Universidade Federal de Minas Gerais
Departamento de Estatística
Programada de Pós-Graduação em Estatística

PREVISÃO DA ARRECADAÇÃO TRIBUTÁRIA FEDERAL POR MEIO DA UTILIZAÇÃO DE ANÁLISE FATORIAL E MODELOS DE SÉRIES TEMPORAIS

Viviane Rezende Ferreira

Belo Horizonte
2015

Viviane Rezende Ferreira

PREVISÃO DA ARRECADAÇÃO TRIBUTÁRIA FEDERAL POR MEIO DA UTILIZAÇÃO DE ANÁLISE FATORIAL E MODELOS DE SÉRIES TEMPORAIS

Monografia apresentada ao Programa de Pós-Graduação Lato-Sensu em Estatística do Departamento de Estatística da Universidade Federal de Minas Gerais como requisito para obtenção do título de Especialista em Estatística.

Orientador: Luís Alberto Medrano Toscano da Universidade Federal Rural do Rio de Janeiro (UFRRJ), do Departamento de Matemática (DEMAT).

Co-orientadora: Ela Mercedes Medrano Toscano da Universidade Federal de Minas Gerais (UFMG), do Departamento de Estatística (DEST-UFMG).

Belo Horizonte

2015

Aprovada pela Banca Examinadora em cumprimento a requisito exigido para obtenção do Título de Especialista em Estatística.

_____

Luís Alberto Medrano Toscano – DEMAT/UFRRJ

Orientador

_____

Ela Mercedes Medrano Toscano – DEST/UFMG

Co-orientadora

_____

Sueli Aparecida Mingoti – DEST/UFMG

Membro da banca, Convidada

_____

Aureliano Angel Bressan – Cepead/UFMG

Membro da banca, Convidado

Belo Horizonte, 06 de julho de 2015

Ao Senhor Jesus Cristo,
pelo seu infinito amor e graça.

AGRADECIMENTOS

Agradeço a Jesus Cristo que me amou por primeiro e que por meio do Seu sangue me faz mais que vencedora em todas as coisas.

Ao meu orientador Luís Alberto pela paciência e horas no skype e telefone que alavancou a realização da pesquisa.

A minha co-orientadora e professora Mercedes, que acreditou em mim e favoreceu a realização deste trabalho em todo tempo.

A Rose, pelo suporte, orientação, ajuda e telefonemas durante todo o curso.

Ao meu noivo, Gabriel, por sempre me incentivar e encorajar nos momentos que foram mais difícil de prosseguir.

Aos meus pais que me deram os firmes alicerces para eu chegar até aqui e contribuíram para esta realização. Obrigada!

RESUMO

Neste trabalho apresenta-se de forma breve e sucinta a modelagem de dez séries de tributos federais brasileiros por meio da extração da componente sazonal e padronização das séries. Para o sucesso foi realizado a combinação de análise fatorial por máxima verossimilhança e séries temporais na estimação de modelos SARIMA ou ARIMA para cada fator ajustado, prevendo assim doze passos a frente de cada fator e reinserindo a agregação sazonal e a realização de estimação intervalar com 95% de confiança.

Palavras-chave: previsão. sazonalidade. séries temporais. análise fatorial.

LISTA DE TABELAS

LISTA DE FIGURAS

LISTA DE SIGLAS

COFINS – Contribuição para Financiamento da Seguridade Social
CPIS – Programa de Integração Social
CSLL – Contribuição Social sobre o Lucro Líquido
IIMP – Imposto de Importação
IOF – Imposto sobre Operações Financeiras
IPI – Imposto sobre Produtos Industrializados
IRPF – Imposto de Renda sobre Pessoa Física
IRPJ – Imposto de Renda sobre Pessoa Jurídica
IRRF – Imposto sobre Renda Retido na Fonte
MF – Modelo Fatorial
MFD – Modelo Fatorial Dinâmico
EPA – Erro Acumulado Percentual
RMSE – Raiz do Erro Quadrático Médio
MPE – Erro Percentual Médio
MAPE – Erro Percentual Absoluto Médio
MAD – Desvio Absoluto Médio

SUMÁRIO

1. INTRODUÇÃO

Os impostos desempenham um papel importante em um país. É através da arrecadação tributária que o Estado consegue financiar-se e prover bens públicos à população, ou seja, é a principal fonte de captação de recursos para financiamento dos encargos governamentais e para propiciar o bem-estar social.

Do ponto de vista econômico, os impostos têm um papel ímpar na economia, pois a receita arrecadada é utilizada para investimentos em obras públicas para a sociedade e para fomentar o desenvolvimento do país. Estes podem incidir sobre a renda (salários, lucros, ganhos de capital) e patrimônio (terrenos, casas, carros e etc) das pessoas físicas e jurídicas. São valores pagos em moeda nacional arrecadado pelo Governo Federal, Estadual e Municipal.

Neste trabalho foram estudados dez tributos federais brasileiros a fim de estimar um modelo de previsão para arrecadação do ano seguinte e o comportamento destes na economia. Os tributos analisados foram: Contribuição para Financiamento da Seguridade Social (CONFINS), Contribuição Social sobre o Lucro Líquido (CSLL), Imposto sobre Operações Financeiras (IOF), Imposto de Importação (IIMP), Imposto sobre Produtos Industrializados (IPI), Imposto de Renda sobre Pessoa Física (IRPF), Imposto de Renda sobre Pessoa Jurídica (IRPJ), Imposto sobre Renda Retido na Fonte (IRRF), Programa de Integração Social (PIS) e Outros Impostos Arrecadados.

A previsão dos tributos é importante para o Estado estimar o que se pretende arrecadar de receita durante um período. De acordo com o montante previsto a ser arrecadado, será fixada as despesas a serem incorridas, assim como a determinação das necessidades de financiamento do Governo. Ademais, a previsão das receitas repercutirá na concessão de créditos suplementares por excesso de arrecadação.

Os impostos são divididos em diretos e indiretos. Os impostos diretos são designados a taxar diretamente o contribuinte sendo que o principal exemplo deste é o Imposto de Renda. Os impostos indiretos, entretanto, são repassados ao contribuinte através do custo adicionado ao produto e o reflexo deste é sentido no preço final dos produtos. Os impostos indiretos são cobrados em todos os bens adquiridos pelo consumidor.

É certo que, mesmo os tributos que parecem distantes na interferência sobre a vida de cada cidadão, na verdade tem impacto no cotidiano do brasileiro, porque é semelhante a uma reação em cadeia dentro do orçamento e da aplicação dos recursos da administração pública que, via de regra, repercute na existência e na vida do cidadão comum, mesmo que indiretamente. Sendo assim, são relevantes e envolvem diretamente o cidadão brasileiro, por isso merecem descrição e contextualização na próxima seção.

1.1. Contextualização

Baseado nos conceitos constitucionais e do Código Tributário Nacional é apresentado breve descrição e contextualização dos tributos federais brasileiros estudados nesta pesquisa.

A Contribuição para Financiamento da Seguridade Social (COFINS) foi instituída pela Lei Complementar 70 de 30/12/1991, a contribuição COFINS, atualmente, é regida pela Lei 9.718/98, com as alterações subsequentes, são contribuintes da COFINS as pessoas jurídicas de direito privado em geral, inclusive as pessoas a elas equiparadas pela legislação do Imposto de Renda, exceto as microempresas e as empresas de pequeno porte optantes pelo Simples Nacional.

A partir de 01.02.1999, com a edição da Lei 9.718/98, a base de cálculo da contribuição é a totalidade das receitas auferidas pela pessoa jurídica, sendo irrelevante o tipo de atividade por ela exercida e a classificação contábil adotada para as receitas. Já a alíquota geral é de 3% (a partir de 01.02.2001) ou 7,6% (a partir de 01.02.2004) na modalidade não cumulativa, entretanto, para determinadas operações, a alíquota é diferenciada e é necessário consultar a Receita Federal.

A Contribuição Social sobre o Lucro Líquido (CSLL) foi instituída pela Lei 7.689/1988, aplicam-se à CSLL as mesmas normas de apuração e de pagamento estabelecidas para o imposto de renda das pessoas jurídicas, mantidas a base de cálculo e as alíquotas previstas na legislação em vigor (Lei 8.981, de 1995, artigo 57).

Desta forma, além do IRPJ, a pessoa jurídica optante pelo Lucro Real, Presumido ou Arbitrado deverá recolher a Contribuição Social sobre o Lucro Presumido (CSLL), também pela forma escolhida. Não é possível, por exemplo, a

empresa optar por recolher o IRPJ pelo Lucro Real e a CSLL pelo Lucro Presumido. Escolhida a opção, deverá proceder a tributação, tanto do IRPJ quanto da CSLL, pela forma escolhida. A alíquota a partir de 01.02.2000 a alíquota é de 9% (nove por cento) e para as entidades financeiras e equiparadas a alíquota é de 15% (quinze por cento).

O Imposto sobre Produtos Industrializados (IPI) incide sobre produtos industrializados, nacionais e estrangeiros as suas disposições estão regulamentadas pelo Decreto 7.212/2010 (RIPI/2010), sendo composto da soma de impostos incidentes sobre três produtos (bebidas-IPIB, veículos-IPIA, fumo-IPIF). O campo de incidência do imposto abrange todos os produtos com alíquota, ainda que zero, relacionados na Tabela de Incidência do IPI (TIPI), observadas as disposições contidas nas respectivas notas complementares, excluídos aqueles a que corresponde a notação "NT" (não-tributado). A partir de 01.05.2009, o período de apuração do IPI incidente na saída dos produtos dos estabelecimentos industriais ou equiparados a industrial, passa a ser mensal, conforme Lei 11.933/2009, que revogou o § 1o do art. 1 da Lei 8.850/1994.

O Programas de Integração Social (CPIS) foi criado pela Lei Complementar 07/1970 são contribuintes do CPIS as pessoas jurídicas de direito privado e as que lhe são equiparadas pela legislação do Imposto de Renda, inclusive empresas prestadoras de serviços, empresas públicas e sociedades de economia mista e suas subsidiárias, excluídas as microempresas e as empresas de pequeno porte submetidas ao regime do Simples Nacional (LC 123/2006).

A partir de 01.02.1999, com a edição da Lei 9.718/1998, a base de cálculo da contribuição é a totalidade das receitas auferidas pela pessoa jurídica, sendo irrelevante o tipo de atividade por ela exercida e a classificação contábil adotada para as receitas.

A alíquota do PIS é de 0,65% ou 1,65% (a partir de 01.12.2002 - na modalidade não cumulativa - Lei 10.637/2002) sobre a receita bruta ou 1% sobre a folha de salários, nos casos de entidades sem fins lucrativos, contudo, para determinadas operações, a alíquota é diferenciada e precisa ser consultada caso a caso.

O Imposto de Renda Pessoa Jurídica (IRPJ) tem características peculiares, sendo os seus contribuintes as pessoas jurídicas e as pessoas físicas a elas equiparadas, domiciliadas no país. Elas devem apurar o IRPJ com base no lucro,

que pode ser real, presumido ou arbitrado. A alíquota do IRPJ é de 15% (quinze por cento) sobre o lucro apurado, com adicional de 10% sobre a parcela do lucro que exceder R$ 20.000,00 / mês.

As entidades submetidas aos regimes de liquidação extrajudicial e de falência sujeitam-se às normas de incidência do imposto aplicáveis às pessoas jurídicas, em relação às operações praticadas durante o período em que perdurarem os procedimentos para a realização de seu ativo e o pagamento do passivo (Lei 9.430/1996, artigo 60). As empresas públicas e as sociedades de economia mista, bem como suas subsidiárias, são contribuintes nas mesmas condições das demais pessoas jurídicas (Constituição Federal, artigo 173 § 1º).

As pessoas físicas brasileiras deverão prestar contas à Receita Federal, apurando o imposto de renda devido segundo as normas do Regulamento do Imposto de Renda. Anualmente, deverão entregar a declaração de seus rendimentos e bens, pagando o imposto devido ou apurando a restituição, se houver. São tributáveis pelo IRPF os rendimentos (como salários, benefícios e remuneração por serviços prestados), ganhos de capital, juros e outras rendas (como aluguéis e direitos autorais) ou proventos (como aposentadoria).

Os dados contidos neste trabalho referente ao IRPF representam a diferença entre o Imposto de Renda Retido na Fonte (IRRF) e o que não foi pago a Receita Federal que devera ser arrecadado. Sendo assim, o pagamento do IRPF diretamente pela pessoa física ocorre quando não há retenção do imposto na fonte (exemplos: rendimentos de aluguéis, de taxistas, etc.): onde o contribuinte deve utilizar o carnê-leão. O contribuinte que recebe rendimentos sujeitos à retenção na fonte de mais de uma fonte pagadora: o contribuinte pode optar entre o Mensalão e a Declaração de Ajuste Anual IRPF. Também o resultado da Declaração de Ajuste Anual que for de imposto a pagar: o contribuinte pode optar por pagar em quotas, e quando houver ganho de capital na alienação de bens e direitos.

O IRRF é a informação prestada pelas empresas, ou melhor, as fontes pagadoras, com o objetivo de informar à Secretaria da Receita Federal do Brasil os rendimentos pagos a pessoas físicas domiciliadas no Brasil, o valor do imposto sobre a renda e contribuições retidos na fonte, dos rendimentos pagos ou creditados para seus beneficiários, o pagamento, crédito, entrega, emprego ou remessa a residentes ou domiciliados no exterior e os pagamentos a plano de assistência à saúde na modalidade de coletivo empresarial.

São contribuintes do Imposto sobre Operações de Crédito, Câmbio e Seguro, ou relativo à Títulos Mobiliários (IOF) as pessoas físicas e as pessoas jurídicas que efetuarem operações de crédito, câmbio e seguro ou relativas a títulos ou valores mobiliários. A cobrança e o recolhimento do imposto são efetuados pelo responsável tributário: a pessoa jurídica que conceder o crédito; as instituições autorizadas a operar em câmbio; as seguradoras ou as instituições financeiras a quem estas encarregarem da cobrança do prêmio de seguro; as instituições autorizadas a operar na compra e venda de títulos ou valores mobiliários.

O imposto sobre a Importação de Produtos Estrangeiros (IIMP) incide sobre a importação de mercadorias estrangeiras e sobre a bagagem de viajante procedente do exterior. No caso de mercadorias estrangeiras, a base de cálculo é o valor aduaneiro e a alíquota está indicada na Tarifa Externa Comum (TEC). No caso da bagagem, a base de cálculo é o valor dos bens que ultrapassem a cota de isenção e a alíquota é de cinquenta por cento.

Já os Outros Impostos citados nesse trabalho referem-se às demais receitas administradas pela Receita Federal e engloba, por exemplo: Contribuição para o FUNDAP, CIDE combustíveis, a Contribuição sobre Movimentação Financeira (CPMF), o Imposto Territorial Rural (ITR), arrecadação de loterias, além de outras receitas administradas.

1.2. Objetivos

O presente estudo tem como objetivo geral a previsão tributária federal, explorando as informações contidas nas inter-relações entre as séries de dez impostos e o passado das séries.

Adicionalmente, também faz parte do objetivo geral comparar os resultados obtidos por meio dos modelos SARIMA com as previsões dos modelos univariados e com o Modelo Fatorial Dinâmico (MFD), ambos artigos realizados por Mendonça e Medrano (2014).

1.2.1. Objetivos Específicos

Este trabalho possui os seguintes objetivos específicos para alcançar o objetivo geral estabelecido:

- Realizar análise das séries de impostos, os tributos federais gerenciados pela Secretaria de Política Econômica (SPE).
- Realizar a análise fatorial, no conjunto das dez séries, reduzindo a complexidade na modelagem das séries de impostos.
- Modelar as séries de fatores usando modelos de séries temporais SARIMA.
- Componente sazonal modelada endogenamente.
- Obter previsões das séries de impostos, através da desagregação das previsões dos fatores estimados pelo método de análise fatorial.
- Comparar os modelos SARIMA estimados com as previsões univariadas e com o Modelo Fatorial Dinâmico (MFD), ambos artigos realizados por Mendonça e Medrano (2014).

1.3. Justificativas

A previsão de tributos federais gerenciados pela Secretaria de Política Econômica (SPE) é o processo por meio do qual se estima o montante de recursos que serão arrecadados em determinado período. A previsão é fundamental no processo orçamentário; somente após a determinação da disponibilidade de recursos é possível iniciar a definição de quais objetivos serão perseguidos pelo governo por meio da implantação de políticas públicas.

Outro ponto importante é que o uso da análise fatorial é importante para a redução de dimensionalidade do modelo. Mendonça e Medrano (2014) realizaram uma aplicação do Modelo Fatorial Dinâmico e dos modelos univariados para a previsão da receita tributária brasileira e nesse trabalho compara-se os modelos e propõe-se que de algum modo, a informação contida nas inter-relações entre os vários tributos seja reduzida para um conjunto menor de fatores e que a componente sazonal das diferentes séries seja modelada endogenamente, o que permite obter

um melhor ajustamento e previsões mais precisas sobre a dinâmica futura dos impostos.

1.4. Estrutura do Trabalho

Além dessa introdução, este trabalho está estruturado em mais 3 capítulos. No capítulo 2 é apresentado a descrição dos modelos de análise fatorial e séries temporais, bem como a metodologia empregada no ajuste das séries. O capítulo 3 contempla as fontes das bases de dados, os softwares utilizados e, além disso, também são apresentados e discutidos os resultados obtidos com os ajustes dos modelos e as previsões geradas. As considerações finais e conclusões do trabalho são apresentadas no capítulo 4. Por fim, finalizando com as referências e os anexos.

2. METODOLOGIA

2.2. Análise Fatorial

A análise fatorial é a principal e a mais antiga técnica de análise multivariada. MENEZES et al. (1978) comentam que a análise fatorial pode ser usada no agrupamento de variáveis ou no agrupamento de unidades de observações. No primeiro caso a matriz de dados iniciais tem as variáveis nas colunas e as unidades de amostra nas linhas. No segundo caso, transpõe-se a matriz anterior, obtendo-se as unidades nas colunas e as variáveis nas linhas.

Se o número de variáveis estudadas é grande, uma estratégia de análise seria a de tentar simplificar, ou melhor, estruturar o conjunto de dados, a partir das inter-relações entre tais variáveis. Tais inter-relações podem ser medidas pelas covariâncias ou pelos coeficientes de correlação entre as variáveis. Duas técnicas estatísticas de análise multivariada são comumente utilizadas para tratar este problema: Análise de Componentes Principais e Análise Fatorial (Johnson e Wichern, 1988).

A Análise Fatorial é um conjunto de métodos estatísticos que, em certas situações, permite "explicar" o comportamento de um número relativamente grande de variáveis observadas, em termos de um número relativamente pequeno de variáveis latentes ou fatores. As variáveis são agrupadas por meio de suas correlações, ou seja, aquelas pertencentes a um mesmo grupo serão fortemente correlacionadas entre si, mas pouco correlacionadas com as variáveis de outro grupo.

Cada grupo de variáveis representará um fator (Johnson e Wichern, 1988), duas variáveis somente serão altamente correlacionadas se elas tiverem altas cargas no mesmo fator. O modelo fatorial pressupõe efeitos aditivos; fatores, variáveis e resíduos normalmente distribuídos, resíduos independentes e relações lineares entre as variáveis (Johnson e Wichern, 1988).

A padronização é um recurso bastante utilizado na estatística e foi usada nesse trabalho, dado que algumas séries de impostos apresenta variância relativamente grande, foi trabalhado com as séries padronizadas para evitar o

problema de ter uma série, com uma variância relativamente grande, influenciando inapropriadamente, a determinação das cargas dos fatores.

A padronização consiste em subtrair a média do conjunto de dados e dividir o resultado pelo desvio padrão do conjunto de dados. Nesta ocasião a padronização das séries foi realizada após a desazonalização das mesmas e uma característica importante de ressaltar é que para as variáveis padronizadas as medidas de assimetria e curtose não se alteram quando a variável ou conjunto de dados é padronizada.

2.2.1 Modelo de Análise Fatorial

Um vetor aleatório $\underline{X}$ com vetor de médias $\underline{\mu} = [\mu_1, \mu_2, \ldots \mu_p]'$, matriz de covariância Σ_{pxp}e matriz de correlação P_{pxp} .Sendo $\underline{Z}_i = \left[\frac{(X_i - \mu_i)}{\sigma_i}\right]$, i = 1, 2, .., p as variáveis originais padronizadas, cujo μ_i e σ_i são a média e o desvio padrão de X_i i = 1, 2, .., p, respectivamente. Assim, a matriz P_{pxp} é a matriz de covariâncias do vetor aleatório $\underline{Z} = [Z_1, Z_2, \ldots Z_p]'$ segundo MINGOTI, 2005:

O modelo fatorial construído a partir da matriz de correlação teórica é formalizado conforme as equações abaixo:

$$\begin{array}{rccccccc} Z_1 = & l_{11}F_1 & + l_{12}F_2 & + \ldots & + l_{1m}F_m & + \varepsilon_1 \\ Z_2 = & l_{21}F_1 & + l_{22}F_2 & + \ldots & + l_{2m}F_m & + \varepsilon_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ Z_p = & l_{p1}F_1 & + l_{22}F_2 & + \ldots & + l_{pm}F_m & + \varepsilon_p \end{array}' \qquad (2.1)$$

Sendo m ≤ p. O modelo (2.1), em notação matricial é expresso por:

$$\underline{Z} = \Lambda\underline{F} + \underline{\varepsilon} \qquad (2.2)$$

Onde:

$$\underline{Z} = \begin{bmatrix} \frac{(X_1 - \mu_1)}{\sigma_1} \\ \frac{(X_2 - \mu_2)}{\sigma_2} \\ \vdots \\ \frac{(X_p - \mu_p)}{\sigma_p} \end{bmatrix}, \quad \underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_p \end{bmatrix}, \quad \underline{F} = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_p \end{bmatrix}, \quad \Lambda_{pxm} = \begin{bmatrix} l_{11} & l_{12} & \ldots & l_{1m} \\ l_{21} & l_{22} & \ldots & l_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ l_{p1} & l_{p2} & \ldots & l_{pm} \end{bmatrix},$$

Assim, $\underline{F}$, com 1 ≤ m ≤ p, é um modelo aleatório com m fatores, chamados de variáveis latentes, que descrevem os elementos da população estudada e não são observáveis. O modelo *AF* assume que as variáveis Z_i estão correlacionadas linearmente com novas variáveis aleatórias F_j, $j=1,2,\cdots,m$, que deverão ser analisadas e identificadas pelo pesquisador.

O vetor $\underline{\varepsilon}$ é um vetor de erros aleatórios correspondentes à variação de Z_i, que não é explicada pelos fatores comuns F_j. O coeficiente ℓ_{ij}, comumente chamado de peso, carregamento ou Carga Fatorial (CF) que é o coeficiente da i-ésima variável padronizada Z_i no j-ésimo fator F_j , e representa o grau de relação linear entre Z_i e F_j, j=1, 2, ⋯, m.

A matriz Λ_{pxm} armazena os coeficientes ℓ_{ij}, ou loading, e é uma matriz de parâmetros a serem estimados, que de acordo com MINGOTI (2005) as informações das p-variáveis originais padronizadas estarão sendo representadas por (p+m) variáveis aleatórias não observáveis, ou seja $\varepsilon' = [\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_p]$ $F' = [F_1, F_2, \ldots F_p]$.

Em resumo, o modelo fatorial implica na imposição de condições que permitem obter estimativas de Λ e ψ. Posteriormente, a matriz de cargas fatoriais (Λ) pode ser submetida à rotação (multiplicação por uma matriz ortogonal), a qual é determinada por critérios de facilidade de interpretação. Obtidas as cargas e as variâncias específicas, os fatores são identificados e comumente calcula-se os valores dos escores fatoriais.

(Johnson e Wichern, 1988) consideram os Métodos do Componente Principal e o da Máxima Verossimilhança como os mais recomendados para análise fatorial. Neste trabalho utilizou-se apenas o da Máxima Verossimilhança.

O método da Máxima Verossimilhança para Análise Fatorial foi introduzido por Lawley, em 1940. A principal vantagem de utilizar o Método da Máxima Verossimilhança para Análise Fatorial é, talvez, a possibilidade de se realizar testes de hipóteses, com o objetivo de testar a adequacidade do modelo. Independente de qual método seja usado, o analista deveria observar as magnitudes dos elementos da matriz residual $\Sigma - (\Lambda \Lambda' + \Psi)$ para um dado número de fatores m. Quanto menores estes elementos, melhor a solução obtida reproduz Σ, mas sem m=p a matriz residual seria nula.

Existe também três terminologias referentes à análise fatorial, ou seja, componentes principais, eixos principais e fatores principais. Elas diferem somente na

composição das variáveis. O eixo principal é usualmente padronizado para ter uma média zero e uma variância igual à variância total considerada. O componente principal é o mesmo eixo principal, exceto que sua média não é padronizada para zero. O fator principal é normalizado para ter uma média zero e variância unitária (KIM, 1975).

