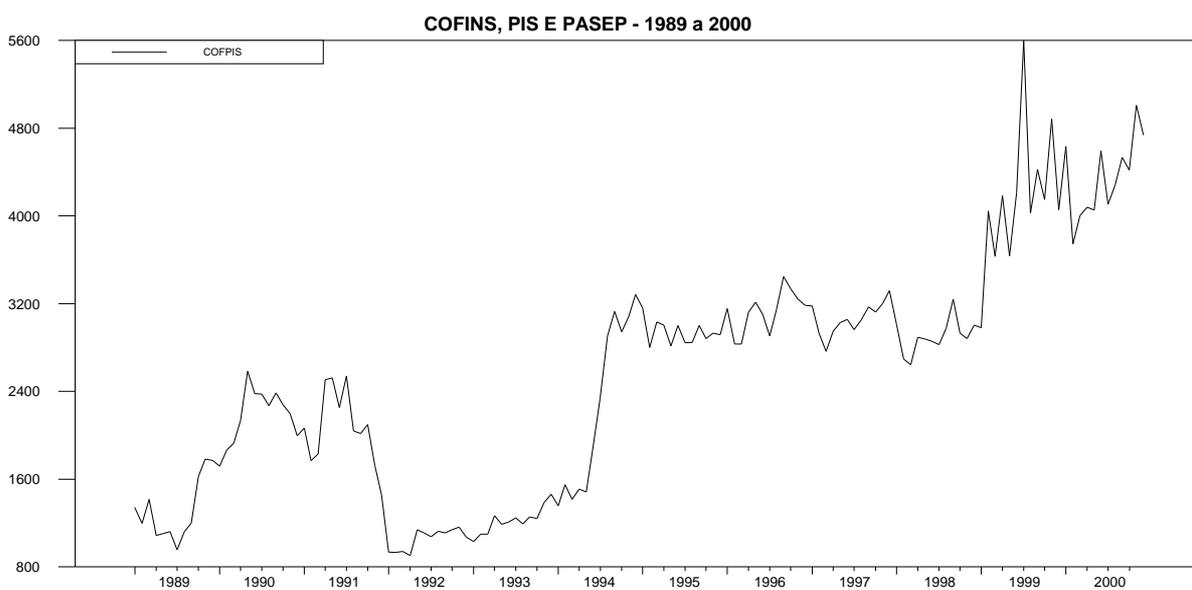


Figura 3.24 – ln da Contribuição para Financiamento da Seguridade Social e da Contribuição para o PIS/PASEP - LCOFPIS\* - jan/1989 a dez/2000

Através da simples observação da série LCOFPIS\*, constata-se que há diversos pontos que destoam do conjunto restante das observações, indicando possíveis "outliers". Procedendo às devidas correções, a série corrigida, LCOFPIS, apresenta-se assim grafada (fig. 3.25):

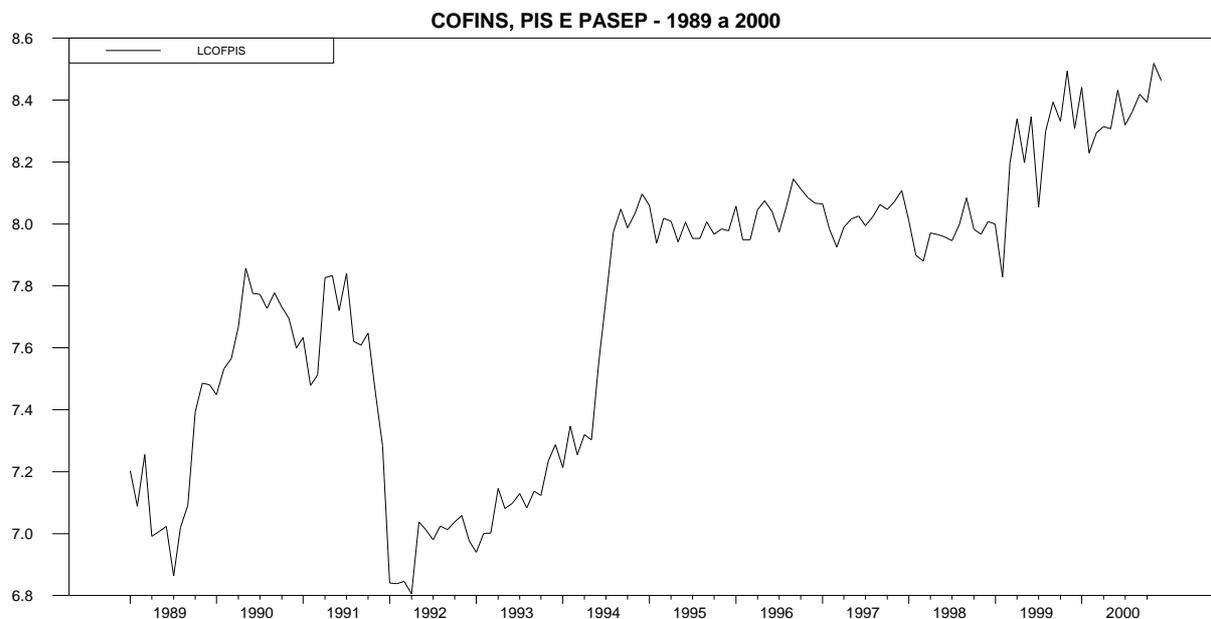


Figura 3.25 – Série Corrigida do ln da Contribuição para Financiamento da Seguridade Social e da Contribuição para o PIS/PASEP - LCOFPIS - jan/1989 a dez/2000

Os dados apresentados na fig. 3.25 mostram nitidamente diversas mudanças estruturais: uma ocorrida em janeiro de 1992, a outra, em julho de 1994, mês de implantação do Plano Real e mais uma em março de 1999.

Em virtude da ocorrência das diversas mudanças estruturais observadas na série COFPIS, não faremos previsões para tal série, já que o período estudado teria que se limitar a março/1999 - dezembro/2001, perfazendo um total de apenas 22 observações, insuficientes para a utilização da modelagem ARIMA.

### 3.2.8 - Outras Contribuições - OCONTR

As demais contribuições federais administradas pela SRF (CSLL, FUNDAF, Seguridade dos Servidores) foram consolidados numa única série, aqui denominada de Outras Contribuições - OCONTR. Tal série está apresentada no gráfico da figura 3.26. A transformação da série através da aplicação do logaritmo natural a todos os seus valores mostra uma nova série, LOCONTR\*, que possui uma variância aproximadamente constante, à exceção de "outliers" que serão corrigidos oportunamente, conforme se observa na fig. 3.27.

Através da simples observação da série LOCONTR\*, constata-se que há alguns poucos pontos que destoam do conjunto restante das observações, indicando possíveis "outliers".

Procedendo às devidas correções, a série corrigida, LOCONTR, apresenta-se grafada na fig. 3.28.

Os dados apresentados na fig. 3.28 mostram nitidamente duas mudanças estruturais, uma ocorrida em janeiro de 1992 e a outra, em julho de 1994, mês de implantação do Plano Real.

Em virtude da ocorrência das mudanças estruturais observadas na série LOCONTR, utilizaremos somente os dados referentes ao último período (jul/94 a dez/2000) para a identificação e estimação do modelo, perfazendo um total de 78 observações, suficientes para a utilização da modelagem ARIMA, mesmo que haja alguma sazonalidade na série. O gráfico da série LOCONTR para o período considerado está, agora, apresentado na figura 3.29.

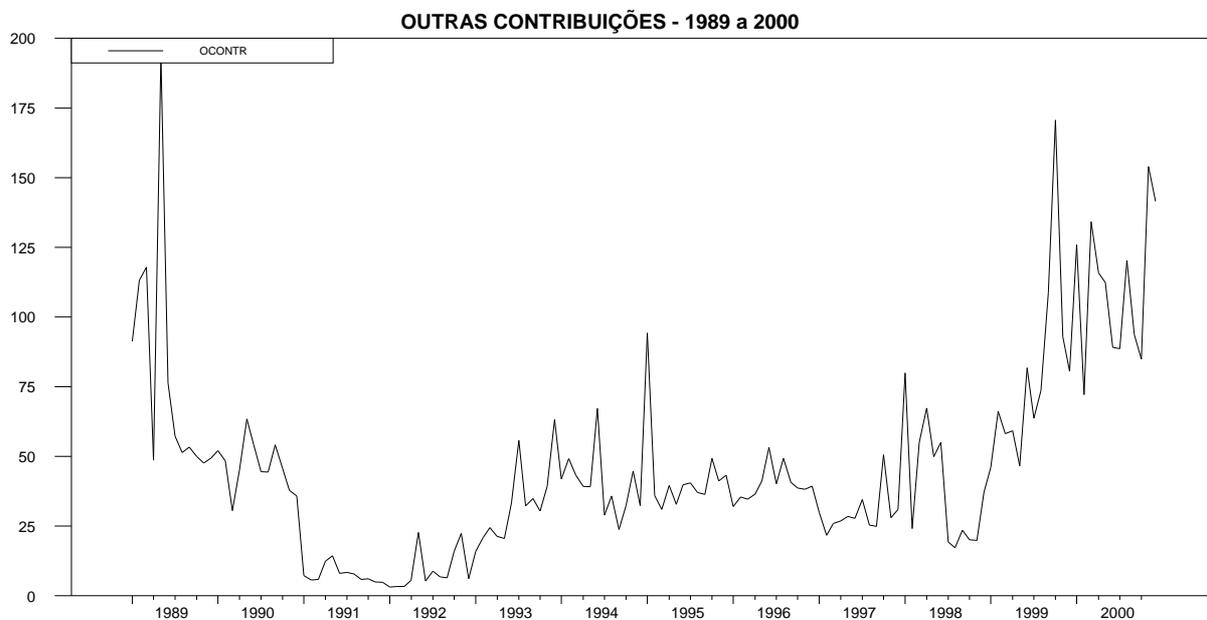


Figura 3.26 – Outras Contribuições - OCONTR - jan/1989 a dez/2000

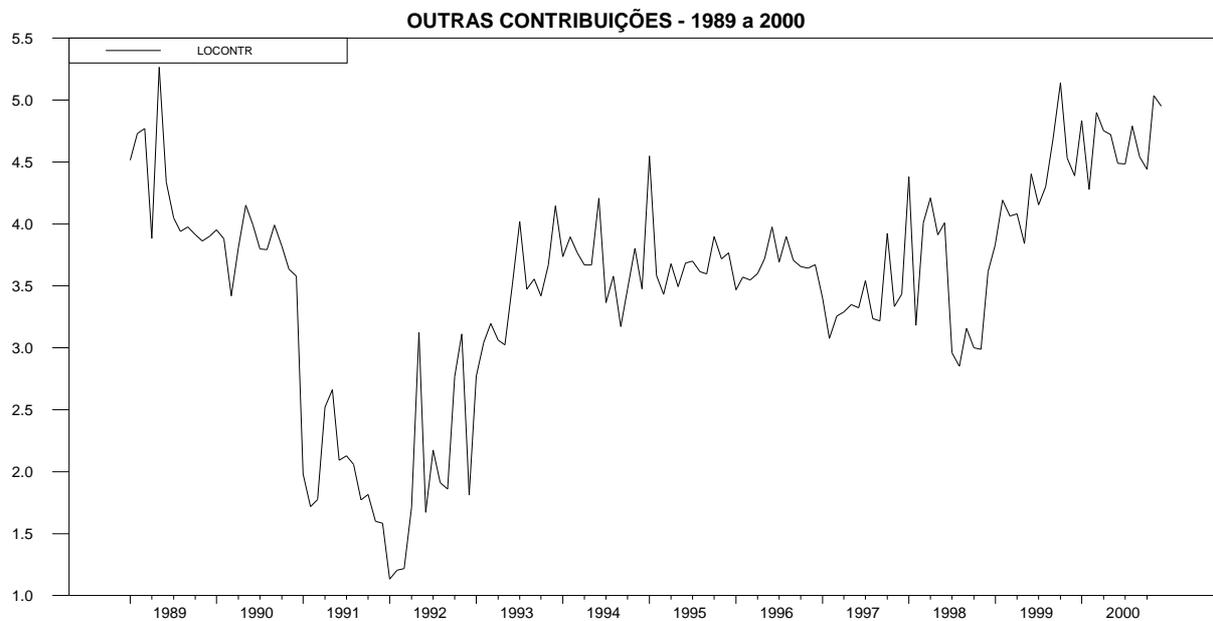


Figura 3.27 – ln da série Outras Contribuições - LOCONTR\* - jan/1989 a dez/2000

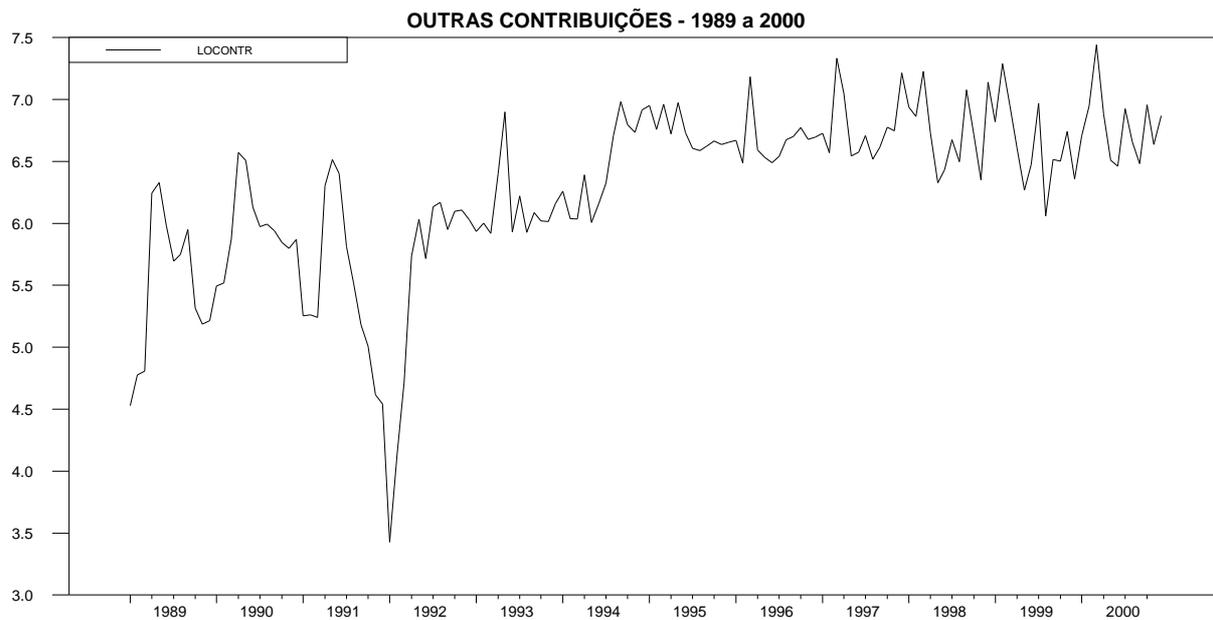


Figura 3.28 – ln da série Outras Contribuições Corrigida - LOCONTR - jan/1989 a dez/2000

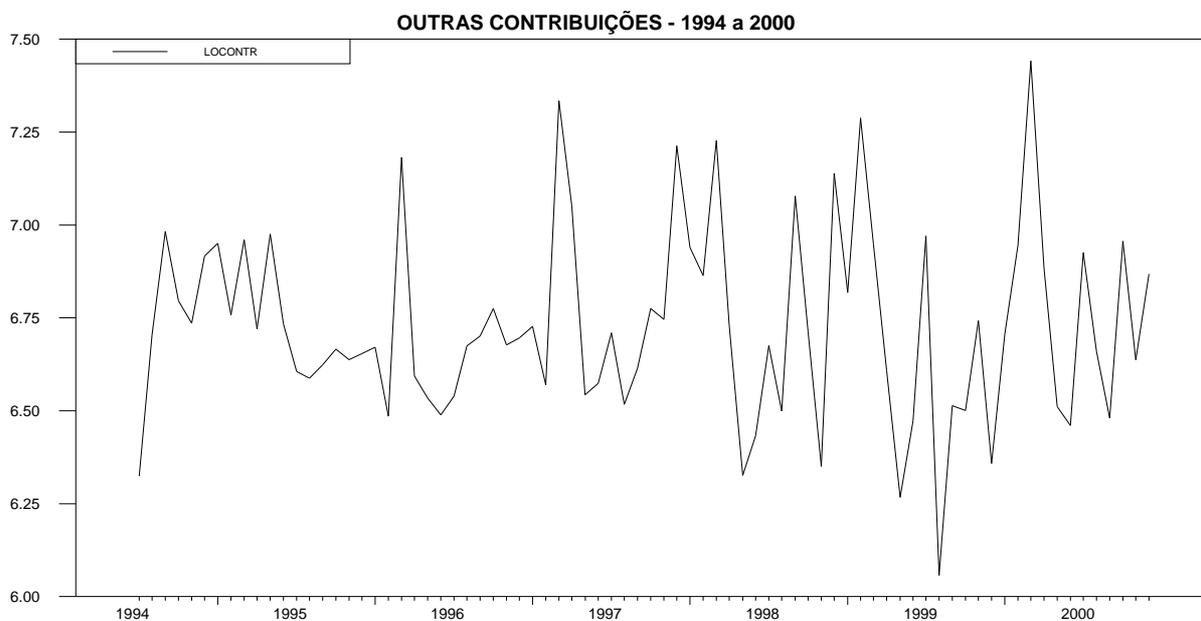


Figura 3.29 – ln da série Outras Contribuições Corrigida - LOCONTR - jul/1994 a dez/2000

### 3.2.9 - Receita Total de Contribuições - RCONT

Esta série representa o somatório de todas as séries referentes às contribuições federais, ou seja, o somatório das três séries de contribuições estudadas até o presente momento. Tal série está apresentada no gráfico da figura 3.30. O gráfico da série logaritimizada, LRCONT\*, está apresentado na figura 3.31.

Através da simples observação da série LRCONT\*, constata-se que há alguns poucos pontos que destoam do conjunto restante das observações, indicando possíveis "outliers". Procedendo às devidas correções, a série corrigida, LRIMP, apresenta-se grafada na fig. 3.32.

Os dados apresentados na fig. 3.32 mostram nitidamente uma mudança estrutural em julho de 1994, mês de implantação do Plano Real.

Em virtude da ocorrência da mudança estrutural observada na série em estudo, utilizaremos somente os dados referentes ao segundo período (jul/94 a dez/2000) para a identificação e estimação do modelo, perfazendo um total de 78 observações, suficientes para a utilização da modelagem ARIMA, mesmo que haja alguma sazonalidade na série. O gráfico da série LRIMP para o período considerado está, agora, apresentado na figura 3.33.

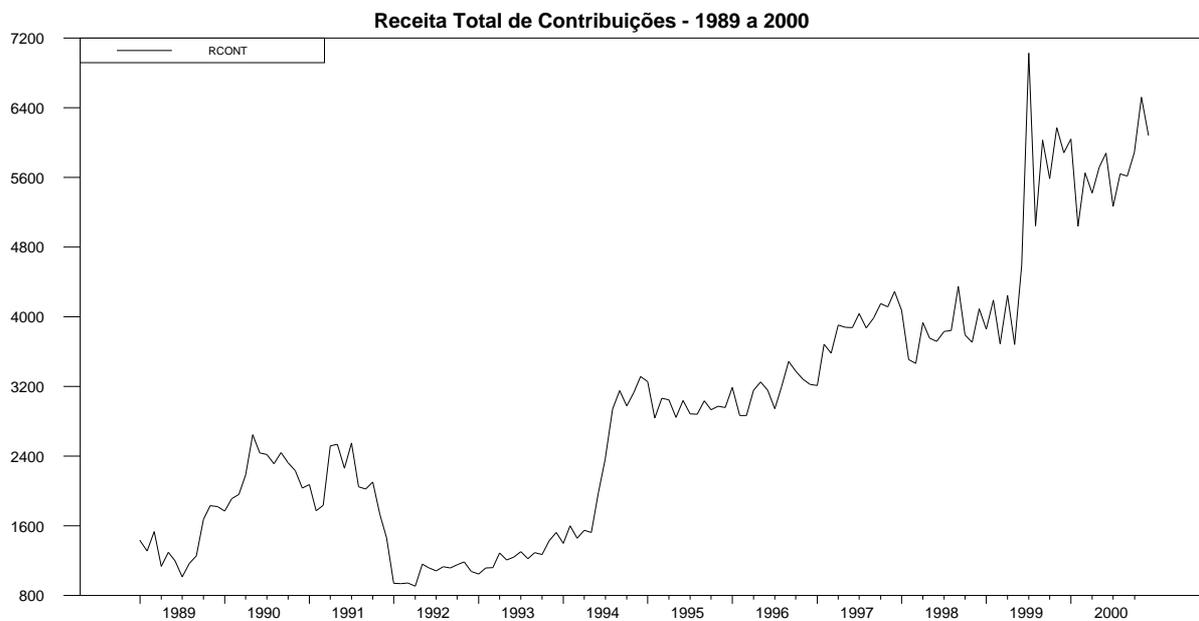


Figura 3.30 – Receita de Contribuições - RCONT - jan/1989 a dez/2000

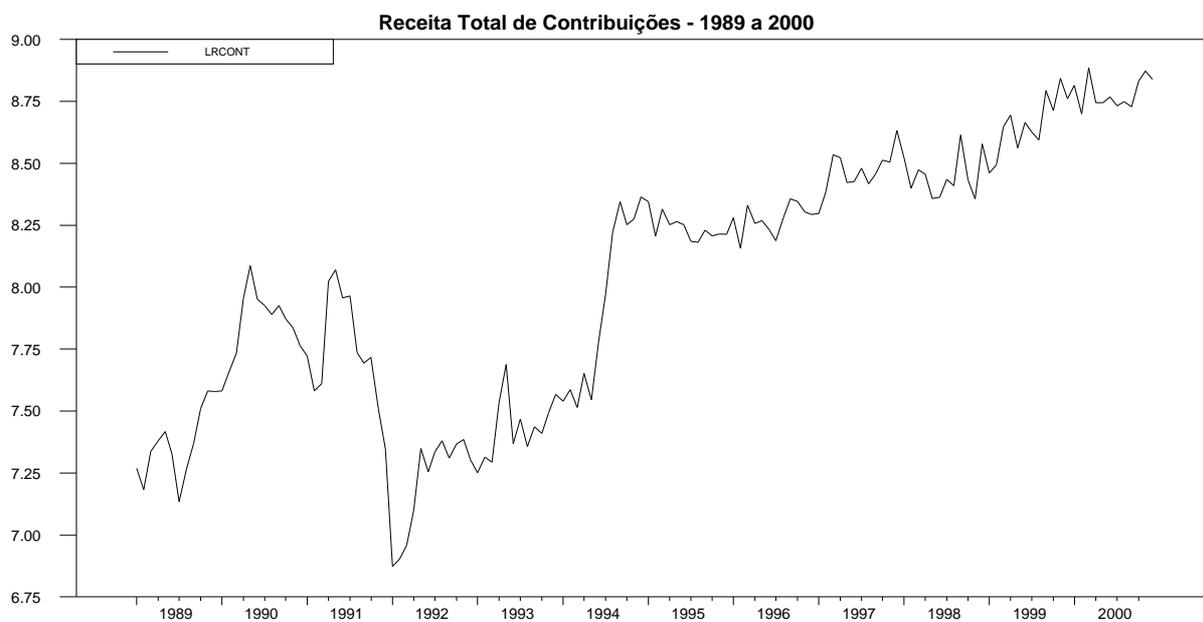


Figura 3.31 – ln da Receita de Contribuições - LRCONT\* - jan/1989 a dez/2000

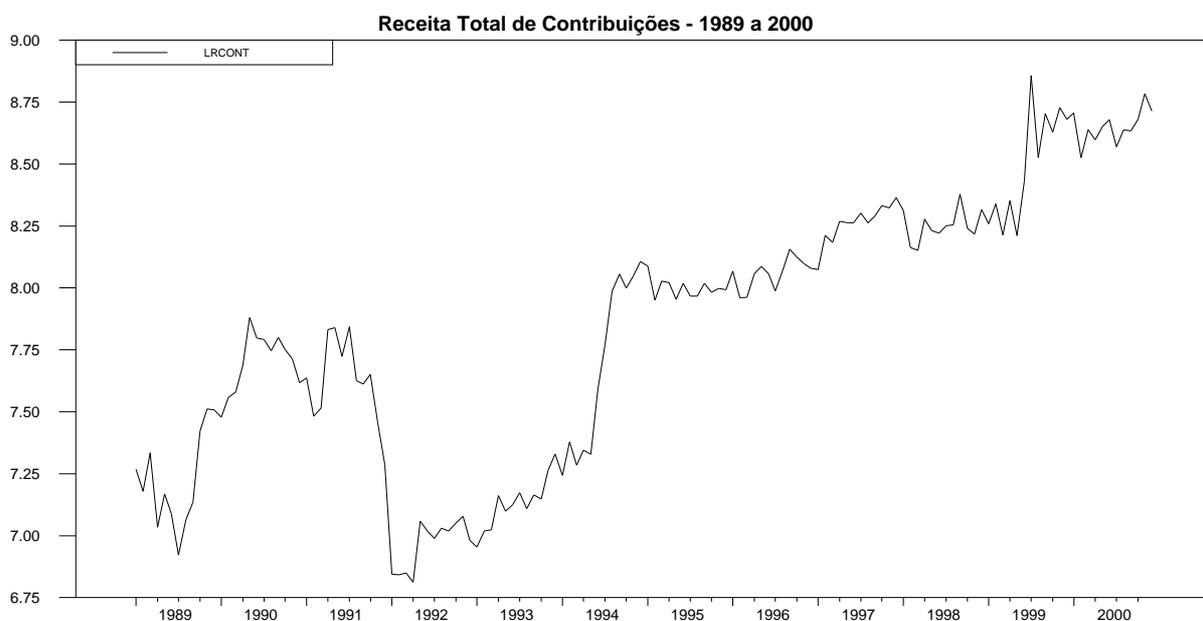


Figura 3.32 – Série In da Receita de Contribuições Corrigida- LRCONT - jan/1989 a dez/2000

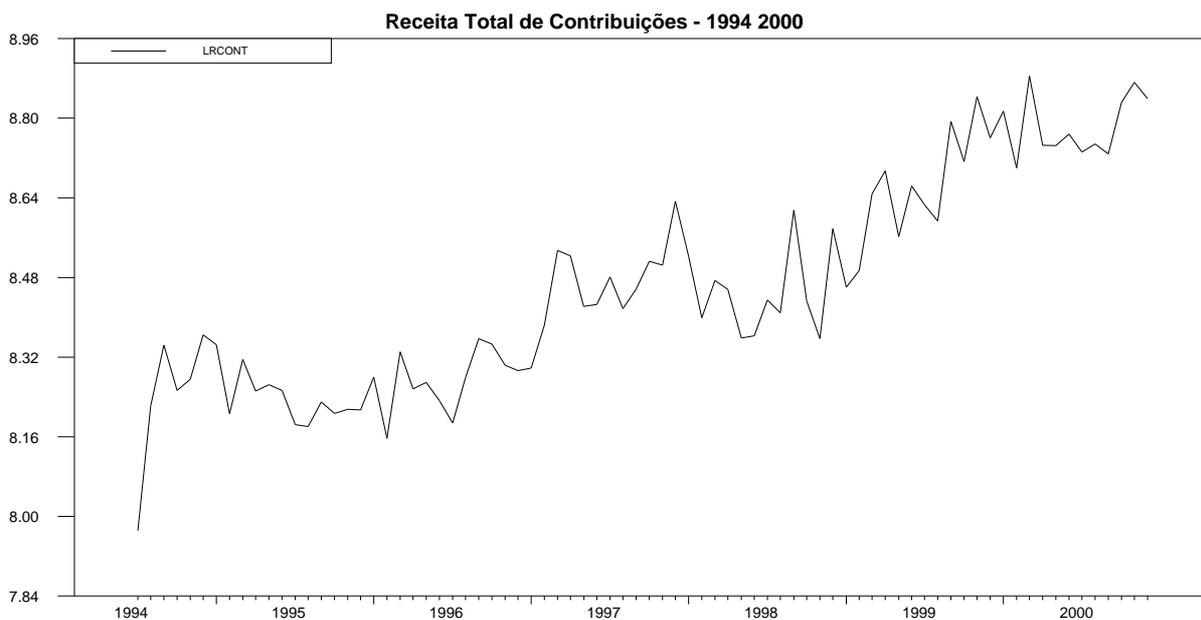


Figura 3.33 – Série In da Receita de Contribuições Corrigida- LRCONT - jul/1994 a dez/2000

### 3.2.10 - Receita Administrada pela SRF - RECADM

Esta série representa o somatório de todas as séries referentes aos impostos e contribuições federais administrados pela SRF, ou seja, o somatório das oito séries estudadas

até o presente momento mais a série Outras Receitas Administradas, não discutida neste trabalho. Tal série está apresentada no gráfico da figura 3.34. O gráfico da série logaritmizada, LRECADM\*, está apresentado na figura 3.35.

Através da simples observação da série LRECADM\*, constata-se que há alguns poucos pontos que destoam do conjunto restante das observações, indicando possíveis "outliers". Tais "outliers" resultaram, obviamente, de "outliers" observados nas séries individuais que impactaram a série agregada.

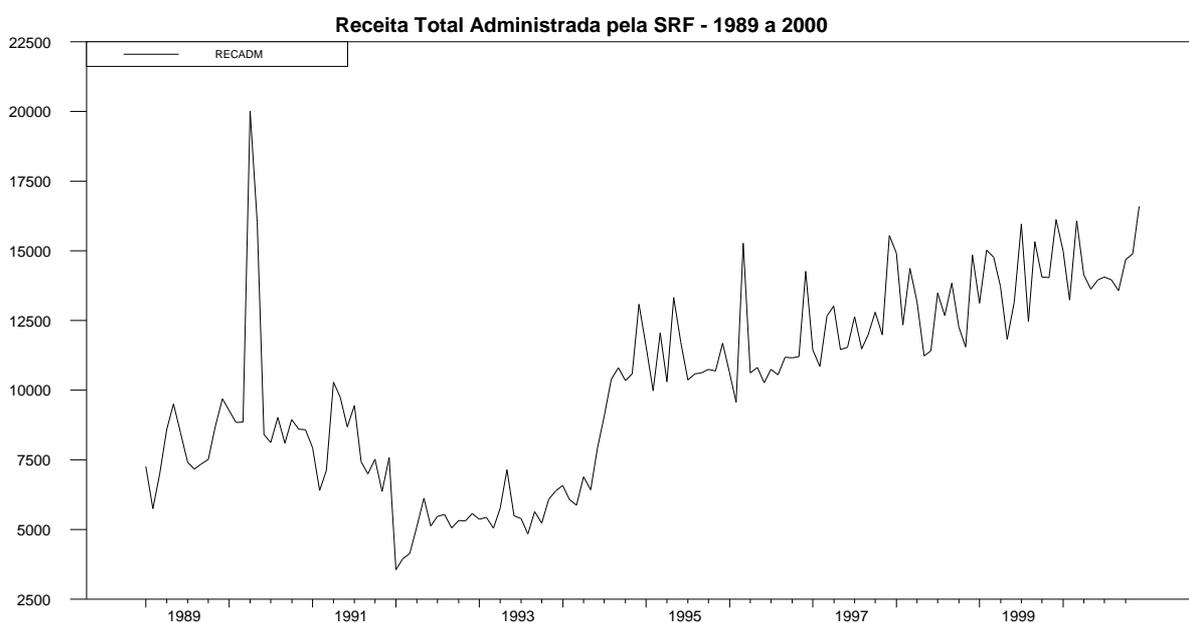


Figura 3.34 – Receita Administrada pela SRF - RECADM - jan/1989 a dez/2000

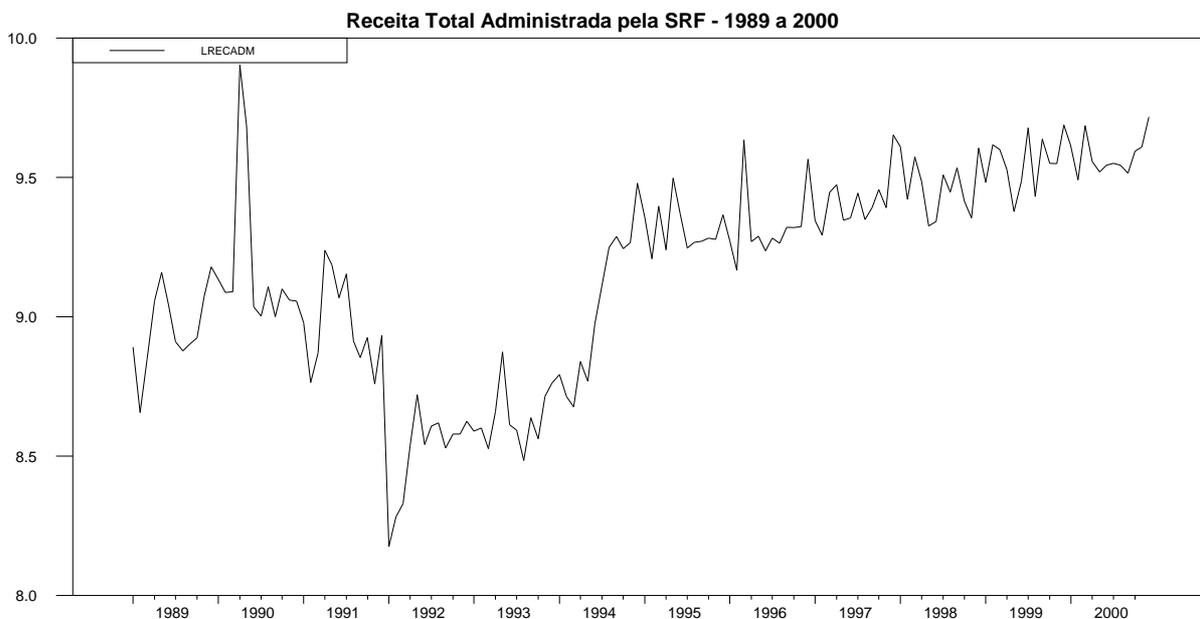


Figura 3.35 – ln da Receita Administrada pela SRF - LRECADM\* - jan/1989 a dez/2000

Os dados apresentados na fig. 3.36 mostram nitidamente uma mudança estrutural em julho de 1994, mês de implantação do Plano Real.

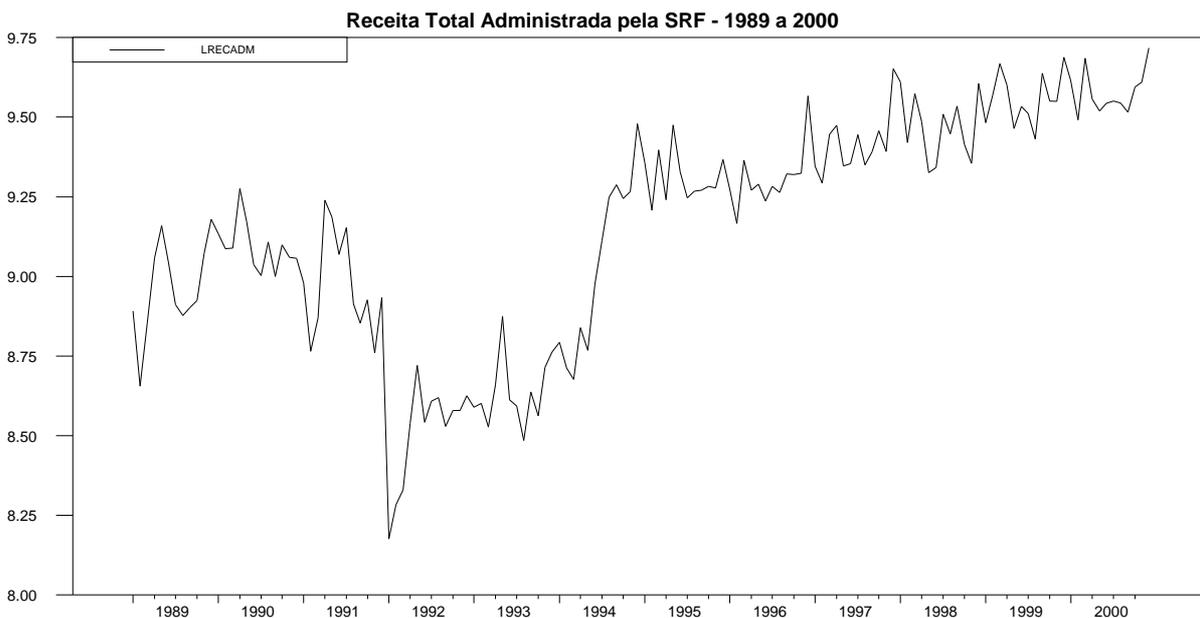


Figura 3.36 – Série ln da Receita Administrada pela SRF Corrigida- LRECADM - jan/1989 a dez/2000

Apesar de a mudança estrutural poder ser facilmente observada através do gráfico da figura 3.32, faremos para esta série o teste-F para mudanças estruturais, apresentado e

discutido por Enders (1996 : pg. 33, 34). Este autor assim discorre sobre os procedimentos para a realização do teste:

1° - Se há um número suficiente de observações, ajustar o mesmo modelo ARMA para cada um dos dois subperíodos analisados pode trazer informações relativas à hipótese de que o processo gerador dos dados está inalterado ao longo do tempo. Para tal, divide-se os dados observados ( $T$ ) em duas subamostras com  $m$  observações na primeira e  $n = T - m$ , na segunda. Calcula-se a Soma dos Quadrados dos Resíduos (RSS) e denota-se por  $\alpha_i(k)$  e  $\beta_j(k)$  os coeficientes  $\alpha_i$  e  $\beta_j$  estimados para a subamostra  $k$ . Utilizando os dados de cada subamostra, estima-se os dois modelos:

$$y_t = \alpha_0(1) + \alpha_1(1)y_{t-1} + \dots + \alpha_p(1)y_{t-p} + \varepsilon_t + \beta_1(1)\varepsilon_{t-1} + \dots + \beta_q(1)\varepsilon_{t-q}, \text{ usando } 1, \dots, m$$

$$y_t = \alpha_0(2) + \alpha_1(2)y_{t-1} + \dots + \alpha_p(2)y_{t-p} + \varepsilon_t + \beta_1(2)\varepsilon_{t-1} + \dots + \beta_q(2)\varepsilon_{t-q}, p/m + 1 \dots T$$

2° - Calcula-se, então, a Soma dos Quadrados dos Resíduos para cada modelo,  $RSS_1$  e  $RSS_2$ , respectivamente.

3° - Para testar a hipótese de que todos os coeficientes são iguais [ $\alpha_i(1) = \alpha_i(2)$ , para todo  $i$  e  $\beta_j(1) = \beta_j(2)$ , para todo  $j$ ], utiliza-se o teste-F na forma:

$$F = \frac{(RSS - RSS_1 - RSS_2) / n}{(RSS_1 + RSS_2) / (T - 2n)}$$

onde  $n =$  número de parâmetros estimados  $= p + q + 1$  (com intercepto) ou  $p + q$  (sem intercepto) e o número de graus de liberdade é dado por  $(n, T - 2n)$ .

4° - Calcula-se, então, o valor da estatística  $F$  e, quanto maior for o valor calculado, mais restritiva é a hipótese de que os dois conjuntos de coeficiente sejam iguais. Ou seja, se o valor da estatística  $F$  calculada for superior ao valor crítico tabelado para os graus de liberdade considerados, há fortes evidências de que a série apresenta mudança estrutural.

Para a série em questão procedemos à identificação, estimação e checagem de diagnóstico de um modelo ARIMA que se ajustasse adequadamente aos dados de toda amostra, tendo encontrado, como melhor modelo, o SARIMA (1,1,1)(0,1,1)<sub>12</sub>. Em seguida, estimamos os coeficientes do modelo para as duas subamostras (jan/1989 a jun/1994 e jul/1994 a dez/2000). Procedeu-se da forma citada nos quatro passos acima, utilizando-se o software RATS, cujo programa se encontra no Apêndice 1, e obteve-se para a estatística  $F$

calculado o valor  $F(3,100) = 4.13183$ , com nível de significância de 0.00831750, indicando haver forte evidência de mudança estrutural em julho de 1994.

Em virtude da comprovação de ocorrência da mudança estrutural observada na série em estudo, utilizaremos somente os dados referentes ao segundo período (jul/94 a dez/2000) para a identificação e estimação do modelo, perfazendo um total de 78 observações, suficientes para a utilização da modelagem ARIMA, mesmo que haja alguma sazonalidade na série. O gráfico da série LRECADM para o período considerado está, agora, apresentado na figura 3.37.

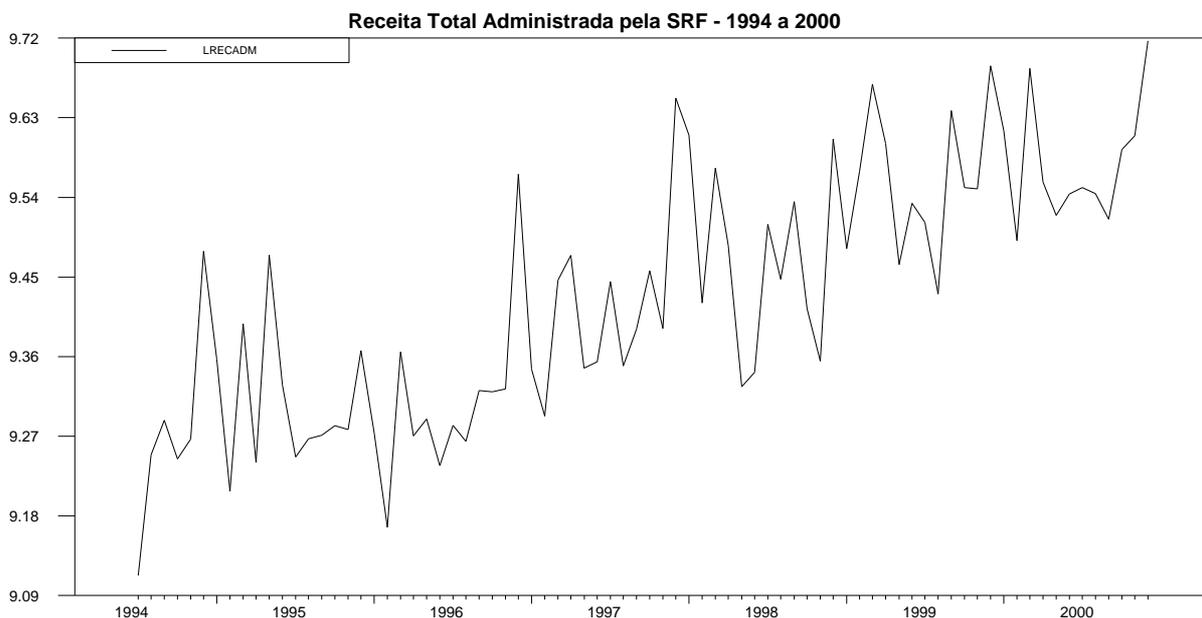


Figura 3.37 – Série ln da Receita Administrada pela SRF Corrigida- LRECADM - jul/1994 a dez/2000

## 4 - RESULTADOS DA MODELAGEM BOX-JENKINS

*Neste capítulo, utilizaremos os procedimentos da metodologia ARIMA para modelar as séries de tributos objetos do presente estudo. Para cada série, proceder-se-á à identificação e à estimação do modelo, à verificação de diagnóstico e, por fim, à previsão dos valores mensais de 2001 para cada modelo selecionado.*

### 4.1 - SÉRIE DO IMPOSTO DE IMPORTAÇÃO (II)

#### 4.1.1 - Identificação

Conforme analisado no capítulo 3, o gráfico da série do Imposto de Importação ( $Z_{II}$ ), apresentado na figura 3.1, mostra uma variância que aumenta com o aumento do nível da série. Para estabilizar esta variância utilizar-se-á o logaritmo natural da série original, de tal forma que o modelo ARIMA seja construído para a série transformada  $Z'_{II} = \ln(Z_{II})$ . O gráfico de  $Z'_{II}$  está apresentado na figura 3.2 para o período jan/1989 a dez/2000 e na figura 3.4 para o período jul/1994 a dez/2000, que será o utilizado no presente estudo.

Para identificação do modelo, utilizar-se-á, inicialmente, as Funções de Autocorrelação Amostral (FACA) e de Autocorrelação Parcial Amostral (FACPA). A figura 4.1 mostra a FACA (em preto) e a FACPA (em azul) para  $Z'_{II}$  com  $d = D = 0$  (sem qualquer diferenciação). Verifica-se que as autocorrelações (FACA) são significativamente diferentes de zero até a defasagem 5, decaindo lentamente até a defasagem 18. Este comportamento sugere que a série  $Z'_{II}$  seja não-estacionária, consistente com o gráfico apresentado na figura 3.2, em que os dados apresentam uma tendência ascendente.

A primeira diferença não sazonal  $W^*_{II} = \nabla Z'_{II}$  ( $d = 1$ ) está apresentada no gráfico 4.2(superior). Observa-se que esta série tem uma média estável ao longo do tempo e, como relatado no capítulo 2, isto indica que a série diferenciada é estacionária. Entretanto, os dados da figura 4.2 (superior) mostram uma forte variação sazonal.

A FACA e a FACPA para a série diferenciada  $W^*_{II} = \nabla Z'_{II}$  ( $d = 1$ ) estão apresentadas na figura 4.2(inferior). Verifica-se ali que a autocorrelação cai a valores insignificamente próximos de zero bastante rapidamente, com exceção dos valores para as defasagens 1 e 10. Isto reforça a conclusão do parágrafo anterior de que a série diferenciada é agora estacionária, nos termos do discutido no capítulo 2. Entretanto, as autocorrelações para as defasagens 10, 12 e 24 são altas e não se observa uma queda significativa da defasagem 12 para a 24. Isto

reforça a observação anteriormente feita de que os dados apresentam forte variação sazonal e sugere uma diferenciação sazonal ( $D = 1$ ) em adição à diferenciação não sazonal ( $d = 1$ ).

