

ISSN 1519-4612

Universidade Federal Fluminense

**TEXTOS PARA DISCUSSÃO**

**UFF/ECONOMIA**

Universidade Federal Fluminense

Faculdade de Economia

Campus do Gragoatá, São Domingos, Bloco F, Niterói/RJ

CEP: 24210-201      Tel.: (0xx21) 2629-9692

<http://economia.uff.br>

Editor: Luiz Fernando Cerqueira; [Icer@uol.com.br](mailto:Icer@uol.com.br); [lf\\_cerqueira@id.uff.br](mailto:lf_cerqueira@id.uff.br).

**Metodologia de Estimação do PIB  
Trimestral Utilizando Procedimentos de  
Cointegração e Filtros de Kalman**

**Luiz Fernando Cerqueira**

**TD 350**

**Janeiro/2021**

**ABSTRACT**

In this paper we present a methodology to estimate the Brazilian GDP's quarterly series from 1960 to 1996. Firstly, the suggested methodology applies Engle–Granger two–step cointegration procedure, involving annual GDP data and some GDP correlated variables. The estimated coefficient of the long–run relationship is used to create a quarterly equation with the GDP and its covariates. From this equation the first quarterly GDP estimate is obtained. The following step consists of improving this estimate with the state space framework, which use the Kalman filter. We estimated univariate models with missing values and benchmarking process, as well as a multivariate SURE with explanatory variables. The estimated model shows appropriate goodness of fit and the diagnostic tests for standardized innovations suggest the assumptions underlying the Gaussian linear model are valid for the data. The paper's major contribution is to present an approach for estimating the quarterly GDP series and developing a methodological framework for its forecasting.

**RESUMO**

Neste artigo apresentamos uma metodologia para estimação da série do PIB trimestral no Brasil para o período 1960–1996. A metodologia parte da estimação da equação estática do procedimento de cointegração de Engle–Granger, envolvendo dados anuais do PIB e variáveis a ele relacionadas. Os coeficientes estimados da regressão são utilizados com os dados trimestrais dessas variáveis para construir uma primeira estimativa da série do PIB trimestral. Em seguida, aprimoramos esses resultados com procedimentos de filtros de Kalman. Estimamos modelos univariados com valores omissos e processos de *benchmarking*, bem como um modelo multivariado SURE com variáveis explicativas. Os modelos apresentam boas performances no que concerne à obediência das hipóteses do modelo gaussiano linear. A principal contribuição do artigo é apresentar uma metodologia consistente e original para a recuperação/estimação e oferecer bases para o desenvolvimento de uma metodologia de previsão/extrapolação do PIB brasileiro trimestral.

**PALAVRAS–CHAVES:** Filtros de Kalman; valores omissos; benchmarking.

**Keywords:** Kalman filters, missing values, benchmarking.

**JEL:** C32, C51, C52.

## METHODOLOGICAL PROCEDURE FOR ESTIMATING BRAZILIAN QUARTERLY GDP SERIES WITH COINTEGRATION AND KALMAN FILTERING APPROACHES

### 1. Introdução

O objetivo deste artigo é descrever e aplicar a metodologia desenvolvida para recuperação e estimação do PIB trimestral no período 1960 a 1996, que utiliza procedimentos de cointegração combinados com métodos de espaço de estado em modelos estruturais de séries de tempo. Dentro deste esquema, estimamos uma relação de longo prazo, com frequência anual, entre o PIB e séries relacionadas. E aplicamos o vetor de cointegração obtido sobre os dados trimestrais dessas variáveis para computar uma primeira estimativa do PIB. A série estimada é encadeada para trás com a série oficial de PIB estimada pelo IBGE a partir de 1980. O passo seguinte foi o de filtrar estes resultados obtidos com modelos estruturais em espaço de estado.