Segundo HARMAN (1968), a solução da fatoração com valores unitários nas diagonais da matriz de correlação pode ser chamada de solução de componente principal, e a solução com comunalidades nas diagonais da matriz de correlação é denominada de solução do fator principal.

Foi observado que a solução pelo método do fator principal requer um conhecimento a *priori* das p comunalidades $h_1^2, h_2^2, \cdots, h_p^2$, para formar a matriz de correlação reduzida R*.

A comunalidade pode ser definida sendo a parte da variância de Z_i explicada pelos fatores do modelo. E a variância específica como a parte da variância de Z_i explicada pelos erros do modelo.

Existem vários métodos para estimar as comunalidades. Os mais comuns, conforme citado por KARSON (1982), são:

a) $\hat{h}_j^2 = 1$ (j = 1, 2, ..., p), ou seja, tomar cada comunalidade como sendo igual a 1. Dessa forma R* = R e a solução pelo método do fator principal seria idêntica à solução pelo método do componente principal.

b) $\hat{h}_j^2 = R^2_{j.1,2,\ldots,j-1,j+1,\ldots,p}$, onde R^2 é o quadrado do coeficiente de correlação múltipla entre a variável X_j e todas as outras. Tipicamente esse valor é calculado po

$$\mathrm{r}1 - \frac{1}{\mathrm{r}^{\mathrm{jj}}}$$

Onde r^{jj} é o j-ésimo elemento da diagonal principal de R^{-1}.

c) $\hat{h}_j^2 = \max_{j'} | r_{jj'} |$ (j ≠ j'), o que significa que $\hat{h}_j^2$ é tomado como o maior valor absoluto da correlação simples entre X_j e as outras variáveis.

d) $\hat{h}_j^2 = \sum_{\substack{j'=1 \\ j' \neq j}}^{p} \frac{r_{jj'}}{p-1}$, assumindo que a média resultante das (p - 1) correlações simples de X_j com as outras variáveis seja positiva.

e) é tomado inicialmente por qualquer um dos quatro métodos acima e uma solução pelo método do fator principal é obtida. A partir dessa solução, $\sum_{k=1}^{m} a_{kj}^2$ é computado para cada j. Esses valores são tomados como novas comunalidades $\hat{h}_j^2$, e uma nova solução é obtida. Esse processo iterativo é mantido até que tenhamos pequenas diferenças nas comunalidades de uma etapa para a outra.

2.2.2 Determinação do Número de Fatores m

Após a realização da análise fatorial, para estimação do número de fatores m faz parte do objetivo deste trabalho encontrar um número relativamente pequeno de fatores que possuam um alto grau de explicação da variabilidade original dos dados e assim encontrar também fatores interpretáveis. Para isso utilizou-se de critérios abordados por MINGOTI (2005):

Critério 1: de posse dos autovalores calculados, é realizado uma análise da proporção da variância total relacionada com cada autovalor, sendo o valor de m determinado por àqueles autovalores que apresentarem maiores proporções explicada da variância total. Foi o critério utilizado nesse trabalho para determinação do número de fatores m.

Critério 2: conhecido por método de Kaiser (1958), o valor de m será igual ao número de autovalores maiores ou iguais a 1.

Critério 3: este critério é realizado através da observação do gráfico scree-plot (Cattell, 1966). No gráfico procura-se o “ponto de salto” que aponta para um ponto de decréscimo de importância da variância total, sendo o valor de m, os autovalores anteriores ao “ponto de salto”.

Valendo-se que uma escolha adequada de m, leva em consideração também a interpretabilidade dos fatores e o princípio da parcimônia (que é a descrição da estrutura de variabilidade com um número reduzido de fatores).

2.2.3 Método da Máxima Verossimilhança

Para o método da Máxima Verossimilhança na Análise Fatorial ou independente de qual método seja usado, o analista do presente estudo deveria observar as magnitudes dos elementos da matriz residual $\Sigma - (\Lambda \Lambda' + \Psi)$ para um dado número de fatores m. Quanto menores estes elementos, melhor a solução obtida reproduz Σ, sendo também melhor a estrutura proposta para X_j.

Além das suposições habituais do modelo fatorial, supõe-se que os vetores aleatórios F (fatores comuns) e ε (fatores específicos) têm distribuição normal multivariada com vetores de média zero e com matrizes de covariâncias Σ_{pxp} e Ψ, respectivamente, além disso também há suposição de normalidade, seguido de que F e ε são mutuamente independentes. Como X é expresso em termos de F e ε, conforme a equação (2.2), temos que o vetor das variáveis observáveis é também normal, com média zero e matriz de covariância Σ_{pxp}. Portanto, as condições para a aplicação do Método da Máxima Verossimilhança na Análise Fatorial são:

a) O vetor X tem uma distribuição normal multivariada, com vetor de médias zero e matriz de covariância Σ_{pxp};

b) $X = \Lambda F + \varepsilon$, onde $E(F) = E(\varepsilon) = \mathbf{0}$, $Var(F) = I$, $Var(\varepsilon) = \Psi = diag(\Psi_i^2)$, $Cov(F,\varepsilon) = \mathbf{0}$;

c) $\Sigma = \Lambda\Lambda' + \Psi$;

d) I é a matriz identidade;

O caminho a ser tomado será o de maximizar a função de verossimilhança de X, com respeito aos p elementos de Λ, onde Λ é uma matriz pxm e os p elementos de Ψ.

Portanto, segundo KARSON (1982), ignorando os termos que não envolvem parâmetros, a função de verossimilhança pode ser escrita como:

$$L = |\Sigma|^{-(n-1)/2} \, e^{-[(n-1)/2]tr(S\Sigma^{-1})}$$

Conforme feito usualmente, é melhor maximizar L maximizando

$$\ln L = -\left[(n-1)/2\right]\left[\ln|\Sigma| + tr(S\Sigma^{-1})\right] \tag{2.3}$$

Quando Σ na equação (2.4) é substituído por $\Lambda\Lambda' + \Psi$, temos,

$$\ln L = -\left[(n-1)/2\right]\left\{\ln|\Lambda\Lambda' + \Psi| + tr\left[S(\Lambda\Lambda' + \Psi)^{-1}\right]\right\} \tag{2.4}$$

O método da máxima verossimilhança envolve determinar os pxm elementos de Λ e os p elementos de Ψ que maximizem conforme a equação (2.4). A indeterminação de Λ, será aqui removida pela imposição da condição de que $\Lambda'\Psi^{-1}\Lambda$ deverá ser diagonal. Se $\hat{\Lambda}$ denotar a matriz dos estimadores de máxima verossimilhança de Λ e $\hat{\Psi}$ denotar o estimador de máxima verossimilhança de Ψ, então $\hat{\Lambda}$ e $\hat{\Psi}$ serão obtidos diferenciando a equação (2.4) em relação a Λ e Ψ; fazendo as p(m + 1) derivadas iguais a zero; e resolvendo o sistema das p(m + 1) equações. Detalhes dessa solução aparecem em Jöreskog (1967) e Lawley e Maxwell (1971), citados por KARSON (1982). As equações que requerem uma solução numérica são:

$$\hat{\Lambda}\left(I + \hat{\Lambda}'\hat{\Psi}^{-1}\hat{\Lambda}\right) = S\hat{\Psi}^{-1}\hat{\Lambda} \tag{2.5}$$

$$\hat{\Psi} = \mathrm{diag}(\mathrm{S} - \hat{\Lambda}\hat{\Lambda}') \tag{2.6}$$

Onde $\hat{\Psi}$ em (2.6) é uma matriz diagonal, cujos elementos são os elementos diagonais da matriz (pxp) $\mathrm{S} - \hat{\Lambda}\hat{\Lambda}'$.

Na estimação do modelo fatorial, foi usado o método de máxima verossimilhança, uma vez que as séries são altamente correlacionadas, e uma das vantagens do método é que permite a avaliação das magnitudes dos elementos da matriz residual, $\Sigma - (\Lambda\Lambda' + \Psi)$ para um determinado número de fatores m, ou seja, quanto menores estes elementos, melhor a solução obtida, sendo também melhor a estrutura proposta para as séries. Sendo assim, baseou-se em Mendonça e Medrano (2014) também na definição do número de fatores em função do menor erro de previsão.

O modelo proposto por Mendonça e Medrano (2014), para modelar o vetor impostos $y_t = (y_{1,t}, y_{2,t}, \ldots, y_{m,t})'$ foi:

$$y_t = Lf_t + S_t + e_t \quad e_t \sim N(0, \Sigma)$$

$$f_t = \Gamma_1 f_{t-1} + \Gamma_2 f_{t-2} + \cdots + \Gamma_p f_{t-p} + v_t \quad v_t \sim N(0, \Delta)$$

Onde S_t é o termo que modela a sazonalidade e f_t segue uma estrutura que mudam no tempo. Uma abordagem Bayesiana foi utilizada para estimar os parâmetros do modelo.

2.3. Séries Temporais

Para realizar a previsão de séries temporais, a discussão de alguns conceitos básicos é essencial. O primeiro deles é o de processo estocástico, um processo estocástico é um conjunto de variáveis aleatórias ordenadas no tempo. Na visualização de variáveis aleatórias contínuas, utiliza-se a representação $Y_{(t)}$, caso sejam discretas, utiliza-se Y_t. Um processo estocástico pode então ser representado por:

$$Y_t = [Y_{t_1}, Y_{t_2}, \ldots, Y_{t_n}]$$

Uma série temporal nada mais é do que uma realização única, das diversas possíveis, do processo estocástico em estudo. A partir desse ponto, a estratégia de análise será a mesma utilizada para dados transversais. Neste, utilizou-se as informações de uma amostra para fazermos inferência sobre a população. Na análise de séries temporais, utilizamos as informações da realização para fazer inferência sobre o processo estocástico, GUAJARATI (2004).

Nas próximas seções serão expostos os demais conceitos básicos que irão permitir a construção de modelos para previsões de séries temporais, conceitos como estacionaridade, processo ruído branco, função de autocorrelação e função de autocorrelação parcial que serão empregados na previsão das séries de impostos.

2.3.1 Descrição e Apresentação das Séries de Tributos

A descrição e apresentação das séries serão em seção posterior por meio de gráficos de linha das séries para observar tendências e variabilidade sazonal, Box-plot por anos, para observar a variabilidade ano por ano, Box-plot por mês, para observar a variabilidade sazonal mensal.

Estatísticas descritivas de cada série, para observar a forma da distribuição, em torno de qual valor está variando a série, com menor ou maior variabilidade, se a distribuição é simétrica ou não.

2.3.2 Estacionaridade

Para que seja possível fazer inferência sobre o processo estocástico, é necessário que o mesmo mantenha suas propriedades estatísticas constantes ao longo do tempo, ou seja, o processo precisa estar em um estado de equilíbrio estatístico, (BOX; JENKINS; REINSEL, 2008). A essa característica dá-se o nome de estacionaridade.

Para fins desse trabalho, e da grande maioria das aplicações, a estacionaridade fraca é suficiente. Um processo estocástico é dito fracamente estacionário, se a sua média e variância são constantes ao longo do tempo, e a covariância entre dois períodos depende apenas da distância entre eles, e não dos períodos específicos em que a covariância foi computada. Assim, se Y_t é um processo estocástico fracamente estacionário, então:

$$E(Y_t) = \mu$$

$$Var(Y_t) = E[(Y_t - \mu)^2] = \sigma^2$$

$$Cov(Y_t, Y_{t-k)}) = E[(Y_t - \mu)(Y_{t-k} - \mu)] = \gamma k$$

2.3.3 Operador de Defasagem

Ao trabalhar com séries temporais, o operador de defasagem B é conveniente, conhecido como operador de retardo ou translação para o passado, bastante usado por Box e Jenkins (2008) na descrição de modelos, aplicado sobre uma série temporal Y_t, o operador “defasa” a mesma em um período e é definido por $B^k\, Y_t = Y_{t-k}$, assim temos que $B^1\, Y_t = Y_{t-1}$, $B^2\, Y_t = Y_{t-2}$.

Em situações normais, será suficiente tomar uma ou duas diferenças para que a série se torne estacionária. Em geral, uma diferenciação de ordem d pode ser expressa como:

$$\Delta d Y_t = (1 - B) d Y_t$$

2.3.4 Função de Autocorrelação e Autocorrelação Parcial

Na análise de séries temporais, o interesse principal é modelar a estrutura de dependência que naturalmente está presente em observações ordenadas no tempo. As funções de autocorrelação e autocorrelação parcial são utilizadas extensivamente para auxiliar na consecução desse objetivo. A correlação entre duas variáveis aleatórias Y_t e Y_{t-k} separadas por k períodos é chamada de autocorrelação de ordem k, e definida como:

$$\rho_k = \frac{Cov\ (Y_t, Y_{t-k})}{Var\ (Y_t) Var(Y_{t-k})} = \frac{\gamma_k}{\gamma_0}$$

A partir da suposição de estacionaridade temos as propriedades da função de autocorrelação sendo $\rho_0 = 1$, para todo k temos $|\rho_k| \leq 1$ e $\rho_{-k} = \rho_k$ com distribuição Normal para $\hat{\rho}_k \sim Normal(0, S_{\hat{\rho}_k})$, onde o desvio padrão S é definido por:

$$S_{\widehat{\rho_k}} \cong \sqrt{\frac{1}{n}(1 + 2\hat{\rho}_1^2, \ldots, +2\hat{\rho}_{k-1}^2}$$

Quando plotado um gráfico ρ_k versus k, obtêm-se a função de autocorrelação (FAC), também chamado de correlograma. A Figura 1 apresenta uma simulação de um ruído branco gaussiano e sua respectiva função de autocorrelação com limites de significância de 5%:

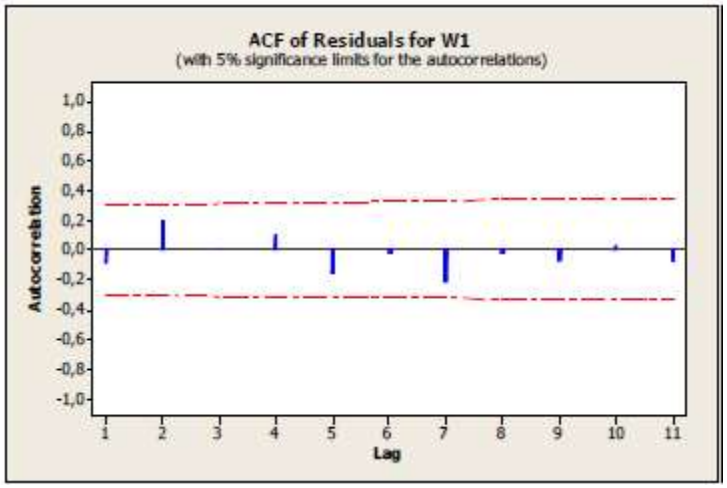


Figura 1: Exemplo de um Ruído Branco Gaussiano

Para efetuar a análise da estrutura de dependência entre Y_t e Y_{t-k}, as informações da função de autocorrelação precisam ser complementadas por um conceito associado, o de autocorrelação parcial. Após a remoção das dependências lineares das variáveis intermediárias $Y_t + 1$, $Y_t + 2$, ... $Y_t + k - 1$, tem-se a correlação condicional, que é a chamada autocorrelação parcial.

A autocorrelação parcial visa superar esse problema. Ela mensura a correlação linear entre Y_t e Y_{t-k} depois de removidos os efeitos das outras defasagens 1, 2, ..., k-1. A autocorrelação parcial no lag k pode ser estimada a partir do coeficiente de regressão:

$$Y_t = \phi_{k1} Y_{\varepsilon-1} + \cdots + \phi_{kk} Y_{\varepsilon-k} + a_t$$

$$\phi_{kk} = Corr\,(Y_t, Y_{t+k}/Y_{t+1}, \ldots, Y_{t+k-1})$$

A estimação da função de autocorrelação parcial com distribuição normal sendo $\hat{\phi}_{kk} \sim Normal\,(0, S_{\widehat{\phi_{11}}})$, por definição o desvio padrão S mensurado por $S_{\phi_{kk}} \cong \sqrt{\frac{1}{n}}$.

$$\hat{\phi}_{11} = \rho_1$$

$$\hat{\phi}_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}}$$

Considerando uma série temporal aleatória, plota-se um gráfico ϕ_{kk} versus k e obtêm-se a função de autocorrelação parcial (FACP). A Figura 2 apresenta a função de autocorrelação parcial do processo ruído branco gaussiano com os limites de significância de 5% de confiança:

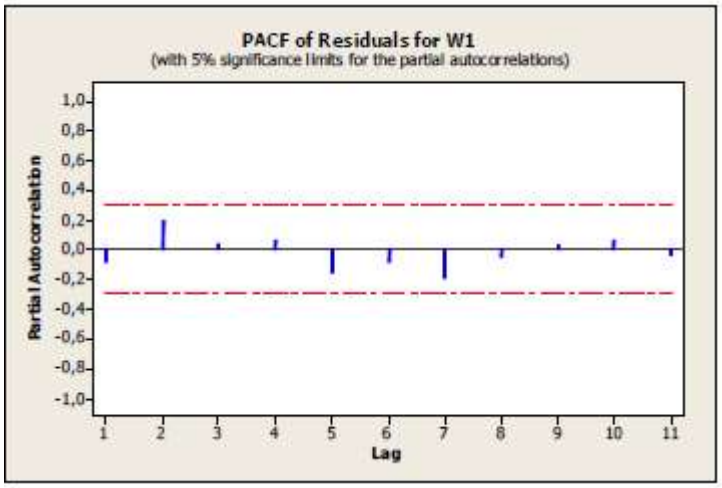


Figura 2: Função de Autocorrelação Parcial de um Ruído Branco Gaussiano

Um dos processos estocásticos mais utilizados na análise de séries temporais é o chamado ruído branco. Ele denota uma sequência de variáveis aleatórias independentes e identicamente distribuídas, com média zero e variância constante e não correlacionadas. Caso as variáveis aleatórias sigam a distribuição normal, é dito que tratar-se de um ruído branco gaussiano, assim satisfazendo $a_t = N(\mu, \sigma_a^2)\ \forall\ t = 1,2, \ldots$. Assim temos $E(a_t) = \mu = 0$, $Var(a_t) = \sigma_a^2$ e $Cov(a_t, a_{t-k}) = 0$.

Durante o processo de modelagem, um dos principais objetivos é encontrar resíduos que se comportem como um processo ruído branco.

2.3.5 Modelos ARMA

Dada uma série temporal Y_t estacionárias que apresenta uma estrutura de dependência serial entre observações, nosso objetivo será encontrar o melhor modelo que descreva esta dependência e a transforma num ruído branco. Um modelo de regressão em series temporais, Y_t pode ser representado por uma função

do passado da série Y_{t-1}, Y_{t-2},... Y_{t-p}, temos um modelo é chamado de modelo auto-regressivo. Neste modelo o suposto básico da independência dos erros (resíduo) pode ser facilmente violado, desde que as variáveis explicativas ou regressoras ($Y_{t-1}, Y_{t-2}, \ldots Y_{t-p}$) usualmente tem uma relação de dependência. Outro modelo de regressão em series temporais, que explique a variabilidade de Y_t pode ser usando os erros defasados no tempo, onde está explicito a relação da dependência entre os sucessivos erros defasados no tempo, é chamado de modelo de médias moveis.

Nos modelos auto-regressivos podem ser efetivamente acoplados os modelos de médias móveis, este modelo é uma forma geral da classe de modelos chamados, modelos auto-regressivos e de médias moveis.

2.3.5.1 Modelos autoregressivos de Ordem p - AR(p)

Em geral, um modelo auto-regressivo de ordem p é definido como segue:

$$Y_t = \theta_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots \quad \phi_p Y_{t-p} + a_t$$

Onde: o termo constante $\theta_0 = \mu(1 - \phi_1 - \phi_2 - .... - \phi_p)$, μ é a média do processo; ϕ_j é o j-ésimo parâmetro auto-regressivo, $\{a_t\}$ é um processo Ruído Branco com media zero $E(a_t) = 0$ e variância constante $Var(a_t) = \sigma^2$. A seguir as restrições dos parâmetros auto-regressivos que definem a condição de estacionaridade dos modelos:

Para p = 1, $-1 < \phi < 1$.

Para p = 2, as seguintes três condições definem a região de estacionaridade:

$$\phi_1 + \phi_2 < 1, \qquad \phi_2 + \phi_1 < 1, \qquad \text{e} \quad -1 < \phi_2 < 1;$$

Para p > 2 as condições para os coeficientes são mais complicadas, uma mistura de decaimento exponencial. Dependendo do sinal dos coeficientes, as autocorrelações podem apresentar decaimento alternado, se as raízes são complexas, as autocorrelações podem apresentar uma mistura de decaimento

exponencial senoidal. A função de autocorrelação parcial (FACP) apresenta as p primeiras correlações significativamente diferentes de zero.

2.3.4.2 Modelo Médias Móveis - MA(q)

Um modelo de médias móveis de ordem q pode ser descrito pela seguinte relação:

$$Y_t - \mu = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \ldots - \theta_q a_{t-q}$$

Onde: μ é a média do processo; θ_j é o j-ésimo parâmetro da componente médias móveis; $\{a_t\}$ é um processo ruído branco com média zero E(a_t)= 0 e variância constante Var(a_t)=σ^2.

A seguir as restrições dos parâmetros θ_i que definem a condição de invertibilidade dos modelos:

Para q = 1, -1< θ < 1.

Para q = 2, as seguintes três condições definem a região de invertibilidade:

$$\theta_1 + \theta_2 < 1, \qquad \theta_2 + \theta_1 < 1, \qquad \text{e} \quad -1 < \theta_2 < 1;$$

Para p > 2 as condições para os coeficientes são mais complicadas.

2.3.5.3 Modelo autoregressivo e medias moveis de ordem (p,q) ARMA(p,q)

Nos modelos auto-regressivos podem ser efetivamente acoplados os modelos de médias móveis:

$$Y_t - \phi_1 Y_{t-1} - \ldots \quad \phi_p Y_{t-p} = \theta_0 + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q}$$

Usando o operador de retardo B temos outra forma de representar o modelos ARMA(p,q):

$$\Phi(B) Y_t = \Theta_0 + \Theta(B) a_t$$

Onde:

Θ_0 é a constante do processo;

$\Phi(B) = (1 - \phi_1 B - \ \ldots \ - \phi_p B^p) = 0$ é o polinômio autorregressivo de ordem "***p***" com raízes fora do círculo unitário, condição de estacionaridade do processo;

$\Theta(B) = (1 - \theta_1 B - \ \ldots \ - \theta_q B^q) = 0$ é o polinômio de medias moveis de ordem "*q"* com raízes fora do círculo unitário, condição de invertibilidade do processo.

$\{a_t\}$ é um processo Ruído branco com média zero E(a_t) = 0 e variância constante Var(a_t) = σ^2 este modelo é uma forma geral da classe de modelos chamados, modelos auto-regressivos e de médias móveis.

2.3.6 Modelos ARIMA

Em situações normais, será suficiente tomar uma ou duas diferenças para que a série se torne estacionária. Então podemos falar que ama diferenciação de ordem "*d"* aplicada a série não estacionaria, $Z_t = (1-B)^d Y_t$ se torna estacionária e segue um modelo ARMA(p,q):

$$\Phi_p(B) Z_t = \Theta_q(B) a_t$$

Então o modelo para a série Y_t é da forma, onde:

$$\Phi_p(B)(1-B)^d Y_t = \Theta_0 + \Theta_q(B) a_t$$

$\Theta_0 = (1 - \phi_1 - \ \ldots \ - \phi_p)$ é a constante do processo;

$\Phi(B) = (1 - \phi_1 B - \ \ldots \ - \phi_p B^p) = 0$ é o polinômio auto-regressivo de ordem "*p"* com raízes fora do círculo unitário, condição de estacionaridade do processo;

$\Theta(B) = (1 - \theta_1 B - \ \ldots \ - \theta_q B^q) = 0$ é o polinômio de medias moveis de ordem "*q"* com raízes fora do círculo unitário, condição de invertibilidade do processo.

$\{a_t\}$ é um processo Ruído branco com média zero E(a_t) = 0 e variância constante Var(a_t) = σ^2 este modelo é uma forma geral da classe de modelos chamados, modelos auto-regressivos integrados de médias móveis ARIMA(p,d,q).

2.3.7 Modelo SARIMA

Os modelos $ARIMA$ visa tratar o problema da sazonalidade. Isso ocorre quando existe autocorrelação significativa entre defasagens sazonais, ou seja, múltiplos de um período qualquer *"S"*. Neste caso, tem-se um modelo $ARIMA$ sazonal ($SARIMA$). Para expressarmos este tipo de modelo, precisamos estender a

notação utilizada nos modelos $ARIMA(p,\ d,\ q)$ para $SARIMA(p,\ d,\ q)(P,\ D,\ Q)s$, onde $(p,\ d,\ q)$ representa a parte não sazonal, $(P,\ D,\ Q)$ a parte sazonal e s uma sazonalidade de s períodos.

