

Universidade de Brasília (UNB)  
Mestrado Profissional em Economia do Setor Público

# **Previsão da Receita Tributária Federal por Base de Incidência**

Fernando Covelli Benelli  
Orientador: Paulo Springer de Freitas

Brasília, abril de 2013



Fernando Covelli Benelli

**Previsão da Receita Tributária Federal por Base  
de Incidência**

Dissertação apresentada ao Departamento de Economia da Universidade de  
Brasília como requisito parcial para a obtenção do título de Mestre em  
Economia do Setor Público.

**Orientador: Paulo Springer de Freitas**

Brasília

2013



## RESUMO

O presente trabalho tem por objetivo aperfeiçoar a utilização de modelos de séries temporais para a previsão da arrecadação total das Receitas Federais, destacando para tal fim o emprego da modelagem multivariada, VAR, e a combinação de previsões. Atualmente, a Receita Federal do Brasil (RFB) emprega, na realização de suas previsões, o chamado Método dos Indicadores, o qual faz uso de procedimentos puramente aritméticos. Trabalhos anteriores têm demonstrado a superioridade, em termos de acurácia preditiva, dos modelos derivados da metodologia de séries temporais univariadas, notadamente a ARIMA, em relação ao método oficial da RFB. Sendo assim, optou-se, neste trabalho, em adotar as previsões da modelagem oficial como *benchmark*, em relação às quais serão avaliadas as previsões advindas da modelagem univariada, ARIMA, e multivariada, VAR, ou multivariada com correção de erro, VEC. Para efeitos de comparação, também foram realizadas combinações das modelagens uni e multivariadas. Foi ainda considerada a inclusão, nos sistemas VAR/VEC, de duas variáveis antecedentes do PIB: Selic, para um semestre à frente, e IBrX-100, para um trimestre à frente. Além da própria série de arrecadação federal, também foram modeladas cinco agregações de tributos, classificados de acordo com a base de incidência. A soma das previsões para cada uma dessas agregações originou uma nova fonte de previsão para o recolhimento total. Ademais, foram estimadas as funções impulso-resposta de cada grupo de incidência, a fim de avaliar a reação do volume de arrecadação a um choque na atividade econômica. As previsões foram efetuadas para 2010, e avaliadas segundo o critério RMSE, raiz do erro médio quadrático de previsão. Os resultados encontrados apontaram a modelagem VAR/VEC como a mais eficiente, em termos de acurácia preditiva, para a maioria dos grupos, aí incluídos os Tributos sobre a Renda, Tributos sobre a Folha de Salários (arrecadação previdenciária), Tributos sobre Transações Financeiras, Outros Tributos e Total das Receitas. Já para os Tributos sobre Bens e Serviços, a combinação das metodologias ARIMA e VAR/VEC mostrou os menores erros de previsão. Enfim, para o Total das Receitas, a modelagem multivariada gerou uma redução de 44% na RMSE e de 89% no erro de previsão anual, em relação ao Método dos Indicadores.

**Palavras-chave:** ARIMA, VAR, VEC, arrecadação, tributos, previsão, base de incidência, Selic, IBrX-100



## ABSTRACT

This work aims to improve the use of time series models to forecast the total collection of Federal Revenue, highlighting for this purpose the use of multivariate modeling, VAR, and the combination of forecasts. Currently, the Internal Revenue Service of Brazil (RFB) employs in making their predictions the Method of Indicators, which makes use of purely arithmetical procedures. Previous works have already shown the superiority, in terms of predictive accuracy, of the models derived from the methodology of univariate time series ARIMA, notably in relation to the official method of RFB. Therefore, this work chooses to adopt the official forecasts as a benchmark, for which shall be evaluated predictions arising from univariate ARIMA and multivariate VAR modeling, or multivariate error correction, VEC. For the purpose of comparison were also performed combinations of uni and multivariate modeling. It was also considered for inclusion in the VAR/ VEC systems, two antecedent variables of GDP: Selic, to a semester ahead, and IBrX-100, for a quarter ahead. Besides the own series of total federal revenues, five aggregates of taxes were also modeled, classified according to their base of incidence. The sum of the forecasts for each of these aggregations originated a new source of forecast for the full collection. Moreover, it was estimated the impulse-response function of each group of incidence, in order to evaluate the reaction of the storage volume to a shock on economic activity. The forecasts were made for 2010, and evaluated according to the RMSE criteria, root mean square error of prediction. The results indicated VAR / VEC models as the most efficient in terms of predictive accuracy, for most groups, including therein Income Taxes, Taxes on Payroll (Social Security revenue) Taxes on Financial Transactions, Other Taxes and Total Revenue. As for the Taxes on Goods and Services, the combination of methodologies ARIMA and VAR / VEC showed the lowest prediction errors. Anyway, for the Total Income, the multivariate modeling generated a 44% reduction in RMSE and 89% in the annual forecast error in relation to the Method of Indicators.

**Keywords:** ARIMA, VAR, VEC, tax collection, forecasts, tax incidence, Selic, IBrX-100.



## Lista de Gráficos

Gráfico 1 – Tributos sobre a Renda (em logaritmo – jan/1999 a dez/2009).....	38
Gráfico 2 – Estatísticas descritivas – Tributos sobre a Renda (em logaritmo).....	39
Gráfico 3 – Tributos sobre a Folha de Salários (em logaritmo – jan/1999 a dez/2009)..	40
Gráfico 4 – Estatísticas descritivas – Tributos sobre a Folha de Salários (em logaritmo).....	40
Gráfico 5 – Tributos sobre Bens e Serviços (em logaritmo – jan/1999 a dez/2009).....	41
Gráfico 6 – Estatísticas descritivas – Tributos sobre Bens e Serviços (em logaritmo).....	42
Gráfico 7 – Tributos sobre Transações Financeiras (em logaritmo – jan/1999 a dez/2009).....	43
Gráfico 8 – Estatísticas descritivas – Tributos sobre Transações Financeiras (em logaritmo).....	43
Gráfico 9 – Outros Tributos (em logaritmo – jan/1999 a dez/2009).....	44
Gráfico 10 – Estatísticas descritivas – Outros Tributos (em logaritmo).....	44
Gráfico 11 – Total das Receitas (em logaritmo – jan/1999 a dez/2009).....	45
Gráfico 12 – Estatísticas descritivas – Total das Receitas(em logaritmo).....	46
Gráfico 13 – Comparação das melhores previsões para Tributos sobre a Renda com os valores observados e as previsões oficiais (a preços de dez/2009 – IPCA; em R\$ milhões).....	54
Gráfico 14 – Resposta pontual e acumulada de Tributos sobre a Renda a um choque de 1% na variável Papelão Ondulado (PO).....	54
Gráfico 15 – Comparação das melhores previsões para Tributos sobre a Folha de Salários com os valores observados e as previsões oficiais (a preços de dez/2009 – IPCA; em R\$milhões).....	59
Gráfico 16 – Resposta pontual e acumulada de Tributos sobre a Folha de Salários a um choque de 1% na variável Papelão Ondulado (PO).....	59
Gráfico 17 – Comparação das melhores previsões para Tributos sobre Bens e Serviços com os valores observados e as previsões oficiais (a preços de dez/2009 – IPCA; em R\$ milhões).....	63

Gráfico 18 – Resposta pontual e acumulada de Tributos sobre Bens e Serviços a um choque de 1% na variável IBC.....	63
Gráfico 19 – Comparação das melhores previsões para Tributos sobre Transações Financeiras com os valores observados e as previsões oficiais (a preços de dez/2009 – IPCA; em R\$ milhões).....	67
Gráfico 20 – Resposta pontual e acumulada de Tributos sobre Transações Financeiras a um choque de 1% na variável Papelão Ondulado (PO).....	67
Gráfico 21 – Comparação das melhores previsões para Outros Tributos com os valores observados e as previsões oficiais (a preços de dez/2009 – IPCA; em R\$ milhões).....	71
Gráfico 22 – Resposta pontual e acumulada de Outros Tributos a um choque de 1% na variável Índice de Vendas no Varejo (IVV).....	71
Gráfico 23 – Comparação das melhores previsões para Total das Receitas com os valores observados e as previsões oficiais (a preços de dez/2009 – IPCA; em R\$ milhões).....	76
Gráfico 24 – Resposta pontual e acumulada de Total das Receitas a um choque de 1% na variável Papelão Ondulado (PO).....	76

## Lista de Tabelas

Tabela 1 - Testes PP e KPSS de Raiz Unitária.....	47
Tabela 2 - Testes para seleção dos componentes determinísticos no modelo ADF (procedimento de Doldado).....	47
Tabela 3 - Teste ADF de Raiz Unitária (procedimento de Doldado).....	48
Tabela 4 - Teste de Dickey-Pantula para múltiplas raízes unitárias.....	48
Tabela 5 - Modelagem ARIMA da série Tributos sobre a Renda - período: 01/2003 a 12/2009.....	50
Tabela 6 - Modelagem VAR/VEC da série Tributos sobre a Renda - período: 01/2003 a 12/2009.....	51
Tabela 7 - Resíduos e previsão da modelagem VAR/VEC da série Tributos sobre a Renda - período: 01/2003 a 12/2009.....	52
Tabela 8 - Teste de cointegração para as variáveis Renda x PO x IBrX.....	53
Tabela 9 - Combinação das previsões dos modelos selecionados para Tributos sobre a Renda.....	53
Tabela 10 - Modelagem ARIMA da série Tributos sobre a Folha de Salários - período: 01/2004 a 12/2009.....	55
Tabela 11 - Modelagem VAR/VEC da série Tributos sobre a Folha de Salários - período: 01/2003 a 12/2009.....	56
Tabela 12 - Resíduos e previsão da modelagem VAR/VEC da série Tributos sobre a Folha de Salários - período: 01/2003 a 12/2009.....	57
Tabela 13 - Teste de cointegração de Johansen para as variáveis Folha x PO x MS.....	58
Tabela 14 - Combinação das previsões dos modelos selecionados para Tributos sobre a Folha de Salários.....	58
Tabela 15 - Modelagem ARIMA da série Tributos sobre Bens e Serviços - período: 01/2004 a 12/2009.....	60
Tabela 16 - Modelagem VAR/VEC da série Tributos sobre a Folha de Salários - período: 01/2003 a 12/2009.....	61
Tabela 17 - Resíduos e previsão da modelagem VAR/VEC da série Tributos sobre a Folha de Salários - período: 01/2003 a 12/2009.....	61
Tabela 18 - Teste de cointegração para as variáveis BS x IBC x Selic.....	62
Tabela 19 - Combinação das previsões dos modelos selecionados para Tributos sobre Bens e Serviços.....	62
Tabela 20 - Modelagem ARIMA da série Tributos sobre Transações Financeiras - período: 01/2002 a 12/2009.....	64

Tabela 21 - Modelagem VAR/VEC da série Tributos sobre Transações Financeiras - período: 01/2002 a 12/2009.....	65
Tabela 22 - Resíduos e erros de previsão da modelagem VAR/VEC da série Tributos sobre Transações Financeiras - período: 01/2002 a 12/2009.....	65
Tabela 23 - Teste de cointegração para as variáveis BS x IBC x Selic.....	66
Tabela 24 - Combinação das previsões dos modelos selecionados para Tributos sobre Transações Financeiras.....	66
Tabela 25 - Modelagem ARIMA da série Outros Tributos - período: 01/2003 a 12/2009.....	68
Tabela 26 - Modelagem VAR/VEC da série Outro Tributos - período: 01/2003 a 12/2009.....	69
Tabela 27 - Resíduos e previsão da modelagem VAR/VEC da série Outro Tributos - período: 01/2003 a 12/2009.....	69
Tabela 28 - Teste de cointegração de Johansen para as variáveis Outros x PO.....	70
Tabela 29 - Combinação das previsões dos modelos selecionados para Outros Tributos.....	70
Tabela 30 - Modelagem ARIMA da série Total das Receitas- período: 01/2002 a 12/2009.....	72
Tabela 31 - Modelagem VAR/VEC da série Total das Receitas- período: 01/2002 a 12/2009.....	73
Tabela 32 - Resíduos e previsão da modelagem VAR/VEC da série Total das Receitas- período: 01/2002 a 12/2009.....	74
Tabela 33 - Teste de cointegração de Johansen para as variáveis Total x PO x IBrX.....	75
Tabela 34 - Combinação das previsões dos modelos selecionados para Total das Receitas.....	75
Tabela 35 - Receita Tributária por Base de Incidência - Detalhe da Agregação – 2009 (R\$ milhões).....	87
Tabela 36 – Testes de Raiz Unitária PP e KPSS para as séries componentes do VAR.....	90
Tabela 37 – Teste ADF de Raiz Unitária (procedimento de Doldado) para as séries componentes do VAR.....	91
Tabela 38 - Teste de Dickey-Pantula para múltiplas raízes unitárias para as séries componentes do VAR.....	92
Tabela 39 - Testes para seleção dos componentes determinísticos no modelo ADF (procedimento de Doldado) para as séries componentes do VAR.....	93
Tabela 40 e 41 – Estimação e previsões do modelo SARIMA (2,1,0)(1,1,0) <sub>6</sub> + 6 dos para a série Tributos sobre a Renda.....	94
Tabela 42 e 43 – Estimação e previsões do modelo VAR em 1º diferença Renda x PO x IBrX para a série Tributos sobre a Renda.....	95

Tabela 44 – Combinação – média aritmética das previsões geradas pelos modelos Indicadores e ARIMA – Tributos sobre a Renda.....	98
Tabelas 45 e 46 – Estimação e previsões do modelo SARIMA (0,1,1)(1,1,0) <sub>12</sub> + do para a série Tributos sobre a Folha de Salários.....	99
Tabelas 47 e 48 – Estimação e previsões do modelo VEC Folha x PO x MS para a série Tributos sobre a Folha de Salários.....	100
Tabela 49 – Combinação – média aritmética das previsões geradas pelos modelos ARIMA e VAR – Tributos sobre a Folha de Salários.....	102
Tabelas 50 e 51 – Estimação e previsões do modelo SARIMA (0,1,0)(0,1,1) <sub>12</sub> para a série Tributos sobre Bens e Serviços.....	103
Tabelas 52 e 53 – Estimação e previsões do modelo VEC BS x IBC para a série Tributos sobre Bens e Serviços.....	104
Tabela 54 – Combinação – média aritmética das previsões geradas pelos modelos ARIMA e VAR – Tributos sobre Bens e Serviços.....	107
Tabelas 55 e 56 – Estimação e previsões do modelo ARIMA – (2,1,0) + 4 dos para a série Tributos sobre Transações Financeiras.....	108
Tabelas 57 e 58 – Estimação e previsões do modelo VAR TF x PO para a série Tributos sobre Transações Financeiras.....	108
Tabela 59 – Combinação – média aritmética das previsões geradas pelos modelos ARIMA e VAR – Tributos sobre Transações Financeiras.....	109
Tabelas 60 e 61 – Estimação e previsões do modelo SARIMA (0,0,1)(0,1,1) <sub>3</sub> para a série Outros Tributos.....	112
Tabelas 62 e 63 – Estimação e previsões do modelo VEC Outros x IVV para a série Outros Tributos.....	113
Tabela 64 – Combinação – média aritmética das previsões geradas pelos modelos vencedores ARIMA e VAR – Outros Tributos.....	116
Tabelas 65 e 66 – Estimação e previsões para o modelo SARIMA (2,1,0)(1,1,0) <sub>12</sub> + 2 dos para a série Total das Receitas.....	117
Tabelas 67 e 68 – Estimação e previsões do modelo vencedor VEC Total x PO x IBrX para a série Total das Receitas.....	118
Tabela 69 – Combinação – média aritmética das previsões geradas pelos modelos Indicadores e ARIMA – Outros Tributos.....	120
Tabelas 70 a 75 – Previsões dos modelos para 2011.....	121
Tabela 76 – Quadro comparativo das previsões efetuadas para 2010 e 2011 (pelo RMSE).....	122
Tabela 77 – Testes de quebra estrutural para a série Tributos sobre a Renda.....	123

Tabela 78 – Testes de quebra estrutural para a série Tributos sobre a Folha de Salários.....	123
Tabela 79 – Testes de quebra estrutural para a série Tributos sobre Bens e Serviços.....	124
Tabela 80 – Testes de quebra estrutural para a série Tributos sobre a Transações Financeiras.....	125
Tabela 81 – Testes de quebra estrutural para a série Outros Tributos.....	125
Tabela 82 – Testes de quebra estrutural para a série Total das Receitas.....	126

## Lista de Abreviaturas e Siglas

ADF	Teste aumentado de Dickey-Fuller para raiz unitária
AIC	Critério de informação de Akaike
AP	Arrecadação previdenciária
AR	Modelo autorregressivo
ARIMA	Modelo autorregressivo integrado de médias móveis
ARMA	Modelo autorregressivo de médias móveis
BIC	Critério de informação bayesiano
BS	Série dos Tributos sobre Bens e Serviços
Cofins	Contribuição Social para o Financiamento da Seguridade Social
CPMF	Contribuição Provisória sobre Movimentação ou Transmissão de Valores e de Créditos e Direitos de Natureza Financeira
Cred	Série das operações de crédito totais do sistema financeiro
CSLL	Contribuição Social sobre o Lucro Líquido
Desemp	Série da taxa de desemprego
FAC	Função de autocorrelação
FACP	Função de autocorrelação parcial
Folha	Série dos Tributos sobre a Folha de Salários
IBC-Br	Série do índice de atividade econômica do Banco Central
IBrX-100	Série do índice Brasil BM&FBOVESPA
II	Imposto de Importação
IOF	Imposto sobre Operações de Crédito, Câmbio e Seguro, ou Relativos a Títulos ou Valores Mobiliários
IPCA	Índice Nacional de Preços ao Consumidor Amplo
IPI	Série do índice de Produção Industrial
IR	Imposto de Renda
IRPJ	Imposto de Renda Pessoa Jurídica

IRRF	Imposto de Renda Retido na Fonte
ITR	Imposto Territorial Rural
IVV	Série do índice de Vendas no Varejo
Juros	Série da taxa média mensal de juros das operações de crédito
KPSS	Teste de raiz unitária de Kwiatkowski-Phillips-Schmidt-Shin
LM	Teste de correlação serial do multiplicador de Lagrange de Breusch-Godfrey
MA	Modelo de médias móveis
MAD	Desvio absoluto médio
MAPE	Erro percentual absoluto médio
MS	Série da massa salarial
PIS/PASEP	Contribuição para os Programas de Integração Social e de Formação do Patrimônio do Servidor Público
PIB	Série do Produto Interno Bruto
PO	Série de expedição de caixas, acessórios e chapas - papelão ondulado
PopO	Série da população ocupada
PP	Teste de raiz unitária de Phillips-Perron
Renda	Série dos Tributos sobre a Renda
Rendim	Série do rendimento médio real do trabalho principal
RFB	Receita Federal do Brasil
RMSE	Raiz do erro quadrático médio
SARIMA	Modelo sazonal autorregressivo integrado de médias móveis
TF	Série dos Tributos sobre Transações Financeiras
VAR	Modelo autorregressivo vetorial
VEC	Modelo autorregressivo vetorial com correção de erro

# SUMÁRIO

<b>INTRODUÇÃO.....</b>	<b>1</b>
<b>1. REVISÃO DA LITERATURA.....</b>	<b>3</b>
<b>2. MÉTODOS ECONÔMICOS.....</b>	<b>9</b>
2.1 Método de Box-Jenkins.....	10
2.1.1 Convergência e Estabilidade.....	10
2.1.2 Generalizando a Metodologia de Solução.....	12
2.1.3 Estacionariedade.....	13
2.1.4 FAC e FACP.....	13
2.1.5 Sazonalidade.....	14
2.1.6 Integração.....	15
2.1.7 Raiz Unitária.....	17
2.1.8 Etapas da Metodologia Box-Jenkins.....	21
2.2 Vetor Autorregressivo (VAR).....	23
2.2.1 Estabilidade e Estacionariedade.....	24
2.2.2 Estimação.....	24
2.2.3 Identificação.....	25
2.2.4 Causalidade e Exogeneidade.....	25
2.2.5 Cointegração.....	26
2.2.6 Vetor de Correção de Erros (VEC).....	27
2.2.7 Teste de Cointegração de Johansen.....	29
2.3 Combinação de Modelos de Previsão.....	30
2.4 Métodos de Comparação de Previsões.....	31
<b>3. APRESENTAÇÃO DOS DADOS.....</b>	<b>33</b>
<b>4. ANÁLISE DAS SÉRIES.....</b>	<b>37</b>
4.1 Tributos sobre a Renda.....	37

4.2 Tributos sobre a Folha de Salários.....	39
4.3 Tributos sobre Bens e Serviços.....	41
4.4 Tributos sobre Transações Financeiras.....	42
4.5 Outros Tributos.....	44
4.6 Total das Receitas.....	45
4.7 Testes de Raiz Unitária.....	46
<b>5. RESULTADOS.....</b>	<b>49</b>
5.1 Tributos sobre a Renda.....	49
5.2 Tributos sobre a Folha de Salários.....	54
5.3 Tributos sobre Bens e Serviços.....	59
5.4 Tributos sobre Transações Financeiras.....	63
5.5 Outros Tributos.....	67
5.6 Total das Receitas.....	71
<b>6. CONCLUSÕES.....</b>	<b>79</b>
<b>7. BIBLIOGRAFIA.....</b>	<b>83</b>
<b>8. ANEXOS.....</b>	<b>87</b>

## Introdução

A atividade de previsão da arrecadação tributária federal constitui instrumento de fundamental importância para a condução da economia do país. Com base nela, determina-se, dentro da proposta orçamentária da União, o volume total de despesas para o exercício seguinte. Esse montante de recursos públicos delimitará grande parte da participação do Estado na economia. Segundo dados do IBGE para 2008, o governo brasileiro, nos três níveis de governo, foi responsável por aproximadamente 11% do emprego e 25% do consumo final de bens e serviços.

Atualmente, a realização dessa estimativa, para fins do Orçamento da União, está a cargo da Receita Federal do Brasil (RFB). Para tanto, o órgão utiliza um método de previsão oficialmente denominado de Método dos Indicadores. Em síntese, trata-se da multiplicação dos valores recolhidos em determinado ano por índices de variação (de quantidade, preços, alíquotas etc.) que buscam projetar os movimentos da economia para o ano seguinte. Nos últimos dez anos, alguns trabalhos acadêmicos, entre eles Melo (2001), Siqueira (2002) e Campos (2009), têm demonstrado a inferioridade desse método, em termos de acurácia preditiva, em relação a outros que empregam a teoria econométrica de séries temporais, notadamente o ARIMA.

O presente trabalho tem por objetivo principal colaborar no aperfeiçoamento da metodologia de previsão da arrecadação total do governo federal. Em vista disso, adota como “benchmark” os resultados obtidos da estimação oficial para esta série, para então compará-los aos provenientes das estimações uni e multivariada. As previsões são realizadas para os meses 2010. Os sistemas da modelagem VAR incluem duas ou três variáveis, em geral uma *proxy* do PIB e uma variável antecedente da atividade econômica, taxa Selic ou IBrX-100 (índice de ações da BM&FBovespa). Em casos particulares, outras variáveis também são testadas para verificar seu poder preditivo.

Em seguida, a série de recolhimento total é dividida em cinco grupos de tributos, agregados de acordo com a base de incidência: Tributos sobre a Renda, Tributos sobre a Folha de Salários, Tributos sobre Bens e Serviços, Tributos sobre Transações Financeiras e Outros Tributos. Dessa forma, obtiveram-se quatro fontes primárias de previsão para a arrecadação total, quais sejam: ARIMA e VAR aplicados sobre a série Total das Receitas e a soma das previsões dos grupos com o emprego de cada metodologia. Além disso, também foram avaliadas combinações dessas fontes, gerando cinco outros modelos de previsão.

Dentre os principais resultados, tem-se que a modelagem multivariada, VAR, mostrou desempenho superior à ARIMA nos dois maiores grupos (em volume de arrecadação): Tributos sobre a Renda e Tributos sobre a Folha de Salários. Nos outros três grupos, Tributos sobre Bens e Serviços, Tributos sobre Transações Financeiras e Outros Tributos, a média aritmética das duas modelagens apresentou as melhores previsões.

Já para a arrecadação total, a previsão mais acurada foi a obtida pela combinação ótima, isto é, pela soma das melhores previsões de cada grupo. Além disso, cabe destacar que nos Tributos sobre a Renda e nos Tributos sobre a Folha de Salários, além do próprio Total das Receitas, observou-se melhora na capacidade preditiva dos modelos VAR com a inclusão da variável antecedente da atividade econômica um trimestre à frente, IBrX-100.

O trabalho encontra-se dividido em 6 capítulos. No Capítulo 1, é apresentada a revisão da literatura. De início, procura-se efetuar uma análise mais detida do Método dos Indicadores, o qual será aplicado na maioria dos trabalhos que serão expostos em seguida. O Capítulo 2 expõe a teoria concernente às metodologias Box-Jenkins e vetores autorregressivos, bem como os critérios utilizados de comparação das previsões e os argumentos que amparam a aplicação das combinações das metodologias. No Capítulo 3, são apresentadas as seis séries de tributos sujeitas à predição, além das 12 séries componentes dos sistemas VAR, aí incluídas as quatro *proxies* do PIB e as duas variáveis antecedentes da atividade econômica (Selic e IBrX-100). No Capítulo 4, efetua-se um breve relato dos principais fatos incidentes nos impostos e contribuições que compõem as série de tributos, com ênfase na detecção de quebras estruturais. No Capítulo 5, são expostos os resultados derivados da aplicação das metodologias ARIMA e VAR, e da combinação destas. Em seguida, as metodologias são comparadas entre si com o Método dos Indicadores, e a seleção do melhor modelo é acompanhada da análise dos critérios empregados. O Capítulo 6 encerra o trabalho com a apresentação das conclusões.

# 1 Revisão da Literatura

O objetivo deste capítulo é apresentar os diferentes métodos de previsão das receitas públicas e suas respectivas aplicações em trabalhos de pesquisa. Numa primeira abordagem, os modelos preditivos podem ser classificados em Aritméticos e Econométricos. O primeiro grupo inclui aqueles cujas previsões são obtidas através da multiplicação de valores passados por índices que representem as mudanças esperadas para o período seguinte. Tal procedimento é atualmente empregado pela Receita Federal do Brasil para a previsão da arrecadação federal e recebe o nome de Método dos Indicadores (estes se referem às projeções de parâmetros macroeconômicos como inflação, taxas de câmbio e juros elaborados pela Secretaria de Política Econômica do Ministério da Fazenda – SPE). No grupo Econométricos, os modelos fazem uso da inferência estatística para realizar suas previsões, dentre esses encontram-se os derivados da teoria de séries temporais, ARIMA e VAR. Estes métodos serão abordados detalhadamente no próximo capítulo.

Outra classificação é utilizada em Santos, Ribeiro e Gobetti (2008). Neste trabalho, os modelos de previsão são divididos em quatro grupos: Indicadores (baseados no Método dos Indicadores), Keynesianos Convencionais, Ateóricos e Intertemporais. No primeiro, as previsões de um tributo para o período presente resultam da multiplicação de valores passados do mesmo tributo por indicadores macroeconômicos. Os Keynesianos Convencionais estimam equações econométricas baseadas numa teoria subjacente relativa aos fatos geradores da arrecadação (geralmente PIB e inflação). Já os Ateóricos utilizam procedimentos exclusivamente estatísticos sobre os valores passados da própria variável sujeita à predição. Destaca-se neste grupo a metodologia proposta por Box e Jenkins (1970) – ARIMA. Por fim, os modelos Intertemporais buscam explicitar as condições de sustentabilidade da dívida pública. Partem do conceito de “restrição intertemporal do governo”, em que ajustes no superávit primário são realizados a fim de manter o endividamento público num nível sustentável, para realizar testes de longo prazo de cointegração das variáveis envolvidas.

Como este, a maioria dos trabalhos sobre previsão de receitas tributárias federais no Brasil, como Melo (2001), Siqueira (2002) e Campos (2009), apresenta como marco de comparação para avaliar a qualidade de seus resultados as estimações oriundas do método de previsão oficial utilizado pela Receita Federal do Brasil (RFB), chamado Método dos Indicadores. Em razão disso, optou-se por apresentar suas principais características, tanto por sua importância na recente literatura acadêmica, como pela referência da aplicação oficial.

Trata-se do mais intuitivo dos métodos de previsão abordados aqui, pois simplesmente consiste em multiplicar os valores realizados de determinado período por índices ou indicadores que representem relevantes variações ocorridas na economia. Com isso, obtêm-se as estimativas para o período seguinte.

Em Melo (2001, p.36), o método é formulado como segue:

$$X_t = X_{t-1}(1 + \Delta P)(1 + \Delta Q)(1 + \Delta L)(1 + \Delta U) \quad (1)$$

em que:

$X_t$  = arrecadação prevista para determinado período do ano t;

$X_{t-1}$  = arrecadação efetiva do mesmo período do ano t-1;

$\Delta P$  = variação percentual do indicador de preços;

$\Delta Q$  = variação percentual do indicador de quantidades;

$\Delta L$  = variação percentual decorrente de alterações da legislação; normalmente significa variação de alíquotas;

$\Delta U$  = variação percentual de qualquer outro indicador que tenha influência na arrecadação e não possa ser enquadrado nos indicadores básicos (preço, quantidade e legislação).<sup>1</sup>

Em razão das deficiências encontradas no Método dos Indicadores, os trabalhos de previsão desenvolvidos em Melo (2001), Siqueira (2002) e Campos (2009) apresentam métodos alternativos a este para efeitos de comparação. Santos, Ribeiro e Gobetti (2008), por exemplo, criticam a dificuldade de calibrar seus indicadores *ex-ante*. Como se pode observar da equação (1), a qualidade da previsão do valor arrecadado será tanto melhor quanto melhores forem as previsões da taxa de inflação, da variação do indicador de quantidades e até das futuras legislações que venham alterar as alíquotas vigentes. O procedimento usual consiste em supor que as informações para o período de previsão sejam idênticas às do período corrente.

Peceguini (2001) recomenda o Método dos Indicadores quando da previsão de séries históricas curtas, ou seja, com poucos dados disponíveis, o que invalidaria a aplicação de modelos econométricos, que requerem mais graus de liberdade. O autor utiliza o método dos indicadores e econométricos para prever a arrecadação anual e mensal do ICMS no Estado de São Paulo, para o ano de 2000, tomando como base a série do tributo de janeiro de 1995 a dezembro de 1999. Como principal resultado, mostra que os modelos econométricos utilizados, a saber, Decomposição Clássica, Linha de Tendência e Autorregressivo obtiveram, nesta ordem, um desempenho superior ao dos Indicadores. Os critérios de seleção adotados foram as correlações entre os valores previstos e efetivos e o número de desvios superiores a 5%.

Um grupo de trabalhos mais recente buscou aplicar modelos ateóricos para a previsão de receitas tributárias. O desenvolvimento desses modelos recebeu um forte

---

<sup>1</sup> Os indicadores citados podem variar de acordo com o tributo a ser previsto. Por exemplo, na previsão do Imposto de Importação o indicador de preços será representado pela variação cambial e o de quantidades pelo volume das importações. Em outro caso, como no Imposto de Renda Pessoa Jurídica, os preços estarão atrelados ao IGP-DI, enquanto as quantidades, ao número de contribuintes.

impulso graças à contribuição de Box e Jenkins (1970), tendo sido sua aplicação largamente difundida, e desde então conhecidos por modelos ARIMA (modelos autorregressivos integrados de média móvel). Como será demonstrado nos resultados dos trabalhos a seguir, a eficiência desse processo de estimação mostrou-se claramente superior àquela dos Indicadores.

Trabalhos de destaque para o Brasil neste campo são Melo (2001), Siqueira (2002) e Guaragna e Mello (2003).

Em Melo (2001), duas metodologias de séries temporais ateóricas (ou univariadas) são utilizadas a fim de se prever o recolhimento total do Imposto de Renda, para o ano de 2000, com base na série do mesmo tributo de julho de 1994 a dezembro de 1999. Um dos métodos é o de alisamento exponencial de Holt-Winters sazonal aditivo, de menor importância no contexto deste trabalho, por tratar-se de um procedimento automático de previsão, que não envolve métodos estatísticos. Mesmo assim, os valores obtidos mostraram melhor desempenho que o dos Indicadores. Na segunda e mais extensa parte do estudo, é empregada a modelagem de Box-Jenkins. Na comparação com o Método dos Indicadores, a modelagem ARIMA permitiu reduzir o erro de previsão médio de 10% para 0,17%, para o período de janeiro a dezembro de 2000. Diante de tais resultados, o autor conclui recomendando fortemente o aprimoramento das técnicas de estimação adotadas pela Receita, e ainda sugere, de acordo com um estudo realizado por Granger e Newbold (1974), a combinação de métodos de previsão distintos a fim de melhorar a precisão das estimativas.

Ainda dentro do quadro metodológico proposto por Melo (2001), ou seja, Indicadores versus ARIMA, Siqueira (2002) amplia para dez o rol dos tributos analisados (nove federais mais o ICMS). O autor encontra um melhor desempenho da modelagem ARIMA em oito das séries estudadas. Em duas delas, Cofins adicionada ao PIS e Outras Contribuições Sociais, o Método dos Indicadores obteve melhores resultados. Entre as justificativas da superioridade do método convencional da RFB em relação ao ARIMA nesses dois últimos casos, o autor destaca: i) presença de quebras estruturais nas séries; ii) alterações legais em variáveis significantes da arrecadação (alíquotas, bases de cálculo etc.); e iii) deficiência dos métodos quantitativos em capturar informações importantes que possam alterar o padrão histórico das séries. Por isso, Siqueira (2002) adota uma postura mais cautelosa do que Melo (2001) em relação à superioridade de um modelo com relação a outro. Prefere a análise do problema caso a caso, onde as características relevantes de cada série histórica possam ser claramente identificadas, permitindo a aplicação do método mais adequado.

Em ambos os trabalhos citados acima, encontramos a defesa dos modelos univariados de séries temporais, particularmente os derivados da metodologia Box e Jenkins, frente àqueles originados de regressão multivariada. Neste último método, é estimada uma equação que relaciona a variável predita – no caso, a arrecadação tributária – com outras variáveis que possam ter influência no volume recolhido. Alertam para o problema do erro de especificação, principalmente com relação à

parametrização, devido ao grande número de variáveis envolvidas, além da presença de correlação entre elas, condição conhecida como multicolinearidade, que comprometeria a eficiência dos estimadores. A aplicação da metodologia VAR, como se verá adiante neste trabalho, procura sanar as deficiências da regressão multivariada.

Guaranga e Mello (2003) também utilizam modelos ateóricos para fazer previsões. Os autores buscaram mesclar metodologias qualitativas e quantitativas (embora esta última seja claramente dominante) a fim de se prever a receita de ICMS para o estado do Rio Grande do Sul. Para tanto, utilizaram-se os valores históricos de 1989 a 1998, e a previsão para os anos de 1999, 2000 e 2001. Adicionalmente, com o intuito de comprovar a robustez do modelo, efetuaram-se previsões para outros doze estados da federação e para o país como um todo. A parte qualitativa da modelagem consistiu na realização de ajustes em determinados meses ou dias que apresentaram um comportamento atípico. Um exemplo desses ajustes foi o acréscimo das antecipações de receita, ocorridas em dezembro de alguns anos da amostra, ao valor arrecadado em janeiro do ano seguinte. Tais ajustes possibilitaram uma suavização das séries analisadas, melhorando o trabalho de previsão.

Os Métodos dos Indicadores e de Box-Jenkins são os mais usuais na literatura de previsão de receitas. A partir dos anos 90, contudo, uma terceira metodologia também passou a ser adotada para este fim: VAR (vetores autorregressivos). O VAR é um sistema de equações simultâneas, em que todas as variáveis são explicadas por valores defasados dela própria e das demais presentes no sistema. Inicialmente foi concebido com a intenção de explicar inter-relações de longo prazo entre as variáveis do sistema. Contudo, sua utilização como instrumento de previsão tem mostrado bons resultados.

No Brasil, podemos encontrar muitos trabalhos que exploram a aplicação do VAR como instrumento de análise teórica ou histórica na área fiscal. Porém, sua utilização como metodologia de previsão de receitas tributárias ainda não é muito difundida. No âmbito das receitas tributárias federais, trabalho de destaque é o de Campos (2009). Três séries são analisadas: Imposto de Importação, Imposto sobre a Renda das Pessoas Jurídicas – Demais Obrigadas a Apuração do Lucro Real e Cofins – Demais Empresas, todas dentro da jurisdição do Estado de São Paulo, no período de 2000 a 2006, e avaliação da previsão em 2007. Sete métodos alternativos aos dos Indicadores são empregadas: dois univariados ou ateóricos, quais sejam, Box-Jenkins (ARIMA) e modelos dinâmicos univariados, e cinco multivariados ou teóricos: função de transferência, autorregressão vetorial (VAR), VAR com correção de erro (VEC), equações simultâneas e modelos estruturais.<sup>2</sup>

---

<sup>2</sup> Dentro do quadro classificatório utilizado por Peceguini (2001), o trabalho acima utiliza diversas modelagens econométricas em contraposição a uma única aritmética (Método dos Indicadores). As modelagens ARIMA, VAR e VAR com correção de erro (VEC) serão apresentadas detalhadamente nos Capítulos seguintes. Entende-se por modelo dinâmico univariado um modelo puramente autorregressivo (AR). A função de transferência equivale a uma regressão multivariada com a restrição de as variáveis independentes serem todas exógenas. O modelo de equações simultâneas pode ser encarado como um precursor do VAR. Nele, inexistente a exigência de as mesmas variáveis aparecerem em cada uma das equações, e com o mesmo número de defasagens. Já os modelos estruturais aplicam a decomposição clássica do modelo de séries temporais em tendência, ciclo e sazonalidade e fazem estimativas desses parâmetros para efeitos de previsão. Para maiores detalhes, ver Enders (1995).

Campos (2009) justifica a escolha da circunscrição de São Paulo por esta corresponder a aproximadamente 43% da arrecadação tributária federal total do país. Para medir a acurácia da previsão foi selecionado o critério da raiz do erro quadrático médio de previsão (RMSE). Os modelos com melhor desempenho para as previsões mensais referentes às séries citadas acima foram, respectivamente: VEC, dinâmico univariado e de equações simultâneas. Ressalte-se que todos apresentarem uma melhora significativa de previsão, com RMSE em média 42% inferior ao Método dos Indicadores e 35% em relação à modelagem ARIMA.

Abaixo é exibido um quadro sintético em que são comparadas as principais características dos trabalhos citados com as do presente estudo:

	Peceguini (2001)	Melo (2001)	Siqueira (2002)	G. e Mello (2003)	Campos (2009)	Presente Trabalho
Tributo sujeito à predição	ICMS (SP)	IR	10 séries de tributos <sup>2</sup>	ICMS (RS)	II, IRPJ, Cofins (SP)	6 séries de tributos federais (agregados por base de incidência) <sup>5</sup>
Período de Previsão	meses de 2000	meses de 2000	meses de 2001	meses de 1999, 2000, 2001	meses de 2007	meses de 2010
Período de Estimação	01/95 a 12/99	07/94 a 12/99	01/89 a 12/00	01/89 a 12/98	01/00 a 12/06	01/02 a 12/09
Tratamento de Quebras Estruturais	Não	Não	Sim	Não	Sim	Sim
Metodologia Aplicada	ARIMA, DC e LT <sup>1</sup>	ARIMA	ARIMA	Alisamento	ARIMA, DU, FT, VAR VEC, ES e MÊ	VAR, VEC e combinações
Benchmark	Indicadores	Indicadores	Indicadores	Margem fixa de erro de 2%	Indicadores	Indicadores
Resultado (erro médio de previsão mensal, na comparação com o benchmark - %)	de 4,7% para 3,4% (ARIMA)	de 10% para 0,17%	ARIMA superior em 8 séries <sup>3</sup>  Arrecadação Total: de 8,5% para 7,7%	1999: 0,03% 2000: 1,25% 2001: 0,08%	II: VEC: 5,7% para 4,6% IRPJ: DU: 19,3% para 11,1% Cofins: ES: 5,6% para 3,8%	VAR/VEC/Combinações superior nas 6 séries  Arrecadação Total: de 6,5% para 2,9%

1 Decomposição Clássica e Linha de Tendência

2 II, IPI, IR, Outros Impostos, Total de Impostos, CPMF, Seguridade Social (Cofins + PIS), Outras Contribuições, Total das Contribuições e Total Administrado pela RFB

3 Exceções: Seguridade Social e Outras Contribuições

4 ARIMA, Dinâmicos Univariados, Função de Transferência, VAR, VEC, Equações Simultâneas e Modelo Estrutural

5 Tributos sobre Renda, Folha de Salários (Arrecadação Previdenciária), Bens e Serviços, Transações Financeiras, Outros Tributos e Total dos Tributos

## Literatura Comparada

Ampliando a revisão da literatura para o contexto internacional, podemos verificar que os métodos de previsão empregados pelos órgãos governamentais na previsão das receitas federais variam bastante. Contudo, nos países industrializados ou de razoável solidez institucional, esses valores assumem destacada importância na política fiscal, devido à sua influência na elaboração do Orçamento. Buettner e Kauder (2010), por exemplo, apresentam uma revisão de práticas e desempenhos preditivos de vários métodos adotados por países selecionados da OCDE. Notam, em primeiro lugar, que a variação na acurácia preditiva entre os países deve-se antes à incerteza dos fundamentos macroeconômicos que ao método escolhido. Erros de predição nos valores do PIB, salientam os autores, geram distorções de expressiva magnitude tanto nos

resíduos quanto no desvio-padrão das receita tributária. Além disso, diferenças na estrutura de taxação, principalmente em relação aos tributos sobre a renda, também parecem influir na qualidade das previsões. Porém, um resultado surpreendente do trabalho foi quantificar os efeitos do nível de independência dos órgãos responsáveis – em relação a ingerências governamentais – sobre a acurácia das previsões. A análise dos autores sugere que a influência do governo tende a gerar desvios significativos das previsões quando comparadas ao seu valor esperado. Como proposta para trabalhos futuros, Buettner e Kauder sugerem uma investigação dos incentivos do governo para exercer influência nas previsões de receita tributária e das possíveis consequências dessa influência.

Uma análise estrita dos métodos de previsão empregados nos Estados Unidos pode ser encontrada em Huntley e Miller (2009). Nesse país, o próprio governo é responsável pela elaboração das estatísticas, como ocorre no Brasil. O órgão encarregado de tal finalidade chama-se CBO (Congressional Budget Office – Secretaria de Orçamento, pertencente ao Poder Legislativo). Todo ano, a CBO realiza projeções de aproximadamente 60 variáveis macroeconômicas e avalia a acurácia das mesmas em comparação com as elaboradas pela OMB (Office of Management and Budget – Secretaria de Administração e Orçamento, órgão ligado à presidência que subsidia a preparação do orçamento) e pelo Blue Chip Consensus, uma média das previsões do setor privado. No estudo em questão, propõe-se a utilização de um vetor autoregressivo (VAR) como método alternativo de avaliação tanto das variáveis nominais que diretamente afetam o Orçamento, quanto das reais, que o fazem de maneira indireta. Como conclusão, as projeções da CBO mostraram desempenho equivalente ou ligeiramente superior às do VAR, com exceção do nível de consumo real, que exibiu um forte viés de baixa.