A motivação inicial era estimar a série do índice do PIB trimestral para o período 1960 a 1979<sup>1</sup> e obter uma série completa para todo o período pós-1960, com o intuito de utilizar a série obtida em estudos posteriores, por exemplo, a estimação de uma demanda por moeda para este período ou subperíodos, em que a renda possui um papel central como variável determinante. Entretanto, na *démarche* empreendida no processo de estimação apareceram dificuldades operacionais, relatadas na seção 3.1, que nos conduziram à estimação para um período mais amplo e a uma metodologia “mais complexa”, diferente da inicialmente pretendida, e que se resumia aos procedimentos tradicionais de cointegração. Acabamos chegando, assim, à escolha de modelos estruturais em estado de espaço.

---

<sup>1</sup> A partir de 1992, a série do Índice de PIB trimestral não dessazonalizado passou a ser divulgada periodicamente pelo IBGE, tendo sido computada retroativamente desde 1980. Essa série passou por diversas revisões. A série utilizada neste artigo esteve disponível no *site* do IPEADATA e se referia àquela calculada pelo IBGE com a “metodologia antiga”. Atualmente, ela aparece no *site* encadeada com a série calculada pela metodologia nova. Esta série não está disponível no *site* do IBGE.

Alguns estudos anteriores ao nosso, utilizando enfoques diversos, procuraram sanar esta deficiência das estatísticas oficiais. Em geral, na ausência de estatísticas coletadas com o objetivo de computar o PIB, a montagem de estimativas trimestrais para o produto real agregado pode adotar três critérios: (i) um trabalho de interpolação dos dados anuais; (ii) um levantamento próprio a partir de amostras de bens e serviços; e (iii) uma combinação dos dois primeiros critérios; ver Contador e Santos Filho (1987).

Existem diversas técnicas para interpolação de séries, desde métodos simples baseados em suposições de comportamento linear ou quadrático, até métodos mais elaborados, que lidam mais apropriadamente com valores omissos, a exemplo do ajuste aos dados de funções *spline* cúbicas e filtro de Hodrick-Prescott, que utilizam suavizações exatas de filtros de Kalman. Ou ainda procedimentos de controle ótimo de funcionais de custo quadráticas, que utilizam soluções recursivas baseadas também em filtros de Kalman. Outros métodos utilizam séries disponíveis em bases mensais ou trimestrais, que possuem forte relação com a variável objeto de interesse. No Brasil, a preferência dos pesquisadores recaiu sobre o uso da interpolação com séries relacionadas com o produto. Neste estudo seguimos também este último caminho.

A primeira etapa da nossa metodologia foi obter uma estimativa do PIB trimestral a partir dos coeficientes de uma regressão, rodada com frequência anual, entre o PIB e a produção de veículos automotores, de cimento, consumo industrial de energia elétrica nas regiões Rio e São Paulo, receita tributária real do tesouro nacional<sup>2</sup>; utilizando as séries em índices base 1980. Se as séries são todas integradas de mesma ordem, faz sentido realizar um teste de cointegração e utilizar o vetor de cointegração associado à equação do PIB como relação de longo prazo de interesse entre estas variáveis. Embora, o tamanho da amostra seja um empecilho para esta estratégia,

---

<sup>2</sup> As informações sobre a obtenção e tratamento dos dados utilizados neste artigo estão descritos detalhadamente em Cerqueira (2007).

especialmente para a implementação do procedimento de Johansen, o teste de cointegração de Engle-Granger por ser menos exigente em termos de número de observações é factível<sup>3</sup>. Assim, os resultados obtidos com a equação estática de Engle-Granger constituem uma informação a priori a ser utilizada. O índice do PIB trimestral é então computado com uma equação em que os coeficientes estimados da equação anual são multiplicados pelos índices trimestrais, respectivos, das séries relacionadas.