Os modelos para uma série temporal Y_t, que apresenta uma determinada tendência, com marcada variabilidade sazonal, com período sazonal *"S",* que após de aplicar diferenciação não sazonal de ordem *d* e diferenciação sazonal (desazonalização) de ordem *D*, se tem uma série estacionária, usando o operador de retardo *B*, ($W_{t-k}=B^kW$) está diferenciação pode ser representado pela relação:

$$W_t = (1-B)^d \left(1-B^S\right)^D Y_t$$

.

Para os processos com marcadas variabilidades sazonais, após uma diferenciação, pode-se apresentar uma estrutura de dependência serial entre observações em períodos sazonais, que pode ser representado por uma combinação do passado da série e/ou uma relação de dependência do passado dos erros. Está variabilidade sazonal pode ser representada por uma componente autorregressivo sazonal de ordem *P*, e uma componente de médias móveis sazonal de ordem *Q*, a equação do modelo para a série diferenciada é:

$$W_t = \Phi_1 W_{t-S} + \quad ... \quad + \Phi_P W_{t-SP} = a_t - \Theta_1 a_{t-S} - \ldots - \ldots - \Theta_Q a_{t-SQ}$$

Usando o operador de defasagem *B*, chegamos a seguinte representação da equação do modelo,

$$\Phi_P(B^S) W_t = \Theta_Q(B^S)\, a_t$$

Onde:

$\Phi_P\left(B^S\right) = \left(1-\phi_1 B^S - ... - \phi_P B^{SP}\right) = 0$, é o polinômio autorregressivo sazonal de ordem P;

$\Theta_Q\left(B^S\right) = \left(1-\theta_1 B^S - \ldots - \theta_Q B^{QS}\right) = 0$, é o polinômio de médias móveis sazonal de ordem Q, com raízes fora do círculo unitário e sem raízes comuns; e $\{a_t\}$ é um processo Ruído Branco com média zero $E(a_t)=0$ e variância constante $Var(a_t)=\sigma^2$. Este modelo é chamando de modelo *Autorregressivo Sazonal Integrado e Médias Móveis Sazonal,* denotado como segue *SARIMA(p,d,q)(P,D,Q)s.*

Um modelo mais geral é considerando que a série diferenciada apresente uma estrutura de dependência serial entre observações em períodos sazonais e não

sazonais para a série não diferenciada, onde a variabilidade da série pode ser representada como segue:

$$\Phi_{\mathrm{p}}(B)\ \ \Phi_{P}\left(B^{S}\right)\ \left(1-B\right)^{d}\left(1-B^{S}\right)^{D}Y_{t}=\Theta_{\mathrm{Q}}(B^{S})\ \ \Theta_{q}\left(B\right)a_{t}$$

Onde:

$\Phi_{p}(B)=\left(1-\phi_{1}B-\phi_{2}B^{2}-\ldots-\phi_{p}B^{p}\right)=0$, é o polinômio autorregressivo de ordem p;

$\Theta_{q}\left(B\right)=\left(1-\theta_{1}B-\theta_{2}B^{2}-\ldots-\theta_{q}B^{q}\right)=0$, é o polinômio de médias móveis de ordem q com raízes fora do círculo unitário e sem raízes comuns, e as componentes sazonais;

$\Phi_{P}\left(B^{S}\right)=\left(1-\phi_{1}B^{S}-\ldots-\phi_{P}B^{SP}\right)=0$, é o polinômio autorregressivo sazonal de ordem P;

$\Theta_{Q}\left(B^{S}\right)=\left(1-\theta_{1}B^{S}-\ldots-\theta_{Q}B^{QS}\right)=0$, é o polinômio de médias móveis sazonal de ordem Q com raízes fora do círculo unitário e sem raízes comuns; e $\{a_{t}\}$ é um processo Ruído Branco com média zero E(a_{t})=0 e variância constante Var(a_{t})=σ^{2} , este modelo é chamando de modelo Autorregressivo Sazonal Integrado e Médias Móveis Sazonal, denotado por *SARIMA(p,d,q)(P,D,Q)$_{s}$*.

2.4. Fluxograma das Etapa

A metodologia aplicada segue um procedimento em três etapas apresentadas na Figura 3. Na primeira etapa: descrição e agregação das séries de impostos, foram analisadas as principais características de cada série de impostos, seguido pelo processo de agregação de séries, aplicando-se a técnica de análise fatorial, via matriz de correlação. Na segunda etapa: modelagem das séries de fatores, a modelagem das séries de fatores foi aplicando-se os modelos SARIMA, e a terceira etapa: previsão de séries de impostos, onde foram calculadas as previsões 12 passos à frente dos fatores latentes, para finalizar com as previsões de cada série de impostos, desagregando as previsões dos fatores, seguido da despadronização e agregação do fator sazonal.

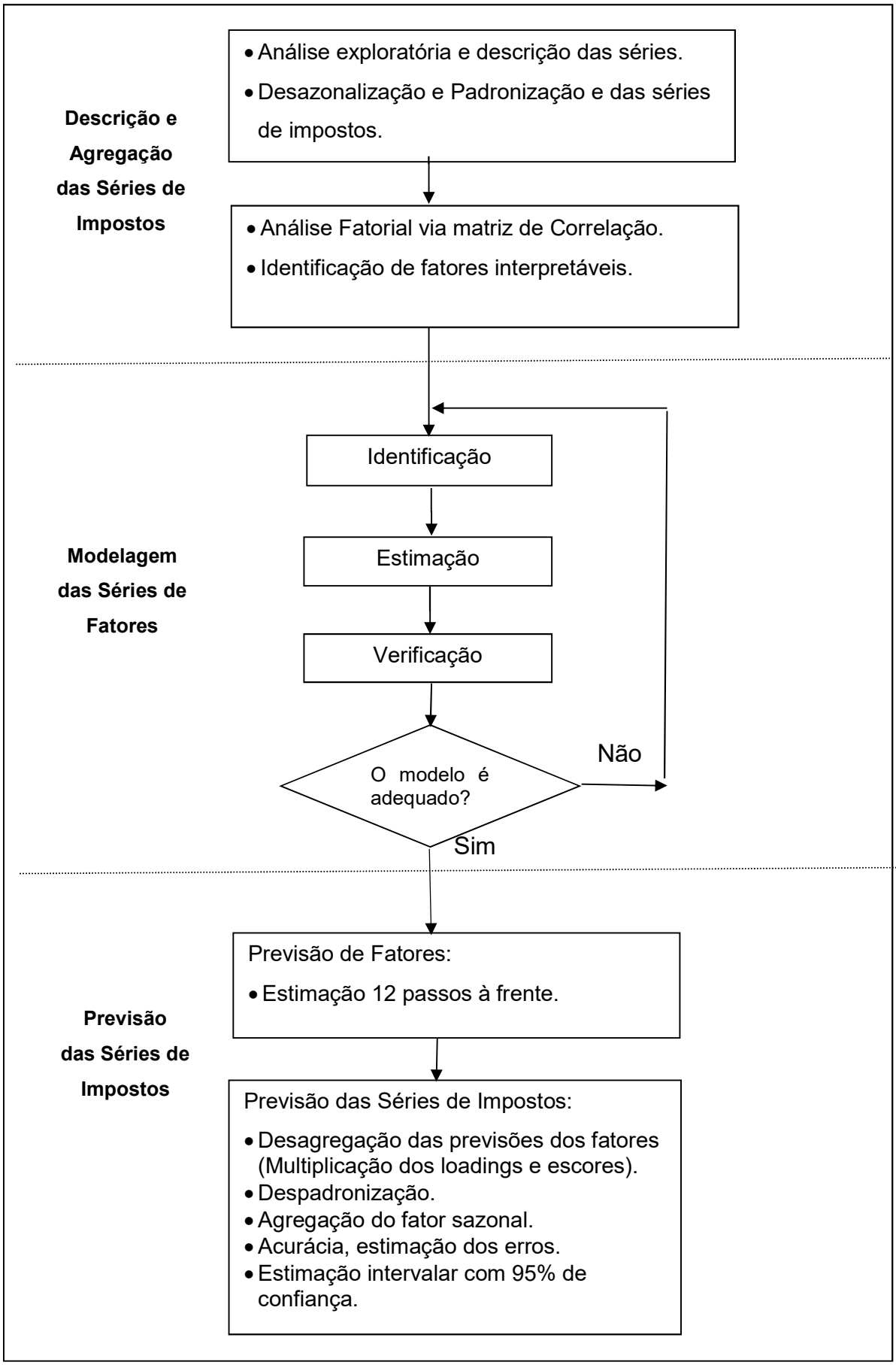


Figura 3: Fluxograma das etapas da metodologia

3. ANÁLISE DOS RESULTADOS

3.1. Base de Dados

A base de dados original consiste em dez impostos diferentes com 156 observações mensais cada um, no período de janeiro de 2001 a dezembro de 2013, sendo estes impostos federais e arrecadados em todo território brasileiro. São agrupados da seguinte forma: dois tributos originam sobre as contribuições previdenciárias, a Contribuição Social sobre o Lucro Líquido (CSLL) e a Contribuição para o PIS/PASEP (CPIS), três são incidentes sobre a renda, o Imposto de Renda sobre Pessoa Jurídica (IRPJ), o Imposto de Renda sobre Pessoa Física (IRPF) e o Imposto de Renda Retido na Fonte (IRRF), outra parcela são incidentes sobre os produtos, o Imposto sobre Importação (IIMP), a Contribuição para o Financiamento da Seguridade Social (COFINS), o Imposto sobre as Operações Financeiras (IOF) e o Imposto sobre Produtos Industrializados (IPI) e por último o item Outros Receitas (Outros Impostos) que é composto de outras receitas tributárias de menor valor.

O período de janeiro de 2001 a dezembro de 2012 foi utilizado para verificação da sazonalidade, ajuste dos modelos e previsões, já o período compreendido entre janeiro de 2013 a dezembro de 2013 foi usado para validação da acurácia das previsões. A base foi fornecida pelo Instituto de Pesquisa Econômica Aplicada (IPEA) e esta é gerenciada pela Secretaria de Política Econômica (SPE).

3.2. Softwares

Para desenvolvimento deste trabalho foi utilizado o software Minitab, na sua versão 16 e também o software R, na sua versão 3.1.0 (2014-04-10).

O Minitab é um software que fornece um ambiente completo para análise de dados, opera simultaneamente com planilhas de dados, tabelas estatísticas, gráficos e textos.

Já o R é ao mesmo tempo uma linguagem de programação e um ambiente para estatística computacional, ele possui uma série de técnicas implementadas e

pacotes estatísticos, bem como a funcionalidade que suportam todas as fases do processo de análise de dados.

3.3. Análise Exploratória

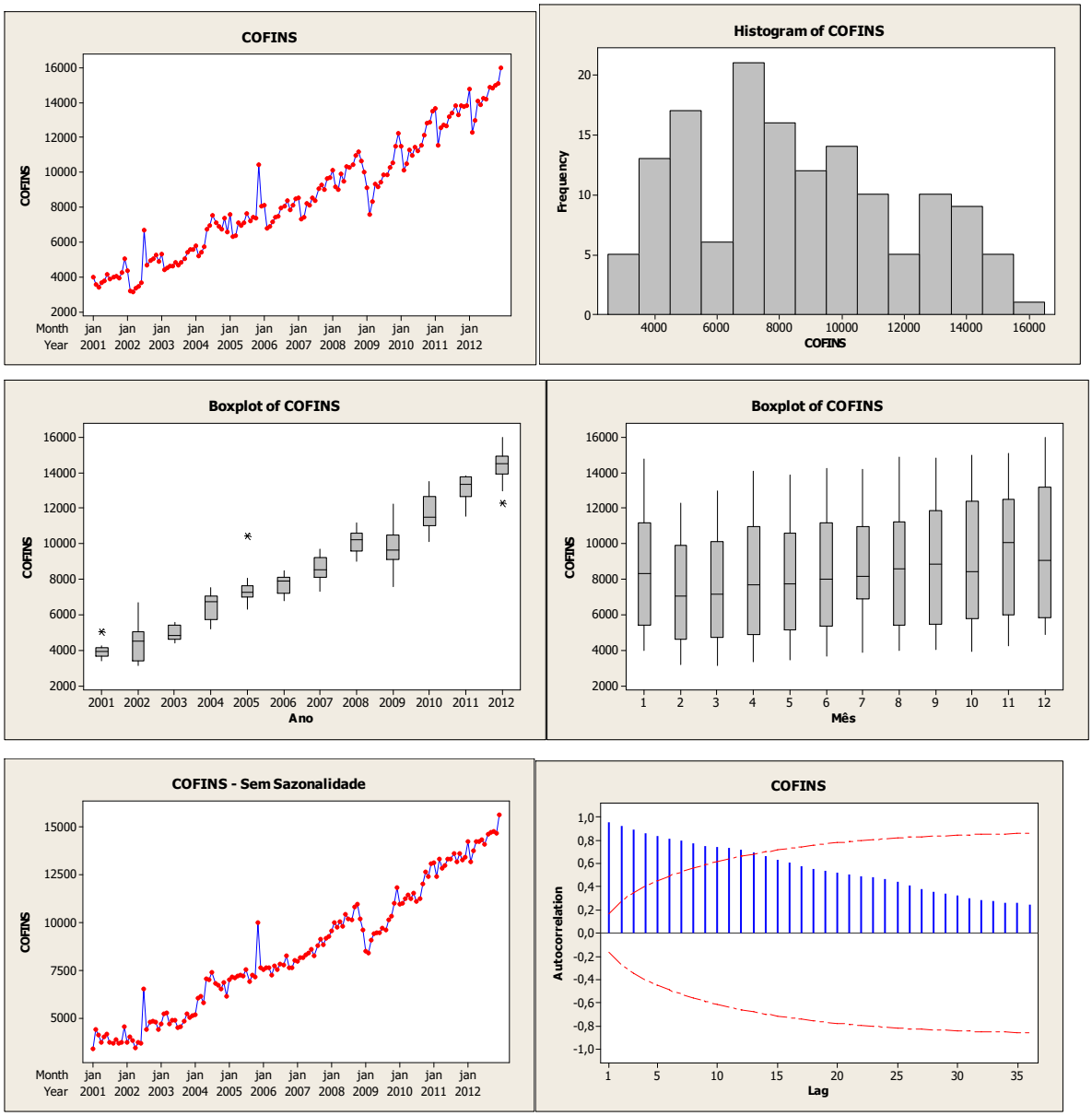
Nesta seção será apresentados para cada série: o gráfico de linha, box-plot por ano, box-plot por mês, histograma da série, a série desazonalizada, correlogramas da série, as estatísticas descritivas de cada série

Para a série COFINS constata-se a existência de tendências com a presença de alguns pontos discrepantes, olhando para os gráficos nota-se uma variação no tempo da série apontando para um efeito sazonal, uma distribuição não simétrica da série e nos correlogramas é visível que a série é não estacionária.

A série CPIS possui tendência crescente, uma leve variabilidade sazonal e chama muito a atenção por um outlier em janeiro de 2011 e os gráficos de box-plot confirmam a expectativa de processos serem estacionários, vide na Figura 5.

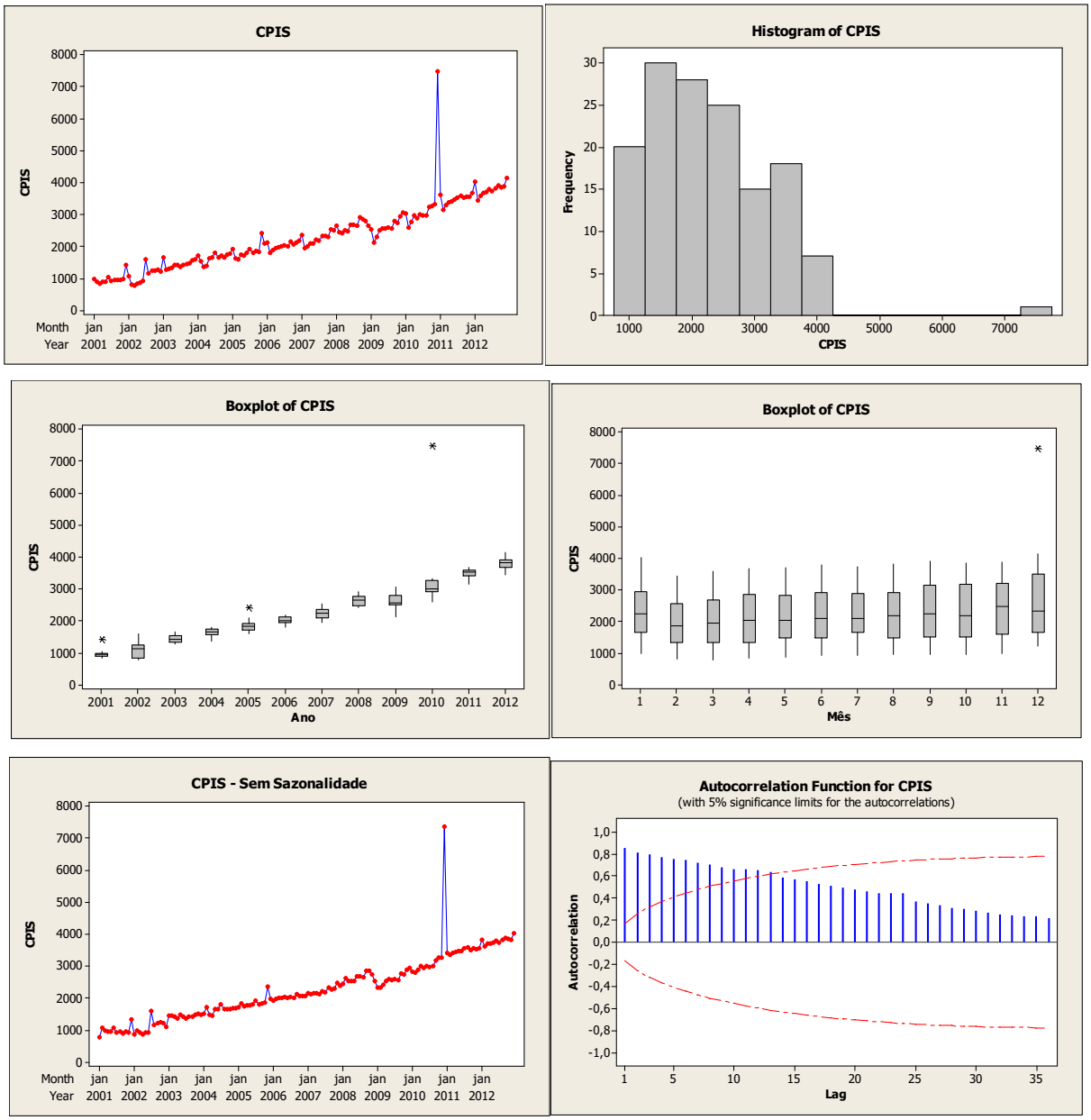
Na série IOF tem-se uma quebra estrutural na média e variância com variabilidade sazonal aditiva, a curtose aponta para um valor não significativo para e por meio dos box-plot e histograma percebe-se que não há simetria nos dados. Na Figura 6 está a descrição da série de impostos IOF.

As séries CSLL, IRRF e IRPF, apresentadas nas figuras 7, 8 e 9, possuem componente sazonal de efeito multiplicativo que não foram trabalhados, as séries possuem tendências, não são estacionárias e as medidas descritivas de curtose e simetria remetem a séries com grande variabilidade.



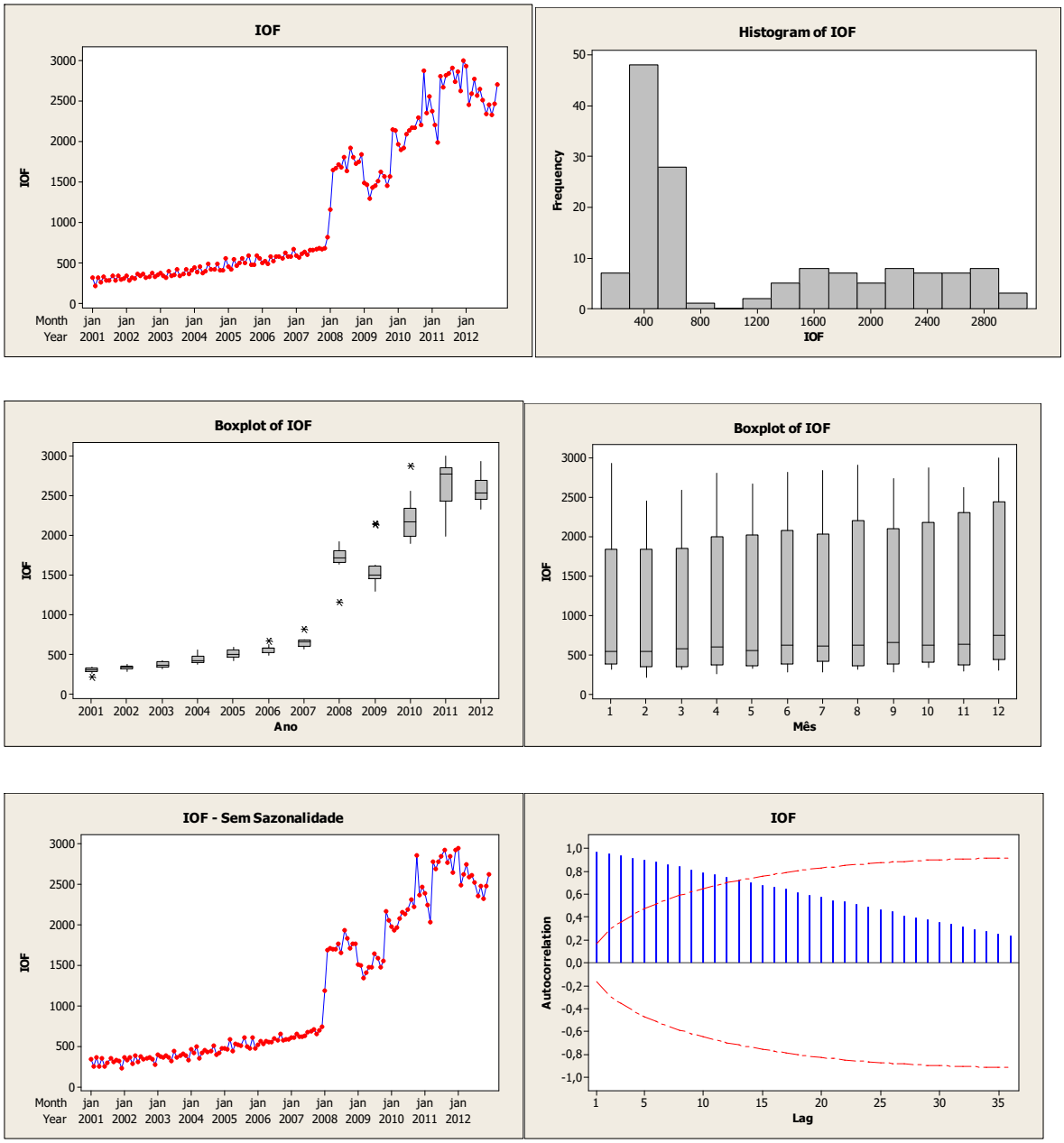
Série	Média	Desvio Padrão	Mediana	Assimetria	Curtose
COFINS	8.538,00	3.356,00	8.102,00	0,27	-0,93

Figura 4: Descrição da Série COFINS



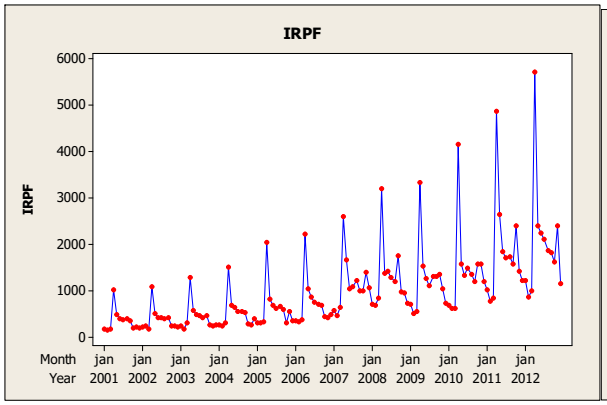
Série	Média	Desvio Padrão	Mediana	Assimetria	Curtose
CPIS	2.265,20	985,70	2.118,50	1,17	4,11