Cabe destacar que as previsões da CBO diferenciam-se do Método dos Indicadores brasileiro sob muitos aspectos. Enquanto este último realiza projeções para o ano seguinte, a CBO o faz para 10 anos à frente. Também divide este período em curto (os dois primeiros anos) e médio prazos. As projeções de curto prazo baseiam-se em vários fatores considerados de grande influência na economia. Modelos econométricos são empregados a fim de se estimar o grau de associação entre variáveis relevantes como importação versus inflação ou PIB versus consumo, bem como para avaliar a consistência dos estimadores empregados. Contudo, a CBO não faz uso de nenhum modelo explícito no cálculo final das previsões (ao contrário do Brasil, cujo modelo é determinístico e de conhecimento geral). Para tanto, incorpora diversas fontes de informação, como projeções do setor privado, indicadores de mercados futuros, previsões orçamentárias e mudanças recentes da política monetária. Feito isso, submete uma versão preliminar de suas previsões ao crivo de conselheiros econômicos externos.

## 2 Métodos Econométricos

Neste capítulo, serão apresentadas as bases teóricas dos dois métodos econométricos adotados neste trabalho. O primeiro deles é o de Box-Jenkins, também designado por ARIMA, de natureza univariada e que servirá como base de comparação para verificar a qualidade das previsões efetuadas. O segundo é o VAR, ao qual se dará mais detida atenção, devido à sua maior complexidade e por ser muito utilizado nas projeções do Capítulo 5.

### 2.1 Método de Box-Jenkins

Até 1978, ano da publicação do trabalho seminal de Box e Jenkins (1978), a metodologia clássica de regressão linear era amplamente utilizada como o método mais confiável de realizar previsões. A boa reputação deste método, também em sua versão mais completa de equações simultâneas, assentava-se no fato de apresentar uma boa aderência à teoria econômica. A ousadia revolucionária dos autores foi a de propor um método de previsão que não fizesse qualquer referência à teoria, ou, como preferiam dizer, que “deixassem os dados falar por si mesmos”. O procedimento de estimação neste método ainda apresenta semelhanças com o modelo clássico, com o emprego dos mínimos quadrados ordinários e de suas necessárias restrições, mas, ao contrário deste, que incluía variáveis distintas numa mesma equação, apenas inclui os valores defasados da própria variável e dos termos de erro.

Posteriormente, Enders (1995) fez uma abordagem moderna e generalizante do assunto. Esse autor enxerga os modelos ARIMA como uma classe particular de equações em diferença lineares estocásticas, ressaltando que muitas das teorias econômicas exibem naturalmente uma representação deste tipo.

Considere a seguinte equação em diferença linear estocástica de ordem  $n$  e com coeficientes constantes:

$$y_t = a_0 + \sum_{i=1}^p a_i y_{t-i} + \sum_{i=0}^q \theta_i \varepsilon_{t-i} \quad (2)$$

Dentro da nomenclatura da teoria de séries temporais, a equação (2) representa um modelo ARMA( $p,q$ ). O índice  $p$  designa o número de defasagens da variável modelada  $y$ , a parte autorregressiva do modelo – AR,  $q$  designa o número de defasagens do termo de erro ( $\varepsilon_t$ , com distribuição ruído branco<sup>3</sup>), a parte de médias móveis – MA.

---

<sup>3</sup> Tipo de distribuição probabilística (i.i.d.) em que cada termo da sequência apresenta média constante igual a zero, variância constante e ausência de autocorrelação serial entre quaisquer períodos.

Vale lembrar que a solução de uma equação em diferença é ela própria também uma equação, e não um número. Ela irá expressar  $y_t$  em termos de  $t$ ,  $\epsilon_t$  e talvez alguma constante como condição inicial de  $y_t$ . Ou seja, a substituição de  $y_t$  por sua solução, no lado direito da equação, deve igualar a substituição das variáveis defasadas de  $y_t$  ( $y_{t-1}$ ,  $y_{t-2}$ , ...) por suas respectivas soluções, também defasadas, no lado esquerdo da equação.

Uma maneira de se resolver a equação (2) é a chamada solução por iteração. O procedimento é simples e pode ser encontrado em Enders (1995, p.10). Basta dizer que, uma vez em posse da condição inicial ( $y_0$ ), podemos facilmente expressar  $y_1$  em termos de  $y_0$  e  $\epsilon_1$ , o que permite fazer a mesma coisa com  $y_2$ , expressando-o em termos  $y_0$ ,  $\epsilon_1$  e  $\epsilon_2$ . Iteração significa justamente repetição, e repetindo essas substituições sucessivas vezes, podemos então encontrar a solução geral de  $y_t$ , que será função de  $t$ ,  $y_0$  e da somatória dos termos de erro de 0 a  $t$ .

### 2.1.1 Convergência e Estabilidade

Para que seja possível estimar uma série, é necessário que ela seja estacionária, o que significa que deve atender a três condições, chamadas condições de estacionariedade. Ao longo do tempo, devem ser constantes: a média, as variações em torno dessa média (variâncias) e as covariâncias entre períodos equidistantes.

Prosseguindo com a abordagem proposta por Enders (1995), ou seja, a de fazer um paralelo entre o modelo ARMA e o método de resolução das equações em diferença, torna-se útil considerar que tipo de comportamento a solução de uma equação em diferença pode assumir. Conforme  $t$  aumenta, ela será convergente quando aproximar-se cada vez mais de um determinado valor; e não convergente em dois casos: explosiva, quando distanciar-se dele, ou de caminho aleatório, quando “vagueia” sem nenhuma tendência explícita. Assim, fica claro que apenas a solução convergente será de interesse para fins de estimação. Pois decorrido algum período de tempo, atingirá seu trecho estacionário, isto é, permanecerá próxima a um dado valor, e isto corresponde à exigência de média constante do modelo estatístico.

Considere agora uma equação como em (2), mas em primeira ordem:

$$y_t = a_0 + a_1 y_{t-1} + \epsilon_t \quad (3)$$

Aplicando a método da iteração e supondo conhecida a condição inicial, temos a seguinte solução:

$$y_t = a_0 \sum_{i=0}^{t-1} a_1^i + a_1^t y_0 + \sum_{i=0}^{t-1} a_1^i \epsilon_{t-i} \quad (4)$$

Observe o primeiro e o terceiro termos do lado direito da equação (4), eles serão decisivos para determinar a natureza da solução. O primeiro é a soma de uma progressão geométrica de razão  $a_1$ , multiplicada pela constante  $a_0$ . O terceiro, a soma

dos termos de erro, cada um deles multiplicados pelo coeficiente  $a_1$  elevado a  $i$ . Sendo assim, teremos os seguintes resultados:

- Se  $|a_1| > 1$ , então os resíduos apresentarão um efeito crescente em relação a  $t$  e a série será não convergente do tipo explosiva.
- Se  $|a_1| = 1$ , então os resíduos se acumularão com o tempo e a série será não convergente do tipo caminho aleatório.
- Se  $|a_1| < 1$ , então os resíduos produzirão efeitos cada vez menores em relação a  $t$  e a série será convergente.

Obviamente, este último caso é o de maior interesse numa estimação. Dessa forma, o requisito de que  $|a_1| < 1$  adquire tal relevância que passa a chamar-se condição de estabilidade de uma série temporal.

### 2.1.2 Generalizando a Metodologia de Solução

O método da iteração para a resolução de equações em diferença é de simples aplicação. Contudo, torna-se extremamente complexo quando um número maior de defasagens é incorporado ao modelo. Felizmente, os métodos de solução de equações em diferença já se encontravam em estado bem avançado quando do surgimento da estimação de séries temporais e puderam prover um método alternativo de solução.

Antes de abordarmos o procedimento em si, cumpre definir o que é a parte homogênea de uma equação em diferença. Trata-se da combinação linear de todas as parcelas que contêm variáveis autorregressivas de  $y_t$ , isto é:

$$y_t^h = a_1 y_{t-1} + a_2 y_{t-2} + \dots \quad (5)$$

Esta parte da equação original é de fundamental importância na obtenção da solução geral. E mesmo antes de conhecida a solução geral, pode-se obter conclusões relevantes a respeito de sua natureza (convergente ou não), apenas em posse da solução homogênea, que se obtém igualando a equação (5) a zero (que então passa a chamar-se equação característica) e encontrando suas raízes (raízes características).

Por analogia ao método da iteração, demonstra-se que também neste método a solução homogênea será convergente apenas quando as raízes características forem, em módulo, menores do que 1. Por existir a possibilidade de as raízes assumirem valores imaginários, emprega-se a expressão mais rigorosa de que as raízes características devem estar dentro do círculo unitário no plano imaginário. Quando a solução homogênea converge, a geral também converge, e vice-versa.

A metodologia para encontrar a solução geral passa, ainda, pela determinação da chamada solução particular. Esta não apresenta maiores dificuldades, pois se trata de qualquer função que solucione a equação original. Por não haver uma metodologia geral

para encontrar a particular, em geral iniciam-se as tentativas sempre pelas funções mais simples, como a constante.

Por fim, demonstra-se que a combinação linear da solução particular com a solução homogênea fornece a solução geral, isto é:

$$y_t = y_t^p + y_t^h \quad (6)$$

A solução geral estará complementemente determinada, ou seja, prescindirá de qualquer constante arbitrária, ao se impor uma condição inicial ao modelo.

A exigência de que as raízes características encontrem-se dentro do círculo unitário implica algumas restrições com relação aos coeficientes da equação em diferença original. Tais resultados são úteis porque facilitam a verificação da condição de estabilidade, sem a necessidade de solucionar a equação e encontrar suas raízes.

É condição suficiente para garantir a condição de estabilidade que:

$$\sum_{i=1}^n |a_i| < 1 \quad (7)$$

### 2.1.3 Estacionariedade

Como visto anteriormente, a estacionariedade é um requisito imprescindível para que uma série temporal seja apropriadamente estimada. Isto significa que as estimativas oriundas de uma série não estacionária apresentarão problemas de viés e de eficiência, e portanto serão estimadores inadequados.

A condição de estabilidade exposta em (7) é uma condição suficiente para garantir a convergência da série, mas não a sua estacionariedade. Além dela, também é exigido que os termos de erro da equação em diferença siga um processo ruído branco. Trata-se de um processo estacionário de média zero e variância  $\sigma^2$ , simbolicamente:  $\varepsilon_t \sim RB(0, \sigma^2)$ . Com elas, pode-se garantir que, ao menos no longo prazo, a média e a variância da série serão invariáveis para cada instante  $t$ , e a autocorrelação entre períodos idênticos não se alterará. Para intuir melhor essa conclusão, basta observar que um dos termos da solução em (4) é  $a_1^t y_0$ . Este termo, que também aparece na fórmula da média da série, deverá ser igualado a zero para que esta seja constante. Assim, para uma dada condição inicial, isso ocorrerá quando  $|a_1| < 1$  e  $t \rightarrow \infty$ . Demonstram-se as propriedades da variância e da autocorrelação de maneira semelhante. Pode-se objetar, dessa forma, que uma série que tenha se iniciado recentemente teria menor probabilidade de ser estacionária. É por esse motivo que uma prática comum consiste em assumir que a série tenha se iniciado num período suficientemente distante da primeira observação.

### 2.1.4 FAC e FACP

Um problema que emerge naturalmente quando se tenta modelar um processo estocástico é o da identificação. Pode-se desconfiar que a simples inspeção gráfica não seja suficiente para determinar se se trata de um processo autorregressivo, de médias móveis ou de ambos. Uma alternativa mais segura consiste em verificar como se comporta a função de autocorrelação (FAC) da série.

A FAC expressa a magnitude da correlação entre os valores de  $y_t$  e suas defasagens  $y_{t-1}, y_{t-2}$ , e pode ser representada por um gráfico de barras bidimensional. A FACP – função de autocorrelação parcial – diferencia-se da FAC por não considerar os efeitos das defasagens intermediárias no cálculo da autocorrelação. Na equação (3), por exemplo, tem-se que o efeito de uma alteração em  $y_1$  é repassado para  $y_2$ , que por sua vez também o repassa para  $y_3$ , e assim sucessivamente. Todos esses efeitos acumulam-se no cálculo da autocorrelação direta. Já na parcial, eles são isolados e descartados, de modo que o valor encontrado represente estritamente a relação entre as duas variáveis.

O processo de identificação consiste primeiramente na determinação da FAC e da FACP teóricas de um dado processo estocástico. Trata-se de um procedimento puramente algébrico, bem sumarizado por Enders (1995). Num processo autorregressivo de ordem 1 – AR(1) – como a equação (3), têm-se as seguintes covariâncias:

$$\gamma_0 = \sigma^2/[1 - (a_1)^2] \quad (8)$$

$$\gamma_s = \sigma^2(a_1)^s/[1 - (a_1)^2] \quad (9)$$

Dividindo-se a equação (9) pela (8), obtêm-se as autocorrelações  $\rho_0 = 1, \rho_1 = a_1, \rho_2 = (a_1)^2, \dots, \rho_s = (a_1)^s$ . Lembrando que uma condição necessária para estacionariedade é  $|a_1| < 1$ , então fica claro o formato de decaimento exponencial desse correlograma.

Para encontrar a FACP entre as defasagens  $t$  e  $t - s$ , os efeitos das defasagens intermediárias devem ser eliminados. A maneira mais direta de se conseguir isso consiste em criar a série  $\{y_t^*\}$  subtraindo a média de  $y$  ( $\mu$ ) de cada observação. Em seguida, efetua-se uma regressão autorregressiva até a defasagem  $s$ :

$$y_t^* = \phi_{s1}y_{t-1}^* + \phi_{s2}y_{t-2}^* + \dots + \phi_{ss}y_{t-s}^* \quad (10)$$

O coeficiente  $\phi_{ss}$  em (10) representa a autocorrelação parcial entre  $y_t$  e  $y_{t-s}$ , ou seja, a autocorrelação entre essas duas variáveis expurgada das autocorrelações intermediárias. Tomando novamente a equação (3), conclui-se que, em um processo AR (1), apenas a autocorrelação  $\phi_{11}$  assumirá algum valor, enquanto os demais coeficientes serão nulos, formando assim um correlograma truncado na primeira defasagem.

Aplicando o mesmo procedimento para os demais processos estocásticos, a tabela seguinte pode ser construída:

Propriedades da FAC e da FACP		
Modelo	FAC	FACP
Ruído Branco	todos $\rho_s = 0$	todos $\Phi_{ss} = 0$
AR(p)	decaimento exponencial ( $a_1 > 0$ ) ou senoidal ( $a_1 < 0$ )	$\Phi_{ss} = 0$ para $s > p$
MA(q)	$\rho_s = 0$ para $s > q$	decaimento exponencial e/ou senoidal
ARMA(p,q)	decaimento exponencial e/ou senoidal depois da defasagem q	decaimento exponencial e/ou senoidal a partir da defasagem p

### 2.1.5 Sazonalidade

Os modelos teóricos acima descritos não levam em conta a importante questão da sazonalidade. Num processo autorregressivo de segunda ordem, por exemplo, os valores presentes são definidos exclusivamente em função dos valores assumidos em dois períodos anteriores mais o termo de erro. Em séries econômicas, contudo, é muito comum a presença de fatores cíclicos que provocam “picos” a períodos fixos de tempo. Em particular, as séries da arrecadação tributária tendem a apresentar uma sazonalidade bem definida, sobretudo em decorrência da própria legislação que regula o tributo. Campos (2009), cita como exemplo o caso do Imposto sobre a Renda da Pessoa Jurídica. A apuração desse tributo é efetuada trimestralmente com base no lucro real, podendo ser recolhido em até três parcelas. Como incidem juros a partir da segunda parcela, os pagamentos se concentram na primeira, conferindo sazonalidade trimestral à série.

Um procedimento freqüente quando se lida com este tipo de série consiste em aplicar técnicas de dessazonalização, como a X-11 ou a X-12, antes de se efetuar qualquer estimação. Enders (1995) elabora duas importantes críticas quanto à este procedimento. Primeiro, afirma que padrões sazonais podem permanecer mesmo após o ajuste, sobretudo quando se utiliza parte da série completa. Segundo, com base num artigo de Bell e Hillmer (1984), aconselha a estimação conjunta dos coeficientes ARMA e sazonais, a qual mostra resultados melhores do que a efetuada diretamente sobre a série ajustada.

## Modelagem

Teoricamente, a representação de modelos puramente sazonais pode assumir formas bem simples. Para uma série mensal com episódios sazonais trimestrais, tem-se que:

$$y_t = a_3 y_{t-3} + \varepsilon_t, \quad \text{onde } |a_3| < 1$$

Não haveria dificuldade na identificação desse processo. O correlograma FAC apresentaria “picos” decrescentes nas defasagens múltiplas de três, e o FACP, um único “pico” na defasagem três.

Ocorre que, na maioria das situações práticas, as séries contêm simultaneamente padrões sazonais e não sazonais. Uma forma de representar essa interação seria a de adicionar o padrão sazonal ao modelo ARMA:

$$y_t = a_1 y_{t-1} + a_3 y_{t-3} + \varepsilon_t$$

Noutra, o termo sazonal é incluído numa multiplicação:

$$(1 - a_1 L)(1 - a_3 L^3)y_t = \varepsilon_t, \text{ onde } L \text{ é o operador de } 1^{\text{a}} \text{ diferença}$$

Que também pode ser escrito como:

$$y_t = a_1 y_{t-1} + a_3 y_{t-3} - a_1 a_3 y_{t-5} + \varepsilon_t$$

Não há justificativa teórica para se preferir uma forma ou outra. A maioria dos pacotes econométricos, no entanto, costuma utilizar a segunda. Sucintamente, isso se deve ao fato de nesta encontrarmos três coeficientes com a estimação de apenas dois ( $a_1$  e  $a_3$ ).

### 2.1.6 Integração

A correta estimação de um modelo ARMA(p,q) tem como pressuposto fundamental a estacionariedade da série  $y_t$ . Se quaisquer dos critérios dessa hipótese não forem atendidos, a estimação perde validade e seus resultados não são mais estatisticamente confiáveis.

Em economia, contudo, as séries temporais geralmente apresentam um padrão não estacionário, o que levou ao desenvolvimento de uma série de técnicas que possam lidar com esse problema.

Para melhor visualizar este ponto, Enders (1995) aplica a decomposição clássica das séries temporais na solução da equação em diferença linear estocástica dada em (4):

$$y_t = \textit{tendência} + \textit{sazonal} + \textit{irregular}$$

Cada um desses três componentes pode conter fontes de não estacionariedade. No subcapítulo anterior, foram apresentadas técnicas para tratar o problema da sazonalidade. O componente irregular algumas vezes pode exibir uma variação crescente, o que viola a condição de homoscedasticidade, ou variância constante ao longo do tempo. Modelos específicos foram desenvolvidos para sanar este problema, chamados autorregressivos de heteroscedasticidade condicional (ARCH), os quais, por não ter havido necessidade de sua aplicação neste trabalho, não serão abordados aqui.

O componente de tendência costuma ser um dos mais fortes geradores de distorções na série. A sua presença pode alterar as médias ou as variâncias condicionais da variável analisada, inviabilizando a estimação. A alternativa que parecia ser a mais simples e eficiente seria a de incluir uma variável  $t$  na equação de regressão. Dessa forma, tanto a estacionariedade seria recuperada, quanto se evitaria a questão da regressão espúria, que ocorre quando o regressor e o regressando exibem uma alta correlação, na verdade reflexo da tendência comum entre eles, e não do grau de associação ou poder explicativo. Séries desse tipo passaram a ser chamadas de processos de tendência estacionária (PTE). Em sua forma mais simples, tem-se:

$$y_t = a_0 + a_1 t + \varepsilon_t$$

Subtraindo o componente de tendência de cada observação  $y_t$  torna clara a estacionariedade da série  $y_t^* = y_t - a_0 - a_1 t$ :

$$y_t^* = \varepsilon_t, \text{ onde } \varepsilon_t \sim RB(0, \sigma^2)$$

Isso aparentemente contornava o problema da tendência e tornava-se um recurso de grande aplicação prática. Um fato estilizado da literatura mostra que a maioria das séries econômicas contém uma clara tendência estacionária (Enders, p.137). Contudo, também é verdadeiro que muitas séries, mesmo sem apresentar esse tipo de tendência, ainda não exibam estacionariedade. De fato, estas ocorrem com frequência, especialmente em séries financeiras, e são chamadas de “caminho aleatório”.

Sabe-se que os processos cujas raízes características encontram-se dentro do círculo unitário possuem soluções convergentes e, portanto, tendem para um valor fixo conforme  $t$  aumenta. Por outro lado, quando uma ou mais dessas raízes for em módulo maior do que um, o comportamento da série será explosivo, inviabilizando a estimação. Na série de “caminho aleatório”, algumas raízes encontram-se no limiar do círculo unitário, ou seja, têm valor absoluto igual a um, enquanto as demais são menores do que um. É justamente essa particularidade que origina o comportamento dispersivo e errante da série, como se verá a seguir.

Um modelo simples de “caminho aleatório” pode ser escrito como:

$$y_t = y_{t-1} + \varepsilon_t$$

Isolando e igualando a zero a parte homogênea, encontra-se a raiz característica de valor igual a um. A solução geral desse processo, para uma dada condição inicial, é:

$$y_t = y_0 + \sum_{i=1}^t \varepsilon_t$$

Demonstra-se que a média condicional dessa série é constante, porém sua variância cresce com  $t$ , o que a torna não estacionária. Vale destacar o comportamento cumulativo dos termos de erro, onde os erros ocorridos no passado afetam o valor presente da série, o que não acontece nos modelos de tendência estacionária. Em razão disso, tais séries costumam apresentar tendências “temporárias”, que variam junto à somatória dos erros, e são chamadas de processos de tendência estocástica.

Agora, efetuando-se uma simples manipulação algébrica:

$$y_t - y_{t-1} = \varepsilon_t$$

Isto é, a primeira diferença de uma série de “caminho aleatório” constitui um processo estacionário, que vem a ser chamado de processo integrado de primeira ordem – I(1). Determina-se a ordem pelo número de vezes que uma série deve ser diferenciada até tornar-se estacionária, quando então atinge a ordem zero – I(0). Demonstra-se que serão necessárias tantas diferenciações quanto o número de raízes unitárias existentes na série (Enders 1995, p. 178). Os processos estacionários por diferenciação recebem o nome de processos de diferença estacionária (PDE).

Em síntese, há dois procedimentos para tornar uma série estacionária: a diferenciação ou a inclusão de uma variável de tendência. A aplicação de um ou de outro irá depender do tipo de tendência presente no processo. Em razão disso, é de suma importância a correta identificação da série. Diferenciar um PTE introduz raiz unitária não invertível no componente MA do modelo, enquanto incluir um termo determinístico de tendência em um PDE não elimina a tendência estocástica (Enders, 1995, p. 180).

Para reconhecer um PTE, um teste estatístico padrão como o t pode ser empregado para avaliar a significância da variável determinística. Contudo, a detecção de um PDE pode se tornar uma tarefa complexa, pelos motivos expostos a seguir.

### 2.1.7 Raiz Unitária

A principal dificuldade na identificação de um PDE consiste no fato de os correlogramas de um processo de raiz unitária e de outro com raízes próximas a um serem praticamente idênticos.

A alternativa mais imediata seria a de se aplicar diretamente um teste t nas raízes da equação. Por exemplo, num processo AR(1) do tipo:

$$y_t = a_1 y_{t-1} + \varepsilon_t \tag{11}$$

Tem-se, sob a hipótese nula, que  $a_1 = 0$ . No caso da hipótese ser rejeitada, a única conclusão possível é a de que a raiz é estatisticamente diferente zero. Dessa forma, a hipótese nula deveria ser alterada  $a_1 = 1$ , se a intenção é detectar a presença de

raiz unitária. Contudo, a aceitação dessa hipótese implica que o processo gerador dessa série é  $y_t = y_{t-1} + \varepsilon_t$ . Em tal situação, demonstra-se que a variância de  $y_t$  tende ao infinito conforme  $t$  aumenta, o que contraria um dos pressupostos básicos da estimação por mínimos quadrados, invalidando seus resultados.

Em vista disso, Dickey e Fuller (Dickey e Fuller, 1976), utilizando experimentos de Monte Carlo, desenvolveram um teste capaz de substituir o teste  $t$  na detecção da presença de raiz unitária, que recebe o nome de teste  $\tau$  (tau).

Por razões de conveniência e praticidade, o procedimento inicia-se a partir de uma versão modificada da equação (11). Subtraindo  $y_{t-1}$  de ambos os lados da equação, vem que:

$$\Delta y_t = \gamma y_{t-1} + \varepsilon_t \quad (12)$$

Agora, testar a hipótese  $a_1 = 1$  equivale a testar se  $\gamma = a_1 - 1 = 0$ . Outros modelos igualmente válidos incluem a adição de um termo de drift ( $a_0$ ) e/ou tendência ( $a_2 t$ ). No entanto, os valores críticos dependem da forma da regressão e do tamanho da amostra utilizada. No caso mais simples da equação acima, aplica-se a estatística  $\tau$ ; na presença de drift, a  $\tau_\mu$ ; e na presença de drift e termo de tendência, a  $\tau_\tau$ .

Noutro estudo, os mesmos autores também propuseram testes conjuntos para os coeficientes dos modelos (Dickey e Fuller, 1981). A hipótese  $\gamma = a_0 = 0$  pode ser testada utilizando-se a estatística  $\phi_1$ . A aceitação dessa hipótese implica que o modelo restrito, ou o modelo sem drift, deva ser adotado no lugar do modelo irrestrito, ou com drift. Para testar o modelo completo (com drift e tendência) contra o modelo simples, isto é,  $\gamma = a_0 = a_2 = 0$ , utiliza-se a estatística  $\phi_2$ ; já contra o modelo com drift, isto é,  $\gamma = a_2 = 0$ , utiliza-se a estatística  $\phi_3$ .

Além destes, o presente trabalho igualmente faz uso de testes de coeficientes condicionados à presença de uma raiz unitária ( $\gamma = 0$ ). No modelo completo, as estatísticas  $\tau_{\alpha\tau}$  e  $\tau_{\beta\tau}$  avaliam a presença dos termos de drift e tendência, respectivamente. No modelo com drift, a estatística  $\tau_{\alpha\mu}$  avalia a presença deste termo.

Muitas vezes a especificação dos modelos mostra-se inadequada para a realização do teste DF. Isto ocorre quando os termos de erro não apresentam uma distribuição do tipo ruído branco. Para sanar o problema, são acrescentados termos defasados em primeira diferença, como segue:

$$\Delta y_t = a_0 + \gamma y_{t-1} + \sum_{i=2}^p \beta_i \Delta y_{t-i+1} + \varepsilon_t \quad (13)$$

O número de termos defasados pode ser determinados por testes de especificação como AIC e BIC. O procedimento mais recomendado, contudo, consiste em aplicar os testes  $t$  e  $F$  num modelo com um grande número de defasagens e ir reduzindo gradativamente esse número até encontrar um coeficiente significativo. Após isso, são efetuados testes de diagnóstico para verificar se os resíduos apresentam

distribuição ruído branco. Quando aplicado neste formato, o teste é chamado de teste aumentado de Dickey-Fuller (ADF). As estatísticas dos testes DF e ADF têm a mesma distribuição assintótica, de maneira que os valores críticos continuam os mesmos.

### **Seleção dos Termos Determinísticos no Teste ADF**

No parágrafo anterior, foi apresentado um método para a determinação do número de defasagens a ser incluídas no teste ADF. A questão agora recai sobre a inclusão dos termos de drift e/ou tendência. Uma inapropriada especificação com relação a esses termos pode levar o poder do teste (definido como a probabilidade de se rejeitar a hipótese nula quando esta é falsa) a zero, enviesando fortemente as estimativas do coeficiente  $\gamma$ , o que compromete as conclusões sobre a presença de raiz unitária. O problema central resume-se a que “os teste de raiz unitária estão condicionados à presença dos regressores determinísticos, e os testes para os regressores determinísticos estão condicionados à presença da raiz unitária” (Enders, 1995, p.255).

É evidente que, realizados os testes para os três tipos de especificação (com drift, com tendência e sem ambos), um resultado simultâneo positivo para a presença de raiz unitária, por exemplo, é garantia suficiente para a aceitação dessa hipótese, uma vez que um dos modelos estará corretamente especificado. Contudo, não pode haver qualquer decisão quando ao menos um dos resultados diferir dos demais.

Em razão disso, optou-se, no presente trabalho, por privilegiar os resultados oriundos do teste ADF conforme o procedimento exposto a seguir, em detrimento de outros testes de raiz unitária como Phillips-Perron (PP) ou Kwiatkowski-Phillips-Schmidt-Shin (KPSS). Estes também são empregados aqui e apresentam algumas vantagens com relação ao ADF, como se verá adiante. No entanto, os resultados foram inconclusivos na maioria das séries analisadas, justamente por diferirem dependendo da especificação adotada.

Abaixo expõe-se uma tabela sintética. A descrição completa do procedimento pode ser encontrada em Enders (1995). Basicamente, constitui-se de sucessivas regressões iniciando do modelo mais completo (com drift, tendência e número de defasagens selecionado pelo critério AIC e BIC). A partir deste, realiza-se o teste de raiz unitária. O poder do teste em rejeitar  $\gamma = 0$  é muito baixo, e caso isto ocorra a conclusão de ausência de raiz unitária é muito segura. Se, pelo contrário, o resultado desse teste for de aceitação da hipótese nula, então o próximo passo consiste no teste dos termos determinísticos. O modelo selecionado será aquele em que os termos de drift e/ou tendência forem significantes na aplicação dos testes condicionais ( $\tau_{\beta\tau}$  e  $\tau_{\alpha\mu}$ ). Em caso de rejeição do regressor, um novo modelo sem a sua presença é estimado, reiniciando o processo. A fim de reforçar as conclusões encontradas, os testes conjuntos de significância também são empregados ( $\phi_1, \phi_2$  e  $\phi_3$ ). Tais testes são construídos exatamente da mesma maneira que os testes-F comuns:

$$\phi_i = \frac{[SQR(restrito) - SQR(irrestrito)]/r}{SQR(irrestrito)/(T-k)} \quad (13A)$$

Onde,

SQR = soma dos quadrados dos resíduos

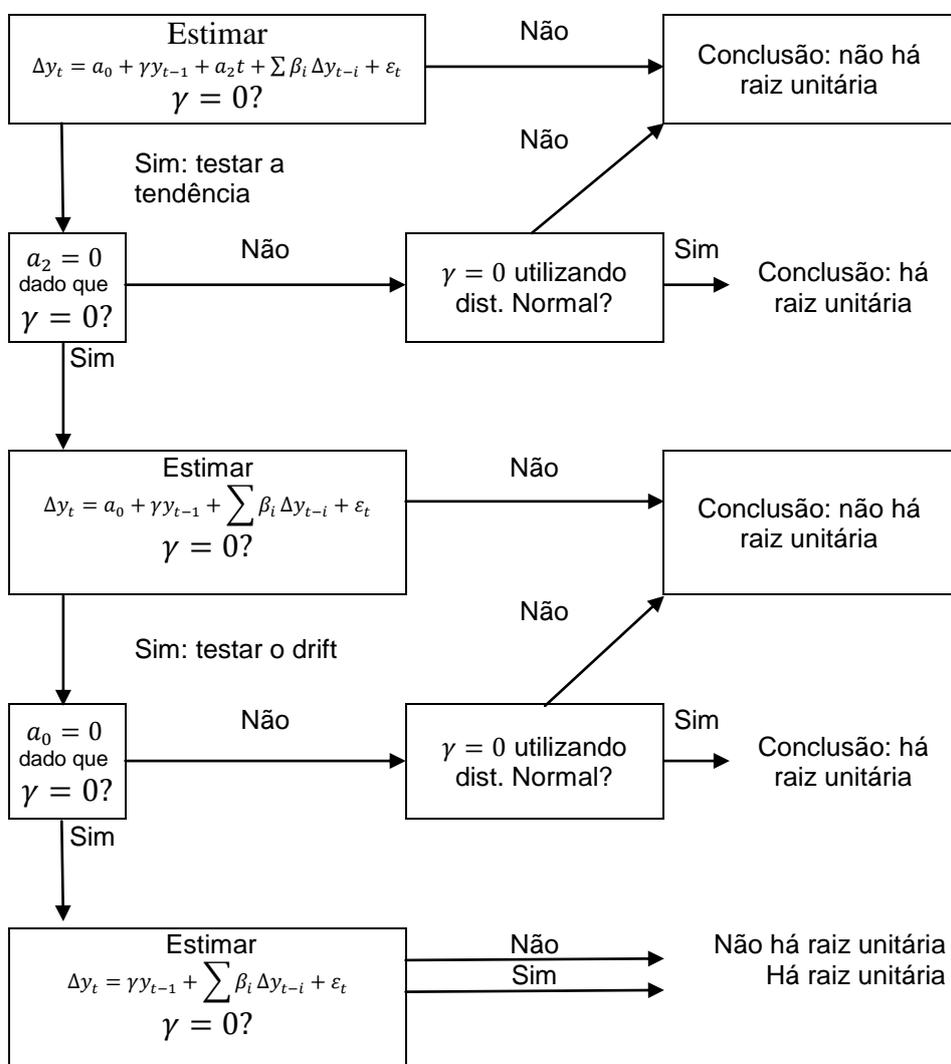
r = número de restrições

T = número de observações

k = número de parâmetros estimados no modelo irrestrito

Uma observação: devido ao fato das distribuições de Dickey-Fuller  $\tau_\tau$  e  $\tau_\mu$  convergirem para a normal, esta será aplicada em alguns testes de raiz unitária.

Tabela 1: Procedimento para a seleção dos termos determinísticos



### Teste para mais de uma raiz unitária

Na maioria das séries econômicas que apresentam tendência estocástica, uma diferenciação é suficiente para torná-las estacionárias, ou seja, contêm apenas uma raiz unitária. Casos de duas raízes unitárias são mais raros, e de três, raríssimos. Dickey e Pantula (1989) desenvolveram uma extensão do teste DF para checar essas situações. Como o presente trabalho lida com um grande número de variáveis, optou-se por aplicar também este teste. Inicia-se testando a hipótese nula de três raízes unitárias através do modelo:

$$\Delta^3 y_t = b_1 \Delta^2 y_{t-1} + \varepsilon_t$$

Os valores críticos são os mesmos do teste ADF, e o teste é efetuado sobre o coeficiente  $b_1$ . A hipótese alternativa considera a presença de duas raízes unitárias. No caso de rejeição da hipótese nula, um novo teste é aplicado, desta vez com duas raízes na hipótese nula e uma na hipótese alternativa. O modelo aplicado é o que segue:

$$\Delta^3 y_t = b_1 \Delta^2 y_{t-1} + b_2 \Delta y_{t-1} + \varepsilon_t$$

O coeficiente avaliado é  $b_2$ . Finalmente, para testar a hipótese de uma raiz unitária contra a estacionariedade da série, é utilizado o modelo mais extenso levando em consideração o coeficiente  $b_3$ :

$$\Delta^3 y_t = b_1 \Delta^2 y_{t-1} + b_2 \Delta y_{t-1} + b_3 y_{t-1} \varepsilon_t$$

### Os testes PP e KPSS

Uma hipótese fundamental dos testes DF é a de que os termos de erro seguem uma distribuição do tipo ruído branco. Para atingir este objetivo, aplica-se a versão aumentada do teste, incluindo defasagens em primeira diferença até que os resíduos sejam não autocorrelacionados com média e variância condicionais constantes.

Phillips e Perron (1988) desenvolveram um teste que permite relaxar esta hipótese, validando os resultados qualquer que seja a distribuição dos resíduos. Com isso, o teste PP também possui a vantagem de dispensar uma precisa especificação ARIMA, como se exige do teste DF, sendo seus resultados independentes do número de defasagens empregado. Vale destacar que, por construção, o teste DF é um caso particular do teste PP, quando a variância dos resíduos equivale à variância dos resíduos de longo prazo. Justamente no cálculo desta última surgem problemas de distorção ligados ao tamanho da amostra, que podem ser minimizados pela ponderação das observações através de uma função conhecida como função janela e da determinação de uma banda de truncagem. No pacote do EViews, emprega-se como procedimento padrão a janela de Bartlett e o critério de Newey-West (1994) para delimitação da banda (Bueno, 2008).

Outra dificuldade ao se empregar o teste DF é o seu baixo poder. A intenção dos autores neste estudo (Kwiatkowski-Phillips-Schmidt-Shin, 1992) seria a de complementar os testes tradicionais nas situações em que a identificação da raiz unitária não é muito clara. Difere dos demais ao propor a hipótese nula de estacionariedade com ou sem a presença de termos determinísticos.

### 2.1.8 Etapas da Metodologia Box-Jenkins

Uma das idéias fundamentais por trás da análise estatística das séries temporais é a pressuposição da existência de um processo estocástico cuja realização é a própria série em estudo. Em outras palavras, o conjunto de dados coletado é apenas uma amostra entre tantas outras possíveis. Nesse contexto, também cabe ressaltar o importante Teorema de Wold, onde se demonstra que qualquer processo estacionário pode ser decomposto em dois componentes lineares, um determinístico e outro estocástico (Bueno, 2008).

A metodologia de Box-Jenkins vem ao encontro desses dois conceitos ao propor um procedimento em 4 etapas com a finalidade de selecionar modelos e realizar previsões de séries temporais. São elas:

1. Identificação: inicia-se com a inspeção do gráfico da série para uma primeira aproximação de *outliers*, quebras estruturais ou valores ausentes. Caso a série exiba uma tendência pronunciada ou pareça vagar sem indício de média constante, então há suspeita de não estacionariedade. Aqui também são aplicados os testes de raiz unitária apresentados anteriormente. Se confirmada a suspeita, a série deve ser diferenciada. Resultados mais precisos são obtidos pela análise dos correlogramas e correlogramas parciais, além dos critérios de informação como AIC e BIC. Com isso, obtêm-se os valores máximos de  $p$  e  $q$ , a ordem de integração  $d$ , bem como a inclusão das variáveis sazonais para a construção dos modelos  $ARIMA(p,d,p)(P,D,Q)_s$ .

2. Estimação: consiste na estimação dos coeficientes autorregressivos, de médias móveis, sazonais e determinísticos dos modelos selecionados. Geralmente é empregado o método dos mínimos quadrados, embora o método da máxima verossimilhança também seja algumas vezes empregado, bem como outros métodos não-paramétricos. Enders (1995) elenca alguns critérios a ser observados nesta etapa: i) parcimônia: na prática, equivale a maximizar a relação  $R^2/(n - gl)$ , onde  $n$  é o tamanho da amostra e  $gl$  é o número de graus de liberdade ou seja, será preferível o modelo que possuir o maior poder explicativo às custas de um menor número de coeficiente ; ii) estacionariedade e invertibilidade: os coeficientes encontrados devem atender essas duas condições. Um modelo ARMA é invertível se pode ser representado por um modelo puramente autorregressivo convergente. Demonstra-se que processos estacionários não

invertíveis não podem ser corretamente estimados; iii) qualidade do ajuste: um bom modelo ajusta-se bem aos dados, o que equivale a possuir um  $R^2$  alto, sem deixar de observar o critério da parcimônia. Também é recomendável observar se os valores dos coeficientes permitem que a série convirja rapidamente, a fim de não comprometer a estabilidade do modelo.

3. Checagem de diagnóstico: consiste primeiramente em verificar se os resíduos constituem uma distribuição do tipo ruído brando, hipótese fundamental de uma estimação apropriada. Para tanto, são aplicados testes de autocorrelação serial e de heteroscedasticidade. Também é aconselhável checar a normalidade dessa distribuição, bem como a presença de *outliers* ou de períodos mal ajustados. Caso haja suspeita de os resíduos não atenderem algumas dessas exigências, todo o processo deve ser reiniciado, a partir mesmo da primeira etapa, identificação. Por isso diz-se que a metodologia Box-Jenkins é um processo iterativo.

4. Previsão: atribuição essencial da modelagem ARMA, e o que a diferencia dos demais métodos econométricos, por ter esta função como seu principal objetivo. Para chegar a esta etapa, assume-se que todas as anteriores foram completadas com êxito e que, em função disso, tanto o processo gerador de dados, como os valores passados e presentes de  $y_t$  e  $\varepsilon_t$ , são conhecidos. Em geral, a modelagem produz previsões de boa qualidade, especialmente nas de curto prazo.

## 2.2 Vetor Autorregressivo (VAR)

Na modelagem clássica de regressão linear, os efeitos de uma variável sobre a outra são considerados unidirecionais. A representação de equação única com variáveis contemporâneas ou defasadas já era suficiente para expressar esse tipo de relação. A observação dos fenômenos econômicos, todavia, mostra frequentemente a existência de reciprocidade entre os termos em estudo. Nesse contexto, a formalização do problema exige a construção de um sistema de múltiplas equações, em que cada variável é expressa como função das outras variáveis presentes, tanto exógenas como endógenas, além dos valores defasados destas últimas. Algumas dificuldades ocorrem na estimação de mínimos quadrados deste sistema, notadamente a inconsistência de seus estimadores, em razão da inevitável correlação entre os erros e as variáveis endógenas. Como se verá adiante, esta dificuldade pode ser contornada com a modificação das equações para uma versão reduzida, que, por sua vez, conduz à questão da identificação, em que se exige que algumas variáveis exógenas e outras defasadas estejam presentes somente em algumas das equações para que a estimação seja possível.

Uma crítica séria à estimação de equações simultâneas foi formulada por Christopher Sims, num artigo de 1980. Neste, o autor considera como deficiência teórica a necessidade de discriminar a priori algumas variáveis como exógenas e outras como endógenas, apontando a presença de subjetividade e arbitrariedade no processo. Logo, tendo em vista a simultaneidade dos fenômenos estudados, as variáveis devem ser todas

tratadas como endógenas, para então as relações de causalidade poderem ser corretamente avaliadas.

A apresentação da modelagem que se segue é baseada em Enders (1995) e, para efeitos didáticos, considerada um sistema formado por duas equações (bivariado):

$$y_t = b_{10} - b_{12}z_t + \gamma_{11}y_{t-1} + \gamma_{12}z_{t-1} + \varepsilon_{yt} \quad (14)$$

$$z_t = b_{20} - b_{21}y_t + \gamma_{22}y_{t-1} + \gamma_{22}z_{t-1} + \varepsilon_{zt} \quad (15)$$

Onde assume-se que (i)  $y_t$  e  $z_t$  são estacionários; (ii)  $\varepsilon_{yt}$  e  $\varepsilon_{zt}$  são resíduos ruído branco não correlacionados, com desvio-padrão  $\sigma_y$  e  $\sigma_z$ , respectivamente. Tal sistema recebe a denominação de VAR na forma estrutural, ou primitivo, pois representa a abordagem mais imediata de um fenômeno de retroalimentação.

Para que possam ser devidamente estimadas, as equações (14) e (15) precisam ser escritas na forma reduzida. Nesta, cada variável é expressa como função de suas próprias defasagens e das defasagens das demais variáveis incluídas no sistema. Assim, fica abolida qualquer relação de contemporaneidade entre  $y_t$  e  $z_t$ . Após alguma manipulação algébrica, tem-se o VAR padrão:

$$y_t = a_{10} + a_{11}y_{t-1} + a_{12}z_{t-1} + e_{1t} \quad (16)$$

$$z_t = b_{20} + a_{21}y_{t-1} + a_{22}z_{t-1} + e_{2t} \quad (17)$$

Os novos resíduos do VAR padrão  $e_{1t}$  e  $e_{2t}$  continuam seguindo a distribuição ruído branco, porém agora são correlacionados entre si.