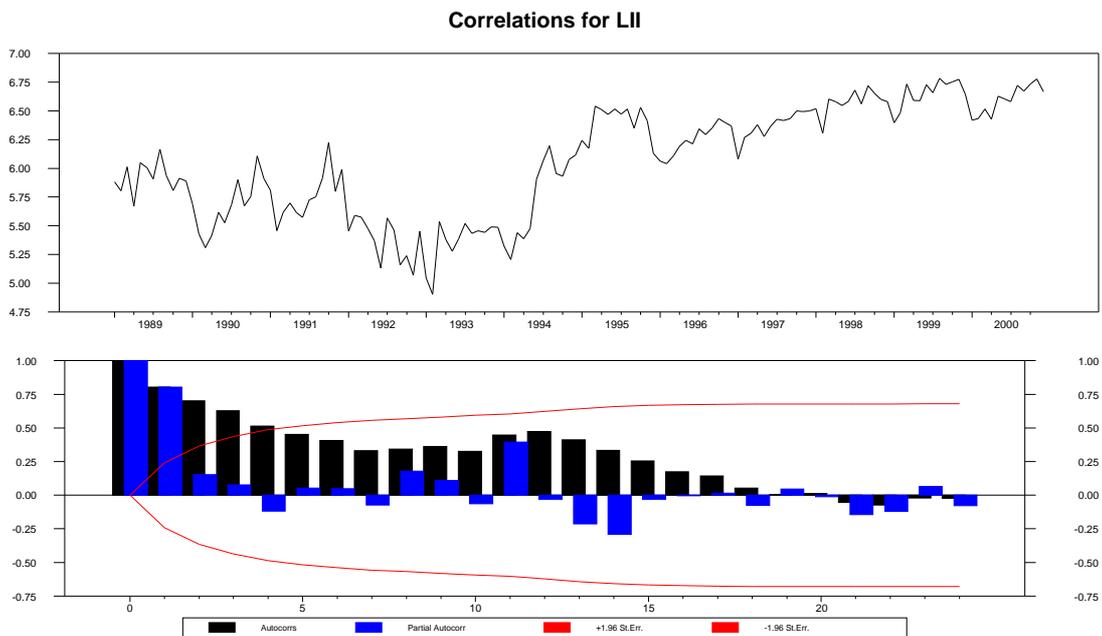


Figura 4.1 – Funções de Autocorrelação Amostral(FACA e FACPA) para a série  $Z'_{II}$

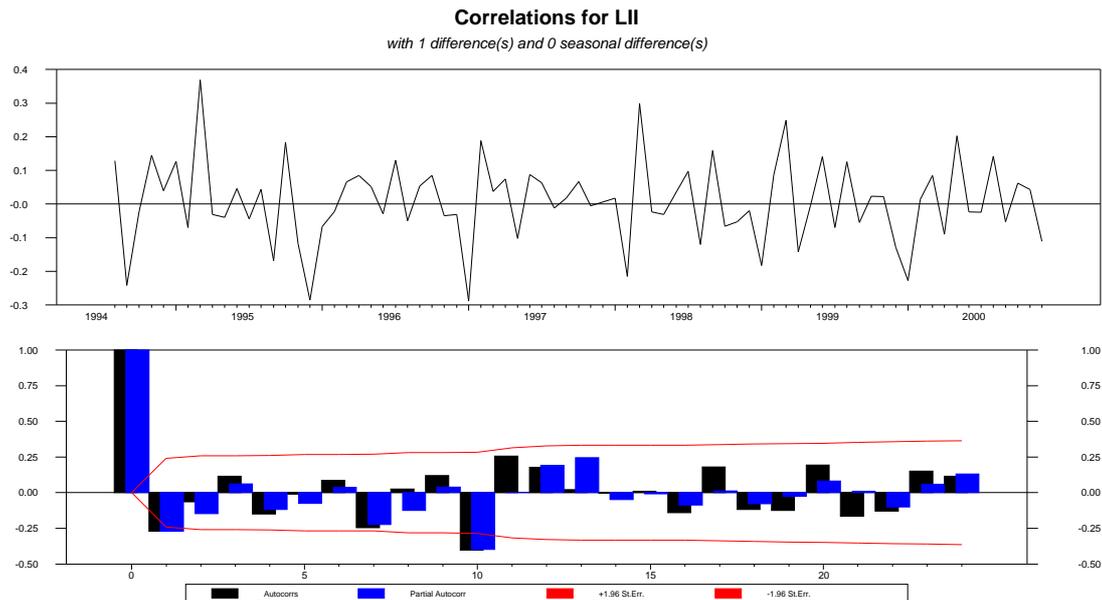


Figura 4.2 – Gráfico da Série Diferenciada  $W^*_{II} = \nabla Z'_{II}$  ( $d = 1$ ) e suas Funções de Autocorrelação Amostral (FACA e FACPA)

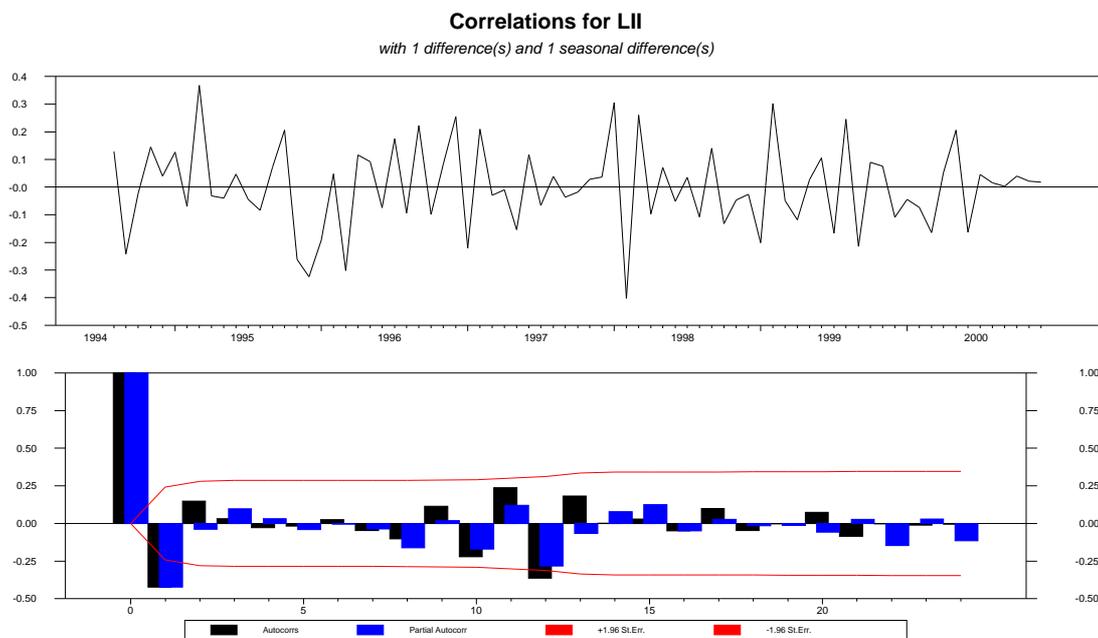


Figura 4.3 – Gráficos da Série Diferenciada  $W_{II} = \nabla_{12}\nabla Z'_{II}$  ( $d = 1, D = 1$ ) e suas Funções de Autocorrelação Amostral (FACA e FACPA)

A figura 4.3 (superior) apresenta o gráfico dos dados após sofrerem as duas diferenciações, sazonal ( $D = 1$ ) e não sazonal ( $d = 1$ ),  $W_{II} = \nabla_{12}W^*_{II} = \nabla_{12}\nabla Z'_{II}$ . Esta nova série,  $W_{II}$ , apresenta uma média estável e não apresenta mais o forte padrão sazonal observado na série  $W^*_{II}$ . A FACA e a FACPA para  $W_{II}$  estão apresentadas na figura 4.3 (inferior), em preto e em azul, respectivamente.

Comparando-se a FACA e a FACPA da série  $W_{II}$  com as FAC e FACP teóricas apresentadas no capítulo 2, conclui-se que remanescem padrões sazonais e não sazonais que sugerem a inclusão de componentes de média móveis (MA) e/ou auto-regressivos (AR). No caso da parte não sazonal do modelo, as funções de autocorrelação sugerem a inclusão de um termo de média móvel - MA(1) - e/ou um termo auto-regressivo - AR(1). No que se refere a parte sazonal do modelo, as funções de autocorrelação também sugerem a inclusão de um termo auto-regressivo - SAR(1) e ou um de média móvel - SMA(1). Assim, combinando os componentes supracitados, poder-se-ia, inicialmente, analisar as seguintes opções de modelos.

**Modelo 1 - ARIMA(0, 1, 0)(0, 1, 1)<sub>12</sub>**

As autocorrelações residuais para os modelos em análise foram obtidas pelo RATS (não mostradas). Observa-se, para este modelo que o ajuste em questão é bom em representar o

padrão sazonal da série (as autocorrelações nas defasagens 12 e 24 provam isso), mostrando ser evidente a inclusão de um termo sazonal, auto-regressivo ou de média móvel. Entretanto, as autocorrelações da primeira defasagem indicam a necessidade de inclusão de termos de médias móveis e/ou auto-regressivos não sazonais. Inicialmente, incluir-se-á um termo de média móvel.

**Modelo 2 - ARIMA(0, 1, 1)(0, 1, 1)<sub>12</sub>**

Agora todas as correlações residuais são significativamente diferentes de zero, indicando ser este um potencial modelo a representar a série em estudo. Tentemos com um termo SAR(1).

**Modelo 3 - ARIMA(0, 1, 1)(1, 1, 0)<sub>12</sub>**

Aqui se observa algumas correlações residuais não significativamente diferentes de zero (autocorrelações das defasagens 7 e 10, por exemplo), indicando não ser este um modelo adequado para representar a série em estudo. Tentemos outra opção.

**Modelo 4 - ARIMA(1, 1, 0)(0, 1, 1)<sub>12</sub>**

Agora todas as correlações residuais são significativamente diferentes de zero, indicando ser este um outro potencial modelo a representar a série em estudo.

**Modelo 5 - ARIMA(1, 1, 0)(1, 1, 1)<sub>12</sub>**

Aqui se observa algumas correlações residuais não significativamente diferentes de zero (autocorrelações das defasagens 3 e 7, por exemplo), indicando não ser este um modelo adequado para representar a série em estudo. Tentemos eliminar neste, o termo sazonal de média forma e analisemos, então, o modelo ARIMA(1, 1, 0)(1, 1, 0)<sub>12</sub>.

O modelo ARIMA(1,1,0)(1,1,0)<sub>12</sub> também não se mostrou adequado.

Adicionalmente, para corroborar as constatações precedentes, optou-se por utilizar os critérios de seleção de Akaike e de Schwartz, quais sejam, o *Akaike Information Criterion*(AIC) e *Schwartz Bayesian Criterion* (SBC ou BIC).

Os valores dos critérios AIC e BIC para cada um dos modelos estudados ( $d = 1$ ;  $D = 1$ ;  $p = 0, 1$ ;  $q = 0, 1$ ;  $P = 0, 1$  e  $Q = 0, 1$ ) estão apresentados na tabela 4.1. Como o critério BIC seleciona modelos mais parcimoniosos que o AIC, utilizaremos aquele para a escolha dos melhores modelos.

Tabela 4.1 – Valores do AIC e do BIC para a Série do Imposto de Importação

MODELO	AIC	BIC	MODELO	AIC	BIC
(0,1,0)(0,1,0) <sub>12</sub>	4,86	4,86	<b>(0,1,1)(1,1,0)<sub>12</sub></b>	-20,73	<b>-16,83</b>
(0,1,0)(0,1,1) <sub>12</sub>	-5,55	-3,60	(0,1,1)(1,1,1) <sub>12</sub>	-22,59	-15,74
(0,1,0)(1,1,0) <sub>12</sub>	-4,72	-2,77	(1,1,0)(0,1,0) <sub>12</sub>	-16,66	-14,71
(0,1,0)(1,1,1) <sub>12</sub>	-7,28	-3,38	<b>(1,1,0)(0,1,1)<sub>12</sub></b>	-21,66	<b>-17,76</b>
(0,1,1)(0,1,0) <sub>12</sub>	-13,80	-11,85	<b>(1,1,0)(1,1,0)<sub>12</sub></b>	-24,01	<b>-20,10</b>
<b>(0,1,1)(0,1,1)<sub>12</sub></b>	<b>-19,37</b>	<b>-15,97</b>	<b>(1,1,0)(1,1,1)<sub>12</sub></b>	-25,62	<b>-19,77</b>

Assim, analisando a tabela 4.1, observa-se que os cinco modelos que apresentam menores valores para o BIC são os seguintes: ARIMA(0,1,1)(0,1,1)<sub>12</sub>, ARIMA(0,1,1)(1,1,0)<sub>12</sub>, ARIMA(1,1,0)(0,1,1)<sub>12</sub>, ARIMA(1,1,0)(1,1,0)<sub>12</sub>, ARIMA(1,1,0)(1,1,1)<sub>12</sub>.

Passa-se, então, para a etapa de estimação dos modelos inicialmente selecionados.

### 4.1.2 – Estimação

As estimativas de máxima verossimilhança dos modelos selecionados foram obtidas através do software RATS e são as seguintes:

$$(1 - B)(1 - B^{12})Z_{II_t} = (1 - 0,492B)(1 - 0,478B^{12})\hat{a}_t \tag{4.1}$$

(3,97)                      (3,51)  
[0,12]                      [0,14]

$$(1 - 0,306B^{12})(1 - B)(1 - B^{12})Z_{II_t} = (1 - 0,548B)\hat{a}_t \tag{4.2}$$

(3,05)    (4,57)  
[0,10]    [0,12]

$$(1 - 0,544B)(1 - B)(1 - B^{12})Z_{II_t} = (1 - 0,461B^{12})\hat{a}_t \tag{4.3}$$

(4,47)    (3,20)  
[0,12]    [0,14]

$$(1 - 0,580B)(1 - 0,317B^{12})(1 - B)(1 - B^{12})Z_{II_t} = \hat{a}_t \tag{4.4}$$

(5,07)              (3,12)  
[0,11]              [0,10]

$$(1 - 0,565B)(1 - 0,266B^{12})(1 - B)(1 - B^{12})Z_{II_t} = (1 - 0,387B^{12})\hat{a}_t \tag{4.5}$$

(4,68)              (2,12)    (2,08)  
[0,12]              [0,13]    [0,19]

Os números entre parênteses são os valores absolutos da estatística  $t$  dos coeficientes e os entre colchetes são os desvios padrões das estimativas. Os coeficientes da parte autorregressiva (AR) respeitam as condições para a estacionariedade do modelo e os coeficientes da parte de média móvel (MA) respeitam as condições de invertibilidade, nos termos do explicitado no capítulo 2.

#### 4.1.3 - Checagem de Diagnóstico

Verifica-se que todos os coeficientes estimados são estatisticamente significativos a 5%. Os resíduos para os cinco modelos acima descritos, juntamente com a FACA e a FACPA residuais para cada um deles foram obtidos pelo RATS, conforme já citado. A FACA residual, em particular, mostra a satisfatoriedade do ajuste para os modelos  $ARIMA(0,1,1)(0,1,1)_{12}$  e  $ARIMA(1,1,0)(0,1,1)_{12}$ , indicando, claramente, que todas as autocorrelações residuais são estatisticamente nulas. Já para o caso dos modelos  $ARIMA(0,1,1)(1,1,0)_{12}$ ,  $ARIMA(1,1,0)(1,1,1)_{12}$  e  $ARIMA(1,1,0)(1,1,0)_{12}$  as autocorrelações residuais para algumas defasagens se mostraram estatisticamente diferentes de zero, conforme já discutido anteriormente.

Reforçando a satisfatoriedade dos ajustes dos modelos  $ARIMA(0,1,1)(0,1,1)_{12}$  e  $ARIMA(1,1,0)(0,1,1)_{12}$ , a estatística  $\chi^2$  se mostrou insignificante ao nível de 5% para os graus de liberdade de cada modelo e tanto os gráficos dos resíduos padronizados ( $a_t/\sigma_a$ ), traçados nas figuras 2.1.1 e 2.1.2 do Apêndice 2, quanto o teste de normalidade de Jarque-Bera ( $\chi^2 = 4,8280$  e  $8,4913$ , com níveis de significância de  $0,089$  e  $0,014$ ) sugerem que os choques devem ser normalmente distribuídos<sup>7</sup>. Devemos observar que tanto no gráfico da figura 2.1.1, quanto no gráfico da figura 2.1.2 somente um valor foi superior a três vezes o desvio padrão (valor para junho/1995). Em uma distribuição perfeitamente normal isso não ocorre, mas a justificativa pode estar no "outlier"<sup>8</sup> que não deve ter sido removido adequadamente pelo processo escolhido.

Assim, utilizaremos os modelos  $ARIMA(0,1,1)(0,1,1)_{12}$  e  $ARIMA(1,1,0)(0,1,1)_{12}$  para a realização das previsões para a série do Imposto de Importação.

---

<sup>7</sup> A estatística  $\chi^2$  e os valores do teste de Jarque-Bera, juntamente com seus níveis de significância, são dados pelo procedimento SARIMA do software RATS.

<sup>8</sup> Verifique que no capítulo 3, subitem 3.2.1, os valores para os meses de maio e junho de 1995 foram considerados "outliers".

#### 4.1.4 - Previsão

As previsões para os dois modelos selecionados estão mostradas abaixo nas tabelas 4.2 e 4.3, juntamente com o valor da raiz quadrada do erro quadrático médio de previsão (RQEQMP). Conforme se pode observar, foram calculados as RQEQMP para previsões para três, seis e doze passos à frente. A análise dos dados a ser realizada no capítulo 5 utilizará essa medida para a comparação dos modelos.

Tabela 4.2 – Previsões do modelo ARIMA (0, 1, 1)(0, 1, 1)<sub>12</sub>

MESES	Previsão do Modelo (LII)	Limite 95% abaixo	Previsão (II <sub>P</sub> ) milhões de R\$	Limite 95% acima	Valor Real Observado (II <sub>R</sub> ) milhões de R\$	ERRO ABS. (II <sub>P</sub> - II <sub>R</sub> )	ERRO PERC. (%)
JAN/01	6,5158	564,42	675,734333	809,00	724,383686	-48,65	-6,72
FEV/01	6,5041	557,86	667,874312	799,59	599,672735	68,20	11,37
MAR/01	6,6836	654,32	799,191027	976,13	864,785447	-65,59	-7,59
ABR/01	6,6239	604,20	752,875594	938,14	779,126172	-26,25	-3,37
MAI/01	6,6755	636,19	792,743727	987,82	898,763557	-106,02	-11,80
JUN/01	6,7244	654,85	832,472343	1058,28	833,518873	-1,05	-0,13
JUL/01	6,7519	659,78	855,683017	1109,76	891,449909	-35,77	-4,01
AGO/01	6,8069	669,75	904,063860	1220,36	951,736057	-47,67	-5,01
SET/01	6,7797	638,87	879,804743	1211,60	773,984209	105,82	13,67
OUT/01	6,8179	650,60	914,063460	1284,21	900,747094	13,32	1,48
NOV/01	6,8209	639,64	916,809768	1314,09	761,757844	155,05	20,35
DEZ/01	6,6535	530,33	775,493810	1133,99	596,130180	179,36	30,09
TOTAL		7460,80	9766,809993	12842,98	9576,055762	190,75	1,99
RAIZ QUADRADA DO EQM - 3 MESES						61,43	
RAIZ QUADRADA DO EQM - 6 MESES						62,25	
RAIZ QUADRADA DO EQM - 12 MESES						88,69	

Tabela 4.3 – Previsões do modelo ARIMA (1, 1, 0)(0, 1, 1)<sub>12</sub>

MESES	Previsão do Modelo (LII)	Limite 95% abaixo	Previsão (II <sub>P</sub> ) milhões de R\$	Limite 95% acima	Valor Real Observado (II <sub>R</sub> ) milhões de R\$	ERRO ABS. (II <sub>P</sub> - II <sub>R</sub> )	ERRO PERC. (%)
JAN/01	6,5225	568,22	680,276954	814,44	724,383686	-44,11	-6,09
FEV/01	6,5067	559,31	669,613044	801,67	599,672735	69,94	11,66
MAR/01	6,6860	655,89	801,111389	978,48	864,785447	-63,67	-7,36
ABR/01	6,6236	604,02	752,649765	937,86	779,126172	-26,48	-3,40
MAI/01	6,6798	638,93	796,159864	992,08	898,763557	-102,60	-11,42
JUN/01	6,7280	657,21	835,474645	1062,10	833,518873	1,96	0,23
JUL/01	6,7480	657,21	852,352352	1105,44	891,449909	-39,10	-4,39
AGO/01	6,8065	669,48	903,702307	1219,87	951,736057	-48,03	-5,05
SET/01	6,7842	641,75	883,772786	1217,07	773,984209	109,79	14,18
OUT/01	6,8225	653,60	918,277837	1290,13	900,747094	17,53	1,95
NOV/01	6,8276	643,94	922,973017	1322,92	761,757844	161,22	21,16
DEZ/01	6,6658	536,89	785,091287	1148,03	596,130180	188,96	31,70
TOTAL		7486,44	9801,455247	12890,09	12726,46	225,40	2,35
RAIZ QUADRADA DO EQM - 3 MESES						60,25	
RAIZ QUADRADA DO EQM - 6 MESES						60,72	
RAIZ QUADRADA DO EQM - 12 MESES						91,29	

## 4.2 - SÉRIE DO IMPOSTO SOBRE PRODUTOS INDUSTRIALIZADOS (IPI)

### 4.2.1 - Identificação

O gráfico da série do Imposto sobre Produtos Industrializados ( $Z_{IPI}$ ), apresentado na figura 3.5, mostra uma variância que cresce com o aumento do nível da série. Para estabilizar esta variância utilizar-se-á o logaritmo natural da série original, de tal forma que o modelo ARIMA seja construído para a série transformada  $Z'_{IPI} = \ln(Z_{IPI})$ . O gráfico de  $Z'_{IPI}$  está apresentado na figura 3.6 para o período jan/1989 a dez/2000 e na figura 3.8 para o período jul/1994 a dez/2000, que será o utilizado no presente estudo.

Para identificação do modelo utilizar-se-á, inicialmente, as Funções de Autocorrelação Amostral (FACA) e de Autocorrelação Parcial Amostral (FACPA). A figura 4.4 mostra a FACA, em preto, e a FACPA, em azul, para  $Z'_{IPI}$  com  $d = D = 0$  (sem qualquer diferenciação). Verifica-se que as autocorrelações (FACA) são significativamente diferentes de zero até a defasagem 6, decaindo lentamente até a defasagem 15. Este comportamento sugere que a série  $Z'_{IPI}$  seja não-estacionária, consistente com o gráfico apresentado na figura 3.6, em que os dados apresentam uma pequena tendência ascendente.

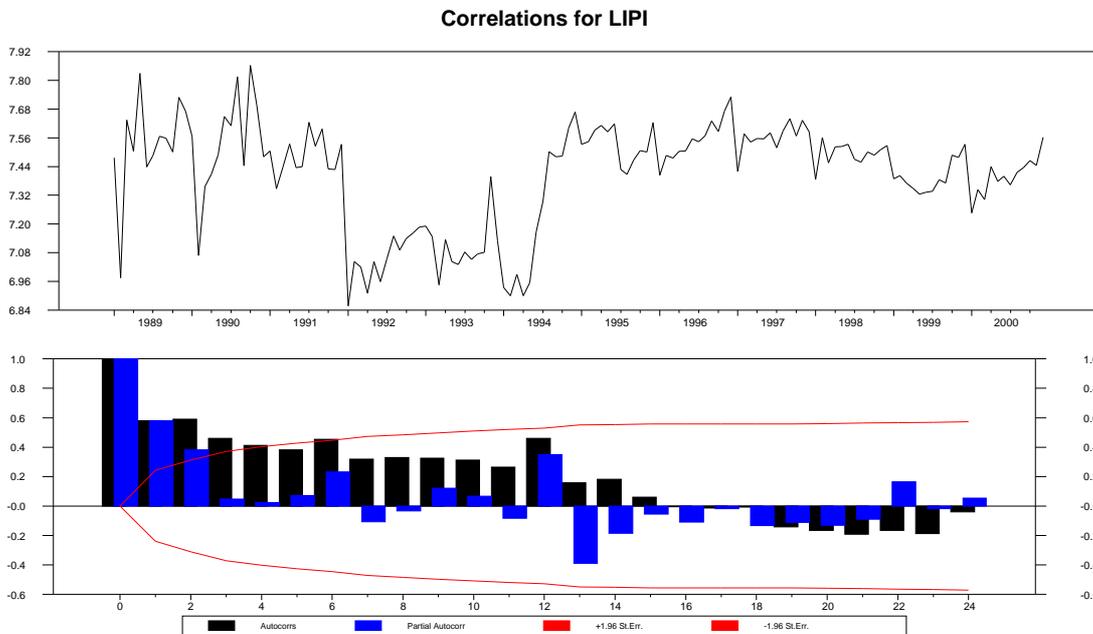


Figura 4.4 – Funções de Autocorrelação Amostral (FACA e FACPA) para a série  $Z'_{IPI}$

A primeira diferença não sazonal  $W^*_{IPI} = \nabla Z'_{IPI}$  ( $d = 1$ ) está apresentada no gráfico da figura 4.5 (superior). Observa-se que esta série tem uma média estável ao longo do tempo e, como relatado no capítulo 2, isto indica que a série diferenciada é estacionária. Entretanto, os dados da figura 4.5 (superior) mostram uma certa variação sazonal.

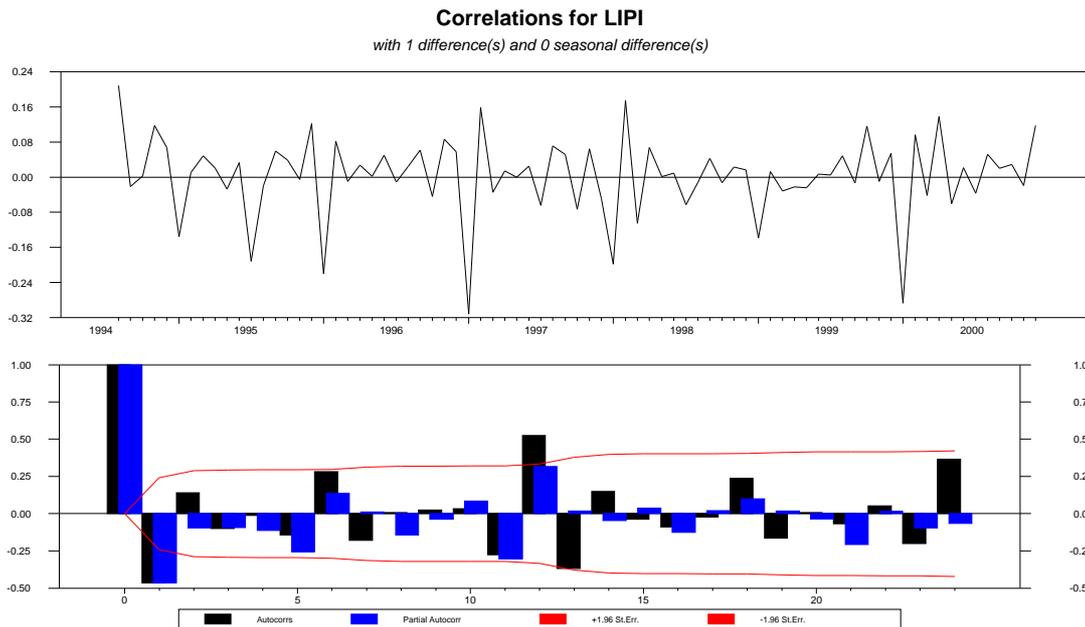


Figura 4.5 – Gráfico da Série Diferenciada  $W^*_{IPI} = \tilde{\nabla} Z'_{IPI}$  ( $d = 1$ ) e suas Funções de Autocorrelação Amostral (FACA e FACPA)

A FACA e a FACPA para a série diferenciada  $W^*_{IPI} = \nabla Z'_{IPI}$  ( $d = 1$ ) estão apresentadas na figura 4.5(inferior). Verifica-se ali que a autocorrelação (FACA) cai a valores insignificativamente próximos de zero bastante rapidamente, salvo alguns valores para as defasagens 1, 6, 12, 13 e 24. Isto reforça a conclusão do parágrafo anterior de que a série diferenciada é agora estacionária, nos termos do discutido no capítulo 2. Entretanto, as autocorrelações para as defasagens 12 e 24 são altas e não se observa uma redução significativa da defasagem 12 para a 24. Isto reforça a observação anteriormente feita de que os dados apresentam certa variação sazonal e sugere uma diferenciação sazonal ( $D = 1$ ) em adição à diferenciação não sazonal ( $d = 1$ ).

A figura 4.6 (superior) apresenta o gráfico dos dados após sofrerem as duas diferenciações, sazonal ( $D = 1$ ) e não sazonal ( $d = 1$ ),  $W_{IPI} = \nabla_{12} W^*_{IPI} = \nabla_{12} \nabla Z'_{IPI}$ . Esta nova série,  $W_{IPI}$ , apresenta uma média estável e não apresenta mais o forte padrão sazonal observado na série  $W^*_{IPI}$ . A FACA e a FACPA para  $W_{IPI}$  estão apresentadas na figura 4.6(inferior), em preto e azul, respectivamente.

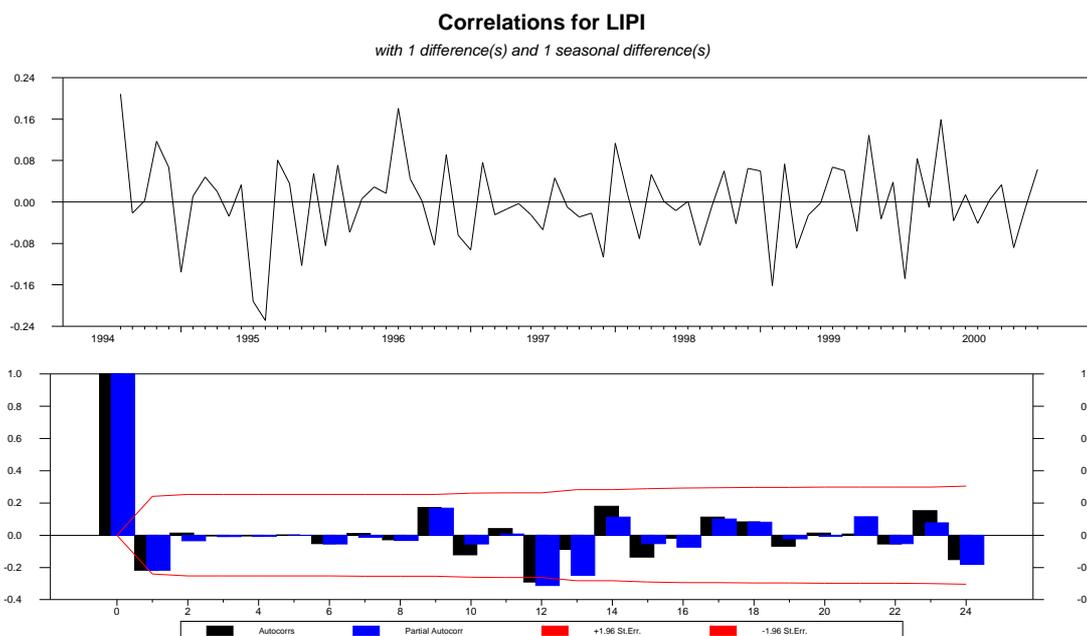


Figura 4.6 – Gráfico da Série Diferenciada  $W_{IPI} = \nabla_{12} \nabla Z'_{IPI}$  ( $d = 1, D = 1$ ) e suas Funções de Autocorrelação Amostral (FACA e FACPA)

Comparando-se a FACA e a FACPA da série  $W_{IPI}$  com as FAC e FACP teóricas apresentadas no capítulo 2, conclui-se que remanescem padrões sazonais e não sazonais que sugerem a inclusão de componentes de média móveis (MA) e/ou auto-regressivos (AR). No

caso da parte não sazonal do modelo, as funções de autocorrelação sugerem a inclusão de um termo de média móvel - MA(1) - e/ou um termo auto-regressivo - AR(1). No que se refere a parte sazonal do modelo, as funções de autocorrelação sugerem a inclusão de um ou dois termos auto-regressivos - SAR(1) ou SAR(2) e um de média móvel - SMA(1). Assim, combinando os componentes supracitados, poder-se-ia, inicialmente, analisar as seguintes opções de modelos.

**Modelo 1 - ARIMA(0, 1, 0)(1, 1, 0)<sub>12</sub>**

As autocorrelações residuais para os modelos em análise foram obtidas pelo RATS (não mostradas). Observa-se correlações residuais não significativamente diferentes de zero (autocorrelações das defasagens 1 e 9), indicando não ser este um modelo adequado para representar a série em estudo. Tentemos a inclusão de um termo auto-regressivo não sazonal - AR(1) e/ou um termo de média móvel não sazonal - MA(1), conforme parece indicar as autocorrelações na primeira defasagem.

**Modelo 2 - ARIMA(1, 1, 0)(1, 1, 0)<sub>12</sub>**

Ainda se observa correlações residuais não significativamente diferentes de zero (autocorrelações da defasagem 9), indicando não ser este, ainda, um modelo adequado para representar a série em estudo.

**Modelo 3 - ARIMA(0, 1, 1)(1, 1, 0)<sub>12</sub>**

Ainda se observa correlações residuais não significativamente diferentes de zero (para as autocorrelações da defasagem 9 novamente), indicando não ser este, ainda, um modelo adequado para representar a série em estudo.

Tentemos o seguinte modelo, então:

**Modelo 4 - ARIMA(0, 1, 1)(0, 1, 1)<sub>12</sub>**

Agora todas as correlações residuais são significativamente diferentes de zero, indicando ser este um potencial modelo a representar a série em estudo. Passa-se, então a analisar outros modelos que tenham um segundo termo auto-regressivo sazonal.

A permanência das autocorrelações residuais não estatisticamente nulas na defasagem 9 nos modelos 2 e 3 está sugerindo a possibilidade de mais um termo sazonal para aqueles modelos. Tentemos mais um termo auto-regressivo sazonal - SAR(2).

**Modelo 5 - ARIMA(1, 1, 0)(2, 1, 0)<sub>12</sub>**

Agora todas as correlações residuais são significativamente diferentes de zero, indicando ser este um potencial modelo a representar a série em estudo. Passa-se, então a analisar outros modelos que tenham um segundo termo auto-regressivo sazonal.

**Modelo 6 - ARIMA(0, 1, 1)(2, 1, 0)<sub>12</sub>**

Novamente todas as correlações residuais são significativamente diferentes de zero, indicando ser este um outro potencial modelo a representar a série em estudo.

**Modelo 7 - ARIMA(1, 1, 1)(2, 1, 0)<sub>12</sub>**

Novamente todas as correlações residuais são significativamente diferentes de zero, indicando ser este um outro potencial modelo a representar a série em estudo.

Adicionalmente, para corroborar as constatações precedentes, optou-se por utilizar os critérios de seleção de Akaike e de Schwartz, quais sejam, o *Akaike Information Criterion* (AIC) e *Schwartz Bayesian Criterion* (SBC ou BIC).

Os valores dos critérios AIC e BIC para cada um dos modelos estudados ( $d = 1$ ;  $D = 1$ ;  $p = 0, 1$ ;  $q = 0, 1$ ;  $P = 0, 1, 2$  e  $Q = 0, 1$ ) estão apresentados na tabela 4.4. Como o critério BIC seleciona modelos mais parcimoniosos que o AIC, utilizaremos aquele para a escolha dos melhores modelos.

Tabela 4.4 – Valores do AIC e do BIC para a Série do Imposto sobre Produtos Industrializados (IPI)

MODELO	AIC	BIC	MODELO	AIC	BIC
(0,1,0)(0,1,0) <sub>12</sub>	-65,31	-65,31	(1,1,0)(1,1,0) <sub>12</sub>	-74,91	-71,53
(0,1,0)(0,1,1) <sub>12</sub>	-70,70	-69,00	(1,1,1)(0,1,0) <sub>12</sub>	-69,90	-66,53
(0,1,0)(1,1,0) <sub>12</sub>	-68,10	-66,42	<b>(1,1,1)(0,1,1)<sub>12</sub></b>	<b>-80,45</b>	<b>-75,38</b>
(0,1,1)(0,1,0) <sub>12</sub>	-25,46	-23,77	(1,1,1)(1,1,0) <sub>12</sub>	-73,80	-68,73
<b>(0,1,1)(0,1,1)<sub>12</sub></b>	<b>-81,42</b>	<b>-78,05</b>	(0,1,0)(2,1,0) <sub>12</sub>	-73,01	-69,63
(0,1,1)(1,1,0) <sub>12</sub>	-75,34	-71,96	(0,1,1)(2,1,0) <sub>12</sub>	-80,05	-74,98
(1,1,0)(0,1,0) <sub>12</sub>	-71,15	-69,46	<b>(1,1,0)(2,1,0)<sub>12</sub></b>	<b>-80,53</b>	<b>-75,47</b>
<b>(1,1,0)(0,1,1)<sub>12</sub></b>	<b>-79,52</b>	<b>-76,14</b>	(1,1,1)(2,1,0) <sub>12</sub>	-78,66	-71,91

Assim, analisando a tabela 4.4, observa-se que os cinco modelos que apresentam menores valores para o BIC são os seguintes: ARIMA(0,1,1)(0,1,1)<sub>12</sub>, ARIMA(1,1,0)(0,1,1)<sub>12</sub>, ARIMA(1,1,1)(0,1,1)<sub>12</sub>, ARIMA(1,1,0)(2,1,0)<sub>12</sub> e ARIMA(0,1,1)(2,1,0)<sub>12</sub>.

Passa-se, então, para a etapa de estimação dos modelos inicialmente selecionados.

### 4.2.2 - Estimação

As estimativas de máxima verossimilhança dos modelos seleccionados foram obtidas através do software RATS e são as seguintes:

$$(1-B)(1-B^{12})Z_{II_t} = (1-0,594B)(1-0,733B^{12})\hat{a}_t \quad (4.6)$$

(4,53)      (4,89)  
[0,13]      [0,15]

$$(1-0,512B)(1-B)(1-B^{12})Z_{II_t} = (1-0,706B^{12})\hat{a}_t \quad (4.7)$$

(3,50)      (4,53)  
[0,15]      [0,16]

$$(1-0,485B)(1-0,498B^{12}-0,406B^{24})(1-B)(1-B^{12})Z_{II_t} = \hat{a}_t \quad (4.8)$$

(3,08)      (3,17)      (2,71)  
[0,16]      [0,16]      [0,15]

$$(1-0,242B)(1-B)(1-B^{12})Z_{II_t} = (1-0,470B)(1-0,799B^{12})\hat{a}_t \quad (4.9)$$

(0,98)      (2,14)      (5,48)  
[0,25]      [0,22]      [0,15]

$$(1-0,467B^{12}-0,372B^{24})(1-B)(1-B^{12})Z_{II_t} = (1-0,489B)\hat{a}_t \quad (4.10)$$

(2,89)      (2,55)      (3,22)  
[0,16]      [0,15]      [0,15]

Os números entre parênteses são os valores absolutos da estatística t dos coeficientes e os entre colchetes são os desvios padrões das estimativas. Os coeficientes da parte auto-regressiva (AR) respeitam as condições para a estacionariedade do modelo e os coeficientes da parte de média móveis (MA) respeitam as condições de invertibilidade, nos termos do explicitado no capítulo 2.

### 4.2.3 - Checagem de Diagnóstico

Verifica-se que, para o terceiro modelo, o coeficiente estimado do termo auto-regressivo não é estatisticamente significativo a 5% (t=0,98). Todos os demais coeficientes são estatisticamente significativos a 5%.

Os resíduos para os quatro modelos restantes, juntamente com a FACA e a FACPA residuais para cada um deles, foram obtidos pelo RATS, conforme já citado anteriormente. A FACA residual, em particular, mostra a satisfatoriedade do ajuste para os modelos  $ARIMA(0,1,1)(0,1,1)_{12}$ ,  $ARIMA(1,1,0)(2,1,0)_{12}$  e  $ARIMA(0,1,1)(2,1,0)_{12}$ , indicando, claramente, que todas as autocorrelações residuais são estatisticamente nulas, conforme já discutido anteriormente. O modelo  $ARIMA(1,1,0)(0,1,1)_{12}$  possui autocorrelações residuais estatisticamente não nulas, devendo ser, portanto, descartado.

Os modelos  $ARIMA(1,1,0)(2,1,0)_{12}$  e  $ARIMA(0,1,1)(2,1,0)_{12}$  apresentaram estatísticas  $\chi^2$  significantes ao nível de 5% para os graus de liberdade de cada modelo e o teste de normalidade de Jarque-Bera ( $\chi^2 = 3,436$  e  $3,062$ , com níveis de significância de  $0,179$  e  $0,216$ ) sugere que os choques não são normalmente distribuídos, clamando para que se descarte tais modelos.

Reforçando a satisfatoriedade do ajuste do modelo  $ARIMA(0,1,1)(0,1,1)_{12}$ , a estatística  $\chi^2$  se mostrou insignificante ao nível de 5% para os graus de liberdade do modelo e tanto o gráfico dos resíduos padronizados ( $a_t/\sigma_a$ ), traçado nas figuras 2.2.1 do Apêndice 2, quanto o teste de normalidade de Jarque-Bera ( $\chi^2 = 19,845$  com níveis de significância de  $0,00005$ ) sugerem que os choques devem ser normalmente distribuídos. Devemos observar que no gráfico da figura 2.2.1 somente dois valores foram superiores a três vezes o desvio padrão (valores para dezembro/1994 e julho/1995). Em uma distribuição perfeitamente normal isso obviamente não deveria ocorrer, mas a justificativa pode estar nos "outliers" que não devem ter sido removidos adequadamente pelo processo escolhido.

Assim, utilizaremos o modelo  $ARIMA(0,1,1)(0,1,1)_{12}$  para a realização das previsões para a série do Imposto sobre Produtos Industrializados.

#### 4.2.4 - Previsão

As previsões para o modelo selecionado estão mostradas abaixo na tabela 4.5, juntamente com o valor da raiz quadrada do erro quadrático médio de previsão (RQEQMP).

Tabela 4 5 – Previsões do modelo ARIMA (1, 1, 0)(2, 1, 0)<sub>12</sub> em milhões de R\$

MÊS	Previsão Calculada (lnF)	Limite 95% abaixo	Previsão (F)	Limite 95% acima	Valor Real Observado(Z)	ERRO (Z - F)	ERRO (%)
Jan/01	7,2980	1233,98	1477,342286	1768,70	1476,73234	0,61	0,04
Fev/01	7,3825	1342,78	1607,603758	1924,65	1614,843449	-7,24	-0,45
Mar/01	7,3409	1262,56	1542,099379	1883,52	1485,339792	56,76	3,82
Abr/01	7,4090	1324,78	1650,774746	2056,99	1782,864213	-132,09	-7,41
Mai/01	7,4062	1321,07	1646,159042	2051,24	1739,109891	-92,95	-5,34
Jun/01	7,4477	1349,78	1715,912006	2181,35	1781,300754	-65,39	-3,67
Jul/01	7,4159	1281,65	1662,204479	2155,76	1740,241092	-78,04	-4,48
Ago/01	7,4077	1221,34	1648,630134	2225,42	1709,468182	-60,84	-3,56
Set/01	7,4262	1219,50	1679,413661	2312,77	1744,75232	-65,34	-3,74
Out/01	7,4601	1236,57	1737,321781	2440,85	1773,602727	-36,28	-2,05
Nov/01	7,4642	1217,07	1744,459422	2500,39	1766,696247	-22,24	-1,26
Dez/01	7,5332	1278,19	1869,076990	2733,12	1913,632225	-44,56	-2,33
TOTAL		15289,28	19980,997684	26234,76	20528,58323	-547,59	-2,67
RAIZ QUADRADA DO ERRO QUADRADO MÉDIO – 3 MESES						33,04	
RAIZ QUADRADA DO ERRO QUADRADO MÉDIO – 6 MESES						74,88	
RAIZ QUADRADA DO ERRO QUADRADO MÉDIO – 12 MESES						65,51	

### 4.3 - SÉRIE DO IMPOSTO DE RENDA TOTAL (IR)

#### 4.3.1 - Identificação

O gráfico da série do Imposto de Renda ( $Z_{IR}$ ), apresentado na figura 3.9, mostra uma variância que cresce com o aumento do nível da série. Conforme já discutido, uma forma eficiente de estabilizar a variância consiste em utilizar o logaritmo natural da série original, de tal forma que o modelo ARIMA seja construído para a série transformada  $Z'_{IR} = \ln(Z_{IR})$ . O gráfico de  $Z'_{IR}$  está apresentado na figura 3.10 para o período jan/1989 a dez/2000 e na figura 3.12 para o período jul/1994 a dez/2000, que será o utilizado no presente estudo.

Da mesma forma que nas subseções anteriores deste capítulo, para identificação do modelo utilizar-se-á, inicialmente, as Funções de Autocorrelação Amostral (FACA) e de Autocorrelação Parcial Amostral (FACPA). A figura 4.7 mostra a FACA, em preto, e a FACPA, em azul, para  $Z'_{IR}$  com  $d = D = 0$  (sem qualquer diferenciação). Verifica-se que as autocorrelações (FACA) são significativamente diferentes de zero até a defasagem 12, decaindo lentamente. Este comportamento sugere que a série  $Z'_{IR}$  seja não-estacionária, consistente com o gráfico apresentado na figura 3.12, em que os dados apresentam uma tendência ascendente.

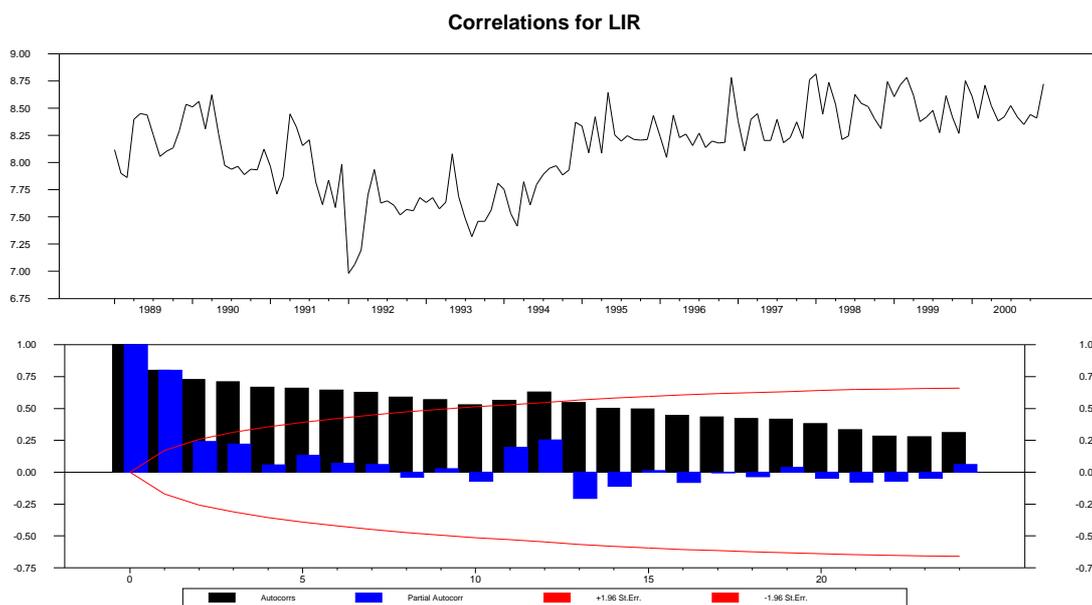


Figura 4.7 – Funções de Autocorrelação Amostral (FACA e FACPA) para a série  $Z'_{IR}$

A primeira diferença não sazonal  $W^*_{IR} = \nabla Z'_{IR}$  ( $d = 1$ ) está apresentada no gráfico 4.8(superior). Observa-se que esta série tem uma média estável ao longo do tempo e, como relatado no capítulo 2, isto indica que a série diferenciada é estacionária. Entretanto, os dados da figura 4.8(superior) mostram uma forte variação sazonal.

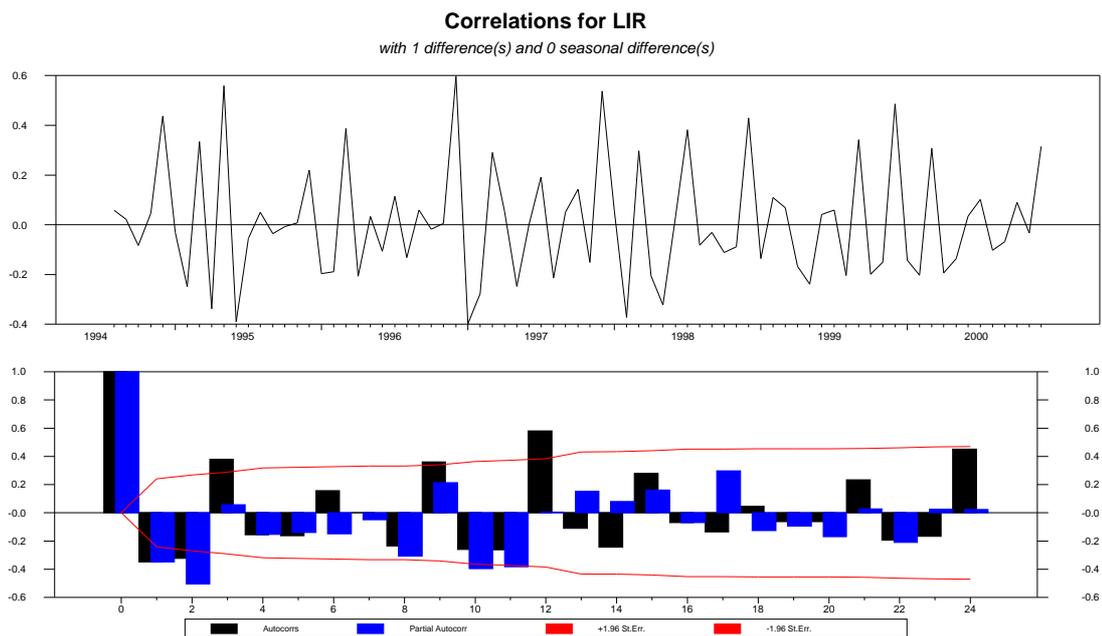


Figura 4.8 – Gráfico da Série Diferenciada  $W^*_{IR} = \nabla Z'_{IR}$  ( $d = 1$ ) e suas Funções de Autocorrelação Amostral (FACA e FACPA)

A FACA e a FACPA para a série diferenciada  $W^*_{IR} = \nabla Z'_{IR}$  ( $d = 1$ ) estão apresentadas na figura 4.8(inferior). Verifica-se ali que a autocorrelação cai a valores insignificamente próximos de zero bastante rapidamente, com exceção de alguns valores próximos às defasagens 1, 12 e 24. Isto reforça a conclusão do parágrafo anterior de que a série diferenciada é agora estacionária, nos termos do discutido no capítulo 2, mas que os dados apresentam forte variação sazonal (as autocorrelações para as defasagens 12 e 24 são altas e não se observa uma queda significativa da defasagem 12 para a 24), sugerindo, pois, uma diferenciação sazonal ( $D = 1$ ) em adição à diferenciação não sazonal ( $d = 1$ ).

A figura 4.9(superior) apresenta o gráfico dos dados após sofrerem as duas diferenciações, sazonal ( $D = 1$ ) e não sazonal ( $d = 1$ ),  $W_{IR} = \nabla_{12} W^*_{IR} = \nabla_{12} \nabla Z'_{IR}$ . Esta nova série,  $W_{IR}$ , apresenta uma média estável e não apresenta mais o forte padrão sazonal

observado na série  $W_{IR}^*$ . A FACA e a FACPA para  $W_{IR}$  estão apresentadas na figura 4.9(inferior), em preto e azul, respectivamente.