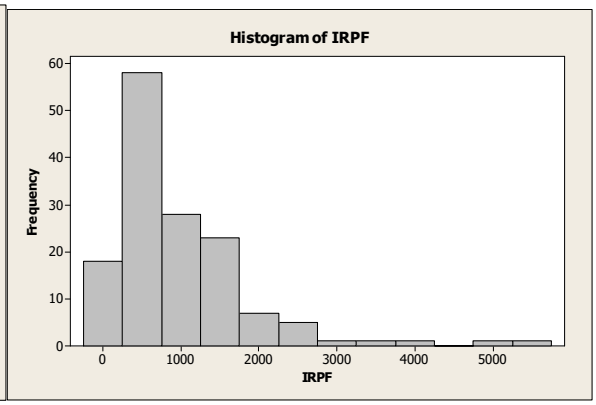
Figura 5: Descrição da Série CPIS

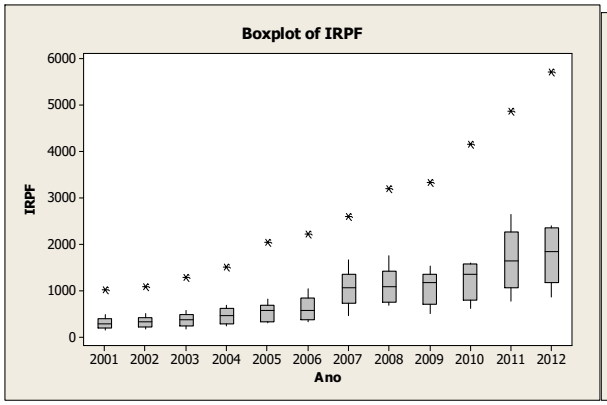


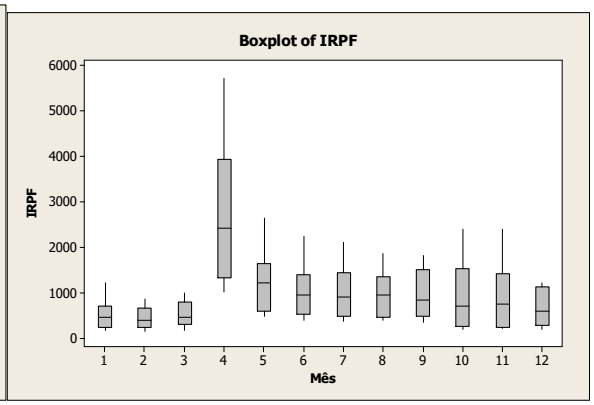
Série	Média	Desvio Padrão	Mediana	Assimetria	Curtose
IOF	1.157,20	901,20	595,40	0,67	-1,15

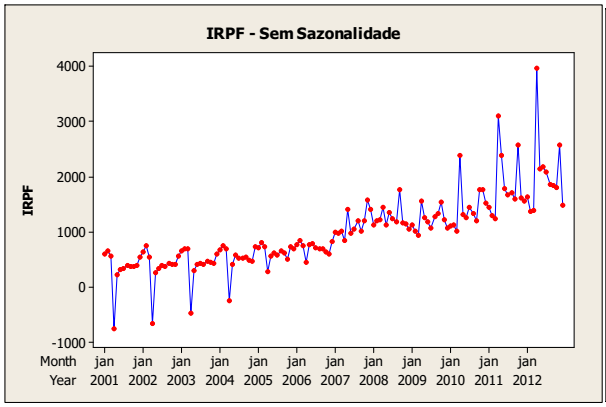
Figura 6: Descrição da Série IOF

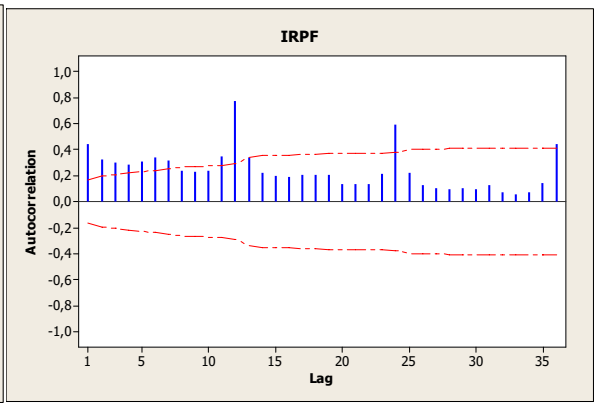






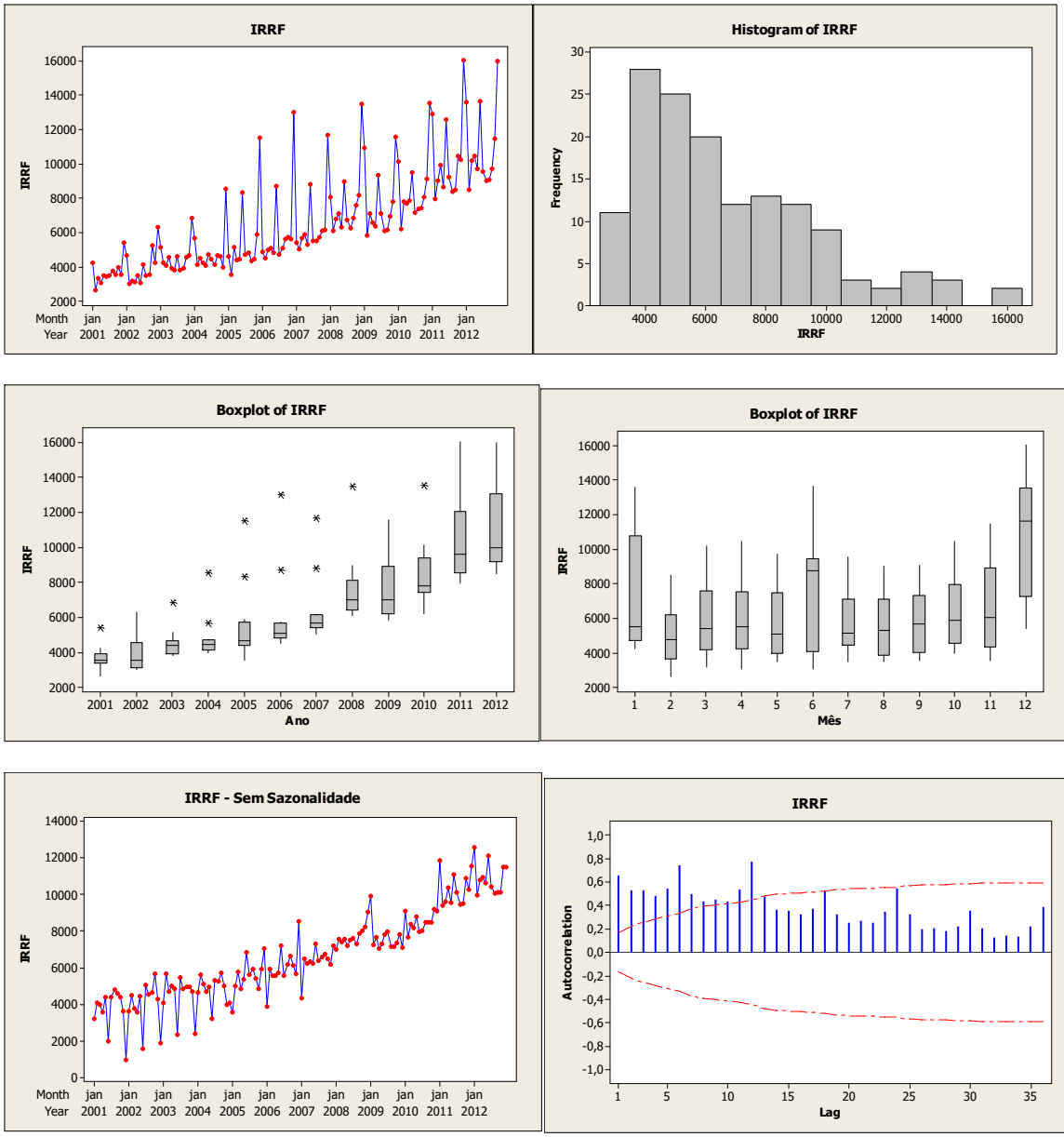






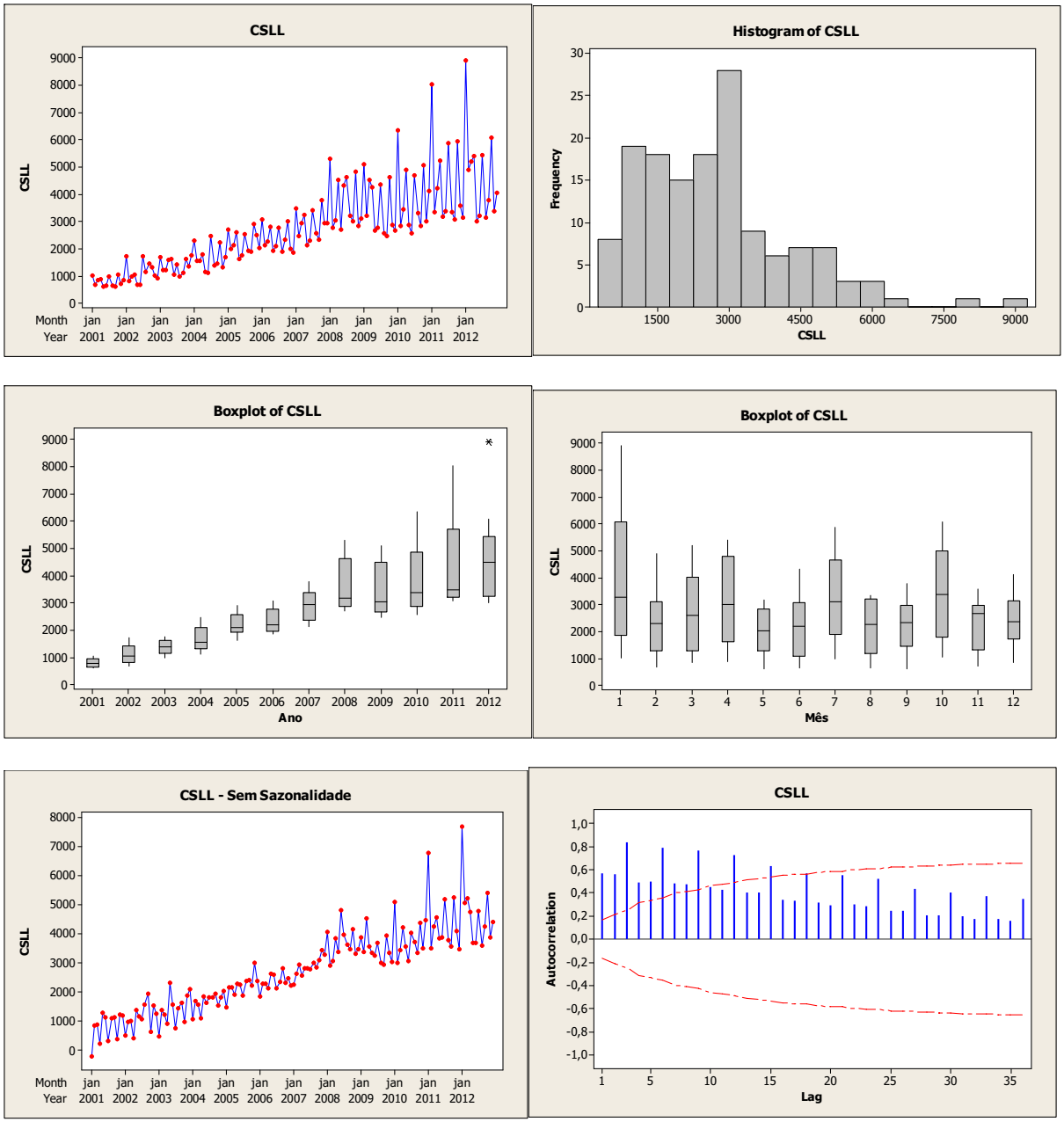
Série	Média	Desvio Padrão	Mediana	Assimetria	Curtose
IRPF	988,70	876,10	698,40	2,42	8,39

Figura 7: Descrição da Série IRPF



Série	Média	Desvio Padrão	Mediana	Assimetria	Curtose
IRRF	6.658,00	2.906,00	5.835,00	1,05	0,65

Figura 8: Descrição da Série IRRF

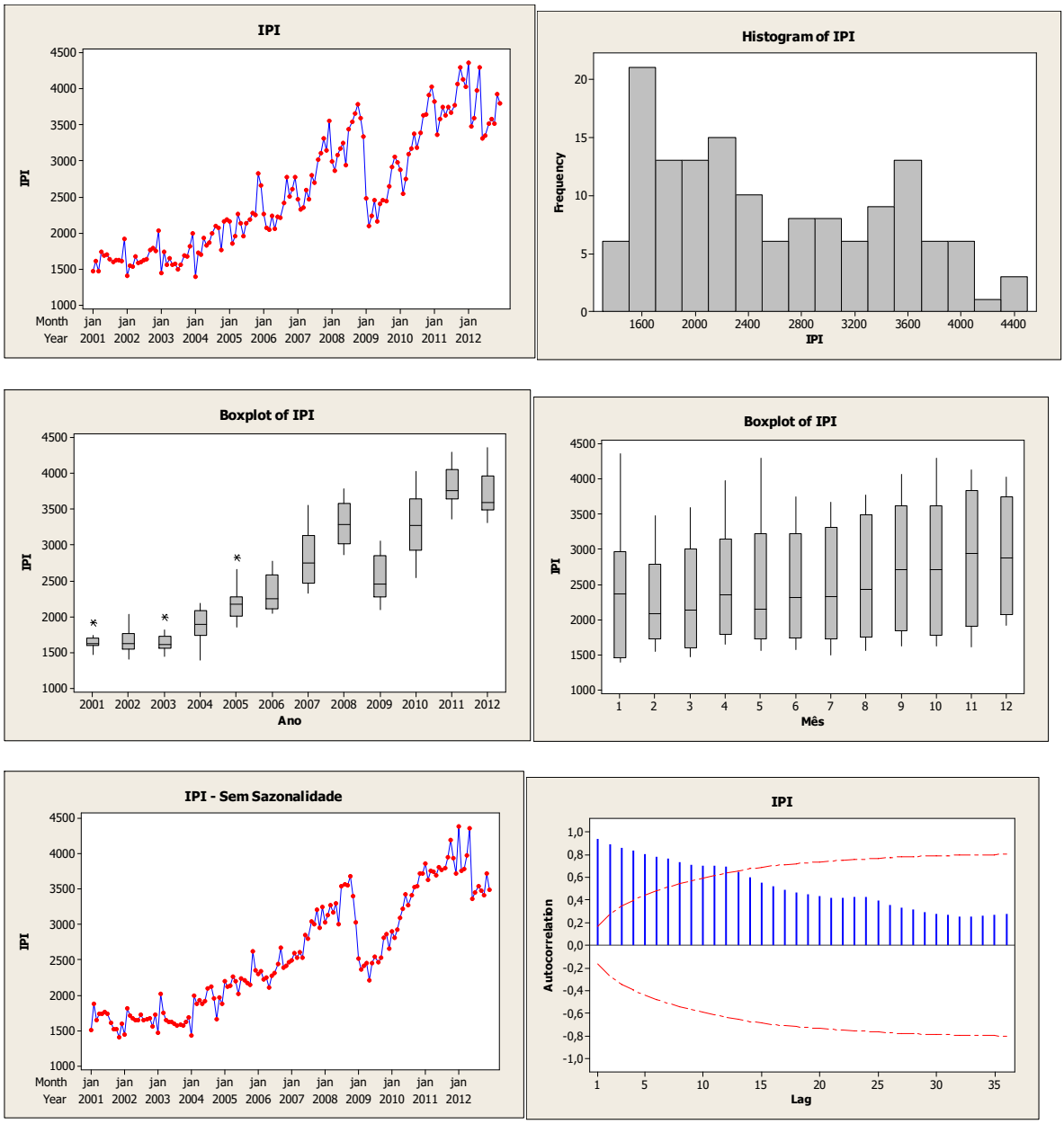


Série	Média	Desvio Padrão	Mediana	Assimetria	Curtose
CSLL	5.208,00	2.925,00	4.930,00	1,05	1,53

Figura 9: Descrição da Série CSLL

No tributo IPI, na figura 10 pelas estatísticas descritivas e observações dos gráficos vê-se uma alteração na tendência ao longo do tempo observado, possivelmente devido às políticas de incentivo de consumo com redução do IPI, varia muito em torno da média. Observando a FAC ao nível de 5% de confiança nota-se que os decaimentos não se aproximam muito do zero, isto ocorre porque as

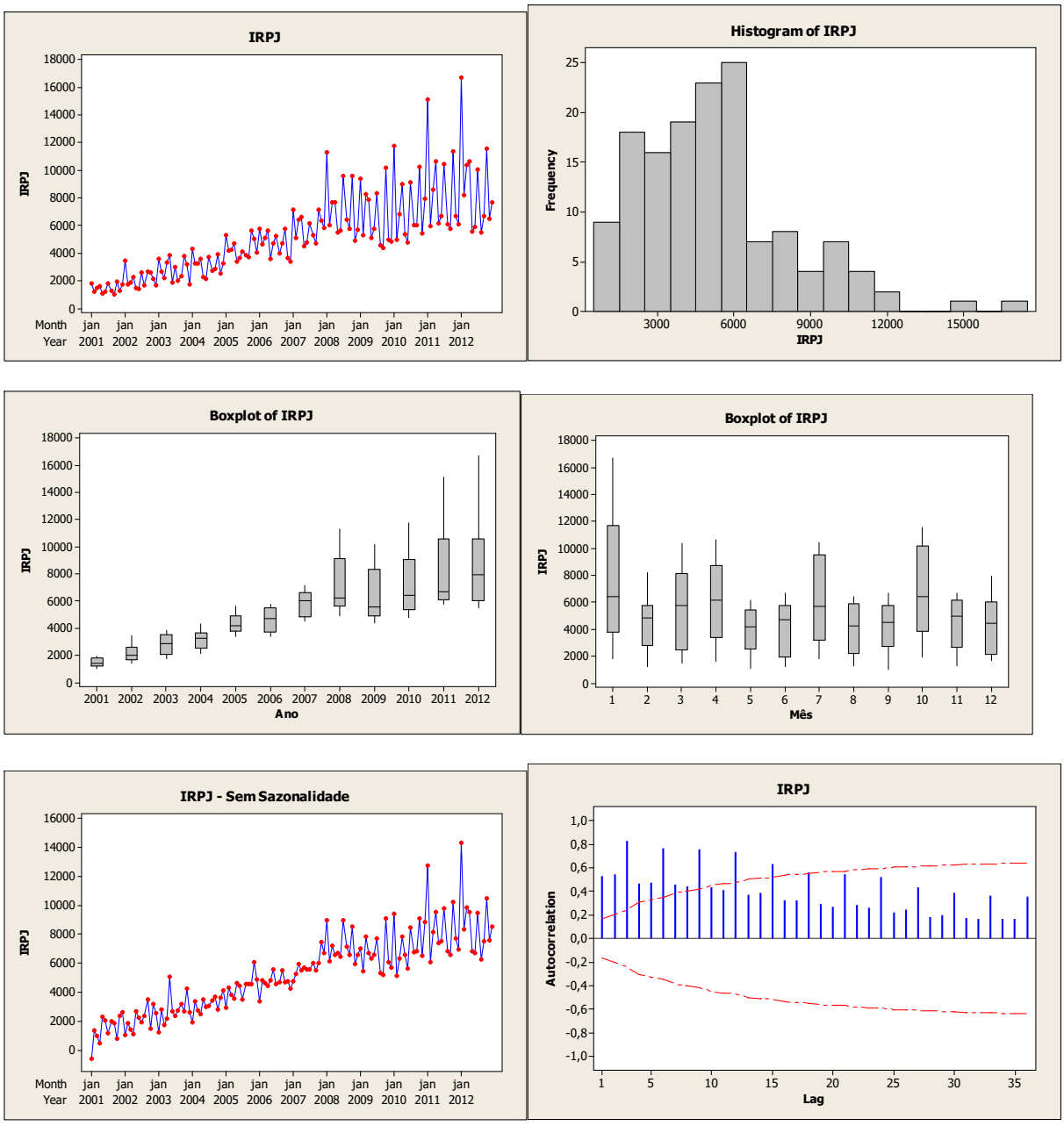
observações de um lado da média tende a ser seguida por um grande número de observações do mesmo lado (devido à tendência), é evidente que as autocorrelações amostrais decaem muito lentamente. Na FACP apresenta uma única autocorrelação significativa na primeira defasagem e as demais são não significativas.



Série	Média	Desvio Padrão	Mediana	Assimetria	Curtose
IPI	2.576,70	829,70	2.431,90	0,38	-1,11

Figura 10: Descrição da Série IPI

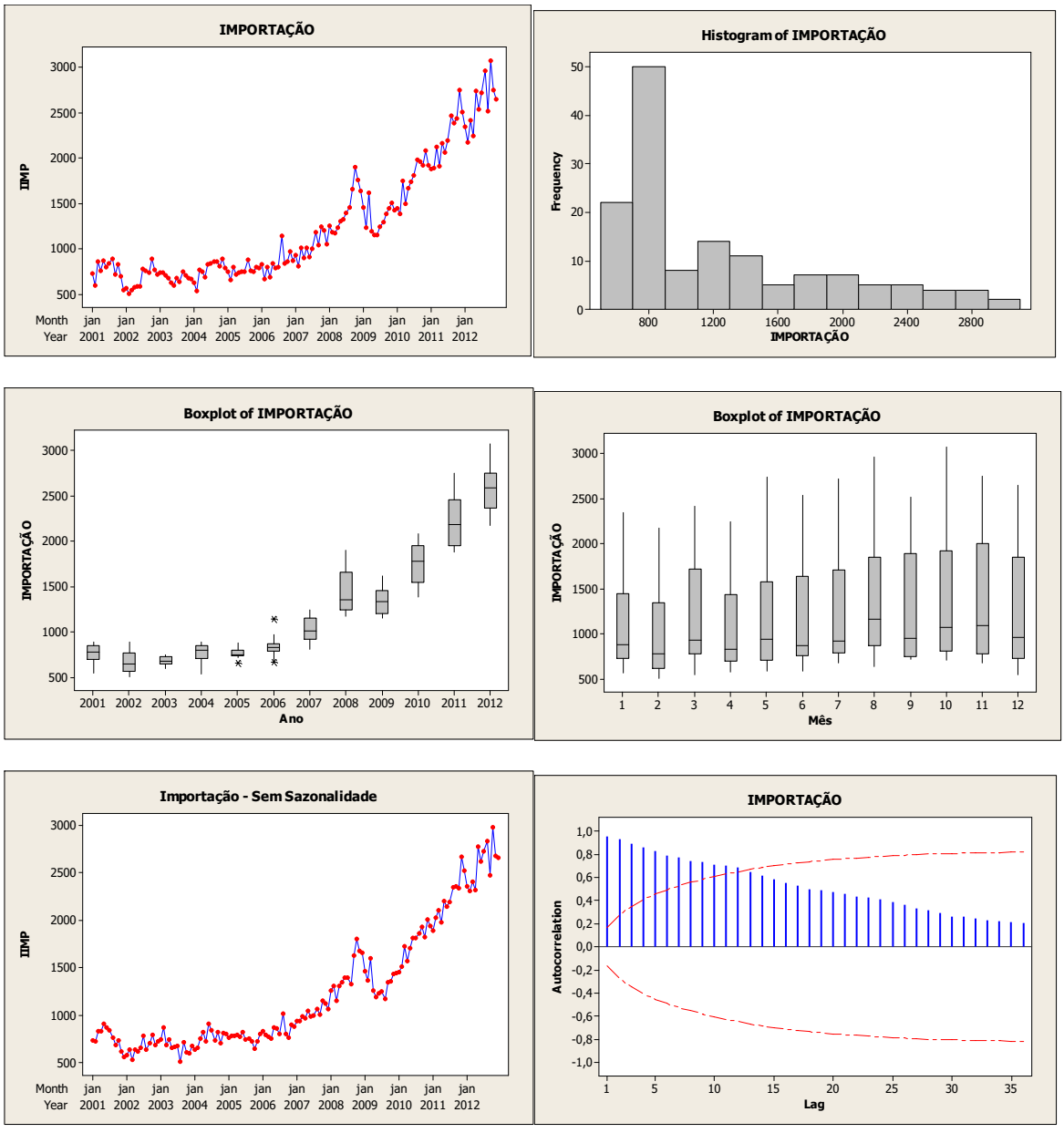
Na figura 11 é apresentado o IRPJ, a série apresenta tendência crescente sem quebra estrutural e sem presença de outliers, com sazonalidade aditiva, pelo box-plot nota-se como o IRPJ aumenta a partir de 2008 e o comportamento da média mês a mês aponta para grandes variações. Observando a FAC ao nível de 5% de confiança nota-se que os decaimentos não se aproximam muito do zero, isto ocorre porque as observações de um lado da média tende a ser seguida por um grande número de observações do mesmo lado (devido à tendência).