### 2.2.1 Estabilidade e Estacionariedade

Nos modelos autorregressivos AR(p) da seção anterior, a estabilidade da equação era garantida pela imposição de que suas raízes situavam-se dentro do círculo unitário. A dedução das condições de estabilidade de um VAR padrão pode ser realizada de maneira semelhante, bastando expressar as duas equações acima em notação matricial:

$$x_t = A_0 + A_1x_{t-1} + e_t \quad (18)$$

Resolvendo a equação (18) por iteração retroativa, e analisando sua solução, mostra-se que a condição de estabilidade neste caso requer que as raízes da expressão  $(1 - a_{11}L)(1 - a_{22}L) - (a_{12}a_{21}L^2)$ , chamadas raízes características inversas, situem-se fora do círculo unitário

### 2.2.2 Estimação

Não é possível estimar um VAR em sua forma primitiva (14) e (15). Devido à existência de feedback, os resíduos  $\varepsilon_{zt}$  correlacionam-se com  $y_t$  e os resíduos  $\varepsilon_{yt}$ , com  $z_t$ , e assim viola-se uma das hipóteses fundamentais da estimação clássica. Como no

caso das equações simultâneas, contorna-se esta dificuldade regredindo as equações em sua forma reduzida, como consta na forma padrão do VAR (16) e (17).

### 2.2.3 Identificação

Na transformação algébrica do VAR estrutural para o VAR padrão, é criado um sistema de equações que estabelece a relação entre os coeficientes  $b_{ii}$  e  $\gamma_{ii}$  do primeiro com os coeficientes  $a_{ii}$  do segundo. Estimando o exemplo acima, são formadas nove equações referentes aos seis coeficientes ( $a_{10}, a_{20}, a_{11}, a_{12}, a_{21}$  e  $a_{22}$ ), mais as variâncias de cada resíduo  $\text{var}(e_{1t})$  e  $\text{var}(e_{2t})$  e a covariância entre eles  $\text{cov}(e_{1t}, e_{2t})$ . No VAR padrão, contudo, há dois coeficientes a mais, menos a covariância (que não precisa ser calculada pois vale 0), totalizando 10 parâmetros a ser estimados. Dessa forma, a menos que se restrinja um dos coeficientes do VAR padrão, não é possível recuperar os parâmetros do VAR original, e o sistema é dito subidentificado.

O tipo de restrição aplicada pode ser sugerida pela teoria econômica. Sims (1980) exemplifica a questão fazendo  $b_{21}=0$ . Isto equivale a cancelar o efeito contemporâneo de  $y_t$  sobre  $z_t$ . Com isso, a estimação dos parâmetros originais torna-se possível, pois o número de equações iguala o número de incógnitas. Também pode-se mostrar algebricamente que ambos os resíduos  $\varepsilon_{yt}$  e  $\varepsilon_{zt}$  afetam o valor contemporâneo de  $y_t$ , mas que apenas  $\varepsilon_{zt}$  afeta o valor contemporâneo de  $z_t$ . Além disso, tem-se que:

$$e_{2t} = \varepsilon_{zt}$$

$$e_{1t} = \varepsilon_{yt} - b_{12}\varepsilon_{zt}$$

A decomposição dos resíduos nesta forma triangular chama-se Decomposição de Choleski. Outras formas também podem ser empregadas, conforme o sentido econômico que se pretenda atribuir ao modelo.

### 2.2.4 Causalidade e Exogeneidade

Na análise clássica de regressão, algumas medidas de relação entre as variáveis do modelo podiam ser calculadas como a  $R^2$  e a covariância. Contudo, nenhuma delas podia fazer qualquer afirmação sobre causalidade. Esta lacuna foi preenchida, em sua forma estatística, por Granger (1969). Diz-se que uma variável Granger causa outra quando valores defasados da primeira entram na equação da segunda. Em nomenclatura econométrica, isto significa que os valores passados  $z_t$  colaboram na previsão de  $y_t$ , mantidas inalteradas as demais informações. Para se testar a hipótese da causalidade de  $z_t$  sobre  $y_t$  no VAR padrão acima, impõe-se a restrição  $a_{12}=0$ . Inversamente, para avaliar se  $y_t$  Granger causa  $z_t$ , testa-se a restrição  $a_{21} = 0$ . Podem ser utilizadas as estatísticas  $t$  e  $F$ .

A exogeneidade exige condições mais fortes para ser atendida. Neste caso, se tanto os valores passados como o valor presente de  $z_t$  não afetarem  $y_t$ , então  $z_t$  é dita uma variável exógena a  $y_t$ , e não entra na equação desta.

### 2.2.5 Cointegração

A análise das relações de equilíbrio de longo prazo constitui um vasto campo de estudos na economia. Importantes teorias como a da função consumo, a da paridade do poder de compra e as de crescimento de longo prazo, entre muitas outras, defendem a idéia de que as variáveis envolvidas ajustam-se no decorrer do tempo de acordo com uma determinada lei. A validação estatística dessas teorias, todavia, apresentava sérias dificuldades. Ela não poderia ser levada a cabo pelo emprego das séries temporais univariadas, pois estas não possuíam qualquer fundamentação teórica, além de sua principal finalidade ser a previsão. A modelagem de equações simultâneas partia da hipótese de um equilíbrio entre forças opostas (a oferta e a demanda de bens, por exemplo), mas nada podia dizer sobre a dinâmica desse equilíbrio, nem se as variáveis realmente se encontravam nesse estado ou qual seria este.

Engle e Granger (1987) criaram então o conceito de cointegração, o qual deu origem a um teste muito difundido de mesmo nome. Formalmente, iniciam considerando que um conjunto de variáveis econômicas encontra-se em equilíbrio de longo prazo quando:

$$\beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_n x_{nt} = 0$$

Para que o equilíbrio seja estatisticamente significativo, os desvios em relação a este devem seguir um processo estacionário. De modo que, em notação matricial, tem-se:

$$\beta x_t = e_t$$

Os componentes do vetor  $x_t = (x_{1t}, x_{2t}, \dots, x_{nt})$  são ditos cointegrados de ordem  $d, b$ , denominados  $x_t \sim CI(d, b)$ , se:

1. Todos os componentes de  $x_t$  são integrados de ordem  $d$ , isto é, são  $I(d)$ .
2. Existe um vetor não nulo  $\beta = (\beta_1, \beta_2, \dots, \beta_n)$  tal que a combinação linear  $\beta x_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_n x_{nt}$  é integrada de ordem  $(d - b)$ , onde  $b > 0$ .

O vetor  $\beta$  é chamado de vetor de cointegração.

Em Enders (1995) e Bueno (2008), encontram-se algumas observações que merecem destaque:

1. O termo longo prazo designa a existência de tendência estocástica nas séries. Se esta for comum a todas as variáveis, diz-se que existe um equilíbrio de longo prazo.
2. O vetor de cointegração não é único. Se  $(\beta_1, \beta_2, \dots, \beta_n)$  é um vetor de cointegração, então, para qualquer valor não nulo de  $\lambda$ ,  $(\lambda\beta_1, \lambda\beta_2, \dots, \lambda\beta_n)$

também é um vetor de cointegração. Usualmente, normaliza-se o vetor fazendo  $\lambda = 1/\beta_1$ .

3. As variáveis devem ser integradas de mesma ordem. Ainda assim, variáveis de mesma ordem podem não cointegrar, significando ausência de equilíbrio de longo prazo.
4. Se o vetor  $x_t$  possui  $n$  componentes, então haverá no máximo  $n - 1$  vetores de cointegração linearmente independentes.
5. Poucas variáveis econômicas possuem ordem de integração superior à unidade. Na literatura de séries temporais, o termo “cointegração” comumente refere-se ao caso CI(1,1).

### 2.2.6 Vetor de Correção de Erros (VEC)

A possibilidade de as variáveis incluídas num modelo VAR serem cointegradas gera modificações importantes na formalização do modelo. Sabe-se que, na presença de cointegração, as tendências estocásticas de cada uma das variáveis são anuladas devido a uma particular combinação linear. Disso se segue que os desvios de curto prazo não podem persistir por longos períodos, caso contrário o equilíbrio de longo prazo não se manteria, e se chegaria à conclusão contraditória de ausência de cointegração.

Considere agora as variáveis do vetor  $x_t$  do subcapítulo anterior. Então uma representação VAR com a presença de uma matriz de correção de erros formaliza-se como:

$$\Delta x_t = \pi_0 + \pi x_{t-1} + \sum \pi_i \Delta x_{t-i} + e_t \quad (19)$$

Onde  $\pi_0$  é o vetor de interceptos ( $n \times 1$ );  $\pi_i$  é a matriz de coeficientes ( $n \times n$ ); e  $\pi$  é uma matriz ( $n \times n$ ), em que pelo menos um dos elementos  $\pi_{jk} \neq 0$ .

Suponha agora que todas as variáveis em  $x_t$  sejam integradas de ordem 1. Segue-se disso que  $\Delta x_t$  é uma série I(0). Do lado direito da equação (19), tem-se que os elementos de  $\pi_0$  e  $\pi_i$  são constantes,  $\Delta x_{t-i}$  e  $e_t$  também são I(0), por hipótese. Logo, para que a igualdade se mantenha, a matriz  $\pi$  deve ser a matriz de cointegração, onde cada linha representa um vetor de cointegração para  $x_t$ .

Em relação à matriz  $\pi$ , cabem ressaltar dois pontos importantes;

1. Se todos seus elementos forem iguais a zero, a equação (19) passa a representar um VAR em primeira diferença tradicional. Nesse caso,  $\Delta x_t$  não responde aos desvios do período anterior do equilíbrio de longo prazo.
2. Se pelo menos um dos  $\pi_{jk}$  diferir de zero,  $\Delta x_t$  responderá a desvios do período anterior do equilíbrio de longo prazo. Logo, a representação de um

VAR em primeira diferença é inapropriada na existência do termo de correção de erro.

O modelo VAR bivariado de primeira ordem do subcapítulo anterior (sem interceptos, por simplicidade) ajudará a compreender melhor a relação entre cointegração e o VEC:

$$y_t = a_{11}y_{t-1} + a_{12}z_{t-1} + \varepsilon_{yt} \quad (20)$$

$$z_t = a_{21}y_{t-1} + a_{22}z_{t-1} + \varepsilon_{zt} \quad (21)$$

As soluções de  $y_t$  e  $z_t$  assumem a forma de equações em diferença de segunda ordem univariadas. Mais importante a se notar é que ambas possuem a mesma equação característica:

$$\lambda^2 - (a_{11} + a_{22})\lambda + (a_{11}a_{22} - a_{12}a_{21}) = 0 \quad (22)$$

Para garantir que  $y_t$  e  $z_t$  sejam CI(1,1), é necessário que uma das raízes características seja igual a um e a outra menor do que um em valor absoluto. Com isso, cada variável terá a mesma tendência estocástica, e será estacionária em primeira diferença.

Atendida a condição acima, as seguintes restrições são impostas aos coeficientes:

$$a_{11} = [(1 - a_{22}) - a_{12}a_{21}]/(1 - a_{22}) \quad (23)$$

$$a_{22} > -1 \quad (24)$$

$$a_{12}a_{21} + (a_{22})^2 < 1 \quad (25)$$

Substituindo a restrição (23) no modelo VAR, com  $a_{12}$  e  $a_{21}$  diferentes de zero e o vetor de cointegração normalizado em relação à variável  $y_t$ , obtém-se, após alguma manipulação algébrica:

$$\Delta y_t = \alpha_y(y_{t-1} - \beta z_{t-1}) + \varepsilon_{yt} \quad (26)$$

$$\Delta z_t = \alpha_z(y_{t-1} - \beta z_{t-1}) + \varepsilon_{zt} \quad (27)$$

Onde:

$$\alpha_y = -a_{12}a_{21}/(1 - a_{22})$$

$$\beta = (1 - a_{22})/a_{21}$$

$$\alpha_z = a_{21}$$

A representação (26) e (27) caracteriza um modelo VEC, que nada mais é do que um VAR em primeira diferença adicionado dos termos de correção de erro  $\alpha_y(y_{t-1} - \beta z_{t-1})$  e  $\alpha_z(y_{t-1} - \beta z_{t-1})$ . No equilíbrio de longo prazo,  $y_{t-1} = \beta z_{t-1}$ , e as variáveis mudam somente em resposta aos choques  $\varepsilon_{yt}$  e  $\varepsilon_{zt}$ . Os parâmetros  $\alpha_y$  e  $\alpha_z$  são as velocidades de ajustamento, e dimensionam o impacto dos desvios de curto do período anterior na trajetória das variáveis. A transformação efetuada acima ilustra o chamado Teorema da Representação de Granger, onde se diz que se um conjunto de variáveis são CI(1,1), então elas possuem uma representação em forma de VEC (Bueno, 2008).

Devido à sua particular configuração do modelo VEC, o conceito de causalidade de Granger precisa ser reformulado. Num sistema cointegrado,  $z_t$  não Granger causa  $y_t$  se os valores defasados de  $\Delta z_{t-i}$  não entram na equação de  $\Delta y_t$  e se, além disso,  $y_t$  não responde aos desvios do equilíbrio de longo prazo, ou seja,  $\alpha_y = 0$ .

### 2.2.7 Teste de Cointegração de Johansen

Um teste de cointegração muito aplicado na literatura recente de séries temporais foi o proposto por Johansen (1988) e Stock e Watson (1988). Para uma primeira abordagem de seu funcionamento, convém retomar o modelo VEC conforme representado pelas equações (1) e (2) e reescrevê-lo na forma matricial:

$$\begin{bmatrix} \Delta y_t \\ \Delta z_t \end{bmatrix} = \begin{bmatrix} a_{11} - 1 & a_{12} \\ a_{21} & a_{22} - 1 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{yt} \\ \varepsilon_{zt} \end{bmatrix} \quad (28)$$

A matriz com os coeficientes do lado direito da equação é a matriz dos vetores de cointegração,  $\pi$ . Note que as linhas não são linearmente independentes. Ao multiplicar a primeira linha por  $-(1 - a_{22})/a_{12}$ , ela fica idêntica à segunda. Assim, o determinante de  $\pi$  é igual a zero quando as variáveis são cointegradas.

Comparando o determinante de  $\pi$  com a equação característica dada em (22), tem-se que se a maior das raízes características for igual a unidade ( $\lambda_1 = 1$ ), então  $\pi$  tem posto igual a unidade e determinante igual a zero. Para que  $\pi$  tenha posto nulo, é necessário que  $a_{11} = 1, a_{22} = 1, e a_{12} = a_{21} = 0$ . Dessa forma, o VAR representado pelas equações (20) e (21) ficaria como  $\Delta y_t = \varepsilon_{yt}$  e  $\Delta z_t = \varepsilon_{zt}$ , o que configura as séries  $y_t$  e  $z_t$  como de raiz unitária, sem qualquer vetor de cointegração. Finalmente, se  $\pi$  possuir posto cheio, nenhuma das raízes é unitária, e as séries  $y_t$  e  $z_t$  são estacionárias.

Em seu aspecto mais importante, esses resultados mostram que o posto da matriz  $\pi$  é igual ao número de vetores de cointegração independentes. Generalizando para a matriz  $\pi$  ( $n \times n$ ) e chamando de  $\rho$  o posto dessa matriz, obtém-se:

- $\rho = 0 \Rightarrow$  séries de raiz unitária não cointegradas  $\Rightarrow$  modelo VAR em primeira diferença.
- $\rho = n \Rightarrow$  séries estacionárias  $\Rightarrow$  modelo VAR padrão

- $0 < \rho < n \Rightarrow$  séries de raiz unitária com um ou mais vetores de cointegração  $\Rightarrow$  modelo VEC

Suponha agora que as  $n$  raízes características da matriz  $\pi$  sejam ordenadas em  $\lambda_1 > \lambda_2 > \dots > \lambda_n$ . Se as variáveis envolvidas não são cointegradas,  $\rho = 0$ , e todas as raízes serão nulas. Como  $\ln(1) = 0$ , cada expressão  $\ln(1 - \lambda_i)$  será igual a zero na ausência de cointegração. Da mesma forma, se  $\rho = 1$ ,  $0 < \lambda_1 < 1$  e a expressão  $\ln(1 - \lambda_i)$  será negativa. Além disso, as demais raízes serão nulas, tal que  $\ln(1 - \lambda_i) = 0$ . Com base nisto, o teste de Johansen faz uso de duas estatísticas calculadas a partir da estimação das raízes características. São elas:

$$\lambda_{traço}(r) = -T \sum_{i=r+1}^n \ln(1 - \hat{\lambda}_i) \quad (29)$$

$$\lambda_{max}(r, r+1) = -T \ln(1 - \hat{\lambda}_{r+1}) \quad (30)$$

Onde:

$\hat{\lambda}_i$  = valores estimados das raízes características

$T$  = número de observações

A estatística (29) testa a hipótese nula de que o número de vetores de cointegração distintos é igual ou menor do que  $r$ , contra uma alternativa geral. A (30), se o número de vetores de cointegração distintos é igual a  $r$ , contra a alternativa  $r+1$ . Da discussão acima, fica claro que ambas serão iguais a zero quando todas  $\lambda_i = 0$ . Quanto mais as estimativas das raízes se distanciarem de zero, maiores serão os valores de  $\lambda_{traço}$  e  $\lambda_{max}$ .

### 2.3 Combinação de modelos de previsão

Neste trabalho serão modeladas cinco séries de tributos agregados, cuja soma equivale à série de total das receitas, que também será modelada separadamente. Uma das avaliações da qualidade da previsão consiste em analisar a acurácia preditiva da série de tributos totais em comparação com a previsão combinada dos demais grupos.

Alguns argumentos sugerem que a utilização da previsão combinada pode gerar melhores resultados (Hollauer, Issler e Notini, 2008):

- a diversificação das previsões leva à diminuição dos erros diversificáveis;
- inclusão de um maior número de variáveis e informações de especificação de outros modelos;
- possibilidade de examinar um maior número de variáveis sem o risco de sobreparametrização dos modelos VAR, que prejudica seu desempenho preditivo.

A combinação pode ser efetuada pelo emprego da média aritmética simples, da média ponderada ou pelo ajuste dos pesos por meio de mínimos quadrados ordinários. No caso, optou-se pela utilização da média aritmética, para que não haja favorecimento de nenhum dos métodos empregados, além de já ter sido observado um desempenho superior da média aritmética em relação aos métodos de mínimos quadrados (Jordan e Savioz, 2002).

#### 2.4 Métodos de comparação de previsões

Os métodos usuais na seleção do melhor modelo preditivo calculam algum tipo de média com base nos erros de previsão, definidos como a diferença entre os valores previstos e observados:

$$e_t = \hat{y}_t - y_t$$

São três os métodos mais difundidos: o desvio absoluto médio (MAD), o erro quadrático médio (MSE) e o erro percentual absoluto médio (MAPE). Os mesmos encontram-se no pacote econométrico EViews como avaliadores padrão e são calculados automaticamente.

O MAD corresponde à média dos erros absolutos:

$$MAD = \frac{\sum |e_t|}{n}$$

onde  $n$  é o número de valores previstos.

O RMSE é a raiz quadrada dos erros quadráticos médios:

$$RMSE = \sqrt{\frac{\sum e_t^2}{n}}$$

E o MAPE corresponde à média dos erros relativos (percentuais):

$$MAPE = \frac{\sum \frac{|e_t|}{|y_t|}}{n} * 100$$

Critério semelhante pode ser aplicado em previsões anualizadas, bastando para isso somar os valores previstos e observados para os doze meses correspondentes. Campos (2008) ressalta a importância dessa medida, muito utilizada para fins de orçamento, e definida como:

$$e_{anual} = \frac{\sum_1^{12} \hat{y}_t - \sum_1^{12} y_t}{\sum_1^{12} y_t} * 100$$

Quanto à preferência de um critério sobre outro, Melo (2001) recomenda a utilização do RMSE quando erros elevados de previsão são inaceitáveis, como é o caso do presente estudo. Por outro lado, acrescenta que o MAD é mais adequado quando erros desse tipo podem ser tolerados. Enquanto o MAPE é indicado na comparação de previsões de séries temporais distintas.



### 3 Apresentação dos Dados

O estudo abrange a análise de cinco grupos de receitas tributárias federais, classificadas segundo a base de incidência, mais o grupo de total das receitas. Tais grupos compreendem dez séries de receitas, entre administradas (a maioria) e não administradas pela Receita Federal do Brasil (RFB), e com períodos de observação variando caso a caso, mas sempre concluindo em dezembro de 2010. Os grupos e as séries que os compõem, de acordo com a denominação adotada pela RFB, e suas respectivas participações no recolhimento total para o ano de 2010, são:

- **Tributos sobre a Renda (Renda) – 31,55%**

Imposto de Renda (IR) – 25,85%

Contribuição Social sobre o Lucro Líquido (CSLL) – 5,70%

- **Tributos sobre a Folha de Salários (Folha) – 33,50%**

Arrecadação Previdenciária (AP) – 28,95%

PIS/Pasep – 4,55%

- **Tributos sobre Bens e Serviços (BS) – 24,92%**

Contribuição para Financiamento da Seguridade Social (Cofins) – 17,34%

Imposto sobre Produtos Industrializados (IPI) – 4,96%

Imposto de Importação (II) – 2,62%

- **Tributos sobre Transações Financeiras (TF) – 3,31%**

Imposto sobre Operações Financeiras (IOF) – 3,31%

- **Outros Tributos – 6,72%**

- **Total das Receitas<sup>4</sup> – 100%**

Procurou-se selecionar as séries de maior peso na arrecadação e que pudessem ser agrupadas de forma a racionalizar e simplificar o trabalho de previsão. Além da aplicação direta das modelagens ARIMA e VAR na série agregada de receitas, também foi realizada a combinação das previsões de cada grupo para se obter o total. Em vista disso, a ordenação acima não corresponde rigorosamente àquela adotada pela RFB em sua análise anual da carga tributária brasileira, como se pode observar pela tabela

---

<sup>4</sup> Os royalties do petróleo não são considerados tributo, mas receita patrimonial. Devido à essa diferenciação jurídica, optou-se por utilizar a denominação Total das Receitas, ao invés de Total dos Tributos.

constante do Anexo A. Só para citar alguns exemplos, a tabela da RFB contém um grupo adicional sob o título de Tributos sobre a Propriedade, e quase a totalidade dos tributos aí elencados pertence ao nível estadual ou municipal, como o IPVA e o IPTU (nesse estudo da RFB são incorporados os três níveis de governo), o que fez com que o grupo fosse retirado. Também ocorre uma discriminação mais minuciosa dos tributos, classificando subcategorias de um mesmo tributo em bases de incidência distintas. Como é o caso do IRPJ, desmembrado em IRPJ-lucro real e IRPJ-lucro presumido, o primeiro incluído na categoria de tributos sobre a renda, e o segundo na de tributos sobre bens e serviços; ou do INSS, que possui uma parcela mínima em grupos outros que não o da Folha de Salários. Evidentemente, tal discriminação atende prioritariamente a fins legais e burocráticos, não sendo adequada a uma análise estritamente econômica. Dessa forma, optou-se em classificar o total do tributo segundo sua base de incidência preponderante. Ademais, muitos tributos facilmente identificáveis pela incidência, como a CPMF, a CIDE-Combustíveis e o ITR foram incluídos na categoria Outros devido a seus valores pouco significativos. Também constam desse grupo as receitas administradas por outros órgãos.

A seguir são apresentadas as demais séries componentes da modelagem VAR. O período de observação das séries situa-se entre janeiro de 2002 e dezembro de 2010.

- Papelão Ondulado – PO – Expedição de caixas, acessórios e chapas – Toneladas – mensal – Associação Brasileira do Papelão Ondulado, Boletim Informativo (ABPO/Boletim) – em toneladas – ABPO12\_PAPEL12.
- Índice de Atividade Econômica do Banco Central – IBC-Br – fonte: Banco Central do Brasil – série nº 17439 – mensal (base 2002=100).
- Índice de Produção Industrial – IPI – fonte: IBGE (Instituto Brasileiro de Geografia e Estatística) – Produção Física industrial – Indústria Geral – Índice de base fixa mensal sem ajuste sazonal (base: média de 2002=100).
- Índice de Vendas no Varejo – IVV – fonte: Banco Central do Brasil – série nº 1455 – Índice volume de vendas no varejo (base 2003=100) – Total – Brasil.
- PIB mensal – fonte: Banco Central do Brasil – série nº 4380 – PIB mensal – Valores Correntes (R\$ milhões) – deflacionado pelo IPCA e transformado em índice.
- População Ocupada - PopO – fonte: IBGE (Instituto Brasileiro de Geografia e Estatística) – mensal – Tabela 25: total de pessoas ocupadas nas 6 regiões metropolitanas (Recife, Salvador Belo Horizonte, São Paulo, Rio de Janeiro e Porto Alegre) com 10 anos ou mais de idade.
- Rendimento – Rendim – fonte: IBGE (Instituto Brasileiro de Geografia e Estatística) – mensal – em reais – Tabela 140: Rendimento médio real do trabalho principal, efetivamente recebido no mês de referência, pelas pessoas de 10 anos ou mais de idade, ocupadas na semana de referência, total das regiões metropolitanas (Recife, Salvador Belo Horizonte, São Paulo, Rio de Janeiro e Porto Alegre).

- Selic - fonte: Banco Central do Brasil - série nº 432 - Taxa de juros - Meta Selic definida pelo Copom – unidade: % a.a. – periodicidade diária – taxas registradas pela média do mês e descontadas da inflação acumulada em 12 meses (IPCA).
- Índice Brasil BM&FBovespa – IBrX-100 – fonte: BM&FBovespa - carteira teórica composta por 100 ações selecionadas entre as mais negociadas na BM&FBovespa, em termos de número de negócios e volume financeiro – periodicidade diária (base fixada em 1000 pontos para o dia 28 de dezembro de 1995) – registrada pela variação no mês.
- Taxa de Desemprego – Desemp – fonte: IBGE (Instituto Brasileiro de Geografia e Estatística) – taxa de desemprego na região metropolitana – Brasil –mensal.
- Crédito – Cred – fonte: Banco Central do Brasil - série nº 2052 - Operações de crédito totais do sistema financeiro – mensal – unidade padrão: milhões – deflacionada pelo IPCA e transformada em índice.
- Juros – fonte: Banco Central do Brasil – série nº 8287 - Taxa média mensal (pré-fixada, pós-fixada e flutuante) das operações de crédito com recursos livres referenciais para taxa de juros - Total geral - %a.a. – mensal.

As quatro primeiras séries da lista acima (PO, IBC-Br, IPI e IVV) constituem *proxies* mensais da atividade econômica, um fator determinante do volume de arrecadação de qualquer tributo. No VAR, todos os grupos são testados com cada uma dessas variáveis, exceto a variável PIB que, como se verá adiante, possui ordem de integração incompatível com as demais séries ( $I(0)$ ). O grupo Tributos sobre Transações Financeiras, excepcionalmente, inclui apenas a variável PO, pois entendeu-se que outros fatores econômicos tinham uma influência mais direta na trajetória do IOF, como o volume de crédito total do sistema financeiro (Cred) e taxa de juros média das operações de crédito (Juros).

As séries de rendimento médio do trabalho (Rendim) e da população ocupada (PopO) são multiplicadas para se obter uma nova série de massa salarial (MS). Como é sabido, esta variável é de fundamental importância na determinação do volume de arrecadação previdenciária, e será incluída na modelagem VAR dos Tributos sobre a Folha de Salários, isoladamente ou em conjunto com alguma medida de atividade econômica. A série Rendim também é empregada na modelagem do grupo Tributos sobre a Renda, isoladamente ou acompanhada de um índice de atividade e/ou taxa de desemprego (Desemp).

As séries IBrX-100 e Selic são empregadas como indicadores financeiros antecedentes do PIB (e de suas *proxies*). Sendo a atividade econômica a variável de maior influência na trajetória da arrecadação, buscou-se incluir na modelagem outras variáveis que pudessem antecipar seu comportamento, com a intenção de melhorar o desempenho preditivo do modelo. Segundo os trabalhos de Estrela e Mishkin (1997) e

Stock e Watson (2001), os indicadores financeiros refletem as expectativas dos agentes sobre a ocorrência de eventos futuros. Além disso, apresentam características apropriadas a bons antecedentes: conformidade ao ciclo de negócios, significância econômica, precisão estatística, tempestividade e pouca necessidade de revisão das séries. Com base nisto, o Banco Central do Brasil, em seu Relatório de Inflação de 2009, desenvolveu o estudo “Indicadores Financeiros Antecedentes do PIB Brasileiro”. De um total de 45 indicadores financeiros, dois se destacaram como antecedentes da atividade econômica: o IBrX-100, para um trimestre à frente, e o spread de debêntures, para três trimestres à frente. Em função da praticidade de manipulação, foi selecionado no presente trabalho o IBrX-100 como antecedente do PIB um trimestre à frente. Com base em outro estudo do Banco Central, “A Defasagem da Transmissão da Política Monetária para Preços”, também foi utilizada a taxa Selic como previsor do estado da economia um semestre à frente.

## 4 Análise das Séries

Todas as séries dos seis grupos analisados apresentaram algum tipo de quebra depois de 2002, principalmente em razão de alterações na legislação de tributos de grande relevância. Por esse motivo, iniciam-se entre janeiro daquele ano e janeiro de 2004, e terminam em dezembro de 2010. Optou-se, dessa forma, em não incluir *dummies* de quebra estrutural na especificação dos modelos. Outras razões também contribuíram para essa escolha: i) presença de um número razoável de observações (96, em média); ii) possibilidade de se trabalhar com um trecho mais homogêneo das séries, aumentando a probabilidade de os resíduos apresentarem propriedades estatísticas desejáveis; e iii) as *dummies* alteram o intercepto das regressões, mas não as relações entre as variáveis (como muitas vezes acontece após uma alteração legislativa).

De acordo com o procedimento usual, todas as séries foram submetidas à transformação logarítmica, que apresenta as vantagens de reduzir a variabilidade dos dados, e de aproximar sua distribuição da normal (Pankratz, 1991).

Os gráficos aqui apresentados expressam os valores em logaritmo das séries deflacionados pelo IPCA. Contudo, vale lembrar que, para fins de modelagem, o logaritmo natural foi aplicado às séries transformadas em índice (base 1994=100). Aquelas cujos dados originais já constavam na forma de índice foram mantidas assim.

A análise gráfica permitirá detectar a presença de possíveis *outliers* ou de quebras estruturais. Também serão empregados procedimentos formais para se testar a presença de quebras estruturais, como os testes de Chow e Quandt-Andrews. O tratamento dos *outliers*, com a adição de *dummies* correspondentes, se fará conjuntamente à especificação do modelo, tanto pela análise gráfica preliminar como pela análise dos resíduos. A modelagem dos *outliers* será efetuada quando sua magnitude for suficiente para gerar distorções na normalidade ou homoscedasticidade dos resíduos. Como se verá adiante, tal fato ocorreu com mais frequência na modelagem VAR, em razão da existência de um número maior de séries envolvidas em cada modelo

Neste capítulo, serão apresentados apenas os gráficos das séries sujeitas à predição, ou seja, dos seis grupos de tributos. O exame das alterações mais importantes ocorridas em cada categoria, para o período de 1999 a 2009, tem por base o trabalho de Ueda (2007). Para cada grupo também serão exibidos os resultados dos testes de raiz unitária. Os mesmos testes para as outras dez séries encontram-se no Anexo B (Tabelas de 36 a 39).

### 4.1 Tributos sobre a Renda

IR – das subcategorias deste tributo, merecem destaque o IRPJ e o IRRF-Rendimentos do Trabalho, responsáveis, respectivamente, por 43% e 29% do total da arrecadação global do IR verificada em 2010. Com relação ao IRPJ, uma importante alteração ocorreu em 2002, quando sua arrecadação elevou-se 85% (em termos reais,

medido pelo IPCA) na comparação com o ano anterior. Tal resultado deveu-se à edição da MP nº 2.222, de 09/2001, que regularizou a cobrança do tributo sobre as receitas financeiras dos fundos de pensão. Com a expressiva expansão da base tributária decorrente desta medida, os fundos de pensão passaram a responder por cerca de 27% do recolhimento total desse imposto. Optou-se então por utilizar, para fins de modelagem, os dados a partir de 2003. Corrobora esta decisão o fato de não ter havido mudanças bruscas na série do IRRF-Rendimentos do Trabalho no período analisado, o que pode ser atribuído à dinâmica mais moderada do mercado de trabalho e às alterações legislativas de pequenas conseqüências no total arrecadado.

CSLL – no ano de 2002, houve um aumento real de 35% na arrecadação dessa contribuição contra o ano anterior. A principal mudança ocorreu em setembro de 2003, quando da publicação da Lei nº 10.864, que determinava a elevação da alíquota de 12% para 35% para as empresas prestadoras de serviços optantes do regime de lucro presumido. Contudo, o impacto da lei no volume arrecadado não foi percebido naquele mesmo ano, em virtude do recolhimento atípico observado em 2002, devido à contribuição extra de R\$ 2,1 bilhões decorrente da alienação de títulos públicos e da variação cambial. Assim sendo, fica claro que o corte para a CSLL poderia ser efetuado em janeiro de 2002. Todavia, para uniformizar o grupo, optou-se pelo corte mais recente, em 2003.

Gráfico 1 – Tributos sobre a Renda (em logaritmo – jan/1999 a dez/2009)

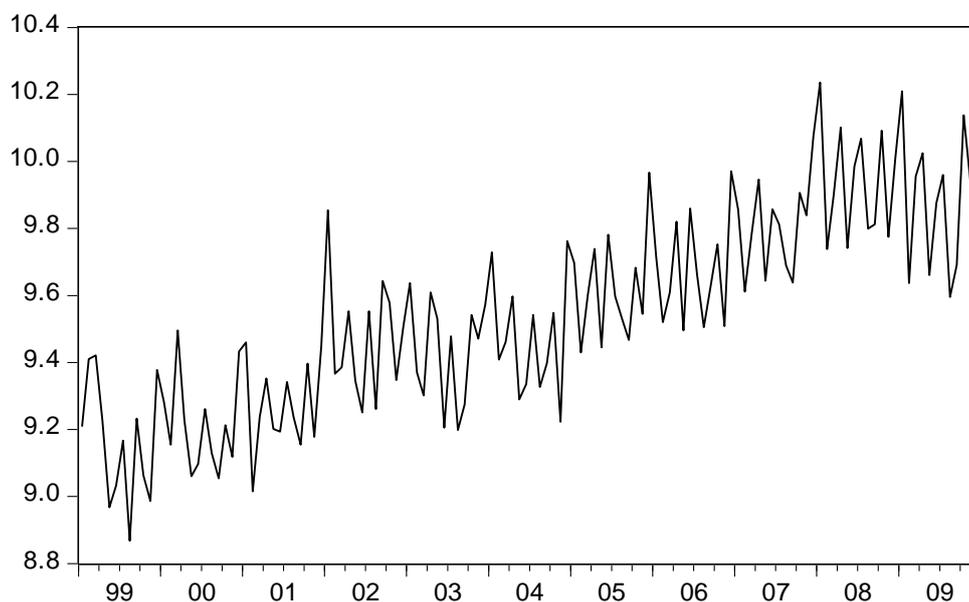
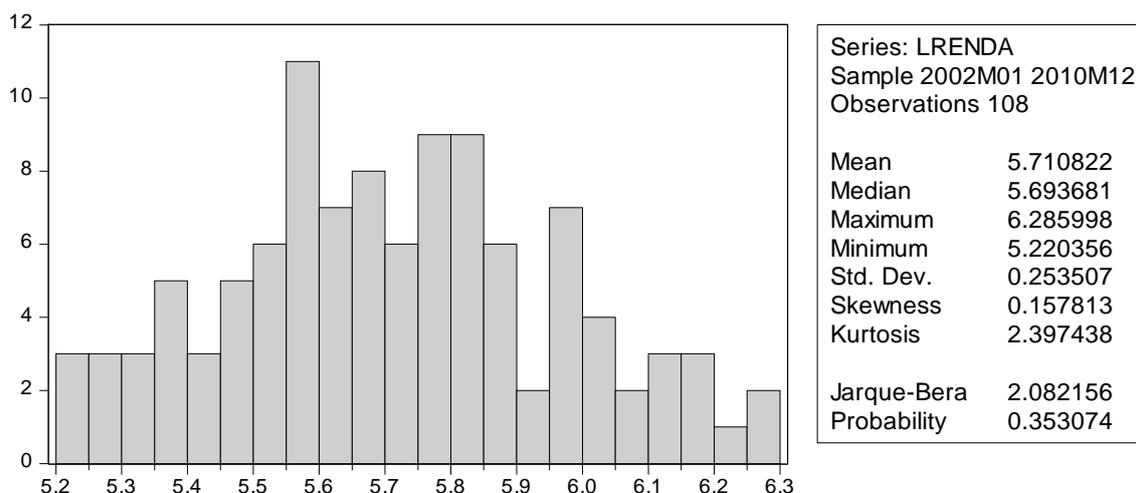


Gráfico 2 – Estatísticas descritivas – Tributos sobre a Renda (em logaritmo)



A sazonalidade da série fica evidente pela existência dos picos trimestrais, característicos dessa classe de tributos. Do início do período até dezembro de 2001, a série parece oscilar em torno de uma média constante de 9,2 (R\$ 9,9 bilhões). A partir daí, assume uma trajetória ascendente que termina em dezembro de 2007, para então novamente oscilar, desta vez com mais vigor, em torno da média de 10,1 (R\$ 24,3 bilhões). Na análise descritiva, o teste de Jarque-Bera não rejeita a hipótese de distribuição normal das observações, enquanto o coeficiente de variação (definida como a divisão do desvio-padrão pela média) foi equivalente a 3,15%. O teste de Chow de quebra estrutural indicou a presença de ponto de quebra na série em janeiro de 2003, com elevada significância (p-valor = 0,01). Já o teste de Quandt-Andrews não apontou presença de quebra ao nível de significância de 5% (os resultados dos testes de quebra estrutural realizados no Eviews encontram-se no apêndice F).

#### 4.2 Tributos sobre a Folha de Salários

Arrecadação Previdenciária – uma quebra importante para esta série ocorre a partir de 2004. Observa-se que a média de arrecadação no período 1995-2003 foi de 4,7% do PIB, crescendo para 5,1% do PIB na média do triênio seguinte, 2004-2006. Esse comportamento é explicado por dois fatores. O primeiro está ligado ao mercado de trabalho, de cuja dinâmica a arrecadação previdenciária é altamente dependente. De fato, no período 2004-2006, houve um incremento da massa salarial de 16,3%, em termos reais, acompanhado de um recuo na taxa de desocupação de 12,3% da população economicamente ativa para 10,0% no final do período. Conjugando essas duas tendências, o excelente desempenho dos rendimentos do trabalho pode ser visto como o principal fator a explicar a quebra constatada no período.

PIS/Pasep – não houve evidências de quebra estrutural no período analisado. Vale destacar, contudo, a ocorrência de uma arrecadação extraordinária de R\$ 4 bilhões

em dezembro de 2010, em função de depósito judicial efetuado por entidade financeira. Este excedente foi retirado do total arrecadado naquele mês, para que não deturpasse a acurácia das previsões.

Gráfico 3 – Tributos sobre a Folha de Salários (em logaritmo – jan/1999 a dez/2009)

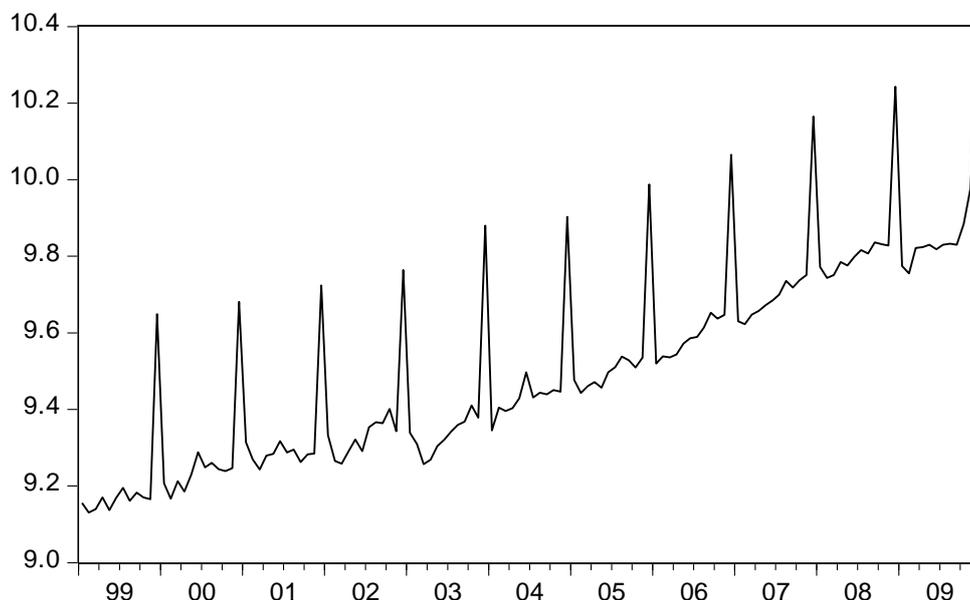
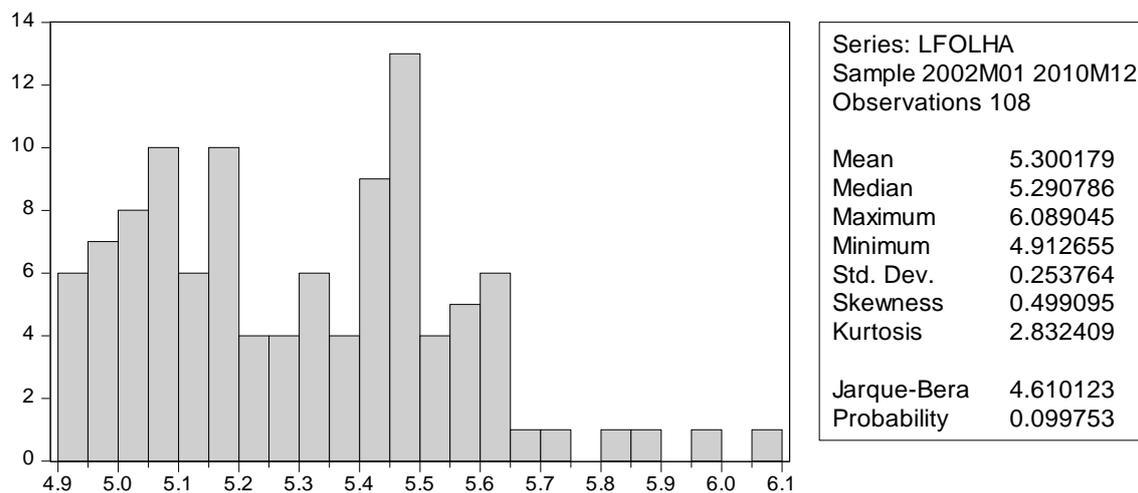


Gráfico 4 – Estatísticas descritivas – Tributos sobre a Folha de Salários (em logaritmo)



A concentração das contribuições previdenciárias no mês de dezembro confere a esta série uma sazonalidade anual. Exibe trajetória ascendente em toda sua trajetória, apresentando a existência de um patamar mais elevado a partir de 2004, pelos motivos acima expostos. As estatísticas descritivas indicam rejeição da hipótese de normalidade,

já o coeficiente de variação foi de 2,73%. Os testes de Chow e de Quandt-Andrews de detecção de quebra estrutural apontaram a inexistência de qualquer quebra na série a partir de 2002.

