

# **Métodos econométricos aplicados à previsão da arrecadação de receitas federais**

3º Lugar

**CELSO VILELA CHAVES CAMPOS\***

\* Engenheiro Eletricista – PUC/MG  
Mestre em Economia – FEA/USP  
Auditor-Fiscal da Receita Federal do Brasil  
Delegacia da Receita Federal – Ribeirão Preto-SP



# **MÉTODOS ECONOMÉTRICOS APLICADOS À PREVISÃO DA ARRECADAÇÃO DE RECEITAS FEDERAIS**

---

## **RESUMO**

### **1 OBJETIVOS BÁSICOS**

O objetivo principal do presente trabalho é oferecer métodos alternativos de previsão da arrecadação tributária federal, baseados em metodologias de séries temporais, inclusive com a utilização de variáveis explicativas, que reflitam a influência do cenário macroeconômico na arrecadação tributária, com o intuito de melhorar a acurácia da previsão da arrecadação.

### **2 METODOLOGIA UTILIZADA**

São aplicadas as metodologias de modelos dinâmicos univariados, multivariados, quais sejam, função de transferência, autoregressão vetorial (VAR), VAR com correção de erro (VEC), equações simultâneas e de modelos estruturais. O trabalho tem abrangência regional e limita-se à análise de três séries mensais da arrecadação, relativas ao Imposto de Importação, Imposto sobre a Renda da Pessoa Jurídica – Demais Obrigadas à Apuração do Lucro Real e Contribuição para o Financiamento da Seguridade Social – Demais Empresas, no âmbito da jurisdição do Estado de São Paulo, no período de 2000 a 2007.

A escolha dos tributos analisados levou em conta a participação de cada um na arrecadação total do estado para o ano de 2007, bem como o âmbito de incidência do tributo, que, no presente estudo é, respectivamente, o setor externo, o lucro e o faturamento das empresas.

Os resultados das previsões dos modelos anteriormente citados são comparados entre si, com a modelagem ARIMA, também desenvolvida para servir de base de comparação, e com o método dos indicadores, atualmente utilizado pela Secretaria da Receita Federal do Brasil (RFB) para previsão anual da arrecadação tributária, por meio da raiz do erro médio quadrático de previsão (RMSE).

Utilizam-se como variáveis explicativas a Taxa de Câmbio Real, o PIB mensal, o Índice de Produção Industrial da Indústria Geral de São Paulo e o Índice de Vendas no Varejo de São Paulo.

### **3 ADEQUAÇÃO DO TRABALHO AOS CRITÉRIOS DE JULGAMENTO**

#### **3.1 Relação custo versus benefícios**

O custo para implementação dessa nova forma de previsão da arrecadação é muito baixo e constitui-se, basicamente, do custo do *software* necessário para as estimações, bem como do treinamento de uma pequena equipe na utilização do *software* e em aspectos básicos de séries temporais. O trabalho de previsão poderia ser coordenado por profissional experiente na área, já existente na RFB.

Os benefícios de uma previsão mais acurada da arrecadação federal são enormes, na medida em que uma das finalidades da RFB, conforme consta de seu Regimento Interno, é a realização da previsão, o acompanhamento, a análise e o controle das receitas sob sua administração, bem como a coordenação e a consolidação das previsões das demais receitas federais, para subsidiar a elaboração da proposta orçamentária da União.

Como se sabe, a estimativa de arrecadação das receitas públicas é condição preliminar para a elaboração do orçamento anual do governo, como forma de orientar e limitar o montante dos gastos públicos.

A informação sobre o montante de receitas disponíveis a cargo do administrador tributário é de fundamental importância, uma vez que funciona como limitador na fixação do montante dos gastos públicos, contribuindo, assim, para o necessário equilíbrio nas contas públicas.

Ademais, as atividades do acompanhamento, da análise e do controle das receitas federais dependem de uma previsão benfeita, com os objetivos de: i) comparar o valor efetivamente arrecadado com o valor previsto, para verificar a eficácia do método adotado; ii) acompanhar as alterações do cenário macroeconômico e analisar sua influência sobre os valores arrecadados, identificando possíveis causas dos desvios do valor arrecadado em relação ao previsto; iii) auxiliar a atividade de fiscalização, entre outros.

### ***3.2 Aumento de produtividade***

O método dos indicadores, atualmente utilizado pela RFB para previsão da arrecadação das receitas federais, embora não seja complexo do ponto de vista teórico, é extremamente trabalhoso, em função, principalmente, do elevado número de tributos a serem previstos, índices de preço quantidade envolvidos, da dificuldade de se obter um índice de alterações na legislação e da necessidade de se excluir as arrecadações atípicas ocorridas. Ademais, o método não tem levado a previsões precisas. Assim, o método proposto no presente trabalho não só melhora a acurácia da previsão, como também aumenta a produtividade, na medida em que necessita de um número bem inferior de servidores ao atualmente empregado nessa atividade.

### ***3.3 Viabilidade de implementação***

A viabilidade técnica justifica-se pelo fato de que a RFB já dispõe de servidores aptos a coordenar e a implementar a metodologia proposta. Quanto ao aspecto político, depende de decisão, por parte da administração superior do órgão, para implementação.

Cabe ressaltar que a implementação da metodologia proposta poderia se dar de forma gradativa, com os dois métodos de previsão (método dos indicadores e método econométrico) coexistindo por um

período de tempo, a fim de se verificar a eficiência e a eficácia da nova metodologia de maneira cabal.

Ademais, conforme enfatizado na literatura, as técnicas utilizadas dependem da natureza do tributo e não há padrão no tempo sobre o qual se possam fundar todas as previsões. Assim, pode acontecer de, em determinadas situações específicas, o método dos indicadores ser mais eficiente que os métodos econométricos propostos, como, por exemplo, quando de uma mudança drástica na legislação de determinado tributo. Neste caso, a coexistência das duas metodologias seria a alternativa mais viável.

### ***3.4 Valorização do servidor***

Previsões da arrecadação federal mais acuradas com certeza contribuirão para a valorização dos servidores envolvidos, com reflexos na capacitação, motivação e crescimento pessoal.

### ***3.5 Melhoria da qualidade do serviço prestado***

Além da valorização dos servidores envolvidos, haverá impactos positivos para o órgão em função da melhoria da qualidade do serviço prestado e do reconhecimento, pela sociedade, da instituição como entidade de excelência no serviço público.

O trabalho pode ser vinculado aos seguintes objetivos gerais do Programa de Trabalho da RFB: 1) subsidiar a formulação da política tributária e de comércio exterior; 2) promover a integração da RFB com órgãos de Estado e organismos nacionais e internacionais; 3) fortalecer a imagem institucional da RFB e promover a conscientização tributária do cidadão; 4) promover o atendimento de excelência ao contribuinte; 5) otimizar o controle e a cobrança do crédito tributário; 6) aprimorar a qualidade e a produtividade do trabalho fiscal; 7) implementar gestão de excelência na RFB.

### **3.6 Promoção da justiça fiscal e social dos tributos**

Destaque-se que a Lei de Responsabilidade Fiscal, que estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal com vistas ao equilíbrio das contas públicas, explicitamente dispõe que a previsão de receitas é um dos requisitos essenciais da responsabilidade na gestão fiscal.

Assim, uma previsão mais acurada contribui para elevar a responsabilidade na gestão fiscal, contribuindo para maior transparência do processo orçamentário e facilitando o acompanhamento e o controle pela sociedade, com reflexo no exercício da cidadania.

## **4 PRINCIPAIS CONCLUSÕES**

Como conclusão final, a aplicação dos métodos desenvolvidos neste trabalho, considerados os melhores modelos de cada série, levou à redução média do RMSE de 42% em relação ao erro cometido pelo método dos indicadores e de 35% em relação à modelagem ARIMA, além da drástica redução média do erro anual de previsão. A técnica foi aplicada a três tributos que são representativos do sistema tributário e, a partir de agora, seria interessante aplicá-la aos demais tributos administrados pela RFB. Como toda técnica, deve ser aprimorada constantemente, em especial com relação à tentativa de escolha de outras variáveis explicativas a serem incluídas nos sistemas multivariados.

A utilização de metodologias de séries temporais para a previsão da arrecadação de receitas federais mostrou ser uma alternativa viável ao método dos indicadores, contribuindo para previsões mais precisas e tornando-se, assim, ferramenta segura de apoio para a tomada de decisões dos gestores.





# MÉTODOS ECONOMÉTRICOS APLICADOS À PREVISÃO DA ARRECAÇÃO DE RECEITAS FEDERAIS

---

## 1 INTRODUÇÃO

Dentre as missões da Secretaria da Receita Federal do Brasil (RFB), destaca-se a de exercer a administração tributária em benefício da sociedade. Uma das formas de se garantir o sucesso dessa missão se faz por meio de uma boa previsão da arrecadação das receitas federais. Entre as finalidades da RFB, conforme dispõe seu Regimento Interno, consta:

*Realizar a previsão, o acompanhamento, a análise e o controle das receitas sob sua administração, bem como coordenar e consolidar as previsões das demais receitas federais, para subsidiar a elaboração da proposta orçamentária da União. (inciso VIII do art. 1º do anexo de seu Regimento Interno, aprovado pela Portaria do Ministro de Estado da Fazenda – MF nº 125, de 4 de março de 2009, publicada no D.O.U. de 6 de março de 2009.)*

Pode-se considerar que subsidiar a elaboração da proposta orçamentária da União é um dos principais objetivos da previsão das receitas federais, principalmente após o advento da Lei Complementar nº 101, de 05/05/2000, conhecida como Lei de Responsabilidade Fiscal, que estabelece, para todos os entes federados (União, estados e municípios), normas de finanças públicas voltadas para a responsabilidade

na gestão fiscal com vistas ao equilíbrio das contas públicas. Nesse sentido, destaca explicitamente a previsão de receitas como um dos requisitos essenciais da responsabilidade na gestão fiscal, sendo vedada a transferência voluntária de receitas para o ente federado que não observar esse mandamento.

A estimativa de arrecadação das receitas públicas é condição preliminar para a elaboração do orçamento anual do governo, como forma de orientar e limitar o montante dos gastos públicos. A informação sobre o montante de receitas disponíveis a cargo do administrador tributário é de fundamental importância, uma vez que funciona como limitador na fixação do montante dos gastos públicos, contribuindo, assim, para o necessário equilíbrio nas contas públicas.

Ademais, as atividades do acompanhamento, da análise e do controle das receitas federais dependem de uma previsão benfeita, com o objetivo de: i) comparar o valor efetivamente arrecadado com o valor previsto, para verificar a eficácia do método adotado; ii) acompanhar as alterações do cenário macroeconômico e analisar sua influência sobre os valores arrecadados, identificando possíveis causas dos desvios do valor arrecadado em relação ao previsto; iii) auxiliar a atividade de fiscalização, entre outros.

A RFB utiliza, como se verá com mais detalhes em 2.2, um método de previsão das receitas federais denominado pelo órgão de “método dos indicadores”, em que, basicamente, a arrecadação prevista é igual à arrecadação do mesmo período do ano anterior multiplicada por índices que reflitam as variações de preço, de quantidade e da legislação no período. A previsão é feita em nível regional, excluindo-se eventuais arrecadações atípicas, e o rateio entre as unidades da RFB é efetuado com base na participação de cada unidade na arrecadação realizada no ano anterior, levando-se em consideração, ainda, eventuais alterações de jurisdição de grandes empresas. A soma das previsões efetuadas por cada uma das regiões fiscais compõe a previsão nacional de arrecadação.

Esse método, embora não seja complexo do ponto de vista teórico, é extremamente trabalhoso, em função, principalmente, do número de tributos a serem previstos (14 tributos que, com as subclassificações,

dão origem a 55 séries); da quantidade de índices de preço e quantidade envolvidos; da dificuldade de se obter um índice de alterações na legislação e da necessidade de se excluir as arrecadações atípicas ocorridas. Ademais, o método não tem levado a previsões precisas, conforme se pode observar na Tabela 1, em que é apresentado o erro de previsão de alguns tributos, dentre os quais se encontram os que são analisados neste trabalho, no período de 2001 a 2007, medido pelo erro percentual absoluto médio (MAPE<sup>1</sup>), que considera o erro relativo de cada previsão.

Tabela 1. Método dos indicadores – Erro percentual absoluto médio da previsão – MAPE – 2001/2007

SÉRIE	2001	2002	2003	2004	2005	2006	2007	Média
II	19,2%	20,0%	16,1%	29,5%	28,4%	11,1%	5,7%	18,6%
IPI_OP	7,8%	8,5%	22,6%	10,9%	9,0%	5,3%	9,0%	10,5%
IRPJ_DOLR	27,2%	31,8%	21,4%	34,1%	20,8%	14,1%	19,3%	24,1%
IRPJ_NOLR	13,7%	4,8%	7,6%	12,2%	8,8%	4,4%	8,5%	8,6%
IRRF_T	6,8%	8,2%	5,5%	5,6%	7,5%	12,5%	3,9%	7,1%
COFINS_DE	4,0%	7,2%	7,8%	8,4%	4,1%	5,5%	5,6%	6,1%
PIS_DE	7,1%	9,0%	14,1%	5,3%	4,4%	5,1%	6,7%	7,4%
CSLL_DE	12,6%	21,5%	18,1%	23,1%	13,0%	10,2%	13,5%	16,0%
RA	4,8%	13,9%	4,0%	8,7%	5,7%	7,1%	10,3%	7,8%

Fonte: Elaboração do autor

Notas: II – Imposto de Importação.

IPI\_OP – Imposto sobre Produtos Industrializados – Outros Produtos.

IRPJ\_DOLR, IRPJ\_NOLR – Imposto sobre a Renda da Pessoa Jurídica – Demais Obrigadas ao Lucro Real e Não Obrigadas ao Lucro Real.

IRRF\_T – Imposto sobre a Renda Retido na Fonte – Rendimentos do Trabalho.

COFINS\_DE – Contribuição para o Financiamento da Seguridade Social – Demais Empresas.

PIS\_DE – Contribuição para o Programa de Integração Social – Demais Empresas.

CSLL\_DE – Contribuição Social sobre o Lucro Líquido – Demais Empresas.

RA – Receita Administrada pela RFB.

1 A definição do MAPE encontra-se no capítulo 3.

Conforme se verifica, há uma tendência de diminuição dos erros de previsão medidos pelo MAPE ao longo dos anos, provavelmente devido à maior estabilidade econômica dos últimos anos. Deve-se ressaltar que o método dos indicadores funciona melhor quanto mais estável for o ambiente econômico. Contudo, os erros de previsão são bastante elevados e variam muito de um ano para outro.

Este trabalho tem abrangência regional e limita-se à análise e à previsão de três séries mensais da arrecadação, relativas ao Imposto de Importação, Imposto sobre a Renda da Pessoa Jurídica e Contribuição para o Financiamento da Seguridade Social, no âmbito da jurisdição do Estado de São Paulo, no período de 2000 a 2007.