Série	Média	Desvio Padrão	Mediana	Assimetria	Curtose
IRPJ	2.707,00	1.530,00	2.570,00	1,09	1,74

Figura 11: Descrição da Série IRPJ

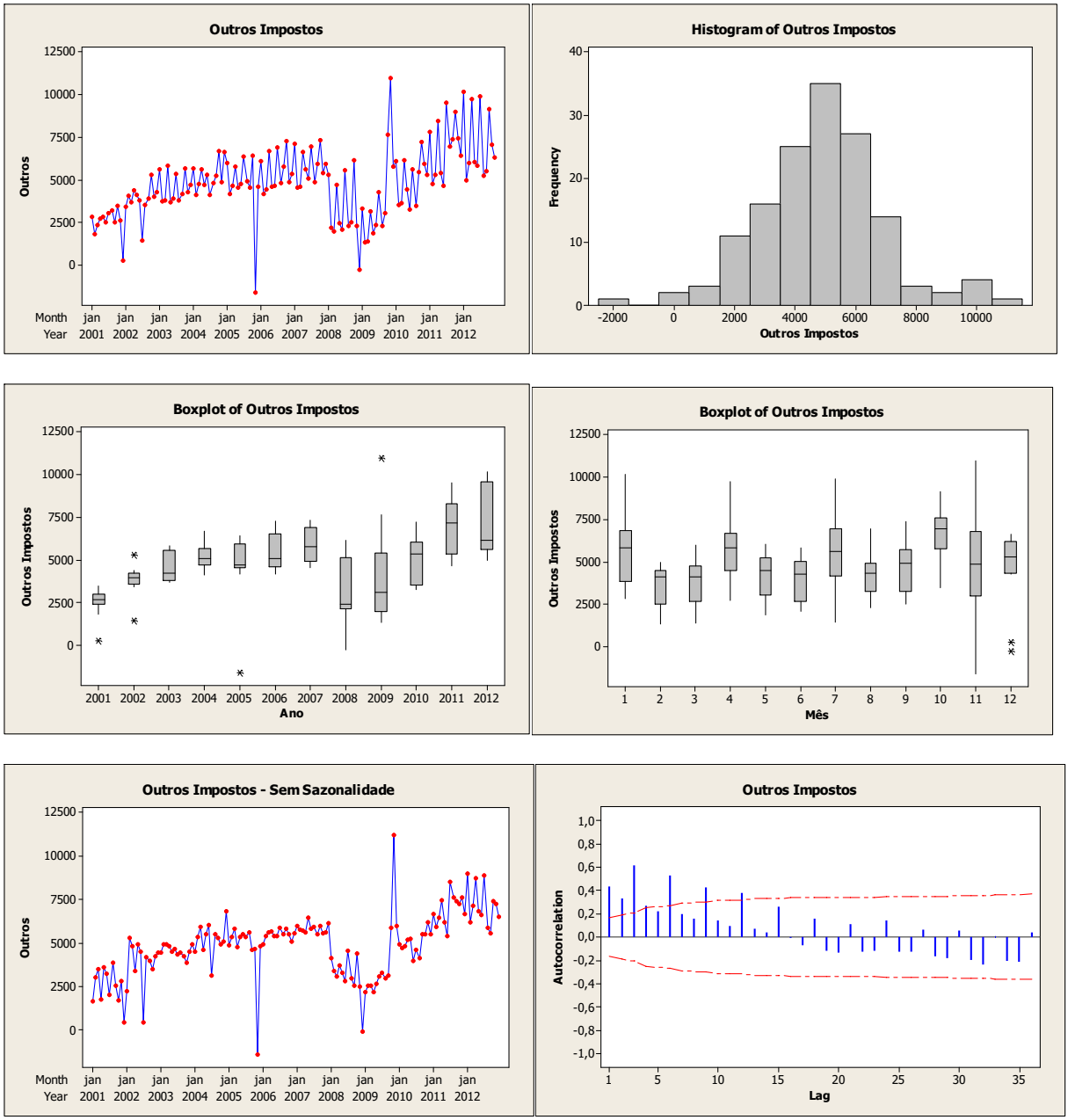
A série IIMP (Importação) também demonstra que não é estacionária, pois apresenta uma tendência crescente de acordo com os gráficos constantes na figura 12, baixa assimetria e alta variabilidade em torno da média, curtose muito baixa e no correlograma FAC ao nível de 5% de confiança nota-se que os decaimentos demoram para decaem lentamente.



Série	Média	Desvio Padrão	Mediana	Assimetria	Curtose
IMPORTAÇÃO	1.236,30	647,50	904,00	1,06	0,02

Figura 12: Descrição da Série IIMP

Por fim, a série Outros Impostos registra uma mudança na tendência ao longo do tempo observado, o processo é não estacionário e alta variabilidade nos dados em torno da média e do desvio-padrão, e também presença de dados discrepantes.



Série	Média	Desvio Padrão	Mediana	Assimetria	Curtose
Outros Impostos	4.857,00	2.025,00	4.773,00	0,17	1,01

Figura 13: Descrição da Série Outros Impostos

Um resumo das características das séries é apresentado na Tabela 1.

Tabela 1: Características das Séries de Tributos no Período de jan/2001 a dez/2012

Série	Tendência	Sazonalidade	Assimetria	Quebra Estrutural na μ?	Mudança variabilidade?	Outliers?
CONFINS	Crescente	Aditiva	0,27	Não	Não	Não
CPIS	Crescente	Aditiva	1,17	Não	Não	**Sim**
IRRF	Crescente	Multiplicativa	1,05	Não	Não	Não
IPI	Crescente	Aditiva	0,38	Não	Não	Não
IOF	Crescente	Aditiva	0,67	**Sim**	**Sim**	Não
IRPF	Crescente	Multiplicativa	**2,42**	Não	Não	Não
CSLL	Crescente	Multiplicativa	1,05	Não	**Sim**	Não
IRPJ	Crescente	Aditiva	**1,74**	Não	**Sim**	Não
IIMP	Crescente	Aditiva	1,06	**Sim**	Não	Não
Outros Impostos	Crescente	Aditiva	1,01	**Sim**	**Sim**	**Sim**

Conforme apresentado na Tabela 1 relata-se a existência de movimentos sazonais nas séries, que pode ser aditiva ou multiplicativa. No caso aditivo, a série mostra uma flutuação sazonal estável, sem levar em consideração o nível médio da série; no caso multiplicativo, o tamanho da flutuação sazonal varia, dependendo do nível médio da série. Existem dois interesses principais no ajuste de séries temporais para variação sazonal: o estudo da sazonalidade propriamente dita e a remoção da sazonalidade da série para depois estudá-la em seus demais aspectos. Esse último é o que se interessa para este trabalho, onde foi removida a sazonalidade de todas as séries de impostos e trabalha-se com as séries desazonalizadas.

São apresentados para cada série: o gráfico de linha, box-plot por ano, box-plot por mês, histograma da série, a série desazonalizada, correlogramas da série, as estatísticas descritivas de cada série e a tabela das correlações entre as séries.

No período de estudo, todas as séries de impostos apresentam uma tendência crescente, com variabilidade sazonalidade aditiva (COFINS, CPIS IPI, IOF, IIMP, IRPJ e OUTROS) ou multiplicativa (CSLL, IRRF e IRPF).

As séries de impostos que apresentam menor assimetria são as séries que não apresentam quebras estruturais e sem dados discrepantes, séries com uma variabilidade mais homogênea, COFINS e IPI. Todas as séries de impostos são altamente correlacionadas, e todas as correlações são significativas, conforme apresentado na Tabela 2 de Matriz de Correlações entre as Séries de Tributos. Sendo assim, rejeita-se a hipótese de correlação zero.

Tabela 2: Matriz de Correlações entre as Séries de Tributos

	CONFINS	CPIS	IOF	CSLL	IRPJ	IRPF	IRRF	IPI	IIMP	OUTROS
COFINS	**1.00000**	0.95107	0.93068	0.87346	0.85872	0.83004	0.94058	0.94614	0.93328	0.58262
CPIS	0.95107	**1.00000**	0.90107	0.83917	0.82762	0.78543	0.89381	0.90542	0.88584	0.54862
IOF	0.93068	0.90107	**1.00000**	0.81785	0.79342	0.80711	0.90830	0.91110	0.90655	0.49565
CSLL	0.87346	0.83917	0.81785	**1.00000**	0.98507	0.72534	0.85698	0.87104	0.78677	0.51252
IRPJ	0.85872	0.82762	0.79342	0.98507	**1.00000**	0.71897	0.84505	0.86461	0.76469	0.52214
IRPF	0.83004	0.78543	0.80711	0.72534	0.71897	**1.00000**	0.78980	0.80306	0.79106	0.52667
IRRF	0.94058	0.89381	0.90830	0.85698	0.84505	0.78980	**1.00000**	0.90387	0.90778	0.55803
IPI	0.94614	0.90542	0.91110	0.87104	0.86461	0.80306	0.90387	**1.00000**	0.87871	0.54530
IIMP	0.93328	0.88584	0.90655	0.78677	0.76469	0.79106	0.90778	0.87871	**1.00000**	0.57164
OUTROS	0.58262	0.54862	0.49565	0.51252	0.52214	0.52667	0.55803	0.54530	0.57164	**1.00000**

3.4. Resultados da Análise Fatorial

A análise fatorial das séries de tributos desazonalizadas e padronizada foram realizadas por meio do método de máxima verossimilhança, indicado para dados distribuídos normalmente e apontam a construção de três fatores tomando como referência a explicação de 85% da variabilidade dos dados.

A análise fatorial nesse trabalho se apresentou promissora em organizar as séries em fatores, compostos por tributos altamente relacionados. Pois uma escolha adequada do valor de m deve, no entanto, levar em consideração a interpretabilidade dos fatores e o princípio da parcimônia, ou seja, a descrição da estrutura de variabilidade do vetor aleatório Z com um número pequeno de fatores, segundo MINGOTI (2005).

Neste estudo se trabalhou com três fatores (m = 3), apesar de ter apenas 1 fator como solução conforme a Tabela 2, a escolha do número de fatores está ligada ao modelo SARIMA de previsão das séries, que de acordo com Mendonça e Medrano (2014) com um fator não seria o suficiente para obter um modelo que prevesse todas as dez séries de tributos, pois já realizaram este estudo anteriormente do número de fatores. Sendo assim, três fatores são capazes de prever o menor erro para os modelos em análise.

A variância total explicada pelo modelo é de 85%. Observando-se as cargas fatoriais (loadings) relacionados com cada fator, e sabendo que estas cargas fatoriais representam a correlação entre fator e as séries de tributos, nota-se que o primeiro fator é altamente correlacionado com todos os impostos, explicando a maior parte da variabilidade dos dados. O segundo e o terceiro fatores explicam pouco dos dados mas são fundamentais na previsão dos tributos por meio dos modelos SARIMA.

Foram testados os métodos de rotação ortogonal na análise fatorial neste trabalho, entretanto não foram significativos e não foi adotado. Porque na rotação ortogonal os fatores passam a ser correlacionados entre si, sendo que a grandeza das correlações dependerá do tipo de rotação utilizada, somado a isso, tal opção

altera as suposições do modelo linear original de análise fatorial, o que é inconsistente com toda a formulação usada na estimação das matrizes das cargas fatoriais e das variâncias específicas, além da modelagem dos modelos SARIMA e ARIMA se tornarem mais difíceis.

Dessa forma, opta-se por manter a análise fatorial com m = 3 e sem rotação ortogonal das séries desazonalizadas e padronizadas, assumindo-se que as séries são distribuídas normalmente.

Tabela 3: Análise Fatorial das Séries de Tributos

Método Máxima Verossimilhança - Matriz de Correlação

Sem Rotação

Variavéis	Fator 1	Fator 2	Fator 3	Comunalidade	Variância Específica
COFINS	0,961	-0,234	-0,094	0,987	0,013
CPIS	0,913	-0,213	-0,090	0,887	0,113
IOF	0,914	-0,299	0,178	0,956	0,044
CSLL	0,972	0,158	0,028	0,970	0,030
IRPJ	0,972	0,219	0,006	0,993	0,007
IRPF	0,828	-0,185	0,015	0,720	0,280
IRRF	0,920	-0,185	0,058	0,883	0,117
IPI	0,938	-0,167	-0,027	0,909	0,091
IIMP	0,880	-0,383	0,136	0,940	0,060
Outros Impostos	0,480	-0,076	-0,143	0,257	0,743
Variância	7,899	0,512	0,093	8,503	-
% Var	0,790	0,051	0,009	0,850	-

Fatores

Variáveis	Fator 1	Fator 2	Fator 3
COFINS	0,253	-0,977	-2,511
CPIS	0,025	-0,093	-0,252
IOF	0,072	-0,375	1,428
CSLL	0,092	0,237	0,266
IRPJ	0,476	1,701	0,322
IRPF	0,010	-0,035	0,018
IRRF	0,027	-0,087	0,174
IPI	0,034	-0,097	-0,100
IIMP	0,052	-0,357	0,815
Outros Impostos	0,002	-0,005	-0,064

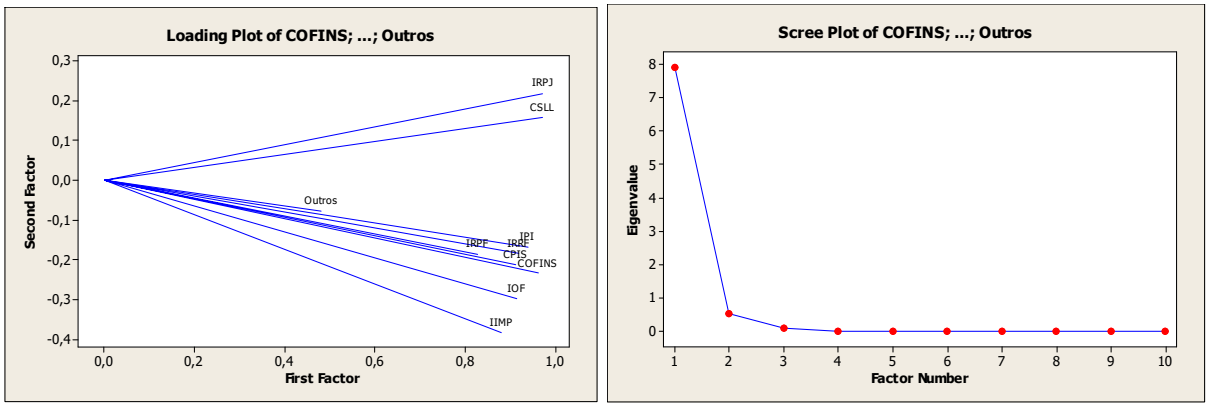


Figura 14: Gráficos da Análise Fatorial das Séries de Impostos

Os escores, para fins do trabalho com modelo SARIMA e ARIMA, serão usados com a nomenclatura de fatores a partir desse momento. Os fatores obtidos por meio da análise fatorial apresentam os seguintes comportamentos:

Fator 1

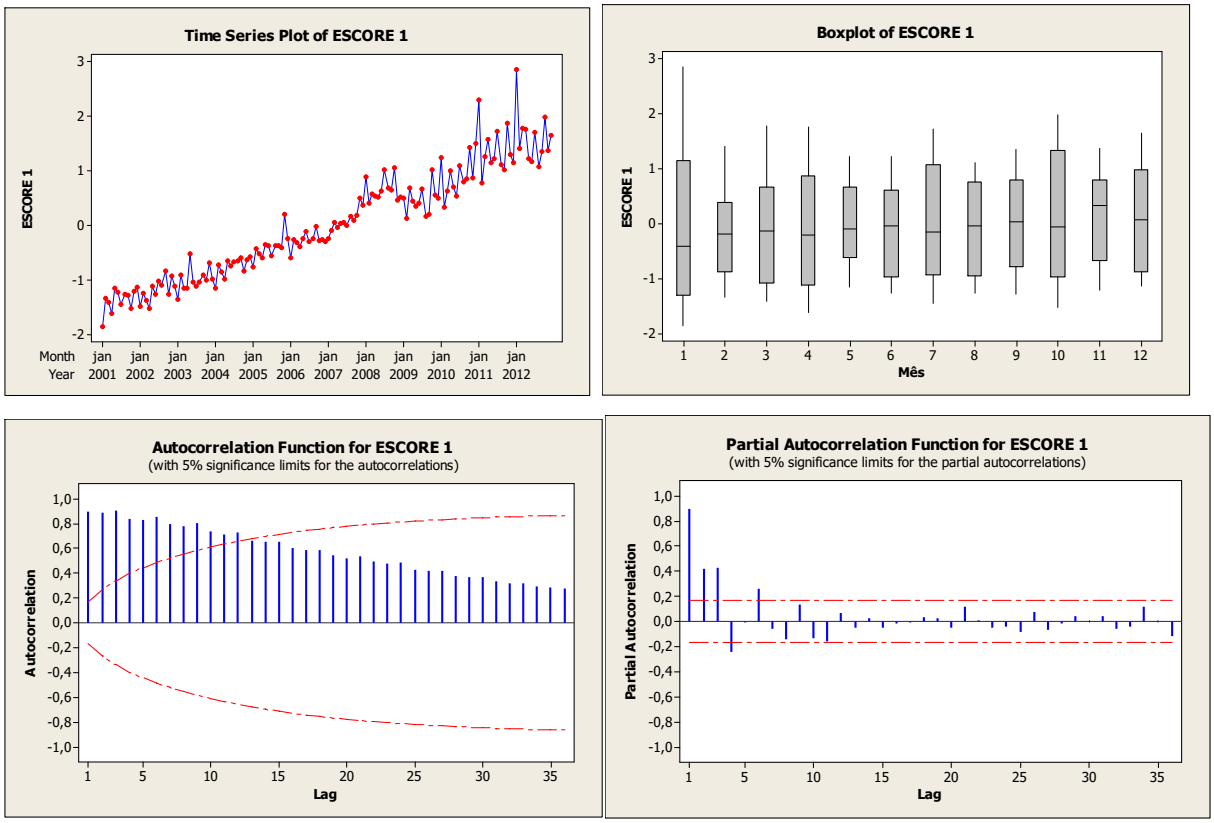


Figura 15: Fator 1 (escore) encontrado na Análise Fatorial

Fator 2

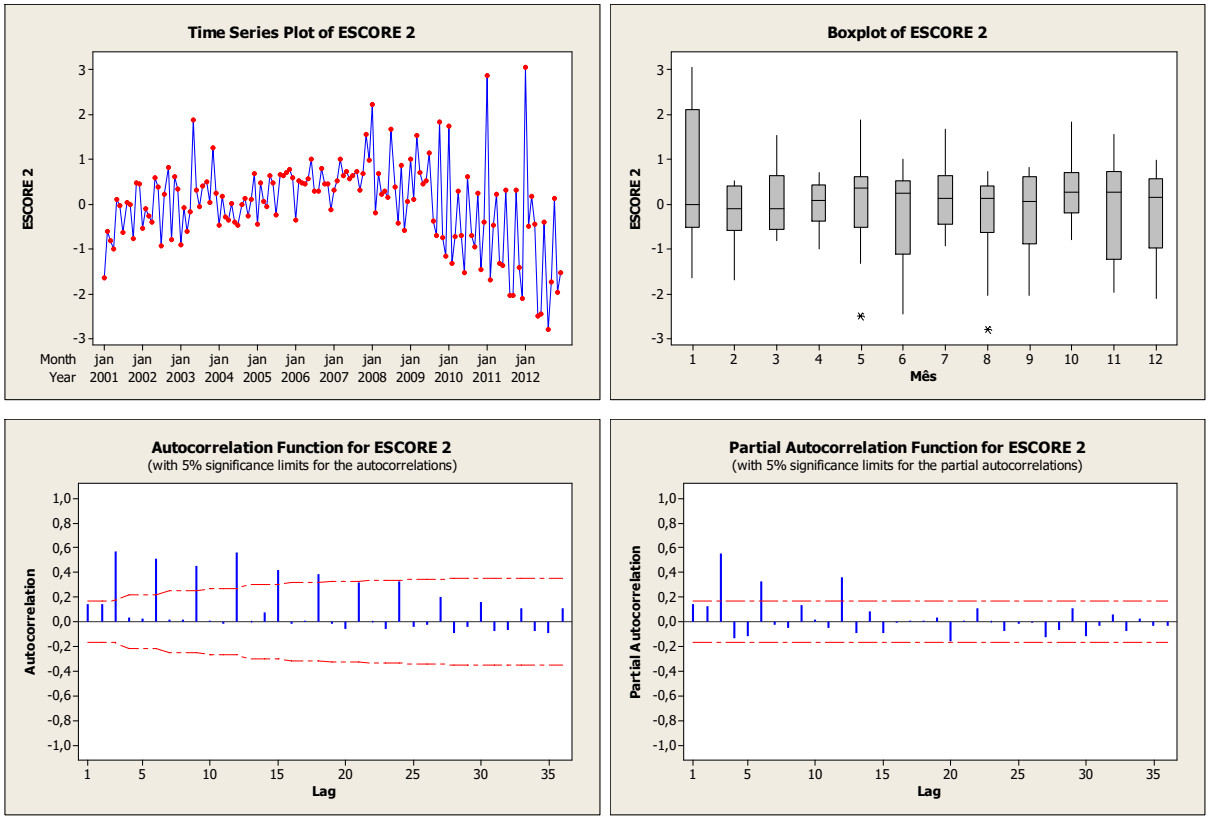


Figura 16: Fator 2 (escore) encontrado na Análise Fatorial

Fator 3

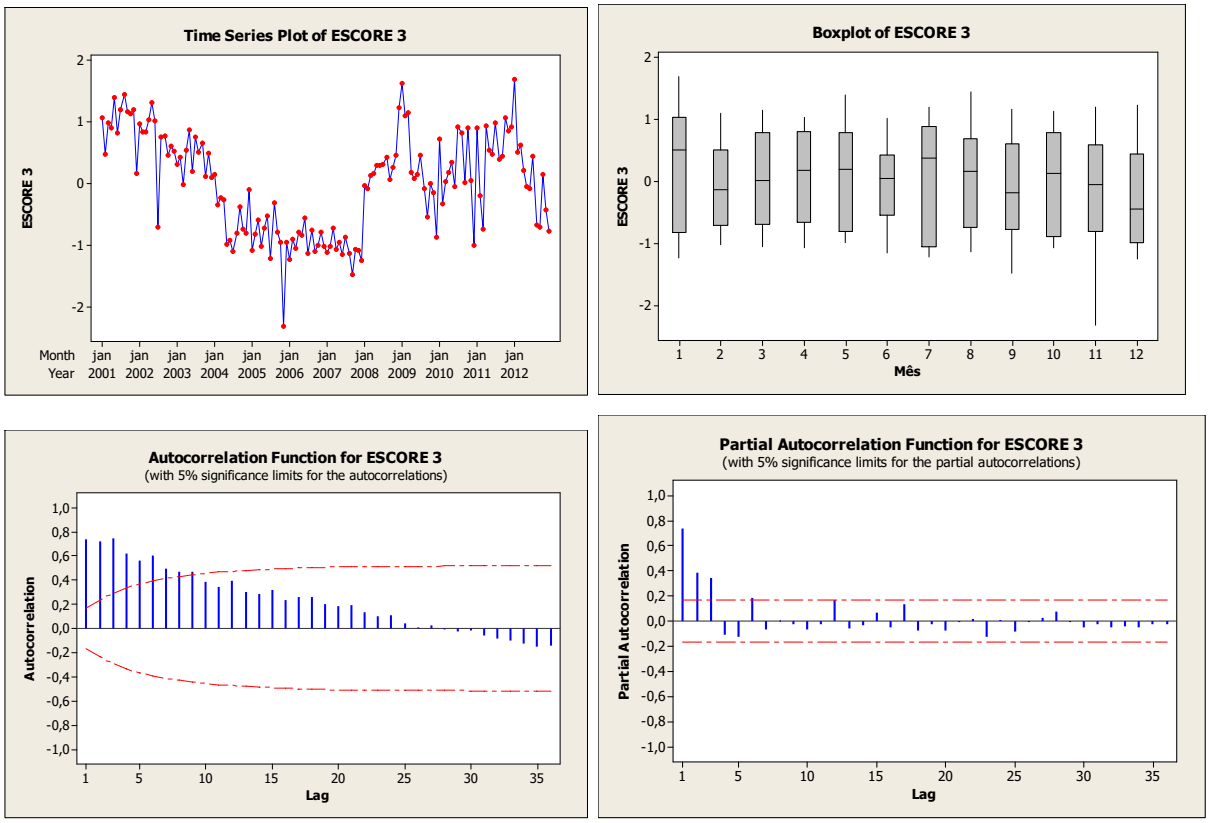


Figura 17: Fator 3 (escore) encontrado na Análise Fatorial

3.5. Ajuste de Modelos SARIMA e ARIMA nos Fatores

Nas séries de fatores trabalhou-se a relação das séries com o passado, se ajustou modelos SARIMA e ARIMA, testando a significância dos coeficientes e verificando as condições de estacionaridade e invertibilidade dos modelos. Os melhores ajustes foram:

FATOR 1 – Modelo SARIMA(2,1,0)(1,1,0)$_{12}$

$$\left(1+0{,}5319B+0{,}4468B^2\right)\left(1+0{,}5436B^{12}\right)\left(1-B\right)\left(1-B^{12}\right)Y_t = a_t$$

$$\begin{aligned} Y_t = {} & 0{,}4681Y_{t-1} - 0{,}9787Y_{t-2} + 0{,}4468Y_{t-3} + 0{,}4564Y_{t-12} + 0{,}2136Y_{t-13} \\ & + 0{,}4469Y_{t-14} - 0{,}2039Y_{t-15} + 0{,}5436Y_{t-24} - 0{,}2545Y_{t-25} - 0{,}5320Y_{t-26} \\ & + 0{,}2429Y_{t-27} - a_t \end{aligned}$$

FATOR 2 - SARIMA(2,1,0)(0,1,1)$_{12}$

$$\left(1+0{,}5829B+0{,}4503B^2\right)\left(1-B\right)\left(1-B^{12}\right)Y_t = \left(1-0{,}3989B^{12}\right)a_t$$

$$\begin{aligned} Y_t = {} & 0{,}4171Y_{t-1} + 0{,}1326Y_{t-2} + 0{,}4503Y_{t-3} + Y_{t-12} + 0{,}4171Y_{t-13} \\ & - 1{,}0332Y_{t-14} - 0{,}4503Y_{t-15} + 0{,}5436Y_{t-24} + a_t - 0{,}3989a_{t-12} \end{aligned}$$

FATOR 3 - ARIMA(2,1,0)

$$\left(1+0{,}6532B+0{,}3960B^2\right)\left(1-B\right)Y_t = a_t$$

$$Y_t = 1{,}6532Y_{t-1} + 0{,}0{,}2572Y_{t-2} + 0{,}3960Y_{t-3} + a_t$$

Na tabela 3, encontra-se um resumo das estatísticas dos modelos e as principais estatísticas de qualidade do ajuste. Os resíduos dos Fatores 1 e 3 não apresentou normalidade, pois a variância se altera a medida que prevê diferentes valores para as séries de impostos mês a mês para o ano de 2013 e isto provavelmente é a causa da não normalidade dos resíduos. Após a análise do ajuste, se pode concluir que os modelos foram capazes de acompanhar razoável bem o comportamento das séries de fatores.