### 4.3 Tributos sobre Bens e Serviços

Cofins – a inclusão, em 2004, das importações em sua base tributável, foi a principal causa do crescimento real de 21% no recolhimento dessa contribuição, em comparação ao ano anterior. Também surtiu considerável impacto no volume arrecadado a edição da Lei nº 10.833, de 12/2003, que estabeleceu a não cumulatividade da Cofins. Considerando que os demais tributos que compõem o grupo, II e IPI, não mostraram alterações expressivas no período, e a própria Cofins já corresponde a cerca de 70% do total arrecadado, foi selecionado o ano de 2004 como data de corte.

Gráfico 5 – Tributos sobre Bens e Serviços (em logaritmo – jan/1999 a dez/2009)

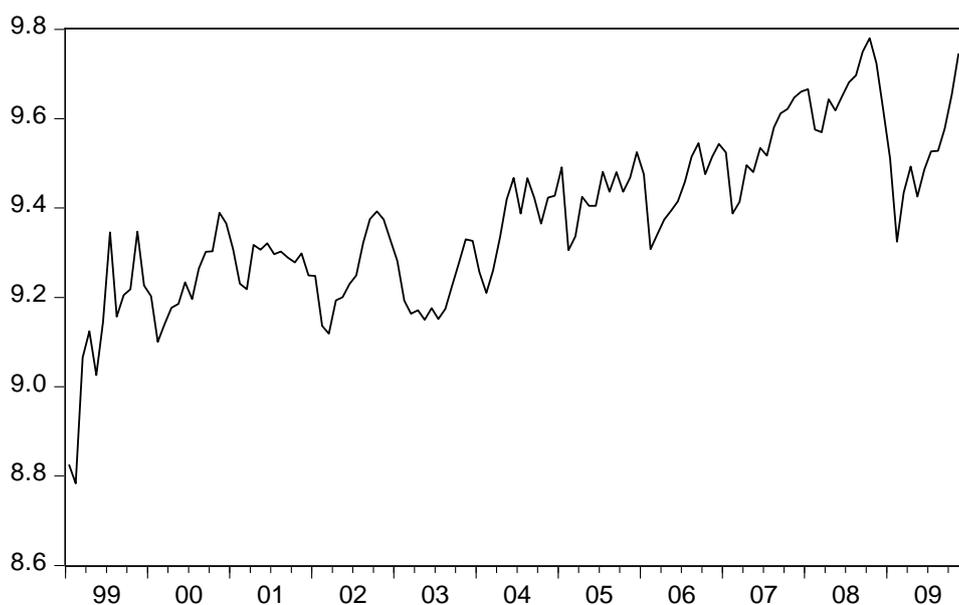
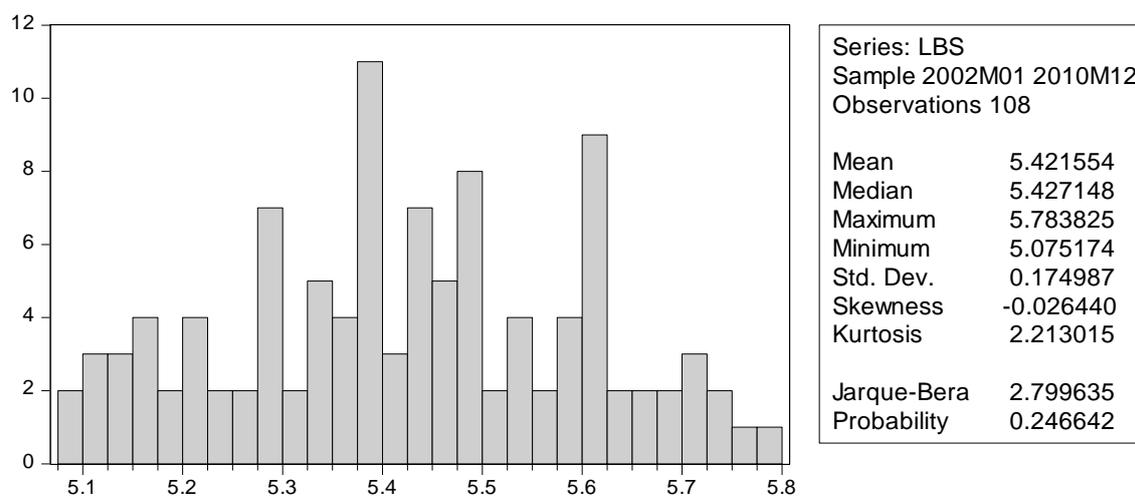


Gráfico 6 – Estatísticas descritivas – Tributos sobre Bens e Serviços (em logaritmo)



Nota-se comportamento bem distinto entre os períodos 1999-2003 e 2004-2009. No primeiro, a série aparenta relativa estabilidade, alternando-se ao redor da média constante de 9,2 (R\$ 9,9 bilhões), com exceção dos primeiros meses de 1999. Em março daquele ano, ocorreu uma quebra estrutural para cima, decorrente da majoração da alíquota da Cofins em 50% (2% para 3%). No período de 2004-2009, verifica-se uma tendência de crescimento contínuo, em virtude das razões expostas acima. As estatísticas descritivas não rejeitam a hipótese de distribuição normal dos dados. O coeficiente de variação atingiu 1,92%, a menor variabilidade entre as séries. O teste de Chow aplicado em janeiro de 2004 indicou fortemente a hipótese de aceitação de quebra estrutural nesta data ( $p$ -valor = 0). Por outro lado, o teste de Quandt-Andrews indicou dezembro de 2008 com probabilidade máxima de ocorrência de quebra estrutural.

#### 4.4 Tributos sobre Transações Financeiras

IOF – até 2008, a série apresentou um comportamento regular, sem grandes mudanças. Nesta data, contudo, o recolhimento deste tributo registrou um crescimento real de 145,7%, na comparação com o ano anterior. Essa variação expressiva está ligada à expansão do crédito, e, principalmente, à alteração da legislação relativa a este imposto. Em janeiro de 2008, através do Decreto nº 6.339, o governo elevou as alíquotas para diversas operações. Tal medida visava compensar a perda de arrecadação decorrente da não prorrogação da CPMF. Um corte nesta data disponibilizaria apenas 24 observações para fins de modelagem, referentes aos anos de 2008 e 2009, número insuficiente para gerar modelos confiáveis, notadamente na especificação VAR. Em vista disso, optou-se por utilizar o período mais longo dentre os selecionados para as séries anteriores (ou seja, 2002), e aplicar uma dummy de nível para os dados observados a partir de 2008, como pode-se observar no gráfico abaixo.

Gráfico 7 – Tributos sobre Transações Financeiras (em logaritmo – jan/1999 a dez/2009)

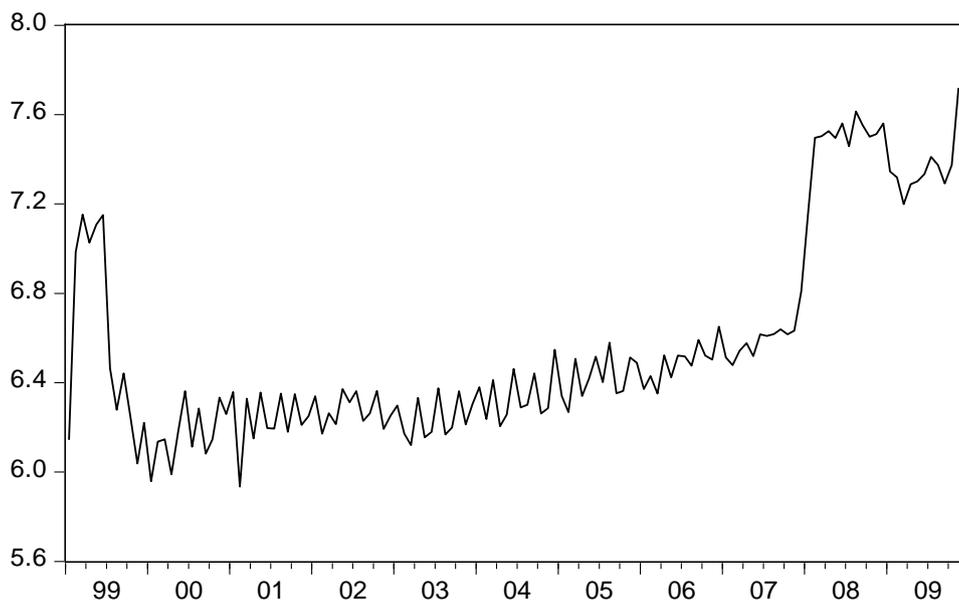
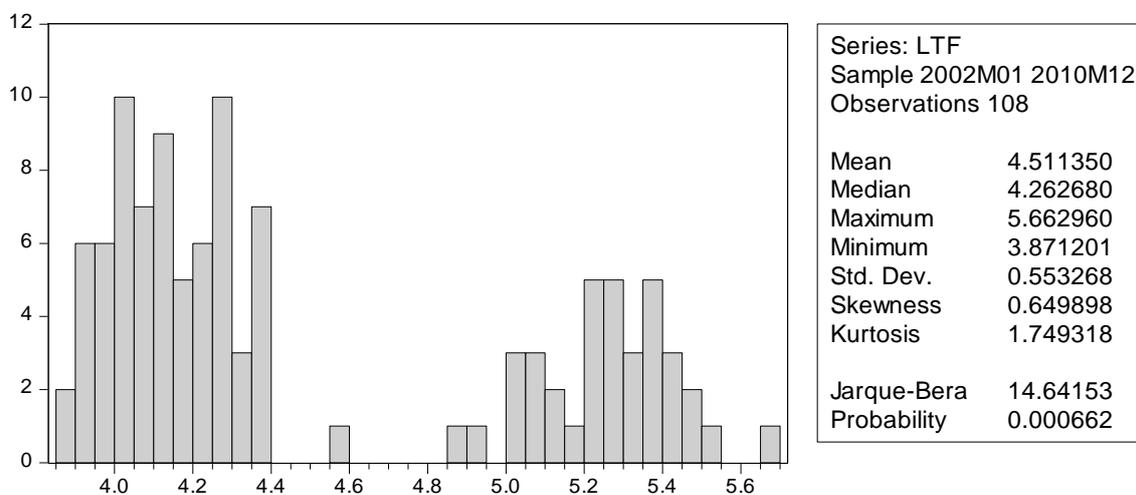


Gráfico 8 – Estatísticas descritivas – Tributos sobre Transações Financeiras (em logaritmo)



Essa mesma quebra de 2008 parece ter afetado as estatísticas descritivas. A hipótese de normalidade foi rejeitada pelo teste Jarque-Bera e o coeficiente de variação alcançou 7,01%, o maior entre as séries. Como esperado pela análise gráfica, o teste de Chow indicou a rejeição da hipótese de quebra estrutural em janeiro de 2002, enquanto o teste de Quandt-Andrews atribui probabilidade máxima de quebra estrutural em janeiro de 2008.

## 4.5 Outros Tributos

Os principais componentes deste grupo são o ITR, a CPMF, as Contribuições para o Fundaf, as Outras Receitas Administradas e as Receitas Administradas por Outros Órgãos. Até 2008, a CPMF respondia por aproximadamente 70% do total do grupo, assumindo apenas valores residuais e pouco significativos desde então. A fim de evitar distorções na previsão desta série, os valores da CPMF foram retirados. Cabe ressaltar que esta ausência não prejudica de forma alguma a previsão para a agregação geral dos tributos (série Total das Receitas), uma vez que esta é efetuada para o ano de 2010, período em que a CPMF não mais existia. O período selecionado para esta série foi de janeiro de 2003 a dezembro de 2009, e baseou-se em sua análise gráfica e buscou seguir os períodos selecionados para as demais séries, pois sua rubrica mais relevante, Receitas Administradas por Outros Órgãos, é composta por diversos tributos indiscriminados pela RFB, o que inviabiliza sua análise histórica.

Gráfico 9 – Outros Tributos (em logaritmo – jan/1999 a dez/2009)

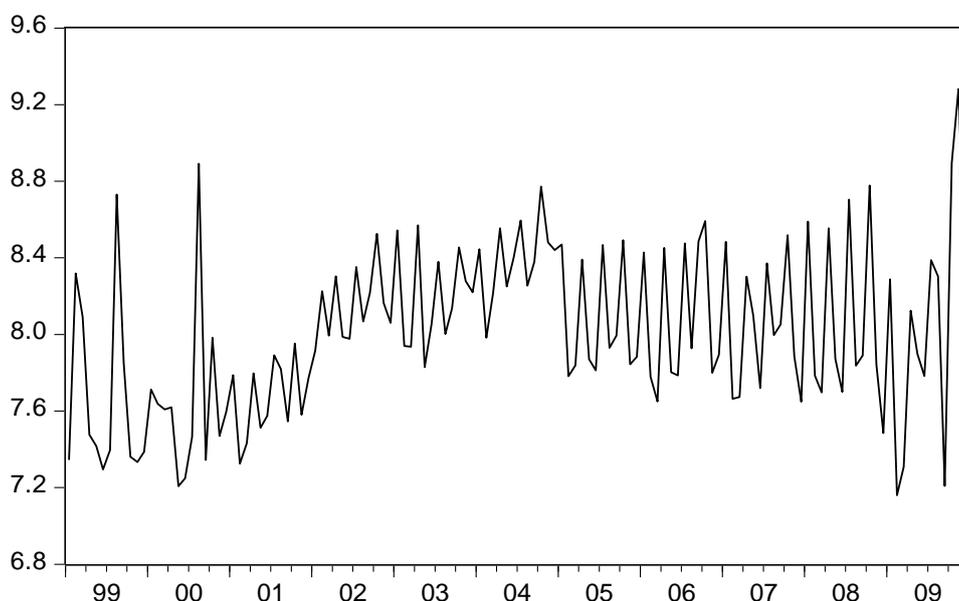
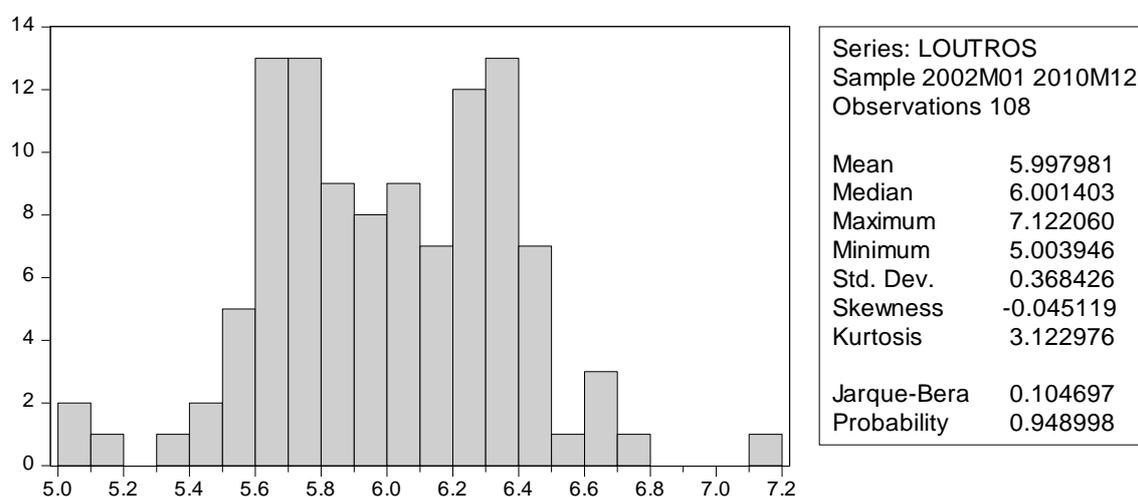


Gráfico 10 – Estatísticas descritivas – Outros Tributos (em logaritmo)



Aplicando-se o teste de Jarque-Bera, a hipótese de distribuição normal não foi rejeitada, enquanto o coeficiente de variação foi de 5,38%. Os testes formais indicaram aceitação da hipótese de quebra estrutural em janeiro de 2003 (Chow) e máxima probabilidade de ocorrência de quebra em fevereiro de 2005 (Quandt-Andrews), ambos com elevado nível de significância ( $p$ -valor = 0).

#### 4.6 Total das Receitas

Resultado da agregação de todos os cinco grupos citados, esta série é influenciada majoritariamente pelos três primeiros: para o ano de 2010, os Tributos sobre a Renda responderam por aproximadamente 32% da arrecadação; os Tributos sobre a Folha de Salários, por 33%; e os Tributos sobre Bens e Serviços, por 25%. De forma que os três grupos atingem por volta de 90% da arrecadação. Cada um deles apresenta nítidas quebras estruturais, pelos motivos expostos acima. Todavia, as magnitudes dessas quebras não se mostraram suficientes para influenciar a trajetória do total arrecadado.

Gráfico 11 – Total das Receitas(em logaritmo – jan/1999 a dez/2009)

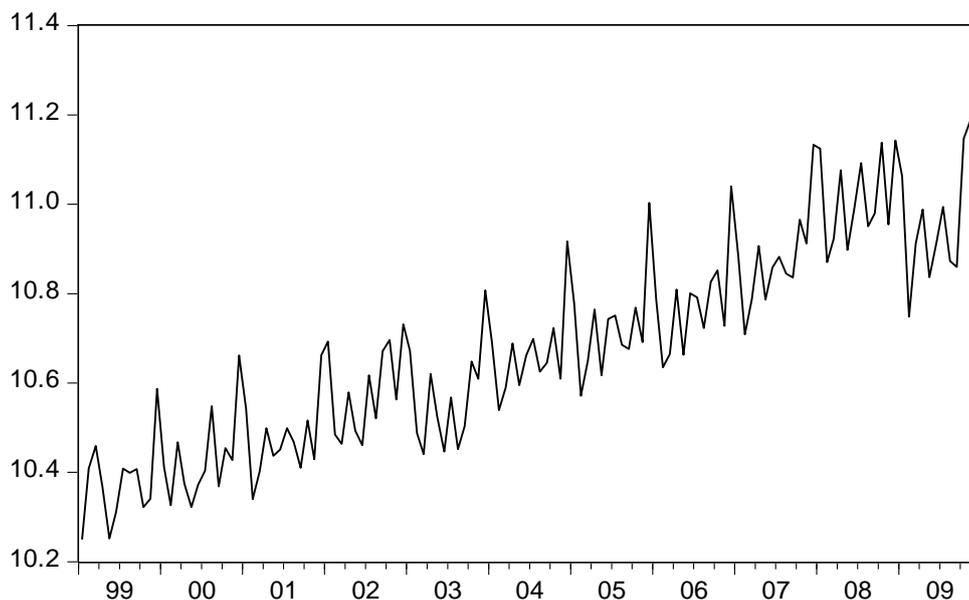
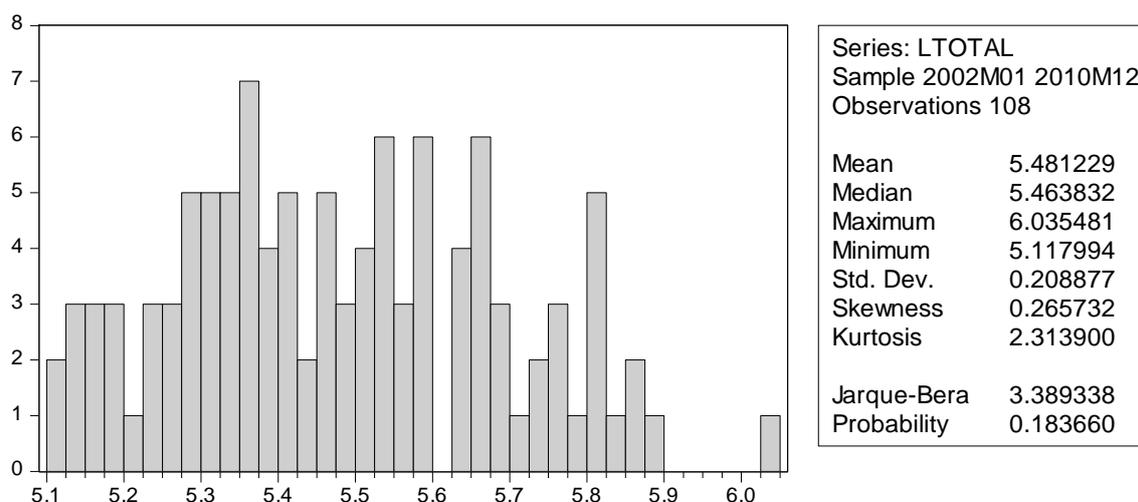


Gráfico 12 – Estatísticas descritivas – Total das Receitas (em logaritmo)



Pela análise gráfica desta série, pode-se notar que esta não exhibe degraus ou picos pronunciados, exceto aqueles ligados à sazonalidade. Optou-se, como de praxe, em selecionar o período mais longo dentre os aplicados para os demais grupos - 2002, tanto para manter semelhante o nível de informação das modelagens, como para evitar distorções na comparação das duas previsões do total arrecado (uma provém exclusivamente da série, outra é resultado da agregação das previsões de cada grupo). Pela análise descritiva, não houve rejeição da hipótese de normalidade dos dados ao nível de significância de 5%, embora ao nível de 10% a hipótese tenha sido rejeitada. O coeficiente de variação foi de 2,15%. De acordo com os testes formais, na data selecionada de corte – janeiro de 2002 – não foi rejeitada a hipótese de presença de quebra estrutural (Chow), enquanto o teste de Quandt-Andrews apontou fevereiro de 2003 como a data com probabilidade máxima de ocorrência de quebra, ambos com elevado nível de significância (p-valor = 0).

#### 4.7 Testes de Raiz Unitária

O procedimento para o teste ADF de raiz unitária, conforme explicitado em 2.1.7, seguiu a sugestão de Doldado et alli (1990). Os valores críticos utilizados foram extraídos de MacKinnon et alli (1999). As significância da tendência e da constante foram analisadas pelas estatísticas  $\Phi_1, \Phi_2, \Phi_3, \tau_{\beta\tau}$  e  $\tau_{\alpha\mu}$  (equação 13A). Neste teste, não se utiliza nenhum critério automático de comprimento de defasagem, sendo o mesmo realizado manualmente pela exclusão das defasagens não significativas. Também são apresentados os resultados dos testes Phillips-Perron (PP) e Kwiatkowski-Phillips-Schmidt-Shin (KPSS), efetuados a partir das definições padrão do software Eviews. Para ambos os testes, emprega-se como procedimento padrão a janela de Bartlett e o critério de Newey-West para delimitação da banda (Bueno, 2008). Pelas razões apresentadas em 2.1.7, na hipótese de conclusões contraditórias entre os testes, será dada preferência ao resultado encontrado no ADF. Adicionalmente, efetua-se o teste de Dickey-Pantula para verificar a presença de mais de uma raiz unitária.

Nas seis séries de tributos analisadas, os testes PP e KPSS são inconclusivos com relação à existência de raiz unitária, com exceção da série de Transações Financeiras (IOF), em que todas as especificações disponíveis indicam a aceitação de H0 (presença de raiz unitária) no teste PP ou rejeição de H0 (presença de estacionariedade) no teste KPSS. As células sombreadas na Tabela 1 indicam os modelos nos quais a hipótese de raiz unitária foi rejeitada.

Tabela 1 - Testes PP e KPSS de Raiz Unitária

	Período	p-valor					
		sem cte. e tend.		cte.		cte. e tend.	
		PP	KPSS	PP	KPSS	PP	KPSS
BS	2004m01 a 2009m12	0,89		0,15	0,00	0,07	0,42
Folha	2004m01 a 2009m12	0,99	-	0,03	0,00	0,00	0,78
Outros	2003m01 a 2009m12	0,57	-	0,00	0,18	0,00	0,52
Renda	2003m01 a 2009m12	0,80	-	0,00	0,00	0,00	0,19
TF	2002m01 a 2009m12	0,96	-	0,94	0,00	0,21	0,01
Total	2002m01 a 2009m12	0,90	-	0,03	0,00	0,00	0,38

H0 (PP): presença de raiz unitária

H0 (KPSS): a série é estacionária

Células sombreadas = rejeição da hipótese de raiz unitária

Aplicando o procedimento de Doldado et alli (1990) para o teste ADF, constata-se a presença de raiz unitária para todas as séries, ao nível de significância de 5%. A tabela 2 exibe uma síntese das estatísticas utilizadas no procedimento. As células tracejadas indicam que não houve necessidade de se testar o regressor determinístico, pois o processo foi concluído antes disso. Os modelos selecionados e os resultados do teste de raiz unitária encontram-se na Tabela 3.

Tabela 2 - Testes para seleção dos componentes determinísticos no modelo ADF (procedimento de Doldado)

	Período		$\Phi_3$	$\Phi_2$	$\Phi_1$	$T_{\beta T}$	$T_{\alpha T}$
BS	2004m01 a 2009m12	Observado	2,42	3,60	2,79	2,20	2,34
		Crítico (5%)	6,73	5,13	4,86	2,77	2,51
Folha	2004m01 a 2009m12	Observado	7,03	4,78	0,11	2,22	0,48
		Crítico (5%)	6,73	5,13	4,86	2,74	2,48
Outros	2003m01 a 2009m12	Observado	0,09	2,47	3,66	-0,42	2,68
		Crítico (5%)	6,49	4,88	4,86	2,77	2,51
Renda	2003m01 a 2009m12	Observado	6,93	5,30	-	2,96	-
		Crítico (5%)	6,61	5,00	-	2,77	-
TF	2002m01 a 2009m12	Observado	1,45	1,46	-	2,83	-
		Crítico (5%)	1,32	1,32	-	2,77	-
Total	2002m01 a 2009m12	Observado	2,55	2,60	0,03	2,76	0,22
		Crítico (5%)	6,61	5,00	4,86	2,77	2,51

H0( $\Phi_3$ ): no modelo  $\Delta y_t = a_0 + \alpha y_{t-1} + a_2 t + e_t$ , testa-se a hipótese conjunta  $\alpha = a_2 = 0$

H0( $\Phi_2$ ): no modelo  $\Delta y_t = a_0 + \alpha y_{t-1} + a_2 t + e_t$ , testa-se a hipótese conjunta  $\alpha = a_2 = a_0 = 0$

H0( $\Phi_1$ ): no modelo  $\Delta y_t = a_0 + \alpha y_{t-1} + e_t$ , testa-se a hipótese conjunta  $\alpha = a_0 = 0$

H0( $T_{\beta T}$ ): no modelo  $\Delta y_t = a_0 + \alpha y_{t-1} + a_2 t + e_t$ , testa-se a hipótese  $a_2 = 0$  dado que  $\alpha = 0$

H0( $T_{\alpha T}$ ): no modelo  $\Delta y_t = a_0 + \alpha y_{t-1} + e_t$ , testa-se a hipótese  $a_0 = 0$  dado que  $\alpha = 0$

Tabela 3 - Teste ADF de Raiz Unitária (procedimento de Doldado) p-valor

	Período	Defasagens	Cte	Tend	p-valor
BS	2004m01 a 2009m12	0	N	N	0,88
Folha	2004m01 a 2009m12	14	N	N	1,00
Outros	2003m01 a 2009m12	3	S	N	0,08
Renda	2003m01 a 2009m12	6	S	S	0,09
TF	2002m01 a 2009m12	0	S	S	0,21
Total	2002m01 a 2009m12	12	N	N	0,94

H0 (ADF): presença de raiz unitária

O teste de Dickey-Pantula indicou que não há mais de uma raiz unitária nas séries analisadas.

Tabela 4 - Teste de Dickey-Pantula para múltiplas raízes unitárias p-valor

	Período	n=3	n=2	n=1
BS	2004m01 a 2009m12	0,00	0,00	0,44
Folha	2004m01 a 2009m12	0,00	0,00	0,48
Outros	2003m01 a 2009m12	0,00	0,00	0,63
Renda	2003m01 a 2009m12	0,00	0,00	0,40
TF	2002m01 a 2009m12	0,00	0,00	0,07
Total	2002m01 a 2009m12	0,00	0,00	0,28

H0: presença de n raízes unitárias

H1: presença de (n-1) raízes unitárias

Os mesmos testes apresentados nas Tabelas 1 a 4 foram aplicados nas demais séries que integrarão a modelagem VAR: PIB, IBC-Br, IVV, IPI, Papelão Ondulado, Selic, Juros das Operações de Crédito, Volume de Crédito, Rendimento Médio, Massa Salarial, Taxa de Desemprego e IBrX-100. Algumas delas são empregadas para períodos distintos, dependendo do grupo de tributos sujeito à previsão. Por exemplo, a série Selic integra tanto um dos modelos do grupo Tributos sobre Bens e Serviços, como um do Total das Receitas. Devido ao critério de se utilizar o período mais longo sem quebras estruturais, o primeiro grupo iniciou-se em 2004, enquanto o segundo em 2002. Desse modo, os testes de raiz unitária foram aplicados para cada uma das extensões empregadas. Em síntese, todas essas séries, incluindo as variações de período, apresentaram uma e apenas uma raiz unitária. Uma importante exceção foi a série do PIB, para qual foi rejeitada a hipótese de raiz unitária para todos os períodos analisados. Por esta razão, o PIB não foi incluído entre as variáveis componentes do VAR, pois optou-se em preservar a mesma ordem de integração para todas as variáveis. As demais medidas de atividade econômica – IBC-Br, IVV, IPI e Papelão Ondulado - mostraram-se adequadas.

## 5 Resultados

Neste capítulo, serão apresentados e comparados os resultados de ambos os métodos empregados neste trabalho. A acurácia preditiva é o principal objetivo dessas modelagens econométricas, e a medida mais apropriada na avaliação deste quesito é o RMSE, conforme visto na seção 2.4. Contudo, um baixo RMSE não garante que o modelo escolhido seja o melhor entre os disponíveis. Afinal, o que se pretende é realizar uma boa previsão para anos vindouros, cujas realizações ainda não são conhecidas. E pequenos erros de previsão para o ano de controle – 2010 – não garantem um desempenho semelhante para os anos seguintes. Dessa forma, as propriedades desejáveis dos resíduos para cada modelo, como normalidade, ausência de autocorrelação e homoscedasticidade, serão decisivos na avaliação de sua adequação, em conjunto aos critérios de seleção de AIC, SBC e HQ. Satisfeitos esses requisitos, passa-se então para a medição do RMSE. As tabelas apresentadas já supõem cumpridas as exigências de significância dos coeficientes estimados (ao nível de 10%) e de invertibilidade dos modelos (as estimações podem ser encontradas no Anexo C).

Conforme expresso no capítulo anterior, as *dummies-outliers* serão empregadas apenas em casos particulares do histórico de cada série (mudanças na legislação, receitas extraordinárias, picos de atividade etc.). Como se verá adiante, a própria agregação dos tributos por base de incidência reduziu a quantidade e a dimensão de *outliers* muito acentuados.

Vale ressaltar que, para as previsões efetuadas neste trabalho, foi empregado o método de previsão dinâmica. Este caracteriza-se por incluir valores previstos pelo próprio modelo nas variáveis defasadas (exceto, é claro, para os meses não sujeitos à previsão, isto é, anteriores a dezembro de 2009), enquanto na versão estática as variáveis defasadas do modelo são preenchidas com os valores observados da série. Tal opção está em consonância com as previsões do método oficial aqui utilizadas, que são elaboradas no final de cada ano para o ano seguinte (no caso em dezembro de 2009 para os meses de 2010).

Para facilitar a comparação dos métodos utilizados, as duas modelagens serão apresentadas conjuntamente, para cada grupo investigado. As estimativas dos modelos com melhor capacidade preditiva, fornecidas pelo pacote econométrico Eviews, bem como os valores previstos para cada mês segundo cada metodologia empregada, encontram-se no Anexo C – Tabelas de 40 a 69.

### 5.1 Tributos sobre a Renda

**ARIMA** - A não estacionariedade da série fica evidente pelo exame gráfico dos dados em logaritmo (Gráfico 1, Capítulo 4), os quais exibem pronunciada tendência ascendente no período considerado (jan/2003 a dez/2009). O correlograma da série em primeira diferença sugere a existência de um componente sazonal no modelo.

Efetuada-se a primeira diferença ( $d=1$ ) conjuntamente à diferença sazonal semestral ( $D=1$ ), tem-se que a função de autocorrelação (FAC) da série resultante é truncada na primeira defasagem, indicando a presença de um componente de média móvel de ordem 1. Por sua vez, a função de autocorrelação parcial (FACP) exibe truncagem na segunda defasagem, o que caracteriza um componente autorregressivo de segunda ordem. Observa-se também por estes correlogramas a persistência de padrão sazonal, notadamente nas defasagens 6 e 12, sugerindo a inclusão de termos autorregressivos e/ou de médias móveis sazonais. Procedeu-se então à estimação dos modelos SARIMA  $(p,d,q)(P,D,Q)_6$ , com  $p \leq 2$ ,  $d = 1$ ,  $q \leq 1$ ,  $P \leq 1$ ,  $D = 1$ ,  $Q \leq 1$ , sempre dentro do período selecionado de janeiro de 2002 a dezembro de 2009. O ano de 2010 foi retirado da amostra para efeitos de avaliação da previsão. A série assim modelada apresentou autocorrelações residuais em algumas defasagens. A análise dos resíduos indicou possíveis correções com a inclusão de *outliers* nas seguintes datas: 03/2003, 04/2003, 06/2003, 12/2003, 05/2004 e 12/2004. Isto feito, os resultados mostraram-se satisfatórios.

Tabela 5 - modelagem ARIMA da série Tributos sobre a Renda - período: 01/2003 a 12/2009

nº	Modelo	Resíduos			Ajuste			Previsão			
		Correl.**	Normal.***	Heteros.****	AIC	SBC	HQ	RMSE	MAE	MAPE (%)	Annual (%)
1	$(2,1,1)(1,1,1)_6 + 6 \text{ do}^*$	N	0.50	0.92	-1.55	-1.24	-1.43	51,46	45,81	12,37	7.74
2	$(2,1,1)(1,1,0)_6 + 6 \text{ do}^*$	5(0,032)	0.16	0.53	-1.56	-1.27	-1.44	46.31	41.98	11.40	5.66
3	$(2,1,0)(1,1,1)_6 + 6 \text{ do}^*$	5(0,042)	0.36	0.72	-1.55	-1.26	-1.44	48.67	43.24	11.70	6.44
<b>4</b>	<b><math>(2,1,0)(1,1,0)_6 + 6 \text{ do}^*</math></b>	<b>4(0,035)</b>	<b>0.07</b>	<b>0.20</b>	<b>-1.56</b>	<b>-1.30</b>	<b>-1.45</b>	<b>44.93</b>	<b>41.61</b>	<b>11.21</b>	<b>4.44</b>

\*do=dummies para outliers (03/2003, 04/2003, 06/2003, 12/2003, 05/2004, 12/2004)

\*\* Estatística Q para a presença de autocorrelação nos resíduos (em p-valor). H0: presença de autocorrelação. N significa probabilidade superior a 10% em todas autocorrelações;

ou defasagem na qual se observa probabilidade inferior a 10%.

\*\*\* Teste de Jarque-Bera para normalidade dos resíduos (em p-valor). H0: presença de normalidade

\*\*\*\* Teste de White com termos cruzados para heteroscedasticidade dos resíduos (em p-valor). H0: presença de homoscedasticidade

O exame dos resultados apontou o modelo nº 4 - SARIMA  $(2,1,0)(1,1,0)_6 + 6$  dos – como o mais adequado. O modelo nº 4 apresentou o menor RMSE, bem como os menores valores dos critérios de seleção.

**VAR/VEC** – Inicia-se com a seleção das variáveis a serem incluídas no sistema. Esta etapa é geralmente amparada por algum supor teórico. Dois conjuntos de modelos foram trabalhados. O primeiro é composto por apenas duas variáveis, a série Renda e uma *proxy* da atividade econômica, teoricamente o principal fator não estrutural de oscilação da arrecadação. Empregaram-se quatro variáveis dessa classe: duas pela ótica do produto: IBC-Br e PO (papelão ondulado); uma pela ótica da renda: Rendimento médio; e a taxa desemprego. O segundo conjunto adiciona uma terceira variável, com o objetivo de captar antecipadamente a atividade econômica. Integram este grupo o IBrX-100 ou a Selic, utilizados como antecedentes da atividade um trimestre e um semestre à frente, respectivamente.

Selecionadas as variáveis, passa-se então para os testes de ordem de integração. São os mesmos utilizados no Subcapítulo 4.6, e constataram a presença de uma e apenas

uma raiz unitária em todas as variáveis utilizadas neste trabalho, caracterizando-as como I(1).

Foram elaborados oito modelos, quatro deles com duas variáveis e quatro com três variáveis, como constam nas Tabelas 6 e 7 abaixo. A seguir, procura-se encontrar o número de defasagens adequados para cada modelo, bem como o eventual emprego de dummies sazonais ou dummies *outliers*. O número de defasagens selecionado, bem como os critérios utilizados encontram-se na Tabela 6.

Tabela 6 - modelagem VAR/VEC da série Tributos sobre a Renda - período: 01/2003 a 12/2009

nº	Variáveis	Modelo	nº de vetores de cointegração*	Defasagens**	Intercepto no VAR e no Vetor***	Tendência no Vetor
1	Renda x PO	VEC	1	5	S	S
2	Renda x Desemp	VAR em 1º diferença	0	12	N	N
3	Renda x Rendim	VEC	1	12	S	S
4	Renda x IBC	VEC	1	7	S	S
5	Renda x Rendim x Desemp	VAR em 1º diferença	0	2	N	N
6	Renda x PO x Selic	VAR em 1º diferença	0	2	N	N
7	Renda x Rendim x Selic	VAR em 1º diferença	0	2	N	N
8	Renda x PO x IBrX	VAR em 1º diferença	0	5	N	N

\* Teste de Johansen para detectar o número de vetores de cointegração. Resultados das estatísticas  $\lambda$ -traço e  $\lambda$ -max sempre coincidentes (quando isto não ocorreu, o modelo foi excluído). Número máximo de vetores de cointegração = nº de variáveis - 1.

\*\* Cinco critérios de informação foram empregados: LR (likelihood ratio test), FPE (final prediction error), AIC (Akaike), SC (Schwarz) e HQ (Hannan-Quinn).

A correta especificação deve gerar resíduos com distribuição ruído branco. Quando mais de um modelo foi selecionado, optou-se pelo mais parcimonioso.

Conforme orientação de Enders (p.396), os testes foram aplicados sobre os dados em nível (VAR tradicional). O número de defasagens foi, então, reduzido em uma unidade por conta da diferenciação dos dados.

\*\*\* A seleção dos componentes determinísticos no VEC seguiu a orientação do manual do Eviews, segundo o qual as séries que apresentarem tendência estocástica modelam-se melhor com intercepto no vetor de cointegração e no VAR; e as que apresentarem tendência determinística, com ambos os interceptos mais um componente de tendência linear no vetor de cointegração (em alguns modelos também foi incluída tendência no VAR). Na modelagem do VAR em 1ª diferença, não foram utilizados componentes determinísticos.

Tabela 7 - resíduos e previsão da modelagem VAR/VEC da série Tributos sobre a Renda - período: 01/2003 a 12/2009

nº	Variáveis	Resíduos				Previsão			
		Port.*	LM*	Normal.**	Heteros.***	RMSE	MAE	MAPE (%)	Annual (%)
1	Renda x PO	N	N	0.84	0.15	48.32	44.64	12.71	11.63
2	Renda x Desemp	6 (0,0234)	N	0.78	0.68	56.23	48.51	13.01	5.26
3	Renda x Rendim	13 a 24	18 (0,02)	0.79	0.43	37.62	31.45	8.47	3.35
4	Renda x IBC	8 a 16	N	0.81	0.50	46.08	41.84	12.11	10.85
5	Renda x Rendim x Desemp	N	N	0.73	0.63	34.73	25.31	7.62	4.62
6	Renda x PO x Selic	vários	7 (0,01)	0.17	0.70	71.14	67.47	18.80	17.58
7	Renda x Rendim x Selic	N	N	0.42	0.45	49.99	43.32	12.63	11.28
<b>8</b>	<b>Renda x PO x IBrX</b>	<b>N</b>	<b>N</b>	<b>0.18</b>	<b>0.50</b>	<b>27.02</b>	<b>20.28</b>	<b>6.19</b>	<b>4.88</b>

\* Estatísticas Portmanteau e LM para a presença de autocorrelação nos resíduos (em p-valor). H0: presença de autocorrelação. N significa probabilidade superior a 10%

em todas as autocorrelações; ou defasagem na qual se observa probabilidade inferior a 10%.

\*\* Teste de Jarque-Bera para normalidade dos resíduos - ortogonalização de Doornik-Hansen (em p-valor). H0: presença de normalidade

\*\*\* Teste de White sem termos cruzados para heteroscedasticidade dos resíduos (em p-valor). H0: presença de homoscedasticidade.

Todos os critérios de seleção apontam o modelo 8 – RendaxPOxIBrX, na representação VAR em 1º diferença – como o melhor entre os disponíveis. Os resíduos de sua modelagem não apresentarem nenhum indício de autocorrelação. Os testes de normalidade e homoscedasticidade foram positivos para estas características. Ademais, a principal medida de previsão, o RMSE, foi de 27,02, o menor entre os modelos analisados e 41% inferior ao apresentado pelo melhor modelo ARIMA. Também mostrou os menores valores para as demais medidas de previsão MAE, MAPE e Anual. Vale destacar que nesta última seu resultado foi de 4,88%, um acréscimo de 10% encontrado na modelagem ARIMA. Interessante notar a presença de cointegração entre a série Renda e todas as medidas de atividade econômica utilizadas, PO, Rendim e IBC-Br, confirmando a esperada relação estável de longo prazo entre a arrecadação e a atividade econômica. Porém, ao se introduzir as variáveis antecedentes da atividade, IBrX e Selic, a relação de cointegração se desfaz em todas as modelagens.

No estágio seguinte, como pode ser visto na tabela seguinte, aplica-se o teste de cointegração de Johansen para os modelos especificados, para que se possa determinar o tipo de modelagem a ser empregada: VAR em 1ª diferença, caso seja rejeitada a hipótese de cointegração; ou VEC, caso contrário. Na aplicação desse teste, cujo funcionamento encontra-se explicitado no Subcapítulo 2.2.7, duas estatísticas são avaliadas:  $\lambda_{\text{traço}}$  e  $\lambda_{\text{max}}$ . No presente trabalho, optou-se por descartar os modelos que não apresentaram resultados coincidentes para as duas estatísticas. Na Tabela 8, são exibidos os p-valores dessas estatísticas para o modelo com melhor ajuste, bem como as respectivas conclusões. Por fim, procede-se ao exame dos resíduos gerados por cada modelo, além da qualidade da previsões.