Portanto, o objetivo principal do trabalho é oferecer métodos alternativos de previsão da arrecadação tributária federal, baseados em metodologias de séries temporais, inclusive com a utilização de variáveis explicativas, que reflitam a influência do cenário macroeconômico na arrecadação tributária, com o intuito de se melhorar a acurácia da previsão da arrecadação.

## **2 MARCO TEÓRICO<sup>2</sup>**

### ***2.1 Receitas dos impostos***

Um modelo simples para a receita tributária em um sistema com imposto de renda (pago tanto pelos produtores quanto pelos trabalhadores) assume que o imposto total arrecadado é uma fração da renda real sujeita ao imposto. Assim, assumindo-se constante a taxa média do imposto, pode-se considerar que a receita tributária total é proporcional à renda sujeita à incidência do imposto.

Basicamente, dado o sistema tributário vigente, um aumento da

---

2 Devido ao espaço disponível, são feitas breves considerações teóricas. Para aprofundamento dos temas, sugere-se consultar os diversos livros que apresentam tratamentos teóricos completos dos modelos, com destaque para Barro (1993), Musgrave e Musgrave (1980), Romer (2006), Enders (2004), Box, Jenkins e Reinsel (2008), Hamilton (1994), Brockwell e Davis (2002), Maddala e Kim (1998), Granger e Newbold (1977) e Harvey (1989).

renda eleva as receitas tributárias, que, por sua vez, no período seguinte, diminuem a renda disponível e, conseqüentemente, o produto. Assim, fica difícil estabelecer uma relação de causa e efeito entre as duas variáveis.

Portanto, na estimação de modelos para previsão da arrecadação do IRPJ, utiliza-se como variável explicativa o PIB (renda) ou alguma variável que tenha relação com o PIB,<sup>3</sup> já que o IRPJ, no caso da série IRPJ\_DOLR, analisada neste trabalho, é um tributo que depende da apuração do lucro real pelas empresas, que, em última instância, está ligado ao nível da atividade econômica, embora esta relação possa não ser tão direta. No caso da Cofins, cuja série tratada é a COFINS\_DE, por se tratar de tributo dependente do faturamento das empresas, que, por sua vez, está diretamente ligado ao nível de atividade econômica, são utilizadas as mesmas variáveis explicativas.

A receita do Imposto de Importação (II) é função da demanda de bens importados, da taxa real de câmbio<sup>4</sup> e da alíquota média dos tributos devidos na importação. A demanda de bens importados, por sua vez, é função da renda doméstica, da taxa real de câmbio e dos tributos devidos na importação. Quanto maior a renda doméstica, maior a demanda por bens, sejam domésticos ou estrangeiros. No que diz respeito à taxa real de câmbio, quanto mais caros os bens domésticos relativamente aos bens estrangeiros, maior a demanda por bens importados. Assim, uma diminuição da taxa real de câmbio conduz ao aumento das importações. No caso dos tributos devidos na importação, quanto maior for o nível de tributação, menor será a demanda por bens importados.

O resultado final sobre a receita do II, decorrente de uma variação no câmbio, depende basicamente das sensibilidades da demanda de importados em relação à taxa de câmbio real (efeito negativo) e em relação à renda (efeito positivo), que, por sua vez, depende positiva-

---

3 Chamada *proxy* do PIB.

4 Na realidade, a receita do II é função da taxa nominal de câmbio. Porém, considerando que: i) a relação entre a taxa nominal e a real é expressa pela relação entre os níveis de preços estrangeiro e doméstico; e ii) a demanda de bens importados é função da taxa real de câmbio; tratar a receita do II função da taxa real de câmbio simplifica a análise e não altera o conceito básico que se pretende mostrar.

mente da taxa de câmbio real. Assim, um aumento da taxa de câmbio tem efeito positivo na renda e negativo na demanda de importados. Porém, o aumento na renda tem efeito positivo nesta última. Portanto, o efeito na demanda de importados é incerto e depende das elasticidades envolvidas. Já a receita do II é influenciada positivamente tanto pela taxa de câmbio quanto pela demanda de importados e, devido à incerteza desta última, não há como saber o efeito final sobre a receita do II, decorrente de um aumento na taxa de câmbio, a menos que se tenha conhecimento das elasticidades envolvidas.

A análise anterior pretende estabelecer o resultado final do processo de alteração da receita tributária do Imposto de Importação em função de uma alteração na taxa de câmbio. Entretanto, considerando-se os efeitos dinâmicos da variação da taxa de câmbio, num primeiro momento, dada uma demanda por bens importados, a diminuição da taxa de câmbio (apreciação) leva à diminuição do II em função da diminuição da base de incidência do imposto, o que pode ser chamado de efeito preço, que neste caso é negativo. Porém, à medida que o tempo passa, a mudança nos preços relativos altera o comércio exterior. A menor taxa de câmbio leva ao aumento da demanda de bens importados, que ficam mais baratos em relação aos nacionais, com o conseqüente aumento do II – tudo isso mantida inalterada a alíquota do imposto, o que pode ser chamado de efeito quantidade, neste caso positivo. Portanto, o efeito final sobre a receita do II vai depender de qual efeito irá prevalecer. Assim, à primeira vista, renda (PIB) e taxa de câmbio real parecem ser boas variáveis explicativas do comportamento do Imposto de Importação, embora não se possa estabelecer uma relação de causa e efeito entre essas variáveis.

## **2.2 Método dos indicadores – análise econométrica**

Segundo Santana (2004), o método dos indicadores, atualmente utilizado pela RFB na elaboração da previsão dos tributos federais, consiste na multiplicação da arrecadação do período anterior (base) por: i) um índice de preço – Efeito-Preço, que representa a variação inflacionária a que está sujeito o fato econômico gerador da arrecadação; ii) um

índice de quantidade – Efeito-Quantidade, que representa a variação real desse fato gerador; iii) um índice de legislação – Efeito-Legislação, que representa o efeito causado na arrecadação por modificações na legislação tributária; e iv) outros índices – Efeito-Residual, que representam quaisquer influências na arrecadação tributária.

Esse método pode ser resumido na seguinte fórmula:

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

em que  $A_t$  é a arrecadação prevista para determinado período do ano  $t$ ;  $A_{t-1}$  é a arrecadação efetiva do mesmo período do ano  $t-1$ ;  $\Delta P$  é a variação percentual do indicador de preços;  $\Delta Q$  é a variação percentual do indicador de quantidades;  $\Delta L$  é a variação percentual decorrente de alterações na legislação (geralmente variação de alíquotas); e  $\Delta U$  é a variação percentual de qualquer outro indicador que tenha influência na arrecadação e não possa ser enquadrado nos indicadores básicos anteriores. Assim, a qualidade da previsão depende, principalmente, da obtenção de bons indicadores de preço e quantidade específicos para cada tributo.

Partindo-se da equação acima, aplicando-se o logaritmo natural ( $\ln$ ) a ambos os lados da equação e lembrando-se, ainda, que para pequenas variações percentuais nos índices,  $\ln(1 + \Delta I\%) \approx \Delta I\%$ , tem-se:

$$\ln A_t = \ln A_{t-1} + \ln(1 + \Delta P_t) + \ln(1 + \Delta Q_t) + \ln(1 + \Delta L_t) + \ln(1 + \Delta U_t),$$

$$\ln A_t \approx \ln A_{t-1} + \Delta P_t + \Delta Q_t + \Delta L_t + \Delta U_t \approx \ln A_{t-1} + \Delta I_t,$$

em que  $\Delta I_t = \Delta P_t + \Delta Q_t + \Delta L_t + \Delta U_t$ . Assim:

$$\ln A_t \approx \ln A_{t-1} + \Delta I_t. \text{ Chamando } \ln A_t = a_t \text{ e } \Delta I_t = i_t,$$

$$a_t \approx a_{t-1} + i_t. \tag{1}$$

A equação anterior assemelha-se a um processo autorregressivo de ordem um,  $AR(1)$ , que, de acordo com Brockwell e Davis (2002), pode ser representado por:

$$X_t = \phi X_{t-1} + Z_t, \tag{2}$$

em que  $\{Z_t\} \sim WN(0, \sigma^2)$ ,  $|\phi| < 1$ , e  $Z_t$  é não correlacionado com  $X_s$ , para todo  $s < t$ . Ou seja, no caso da equação (1),  $f = 1$ , e o choque  $Z_t$  é representado pela soma das variações percentuais nos diversos índices presentes no método, podendo assumir valores tanto positivos quanto negativos. Este caso, em que  $\phi = 1$ , representa um processo não estacionário, com presença de raiz unitária, chamado processo *random walk* (RW), ou passeio aleatório, cuja solução, considerando-se que  $a_0$  é a condição inicial, é dada por:

$$a_t = a_0 + \sum_{j=1}^t i_j.$$

Ou seja, trata-se de um processo totalmente aleatório, com média constante, dada por  $a_0$ . Ademais, o choque  $i_t$  tem efeito permanente sobre  $a_t$ , o que se reflete diretamente nas previsões de  $a_{t+s}$ ,  $s > 0$ . Sua variância não é constante, sendo função direta de  $t$ , e aproxima-se do infinito à medida que  $t$  tende para infinito. Assim, um processo RW vagueia sem exibir tendência alguma de crescimento ou decrescimento.

Portanto, conforme destaca Melo (2001), o método dos indicadores utilizado pela Secretaria da Receita Federal não está reproduzindo um processo autorregressivo causal, estacionário, e suas previsões não são confiáveis, uma vez que as condições básicas de estacionariedade não são satisfeitas. A circunstância que atenua o erro de previsão obtido pelo método dos indicadores é o fato de que as previsões são feitas anualmente, para um curto horizonte de tempo. Tal método, aplicado a previsões para longos períodos de tempo, certamente levaria a erros inaceitáveis.

### 2.3 Modelagens de séries temporais

A seguir são apresentadas apenas as equações básicas dos modelos.

► *Modelo Box-Jenkins – ARIMA: ARMA (p,q)*

$$y_t = \varphi_0 + \sum_{i=1}^p \varphi_i y_{t-i} + \sum_{i=0}^q \theta_i \varepsilon_{t-i}, \quad (3)$$



em que  $\phi_0$  é uma constante,  $\phi_1, \dots, \phi_p$  são os coeficientes da parte autorregressiva,  $\theta_0 = 1$ ,  $\theta_1, \dots, \theta_q$  são os coeficientes da parte média móvel,  $\varepsilon_1, \dots, \varepsilon_{t-q}$  são as componentes aleatórias do modelo e  $y_t$  é o processo modelado.

► *Modelo de função de transferência – TFM*

$$y_t = a_0 + A(L)y_{t-1} + C(L)z_t + B(L)\varepsilon_t, \quad (4)$$

em que  $A(L)$ ,  $B(L)$  e  $C(L)$  são polinômios no operador defasagem  $L$ . As sequências  $\{y_t\}$  e  $\{z_t\}$  representam as séries das variáveis endógena e exógena, respectivamente. O objetivo é estimar o parâmetro  $a_0$  e os parâmetros dos polinômios  $A(L)$ ,  $B(L)$  e  $C(L)$ . O polinômio  $C(L)$  é chamado função de transferência, na medida em que revela como o movimento na variável exógena  $z_t$  afeta  $\{y_t\}$ . Os coeficientes de  $C(L)$  são chamados de pesos da função de transferência.

► *Modelo dinâmico de múltiplas equações – VAR/VEC/SEM*

Modelos TFM pressupõem que a trajetória da variável dependente seja influenciada pela trajetória de uma variável independente ou exógena. Porém, muitos sistemas econômicos apresentam realimentação, ou seja, a trajetória da variável dita independente é afetada pela trajetória da variável dependente e a hipótese básica para a funcionalidade do modelo TFM é violada. Nesse caso, uma alternativa é trabalhar com um sistema de equações que trata todas as variáveis simetricamente, sem fazer distinção entre variável dependente e independente, o que é chamado de autorregressão vetorial (VAR). Um VAR de primeira ordem formado por duas variáveis, em que a trajetória de  $\{y_t\}$  seja afetada pelo presente e passado de  $\{z_t\}$  e que a trajetória de  $\{z_t\}$  seja afetada pelo presente e passado de  $\{y_t\}$ , é expresso, na forma VAR padrão, por:

$$\begin{cases} y_t = a_{10} + a_{11}y_{t-1} + a_{12}z_{t-1} + e_{1t}, \\ z_t = a_{20} + a_{21}y_{t-1} + a_{22}z_{t-1} + e_{2t} \end{cases}. \quad (5)$$

No sistema anterior,  $e_{1t}$  e  $e_{2t}$ , que se compõem dos dois choques  $\varepsilon_{yt}$  e  $\varepsilon_{zt}$ , são individualmente não correlacionados serialmente, porém correlacionados entre si.

### ► *Modelo estrutural – EM/EMX*

Modelos estruturais de séries de tempo são modelos formulados diretamente em termos de seus componentes de interesse, quais sejam, tendência, ciclo, sazonalidade e componente irregular, os quais têm uma interpretação direta. Além de terem considerável apelo intuitivo, proveem uma clara ligação com modelos de regressão. Um resumo dos principais modelos estruturais e suas propriedades pode ser encontrado em Harvey (1989). O modelo estrutural básico (BSM – modelo F – Harvey (1989), p. 510-511), com a inclusão opcional de variáveis exógenas, é expresso por:

$$y_t = \mu_t + \gamma_t + x_t' \delta + \varepsilon_t, \quad t = 1, \dots, T, \quad (6)$$

em que  $\mu_t$ ,  $\gamma_t$  e  $\varepsilon_t$  são os componentes de tendência, sazonalidade e irregular;  $x_t$  é um vetor ( $k \times 1$ ) de variáveis explicativas, que devem ser fracamente exógenas,<sup>5</sup> e  $\delta$  é o vetor associado de parâmetros. Evidentemente, o modelo pode ser manipulado para acrescentar ou excluir componentes. Se valores defasados de  $x_t$  entram na equação anterior, a variável  $x_t$  é chamada de indicador de direção (*leading indicator*).

## **2.4 Raiz unitária**

Para aplicação na modelagem de séries temporais,<sup>6</sup> é necessário que a série temporal seja estacionária. Assim, pode ser preciso aplicar o operador diferença ( $\Delta$ ) à série. Para séries que apresentam componente de tendência, é importante diferenciar entre aquelas que apresentam tendência estacionária e aquelas cuja tendência é estocástica. No primeiro caso, os desvios em relação à tendência determinista são temporários, seguindo um processo estacionário autorregressivo de médias móveis (ARMA), e o modelo é chamado tendência estacionária

---

5 De maneira análoga ao modelo TFM, quando o propósito é a previsão, a variável explicativa deve satisfazer a exigência mais rigorosa da exogeneidade forte (ver seção 2.5).

6 A estacionariedade da série temporal é requerida em todas as modelagens aplicadas neste trabalho, com exceção dos modelos estruturais.

(TS). No segundo caso, os choques têm impacto permanente na média condicional da série, não seguindo um processo estacionário em relação à tendência determinista ou à média do processo, e os modelos são chamados tendência estocástica (diferença estacionária – DS).