Tabela 4: Resumo das Estatísticas do Ajuste de Modelos nos Fatores (escores)

Estatísticas	Fator 1 SARIMA(2,1,0)(1,1,0)$_{12}$	Fator 2 SARIMA(2,1,0)(0,1,1)$_{12}$	Fator 3 ARIMA(2,1,0)
Estatísticas do Modelo			
ϕ_1	-0,5319	-0,5829	-0,6532
ϕ_2	-0,4468	-0,4503	-0,3960
Φ_1	-0,5436	-	-
Θ_1	-	0,3989	-
SSR	6,44879	62,3036	31,9172
Estatísticas dos Resíduos			
Média	-0,0010	-0,0237	-0,0203
Desvio padrão	0,2227	0,6919	0,4737
Assimetria	0,70	0,09	-0,57
Curtose	2,08	0,80	1,82
Teste de Normalidade (Valor_P)	1,420 (< 0,005)	0,694 (0,069)	0,832 (0,031)

3.6. Previsão das Séries de Impostos

Seguindo o Fluxograma (Figura 3), a previsão das séries dos tributos federais foi realizada por etapas, a previsão de doze passos à frente de cada fator, depois a desagregação dos fatores, por meio da multiplicação dos loadings e escores, assim o resultado obtido foram levados para a despadronização das séries, ou seja, a média e o desvio-padrão foram novamente acrescentados nas variáveis, onde também ocorreu a agregação do fator sazonal novamente às séries.

Para a comparação da acurácia das previsões das séries de impostos, foi separado da amostra inicial de doze observações referentes ao período de janeiro de 2013 a dezembro de 2013.

As previsões das séries de impostos $Y_{T+h}, h \geq 1$, com origem no instante *T* é denotado por, $\hat{Y}_T(h)$, e calculamos os erros de previsão $e_T(h) = Y_{T+h} - \hat{Y}_T(h)$, e assumiremos que a $E(e_t) = 0$, a $Var(e_t) = \sigma_e^2$ para todo *t* e que $E(e_t e_s) = 0, t \neq s$, também iremos a supor que os erros de previsão seguem uma distribuição normal, então podemos determinar um intervalo com 95% de confiança para Y_{T+h},

$$\hat{Y}_t(h) - 1{,}96S(h) \leq Y_{T+h} \leq \hat{Y}_t(h) + 1{,}96S(h)$$

Onde

$S(h) = \sqrt{\dfrac{\sum_{t=m}^{n-h} e_T^2(h)}{T-h+1}}$ é o desvio padrão dos erros de previsão h passos a frente.

As previsões para a série CONFINS para o ano de 2013, com origem em dezembro de 2012, com os respectivos intervalos de confiança, na figura 18, que indica um bom ajustamento do modelo, uma vez que os dados da série CONFINS real está está dentro a estimação intervalar.

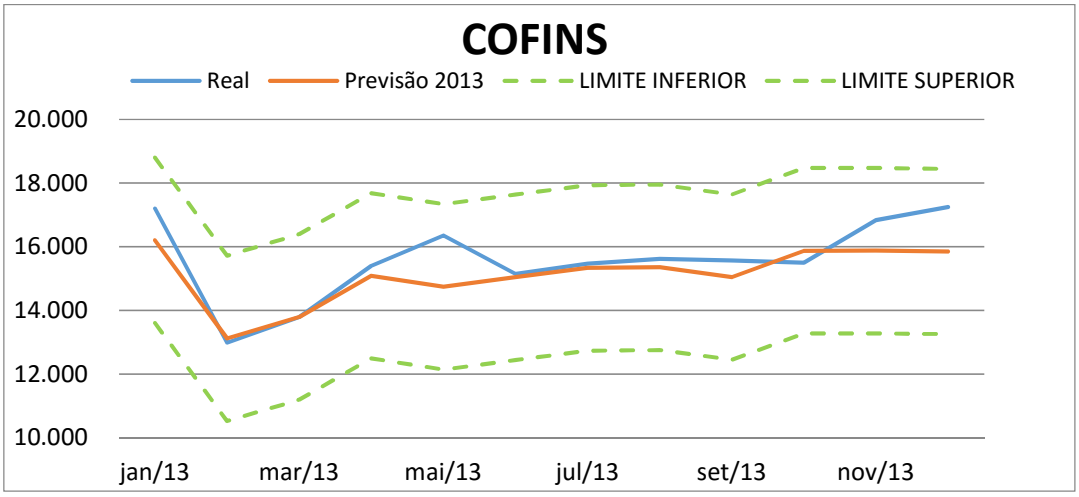


Figura 18: Estimação Intervalar da série de imposto COFINS

No Anexo 4, onde se apresenta a estimação intervalar das outras séries de impostos, onde se observa um bom ajustamento dos modelos, para as séries CPIS, IRRF, IRPJ, CPI e CSLL, uma vez que os dados das séries estão dentro a estimação intervalar, entretanto as séries IOF que apresenta uma quebra estrutural, IRPF mostra um outlier, IIMP que registra uma mudança na tendência difícil de prever e Outros Impostos que apresenta alta variabilidade nos dados, nos mostsra

que o modelo não é adequado, uma vez que as os dados das séries no período de validação estão fora da estimação intervalar.

Confrontaram-se os erros das previsões obtidas por meio do modelo fatorial (MF), com aquelas geradas pelo modelo fatorial dinâmico (MFD), para cada imposto separadamente e verificou-se que o MF traz ganhos consideráveis em termos de eficiência e previsão para a maioria dos impostos. Na comparação dos resultados do MF e MFD, merece destaque o fato de ter obtido performance melhor no MF para COFINS, CPIS, CSLL, IPI, IOF e Outros Impostos, é possível notar que MF foi muito bom pois foi capaz de suavizar e trabalhar os meses com sazonalidade.

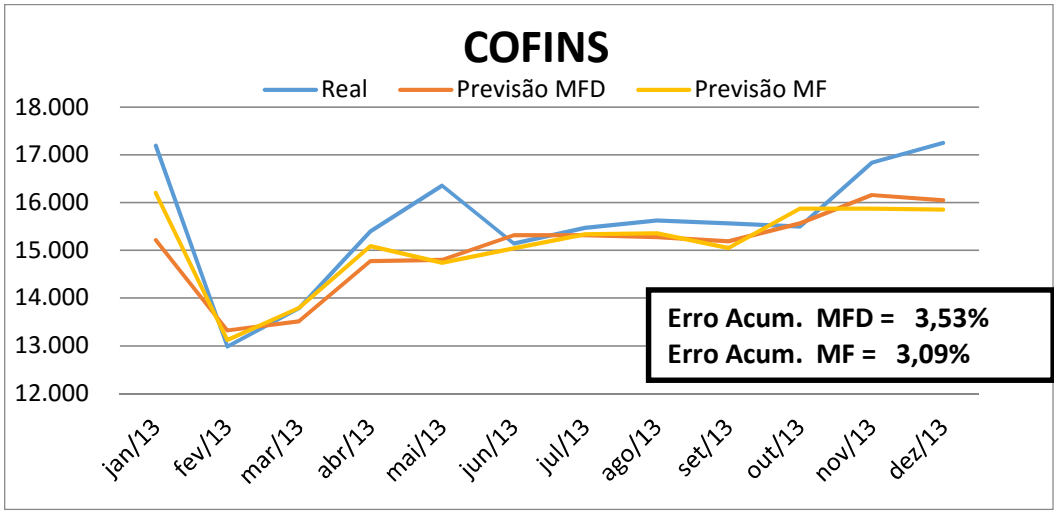


Figura 19: Comparação do MFD e MF da série de imposto COFINS

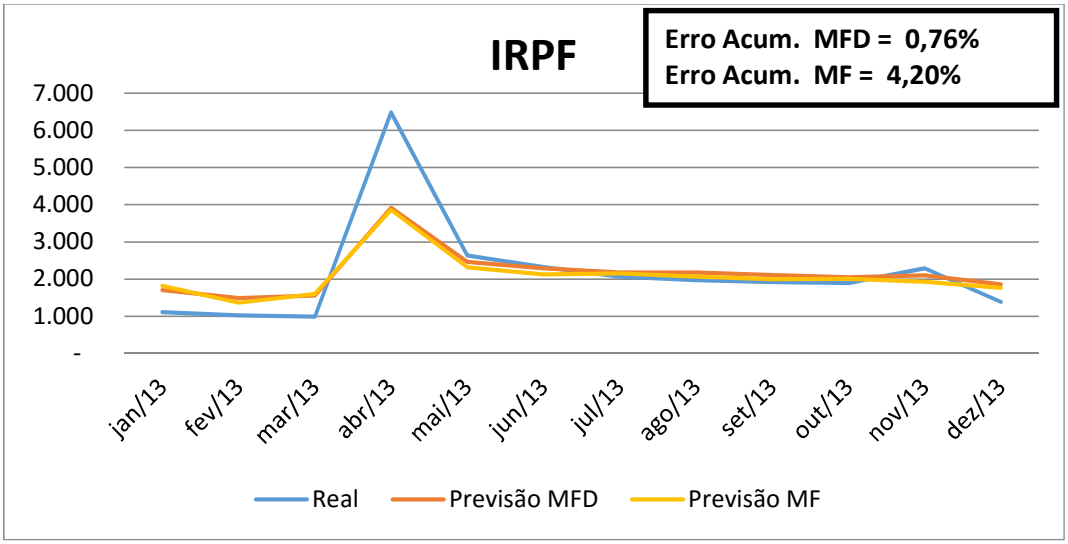


Figura 20: Comparação do MFD e MF da série de imposto IRPF

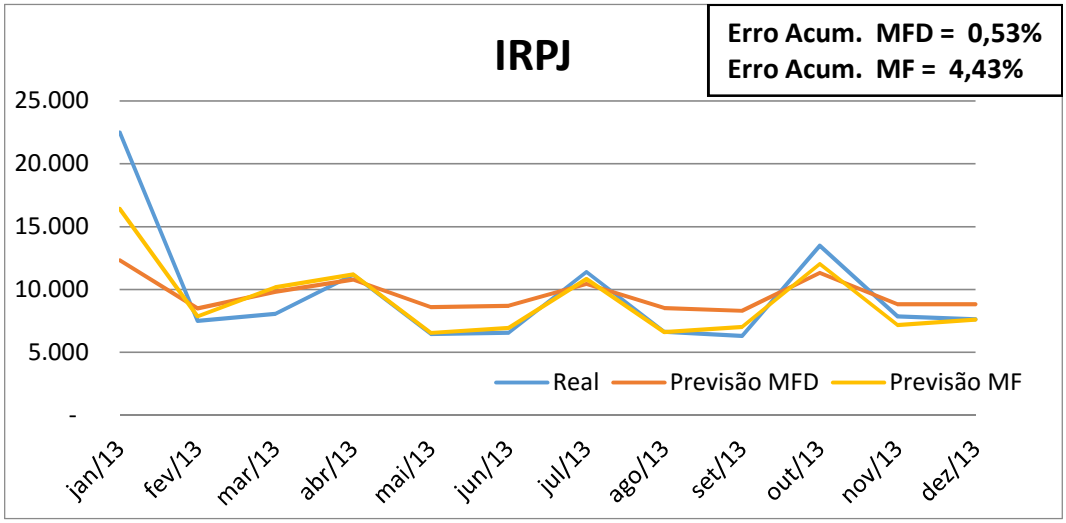


Figura 21: Comparação do MFD e MF da série de imposto IRPJ

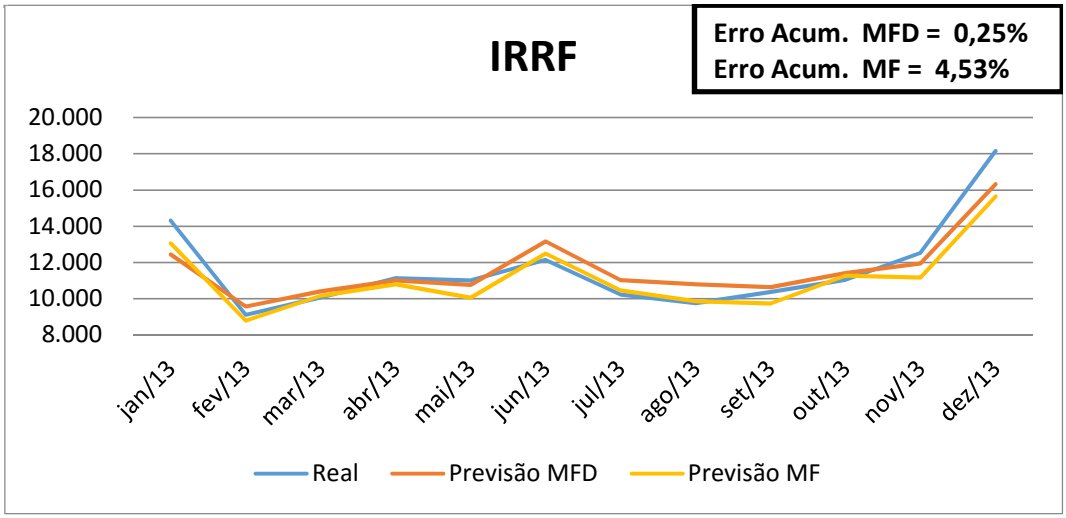


Figura 22: Comparação do MFD e MF da série de imposto IRRF

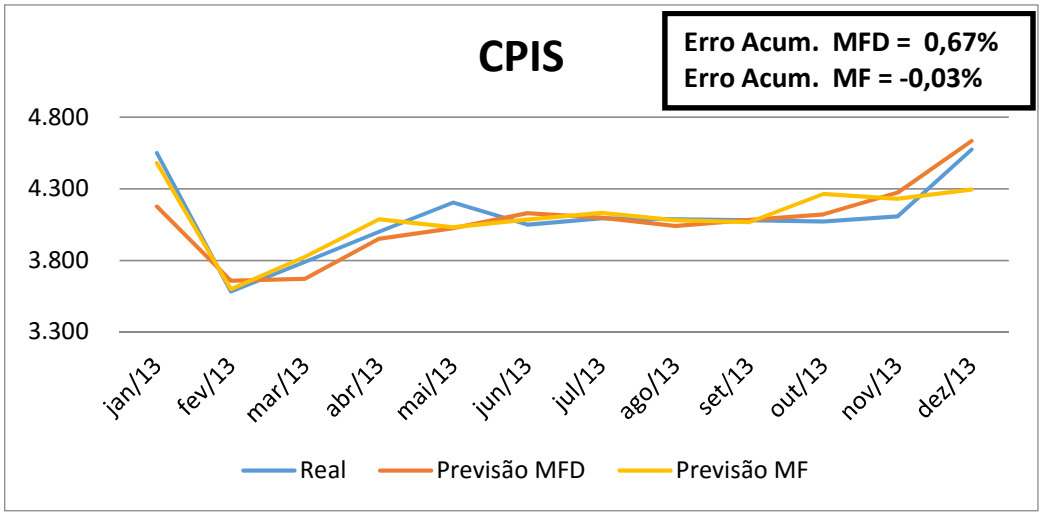


Figura 23: Comparação do MFD e MF da série de imposto CPIS

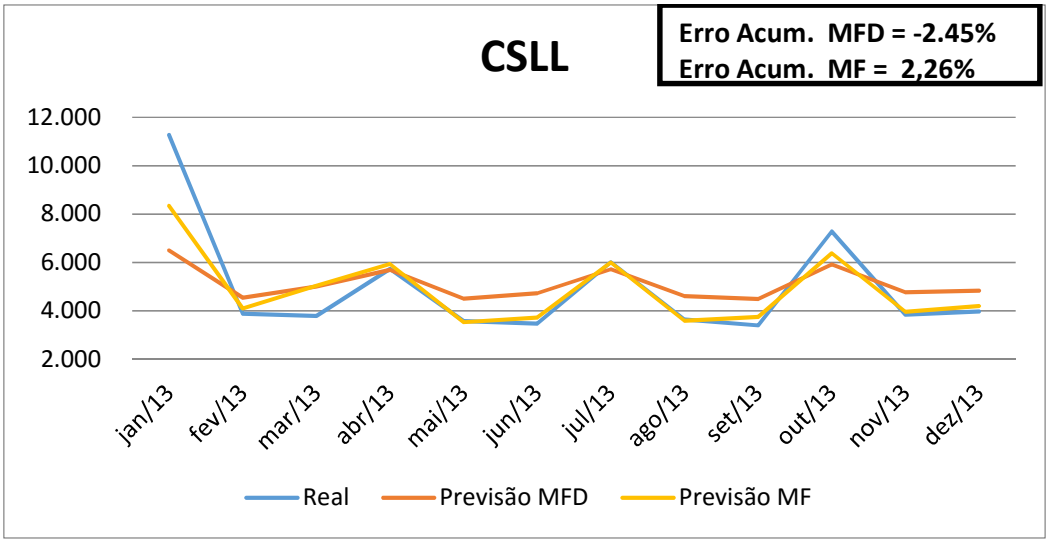


Figura 24: Comparação do MFD e MF da série de imposto CSLL

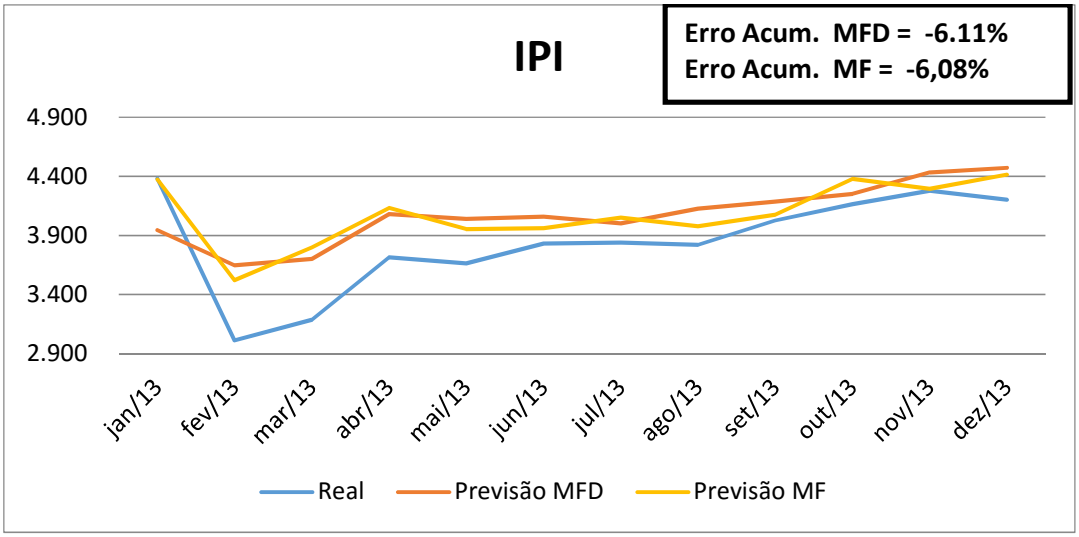


Figura 25: Comparação do MFD e MF da série de imposto IPI

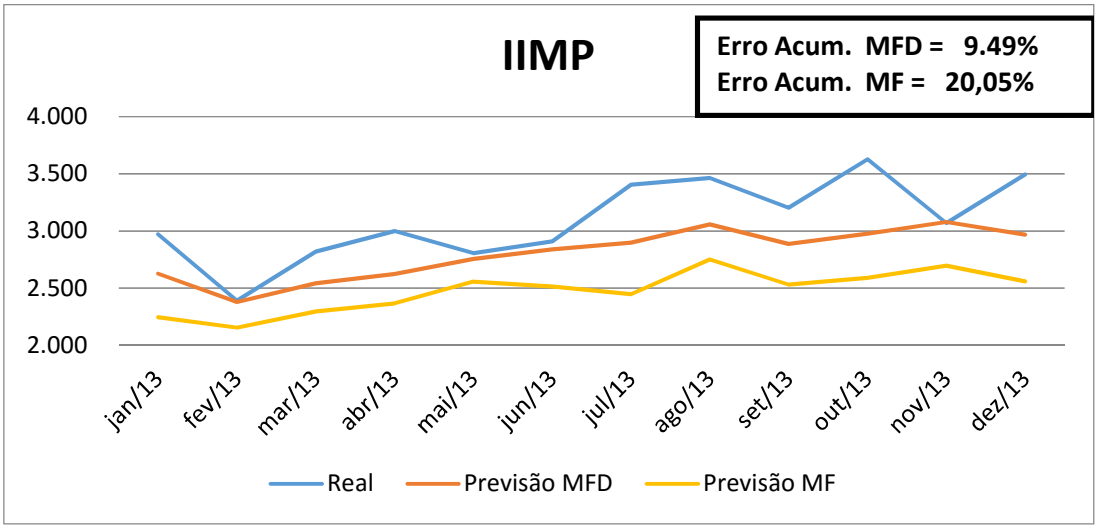


Figura 26: Comparação do MFD e MF da série de imposto Importações

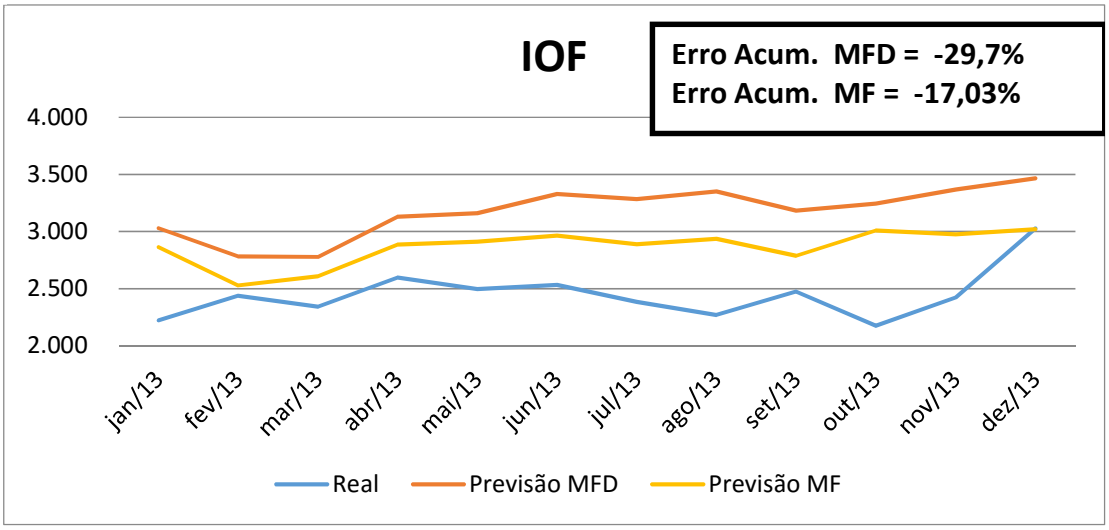


Figura 27: Comparação do MFD e MF da série de imposto IOF

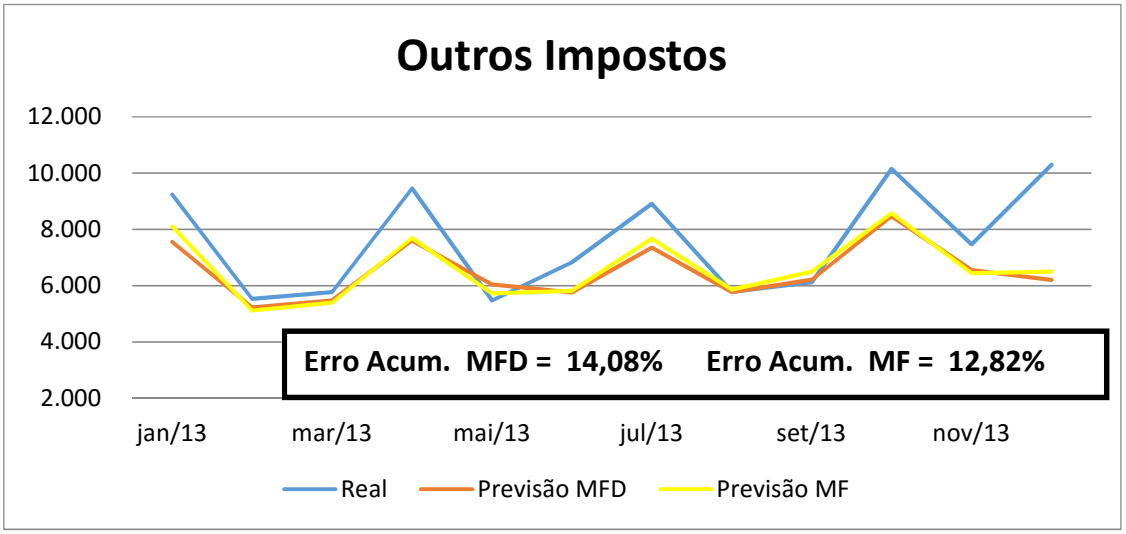


Figura 28: Comparação do MFD e MF da série de imposto Outros Impostos

Foram calculadas diversas estatísticas dos erros de previsão no período de validação, Seja $\epsilon_t = Y_t - F_t$ e p_t o erro de previsão percentual definido como $p_t = 100\left(\frac{\epsilon_t}{Y_t}\right)$. Assim definem-se as seguintes medidas:

Erro Percentual Acumulado (EPA) = $EPA = \frac{Valor\ Realizado\ 2013 - Previsão\ 2013}{Valor\ Realizado\ 2013}$

Raiz do Erro Quadrático Médio (RMSE) RM $\sqrt{\frac{\sum_{t=1}^{n} \epsilon_t^2}{n}}$

Erro Percentual Médio (MPE) $MPE = \frac{\sum_{t=1}^{n} p_t}{n}$

Erro Absoluto Percentual Médio (MAPE) $MAPE = \frac{\sum_{t=1}^{n} |p_t|}{n}$

Erro Absoluto da Média (MAD) $MAD = \frac{\sum_{t=k-1}^{n} |p_t|}{n}$

dentre eles destaca-se o erro acumulado percentual (EPA), que é utilizado pela Secretaria de Política Econômica para acompanhar a arrecadação tributária no Brasil.