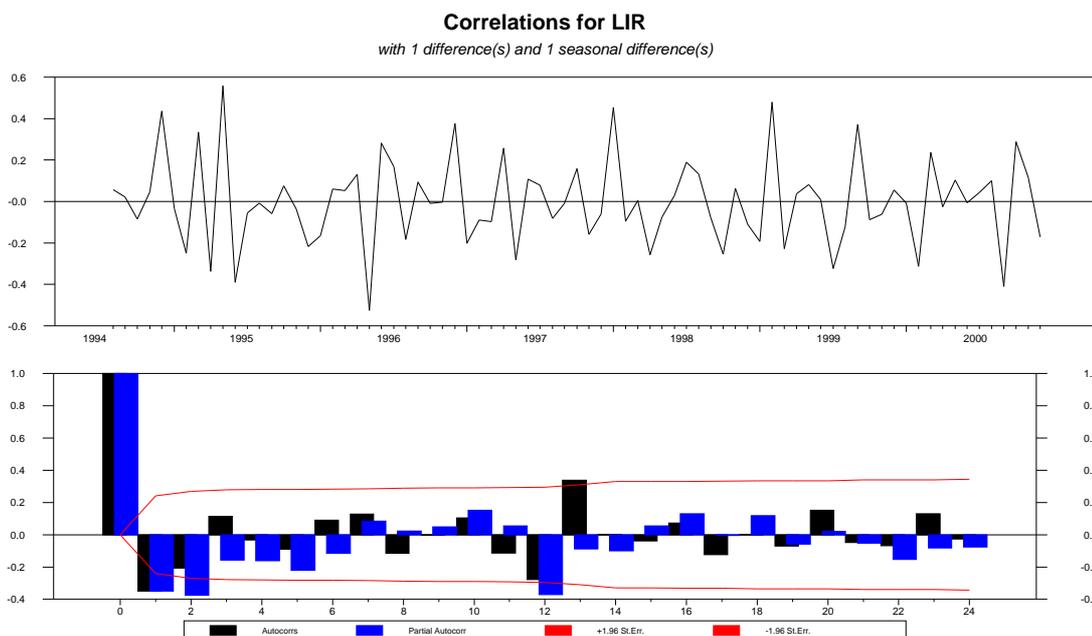


Figura 4.9 – Gráfico da Série Diferenciada  $W_{IR} = \nabla_{12}\nabla Z'_{IR}$  ( $d = 1, D = 1$ ) e suas Funções de Autocorrelação Amostral (FACA e FACPA)

Comparando-se a FACA e a FACPA da série  $W_{IR}$  com as FAC e FACP teóricas apresentadas no capítulo 2, conclui-se que remanescem padrões sazonais e não sazonais que sugerem a inclusão de componentes de média móveis (MA) e/ou auto-regressivos (AR). No caso da parte não sazonal do modelo, as funções de autocorrelação sugerem a inclusão de um termo de média móvel - MA(1) - e/ou um ou dois termos auto-regressivos - AR(1) e AR(2). No que se refere a parte sazonal do modelo, as funções de autocorrelação sugerem a inclusão de um termo auto-regressivo - SAR(1) ou de um termo de média móvel - SMA(1). Assim, combinando os componentes supracitados, poder-se-ia, inicialmente, analisar as seguintes opções de modelos:

**Modelo 1 - ARIMA(0, 1, 0)(0, 1, 1)<sub>12</sub>**

As autocorrelações residuais para os modelos em análise foram obtidas pelo RATS (não mostradas) Observa-se correlações residuais não significativamente diferentes de zero (autocorrelações das defasagens 1 e 2), indicando não ser este um modelo adequado para representar a série em estudo. Tentemos a inclusão de um termo auto-regressivo não sazonal -

AR(1) e/ou um termo de média móvel não sazonal - MA(1), conforme parece indicar as autocorrelações na primeira defasagem.

**Modelo 2 - ARIMA(1, 1, 0)(0, 1, 1)<sub>12</sub>**

Ainda se observa correlações residuais não significativamente diferentes de zero (autocorrelações da defasagem 2), indicando não ser este, ainda, um modelo adequado para representar a série em estudo. Tentemos incluir um termo de média móvel não sazonal.

**Modelo 3 - ARIMA(0, 1, 1)(0, 1, 1)<sub>12</sub>**

Agora, todas as correlações residuais são significativamente diferentes de zero, indicando ser este um potencial modelo a representar a série em estudo.

Verifiquemos a inclusão de um termo auto-regressivo sazonal.

**Modelo 4 - ARIMA(0, 1, 1)(1, 1, 0)<sub>12</sub>**

Aqui se observa uma correlação residual não significativamente diferente de zero (autocorrelação da defasagem 7, em preto), indicando que, provavelmente, este modelo não seja adequado para representar a série em estudo. Tentemos incluir um termo auto-regressivo não sazonal a este modelo.

**Modelo 5 - ARIMA(1, 1, 1)(1, 1, 0)<sub>12</sub>**

Aqui se observa correlações residuais não significativamente diferentes de zero (autocorrelação da defasagem 7, por exemplo), indicando que, provavelmente, este modelo também não seja adequado para representar a série em estudo. Tentemos incluir um termo auto-regressivo não sazonal a este modelo. Tentemos o seguinte:

**Modelo 6 - ARIMA(1, 1, 1)(0, 1, 1)<sub>12</sub>**

Agora, mais uma vez se verifica que todas as correlações residuais são significativamente diferentes de zero, indicando ser este um outro potencial modelo a representar a série em estudo. Por fim tentemos a inclusão de um segundo termo auto-regressivo ao modelo 2.

**Modelo 7 - ARIMA(2, 1, 0)(0, 1, 1)<sub>12</sub>**

Aqui, novamente se verifica que todas as correlações residuais são significativamente diferentes de zero, indicando ser este um outro potencial modelo a representar a série em estudo.

Para se determinar quais os modelos que melhor se ajustam aos dados, optou-se por utilizar os critérios de seleção de Akaike e de Schwartz, quais sejam, o *Akaike Information Criterion (AIC)* e *Schwartz Bayesian Criterion (SBC ou BIC)*.

Os valores dos critérios AIC e BIC para cada um dos modelos estudados ( $d = 1$ ;  $D = 1$ ;  $p = 0, 1, 2$ ;  $q = 0, 1$ ;  $P = 0, 1$ ; e  $Q = 0, 1$ ) estão apresentados na tabela 4.6 3. Como o critério BIC seleciona modelos mais parcimoniosos que o AIC, utilizaremos aquele para a escolha dos melhores modelos.

Tabela 4.6 – Valores do AIC e do BIC para a Série do Imposto de Renda Total (IR)

MODELO	AIC	BIC	MODELO	AIC	BIC
$(0,1,0)(0,1,0)_{12}$	33,44	33,44	$(1,1,0)(1,1,0)_{12}$	24,89	28,76
$(0,1,0)(0,1,1)_{12}$	14,46	16,40	$(1,1,1)(0,1,0)_{12}$	12,50	16,37
$(0,1,0)(1,1,0)_{12}$	29,07	31,00	<b><math>(1,1,1)(0,1,1)_{12}</math></b>	<b>-4,49</b>	<b>1,31</b>
$(0,1,1)(0,1,0)_{12}$	14,12	16,06	<b><math>(1,1,1)(1,1,0)_{12}</math></b>	<b>1,03</b>	<b>6,83</b>
<b><math>(0,1,1)(0,1,1)_{12}</math></b>	<b>2,90</b>	<b>6,76</b>	$(2,1,0)(0,1,0)_{12}$	23,24	27,10
<b><math>(0,1,1)(1,1,0)_{12}</math></b>	<b>11,74</b>	<b>15,61</b>	<b><math>(2,1,0)(0,1,1)_{12}</math></b>	<b>9,45</b>	<b>15,24</b>
$(1,1,0)(0,1,0)_{12}$	28,55	30,48	$(2,1,0)(1,1,0)_{12}$	19,67	25,46
$(1,1,0)(0,1,1)_{12}$	12,02	15,89			

Assim, analisando a tabela 4.6, observa-se que os cinco modelos que apresentam menores valores para o BIC são os seguintes:  $ARIMA(0,1,1)(0,1,1)_{12}$ ,  $ARIMA(0,1,1)(1,1,0)_{12}$ ,  $ARIMA(1,1,1)(0,1,1)_{12}$ ,  $ARIMA(1,1,1)(1,1,0)_{12}$  e  $ARIMA(2,1,0)(0,1,1)_{12}$ .

Passa-se, então, para a etapa de estimação dos modelos selecionados.

### 4.3.2 - Estimação

As estimativas de máxima verossimilhança dos modelos selecionados foram obtidas através do software RATS e são as seguintes:

$$(1-B)(1-B^{12})Z_{it} = (1-0,771B)(1-0,663B^{12})\hat{a}_t \quad (4.11)$$

(8,41)	(5,570)
[0,09]	[0,12]

$$(1-0,263B^{12})(1-B)(1-B^{12})Z_{it} = (1-0,830B)\hat{a}_t \quad (4.12)$$

(2,09)	(10,36)
[0,13]	[0,08]

$$(1-0,564B)(1-B)(1-B^{12})Z_{it} = (1-1,148B)(1-0,699B^{12})\hat{a}_t \quad (4.13)$$

(88,73)	(9,42)	(5,81)
[0,006]	[0,12]	[0,12]

$$(1-0,380B^{12}-0,292B^{24})(1-B)(1-B^{12})Z_{it} = (1-0,673B^{12})\hat{a}_t \quad (4.14)$$

(2,74)	(2,16)	(5,68)
[0,14]	[0,14]	[0,12]

$$(1+0,376B)(1-0,367B^{12})(1-B)(1-B^{12})Z_{it} = (1-1,182B)\hat{a}_t \quad (4.15)$$

(3,61)	(2,94)	(10,58)
[0,10]	[0,12]	[0,11]

Os números entre parênteses são os valores absolutos da estatística t dos coeficientes e os entre colchetes são os desvios padrões das estimativas. Os coeficientes da parte autorregressiva (AR) respeitam as condições para a estacionariedade do modelo e os coeficientes da parte de média móveis (MA) respeitam as condições de invertibilidade, nos termos do explicitado no capítulo 2

### 4.3.3 - Checagem de Diagnóstico

Verifica-se que todos os coeficientes estimados são estatisticamente significativos a 5%. Os resíduos para os cinco modelos acima descritos, juntamente com a FACA e a FACPA residuais para cada um deles, foram obtidos pelo RATS, conforme já citado anteriormente. A

FACA residual, em particular, mostra a satisfatoriedade do ajuste para os modelos  $ARIMA(0,1,1)(0,1,1)_{12}$ ,  $ARIMA(1,1,1)(0,1,1)_{12}$  e  $ARIMA(2,1,0)(0,1,1)_{12}$ , indicando, claramente, que todas as autocorrelações residuais são estatisticamente nulas. Já para o caso dos modelos  $ARIMA(0,1,1)(1,1,0)_{12}$  e  $ARIMA(1,1,1)(1,1,0)_{12}$  as autocorrelações residuais para as algumas defasagens se mostraram estatisticamente diferentes de zero, conforme já discutido anteriormente.

Reforçando a satisfatoriedade dos ajustes dos modelos  $ARIMA(0,1,1)(0,1,1)_{12}$  e  $ARIMA(2,1,0)(0,1,1)_{12}$ , a estatística  $\chi^2$  se mostrou insignificante ao nível de 5% para os graus de liberdade de cada modelo e tanto os gráficos dos resíduos padronizados ( $a_t/\sigma_a$ ), traçados nas figuras 2.3.1 e 2.3.2 do Apêndice 2, quanto o teste de normalidade de Jarque-Bera sugerem que os choques devem ser normalmente distribuídos. Devemos observar que tanto no gráfico da figura 2.3.1, quanto no gráfico da figura 2.3.2 nenhum valor foi superior a três vezes o desvio padrão. Isso reforça que os choques apresentam uma distribuição perfeitamente normal.

É importante ressaltar que o segundo termo auto-regressivo não sazonal – AR(2) – do modelo  $ARIMA(2,1,0)(0,1,1)_{12}$  introduz um ciclo estocástico nas previsões. O período médio deste ciclo é de 3,25 meses.<sup>9</sup> Conforme discutido em Bell e Hillmer (1983), quando se encontra um ciclo estocástico em dados mensais acumulados com um período médio de cerca de três meses, isto sugere uma forte evidência de que os dados apresentam variação de calendário (“trading day variation”). Isto pode ser explorado em trabalhos posteriores, visando à melhoria das previsões.

#### 4.3.4 - Previsão

As previsões para os dois modelos selecionados estão mostradas abaixo nas tabelas 4.7 e 4.8, juntamente com o valor da raiz quadrada do erro quadrático médio de previsão (RQEQMP). Conforme se pode observar, foram calculados as RQEQMP para previsões para três, seis e doze passos à frente. A análise dos resultados a ser realizada no capítulo 5 utilizará essa medida para a comparação dos modelos.

---

<sup>9</sup> O período médio de um ciclo estocástico é calculado por:

$$p^* = \frac{360}{\cos^{-1} [f_1 / 2(-f_2)^{1/2}]}, \text{ onde } f_1 \text{ e } f_2 \text{ são os coeficientes dos termos auto-regressivos e } \cos^{-1} \text{ é dado em graus.}$$

Tabela 4 7 – Previsões do modelo ARIMA (0, 1, 1)(0, 1, 1)<sub>12</sub>

MESES	Previsão do Modelo (LIR)	Limite 95% abaixo	Previsão (IR <sub>P</sub> ) milhões de R\$	Limite 95% acima	Valor Real Observado (IR <sub>R</sub> ) milhões de R\$	ERRO ABS. (IR <sub>P</sub> - IR <sub>R</sub> )	ERRO PERC. (%)
JAN/01	8,6226	4548,71	5555,812744	6785,89	6224,726267	-668,91	-10,75
FEV/01	8,4125	3613,75	4503,003961	5611,09	3993,638758	509,37	12,75
MAR/01	8,4258	3662,13	4563,293954	5686,21	5051,099017	-487,81	-9,66
ABR/01	8,6154	4339,00	5515,954554	7012,15	5840,220643	-324,27	-5,55
MAI/01	8,4266	3521,35	4566,946050	5923,01	5213,429848	-646,48	-12,40
JUN/01	8,4188	3356,99	4531,462436	6116,83	5256,170647	-724,71	-13,79
JUL/01	8,5443	3730,51	5137,387702	7074,84	6031,905072	-894,52	-14,83
AGO/01	8,5668	3739,85	5254,289134	7382,00	5758,03735	-503,75	-8,75
SET/01	8,4738	3340,25	4787,674108	6862,31	5324,477202	-536,80	-10,08
OUT/01	8,6598	3943,41	5766,381271	8432,09	6652,519584	-886,14	-13,32
NOV/01	8,3877	2944,52	4392,702850	6553,14	5585,62473	-1192,92	-21,36
DEZ/01	8,7931	4329,04	6588,625288	10027,63	7632,816612	-1044,19	-13,68
TOTAL		45069,49	61163,534052	83467,21	68564,66573	-7401,13	-10,79
RAIZ QUADRADA DO EQM - 3 MESES						561,20	
RAIZ QUADRADA DO EQM - 6 MESES						576,36	
RAIZ QUADRADA DO EQM - 12 MESES						743,52	

Tabela 4 8 – Previsões do modelo ARIMA (2, 1, 0)(0, 1, 1)<sub>12</sub>

MESES	Previsão do Modelo (LIR)	Limite 95% abaixo	Previsão (IR <sub>P</sub> ) em milhões de R\$	Limite 95% acima	Valor Real Observado (IR <sub>R</sub> )	ERRO ABS. (IR <sub>P</sub> - IR <sub>R</sub> )	ERRO PERC. (%)
JAN/01	8,6273	4570,14	5581,986525	6817,85	6224,726267	-642,74	-10,33
FEV/01	8,3956	3553,19	4427,542639	5517,06	3993,638758	433,90	10,86
MAR/01	8,4128	3614,83	4504,355065	5612,77	5051,099017	-546,74	-10,82
ABR/01	8,6222	4368,61	5553,590864	7060,00	5840,220643	-286,63	-4,91
MAI/01	8,4154	3482,13	4516,081626	5857,04	5213,429848	-697,35	-13,38
JUN/01	8,4139	3340,58	4509,312582	6086,94	5256,170647	-746,86	-14,21
JUL/01	8,5423	3723,06	5127,123194	7060,70	6031,905072	-904,78	-15,00
AGO/01	8,4145	3211,52	4512,018981	6339,15	5758,03735	-1246,02	-21,64
SET/01	8,4716	3332,91	4777,152802	6847,23	5324,477202	-547,32	-10,28
OUT/01	8,6612	3948,93	5774,459858	8443,90	6652,519584	-878,06	-13,20
NOV/01	8,3832	2931,30	4372,980097	6523,72	5585,62473	-1212,64	-21,71
DEZ/01	8,7885	4309,17	6558,387213	9981,61	7632,816612	-1074,43	-14,08
TOTAL		36645,80	44386,36	60214,99	82147,98	-8349,67	-12,18
RAIZ QUADRADA DO EQM - 3 MESES						547,82	
RAIZ QUADRADA DO EQM - 6 MESES						581,17	
RAIZ QUADRADA DO EQM - 12 MESES						821,24	

## 4.4 - SÉRIE OUTROS IMPOSTOS (OI)

### 4.4.1 - Identificação

Conforme analisado no capítulo 3, o gráfico da série Outros Impostos ( $Z_{OI}$ ), apresentado na figura 3.13, não mostra uma variância que cresce significativamente com o aumento do nível da série. Desta forma, não sendo necessário estabilizar a variância, não necessitaremos utilizar o logaritmo natural da série original. O gráfico de  $Z_{OI}$  está apresentado na figura 3.15 para o período fev/1995 a dez/2000, que será o utilizado no presente estudo.

Da mesma forma que nas subseções anteriores deste capítulo, para identificação do modelo utilizar-se-á, inicialmente, as Funções de Autocorrelação Amostral (FACA) e de Autocorrelação Parcial Amostral (FACPA). A figura 4.10 mostra a FACA, em preto, e a FACPA, em azul, para  $Z'_{OI}$  com  $d = D = 0$  (sem qualquer diferenciação). Verifica-se que as autocorrelações (FACA) são significativamente diferentes de zero até a defasagem 4, decaindo abruptamente a partir de então. Este comportamento sugere que a série  $Z'_{OI}$  possa ser estacionária, consistente com o gráfico apresentado na figura 3.15, em que os dados não apresentam uma tendência significativa (estacionariedade da média). Resta saber sobre a estacionariedade da variância.

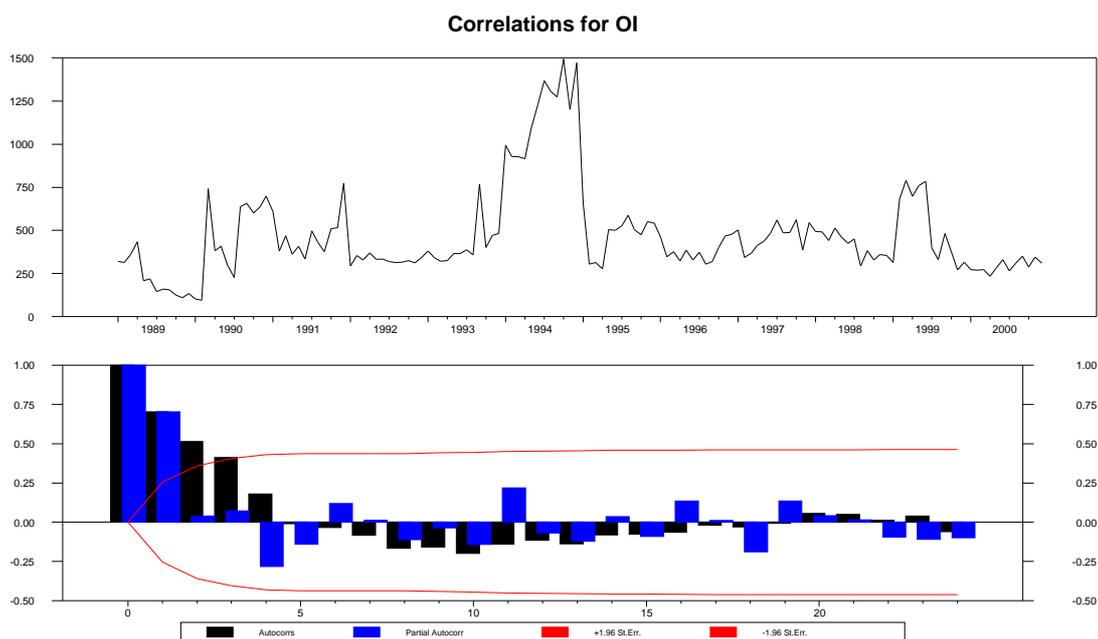


Figura 4.10 – Funções de Autocorrelação Amostral (FACA e FACPA) para a série  $Z_{OI}$

A primeira diferença não sazonal  $W^*_{OI} = \nabla Z_{OI}$  ( $d = 1$ ) está apresentada no gráfico 4.11(superior). Observa-se que esta série tem uma média estável ao longo do tempo e, como relatado no capítulo 2, isto indica que a série diferenciada é estacionária. Neste caso, os dados da figura 4.11(superior) não mostram uma forte evidência de que haja alguma variação sazonal. De qualquer forma, far-se-á uma diferenciação sazonal para verificação do comportamento dessa nova série.

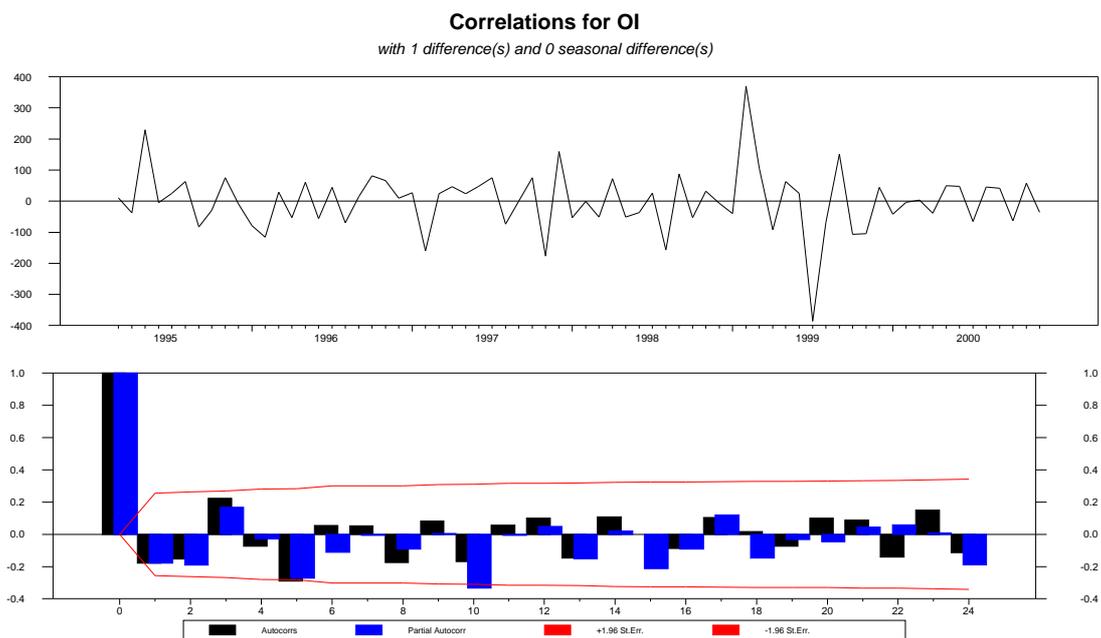


Figura 4.11 – Gráfico da Série Diferenciada  $W^*_{OI} = \nabla Z_{OI}$  ( $d = 1$ ) e suas Funções de Autocorrelação Amostral (FACA e FACPA)

A FACA para a série diferenciada  $W^*_{OI} = \nabla Z'_{OI}$  ( $d = 1$ ) estão apresentadas na figura 4.11(inferior). Verifica-se ali que a autocorrelação cai a valores insignificamente próximos de zero bastante rapidamente. Isto reforça a conclusão do parágrafo anterior de que a série diferenciada é agora estacionária, nos termos do discutido no capítulo 2. Entretanto, as autocorrelações para algumas ainda são altas e se observa um aumento significativo da defasagem 12 para a 24. Isto pode estar sugerindo que os dados apresentam alguma variação sazonal e clama por uma diferenciação sazonal ( $D = 1$ ) em adição à diferenciação não sazonal ( $d = 1$ ), como forma de extinguir a dúvida.

A figura 4.12(superior) apresenta o gráfico dos dados após sofrerem as duas diferenciações, sazonal ( $D = 1$ ) e não sazonal ( $d = 1$ ),  $W_{OI} = \nabla_{12} W^*_{OI} = \nabla_{12} \nabla Z_{OI}$ . Esta nova

série,  $W_{OI}$ , apresenta uma média estável, mas apresenta um padrão sazonal superior ao observado na série  $W^*_{OI}$ . A FACA e a FACPA para  $W_{OI}$  estão apresentadas na figura 4.12(inferior), em preto e azul, respectivamente. Tudo indica que não seja necessária uma diferenciação sazonal para a série em estudo.

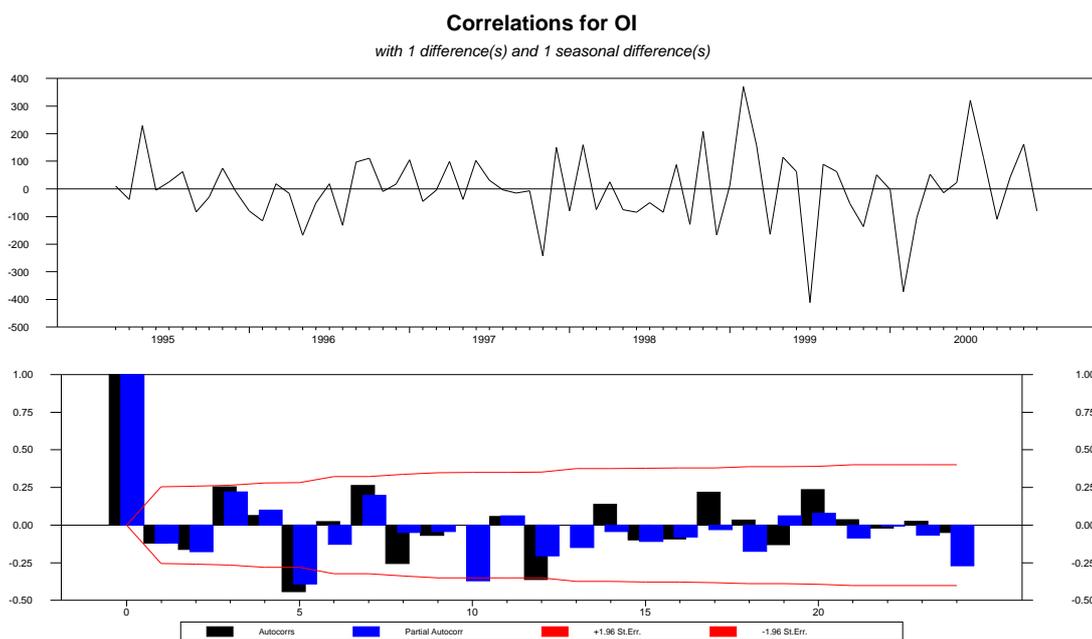


Figura 4.12 – Gráfico da Série Diferenciada  $W_{OI} = \nabla_{12}\nabla Z_{OI}$  ( $d = 1, D = 1$ ) e suas Funções de Autocorrelação Amostral (FACA e FACPA)

Assim, comparando-se a FACA e a FACPA da série  $W^*_{OI}$  com as FAC e FACP teóricas apresentadas no capítulo 2, conclui-se que há evidências de outros padrões não sazonais que sugerem a inclusão de componentes de média móveis (MA) e/ou auto-regressivos (AR), mas nada que justifique a inclusão de outros termos sazonais além de, no máximo, uma única diferenciação.

**Modelo 1 - ARIMA(0, 1, 1)**

As autocorrelações residuais para os modelos em análise foram obtidas pelo RATS (não mostradas). Observa-se correlações residuais não significativamente diferentes de zero (autocorrelações das defasagens 5 e 10), indicando que talvez este modelo não seja adequado para representar a série em estudo. Tentemos a inclusão de um termo auto-regressivo não sazonal - AR(1).

**Modelo 2 - ARIMA(1, 1, 0)**

Ainda se observa correlações residuais que se mostram não significativamente diferentes de zero (autocorrelações das defasagens 5 e 10), mas se pode verificar se este modelo é adequado para representar a série em estudo. Tentemos incluir um termo de média móvel não sazonal ao modelo em questão.

**Modelo 3 - ARIMA(1, 1, 1)**

Ainda se observa correlações residuais que se mostram não significativamente diferentes de zero (autocorrelações das defasagens 5 e 10 novamente), mas se pode verificar se este modelo é adequado para representar a série em estudo. Tentemos incluir mais um termo auto-regressivo não sazonal ao modelo em questão.

**Modelo 4 - ARIMA(2, 1, 1)**

Agora se observa que as correlações residuais se mostram significativamente diferentes de zero (apenas a autocorrelação da defasagem 5 está alta). Então, pode-se verificar se este modelo também se apresenta como adequado para representar a série em estudo. Tentemos incluir mais um termo de média móvel não sazonal ao modelo 3.

**Modelo 5 - ARIMA(1, 1, 2)**

Mais uma vez se observa que as correlações residuais se mostram significativamente diferentes de zero, mas a persistência de uma alta correlação na 5ª defasagem pode estar clamando por uma diferenciação sazonal ou a inclusão de um termo AR(5). Tentemos inicialmente a diferenciação.

**Modelo 6 - ARIMA(0, 1, 1)(0, 1, 0)<sub>12</sub>**

Verifica-se uma piora na situação, indicando que devemos realmente abandonar a diferenciação sazonal. Tentemos a inclusão de um termo auto-regressivo na 5ª defasagem.

**Modelo 7 - ARIMA(//1,5//, 1, 0)<sup>10</sup>**

Apesar de ser um modelo não muito parcimonioso, suas funções de autocorrelação se mostraram bem comportadas, indicando que podemos investir neste modelo como um dos que pode representar adequadamente os dados em questão.

---

<sup>10</sup> Trata-se da representação do RATS para o modelo  $(1 - a_1 B - a_5 B^5)(1 - B).Z_{01} = (1 - b_1 B).a_t$ .

Para se determinar quais os modelos que melhor se ajustam aos dados, optou-se por utilizar os critérios de seleção de Akaike e de Schwartz, quais sejam, o *Akaike Information Criterion*(AIC) e *Schwartz Bayesian Criterion* (SBC ou BIC).

Os valores dos critérios AIC e BIC para cada um dos modelos estudados (d = 1; p = 0, 1, 2 e //5//; q = 0, 1, 2;) estão apresentados na tabela 4.9. Como o critério BIC seleciona modelos mais parcimoniosos que o AIC, utilizaremos aquele para a escolha dos três melhores modelos.

Tabela 4.9 – Valores do A.I.C e do B.I.C para a Série Outros Impostos (OI)

MODELO	AIC	BIC	MODELO	AIC	BIC
<b>(0,1,1)</b>	<b>702</b>	<b>704</b>	(0,1,1)(0,1,0) <sub>12</sub>	735	737
<b>(1,1,0)</b>	<b>702</b>	<b>704</b>	(1,1,0) (0,1,0) <sub>12</sub>	735	737
(1,1,1)	704	708	(1,1,1) (0,1,0) <sub>12</sub>	729	733
(2,1,1)	702	707	(2,1,1) (0,1,0) <sub>12</sub>	737	743
(1,1,2)	701	707	(1,1,2) (0,1,0) <sub>12</sub>	737	743
(//1,5//,1,1)	700	706	(//1,5//,1,1) (0,1,0) <sub>12</sub>	726	732
<b>(//1,5//,1,0)</b>	<b>700</b>	<b>703</b>	(//1,5//,1,0) (0,1,0) <sub>12</sub>	725	729
(//1,5//,1,2)	701	709	(//1,5//,1,2) (0,1,0) <sub>12</sub>	728	735

Assim, analisando a tabela 4.9, observa-se que os modelos que apresentam menores valores para o BIC são os seguintes: ARIMA(0,1,1), ARIMA(1,1,0) e ARIMA(//1,5//,1,0).

Passa-se, então, para a etapa de estimação dos modelos selecionados.

#### 4.4.2 - Estimação

As estimativas de máxima verossimilhança dos três modelos selecionados foram obtidas através do software RATS e são as seguintes:

$$(1 - B)Z_{it} = (1 - 0,222B) \hat{a}_t \tag{4.16}$$

(1,82)  
[0,12]

$$(1 - 0,162B)(1 - B)Z_{it} = \hat{a}_t \tag{4.17}$$

(1,31)  
[0,12]

$$(1 - 0,186B - 0,275B^5)(1 - B)Z_{it} = \hat{a}_t \tag{4.18}$$

(1,86) (2,38)  
[0,12] [0,12]

Os números entre parênteses são os valores absolutos da estatística  $t$  dos coeficientes e os entre colchetes são os desvios padrões das estimativas. Os coeficientes da parte autorregressiva (AR) respeitam as condições para a estacionariedade do modelo e os coeficientes da parte de média móveis (MA) respeitam as condições de invertibilidade, nos termos do explicitado no capítulo 2.

#### 4.4.3 - Checagem de Diagnóstico

Verifica-se que nem todos os coeficientes estimados são estatisticamente significativos a 5%. Este é o caso do coeficiente do 2º modelo. Resta, para análise, somente o primeiro e o terceiro modelos. Os resíduos para estes modelos, juntamente com a FACA e a FACPA residuais, foram obtidos pelo RATS, conforme mencionado anteriormente. Para o primeiro modelo a FACA residual, em particular, mostra uma alta correlação na 5ª defasagem, indicando que o modelo não deve representar bem os dados em questão, conforme já discutido anteriormente. Resta, portanto, apenas o 3º modelo, após toda essa árdua tarefa de identificação e estimação. A FACA e a FACPA parecem mostrar que o ajuste com este modelo em particular é satisfatório. Utilizaremos-lo, pois, para fins de previsão, cuja análise minuciosa será desenvolvida no próximo capítulo.

Para reforçar a possível satisfatoriedade do ajuste do modelo  $ARIMA(1,1,1, 1, 0)$ , a estatística  $\chi^2$  se mostrou insignificante ao nível de 5% para os graus de liberdade adotado e tanto os gráficos dos resíduos padronizados ( $a_t/\sigma_a$ ), traçados na figura 2.4.1 do Apêndice 2, quanto o teste de normalidade de Jarque-Bera ( $\chi^2 = 62,48$ ) sugerem que os choques devem ser normalmente distribuídos. Devemos observar que no gráfico da figura 2.4.1 somente dois valores foram superiores a três vezes o desvio padrão (fevereiro e julho/1999). Em uma distribuição perfeitamente normal isso não ocorre, mas a justificativa pode estar nos "outliers" que não devem ter sido removidos adequadamente pelo processo escolhido.

#### 4.4.4 – Previsão

As previsões para o modelo selecionado estão mostradas abaixo na tabela 4.10, juntamente com o valor da raiz quadrada do erro quadrático médio de previsão (RQEQMP). Conforme se pode observar, foram calculados as RQEQMP para previsões para três, seis e doze passos à frente. A análise dos resultados a ser realizada no capítulo 5 utilizará essa medida para a comparação dos modelos.

Tabela 4 10 – Previsões do modelo ARIMA (//1,5//, 1, 0)

MESES	Limite 95% abaixo	Previsão (I <sub>p</sub> ) milhões de R\$	Limite 95% acima	Valor Real Observado (I <sub>R</sub> ) milhões de R\$	ERRO ABS. (I <sub>p</sub> - I <sub>R</sub> )	ERRO PERC. (%)
JAN/01	265,37	303,370000	341,37	332,1853622	-28,82	-8,67
FEV/01	247,14	293,140000	339,14	217,9645026	75,18	34,49
MAR/01	258,41	312,410000	366,41	327,0420982	-14,63	-4,47
ABR/01	234,91	292,910000	350,91	282,7306577	10,18	3,60
MAI/01	244,36	306,360000	368,36	354,9733344	-48,61	-13,69
JUN/01	235,44	305,440000	375,44	302,6540003	2,79	0,92
JUL/01	232,42	308,420000	384,42	310,6757221	-2,26	-0,73
AGO/01	224,57	302,570000	380,57	370,4110413	-67,84	-18,32
SET/01	227,02	309,020000	391,02	421,0535141	-112,03	-26,61
OUT/01	218,12	304,120000	390,12	398,2988742	-94,18	-23,65
NOV/01	215,29	305,290000	395,29	349,0111388	-43,72	-12,53
DEZ/01	208,25	304,250000	400,25	362,1735324	-57,92	-15,99
TOTAL		3647,300000	-	4029,173778	-381,87	-9,48
RAIZ QUADRADA DO EQM - 3 MESES						47,24
RAIZ QUADRADA DO EQM - 6 MESES						39,09
RAIZ QUADRADA DO EQM - 12 MESES						58,07

### 4.5 - SÉRIE RECEITA TOTAL DE IMPOSTOS (RIMP)

#### 4.5.1 - Identificação

O gráfico da série Receita Total de Impostos ( $Z_{RIMP}$ ), apresentado na figura 3.16, mostra uma variância que cresce com o aumento do nível da série. Conforme já efetuado em subseções anteriores, uma forma eficiente de estabilizar a variância consiste em utilizar o logaritmo natural da série original, de tal forma que o modelo ARIMA seja construído para a série transformada  $Z'_{RIMP} = \ln(Z_{RIMP})$ . O gráfico de  $Z'_{RIMP}$  está apresentado na figura 3.17 para o período jan/1989 a dez/2000 e na figura 3.19 para o período jul/1994 a dez/2000, que será o utilizado no presente estudo.

Para identificação do modelo utilizar-se-á as Funções de Autocorrelação Amostral e de Autocorrelação Parcial Amostral. A figura 4.13 mostra a FACA e a FACPA para  $Z'_{RIMP}$  sem qualquer diferenciação. Verifica-se que as autocorrelações (FACA) são significativamente diferentes de zero para diversas defasagens, decaindo abruptamente para alguns valores, indicando um padrão desordenado. Este comportamento sugere uma forte evidência de que a série  $Z'_{RIMP}$  seja não-estacionária, consistente com o gráfico apresentado na figura 3.17, em que os dados apresentam uma pequena tendência ascendente no período pós Real..

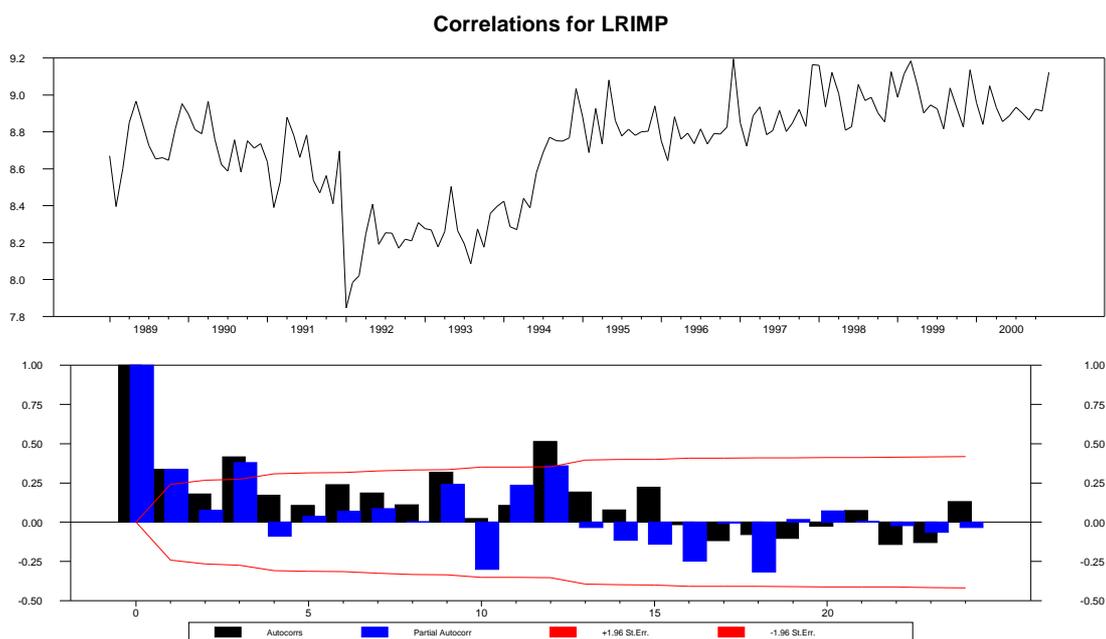


Figura 4.13 – Funções de Autocorrelação Amostral (FACA e FACPA) para a série  $Z'_{RIMP}$

A primeira diferença não sazonal  $W^*_{RIMP} = \nabla Z'_{RIMP}$  ( $d = 1$ ) está apresentada no gráfico 4.14(superior). Observa-se que esta série tem uma média estável ao longo do tempo e, como relatado no capítulo 2, isto indica que a série diferenciada é estacionária. Entretanto, os dados da figura 4.14(superior) mostram uma forte variação sazonal.

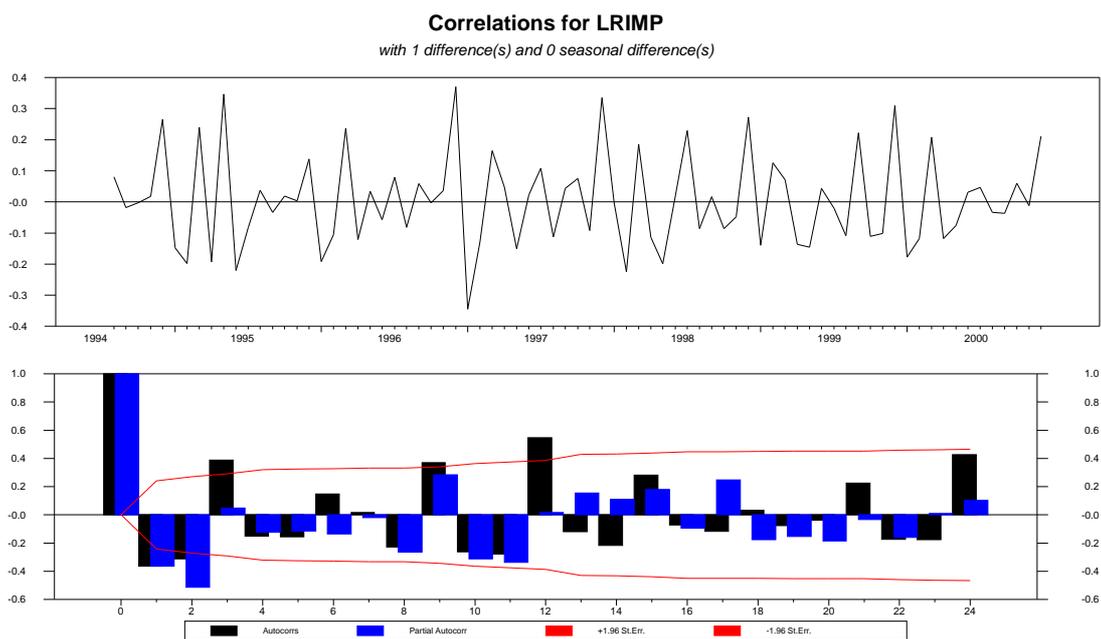


Figura 4.14 – Gráfico da Série Diferenciada  $W^*_{RIMP} = \nabla Z'_{RIMP}$  ( $d = 1$ ) e suas Funções de Autocorrelação Amostral (FACA e FACPA)

A FACA para a série diferenciada  $W^*_{RIMP} = \nabla Z'_{RIMP}$  ( $d = 1$ ) estão apresentadas na figura 4.14(inferior). Verifica-se ali que a autocorrelação cai a valores insignificamente próximos de zero bastante rapidamente, exceto para algumas poucas defasagens. Isto reforça a conclusão do parágrafo anterior de que a série diferenciada é agora estacionária, nos termos do discutido no capítulo 2. Entretanto, as autocorrelações para as defasagens 12 e 24 são altas e não se observa uma queda significativa da defasagem 12 para a 24. Isto reforça a observação anteriormente feita de que os dados apresentam forte variação sazonal e sugere uma diferenciação sazonal ( $D = 1$ ) em adição à diferenciação não sazonal ( $d = 1$ ).

A figura 4.15 (superior) apresenta o gráfico dos dados após sofrerem as duas diferenciações, sazonal ( $D = 1$ ) e não sazonal ( $d = 1$ ),  $W_{RIMP} = \nabla_{12} W^*_{RIMP} = \nabla_{12} \nabla Z'_{RIMP}$ . Esta nova série,  $W_{RIMP}$ , apresenta uma média estável e não apresenta mais o forte padrão sazonal observado na série  $W^*_{RIMP}$ . A FACA e a FACPA para  $W_{RIMP}$  estão apresentadas na figura 4.15 (inferior).

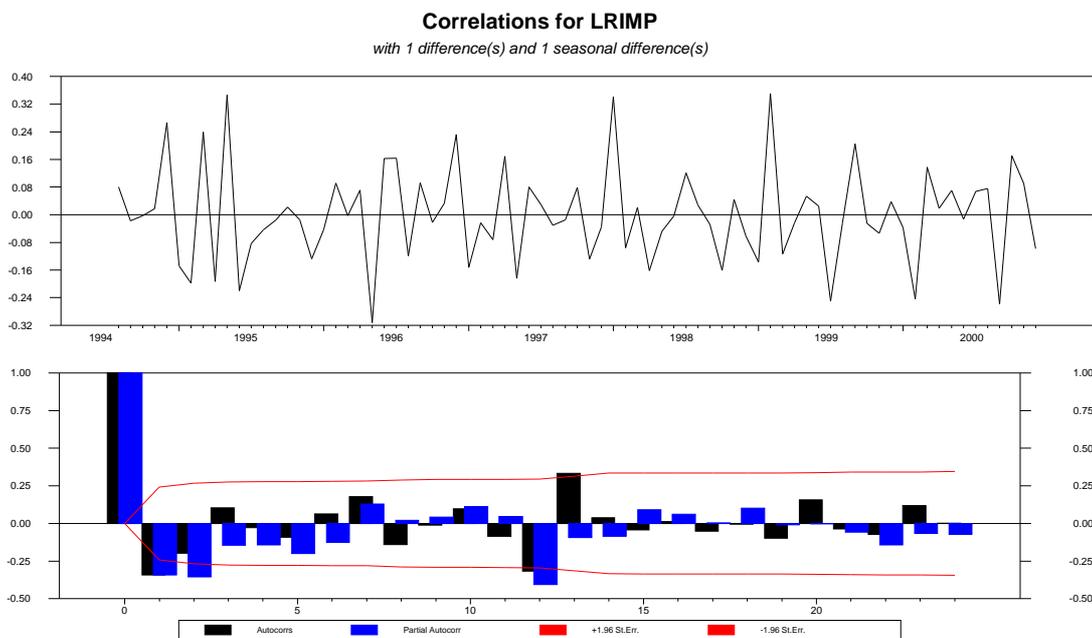


Figura 4.15 – Gráfico da Série Diferenciada  $W_{RIMP} = \nabla_{12} \nabla Z_{RIMP}$  ( $d = 1, D = 1$ ) e suas Funções de Autocorrelação Amostral (FACA e FACPA)

Comparando-se a FACA e a FACPA da série  $W_{RIMP}$  com as FAC e FACP teóricas apresentadas no capítulo 2, conclui-se que remanescem padrões sazonais e não sazonais que sugerem a inclusão de componentes de média móvels (MA) e/ou auto-regressivos (AR). No caso da parte não sazonal do modelo, as funções de autocorrelação sugerem a inclusão de um termo de média móvel - MA(1) - e/ou um ou dois termos auto-regressivos - AR(1) e AR(2). No que se refere a parte sazonal do modelo, as funções de autocorrelação sugerem a inclusão de um termo auto-regressivo - SAR(1) ou de um termo de média móvel - SMA(1). Assim, combinando os componentes supracitados, poder-se-ia, inicialmente, analisar as seguintes opções de modelos:

**Modelo 1 - ARIMA(0, 1, 0)(0, 1, 1)<sub>12</sub>**

As autocorrelações residuais para os modelos em análise foram obtidas pelo RATS (não mostradas). Observa-se correlações residuais não significativamente diferentes de zero (autocorrelações das defasagens 1 e 2), indicando não ser este um modelo adequado para representar a série em estudo. Tentemos a inclusão de um termo auto-regressivo não sazonal - AR(1) e/ou um termo de média móvel não sazonal - MA(1), conforme parece indicar as autocorrelações na primeira defasagem.

**Modelo 2 - ARIMA(1, 1, 0)(0, 1, 1)<sub>12</sub>**

Ainda se observa correlações residuais não significativamente diferentes de zero (autocorrelações da defasagem 2), indicando não ser este, ainda, um modelo adequado para representar a série em estudo. Tentemos incluir um termo de média móvel não sazonal.

**Modelo 3 - ARIMA(0, 1, 1)(0, 1, 1)<sub>12</sub>**

Agora, todas as correlações residuais são significativamente diferentes de zero, indicando ser este um potencial modelo a representar a série em estudo. Verifiquemos a inclusão de um termo auto-regressivo sazonal.