A correta identificação da tendência presente na série, se determinista ou estocástica, é fundamental para a aplicação do procedimento correto a fim de se remover a tendência e obter-se assim uma série estacionária. Há duas maneiras de se remover a tendência: para uma série TS, a remoção faz-se por meio da inclusão de um componente dependente do tempo na equação de regressão; para uma série DS, ou seja, com raiz unitária, aplicar a diferença à série conduzirá à estacionariedade. Portanto, percebe-se a importância dos testes de raiz unitária como pré-requisito para a correta estimação dos modelos.

## 2.5 Exogeneidade

Enders (2004) destaca que o ponto fundamental na análise de função de transferência é que se admite que  $\{z_t\}$  é um processo exógeno que se desenvolve independentemente de  $\{y_t\}$ , ou seja,  $\{y_t\}$  não tem efeito sobre  $\{z_t\}$ . A partir da equação (4), considerando  $C(L) = c_0 + c_1L + c_2L^2 + \dots$ , se  $c_0 = 0$ , o valor contemporâneo de  $z_t$  não afeta  $y_t$ . Assim, para a validade do modelo TFM,<sup>7</sup> é necessário que haja causalidade no sentido de  $\{z_t\}$  para  $\{y_t\}$ , sem a possibilidade de haver realimentação entre a variável dependente e as variáveis explicativas.

Para o entendimento dessa questão, três conceitos precisam ficar bem entendidos, quais sejam, exogeneidade fraca, causalidade de Granger e exogeneidade forte.

Segundo Harvey (1989), no caso de séries de tempo, com a variável dependente  $y = [y_1, \dots, y_T]'$  e a variável exógena designada pelo vetor  $X = [x_1, \dots, x_T]'$ , a função densidade de probabilidade conjunta (pdf conjunta) é dada por:

---

7 Os conceitos aqui apresentados também se aplicam ao modelo estrutural com inclusão de variáveis exógenas.

$$L(y, X; \lambda) = \prod_{t=1}^T p(y_t, x_t | Y_{t-1}, X_{t-1}; \lambda),$$

em que  $\lambda$  é o conjunto completo de parâmetros sobre o qual a pdf conjunta de  $y$  e  $X$  dependem. Considere que o modelo possa ser reparametrizado em função de um novo conjunto de parâmetros  $\theta^*$ , de tal forma que  $\theta^* = [\theta', \theta'_x]'$ , em que  $\theta$  e  $\theta_x$  tem variação livre (*variation-free*), ou seja, para qualquer valor de  $\theta_x$ ,  $\theta$  pode assumir qualquer valor no seu espaço de parâmetros e vice-versa. Suponha, ainda, que os parâmetros de interesse são funções apenas de  $\theta$ . A variável  $x_t$  será fracamente exógena com relação a  $\theta$  se a equação anterior puder ser fatorada em:

$$L(y, X; \theta^*) = \left[ \prod_{t=1}^T p(y_t | Y_{t-1}, X_t; \theta) \right] \left[ \prod_{t=1}^T p(x_t | X_{t-1}, Y_{t-1}; \theta_x) \right]$$

Assim, o segundo termo anterior, a distribuição marginal de  $x_t$ , pode ser ignorado já que não contém informação em  $\theta$  (parâmetros de interesse). Portanto, inferência sobre os parâmetros de interesse  $\theta$  pode ser efetuada por meio da distribuição condicional de  $y_t$ , o primeiro termo da equação anterior. Isso equivale a considerar a variável explicativa como se fosse fixa em repetidas amostras, mesmo que ela seja gerada por um processo estocástico.

A causalidade de Granger estabelece que uma variável  $x$  Granger causa uma variável  $y$ , se  $y_t$  puder ser mais bem prevista com o uso da informação do passado de  $x$ , do que sem a utilização dessa informação, mantidas as demais informações idênticas. Dizer que  $X_{t-1}$  não Granger causa  $y_t$  equivale a dizer que a distribuição de  $y_t$  é independente dos valores passados de  $x_t$ .

Ademais, conforme estabelece Covey e Bessler (1992), a causalidade de  $X$  para  $Y$  requer alguma teoria ou crença de que  $X$  causa  $Y$  e, ainda, evidências de que a adição de  $X$  no conjunto de informações resulta em melhora da *performance* preditiva dos valores futuros de  $Y$ . Melhora do poder preditivo sem uma teoria que sustente o modelo é insuficiente, pois é possível obter causalidade espúria entre duas va-

riáveis porque uma terceira variável, que causa ambas, foi deixada de fora do conjunto de informações.

Quando o interesse está na previsão da variável  $y_t$ , é necessário estabelecer o conceito de exogeneidade forte das variáveis explicativas. Um vetor de variáveis explicativas  $x_t$  é fortemente exógeno com relação a  $y_t$ , para o conjunto de parâmetros  $\theta$ , se  $x_t$  é fracamente exógeno com relação a  $\theta$  e  $y_t$  não causar  $x_t$  no sentido de Granger. Ou seja,  $x_t$  não pode ser influenciado pelos valores contemporâneos ou passados de  $y_t$ . Saliente-se que exogeneidade forte implica exogeneidade fraca, mas não o inverso.

É importante notar que a afirmação “X Granger causa Y” não implica que Y é o efeito ou resultado de X. A causalidade de Granger mede a precedência de uma variável em relação à outra e não pode, por si só, indicar causalidade, no uso mais comum do termo. Note-se, também, que a causalidade nos dois sentidos é frequente, ou seja, podemos ter “X Granger causa Y” e “Y Granger causa X” (EVIEWES 5 USER’S GUIDE, p. 376).

## 2.6 Cointegração

Quando o modelo VAR (equação (5)) contém variáveis não estacionárias, é possível haver uma combinação linear das variáveis que seja estacionária. Nesse caso as variáveis são cointegradas, e estimar um VAR em diferenças conduz a resultados incorretos. Cointegração refere-se à combinação linear de variáveis não estacionárias e integradas de mesma ordem, além do que o número máximo de vetores de cointegração é de  $n - 1$ , em que  $n$  é o número de variáveis envolvidas.

A principal característica de variáveis cointegradas é que suas trajetórias no tempo são influenciadas pela extensão de qualquer desvio do equilíbrio de longo prazo, ou seja, a dinâmica da relação de curto prazo, chamada de representação de correção de erro, é influenciada pelo desvio da relação de longo prazo. Um vetor  $x_t = (x_{1t}, x_{2t}, \dots, x_{nt})'$  tem uma representação de correção de erro (VEC) se puder ser expresso na forma:

$$\Delta x_t = \pi_0 + \pi x_{t-1} + \pi_1 \Delta x_{t-1} + \dots + \pi_p \Delta x_{t-p} + \varepsilon_t, \quad (7)$$

em que:  $\pi_0$  = vetor ( $n \times 1$ ) de interceptos com elementos  $\pi_{i0}$ ;  $\pi_1$  = matriz ( $n \times n$ ) de coeficientes com elementos  $\pi_{jk}(i)$ ;  $\pi$  = matriz com elementos  $\pi_{jk}$  tal que um ou mais  $\pi_{jk} \neq 0$ ;  $\varepsilon_t$  = vetor ( $n \times 1$ ) com elementos  $\varepsilon_{it}$ , tal que  $\varepsilon_{it}$  pode ser correlacionado com  $\varepsilon_{jt}$ .

Supondo que todas variáveis em  $x_t$  sejam integradas de ordem 1, ou seja,  $I(1)$ , se há uma representação de correção de erro como na equação (7), necessariamente há uma combinação linear de variáveis  $I(1)$  que é estacionária. O termo  $\pi x_{t-1}$  é estacionário já que todos os outros termos da equação (7) também o são. Cada linha de  $\pi$  é um vetor de cointegração de  $x_t$ . Se todos os elementos de  $\pi$  são iguais a zero, a equação (7) transforma-se num VAR tradicional em primeira diferença e não há representação de correção de erro, desde que  $\Delta x_t$  não responde ao desvio do período anterior do equilíbrio de longo prazo. Entretanto, se um ou mais elementos  $\pi_{jk}$  diferem de zero,  $\Delta x_t$  responde ao desvio do período anterior do equilíbrio de longo prazo e, portanto, estimar  $x_t$  como um VAR em primeira diferença não é apropriado se  $x_t$  tem uma representação de correção de erro.

Em sistemas cointegrados, geralmente ambas as variáveis respondem aos desvios do equilíbrio de longo prazo. Porém, é possível que uma das variáveis  $\{z_t\}$  não responda ao desvio do equilíbrio de longo prazo e todo o ajuste seja feito pela outra variável  $\{y_t\}$ . Nesse caso,  $\{z_t\}$  é considerada fracamente exógena.

De acordo com Hendry (1995), processos integrados podem ser expressos em termos de variáveis em diferenças e relações de cointegração sem perda de informação, mas a redução na dimensão do vetor de parâmetros pode trazer ganho. Assim, para se obter um modelo mais parcimonioso, após o VEC, o modelo é reduzido para  $I(0)$  (estacionário), formando um modelo de equações simultâneas (SEM), em diferenças e com as relações de cointegração (identidades), a partir do qual restrições podem ser impostas em cada equação individualmente, bem como sobre as identidades, e os testes de adequação do modelo são efetuados. Após a redução, a análise estatística convencional pode ser implementada. Caso os parâmetros estimados para as identidades

de determinada equação (variável dependente  $y_2$ , por exemplo) sejam iguais a zero, a respectiva variável é considerada fracamente exógena. Se, adicionalmente, valores passados das outras variáveis ( $y_1, y_3, \dots$ ) podem ser retirados da equação de  $y_2$  sem perda de informação, ou seja,  $y_1, y_3$ , etc. não causam  $y_2$  no sentido de Granger, então  $y_2$  é fortemente exógena. Nesse caso, alternativamente, pode-se estimar um modelo TFM em que  $y_2$  é inserida como variável exógena, tornando os modelos mais simples de estimar, devido à diminuição do número de parâmetros, e previsões dinâmicas podem ser efetuadas.

### 3 PROCEDIMENTOS METODOLÓGICOS

No presente trabalho são analisadas as séries de três tributos, compreendendo o período de 2000 a 2007, no âmbito da jurisdição da 8ª Região Fiscal (8ª RF), que corresponde ao Estado de São Paulo. Os tributos a serem analisados, segundo a classificação adotada pela RFB, são o Imposto de Importação (II), o Imposto sobre a Renda da Pessoa Jurídica – Demais Obrigadas ao Lucro Real (IRPJ\_DOLR) e a Contribuição para o Financiamento da Seguridade Social – Demais Empresas (COFINS\_DE).<sup>8</sup>

A escolha dos tributos a serem analisados levou em conta a participação de cada um na arrecadação total do estado para o ano de 2007, bem como o âmbito de incidência do tributo, que no presente estudo é o setor externo, o lucro e o faturamento das empresas, para o II, o IRPJ\_DOLR e a COFINS\_DE, respectivamente. Os tributos especificados foram responsáveis por aproximadamente 24% da arrecadação da 8ª RF, que, por sua vez, é responsável por aproximadamente 43% da arrecadação tributária total do país.

Para cada uma das séries anteriormente especificadas, são efetuadas previsões da arrecadação por meio das seguintes modelagens: i) Box-Jenkins (ARIMA); ii) modelo dinâmico univariado (DR) e modelo

---

8 As séries do IRPJ\_DOLR e da COFINS\_DE excluem a arrecadação das empresas do setor financeiro, que, supõe-se, possuem dinâmica própria e diferenciada das demais empresas.

de função de transferência (TFM), em que variáveis exógenas são incluídas no modelo dinâmico univariado; iii) modelos dinâmicos de múltiplas equações: autorregressão vetorial (VAR), modelo VAR com correção de erros (VEC) e modelos de equações simultâneas (SEM); iv) modelo estrutural (EM e EMX) sem e com inclusão de variáveis exógenas, respectivamente.

Para conclusão final, os métodos de previsão são comparados entre si e com a previsão efetuada pela RFB, em que se utilizou o método dos indicadores. Os dados reais da arrecadação de 2007 serviram de base de comparação.

Os dados das séries de arrecadação dos tributos envolvidos na análise foram obtidos diretamente da Receita Federal do Brasil.<sup>9</sup> Quanto às séries das variáveis explicativas, foram utilizadas as seguintes:<sup>10</sup>

- ▶ **Taxa Real de Câmbio (TCR):** Ipea (Instituto de Pesquisa Econômica Aplicada – [www.ipea.gov.br](http://www.ipea.gov.br)) – Taxa de câmbio – efetiva real – INPC – exportações – índice (média 2000 = 100) – Ref: IPEA – GAC12\_TCRXTINPC12;
- ▶ **PIB mensal (PIB):** fonte: Bacen (Banco Central do Brasil – [www.bcb.gov.br](http://www.bcb.gov.br)) – Ref.: série nº 4380 – PIB mensal – Brasil – Valores correntes (R\$ milhões) – deflacionado pelo IER<sup>11</sup> e transformado em índice;
- ▶ **Índice de Produção Industrial – IPI\_g\_sp:** IBGE (Instituto Brasileiro de Geografia e Estatística – [www.ibge.gov.br](http://www.ibge.gov.br)) – Produção física industrial – Indústria geral – São Paulo – Índice de base fixa mensal sem ajuste sazonal (Base: média de 2002 = 100);
- ▶ **Índice de Vendas no Varejo – IVV\_sp:** Bacen (Banco Central do Brasil – [www.bcb.gov.br](http://www.bcb.gov.br)) – Ref.: série nº 1475 – Índice volume de vendas no varejo (2003=100) – Total – São Paulo.

---

9 Sistema *Data Warehouse* – DW Arrecadação.

10 Todos os números-índice tiveram a base alterada para janeiro de 2000 = 100.

11 IER – Índice Específico da Receita (ponderado 55% IPCA – 45% IGP-DI) – índice de preço utilizado pela RFB na elaboração da previsão pelo método dos indicadores, que também será usado neste trabalho como índice de deflação das séries dos tributos analisados.



Para cada uma das séries das variáveis explicativas são efetuadas previsões dessas variáveis para o ano de 2007, por meio das modelagens ARIMA, dinâmica univariada e estrutural.<sup>12</sup> Esses valores são incorporados aos modelos de função de transferência e estrutural com variáveis explicativas, das variáveis dependentes de arrecadação.

Nos modelos multivariados para o Imposto sobre a Renda da Pessoa Jurídica e a Cofins, além da série do próprio tributo, são utilizadas as séries do PIB ou as *proxys* do PIB, IPI\_g\_sp ou IVV\_sp, individualmente para cada especificação. Para a série do Imposto de Importação, além destas, também consideradas individualmente, é utilizada, conjuntamente, a série da TCR, em cada especificação.