A natureza restritiva da já convencional metodologia raiz unitária/cointegração não se coaduna com o uso de modelos que admitem uma heterogeneidade temporal mais geral; ver Caporale e Pittis (2002). No período em estudo, as mudanças foram endêmicas. Neste caso, os processos econômicos são modelados mais adequadamente na forma de espaços de estado, que possuem uma maior flexibilidade para tratar uma gama variada de problemas na análise de séries temporais. Nesta abordagem, assume-se que a evolução da série no tempo é determinada por uma série não observada de vetores  $\alpha_t$ 's, com os quais está associada uma série de observações  $y_t$ 's. A relação entre os  $\alpha_t$ 's e os  $y_t$ 's é especificada por um modelo de espaço de estado. O propósito da análise de espaço de estado é inferir as propriedades relevantes dos  $\alpha_t$ 's a partir do conhecimento das observações. Nesses modelos, as observações são constituídas por tendência, sazonais, ciclos, componentes de regressão, mais um termo irregular. Os componentes são geralmente representados por formas de modelos de *random walk* e são modelados previamente antes de serem colocados junto às observações no modelo de espaço de estado. Outra característica dos modelos de espaço de estado é que eles são flexíveis. Devido à natureza recursiva dos modelos e das técnicas computacionais usadas para

---

<sup>3</sup> O que se deve ao fato de que defasagens – nesse caso apenas da variável dependente – são exigidas somente no segundo estágio do procedimento, quando é aplicado um teste de raiz unitária sobre os resíduos da equação estática estimada no primeiro estágio. E com frequência anual, o número de defasagens requerido é pequeno. Embora os seus resultados devam ser tomados com cautela porque é um teste, como o de Johansen, idealizado para grandes amostras.

analisá-los, é imediato permitir mudanças conhecidas na estrutura do sistema ao longo do tempo; ver Durbin e Koopman (2004).

Modelos de espaço de estado são bastante gerais, cobrindo uma ampla variedade de modelos. Observações multivariadas podem ser tratadas como extensões da teoria univariada. Pode-se permitir a presença de valores omissos. Variáveis explicativas podem ser incorporadas ao modelo. Além disso, se necessário, pode-se permitir que os coeficientes da regressão variem estocasticamente ao longo do tempo. Ajustes e variações de calendário também podem ser incluídos. Da mesma forma, observações mensais e trimestrais podem ser ajustadas para terem consistência com observações anuais totais – este processo de harmonização das estimativas trimestrais com o total anual recebe o nome de *benchmarking*, e é muitas vezes utilizado por agências produtoras de dados oficiais.

A etapa seguinte da metodologia consiste, então, em aprimorar a estimativa inicial através de procedimentos de filtros de Kalman. Utilizamos três modelos distintos para decomposição em espaço de estado da série do PIB em seus componentes estruturais. Um modelo univariado com valores omissos e um multivariado SUTSE, em que a obtenção do PIB filtrado por cada modelo, respectivamente, é feita sem adicionar o ruído gerado pelo termo irregular aos demais componentes. E um terceiro modelo, univariado, com *benchmarking*, harmonizando os dados de frequência trimestral com os anuais, através da extração do erro de medida existente nos valores trimestrais. Até o momento da finalização da presente versão, não temos conhecimento de trabalhos semelhantes na literatura econômica brasileira.

O artigo está organizado da seguinte forma. Na seção 2 é feita uma revisão da literatura brasileira sobre estimações do PIB trimestral. Na seção 3 a metodologia é descrita em seus vários aspectos: na subseção 3.1 é descrita a primeira etapa da

metodologia que utiliza o procedimento de cointegração, na 3.2 é feita uma breve resenha sobre o assunto junto com uma rápida apresentação sobre modelos estruturais com *benchmarking*; na 3.4 são apresentados os procedimentos da segunda etapa da metodologia. Na seção 4 são apresentados os resultados dos modelos estimados em espaço de estado. Na seção 5 oferece-se uma conclusão. No apêndice A apresentamos os gráficos das series empregadas. No apêndice B aparecem as matrizes desenhadas para se estimar o modelo de *benchmarking*.

## 2. Revisão da Literatura

Na ausência de estatísticas oficiais coletadas com o objetivo de computar o PIB, a montagem de estimativas trimestrais para o produto real agregado pode adotar três critérios: (i) interpolação dos dados anuais; (ii) levantamento próprio a partir de amostras de bens e serviços; e (iii) combinação dos dois primeiros critérios; cf. Contador e Santos Filho (1987). No Brasil, a preferência dos pesquisadores recaiu sobre o uso da interpolação com séries relacionadas ao produto<sup>4</sup>.