**Modelo 4 - ARIMA(0, 1, 1)(1, 1, 0)<sub>12</sub>**

Aqui se observa uma correlação residual não significativamente diferente de zero (autocorrelação da defasagem 7, em preto), indicando que, provavelmente, este modelo não seja adequado para representar a série em estudo. Tentemos incluir um termo auto-regressivo não sazonal a este modelo.

**Modelo 5 - ARIMA(1, 1, 1)(1, 1, 0)<sub>12</sub>**

Aqui se observa correlações residuais não significativamente diferentes de zero (autocorrelações das defasagens 7 e 13, por exemplo), indicando que, provavelmente, este modelo também não seja adequado para representar a série em estudo. Tentemos incluir um termo auto-regressivo não sazonal a este modelo. Tentemos o seguinte:

**Modelo 6 - ARIMA(1, 1, 1)(0, 1, 1)<sub>12</sub>**

Agora, mais uma vez se verifica que todas as correlações residuais são significativamente diferentes de zero, indicando ser este um outro potencial modelo a representar a série em estudo. Por fim tentemos a inclusão de um segundo termo auto-regressivo ao modelo 2.

**Modelo 7 - ARIMA(2, 1, 0)(0, 1, 1)<sub>12</sub>**

Aqui, novamente se verifica que todas as correlações residuais são significativamente diferentes de zero, indicando ser este um outro potencial modelo a representar a série em estudo.

Para se determinar quais os modelos que melhor se ajustam aos dados, optou-se por utilizar os critérios de seleção de Akaike e de Schwartz, quais sejam, o *Akaike Information Criterion (AIC)* e *Schwartz Bayesian Criterion (SBC ou BIC)*.

Os valores dos critérios AIC e BIC para cada um dos modelos estudados ( $d = 1; D = 1; p = 0, 1, 2; q = 0, 1; P = 0, 1; e Q = 0, 1$ ) estão apresentados na tabela 4.11.

Tabela 4.11 – Valores do AIC e do BIC para a Série Total de Impostos (RIMP)

MODELO	AIC	BIC	MODELO	AIC	BIC
$(0,1,0)(0,1,0)_{12}$	-7,62	-7,62	$(1,1,0)(1,1,0)_{12}$	-17,32	-13,46
$(0,1,0)(0,1,1)_{12}$	-28,43	-26,50	$(1,1,1)(0,1,0)_{12}$	-24,65	-20,74
$(0,1,0)(1,1,0)_{12}$	-13,77	-11,84	<b><math>(1,1,1)(0,1,1)_{12}</math></b>	<b>-39,20</b>	<b>-33,41</b>
$(0,1,1)(0,1,0)_{12}$	-24,81	-22,88	$(1,1,1)(1,1,0)_{12}$	-27,48	-21,68
<b><math>(0,1,1)(0,1,1)_{12}</math></b>	<b>-37,27</b>	<b>-33,40</b>	$(2,1,0)(0,1,0)_{12}$	-16,81	-12,95
$(0,1,1)(1,1,0)_{12}$	-27,16	-23,30	<b><math>(2,1,0)(0,1,1)_{12}</math></b>	<b>-32,34</b>	<b>-26,55</b>
<b><math>(1,1,0)(0,1,1)_{12}</math></b>	<b>-30,62</b>	<b>-26,75</b>	$(2,1,0)(1,1,0)_{12}$	-21,26	-15,47

Assim, analisando a tabela 4.11, observa-se que os quatro modelos que apresentam menores valores para o BIC são os seguintes: ARIMA(0,1,1)(0,1,1)<sub>12</sub>, ARIMA(1,1,0)(1,1,0)<sub>12</sub>, ARIMA(1,1,1)(0,1,1)<sub>12</sub> e ARIMA (2,1,0)(0,1,1)<sub>12</sub>.

Passa-se, então, para a etapa de estimação dos modelos selecionados.

#### 4.5.2 - Estimação

As estimativas de máxima verossimilhança dos modelos selecionados foram obtidas através do software RATS e são as seguintes:

$$(1-B)(1-B^{12})Z_{It} = (1-0,737B)(1-0,699B^{12})\hat{a}_t \tag{4.19}$$

(7,45)      (6,00)  
[0,10]      [0,12]

$$(1+0,295B)(1-B)(1-B^{12})Z_{It} = (1-0,857B)(1-0,715B^{12})\hat{a}_t \tag{4.20}$$

(1,99)                      (9,15)      (6,16)  
[0,15]                      [0,09]      [0,12]

$$(1-0,276B)(1-B)(1-B^{12})Z_{It} = (1-0,752B^{12})\hat{a}_t \tag{4.21}$$

(2,04)                                      (6,75)  
[0,14]                                      [0,11]

$$(1-0,360B^{12}-0,263B^{24})(1-B)(1-B^{12})Z_{It} = (1-0,707B)\hat{a}_t \tag{4.22}$$

(2,59)      (1,94)                                      (6,04)  
[0,14]      [0,14]                                      [0,12]

Os números entre parênteses são os valores absolutos da estatística  $t$  dos coeficientes e os entre colchetes são os desvios padrões das estimativas. Os coeficientes da parte autorregressiva (AR) respeitam as condições para a estacionariedade do modelo e os coeficientes da parte de média móveis (MA) respeitam as condições de invertibilidade, nos termos do explicitado no capítulo 2.

#### 4.5.3 - Checagem de Diagnóstico

Verifica-se que todos os coeficientes estimados são estatisticamente significativos a 5%. Os resíduos para os quatro modelos acima descritos, juntamente com a FACA e a FACPA residuais para cada um deles, foram obtidos pelo RATS, conforme já mencionado. A FACA residual, em particular, mostra a satisfatoriedade do ajuste para os modelos  $ARIMA(0,1,1)(0,1,1)_{12}$ ,  $ARIMA(1,1,1)(0,1,1)_{12}$  e  $ARIMA(2,1,0)(0,1,1)_{12}$ , indicando, claramente, que todas as autocorrelações residuais são estatisticamente nulas. Já para o caso do modelo  $ARIMA(1,1,0)(0,1,1)_{12}$  as autocorrelações residuais para as algumas defasagens se mostraram estatisticamente diferentes de zero, conforme já discutido anteriormente.

Reforçando a satisfatoriedade dos ajustes dos modelos  $ARIMA(0,1,1)(0,1,1)_{12}$ ,  $ARIMA(1,1,1)(0,1,1)_{12}$  e  $ARIMA(2,1,0)(0,1,1)_{12}$ , a estatística  $\chi^2$  se mostrou insignificante ao nível de 5% para os graus de liberdade de cada modelo e tanto os gráficos dos resíduos padronizados ( $a_t/\sigma_a$ ), traçados nas figuras 2.5.1, 2.5.2 e 2.5.3 do Apêndice 2, quanto o teste de normalidade de Jarque-Bera sugerem que os choques devem ser normalmente distribuídos. Devemos observar que tanto no gráfico da figura 2.5.1, quanto nos demais (figura 2.5.2 e 2.5.3) nenhum valor foi superior a três vezes o desvio padrão. Isso reforça que os choques apresentam uma distribuição perfeitamente normal.

#### 4.5.4 - Previsão

As previsões para os três modelos selecionados estão mostradas abaixo nas tabelas 4.12 a 4.14, juntamente com o valor da raiz quadrada do erro quadrático médio de previsão (RQEQMP). Conforme se pode observar, foram calculados as RQEQMP para previsões para três, seis e doze passos à frente. A análise dos resultados a ser realizada no capítulo 5 utilizará essa medida para a comparação dos modelos.

Tabela 4 12 – Previsões do modelo ARIMA (0, 1, 1)(0, 1, 1)<sub>12</sub>

MESES	Previsão do Modelo (LRIMP)	Limite 95% abaixo	Previsão (I <sub>P</sub> ) milhões de R\$	Limite 95% acima	Valor Real Observado (I <sub>R</sub> ) milhões de R\$	ERRO ABS. (I <sub>P</sub> - I <sub>R</sub> )	ERRO PERC. (%)
JAN/01	9,1283	7393,08	9212,347659	11479,29	8758,027655	454,32	5,19
FEV/01	8,8400	5431,66	6904,992640	8777,97	6426,119445	478,87	7,45
MAR/01	8,9102	5711,29	7407,142379	9606,55	7728,266353	-321,12	-4,16
ABR/01	8,9809	5889,34	7949,783702	10731,09	8684,941686	-735,16	-8,46
MAI/01	8,9073	5363,11	7385,692783	10171,04	8206,27663	-820,58	-10,00
JUN/01	8,9093	5267,44	7400,478949	10397,29	8173,644274	-773,17	-9,46
JUL/01	8,9677	5473,64	7845,536105	11245,24	8974,271795	-1128,74	-12,58
AGO/01	8,9035	5031,63	7357,680407	10759,02	8789,652631	-1431,97	-16,29
SET/01	8,9493	5163,14	7702,498225	11490,78	8264,267245	-561,77	-6,80
OUT/01	9,0987	5876,40	8943,658387	13611,90	9725,168279	-781,51	-8,04
NOV/01	8,9071	4755,70	7384,215792	11465,53	8463,08996	-1078,87	-12,75
DEZ/01	9,1628	6141,35	9535,719748	14806,18	10504,75255	-969,03	-9,22
TOTAL		67675,56	67497,80	95029,75	134541,86	-7668,73	-7,47
RAIZ QUADRADA DO EQM - 3 MESES						423,81	
RAIZ QUADRADA DO EQM - 6 MESES						625,89	
RAIZ QUADRADA DO EQM - 12 MESES						851,68	

Tabela 4 13 – Previsões do modelo ARIMA (1, 1, 1)(0, 1, 1)<sub>12</sub>

MESES	Previsão do Modelo (LRIMP)	Limite 95% abaixo	Previsão (I <sub>P</sub> ) em milhões de R\$	Limite 95% acima	Valor Real Observado (I <sub>R</sub> )	ERRO ABS. (I <sub>P</sub> - I <sub>R</sub> )	ERRO PERC. (%)
JAN/01	9,1365	7453,95	9288,199477	11573,81	8758,027655	530,17	6,05
FEV/01	8,8395	5428,94	6901,541007	8773,58	6426,119445	475,42	7,40
MAR/01	8,9098	5709,00	7404,180114	9602,70	7728,266353	-324,09	-4,19
ABR/01	8,9840	5907,63	7974,466270	10764,40	8684,941686	-710,48	-8,18
MAI/01	8,9098	5376,54	7404,180114	10196,50	8206,27663	-802,10	-9,77
JUN/01	8,9124	5283,80	7423,456030	10429,57	8173,644274	-750,19	-9,18
JUL/01	8,9710	5491,74	7871,469140	11282,41	8974,271795	-1102,80	-12,29
AGO/01	8,9066	5047,26	7380,524607	10792,43	8789,652631	-1409,13	-16,03
SET/01	8,9526	5180,21	7727,958455	11528,76	8264,267245	-536,31	-6,49
OUT/01	9,1025	5898,78	8977,708944	13663,73	9725,168279	-747,46	-7,69
NOV/01	8,9100	4769,52	7405,661098	11498,82	8463,08996	-1057,43	-12,49
DEZ/01	9,1660	6161,03	9566,282926	14853,64	10504,75255	-938,47	-8,93
TOTAL		67660,41	67708,39	95325,63	134960,35	-7372,85	-7,18
RAIZ QUADRADA DO EQM - 3 MESES						451,72	
RAIZ QUADRADA DO EQM - 6 MESES						622,23	
RAIZ QUADRADA DO EQM - 12 MESES						835,17	

Tabela 4 14 – Previsões do modelo ARIMA (2, 1, 0)(0, 1, 1)<sub>12</sub>

MESES	Previsão do Modelo (LRIMP)	Limite 95% abaixo	Previsão (I <sub>p</sub> ) em milhões de R\$	Limite 95% acima	Valor Real Observado (I <sub>R</sub> )	ERRO ABS. (I <sub>p</sub> - I <sub>R</sub> )	ERRO PERC. (%)
JAN/01	9,1301	7406,40	9228,944818	11499,97	8758,027655	470,92	5,38
FEV/01	8,8302	5378,69	6837,654210	8692,36	6426,119445	411,53	6,40
MAR/01	8,9205	5770,42	7483,830209	9706,00	7728,266353	-244,44	-3,16
ABR/01	8,9832	5902,91	7968,089248	10755,80	8684,941686	-716,85	-8,25
MAI/01	8,9079	5366,33	7390,125528	10177,15	8206,27663	-816,15	-9,95
JUN/01	8,9109	5275,88	7412,329193	10413,93	8173,644274	-761,32	-9,31
JUL/01	8,9696	5484,05	7860,456794	11266,62	8974,271795	-1113,82	-12,41
AGO/01	8,9548	5296,49	7744,978679	11325,36	8789,652631	-1044,67	-11,89
SET/01	8,9513	5173,48	7717,918637	11513,78	8264,267245	-546,35	-6,61
OUT/01	9,0897	5823,75	8863,526595	13489,95	9725,168279	-861,64	-8,86
NOV/01	8,9084	4761,89	7393,821515	11480,44	8463,08996	-1069,27	-12,63
DEZ/01	9,1645	6151,80	9551,944258	14831,37	10504,75255	-952,81	-9,07
TOTAL		67743,02	67792,09	95453,6196 84	135152,74	-7244,86	-7,05
RAIZ QUADRADA DO EQM - 3 MESES						387,67	
RAIZ QUADRADA DO EQM - 6 MESES						606,97	
RAIZ QUADRADA DO EQM - 12 MESES						797,63	

## 4.6 - SÉRIE DA CPMF

### 4.6.1 - Identificação

Conforme visto no capítulo 3, o gráfico da série da Contribuição Provisória sobre Movimentação Financeira ( $Z_{CPMF}$ ), apresentado na figura 3.20, também mostra uma variância que cresce com o aumento do nível da série. Conforme já cansativamente comentado nas seções anteriores, uma forma eficiente de estabilizar a variância consiste em utilizar o logaritmo natural da série original, de tal forma que o modelo ARIMA seja construído para a série transformada  $Z'_{CPMF} = \ln(Z_{CPMF})$ . O gráfico de  $Z'_{CPMF}$  está apresentado na figura 3.22 para o período fev/1997 a dez/2000, que será o utilizado no presente estudo.

Para identificação do modelo utilizar-se-á, inicialmente, as Funções de Autocorrelação Amostral (FACA) e de Autocorrelação Parcial Amostral (FACPA). A figura 4.16 mostra a FACA, em preto, e a FACPA, em azul, para  $Z'_{CPMF}$  com  $d = D = 0$  (sem qualquer diferenciação). Verifica-se que as autocorrelações (FACA) são significativamente diferentes de zero para diversas defasagens, decaindo lentamente. Este comportamento sugere uma forte evidência de que a série  $Z'_{CPMF}$  seja não-estacionária, consistente com o gráfico apresentado na figura 3.22, em que os dados apresentam uma pequena tendência ascendente.

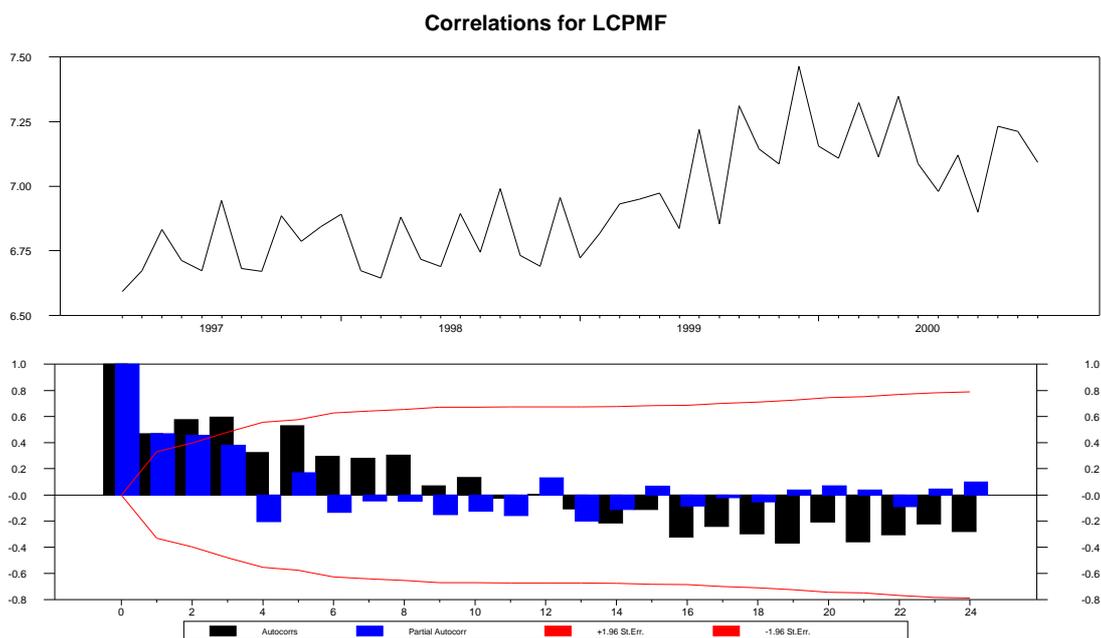


Figura 4.16 – Funções de Autocorrelação Amostral (FACA e FACPA) para a série  $Z'_{CPMF}$

A primeira diferença não sazonal  $W^*_{CPMF} = \nabla Z'_{CPMF}$  ( $d = 1$ ) está apresentada no gráfico 4.17(superior). Observa-se que esta série tem uma média estável ao longo do tempo e, como relatado no capítulo 2, isto indica que a série diferenciada é estacionária. Entretanto, os dados da figura 4.17 (superior) mostram que pode haver alguma variação sazonal.

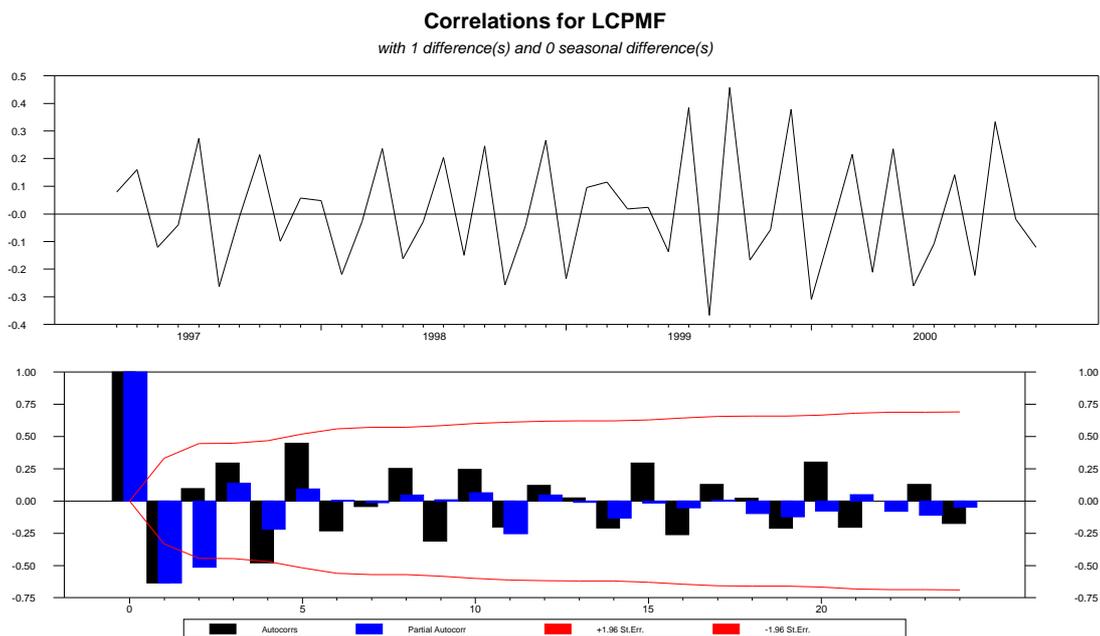


Figura 4.17 – Gráfico da Série Diferenciada  $W^*_{CPMF} = \nabla Z'_{CPMF}$  ( $d = 1$ ) e suas Funções de Autocorrelação Amostral (FACA e FACPA)

A FACA e a FACPA para a série diferenciada  $W^*_{CPMF} = \nabla Z'_{CPMF}$  ( $d = 1$ ) são mostradas na figura 4.17(inferior). Verifica-se ali que a autocorrelação cai a valores insignificamente próximos de zero bastante rapidamente. Isto reforça a conclusão do parágrafo anterior de que a série diferenciada é agora estacionária, nos termos do discutido no capítulo 2. Entretanto, as autocorrelações para algumas defasagens são altas, com picos alternados em intervalos mais ou menos equidistantes. Isto reforça a observação anteriormente feita de que os dados podem apresentar alguma variação sazonal e sugere verificar se uma diferenciação sazonal ( $D = 1$ ), em adição à diferenciação não sazonal ( $d = 1$ ), melhoraria o ajuste.

A figura 4.18(superior) apresenta o gráfico dos dados após sofrerem as duas diferenciações, sazonal ( $D = 1$ ) e não sazonal ( $d = 1$ ),  $W_{CPMF} = \nabla_{12} W^*_{CPMF} = \nabla_{12} \nabla Z'_{CPMF}$ . Esta nova série,  $W_{CPMF}$ , apresenta uma média estável e não apresenta mais um padrão sazonal

observado na série  $W_{CPMF}^*$ . A FACA e a FACPA para  $W_{CPMF}$  estão apresentadas na figura 4.18(inferior).

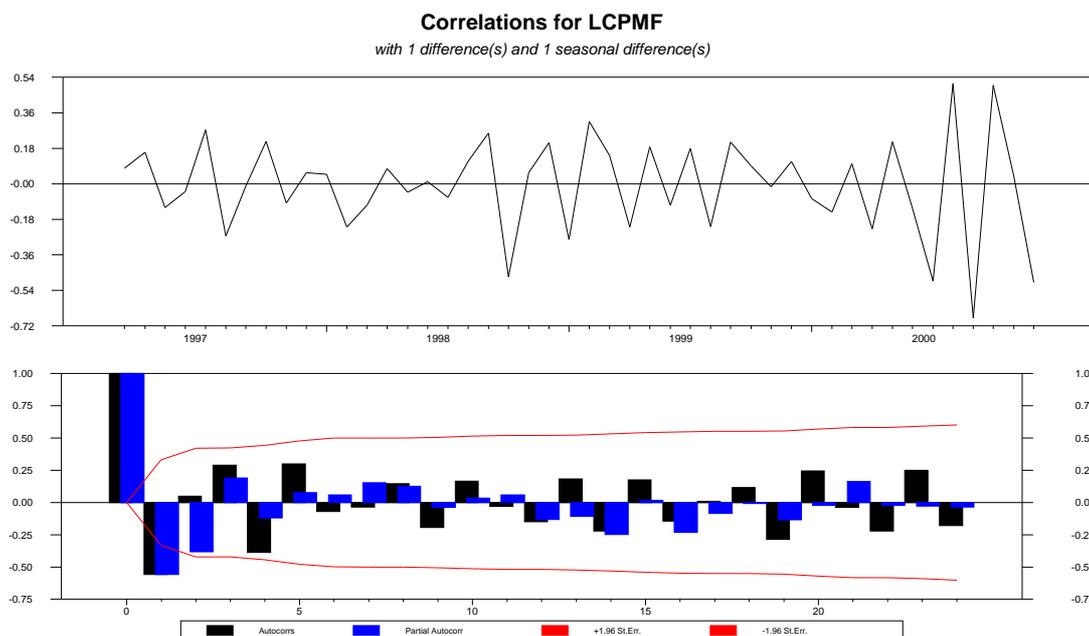


Figura 4.18 – Gráfico da Série Diferenciada  $W_{CPMF} = \nabla_{12} \nabla Z_{CPMF}$  ( $d = 1, D = 1$ ) e suas Funções de Autocorrelação Amostral(FACA e FACPA)

Comparando-se a FACA e a FACPA da série  $W_{CPMF}$  com as FAC e FACP teóricas apresentadas no capítulo 2, conclui-se que remanescem padrões não sazonais que sugerem a inclusão de componentes de média móveis (MA) e/ou auto-regressivos (AR), mas a diferenciação sazonal parece não ter deixado qualquer vestígio de sazonalidade que justifique a inclusão de algum termo sazonal a mais. No caso da parte não sazonal do modelo, as funções de autocorrelação sugerem a inclusão de um ou dois termos de média móvel - MA(1) e MA(2) - e/ou um ou dois termos auto-regressivos - AR(1) e AR(2). No que se refere a parte sazonal do modelo, acrescentaremos apenas a possibilidade de um termo de média móvel - SMA(1), para verificar se há alguma melhora no ajuste. Assim, combinando os componentes supracitados, poder-se-ia, inicialmente, analisar as seguintes opções de modelos:

**Modelo 1 - ARIMA(0, 1, 1)**

As autocorrelações residuais para os modelos em análise foram obtidas pelo RATS (não mostradas). Observa-se correlações residuais não significativamente diferentes de zero (autocorrelações das defasagens 1 e 3, por exemplo), indicando não ser este um modelo

adequado para representar a série em estudo. Tentemos a inclusão de um termo auto-regressivo não sazonal - AR(1).

#### **Modelo 2 - ARIMA(1, 1, 1)**

Ainda se observa correlações residuais altas (autocorrelações das defasagens 3 e 4), mas que são estatisticamente diferentes de zero, indicando que este modelo pode ser adequado para representar a série em estudo. Tentemos a inclusão de um segundo termo auto-regressivo não sazonal - AR(2).

#### **Modelo 3 - ARIMA(2, 1, 0)**

Observa-se, agora que todas as correlações residuais são significativamente diferentes de zero, indicando ser este um potencial modelo a representar a série em estudo. Tentemos a inclusão de um termo de média móvel não sazonal - MA(1) a este modelo..

#### **Modelo 4 - ARIMA(2, 1, 1)**

Observa-se, mais uma vez, que todas as correlações residuais são significativamente diferentes de zero, indicando ser este um outro potencial modelo a representar a série em estudo. Tentemos a inclusão de um segundo termo de média móvel não sazonal - MA(2) ao primeiro modelo.

#### **Modelo 5 - ARIMA(0, 1, 2)**

Observa-se correlações residuais não significativamente diferentes de zero (autocorrelações da defasagem 4), indicando não ser este um modelo adequado para representar a série em estudo. Tentemos a inclusão de uma diferenciação sazonal ( $D = 1$ ) no modelo que apresentou melhor ajuste – ARIMA (2, 1, 0).

#### **Modelo 6 - ARIMA(2, 1, 0)(0, 1, 0)<sub>12</sub>**

Observa-se, mais uma vez, que todas as correlações residuais são significativamente diferentes de zero, indicando ser este um outro potencial modelo a representar a série em estudo. Tentemos, por fim, a inclusão de um s termo de média móvel sazonal - SMA(1) ao modelo anterior.

#### **Modelo 7 - ARIMA(2, 1, 0)(0, 1, 1)<sub>12</sub>**

Observa-se, mais uma vez, que todas as correlações residuais são significativamente diferentes de zero, indicando ser este um outro potencial modelo a representar a série em estudo.

Para se determinar quais os modelos que melhor se ajustam aos dados, optou-se por utilizar os critérios de seleção de Akaike e de Schwartz, quais sejam, o *Akaike Information Criterion*(AIC) e *Schwartz Bayesian Criterion* (SBC ou BIC).

Os valores dos critérios AIC e BIC para cada um dos modelos estudados ( $d = 1$ ;  $D = 1$ ;  $p = 0, 1, 2$ ;  $q = 0, 1, 2$ ;  $P = 0$ ; e  $Q = 0, 1$ ) estão apresentados na tabela 4.15. Como o critério BIC seleciona modelos mais parcimoniosos que o AIC, utilizaremos aquele para a escolha dos três melhores modelos.

Tabela 4.15 – Valores do AIC e do BIC para a Série da CPMF

MODELO	AIC	BIC	MODELO	AIC	BIC
(0,1,0)(0,1,0) <sub>12</sub>	28,17	28,17	(2,1,0)(0,1,0) <sub>12</sub>	6,60	9,53
(0,1,0)(0,1,1) <sub>12</sub>	25,19	26,66	(2,1,0)(0,1,1) <sub>12</sub>	-1,81	2,59
(0,1,0)	14,54	14,54	<b>(2,1,0)</b>	<b>-11,53</b>	<b>-8,59</b>
(0,1,1)(0,1,0) <sub>12</sub>	13,95	15,41	(2,1,1)(0,1,0) <sub>12</sub>	-5,44	-1,04
(0,1,1)(0,1,1) <sub>12</sub>	7,36	10,30	(2,1,1)(0,1,1) <sub>12</sub>	-1,83	4,03
<b>(0,1,1)</b>	<b>-2,60</b>	<b>-1,13</b>	<b>(2,1,1)</b>	<b>-11,14</b>	<b>-6,74</b>
(1,1,0)(0,1,0) <sub>12</sub>	15,39	16,85	(0,1,2)(0,1,0) <sub>12</sub>	10,53	13,46
(1,1,0)(0,1,1) <sub>12</sub>	13,32	16,26	(0,1,2)(0,1,1) <sub>12</sub>	5,56	9,96
(1,1,0)	-2,21	-0,74	<b>(0,1,2)</b>	<b>-5,75</b>	<b>-2,81</b>
(1,1,1)(0,1,0) <sub>12</sub>	12,12	15,05	(1,1,2)(0,1,0) <sub>12</sub>	17,45	21,85
(1,1,1)(0,1,1) <sub>12</sub>	6,74	11,14	(1,1,2)(0,1,1) <sub>12</sub>	7,45	13,31
<b>(1,1,1)</b>	<b>-6,33</b>	<b>-3,40</b>	(1,1,2)	-5,36	-0,967

Assim, analisando a tabela 4.15, observa-se que os modelos que apresentam menores valores para o BIC são os seguintes: ARIMA(1,1,1), ARIMA(2,1,0), ARIMA(2,1,1) e ARIMA(0,1,2).

Passa-se, então, para a etapa de estimação dos modelos selecionados.

### 4.6.2 - Estimação

As estimativas de máxima verossimilhança dos modelos selecionados foram obtidas através do software RATS e são as seguintes:

$$\begin{matrix} (1 - 0,419B)(1 - B)Z_{II_t} = (1 - 0,537B)\hat{a}_t & (4.23) \\ (2,35) & (3,22) \\ [0,18] & [0,17] \end{matrix}$$

$$\begin{matrix} (1 - 0,978B - 0,571B^2)(1 - B)Z_{II_t} = \hat{a}_t & (4.24) \\ (7,85) & (4,59) \\ [0,12] & [0,12] \end{matrix}$$

$$\begin{matrix} (1 - 1,130B - 0,670B^2)(1 - B)Z_{II_t} = (1 + 0,231B)\hat{a}_t & (4.25) \\ (6,10) & (4,84) & (0,93) \\ [0,19] & [0,14] & [0,25] \end{matrix}$$

$$\begin{matrix} (1 - B)Z_{II_t} = (1 - 1,130B - 0,670B^2)\hat{a}_t & (4.26) \\ (6,80) & (2,71) \\ [0,14] & [0,15] \end{matrix}$$

Os números entre parênteses são os valores absolutos da estatística t dos coeficientes e os entre colchetes são os desvios padrões das estimativas. Os coeficientes da parte autorregressiva (AR) respeitam as condições para a estacionariedade do modelo e os coeficientes da parte de média móvels (MA) respeitam as condições de invertibilidade, nos termos do explicitado no capítulo 2

### 4.6.3 - Checagem de Diagnóstico

Verifica-se que nem todos os coeficientes estimados são estatisticamente significativos a 5%. O coeficiente do termo de média móvel do modelo ARIMA(2,1,1) não foi estatisticamente significativo. Os resíduos para os quatro modelos acima descritos, juntamente com a FACA e a FACPA residuais para cada um deles, foram obtidos pelo RATS, conforme já citado anteriormente. A FACA residual, em particular, mostra a satisfatoriedade do ajuste para os modelos ARIMA(1,1,1) e ARIMA(2,1,0), indicando, claramente, que todas as

autocorrelações residuais são estatisticamente nulas. Já para o caso do modelo ARIMA(0,1,2) as autocorrelações residuais para as algumas defasagens se mostraram estatisticamente diferentes de zero, conforme já discutido anteriormente.

Reforçando a satisfatoriedade dos ajustes dos modelos ARIMA(1,1,1) e ARIMA(2,1,0), a estatística  $\chi^2$  se mostrou insignificante ao nível de 5% para os graus de liberdade de cada modelo e tanto os gráficos dos resíduos padronizados ( $a_t/\sigma_a$ ), traçados nas figuras 2.6.1 e 2.6.2 do Apêndice 2, quanto o teste de normalidade de Jarque-Bera sugerem que os choques devem ser normalmente distribuídos. Devemos observar que tanto no gráfico da figura 2.6.1 quanto no da figura 2.6.2 nenhum valor foi superior a três vezes o desvio padrão. Isso reforça que os choques apresentam uma distribuição perfeitamente normal.

É importante ressaltar que o segundo termo auto-regressivo não sazonal – AR(2) – do modelo ARIMA(2,1,0) introduz um ciclo estocástico nas previsões. O período médio deste ciclo é de 2,76 meses. Conforme discutido em Bell e Hillmer (1983), quando se encontra um ciclo estocástico em dados mensais acumulados com um período médio de cerca de três meses, isto sugere uma forte evidência de que os dados apresentam variação de calendário (“trading day variation”).

#### 4.5.4 - Previsão

As previsões para os dois modelos selecionados estão mostradas abaixo nas tabelas 4.16 e 4.17, juntamente com o valor da raiz quadrada do erro quadrático médio de previsão (RQEQMP). Conforme se pode observar, foram calculados as RQEQMP para previsões para três, seis e doze passos à frente. A análise dos resultados a ser realizada no capítulo 5 utilizará essa medida para a comparação dos modelos.

Tabela 4 16 – Previsões do modelo ARIMA (2, 1, 0)

MESES	Previsão do Modelo (LCPMF)	Limite 95% abaixo	Previsão (C <sub>P</sub> ) milhões de R\$	Limite 95% acima	Valor Real Observado (C <sub>R</sub> ) milhões de R\$	ERRO ABS. (C <sub>P</sub> - C <sub>R</sub> )	ERRO PERC. (%)
JAN/01	7,2209	1097,62	1367,719454	1704,28	1329,566693	38,15	2,87
FEV/01	7,0102	871,49	1107,876058	1408,39	791,7137668	316,16	39,93
MAR/01	7,1459	978,38	1268,892815	1645,67	1400,175046	-131,28	-9,38
ABR/01	7,1960	988,31	1334,083743	1800,82	1362,009016	-27,93	-2,05
MAI/01	7,1772	950,70	1309,237257	1802,99	1794,427711	-485,19	-27,04
JUN/01	7,1665	921,96	1295,303099	1819,83	1427,064128	-131,76	-9,23
JUL/01	7,1795	915,53	1312,251968	1880,89	1529,937263	-217,69	-14,23
AGO/01	7,1615	881,39	1288,842748	1884,65	1841,00157	-552,16	-29,99
SET/01	7,1717	872,79	1302,056218	1942,44	1456,858808	-154,80	-10,63
OUT/01	7,1720	855,77	1302,446893	1982,27	1927,263244	-624,82	-32,42
NOV/01	7,1659	833,72	1294,526150	2010,02	1574,091398	-279,57	-17,76
DEZ/01	7,1717	838,57	1302,056218	2021,71	1757,27512	-455,22	-25,90
TOTAL		10809,01	11006,24	15485,29	21903,97	-2706,09	-14,88
RAIZ QUADRADA DO EQM - 3 MESES						198,87	
RAIZ QUADRADA DO EQM - 6 MESES						249,06	
RAIZ QUADRADA DO EQM - 12 MESES						344,51	

Tabela 4 27 – Previsões do modelo ARIMA (1, 1, 1)

MESES	Previsão do Modelo (LCPMF)	Limite 95% abaixo	Previsão (C <sub>P</sub> ) em milhões de R\$	Limite 95% acima	Valor Real Observado (C <sub>R</sub> )	ERRO ABS. (C <sub>P</sub> - C <sub>R</sub> )	ERRO PERC. (%)
JAN/01	7,1635	1036,39	1291,423013	1609,21	1329,566693	-38,14	-2,87
FEV/01	7,0019	864,28	1098,718742	1396,75	791,7137668	307,00	38,78
MAR/01	7,1388	971,46	1259,915583	1634,02	1400,175046	-140,26	-10,02
ABR/01	7,1352	930,01	1255,388041	1694,60	1362,009016	-106,62	-7,83
MAI/01	7,1398	915,80	1261,176129	1736,80	1794,427711	-533,25	-29,72
JUN/01	7,1361	894,35	1256,518399	1765,34	1427,064128	-170,55	-11,95
JUL/01	7,1362	876,73	1256,644057	1801,18	1529,937263	-273,29	-17,86
AGO/01	7,1566	877,08	1282,542866	1875,44	1841,00157	-558,46	-30,33
SET/01	7,1456	850,31	1268,512204	1892,40	1456,858808	-188,35	-12,93
OUT/01	7,1398	828,65	1261,176129	1919,46	1927,263244	-666,09	-34,56
NOV/01	7,1362	809,32	1256,644057	1951,20	1574,091398	-317,45	-20,17
DEZ/01	7,1362	809,32	1256,644057	1951,20	1757,27512	-500,63	-28,49
TOTAL		10472,95	10663,72	15005,30	21227,61	-3186,08	-17,51
RAIZ QUADRADA DO EQM - 3 MESES						196,11	
RAIZ QUADRADA DO EQM - 6 MESES						270,86	
RAIZ QUADRADA DO EQM - 12 MESES						371,63	

### 4.7 - SÉRIE RECEITA DE OUTRAS CONTRIBUIÇÕES (OCONTR)

#### 4.7.1 - Identificação

Conforme analisado no capítulo 3, o gráfico da série Receita de Outras Contribuições ( $Z_{OCONTR}$ ), apresentado na figura 3.26, mostra uma variância que cresce com o aumento do nível da série. Para estabilizar esta variância utilizar-se-á o logaritmo natural da série original, de tal forma que o modelo ARIMA seja construído para a série transformada  $Z'_{OCONTR} = \ln(Z_{OCONTR})$ . O gráfico de  $Z'_{OCONTR}$  está apresentado na figura 3.27 para o período jan/1989 a dez/2000 e na figura 3.29 para o período jul/1994 a dez/2000, que será o utilizado no presente estudo.

Para identificação do modelo utilizar-se-á, inicialmente, as Funções de Autocorrelação Amostral (FACA) e de Autocorrelação Parcial Amostral (FACPA). A figura 4.19 mostra a FACA, em preto, e a FACPA, em azul, para  $Z'_{OCONTR}$  com  $d = D = 0$  (sem qualquer diferenciação). Verifica-se ali que a autocorrelação cai a valores insignificamente próximos de zero bastante rapidamente, salvo alguns raros valores, como para a defasagem 12, por exemplo. Este comportamento sugere uma forte evidência de que a série  $Z'_{OCONTR}$  seja estacionária, consistente com o gráfico apresentado na figura 3.29, em que os dados não apresentam uma tendência visível.

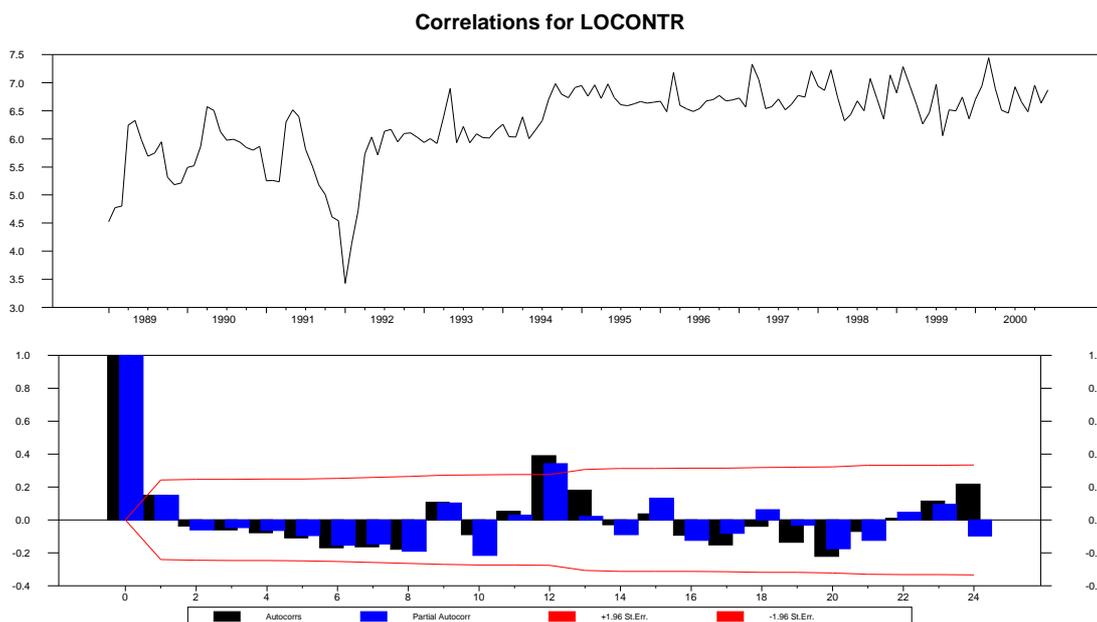


Figura 4.19 – Funções de Autocorrelação Amostral (FACA e FACPA) para a série  $Z'_{OCONTR}$

Apesar de haver evidência de estacionariedade não sazonal, verificar-se-á o comportamento da série diferenciada. A primeira diferença não sazonal  $W^*_{OCONTR} = \nabla Z'_{OCONTR}$  ( $d = 1$ ) está apresentada no gráfico 4.20 (superior). Observa-se que esta série tem uma média estável ao longo do tempo e, como relatado no capítulo 2, isto indica que a série diferenciada é estacionária. Entretanto, os dados da figura 4.20 (inferior) mostram uma forte variação sazonal.

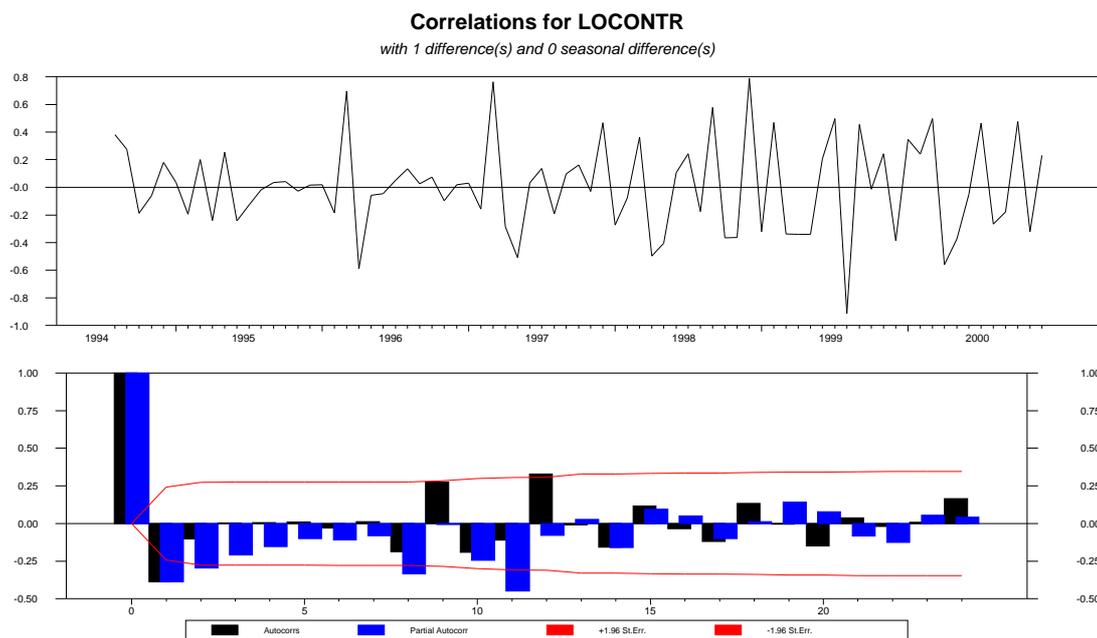


Figura 4.20 – Gráfico da Série Diferenciada  $W^*_{OCONTR} = \nabla Z'_{OCONTR}$  ( $d = 1$ ) e suas Funções de Autocorrelação Amostral (FACA e FACPA )

A FACA e a FACPA para a série diferenciada  $W^*_{OCONTR} = \nabla Z'_{OCONTR}$  ( $d = 1$ ) são mostradas na figura 4.20(inferior). Verifica-se ali que a autocorrelação cai a valores insignificamente próximos de zero bastante rapidamente, salvo alguns valores. Isto reforça a conclusão do parágrafo anterior de que a série diferenciada seja estacionária, nos termos do discutido no capítulo 2. Entretanto, as autocorrelações para algumas defasagens são altas, com picos alternados. Isto reforça a observação anteriormente feita de que os dados podem apresentar alguma variação sazonal e sugere verificar se uma diferenciação sazonal ( $D = 1$ ), em adição à diferenciação não sazonal ( $d = 1$ ), melhoraria o ajuste.

A figura 4.21(superior) apresenta o gráfico dos dados após sofrerem as duas diferenciações, sazonal ( $D = 1$ ) e não sazonal ( $d = 1$ ),  $W_{OCONTR} = \nabla_{12} W^*_{OCONTR} =$

$\nabla_{12}\nabla Z'_{OCONTR}$ . Esta nova série,  $W_{OCONTR}$ , apresenta uma média estável e não apresenta mais um padrão sazonal observado na série  $W^*_{OCONTR}$ . A FACA e a FACPA para  $W_{OCONTR}$  estão apresentadas na figura 4.21(inferior).

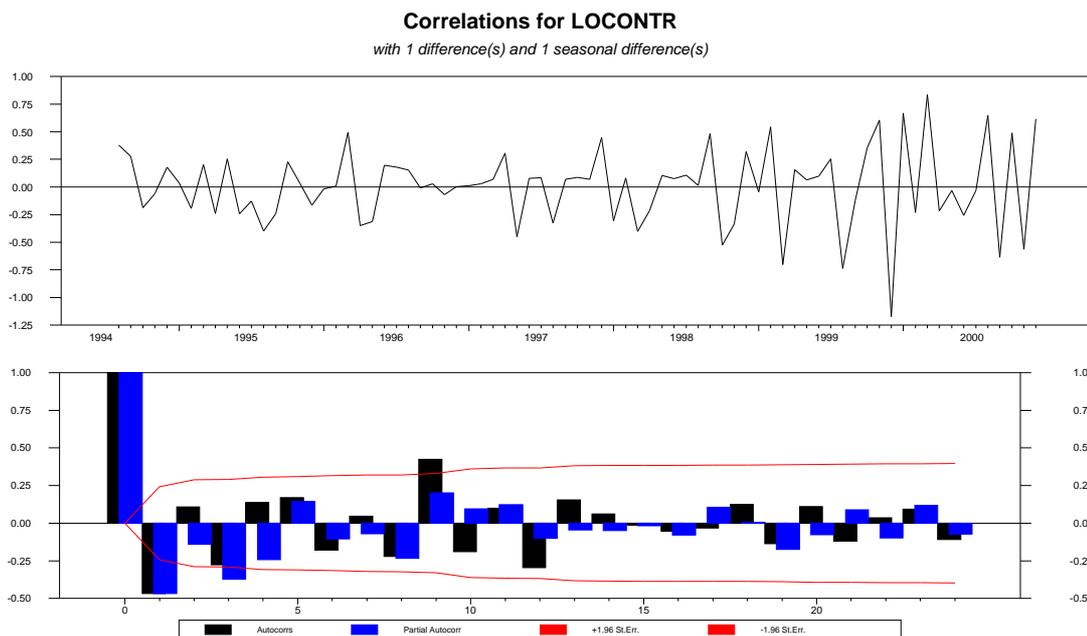


Figura 4.21 – Gráfico da Série Diferenciada  $W_{OCONTR} = \nabla_{12}\nabla Z'_{OCONTR}$  ( $d = 1, D = 1$ ) e suas Funções de Autocorrelação Amostral(FACA e FACPA)

Comparando-se a FACA e a FACPA da série  $W_{OCONTR}$  com as FAC e FACP teóricas apresentadas no capítulo 2, conclui-se que remanescem padrões sazonais e não sazonais que sugerem a inclusão de componentes de média móveis (MA) e/ou auto-regressivos (AR). No caso da parte não sazonal do modelo, as funções de autocorrelação sugerem a inclusão de um termo de média móvel - MA(1) - e/ou um termo auto-regressivo - AR(1). No que se refere a parte sazonal do modelo, acrescentaremos apenas a possibilidade de um termo de média móvel - SMA(1) ou um termo auto-regressivo – SAR(1). Assim, combinando os componentes supracitados, poder-se-ia, inicialmente, analisar as seguintes opções de modelos:

**Modelo 1 - ARIMA(0, 1, 0)(0, 1, 1)<sub>12</sub>**

As autocorrelações residuais para os modelos em análise foram obtidas pelo RATS. Observa-se correlações residuais não significativamente diferentes de zero (autocorrelações das defasagens 1 e 3, por exemplo), indicando não ser este um modelo adequado para representar a série em estudo. Tentemos a inclusão de um termo de média móvel não sazonal - MA(1).