Para estimação dos modelos foram utilizados os seguintes *softwares* e métodos de estimação: i) ARIMA: E-views (versão 5.0) – OLS – *Ordinary Least Square*; ii) modelos univariados dinâmicos – DR: *PcGive* (versão 10.0b) – OLS – *Ordinary Least Square*; iii) modelos de função de transferência – TFM: *PcGive* (versão 10.0b) – OLS – *Ordinary Least Square*; iv) modelos multivariados dinâmicos – VAR/VEC/SEM: *PcGive* (versão 10.0b) – URF\_VAR: OLS – *Ordinary Least Square*; VEC: *Reduced Rank Regression*; SEM: FIML – *Full Information Maximum Likelihood*; v) modelos estruturais – EM e EMX: *Stamp (Structural Time Series Analyser, Modeller and Predictor* – versão 6.20) – *Maximum Likelihood*.<sup>13</sup>

Na especificação ARIMA, para permitir a comparação entre os diversos modelos por meio dos critérios de ajustamento AIC e BIC, as estimações foram efetuadas sobre o mesmo intervalo de dados. Ocorre que, quando o modelo estimado inclui variáveis em diferença, especialmente diferença sazonal, isso leva à perda de muitas observações no início da série.<sup>14</sup> Assim, o intervalo de estimação da série é determinado em função do modelo com maior número de parâmetros a ser estimado, sendo este um dos motivos da diferença na extensão do intervalo de

---

12 Isso se mostrou necessário devido ao fato de que, numa situação real de aplicação do método à previsão da arrecadação, no momento da previsão, os valores reais das variáveis explicativas não são conhecidos.

13 *PcGive* e *Stamp* são módulos do *software GiveWin* (versão 2.02).

14 Para série mensal, a primeira diferença sazonal leva à perda de 12 observações.

estimação para as três séries. Na especificação dos demais modelos (DR, TFM, VAR/VEC/SEM, EM) também é mantido o mesmo intervalo de dados, porém diferentemente do intervalo de estimação da especificação ARIMA, já que não há perda de informação devido à diferença sazonal presente neste último.

Na modelagem ARIMA os *outliers* não foram tratados. Deve-se enfatizar que o objetivo do presente trabalho é desenvolver metodologias alternativas ao método dos indicadores da RFB para previsão da arrecadação federal. A modelagem ARIMA para previsão de receitas federais, que é uma das metodologias alternativas, já foi elaborada de forma bem mais detalhada por dois outros trabalhos, Melo (2001) e Siqueira (2002), não constituindo, portanto, o objetivo mais importante aqui. Assim, a procura do modelo ARIMA ideal não foi aprofundada como deveria, por exemplo, com o tratamento dos *outliers* e das quebras estruturais. O foco foi encontrar modelos adequados do ponto de vista econométrico e que servissem de base de comparação com o método dos indicadores e com os demais modelos elaborados neste trabalho.

Em todas as modelagens desenvolvidas são efetuados vários testes, a exemplo dos testes de autocorrelação, normalidade e heteroscedasticidade dos resíduos, até se chegar ao modelo mais adequado. A modelagem dinâmica multivariada é usada, também, para se testar a exogeneidade das variáveis explicativas.

Para identificação da raiz unitária nas séries são utilizados os testes Ng-Perron (NgP) e Elliott-Rothenberg-Stock DF-GLS (DF-GLS (ERS)). Esses testes confirmam a presença de raiz unitária, tanto para as séries de arrecadação, quanto para as séries das variáveis explicativas.

Para se verificar a presença e a quantidade de relações de cointegração  $I(1)$ , foi aplicada a metodologia de Johansen. Todas as combinações de LII/LTCR<sup>15</sup> com LIPI, LPIB ou LIVV apresentaram duas relações de cointegração, ou seja, existem duas combinações lineares, entre as três variáveis, que são estacionárias. LCF combinado com LIPI, LPIB ou LIVV apresentou uma relação de cointegração. O mesmo aconteceu com

---

15 Ver nota 17 para identificação das variáveis.

as combinações LLR/LIPI e LLR/LPIB. A única exceção foi LLR/LIVV, que não apresentou relação de cointegração. Não foram encontradas relações de cointegração I(2) entre as variáveis.

Os modelos estimados consideram dados até dezembro de 2006. Os valores de 2007 são deixados de fora da amostra para permitir a comparação com os valores previstos. Os métodos mais utilizados para medir a acurácia da previsão são o desvio absoluto médio (MAD), a raiz do erro quadrático médio (RMSE) e o erro percentual absoluto médio (MAPE). Todos utilizam os resíduos em seus cálculos, definido como a diferença entre o valor previsto pelo modelo e o valor real, ou seja:  $e_t = P_t - R_t$ . Assim, os erros de previsão são definidos pelas seguintes fórmulas:

$$MAD = \frac{\sum |e_t|}{n}, \quad RMSE = \sqrt{\frac{\sum e_t^2}{n}}, \quad MAPE = \frac{\sum \left| \frac{e_t}{R_t} \right|}{n} * 100, \quad E_{anual} = \frac{\sum_1^{12} P_t - \sum_1^{12} R_t}{\sum_1^{12} R_t} * 100.$$

em que  $n$  é o número de valores previstos, obtidos dos dados passados.<sup>16</sup> O MAD é definido como a média dos valores absolutos de cada resíduo; o RMSE é a raiz quadrada da média dos valores quadráticos de cada resíduo; o MAPE considera o erro relativo de cada previsão. Nas tabelas de resultados dos modelos estimados também é apresentado o erro de previsão anual ( $E_{anual}$ ), que pretende mensurar qual o erro percentual anual da previsão, bem como se o valor previsto foi superior ou inferior ao valor real. Essa medida é importante, pois, quando se trata de estimar a receita para fins de orçamento, e mesmo para a avaliação do desempenho anual da arrecadação, efetuado pela Secretaria da Receita Federal do Brasil, o que realmente importa é uma previsão anual benfeita.

Melo (2001) destaca que, se erros elevados de previsão são inaceitáveis, então o uso do RMSE faz-se necessário – enquanto, se for possível ignorar alguns erros elevados, o MAD funciona melhor.

<sup>16</sup> Neste trabalho,  $n = 12$ .

O MAPE, por se tratar de uma medida percentual, é utilizado para comparar a acurácia de duas séries temporais diferentes. Neste trabalho, embora constem nos quadros comparativos as quatro medidas descritas, o RMSE é utilizado como principal critério de acurácia para as comparações dos métodos de previsão apresentados. Para que os erros de previsão dos modelos em nível pudessem ser comparados com os erros dos modelos em diferenças, no caso destes últimos, foi efetuada a transformação da previsão da variável em diferença para previsão em nível.

## 4 ANÁLISE DAS SÉRIES

### 4.1 Gráficos

O Gráfico 1 apresenta as séries de arrecadação após transformação logarítmica<sup>17</sup> e deflacionadas pelo IER. Os valores referem-se à arrecadação divulgada pela RFB, que é a base de comparação utilizada pelo órgão para se apurar a acurácia da previsão. Compõem-se das seguintes rubricas:<sup>18</sup> arrecadação apropriada no dia + arrecadação apropriada em atraso no decêndio + saldo das retificações + Cide + Refis + retenção pessoa jurídica de direito privado + Paes + Estorno Paes + TPAI + RPEM + valor retenção de órgão público. Para o Imposto de Importação, o início da série ocorre em janeiro de 2000, enquanto para as séries do Imposto sobre Renda da Pessoa Jurídica e da Cofins, o início dá-se em agosto de 2000.<sup>19</sup> Uma análise mais detalhada das séries encontra-se em 4.2, que trata da detecção inicial de *outliers*.

---

17 Transformação Box-Cox ou logarítmica, com  $\lambda = 0$  com o objetivo de estabilização da variância. As séries transformadas do II, IRPJ\_DOLR e COFINS\_DE são chamadas LII, LLR e LCF. LIVV, LPIB, LIPI e LTCR referem-se às transformações logarítmicas das variáveis explicativas.

18 Cide – Contribuição de Intervenção no Domínio Econômico – Combustíveis – Lei nº 10.336/2001. Refis/Paes – programas de parcelamento instituídos pelas Leis nºs 9.964/2000 e 10.864/2003, respectivamente. TPAI – Tributação do Patrimônio de Afetação – Lei nº 10.931/2004. RPEM – Retenção de CSLL, Cofins e Pis/Pasep, efetuada por órgão público dos estados e municípios nos pagamentos efetuados a pessoas jurídicas de direito privado pelo fornecimento de bens e serviços – Lei nº 10.833/2003, art. 33.

19 Início da série da arrecadação divulgada pela RFB (DW Arrecadação). No caso do II, dispunha-se dos valores desde janeiro de 2000.

O Gráfico 2 apresenta as séries das variáveis explicativas, em número índice. Destacam-se na análise gráfica a alta variabilidade até 2004 e a tendência de queda a partir de 2003 da TCR, o forte componente sazonal presente nas séries IPI\_g\_sp, PIB e, principalmente, IVV\_sp, além da clara tendência ascendente nessas séries a partir de 2003. Conforme era de esperar, pois se trata do próprio PIB ou *proxys* deste, as séries PIB, IVV\_sp e IPI\_g\_sp mostram forte correlação positiva. Os coeficientes de correlação, quando se consideram as séries dessazonalizadas, que foram utilizadas para detecção inicial de *outliers*, conforme se verá mais adiante, e após transformação logarítmica, são de 0,76, 0,79 e 0,87 entre LIPI/LIVV, LPIB/LIVV e LPIB/LIPI, respectivamente.

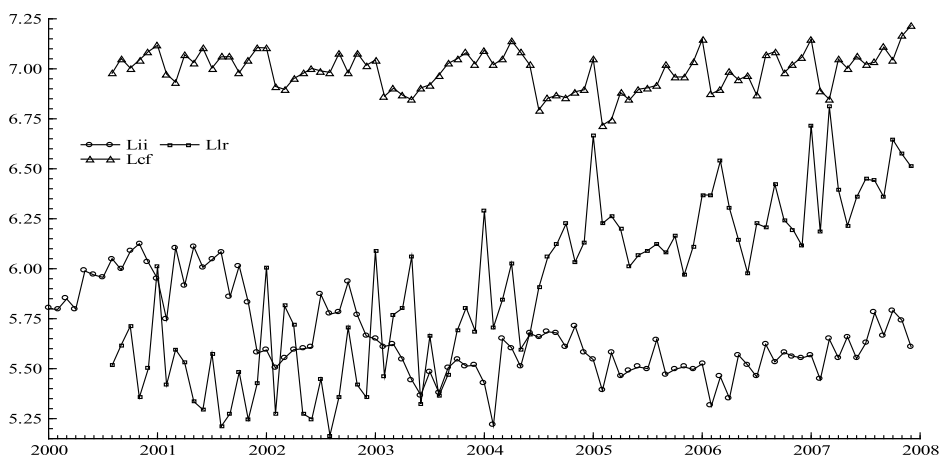


Gráfico 1. Séries LII, LLR e LCF

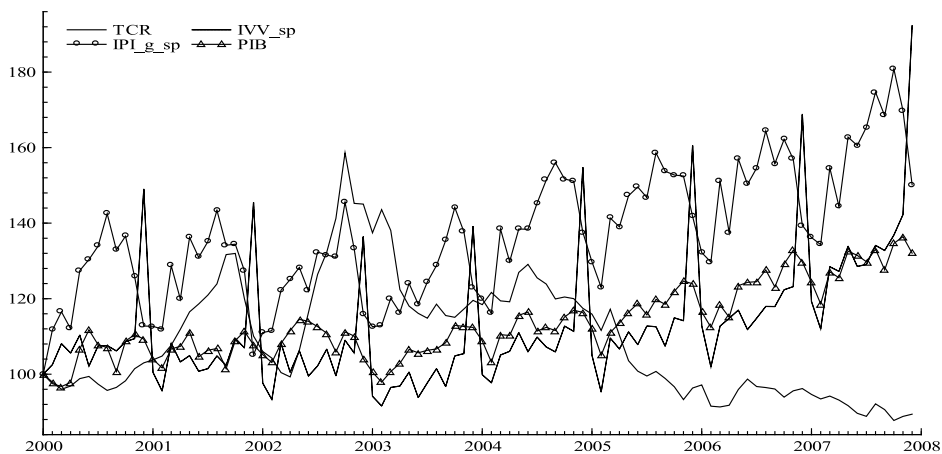


Gráfico 2. Séries TCR, IVV\_sp, IPI\_g\_sp e PIB – Índices

#### 4.2 Detecção inicial de outliers

Segundo Harvey e Koopman (1992), a presença de *outliers* (valores fora do padrão) e as quebras estruturais podem ser identificadas por meio da utilização dos resíduos auxiliares, que são obtidos da estimação de modelos estruturais. Os dois principais tipos de *outliers* encontrados são o aditivo (AO), que afeta somente uma determinada observação, e a mudança estrutural (IO ou LS), que afeta o valor da série a partir de uma determinada observação. Destaque-se que a presença de *outliers* pode causar sérios problemas de viés na estimação de modelos ARMA.

No presente trabalho utilizou-se o modelo básico estrutural<sup>20</sup> como especificação inicial para a detecção de *outliers*, por meio da análise dos resíduos auxiliares, levando-se em consideração, ainda, que os dados foram previamente ajustados sazonalmente, conforme recomendação de Koopman et alii (2000). A análise dos resíduos dos modelos estimados também pode indicar a presença de *outliers* e quebras estruturais, mas

20 Geralmente o modelo básico estrutural com pequenas alterações na especificação apresentou resultados satisfatórios em termos da adequação do modelo. A exceção foi o modelo para a variável explicativa Taxa Real de Câmbio, em que foi necessária a introdução de defasagens da variável dependente para que se alcançasse a normalidade dos resíduos.

geralmente não fornece uma clara indicação da fonte do problema. Em todas as séries detectou-se algum tipo de *outlier*, conforme pode ser verificado na Tabela 2.

Deve-se enfatizar que esses modelos básicos pretendem fornecer um ponto de partida para a identificação de *outliers*, principalmente para utilização em modelagens outras que não a estrutural. Quando da estimação de determinado modelo, a identificação dos *outliers* desenvolve-se conjuntamente à especificação do modelo, não sendo necessariamente utilizados todos os identificados na Tabela 2, bem como podem surgir outros.

Visando a obter um melhor conhecimento das séries de arrecadação analisadas, a seguir é apresentada análise gráfica (relativa ao Gráfico 1), juntamente com os resultados dos *outliers* e das quebras estruturais encontrados (vide Tabela 2) e a análise da arrecadação efetuada pela RFB.

A análise gráfica da série do Imposto de Importação parece mostrar comportamento bem distinto entre os períodos 2000–2002/2003 e 2003/2004–2006. O método de detecção de *outliers* aponta mudanças mais pronunciadas no nível da série em novembro/dezembro de 2001, em maio de 2003 e em março de 2004. A primeira queda, que foi negativa, está provavelmente associada à crise provocada pelos atentados terroristas a Nova York em setembro de 2001. Quanto à segunda, também negativa, a regulamentação do Imposto de Importação sofreu alteração por meio do Decreto nº 4.543, de 26 de dezembro de 2002, em vigor a partir de janeiro de 2003 (novo Regulamento Aduaneiro revoga o Decreto nº 91.030/85) e esta, juntamente com a diminuição da demanda por força do desaquecimento econômico e com a forte valorização cambial, ocorridas no início de 2003,<sup>21</sup> pode ter sido a causa dessa mudança estrutural. Por fim, a alteração de março de 2004 é positiva e possivelmente se deve à recuperação da atividade econômica.