Na Tabela 5, se apresenta um resumo comparativo das estatísticas dos erros de previsão das séries de impostos do MF e do MFD, do trabalho de Mendonça e Medrano (2014). Na literatura existem diversas medidas de acurácia propostas para avaliar a performance de previsões, no entanto, o objetivo desta seção não é discutir os pontos positivos e negativos de cada uma delas, mas tão somente apresentar aquelas utilizadas para este fim.

Tabela 5: Resumo das Estatísticas dos Erros de Previsão das Séries dos Impostos

SÉRIES	**EPA**		**RMSE**		**MPE**		**MAPE**		**MAD**	
	MF	**MFD**	**MF**	**MFD**	**MF**	**MFD**	**MF**	**MFD**	**MF**	**MFD**
COFINS	**3,09**	3,53	**R$ 768**	R$ 872	**2.88**	3,29	**3.47**	3,99	**R$ 568**	R$ 647
CPIS	**0,03**	0,67	**R$ 121**	R$ 140	-0,07	**0,59**	**2,11**	2,40	**R$ 89**	R$ 100
IOF	**-17.3**	- 29,70	**R$ 478**	R$ 766	**-17.76**	-30,37	**17.80**	30,37	**R$ 418**	R$ 727
IRPJ	4,43	**0,53**	**R$ 1.938**	R$ 3.323	**0,17**	-9,65	**8,51**	21,85	**R$ 1.043**	R$ 2.232
CSLL	2,26	-2,45	**R$ 970**	R$ 1.654	**-2,21**	-12,47	**9,13**	23,54	**R$ 546**	R$ 1.198
IRPF	4,20	**0,76**	R$ 831	**R$ 810**	**-10,62**	- 14,02	23,43	**23,28**	R$ 491	**R$ 476**
IRRF	4,53	**0,25**	R$ 979	**R$ 940**	3,78	**- 0,84**	**5,35**	6,05	**R$ 701**	R$ 749
IPI	**-6,08**	-6,11	**R$ 298**	R$ 346	**-6,64**	- 6,79	**6,67**	8,46	**R$ 235**	R$ 308
IIMP	20,05	**9,49**	R$ 674	**R$ 362**	19,47	**8,99**	19,47	**9,04**	R$ 621	**R$ 296**
Outros	**12,82**	14,8	**R$ 1.459**	R$ 1.603	**10,69**	11,55	**12,67**	13,55	**R$ 1.089**	R$ 1.180

4. CONCLUSÃO

O presente trabalho teve como objetivo comparar a aplicação do modelo fatorial dinâmico de Mendonça e Medrano (2014) com modelo fatorial para previsão de arrecadação de uma amostra substancial da carga tributária bruta brasileira composta por dez tributos importantes. A base é composta de dados mensais no período de janeiro de 2001 a dezembro de 2012 sendo que a previsão foi feita fora da amostra para o ano de 2013.

Os resultados apresentados nesse trabalho mostram a existência de uma larga variedade nas séries de tributos federais que são bons candidatos ao ajustamento sazonal e padronização. A extração da componente sazonal das séries e sua modelagem endogenamente permite a obtenção de melhores estimativas ajustadas aos dados e predições mais confiáveis, uma vez que a sazonalidade é uma característica inerente a todas as séries de impostos trabalhadas.

Portanto conclui-se que a padronização das séries e despadronização dos fatores, a desazonalização e a agregação do fator sazonal produzem ganhos significativos na estimação de modelos com melhores acurácias. De modo a testar a qualidade da previsão fez-se uso de uma gama de critérios estatísticos que evidenciaram que as previsões do modelo fatorial foram bastante razoáveis. Baseado em estudos preliminares, acredita-se que as previsões fornecidas por essa modelagem são promissoras numa comparação formal com outros modelos geralmente usados para modelar a arrecadação tributária no Brasil.

Por fim, o assunto é realmente vasto, e deve ser visto como um subsídio aos estudiosos de análise fatorial conjugada com séries temporais na formulação e implementação de previsões e ações estratégicas na arrecadação dos tributos.

REFERÊNCIAS BIBLIOGRÁFICAS

BRASIL. Ministério da Fazenda. RFB. Receita Federal do Brasil: MF, 2015. Disponível em <http://www.receita.fazenda.gov.br/> Acesso em: junho de 2015.

BRASIL. Ministério da Previdência Social. MPS, 2015. Disponível em <http://www.previdenciasocial.gov.br/> Acesso em: junho de 2015.

BOX G. E. P.; JENKINS G. M.; REINSEL G. C., Time Series Analysis - Forecasting and Control. 4. ed. Willey. 2008.

FAVERO, L.P.; BEIFIORE, P.; DA SILVA, F.L.; CHAN, B.L.. **Análise de dados: modelagem multivariada para tomada de decisões**. Rio de Janeiro: Elsevier, 2009.

GUJARATI, D. N. Basic Econometrics. 4. ed. New York: McGraw Hill, 2004. 1002 p.

LEVINE, D. M., STEPHAN, D. F., KREHBIELI, T. C., BERENSON, M. L. **Estatística – Teoria e aplicações usando o Microsoft Excel em português**. 6. ed. São Paulo: LTC, 2000.

HARMAN, H. H. **Modern factor analysis**. Chicago, Tha University of Chicago Press, 1968. 474 p.

JOHNSON, R. A., WICHERN, D. W. **Applied multivariate statistical analysis**. Englewood Cliffs, N. J., Prentice-Hall, Inc., 1988. 607 p.

KARSON, M.J. **Multivariate statistical methods**. Ames, Iowa, The Iowa State University Press, 1982. 307 p.

KIM, J.O. Factor analysis. In: NIEH, H.; HULL, C.H.; JENKINS, J.C.; STENBRENNER, K.; BENJ, D.H. (eds.). **Statistical package for social sciences**. New York. Megal Hill, 1975. v 5. p 468-514.

MENDONÇA Mário Jorge e MEDRANO, Luís Alberto. **Aplicação do Modelo Fatorial Dinâmico para Previsão da Receita Tributária no Brasil**. Instituto de Pesquisa Econômica Aplicada. Artigo., 2014.

MENDONÇA, M. J. C. ; MEDRANO, L. ; Sachsida . Um Modelo Econométrico para Previsão de Impostos no Brasil. Economia Aplicada (Impresso), v. 17, p. 295-329, 2013.

MENEZES, A.C.F.; FAISSOL, S.; FERREIRA, M.L. Análise da matriz geográfica: estruturas e inter-relações. In: IBGE. **Tendências atuais na geografia urbano-regional: teorização e quantificação**. Rio de Janeiro, 1978. p. 67-109.

MORETTIN, P. A; TOLOI C. M. C. **Análise de Séries Temporais**. 2. Ed.. São Paulo: Edgard Blücher, 2006. 784p.

MORETTIN, L. G. **Estatística básica**: volume 2 : **inferência**. São Paulo: Makron Books, 2000. 182p.

MINGOTI, S. A. **Análise de dados através de métodos de estatística multivariada**: uma abordagem aplicada. Belo Horizonte: Editora UFMG, 2005. 297p.

SOUZA, A. L. (2005). Análise Fatorial: uma introdução. Artigo. MANEJO FLORESTAL – DEF/UFV. Minas Gerais.

VERGARA, S. C. **Projetos e relatórios de pesquisa em administração**. 5. ed. São Paulo: Atlas, 2004. 96p.

ANEXO 1 – Saída Minitab: Análise Fatorial

Factor Analysis: COFINS; CPIS; IOF; CSLL; IRPJ; IRPF; IRRF; IPI; IIMP; Outros

Maximum Likelihood Factor Analysis of the Correlation Matrix

Unrotated Factor Loadings and Communalities

Variable	Factor1	Factor2	Factor3	Communality
COFINS	0,961	-0,234	-0,094	0,987
CPIS	0,913	-0,213	-0,090	0,887
IOF	0,914	-0,299	0,178	0,956
CSLL	0,972	0,158	0,028	0,970
IRPJ	0,972	0,219	0,006	0,993
IRPF	0,828	-0,185	0,015	0,720
IRRF	0,920	-0,185	0,058	0,883
IPI	0,938	-0,167	-0,027	0,909
IIMP	0,880	-0,383	0,136	0,940
Outros	0,480	-0,076	-0,143	0,257
Variance	7,8987	0,5116	0,0927	8,5030
Var	0,790	0,051	0,009	0,850

Factor Score Coefficients

Variable	Factor1	Factor2	Factor3
COFINS	0,253	-0,977	-2,511
CPIS	0,025	-0,093	-0,252
IOF	0,072	-0,375	1,428
CSLL	0,092	0,237	0,266
IRPJ	0,476	1,701	0,322
IRPF	0,010	-0,035	0,018
IRRF	0,027	-0,087	0,174
IPI	0,034	-0,097	-0,100
IIMP	0,052	-0,357	0,815
Outros	0,002	-0,005	-0,064

ANEXO 2 – Script R para Desagregação das Previsões dos Fatores

```
############################################################
# Monografia – Especialização em Estatística               #
# Script para Multiplicação de Matrizes                    #
############################################################


# Multiplicação dos Loadings e Escores
> setwd("C:/Users/Public/Documents/Pós-Graduação -
Estatística/Monografia/Matrizes para Multiplicação")
> L= read.csv(file="Loadings.csv",head=F,sep=";",dec = ",")
> f= read.csv(file="Escores.csv",head=F,sep=";",dec = ",")
> h=12
> L=data.matrix(L)
> f=data.matrix(f)
> y_prev=L%*%t(f)
> y_prev
> data.matrix(L)
> t(y_prev)



# Matriz de Resíduos
> setwd("C:/Users/Public/Documents/Pós-Graduação -
Estatística/Monografia/Matrizes para Multiplicação")
> L= read.csv(file="Loadings.csv",head=F,sep=";",dec = ",")
> A = read.csv(file="Correlacao.csv",head=F,sep=";",dec = ",")
> h=12
> L=data.matrix(L)
> A=data.matrix(A)
> K=L%*%t(L)
> K
> Residuos=A-K
> Residuos
```

ANEXO 3 – Saída Minitab: Ajuste de Modelos nos Fatores

FATOR 1 – Modelo SARIMA(2,1,0)(1,1,0)$_{12}$

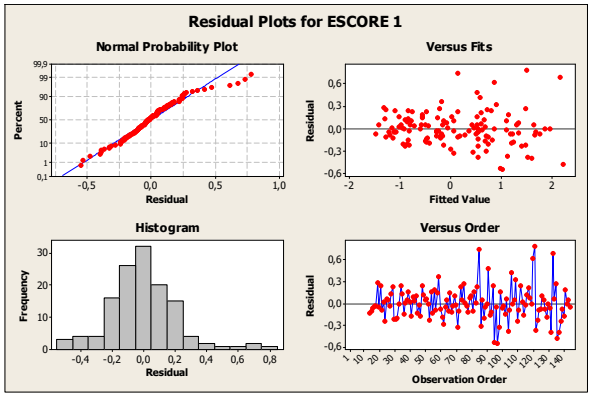
Final Estimates of Parameters

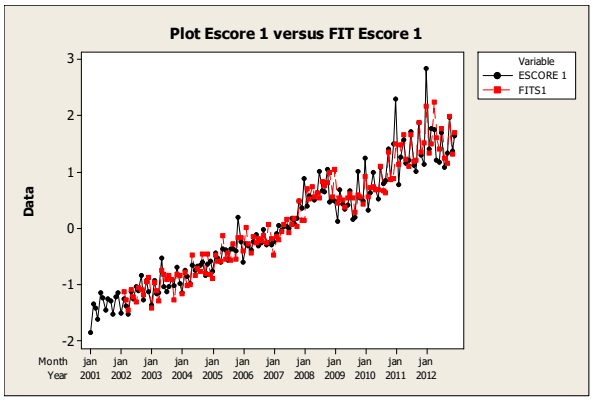
Type		Coef	SE Coef	T	P
AR	1	-0,5319	0,0789	-6,74	0,000
AR	2	-0,4468	0,0788	-5,67	0,000
SAR	12	-0,5436	0,0802	-6,78	0,000

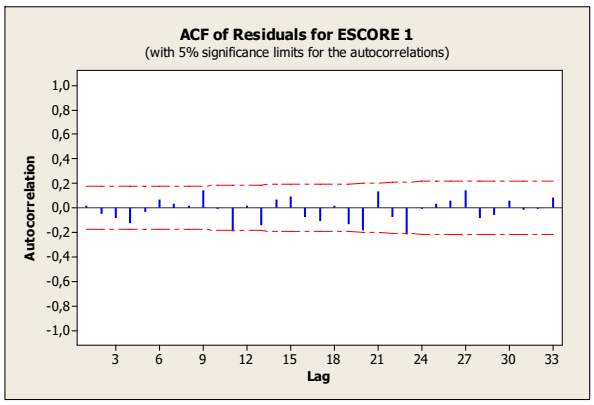
Differencing: 1 regular, 1 seasonal of order 12
Number of observations: Original series 144, after differencing 131
Residuals: SS = 6,44879 (backforecasts excluded)
MS = 0,05038 DF = 128

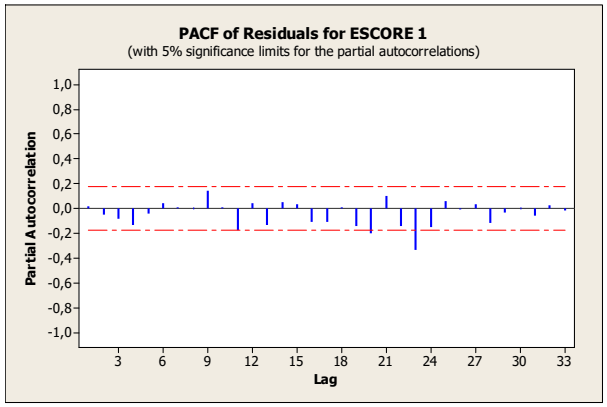
Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	12,8	40,7	53,8	66,4
DF	9	21	33	45
P-Value	0,173	0,006	0,013	0,020









FATOR 2 - SARIMA(2,1,0)(0,1,1)$_{12}$

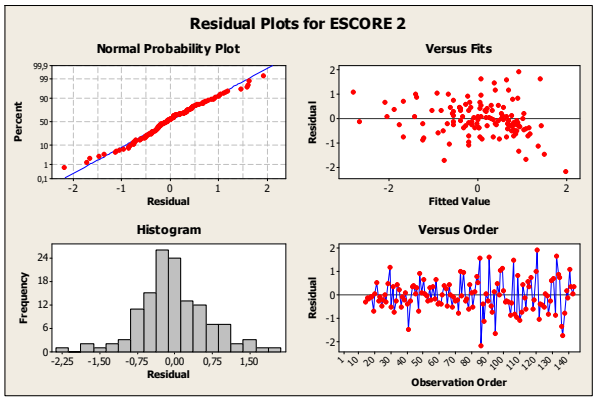
Final Estimates of Parameters

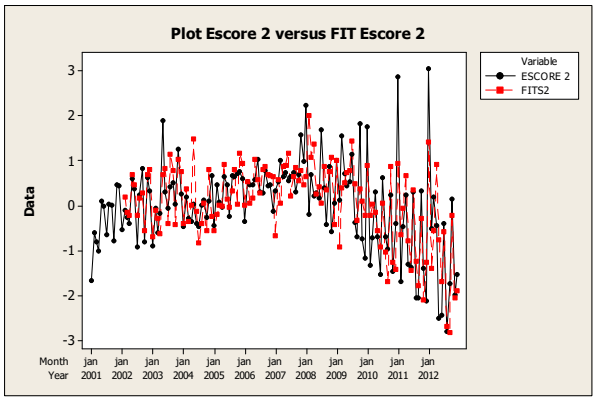
Type		Coef	SE Coef	T	P
AR	1	-0,5829	0,0811	-7,19	0,000
AR	2	-0,4503	0,0810	-5,56	0,000
SMA	12	0,3989	0,0969	4,12	0,000

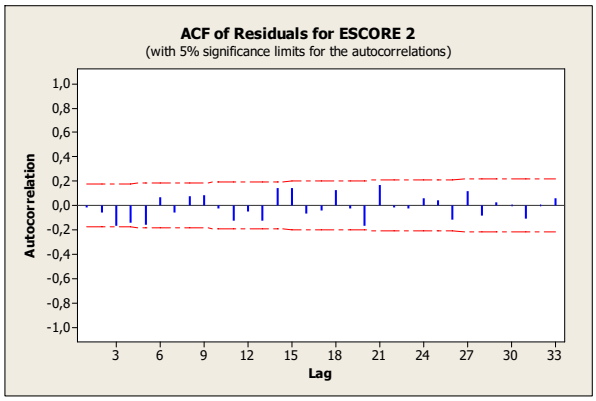
Differencing: 1 regular, 1 seasonal of order 12
Number of observations: Original series 144, after differencing 131
Residuals: SS = 62,3036 (backforecasts excluded)
MS = 0,4867 DF = 128

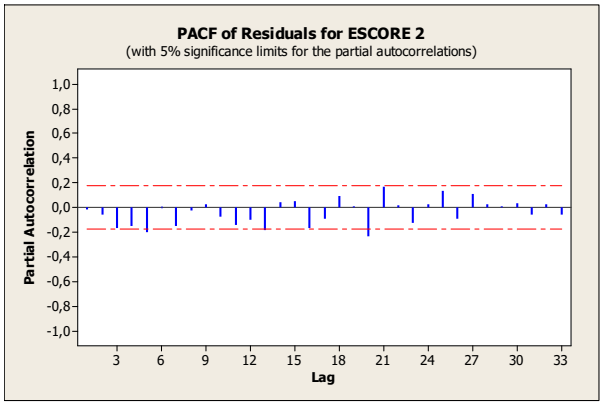
Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	15,9	37,1	51,1	68,8
DF	9	21	33	45
P-Value	0,069	0,016	0,023	0,013









FATOR 3 - SARIMA(1,0,1)(0,1,1)$_{12}$

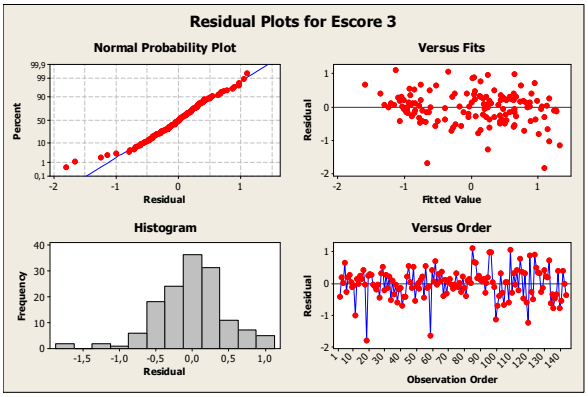
Final Estimates of Parameters

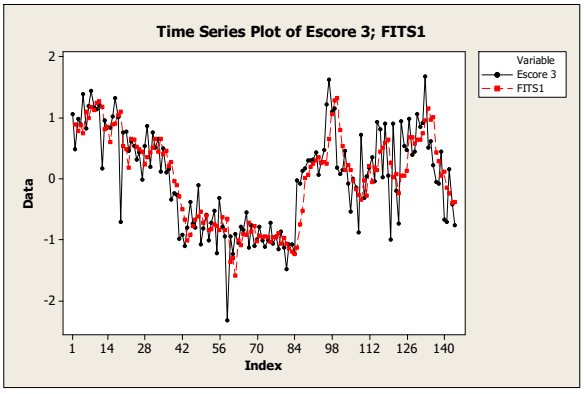
Type		Coef	SE Coef	T	P
AR	1	-0,6532	0,0778	-8,40	0,000
AR	2	-0,3960	0,0781	-5,07	0,000

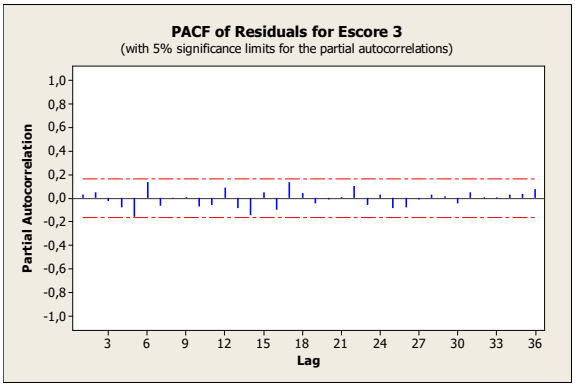
Differencing: 1 regular difference
Number of observations: Original series 144, after differencing 143
Residuals: SS = 31,9172 (backforecasts excluded)
MS = 0,2264 DF = 141

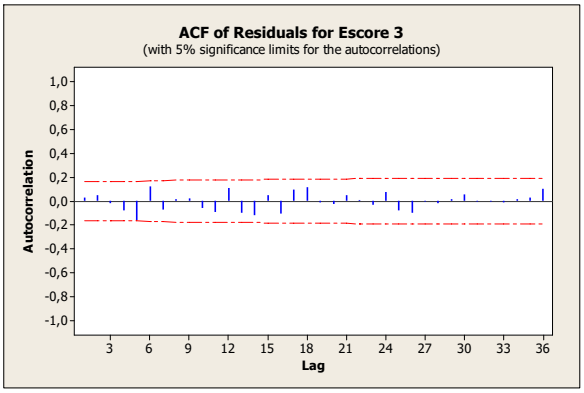
Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	12,0	23,6	29,3	41,6
DF	10	22	34	46
P-Value	0,285	0,369	0,698	0,659