**Modelo 2 - ARIMA(0, 1, 1)(0, 1, 1)<sub>12</sub>**

Observa-se que as correlações residuais são significativamente diferentes de zero, indicando ser este um potencial modelo a representar a série em estudo. Tentemos a inclusão de um termo auto-regressivo não sazonal - AR(1).

**Modelo 3 - ARIMA(1, 1, 1)(0, 1, 1)<sub>12</sub>**

Observa-se que as correlações residuais são significativamente diferentes de zero, indicando ser este um outro potencial modelo a representar a série em estudo. Como a diferenciação sazonal por si parece eliminar a estacionariedade, passa-se aos modelos:

**Modelo 4 - ARIMA(0, 0, 0)(0, 1, 1)<sub>12</sub>**

Observa-se que as correlações residuais são significativamente diferentes de zero, indicando ser este mais um potencial modelo a representar a série em estudo. Tentemos a inclusão de um termo auto-regressivo sazonal - AR(1).

**Modelo 5 - ARIMA(0, 0, 0)(1, 1, 0)<sub>12</sub>**

Observa-se que as correlações residuais são significativamente diferentes de zero, indicando ser este mais um potencial modelo a representar a série em estudo. Tentemos a inclusão conjunta de um termo SAR(1) e um SMA(1).

**Modelo 6 - ARIMA(0, 0, 0)(1, 1, 1)<sub>12</sub>**

Observa-se que as correlações residuais são significativamente diferentes de zero, indicando ser este mais um potencial modelo a representar a série em estudo.

Para se determinar quais os modelos que melhor se ajustam aos dados, optou-se por utilizar os critérios de seleção de Akaike e de Schwartz, quais sejam, o *Akaike Information Criterion*(AIC) e *Schwartz Bayesian Criterion* (SBC ou BIC).

Os valores dos critérios AIC e BIC para cada um dos modelos estudados ( $d = 1$ ;  $D = 1$ ;  $p = 0, 1$ ;  $q = 0, 1$ ;  $P = 0, 1$  e  $Q = 0, 1$ ) estão apresentados na tabela 4.18 da página seguinte. Como o critério BIC seleciona modelos mais parcimoniosos que o AIC, utilizaremos aquele para a escolha dos três melhores modelos.

Assim, analisando a tabela 4.18, observa-se que os modelos que apresentam menores valores para o BIC são os seguintes: ARIMA (0,1,1)(0,1,1)<sub>12</sub>, ARIMA(1,1,1)(0,1,1)<sub>12</sub>, ARIMA(0,0,0)(0,1,1)<sub>12</sub> e ARIMA(0,0,0)(1,1,0)<sub>12</sub>.

Passa-se, então, para a etapa de estimação dos modelos selecionados.

Tabela 4.18 – Valores do AIC e do BIC para a Série Outras Contribuições (OCONTR)

MODELO	AIC	BIC	MODELO	AIC	BIC
(0,1,0)(0,1,0) <sub>12</sub>	106,65	106,65	(1,1,0)(1,1,0) <sub>12</sub>	85,97	89,84
(0,1,0)(0,1,1) <sub>12</sub>	93,76	95,69	(1,1,1)(0,1,0) <sub>12</sub>	79,36	83,22
(0,1,0)(1,1,0) <sub>12</sub>	100,77	102,70	<b>(1,1,1)(0,1,1)<sub>12</sub></b>	<b>65,73</b>	<b>71,53</b>
(0,1,1)(0,1,0) <sub>12</sub>	77,55	79,48	(1,1,1)(1,1,0) <sub>12</sub>	73,71	79,50
<b>(0,1,1)(0,1,1)<sub>12</sub></b>	<b>63,77</b>	<b>67,63</b>	(0,1,0) <sub>12</sub>	74,96	74,96
(0,1,1)(1,1,0) <sub>12</sub>	71,71	75,58	<b>(0,1,1)<sub>12</sub></b>	<b>60,47</b>	<b>62,40</b>
(1,1,0)(0,1,0) <sub>12</sub>	91,43	93,37	<b>(1,1,0)<sub>12</sub></b>	<b>67,03</b>	<b>68,96</b>
(1,1,0)(0,1,1) <sub>12</sub>	77,37	81,23	(2,1,0) <sub>12</sub>	67,14	71,98

### 4.7.2 - Estimação

As estimativas de máxima verossimilhança dos modelos selecionados foram obtidas através do software RATS e são as seguintes:

$$(1 - B)(1 - B^{12})Z_{it} = (1 - 0,909B)(1 - 0,711B^{12})\hat{a}_t \quad (4.27)$$

(15,20)      (5,48)  
[0,06]      [0,13]

$$(1 - 0,031B)(1 - B)(1 - B^{12})Z_{it} = (1 - 0,902B)(1 - 0,714B^{12})\hat{a}_t \quad (4.28)$$

(0,19)                      (15,20)      (5,48)  
[0,16]                      [0,06]      [0,13]

$$(1 - B^{12})Z_{it} = (1 - 0,691B^{12})\hat{a}_t \quad (4.29)$$

(5,34)  
[0,13]

$$(1 - 0,464B^{12})(1 - B^{12})Z_{it} = \hat{a}_t \quad (4.30)$$

(3,27)  
[0,14]

Os números entre parênteses são os valores absolutos da estatística t dos coeficientes e os entre colchetes são os desvios padrões das estimativas. Os coeficientes da parte auto-regressiva (AR) respeitam as condições para a estacionariedade do modelo e os coeficientes

da parte de média móveis (MA) respeitam as condições de invertibilidade, nos termos do explicitado no capítulo 2

### 4.7.3 - Checagem de Diagnóstico

Verifica-se que nem todos os coeficientes estimados são estatisticamente significativos a 5%. O coeficiente do termo auto-regressivo do modelo  $ARIMA(1,1,1)(0,1,1)_{12}$  não foi estatisticamente significativo. Os resíduos para os quatro modelos acima descritos, juntamente com a FACA e a FACPA residuais para cada um deles, foram obtidos pelo RATS. A FACA residual, em particular, mostra a satisfatoriedade do ajuste para todos os modelos, indicando, claramente, que todas as autocorrelações residuais são estatisticamente nulas. O modelo  $ARIMA(1,1,1)(0,1,1)_{12}$  será descartado pelo teste de significância, conforme já discutido anteriormente.

Reforçando a satisfatoriedade do ajuste do modelo  $ARIMA(0,1,1)(0,1,1)_{12}$ , a estatística  $\chi^2$  se mostrou insignificante ao nível de 5% para os graus de liberdade de cada modelo e tanto o gráfico dos resíduos padronizados ( $a_t/\sigma_a$ ), traçado na figura 2.7.1 do Apêndice 2, quanto o teste de normalidade de Jarque-Bera sugerem que os choques devem ser normalmente distribuídos. Devemos observar que no gráfico da figura 2.7.1 nenhum valor foi superior a três vezes o desvio padrão. Isso reforça que os choques apresentam uma distribuição perfeitamente normal. Os demais modelos,  $ARIMA(0,1,1)_{12}$  e  $ARIMA(1,1,0)_{12}$ , apresentaram gráficos dos resíduos padronizados ( $a_t/\sigma_a$ ), traçados nas figura 2.7.2 e 2.7.3 do Apêndice 2, que sugerem que os resíduos não sejam normalmente distribuídos, evidenciando que tais modelos não se mostram adequados para a previsão.

### 4.7.4 - Previsão

As previsões para o único modelo selecionado estão mostradas abaixo na tabelas 4.19, juntamente com o valor da raiz quadrada do erro quadrático médio de previsão (RREQMP). Conforme se pode observar, foram calculados as RREQMP para previsões para três, seis e doze passos à frente. A análise dos resultados a ser realizada no capítulo 5 utilizará essa medida para a comparação dos modelos.

Tabela 4 39 – Previsões do modelo ARIMA (2, 1, 0)

MESES	Previsão do Modelo (LOCONTR)	Limite 95% abaixo	Previsão (C <sub>P</sub> ) milhões de R\$	Limite 95% acima	Valor Real Observado (C <sub>R</sub> ) milhões de R\$	ERRO ABS. (C <sub>P</sub> - C <sub>R</sub> )	ERRO PERC. (%)
JAN/01	6,9022	751,60	994,460123	1315,80	1358,367148	-363,91	-26,79
FEV/01	7,1128	927,78	1227,579962	1624,25	989,8152434	237,76	24,02
MAR/01	7,3956	1182,75	1628,801912	2243,07	1184,473848	444,33	37,51
ABR/01	6,9366	747,40	1029,264759	1417,43	1224,152773	-194,89	-15,92
MAI/01	6,5839	504,67	723,354919	1036,81	935,7180963	-212,36	-22,70
JUN/01	6,6165	521,39	747,324877	1071,16	1016,474392	-269,15	-26,48
JUL/01	7,0160	762,04	1114,320410	1629,45	1392,721092	-278,40	-19,99
AGO/01	6,6252	515,53	753,854968	1102,35	1017,04048	-263,19	-25,88
SET/01	6,7303	561,32	837,398448	1249,25	945,4651254	-108,07	-11,43
OUT/01	6,9326	687,18	1025,155923	1529,35	1494,6722	-469,52	-31,41
NOV/01	6,7806	578,59	880,596924	1340,23	1257,855542	-377,26	-29,99
DEZ/01	6,9301	671,89	1022,596234	1556,35	1467,789037	-445,19	-30,33
TOTAL		8412,16	11984,709460	17115,51	14284,54498	-2299,84	-16,10
RAIZ QUADRADA DO EQM - 3 MESES						358,88	
RAIZ QUADRADA DO EQM - 6 MESES						300,53	
RAIZ QUADRADA DO EQM - 12 MESES						324,21	

### 4.8 - SÉRIE RECEITA TOTAL DE CONTRIBUIÇÕES (RCONT)

#### 4.8.1 - Identificação

Conforme analisado no capítulo 3, o gráfico da série Receita Total de Contribuições ( $Z_{RCONT}$ ), apresentado na figura 3.30, mostra uma variância que cresce com o aumento do nível da série. Para estabilizar esta variância utilizar-se-á o logaritmo natural da série original, de tal forma que o modelo ARIMA seja construído para a série transformada  $Z'_{RCONT} = \ln(Z_{RCONT})$ . O gráfico de  $Z'_{RCONT}$  está apresentado na figura 3.31 para o período jan/1989 a dez/2000 e na figura 3.33 para o período jul/1994 a dez/2000, que será o utilizado no presente estudo.

Para identificação do modelo utilizar-se-á, inicialmente, as Funções de Autocorrelação Amostral (FACA) e de Autocorrelação Parcial Amostral (FACPA). A figura 4.22 mostra a FACA, em preto, e a FACPA, em azul, para  $Z'_{RCONT}$  com  $d = D = 0$  (sem qualquer diferenciação). Verifica-se que as autocorrelações (FACA) são significativamente diferentes de zero até a defasagem 7, decaindo lentamente até a defasagem 18. Este comportamento sugere que a série  $Z'_{RCONT}$  seja não-estacionária, consistente com o gráfico apresentado na figura 3.33, em que os dados apresentam uma tendência ascendente.

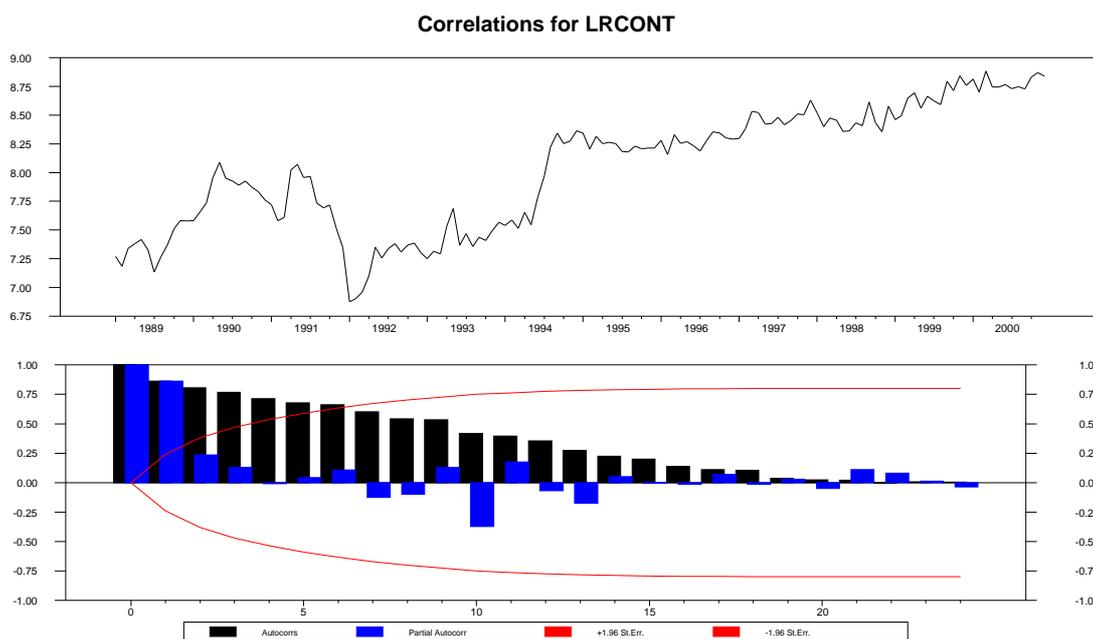


Figura 4.22 – Funções de Autocorrelação Amostral (FACA e FACPA) para a série  $Z'_{RCONT}$

A primeira diferença não sazonal  $W^*_{RCONT} = \nabla Z'_{RCONT}$  ( $d = 1$ ) está apresentada no gráfico 4.23 (superior). Observa-se que esta série tem uma média estável ao longo do tempo e, como relatado no capítulo 2, isto indica que a série diferenciada é estacionária. Entretanto, os dados da figura 4.23 (superior) mostram uma forte variação sazonal.

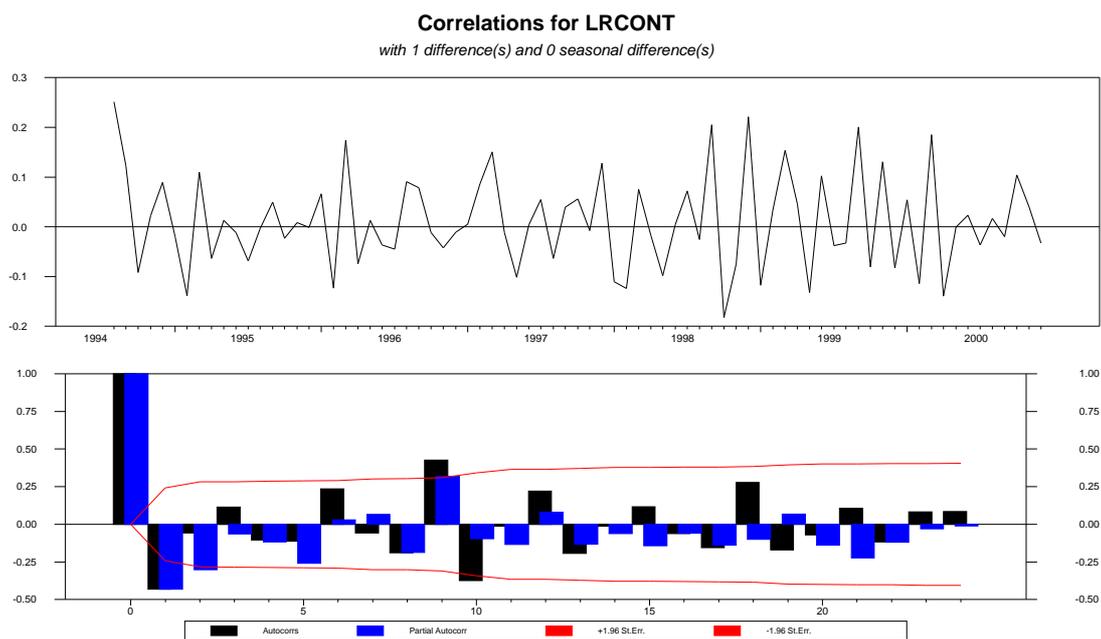


Figura 4.23 – Gráfico da Série Diferenciada  $W^*_{RCONT} = \nabla Z'_{RCONT}$  ( $d = 1$ ) e suas Funções de Autocorrelação Amostral (FACA e FACPA)

A FACA para a série diferenciada  $W^*_{RCONT} = \nabla Z'_{RCONT}$  ( $d = 1$ ) estão apresentadas na figura 4.23(inferior). Verifica-se ali que a autocorrelação cai a valores insignificamente próximos de zero bastante rapidamente, exceto para algumas defasagens. Isto reforça a conclusão do parágrafo anterior de que a série diferenciada é agora estacionária, nos termos do discutido no capítulo 2. Entretanto, as autocorrelações para as defasagens 12 e 24 são altas e não se observa uma queda significativa da defasagem 12 para a 24. Isto reforça a observação anteriormente feita de que os dados apresentam forte variação sazonal e sugere uma diferenciação sazonal ( $D = 1$ ) em adição à diferenciação não sazonal ( $d = 1$ ).

A figura 4.24(superior) apresenta o gráfico dos dados após sofrerem as duas diferenciações, sazonal ( $D = 1$ ) e não sazonal ( $d = 1$ ),  $W_{RCONT} = \nabla_{12} W^*_{RCONT} = \nabla_{12} \nabla Z'_{RCONT}$ . Esta nova série,  $W_{RCONT}$ , apresenta uma média estável e não apresenta mais o forte padrão sazonal observado na série  $W^*_{RCONT}$ . A FACA e a FACPA para  $W_{RCONT}$  está apresentada na figura 4.24(inferior).

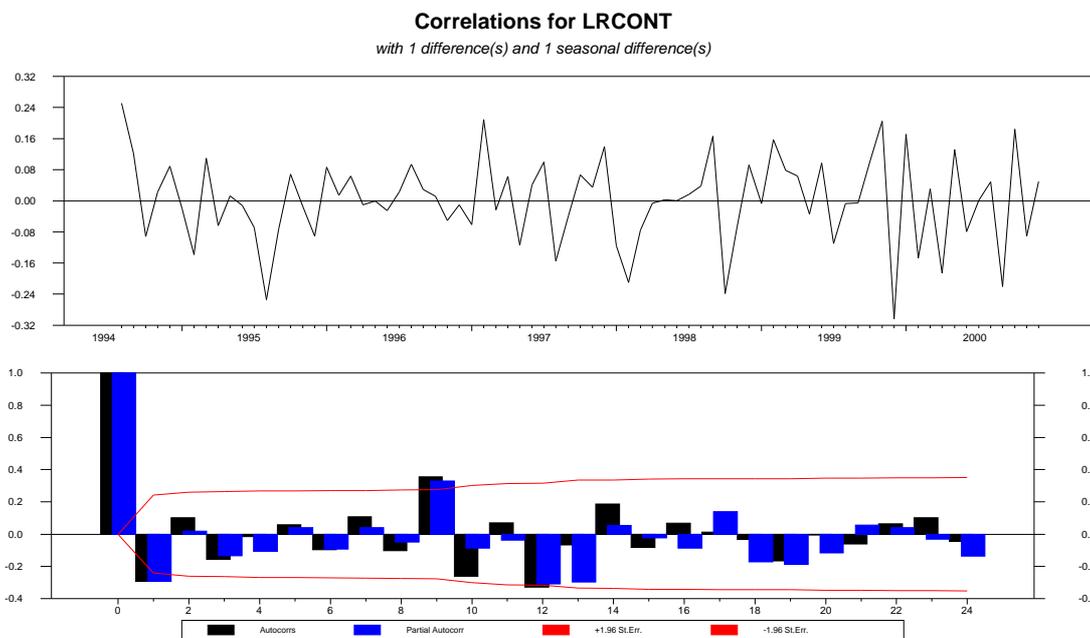


Figura 4.24 – Gráfico da Série Diferenciada  $W_{RCONT} = \nabla_{12} \nabla Z_{RCONT}$  ( $d = 1, D = 1$ ) e suas Funções de Autocorrelação Amostral (FACA e FACPA)

Comparando-se a FACA e a FACPA da série  $W_{RCONT}$  com as FAC e FACP teóricas apresentadas no capítulo 2, conclui-se que remanescem padrões sazonais e não sazonais que sugerem a inclusão de componentes de média móvels (MA) e/ou auto-regressivos (AR). No caso da parte não sazonal do modelo, as funções de autocorrelação sugerem a inclusão de um ou dois termos de média móvel - MA(1)/MA(2) - e/ou um ou dois termos auto-regressivos - AR(1) e AR(2). No que se refere a parte sazonal do modelo, as funções de autocorrelação sugerem a inclusão de um termo auto-regressivo - SAR(1) ou de um termo de média móvel - SMA(1). Assim, combinando os componentes supracitados, poder-se-ia, inicialmente, analisar as seguintes opções de modelos:

**Modelo 1 - ARIMA(0, 1, 0)(0, 1, 1)<sub>12</sub>**

As autocorrelações residuais para os modelos em análise foram obtidas pelo RATS (não mostradas). Observa-se correlações residuais não significativamente diferentes de zero (autocorrelações das defasagens 9), indicando não ser este um modelo adequado para representar a série em estudo. Tentemos a inclusão de um termo auto-regressivo não sazonal - AR(1) e/ou um termo de média móvel não sazonal - MA(1), conforme parece indicar as autocorrelações na primeira defasagem.

**Modelo 2 - ARIMA(1, 1, 0)(0, 1, 1)<sub>12</sub>**

Ainda se observa correlações residuais não significativamente diferentes de zero (autocorrelações das defasagens 9 e 10), indicando não ser este, ainda, um modelo adequado para representar a série em estudo. Tentemos incluir um termo de média móvel não sazonal.

**Modelo 3 - ARIMA(0, 1, 1)(0, 1, 1)<sub>12</sub>**

As autocorrelações da defasagem 9 teimam em persistirem elevadas, indicando não ser este, ainda, um modelo adequado para representar a série em estudo. Verifiquemos a inclusão de um termo auto-regressivo sazonal.

**Modelo 4 - ARIMA(0, 1, 1)(1, 1, 1)<sub>12</sub>**

Agora, verifica-se que todas as correlações residuais são significativamente nulas, indicando ser este um potencial modelo a representar a série em estudo. Tentemos o seguinte modelo: ARIMA(1, 1, 0)(0, 1, 1)<sub>12</sub>.

**Modelo 5 - ARIMA(1, 1, 0)(0, 1, 1)<sub>12</sub>**

Aqui também se observa correlações residuais não significativamente diferentes de zero (autocorrelações das defasagens 9 e 10, por exemplo), indicando que, provavelmente, este modelo também não seja adequado para representar a série em estudo. Tentemos incluir mais um termo auto-regressivo não sazonal a este modelo. Tentemos, então, o seguinte ajuste:

**Modelo 6 - ARIMA(2, 1, 1)(1, 1, 1)<sub>12</sub>**

Agora, mais uma vez se verifica que todas as correlações residuais são significativamente nulas, indicando ser este um outro potencial modelo a representar a série em estudo. Por fim tentemos a inclusão de um segundo termo de média móvel não-sazonal..

**Modelo 7 - ARIMA(2, 1, 0)(0, 1, 1)<sub>12</sub>**

Aqui também se observa correlações residuais não significativamente diferentes de zero (autocorrelações das defasagens 11 e 13, por exemplo), indicando que, provavelmente, este modelo também não seja adequado para representar a série em estudo.

Para se determinar quais os modelos que melhor se ajustam aos dados, optou-se por utilizar os critérios de seleção de Akaike e de Schwartz, quais sejam, o *Akaike Information Criterion (AIC)* e *Schwartz Bayesian Criterion (SBC ou BIC)*.

Os valores dos critérios AIC e BIC para cada um dos modelos estudados ( $d = 1$ ;  $D = 1$ ;  $p = 0, 1, 2$ ;  $q = 0, 1, 2$ ;  $P = 0, 1$ ; e  $Q = 0, 1$ ) estão apresentados na tabela 4.20.



Os números entre parênteses são os valores absolutos da estatística  $t$  dos coeficientes e os entre colchetes são os desvios padrões das estimativas. Os coeficientes da parte auto-regressiva (AR) respeitam as condições para a estacionariedade do modelo e os coeficientes da parte de média móveis (MA) respeitam as condições de invertibilidade, nos termos do explicitado no capítulo 2.

#### 4.8.3 - Checagem de Diagnóstico

Verifica-se que nem todos os coeficientes estimados são estatisticamente significativos a 5%. Os coeficientes dos primeiros termos auto-regressivos dos modelos  $ARIMA(0,1,1)(1,1,1)_{12}$  e  $ARIMA(2,1,1)(1,1,1)_{12}$  não foram estatisticamente significativos. Os resíduos para os cinco modelos acima descritos, juntamente com a FACA e a FACPA residuais para cada um deles, foram obtidos pelo RATS. A FACA residual, em particular, apesar de mostrar altas correlações residuais para algumas defasagens não pode, por si só, descartar a satisfatoriedade do ajuste para todos os três modelos restantes.

Para reforçar possível satisfatoriedade dos ajustes dos três modelos remanescentes, a estatística  $\chi^2$  se mostrou insignificante ao nível de 5% para os graus de liberdade de cada modelo e tanto o gráfico dos resíduos padronizados ( $a_t/\sigma_a$ ), traçado nas figuras 2.8.1, 2.8.2 e 2.8.3 do Apêndice 2, quanto o teste de normalidade de Jarque-Bera sugerem que os choques devem ser normalmente distribuídos. Devemos observar que no gráfico da figura 2.8.1 apenas um valor foi superior a três vezes o desvio padrão (valor para agosto/1994). Em uma distribuição perfeitamente normal isso não ocorre, mas a justificativa pode estar no "outlier" que não deve ter sido removido adequadamente pelo processo escolhido. Para os outros dois modelos, nenhum valor foi superior a três vezes o desvio padrão. Isso reforça que os choques apresentam uma distribuição perfeitamente normal. Utilizaremos, assim, os modelos  $ARIMA(1,1,0)(0,1,1)_{12}$  e  $ARIMA(1,1,2)(1,1,1)_{12}$  para a previsão.

#### 4.8.4 - Previsão

As previsões para os dois modelos selecionados estão mostradas abaixo nas tabelas 4.21 e 4.22, juntamente com o valor da raiz quadrada do erro quadrático médio de previsão (RQEQMP). Conforme se pode observar, foram calculados as RQEQMP para previsões para três, seis e doze passos à frente. A análise dos resultados a ser realizada no capítulo 5 utilizará essa medida para a comparação dos modelos.

Tabela 4 24 – Previsões do modelo ARIMA (1, 1, 0)(0, 1, 1)<sub>12</sub>

MESES	Previsão do Modelo (LRCONT)	Limite 95% abaixo	Previsão (C <sub>P</sub> ) milhões de R\$	Limite 95% acima	Valor Real Observado (C <sub>R</sub> ) milhões de R\$	ERRO ABS. (C <sub>P</sub> - C <sub>R</sub> )	ERRO PERC. (%)
JAN/01	8,8596	6000,51	7041,665515	8263,47	7691,299726	-649,63	-8,45
FEV/01	8,7646	5348,65	6403,500033	7666,38	6253,134291	150,37	2,40
MAR/01	8,9030	6020,95	7354,002487	8982,20	6892,246507	461,76	6,70
ABR/01	8,8720	5721,58	7129,525775	8883,94	7504,80402	-375,28	-5,00
MAI/01	8,8241	5345,98	6796,071475	8639,50	7520,087136	-724,02	-9,63
JUN/01	8,8490	5372,24	6967,418067	9036,25	7600,850542	-633,43	-8,33
JUL/01	8,8450	5140,99	6939,604060	9367,49	8032,396147	-1092,79	-13,60
AGO/01	8,8611	5120,97	7052,235939	9711,83	8122,945643	-1070,71	-13,18
SET/01	8,9290	5372,24	7547,713887	10604,14	7747,376444	-199,66	-2,58
OUT/01	8,9213	5225,47	7489,819669	10735,38	8715,791911	-1225,97	-14,07
NOV/01	8,9503	5272,71	7710,204576	11274,51	8573,023346	-862,82	-10,06
DEZ/01	8,9671	5255,87	7840,830195	11697,14	8797,189005	-956,36	-10,87
TOTAL		61712,38	65198,15	86272,59	114862,24	-7178,55	-7,68
RAIZ QUADRADA DO EQM - 3 MESES						468,28	
RAIZ QUADRADA DO EQM - 6 MESES						536,05	
RAIZ QUADRADA DO EQM - 12 MESES						778,18	

Tabela 4 22 – Previsões do modelo ARIMA (1, 1, 2)(1, 1, 1)<sub>12</sub>

MESES	Previsão do Modelo (LRCONTR)	Limite 95% abaixo	Previsão (C <sub>P</sub> ) milhões de R\$	Limite 95% acima	Valor Real Observado (C <sub>R</sub> ) milhões de R\$	ERRO ABS. (C <sub>P</sub> - C <sub>R</sub> )	ERRO PERC. (%)
JAN/01	8,8912	6193,16	7267,735235	8528,77	7691,299726	-423,56	-5,51
FEV/01	8,7567	5306,56	6353,111679	7606,06	6253,134291	99,98	1,60
MAR/01	8,8934	5963,42	7283,741853	8896,38	6892,246507	391,50	5,68
ABR/01	8,8698	5709,00	7113,858059	8864,41	7504,80402	-390,95	-5,21
MAI/01	8,8195	5321,44	6764,881339	8599,85	7520,087136	-755,21	-10,04
JUN/01	8,8447	5349,19	6937,522491	8997,48	7600,850542	-663,33	-8,73
JUL/01	8,8442	5136,87	6934,054597	9359,99	8032,396147	-1098,34	-13,67
AGO/01	8,8701	5167,27	7115,992537	9799,63	8122,945643	-1006,95	-12,40
SET/01	8,9304	5379,77	7558,288087	10619,00	7747,376444	-189,09	-2,44
OUT/01	8,9143	5189,02	7437,574005	10660,49	8715,791911	-1278,22	-14,67
NOV/01	8,9418	5228,08	7644,945580	11179,09	8573,023346	-928,08	-10,83
DEZ/01	8,9604	5220,77	7788,472228	11619,04	8797,189005	-1008,72	-11,47
TOTAL		61681,84	65164,56	86200,18	114730,19	-7250,97	-7,76
RAIZ QUADRADA DO EQM - 3 MESES						337,97	
RAIZ QUADRADA DO EQM - 6 MESES						500,97	
RAIZ QUADRADA DO EQM - 12 MESES						778,60	

## 4.9 - SÉRIE DA RECEITA TOTAL ADMINISTRADA PELA SRF (RECADM)

### 4.9.1 - Identificação

Conforme analisado no capítulo 3, o gráfico da série da Receita Total Administrada pela SRF ( $Z_{\text{RECADM}}$ ), apresentado na figura 3.34, mostra uma variância que cresce com o aumento do nível da série. Para estabilizar esta variância utilizar-se-á o logaritmo natural da série original, de tal forma que o modelo ARIMA seja construído para a série transformada  $Z'_{\text{RECADM}} = \ln(Z_{\text{RECADM}})$ . O gráfico de  $Z'_{\text{RECADM}}$  está apresentado na figura 3.35 para o período jan/1989 a dez/2000 e na figura 3.37 para o período jul/1994 a dez/2000, que será o utilizado no presente estudo.

Para identificação do modelo utilizar-se-á, inicialmente, as Funções de Autocorrelação Amostral (FACA) e de Autocorrelação Parcial Amostral (FACPA). A figura 4.25 mostra a FACA, em preto, e a FACPA, em azul, para  $Z'_{\text{RECADM}}$  com  $d = D = 0$  (sem qualquer diferenciação). Verifica-se que as autocorrelações (FACA) são significativamente diferentes de zero até a defasagem 7, decaindo lentamente até a defasagem 18. Este comportamento sugere que a série  $Z'_{\text{RECADM}}$  seja não-estacionária, consistente com o gráfico apresentado na figura 3.37, em que os dados apresentam uma tendência ascendente.

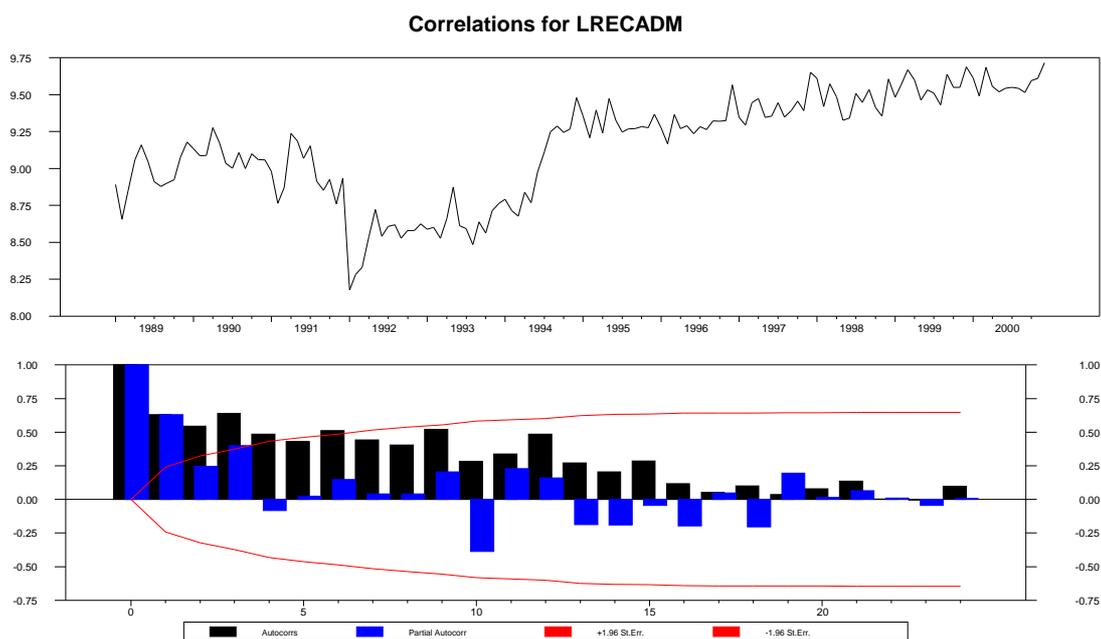


Figura 4.25 – Funções de Autocorrelação Amostral (FACA e FACPA) para a série  $Z'_{\text{RECADM}}$

A primeira diferença não sazonal  $W^*_{RECADM} = \nabla Z'_{RECADM}$  ( $d = 1$ ) está apresentada no gráfico 4.26 (superior). Observa-se que esta série tem uma média estável ao longo do tempo e, como relatado no capítulo 2, isto indica que a série diferenciada é estacionária. Entretanto, os dados da figura 4.26 (superior) mostram uma forte variação sazonal.

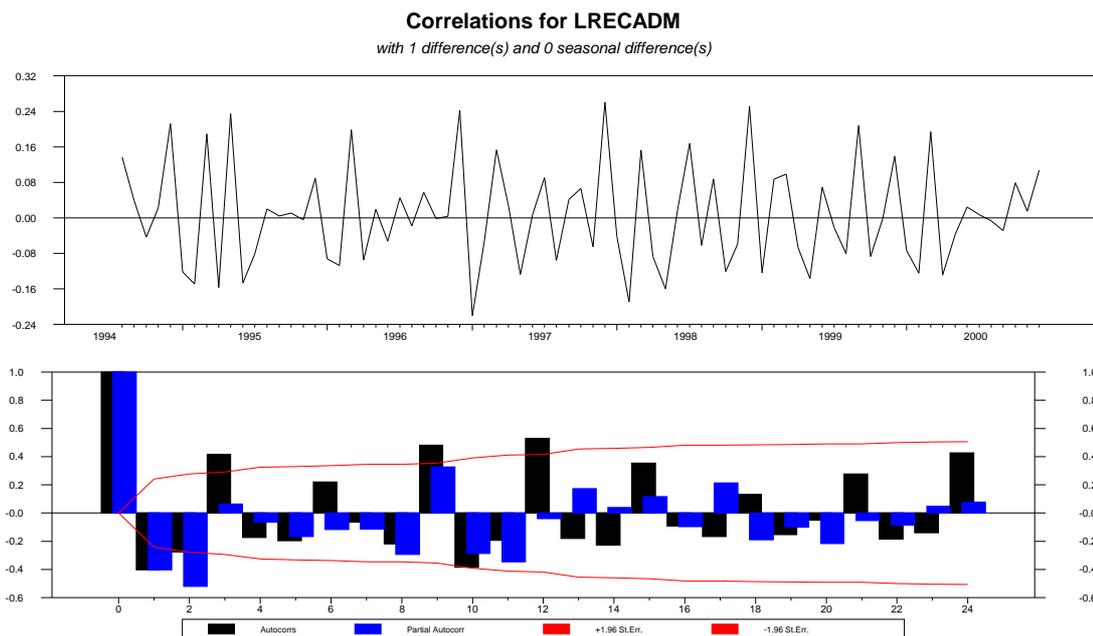


Figura 4.26 – Gráfico da Série Diferenciada  $W^*_{RECADM} = \nabla Z'_{RECADM}$  ( $d = 1$ ) e suas Funções de Autocorrelação Amostral (FACA e FACPA)

A FACA para a série diferenciada  $W^*_{RECADM} = \nabla Z'_{RECADM}$  ( $d = 1$ ) estão apresentadas na figura 4.26(inferior). Verifica-se ali que a autocorrelação cai a valores insignificamente próximos de zero bastante rapidamente, salvo para alguns valores. Isto reforça a conclusão do parágrafo anterior de que a série diferenciada é agora estacionária, nos termos do discutido no capítulo 2. Entretanto, as autocorrelações para as defasagens 12 e 24 são altas e não se observa uma queda significativa da defasagem 12 para a 24. Isto reforça a observação anteriormente feita de que os dados apresentam forte variação sazonal e sugere uma diferenciação sazonal ( $D = 1$ ) em adição à diferenciação não sazonal ( $d = 1$ ).

A figura 4.27(superior) apresenta o gráfico dos dados após sofrerem as duas diferenciações, sazonal ( $D = 1$ ) e não sazonal ( $d = 1$ ),  $W_{RECADM} = \nabla_{12} W^*_{RECADM} = \nabla_{12} \nabla Z'_{RECADM}$ . Esta nova série,  $W_{RECADM}$ , apresenta uma média estável e não apresenta mais

o forte padrão sazonal observado na série  $W^*_{RECADM}$ . A FACA e a FACPA para  $W_{RECADM}$  estão apresentadas na figura 4.27 (inferior).

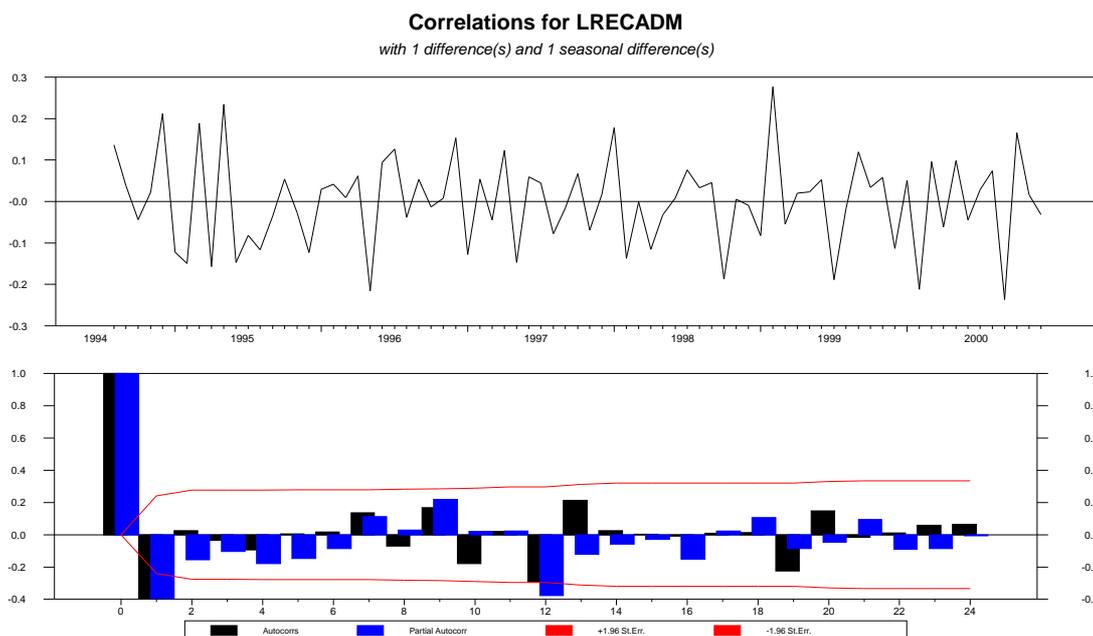


Figura 4.27 – Gráfico da Série Diferenciada  $W_{RECADM} = \nabla_{12}\nabla Z_{RECADM}$  ( $d = 1, D = 1$ ) e suas Funções de Autocorrelação Amostral (FACA e FACPA)

Comparando-se a FACA e a FACPA da série  $W_{RECADM}$  com as FAC e FACP teóricas apresentadas no capítulo 2, conclui-se que remanescem padrões sazonais e não sazonais que sugerem a inclusão de componentes de média móveis (MA) e/ou auto-regressivos (AR). No caso da parte não sazonal do modelo, as funções de autocorrelação sugerem a inclusão de um termo de média móvel - MA(1) e/ou um termo auto-regressivo - AR(1). No que se refere a parte sazonal do modelo, as funções de autocorrelação sugerem a inclusão de um termo auto-regressivo - SAR(1) E/ou de um termo de média móvel - SMA(1). Assim, combinando os componentes supracitados, poder-se-ia, inicialmente, analisar as seguintes opções de modelos:

**Modelo 1 - ARIMA(0, 1, 0)(0, 1, 1)<sub>12</sub>**

As autocorrelações residuais para os modelos em análise foram obtidas pelo RATS. Observa-se correlações residuais não significativamente diferentes de zero (autocorrelações das defasagens 1 e 9), indicando não ser este um modelo adequado para representar a série em estudo. Tentemos a inclusão de um termo auto-regressivo não sazonal - AR(1) e/ou um

termo de média móvel não sazonal - MA(1), conforme parece indicar as autocorrelações na primeira defasagem.

**Modelo 2 - ARIMA(1, 1, 0)(0, 1, 1)<sub>12</sub>**

Ainda se observa correlações residuais não significativamente diferentes de zero (autocorrelações das defasagens 2 e 4, por exemplo), indicando não ser este, ainda, um modelo adequado para representar a série em estudo. Tentemos, então, a opção de um termo de média móvel não sazonal.

**Modelo 3 - ARIMA(0, 1, 1)(0, 1, 1)<sub>12</sub>**

As autocorrelações da defasagem 10 teimam em persistirem elevadas, indicando não ser este, ainda, um modelo adequado para representar a série em estudo. Verifiquemos a inclusão de um termo auto-regressivo sazonal.

**Modelo 4 - ARIMA(0, 1, 1)(1, 1, 1)<sub>12</sub>**

Agora, verifica-se que todas as correlações residuais são significativamente nulas, indicando ser este um potencial modelo a representar a série em estudo. Tentemos o seguinte modelo: ARIMA(1, 1, 0)(0, 1, 1)<sub>12</sub>.

**Modelo 5 - ARIMA(1, 1, 0)(0, 1, 1)<sub>12</sub>**

Aqui também se observa correlações residuais não significativamente diferentes de zero (autocorrelações das defasagens 2 e 4, por exemplo), indicando que, provavelmente, este modelo também não seja adequado para representar a série em estudo. Tentemos incluir um termo de média móvel não-sazonal a este modelo.

**Modelo 6 - ARIMA(1, 1, 1)(0, 1, 1)<sub>12</sub>**

As autocorrelações da defasagem 10 voltam a ficar elevadas, indicando não ser este, ainda, um modelo adequado para representar a série em estudo. Por fim tentemos a inclusão de um termo auto-regressivo sazonal..

**Modelo 7 - ARIMA(1, 1, 1)(1, 1, 1)<sub>12</sub>**

Aqui também se verifica que todas as correlações residuais são significativamente nulas, indicando ser este um potencial modelo a representar a série em estudo.

Para se determinar quais os modelos que melhor se ajustam aos dados, optou-se por utilizar os critérios de seleção de Akaike e de Schwartz, quais sejam, o *Akaike Information Criterion(AIC)* e *Schwartz Bayesian Criterion (SBC ou BIC)*.



Os números entre parênteses são os valores absolutos da estatística  $t$  dos coeficientes e os entre colchetes são os desvios padrões das estimativas. Os coeficientes da parte auto-regressiva (AR) respeitam as condições para a estacionariedade do modelo e os coeficientes da parte de média móveis (MA) respeitam as condições de invertibilidade, nos termos do explicitado no capítulo 2.

#### 4.9.3 - Checagem de Diagnóstico

Verifica-se que nem todos os coeficientes estimados são estatisticamente significativos a 5%. Os coeficientes dos termos auto-regressivos sazonais dos modelos  $ARIMA(0,1,1)(1,1,1)_{12}$  e  $ARIMA(1,1,1)(1,1,1)_{12}$  não foram estatisticamente significativos. Os resíduos para os quatro modelos acima descritos, juntamente com a FACA e a FACPA residuais para cada um deles, foram obtidos pelo RATS. A FACA residual, em particular, mostra a satisfatoriedade do ajuste para o modelo  $ARIMA(1,1,1)(0,1,1)_{12}$ , indicando, claramente, que todas as autocorrelações residuais são estatisticamente nulas. Já para o caso do modelo  $ARIMA(0,1,1)(0,1,1)_{12}$  as autocorrelações residuais para as algumas defasagens se mostraram estatisticamente diferentes de zero, conforme já discutido anteriormente..

Para reforçar possível satisfatoriedade do ajuste do modelo remanescente, a estatística  $\chi^2$  se mostrou insignificante ao nível de 5% para os graus de liberdade considerados (51) e tanto o gráfico dos resíduos padronizados ( $a_t/\sigma_a$ ), traçado na figura 2.9.1 do Apêndice 2, quanto o teste de normalidade de Jarque-Bera sugerem que os choques devem ser normalmente distribuídos. Devemos observar que no gráfico da figura 2.9.1 nenhum valor foi superior a três vezes o desvio padrão. Isso reforça que os choques apresentam uma distribuição perfeitamente normal. Utilizaremos, assim, o modelo  $ARIMA(1,1,1)(0,1,1)_{12}$  para as previsões.

#### 4.9.4 - Previsão

As previsões para o único modelo selecionado estão mostradas abaixo nas tabelas 4.24, juntamente com o valor da raiz quadrada do erro quadrático médio de previsão (RQEQMP). Conforme se pode observar, foram calculados as RQEQMP para previsões para três, seis e doze passos à frente. A análise dos resultados a ser realizada no capítulo 5 utilizará essa medida para a comparação dos modelos.

Tabela 4.24 – Previsões do modelo ARIMA (1, 1, 1)(0, 1, 1)<sub>12</sub>

MESES	Previsão do Modelo (LRECADM)	Limite 95% abaixo	Previsão (T <sub>P</sub> ) milhões de R\$	Limite 95% acima	Valor Real Observado (T <sub>R</sub> ) milhões de R\$	ERRO ABS. (T <sub>P</sub> – T <sub>R</sub> )	ERRO PERC. (%)
JAN/01	9,6377	13065,10	15332,039457	17992,31	16604,71171	-1272,67	-7,66
FEV/01	9,4555	10673,29	12778,252736	15298,35	12789,80791	-11,56	-0,09
MAR/01	9,6705	12971,37	15843,268678	19351,01	14747,18703	1096,08	7,43
ABR/01	9,6434	12374,58	15419,681625	19214,11	16338,92528	-919,24	-5,63
MAI/01	9,5839	11428,89	14528,971943	18469,94	15861,73698	-1332,77	-8,40
JUN/01	9,5930	11305,00	14661,788989	19015,32	15888,85745	-1227,07	-7,72
JUL/01	9,6275	11242,99	15176,447522	20486,06	17189,73709	-2013,29	-11,71
AGO/01	9,6301	11049,05	15215,957627	20954,32	17065,84892	-1849,89	-10,84
SET/01	9,6480	11025,87	15490,775551	21763,73	16134,57741	-643,80	-3,99
OUT/01	9,7158	11565,71	16577,472925	23760,98	18600,69309	-2023,22	-10,88
NOV/01	9,6363	10470,32	15310,589620	22388,44	17186,80717	-1876,22	-10,92
DEZ/01	9,7968	12049,76	17976,129178	26817,23	19462,71818	-1486,59	-7,64
TOTAL		131793,56	139221,95	184311,38	245511,80	-13560,23	-6,85
RAIZ QUADRADA DO EQM – 3 MESES						969,75	
RAIZ QUADRADA DO EQM – 6 MESES						1076,12	
RAIZ QUADRADA DO EQM – 12 MESES						1433,58	

## 5 – DISCUSSÃO DOS RESULTADOS

*Neste capítulo, discutir-se-á os resultados obtidos no capítulo anterior, comparando as previsões obtidas pelo método de Box-Jenkins com aquelas resultantes do método de indicadores utilizado pela Secretaria da Receita Federal, determinando-se quais seriam os melhores modelos para previsão dos valores futuros de cada uma das séries tributárias aqui estudadas.*

Como medida de acurácia global das 12 previsões mensais, optou-se por utilizar a Raiz Quadrada do Erro Quadrado Médio de Previsão (RQEQMP) apresentada no capítulo 2. Para verificar a precisão dos métodos em relação ao número de passos à frente que se deseja prever, calculou-se a RQEQMP para três, seis e doze meses à frente. Com o intuito de simplesmente avaliar a qualidade individual das previsões (mês a mês e anual), calculou-se o erro da previsão para cada mês de 2001 e para toda a arrecadação anual. O erro de previsão é definido como a diferença percentual entre o valor observado e o previsto, ou seja:

$$\text{ERRO(\%)} = \frac{(\text{VALOR PREVISTO} - \text{VALOR OBSERVADO})}{\text{VALOR OBSERVADO}} \times 100$$

### 5.1 – SÉRIE DO IMPOSTO DE IMPORTAÇÃO – II

A tabela 5.1 abaixo traz os valores observados da série do Imposto de Importação (valores reais corrigidos pelo IGP-DI base dez/2000), juntamente com os valores da previsão elaborada pela SRF, os valores das previsões obtidas através da metodologia de Box-Jenkins, desenvolvida no capítulo 4 e, por fim, os valores para as medidas de acurácia (RQEQMP) para 3, 6 e 12 meses à frente.