---

21 Lembre-se que no segundo semestre de 2002 houve espantosa desvalorização cambial devido à instabilidade decorrente do processo eleitoral e, num primeiro momento, a valorização do câmbio, ocorrida no início de 2003, diminui a base de cálculo e, conseqüentemente, a arrecadação do II.

Tabela 2. Detecção inicial de *outliers* – variáveis principais e explicativas

Série (*1)	Mês/Ano	Tipo de outlier (*2)	Valor resíduo auxiliar (*3)	Estatística (*4)	
LII	09/2001	LS	-2,2219	0,0145	
	11/2001	LS	-2,6268	0,0051	
	12/2001	LS	-2,6573	0,0047	
	07/2002	LS	2,0910	0,0198	
	07/2002	AO	2,1632	0,0167	
	03/2003	LS	-2,1223	0,0184	
	05/2003	LS	-2,7202	0,0040	
	08/2003	AO	-2,1477	0,0173	
	02/2004	AO	-2,2805	0,0126	
	03/2004	LS	2,5984	0,0055	
LLR	05/2003	AO	4,2271	0,0000	
	06/2003	LS	-2,3028	0,0120	
	06/2004	LS	2,1475	0,0175	
	07/2004	LS	2,3597	0,0104	
	08/2004	LS	2,8123	0,0031	
	01/2006	AO	-2,4250	0,0088	
	02/2006	LS	2,0115	0,0239	
	LCF	04/2003	LS	-2,0440	0,0222
02/2004		LS	2,5264	0,0068	
06/2004		LS	-2,5290	0,0067	
07/2004		LS	-3,4958	0,0004	
07/2004		AO	-2,0777	0,0205	
01/2005		AO	2,5842	0,0058	
02/2005		LS	-2,0771	0,0206	
08/2006		LS	2,4313	0,0087	
08/2006		AO	2,1896	0,0158	
LTCR		11/2001	LS	-2,6094	0,0055
	10/2002	AO	3,1343	0,0012	
	11/2002	LS	-2,4420	0,0085	
	11/2002	AO	-3,2917	0,0008	
	02/2003	AO	2,5205	0,0069	
	03/2005	AO	2,4074	0,0092	
	04/2005	LS	-2,3789	0,0099	
LPIB	03/2000	LS	-2,3234	0,0113	
	05/2000	LS	2,4861	0,0074	
	06/2000	LS	2,5217	0,0068	
	06/2000	AO	2,6259	0,0051	
	09/2000	AO	-2,0996	0,0194	
	06/2001	LS	-3,0621	0,0015	
	06/2001	AO	-3,0034	0,0018	
	12/2002	LS	-2,1288	0,0181	
	09/2003	LS	2,1546	0,0170	
	LIW	02/2000	AO	2,1376	0,0177
		03/2002	AO	2,5930	0,0056
		11/2002	LS	-2,0998	0,0194
		12/2002	LS	-2,9287	0,0022
01/2004		LS	2,6425	0,0049	
02/2004		LS	2,5052	0,0071	
08/2004		LS	-2,0401	0,0222	
02/2005		AO	-2,1748	0,0162	
LIPI	01/2000	AO	-2,1465	0,0174	
	02/2000	LS	2,1415	0,0176	
	02/2000	AO	2,1680	0,0165	
	10/2001	LS	-2,0684	0,0208	
	12/2001	AO	-2,6386	0,0050	
	04/2002	AO	2,4491	0,0082	
	10/2002	LS	2,1358	0,0178	
	10/2002	AO	2,2615	0,0132	
	09/2003	LS	2,5351	0,0065	

Fonte: Elaboração do autor

Notas: (\*1) Para séries com forte componente sazonal, trata-se da série ajustada sazonalmente.

(\*2) AO: additive outlier; LS: level shift.

(\*3) Valor do resíduo auxiliar do componente irregular (AO) ou do componente de nível (LS). Somente estão mostrados os períodos em que o valor (em módulo) foi superior a 2, conforme Harvey e Koopman (1992).

(\*4) Estatística de distribuição padrão (não pode ser usada como teste padrão devido à correlação serial presente nos resíduos auxiliares).



A análise gráfica da série do Imposto sobre a Renda da Pessoa Jurídica parece mostrar comportamento bem distinto entre os períodos 2000–2003 (nível aproximadamente constante) e 2004–2006 (com tendência ascendente). O método de detecção de *outliers* aponta mudança mais pronunciada no nível da série em agosto de 2004. A causa dessa alteração (positiva) também parece dever-se à recuperação da atividade econômica. Apresenta, ainda, forte *outlier* do tipo aditivo em maio de 2003, devido ao fato de que diversas empresas interromperam longas sequências de ausências de recolhimento, aumentando a arrecadação daquele mês, conforme consta na análise da arrecadação efetuada pela RFB.

Quanto à série da Cofins, a análise gráfica parece mostrar comportamento bem distinto entre os períodos 2000–meados de 2004 (nível aproximadamente constante) e o segundo semestre de 2004–2006 (com tendência ascendente). O método de detecção de *outliers* aponta mudança (negativa) mais pronunciada no nível da série em julho de 2004. No ano de 2004, a legislação da Cofins sofreu muitas alterações, principalmente em função da instituição da Cofins não cumulativa<sup>22</sup> e da incidência sobre importações.<sup>23</sup> Assim, a Cofins sobre o faturamento apresentou tendência de declínio ao longo de 2004, em virtude da instituição da incidência não cumulativa e da incidência sobre importações. Já a Cofins não cumulativa apresentou tendência de estabilização (apenas oscilações sazonais) no 2º semestre, quando já se faziam sentir os efeitos da incidência sobre importações (desembaraço de mercadorias por filiais) e de outras alterações de legislação relacionadas à desoneração de cadeias produtivas (Lei nº 10.925/2004 e MP nº 183/2004) e ao estímulo aos investimentos (MP nº 219/2004). Assim, em função das profundas e diversas alterações legislativas, o imposto oscilou bastante em 2004, com tendência de queda na arrecadação, que passa a crescer novamente, com tendência ascendente, a partir de 2005, graças, provavelmente, à recuperação da atividade econômica.

---

22 Lei nº 10.833, de 29 de dezembro de 2003, com vigência a partir de março de 2004, que foi projetada de sorte que mantivesse os níveis de arrecadação. Dessa maneira, para compensar diversas deduções permitidas à base de cálculo, promoveu-se elevação da alíquota de 3% para 7,6%.

23 Lei nº 10.865, de 30 de abril de 2004, com vigência a partir de maio de 2004.

### **4.3 Testes de exogeneidade das variáveis explicativas**

Os testes de exogeneidade das variáveis explicativas – tendo por base a causalidade de Granger em sistemas cointegrados para a identificação das variáveis mais apropriadas às modelagens de função de transferência e estrutural com variáveis exógenas – foram efetuados a partir da análise dos coeficientes e da velocidade de ajustamento (parâmetro  $\alpha$ ) dos modelos VEC\_SEM e são apresentados na Tabela 3.

A tabela indica o modelo VEC\_SEM resultado da redução para I(0) dos modelos URF\_VAR mais adequados, o posto da matriz  $\Pi$ , que corresponde ao número de relações de cointegração presentes no modelo, a velocidade de ajustamento da variável explicativa e se algum *lag* da variável principal está presente na equação da variável explicativa. Caso o(s) parâmetro(s)  $\alpha$  seja(m) igual(is) a zero, a variável é dita fracamente exógena. Se, adicionalmente, nenhum *lag* da variável principal estiver presente na equação da respectiva variável explicativa, a variável é dita fortemente exógena, e um modelo de função de transferência ou estrutural, mais simples e com menos parâmetros, pode ser estimado com a inclusão da variável explicativa como variável exógena. Caso contrário, ou seja, o(s) parâmetro(s)  $\alpha$  seja(m) diferente(s) de zero, independentemente da condição do *lag* da variável principal estar presente na equação da variável explicativa, a variável é endógena e o modelo VEC\_SEM é o mais adequado para se efetuar previsões da variável principal.

Tabela 3. Resultados dos testes de exogeneidade das variáveis explicativas em modelos multivariados

Modelo VEC/SEM	Posto da matriz $\Pi$ (*1)	Variável explicativa	Velocidade de ajustamento (*2)		Var. principal → Var. explicativa (*3)	Conclusão (*4)
			$\alpha_1$	$\alpha_2$		
DLII/DLIPI/DLTCR VEC/SEM Nº 1 – Tab. 6	2	DLIPI	0,29 (0,000)	-0,027 (0,000)	SIM	endógena
		DLTCR	0	-0,026 (0,000)	SIM	endógena
DLII/DLIWV/DLTCR VEC/SEM Nº 2 – Tab. 6	2	DLIWV	0	-0,12 (0,004)	SIM	endógena
		DLTCR	0	-0,31 (0,000)	SIM	endógena
DLII/DLPIB/DLTCR VEC/SEM Nº 3 – Tab. 6	2	DLPIB	-0,066 (0,000)	-0,44 (0,000)	NÃO	endógena
		DLTCR	-0,17 (0,000)	0,80 (0,000)	SIM	endógena
DLLR/DLIPI VEC/SEM Nº 1 – Tab. 6	1	DLIPI	0,054 (0,018)	—	SIM	endógena
DLLR/DLPIB VEC/SEM Nº 2 – Tab. 6	1	DLPIB	0	—	NÃO	exógena forte
DLLR/DLIWV VEC/SEM Nº 3 – Tab. 6	0	DLIWV	—	—	SIM	endógena
DLCF/DLIWV VEC/SEM Nº 1 – Tab. 6	1	DLIWV	0	—	SIM	exógena fraca
DLCF/DLPIB VEC/SEM Nº 2 – Tab. 6	1	DLPIB	0	—	NÃO	exógena forte
DLCF/DLIPI VEC/SEM Nº 3 – Tab. 6	1	DLIPI	0,34 (0,000)	—	SIM	endógena

Fonte: Elaboração do autor

Notas: (\*1) Posto da matriz  $\Pi$  do modelo URF\_VAR correspondente.

(\*2) Coeficiente e t-prob da velocidade de ajustamento ( $\alpha$ ) presente na equação da respectiva variável dependente (explicativa), no modelo VEC\_SEM.

(\*3) Indica se algum lag da variável principal está presente na equação da variável explicativa.

(\*4) Indica a condição da variável explicativa no modelo.

## 5 RESULTADOS

A seguir são apresentados os resultados das diversas modelagens tratadas no presente trabalho. Procurou-se dar ênfase aos resultados das previsões, bem como aos principais testes de adequação dos modelos, cujas especificações estão resumidas. Os erros de previsão obtidos pelo método dos indicadores também constam de cada tabela, para efeito de comparação. A Tabela 9, exibida no Apêndice, apresenta os valores reais e previstos dos melhores modelos de cada metodologia.

### 5.1 Modelagem Box-Jenkins – ARIMA

Os resultados da modelagem ARIMA encontram-se na Tabela 4. São apresentados somente os dois melhores resultados para cada série de arrecadação. A escolha dos melhores modelos baseou-se nos critérios de ajustamento (AIC e BIC), significância dos coeficientes estimados, normalidade dos resíduos, não correlação dos resíduos, estacionariedade e invertibilidade do modelo e, finalmente, nos menores erros de previsão RMSE.

- ▶ **Imposto de Importação:** os resultados apontam o modelo SARIMA  $(0,1,1) \times (0,0,1)_{12}$  como o melhor, com RMSE 15% inferior ao do método dos indicadores.
- ▶ **Imposto sobre a Renda da Pessoa Jurídica:** os resultados apontam o modelo SARIMA  $(1,1,0) \times (0,1,1)_{12}$  como o melhor, com RMSE aproximadamente 20% inferior ao do método dos indicadores. Destaque-se que o modelo nº 2, que é sem componente estocástico, somente com constante e tendência determinista, apresentou ótimo resultado do RMSE, 16% inferior ao modelo RFB e erro anual de apenas 1,7%.
- ▶ **Cofins:** os resultados apontam o modelo SARIMA  $((1,10),0,(12)) \times (0,1,0)_{12}$  com constante e tendência determinista como o melhor, porém com o RMSE pouco superior ao do método dos indicadores. Houve dificuldade de se encontrar normalidade nos resíduos das regressões.

Tabela 4. Modelagem ARIMA – II/IRPJ\_DOLR/COFINS\_DE

Modelos	SARIMA (p,d,q) x (P,D,Q) s=12 c = cte ; t = tend. determinista	Normal. resíduos (*1)	Correlação serial dos resíduos			Estac. e invert.	Erros de previsão			
			p (*2)	Estat. Q (Ljung-Box) (prob) (*3)	Estat. Breusch-Godfrey LM (prob)		MAD	RMSE	MAPE (%)	Annual (%)
<b>Série: II – Intervalo dos dados: 04/2002 a 12/2006</b>										
1	(0,1,1) x (0,0,1)	0,67	3	3,12 (0,08)	5,87 (0,12)	-/sim	15,0	19,3	5,4	2,4
2	(1,1,0) x (0,0,1)	0,89	3	2,63 (0,11)	5,58 (0,13)	sim/sim	18,2	23,0	6,6	4,7
<b>RFB</b>							<b>15,6</b>	<b>22,6</b>	<b>5,7</b>	<b>2,5</b>
<b>Série: IRPJ_DOLR – Intervalo dos dados: 10/2002 a 12/2006</b>										
1	(1,1,0) x (0,1,1)	0,16	3	3,28 (0,07)	3,63 (0,30)	sim/sim	97,9	123,0	14,0	-13,6
2	(0,0,0) x (0,0,0) + c + t	0,27	1	1,23 (0,27)	1,19 (0,27)	-/-	95,8	127,9	14,1	-1,7
<b>RFB</b>							<b>132,6</b>	<b>152,2</b>	<b>19,3</b>	<b>-17,3</b>
<b>Série: COFINS_DE – Intervalo dos dados: 06/2002 a 12/2006</b>										
1	((1,10),0,(12)) x (0,1,0) + c + t	0,83	4	1,54 (0,21)	1,85 (0,76)	sim/sim	73,7	89,4	6,2	-5,5
2	((10),1,1) x (1,0,1)	0,19	5	4,32 (0,04)	4,35 (0,50)	sim/sim	78,6	98,7	6,7	-4,3
<b>RFB</b>							<b>66,3</b>	<b>85,0</b>	<b>5,6</b>	<b>-4,5</b>

Fonte: Elaboração do autor

Notas: (\*1) Probabilidade do teste Jarque-Bera para normalidade dos resíduos.