Tabela 8 - Teste de cointegração para as variáveis Renda x PO x IBrX

Teste de Cointegração:	HO	H1	p-valor		HO	H1	p-valor
	r=0	r>0	0.15		r=0	r=1	0.32
$\lambda$ traço	r?1	r>1	0.26	$\lambda$ máximo	r=1	r=2	0.25
	r?2	r>2	0.55		r=2	r=3	0.55
Conclusão: ausência de vetor de cointegração				Conclusão: ausência de vetor de cointegração			

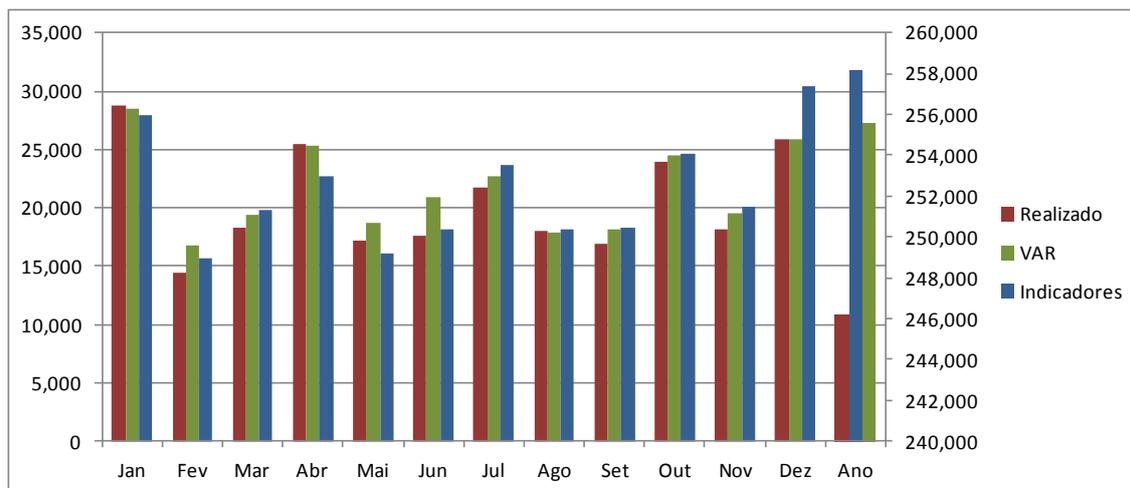
**Combinação das Previsões** – Pela análise dos valores previstos (Tabela 9), pode-se notar que, em sete dos doze meses estimados, a modelagem VAR/VEC forneceu estimações mais próximas dos dados observados. Além disso, em nove dos doze meses avaliados as previsões das duas modelagens (ARIMA e VAR/VEC) exibiram erros com sinais coincidentes, ou seja, para um dado mês, ambas foram simultaneamente maiores ou menores que o valor realizado. Não houve sinal de viés preponderante nas previsões econométricas, tendo em vista que em 5 dos 12 meses as estimações foram inferiores ao valor realizado, ocorrendo o inverso nos demais meses. Em razão desses fatos, era de se esperar que a combinação das previsões através da média aritmética exibisse uma acurácia de qualidade intermediária entre as já apresentadas, como consta na Tabela 5. Cabe destacar, a ligeira melhora na medida de previsão anual que, em valores absolutos, recuou 15% em relação ao VAR/VEC. Com relação à previsão oficial, o modelo selecionado conseguiu reduzir o RMSE em 25%, enquanto o erro de previsão anual sofreu um acréscimo de 28%.

Tabela 9 - Combinação das previsões dos modelos selecionados para Tributos sobre a Renda

Metodologia	Modelo	Previsão				Previsão Mensal		
		RMSE	MAE	MAPE (%)	Anual (%)	Nº de meses com a melhor previsão	Erro (%) máximo	Erro (%) mínimo
Indicadores	Indicadores	35.82	29.28	7.50	3.80	3	17,13	0.91
ARIMA	(2,1,0)(1,1,0) <sub>6</sub> + 6 do	44.93	41.61	11.21	4.44	0	24.27	1.98
<b>VAR</b>	<b>Renda x PO x IBrX</b>	<b>27.02</b>	<b>20.28</b>	<b>6.19</b>	<b>4.88</b>	<b>7</b>	<b>18.73</b>	<b>0.04</b>
Combinação Ótima*	Indicadores x ARIMA	30.70	27.80	7.59	4.12	2	12.78	0.17

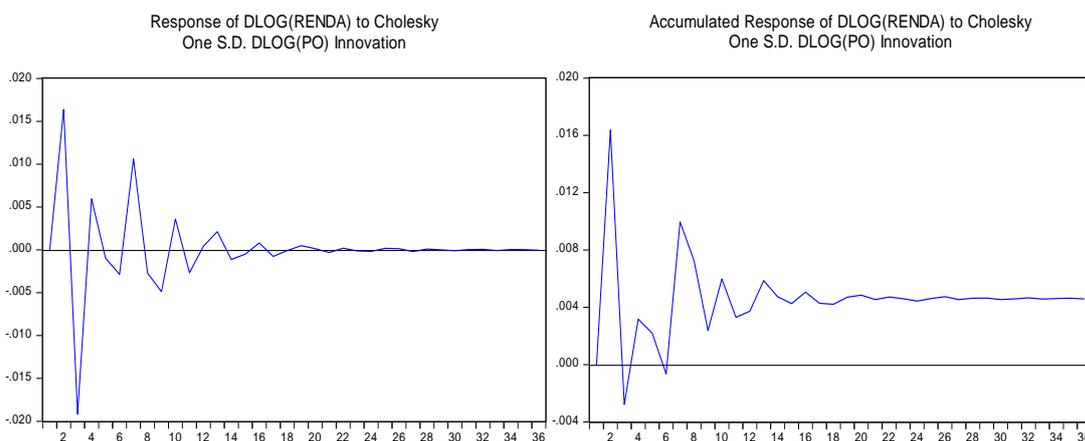
\* Corresponde à combinação com os melhores resultados entre as 6 possíveis (entre Indicadores, ARIMA e VAR)

Gráfico 13 – Comparação das melhores previsões para Tributos sobre a Renda com os valores observados e as previsões oficiais (a preços de dez/2009 – IPCA; em R\$ milhões)



**Função Impulso-Resposta** – O impacto da atividade econômica sobre a arrecadação desse grupo atinge a resposta de maior magnitude no momento imediatamente posterior ao choque (1,6%). A partir daí, as receitas passam a apresentar picos trimestrais, seguindo a própria sazonalidade da série (os dois maiores tributos que a compõem, em volume de recolhimento, são justamente o IRPJ e a CSLL, com prazos de recolhimento trimestral). Por sua vez, as respostas acumuladas exibem estabilidade a partir do ano seguinte, quando aproximadamente metade da magnitude do choque (0,46%) é absorvida de maneira definitiva pela arrecadação.

Gráfico 14 – Resposta pontual e acumulada de Tributos sobre a Renda a um choque de 1% na variável Papelão Ondulado (PO)



## 5.2 Tributos sobre a Folha de Salários

**ARIMA** – Tanto o aspecto gráfico da série (Gráfico 2, Capítulo 4), exibindo tendência crescente (principalmente no intervalo relevante de jan/2004 a dez/2009), como o decaimento suave da FAC dos dados em nível, indicam a presença de não

estacionariedade da série. Diferenciando-se a série, o novo correlograma exibe a FAC truncada na primeira defasagem, o mesmo ocorrendo na FACP, porém na segunda defasagem. Também se nota um forte padrão sazonal, caracterizado pelos picos na 12<sup>a</sup> defasagem e seus múltiplos, refletindo a concentração dos recolhimentos previdenciários em dezembro de cada ano. Após a dupla diferenciação ( $d=1$  e  $D=1$ ), a sazonalidade anual é consideravelmente suavizada, sem, contudo, desaparecer de todo, o que sugere a inclusão de componentes sazonais de média móvel ou autorregressivos. As truncagens permanecem as mesmas citadas anteriormente, indicando a presença de componentes de média móvel de primeira ordem e autorregressivos de segunda ordem. Testaram-se, então, variações do modelo SARIMA  $(p,d,q)(P,D,Q)_{12}$ , com  $p \leq 2$ ,  $d = 1$ ,  $q \leq 1$ ,  $P \leq 1$ ,  $D = 1$ ,  $Q \leq 1$ .

Tabela 10 - modelagem ARIMA da série Tributos sobre a Folha de Salários - período: 01/2004 a 12/2009

n°	Modelo	Resíduos			Ajuste			Previsão			
		Correl.**	Normal.***	Heteros.****	AIC	SBC	HQ	RMSE	MAE	MAPE (%)	Anual (%)
1	$(2,1,1)(1,1,0)_{12} + do^*$	N	0.97	0.19	-4.09	-3.93	-4.83	9.44	7.44	2.61	2.46
2	$(0,1,1)(1,1,0)_{12} + do^*$	N	0.28	0.89	-4.07	-3.98	-4.03	10.38	8.17	2.88	2.81
3	$(0,1,1)(0,1,1)_{12} + do^*$	N	0.97	0.02	-4.52	-4.42	-4.48	9.16	6.02	1.98	2.13
4	$(2,1,0)(0,1,1)_{12} + do^*$	4 a 15	0.94	0.10	-4.49	-4.37	4.44	11.83	8.13	2.69	2.87

\*do=dummy para outlier (11/2009)

\*\* Estatística Q para a presença de autocorrelação nos resíduos (em p-valor). H0: presença de autocorrelação. N significa probabilidade superior a 10% em todas autocorrelações;

ou defasagem na qual se observa probabilidade inferior a 10%.

\*\*\* Teste de Jarque-Bera para normalidade dos resíduos (em p-valor). H0: presença de normalidade

\*\*\*\* Teste de White com termos cruzados para heteroscedasticidade dos resíduos (em p-valor). H0: presença de homoscedasticidade

A aplicação do teste de Jarque-Bera nos modelos originalmente estimados sugeriu a rejeição da hipótese de normalidade dos resíduos. Analisando a distribuição destes, percebeu-se a necessidade de inclusão de uma dummy *outlier* para novembro de 2009. Com isso, a hipótese deixou de ser rejeitada. O modelo n°4, apesar dos bons valores de ajuste, exibiu diversas autocorrelações nos resíduos. Já o n°3 alcançou o menor RMSE, mas também apresentou uma importante deficiência: ausência de homoscedasticidade nos resíduos. O modelo n°1 – SARIMA  $(2,1,1)(1,1,0)_{12} + do$  – foi considerado o melhor.

**VAR/VEC** – A variável sujeita à predição é Folha, agregação das arrecadações da Previdência Social e do PIS/Pasep, no período de jan/2004 a dez/2009, sendo a qualidade da previsão avaliada sobre o ano de 2010. Foram elaborados sistemas de duas e três variáveis. No primeiro, buscou-se avaliar o desempenho das *proxies* de atividade econômica no desempenho da previsão de Folha, empregando-se as séries IPI, IBC-Br, IVV e Papelão Ondulado. A teoria econômica também aponta, além destas, a variável massa salarial (MS) como fator determinante da arrecadação previdenciária. Os sistemas trivariados dividem-se em dois grupos: o que insere antecedentes da atividade econômica um trimestre e um semestre à frente, respectivamente IBrX-100 e Selic; e o que insere conjuntamente as *proxies* do PIB e a variável MS. Os testes de raiz unitária

do Subcapítulo 4.6 indicaram a presença de uma e apenas uma raiz unitária nestas séries, permitindo que os testes de cointegração fossem conduzidos adequadamente.

Tabela 11 - modelagem VAR/VEC da série Tributos sobre a Folha de Salários - período: 01/2003 a 12/2009

nº	Variáveis	Modelo	nº de vetores de cointegração*	Defasagens**	Intercepto no VAR e no Vetor***	Tendência no Vetor
1	Folha x IPI	VEC	1	3	S	S
2	Folha x IVV	VAR em 1º diferença	0	6	N	N
3	Folha x IBC-Br	VAR em 1º diferença	0	11	N	N
4	Folha x PO	VAR em 1º diferença	0	2	N	N
5	Folha x MS	VEC	1	1	S	S
6	Folha x IPI x MS	VEC	1	3	S	S
7	Folha x IVV x MS	VEC	1	5	S	S
8	Folha x PO x MS	VEC	1	2	S	S
9	Folha x IPI x Selic	VEC	1	2	S	S + tend no VAR
10	Folha x PO x Selic	VEC	1	2	S	N
11	Folha x IPI x IBrX	VEC	1	5	S	S
12	Folha x PO x IBrX	VAR em 1º diferença	0	2	N	N

\* Teste de Johansen para detectar o número de vetores de cointegração. Resultados das estatísticas  $\lambda$ -traço e  $\lambda$ -max sempre coincidentes (quando isto não ocorreu, o modelo foi excluído). Número máximo de vetores de cointegração = nº de variáveis - 1.

\*\* Cinco critérios de informação foram empregados: LR (likelihood ratio test), FPE (final prediction error), AIC (Akaike), SC (Schwarz) e HQ (Hannan-Quinn).

A correta especificação deve gerar resíduos com distribuição ruído branco. Quando mais de um modelo foi selecionado, optou-se pelo mais parcimonioso.

Conforme orientação de Enders (p.396), os testes foram aplicados sobre os dados em nível (VAR tradicional). O número de defasagens foi, então, reduzido em uma unidade por conta da diferenciação dos dados.

\*\*\* A seleção dos componentes determinísticos no VEC seguiu a orientação do manual do Eviews, segundo o qual as séries que apresentarem tendência estocástica modelam-se melhor com intercepto no vetor de cointegração e no VAR; e as que apresentarem tendência determinística, com ambos os interceptos mais um componente de tendência linear no vetor de cointegração (em alguns modelos também foi incluída tendência no VAR). Na modelagem do VAR em 1ª diferença, não foram utilizados componentes determinísticos.

Tabela 12 - resíduos e previsão da modelagem VAR/VEC da série Tributos sobre a Folha de Salários - período: 01/2003 a 12/2009

nº	Variáveis	Resíduos				Previsão			
		Port.*	LM*	Normal.**	Heteros.***	RMSE	MAE	MAPE (%)	Anual (%)
1	Folha x IPI	7 a 13	1 (0,0125)	0.23	0.35	4.57	3.38	1.25	0.90
2	Folha x IVV	7 a 9	3 (0,016)	0.24	0.65	4.64	3.71	-1.02	1.27
3	Folha x IBC-Br	7 a 15	10 (0,003)	0.38	0.34	6.86	6.23	2.17	-1.43
4	Folha x PO	todos	2 e 7	0.29	0.12	5.00	4.09	1.54	1.39
5	Folha x MS	N	N	0.36	0.24	8.31	7.73	2.89	2.73
6	Folha x IPI x MS	4 (0,015)	N	0.53	0.48	4.75	3.30	1.10	-0.20
7	Folha x IVV x MS	6 a 13	N	0.99	0.54	7.73	6.48	2.26	-1.78
<b>8</b>	<b>Folha x PO x MS</b>	<b>N</b>	<b>N</b>	<b>0.17</b>	<b>0.55</b>	<b>4.05</b>	<b>3.12</b>	<b>1.15</b>	<b>0.78</b>
9	Folha x IPI x Selic	3 a 6	N	0.17	0.86	6.64	5.17	1.81	1.09
10	Folha x PO x Selic	3 a 6	12 (0,01)	0.55	0.75	6.96	6.19	2.24	1.73
11	Folha x IPI x IBrX	todos	N	0.52	0.48	3.09	2.66	0.93	-0.14
12	Folha x PO x IBrX	N	N	0.24	0.08	5.20	4.24	1.55	1.20

\* Estatísticas Portmanteau e LM para a presença de autocorrelação nos resíduos (em p-valor). H0: presença de autocorrelação. N significa probabilidade superior a 10%

em todas as autocorrelações; ou defasagem na qual se observa probabilidade inferior a 10%.

\*\* Teste de Jarque-Bera para normalidade dos resíduos - ortogonalização de Doornik-Hansen (em p-valor). H0: presença de normalidade

\*\*\* Teste de White sem termos cruzados para heteroscedasticidade dos resíduos (em p-valor). H0: presença de homoscedasticidade.

Em todos os modelos considerados (Tabela 12) não houve rejeição das hipóteses de homoscedasticidade e distribuição normal dos resíduos. Um modelo que mostrou boa capacidade preditiva foi o nº11, de variáveis cointegradas. Porém, a detecção de autocorrelações residuais mostrou resultados insatisfatórios, sendo verificada a presença dessa característica em todas as defasagens avaliadas. O modelo nº8 – VEC de Folha x PO x MS – foi selecionado como o mais adequado, com RMSE de 4,05, 57% inferior à encontrada na modelagem ARIMA, enquanto o erro de previsão anual foi reduzido em 68%, passando de 2,46% para 0,78%. Dessa forma, prevaleceu a interpretação tradicional de que a arrecadação previdenciária é majoritariamente influenciada pelo nível de atividade e pela massa salarial, a qual, na modelagem, mostrou melhores resultados do que o nível de atividade isoladamente, ou do que a combinação deste com algum antecedente.

Na tabela a seguir constam os valores assumidos pelas estatísticas pertinentes ao teste de cointegração.

Tabela 13 - Teste de cointegração de Johansen para as variáveis Folha x PO x MS

**Teste de Cointegração:**

	HO	H1	p-valor		HO	H1	p-valor
$\lambda$ traço	r=0	r>0	0,03	$\lambda$ máximo	r=0	r=1	0,01
	r≤1	r>1	0,80		r=1	r=2	0,72
	r≤2	r>2	0,92		r=2	r=3	0,92
Conclusão: presença de 1 vetor de cointegração				Conclusão: presença de 1 vetor de cointegração			

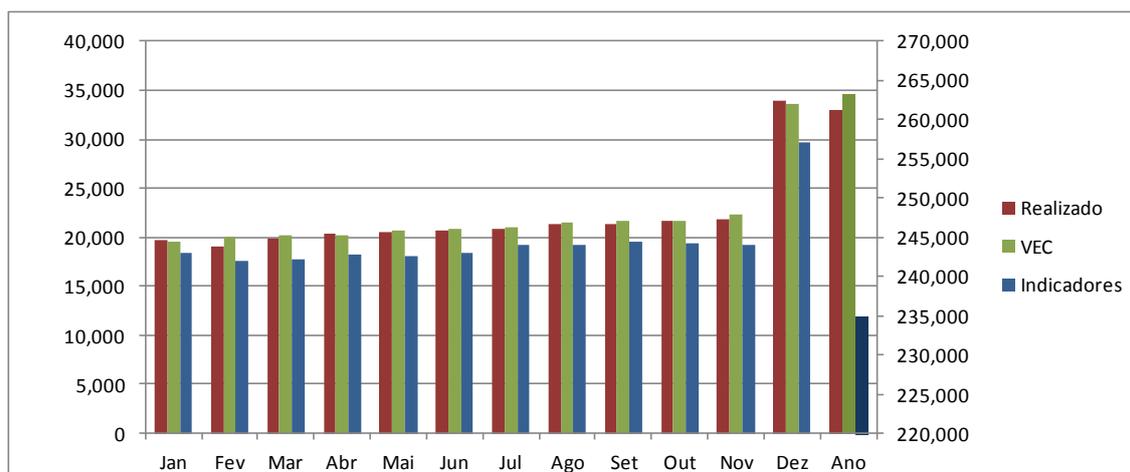
**Combinação das Previsões** – As medidas de erros de previsão percentual (MAPE) dos modelos econométricos giram em torno 1,5%. De fato, observando as tabelas mensais dos valores preditos (Anexo C), pode-se notar que o comportamento desta série permite modelagens muito precisas, pois os erros verificados para cada observação foram muito baixos. Também não houve superioridade de uma metodologia em relação à outra neste quesito, tendo em vista que aproximadamente um terço das realizações foram melhor previstas por cada uma delas. No agregado, contudo, a modelagem VAR/VEC apresentou melhores resultados. Tal fato pode ser explicado pela presença de um erro de aproximadamente 8% para o mês de novembro na modelagem ARIMA, o que também justifica o desempenho inferior da Combinação em relação ao VAR/VEC. Com relação ao viés das previsões, o Método dos Indicadores apresentou todos os resultados negativos, ou seja, inferiores ao observado, enquanto os métodos estatísticos (ARIMA e VAR) foram majoritariamente positivos. Ambos os modelos de séries temporais geraram uma drástica redução nos erros de previsão, quando comparados ao método dos Indicadores. No caso do VEC, a diminuição foi de 95% para a estimativa anual, e de 86% para o RMSE.

Tabela 14 - Combinação das previsões dos modelos selecionados para Tributos sobre a Folha de Salários

Metodologia	Modelo	Previsão Anual				Previsão Mensal		
		RMSE	MAE	MAPE (%)	Anual (%)	com a melhor previsão	Erro (%) máximo	Erro (%) mínimo
Indicadores	Indicadores	30.00	28.58	11.09	11.22	0	12.49	6.68
ARIMA	(0,1,1)(1,1,0) <sub>12</sub> + do	9.44	7.44	2.61	2.46	2	9.33	0.21
<b>VEC</b>	<b>Folha x PO x MS</b>	<b>4.05</b>	<b>3.12</b>	<b>1.15</b>	<b>0.78</b>	<b>9</b>	<b>4.35</b>	<b>0.22</b>
Combinação Ótima*	ARIMA x VEC	6.10	4.71	1.70	1.62	1	5.42	0.74

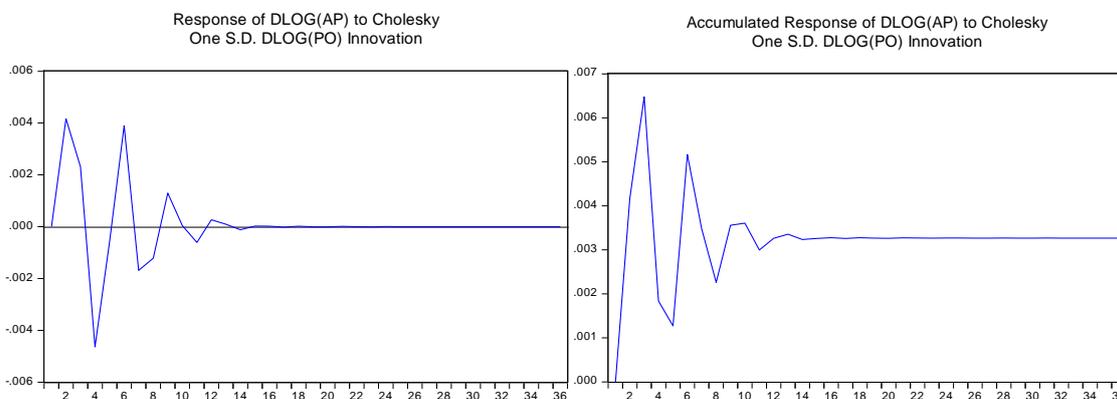
\* Corresponde à combinação com os melhores resultados entre as 6 possíveis (entre Indicadores, ARIMA e VAR)

Gráfico 15 – Comparação das melhores previsões para Tributos sobre a Folha de Salários com os valores observados e as previsões oficiais (a preços de dez/2009 – IPCA; em R\$ milhões)



**Função Impulso-Resposta** – Neste grupo, as respostas de maior magnitude ocorrem nos primeiros seis meses após o choque, em consonância com a periodicidade mensal dos recolhimentos previdenciários. Já no início do segundo semestre, as respostas são suavizadas, até desaparecerem de todo ao final de doze meses. No acumulado, o mesmo padrão se repete até se estabilizar numa variação de longo prazo de 3%.

Gráfico 16 – Resposta pontual e acumulada de Tributos sobre a Folha de Salários a um choque de 1% na variável Papelão Ondulado (PO)



### 5.3 Tributos sobre Bens e Serviços

**ARIMA** - A análise gráfica da série parece mostrar que a partir de 2004 houve uma suave elevação na média (Gráfico 3, Capítulo 4). Já a variância exibiu um crescimento mais facilmente notável ao longo do tempo. Em conjunto com o brando decaimento da FAC, evidencia-se a não estacionariedade da série. O correlograma da série em primeira diferença mostrou uma única truncagem na 12ª defasagem, tanto na FAC como na FACP, sugerindo a ausência de qualquer componente autorregressivo ou de média móvel, exceto o sazonal. A segunda diferenciação para sazonalidade não

gerou melhoras significativas nos picos observados. Dessa forma, optou-se por modelar a série com e sem a diferenciação sazonal. Os modelos foram estimados com base na especificação SARIMA  $(p,d,q)(P,D,Q)_s$ , com  $p=0$ ,  $d = 1$ ,  $q = 0$ ,  $P \leq 1$ ,  $D = 1$ ,  $Q \leq 1$  e  $s=12$  ou  $s=4$ .

Tabela 15 - modelagem ARIMA da série Tributos sobre Bens e Serviços - período: 01/2004 a 12/2009

nº	Modelo	Resíduos			Ajuste			Previsão			
		Correl.*	Normal.**	Heteros.***	AIC	SBC	HQ	RMSE	MAE	MAPE (%) Anual (%)	
1	(0,1,0)(1,1,1) <sub>4</sub>	N	0.00	0.97	-2.72	-2.66	-2.70	17.02	15.23	5.48	0.71
2	(0,1,0)(0,1,2) <sub>4</sub>	N	0.14	0.88	-2.76	-2.71	-2.74	19.61	16.75	5.80	4.41
3	(0,0,0)(1,0,2) <sub>4</sub>	N	0.00	0.76	-2.81	-2.72	-2.78	41.11	36.42	12.45	-12.79
<b>4</b>	<b>(0,1,0)(0,1,1)<sub>12</sub></b>	<b>N</b>	<b>0.95</b>	<b>0.37</b>	<b>-3.20</b>	<b>-3.17</b>	<b>-3.19</b>	<b>4.88</b>	<b>3.61</b>	<b>1.25</b>	<b>0.43</b>
5	(0,1,0)(1,1,0) <sub>12</sub>	N	0.67	0.31	-2.87	-2.84	-2.86	8.63	6.71	2.27	0.61
6	(0,1,0)(1,0,1) <sub>12</sub>	N	0.96	0.41	-3.18	-3.13	-3.16	7.86	6.49	2.34	1.00

\* Estatística Q para a presença de autocorrelação nos resíduos (em p-valor). H0: presença de autocorrelação. N significa probabilidade superior a 10% em todas autocorrelações;

ou defasagem na qual se observa probabilidade inferior a 10%.

\*\* Teste de Jarque-Bera para normalidade dos resíduos (em p-valor). H0: presença de normalidade

\*\*\* Teste de White com termos cruzados para heteroscedasticidade dos resíduos (em p-valor). H0: presença de homoscedasticidade

Os modelos nº1 e nº3 da Tabela 15 foram automaticamente descartados por não exibirem sinais de normalidade nos resíduos. O modelo nº4 – SARIMA  $(0,1,0)(0,1,1)_{12}$  foi eleito o melhor, levando-se em conta tanto os três critérios de ajuste adotados, bem como as medidas de comparação das previsões.

**VAR/VEC** – Realizaram-se previsões para todos os meses de 2010 da série BS, Tributos sobre Bens e Serviços, que corresponde à soma dos valores recolhidos dos tributos incidentes sobre esta base de arrecadação - IPI, Cofins e II. De um total de oito modelos estimados, metade compõe-se de sistemas bivariados que incluem alguma *proxy* do PIB (que, vale lembrar, não foi incluído por apresentar ordem de integração distinta das demais séries, de acordo com os testes de raiz unitária); na outra metade são selecionadas duas *proxies* que mostraram melhores resultados na modelagem com a série BS, e para cada uma delas são construídos sistemas de três variáveis que incluem os antecedentes do nível de atividade, Selic e IBrX-100.

Tabela 16 - Modelagem VAR/VEC da série Tributos sobre a Folha de Salários - período: 01/2003 a 12/2009

nº	Variáveis	Modelo	nº de vetores de cointegração*	Defasagens**	Intercepto no VAR e no Vetor***	Tendência no Vetor
1	BS x PO	VAR em 1ªdiferença	0	2	N	N
2	BS x IVV	VAR em 1ªdiferença	0	5	N	N
3	BS x IBC	VEC	1	1	S	S + tend no VAR
4	BS x IPI	VAR em 1ªdiferença	0	2	N	N
5	BS x IBC x Selic	VEC	1	1	S	S + tend no VAR
6	BS x IPI x Selic	VEC	2	2	S	S + tend no VAR
7	BS x IBC x IBrX	VEC	1	1	S	S + tend no VAR
8	BS x IPI x IBrX	VEC	1	1	S	S + tend no VAR

\* Teste de Johansen para detectar o número de vetores de cointegração. Resultados das estatísticas  $\lambda$ -traço e  $\lambda$ -max sempre coincidentes (quando isto não ocorreu, o modelo foi excluído). Número máximo de vetores de cointegração = nº de variáveis - 1.

\*\* Cinco critérios de informação foram empregados: LR (likelihood ratio test), FPE (final prediction error), AIC (Akaike), SC (Schwarz) e HQ (Hannan-Quinn).

A correta especificação deve gerar resíduos com distribuição ruído branco. Quando mais de um modelo foi selecionado, optou-se pelo mais parcimonioso.

Conforme orientação de Enders (p.396), os testes foram aplicados sobre os dados em nível (VAR tradicional). O número de defasagens foi, então, reduzido em uma unidade por conta da diferenciação dos dados.

\*\*\* A seleção dos componentes determinísticos no VEC seguiu a orientação do manual do Eviews, segundo o qual as séries que apresentem tendência estocástica modelam-se melhor com intercepto no vetor de cointegração e no VAR; e as que apresentem tendência determinística, com ambos os interceptos mais um componente de tendência linear no vetor de cointegração (em alguns modelos também foi incluída tendência no VAR). Na modelagem do VAR em 1ªdiferença, não foram utilizados componentes determinísticos.

Tabela 17 - Resíduos e previsão da modelagem VAR/VEC da série Tributos sobre a Folha de Salários - período: 01/2003 a 12/2009

nº	Variáveis	Resíduos				Previsão			
		Port.*	LM*	Normal.**	Heteros.***	RMSE	MAE	MAPE (%)	Anual (%)
1	BS x PO	N	10 e 17	0.62	0.44	13.96	12.76	4.56	4.48
2	BS x IVV	N	N	0.57	0.37	13.69	11.71	3.99	4.11
<b>3</b>	<b>BS x IBC</b>	<b>N</b>	<b>10 e 24</b>	<b>0.12</b>	<b>0.79</b>	<b>4.71</b>	<b>3.71</b>	<b>1.30</b>	<b>0.47</b>
4	BS x IPI	N	3 e 20	0.56	0.83	6.59	5.75	2.05	1.41
5	BS x IBC x Selic	N	2 (0,01)	0.77	0.86	8.22	6.88	2.35	-1.77
6	BS x IPI x Selic	vários	1 e 4	0.67	0.25	9.81	8.44	3.02	2.95
7	BS x IBC x IBrX	16 em diante	N	0.64	0.10	11.31	8.28	2.82	-0.85
8	BS x IPI x IBrX	N	1 (0,019)	0.24	0.06	13.15	11.68	4.19	4.10

\* Estatísticas Portmanteau e LM para a presença de autocorrelação nos resíduos (em p-valor). H0: presença de autocorrelação. N significa probabilidade superior a 10% em todas as autocorrelações; ou defasagem na qual se observa probabilidade inferior a 10%.

\*\* Teste de Jarque-Bera para normalidade dos resíduos - ortogonalização de Doornik-Hansen (em p-valor). H0: presença de normalidade

\*\*\* Teste de White sem termos cruzados para heteroscedasticidade dos resíduos (em p-valor). H0: presença de homoscedasticidade.

Como se pode observar na Tabela 17, as *proxies* de atividade econômica que exibiram a melhor capacidade preditiva foram IBC e IPI. Para cada uma delas aplicaram-se as variáveis antecedentes Selic e IBrX-100, obtendo-se quatro novos modelos. O modelo nº3 - BS x IBC, na representação VEC – foi considerado o melhor

entre os disponíveis, atingindo RMSE de 4,71, contra o valor de 6,59 do principal concorrente, o nº 4, que apresentou autocorrelações significativas em duas defasagens, além de resultados inferiores nos dois outros critérios de avaliação das previsões, MAE e MAPE. A tabela seguinte apresenta os resultados dos testes de cointegração:

Tabela 18 - Teste de cointegração para as variáveis BS x IBC

Teste de Cointegração:	HO	H1	p-valor		HO	H1	p-valor
$\lambda$ traço	r=0	r>0	0.00	$\lambda$ máximo	r=0	r=1	0.00
	r≤1	r>1	0.32		r=1	r=2	0.32
Conclusão: presença de 1 vetor de cointegração				Conclusão: presença de 1 vetor de cointegração			

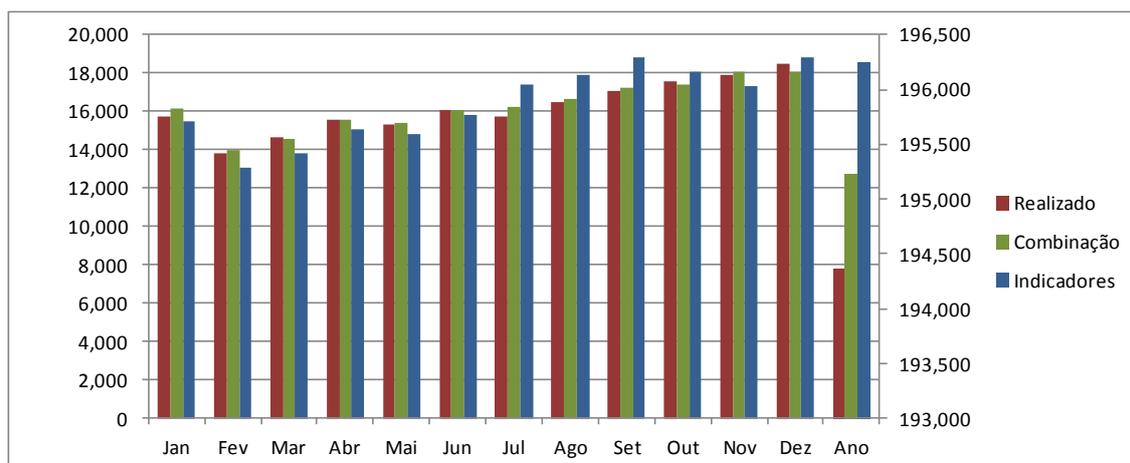
**Combinação das previsões** – neste caso, em metade dos meses as previsões de cada modelagem exibiram erros com sinais contrários (Anexo C), gerando uma melhora considerável na acurácia preditiva com a aplicação da média aritmética, além de evitar a existência de algum viés preponderante. O valor do RMSE foi reduzido em 3% e 6% em relação ao encontrado nas modelagens VEC e ARIMA, respectivamente. A medida dos erros percentuais (MAPE) mostrou uma queda mais suave, atingindo 0,8% de redução na comparação com a melhor modelagem (ARIMA). Analisando os valores preditos para o mês, contudo, o VEC superou os demais, obtendo melhores resultados em 8 meses, seguido pelo ARIMA, com 2, e a Combinação, com 1. Comparando ao método oficial, os valores preditos pela Combinação conseguiram reduzir em 72% o RMSE e em 53% o erro de previsão anual.

Tabela 19 - Combinação das previsões dos modelos selecionados para Tributos sobre Bens e Serviços

Metodologia	Modelo	Previsão Anual				Previsão Mensal		
		RMSE	MAE	MAPE (%)	Anual (%)	com a melhor previsão	Erro (%) máximo	Erro (%) mínimo
Indicadores	Indicadores	16.34	13.72	4.72	-0.96	0	10.62	1.22
ARIMA	(0,1,0)(0,1,1) <sub>12</sub>	4.88	3.61	1.25	0.43	2	9.33	0.21
VEC	BS x IBC	4.71	3.71	1.30	0.47	8	4.35	0.22
<b>Combinação Ótima*</b>	<b>ARIMA x VEC</b>	<b>4.58</b>	<b>3.57</b>	<b>1.24</b>	<b>0.45</b>	<b>1</b>	<b>5.42</b>	<b>0.27</b>

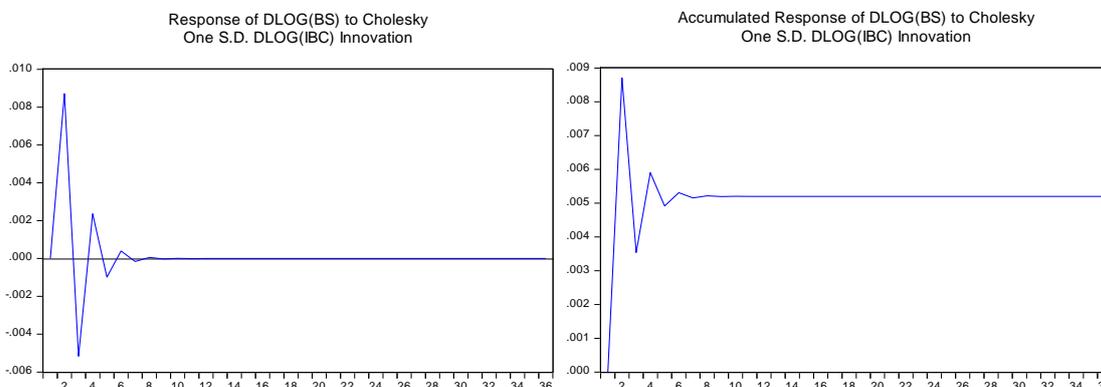
\* Corresponde à combinação com os melhores resultados entre as 6 possíveis (entre Indicadores, ARIMA e VAR)

Gráfico 17 – Comparação das melhores previsões para Tributos sobre Bens e Serviços com os valores observados e as previsões oficiais (a preços de dez/2009 – IPCA; em R\$ milhões)



**Função Impulso-Resposta** – a base de incidência dos tributos que compõem este grupo – circulação de bens e serviços – é a que apresenta maior velocidade de expansão com o incremento da atividade econômica. Por essa razão – e também devido à frequência mensal de recolhimento da Cofins (tributo que corresponde a cerca de 70% do total do grupo) – a resposta do volume de arrecadação foi intensa logo nos primeiros três meses após o choque, bem como efêmera, em comparação aos demais grupos, quase neutralizando-se ao final de seis meses.

Gráfico 18 – Resposta pontual e acumulada de Tributos sobre Bens e Serviços a um choque de 1% na variável IBC



## 5.4 Tributos sobre Transações Financeiras

**ARIMA** - Como se pode observar pelo Gráfico 4 do Capítulo 4, a série apresenta uma acentuada quebra de nível em janeiro de 2008, pelos motivos já expostos. Uma primeira tentativa de modelagem poderia ser a de aplicar uma dummy de nível para esta data e então avaliar a inclusão dos componentes autorregressivos e/ou de médias móveis. Contudo, a análise dos dados antes e depois da quebra mostrou a necessidade de diferenciação para que a estacionariedade fosse alcançada. Efetuando-se esse procedimento para o período completo da série (jan/2002 a dez/2009), a quebra de

nível ressurgiu como um *outlier*, que passa a ser modelado com outro tipo de dummy. O correlograma da série diferenciada indica a inclusão de componentes autorregressivos até a terceira ordem (truncagem na FACP na terceira defasagem) e de médias móveis de primeira ordem (truncagem na FAC na primeira defasagem). Não houve necessidade de diferenciação sazonal. Os modelos foram estimados com base na especificação SARIMA (p,d,q), com  $p \leq 3$ ,  $d = 1$ ,  $q \leq 1$ .

Tabela 20 - modelagem ARIMA da série Tributos sobre Transações Financeiras - período: 01/2002 a 12/2009

nº	Modelo	Resíduos			Ajuste			Previsão			
		Correl.**	Normal.***	Heteros.****	AIC	SBC	HQ	RMSE	MAE	MAPE (%) Anual (%)	
1	((1,3),1,0)	N	0.00	0.13	-1.44	-1.39	-1.42	41.53	39.21	18.11	15.84
2	(2,1,1)	N	0.00	0.01	-1.37	-1.29	-1.34	24.84	16.57	7.00	-2.08
<b>3</b>	<b>(2,1,0) + 4 dos*</b>	<b>3 (0,016)</b>	<b>0.92</b>	<b>0.58</b>	<b>-2.31</b>	<b>-2.15</b>	<b>-2.24</b>	<b>25.35</b>	<b>20.23</b>	<b>9.00</b>	<b>2.90</b>
4	(2,1,(3)) + 4 dos*	N	0.93	0.90	-2.31	-2.13	-2.24	25.42	20.47	9.13	3.16

\*do=dummies para outliers (01/2008, 02/2008, 01/09, 11/2009)

\*\* Estatística Q para a presença de autocorrelação nos resíduos (em p-valor). H0: presença de autocorrelação. N significa probabilidade superior a 10% em todas autocorrelações;

ou defasagem na qual se observa probabilidade inferior a 10%.

\*\*\* Teste de Jarque-Bera para normalidade dos resíduos (em p-valor). H0: presença de normalidade

\*\*\*\* Teste de White com termos cruzados para heteroscedasticidade dos resíduos (em p-valor). H0: presença de homoscedasticidade

O modelo nº3 da Tabela 20 – ARIMA (2,1,0) + 4 dos – foi selecionado como o melhor entre os disponíveis. Apresentou um ajuste superior, bem como menores erros de previsão, em média, que os obtidos pelo modelo nº4. Neste último foi possível corrigir a presença de autocorrelação observada na terceira defasagem do nº3, mas à custa de uma especificação degenerada pela inclusão de um componente isolado de médias móveis. Os demais modelos foram desconsiderados devido à rejeição das hipóteses de normalidade e homoscedasticidade.

**VAR/VEC** – A série sujeita à predição é a TF – Tributos sobre Transações Financeiras. A modelagem foi realizada sobre o período de janeiro de 2002 a dezembro de 2009, e as previsões avaliadas para o ano de 2010. Esse grupo de tributos é composto por apenas pela série do IOF. Como é sabido, a maior parte da arrecadação desse imposto provém das operações de crédito, que por sua vez estão ligadas à taxa média de juros em vigor no mercado e ao nível da atividade econômica. Por esta razão, foram construídos modelos que procuram incorporar esses fatores em sua variáveis. Do total de seis estimados, dois são bivariados e associam TF com Cred e PO. Os demais adicionam uma terceira variável que possa compensar a redução nos graus de liberdade com uma melhora na capacidade preditiva do modelo. São elas: Juros, Selic e IBrX-100.

Tabela 21 - Modelagem VAR/VEC da série Tributos sobre Transações Financeiras - período: 01/2002 a 12/2009

nº	Variáveis	Modelo	nº de vetores de cointegração*	Defasagens**	Intercepto no VAR e no Vetor***	Tendência no Vetor
1	TF x Cred	VEC	1	3	S	S + tend no VAR
2	TF x PO	VAR em 1ª diferença	0	5	N	N
3	TF x PO x Juros	VAR em 1ª diferença	0	3	N	N
4	TF x Cred x Juros	VEC	1	3	S	S
5	TF x PO x Selic	VEC	1	6	S	S + tend no VAR
6	TF x PO x IBrX	VEC	1	3	S	S + tend no VAR

\* Teste de Johansen para detectar o número de vetores de cointegração. Resultados das estatísticas  $\lambda$ -traço e  $\lambda$ -max sempre coincidentes (quando isto não ocorreu, o modelo foi excluído). Número máximo de vetores de cointegração = nº de variáveis - 1.

\*\* Cinco critérios de informação foram empregados: LR (likelihood ratio test), FPE (final prediction error), AIC (Akaike), SC (Schwarz) e HQ (Hannan-Quinn).

A correta especificação deve gerar resíduos com distribuição ruído branco. Quando mais de um modelo foi selecionado, optou-se pelo mais parcimonioso.

Conforme orientação de Enders (p.396), os testes foram aplicados sobre os dados em nível (VAR tradicional). O número de defasagens foi, então, reduzido em uma unidade por conta da diferenciação dos dados.

\*\*\* A seleção dos componentes determinísticos no VEC seguiu a orientação do manual do Eviews, segundo o qual as séries que apresentem tendência estocástica modelam-se melhor com intercepto no vetor de cointegração e no VAR; e as que apresentem tendência determinística, com ambos os interceptos mais um componente de tendência linear no vetor de cointegração (em alguns modelos também foi incluída tendência no VAR). Na modelagem do VAR em 1ª diferença, não foram utilizados componentes determinísticos.