A análise da tabela 5.1 mostra, primeiramente, que os dois modelos ARIMA selecionados produziram resultados globais melhores que o modelo de indicadores utilizado pela SRF, independentemente do número de meses considerados na previsão (3, 6 ou 12 passos à frente). Entre os dois modelos Box-Jenkins, o modelo ARIMA(1,1,0)(0,1,1)<sub>12</sub> apresentou melhores resultados para 3 meses (60,25<61,43) e 6 meses (60,72<62,25) à frente, enquanto o modelo ARIMA(0,1,1)(0,1,1)<sub>12</sub> foi superior para 12 passos à frente (88,69<91,29). De maneira geral, verifica-se que a medida RQEQMP cresce com o aumento de passos à frente a serem previstos, comprovando a afirmação feita no capítulo 2 de que os modelos ARIMA são mais precisos para poucos passos à frente.

Em relação às previsões mensais, consideradas individualmente, verifica-se que os dois modelos de Box-Jenkins apresentaram valores mais próximos dos observados em todos os meses de 2001, à exceção de fevereiro, setembro, novembro e dezembro. No que se refere à arrecadação anual do Imposto de Importação, os modelos propostos no presente estudo se mostraram mais exatos, com erros que giram em torno de 2%, para mais. Aqui, é conveniente relatar que o modelo da SRF subestimou as receitas em 4,07%, enquanto os modelos considerados superestimaram-na em, aproximadamente, 2%.

De uma forma geral, os dois modelos aqui propostos têm uma boa capacidade preditiva, sendo indiferente a escolha de qualquer um deles para a previsão da série em análise. Entretanto, como o modelo  $ARIMA(1,1,0)(0,1,1)_{12}$  apresentou um BIC menor (-17,76 contra -15,97 do outro), optar-se-á por ele.

Assim, após feitas as considerações pertinentes, decidiu-se pela utilização do modelo  $ARIMA(1,1,0)(0,1,1)_{12}$  para a previsão da série do Imposto de Importação. A estimação do modelo em referência está apresentada no capítulo 4 pela equação 4.3. O gráfico das previsões deste modelo está mostrado na figura 5.1.

Tabela 5.1 – Previsões para a Série do Imposto de Importação

MÊS	VALOR OBSERVADO (*)	PREVISÕES					
		SRF		$ARIMA(0,1,1)(0,1,1)_{12}$		$ARIMA(1,1,0)(0,1,1)_{12}$	
		VALOR (*)	ERRO (%)	VALOR (*)	ERRO (%)	VALOR (*)	ERRO (%)
JAN/01	724,383686	630,6781	-12,94	675,734333	-6,72	680,276954	-6,09
FEV/01	599,672735	543,7117	-9,33	667,874312	11,37	669,613044	11,66
MAR/01	864,785447	738,5365	-14,60	799,191027	-7,59	801,111389	-7,36
ABR/01	779,126172	713,9219	-8,37	752,875594	-3,37	752,649765	-3,40
MAI/01	898,763557	751,3793	-16,40	792,743727	-11,80	796,159864	-11,42
JUN/01	833,518873	741,4191	-11,05	832,472343	-0,13	835,474645	0,23
JUL/01	891,449909	786,3775	-11,79	855,683017	-4,01	852,352352	-4,39
AGO/01	951,736057	865,8299	-9,03	904,063860	-5,01	903,702307	-5,05
SET/01	773,984209	772,5649	-0,18	879,804743	13,67	883,772786	14,18
OUT/01	900,747094	926,4824	2,86	914,063460	1,48	918,277837	1,95
NOV/01	761,757844	889,6815	16,79	916,809768	20,35	922,973017	21,16
DEZ/01	596,130180	765,4334	28,40	775,493810	30,09	785,091287	31,70
TOTAL	9576,055762	9126,016	-4,70	9766,809993	1,99	9801,455247	2,35
<b>RQEQMP – 3 MESES</b>		96,35208		61,43		60,25	
<b>RQEQMP – 6 MESES</b>		101,9046		62,25		60,72	
<b>RQEQMP – 12 MESES</b>		102,64		88,69		91,29	

(\*) Valores expressos em Milhões de R\$.

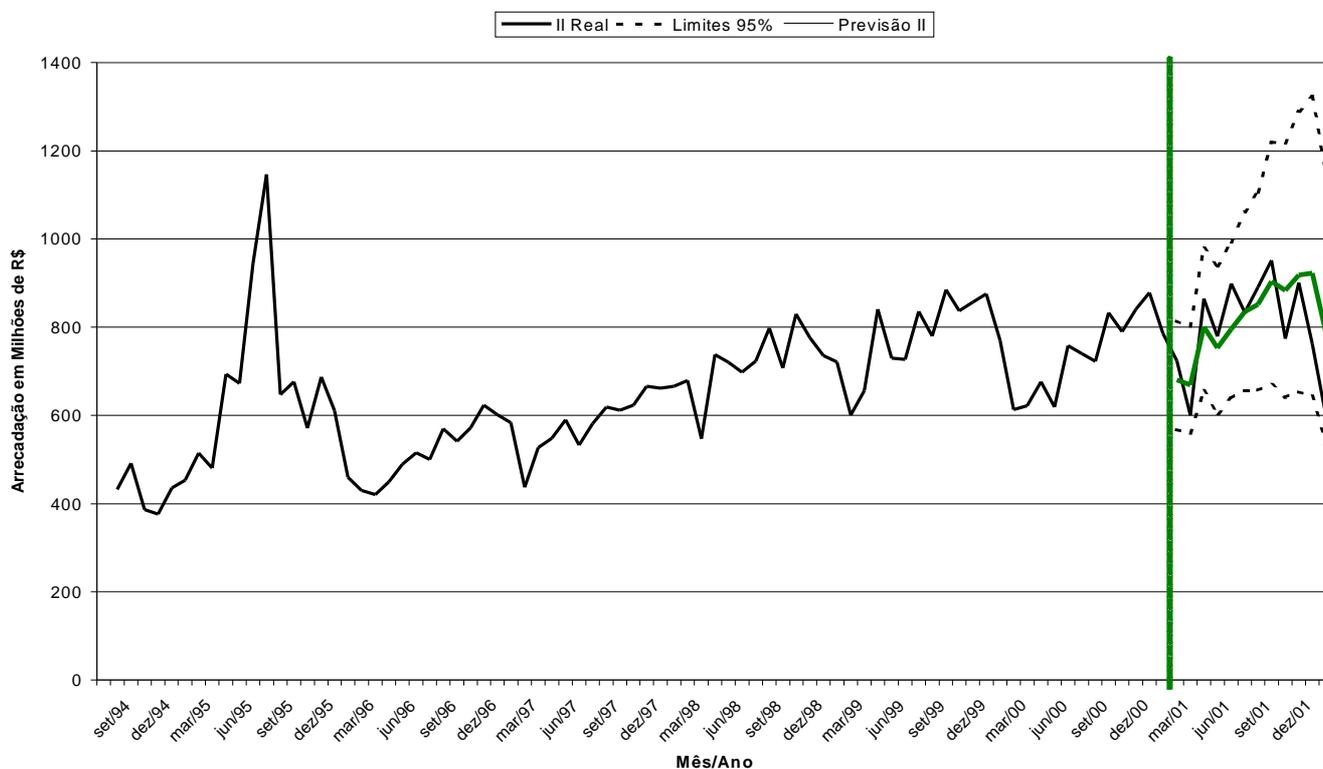


Fig. 5.1 – Previsões do Modelo ARIMA Selecionado para o Imposto de Importação

### 5.2 - SÉRIE DO IMPOSTO SOBRE PRODUTOS INDUSTRIALIZADOS - IPI

A tabela 5.2 abaixo traz os valores observados da série do Imposto sobre Produtos Industrializados (valores reais corrigidos pelo IGP-DI base dez/2000), juntamente com os valores da previsão elaborada pela SRF e os valores das previsões obtidas através da metodologia de Box-Jenkins, desenvolvida no capítulo 4, e, por fim, os valores para as medidas de acurácia (RQE QMP) para 3, 6 e 12 meses à frente.

A análise da tabela 5.2 mostra, primeiramente, que o modelo ARIMA selecionado produziu resultados globais melhores que o modelo de indicadores utilizado pela SRF, independentemente do número de meses considerados na previsão (3, 6 ou 12 passos à frente). Esperava-se que a medida RQE QMP crescesse com o aumento de passos à frente a serem previstos, comprovando a afirmação feita no capítulo 2 de que os modelos ARIMA são mais precisos para poucos passos à frente, mas ocorreu que o valor da RQE QMP para 12 passos à frente foi inferior ao valor para 6 passos à frente.

Tabela 5.2 – Previsões para a Série do Imposto sobre Produtos Industrializados

MÊS	VALOR OBSERVADO (*)	PREVISÕES			
		SRF		ARIMA(1,1,0)(2,1,0) <sub>12</sub>	
		VALOR (*)	ERRO (%)	VALOR (*)	ERRO (%)
JAN/01	1476,73234	1562,591	5,81	1477,342286	0,04
FEV/01	1614,843449	1632,874	1,12	1607,603758	-0,45
MAR/01	1485,339792	1623,93	9,33	1542,099379	3,82
ABR/01	1782,864213	1908,579	7,05	1650,774746	-7,41
MAI/01	1739,109891	1730,003	-0,52	1646,159042	-5,34
JUN/01	1781,300754	1764,317	-0,95	1715,912006	-3,67
JUL/01	1740,241092	1743,035	0,16	1662,204479	-4,48
AGO/01	1709,468182	1788,938	4,65	1648,630134	-3,56
SET/01	1744,75232	1841,721	5,56	1679,413661	-3,74
OUT/01	1773,602727	1752,683	-1,18	1737,321781	-2,05
NOV/01	1766,696247	1759,578	-0,40	1744,459422	-1,26
DEZ/01	1913,632225	1952,082	2,00	1869,076990	-2,33
<b>TOTAL</b>	<b>20528,58323</b>	<b>21060,33</b>	<b>2,59</b>	<b>19980,997684</b>	<b>-2,67</b>
<b>RQEOMP – 3 MESES</b>		94,70		33,04	
<b>RQEOMP – 6 MESES</b>		84,73		74,88	
<b>RQEOMP – 12 MESES</b>		71,16		65,51	

(\*) Valores expressos em Milhões de R\$.

Em relação às previsões mensais, consideradas individualmente, verifica-se que o modelo de Box-Jenkins apresentou valores mais próximos dos observados nos meses janeiro, fevereiro, março, agosto e setembro de 2001. Nos demais meses, a previsão da SRF se aproximou mais do observado. No que se refere à arrecadação anual do Imposto sobre Produtos Industrializados, o modelo ARIMA proposto no presente estudo se mostrou um pouco menos exato que o método de indicadores da SRF, com erros que giram em torno de 2%, para mais ou para menos. Aqui, é conveniente relatar que o modelo da SRF superestimou as receitas em 2,59%, enquanto o modelo considerado subestimou-as em 2,67%.

Assim, após feitas as considerações pertinentes, decidiu-se pela utilização do modelo ARIMA(1,1,0)(2,1,0)<sub>12</sub> para a previsão da série do Imposto sobre Produtos Industrializados. A estimação do modelo em referência está apresentada no capítulo 4 pela equação 4.9. O gráfico das previsões deste modelo está mostrado na figura 5.2.

### 5.3 - SÉRIE DO IMPOSTO DE RENDA TOTAL - IR

A tabela 5.3 abaixo traz os valores observados da série do Imposto de Renda Total (valores reais corrigidos pelo IGP-DI base dez/2000), juntamente com os valores da previsão elaborada pela SRF, os valores das previsões obtidas através da metodologia de Box-Jenkins,

desenvolvida no capítulo 4 e, por fim, os valores para as medidas de acurácia (RQEQMP) para 3, 6 e 12 meses à frente.

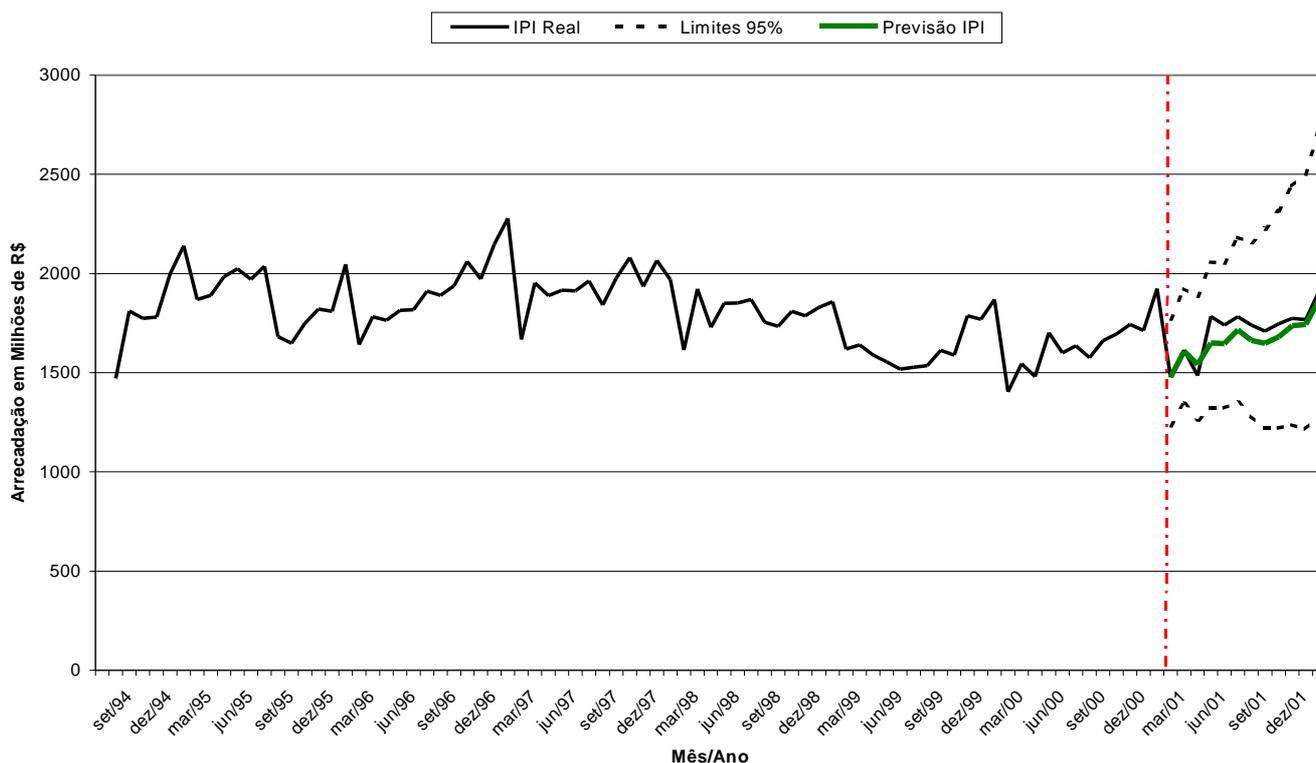


Fig. 5.2 – Previsões do Modelo ARIMA selecionado para o Imposto sobre Produtos Industrializados

A análise da tabela 5.3 mostra, primeiramente, que os dois modelos ARIMA selecionados produziram resultados globais melhores que o modelo de indicadores utilizado pela SRF, independentemente do número de meses considerados na previsão (3, 6 ou 12 passos à frente). Entre os dois modelos Box-Jenkins, o modelo ARIMA(2,1,0)(0,1,1)<sub>12</sub> apresentou melhores resultados para 3 meses à frente (547,82<561,20), enquanto o modelo ARIMA(0,1,1)(0,1,1)<sub>12</sub> foi superior para 6 (576,36<581,17) e 12 passos à frente (743,52<821,24). De maneira geral, verifica-se que a medida RQEQMP cresce com o aumento de passos à frente a serem previstos, comprovando a afirmação feita no capítulo 2 de que os modelos ARIMA são mais precisos para poucos passos à frente.

Em relação às previsões mensais, consideradas individualmente, verifica-se que os dois modelos de Box-Jenkins apresentaram valores mais próximos dos observados em todos os meses de 2001, à exceção de maio, julho, agosto e novembro. No que se refere à arrecadação anual do Imposto de Renda, o modelo ARIMA(0,1,1)(0,1,1)<sub>12</sub> proposto no presente estudo se mostrou mais exato que o da SRF, com erro de 10,79%, para menos, enquanto o modelo

ARIMA(2,1,0)(0,1,1)<sub>12</sub> se mostrou menos exato, com erro de 12,18% para menos. Aqui, é conveniente relatar que tanto o modelo da SRF quanto os modelos considerados subestimaram a arrecadação anual em mais de 10%.

De uma forma geral, os dois modelos aqui propostos têm uma boa capacidade preditiva, sendo indiferente a escolha de qualquer um deles para a previsão da série em análise. Entretanto, como o modelo ARIMA(0,1,1)(0,1,1)<sub>12</sub> apresentou um BIC menor (6,76 contra 15,24 do outro), optar-se-á por ele.

Assim, após feitas as considerações pertinentes, decidiu-se pela utilização do modelo ARIMA(0,1,1)(0,1,1)<sub>12</sub> para a previsão da série do Imposto de Renda Total. A estimação do modelo em referência está apresentada no capítulo 4 pela equação 4.11. O gráfico das previsões deste modelo está mostrado na figura 5.3.

Tabela 5.3 – Previsões para a Série do Imposto de Renda Total

MÊS	VALOR OBSERVADO (*)	PREVISÕES					
		SRF		ARIMA(0,1,1)(0,1,1) <sub>12</sub>		ARIMA(2,1,0)(0,1,1) <sub>12</sub>	
		VALOR (*)	ERRO (%)	VALOR (*)	ERRO (%)	VALOR (*)	ERRO (%)
JAN/01	6224,726267	5541,138	-10,98	5555,812744	-10,75	5581,986525	-10,33
FEV/01	3993,638758	4559,131	14,16	4503,003961	12,75	4427,542639	10,86
MAR/01	5051,099017	5730,808	13,46	4563,293954	-9,66	4504,355065	-10,82
ABR/01	5840,220643	5382,062	-7,84	5515,954554	-5,55	5553,590864	-4,91
MAI/01	5213,429848	4937,309	-5,30	4566,946050	-12,40	4516,081626	-13,38
JUN/01	5256,170647	4370,216	-16,86	4531,462436	-13,79	4509,312582	-14,21
JUL/01	6031,905072	5268,825	-12,65	5137,387702	-14,83	5127,123194	-15,00
AGO/01	5758,03735	5004,193	-13,09	5254,289134	-8,75	4512,018981	-21,64
SET/01	5324,477202	4496,385	-15,55	4787,674108	-10,08	4777,152802	-10,28
OUT/01	6652,519584	5499,774	-17,33	5766,381271	-13,32	5774,459858	-13,20
NOV/01	5585,62473	4422,589	-20,82	4392,702850	-21,36	4372,980097	-21,71
DEZ/01	7632,816612	5825,631	-23,68	6588,625288	-13,68	6558,387213	-14,08
TOTAL	68564,66573	61038,06	-10,98	61163,534052	-10,79	60214,991446	-12,18
<b>RQEQMP – 3 MESES</b>		645,26		561,20		547,82	
<b>RQEQMP – 6 MESES</b>		621,85		576,36		581,17	
<b>RQEQMP – 12 MESES</b>		917,60		743,52		821,24	

(\*) Valores expressos em Milhões de R\$.

#### 5.4 - SÉRIE OUTROS IMPOSTOS - OI

A tabela 5.4 abaixo traz os valores observados da série Outros Impostos (valores reais corrigidos pelo IGP-DI base dez/2000), juntamente com os valores da previsão elaborada pela SRF e os valores das previsões obtidas através da metodologia de Box-Jenkins, desenvolvida

no capítulo 4, e, por fim, os valores para as medidas de acurácia (RQEQMP) para 3, 6 e 12 meses à frente.

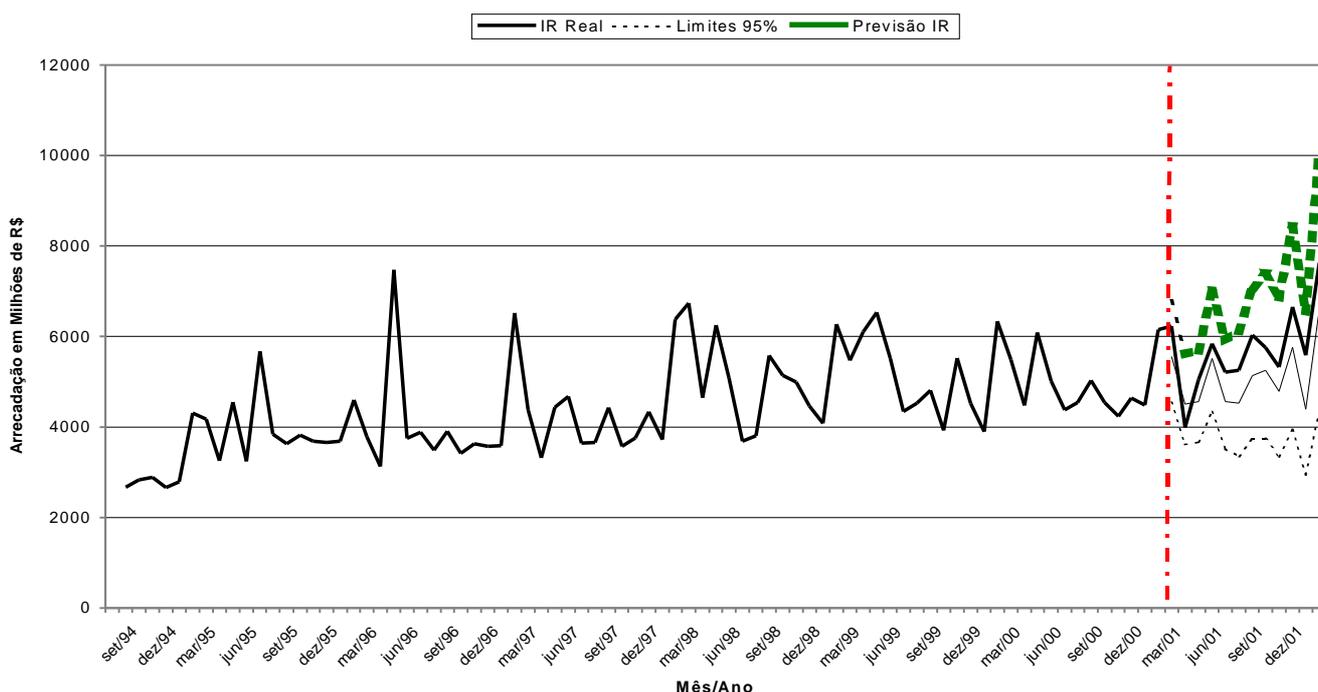


Fig. 5.3 – Previsões do Modelo ARIMA Selecionado para o Imposto de Renda

A análise da tabela 5.4 mostra, primeiramente, que o modelo ARIMA selecionado produziu resultados globais melhores que o modelo de indicadores utilizado pela SRF para o caso de previsões 3 e 6 passos à frente. No entanto, para 12 meses teve um resultado inferior. Esperava-se que a medida RQEQMP crescesse com o aumento de passos à frente a serem previstos, comprovando a afirmação feita no capítulo 2 de que os modelos ARIMA são mais precisos para poucos passos à frente, mas ocorreu que o valor da RQEQMP para 6 passos à frente foi inferior ao valor para 3 passos à frente.

Em relação às previsões mensais, consideradas individualmente, verifica-se que o modelo de Box-Jenkins apresentou valores mais próximos dos observados nos meses janeiro e março a julho de 2001. Nos demais meses, a previsão da SRF se aproximou mais do observado. Observa-se um erro gigantesco para o mês de fevereiro, tendo tal discrepância, obviamente, contribuído para piorar a medida de acurácia do modelo. No que se refere à arrecadação anual dos Outros Impostos, o modelo ARIMA proposto no presente estudo se mostrou menos exato que o método de indicadores da SRF, com erros que giram em torno de

9,5%, para menos. Aqui, é conveniente relatar que tanto o modelo da SRF quanto o modelo considerado subestimaram a arrecadação anual.

Tabela 5 4 – Previsões para a Série Outros Impostos

MÊS	VALOR OBSERVADO (*)	PREVISÕES			
		SRF		ARIMA(//1,5//,1,0)	
		VALOR (*)	ERRO (%)	VALOR (*)	ERRO (%)
JAN/01	332,1853622	254,7343	-23,32	303,370000	-8,67
FEV/01	217,9645026	230,9267	5,95	293,140000	34,49
MAR/01	327,0420982	243,0585	-25,68	312,410000	-4,47
ABR/01	282,7306577	244,2568	-13,61	292,910000	3,60
MAI/01	354,9733344	292,5542	-17,58	306,360000	-13,69
JUN/01	302,6540003	282,126	-6,78	305,440000	0,92
JUL/01	310,6757221	309,2421	-0,46	308,420000	-0,73
AGO/01	370,4110413	397,0282	7,19	302,570000	-18,32
SET/01	421,0535141	482,819	14,67	309,020000	-26,61
OUT/01	398,2988742	427,5683	7,35	304,120000	-23,65
NOV/01	349,0111388	324,1385	-7,13	305,290000	-12,53
DEZ/01	362,1735324	326,3125	-9,90	304,250000	-15,99
TOTAL	4029,173778	3814,765	-5,32	3647,300000	-9,48
<b>RQEQMP – 3 MESES</b>		66,38		47,24	
<b>RQEQMP – 6 MESES</b>		56,30		39,09	
<b>RQEQMP – 12 MESES</b>		46,82		58,07	

(\*) Valores expressos em Milhões de R\$.

Assim, após feitas as considerações pertinentes, verifica-se que o modelo ARIMA(//1,5//,1,0) é bastante adequado para previsões até 6 passos à frente, sendo inadequado para a previsão global dos 12 meses. A estimação do modelo ARIMA em referência está apresentada no capítulo 4 pela equação 4.18. As previsões estão apresentadas no gráfico da figura 5.4.

É importante ressaltar que a SRF faz a previsão individual dos impostos que compõem a série aqui denominada Outros Impostos, e, obviamente, isso deve ter influenciado demasiadamente os resultados.

## 5.5 - SÉRIE RECEITA TOTAL DE IMPOSTOS - RIMP

A tabela 5.5 abaixo traz os valores observados da série Receita Total de Impostos (valores reais corrigidos pelo IGP-DI base dez/2000), juntamente com os valores da previsão elaborada pela SRF, os valores das previsões obtidas através da metodologia de Box-Jenkins, desenvolvida no capítulo 4 e, por fim, os valores para as medidas de acurácia (RQEQMP) para 3, 6 e 12 meses à frente.

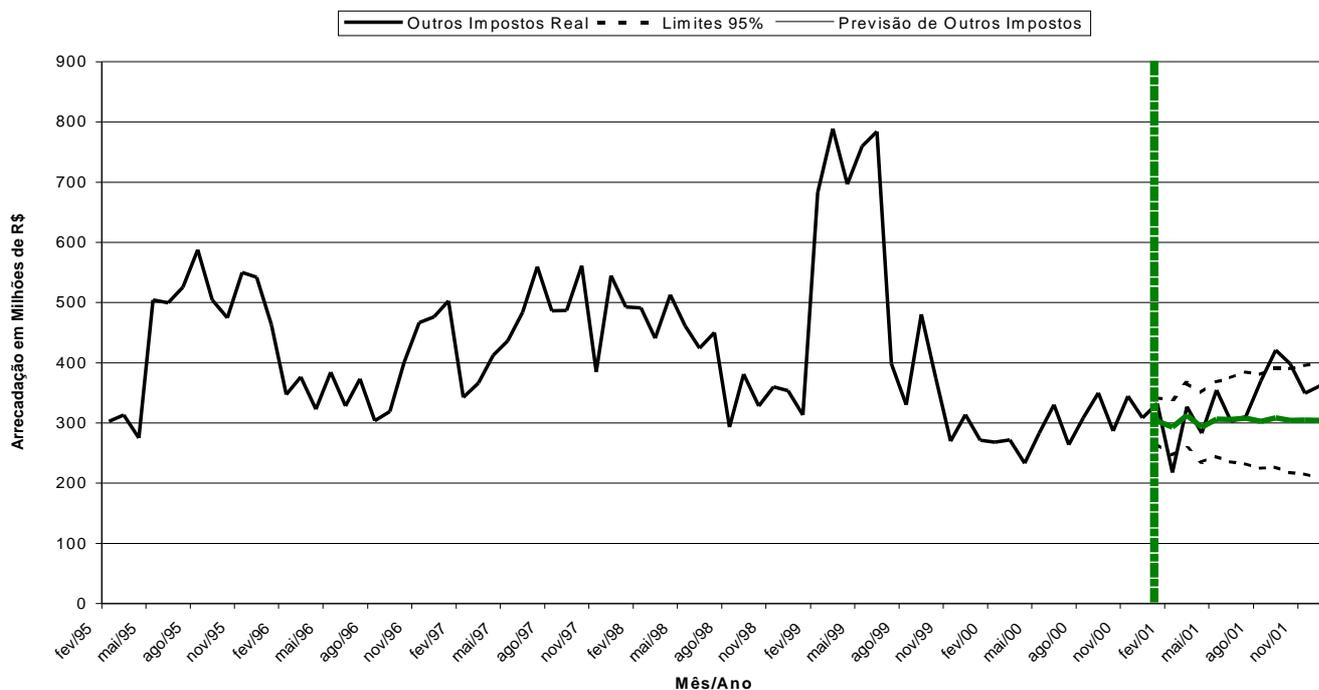


Fig. 5.4 – Previsões do Modelo ARIMA Selecionado para a Série Outros Impostos

A análise da tabela 5.5 mostra, primeiramente, que os três modelos ARIMA selecionados produziram resultados globais melhores que o modelo de indicadores utilizado pela SRF, independentemente do número de meses considerados na previsão (3, 6 ou 12 passos à frente). Entre os três modelos Box-Jenkins, o modelo ARIMA(2,1,0)(0,1,1)<sub>12</sub> apresentou melhores resultados para 3, 6 e 12 passos à frente. De maneira geral, verifica-se que a medida RREQMP cresce com o aumento de passos à frente a serem previstos, comprovando a afirmação feita no capítulo 2 de que os modelos ARIMA são mais precisos para poucos passos à frente.

Em relação às previsões mensais, consideradas individualmente, verifica-se que os três modelos de Box-Jenkins apresentaram valores mais próximos dos observados nos meses de janeiro a março, junho, setembro, outubro e dezembro de 2001. Nos demais meses, a previsão da SRF se aproximou mais do observado. No que se refere à arrecadação anual da Receita Total de Impostos, o modelo ARIMA(2,1,0)(0,1,1)<sub>12</sub> proposto no presente estudo se mostrou mais exato que o da SRF, com erro de 7,05% para menos, enquanto os demais modelos se mostraram menos exatos, com erros de 7,47% e 7,18% para menos. Aqui, é conveniente relatar que tanto o modelo da SRF quanto os modelos considerados subestimaram a arrecadação anual em mais de 7 %.

De uma forma geral, os três modelos aqui propostos têm uma boa capacidade preditiva, sendo indiferente a escolha de qualquer um deles para a previsão da série em análise. Entretanto, como o modelo  $ARIMA(1,1,1)(0,1,1)_{12}$  apresentou um BIC menor (-33,41 contra -33,40 e -26,55 dos outros), optar-se-á por ele.

Assim, após feitas as considerações pertinentes, decidiu-se pela utilização do modelo  $ARIMA(1,1,1)(0,1,1)_{12}$  para a previsão da série da Receita Total de Impostos. A estimação do modelo em referência está apresentada no capítulo 4 pela equação 4.20. As previsões estão apresentadas no gráfico da figura 5.5.

Tabela 5.5 – Previsões para a Série Receita Total de Impostos

MÊS	VALOR OBS. (*)	PREVISÕES							
		SRF		ARIMA(0,1,1)(0,1,1)		ARIMA(1,1,1)(0,1,1)		ARIMA(2,1,0)(0,1,1)	
		VALOR (*)	ERRO (%)	VALOR (*)	ERRO (%)	VALOR (*)	ERRO (%)	VALOR (*)	ERRO (%)
JAN	8758,03	7989,171	-8,78	9212,35	5,19	9288,20	6,05	9228,94	5,38
FEV	6426,12	6966,757	8,41	6904,99	7,45	6901,54	7,40	6837,65	6,40
MAR	7728,27	8336,403	7,87	7407,14	-4,16	7404,18	-4,19	7483,83	-3,16
ABR	8684,94	8248,881	-5,02	7949,78	-8,46	7974,47	-8,18	7968,09	-8,25
MAI	8206,28	7711,288	-6,03	7385,69	-10,00	7404,18	-9,77	7390,13	-9,95
JUN	8173,64	7158,173	-12,42	7400,48	-9,46	7423,46	-9,18	7412,33	-9,31
JUL	8974,27	8107,517	-9,66	7845,54	-12,58	7871,47	-12,29	7860,46	-12,41
AGO	8789,65	8056,072	-8,35	7357,68	-16,29	7380,52	-16,03	7744,98	-11,89
SET	8264,27	7593,624	-8,11	7702,50	-6,80	7727,96	-6,49	7717,92	-6,61
OUT	9725,17	8606,685	-11,50	8943,66	-8,04	8977,71	-7,69	8863,53	-8,86
NOV	8463,09	7396,101	-12,61	7384,22	-12,75	7405,66	-12,49	7393,82	-12,63
DEZ	10504,75	8869,626	-15,57	9535,72	-9,22	9566,28	-8,93	9551,94	-9,07
TOT.	102698,48	95040,3	-7,46	95029,75	-7,47	95325,63	-7,18	95453,62	-7,05
<b>RQEQMP – 3 M</b>		646,34		423,81		451,72		387,67	
<b>RQEQMP – 6 M</b>		673,25		625,89		622,23		606,97	
<b>RQEQMP – 12 M</b>		890,78		851,68		835,17		797,63	

(\*) Valores expressos em Milhões de R\$.

## 5.6 - SÉRIE DA CPMF

A tabela 5.6 abaixo traz os valores observados da série da Contribuição Provisória sobre Movimentação Financeira (valores reais corrigidos pelo IGP-DI base dez/2000), juntamente com os valores da previsão elaborada pela SRF, os valores das previsões obtidas através da metodologia de Box-Jenkins, desenvolvida no capítulo 4 e, por fim, os valores para as medidas de acurácia (RQEQMP) para 3, 6 e 12 meses à frente.

A análise da tabela 5.6 mostra, primeiramente, que os dois modelos ARIMA selecionados produziram resultados globais melhores que o modelo de indicadores utilizado

pela SRF, independentemente do número de meses considerados na previsão (3, 6 ou 12 passos à frente). Entre os dois modelos Box-Jenkins, o modelo ARIMA(2,1,0) apresentou melhores resultados para 6 meses ( $249,06 < 270,86$ ) e 12 meses ( $344,51 < 371,63$ ) à frente, enquanto o modelo ARIMA(1,1,1) foi superior para 3 passos à frente ( $196,11 < 198,87$ ). De maneira geral, verifica-se que a medida RREQMP cresce com o aumento de passos à frente a serem previstos, comprovando a afirmação feita no capítulo 2 de que os modelos ARIMA são mais precisos para poucos passos à frente.

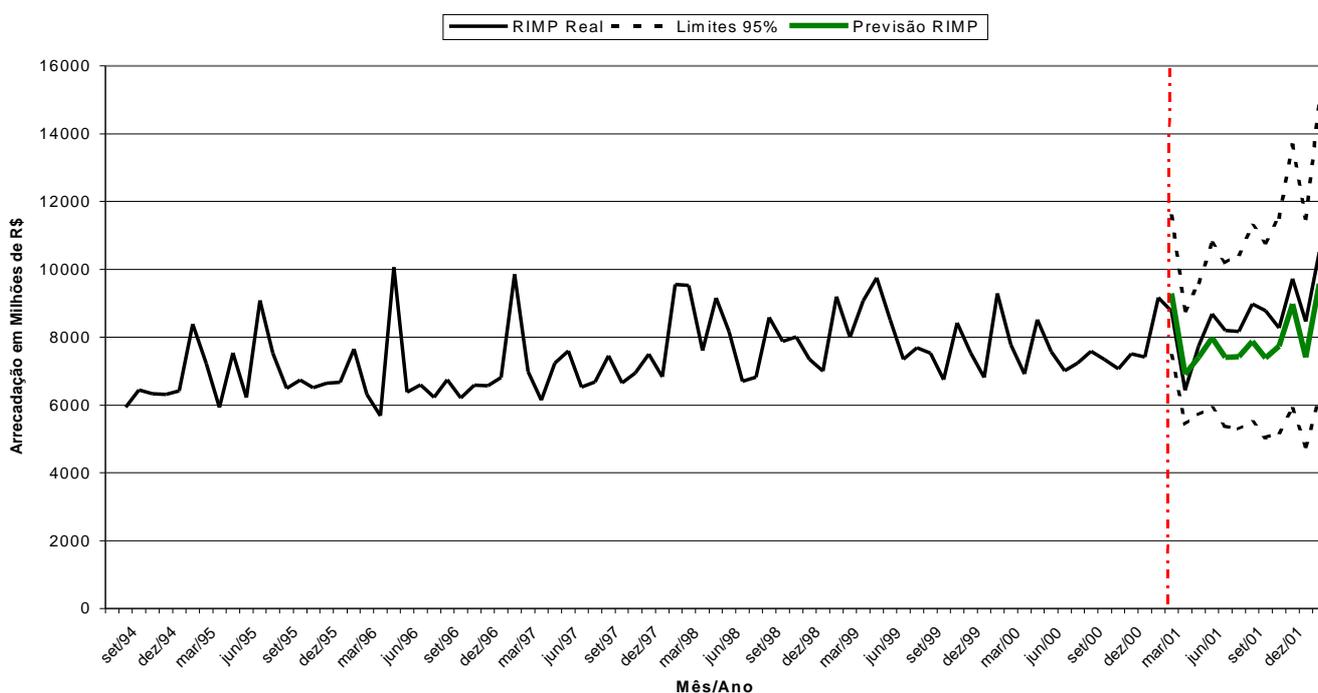


Fig. 5.5 – Previsões do Modelo ARIMA Selecionado para a Receita Total de Impostos

Em relação às previsões mensais, consideradas individualmente, verifica-se que os dois modelos de Box-Jenkins apresentaram valores mais próximos dos observados em todos os meses de 2001, à exceção de fevereiro, agosto e outubro. No que se refere à arrecadação anual da CPMF, os modelos propostos no presente estudo se mostraram mais exatos, com erros que giram em torno de 15% e 18%, para menos. Aqui, é conveniente relatar que o modelo da SRF subestimou as receitas em mais de 20%, enquanto os modelos considerados subestimaram-nas em 14,88% e 17,51%. Observa-se erros de previsões enormes, como por exemplo nos meses de fevereiro, agosto e outubro, sempre no sentido de subestimar a receita. Isso comprova que o crescimento da arrecadação desta importante fonte de recursos superou todas as expectativas da Receita Federal.

Tabela 5 6 – Previsões para a Série da Contribuição Provisória sobre Movimentação Financeira - CPMF

MÊS	VALOR OBSERVADO (*)	PREVISÕES					
		SRF		ARIMA(2,1,0)		ARIMA(1,1,1)	
		VALOR (*)	ERRO (%)	VALOR (*)	ERRO (%)	VALOR (*)	ERRO (%)
JAN/01	1329,566693	1285,757	-3,30	1367,719454	2,87	1291,423013	-2,87
FEV/01	791,7137668	1036,629	30,93	1107,876058	39,93	1098,718742	38,78
MAR/01	1400,175046	1017,604	-27,32	1268,892815	-9,38	1259,915583	-10,02
ABR/01	1362,009016	1108,698	-18,60	1334,083743	-2,05	1255,388041	-7,83
MAI/01	1794,427711	1308,024	-27,11	1309,237257	-27,04	1261,176129	-29,72
JUN/01	1427,064128	1101,605	-22,81	1295,303099	-9,23	1256,518399	-11,95
JUL/01	1529,937263	1148,047	-24,96	1312,251968	-14,23	1256,644057	-17,86
AGO/01	1841,00157	1380,504	-25,01	1288,842748	-29,99	1282,542866	-30,33
SET/01	1456,858808	1191,496	-18,21	1302,056218	-10,63	1268,512204	-12,93
OUT/01	1927,263244	1545,075	-19,83	1302,446893	-32,42	1261,176129	-34,56
NOV/01	1574,091398	1151,289	-26,86	1294,526150	-17,76	1256,644057	-20,17
DEZ/01	1757,27512	1205,389	-31,41	1302,056218	-25,90	1256,644057	-28,49
TOTAL	18191,38376	14480,12	-20,40	15485,292621	-14,88	15005,303277	-17,51
<b>RQEOMP – 3 MESES</b>		263,48		198,87		196,11	
<b>RQEOMP – 6 MESES</b>		320,14		249,06		270,86	
<b>RQEOMP – 12 MESES</b>		373,35		344,51		371,63	

(\*) Valores expressos em Milhões de R\$.

De uma forma geral, os dois modelos aqui propostos têm uma boa capacidade preditiva, sendo indiferente a escolha de qualquer um deles para a previsão da série em análise. Entretanto, como o modelo ARIMA(2,1,0) apresentou um BIC menor (-8,59 contra -3,40 do outro), optar-se-á por ele.

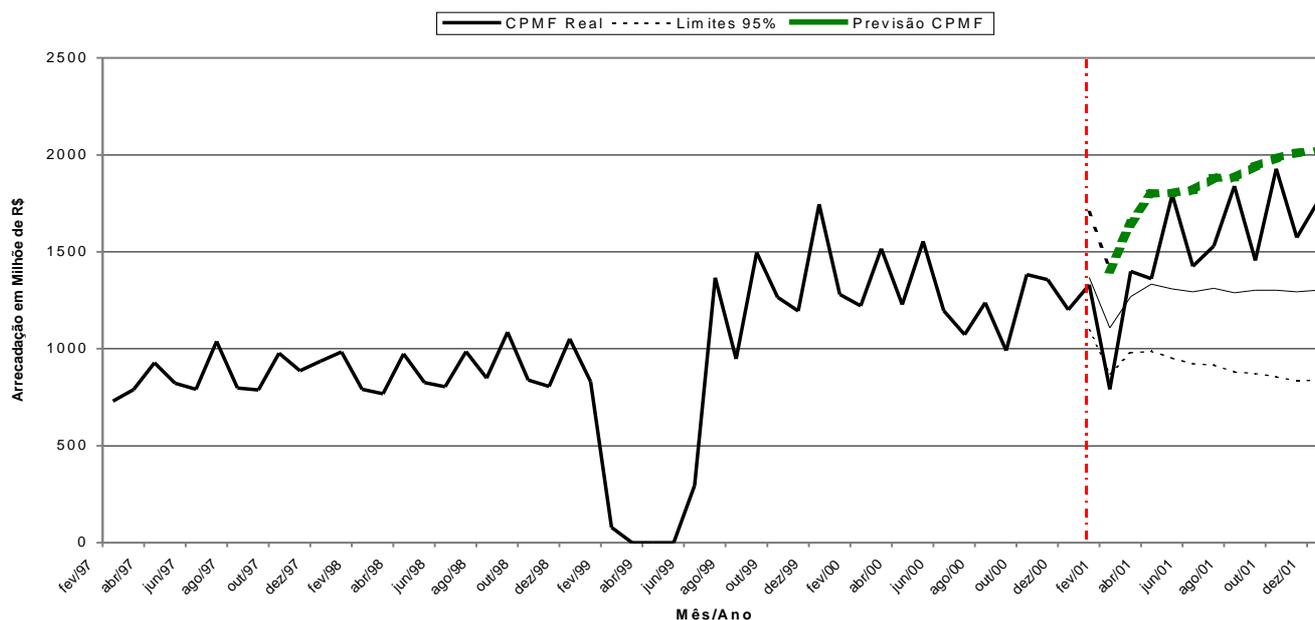


Fig. 5.6 – Previsões do Modelo ARIMA Selecionado para a CPMF

### 5.7 - SÉRIE RECEITA DE OUTRAS CONTRIBUIÇÕES - OCONTR

A tabela 5.7 abaixo traz os valores observados da série Receita de Outras Contribuições, nos termos definidos no capítulo 3 (valores reais corrigidos pelo IGP-DI base dez/2000), juntamente com os valores da previsão elaborada pela SRF e os valores das previsões obtidas através da metodologia de Box-Jenkins, desenvolvida no capítulo 4, e, por fim, os valores para as medidas de acurácia (RREQMP) para 3, 6 e 12 meses à frente.

A análise da tabela 5.7 mostra, primeiramente, que o modelo ARIMA selecionado produziu resultados globais piores que o modelo de indicadores utilizado pela SRF, independentemente do número de meses considerados na previsão (3, 6 ou 12 passos à frente). Esperava-se que a medida RREQMP crescesse com o aumento de passos à frente a serem previstos, comprovando a afirmação feita no capítulo 2 de que os modelos ARIMA são mais precisos para poucos passos à frente, mas ocorreu que o valor da RREQMP para 12 passos à frente foi inferior ao valor para 3 passos à frente.

Tabela 5.7 – Previsões para a Série Receita de Outras Contribuições

MÊS	VALOR OBSERVADO (*)	PREVISÕES			
		SRF		ARIMA(2,1,0)	
		VALOR (*)	ERRO (%)	VALOR (*)	ERRO (%)
JAN/01	1358,367148	976,9228	-28,08	994,460123	-26,79
FEV/01	989,8152434	1123,181	13,47	1227,579962	24,02
MAR/01	1184,473848	1454,588	22,80	1628,801912	37,51
ABR/01	1224,152773	1207,115	-1,39	1029,264759	-15,92
MAI/01	935,7180963	908,9411	-2,86	723,354919	-22,70
JUN/01	1016,474392	1021,134	0,46	747,324877	-26,48
JUL/01	1392,721092	1459,675	4,81	1114,320410	-19,99
AGO/01	1017,04048	1196,136	17,61	753,854968	-25,88
SET/01	945,4651254	1143,846	20,98	837,398448	-11,43
OUT/01	1494,6722	1528,553	2,27	1025,155923	-31,41
NOV/01	1257,855542	1160,757	-7,72	880,596924	-29,99
DEZ/01	1467,789037	1478,457	0,73	1022,596234	-30,33
TOTAL	14284,54498	14659,31	2,62	11984,709460	-16,10
<b>RREQMP – 3 MESES</b>		280,62		358,88	
<b>RREQMP – 6 MESES</b>		198,86		300,53	
<b>RREQMP – 12 MESES</b>		164,29		324,21	

(\*) Valores expressos em Milhões de R\$.

Em relação às previsões mensais, consideradas individualmente, verifica-se que o modelo de Box-Jenkins apresentou valor mais próximo do observado apenas no mês de setembro de 2001. Em todos os demais meses, a previsão da SRF se aproximou mais do observado. No que se refere à arrecadação anual de Outras Contribuições, o modelo ARIMA

proposto no presente estudo se mostrou muito menos exato que o método de indicadores da SRF, com erro que gira em torno de 16%, para menos. Aqui, é conveniente relatar que o modelo da SRF superestimou as receitas em 2,62%, enquanto o modelo considerado subestimou-as em 16,10%.

Assim, após feitas as considerações pertinentes, verifica-se que os modelos de Box-Jenkins se mostraram inadequados para a previsão da série em estudo. O gráfico da série, apresentado na figura 3.30, mostra uma possível justificativa para a não adequação da metodologia ARIMA para fins de previsão da série. Trata-se da presença de diversos “outliers”, além de mudanças estruturais bastante amplas. Se se optasse pela eliminação de todos os pontos discrepantes da série, ter-se-ia um número muito baixo de observações para se trabalhar, também inviabilizando a modelagem. A tentativa de prever os valores futuros desta série, mesmo desconsiderando suas variações, serviu para mostrar que o campo da previsão econômica é bastante amplo, provando que não existirá, obviamente, uma única metodologia que seja adequada à previsão de qualquer série econômica, demonstrando a necessidade de se ampliar a pesquisa na área.