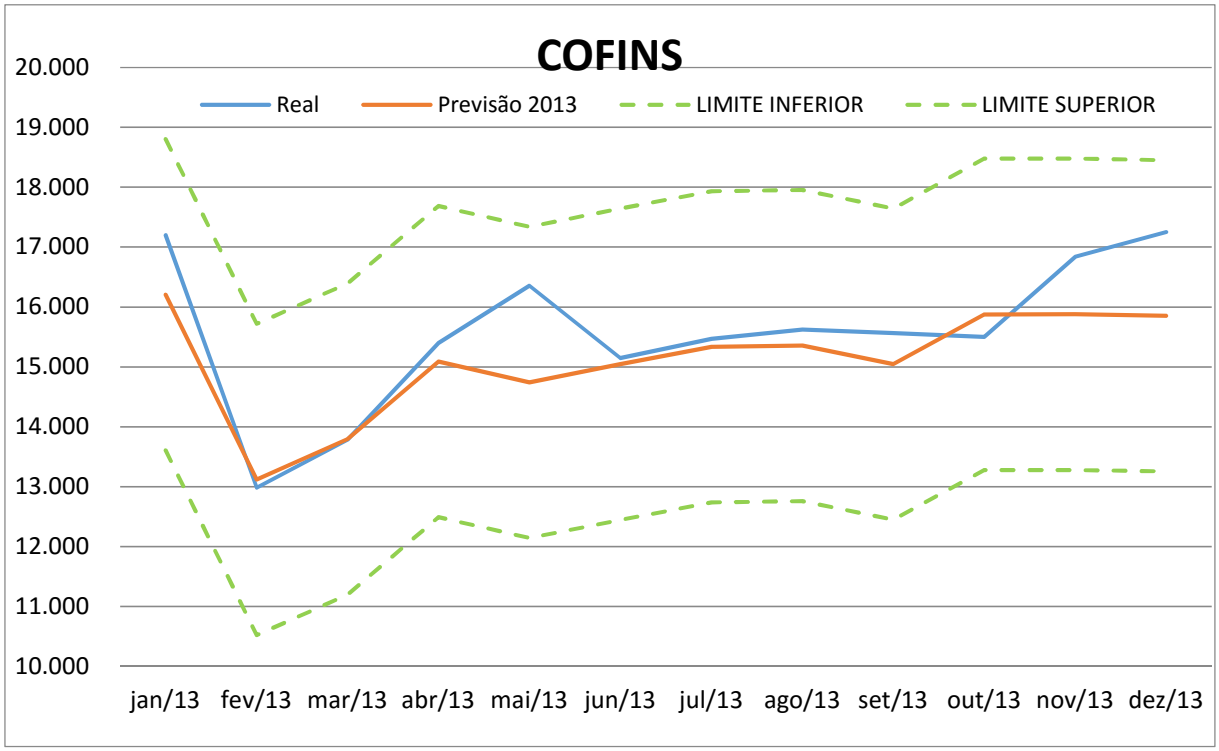




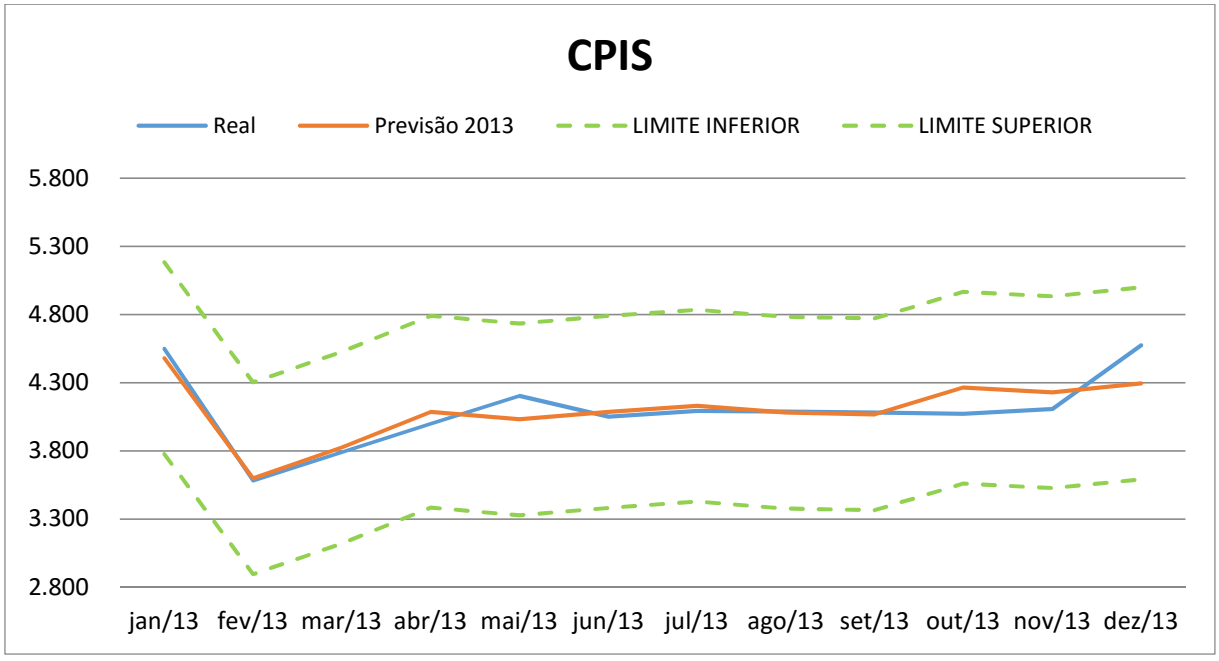


ANEXO 4 – Gráficos das Previsões Intervalares das Séries de Tributos

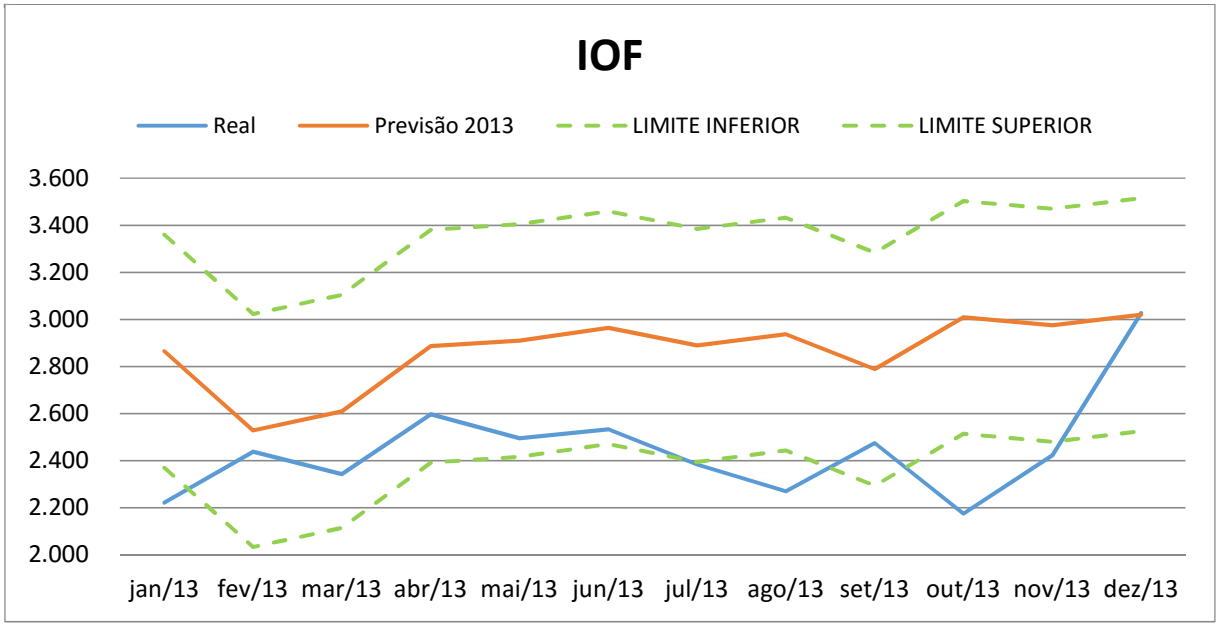
COFINS



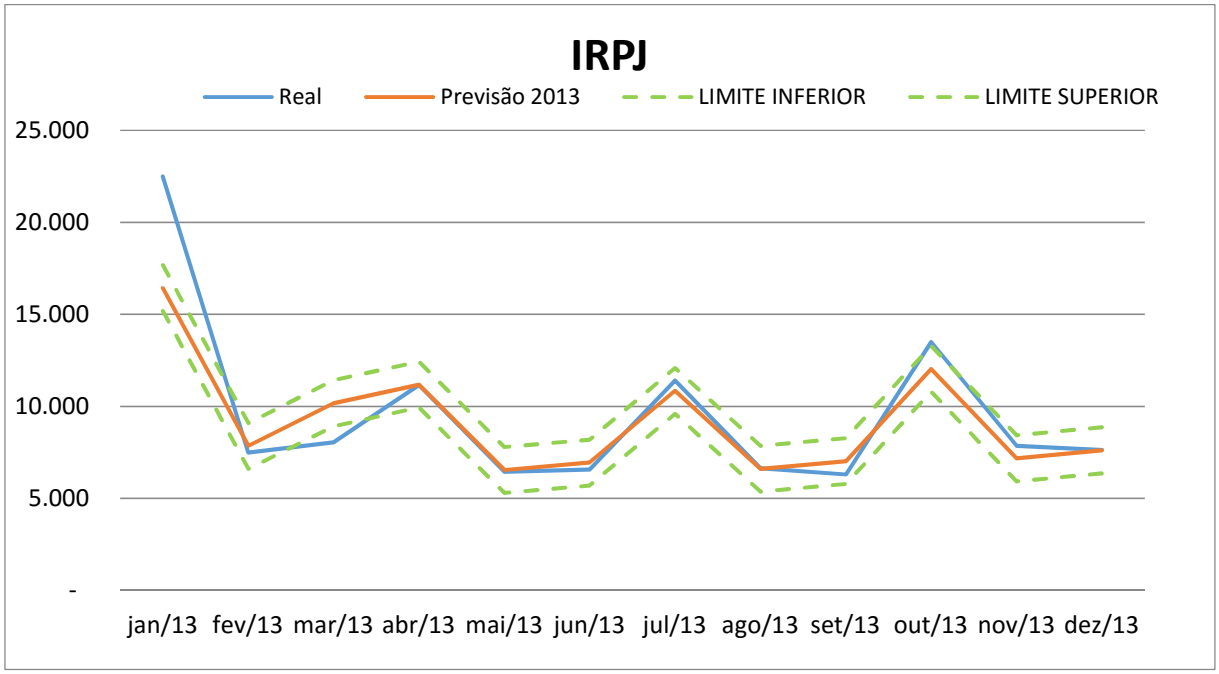
CPIS



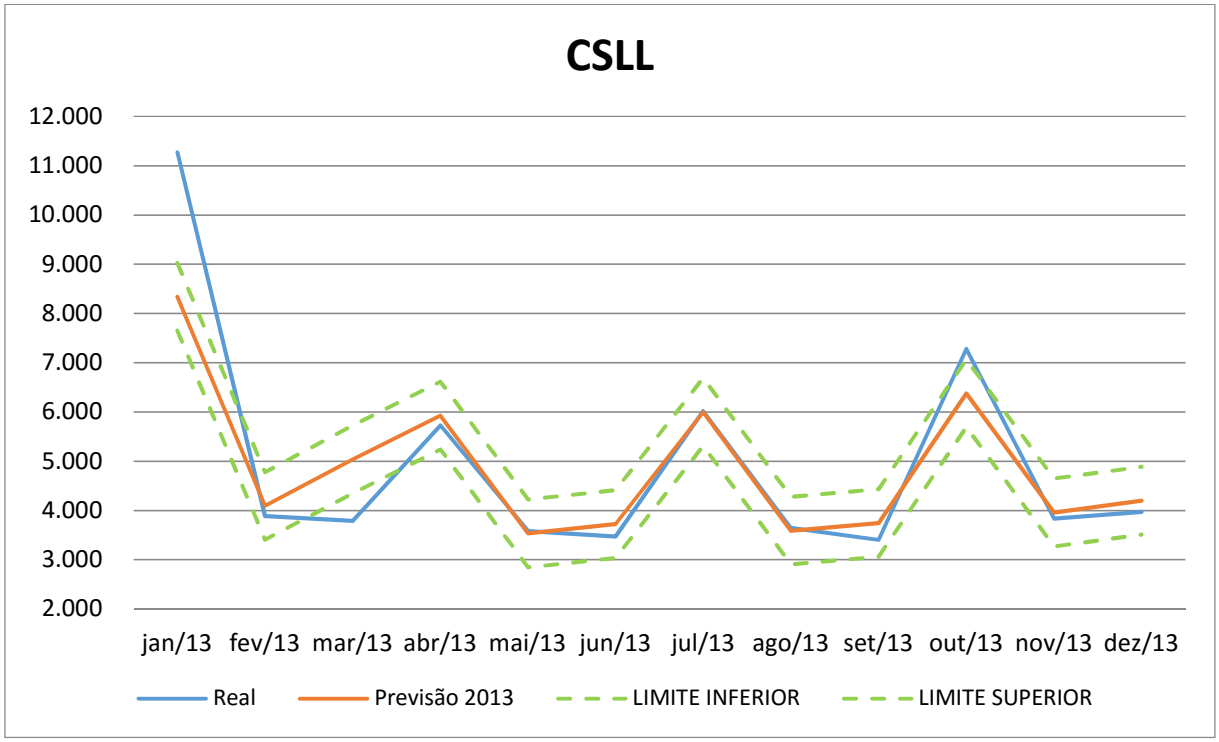
IOF


IOF
Real
Previsão 2013
LIMITE INFERIOR
LIMITE SUPERIOR
3.600
3.400
3.200
3.000
2.800
2.600
2.400
2.200
2.000
jan/13
fev/13
mar/13
abr/13
mai/13
jun/13
jul/13
ago/13
set/13
out/13
nov/13
dez/13

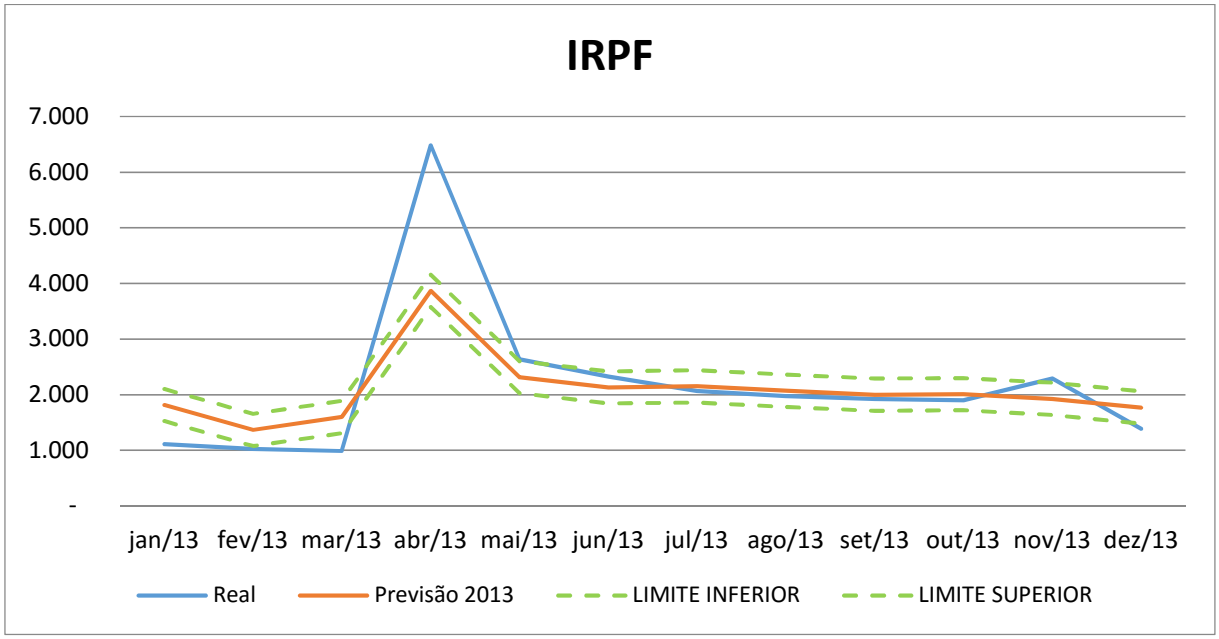
IRPJ


IRPJ
Real
Previsão 2013
LIMITE INFERIOR
LIMITE SUPERIOR
25.000
20.000
15.000
10.000
5.000
-
jan/13
fev/13
mar/13
abr/13
mai/13
jun/13
jul/13
ago/13
set/13
out/13
nov/13
dez/13

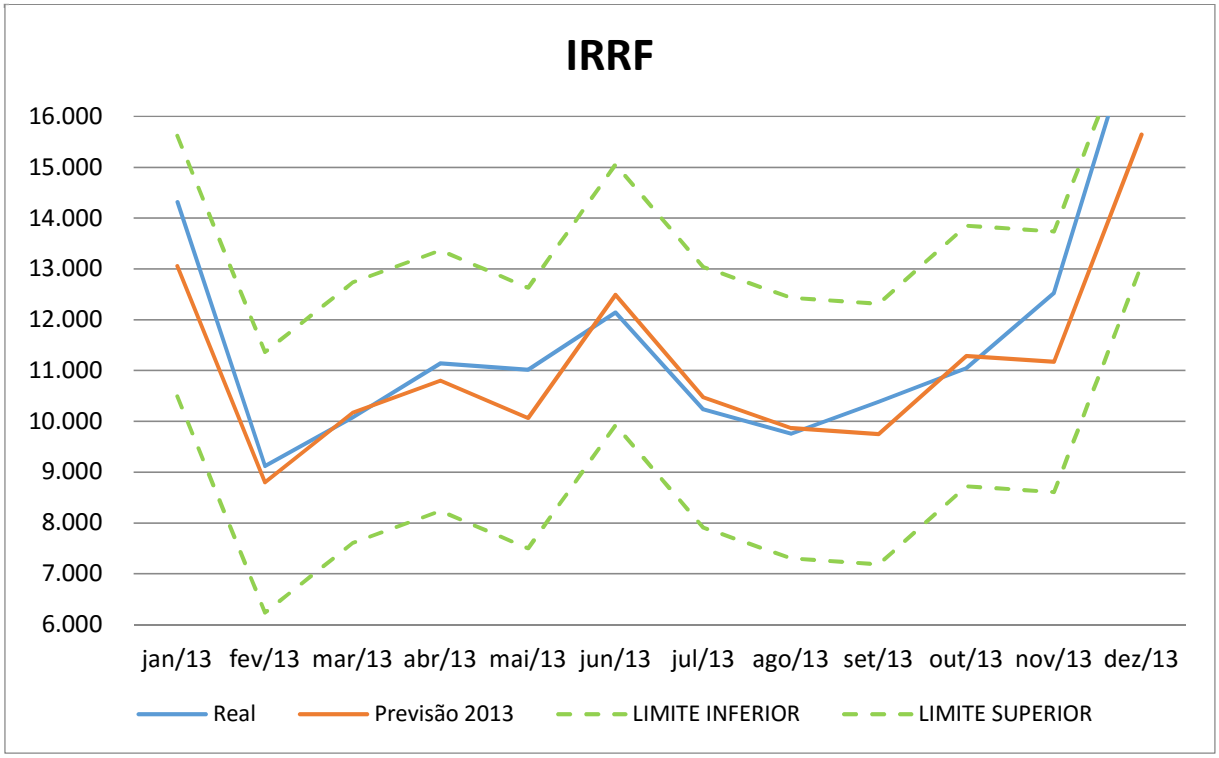
CSLL


CSLL
12.000
11.000
10.000
9.000
8.000
7.000
6.000
5.000
4.000
3.000
2.000
jan/13 fev/13 mar/13 abr/13 mai/13 jun/13 jul/13 ago/13 set/13 out/13 nov/13 dez/13
Real
Previsão 2013
LIMITE INFERIOR
LIMITE SUPERIOR

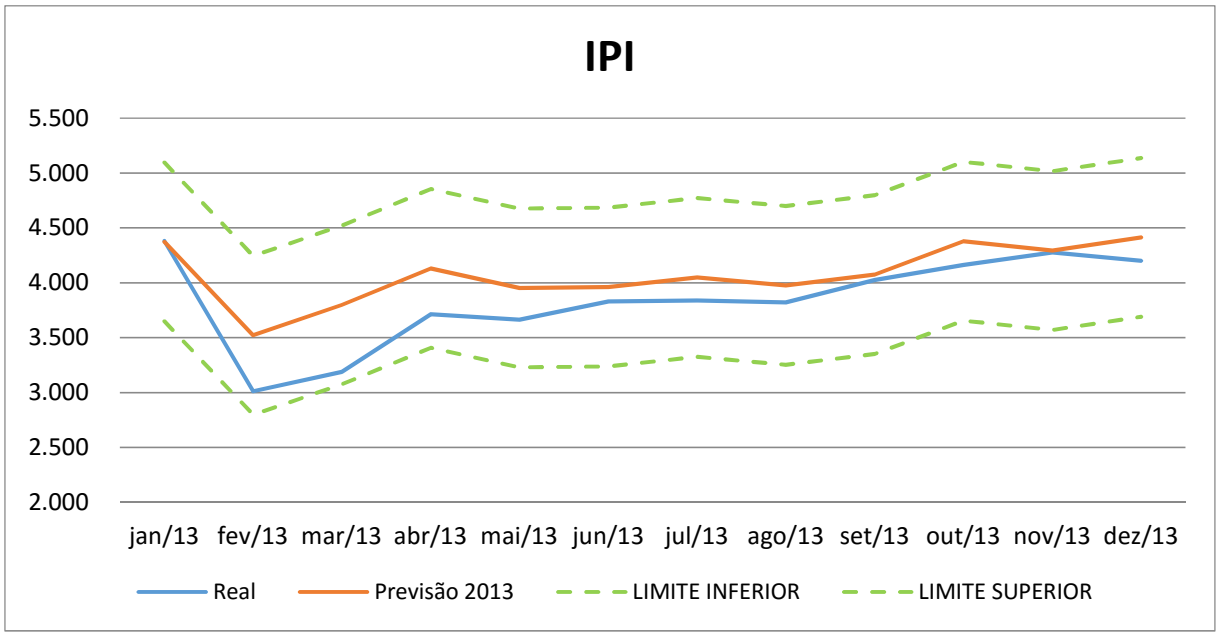
IRPF


IRPF
7.000
6.000
5.000
4.000
3.000
2.000
1.000
-
jan/13 fev/13 mar/13 abr/13 mai/13 jun/13 jul/13 ago/13 set/13 out/13 nov/13 dez/13
Real
Previsão 2013
LIMITE INFERIOR
LIMITE SUPERIOR

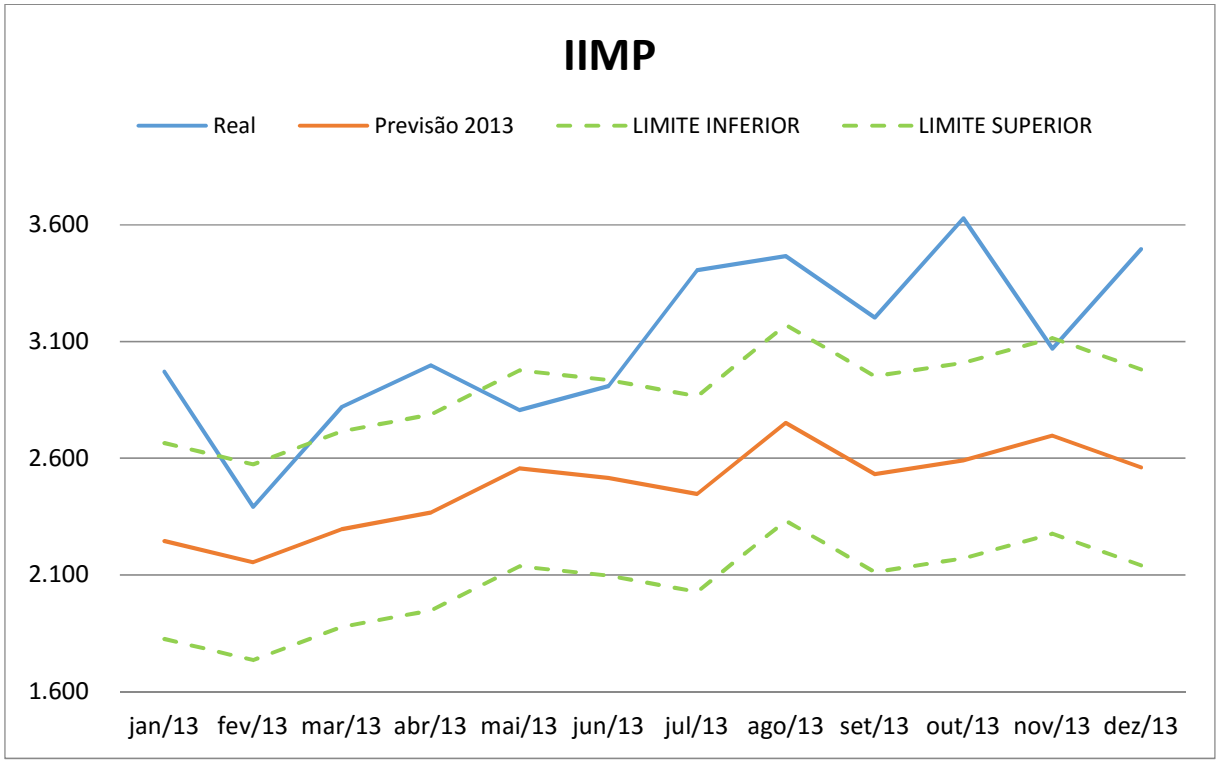
IRRF


IRRF
16.000
15.000
14.000
13.000
12.000
11.000
10.000
9.000
8.000
7.000
6.000
jan/13 fev/13 mar/13 abr/13 mai/13 jun/13 jul/13 ago/13 set/13 out/13 nov/13 dez/13
Real
Previsão 2013
LIMITE INFERIOR
LIMITE SUPERIOR

IPI


IPI
5.500
5.000
4.500
4.000
3.500
3.000
2.500
2.000
jan/13 fev/13 mar/13 abr/13 mai/13 jun/13 jul/13 ago/13 set/13 out/13 nov/13 dez/13
Real
Previsão 2013
LIMITE INFERIOR
LIMITE SUPERIOR

Importações



Outros Impostos