Entre os estudos que utilizam variáveis relacionadas com o PIB como forma de se obter valores interpolados cabe destacar Cardoso (1981), o qual serviu de base para estudos posteriores na mesma linha metodológica, como em Rossi (1988), Pastore (1991), Pinotti (1991) e Cerqueira (1993). Neste artigo o autor obtém estimativas trimestrais para o PIB no período 1966 a 1979 a partir dos coeficientes de uma regressão do PIB anual contra a produção anual de cimento e de veículos automotores, o consumo industrial de energia elétrica no eixo Rio-São Paulo, e a arrecadação tributária real do Tesouro Nacional. No entender de Contador e Santos Filho (1987), o trabalho de Cardoso (1981), possui deficiências no que concerne à excessiva participação da

---

<sup>4</sup> Para uma análise detalhada, ver Contador e Santos Filho (1987).

produção industrial e a imposição de que a sazonalidade do produto real estimado seja a mesma apresentada pela combinação das variáveis. Esses autores, na construção de sua série de PIB trimestral para o período 1973 a 1985, combinam amostragem com interpolação e utilizam séries desagregadas para o produto do setor agropecuário, indústria, comércio e transportes. Contudo, a metodologia empregada é diferente daquela utilizada pelo IBGE. As séries desagregadas usadas se diferem, assim como os pesos associados a cada uma de modo a compor a série agregada<sup>5</sup>.

Ao lado dos problemas apontados, os dados trimestrais obtidos através desta metodologia apresentam alguns problemas dignos de nota. Há descontinuidade dos dados utilizados para os setores agropecuário, de comércio e o produto do governo, assim como a dificuldade de obtê-los. Há também um certo empirismo na escolha das *proxies* das variáveis utilizadas, o que torna o procedimento árduo de se replicar e estendido para períodos maiores e anteriores a 1974. Por essas razões, esse procedimento não será empregado.

Embora à primeira vista as estimativas que utilizam variáveis relacionadas ao PIB possam falhar, por não empregarem um número de variáveis suficientemente representativo de setores produtivos da economia, uma combinação linear dessas variáveis – disponíveis – pode constituir uma boa *proxy* do PIB. Assim sendo é lícito que utilizemos como ponto de partida da nossa metodologia de recuperação e estimação do PIB uma estimativa inicial obtida por essa técnica baseada na estimação do PIB trimestral através de variáveis relacionadas a ele.

---

<sup>5</sup> Nakane (1994) procura corrigir essas deficiências utilizando as mesmas ponderações que o IBGE, de modo a tornar as séries comparáveis. Porém, como sugerido pelo próprio autor, a série estimada apresenta distorções devido às diferenças do padrão de sazonalidade entre ela e a série do IBGE.



### 3. Metodologia

#### 3.1 Primeira Etapa: Recuperação e Estimação do PIB Trimestral através de Técnicas de Cointegração

A idéia inicial era fazer a estimativa do PIB trimestral apenas para o período anterior a 1980<sup>6</sup>. A metodologia de estimação partiria do teste de cointegração entre o PIB e as variáveis a ele relacionadas, com frequência anual, utilizando o procedimento de Johansen. Os coeficientes do termo de correção de erros estimado seriam então aplicados sobre as séries em frequência trimestral, gerando a série do PIB. As variáveis incluídas seriam: PIB, produção de veículos automotores e cimento, consumo industrial de energia elétrica nas regiões Rio e São Paulo, receita tributária real, produção de petróleo e gás e de aço em lingotes. As séries estão expressas em índices base 1980.

No entanto, o pequeno tamanho da amostra (16 observações) se revelou desde logo um entrave. O pequeno número de graus de liberdade impediu a colocação de um número maior de defasagens para a correção da correlação serial nos resíduos, o que invalidava o teste de cointegração, o qual, ademais, requer “grandes amostras”. Além disso, a possível presença de séries  $I(2)$ , no período 1965 a 1980, poderia estar causando o excesso de correlação serial. A solução encontrada para dar continuidade aos experimentos foi reduzir o número de variáveis e aumentar o tamanho da amostra. Assim, alongou-se a amostra até o limite da disponibilidade de dados, a qual passou então a cobrir o período 1960 a 1996<sup>7</sup>. O número de variáveis foi reduzido para cinco. A produção de aço foi retirada por ser intensiva em eletricidade, o que estava acarretando inversão de sinais nos coeficientes do mecanismo de correção de erros (ECM). Além disso, sua presença aumentava a correlação serial dos resíduos. A produção de petróleo,