### **5.8 - SÉRIE RECEITA TOTAL DE CONTRIBUIÇÕES - RCONT**

A tabela 5.8 abaixo traz os valores observados da série Receita Total de Contribuições (valores reais corrigidos pelo IGP-DI base dez/2000), juntamente com os valores da previsão elaborada pela SRF, os valores das previsões obtidas através da metodologia de Box-Jenkins, desenvolvida no capítulo 4 e, por fim, os valores para as medidas de acurácia (RQEQMP) para 3, 6 e 12 meses à frente.

A análise da tabela 5.8 mostra, primeiramente, que os dois modelos ARIMA selecionados produziram resultados globais melhores que o modelo de indicadores utilizado pela SRF, independentemente do número de meses considerados na previsão (3, 6 ou 12 passos à frente). Entre os dois modelos Box-Jenkins, o modelo  $ARIMA(1,1,2)(1,1,1)_{12}$  apresentou melhores resultados para 3 e 6 meses à frente, enquanto o modelo  $ARIMA(1,1,0)(0,1,1)_{12}$  foi superior para 12 passos à frente. De maneira geral, verifica-se que a medida RQEQMP cresce com o aumento de passos à frente a serem previstos, comprovando a afirmação feita no capítulo 2 de que os modelos ARIMA são mais precisos para poucos passos à frente.

Em relação às previsões mensais, consideradas individualmente, verifica-se que os dois modelos de Box-Jenkins apresentaram valores mais próximos dos observados nos meses de

janeiro, fevereiro, abril, setembro e novembro de 2001, enquanto o método da Receita Federal se mostrou melhor nos demais meses. No que se refere à arrecadação anual da Receita Total de Contribuições, os modelo ARIMA propostos no presente estudo se mostrou mais exato que o da SRF, com erro de cerca de 7,5%, para menos. Aqui, é conveniente relatar que tanto o modelo da SRF quanto os modelos considerados subestimaram a arrecadação anual em mais de 7%.

De uma forma geral, os dois modelos aqui propostos têm uma boa capacidade preditiva, sendo que o ARIMA(1,1,2)(1,1,1)<sub>12</sub> foi melhor no curto prazo. Mas permaneceria indiferente a escolha de qualquer um deles para a previsão da série em análise. Entretanto, como o modelo ARIMA(1,1,2)(1,1,1)<sub>12</sub> apresentou um BIC menor (-48,57 contra -44,42), optar-se-á por ele.

Assim, após feitas as considerações pertinentes, decidiu-se pela utilização do modelo ARIMA(1,1,2)(1,1,1)<sub>12</sub> para a previsão da série Receita Total de Contribuições. A estimação do modelo em referência está apresentada no capítulo 4 pela equação 4.34. As previsões estão apresentadas no gráfico da figura 5.7.

Tabela 5 8 – Previsões para a Série Receita Total de Contribuições - RCONT

MÊS	VALOR OBSERVADO (*)	PREVISÕES					
		SRF		ARIMA(1,1,0)(0,1,1) <sub>12</sub>		ARIMA(1,1,2)(1,1,1) <sub>12</sub>	
		VALOR (*)	ERRO (%)	VALOR (*)	ERRO (%)	VALOR (*)	ERRO (%)
JAN/01	7691,299726	6853,543	-10,89	7041,665515	-8,45	7267,735235	-5,51
FEV/01	6253,134291	6415,141	2,59	6403,500033	2,40	6353,111679	1,60
MAR/01	6892,246507	6597,756	-4,27	7354,002487	6,70	7283,741853	5,68
ABR/01	7504,80402	6938,929	-7,54	7129,525775	-5,00	7113,858059	-5,21
MAI/01	7520,087136	6884,004	-8,46	6796,071475	-9,63	6764,881339	-10,04
JUN/01	7600,850542	7067,691	-7,01	6967,418067	-8,33	6937,522491	-8,73
JUL/01	8032,396147	7197,564	-10,39	6939,604060	-13,60	6934,054597	-13,67
AGO/01	8122,945643	7447,995	-8,31	7052,235939	-13,18	7115,992537	-12,40
SET/01	7747,376444	7403,322	-4,44	7547,713887	-2,58	7558,288087	-2,44
OUT/01	8715,791911	7845,219	-9,99	7489,819669	-14,07	7437,574005	-14,67
NOV/01	8573,023346	7411,678	-13,55	7710,204576	-10,06	7644,945580	-10,83
DEZ/01	8797,189005	7889,863	-10,31	7840,830195	-10,87	7788,472228	-11,47
TOTAL	93451,14472	85952,71	-8,02	86272,591678	-7,68	86200,177690	-7,76
<b>RQEOMP – 3 MESES</b>		521,15		468,28		337,97	
<b>RQEOMP – 6 MESES</b>		551,34		536,05		500,97	
<b>RQEOMP – 12 MESES</b>		708,56		778,18		778,60	

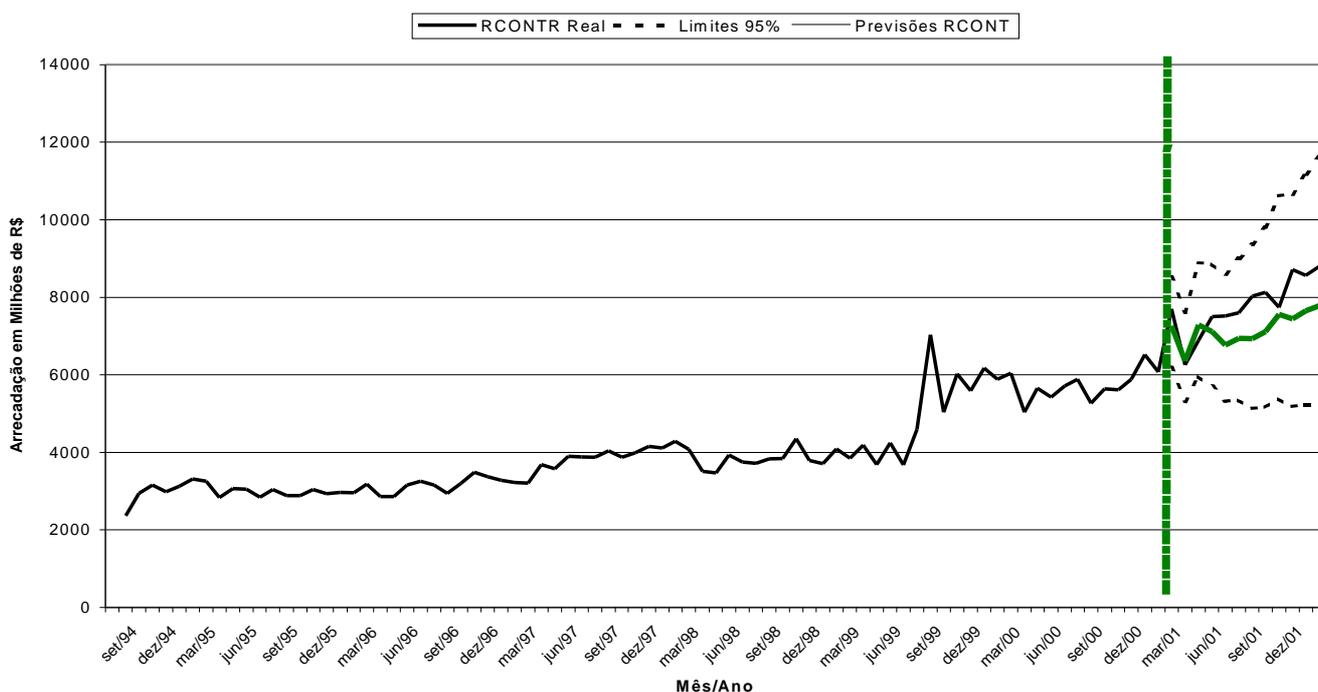


Fig. 5.7- Previsões do Modelo ARIMA Selecionado para a Receita Total de Contribuições

### 5.9 - SÉRIE RECEITA TOTAL ADMINISTRADA PELA SRF - RECADM

A tabela 5.9 abaixo traz os valores observados da série Receita Total Administrada pela SRF (valores reais corrigidos pelo IGP-DI base dez/2000), juntamente com os valores da previsão elaborada pela SRF e os valores das previsões obtidas através da metodologia de Box-Jenkins, desenvolvida no capítulo 4, e, por fim, os valores para as medidas de acurácia (RQEQMP) para 3, 6 e 12 meses à frente.

A análise da tabela 5.9 mostra, primeiramente, que o modelo ARIMA selecionado produziu resultados globais melhores que o modelo de indicadores utilizado pela SRF, independentemente do número de meses considerados na previsão (3, 6 ou 12 passos à frente). Aqui também se verifica que a medida RQEQMP cresce com o aumento de passos à frente a serem previstos, comprovando a afirmação feita no capítulo 2 de que os modelos ARIMA são mais precisos para poucos passos à frente.

Tabela 5.9 – Previsões para a Série Receita Total Administrada pela SRF

MÊS	VALOR OBSERVADO (*)	PREVISÕES			
		SRF		ARIMA(1,1,1)(0,1,1) <sub>12</sub>	
		VALOR (*)	ERRO (%)	VALOR (*)	ERRO (%)
JAN/01	16604,71171	14982,71	-9,77	15332,039457	-7,66
FEV/01	12789,80791	13510,96	5,64	12778,252736	-0,09
MAR/01	14747,18703	15073,4	2,21	15843,268678	7,43
ABR/01	16338,92528	15319,91	-6,24	15419,681625	-5,63
MAI/01	15861,73698	14722,48	-7,18	14528,971943	-8,40
JUN/01	15888,85745	14376,61	-9,52	14661,788989	-7,72
JUL/01	17189,73709	15448,69	-10,13	15176,447522	-11,71
AGO/01	17065,84892	15652,24	-8,28	15215,957627	-10,84
SET/01	16134,57741	15146,13	-6,13	15490,775551	-3,99
OUT/01	18600,69309	16619,72	-10,65	16577,472925	-10,88
NOV/01	17186,80717	14958,1	-12,97	15310,589620	-10,92
DEZ/01	19462,71818	16926,54	-13,03	17976,129178	-7,64
TOTAL	197871,6082	182737,5	-7,65	184311,375852	-6,85
<b>RQEQMP – 3 MESES</b>		1042,01		969,75	
<b>RQEQMP – 6 MESES</b>		1146,05		1076,12	
<b>RQEQMP – 12 MESES</b>		1560,20		1433,58	

(\*) Valores expressos em Milhões de R\$.

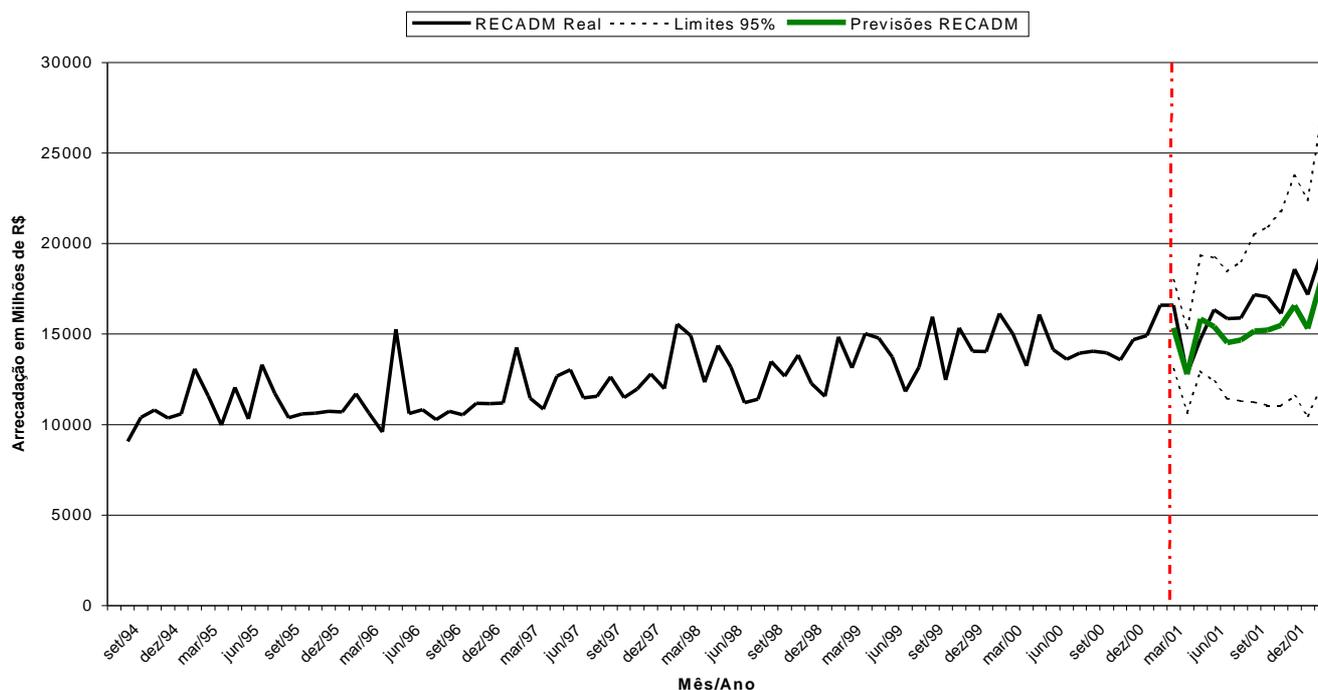


Fig. 5.8 – Previsões do Modelo ARIMA Selecionado para a Arrecadação Total

Em relação às previsões mensais, consideradas individualmente, verifica-se que o modelo de Box-Jenkins apresentou valores mais próximos dos observados em todos os meses

de 2001, à exceção de março, maio, julho, agosto e outubro. No que se refere à arrecadação anual da Receita Total Administrada pela SRF, o modelo ARIMA proposto no presente estudo se mostrou um pouco mais exato que o método de indicadores da SRF, com erros que giram em torno de 7%, para menos. Aqui, é conveniente relatar que tanto o modelo da SRF quanto o modelo ARIMA selecionado subestimaram a arrecadação total em cerca de 7%.

Assim, após feitas as considerações pertinentes, decidiu-se pela utilização do modelo ARIMA(1,1,1)(0,1,1)<sub>12</sub> para a previsão da série Receita Total Administrada pela SRF. A estimação do modelo em referência está apresentada no capítulo 4 pela equação 4.38. As previsões estão apresentadas abaixo no gráfico da figura 5.8.

## 6 – CONCLUSÃO E SUGESTÕES DE TRABALHOS FUTUROS

*Este capítulo final apresenta as conclusões gerais, resume os resultados e as contribuições deste trabalho, sugerindo, ao final, algumas possibilidades de estudos futuros.*

### 6.1 - CONCLUSÕES GERAIS

O objetivo principal deste trabalho foi aplicar a metodologia de Box-Jenkins (modelos ARIMA) na previsão de séries temporais, mais precisamente de séries tributárias federais. A principal motivação para o uso de *Modelos ARIMA* na previsão de séries de receitas públicas foi buscar melhores resultados que os apresentados pelos procedimentos convencionais adotados pela SRF, que na sua maioria são iminentemente qualitativos, possuindo dificuldades no tratamento de dados reais compostos de complexas relações não muito bem identificadas.

Inicialmente, desenvolveu-se uma revisão da literatura referente ao tratamento estatístico de séries temporais, incluindo, além dos conceitos gerais inerentes à análise das séries temporais, toda a teoria por trás do desenvolvimento da metodologia de Box-Jenkins, com a adição de discussões recentes em relação à confiabilidade e exatidão do método em questão. Contribuindo para ampliar a discussão iniciada na década de 70, Newbold & Granger (1974) realizaram diversas comparações entre os vários métodos de previsão com modelos univariados, onde mostraram a superioridade da modelagem ARIMA em relação a métodos de modelagem automático.

Conforme justificado no escopo do trabalho, a escolha de modelos de séries temporais univariados, particularmente aquele baseado na Metodologia de Box-Jenkins, em detrimento de modelos econométricos de regressão, se deu em função de que, em geral, a modelagem de uma regressão requer um amplo conhecimento das relações entre as variáveis do modelo, o que, no caso específico da arrecadação tributária, poderia gerar modelos subparametrizados ou superparametrizados, uma vez que tais séries são dependentes de muitas variáveis econômicas, algumas delas não totalmente identificáveis. Assim, a opção por equações de regressão poderia gerar modelos não muito bem especificados, dada a dificuldade de inter-relacionar todas as variáveis envolvidas. Em defesa da opção feita neste trabalho, verifica-se, através de conclusões do trabalho de Gourieroux & Monfort (1997), que, em uma comparação entre os modelos ARIMA e os de regressão, o erro relativo de previsão da modelagem ARIMA apresenta-se inferior em cerca de 2,5 pontos percentuais se comparados com os

obtidos pelos modelos de regressão.

Os resultados obtidos demonstraram a superioridade do método escolhido em relação ao método dos indicadores utilizado pela SRF, gerando previsões mais acuradas para 8 (oito) das 10 (dez) séries analisadas.

Verificou-se que o método atualmente utilizado pela Secretaria da Receita Federal, conhecido como método de indicadores, apesar de não fornecer previsões que sejam estatisticamente confiáveis, mostrou-se superior ao método em estudo para o caso de duas séries específicas, quais sejam, as séries da COFINS mais PIS (COFPIS) e Outras Contribuições Sociais (OCONTR). A justificativa para a melhor performance do método dos indicadores para estes casos particulares está no fato de que as séries em questão apresentaram um comportamento estrutural bastante variável, em função de, entre outros fatores, terem sido promovidas diversas alterações legais em variáveis significantes da arrecadação (alíquotas, bases de cálculo, etc.), que não foram, e não são, devidamente capturadas pela metodologia ARIMA, mas que são perfeitamente identificáveis na previsão quantitativa/qualitativa realizada pela Receita Federal. Isso se dá porque os métodos quantitativos não podem capturar informações importantes sobre os fatores que provavelmente estão alterando os padrões históricos da série, quando as circunstâncias econômicas, políticas ou administrativas estão em fluxo intenso.

Obviamente, não há nenhum método único para se projetar receitas. Assim sendo, os diferentes métodos pesquisados tendem a trabalhar melhor dependendo do tipo de receita que pretendem prever. Da mesma forma, não há nenhuma estrutura padrão no tempo sobre a qual se pode fundar todas as previsões. Elas variam de técnicas qualitativas relativamente informais a técnicas quantitativas altamente sofisticadas. Na previsão de receitas, mais sofisticado não significa necessariamente mais exato. Em geral, os responsáveis pela previsão usam uma variedade de técnicas, reconhecendo que algumas apresentam melhores resultados do que outras, dependendo da natureza da fonte de receita.

Métodos quantitativos mais simples também executam previsões tão boas quanto os métodos mais complexos (Makridakis, 1984). Finalmente, a abordagem de séries temporais normalmente apresenta melhor performance que a baseada na modelagem causal, ao menos no curto prazo, dado a incerteza associada a captura de todos os fatores econômicos relevantes que influenciam a geração da receita (Frank, 1993).

Finalmente, a previsão de receitas está intimamente ligada ao processo de política pública e, portanto, sujeita ao escrutínio e até mesmo à pressão política, dificultando a utilização de métodos puramente estatísticos/matemáticos.

Este breve estudo sobre modelos de previsão de receitas demonstram o fato de que a previsão é um campo amplo e inóspito da economia. As idiossincrasias e as variações não podem ser representadas completamente em um simples estudo. Para aqueles que lidam com as finanças públicas, o tópico é de uma importância crescente, especialmente em nível federal. Enquanto algumas destas técnicas provavelmente estão além da potencialidade dos gerentes governamentais, os desenvolvimentos de softwares e o auxílio de computadores disponíveis nas instituições aumentam a plausibilidade de se utilizar estas ferramentas, mesmo em unidades menores do governo.

## **6.2 - TRABALHOS FUTUROS**

Conforme já confessado, este trabalho não pretende, de forma alguma, esgotar o tema relativo à previsão da arrecadação tributária, nem mesmo inferir que o método aqui adotado seja definitivamente o melhor caminho para a Receita Federal em termos de metodologia de previsão. O intuito principal foi mostrar às autoridades tributárias que existem maneiras simples de se melhorar a qualidade da previsão elaborada pela Receita Federal, com investimento razoavelmente baixo, concentrado na qualificação de seus recursos humanos e no desenvolvimento de atividades de pesquisa.

Não se pretende aqui, também, recomendar o sepultamento do método dos indicadores, já que, em algumas situações particulares, ele continua sendo uma alternativa viável, até que se desenvolva métodos específicos para cada processo de previsão requerido. Adicionalmente, ele pode ser empregado para eventuais ajustes nas previsões estatísticas, dada a possibilidade de incorporação de informações que dificilmente são capturadas pelos métodos quantitativos mais sofisticados.

Como sugestão de trabalhos futuros, indicamos os seguinte caminhos:

- verificar se a combinação de métodos de previsão pode gerar resultados mais precisos que a utilização dos métodos individualmente. Assim, poder-se-ia avaliar, por exemplo, a combinação do método de Box-Jenkins com outros métodos e verificar a precisão das previsões geradas. Newbold & Granger (1974) verificaram que a combinação de métodos de previsão pode sempre gerar resultados ou iguais ou mais precisos que a utilização dos métodos de forma individual.

- Poder-se-ia, como já recomendado por Melo (2001), incluir variáveis exógenas no modelo ARIMA e estudar a classe de modelos ARMAX.
- Uma forma de tentar aproveitar a quantidade de informações relevantes que a SRF detém seria através da utilização de modelos bayesianos, introduzidos na década de 70, e que se caracterizam pela possibilidade de incorporação de informações subjetivas no processo de modelagem da série.
- Como as técnicas estatísticas para predição e tomada de decisão não têm bons resultados para muitas aplicações que possuem tamanho de amostras restritos e com alguma não-linearidade no conjunto de dados, uma alternativa que se apresenta é a utilização de Redes Neurais Artificiais (RNA). As RNA, dentre as muitas técnicas que surgiram, ganharam especial atenção pela sua habilidade de aprendizado e sua capacidade de generalização, associação e busca paralela. Estas qualidades as tornam capazes de identificar e assimilar as características mais marcantes das séries, tais como sazonalidade, periodicidade, tendências, entre outras, na maioria das vezes camufladas por ruídos; sem necessitar do laborioso passo da formulação teórica, imprescindível para os procedimentos estatísticos. Além disso, as RNA também têm se destacado pelos seguintes aspectos: sua capacidade em lidar com dados não lineares; sua robustez, pois as RNA são capazes de se auto corrigir mesmo depois de previsões erradas; e sua facilidade de utilização, já que existem várias opções de softwares para simulação de RNA, dedicado a pessoas com pouca experiência, tanto em redes neurais como em séries temporais. É interessante ressaltar, no entanto, que um bom conhecimento sobre RNA e o domínio da área de aplicação, no caso de séries temporais, são fatores que têm muita importância no desenvolvimento de aplicações bem sucedidas, principalmente pelo fato de que os projetos de RNA são fortemente baseados na experimentação e na experiência do projetista, que precisa ter bom senso e sensibilidade para fazer os ajustes adequados nos parâmetros de aprendizado, a fim de se obter melhores resultados.
- Além das alternativas de modelagem supracitadas, pode-se também empreender estudos na tentativa de implementar melhorias na qualidade das previsões aqui elaboradas, através da inclusão de métodos que conferissem maior robustez às estimações obtidas, tais como: implementação de filtros de Kalman para a detecção de outros “outliers” não identificados ou removidos pelo método aqui adotado; a

utilização de modelos no espaço de estados que permita a possibilidade de variação no tempo dos parâmetros estimados; a modelagem dos ciclos estocásticos, observados em algumas séries, através de modelos de “Trading Day Variations”, entre outros.

Finalmente, recomenda-se que a Secretaria da Receita Federal desenvolva uma política consistente de pesquisa e desenvolvimento, incluindo na lista de prioridades, entre tantas outras, a área de previsão da arrecadação tributária. Tal iniciativa, lançaria essa respeitada Instituição no campo do debate técnico-científico, onde contribuiria de forma decisiva para a definição de políticas públicas estratégicas para o desenvolvimento do país, particularmente aquelas relativas à política tributária.

**REFERÊNCIAS BIBLIOGRÁFICAS**

- AGIAKLOGLOU, C. & NEWBOLD, P. The balance between size and power in Dickey-Fuller tests with data dependent rules for the choice of truncation lag. *Economics Letters*, 52, 229-234, 1996.
- ANSLEY, C.F. & NEWBOLD, P. Finite sample properties of estimators for autoregressive moving average models. *Journal of Econometrics*, 13, 159-183, 1980.
- BELL, R. W. & HILLMER S. C. Modeling time series with calendar variation. *Journal of the American Statistical Association*, 78(383), 526-534, september/1983.
- BOWERMAN, B.L. & O'CONNELL, R.T. *Time Series Forecasting, Unified Concepts and Computer Implementation*. 2<sup>nd</sup> Edition. Boston, Duxbury Press, 1987.
- BOX, G. E. P. & JENKINS, G. M. *Time Series Analysis: Forecasting and Control*. 1<sup>st</sup> Edition. New Jersey, Prentice Hall, 1976.
- BOX, G. E. P.; JENKINS, G. M. & REINSEL, G. C. *Time Series Analysis: Forecasting and Control*. 3<sup>rd</sup> Edition. New Jersey, Prentice Hall, 1994.
- BROCKWELL, P.J. & DAVIS, R.A. *Introduction to Time Series and Forecasting*. New York, Springer-Verlag, 1996.
- CHANG, I., TIAO, G. C. & CHEN, C. Estimation of Time Series Parameters in the Presence of Outliers. *Technometrics*, 30, 193-204, 1988.
- CIRINCIONE, C., GURRIERI, G.A., & VAN DE SANDE, B. "Municipal Government Revenue Forecasting: Issues of Method and Data. *Public Budgeting & Finance*, 19(1): 26-46. 1999.
- CLEVELAND, W. S.; DEVLIN, S. J. & TERPENNING, I. J. The SABL Seasonal and Calendar Adjustment Procedures. In: ANDERSON, O. 2., E2. *Time Series Analysis: Theory and Practice*. North-Holland Publishing Company, Amsterdam, 1982. Vol. 1, 539-564.

- COOPER, R.L. The predictive performance of quarterly econometric models of the United States. In: Hickman, B.G. *Econometric Models of Cyclical Behavior*. New York, Columbia University Press, 1972.
- CORDEIRO JR., H. H. Testes de Linearidade em Séries Temporais, 1998. 128p. (Mestrado em Estatística – Universidade Federal de Pernambuco - UFPE).
- CRIBARI-NETO, F. *Previsão de Séries Temporais*. Mini-Curso ministrado na SRF. Brasília, 2000.
- DIEBOLD, F. X. & WATSON, M. W. New Developments in Economic Forecasting. *Journal of Applied Econometrics*, 11, 453-594, 1996.
- ENDERS, W. *Applied Econometric Time Series*. New York, John Wiley & Sons, 1995.
- ENDERS, W. *Rats Handbook for Econometric Time Series*. New York, John Wiley & Sons, 1996.
- ESAF - ESCOLA DE ADMINISTRAÇÃO FAZENDÁRIA. Programação da Arrecadação: Previsão, Acompanhamento e Análise. In: PROGRAMA DE FORMAÇÃO AUDITOR-FISCAL DO TESOIRO NACIONAL, Brasília, 1999. *Política e Administração Tributária - Arrecadação e Cobrança*. Volume 1.
- FOX, A. J. Outliers in time series. *Journal of the Royal Statistical Society, Series B*, 34, 350-63, 1972.
- FRANK, H.A. *Budgetary Forecasting in Local Government: New Tools and Techniques*. Westport, CT: Quorum Books. 1993.
- GOURIEROUX, C. & MONFORT, A. *Time Series and Dynamics Models*. New York, Cambridge University Press, 1997.
- GRANGER, C.W.J. & NEWBOLD, P. *Forecasting Economic Time Series*. New York, Academic Press, 1986.
- GRANGER, C.W.J. & NEWBOLD, P. Spurious regressions in econometrics.. *Journal of Econometrics, Vol.2*. pp. 111-120, 1974.

- GUAJARDO, S. A. & MIRANDA, R. An Elected Official's Guide to Revenue Forecasting. Chicago: Government Finance Officers Association, 2000.
- GUJARATI, 2. N. *Econometria Básica*. 3ª E2. São Pulo, Makron Books, 2000.
- HAMILTON, J. 2. *Time Series Analysis*. New Jersey, Princeton University Press, 1994.
- HARVEY, A. C. *Time Series Models*. 2<sup>nd</sup> Edition. Cambridge, Massachusetts, MIT Press, 1994.
- HILL, R.C., GRIFFITHS, W.E. & JUDGE, G.G. *Econometria*. São Paulo, Editora Saraiva, 1999.
- HILLMER, S. C., BELL, W. R. & TIAO, G. C. Modelling Considerations in the Seasonal Adjustment of economic Time series. In: ZELLNER, A., E2. *Time Applied Time Series Analysis of Economic Data*. US Department of Commerce - Bureau of the Census, Washington, DC, 1983, 74-100.
- KVANLI, A.H, GUYNES, C.S. & PAVUR, R.J. *Introduction to Business Statistics, A Computer Integrated Approach* 4<sup>th</sup> Edition. St.Paul: West Publishing Company, 1996.
- MAKRIDAKIS, S. & WHEELWRIGHT, S.C. *Forecasting Methods for Management*. (5<sup>th</sup> e2.). New York: John Wiley & Sons. 1989.
- MAKRIDAKIS, S. & WHEELWRIGHT, S.C. *The Handbook of Forecasting: A Manager's Guide*. (2<sup>nd</sup> e2.). New York: John Wiley & Sons. 1987.
- MAKRIDAKIS, S., ANDERSEN, A., CARBONE, R., FILDES, R., HIBON, M., LEWANDOWSKI, R., NEWTON, J., PARZEN, E., & WINKLER, R. *The Forecasting Accuracy of Major Time Series Methods*. New York: John Wiley & Sons. 1984.
- MARTIN, R. 2.; SAMAROV, A. & VANDAELE, W. Robust Methods for ARIMA Models. In: ANDERSON, O. 2., E2. *Time Series Analysis: Theory and Practice*. North-Holland Publishing Company, Amsterdam, 1982. Vol. 1, 755-756.
- MELO, B. S. V. de. Modelo de Previsão para a Arrecadação Tributária. Brasília, 2001. 119p. (Mestrado – Universidade de Brasília - UnB)

- MILLS, T.C. *Time Series Techniques for Economists*. Cambridge, Cambridge University Press, 1990.
- MONTGOMERY, J. C.; JOHNSON, L. A. & GARDINER, J. S. *Forecasting & Time Series Analysis*. 2<sup>nd</sup> Edition. New York, McGraw-Hill, Inc., 1990.
- NELSON, C. R. *Applied Time Series Analysis For Managerial Forecasting*. San Francisco, Holden-Day, Inc., 1973.
- NEWBOLD, P. & GRANGER, C.W.J. Experience with forecasting univariate time series and the combination of forecasts. *Journal of the Royal Statistics Society A*, 137, 131-146, 1974.
- PANDIT, S. M. & WU, S-M. *Time Series and System Analysis With Applications*. New York, John Wiley & Sons, 1983.
- PANKRATZ, A. *Forecasting With Dynamic Regression Models*. New York, John Wiley & Sons, 1991.
- PANKRATZ, A. *Forecasting With Univariate Box-Jenkins Models*. New York, John Wiley & Sons, 1983.
- PICCOLO, L. A Comparison of some Alternative Decomposition Methods for ARMA Models. In: ANDERSON, O. L., Ed. *Time Series Analysis: Theory and Practice*. North-Holland Publishing Company, Amsterdam, 1982. Vol. 1, 565-582.
- PINDYCK, R.S. & RUBINFELD, D.L. *Econometry Models & Economic Forecasts*, 2<sup>nd</sup> Edition. New York, McGraw-Hill Book Company, 1981.
- POSKITT, D.S. & TREMAYNE, A.R. Determining a portfolio of linear time series models. *Biometrika*, 74, 125-137, 1987.
- PUKKILA, T. M. On the Identification of ARMA(p, q) Models. In: ANDERSON, O. L., Ed. *Time Series Analysis: Theory and Practice*. North-Holland Publishing Company, Amsterdam, 1982. Vol. 1, 81-103.
- TSAY, R. S. Time Series Model Specification in the Presence of Outliers. *Journal of the American Statistical Association*, 81, 132-41, 1986.

- VENABLES, W.N. & RIPLEY, B.2. (1999). *Modern Applied Statistics with S-PLUS*, 3<sup>rd</sup> Edition. New York, Springer-Verlag, 1999.
- WEI, W. W. S. *Time Series Analysis*.. New York, Addison-Wesley Publishing Compny, Inc., 1990.
- WONG, J.2. "Local Government Revenue Forecasting: Using Regression and Econometric Revenue Forecasting in a Medium-Sized City." *Public Budgeting & Financial Management*, 7(3): 315-335. 1995.

## APÊNDICE 1

### Programa de Análise e Modelagem dos Dados no RATS

Será listado aqui um resumo do programa, com os principais procedimentos adotados, já que o programa completo se apresenta com 62 páginas, inviabilizando sua apresentação completa.

```

*                               PROGRAMA DISSERTAÇÃO.PRG
*
*                               ENTRADA DOS DADOS
*
CAL 1989 1 12
ALL 8 2000:12
OPEN DATA C:\WINRATS\DISSERTAÇÃO\DADOSFINAIS2000.TXT
DATA(FORMAT=PRN,ORG=OBS) / II IPI IR OI RIMP COFPIS CPMF OCONTR RCONT OREC RECADM
*
*                               TRANSFORMAÇÃO DOS DADOS
*
SET LII = LOG(II)
SET LIPI = LOG(IPI)
SET LIR = LOG(IR)
SET LOI = LOG(OI)
SET LRIMP = LOG(RIMP)
SET LCOFPIS = LOG(COFPIS)
SET LCPMF = LOG(CPMF)
SET LOCONTR = LOG(OCONTR)
SET LRCONT = LOG(RCONT)
SET LOREC = LOG(OREC)
SET LRECADM = LOG(RECADM)
*
*                               GRÁFICOS DAS SÉRIES ORIGINAIS E DAS TRANSFORMADAS
*
*                               GRÁFICOS INDIVIDUAIS 1 (1989 A 2000)
*
GRAPH(HEADER='IMPOSTO DE IMPORTAÇÃO - 1989 A 2000',KEY=UPLEFT) 1
# II
*
GRAPH(HEADER='IMPOSTO SOBRE PRODUTOS INDUSTRIALIZADOS - 1989 A 2000',KEY=UPLEFT)
1
# IPI
*
GRAPH(HEADER='IMPOSTO DE RENDA TOTAL - 1989 A 2000',KEY=UPLEFT) 1
# IR
*
GRAPH(HEADER='OUTROS IMPOSTOS - 1989 A 2000',KEY=UPLEFT) 1
# OI
*
GRAPH(HEADER='RECEITA TOTAL DE IMPOSTOS - 1989 A 2000',KEY=UPLEFT) 1
# RIMP
*
GRAPH(HEADER='COFINS, PIS E PASEP - 1989 A 2000',KEY=UPLEFT) 1

```

```
# COFPIS
*
GRAPH(HEADER='CPMF - 1989 A 2000',KEY=UPLEFT) 1
# CPMF
*

GRAPH(HEADER='OUTRAS CONTRIBUIÇÕES - 1989 A 2000',KEY=UPLEFT) 1
# OCONTR
*
GRAPH(HEADER='RECEITA TOTAL DE CONTRIBUIÇÕES - 1989 A 2000',KEY=UPLEFT) 1
# RCONT
*
GRAPH(HEADER='RECEITA TOTAL ADMINISTRADA PELA SRF - 1989 A 2000',KEY=UPLEFT) 1
# RECADM
*
*           GRÁFICOS INDIVIDUAIS LOGARITMIZADOS (1989 A 2000)
*
GRAPH(HEADER='LOGARITMO DA SÉRIE IMPOSTO DE IMPORTAÇÃO - 1989 A
2000',KEY=UPLEFT) 1
# LII
*
GRAPH(HEADER='LOGARITMO DA SÉRIE IMPOSTO SOBRE PRODUTOS INDUSTRIALIZADOS -
1989 A 2000',KEY=UPLEFT) 1
# LIPI
*
GRAPH(HEADER='LOGARITMO DA SÉRIE IMPOSTO DE RENDA TOTAL - 1989 A
2000',KEY=UPLEFT) 1
# LIR
*
GRAPH(HEADER='LOGARITMO DA SÉRIE OUTROS IMPOSTOS - 1989 A 2000',KEY=UPLEFT) 1
# LOI
*
GRAPH(HEADER='LOGARITMO DA SÉRIE RECEITA TOTAL DE IMPOSTOS - 1989 A
2000',KEY=UPLEFT) 1
# LRIMP
*
GRAPH(HEADER='LOGARITMO DA SÉRIE COFINS, PIS E PASEP - 1989 A 2000',KEY=UPLEFT) 1
# LCOFPIS
*
GRAPH(HEADER='LOGARITMO DA SÉRIE CPMF - 1989 A 2000',KEY=UPLEFT) 1
# LCPMF
*
GRAPH(HEADER='LOGARITMO DA SÉRIE OUTRAS CONTRIBUIÇÕES - 1989 A 2000',KEY=UPLEFT)
1
# LOCONTR
*
GRAPH(HEADER='LOGARITMO DA SÉRIE RECEITA TOTAL DE CONTRIBUIÇÕES - 1989 A
2000',KEY=UPLEFT) 1
# LRCONT
*
GRAPH(HEADER='LOGARITMO DA SÉRIE RECEITA TOTAL ADMINISTRADA PELA SRF - 1989 A
2000',KEY=UPLEFT) 1
# LRECADM
*
*
*
*
*****
**
```

```

*
*           ENTRADA DOS DADOS APÓS CORREÇÃO DOS OUTLIERS
*
CAL 1989 1 12
ALL 8 2000:12
OPEN DATA C:\WINRATS\DISSERTAÇÃO\DADOSFINAIS2000-OUTLIERS.TXT
DATA(FORMAT=PRN,ORG=OBS) / II IPI IR OI RIMP COFPIS CPMF OCONTR RCONT OREC RECADM
*
*           TRANSFORMAÇÃO DOS DADOS
*
SET LII = LOG(II)
SET LIPI = LOG(IPI)
SET LIR = LOG(IR)
SET LOI = LOG(OI)
SET LRIMP = LOG(RIMP)
SET LCOFPIS = LOG(COFPIS)
SET LCPMF = LOG(CPMF)
SET LOCONTR = LOG(OCONTR)
SET LRCONT = LOG(RCONT)
SET LOREC = LOG(OREC)
SET LRECADM = LOG(RECADM)
*
*
*           GRÁFICOS INDIVIDUAIS (PERÍODOS QUE SERÃO ANALISADOS)
*
GRAPH(HEADER='IMPOSTO DE IMPORTAÇÃO - 1994 A 2000',KEY=UPLEFT) 1
# LII 1994:7 2000:12
*
GRAPH(HEADER='IMPOSTO SOBRE PRODUTOS INDUSTRIALIZADOS - 1994 A 2000',KEY=UPLEFT) 1
# LIPI 1994:7 2000:12
*
GRAPH(HEADER='IMPOSTO DE RENDA TOTAL - 1994 A 2000',KEY=UPLEFT) 1
# LIR 1994:7 2000:12
*
GRAPH(HEADER='OUTROS IMPOSTOS - 1994 A 2000',KEY=UPLEFT) 1
# OI 1994:7 2000:12
*
GRAPH(HEADER='RECEITA TOTAL DE IMPOSTOS - 1994 A 2000',KEY=UPLEFT) 1
# LRIMP 1994:7 2000:12
*
GRAPH(HEADER='COFINS, PIS E PASEP - 1994 A 2000',KEY=UPLEFT) 1
# LCOFPIS 1994:7 2000:12
*
GRAPH(HEADER='CPMF - 1997 A 2000',KEY=UPLEFT) 1
# LCPMF 1997:2 2000:12
*
GRAPH(HEADER='OUTRAS CONTRIBUIÇÕES - 1994 A 2000',KEY=UPLEFT) 1
# LOCONTR 1994:7 2000:12
*
GRAPH(HEADER='RECEITA TOTAL DE CONTRIBUIÇÕES - 1994 A 2000',KEY=UPLEFT) 1
# LRCONT 1994:7 2000:12
*
GRAPH(HEADER='RECEITA TOTAL ADMINISTRADA PELA SRF - 1994 A 2000',KEY=UPLEFT) 1
# LRECADM 1994:7 2000:12
*
*
*
*

```

```

*      OS PROCEDIMENTOS ABAIXO SERÃO EFETUADOS PARA CADA UMA DAS SÉRIES
*
*      (basta substituir LRECADM pelas demais séries)
*
*      ESTÁGIO DE IDENTIFICAÇÃO
*
SOURCE(NOECHO) C:\WINRATS\EXERCÍCIOS\SARIMA.SRC
@SARIMA LRECADM 1994:7 2000:12
*
*      CÁLCULO DOS VALORES DOS CRITÉRIOS AIC E BIC PARA OS MODELOS ESTIMADOS
*
*      (após a realização da rotina acima para os diversos modelos, faremos a estimação
*      de cada um deles e calcularemos AIC e BIC para podermos compará-los)
*
*      BIC E AIC PARA LII
*
*      (esse procedimento será repetido para todas as séries, bastando substituir LII pelas respectivas séries)
*
*      MODELO 1
*
BOXJENK(DIFFS=1,SDIFFS=1) LII 1996:9 2000:12 RESID1
*
* CÁLCULO DO AIC E SBC *
*
COMPUTE SBC1 = %NOBS*LOG(%RSS) + %NREG*LOG(%NOBS)
COMPUTE AIC1 = %NOBS*LOG(%RSS) + 2*%NREG
DISPLAY 'AIC1' AIC1 'SBC1' SBC1
*
*      MODELO 2
*
BOXJENK(DIFFS=1,SDIFFS=1,SMA=1) LII 1996:9 2000:12 RESID1
*
* CÁLCULO DO AIC E SBC *
*
COMPUTE SBC1 = %NOBS*LOG(%RSS) + %NREG*LOG(%NOBS)
COMPUTE AIC1 = %NOBS*LOG(%RSS) + 2*%NREG
DISPLAY 'AIC1' AIC1 'SBC1' SBC1
*
*      MODELO 3
*
BOXJENK(DIFFS=1,SAR=1,SDIFFS=1) LII 1996:9 2000:12 RESID1
*
* CÁLCULO DO AIC E SBC *
*
COMPUTE SBC1 = %NOBS*LOG(%RSS) + %NREG*LOG(%NOBS)
COMPUTE AIC1 = %NOBS*LOG(%RSS) + 2*%NREG
DISPLAY 'AIC1' AIC1 'SBC1' SBC1
*
*      MODELO 4
*
BOXJENK(DIFFS=1,SAR=1,SDIFFS=1,SMA=1) LII 1996:9 2000:12 RESID1
*
DISPLAY %NOBS
* CÁLCULO DO AIC E SBC *
*
COMPUTE SBC1 = %NOBS*LOG(%RSS) + %NREG*LOG(%NOBS)
COMPUTE AIC1 = %NOBS*LOG(%RSS) + 2*%NREG
DISPLAY 'AIC1' AIC1 'SBC1' SBC1
*

```

```

*
*           MODELO 5
*
BOXJENK(DIFFS=1,MA=1,SDIFFS=1) LII 1996:9 2000:12 RESID1
*
* CÁLCULO DO AIC E SBC *
*
COMPUTE SBC1 = %NOBS*LOG(%RSS) + %NREG*LOG(%NOBS)
COMPUTE AIC1 = %NOBS*LOG(%RSS) + 2*%NREG
DISPLAY 'AIC1' AIC1 'SBC1' SBC1
*
*           MODELO 6
*
BOXJENK(DIFFS=1,MA=1,SDIFFS=1,SMA=1) LII 1996:9 2000:12 RESID1
*
* CÁLCULO DO AIC E SBC *
*
COMPUTE SBC1 = %NOBS*LOG(%RSS) + %NREG*LOG(%NOBS)
COMPUTE AIC1 = %NOBS*LOG(%RSS) + 2*%NREG
DISPLAY 'AIC1' AIC1 'SBC1' SBC1
*
*           MODELO 7
*
BOXJENK(DIFFS=1,MA=1,SAR=1,SDIFFS=1) LII 1996:9 2000:12 RESID1
*
* CÁLCULO DO AIC E SBC *
*
COMPUTE SBC1 = %NOBS*LOG(%RSS) + %NREG*LOG(%NOBS)
COMPUTE AIC1 = %NOBS*LOG(%RSS) + 2*%NREG
DISPLAY 'AIC1' AIC1 'SBC1' SBC1
*
*           MODELO 8
*
BOXJENK(DIFFS=1,MA=1,SAR=1,SDIFFS=1,SMA=1) LII 1996:9 2000:12 RESID1
*
* CÁLCULO DO AIC E SBC *
*
COMPUTE SBC1 = %NOBS*LOG(%RSS) + %NREG*LOG(%NOBS)
COMPUTE AIC1 = %NOBS*LOG(%RSS) + 2*%NREG
DISPLAY 'AIC1' AIC1 'SBC1' SBC1
*
*           MODELO 9
*
BOXJENK(AR=1,DIFFS=1,SDIFFS=1) LII 1996:9 2000:12 RESID1
*
* CÁLCULO DO AIC E SBC *
*
COMPUTE SBC1 = %NOBS*LOG(%RSS) + %NREG*LOG(%NOBS)
COMPUTE AIC1 = %NOBS*LOG(%RSS) + 2*%NREG
DISPLAY 'AIC1' AIC1 'SBC1' SBC1
*
*           MODELO 10
*
BOXJENK(AR=1,DIFFS=1,SDIFFS=1,SMA=1) LII 1996:9 2000:12 RESID1
*
* CÁLCULO DO AIC E SBC *
*
COMPUTE SBC1 = %NOBS*LOG(%RSS) + %NREG*LOG(%NOBS)
COMPUTE AIC1 = %NOBS*LOG(%RSS) + 2*%NREG
DISPLAY 'AIC1' AIC1 'SBC1' SBC1
*

```

```

*
*           MODELO 11
*
BOXJENK(AR=1,DIFFS=1,SAR=1,SDIFFS=1) LII 1996:9 2000:12 RESID1
*
* CÁLCULO DO AIC E SBC *
*
COMPUTE SBC1 = %NOBS*LOG(%RSS) + %NREG*LOG(%NOBS)
COMPUTE AIC1 = %NOBS*LOG(%RSS) + 2*%NREG
DISPLAY 'AIC1' AIC1 'SBC1' SBC1
*
*           MODELO 12
*
BOXJENK(AR=1,DIFFS=1,SAR=1,SDIFFS=1,SMA=1) LII 1996:9 2000:12 RESID1
*
* CÁLCULO DO AIC E SBC *
*
COMPUTE SBC1 = %NOBS*LOG(%RSS) + %NREG*LOG(%NOBS)
COMPUTE AIC1 = %NOBS*LOG(%RSS) + 2*%NREG
DISPLAY 'AIC1' AIC1 'SBC1' SBC1
*
*           MODELO 13
*
BOXJENK(AR=1,DIFFS=1,MA=1,SDIFFS=1) LII 1996:9 2000:12 RESID1
*
* CÁLCULO DO AIC E SBC *
*
COMPUTE SBC1 = %NOBS*LOG(%RSS) + %NREG*LOG(%NOBS)
COMPUTE AIC1 = %NOBS*LOG(%RSS) + 2*%NREG
DISPLAY 'AIC1' AIC1 'SBC1' SBC1
*
*           MODELO 14
*
BOXJENK(AR=1,DIFFS=1,MA=1,SDIFFS=1,SMA=1) LII 1996:9 2000:12 RESID1
*
* CÁLCULO DO AIC E SBC *
*
COMPUTE SBC1 = %NOBS*LOG(%RSS) + %NREG*LOG(%NOBS)
COMPUTE AIC1 = %NOBS*LOG(%RSS) + 2*%NREG
DISPLAY 'AIC1' AIC1 'SBC1' SBC1
*
*           MODELO 15
*
BOXJENK(AR=1,DIFFS=1,MA=1,SAR=1,SDIFFS=1) LII 1996:9 2000:12 RESID1
*
* CÁLCULO DO AIC E SBC *
*
COMPUTE SBC1 = %NOBS*LOG(%RSS) + %NREG*LOG(%NOBS)
COMPUTE AIC1 = %NOBS*LOG(%RSS) + 2*%NREG
DISPLAY 'AIC1' AIC1 'SBC1' SBC1
*
*           MODELO 16
*
BOXJENK(AR=1,DIFFS=1,MA=1,SAR=1,SDIFFS=1,SMA=1) LII 1996:9 2000:12 RESID1
* CÁLCULO DO AIC E SBC *
*
COMPUTE SBC1 = %NOBS*LOG(%RSS) + %NREG*LOG(%NOBS)
COMPUTE AIC1 = %NOBS*LOG(%RSS) + 2*%NREG
DISPLAY 'AIC1' AIC1 'SBC1' SBC1
*

```

**GRÁFICO DOS RESÍDUOS PADRONIZADOS II**

(esse procedimento também será repetido para cada uma das séries, substituindo-se os modelos abaixo pelos selecionados para cada série, recalculando-se os desvios padrões dos resíduos, bem como o desvio padrão da estimativa.)