(\*2) Ordem do lag. Na estatística Q refere-se à ordem que apresenta probabilidade inferior a 10%. Caso todas as probabilidades sejam superiores a 10%, refere-se à ordem da primeira autocorrelação computada. Este mesmo valor de p é usado na estatística Breusch-Godfrey LM.

(\*3) Estatística ajustada pelos números de coeficientes ARMA estimados.

## 5.2 Modelos dinâmicos univariados e de função de transferência<sup>24</sup>

Os resultados dos modelos dinâmicos univariados (DR) e de função de transferência (TFM) são apresentados na Tabela 5. Todos os

24 Os modelos de função de transferência são modelos dinâmicos de uma única equação, embora envolvam mais de uma série. Assim, por sua semelhança com os modelos dinâmicos univariados e por terem sido estimados pelos mesmos métodos e programas destes, são tratados no mesmo tópico.

modelos estão em diferenças devido à presença de raiz unitária e levam em consideração o resultado dos testes de exogeneidade das variáveis explicativas. Assim, somente foram estimados modelos TFM quando a variável explicativa se mostrou fortemente exógena, conforme resultado apresentado na Tabela 3.

- ▶ **Imposto de Importação:** o modelo nº 1 não consegue aproximar a previsão obtida pelo método dos indicadores, além de apresentar significância no nível de 10% no *Reset* teste, o que indica má especificação do modelo. Não se conseguiu chegar a resultado melhor com a modelagem dinâmica univariada.
- ▶ **Imposto sobre a Renda da Pessoa Jurídica:** o modelo nº 1 apresenta ótimos resultados, tanto com relação à congruência do modelo quanto com relação às ótimas previsões para 2007, com RMSE e erro anual de previsão 46% e 71% inferiores, respectivamente, ao obtido pelo método dos indicadores. Adite-se que este foi o melhor resultado obtido para o IRPJ, superando todas as metodologias desenvolvidas neste trabalho. Tendo em vista os resultados dos testes de exogeneidade das variáveis explicativas, estimou-se modelo de função de transferência com DLPIB como variável exógena, que, porém, não apresentou bons resultados, já que nenhuma defasagem de DLPIB se mostrou significativa.
- ▶ **Cofins:** o modelo nº 1 apresenta resultado próximo ao apresentado pelo método dos indicadores. Quando se acrescenta DLPIB como variável exógena (modelo TFM nº 2), o coeficiente da primeira defasagem, DLPIB\_1, mostra-se bastante significativo e há sensível melhora nos erros de previsão, com RMSE 15% inferior ao obtido pelo método dos indicadores. O modelo nº 3, função de transferência com DLIVV como variável exógena, melhora ainda mais os resultados, com RMSE 21% inferior ao obtido pelo método dos indicadores e erro anual de previsão de apenas 1%.

Tabela 5. Modelos dinâmicos univariados e de função de transferência – DR/TFM – II/IRPJ\_DOLR/COFINS\_DE

Modelo		Testes					Erros de previsão			
Nº	Especificação (*1)	CP (*2)	AR (*3)	N (*4)	H (*5)	RESET (*6)	MAD	RMSE	MAPE (%)	Anual (%)
<b>Série: II – Intervalo dos dados: 08/2000 a 12/2006</b>										
1	DLII	0,52 (0,89)	1,27 (0,29)	0,95 (0,62)	1,13 (0,36)	3,05 (0,09)	27,1	31,2	9,3	-9,4
	<b>Modelo RFB</b>						15,6	22,6	5,7	2,5
<b>Série: IRPJ_DOLR – Intervalo dos dados: 02/2001 a 12/2006</b>										
1	DLLR	1,51 (0,15)	1,03 (0,41)	0,89 (0,64)	0,70 (0,79)	1,09 (0,30)	72,8	82,7	11,1	-5,0
	<b>Modelo RFB</b>						132,6	152,2	19,3	-17,3
<b>Série: COFINS_DE – Intervalo dos dados: 03/2001 a 12/2006</b>										
1	DLCF	0,85 (0,60)	0,23 (0,95)	1,17 (0,56)	0,46 (0,91)	0,43 (0,51)	71,9	90,6	6,0	-5,0
2	DLCF	0,86	1,28	1,96	0,77	1,48	55,3	72,5	4,6	-3,2
(*7)	(DLPIB_1)	(0,59)	(0,28)	(0,38)	(0,63)	(0,23)	62,6	80,4	5,2	-3,8
3	DLCF	1,04	0,71	1,35	0,32	0,18	47,2	67,0	4,1	-1,0
(*7)	(DLIW , DLIWV_1)	(0,42)	(0,62)	(0,51)	(0,97)	(0,67)	53,5	73,7	4,6	-2,3
	<b>Modelo RFB</b>						66,3	85,0	5,6	-4,5

Fonte: Elaboração do autor

Notas: (\*1) A variável entre parênteses indica a variável, com respectiva defasagem, utilizada como exógena nos modelos TFM.

(\*2) Teste Chow de constância de parâmetros. Segue uma distribuição  $F(h, T - k)$ , em que  $h$  é o horizonte de previsão (12),  $T$  o número de observações e  $k$  o número de parâmetros.

(\*3) Teste de autocorrelação dos resíduos, com cinco defasagens segundo distribuição  $F(s, T - k - s)$ , em que  $s$  é o número de defasagens utilizadas.

(\*4) Teste de normalidade dos resíduos. Segue distribuição  $\chi^2(2)$ .

(\*5) Teste de heteroscedasticidade. Teste F, baseado em White (1980).

(\*6) Teste *Reset* (*Ramsey Regression Equation Specification Error Test*). Testa se o modelo está bem especificado. Segue distribuição  $F(1, T - k - 1)$ .

(\*7) Teste Chow de constância de parâmetros. Segue uma distribuição  $F(h, T - k)$ , em que  $h$  é o horizonte de previsão (12),  $T$  o número de observações e  $k$  o número de parâmetros.

Entretanto, o primeiro resultado dos erros de previsão para os modelos de função de transferência nº 2 e nº 3 utilizam os dados reais das variáveis exógenas, DLPIB e DLIVV, respectivamente, para o ano de 2007. Nesse caso, constata-se a utilidade dessas variáveis, pois os erros de previsão diminuem. Assim, num segundo momento, foi efetuada a previsão dessas variáveis explicativas para o ano de 2007, por meio das diversas técnicas utilizadas neste trabalho, conforme explicitado no Capítulo 3. A incorporação desses valores previstos aos modelos de função de transferência levou ao segundo resultado apresentado para os erros de previsão.

A melhor previsão para o PIB em 2007, obtida por meio de modelagem dinâmica univariada, obteve variação positiva de 4,7% em relação a 2006, enquanto o PIB real de 2007 foi 5,2% superior ao de 2006. Essa previsão 10% inferior ao valor real fez com que o erro anual de previsão aumentasse de 3,2% para 3,8%, ou seja, um aumento de 0,6 ponto percentual, porém ainda mantendo-se inferior aos obtidos pelo método dos indicadores e pelo modelo dinâmico univariado nº 1, que foram de 4,5% e 5,0%, respectivamente.

Com relação ao IVV\_sp, o melhor resultado foi obtido pela modelagem ARIMA e prevê variação positiva para o ano de 2007 de 8,7% em relação a 2006. A variação real foi de 12,5%, ou seja, 44% superior, o que elevou o erro anual de previsão de 1,0% para 2,3%, mantendo-se, também, inferior ao obtido pelo método dos indicadores e pelo modelo dinâmico univariado nº 1.

### **5.3 Modelos dinâmicos de múltiplas equações**

Os resultados dos modelos dinâmicos de múltiplas equações são apresentados na Tabela 6. A tabela apresenta, para cada especificação das variáveis envolvidas, o resultado do modelo VEC\_SEM reduzido para  $I(0)$ , que é o modelo final apropriado para se efetuar previsões dinâmicas.<sup>25</sup>

---

<sup>25</sup> Os resultados dos modelos VAR irrestrito (URF\_VAR), do modelo VAR cointegrado (VEC) ou do modelo URF\_VAR em diferenças (quando não há presença de relações de cointegração entre as variáveis) não são mostrados devido à falta de espaço.



Tabela 6. Modelos dinâmicos multivariados – VAR/VEC/SEM – II/IRPJ\_DOLR/COFINS\_DE

Modelo			Testes				Erros de Previsão			
Nº	Especificação (*1)	Modelo final (*2)	CP (*3)	AR (*4)	N (*5)	H (*6)	MAD	RMSE	MAPE (%)	Anual (%)
<b>Série: II – Intervalo dos dados: 07/2000 a 12/2006</b>										
1	DLII/DLIPI/DLTCR	VEC_SEM	25,13 (0,91)	0,96 (0,54)	0,91 (0,99)	0,56 (1,00)	11,0	14,2	4,0	0,7
2	DLII/DLIVV/DLTCR	VEC_SEM	29,38 (0,77)	1,57 (0,03)	3,06 (0,80)	0,69 (0,98)	13,0	14,5	4,6	-1,2
3	DLII/DLPIB/DLTCR	VEC_SEM	33,73 (0,58)	0,83 (0,75)	1,58 (0,95)	0,55 (1,00)	13,1	16,5	4,7	0,4
	<b>Modelo RFB</b>						<b>15,6</b>	<b>22,6</b>	<b>5,7</b>	<b>2,5</b>
<b>Série: IRPJ_DOLR – Intervalo dos dados: 02/2001 a 12/2006</b>										
1	DLLR/DLIPI	VEC_SEM	17,98 (0,80)	1,17 (0,29)	1,58 (0,81)	0,63 (0,96)	85,5	91,9	13,0	-3,3
2	DLLR/DLPIB	VEC_SEM	35,10 (0,07)	1,02 (0,45)	1,15 (0,89)	0,96 (0,55)	97,7	111,2	14,4	-5,4
3	DLLR/DLIVV	SEM	30,60 (0,17)	1,08 (0,38)	2,27 (0,69)	0,83 (0,78)	122,5	148,2	18,0	-17,4
	<b>Modelo RFB</b>						<b>132,6</b>	<b>152,2</b>	<b>19,3</b>	<b>-17,3</b>
<b>Série: COFINS_DE – Intervalo dos dados: 02/2001 a 12/2006</b>										
1	DLCF/DLIVV	VEC_SEM	58,04 (0,00)	0,67 (0,85)	1,40 (0,84)	0,89 (0,68)	34,6	48,5	3,1	0,4
2	DLCF/DLPIB	VEC_SEM	18,04 (0,80)	0,42 (0,98)	0,75 (0,95)	0,77 (0,83)	54,7	67,1	5,0	4,1
3	DLCF/DLIPI	VEC_SEM	17,06 (0,85)	0,82 (0,68)	4,43 (0,35)	0,61 (0,97)	77,5	90,1	6,5	-5,4
	<b>Modelo RFB</b>						<b>66,3</b>	<b>85,0</b>	<b>5,6</b>	<b>-4,5</b>

Fonte: Elaboração do autor

Notas: (\*1) Variáveis utilizadas no modelo.

(\*2) VEC\_SEM = modelo de representação de correção de erro – modelo de equações simultâneas.

(\*3) Teste de constância de parâmetros usando  $V[e]$  (VEC\_SEM), definida como a matriz de variância dos erros de previsão, a qual leva em conta a incerteza dos parâmetros. Segue distribuição  $\chi^2(nh)$ , em que  $n$  é o número de variáveis dependentes e  $h$  o horizonte de previsão.(\*4) Teste vetorial de autocorrelação dos resíduos, com cinco defasagens segundo distribuição  $F(np, Nr-q)$ , em que  $n$  é a dimensão do sistema;  $p = ns$ ;  $s$  é o número de defasagens utilizadas (5);  $N = T - k - p - (n - p + 1)/2$ ;  $r = [(n^2p^2 - 4) / (n^2 + p^2 - 5)]^{1/2}$ ;  $q = (np)/2 - 1$ ;  $T$  é o número de observações;  $k$  é o número de regressores no sistema original. Para maiores detalhes, ver Doornik (1996).(\*5) Teste vetorial de normalidade dos resíduos. Segue distribuição  $\chi^2(2n)$ , em que  $n$  é o número de variáveis dependentes.(\*6) Teste vetorial de heteroscedasticidade. Segue distribuição  $F(h, Ns-q)$ . Para detalhes, ver Doornik (1996).

► **Imposto de Importação:** na especificação URF\_VAR, que origina o modelo nº 1 da Tabela 6, a *proxy* do PIB utilizada é o Índice de Produção Industrial. O URF\_VAR contém quatro *lags*, constante, tendência, sazonalidade, *outliers* do tipo AO em 10/2002, em 11/2002, em 02/2003 e do tipo LS em 11/2001, em 05/2002, em 01/2003 e em 04/2005 para o LTCR, do tipo AO em 12/2001 e LS em 09/2003 para o LIPI e do tipo AO em 04/2004 para o LII. O sistema apresenta duas relações de cointegração, sem que se possa considerar LIPI ou LTCR como variáveis exógenas. O modelo nº 1, redução final para  $I(0)$ , apresenta ótimos resultados de previsão, com RMSE 37% inferior ao do método dos indicadores e erro anual de previsão de apenas 0,7%.

A especificação URF\_VAR, que origina o modelo nº 2 da Tabela 6, envolve LII/LIVV/LTCR, contendo três *lags*, constante, tendência, sazonalidade e *outliers* do tipo AO em 10/2002, em 11/2002, em 02/2003 e em 02/2004 e do tipo LS em 11/2001, em 05/2002, em 01/2003 e em 04/2005. Com exceção de 02/2004, relativa ao LIVV, todos os demais foram incluídos em função da análise dos resíduos de LTCR. O sistema apresenta duas relações de cointegração, sem que se possa considerar LIVV ou LTCR como variáveis exógenas. O modelo nº 2, redução final para  $I(0)$ , também apresenta ótimos resultados de previsão, com RMSE 36% inferior ao do método dos indicadores e erro anual de previsão de apenas 1,2%.

Por fim, no modelo nº 3 da Tabela 6, utiliza-se o próprio PIB. O URF\_VAR contém três *lags*, constante, sazonalidade, os mesmos *outliers* para LTCR e do tipo AO em 09/2000, em 06/2001, em 06/2005 e em 04/2006 para o LPIB. O sistema possui duas relações de cointegração, sem que se possa considerar LPIB ou LTCR como variáveis exógenas. O modelo nº 3, redução final para  $I(0)$ , apresenta bons resultados de previsão, um pouco superiores ao modelo com LIVV e LIPI, com RMSE 27% inferior ao do método dos indicadores e erro anual de previsão de apenas 0,4%.

Assim, para o Imposto de Importação, o melhor sistema, tanto em termos de congruência do modelo, quanto com relação aos erros de previsão, é o formado pelo modelo nº 1, com o Índice de Produção Industrial como variável explicativa, que, além de superar o método dos

indicadores, apresenta RMSE 26%, 54% e 50% inferior ao dos melhores modelos ARIMA, dinâmico univariado e estrutural, respectivamente.