Tabela 22 - Resíduos e erros de previsão da modelagem VAR/VEC da série Tributos sobre Transações Financeiras - período: 01/2002 a 12/2009

nº	Variáveis	Resíduos				Previsão			
		Port.*	LM*	Normal.**	Heteros.***	RMSE	MAE	MAPE (%)	Anual (%)
1	TF x Cred	todos	vários	0.69	0.31	34.68	32.84	14.90	12.23
2	<b>TF x PO</b>	<b>N</b>	<b>1 (0,04)</b>	<b>0.24</b>	<b>0.28</b>	<b>19.79</b>	<b>14.61</b>	<b>6.38</b>	<b>2.52</b>
3	TF x PO x Juros	4 e 5	1 (0,05)	0.11	0.13	21.44	13.13	5.52	-2.35
4	TF x Cred x Juros	N	N	0.11	0.93	30.78	29.51	13.32	10.62
5	TF x PO x Selic	todos	1 e 12	0.92	0.42	39.27	29.32	12.30	-7.19
6	TF x PO x IBrX	vários	1 (0,012)	0.43	0.77	21.70	18.64	8.28	5.07

\* Estatísticas Portmanteau e LM para a presença de autocorrelação nos resíduos (em p-valor). H0: presença de autocorrelação. N significa probabilidade superior a 10% em todas as autocorrelações; ou defasagem na qual se observa probabilidade inferior a 10%.

\*\* Teste de Jarque-Bera para normalidade dos resíduos - ortogonalização de Doornik-Hansen (em p-valor). H0: presença de normalidade

\*\*\* Teste de White sem termos cruzados para heteroscedasticidade dos resíduos (em p-valor). H0: presença de homoscedasticidade.

A inclusão das variáveis financeiras nos modelos com 3 variáveis não gerou melhorias significativas na qualidade de previsão. Exceção importante foi o modelo nº3 – TF x PO x Juros – que apresentou as previsões mais acuradas entre os modelos trivariados. Contudo, a retirada da variável Juros trouxe melhorias consideráveis a este modelo, eliminando as autocorrelações existentes nas defasagens 4 e 5, e ainda reduzindo o RMSE em 8%. Comparando com a modelagem ARIMA, sua superioridade foi ainda mais evidente, gerando uma redução de 22% no RMSE e de 13% na previsão anual.

A tabela a seguir mostra os resultados dos testes de cointegração para o modelo selecionado:

Tabela 23 - Teste de cointegração para as variáveis TF x PO

Teste de Cointegração:	HO	H1	p-valor		HO	H1	p-valor
$\Delta$ <i>traço</i>	r=0	r>0	0.45	$\Delta$ <i>máximo</i>	r=0	r=1	0.37
	r≤1	r>1	0.82		r=1	r=2	0.82
Conclusão: ausência de cointegração				Conclusão: ausência de cointegração			

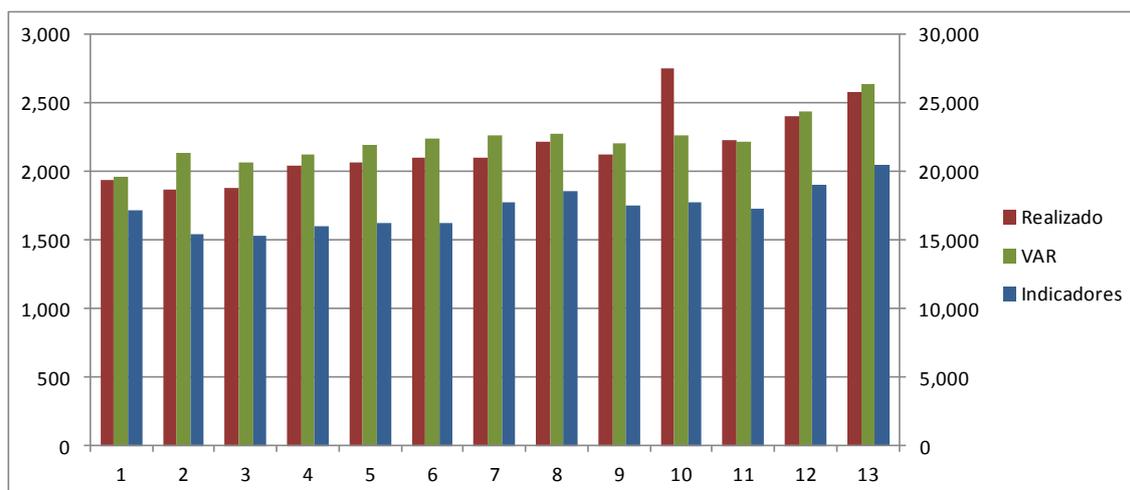
**Combinação das Previsões** – As previsões apresentaram viés negativo nos Indicadores (inferiores ao observado), positivo no VAR e sem viés no ARIMA. A RMSE da Combinação foi de 21,83, um acréscimo de 10% na comparação com a melhor modelagem (VEC). A média percentual (MAPE) do modelo VEC também mostrou melhor desempenho, atingindo 6,38%, uma queda de 14% em relação à Combinação. As previsões apresentaram viés negativo nos Indicadores, positivo no VAR e sem viés no ARIMA. Além disso, também superou os demais ao fornecer melhores previsões em oito dos meses estimados. Por fim, gerou uma acentuada melhora da qualidade preditiva em relação aos Indicadores: o erro de previsão para o acumulado no ano caiu 90%; já para o RMSE, a redução foi de 60%.

Tabela 24 - Combinação das previsões dos modelos selecionados para Tributos sobre Transações Financeiras

Metodologia	Modelo	Previsão Anual				Previsão Mensal		
		RMSE	MAE	MAPE (%)	Anual (%)	com a melhor previsão	Erro (%) máximo	Erro (%) mínimo
Indicadores	Indicadores	49.83	46.16	25.82	25.88	0	35.37	11.49
ARIMA	(2,1,0) + 4 dos	25.35	20.23	9.00	2.90	4	19.48	0.54
<b>VAR</b>	<b>TF x PO</b>	<b>19.79</b>	<b>14.61</b>	<b>6.38</b>	<b>2.52</b>	<b>8</b>	<b>17.87</b>	<b>0.99</b>
Combinação Ótima*	ARIMA x VEC	21.83	17.00	7.52	2.71	0	18.67	1.05

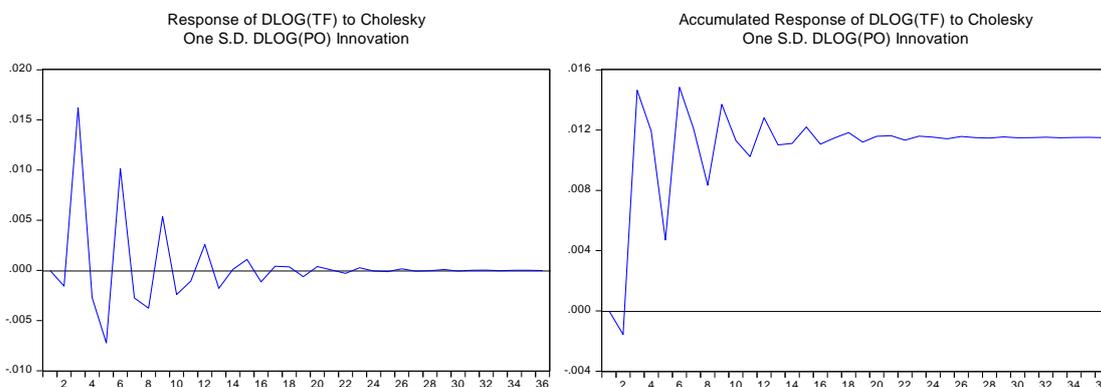
\* Corresponde à combinação com os melhores resultados entre as 6 possíveis (entre Indicadores, ARIMA e VAR)

Gráfico 19 – Comparação das melhores previsões para Tributos sobre Transações Financeiras com os valores observados e as previsões oficiais (a preços de dez/2009 – IPCA; em R\$ milhões)



**Função Impulso-Resposta** – o choque efetuado pela atividade econômica sobre o recolhimento do IOF (único componente deste grupo) atinge a resposta de maior magnitude dois períodos adiante para, em seguida, exibir um decaimento gradual com picos trimestrais até o término do segundo ano. Neste grupo, a resposta acumulada da arrecadação tende a estabilizar-se com um ano de atraso em relação aos demais grupos, e num patamar também mais elevado, de aproximadamente 1,2%.

Gráfico 20 – Resposta pontual e acumulada de Tributos sobre Transações Financeiras a um choque de 1% na variável Papelão Ondulado (PO)



## 5.5 Outros Tributos

**ARIMA** - No gráfico desta série para o período considerado (Gráfico 5, Capítulo 4, jan/2003 a dez/2009), os dados aparentam possuir um comportamento estacionário, oscilando em torno da média de 8,2, em logaritmo. O correlograma confirma em parte esta suspeita, pois também indica a presença de um padrão sazonal trimestral. Tanto a FAC como a FACP apresentam picos periódicos a cada três defasagens, sem nenhuma truncagem inicial. Dessa forma, optou-se por aplicar apenas a diferenciação sazonal, o que praticamente eliminou a sazonalidade da série, fazendo

surgir, por outro lado, uma truncagem na primeira defasagem da FAC e da FACP. A partir daí, efetuou-se a modelagem na representação SARIMA  $(p,d,q)(P,D,Q)_3$ , com  $p \leq 1$ ,  $d = 0$ ,  $q \leq 1$ ,  $P \leq 1$ ,  $D = 1$ ,  $Q \leq 1$ .

Tabela 25 - modelagem ARIMA da série Outros Tributos - período: 01/2003 a 12/2009

nº	Modelo	Resíduos			Ajuste			Previsão			
		Correl.*	Normal.**	Heteros.***	AIC	SBC	HQ	RMSE	MAE	MAPE (%)	Annual (%)
1	$(0,0,1)(0,1,1)_3$	N	0.00	0.00	0.18	0.24	0.20	130.02	99.66	21.12	-13.88
2	$(1,0,0)(1,1,0)_3$	N	0.00	0.00	0.20	0.26	0.23	529.40	438.38	96.68	81.27
3	$(1,0,0)(0,1,1)_3$	N	0.00	0.00	0.16	0.22	0.18	196.67	178.80	39.45	16.28
4	$(0,0,1)(1,1,0)_3$	3 (0,039)	0.00	0.00	0.26	0.31	0.28	364.90	371.78	68.51	49.11

\* Estatística Q para a presença de autocorrelação nos resíduos: N significa probabilidade inferior a 10% em todas autocorrelações

ou defasagem na qual se observa probabilidade inferior a 10%.

\*\* Teste de Jarque-Bera para normalidade dos resíduos. H0: presença de normalidade

\*\*\* Teste de White com termos cruzados para heteroscedasticidade dos resíduos. H0: presença de homoscedasticidade

Em todos os modelos da Tabela 25 houve rejeição das hipóteses de normalidade e de homoscedasticidade, mesmo após algumas tentativas de se incluir *outliers*, que apenas aumentaram os erros de previsão sem contribuir numa distribuição mais adequada dos resíduos. Dos modelos discriminados, o mais apropriado foi o nº 1 – SARIMA  $(0,0,1)(0,1,1)_3$ , que, apesar de possuir um ajuste ligeiramente inferior ao de seu principal concorrente, o nº3, mostrou desempenho superior a este nas medidas de acurácia preditiva.

**VAR/VEC** – a série Outros Tributos é composta por impostos e contribuições administrados pela RFB que não foram discriminados nos grupos acima, além dos tributos administrados pelos demais órgãos da administração federal. Na modelagem VAR desta série, procedeu-se à metodologia padrão de incluir, de uma lado, *proxies* da atividade econômica (sistemas bivariados) e, de outro, adicionar uma terceira variável de antecedente da atividade conjuntamente às duas *proxies* de melhor desempenho, totalizando oito modelos. O período utilizado para estimação foi de janeiro de 2003 a dezembro de 2009, sendo as previsões efetuadas para o ano de 2010.

Tabela 26 - Modelagem VAR/VEC da série Outro Tributos - período: 01/2003 a 12/2009

nº	Variáveis	Modelo	nº de vetores de cointegração*	Defasagens**	Intercepto no VAR e no Vetor***	Tendência no Vetor
1	Outros x IPI	VEC	1	12	S	S
2	Outros x IBC	VEC	1	3	S	S
3	Outros x PO	VAR em 1º diferença	0	5	N	N
4	Outros x IVV	VAR em 1º diferença	0	6	N	N
5	Outros x IBC x Selic	VAR em 1º diferença	0	2	N	N
6	Outros x IVV x Selic	VAR em 1º diferença	0	4	N	N
7	Outros x IBC x IBrX	VAR em 1º diferença	0	2	N	N
8	Outros x IVV x IBrX	VAR em 1º diferença	0	6	N	N

\* Teste de Johansen para detectar o número de vetores de cointegração. Resultados das estatísticas  $\lambda$ -traço e  $\lambda$ -max sempre coincidentes (quando isto não ocorreu, o modelo foi excluído). Número máximo de vetores de cointegração = nº de variáveis - 1.

\*\* Cinco critérios de informação foram empregados: LR (likelihood ratio test), FPE (final prediction error), AIC (Akaike), SC (Schwarz) e HQ (Hannan-Quinn).

A correta especificação deve gerar resíduos com distribuição ruído branco. Quando mais de um modelo foi selecionado, optou-se pelo mais parcimonioso.

Conforme orientação de Enders (p.396), os testes foram aplicados sobre os dados em nível (VAR tradicional). O número de defasagens foi, então, reduzido em uma unidade por conta da diferenciação dos dados.

\*\*\* A seleção dos componentes determinísticos no VEC seguiu a orientação do manual do Eviews, segundo o qual as séries que apresentem tendência estocástica modelam-se melhor com intercepto no vetor de cointegração e no VAR; e as que apresentem tendência determinística, com ambos os interceptos mais um componente de tendência linear no vetor de cointegração (em alguns modelos também foi incluída tendência no VAR). Na modelagem do VAR em 1ª diferença, não foram utilizados componentes determinísticos.

Tabela 27 - Resíduos e previsão da modelagem VAR/VEC da série Outro Tributos - período: 01/2003 a 12/2009

nº	Variáveis	Resíduos				Previsão			
		Port.*	LM*	Normal.**	Heteros.***	RMSE	MAE	MAPE (%)	Annual (%)
1	Outros x IPI	6 (0,032)	N	0.74	0.74	359.84	280.32	56.2	46.05
2	Outros x IBC	4 e 9	N	0.39	0.45	153.68	114.96	22.28	-15.84
3	Outros x PO	N	N	0.57	0.27	226.16	225.49	47.8	39.22
<b>4</b>	<b>Outros x IVV</b>	<b>N</b>	<b>N</b>	<b>0.32</b>	<b>0.67</b>	<b>116.93</b>	<b>98.33</b>	<b>16.92</b>	<b>-8.72</b>
5	Outros x IBC x Selic	vários	7 (0,04)	0.11	0.78	316.46	314.97	75.24	41.81
6	Outros x IVV x Selic	vários	3 (0,035)	0.3	0.5	230.16	202.57	42.72	37.32
7	Outros x IBC x IBrX	19 e 20	N	0.27	0.96	255.57	217.35	47.73	41.98
8	Outros x IVV x IBrX	todos	N	0.83	0.65	173.64	150.73	31.82	17.8

\* Estatísticas Portmanteau e LM para a presença de autocorrelação nos resíduos (em p-valor). H0: presença de autocorrelação. N significa probabilidade superior a 10% em todas as autocorrelações; ou defasagem na qual se observa probabilidade inferior a 10%.

\*\* Teste de Jarque-Bera para normalidade dos resíduos - ortogonalização de Doornik-Hansen (em p-valor). H0: presença de normalidade

\*\*\* Teste de White sem termos cruzados para heteroscedasticidade dos resíduos (em p-valor). H0: presença de homoscedasticidade.

A inclusão das variáveis antecedentes – Selic e IBrX – não gerou melhoras na capacidade preditiva dos modelos (Tabela 27). Além disso, também causou distorções na distribuição dos resíduos, o que se refletiu na presença de autocorrelações residuais em várias defasagens. Entre os modelos estimados, os melhores resultados foram

alcançados pelo nº 4 – especificação VAR em 1ª diferença das variáveis Outros x IVV. Neste, houve rejeição da hipótese de autocorrelação nos dois testes empregados, LM e Portmanteau. Em relação à modelagem ARIMA, o presente modelo reduziu a RMSE em 11%, enquanto os erros absolutos mantiveram-se praticamente inalterados (98,33 do VAR contra 99,66 do ARIMA). Já no acumulado do ano, o VAR foi 20% inferior.

A Tabela 28 apresente os testes de cointegração do modelo VAR selecionado:

Tabela 28 - Teste de cointegração de Johansen para as variáveis Outros x IVV

Teste de Cointegração:	HO	H1	p-valor		HO	H1	p-valor
	<u><math>\lambda</math> traço</u>	r=0	r>0		0.01	<u><math>\lambda</math> máximo</u>	r=0
	r?1	r>1	0.55		r=1	r=2	0.55
Conclusão: presença de 1 vetor de cointegração				Conclusão: presença de 1 vetor de cointegração			

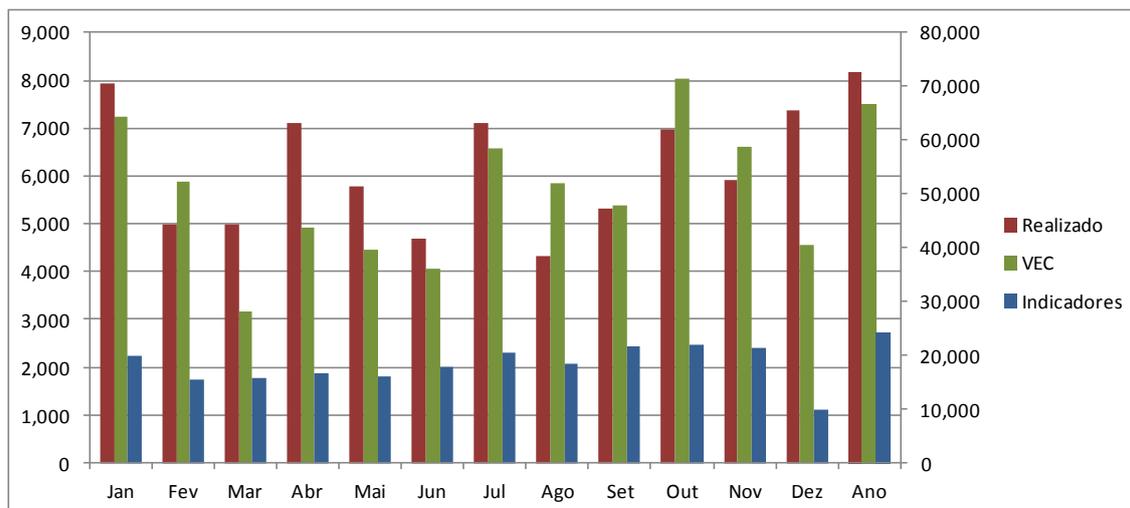
**Combinação das previsões** – Na RMSE, o VAR mostrou um desempenho superior aos demais, alcançando o valor de 116,93, o que corresponde a uma redução de 62% ao encontrado pela Combinação, e de 80% aos Indicadores. Característica dessa série é sua irregularidade, em razão da heterogeneidade dos tributos, o que se refletiu nos elevados erros percentuais das quatro modelagens, bem superiores aos apresentados nos grupos anteriores. Nos modelos econométricos, contudo, as previsões não exibiram qualquer sinal de viés, enquanto nos Indicadores os valores estimados foram, em sua totalidade, inferiores aos observados.

Tabela 29 - Combinação das previsões dos modelos selecionados para Outros Tributos

Metodologia	Modelo	Previsão Anual				Previsão Mensal		
		RMSE	MAE	MAPE (%)	Anual (%)	Nº de meses com a	Erro (%) máximo	Erro (%) mínimo
Indicadores	Indicadores	591.35	518.59	146.11	289.97	0	277.27	108.13
ARIMA	(0,0,1)(0,1,1) <sub>3</sub>	130.02	99.66	21.12	-54.13	6	55.83	3.54
<b>VEC</b>	<b>Outros x IVV</b>	<b>116.93</b>	<b>98.33</b>	<b>16.92</b>	<b>-8.72</b>	<b>6</b>	<b>36.62</b>	<b>1.07</b>
Combinação Ótima*	ARIMA x VEC	304.37	273.24	20.70	-54.21	0	36.39	5.89

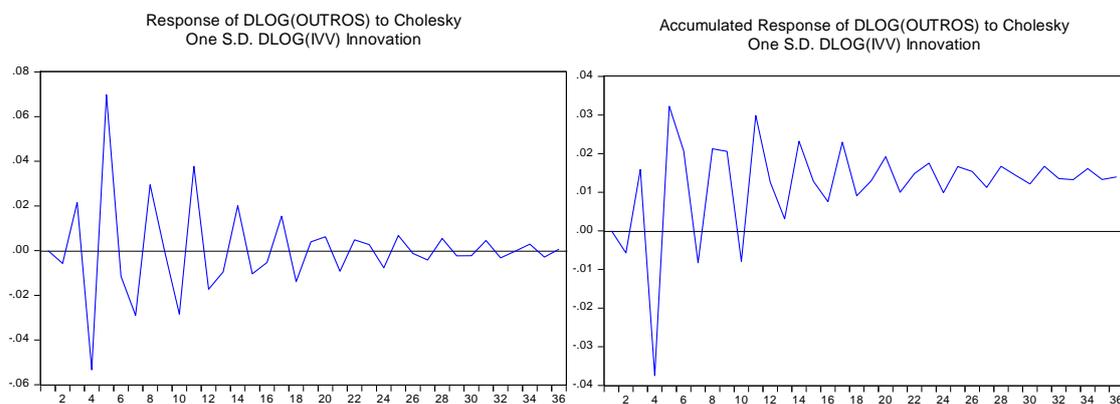
\* Corresponde à combinação com os melhores resultados entre as 6 possíveis (entre Indicadores, ARIMA e VAR)

Gráfico 21 – Comparação das melhores previsões para Outros Tributos com os valores observados e as previsões oficiais (a preços de dez/2009 – IPCA; em R\$ milhões)



**Função Impulso-Resposta** – Como nos demais grupos analisados, a magnitude das respostas ao impulso apresenta um decaimento gradativo ao longo do tempo. Contudo, particularmente neste caso, as respostas ainda exibem significância mesmo após passados os três anos compreendidos na análise. Tal fato pode ser explicado pela variedade de tributos que compõe esse grupo, cada qual reagindo de maneira muito diversa a um choque na atividade econômica. Mesmo na análise da resposta acumulada, o recolhimento não atinge um patamar estável no período.

Gráfico 22 – Resposta pontual e acumulada de Outros Tributos a um choque de 1% na variável Índice de Vendas no Varejo (IVV)



## 5.6 Total das Receitas

**ARIMA** - O gráfico da série (Gráfico 6, Capítulo 4) para o período considerado (jan/2002 a dez/2009) apresenta trajetória ascendente com picos sazonais. A ausência de estacionariedade é sugerida pelo correlograma, que mostra um lento decaimento na FAC. Com a primeira diferenciação, o novo correlograma conserva apenas o padrão sazonal observado no gráfico, que então é suavizado após a diferenciação sazonal, sem

contudo desaparecer de todo, sugerindo a inclusão de componentes sazonais autorregressivos e/ou de médias móveis. A FAC e a FACP da série assim transformada exibem truncagem na primeira e segunda defasagens, respectivamente, indicando a inserção de um componente de médias móveis de primeira ordem e dois componentes autorregressivos de segunda ordem. Testaram-se, então, modelos com a seguinte especificação:  $(p,d,q)(P,D,Q)_{12}$ , com  $p \leq 2$ ,  $d = 1$ ,  $q \leq 1$ ,  $P \leq 1$ ,  $D = 1$ ,  $Q \leq 1$ .

Tabela 30 - modelagem ARIMA da série Total das Receitas - período: 01/2002 a 12/2009

Modelo	Resíduos			Ajuste			Previsão			
	Correl.**	Normal.***	Heteros.****	AIC	SBC	HQ	RMSE	MAE	MAPE (%)	Annual (%)
1 (2,1,0)(0,1,0) <sub>12</sub>	11 a 24	0.00	0.77	-2.51	-2.46	-2.49	27.32	19.04	5.96	0.46
2 (2,1,0)(1,1,0) <sub>12</sub>	11 a 14	0.00	0.00	-2.59	-2.51	-2.55	22.76	16.08	5.11	3.84
3 (2,1,0)(0,1,1) <sub>12</sub> + 3 dos*	11 a 13	0.13	0.10	-2.93	-2.82	-2.89	25.45	13.98	4.38	1.03
<b>4 (2,1,0)(1,1,0)<sub>12</sub> + 3 dos*</b>	<b>11</b>	<b>0.54</b>	<b>0.13</b>	<b>-2.75</b>	<b>-2.62</b>	<b>-2.69</b>	<b>25.21</b>	<b>16.04</b>	<b>5.01</b>	<b>1.38</b>
5 (2,1,(11))(1,1,0) <sub>12</sub> + 3 dos*	N	0.49	0.32	-3.15	-3.00	-3.09	36.18	27.70	8.80	6.33
6 (2,1,(11))(0,1,1) <sub>12</sub> + 3 dos*	N	0.08	0.19	-3.40	-3.24	-3.33	26.37	19.38	5.93	-2.46

\*do=dummies para outliers (09/2002, 11/2009 e 12/2009)

\*\* Estatística Q para a presença de autocorrelação nos resíduos (em p-valor). H0: presença de autocorrelação. N significa probabilidade superior a 10% em todas autocorrelações;

ou defasagem na qual se observa probabilidade inferior a 10%.

\*\*\* Teste de Jarque-Bera para normalidade dos resíduos (em p-valor). H0: presença de normalidade

\*\*\*\* Teste de White com termos cruzados para heteroscedasticidade dos resíduos (em p-valor). H0: presença de homoscedasticidade

Nos dois primeiros modelos da Tabela 30, observa-se a presença de autocorrelação em algumas defasagens. A análise dos resíduos permitiu identificar a presença de alguns *outliers* que estavam gerando as autocorrelações, bem como distorcendo a distribuição dos resíduos, o que se refletiu na rejeição das hipóteses de normalidade e homoscedasticidade. Com a inclusão das dummies apropriadas, os problemas foram corrigidos, e a distribuição dos resíduos passou a exibir um comportamento mais adequado. O modelo nº4 – SARIMA (2,1,0)(1,1,0)<sub>12</sub> com dois *outliers* em 09/2002 e 11/2009 – foi selecionado como o melhor entre os disponíveis. Os modelos nº5 e nº6 apresentaram ajuste superior nos três critérios analisados, porém as médias dos erros de previsão foram maiores, o que não justificaria a presença de um termo degenerado de médias móveis na 11ª defasagem.

**VAR/VEC** – Esta série corresponde ao somatório de todos os demais grupos analisados e equivale ao total da arrecadação apresentado pela RFB em sua nota mensal. Sabe-se que os grupos de maior peso que a compõe são os Tributos sobre a Renda, os Tributos sobre a Folha de Salários e os Tributos sobre Bens e Serviços, cada um responsável por, respectivamente, 31,55%, 33,50% e 24,92% do total arrecadado. Visto que não há uma destacada predominância de qualquer um destes, optou-se pela inclusão, nos sistemas elaborados para a previsão da série Total, de algumas variáveis que foram utilizadas em suas modelagens. Dessa forma, nos sistemas de duas variáveis aparece, além das quatro *proxies* da atividade econômica, a variável de rendimentos médios nominais (Rendim), empregada originalmente na previsão dos Tributos sobre a Renda; nos sistemas trivariados, em conjunto às especificações do procedimento padrão (duas melhores *proxies* combinadas aos antecedentes de atividade Selic e IBRx-100), são

inseridos mais dois modelos contendo a variável de massa salarial (MS), extraída das modelagens dos Tributos sobre a Folha de Salários. Adicionalmente, estimou-se um modelo de quatro variáveis, onde se empregam séries advindas dos dois grupos já citados – MS e Desemp –, para verificar se a relação de longo prazo entre massa salarial e desemprego pode auxiliar na previsão da série Total. O período de estimação está compreendido entre janeiro de 2002 e dezembro de 2009, e as previsões efetuadas para 2010.

Tabela 31 - Modelagem VAR/VEC da série Total das Receitas - período: 01/2002 a 12/2009

nº	Variáveis	Modelo	nº de vetores de cointegração*	Defasagens**	Intercepto no VAR e no Vetor***	Tendência no Vetor
1	Total x Rendim	VAR em 1ª diferença	0	6	N	N
2	Total x IBC	VEC	1	11	S	S + tend no VAR
3	Total x IVV	VAR em 1ª diferença	0	6	N	N
4	Total x IPI	VAR em 1ª diferença	0	6	N	N
5	Total x PO	VAR em 1ª diferença	0	3	N	N
6	Total x MS x PO	VAR em 1ª diferença	0	2	N	N
7	Total x MS x IPI	VAR em 1ª diferença	0	3	N	N
8	Total x IPI x Selic	VEC	1	3	S	N
9	Total x PO x Selic	VAR em 1ª diferença	0	2	N	N
10	Total x IPI x IBrX	VEC	1	5	S	N
11	Total x PO x IBrX	VEC	1	2	S	S
12	Total x MS x IPI x Desemp	VAR em 1ª diferença	0	3	N	N

\* Teste de Johansen para detectar o número de vetores de cointegração. Resultados das estatísticas  $\lambda$ -traço e  $\lambda$ -max sempre coincidentes (quando isto não ocorreu, o modelo foi excluído). Número máximo de vetores de cointegração = nº de variáveis - 1.

\*\* Cinco critérios de informação foram empregados: LR (likelihood ratio test), FPE (final prediction error), AIC (Akaike), SC (Schwarz) e HQ (Hannan-Quinn).

A correta especificação deve gerar resíduos com distribuição ruído branco. Quando mais de um modelo foi selecionado, optou-se pelo mais parcimonioso.

Conforme orientação de Enders (p.396), os testes foram aplicados sobre os dados em nível (VAR tradicional). O número de defasagens foi, então, reduzido em uma unidade por conta da diferenciação dos dados.

\*\*\* A seleção dos componentes determinísticos no VEC seguiu a orientação do manual do Eviews, segundo o qual as séries que apresentarem tendência estocástica modelam-se melhor com intercepto no vetor de cointegração e no VAR; e as que apresentarem tendência determinística, com ambos os interceptos mais um componente de tendência linear no vetor de cointegração (em alguns modelos também foi incluída tendência no VAR). Na modelagem do VAR em 1ª diferença, não foram utilizados componentes determinísticos.

Tabela 32 - Resíduos e previsão da modelagem VAR/VEC da série Total das Receitas - período: 01/2002 a 12/2009

nº	Variáveis	Resíduos				Previsão			
		Port.*	LM*	Normal.**	Heteros.***	RMSE	MAE	MAPE (%)	Anual (%)
1	Total x Rendim	7 e 8	N	0.38	0.53	15.87	13.47	4.38	2.98
2	Total x IBC	13 a 24	N	0.70	0.30	94.46	87.74	27.58	27.73
3	Total x IVV	7 a 13	N	0.90	0.47	13.25	9.95	3.15	0.40
4	Total x IPI	7 a 9	1 (0,047)	0.61	0.78	17.05	14.80	4.93	3.60
5	Total x PO	N	N	0.54	0.27	29.61	28.11	9.30	8.88
6	Total x MS x PO	N	10 (0,04)	0.74	0.65	29.92	28.68	9.34	9.06
7	Total x MS x IPI	vários	14 e 15	0.14	0.34	21.28	19.73	6.55	5.79
8	Total x IPI x Selic	vários	1 e 11	0.07	0.08	24.38	22.24	7.35	7.03
9	Total x PO x Selic	vários	11 (0,017)	0.12	0.54	29.17	27.63	9.07	8.73
10	Total x IPI x IBrX	todos	N	0.16	0.67	13.90	11.53	3.85	2.50
<b>11</b>	<b>Total x PO x IBrX</b>	<b>N</b>	<b>N</b>	<b>0.41</b>	<b>0.58</b>	<b>15.49</b>	<b>10.07</b>	<b>3.10</b>	<b>-0.85</b>
12	Total x MS x IPI x Desemp	todos	12 (0,02)	0.55	0.32	24.77	21.44	7.23	5.98

\* Estatísticas Portmanteau e LM para a presença de autocorrelação nos resíduos (em p-valor). H0: presença de autocorrelação. N significa probabilidade superior a 10% em todas as autocorrelações; ou defasagem na qual se observa probabilidade inferior a 10%.

\*\* Teste de Jarque-Bera para normalidade dos resíduos - ortogonalização de Doornik-Hansen (em p-valor). H0: presença de normalidade

\*\*\* Teste de White sem termos cruzados para heteroscedasticidade dos resíduos (em p-valor). H0: presença de homoscedasticidade.

A série de papelão ondulado (PO), embora não apresentando os menores valores nos critérios de qualidade de seleção, foi a única *proxy* da atividade econômica cujos resíduos atenderam às propriedades estatísticas requeridas (ausência de autocorrelação e de heteroscedasticidade e distribuição normal). Com o emprego da terceira variável IBrX-100, antecedente para um trimestre à frente da atividade econômica, tais propriedades foram preservadas, além de ser obtida uma melhora considerável na acurácia preditiva, reduzindo a RMSE do modelo bivariado nº5 em 48%, e o acumulado do ano em 66%. Dessa forma, o modelo nº 11 – variáveis Total x PO x IBrX, na representação VEC – foi o melhor avaliado dentre os disponíveis. Vale destacar que a mesma série auxiliar (PO) também obteve resultados superiores na previsão de três outros grupos: Tributos sobre a Renda, onde o modelo selecionado continha as mesmas variáveis auxiliares que o do presente grupo; Tributos sobre a Folha de Salários, adicionando-se massa salarial (MS) à série PO; e Outro Tributos, apenas PO.

Na tabela seguinte, são exibidos os resultados dos testes de cointegração do modelo VEC com melhor capacidade preditiva.

Tabela 33 - Teste de cointegração de Johansen para as variáveis Total x PO x IBrX

Teste de Cointegração:	HO	H1	p-valor		HO	H1	p-valor
		r=0	r>0	0,00		r=0	r=1
$\lambda$ traço	r≤1	r>1	0,20	$\lambda$ máximo	r=1	r=2	0,21
	r≤2	r>2	0,44		r=2	r=3	0,44
Conclusão: presença de 1 vetor de cointegração				Conclusão: presença de 1 vetor de cointegração			

**Combinação das Previsões** – o segundo e terceiro modelos da Tabela 34 exibem os resultados da aplicação das metodologias ARIMA e VEC. Nota-se a expressiva superioridade deste último na previsão do total arrecadado, apresentando RMSE de 15,49, contra 25,21 do ARIMA, uma redução de 39%. Além disso, a média dos erros percentuais (MAPE) caiu quase pela metade, de 5,01% para 3,10%. O modelo seguinte, média dos anteriores, mostrou melhoras em relação ao VEC, quando avaliado pela RMSE, que foi reduzida de 15,49 para 13,92. Contudo, o VEC obteve melhor desempenho em todos os demais critérios, incluindo a previsão anual, reduzida em 60%. Com relação à metodologia oficial (Indicadores), este modelo reduziu em 89% o erro de previsão anual, enquanto a RMSE sofreu um decréscimo de 44%. Mesmo as previsões do VEC tendo sido, em sua maioria, ligeiramente superiores aos valores observados (contrapondo-se ao viés negativo dos Indicadores e à ausência de viés do ARIMA), a magnitude reduzida dos erros reforça a seleção dessa modelagem.

Tabela 34 - Combinação das previsões dos modelos selecionados para Total das Receitas

Metodologia	Modelo	Previsão Anual				Previsão Mensal		
		RMSE	MAE	MAPE (%)	Anual (%)	Nº de meses com a	Erro (%) máximo	Erro (%) mínimo
Indicadores	Indicadores	27.24	23.38	8.20	7.51	1	19,04	1,69
ARIMA	(2,1,0)(1,1,0) <sub>12</sub> + 3 dos	25.21	16.04	5.01	1.38	2	24.94	0.21
<b>VEC</b>	<b>Total x PO x IBrX</b>	<b>15.49</b>	<b>10.07</b>	<b>3.10</b>	<b>-0.85</b>	<b>5</b>	<b>13.08</b>	<b>0.14</b>
Combinação Ótima*	Indicadores x VEC	13.92	11.51	2.92	-2.15	4	5.87	0.62
Combinação	Soma Ótima**	15.96	13.80	4.62	3.83	0	11.42	1.33

\* Corresponde à combinação com os melhores resultados entre as 6 possíveis (entre Indicadores, ARIMA e VAR)

\*\* Corresponde à soma das melhores previsões entre as metodologias ARIMA, VAR e Média Aritmética (MA) para cada grupo de tributos. Dessa forma, foram selecionados:

Tributos sobre a Renda: VAR

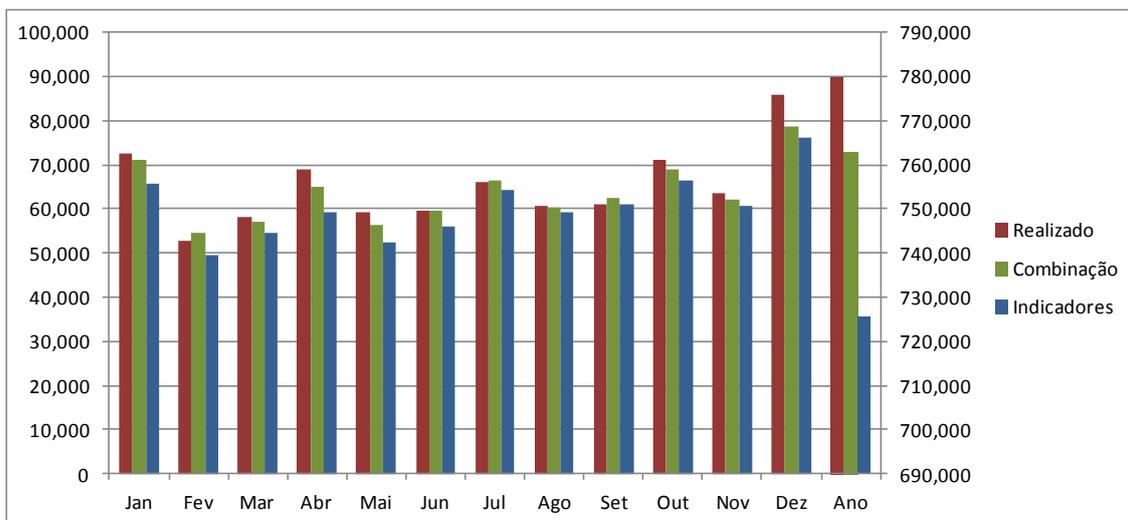
Tributos sobre a Folha de Salários: VAR

Tributos sobre Bens e Serviços: Combinação

Tributos sobre Transações Financeiras: VAR

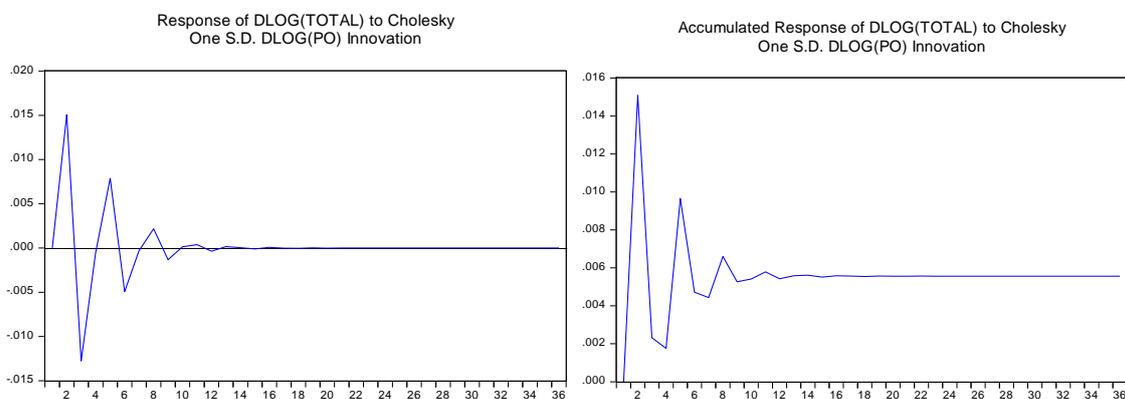
Outros Tributos: VAR

Gráfico 23 – Comparação das melhores previsões para Total das Receitas com os valores observados e as previsões oficiais (a preços de dez/2009 – IPCA; em R\$ milhões)



**Função Impulso-Resposta** – as reações de maior relevância concentram-se nos quatro primeiros meses após o choque, o que pode ser explicado pela preponderância do prazo de recolhimento mensal dentre os tributos de maior volume de arrecadação, como Previdência, Cofins e IRPF, os quais acabam se sobrepondo a importantes tributos sobre a renda de periodicidade trimestral, como IRPJ e CSLL. Passados esses meses, as respostas suavizam-se até estabilizarem-se totalmente a partir do 12º mês.

Gráfico 24 – Resposta pontual e acumulada de Total das Receitas a um choque de 1% na variável Papelão Ondulado (PO)







## 6 Conclusões

O presente trabalho vem confirmar, para o caso da receita tributária federal global, a constatação de estudos anteriores – Melo (2001), Siqueira (2002) e Campos(2009) – de que a metodologia oficialmente empregada pela RFB para a previsão da arrecadação, o chamado Método dos Indicadores, tem mostrado um desempenho significativamente inferior ao dos métodos econométricos de séries temporais, notadamente ARIMA e VAR. O fato de as previsões, nestes estudos, terem sido efetuadas para anos distintos (2000, 2001, 2007 e 2010) reforça ainda mais essa conclusão.

Nas seis séries de tributos analisadas neste estudo, a metodologia VAR/VEC mostrou desempenho superior em cinco casos, entre eles o Total das Receitas. Em seguida, veio a combinação dessa com ARIMA, no grupo Tributos sobre Bens e Serviços. O uso exclusivo da metodologia ARIMA ou o emprego do Método dos Indicadores não apresentou os melhores resultados em nenhum grupo.

No grupo Total das Receitas, foram avaliadas, além das modelagens ARIMA e VEC sobre a própria série, a combinação destas e a soma das melhores previsões efetuadas para cada grupo. De acordo com os critérios de seleção utilizados, sobressaiu-se a metodologia VEC. Sua RMSE foi 39% inferior à ARIMA e 44% inferior aos Indicadores. Além disso, mostrou drástica redução do erro de previsão anual, por volta de 40%, em relação à ARIMA e de 89%, em relação aos Indicadores. No sistema VAR bivariado, o melhor comportamento da distribuição residual foi observado com a *proxy* PO – papelão ondulado. A inclusão da terceira variável IBrX-100, antecedente da atividade econômica, mostrou bom desempenho, diminuindo a RMSE em 48%.

Nos Tributos sobre a Renda e nos Tributos sobre a Folha de Salários, a RMSE, em relação à ARIMA, foi reduzida em, respectivamente, 40% e 58%, além da drástica diminuição no erro de previsão anual, por volta de 70%, e no erro mensal máximo, de 50%. Já na comparação com o Método dos Indicadores, diminui-se a RMSE em 25% e 86% para os respectivos grupos. A combinação das metodologias pela média aritmética não gerou melhoras na qualidade da previsão. Em ambos os casos, a *proxy* do PIB que mostrou o melhor ajuste em termos da distribuição dos resíduos foi PO – emissão de papelão ondulado. Adicionando-se a esta a variável IBrX-100, antecedente da atividade econômica um trimestre à frente, a média dos erros de previsão nos Tributos sobre a Renda sofreu um decréscimo de aproximadamente 45%. Nos Tributos sobre a Folha de Salários, não houve melhoras na capacidade preditiva dos modelos após a inclusão das variáveis antecedentes. Mesmo assim, o sistema trivariado ainda mostrou desempenho superior ao bivariado, mas com a inserção da variável MS – massa salarial – no lugar da antecedente. Vale ressaltar que, juntos, esses dois grupos são responsáveis por aproximadamente 65% do total arrecadado.