**MODELO 1**

BOXJENK(DIFFS=1,MA=1,SDIFFS=1,SMA=1,DEFINE=EQII1,ITERATIONS=50) LII 1994:7 2000:12  
RESID1

SET RESID11 = (RESID1)/0.136  
GRAPH(STYLE=BAR) 1  
#RESID11

**MODELO 2**

BOXJENK(AR=1,DIFFS=1,SDIFFS=1,SMA=1,DEFINE=EQII2,ITERATIONS=50) LII 1994:7 2000:12  
RESID2

SET RESID21 = (RESID2)/0.134  
GRAPH(STYLE=BAR) 1  
#RESID21

**PREVISÃO - II**

(esse procedimento também será repetido para cada uma das séries, substituindo-se os modelos abaixo pelos selecionados para cada série, recalculando-se as previsões e os desvios.)

**MODELO 1**

SOURCE(NOECHO) C:\WINRATS\BJFORE.SRC  
@BJFORE(NOCONSTANT,DIFFS=1,MAS=1,SDIFFS=1,SMA=1) LII 2001:1 2001:12 IIFORE

GRAPH(HEADER='PREVISÃO PARA A SÉRIE LII') 2  
# LII 1994:7 2000:12  
# IIFORE 2001:1 2001:12

PRINT / IIFORE

ERRORS 1 12  
# EQII1 STDERRORS

**MODELO 2**

SOURCE(NOECHO) C:\WINRATS\BJFORE.SRC  
@BJFORE(NOCONSTANT,ARS=1,DIFFS=1,SDIFFS=1,SMA=1) LII 2001:1 2001:12 IIFORE

GRAPH(HEADER='PREVISÃO PARA A SÉRIE LII') 2  
# LII 1994:7 2000:12  
# IIFORE 2001:1 2001:12

PRINT / IIFORE

ERRORS 1 12  
# EQII2 STDERRORS

\*\*\*\*\*

## APÊNDICE 2

### 2.1 - Série do Imposto de Importação

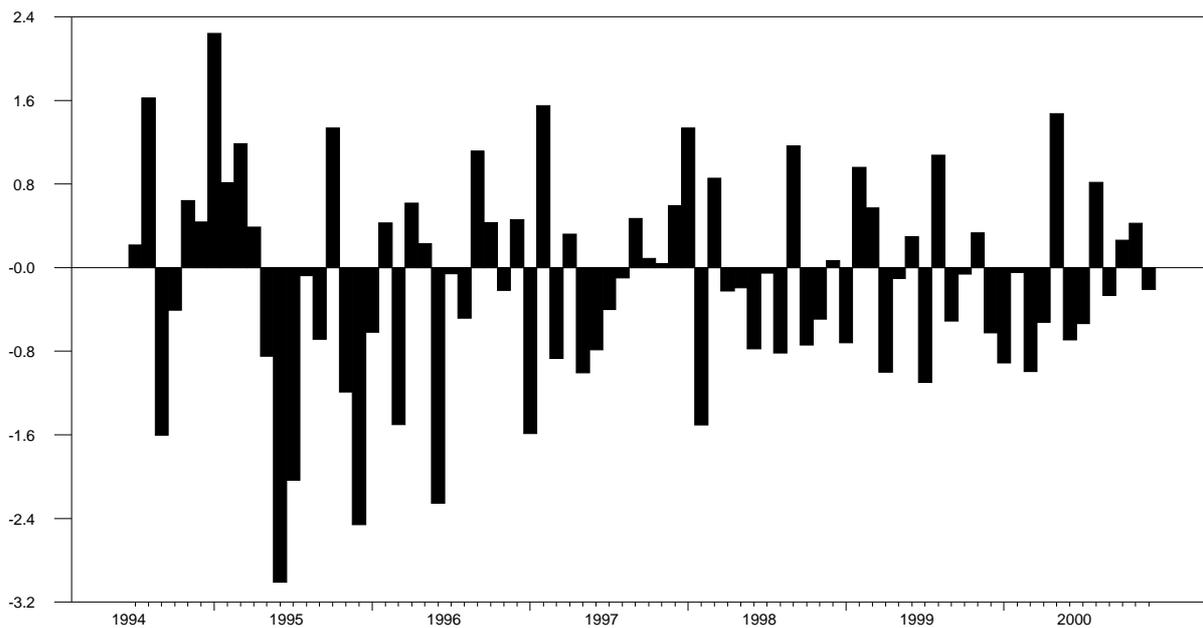


Figura 2.1.1 – Gráfico dos Resíduos Padronizados ( $a_t/\sigma_a$ ) para o Modelo ARIMA (1,1,0)(0,1,1)<sub>12</sub>

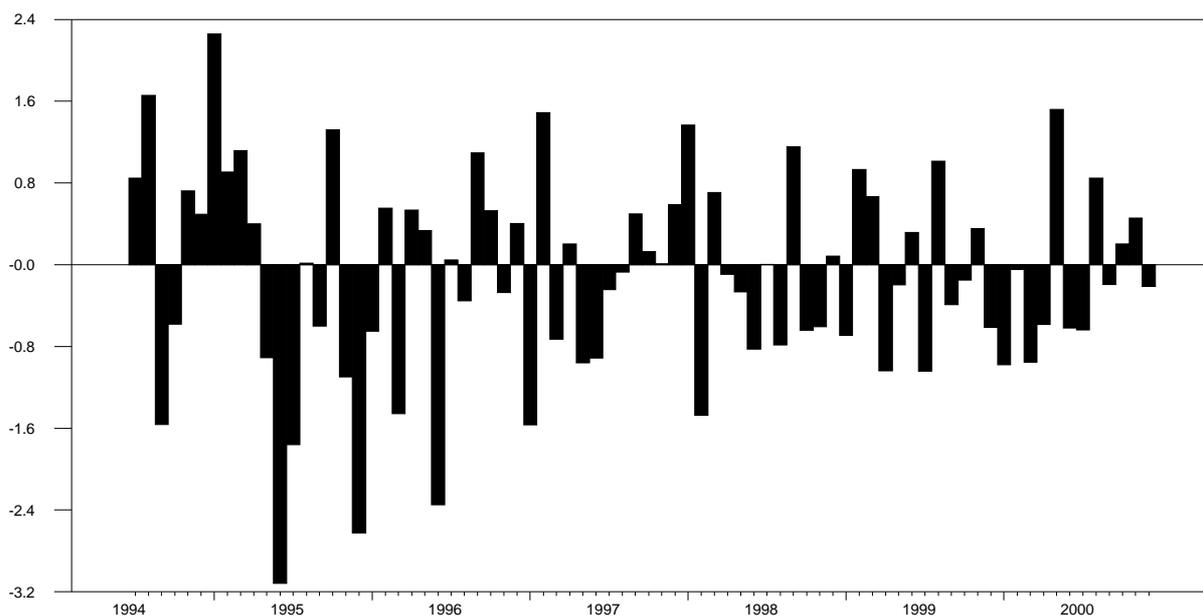


Figura 2.1.2 – Gráfico dos Resíduos Padronizados ( $a_t/\sigma_a$ ) para o Modelo ARIMA (1,1,0)(1,1,1)<sub>12</sub>

### 2.2 - Série do Imposto sobre Produtos Industrializados

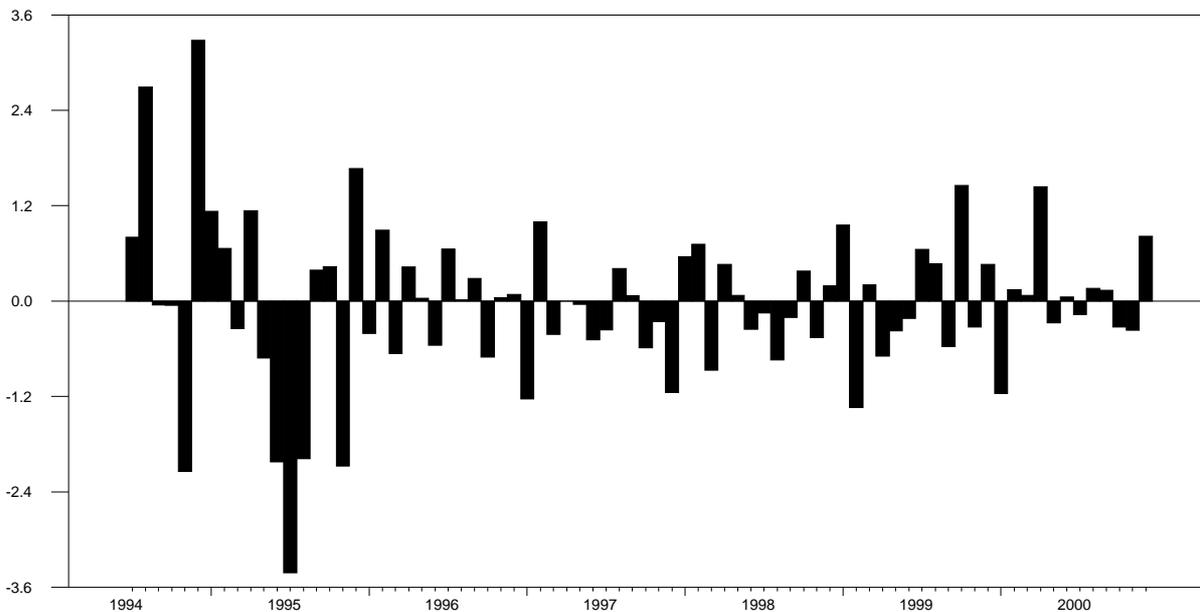


Figura 2.2.1 – Gráfico dos Resíduos Padronizados ( $a_t/\sigma_a$ ) para o Modelo ARIMA (0,1,1)(0,1,1)<sub>12</sub>

### 2.3 - Série do Imposto sobre a Renda

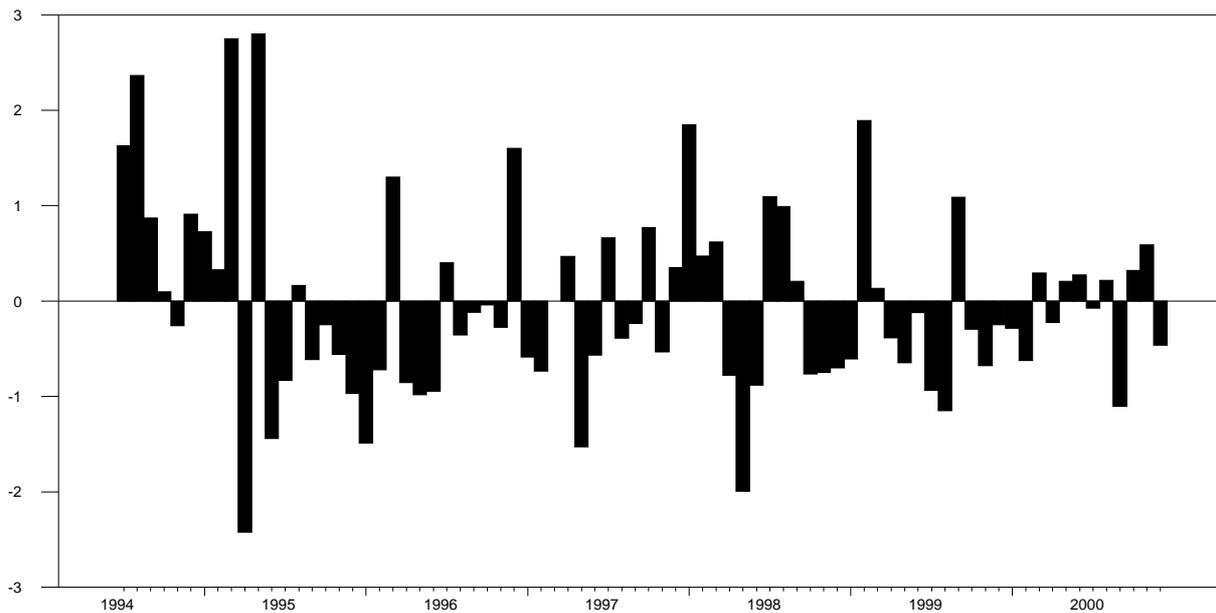


Figura 2.3.1 – Gráfico dos Resíduos Padronizados ( $a_t/\sigma_a$ ) para o Modelo ARIMA (0,1,1)(0,1,1)<sub>12</sub>

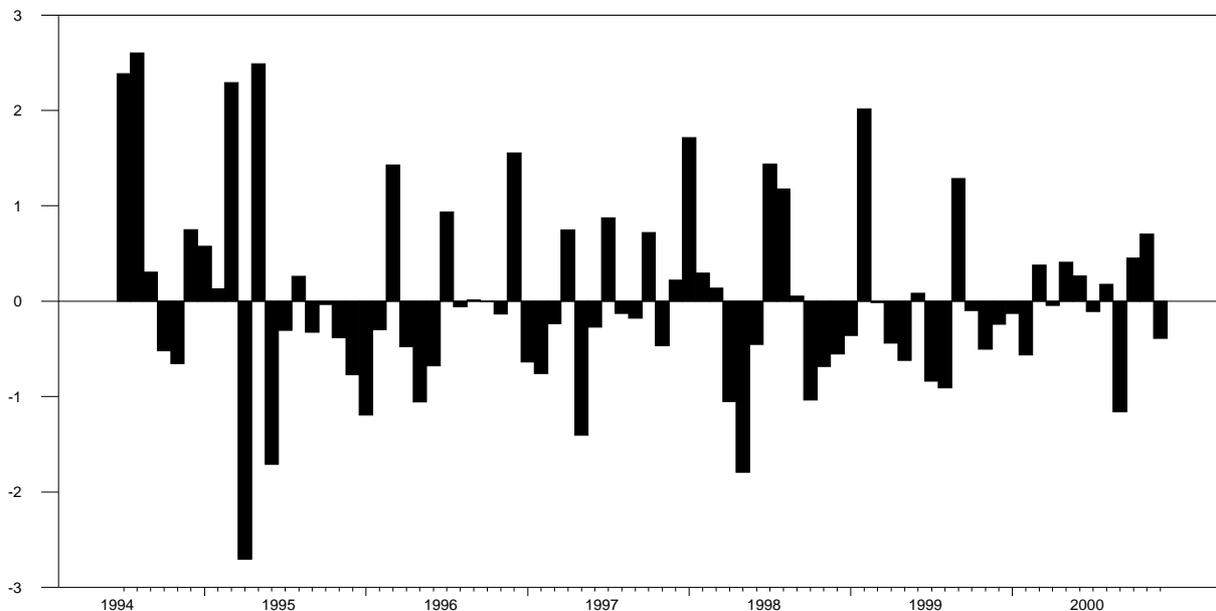


Figura 2.3.2 – Gráfico dos Resíduos Padronizados ( $a_t/\sigma_a$ ) para o Modelo ARIMA (2,1,0)(0,1,1)<sub>12</sub>

**2.4 - Série Outros Impostos**

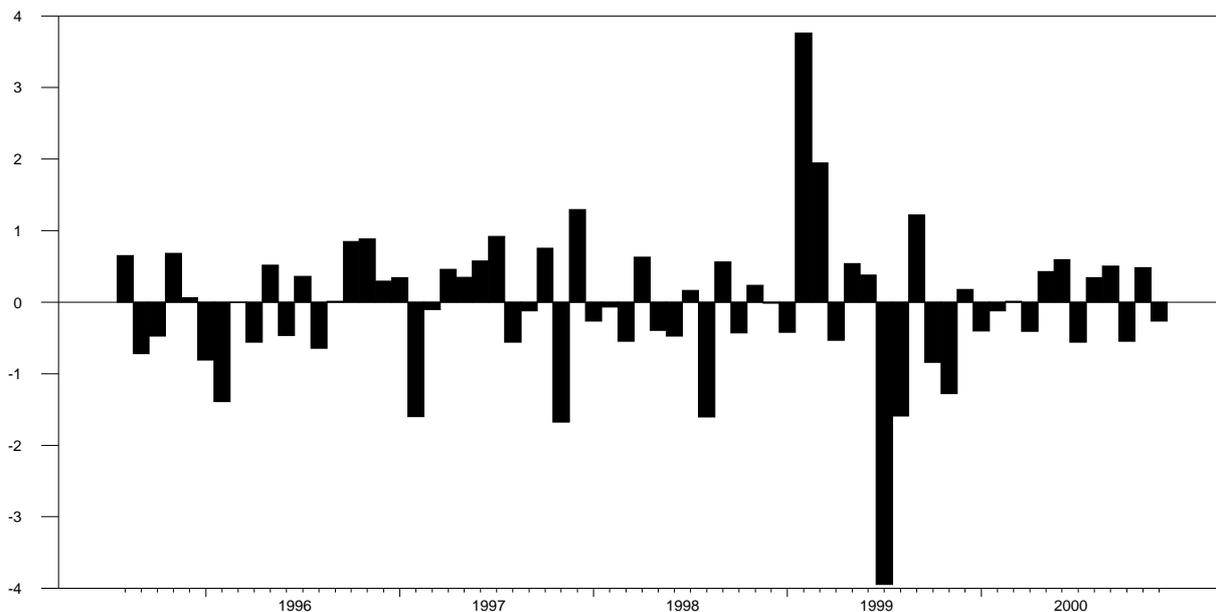


Figura 2.4.1 – Gráfico dos Resíduos Padronizados ( $a_t/\sigma_a$ ) para o Modelo ARIMA(//1,5//,1,0)

### 2.5 - Série Receita Total de Impostos

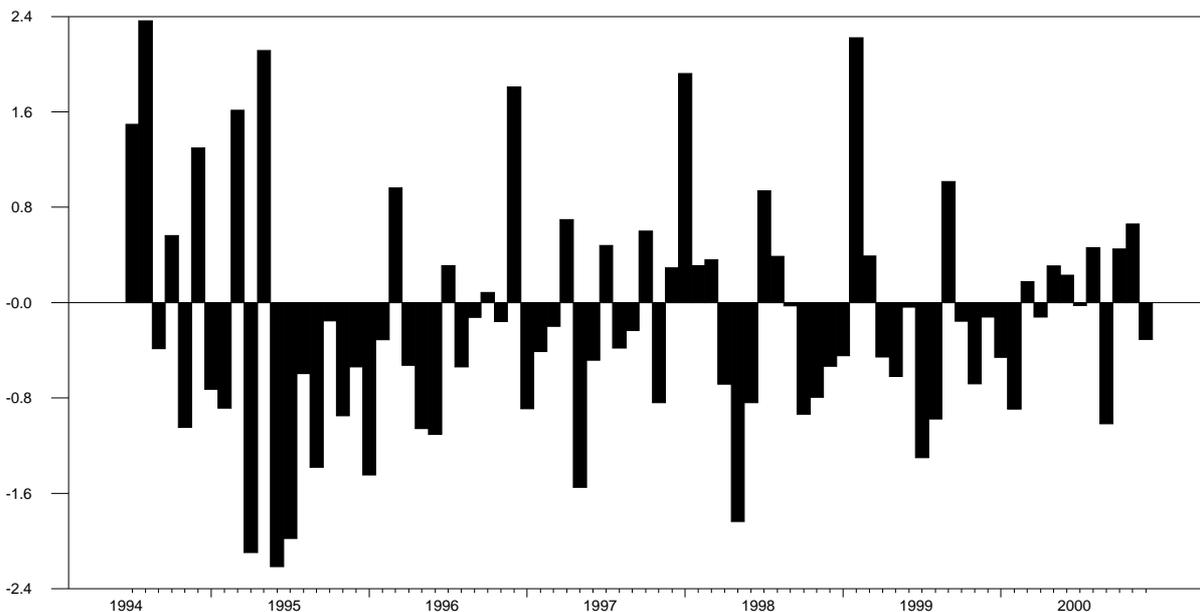


Figura 2.5.1 – Gráfico dos Resíduos Padronizados ( $a_t/\sigma_a$ ) para o Modelo ARIMA (0,1,1)(0,1,1)<sub>12</sub>

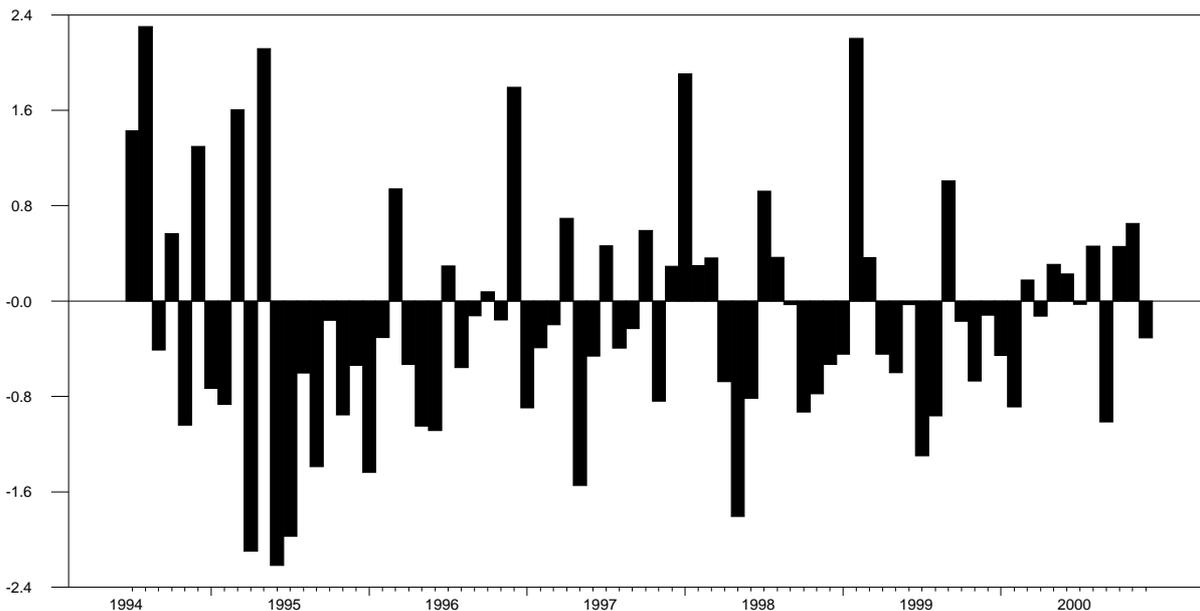


Figura 2.5.2 – Gráfico dos Resíduos Padronizados ( $a_t/\sigma_a$ ) para o Modelo ARIMA (1,1,1)(0,1,1)<sub>12</sub>

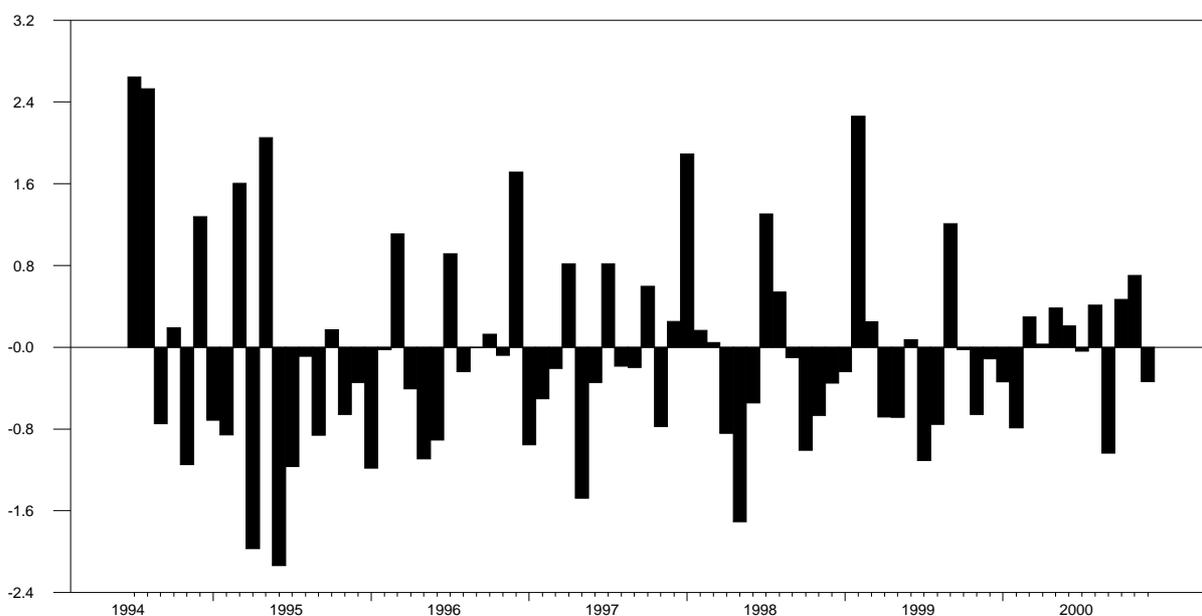


Figura 2.5.3 – Gráfico dos Resíduos Padronizados ( $a_t/\sigma_a$ ) para o Modelo ARIMA (2,1,0)(0,1,1)<sub>12</sub>

**2.6 - Série da Contribuição Provisória sobre Movimentação Financeira (CPMF)**

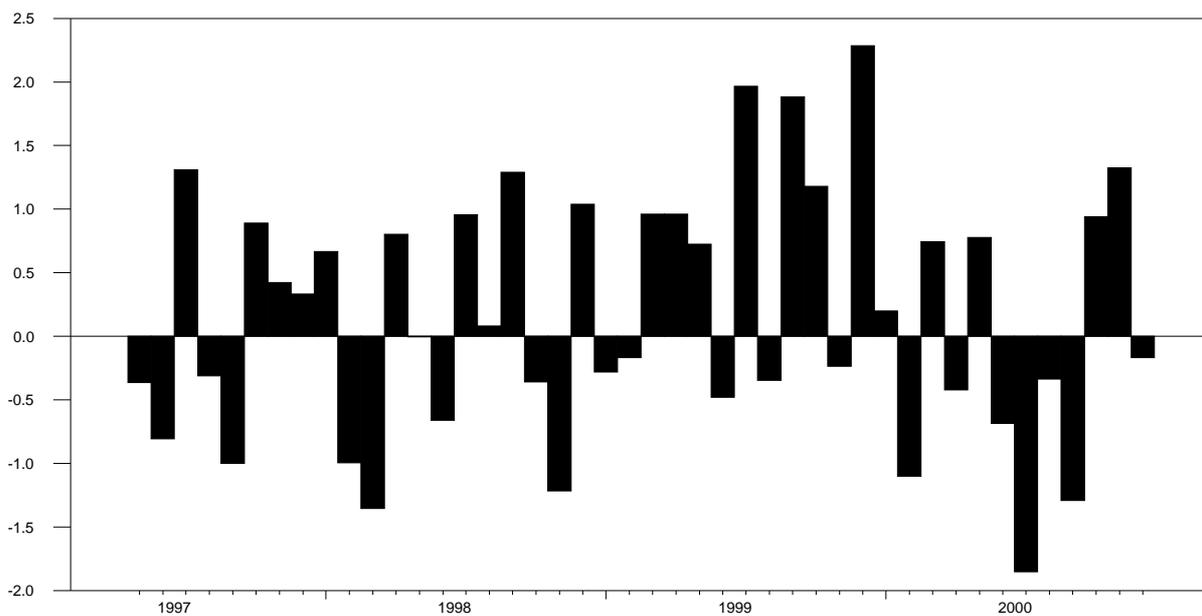


Figura 2.6.1 – Gráfico dos Resíduos Padronizados ( $a_t/\sigma_a$ ) para o Modelo ARIMA (1,1,1)

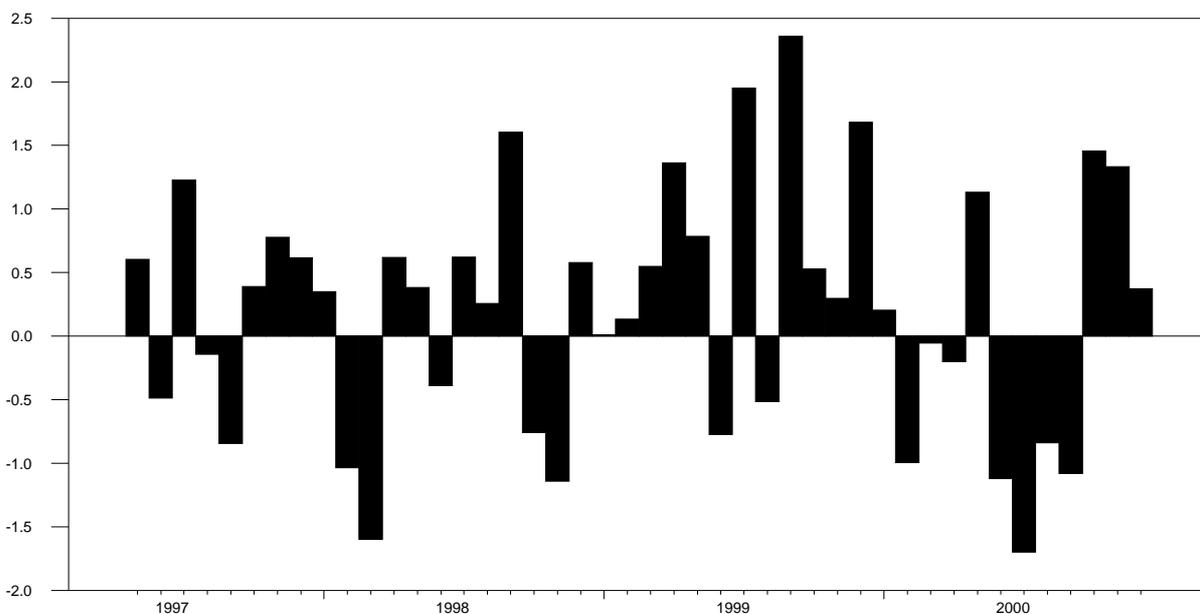


Figura 2.6.2 – Gráfico dos Resíduos Padronizados ( $a_t/\sigma_a$ ) para o Modelo ARIMA (2,1,0)

**2.7 - Série Outras Contribuições (OCONTR)**

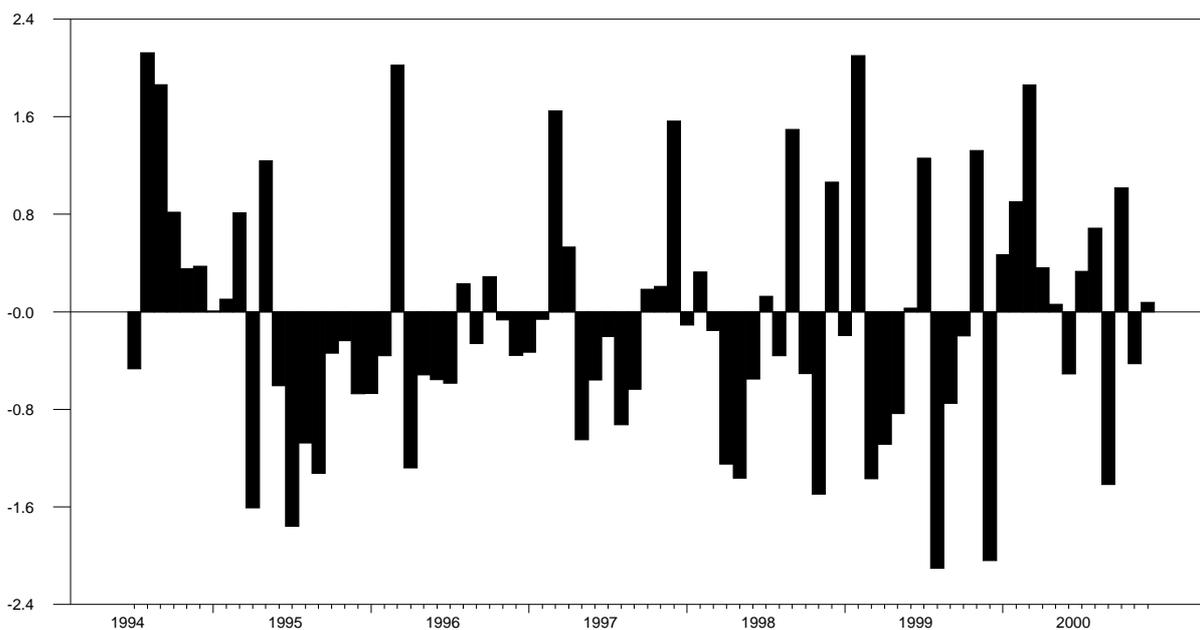


Figura 2.7.1 – Gráfico dos Resíduos Padronizados ( $a_t/S_a$ ) para o Modelo ARIMA(0,1,1)(0,1,1)<sub>12</sub>

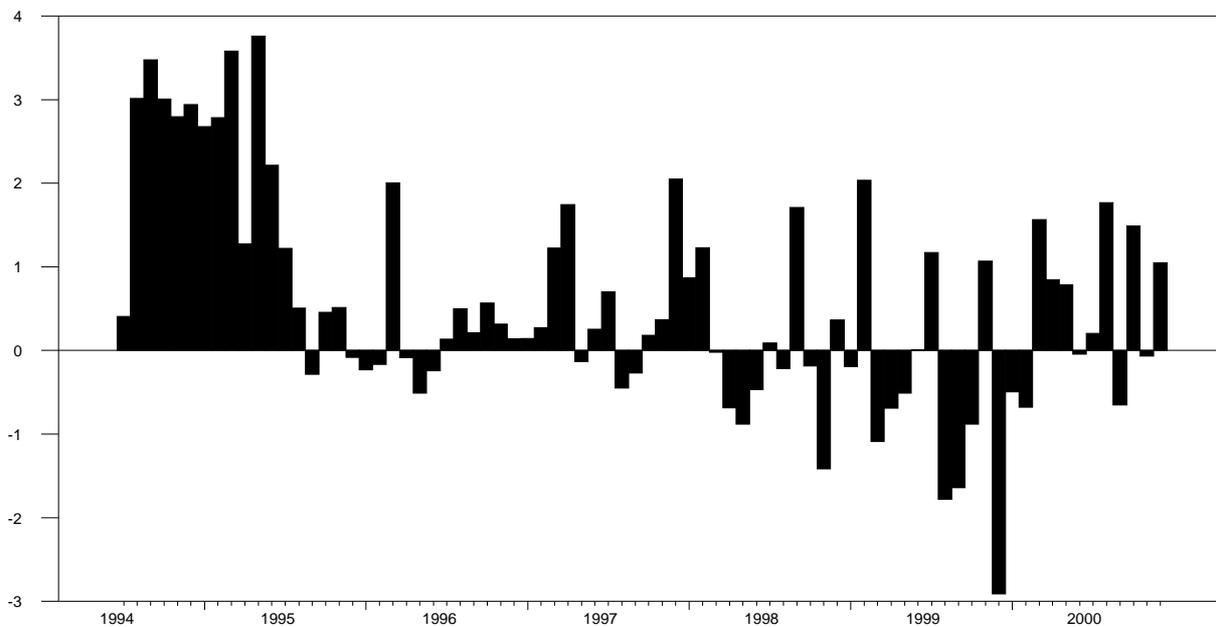


Figura 2.7.2 – Gráfico dos Resíduos Padronizados ( $a_t/s_a$ ) para o Modelo ARIMA (0,1,1)<sub>12</sub>

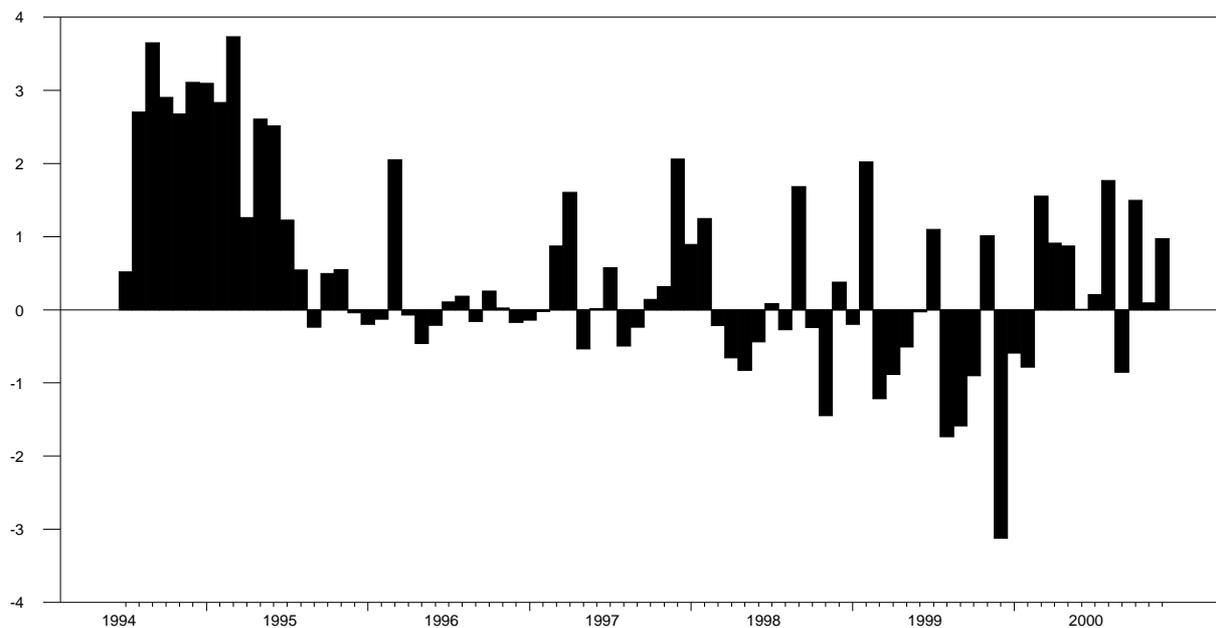


Figura 2.7.3 – Gráfico dos Resíduos Padronizados ( $a_t/s_a$ ) para o Modelo ARIMA (1,1,0)<sub>12</sub>

**2.8 - Série Receita Total de Contribuições (RCONT)**

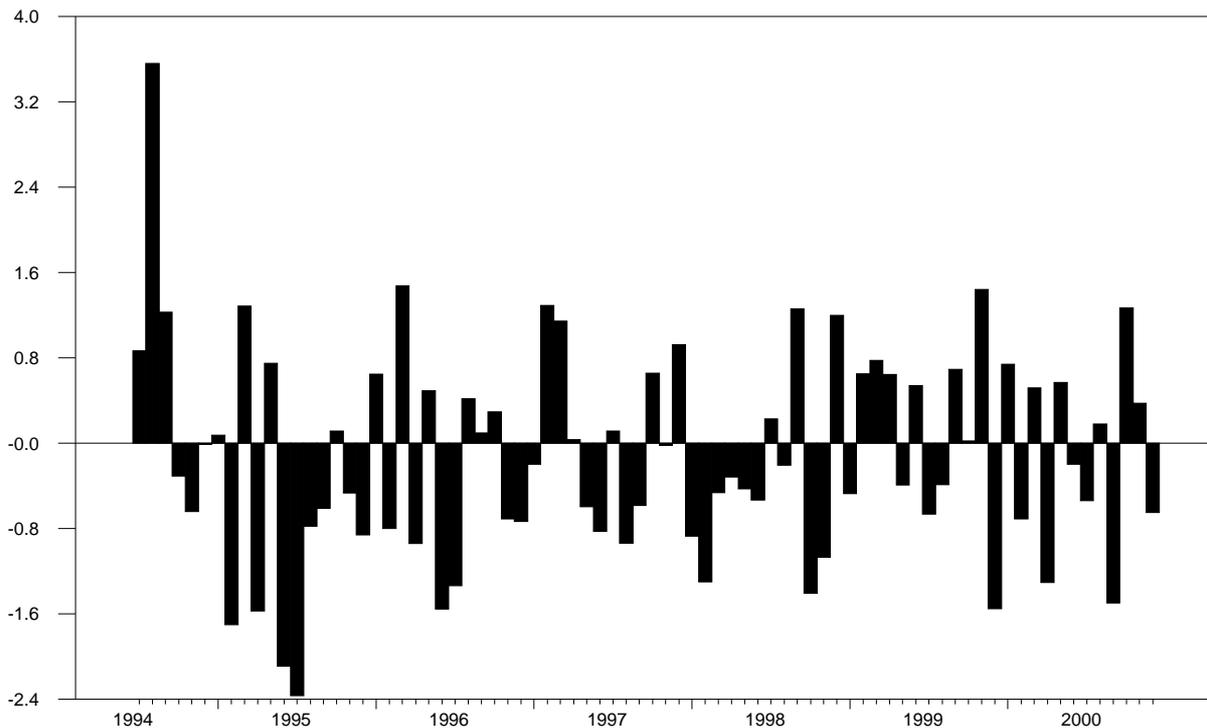


Figura 2.8.1 – Gráfico dos Resíduos Padronizados ( $a_t/s_a$ ) para o Modelo ARIMA(0,1,1)(0,1,1)<sub>12</sub>

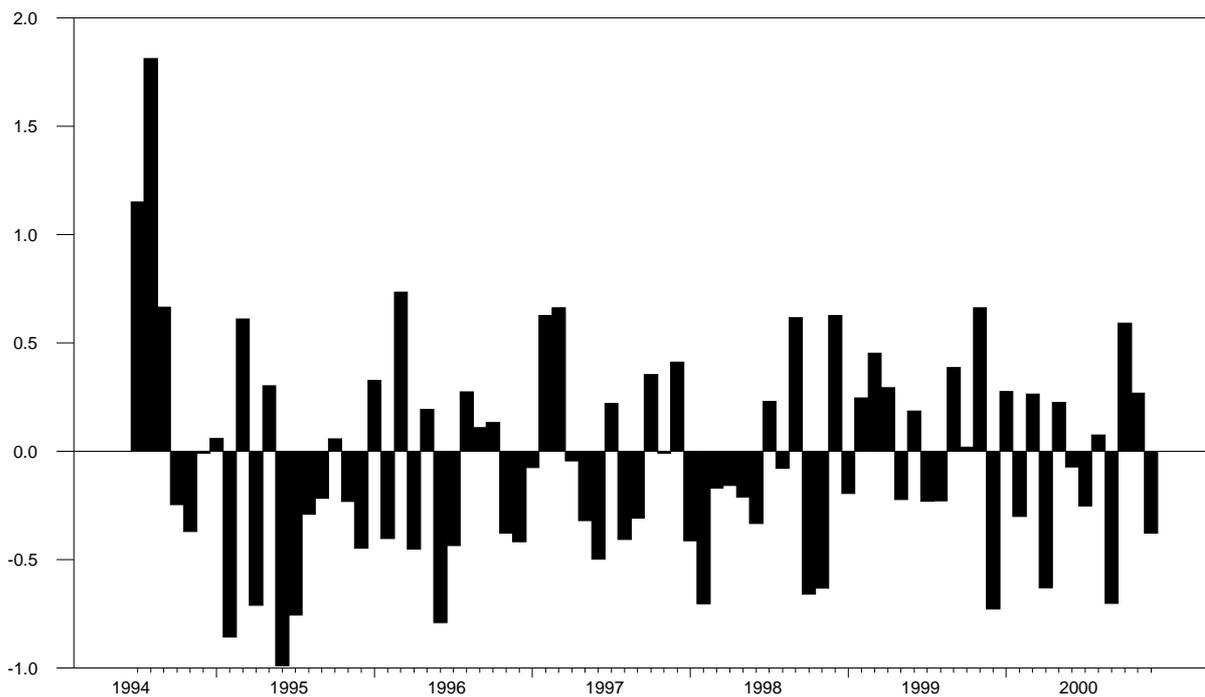


Figura 2.8.2 – Gráfico dos Resíduos Padronizados ( $a_t/s_a$ ) para o Modelo ARIMA(1,1,0)(0,1,1)<sub>12</sub>

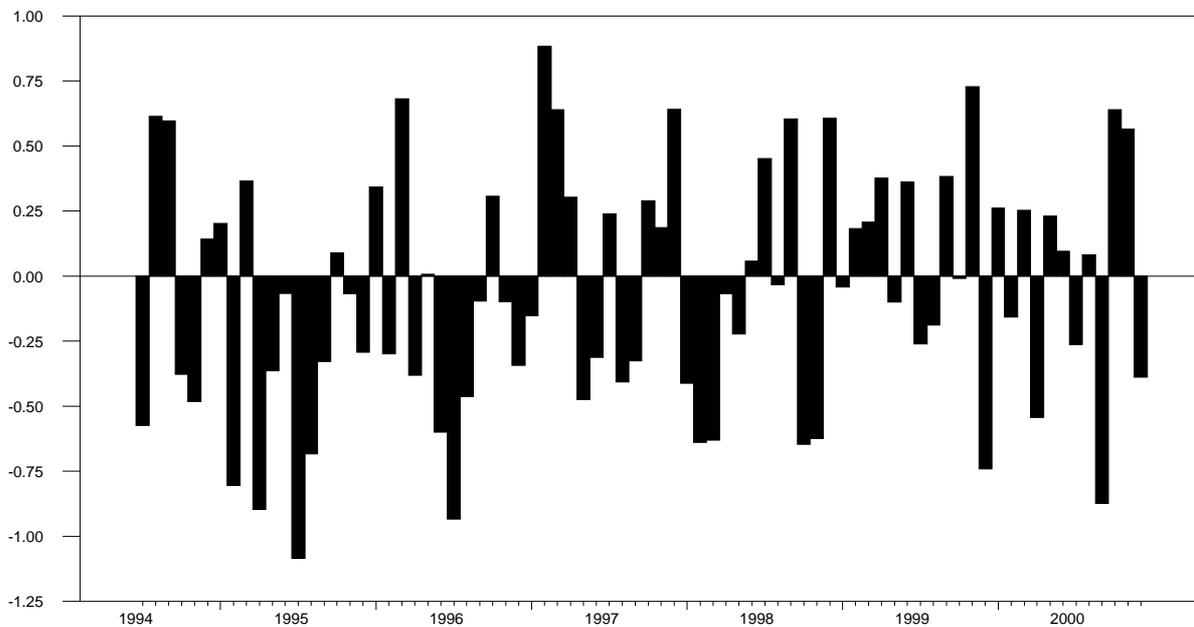


Figura 2.8.3 – Gráfico dos Resíduos Padronizados ( $a_t/s_a$ ) para o Modelo  $ARIMA(1,1,2)(1,1,1)_{12}$

**2.9 - Série Receita Total Administrada pela SRF(RECADM)**

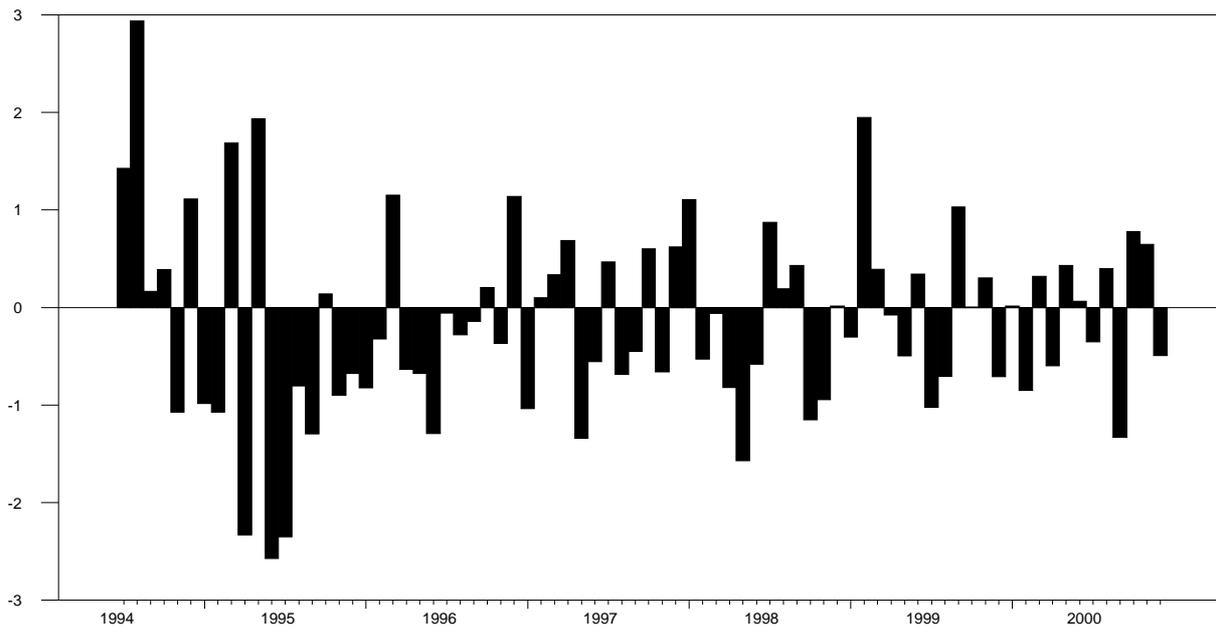


Figura 2.9.1 – Gráfico dos Resíduos Padronizados ( $a_t/s_a$ ) para o Modelo  $ARIMA(1,1,1)(0,1,1)_{12}$