► **Imposto sobre a Renda da Pessoa Jurídica:** no modelo nº 1 da Tabela 6, a variável utilizada como *proxy* do PIB é o Índice de Produção Industrial. O modelo URF\_VAR contém quatro *lags*, constante, tendência, sazonalidade e *outlier* do tipo AO em 05/2003 para o LLR. O sistema apresenta uma relação de cointegração e DLPI não pode ser considerada exógena. O modelo nº 1, redução final para I(0), apresenta os melhores resultados de previsão, com RMSE 40% inferior ao do método dos indicadores e erro anual de previsão de apenas 3,3%, 81% inferior ao do método dos indicadores.

No modelo nº 2 da Tabela 6, PIB é utilizada no VAR. O modelo URF\_VAR contém quatro *lags*, constante, tendência, sazonalidade e *outliers* do tipo AO em 05/2003 e em 01/2006, do tipo LS em 08/2004 para o LLR e do tipo AO em 01/2001 para o LPIB. O sistema apresenta uma relação de cointegração. O modelo nº 2, redução final para I(0), apresenta bons resultados de previsão, com RMSE 27% inferior ao do método dos indicadores e erro anual de previsão de apenas 5,4%, ou seja, quase 70% inferior ao do método dos indicadores. Nesse modelo, o fato de  $\alpha_{1\_DLPIB} = 0$ , juntamente com o fato de que defasagens de DLLR não entram na equação de DLPIB, faz desta última uma variável exógena forte.

No modelo nº 3 da Tabela 6, a *proxy* do PIB passa a ser o Índice de Vendas no Varejo. O modelo URF\_VAR contém seis *lags*, tendência, sazonalidade e um *outlier* do tipo AO em 05/2003 para o LLR. O sistema apresenta posto da matriz  $\Pi$  igual a zero, ou seja, não há relação de cointegração entre LLR e LIVV. Nesse caso estimou-se um modelo VAR em diferenças, com cinco *lags* das variáveis em diferenças. O modelo nº 3, redução final do modelo VAR em diferenças, com algumas restrições impostas, não apresenta resultados satisfatórios em termos de previsão. Ainda assim, o RMSE é ligeiramente inferior ao do método dos indicadores. Nesse modelo, DLLR causa DLIVV no sentido de Granger.

Portanto, para o Imposto sobre a Renda da Pessoa Jurídica, o melhor sistema, tanto em termos de congruência do modelo, quanto com

relação aos erros de previsão, é o representado pelo modelo nº 1, com o Índice de Produção Industrial como variável explicativa, que, além de superar o método dos indicadores, apresenta RMSE 25% inferior ao do modelo ARIMA. Entretanto, o modelo dinâmico univariado em diferenças apresentou o melhor resultado para esse tributo.

► **Cofins:** no modelo nº 1 da Tabela 6, a *proxy* do PIB utilizada é o Índice de Vendas no Varejo. O modelo URF\_VAR contém cinco *lags*, tendência, sazonalidade e um *outlier* do tipo LS em 07/2004 para o LCF. O sistema apresenta uma relação de cointegração. O modelo nº 1, redução final para  $I(0)$ , exhibe ótimos resultados de previsão, com RMSE e erro anual de previsão 43% e 91% inferiores aos do método dos indicadores, respectivamente. Nesse modelo, o fato de  $\alpha_{1\_DLIVV} = 0$  faz de DLIVV uma variável fracamente exógena, já que o *lag* 3 de DLCF é significativo na equação de DLIVV. O problema desse modelo está no teste de constância de parâmetros, que se mostrou altamente significativo desde o início da modelagem do URF\_VAR.

No modelo nº 2 da Tabela 6, PIB é a variável explicativa utilizada. O modelo URF\_VAR contém quatro *lags*, sazonalidade e um *outlier* do tipo AO em 06/2001 para LPIB e do tipo LS em 07/2004 para o LCF. O sistema apresenta uma relação de cointegração. O modelo nº 2, redução final para  $I(0)$ , exhibe bons resultados de previsão, com RMSE 21% inferior ao do método dos indicadores. Nesse modelo, o fato de  $\alpha_{1\_DLPIB} = 0$ , juntamente com o fato de que defasagens de DLCF não entram na equação de DLPIB, faz desta última uma variável exógena forte.

No modelo nº 3 da Tabela 6, a variável utilizada como *proxy* do PIB é o Índice de Produção Industrial. O modelo URF\_VAR contém quatro *lags*, constante, tendência, sazonalidade e um *outlier* do tipo LS em 07/2004 para o LCF. O sistema apresenta uma relação de cointegração e DLIPI não pode ser considerada exógena. O modelo nº 3, redução final para  $I(0)$ , exhibe resultados de previsão próximos aos alcançados pelo método dos indicadores, porém sem superá-lo.

Dessa forma, para a Cofins, o melhor sistema, com relação aos erros de previsão, é o formado pelo modelo nº 1, com o Índice de Vendas no Varejo como variável explicativa, que, além de superar o método dos

indicadores, apresenta RMSE 46%, 28% e 12% inferior ao dos modelos ARIMA, dinâmico univariado e estrutural, respectivamente. Deve-se enfatizar, também, o baixíssimo erro de previsão anual do modelo, de apenas 0,4%.

#### 5.4 Modelos estruturais

Os modelos estruturais foram estimados por meio do *software Stamp (Structural Time Series Analyser, Modeller and Predictor)*, que é um módulo do *GiveWin*. A partir do modelo básico estrutural (BSM – Modelo F – Harvey (1989), p. 510-511), foi implementada uma rotina para detecção de *outliers* com base numa série ajustada sazonalmente. Identificados os *outliers*, tanto do tipo AO quanto LS, somente os mais significativos foram inseridos no modelo estrutural da série original (sem ajuste sazonal) e efetuada a estimação. Os resultados, sem e com a inclusão de variáveis exógenas, encontram-se na Tabela 7.

- ▶ **Imposto de Importação:** os resultados dos modelos estruturais não foram satisfatórios. O modelo nº 1 apresenta erros de previsão bastante superiores aos dos demais modelos tratados neste trabalho. Foram detectados *outliers* mais significativos, do tipo LS, nos meses de dezembro de 2001, maio de 2003 e março de 2004, os quais, quando incorporados aos modelos, resultaram em previsões menos acuradas. O modelo nº 2, com a incorporação da terceira defasagem de LII como variável explicativa, praticamente não melhora o poder preditivo em relação ao modelo nº 1, com alguma melhora nos critérios de ajustamento AIC e BIC e diminuição da autocorrelação dos resíduos.
- ▶ **Imposto sobre a Renda da Pessoa Jurídica:** nesse caso, os *outliers* mais significativos, já incorporados ao modelo nº 1, em função da não normalidade dos resíduos do modelo básico sem eles, são do tipo AO no mês de maio de 2003 e LS no mês de agosto de 2004. Esse modelo já apresenta resultado superior ao do método dos indicadores e similar ao ARIMA. O modelo nº 2, com a incorporação da segunda

defasagem de LIVV como variável explicativa<sup>26</sup> e somente o *outlier* do tipo AO de maio de 2003, melhora bastante o resultado, com RMSE 45% inferior ao do método dos indicadores e da mesma ordem de grandeza do RMSE do modelo dinâmico univariado, que é o melhor resultado para essa série. A inclusão do PIB como variável exógena não mostrou bons resultados.

- ▶ **Cofins:** o modelo básico, nº 1, sem *outliers* ou variáveis exógenas, já apresenta resultado do erro de previsão RMSE 18% inferior ao do método dos indicadores. O modelo nº 2, com a inclusão do *outlier* do tipo LS no mês de julho de 2004, melhora sensivelmente o poder preditivo em relação ao modelo básico nº 1. O modelo nº 3, em que se acrescenta a primeira defasagem de LCF, introduz pequena melhora ao modelo anterior, com RMSE 35% inferior ao do método dos indicadores e pouco superior ao do modelo de equações simultâneas nº 1 (DLCF/DLIVV) da Tabela 6, que apresenta o melhor poder preditivo entre todas as modelagens. A inclusão das variáveis exógenas, PIB ou Índice de Vendas no Varejo, não conduz a resultados satisfatórios.

---

<sup>26</sup> Embora o resultado do teste de exogeneidade das variáveis explicativas, apresentado em 4.3, indique que IVV\_sp deve ser tratada como endógena, pelo fato de não haver relação de cointegração entre LLR e LIVV, foi estimado este modelo EMX com LIVV como variável exógena.

Tabela 7. Modelos estruturais – EM e EMX – II/IRPJ\_DOLR/COFINS\_DE

Mod. nº	Var. dep.	Componentes do modelo estrutural (*1)						Testes Estat. (t-prob) (*2)				Erros de previsão			
		N	T	S	I	Outlier	Variável X_L	N	Estatística Q (P,d)	H	MAD	RMSE	MAPE (%)	E_anual (%)	
<b>Série: II – Intervalo dos dados: 07/2000 a 12/2006</b>															
1	LII	SL	NS	FS	S	-	-	0,18 (0,91)	Q (7,6)	10,72 (0,10)	0,40 (0,99)	27,5	30,1	9,5	-9,7
2	LII	SL	NS	FS	S	-	LII_3	0,35 (0,84)	Q (7,3)	3,61 (0,31)	0,52 (0,95)	25,5	28,3	8,8	-9,0
<b>RFB</b>												<b>15,6</b>	<b>22,6</b>	<b>5,7</b>	<b>2,5</b>
<b>Série: IRPJ_DOLR – Intervalo dos dados: 02/2001 a 12/2006</b>															
1	LLR	SL	FS	TS	S	I200305 L200408	-	0,07 (0,96)	Q (8,6)	10,35 (0,11)	1,83 (0,10)	104,5	123,8	15,4	-11,6
2(*3)	LLR	SL	FS	TS	S	I200305	LIVV_2	1,33 (0,51)	Q (8,4)	3,92 (0,42)	1,84 (0,10)	67,9 78,3	83,5 92,4	10,5 11,9	-1,6 -4,8
<b>RFB</b>												<b>132,6</b>	<b>152,2</b>	<b>19,3</b>	<b>-17,3</b>
<b>Série: COFINS_DE – Intervalo dos dados: 02/2001 a 12/2006</b>															
1	LCF	SL	FS	FS	S	-	-	0,22 (0,90)	Q (7,6)	3,42 (0,75)	1,25 (0,30)	54,3	69,7	4,6	-3,2
2	LCF	SL	SL	FS	S	L200407	-	0,23 (0,89)	Q (8,6)	6,09 (0,41)	1,20 (0,34)	44,1	57,1	3,8	-1,6
3	LCF	SL	SL	FS	S	L200407	LCF_1	0,25 (0,88)	Q (8,5)	5,20 (0,39)	1,16 (0,36)	43,8	55,2	3,8	-1,2
<b>RFB</b>												<b>66,3</b>	<b>85,0</b>	<b>5,6</b>	<b>-4,5</b>

Fonte: elaboração do autor

Notas: (\*1) N = Nível; SL (estocástico); FL (fixo); NL (ausente).

T = tendência ou inclinação; SS (estocástica); FS (fixa); NS (ausente).

S = componente sazonal; DS (dummy); TS (trigonométrico); FS (fixo); NS (ausente).

I = componente irregular; S (sim); N (não).

Outlier: identificação do tipo do outlier, I para tipo aditivo (AO) e L para alteração no nível (LS), seguido do ano e do mês de ocorrência.

Variável X\_L: identificação da variável explicativa (X) com respectiva defasagem (L).

(\*2) N: teste de normalidade dos resíduos de Doornik e Hansen (1994). Segue distribuição  $\chi^2(2)$ .Q(P,d): estatística Q de Box-Ljung com base nas primeiras P autocorrelações dos resíduos, distribuída aproximadamente como  $\chi^2(d)$ , em que  $d = P + 1$  – número de parâmetros do modelo. Estatística ajustada pelo número de parâmetros do modelo.

H: teste de heteroscedasticidade não paramétrico com distribuição F(h,h), em que h é o inteiro mais próximo de T/3.

(\*3) Trata-se de modelo EMX, em que o primeiro resultado dos erros de previsão utiliza os dados reais da variável exógena LIVV\_2 para o ano de 2007, enquanto no segundo resultado são utilizados os dados previstos desta mesma variável, conforme detalhado no Capítulo 3.

## 6 CONCLUSÕES

O objetivo geral do presente trabalho foi apresentar métodos alternativos ao método dos indicadores, utilizado pela Secretaria da Receita Federal do Brasil para a previsão da arrecadação tributária federal, por meio de metodologias de séries temporais. Foram utilizadas as metodologias ARIMA, modelos dinâmicos univariados e de função de transferência, modelos dinâmicos de múltiplas equações e modelos estruturais. As séries de arrecadação analisadas referem-se ao Imposto de Importação, ao Imposto sobre a Renda da Pessoa Jurídica – Demais Obrigadas à Apuração do Lucro Real e à Contribuição para o Financiamento da Seguridade Social – Demais Empresas. A escolha dos tributos analisados levou em conta a participação de cada um na arrecadação total do Estado de São Paulo para o ano de 2007, bem como o âmbito de incidência dos tributos, que no presente estudo é, respectivamente, o setor externo, o lucro e o faturamento das empresas.

Foram estimados diversos modelos em cada uma das metodologias propostas. A Tabela 8 apresenta o resultado da comparação dos erros de previsão entre si e com o método dos indicadores. A medida de acurácia da previsão faz-se por meio da utilização da raiz do erro quadrático médio da previsão (RMSE) do melhor modelo de cada metodologia, para o ano de 2007. São indicados, também, os erros anuais de previsão, devido à importância em relação ao orçamento anual, bem como em relação às metas de arrecadação a serem cumpridas pela RFB.

Tabela 8. Comparação do RMSE e do erro anual de previsão do melhor modelo de cada metodologia

Série	MI/RFB	ARIMA		DR/TFM		VEC/SEM		EM/EMX	
	RMSE E_anual	Mod. (tab)	RMSE E_anual	Mod. (tab)	RMSE E_anual	Mod. (tab)	RMSE E_anual	Mod. (tab)	RMSE E_anual
II	22,6 +2,5%	1 (4)	19,3 +2,4%	1 (5)	31,2 -9,4%	<b>1 (6)</b>	<b>14,2</b> <b>+0,7%</b>	2 (7)	28,3 -9,0%
IRPJ_DOLR	152,2 -17,3%	1 (4)	123,0 -13,6%	<b>1 (5)</b>	<b>82,7</b> <b>-5,0%</b>	1 (6)	91,9 -3,3%	2 (7)	83,5 -1,6%
COFINS_DE	85,0 -4,5%	1 (4)	89,4 -5,5%	3 (5)	67,0 -1,0%	1 (6)	<b>48,5</b> <b>+0,4%</b>	3 (7)	55,2 -1,2%

Fonte: Elaboração do autor



Para a série do Imposto de Importação, o modelo nº 1 da Tabela 6, modelo dinâmico multivariado VEC\_SEM, formado por DLII/DLIPI/DLTCR, apresentou o melhor resultado, com RMSE 37% inferior ao do método dos indicadores e 26% inferior ao do modelo ARIMA, além de o erro de previsão anual ser de apenas 0,7%, em face dos 2,5% apresentados pelo método dos indicadores. Em seguida aparece o modelo nº 1 da Tabela 4, ARIMA, com RMSE cerca de 15% inferior ao do método dos indicadores. As modelagens dinâmica univariada e estrutural não apresentaram resultados satisfatórios, sem conseguir alcançar os erros de previsão obtidos pela RFB.