---

<sup>6</sup> O método de relacionar séries indicadoras à série anual do PIB, o *benchmark*, não representa um processo de *benchmarking* estrito senso. Embora resolva um problema de valores omissos existente nas estatísticas oficiais no período 1960 a 1979. O procedimento descrito nesta seção se inspira em Chow e Lin (1971).

<sup>7</sup> As séries de energia elétrica e a de receita tributária só estão disponíveis a partir de 1960. O truncamento em 1996 deveu-se ao fato de a série do Índice do PIB trimestral metodologia antiga do IBGE ter sido interrompida no terceiro trimestre de 1997. Para não termos que trabalhar com fração de ano a amostra foi truncada no final de 1996.

por sua vez, foi retirada em função do seu crescimento muito acelerado após 1980, o que levou a série a se afastar das demais.

Os estudos realizados com o procedimento de Johansen para o período 1960 a 1996 não apresentaram resultados aceitáveis. Embora os testes de raiz unitária tenham indicado que as cinco séries utilizadas são  $I(1)$ , os resíduos do vetor de correção de erros (VECM) apresentaram elevada correlação serial, que não foi possível corrigir com um polinômio de lags de ordem cinco, número máximo imposto pelo tamanho da amostra ( $T=35$ ). Ademais, o VECM não mostrou boas características de estabilidade, apresentando, mesmo após a restrição sobre o espaço dos vetores de cointegração, raízes muito próximas a **um**. Assim, essa estratégia teve de ser abandonada<sup>8</sup>.

Um procedimento alternativo é o teste de dois estágios de Engle-Granger. Nessa hipótese, a equação estática é usada para fornecer os coeficientes de longo prazo e o teste de raiz unitária aplicado sobre os resíduos da equação indica a existência de uma relação estável de longo prazo entre as variáveis.

Os testes realizados estão reportados na tabela 1. O índice do PIB é a variável dependente. Do lado direito da equação aparece um termo constante (C), uma tendência (TT), e os índices de consumo de energia elétrica (IEES), de produção de veículos (IAUTO), da receita tributária real (IRTNRS) e da produção de cimento (ICIM). Na parte inferior da tabela 1, reportam-se as estatísticas  $R^2$ , DW e SER para a equação estática e, o resultado do teste de raiz unitária dos resíduos junto com os *p-values* dos testes *portmanteau* para a presença de correlação serial e o valor do critério de informação bayesiano.

---

<sup>8</sup> Cabe mencionar que, nos experimentos realizados, algumas variáveis apareceram com sinais negativos, o que se deveu provavelmente à colinearidade existente entre certas séries, como energia elétrica com aço e cimento. Entretanto, as séries de PIB estimadas com os coeficientes do ECM se apresentaram semelhantes às obtidas com a estática de Engle-Granger.

TABELA 1: Teste de Cointegração de Engle-Granger  
Variável Dependente: PIB

Variáveis		Coef.		
C		0.112423		
TT		0.007836		
IEES		0.283306		
IAUTO		0.108377		
IRTNRS		0.098286		
ICIM		0.231170		

$R^2=0.9984$	DW=1.5793	SER=0.0164	$t_{\hat{\alpha}}=-4.6341\ddagger$	lags=0
Q(1)=0.931	Q(4)=0.228	Q(8)=0.479	Q(12)=0.126	SIC=-5.4545

Nota: O símbolo ( $\ddagger$ ) representa rejeição da hipótese nula de uma raiz unitária ao nível de significância de 10%.