Nos Tributos sobre Bens e Serviços, a combinação das metodologias pela média aritmética das previsões reduziu a RMSE do VAR, do ARIMA e dos Indicadores em, respectivamente, 4%, 6% e 63%. O erro de previsão anual foi mantido praticamente inalterado nas modelagens efetuadas (0,45%), porém 56% inferior ao dos Indicadores. A *proxy* do PIB que melhor ajustou-se à esse grupo foi o IBC-Br, índice de atividade econômica elaborado pelo Banco Central. A variável antecedente do PIB para um semestre à frente, Selic, não gerou expressiva melhora na capacidade preditiva do modelo, porém, com ela, os resíduos exibiram uma distribuição mais adequada.

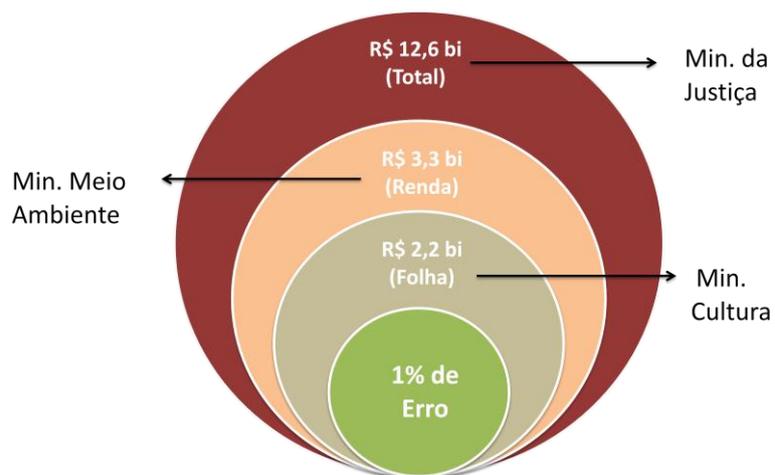
Nos Tributos sobre Transações Financeiras, a metodologia VAR/VEC reduziu a RMSE em 21%, em relação à ARIMA, em 10%, à Combinação e em 60%, aos Indicadores. Por sua vez, o erro de previsão anual sofreu uma queda de 13%, 7% e 90%, respectivamente. Neste grupo, os antecedentes do PIB não exibiram melhoras na acurácia preditiva dos modelos VAR. Enquanto a *Proxy* da atividade com melhor desempenho foi PO – papelão ondulado.

Nos Outros Tributos, o VEC mostrou RMSE menor que nas outras modelagens, em aproximadamente 15%, e em 80% quando comparada ao ARIMA e aos Indicadores, respectivamente. O erro de previsão anual queda acentuada, tanto em relação à ARIMA, 84%, quanto aos Indicadores, 97%. Neste grupo em particular, a série de volume de vendas no varejo, IVV, *proxy* do PIB, exibiu os melhores valores de ajuste no VAR. Por outro lado, a inclusão da terceira variável antecedente, tanto a Selic como a IBrX-100, gerou distorções na distribuição dos resíduos, indicando sua exclusão.

Uma primeira sugestão para trabalhos futuros poderia ser a de testar outras variáveis como antecedentes da atividade econômica, inclusive a série de previsão do PIB elaborada pelo relatório Focus do Banco Central. Cabe destacar que o nível de atividade constitui um importante fator na determinação do volume de arrecadação; portanto, uma previsão de qualidade dessa variável influirá positivamente no aperfeiçoamento dos modelos preditivos de receitas públicas.

Outra sugestão consiste no emprego de outros métodos de combinação de previsões. Entre algumas sugestões, destacam-se a de atribuir maior peso às previsões com melhor performance (Granger & Newbold, 1974) e a de aplicar métodos de regressão com restrição de peso (Diebold & Lopez, 1996).

Por fim, a principal recomendação que se depreende desse trabalho, para efeitos de política fiscal, é a de que a previsão das receitas públicas federais possa ser efetuada por outras metodologias além da oficialmente empregada. A fim de clarificar a importância da precisão dessas informações, foi elaborado o diagrama a seguir, onde são efetuadas aproximações do volume de recursos, em cada grupo de tributos, correspondente a um erro de 1% na estimativa das precisões, comparando-as ao orçamento de alguns ministérios para 2011:



Volume de arrecadação correspondente a 1% de erro na previsão

A qualidade das previsões constitui um instrumento basilar na elaboração do Orçamento da União e, desse modo, repercute no planejamento de todas as ações do governo. Nesse sentido, indica-se a adoção de métodos econométricos de previsão, notadamente o VAR, como já vem sendo realizada por alguns países desenvolvidos e de estrutura institucional mais madura.



## Bibliografia

Banco Central do Brasil. **A Defasagem da Transmissão da Política Monetária para Preços**. Relatório de Inflação, junho de 2009.

Banco Central do Brasil. **Indicadores Financeiros Antecedentes do PIB Brasileiro**. Relatório de Inflação, setembro de 2009.

Banco Central do Brasil. **Índice de Atividade Econômica do Banco Central (IBC-Br)**. Relatório de Inflação, março de 2010.

Box, G.E.P., & Jenkins, G.M. **Time Series Analysis, Forecasting and Control**. San Francisco: Holden Day, 1970.

Bueno, R.L.S. **Econometria de Séries Temporais**. Cengage Learning, 2008

Campos, C.V.C. **Previsão da Arrecadação de Receitas Federais: Aplicações de Modelos de Séries Temporais para o Estado de São Paulo**. Dissertação apresentada ao Programa de Pós-Graduação em Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo. Ribeirão Preto, 2009.

Dickey, D. & Pantula, S.G. Determining the Order of Differencing in Autoregressive Processes. **Journal of Business & Economics Statistics**, v.5, nº 5 – October, 1987.

Diebold, F.X. and Lopez J.A. (1996). “Forecast Evaluation and Combination”. **NBER Technical Working Paper**, 192.

Doldado, J.; Jenkinson, T. & Sosvilla-Rivero, S. Cointegration and Unit Roots. **Journal of Economic Surveys**. 4, 1990.

Duarte, A., Issler, J. e Spacov, A. **Indicadores Coincidentes de Atividade Econômica e uma Cronologia de Recessões para o Brasil**. EPGE-FGV, Ensaios Econômicos n.527, 2004.

Enders, W. **Applied econometric time series**. 1.ed. New Jersey: Wiley, 1995.

Engle, R.F. & Granger, C.W.J. Cointegration and Error Correction: Representation, Estimation and Testing. **Econometrica**, 55, 251-276.

Estrela, A. e Mishkin, F. (1997). The Predictive Power of the Term Structure of Interest Rates in Europe and the United States: Implications for the European Central Bank. **European Economic Review**, 41, 1375-1401, 1997.

Granger, C.W.J. and Newbold, P. (1974). “Experience with Forecasting Univariate Time Series and the Combination of Forecasts”. **Journal of the Royal Statistical Society**, Series A(General), v. 137, n. 2, p. 131-165.

Granger, C.W.J.; Newbold, P. **Forecasting Economic Time Series**. Orlando: Academic Press, 1977.

Guaragna, P.R.S.; Mello, M.R. **Um Modelo de Previsão para a Arrecadação do ICMS**. Brasília: ESAF, 2002, 46 p. Monografia premiada em 2º lugar no VII Prêmio Tesouro Nacional – 2002. Tributação, Orçamentos e Sistemas de Informação sobre Administração Financeira Pública. Disponível no sítio [http://www.tesouro.fazenda.gov.br/Premio\\_TN/VIIPremio/conteudo\\_mono\\_pr7\\_tema3.html](http://www.tesouro.fazenda.gov.br/Premio_TN/VIIPremio/conteudo_mono_pr7_tema3.html).

Hollauer, G. e Issler, J. **Construção de Indicadores Antecedentes para a Atividade Industrial Brasileira e Comparação de Metodologias**. IPEA, Texto para Discussão n. 1191, 2006.

Hollauer, G.; Issler, J.V.; Notini, H.H. Prevendo o Crescimento da Produção Industrial usando um Número Limitado de Combinações. **Economia Aplicada**, São Paulo, v.12, nº 2, p. 177-198, 2008.

Johansen, S. & Juselius, K. Maximum Likelihood Estimation and Inference on Cointegration with Application on the Demand for Money. **Oxford Bulletin of Economics and Statistics**, 52, 1990.

Johansen, S. Statistical Analysis of Cointegration Vector. **Journal of Economic Dynamics and Control**, v.12, 1998.

MacKinnon, J.G. Critical values for Co-integration tests. Em: Granger, C. & Engle, R. (org.). **Long-run Economic Relationships**. Oxford University Press, 1992.

Melo, B.S.V. **Modelo de Previsão para a Arrecadação Tributária**. Brasília: ESAF, 2001, 97p. Monografia vencedora em 1º lugar no VI Prêmio Tesouro Nacional – 2001. Orçamentos e Sistemas de Informação sobre a Administração Financeira Pública. Disponível no sítio [http://www.tesouro.fazenda.gov.br/Premio\\_TN/conteudo\\_mono\\_pr6tema3.html](http://www.tesouro.fazenda.gov.br/Premio_TN/conteudo_mono_pr6tema3.html).

Newbold, P. & Granger, C.W.J. Experience with forecasting univariate time series and the combination of forecasts. **Journal of the Royal Statistics Society**, 137, 131-146, 1974.

Pankratz, A. **Forecasting with Dynamic Regression Models**. New York: John Wiley & Sons Inc., 1991.

Peceguini, E.E. **Análise Comparativa de Métodos de Previsão Aplicados à Arrecadação do ICMS – Estado de São Paulo**. Brasília: ESAF, 2001, 47p. Monografia premiada em 2º lugar no VI Prêmio Tesouro Nacional – 2001. Orçamentos e Sistemas de Informação sobre a Administração Financeira Pública. Disponível no sítio [http://www.tesouro.fazenda.gov.br/Premio\\_TN/conteudo\\_mono\\_pr6\\_tema3.html](http://www.tesouro.fazenda.gov.br/Premio_TN/conteudo_mono_pr6_tema3.html).

Phillips, P. & Perron, P. Testing for a Unit Root in Time Series Regression. **Biometrika**, 75, 1988.

Receita Federal do Brasil. **Carga Tributária no Brasil 2009 – Análise por Tributos e Bases de Incidência**. Estudos Tributários nº 21, 2010.

Santos, C.H.; Ribeiro, M.B.; Gobetti, S.W. **A Evolução da Carga Tributária Bruta Brasileira no período 1995-2007: Tamanho, Composição e Especificações Econométricas Agregadas**. IPEA, Texto para Discussão nº 1350, 2008.

Sims, C. Macroeconomics and Reality. **Econometrica**, v.48, nº 1, p. 1-48, 1980.

Siqueira, M.L. **Melhorando a Previsão de Arrecadação Tributária Federal Através da Utilização de Modelos e Séries Temporais**. Brasília: ESAF, 2002. Monografia premiada em 1º lugar no VII Prêmio Tesouro Nacional – 2002. Tributação, Orçamentos e Sistemas de Informação sobre Administração Financeira Pública. Disponível no sítio [http://www.tesouro.fazenda.gov.br/Premio\\_TN/VIIIPremio/conteudo\\_mono\\_pr7.html](http://www.tesouro.fazenda.gov.br/Premio_TN/VIIIPremio/conteudo_mono_pr7.html).

Stock, J. e Watson, M. **Forecasting Output and Inflation: the Role of Asset Prices**. Working Paper, NBER n. 8180, 2001.

Ueda, E.M. **Carga Tributária Bruta no Brasil – 1995/2006 (MIMEO)**. Banco MIMÉO, 2007.

Ueda, E.M. **Termos de Troca e Arrecadação Pública (MIMEO)**. Banco Central do Brasil, 2009.

White, H. A Heteroskedastic-consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. **Econometrica**, Princeton, v.48, n.4, p.817-838, 1980.



## ANEXO A

Tabela 35 - Receita Tributária por Base de Incidência - Detalhe da Agregação – 2009  
(R\$ milhões)

<b>0000 Total da Receita Tributária.....</b>	<b>1.055.407,07</b>
<b>1000 Tributos sobre a Renda.....</b>	<b>209.779,26</b>
<b>1100 Pessoa Física.....</b>	<b>73.650,20</b>
1100.01 Imposto de Renda das Pessoas Físicas.....	13.913,31
1100.02 Imposto de Renda Retido na Fonte - Trabalho Assalariado.....	42.626,88
1100.03 Imposto de Renda Retido na Fonte - gov. estaduais, suas fundações e autarquias.....	12.461,16
1100.04 Imposto de Renda Retido na Fonte - gov. municipais, suas fundações e autarquias.....	4.648,86
<b>1200 Pessoa Jurídica.....</b>	<b>95.827,44</b>
1200.01 Imposto de Renda das Pessoas Jurídicas (regime do lucro real).....	62.169,57
1200.02 Contribuição Social sobre o Lucro Líquido (regime do lucro real).....	33.657,87
<b>1900 Outros.....</b>	<b>40.301,62</b>
<b>1910 Não residentes .....</b>	<b>10.896,40</b>
1910.02 Imposto de Renda Retido na Fonte - Não Residentes.....	10.896,40
<b>1920 Retenções não Alocáveis.....</b>	<b>29.405,22</b>
1920.01 Imposto de Renda Retido na Fonte - Capital.....	23.278,91
1920.02 Imposto de Renda Retido na Fonte - Outros.....	6.126,31
<b>2000 Tributos sobre a Folha de Salários.....</b>	<b>274.938,99</b>
<b>2100 Previdência Social.....</b>	<b>194.288,36</b>
<b>2110 Empregador.....</b>	<b>125.054,55</b>
2110.01 Contribuição para o INSS - Cota Patronal.....	100.312,33
2110.02 Contribuição para a Seguridade Social do Servidor Público - Parcela Governo.....	10.952,85
2110.03 Previdência dos Estados - Parcela Governo.....	11.049,95
2110.04 Previdência dos Municípios - Parcela Governo.....	2.739,43
<b>2120 Empregado.....</b>	<b>60.216,73</b>
2120.01 Contribuição para o INSS - Cota Empregado.....	37.812,85
2120.02 Contribuição para a Seguridade Social do Servidor Público - Parcela Servidor.....	7.557,99
2120.03 Previdência dos Estados - Parcela Servidor.....	6.077,47
2120.04 Previdência dos Municípios - Parcela Servidor.....	1.506,68
2120.05 SIMPLES - Recolhimento em GPS.....	7.261,74
<b>2130 Autônomo.....</b>	<b>6.950,16</b>
2130.01 Contribuição para o INSS - Autônomo /1.....	6.950,16
<b>2190 Outros.....</b>	<b>2.066,91</b>
2190.04 Reclamatória Trabalhista.....	1.898,92
2190.05 INSS - Outras Receitas.....	167,99
<b>2200 Seguro Desemprego.....</b>	<b>54.725,95</b>
2200.01 Contribuição para o Fundo de Garantia do Tempo de Serviço (FGTS).....	54.725,95
<b>2900 Outros.....</b>	<b>25.924,67</b>
2900.01 Salário Educação (via INSS e FNDE).....	9.685,19
2900.02 Contribuições para o Sistema S.....	8.609,23
2900.03 Contribuições para o PIS sobre a folha de pagamento.....	471,10
2900.04 Pasep.....	5.163,25
2900.05 Contribuição para o Custeio das Pensões Militares.....	1.681,26
2900.06 Cota-Parte Contrib. Sindical.....	314,63
<b>3000 Tributos sobre a Propriedade.....</b>	<b>35.573,49</b>
<b>3100 Propriedade Imobiliária (urbana e rural) .....</b>	<b>12.669,35</b>
3100.01 Imposto Territorial Rural (ITR).....	434,23
3100.02 Imposto sobre a Propriedade Territorial Urbana (IPTU).....	12.235,12
<b>3200 Propriedade de Veículos Automotores.....</b>	<b>17.567,21</b>
3200.01 Imposto sobre a Propriedade de Veículos Automotores (IPVA).....	17.567,21
<b>3300 Transferências Patrimoniais.....</b>	<b>5.336,93</b>
3300.01 Imposto sobre a Transmissão Causa Mortis e Doação (ITCD).....	1.590,35
3300.02 Imposto sobre a Transmissão de Bens Imóveis (ITBI).....	3.746,58

<b>4000 Tributos sobre Bens e Serviços.....</b>	<b>499.872,57</b>
<b>4100 Gerais .....</b>	<b>340.122,15</b>
<b>4110 Não Cumulativos.....</b>	<b>227.380,74</b>
4110.01 ICMS - Exceto Seletivos.....	110.793,01
4110.02 Imposto sobre Produtos Industrializados (IPI) - Exceto Seletivos.....	19.773,60
4110.03 Cofins-Não Cumulativa.....	79.304,43
4110.04 PIS-Não Cumulativo.....	17.509,70
<b>4120 Cumulativos.....</b>	<b>112.741,41</b>
4120.01 Cofins-Cumulativa.....	32.790,83
4120.02 PIS-Cumulativo.....	7.061,63
4120.03 Simples.....	12.811,88
4120.04 CSLL-Lucro Presumido.....	8.642,72
4120.05 IRPJ-Lucro Presumido.....	15.000,27
4120.06 Imposto sobre Serviços (ISS).....	22.354,48
4120.07 INSS - Clubes de Futebol.....	58,99
4120.08 INSS - Comercialização da Produção Rural.....	3.019,96
4120.10 Finsocial.....	15,23
4120.11 Contribuição sobre a Receita de Concursos e Prognósticos.....	2.497,24
4120.14 Simples Nacional.....	7.339,38
4120.15 CIDE Remessas.....	1.148,81
<b>4200 Seletivos.....</b>	<b>125.906,92</b>
<b>4210 Automóveis.....</b>	<b>20.666,17</b>
4210.01 IPI - Automóveis.....	2.110,09
4210.02 ICMS - Automóveis.....	18.556,09
<b>4220 Bebidas.....</b>	<b>10.633,43</b>
4220.01 IPI - Bebidas.....	2.318,06
4220.02 ICMS - Bebidas.....	8.315,37
<b>4230 Combustíveis.....</b>	<b>39.457,60</b>
4230.01 CIDE - Combustíveis.....	4.911,41
4230.02 ICMS - Combustíveis.....	34.546,19
<b>4240 Energia Elétrica.....</b>	<b>21.747,22</b>
4240.01 ICMS - Energia Elétrica.....	21.747,22
<b>4250 Tabaco.....</b>	<b>6.260,55</b>
4250.01 IPI - Tabaco.....	3.332,63
4250.02 ICMS - Tabaco.....	2.927,92
<b>4260 Telecomunicações.....</b>	<b>27.141,94</b>
4260.01 ICMS - Telecomunicações.....	27.141,94
<b>4300 Comércio exterior.....</b>	<b>15.895,41</b>
4300.01 Imposto sobre Importação.....	15.832,01
4300.02 Imposto sobre Exportação.....	63,40
<b>4400 Taxas - Prest.Serviços e Poder Polícia.....</b>	<b>16.213,42</b>
4400.01 Taxas Federais.....	4.989,17
4400.02 Taxas Estaduais.....	7.938,36
4400.03 Taxas Municipais.....	3.285,89
<b>4500 Contribuições Econômicas.....</b>	<b>1.734,67</b>
4500.01 FUNDAF.....	179,82
4500.02 CONDECINE.....	44,15
4500.03 AFRMM .....	1.510,71
<b>5000 Tributos sobre Transações Financeira.....</b>	<b>19.200,60</b>
<b>5100 Tributos sobre Débitos e Créditos Bancários .....</b>	<b>(24,14)</b>
5100.01 Contribuição Provisória sobre Movimentação Financeira (CPMF).....	(24,25)
5100.02 Imposto Provisório sobre Movimentação Financeira (IPMF) (parcela residual).....	0,11
<b>5200 Outros 19.224,74</b>	
5200.01 IOF Seguros.....	2.453,19
5200.02 IOF Oper. Crédito.....	13.360,32

5200.03 IOF Oper. Câmbio.....	2.714,83
5200.04 IOF Aplic. Financ.....	439,87
5200.05 IOF Outros.....	256,53
<b>9000 Outros Tributos.....</b>	<b>16.042,16</b>

## ANEXO B

Tabela 36 – Testes de Raiz Unitária PP e KPSS para as séries componentes do VAR

Obs.: as células sombreadas indicam rejeição da hipótese de presença de raiz unitária

LOG	Período	sem cte. e tend.		cte.		p-valor	
		PP	KPSS	PP	KPSS	cte. e tend.	
						PP	KPSS
Cred	2002m01 a 2009m12	1,00	-	1,00	0,00	0,48	0,00
Desemp	2003m01 a 2009m12	0,21	-	0,84	0,00	0,03	0,48
IBC	2003m01 a 2009m12	1,00	-	0,22	0,00	0,00	0,22
IBRX (2002)	2002m01 a 2009m12	0,99	-	0,76	0,00	0,73	0,02
IBRX (2004)	2004m01 a 2009m12	0,98	-	0,45	0,00	0,68	0,02
IPI (2002)	2002m01 a 2009m12	0,82	-	0,01	0,00	0,00	0,18
IPI (2004)	2004m01 a 2009m12	0,79	-	0,01	0,03	0,02	0,20
IVV (2002)	2002m01 a 2009m12	1,00	-	0,04	0,00	0,00	0,03
IVV (2004)	2004m01 a 2009m12	1,00	-	0,05	0,00	0,00	0,19
Juros	2002m01 a 2009m12	0,28	-	0,72	0,00	0,32	0,42
Selic (2002)	2001m07 a 2009m07	0,14	-	0,84	0,02	0,90	0,02
Selic (2004)	2003m07 a 2009m07	0,35	-	0,77	0,03	0,88	0,00
MS (2002)	2002m01 a 2009m12	0,97	-	0,47	0,00	0,00	0,00
MS (2004)	2004m01 a 2009m12	0,99	-	0,53	0,00	0,00	1,00
PIB (2002)	2002m01 a 2009m12	1,00	-	0,94	0,00	0,01	0,12
PIB (2004)	2004m01 a 2009m12	0,99	-	0,82	0,00	0,02	0,46
PO	2002m01 a 2009m12	0,83	-	0,00	0,01	0,00	0,58
Rendim	2004m01 a 2009m12	0,98	-	0,05	0,00	0,98	1,00

H0 (PP): presença de raiz unitária

H0 (KPSS): a série é estacionária

Tabela 37 – Teste ADF de Raiz Unitária (procedimento de Doldado) para as séries componentes do VAR

Obs.: as células sombreadas indicam rejeição da hipótese de presença de raiz unitária

	Defasagens	Cte	Tend	p-valor
Cred	1	S	S	0,37
Desemp	0	N	N	0,79
IBC	13	S	S	0,11
IBRX (2002)	1	N	N	0,99
IBRX (2004)	1	N	N	0,97
IPI (2002)	12	S	S	0,10
IPI (2004)	12	N	N	0,83
IVV (2002)	15	N	N	1,00
IVV (2004)	14	S	S	0,20
Juros	1	N	N	0,14
Selic (2002)	1	N	N	0,18
Selic (2004)	1	N	N	0,43
MS (2002)	13	S	S	0,61
MS (2004)	13	N	N	0,91
PIB (2002)	8	S	S	0,01
PIB (2004)	3	S	S	0,04
PO	12	S	S	0,09
Rendim (2004)	12	S	S	0,14

H0: presença de raiz unitária

Tabela 38 - Teste de Dickey-Pantula para múltiplas raízes unitárias para as séries componentes do VAR

Obs.: as células sombreadas indicam rejeição da hipótese de presença de raiz unitária

	p-valor		
	n=3	n=2	n=1
Cred	0,00	0,00	0,00
Desemp	0,00	0,00	0,36
IBC	0,00	0,00	0,32
IBRX (2002)	0,00	0,00	0,07
IBRX (2004)	0,00	0,00	0,16
IPI (2002)	0,00	0,00	0,81
IPI (2004)	0,00	0,00	0,82
IVV (2002)	0,00	0,00	0,54
IVV (2004)	0,00	0,00	0,45
Juros	0,00	0,00	0,40
Selic (2002)	0,00	0,00	0,17
Selic (2004)	0,00	0,00	0,44
MS (2002)	0,00	0,00	0,37
MS (2004)	0,00	0,00	0,40
PIB (2002)	0,00	0,00	0,18
PIB (2004)	0,00	0,00	0,23
PO	0,00	0,00	0,88
Rendim (2004)	0,00	0,00	0,60

H0: presença de n raízes unitárias

H1: presença de (n-1) raízes unitárias

Tabela 39 - Testes para seleção dos componentes determinísticos no modelo ADF (procedimento de Doldado) para as séries componentes do VAR

	Período		$\Phi_3$	$\Phi_2$	$\Phi_1$	$T_{\beta T}$	$T_{\alpha\mu}$
Cred	2002m01 a 2009m12	Observado	5,12	4,78	-	3,20	-
		Crítico (5%)	6,49	4,88	-	2,79	-
Desemp	2003m01 a 2009m12	Observado	6,67	4,78	0,35	-3,65	0,83
		Crítico (5%)	6,49	4,88	4,86	2,77	2,51
IBC	2003m01 a 2009m12	Observado	5,84	-8,47	-	2,94	-
		Crítico (5%)	6,73	5,13	-	2,77	-
IBRX (2002)	2002m01 a 2009m12	Observado	1,23	1,36	0,78	1,56	1,24
		Crítico (5%)	6,49	4,88	4,86	2,79	2,54
IBRX (2004)	2004m01 a 2009m12	Observado	0,84	1,63	1,58	1,29	1,76
		Crítico (5%)	6,73	5,13	4,86	2,74	2,48
IPI (2002)	2002m01 a 2009m12	Observado	3,83	3,45	-	2,77	-
		Crítico (5%)	6,49	4,88	-	2,77	-
IPI (2004)	2004m01 a 2009m12	Observado	3,70	3,87	1,77	2,72	1,85
		Crítico (5%)	6,73	5,13	4,86	2,74	2,48
IVV (2002)	2002m01 a 2009m12	Observado	1,76	1,17	0,00	2,76	0,00
		Crítico (5%)	6,49	4,88	4,86	2,77	2,51
IVV (2004)	2004m01 a 2009m12	Observado	7,02	4,91	-	2,84	-
		Crítico (5%)	6,73	5,13	-	2,74	-
Juros	2002m01 a 2009m12	Observado	9,53	6,56	0,07	-2,35	0,38
		Crítico (5%)	6,49	4,88	4,86	2,79	2,54
Selic (2002)	2001m07 a 2009m12	Observado	1,25	1,04	0,29	-1,58	0,75
		Crítico (5%)	6,49	4,88	4,86	2,79	2,54
Selic (2004)	2003m07 a 2009m12	Observado	0,92	1,16	0,80	-1,35	1,25
		Crítico (5%)	6,73	5,13	4,86	2,74	2,48
MS (2002)	2002m01 a 2009m12	Observado	10,25	11,01	-	4,53	-
		Crítico (5%)	6,49	4,88	-	2,77	-
MS (2004)	2004m01 a 2009m12	Observado	1,74	1,59	0,57	1,87	1,05
		Crítico (5%)	6,73	5,13	4,86	2,74	2,48
PIB (2002)	2002m01 a 2009m12	Observado	-13,08	-7,34	0,00	4,24	-0,06
		Crítico (5%)	6,49	4,88	4,86	2,79	2,51
PIB (2004)	2004m01 a 2009m12	Observado	5,88	4,47	-	3,43	-
		Crítico (5%)	6,73	5,13	-	2,77	-
PO	2002m01 a 2009m12	Observado	4,18	3,55	-	2,89	-
		Crítico (5%)	6,49	4,88	-	2,77	-
Rendim	2004m01 a 2009m12	Observado	4,79	7,52	-	3,10	-
		Crítico (5%)	6,73	5,13	-	2,74	-

$H_0(\Phi_3)$ : no modelo  $\Delta y_t = a_0 + \alpha y_{t-1} + a_2 t + e_t$ , testa-se a hipótese conjunta  $\alpha = a_2 = 0$

$H_0(\Phi_2)$ : no modelo  $\Delta y_t = a_0 + \alpha y_{t-1} + a_2 t + e_t$ , testa-se a hipótese conjunta  $\alpha = a_2 = a_0 = 0$

$H_0(\Phi_1)$ : no modelo  $\Delta y_t = a_0 + \alpha y_{t-1} + e_t$ , testa-se a hipótese conjunta  $\alpha = a_0 = 0$

$H_0(T_{\beta T})$ : no modelo  $\Delta y_t = a_0 + \alpha y_{t-1} + a_2 t + e_t$ , testa-se a hipótese  $a_2 = 0$  dado que  $\alpha = 0$

$H_0(T_{\alpha\mu})$ : no modelo  $\Delta y_t = a_0 + \alpha y_{t-1} + e_t$ , testa-se a hipótese  $a_0 = 0$  dado que  $\alpha = 0$

## ANEXO C

Tabela 40 e 41 – Estimação e previsões do modelo vencedor SARIMA (2,1,0)(1,1,0)<sub>6</sub> + 6 dos para a série Tributos sobre a Renda

Dependent Variable: DLOG(RENDA,1,6)  
 Method: Least Squares  
 Date: 04/01/13 Time: 19:12  
 Sample: 2003M01 2009M12  
 Included observations: 84  
 Convergence achieved after 12 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D2003M03	-0.202843	0.089017	-2.278702	0.0255
D2003M04	0.329877	0.090608	3.640686	0.0005
D2003M06	-0.254107	0.085436	-2.974242	0.0039
D2003M12	0.311843	0.085498	3.647361	0.0005
D2004M05	-0.267039	0.078595	-3.397657	0.0011
D2004M12	0.223358	0.074967	2.979426	0.0039
AR(1)	-0.599680	0.112854	-5.313782	0.0000
AR(2)	-0.333401	0.116011	-2.873873	0.0053
SAR(6)	-0.671106	0.085530	-7.846486	0.0000
R-squared	0.700886	Mean dependent var		-0.002215
Adjusted R-squared	0.668980	S.D. dependent var		0.183380
S.E. of regression	0.105506	Akaike info criterion		-1.559136
Sum squared resid	0.834868	Schwarz criterion		-1.298691
Log likelihood	74.48372	Hannan-Quinn criter.		-1.454440
Durbin-Watson stat	1.900714			
Inverted AR Roots	.81-.47i	.81+.47i	.00+.94i	-.00-.94i
	-.30+.49i	-.30-.49i	-.81-.47i	-.81+.47i

## Renda - ARIMA em R\$ milhões

Período: 01/03 a 12/09

(2,1,0)(1,1,0) <sub>6</sub> + 6 do	Realizado	Previsão
2010M01	28,716	26,751
2010M02	14,417	16,496
2010M03	18,256	20,171
2010M04	25,496	24,992
2010M05	17,128	18,326
2010M06	17,599	21,097
2010M07	21,732	24,300
2010M08	17,982	16,280
2010M09	16,891	18,513
2010M10	23,887	26,787
2010M11	18,174	21,196
2010M12	25,917	22,208

Tabela 42 e 43 – Estimação e previsões do modelo vencedor VAR em 1º diferença  
Renda x PO x IBRX para a série Tributos sobre a Renda

Vector Autoregression Estimates  
Date: 09/02/11 Time: 16:06  
Sample (adjusted): 2003M06 2009M12  
Included observations: 79 after adjustments  
Standard errors in ( ) & t-statistics in [ ]

	LOG(RENDA)	LOG(PO)	LOG(IBRX)
LOG(RENDA(-1))	-0.181581 (0.12152) [-1.49428]	-0.054551 (0.04956) [-1.10076]	0.257791 (0.09291) [ 2.77466]
LOG(RENDA(-2))	0.158452 (0.12921) [ 1.22632]	-0.002408 (0.05269) [-0.04570]	0.051138 (0.09879) [ 0.51765]
LOG(RENDA(-3))	0.249309 (0.13282) [ 1.87705]	0.011788 (0.05417) [ 0.21763]	-0.158566 (0.10155) [-1.56145]
LOG(RENDA(-4))	0.014861 (0.12909) [ 0.11512]	-0.018395 (0.05264) [-0.34941]	0.013764 (0.09870) [ 0.13946]
LOG(RENDA(-5))	-0.200735 (0.12495) [-1.60647]	0.015442 (0.05096) [ 0.30303]	0.011915 (0.09554) [ 0.12472]
LOG(PO(-1))	0.527542 (0.34243) [ 1.54058]	0.792364 (0.13965) [ 5.67391]	0.185435 (0.26181) [ 0.70827]
LOG(PO(-2))	-0.604585 (0.42409) [-1.42561]	-0.186655 (0.17295) [-1.07923]	-0.014207 (0.32425) [-0.04382]
LOG(PO(-3))	0.077758 (0.43200) [ 0.18000]	0.281274 (0.17618) [ 1.59653]	0.292730 (0.33029) [ 0.88627]
LOG(PO(-4))	-0.612255 (0.41325) [-1.48154]	-0.161037 (0.16853) [-0.95552]	-0.239708 (0.31596) [-0.75866]
LOG(PO(-5))	-0.105716 (0.34138) [-0.30967]	-0.032848 (0.13922) [-0.23594]	0.094052 (0.26101) [ 0.36034]
LOG(IBRX(-1))	0.223815 (0.16860) [ 1.32747]	0.109329 (0.06876) [ 1.59001]	1.234772 (0.12891) [ 9.57857]
LOG(IBRX(-2))	-0.169185 (0.26126) [-0.64757]	-0.057233 (0.10655) [-0.53716]	-0.331507 (0.19975) [-1.65957]
LOG(IBRX(-3))	0.365847	0.091132	0.117791

	(0.26638)	(0.10863)	(0.20367)
	[ 1.37341]	[ 0.83889]	[ 0.57835]
LOG(IBRX(-4))	-0.371424	-0.244396	-0.303162
	(0.25042)	(0.10213)	(0.19146)
	[-1.48321]	[-2.39308]	[-1.58339]
LOG(IBRX(-5))	0.327083	0.140976	0.165374
	(0.16100)	(0.06566)	(0.12310)
	[ 2.03159]	[ 2.14710]	[ 1.34346]
C	6.494001	1.483881	-1.837865
	(2.08984)	(0.85228)	(1.59783)
	[ 3.10742]	[ 1.74107]	[-1.15022]
@SEAS(1)	0.439327	-0.012517	-0.011212
	(0.08430)	(0.03438)	(0.06445)
	[ 5.21145]	[-0.36409]	[-0.17396]
@SEAS(2)	0.039637	-0.061711	-0.013528
	(0.08143)	(0.03321)	(0.06226)
	[ 0.48678]	[-1.85835]	[-0.21729]
@SEAS(3)	0.109160	0.102443	0.192230
	(0.10894)	(0.04443)	(0.08329)
	[ 1.00200]	[ 2.30577]	[ 2.30782]
@SEAS(4)	0.124603	-0.049528	0.076232
	(0.08412)	(0.03430)	(0.06431)
	[ 1.48134]	[-1.44381]	[ 1.18533]
@SEAS(5)	0.042905	0.049502	0.007073
	(0.07854)	(0.03203)	(0.06005)
	[ 0.54630]	[ 1.54552]	[ 0.11779]
@SEAS(6)	0.080324	-0.070709	0.021581
	(0.07280)	(0.02969)	(0.05566)
	[ 1.10342]	[-2.38180]	[ 0.38774]
@SEAS(7)	0.185215	0.013458	0.148941
	(0.09095)	(0.03709)	(0.06954)
	[ 2.03644]	[ 0.36282]	[ 2.14185]
@SEAS(9)	0.009068	-0.012752	0.072460
	(0.07877)	(0.03213)	(0.06023)
	[ 0.11511]	[-0.39695]	[ 1.20309]
@SEAS(10)	0.211816	0.030133	0.079728
	(0.06604)	(0.02693)	(0.05050)
	[ 3.20721]	[ 1.11878]	[ 1.57891]
@SEAS(11)	0.084846	-0.030857	-0.015550
	(0.05429)	(0.02214)	(0.04151)
	[ 1.56279]	[-1.39364]	[-0.37460]
@SEAS(12)	0.388539	-0.114388	0.054247
	(0.07180)	(0.02928)	(0.05489)
	[ 5.41159]	[-3.90661]	[ 0.98820]
D2005M07	-0.053495	-0.034293	-0.233330
	(0.09592)	(0.03912)	(0.07334)
	[-0.55769]	[-0.87662]	[-3.18147]

D2009M11	0.293430 (0.09830) [ 2.98506]	0.040010 (0.04009) [ 0.99803]	-0.002161 (0.07516) [-0.02875]
R-squared	0.925075	0.900190	0.990904
Adj. R-squared	0.883117	0.844297	0.985811
Sum sq. resids	0.343895	0.057196	0.201032
S.E. equation	0.082933	0.033822	0.063408
F-statistic	22.04768	16.10546	194.5411
Log likelihood	102.6601	173.5173	123.8666
Akaike AIC	-1.864813	-3.658665	-2.401686
Schwarz SC	-0.995016	-2.788868	-1.531888
Mean dependent	5.723427	4.645069	5.845098
S.D. dependent	0.242579	0.085713	0.532315
Determinant resid covariance (dof adj.)		3.08E-08	
Determinant resid covariance		7.81E-09	
Log likelihood		401.1087	
Akaike information criterion		-7.952119	
Schwarz criterion		-5.342727	

**Renda** em R\$ milhões

Período: 01/03 a 12/09

**Modelagem VAR em 1ª diferença RendaxPOxIBRX**

	Realizado	Previsão
2010M01	28716	28902
2010M02	14417	15038
2010M03	18256	17993
2010M04	25496	25141
2010M05	17128	18884
2010M06	17599	18213
2010M07	21732	20713
2010M08	17982	21430
2010M09	16891	18649
2010M10	23887	25165
2010M11	18174	20696
2010M12	25917	27585

Tabela 44 – Combinação – média aritmética das previsões geradas pelos modelos Indicadores e ARIMA – Tributos sobre a Renda

Renda	em R\$ milhões	
Período: 01/03 a 12/09		
Combinação - Média Aritmética - Indicadores x ARIMA		
	Realizado	Previsão
2010M01	28716	27826
2010M02	14417	15767
2010M03	18256	19082
2010M04	25496	25067
2010M05	17128	18605
2010M06	17599	19655
2010M07	21732	22506
2010M08	17982	18855
2010M09	16891	18581
2010M10	23887	25976
2010M11	18174	20946
2010M12	25917	24897

Tabelas 45 e 46 – Estimação e previsões do modelo SARIMA (0,1,1)(1,1,0)<sub>12</sub> + do para a série Tributos sobre a Folha de Salários

Dependent Variable: DLOG(FOLHA,1,12)

Method: Least Squares

Date: 09/05/11 Time: 15:15

Sample: 2004M01 2009M12

Included observations: 72

Convergence achieved after 9 iterations

MA Backcast: 2003M12

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D2009M11	0.072378	0.025687	2.817663	0.0063
AR(12)	-0.450414	0.108740	-4.142120	0.0001
MA(1)	-0.687702	0.088507	-7.769995	0.0000
R-squared	0.493552	Mean dependent var		-0.000534
Adjusted R-squared	0.478872	S.D. dependent var		0.042879
S.E. of regression	0.030954	Akaike info criterion		-4.071865
Sum squared resid	0.066112	Schwarz criterion		-3.977004
Log likelihood	149.5871	Hannan-Quinn criter.		-4.034101
Durbin-Watson stat	1.728433			
Inverted AR Roots	.90+.24i .24+.90i	.90-.24i .24-.90i	.66-.66i -.24-.90i	.66-.66i -.24+.90i
Inverted MA Roots	-.66-.66i	-.66+.66i	-.90-.24i	-.90+.24i
	.69			

Folha em R\$ milhões

Período: 2004m01 a 2009m12

(0,1,1)(1,1,0) <sub>12</sub> + do	Realizado	Previsão
2010M01	19,638	20,340
2010M02	19,127	19,733
2010M03	19,836	20,356
2010M04	20,373	20,948
2010M05	20,490	20,747
2010M06	20,661	20,924
2010M07	20,845	21,277
2010M08	21,377	21,155
2010M09	21,394	21,440
2010M10	21,654	21,958
2010M11	21,910	23,953
2010M12	33,854	34,753

Tabelas 47 e 48 – Estimação e previsões do modelo VEC Folha x PO x MS para a série Tributos sobre a Folha de Salários

Vector Error Correction Estimates

Date: 09/09/11 Time: 17:29

Sample: 2004M11 2009M12

Included observations: 62

Standard errors in ( ) & t-statistics in [ ]

Cointegrating Eq:	CointEq1		
LAP(-1)	1.000000		
LPO(-1)	-0.158529 (0.13910) [-1.13971]		
LMS(-1)	-1.834144 (0.36585) [-5.01342]		
@TREND(94M07)	0.001964 (0.00186) [ 1.05804]		
C	4.257707		
Error Correction:	D(LAP)	D(LPO)	D(LMS)
CointEq1	-0.159164 (0.15818) [-1.00621]	-0.145593 (0.24310) [-0.59890]	0.433300 (0.09115) [ 4.75385]
D(LAP(-1))	-0.363405 (0.21091) [-1.72300]	0.064984 (0.32415) [ 0.20048]	-0.323595 (0.12153) [-2.66261]
D(LAP(-2))	-0.405361 (0.20419) [-1.98525]	-0.149447 (0.31381) [-0.47624]	-0.235096 (0.11766) [-1.99816]
D(LPO(-1))	0.100370 (0.11061) [ 0.90743]	-0.319483 (0.16999) [-1.87940]	-0.012612 (0.06374) [-0.19788]
D(LPO(-2))	0.151610 (0.11007) [ 1.37738]	-0.169180 (0.16916) [-1.00009]	0.149168 (0.06343) [ 2.35186]
D(LMS(-1))	-0.069155 (0.27311) [-0.25321]	0.016474 (0.41973) [ 0.03925]	-0.075452 (0.15737) [-0.47945]
D(LMS(-2))	-0.207667 (0.22587) [-0.91942]	0.324813 (0.34713) [ 0.93571]	0.091734 (0.13015) [ 0.70483]
C	0.034896 (0.01221) [ 2.85841]	0.010852 (0.01876) [ 0.57841]	0.010813 (0.00703) [ 1.53717]