Para o Imposto sobre a Renda da Pessoa Jurídica, todas as metodologias mostraram-se superiores ao método dos indicadores. O modelo ARIMA nº 1 da Tabela 4 já apresenta melhora em relação ao modelo da RFB, com RMSE quase 20% inferior. As demais metodologias apresentam resultados semelhantes, com sensível melhora tanto no RMSE quanto no erro anual de previsão. O modelo de equações simultâneas nº 1 da Tabela 6, formado por DLLR/DLIPI, apresenta RMSE 40% inferior ao do método dos indicadores. Já os modelos nº 2 da Tabela 7, estrutural com Índice de Vendas no Varejo como variável exógena, e nº 1 da Tabela 5, dinâmico univariado, exibem resultados semelhantes, com RMSE 46%, 33% e 10% inferiores aos do método dos indicadores, do modelo ARIMA e de equações simultâneas, respectivamente. Ressalte-se que o modelo estrutural apresenta um erro de previsão anual de apenas 1,6% perante os 17,3% apresentados pelo método dos indicadores, ou seja, redução de 91%.

Finalmente, para a série da Contribuição para o Financiamento da Seguridade Social, o modelo ARIMA nº 1 da Tabela 4 apresenta RMSE aproximadamente igual ao do método dos indicadores, porém sem superá-lo. Em seguida aparece o modelo nº 3 da Tabela 5, função de transferência com Índice de Vendas no Varejo como variável exógena, com RMSE 21% inferior. O modelo estrutural nº 3 da Tabela 7 melhora ainda mais o resultado, com o RMSE 35% inferior ao do método dos indicadores. Como melhor modelo aparece o de nº 1 da Tabela 6, modelo de equações simultâneas com DLCF/DLIVV, com RMSE 43%, 46%, 28% e 12% inferior ao do método dos indicadores e dos modelos

ARIMA, função de transferência e estrutural, respectivamente. Ademais, o erro anual de previsão é de apenas 0,4%, ou seja, 91% inferior ao erro obtido pelo método dos indicadores.

Quanto à utilização das variáveis explicativas, tanto na modelagem de múltiplas equações, quanto nos modelos de função de transferência, chega-se à conclusão de que, para o Imposto de Importação, nos modelos VAR/VEC/SEM, com as variáveis explicativas Índice de Vendas no Varejo ou Índice de Produção Industrial e Taxa de Câmbio Real, os resultados foram muito bons e semelhantes. A utilização do PIB nos modelos também oferece bons resultados, embora um pouco inferiores aos demais. Os resultados dos testes de exogeneidade das variáveis explicativas não permitem considerar a Taxa de Câmbio Real, o Produto Interno Bruto, o Índice de Vendas no Varejo ou o Índice de Produção Industrial como variáveis exógenas fortes e, assim, essas variáveis não são adequadas às metodologias de função de transferência e estrutural com variáveis exógenas. O pré-requisito de validade desses modelos é o de que as variáveis explicativas possuam exogeneidade forte, ou seja, exogeneidade fraca juntamente com não causalidade de Granger no sentido da variável dependente para a variável explicativa. Todas as combinações de variáveis utilizadas, Imposto de Importação e Taxa de Câmbio Real com PIB, Índice de Vendas no Varejo ou Índice de Produção Industrial, apresentaram duas relações de cointegração, o que indica relações de longo prazo entre essas variáveis.

Ainda com relação ao Imposto de Importação, de maneira geral, a modelagem de múltiplas equações (VAR/VEC/SEM) apresentou desempenho superior ao das modelagens ARIMA, dinâmica univariada e estrutural. É importante destacar a sensível melhora nas previsões dos modelos com a inclusão das variáveis explicativas, principalmente em relação aos modelos dinâmico univariado e estrutural.

No caso do Imposto sobre a Renda da Pessoa Jurídica, o Índice de Produção Industrial parece ser a melhor variável a utilizar na modelagem multivariada, seguido do PIB. Dados de 2007 da arrecadação do IRPJ\_DOLR no Estado de São Paulo indicam que o setor das indústrias de transformação foi responsável por 47% da arrecadação e esta

é, possivelmente, a explicação para o Índice de Produção Industrial apresentar melhor resultado do que o PIB na modelagem multivariada. IRPJ/IVV\_sp foi a única combinação de variáveis que não apresentou relação de cointegração, e o VAR em diferenças estimado para essa relação não exibiu bons resultados. As demais combinações, IRPJ/PIB e IRPJ/IPI\_g\_sp, mostram relação de longo prazo entre as variáveis. O modelo estrutural, com a utilização do Índice de Vendas no Varejo como variável exógena, apresentou resultado satisfatório em relação ao método dos indicadores e à modelagem ARIMA. Porém, a modelagem univariada superou todas as demais. Isso pode ter acontecido devido ao fato de que o IRPJ depende do lucro real apurado pelas empresas, que pode não ter relação direta com o nível da atividade econômica representado pelas variáveis PIB, IVV\_sp e IPI\_g\_sp.

Para a Cofins, o Índice de Vendas no Varejo e o PIB parecem ser as melhores alternativas tanto para os modelos de múltiplas equações quanto para os modelos de função de transferência. Como a Cofins depende do faturamento das empresas, parece intuitivo que essas duas variáveis tenham apresentado os melhores resultados. O Índice de Produção Industrial, por sua vez, mede a produção industrial, que pode não ter relação direta com o faturamento das empresas. Todas as combinações de variáveis, Cofins/PIB, Cofins/IVV\_sp e Cofins/IPI\_g\_sp, mostram relação de longo prazo.

No que diz respeito aos erros de previsão, verifica-se na Tabela 1 que há uma tendência de diminuição dos erros de previsão cometidos pelo método dos indicadores, medidos pelo MAPE, ao longo dos anos, provavelmente devido à maior estabilidade econômica dos últimos anos. Assim, merece atenção o fato de que os erros médios do II, IRPJ\_DE e COFINS\_DE são maiores que os erros cometidos no ano de 2007, base de comparação, especialmente no caso do II, em que a média do MAPE no período 2001 a 2007 é de 18,6%, enquanto o erro de 2007 é de apenas 5,7%. Ou seja, deve-se levar em consideração que os valores previstos pelos métodos desenvolvidos neste trabalho estão sendo comparados com previsões efetuadas pelo método dos indicadores para o ano de 2007, em que os erros de previsão ficaram abaixo da média do período 2001-2007, o que reforça ainda mais os resultados alcançados.

Ainda com relação ao desempenho do método dos indicadores, quanto mais estável o ambiente econômico, melhor será sua *performance*, que é totalmente dependente da arrecadação do mesmo período do ano anterior, além de uma boa previsão dos indicadores a serem utilizados. Por sua vez, metodologias de séries temporais geralmente trabalham com uma história maior dos dados, portanto, com o decorrer dos anos, quanto mais informação for acrescentada à série, e desde que adequadamente tratada, mais robustos e melhores se tornarão os resultados.

Outro ponto que merece consideração diz respeito ao tratamento e à identificação das arrecadações atípicas (*outliers*). Uma identificação inicial dos *outliers* e das quebras estruturais foi efetuada por meio da utilização dos resíduos auxiliares obtidos da estimação de modelos estruturais. Porém, os *outliers* identificados não foram incorporados aos modelos ARIMA, e esta pode ter sido uma das causas da não obtenção de um melhor resultado desses modelos. É importante salientar que, mesmo sem tratamento mais profundo, os modelos ARIMA superam ou igualam o método dos indicadores.

Por fim, merece destaque o fato de que as séries de arrecadação utilizadas no trabalho se referem à arrecadação divulgada pela RFB, que é a base de comparação utilizada pelo órgão para se apurar a acurácia da previsão, e este foi o motivo da escolha dessa série, ou seja, ter uma base de comparação com a previsão efetuada pela RFB. Entretanto, essa arrecadação compõe-se de muitas rubricas, tais como retificações, restituições, compensações, Refis, Paes, etc. Muitas dessas rubricas podem ser consideradas atípicas, como é o caso, por exemplo, do Refis e do Paes, que são programas de refinanciamento de dívidas antigas, além de restituições e compensações.<sup>27</sup> Assim, o fato de se usar a arrecadação divulgada pode ter contribuído para o surgimento de *outliers* nas séries. Na série do Imposto de Importação, esse fato parece não ter muita relevância, com as duas arrecadações sendo bastante próximas. Contudo, para as séries do Imposto sobre a

---

<sup>27</sup> Restituições são consideradas arrecadações negativas, enquanto a compensação entre tributos de diferentes espécies implica arrecadação positiva para o tributo devido e arrecadação negativa para o tributo com crédito.

Renda da Pessoa Jurídica e da Cofins, há meses em que a diferença chega a 10%, o que é significativo e pode alterar o resultado. Assim, uma opção de continuação do trabalho, que me parece apropriada, é a utilização da série da arrecadação bruta classificada. Esta é o somatório da arrecadação oriunda dos Documentos de Arrecadação das Receitas Federais (Darf) após a classificação por um sistema interno, que, no caso de pagamentos unificados, como é o caso, por exemplo, do Simples, retenção de órgãos públicos e dívida ativa da União, “explode” o código de recolhimento em seus diversos componentes.<sup>28</sup> Portanto, essa série está livre de várias fontes de *outliers*. Ademais, para as séries IRPJ\_DE e COFINS\_DE, o início da série da arrecadação divulgada ocorre em agosto de 2000, enquanto para a arrecadação bruta classificada o início se dá em janeiro do mesmo ano, o que proporciona um maior número de observações.

Como conclusão final, a aplicação dos métodos desenvolvidos neste trabalho, considerados os melhores modelos de cada série, levou à redução média do RMSE de 42% em relação ao erro cometido pelo método dos indicadores e de 35% em relação à modelagem ARIMA, além da drástica redução do erro anual de previsão. A técnica foi aplicada a três tributos que são representativos do sistema tributário e, a partir de agora, seria interessante aplicá-la aos demais tributos administrados pela RFB. Como toda técnica, deve ser aprimorada constantemente, em especial com relação à tentativa de escolha de outras variáveis explicativas a serem incluídas nos sistemas multivariados.

Este trabalho mostrou que a utilização de metodologias de séries temporais para a previsão da arrecadação de receitas federais é uma alternativa ao método dos indicadores atualmente utilizado pela Receita Federal do Brasil, contribuindo para previsões mais precisas, tornando-se ferramenta segura de apoio para a tomada de decisões dos gestores.

---

28 Por exemplo, o Simples recolhido no código 6106 é explodido nos códigos 7104 – IRPJ, 7606 – Cofins, 7200 – PIS e 7307 – CSSL.



## REFERÊNCIAS

---

- BARRO, R.J. *Macroeconomics*. 4. ed. New York: John Wiley & Son Inc., 1993.
- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. *Time series analysis: forecasting and control*. 4. ed. New Jersey: Wiley, 2008.
- BROCKWELL, P. J.; DAVIS, R. A. *Introduction to time series and forecasting*. 2. ed. New York: Springer, 2002.
- COVEY, T.; BESSLER, D. A. Testing for Granger's full causality. *The Review of Economics and Statistics*, Cambridge (USA), v. 74, n. 1, p. 146-153, Feb. 1992.
- DOORNIK, J. A. *Testing vector error autocorrelation and heteroscedasticity in dynamic models*. Oxford, UK: Nuffield College, Sept. 1996.
- DOORNIK, J. A.; HENDRY, D. F. *Modelling dynamic systems using PcGive 10*. Vol. II. London: Timberlake Consultants Ltd., 2001.
- ENDERS, W. *Applied econometric time series*. 2. ed. USA: Wiley, 2004.
- GRANGER, C. W. J.; NEWBOLD, P. *Forecasting economic time series*. Orlando: Academic Press, 1977.
- HAMILTON, J. D. *Time series analysis*. Princeton: Princeton University Press, 1994.
- HARVEY, A. C. *Forecasting, structural time series models and the Kalman filter*. Cambridge: Cambridge University Press, 1989.

HARVEY, A. C.; KOOPMAN, S. J. Diagnostic checking of unobserved-components time series models. *Journal of Business & Economic Statistics*, Alexandria (USA), v. 10, n. 4, p. 377-389, Oct. 1992.

HENDRY, D. F. *Dynamic econometrics*. New York: Oxford University Press, 1995.

KOOPMAN, S. J.; HARVEY, A. C.; DOORNIK, J. A.; SHEPHARD, N. *Stamp – Structural Time Series Analyser, Modeller and Predictor*. 3. ed. London: Timberlake Consultants Ltd., 2000.

MADDALA, G. S.; KIM, I. M. *Unit roots, cointegration, and structural change*. 1. ed. Cambridge: Cambridge University Press, 1998.

MELO, B. S. V. Modelo de previsão para arrecadação tributária. Brasília: Esaf, 2001. 97 p. *VI Prêmio Tesouro Nacional – 2001: orçamentos e sistemas de informação sobre a administração financeira pública*. Disponível em: <[http://www.tesouro.fazenda.gov.br/Premio\\_TN/conteudo\\_mono\\_pr6tema3.html](http://www.tesouro.fazenda.gov.br/Premio_TN/conteudo_mono_pr6tema3.html)>. Acesso em: 15 set. 2007.

MUSGRAVE, R. A.; MUSGRAVE, P. B. Finanças públicas: teoria e prática. *Public finance in theory and practice*. Tradução de Carlos Alberto Primo Braga. Rio de Janeiro: Campus; São Paulo: Edusp, 1980.

ROMER, D. *Advanced macroeconomics*. 3. ed. Boston: Mc Graw-Hill Irwin, 2006.

SIQUEIRA, M. L. Melhorando a previsão de arrecadação tributária federal através da utilização de modelos de séries temporais. Brasília: Esaf, 2002. 84 p. *VII Prêmio Tesouro Nacional – 2002: tributação, orçamentos e sistemas de informação sobre a administração financeira pública*. Disponível em: <[http://www.tesouro.fazenda.gov.br/Premio\\_TN/VIIIPremio/conteudo\\_catalogos\\_pr7.html](http://www.tesouro.fazenda.gov.br/Premio_TN/VIIIPremio/conteudo_catalogos_pr7.html)>. Acesso em: 15 set. 2007.

TSAY, R. S. Time series model specification in the presence of outliers. *Journal of the American Statistical Association*, Alexandria (USA), v. 81, n. 393, p. 132-141, Mar. 1986.

WHITE, H. A Heteroskedastic-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, Princeton, v. 48, n. 4, p. 817-838, May 1980.