@SEAS(1)	-0.263457 (0.08441) [-3.12112]	-0.084621 (0.12973) [-0.65230]	-0.090706 (0.04864) [-1.86487]
@SEAS(2)	-0.004135 (0.10725) [-0.03855]	-0.013321 (0.16484) [-0.08081]	-0.074157 (0.06180) [-1.19990]
@SEAS(3)	-0.235989 (0.10244) [-2.30364]	0.120060 (0.15744) [ 0.76258]	-0.103743 (0.05903) [-1.75748]
@SEAS(4)	-0.028514 (0.02116) [-1.34729]	0.003862 (0.03253) [ 0.11872]	-0.001935 (0.01219) [-0.15863]
@SEAS(5)	-0.028421 (0.02128) [-1.33555]	0.052007 (0.03270) [ 1.59020]	-0.029861 (0.01226) [-2.43523]
@SEAS(6)	-0.010186 (0.01637) [-0.62203]	-0.010813 (0.02517) [-0.42966]	0.001348 (0.00944) [ 0.14291]
@SEAS(7)	-0.014698 (0.01670) [-0.87994]	-0.008036 (0.02567) [-0.31305]	-0.008513 (0.00962) [-0.88452]
@SEAS(9)	-0.012282 (0.01538) [-0.79855]	-0.000102 (0.02364) [-0.00434]	-0.002730 (0.00886) [-0.30805]
@SEAS(10)	-0.017570 (0.01558) [-1.12789]	0.043172 (0.02394) [ 1.80323]	-0.001310 (0.00898) [-0.14597]
@SEAS(11)	-0.012803 (0.01580) [-0.81033]	-0.036513 (0.02428) [-1.50370]	0.076165 (0.00910) [ 8.36599]
@SEAS(12)	0.376490 (0.02520) [ 14.9388]	-0.144836 (0.03873) [-3.73940]	0.195700 (0.01452) [ 13.4761]
R-squared	0.987722	0.784834	0.979222
Adj. R-squared	0.982583	0.694764	0.970524
Sum sq. resids	0.024919	0.058858	0.008274
S.E. equation	0.024073	0.036997	0.013871
F-statistic	192.1821	8.713625	112.5804
Log likelihood	154.4225	127.7784	188.6008
Akaike AIC	-4.368468	-3.508980	-5.470993
Schwarz SC	-3.716604	-2.857117	-4.819130
Mean dependent	0.013900	0.001184	0.007888
S.D. dependent	0.182408	0.066965	0.080795
Determinant resid covariance (dof adj.)		1.27E-10	
Determinant resid covariance		4.23E-11	
Log likelihood		476.5861	
Akaike information criterion		-13.40600	
Schwarz criterion		-11.31318	

Folha	em R\$ milhões	
Período: 2004m01 a 2009m12		
Modelagem VEC Folha x PO x MS		
	Realizado	Previsão
2010M01	19638	19638
2010M02	19127	19114
2010M03	19836	20351
2010M04	20373	20275
2010M05	20490	20899
2010M06	20661	20720
2010M07	20845	20727
2010M08	21377	21547
2010M09	21394	21411
2010M10	21654	21789
2010M11	21910	21829
2010M12	33854	34244

Tabela 49 – Combinação – média aritmética das previsões geradas pelos modelos ARIMA e VEC – Tributos sobre a Folha de Salários

Folha	em R\$ milhões	
Período: 2004m01 a 2009m12		
Combinação - Média Aritmética - ARIMA x VEC		
	Realizado	Previsão
2010M01	19638	18982
2010M02	19127	18364
2010M03	19836	19083
2010M04	20373	19221
2010M05	20490	19524
2010M06	20661	19583
2010M07	20845	19980
2010M08	21377	20378
2010M09	21394	20510
2010M10	21654	20567
2010M11	21910	20501
2010M12	33854	31982

Tabelas 50 e 51 – Estimação e previsões do modelo vencedor SARIMA (0,1,0)(0,1,1)<sub>12</sub> para a série Tributos sobre Bens e Serviços

Dependent Variable: DLOG(BS,1,12)  
 Method: Least Squares  
 Date: 08/19/11 Time: 17:41  
 Sample: 2004M01 2009M12  
 Included observations: 72  
 Convergence achieved after 7 iterations  
 MA Backcast: 2003M01 2003M12

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(12)	-0.848235	0.039674	-21.38018	0.0000
R-squared	0.453589	Mean dependent var		0.001659
Adjusted R-squared	0.453589	S.D. dependent var		0.065637
S.E. of regression	0.048519	Akaike info criterion		-3.199940
Sum squared resid	0.167139	Schwarz criterion		-3.168319
Log likelihood	116.1978	Hannan-Quinn criter.		-3.187351
Durbin-Watson stat	2.031573			
Inverted MA Roots	.99	.85+.49i	.85-.49i	.49+.85i
	.49-.85i	.00-.99i	-.00+.99i	-.49-.85i
	-.49+.85i	-.85-.49i	-.85+.49i	-.99

**BS** em R\$ milhões

Período: 01/04 a 12/09

(0,1,0)(0,1,1) <sub>12</sub>	Realizado	Previsão
2010M01	15,693	16,163
2010M02	13,826	14,081
2010M03	14,677	14,698
2010M04	15,565	15,576
2010M05	15,291	15,321
2010M06	16,061	15,977
2010M07	15,705	16,206
2010M08	16,478	16,586
2010M09	17,096	17,159
2010M10	17,595	17,405
2010M11	17,896	18,088
2010M12	18,480	17,940

Tabelas 52 e 53 – Estimação e previsões do modelo vencedor VEC BS x IBC x Selic para a série Tributos sobre Bens e Serviços

Vector Error Correction Estimates  
 Date: 09/08/11 Time: 15:29  
 Sample (adjusted): 2004M04 2009M12  
 Included observations: 69 after adjustments  
 Standard errors in ( ) & t-statistics in [ ]

Cointegrating Eq:	CointEq1
LOG(BS(-1))	1.000000
LOG(IBC(-1))	-2.817759 (0.23250) [-12.1192]
SELICAA6(-1)	0.002355 (0.00279) [ 0.84302]
@TREND(94M07)	0.005880 (0.00075) [ 7.85186]
C	7.101481

Error Correction:	D(LOG(BS))	D(LOG(IBC))	D(SELICAA6)
CointEq1	-0.740891 (0.17724) [-4.18012]	0.021499 (0.07867) [ 0.27328]	2.805304 (1.20089) [ 2.33603]
D(LOG(BS(-1)))	0.142414 (0.13840) [ 1.02900]	-0.035377 (0.06143) [-0.57589]	-1.433415 (0.93772) [-1.52862]
D(LOG(BS(-2)))	0.090126 (0.10995) [ 0.81971]	-0.004222 (0.04880) [-0.08650]	-0.999223 (0.74495) [-1.34133]
D(LOG(IBC(-1)))	-1.042097 (0.50663) [-2.05692]	-0.501185 (0.22487) [-2.22874]	3.670818 (3.43262) [ 1.06939]
D(LOG(IBC(-2)))	-0.610204 (0.38520) [-1.58414]	-0.200591 (0.17097) [-1.17323]	0.083363 (2.60986) [ 0.03194]
D(SELICAA6(-1))	0.021534 (0.01235) [ 1.74431]	-0.004506 (0.00548) [-0.82226]	0.602565 (0.08365) [ 7.20373]
D(SELICAA6(-2))	-0.020782 (0.01061) [-1.95961]	0.001727 (0.00471) [ 0.36685]	-0.070795 (0.07185) [-0.98527]
C	-0.007390 (0.02040) [-0.36223]	0.016377 (0.00906) [ 1.80854]	0.054802 (0.13823) [ 0.39646]

@SEAS(1)	0.039111 (0.03212) [ 1.21762]	-0.065517 (0.01426) [-4.59539]	-0.493521 (0.21763) [-2.26770]
@SEAS(2)	-0.033758 (0.03981) [-0.84804]	-0.064949 (0.01767) [-3.67585]	-0.419983 (0.26971) [-1.55714]
@SEAS(3)	0.111738 (0.04113) [ 2.71676]	0.090709 (0.01826) [ 4.96881]	-0.466690 (0.27867) [-1.67472]
@SEAS(4)	0.050547 (0.05481) [ 0.92215]	0.040394 (0.02433) [ 1.66027]	-0.109264 (0.37139) [-0.29420]
@SEAS(5)	-0.035106 (0.04871) [-0.72065]	-0.001805 (0.02162) [-0.08347]	0.476217 (0.33006) [ 1.44283]
@SEAS(6)	-0.038620 (0.02596) [-1.48757]	-0.034739 (0.01152) [-3.01466]	0.174142 (0.17590) [ 0.99000]
@SEAS(8)	0.014187 (0.02502) [ 0.56711]	0.000744 (0.01110) [ 0.06700]	0.041757 (0.16949) [ 0.24636]
@SEAS(9)	0.027608 (0.02748) [ 1.00484]	-0.022126 (0.01220) [-1.81434]	-0.035584 (0.18616) [-0.19115]
@SEAS(10)	0.014080 (0.02467) [ 0.57064]	-0.005926 (0.01095) [-0.54111]	-0.107740 (0.16718) [-0.64446]
@SEAS(11)	0.047648 (0.02873) [ 1.65866]	-0.022673 (0.01275) [-1.77816]	-0.649697 (0.19464) [-3.33802]
@SEAS(12)	0.074608 (0.03308) [ 2.25554]	-0.024655 (0.01468) [-1.67927]	0.052434 (0.22412) [ 0.23396]
D2004M04	-0.027771 (0.04686) [-0.59268]	0.008826 (0.02080) [ 0.42438]	2.900467 (0.31747) [ 9.13610]
D2008M11	-0.054753 (0.04125) [-1.32725]	-0.057832 (0.01831) [-3.15841]	0.242671 (0.27951) [ 0.86822]
D2008M12	-0.081857 (0.04329) [-1.89083]	-0.071498 (0.01922) [-3.72088]	-0.366383 (0.29332) [-1.24910]
D2005M11	0.026645 (0.04016) [ 0.66347]	0.001908 (0.01783) [ 0.10703]	0.986905 (0.27210) [ 3.62702]
R-squared	0.815902	0.914884	0.807037

Adj. R-squared	0.727855	0.874176	0.714750
Sum sq. resids	0.055957	0.011024	2.568771
S.E. equation	0.034878	0.015481	0.236311
F-statistic	9.266676	22.47449	8.744880
Log likelihood	147.6394	203.6840	15.62166
Akaike AIC	-3.612736	-5.237216	0.213865
Schwarz SC	-2.868034	-4.492514	0.958567
Mean dependent	0.006812	0.002233	-0.031014
S.D. dependent	0.066857	0.043643	0.442457

Determinant resid covariance (dof adj.)	1.38E-08
Determinant resid covariance	4.09E-09
Log likelihood	372.6381
Akaike information criterion	-8.685161
Schwarz criterion	-6.321541

**BS** em R\$ milhões

Período: 01/04 a 12/09

Modelagem VEC BSxIBC

	Realizado	Previsão
2010M01	15,693	15669
2010M02	13,826	14310
2010M03	14,677	14572
2010M04	15,565	14981
2010M05	15,291	15053
2010M06	16,061	15871
2010M07	15,705	15401
2010M08	16,478	16680
2010M09	17,096	17275
2010M10	17,595	17457
2010M11	17,896	18100
2010M12	18,480	18408

Tabela 54 – Combinação – média aritmética das previsões geradas pelos modelos ARIMA e VEC – Tributos sobre Bens e Serviços

BS	em R\$ milhões	
Período: 01/04 a 12/09		
Combinação - Média Aritmética - ARIMA x VEC		
	Realizado	Previsão
2010M01	15,693	15916
2010M02	13,826	14195
2010M03	14,677	14635
2010M04	15,565	15279
2010M05	15,291	15187
2010M06	16,061	15924
2010M07	15,705	15803
2010M08	16,478	16633
2010M09	17,096	17217
2010M10	17,595	17431
2010M11	17,896	18094
2010M12	18,480	18174

Tabelas 55 e 56 – Estimação e previsões do modelo ARIMA – (2,1,0) + 4 dos para a série Tributos sobre Transações Financeiras

Dependent Variable: DLOG(TF)  
 Method: Least Squares  
 Date: 09/05/11 Time: 18:13  
 Sample: 2002M01 2009M12  
 Included observations: 96  
 Convergence achieved after 6 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D2008M01	0.490868	0.067003	7.326114	0.0000
D2008M02	0.305235	0.066880	4.563936	0.0000
D2009M01	-0.233095	0.049642	-4.695542	0.0000
D2009M11	0.352627	0.055337	6.372315	0.0000
AR(1)	-0.893499	0.079690	-11.21223	0.0000
AR(2)	-0.664201	0.078450	-8.466506	0.0000
R-squared	0.691982	Mean dependent var		0.014773
Adjusted R-squared	0.674870	S.D. dependent var		0.129958
S.E. of regression	0.074102	Akaike info criterion		-2.306286
Sum squared resid	0.494200	Schwarz criterion		-2.146014
Log likelihood	116.7017	Hannan-Quinn criter.		-2.241501
Durbin-Watson stat	2.046436			
Inverted AR Roots	-.45-.68i	-.45+.68i		

TF	em R\$ milhões	
Período: 01/02 a 12/09		
(2,1,0) + 4 dos	Realizado	Previsão
2010M01	1944	2,244
2010M02	1868	2,214
2010M03	1879	2,163
2010M04	2038	2,237
2010M05	2070	2,205
2010M06	2098	2,182
2010M07	2107	2,217
2010M08	2221	2,209
2010M09	2124	2,199
2010M10	2754	2,218
2010M11	2235	2,203
2010M12	2407	2,201

Tabelas 57 e 58 – Estimação e previsões do modelo VAR TF x PO para a série Tributos sobre Transações Financeiras

Vector Autoregression Estimates

Date: 04/04/13 Time: 10:55

Sample (adjusted): 2002M04 2009M12

Included observations: 93 after adjustments

Standard errors in ( ) & t-statistics in [ ]

	LOG(TF)	LOG(PO)
LOG(TF(-1))	0.414806 (0.08544) [ 4.85476]	0.036339 (0.03687) [ 0.98561]
LOG(TF(-2))	0.251764 (0.09886) [ 2.54664]	-0.065314 (0.04266) [-1.53108]
LOG(TF(-3))	0.314017 (0.08474) [ 3.70550]	0.043945 (0.03657) [ 1.20175]
LOG(PO(-1))	-0.169009 (0.27153) [-0.62242]	0.776831 (0.11717) [ 6.63007]
LOG(PO(-2))	0.730505 (0.33600) [ 2.17409]	-0.106020 (0.14499) [-0.73124]
LOG(PO(-3))	-0.436799 (0.27113) [-1.61103]	0.198293 (0.11699) [ 1.69491]
C	-0.483424 (0.79116) [-0.61103]	0.559891 (0.34139) [ 1.64003]
@SEAS(1)	-0.064678 (0.05740) [-1.12674]	-0.041554 (0.02477) [-1.67762]
@SEAS(2)	-0.020965 (0.06460) [-0.32453]	-0.080087 (0.02788) [-2.87306]
@SEAS(3)	0.011056 (0.05003) [ 0.22099]	0.090781 (0.02159) [ 4.20512]
@SEAS(4)	0.068258 (0.06095) [ 1.11997]	-0.030116 (0.02630) [-1.14517]
@SEAS(5)	-0.055556 (0.05259) [-1.05633]	0.030268 (0.02269) [ 1.33373]

@SEAS(6)	0.071963 (0.04843) [ 1.48585]	-0.045678 (0.02090) [-2.18571]
@SEAS(8)	0.015390 (0.04916) [ 0.31303]	0.003006 (0.02121) [ 0.14167]
@SEAS(9)	-0.028044 (0.04757) [-0.58948]	-0.008409 (0.02053) [-0.40963]
@SEAS(10)	-0.020782 (0.04640) [-0.44788]	0.037926 (0.02002) [ 1.89418]
@SEAS(11)	-0.022526 (0.05202) [-0.43307]	-0.041028 (0.02244) [-1.82793]
@SEAS(12)	0.057035 (0.04619) [ 1.23465]	-0.128135 (0.01993) [-6.42822]
D2008M01	0.494151 (0.09775) [ 5.05502]	0.065532 (0.04218) [ 1.55359]
D2008M02	0.696489 (0.10548) [ 6.60295]	0.002583 (0.04552) [ 0.05674]
D2009M11	0.358120 (0.10039) [ 3.56723]	0.003831 (0.04332) [ 0.08844]
<hr/>		
R-squared	0.972676	0.852440
Adj. R-squared	0.965086	0.811452
Sum sq. resids	0.580239	0.108038
S.E. equation	0.089771	0.038737
F-statistic	128.1505	20.79693
Log likelihood	104.1152	182.2799
Akaike AIC	-1.787425	-3.468386
Schwarz SC	-1.215547	-2.896508
Mean dependent	4.412036	4.636469
S.D. dependent	0.480435	0.089209
<hr/>		
Determinant resid covariance (dof adj.)		1.18E-05
Determinant resid covariance		7.10E-06
Log likelihood		287.3670
Akaike information criterion		-5.276710
Schwarz criterion		-4.132956
<hr/>		

TF	em R\$ milhões	
Período: 01/02 a 12/09		
Combinação - Média Aritmética - ARIMA x VAR		
	Realizado	Previsão
2010M01	1944	2098
2010M02	1868	1912
2010M03	1879	1968
2010M04	2038	2138
2010M05	2070	2122
2010M06	2098	2141
2010M07	2107	2101
2010M08	2221	2205
2010M09	2124	2082
2010M10	2754	2426
2010M11	2235	2008
2010M12	2407	2268

Tabela 59 – Combinação – média aritmética das previsões geradas pelos modelos ARIMA e VAR – Tributos sobre Transações Financeiras

TF	em R\$ milhões	
Período: 01/02 a 12/09		
Combinação - Média Aritmética - ARIMA x VEC		
	Realizado	Previsão
2010M01	1944	2098
2010M02	1868	1912
2010M03	1879	1968
2010M04	2038	2138
2010M05	2070	2122
2010M06	2098	2141
2010M07	2107	2101
2010M08	2221	2205
2010M09	2124	2082
2010M10	2754	2426
2010M11	2235	2008
2010M12	2407	2268

Tabelas 60 e 61 – Estimação e previsões do modelo SARIMA (0,0,1)(0,1,1)<sub>3</sub> para a série Outros Tributos

Dependent Variable: DLOG(OUTROS,0,3)

Method: Least Squares

Date: 09/26/11 Time: 16:55

Sample: 2003M01 2009M12

Included observations: 84

Convergence achieved after 8 iterations

MA Backcast: 2002M09 2002M12

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1)	0.477270	0.098630	4.838983	0.0000
SMA(3)	-0.770149	0.093477	-8.238929	0.0000
R-squared	0.282702	Mean dependent var		0.020232
Adjusted R-squared	0.273954	S.D. dependent var		0.307134
S.E. of regression	0.261704	Akaike info criterion		0.180314
Sum squared resid	5.616084	Schwarz criterion		0.238191
Log likelihood	-5.573186	Hannan-Quinn criter.		0.203580
Durbin-Watson stat	1.959215			
Inverted MA Roots	.92	-.46-.79i	-.46+.79i	-.48

Outros em R\$ milhões

Período: 01/03 a 12/09

(0,0,1)(0,1,1) <sub>3</sub>	Realizado	Previsão
2010M01	6,483	5,659
2010M02	3,481	3,633
2010M03	3,562	2,274
2010M04	5,601	5,210
2010M05	4,300	3,638
2010M06	3,220	2,271
2010M07	5,527	5,206
2010M08	2,746	3,634
2010M09	3,702	2,275
2010M10	5,398	5,206
2010M11	3,237	3,632
2010M12	5,144	2,272

Tabelas 62 e 63 – Estimação e previsões do modelo VEC Outros x IVV para a série Outros Tributos

Vector Error Correction Estimates

Date: 04/04/13 Time: 11:15

Sample: 2003M01 2009M12

Included observations: 84

Standard errors in ( ) & t-statistics in [ ]

Cointegrating Eq:	CointEq1
LOG(OUTROS(-1))	1.000000
LOG(IVV(-1))	-7.677882 (1.49441) [-5.13772]
@TREND(94M07)	0.050686 (0.00899) [ 5.64023]
C	23.68133

Error Correction:	D(LOG(OUTROS))	D(LOG(IVV))
CointEq1	-0.441901 (0.13887) [-3.18208]	0.044719 (0.01586) [ 2.81923]
D(LOG(OUTROS(-1)))	-0.142587 (0.15136) [-0.94202]	-0.028923 (0.01729) [-1.67294]
D(LOG(OUTROS(-2)))	-0.297770 (0.14560) [-2.04508]	-0.039685 (0.01663) [-2.38619]
D(LOG(OUTROS(-3)))	0.231315 (0.14422) [ 1.60387]	-0.013716 (0.01647) [-0.83264]
D(LOG(OUTROS(-4)))	-0.098113 (0.15452) [-0.63495]	-0.026523 (0.01765) [-1.50277]
D(LOG(OUTROS(-5)))	-0.110674 (0.14389) [-0.76914]	-0.028868 (0.01644) [-1.75640]
D(LOG(OUTROS(-6)))	-0.051716 (0.12857) [-0.40223]	-0.021827 (0.01469) [-1.48628]
D(LOG(IVV(-1)))	-1.744648 (1.06745) [-1.63441]	-0.471010 (0.12193) [-3.86311]
D(LOG(IVV(-2)))	1.080632	-0.342521

	(1.10589) [ 0.97716]	(0.12632) [-2.71162]
D(LOG(IVV(-3)))	-1.135928 (1.16703) [-0.97335]	-0.062625 (0.13330) [-0.46981]
D(LOG(IVV(-4)))	2.739569 (1.15245) [ 2.37717]	0.131617 (0.13163) [ 0.99987]
D(LOG(IVV(-5)))	0.813830 (1.13083) [ 0.71967]	0.129620 (0.12917) [ 1.00352]
D(LOG(IVV(-6)))	1.271302 (1.05000) [ 1.21077]	0.352857 (0.11993) [ 2.94214]
C	0.566960 (0.41125) [ 1.37862]	0.070775 (0.04697) [ 1.50670]
@SEAS(1)	-0.873317 (0.56663) [-1.54126]	-0.186729 (0.06472) [-2.88515]
@SEAS(2)	-1.911582 (0.72451) [-2.63846]	-0.226735 (0.08275) [-2.73987]
@SEAS(3)	0.384993 (0.64149) [ 0.60015]	-0.123723 (0.07327) [-1.68854]
@SEAS(4)	-1.444034 (0.61696) [-2.34056]	-0.159542 (0.07047) [-2.26397]
@SEAS(6)	-0.332468 (0.61119) [-0.54397]	-0.127670 (0.06981) [-1.82879]
@SEAS(7)	-0.215788 (0.68988) [-0.31279]	0.052301 (0.07880) [ 0.66373]
@SEAS(8)	-0.385685 (0.41286) [-0.93418]	-0.052147 (0.04716) [-1.10581]
@SEAS(9)	-0.753207 (0.44363) [-1.69783]	-0.113990 (0.05067) [-2.24958]
@SEAS(10)	-0.179696 (0.39614) [-0.45362]	-0.046489 (0.04525) [-1.02744]
@SEAS(11)	-0.680649 (0.43473) [-1.56569]	-0.077281 (0.04966) [-1.55634]

@SEAS(12)	-0.764159 (0.41934) [-1.82228]	0.284012 (0.04790) [ 5.92954]
D2006M09	0.618276 (0.17350) [ 3.56361]	0.024031 (0.01982) [ 1.21266]
D2008M01	0.251037 (0.16835) [ 1.49116]	0.046840 (0.01923) [ 2.43585]
D2008M03	-0.107782 (0.17155) [-0.62830]	0.031926 (0.01959) [ 1.62934]
D2009M09	-0.781251 (0.18596) [-4.20110]	0.016749 (0.02124) [ 0.78851]
D2009M10	0.841411 (0.21236) [ 3.96217]	0.046512 (0.02426) [ 1.91752]
D2009M11	0.835274 (0.23154) [ 3.60743]	-0.023886 (0.02645) [-0.90317]
D2009M12	0.465973 (0.27220) [ 1.71189]	-0.014177 (0.03109) [-0.45600]
<hr/>		
R-squared	0.956910	0.991611
Adj. R-squared	0.931222	0.986609
Sum sq. resids	1.135550	0.014815
S.E. equation	0.147775	0.016879
F-statistic	37.25114	198.2692
Log likelihood	61.56456	243.8125
Akaike AIC	-0.703918	-5.043155
Schwarz SC	0.222107	-4.117130
Mean dependent	0.002539	0.005466
S.D. dependent	0.563479	0.145864
<hr/>		
Determinant resid covariance (dof adj.)		5.55E-06
Determinant resid covariance		2.13E-06
Log likelihood		310.1741
Akaike information criterion		-5.789860
Schwarz criterion		-3.850995
<hr/>		

Outros		em R\$ milhões	
Período: 01/03 a 12/09			
Modelagem VEC OutrosxIVV			
	Realizado	Previsão	
2010M01	6,483	5907	
2010M02	3,481	4099	
2010M03	3,562	2257	
2010M04	5,601	3873	
2010M05	4,300	3308	
2010M06	3,220	2795	
2010M07	5,527	5112	
2010M08	2,746	3706	
2010M09	3,702	3741	
2010M10	5,398	6225	
2010M11	3,237	3620	
2010M12	5,144	3187	

Tabela 64 – Combinação – média aritmética das previsões geradas pelos modelos ARIMA e VEC – Outros Tributos

Outros		em R\$ milhões	
Período: 01/03 a 12/09			
Combinação - Média Aritmética - ARIMA x VEC			
	Realizado	Previsão	
2010M01	6,483	5783	
2010M02	3,481	3866	
2010M03	3,562	2265	
2010M04	5,601	4542	
2010M05	4,300	3473	
2010M06	3,220	2533	
2010M07	5,527	5159	
2010M08	2,746	3670	
2010M09	3,702	3008	
2010M10	5,398	5716	
2010M11	3,237	3626	
2010M12	5,144	2730	

Tabelas 65 e 66 – Estimação e previsões para o modelo SARIMA (2,1,0)(1,1,0)<sub>12</sub> + 3 dos para a série Total das Receitas

Dependent Variable: DLOG(TOTAL,1,12)

Method: Least Squares

Date: 09/10/11 Time: 13:17

Sample: 2002M01 2009M12

Included observations: 96

Convergence achieved after 10 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D2009M11	0.178817	0.054275	3.294623	0.0014
D2002M09	0.142619	0.047456	3.005294	0.0034
AR(1)	-0.562962	0.111028	-5.070434	0.0000
AR(2)	-0.365493	0.106489	-3.432224	0.0009
SAR(12)	-0.336358	0.106260	-3.165441	0.0021
R-squared	0.361796	Mean dependent var		0.000719
Adjusted R-squared	0.333744	S.D. dependent var		0.073124
S.E. of regression	0.059687	Akaike info criterion		-2.748725
Sum squared resid	0.324191	Schwarz criterion		-2.615166
Log likelihood	136.9388	Hannan-Quinn criter.		-2.694738
Durbin-Watson stat	1.811406			
Inverted AR Roots	.88-.24i	.88+.24i	.65+.65i	.65+.65i
	.24+.88i	.24-.88i	-.24+.88i	-.24-.88i
	-.28+.54i	-.28-.54i	-.65+.65i	-.65+.65i
	-.88-.24i	-.88+.24i		

Total em R\$ milhões

Período: 01/02 a 12/09

(2,1,0)(1,1,0) <sub>12</sub> + 3 dos	Realizado	Previsão
2010M01	72474	71,864
2010M02	52718	52,830
2010M03	58210	59,951
2010M04	69074	66,724
2010M05	59280	56,581
2010M06	59639	61,652
2010M07	65915	67,204
2010M08	60804	59,142
2010M09	61206	59,109
2010M10	71288	75,474
2010M11	63452	79,277
2010M12	85802	80,838

Tabelas 67 e 68 – Estimação e previsões do modelo VEC Total x PO x IBrX para a série Total das Receitas

Vector Error Correction Estimates

Date: 09/08/11 Time: 11:36

Sample (adjusted): 2002M04 2009M12

Included observations: 93 after adjustments

Standard errors in ( ) & t-statistics in [ ]

Cointegrating Eq:	CointEq1		
LOG(TOTAL(-1))	1.000000		
LOG(PO(-1))	-0.131164 (0.19517) [-0.67204]		
LOG(IBRX(-1))	-0.045327 (0.04664) [-0.97176]		
@TREND(94M07)	-0.004558 (0.00103) [-4.41164]		
C	-3.949524		

Error Correction:	D(LOG(TOTAL))	D(LOG(PO))	D(LOG(IBRX))
CointEq1	-0.298474 (0.12616) [-2.36579]	-0.197518 (0.11965) [-1.65078]	-0.582213 (0.16770) [-3.47171]
D(LOG(TOTAL(-1)))	-0.621929 (0.13554) [-4.58865]	0.086528 (0.12854) [ 0.67315]	0.636452 (0.18016) [ 3.53267]
D(LOG(TOTAL(-2)))	-0.357989 (0.10815) [-3.31011]	0.126523 (0.10257) [ 1.23354]	0.324207 (0.14376) [ 2.25521]
D(LOG(PO(-1)))	0.224945 (0.13552) [ 1.65983]	-0.210729 (0.12853) [-1.63954]	-0.344436 (0.18014) [-1.91200]
D(LOG(PO(-2)))	-0.052451 (0.13717) [-0.38237]	-0.349319 (0.13009) [-2.68513]	-0.013095 (0.18234) [-0.07182]
D(LOG(IBRX(-1)))	0.098687 (0.07625) [ 1.29426]	0.015194 (0.07231) [ 0.21011]	0.090309 (0.10136) [ 0.89101]
D(LOG(IBRX(-2)))	0.117069 (0.07389) [ 1.58443]	0.068756 (0.07007) [ 0.98119]	-0.119164 (0.09821) [-1.21330]
C	-0.017729 (0.02370) [-0.74817]	0.051849 (0.02247) [ 2.30713]	-0.048511 (0.03150) [-1.54010]

@SEAS(1)	0.112280 (0.03907) [ 2.87393]	-0.034712 (0.03705) [-0.93683]	0.039817 (0.05193) [ 0.76672]
@SEAS(2)	-0.154593 (0.05052) [-3.05977]	-0.129442 (0.04792) [-2.70140]	0.140031 (0.06716) [ 2.08505]
@SEAS(3)	-0.118770 (0.04506) [-2.63585]	0.082158 (0.04273) [ 1.92255]	0.200683 (0.05990) [ 3.35056]
@SEAS(4)	0.037757 (0.03753) [ 1.00606]	-0.050453 (0.03559) [-1.41752]	0.112134 (0.04989) [ 2.24783]
@SEAS(6)	0.003990 (0.03502) [ 0.11392]	-0.106773 (0.03321) [-3.21466]	0.060022 (0.04655) [ 1.28934]
@SEAS(7)	0.068311 (0.02827) [ 2.41623]	-0.026149 (0.02681) [-0.97524]	0.057478 (0.03758) [ 1.52948]
@SEAS(8)	-0.013317 (0.02890) [-0.46077]	-0.046230 (0.02741) [-1.68660]	0.026139 (0.03842) [ 0.68040]
@SEAS(9)	-0.028328 (0.03113) [-0.91011]	-0.053314 (0.02952) [-1.80603]	0.048526 (0.04137) [ 1.17283]
@SEAS(10)	0.116029 (0.02714) [ 4.27457]	0.003403 (0.02574) [ 0.13217]	0.034963 (0.03608) [ 0.96901]
@SEAS(11)	-0.002095 (0.02817) [-0.07439]	-0.081039 (0.02671) [-3.03351]	0.024782 (0.03744) [ 0.66187]
@SEAS(12)	0.228533 (0.02877) [ 7.94290]	-0.174400 (0.02729) [-6.39130]	0.076623 (0.03825) [ 2.00347]
D2009M01	-0.011228 (0.04801) [-0.23384]	-0.007238 (0.04554) [-0.15896]	-0.294646 (0.06382) [-4.61669]
D2009M11	0.254043 (0.04803) [ 5.28951]	0.016334 (0.04555) [ 0.35861]	-0.001056 (0.06384) [-0.01653]
D2002M09	0.170065 (0.04715) [ 3.60709]	-0.028894 (0.04471) [-0.64620]	-0.027996 (0.06267) [-0.44672]
R-squared	0.928087	0.741619	0.482515
Adj. R-squared	0.906817	0.665196	0.329456
Sum sq. resids	0.120141	0.108060	0.212277
S.E. equation	0.041135	0.039012	0.054679
F-statistic	43.63364	9.704181	3.152481

Log likelihood	177.3424	182.2702	150.8732
Akaike AIC	-3.340696	-3.446672	-2.771467
Schwarz SC	-2.741586	-2.847562	-2.172357
Mean dependent	0.008005	0.000790	0.019487
S.D. dependent	0.134756	0.067423	0.066774

Determinant resid covariance (dof adj.)	7.00E-09
Determinant resid covariance	3.11E-09
Log likelihood	514.9221
Akaike information criterion	-9.568218
Schwarz criterion	-7.661960

<b>Total</b>	<b>em R\$ milhões</b>	
--------------	-----------------------	--

Período: 01/02 a 12/09

<b>Modelagem VEC Total x PO x IBrX</b>
--

	Realizado	Previsão
2010M01	72474	73889
2010M02	52718	53197
2010M03	58210	57836
2010M04	69074	72601
2010M05	59280	63391
2010M06	59639	59755
2010M07	65915	67690
2010M08	60804	62473
2010M09	61206	65999
2010M10	71288	72856
2010M11	63452	64523
2010M12	85802	86327

Tabela 69 – Combinação – média aritmética das previsões geradas pelos modelos Indicadores e VEC – Total das Receitas

<b>Total</b>	<b>em R\$ milhões</b>	
--------------	-----------------------	--

Período: 01/02 a 12/09

<b>Combinação - Média Aritmética - Indicadores x VEC</b>
--

	Realizado	Previsão
2010M01	72474	69773
2010M02	52718	51439
2010M03	58210	56287
2010M04	69074	65993
2010M05	59280	57932
2010M06	59639	57891
2010M07	65915	66008
2010M08	60804	60822
2010M09	61206	63460
2010M10	71288	69610
2010M11	63452	62587
2010M12	85802	81297

## ANEXO D

Tabela 70 - Previsões dos modelos selecionados para Tributos sobre a Renda - 2011

Metodologia	Modelo	Previsão			
		RMSE	MAE	MAPE (%)	Anual (%)
ARIMA	$(2,1,0)(1,1,1)_6 + 6 \text{ do}^*$	56.99	39.23	8.16	-2.49
VAR	Renda x PO x IBrX	77.92	53.48	10.63	-5.85
Combinação	Média Aritmética	60.93	41.10	8.40	-2.67

Tabela 71 - Previsões dos modelos selecionados para Tributos sobre a Folha de Salários - 2011

Metodologia	Modelo	Previsão Anual			
		RMSE	MAE	MAPE (%)	Anual (%)
ARIMA	$(0,1,1)(1,1,0)_{12} + \text{do}^*$	11.01	8.92	2.85	0.33
VEC	Folha x PO x MS	8.65	7.12	2.26	1.34
Combinação	Média Aritmética	9.48	7.90	2.51	0.83

Tabela 72 - Previsões dos modelos selecionados para Tributos sobre Bens e Serviços - 2011

Metodologia	Modelo	Previsão Anual			
		RMSE	MAE	MAPE (%)	Anual (%)
ARIMA	$(0,1,0)(0,1,1)_{12}$	13.71	10.60	3.37	2.01
VEC	BS x IBC x Selic	12.89	11.34	3.70	-2.13
Combinação	Média Aritmética	11.61	9.00	2.90	-0.06

Tabela 73 - Previsões dos modelos selecionados para Tributos sobre Transações Financeiras - 2011

Metodologia	Modelo	Previsão Anual			
		RMSE	MAE	MAPE (%)	Anual (%)
ARIMA	$(2,1,0) + 4 \text{ dos}$	25.60	19.83	8.57	1.04
VEC	TF x PO	25.38	21.60	8.90	-2.53
Combinação	Média Aritmética	24.92	20.15	8.50	-0.74

Tabela 74 - Previsões dos modelos selecionados para Outros Tributos - 2011

Metodologia	Modelo	Previsão Anual			
		RMSE	MAE	MAPE (%)	Anual (%)
ARIMA	$(0,0,1)(0,1,1)_3$	145.61	119.16	17.74	-16.40
VEC	Outros x IVV	126.07	110.80	16.62	-5.25
Combinação	Média Aritmética	119.31	103.43	15.46	-10.82

Tabela 75 - Previsões dos modelos selecionados para Total dos Tributos - 2011

Metodologia	Modelo	Previsão Anual			
		RMSE	MAE	MAPE (%)	Anual (%)
ARIMA	$(2,1,0)(1,1,0)_{12} + 2 \text{ dos}$	23.92	22.04	6.30	4.02
VEC	Total x PO x IBrX	22.69	17.98	2.64	-4.87
Combinação	Soma Ótima	22.03	18.58	5.20	-4.78

Tabela 76 - Quadro comparativo das previsões efetuadas para 2010 e 2011 (pelo RMSE)

	ARIMA	VAR	Combinação
Renda - 2010	45	27	30
Renda - 2011	57	78	60
Folha - 2010	9	4	6
Folha - 2011	11	9	9
Bens e Serviços - 2010	5	5	5
Bens e Serviços - 2011	13	13	11
Transações Fin. - 2010	25	20	22
Transações Fin. - 2011	26	25	25
Outros Total - 2010	130	117	304
Outros Total - 2011	146	126	119
Total - 2010	25	16	14
Total - 2011	24	23	22

## ANEXO E

Tabela 77 – Testes de quebra estrutural para a série Tributos sobre a Renda

Chow Breakpoint Test: 2003M01

Null Hypothesis: No breaks at specified breakpoints

Varying regressors: All equation variables

Equation Sample: 2002M01 2010M12

F-statistic	5.975067	Prob. F(1,106)	0.0162
Log likelihood ratio	5.922411	Prob. Chi-Square(1)	0.0149
Wald Statistic	5.975067	Prob. Chi-Square(1)	0.0145

Quandt-Andrews unknown breakpoint test

Null Hypothesis: No breakpoints within 5% trimmed data

Varying regressors: All equation variables

Equation Sample: 2002M01 2010M12

Test Sample: 2002M07 2010M07

Number of breaks compared: 97

Statistic	Value	Prob.
Maximum LR F-statistic (2009M02)	8.252679	0.0911
Maximum Wald F-statistic (2009M02)	8.252679	0.0911
Exp LR F-statistic	2.105519	0.0473
Exp Wald F-statistic	2.105519	0.0473
Ave LR F-statistic	2.795788	0.0424
Ave Wald F-statistic	2.795788	0.0424

Note: probabilities calculated using Hansen's (1997) method

Tabela 78 – Testes de quebra estrutural para a série Tributos sobre a Folha de Salários

Chow Breakpoint Test: 2004M01

Null Hypothesis: No breaks at specified breakpoints

Varying regressors: All equation variables

Equation Sample: 2002M01 2010M12

F-statistic	1.513125	Prob. F(1,106)	0.2214
Log likelihood ratio	1.530775	Prob. Chi-Square(1)	0.2160
Wald Statistic	1.513125	Prob. Chi-Square(1)	0.2187

Quandt-Andrews unknown breakpoint test

Null Hypothesis: No breakpoints within 15% trimmed data

Varying regressors: All equation variables

Equation Sample: 2002M01 2010M12

Test Sample: 2003M06 2009M08

Number of breaks compared: 75

Statistic	Value	Prob.
Maximum LR F-statistic (2009M08)	2.275749	0.7179
Maximum Wald F-statistic (2009M08)	2.275749	0.7179
Exp LR F-statistic	0.356394	0.5755
Exp Wald F-statistic	0.356394	0.5755
Ave LR F-statistic	0.644615	0.5354
Ave Wald F-statistic	0.644615	0.5354

Note: probabilities calculated using Hansen's (1997) method

Tabela 79 - Testes de quebra estrutural para a série Tributos sobre Bens e Serviços

Chow Breakpoint Test: 2004M01

Null Hypothesis: No breaks at specified breakpoints

Varying regressors: All equation variables

Equation Sample: 2002M01 2010M12

F-statistic	21.44762	Prob. F(1,106)	0.0000
Log likelihood ratio	19.90077	Prob. Chi-Square(1)	0.0000
Wald Statistic	21.44762	Prob. Chi-Square(1)	0.0000

Quandt-Andrews unknown breakpoint test

Null Hypothesis: No breakpoints within 5% trimmed data

Varying regressors: All equation variables

Equation Sample: 2002M01 2010M12

Test Sample: 2002M07 2010M07

Number of breaks compared: 97

Statistic	Value	Prob.
Maximum LR F-statistic (2008M12)	66.21978	0.0000
Maximum Wald F-statistic (2008M12)	66.21978	0.0000
Exp LR F-statistic	29.00826	0.0000
Exp Wald F-statistic	29.00826	0.0000
Ave LR F-statistic	26.90247	0.0000
Ave Wald F-statistic	26.90247	0.0000

Note: probabilities calculated using Hansen's (1997) method

Tabela 80 - Testes de quebra estrutural para a série Tributos sobre Transações Financeiras

Chow Breakpoint Test: 2002M01

Null Hypothesis: No breaks at specified breakpoints

Varying regressors: All equation variables

Equation Sample: 2001M01 2010M12

F-statistic	0.929369	Prob. F(1,118)	0.3370
Log likelihood ratio	0.941418	Prob. Chi-Square(1)	0.3319
Wald Statistic	0.929369	Prob. Chi-Square(1)	0.3350

Quandt-Andrews unknown breakpoint test

Null Hypothesis: No breakpoints within 5% trimmed data

Varying regressors: All equation variables

Equation Sample: 2001M01 2010M12

Test Sample: 2001M07 2010M07

Number of breaks compared: 109

Statistic	Value	Prob.
Maximum LR F-statistic (2008M01)	734.0868	0.0000
Maximum Wald F-statistic (2008M01)	734.0868	0.0000
Exp LR F-statistic	362.3521	0.0000
Exp Wald F-statistic	362.3521	0.0000
Ave LR F-statistic	92.69854	0.0000
Ave Wald F-statistic	92.69854	0.0000

Note: probabilities calculated using Hansen's (1997) method

Tabela 81 - Testes de quebra estrutural para a série Outros Tributos

Chow Breakpoint Test: 2003M01

Null Hypothesis: No breaks at specified breakpoints

Varying regressors: All equation variables

Equation Sample: 2002M01 2010M12

F-statistic	6.847678	Prob. F(1,106)	0.0102
Log likelihood ratio	6.760782	Prob. Chi-Square(1)	0.0093
Wald Statistic	6.847678	Prob. Chi-Square(1)	0.0089

Quandt-Andrews unknown breakpoint test  
 Null Hypothesis: No breakpoints within 5% trimmed data  
 Varying regressors: All equation variables  
 Equation Sample: 2002M01 2010M12  
 Test Sample: 2002M07 2010M07  
 Number of breaks compared: 97

Statistic	Value	Prob.
Maximum LR F-statistic (2005M02)	39.63605	0.0000
Maximum Wald F-statistic (2005M02)	39.63605	0.0000
Exp LR F-statistic	15.80432	0.0000
Exp Wald F-statistic	15.80432	0.0000
Ave LR F-statistic	13.93344	0.0000
Ave Wald F-statistic	13.93344	0.0000

Note: probabilities calculated using Hansen's (1997) method

Tabela 82 - Testes de quebra estrutural para a série Total das Receitas

Chow Breakpoint Test: 2002M01  
 Null Hypothesis: No breaks at specified breakpoints  
 Varying regressors: All equation variables  
 Equation Sample: 2001M01 2010M12

F-statistic	13.33894	Prob. F(1,118)	0.0004
Log likelihood ratio	12.85160	Prob. Chi-Square(1)	0.0003
Wald Statistic	13.33894	Prob. Chi-Square(1)	0.0003

Quandt-Andrews unknown breakpoint test  
 Null Hypothesis: No breakpoints within 5% trimmed data  
 Varying regressors: All equation variables  
 Equation Sample: 2001M01 2010M12  
 Test Sample: 2001M07 2010M07  
 Number of breaks compared: 109

Statistic	Value	Prob.
Maximum LR F-statistic (2003M02)	34.05953	0.0000
Maximum Wald F-statistic (2003M02)	34.05953	0.0000
Exp LR F-statistic	13.01918	0.0000
Exp Wald F-statistic	13.01918	0.0000
Ave LR F-statistic	12.04102	0.0000
Ave Wald F-statistic	12.04102	0.0000

Note: probabilities calculated using Hansen's (1997) method