Ao nível de 10%, rejeita-se a hipótese nula de não cointegração entre as variáveis. O teste apresenta resíduos sem correlação serial. Como esperado, o  $R^2$  está bem próximo de **um**, caracterizando um ajuste perfeito, eliminando a necessidade de correção para o viés de equações simultâneas ou correlação serial; ver Engle-Granger (1991)<sup>9</sup>. Ademais, a multicolinearidade existente entre as variáveis colaborou no sentido de obter este ajuste. Por outro lado, os estudos sobre os componentes sazonais estruturais através de modelos em espaço de estado e de Holt-Winters indicaram que a série estimada assim como a do IBGE possuem padrões de sazonalidade muito semelhantes, cujas amplitudes crescem progressivamente e, com a série estimada apresentando uma amplitude inferior à da série do IBGE após 1980. O que não tem implicações para o nosso estudo dado que a série que iremos usar assume entre 1980-1996 os valores da própria série do IBGE.

Talvez a representatividade das variáveis utilizadas seja questionável, porém um pouco de reflexão indica que elas representam de forma adequada os principais setores produtivos da economia. De fato, o setor automobilístico é, desde os anos 60, o carro-chefe da economia. O consumo de cimento, por sua vez, representa de forma adequada

<sup>9</sup> Contudo, sempre se pode argumentar sobre a presença de regressores endógenos, que tornariam as estimativas linearmente viesadas e com distribuição limite assimétrica e contendo parâmetros de *nuisance*. A solução dessas deficiências proposta por Philips e Hansen (1990) é utilizar os estimadores *fully modified LS*. Os experimentos realizados com este procedimento e estatisticamente válidos não apresentaram resultados diferentes dos obtidos com as estimativas de EG. Assim sendo, não há porque se preocupar com endogeneidade.

o nível de atividade da construção civil, responsável pela geração de grande parte do emprego na economia. Da mesma forma, o consumo industrial de energia elétrica na principal região econômica do país, é um indicador relevante do nível de atividade da economia. Por fim, é evidente que a receita tributária federal representa adequadamente as atividades do setor governamental, especialmente quando se tem em conta que parte dela é repassada a estados e municípios e ela é inteiramente gasta, pelo menos em grande parte do período de interesse<sup>10</sup>.

Sendo assim, a série obtida como uma combinação linear dessas variáveis, em que os escalares são os coeficientes da equação acima, fornece uma *proxy* inicial bastante razoável para o índice do PIB trimestral. Especialmente que ela recupera os *gaps* existentes nas estatísticas do período 1960 a 1979, onde as observações diretas não estão disponíveis. Na figura 1 aparece a série estimada com os coeficientes da equação de Engle-Granger para o período (YQEG), juntamente com esta série encadeada (para trás) a partir de 1980.1 com a série do IBGE (YQPROXY). É esta última que será filtrada na segunda etapa da metodologia proposta e será denominada de *proxy* do PIB – seja em índice ou em valor.

### 3.2 Modelos Estruturais em Espaço de Estado<sup>11</sup> com *Benchmarking*

Um problema comum em estatísticas oficiais é o ajustamento de observações mensais e trimestrais, obtidas através de levantamentos ou amostragem e, sujeitas a erros, com os dados anuais totais obtidos através de censos ou levantamentos mais detalhados, e que são por hipótese assumidos livres dos erros de amostragem. O total anual é chamado de *benchmark*, e o processo de harmonização das estimativas com o total anual de *benchmarking*. Mais especificamente, *benchmarking*, “calendarização”,

---

<sup>10</sup> Os dados sobre o setor de transportes não foram incluídos por não termos obtido uma série de tamanho adequado.

<sup>11</sup> Para o leitor interessado em modelos de espaço de estado e filtros de Kalman ver Durbin & Koopman (2004) e Harvey (1990).

ou harmonização é um processo que consiste em combinar de forma “ótima” duas fontes de medição de uma mesma série temporal, geralmente obtidas a partir de frequências distintas onde a série de frequência mais baixa, a série de *benchmark*, é assumida como mais confiável estatisticamente. No processo de *benchmarking* procura-se ajustar a série de frequência mais alta à série *benchmark*.

FIGURA 1: Índice do PIB Trimestral



Fonte: IBGE; Ipeadata; cálculos nossos.

Existem duas metodologias principais de aplicar o benchmarking a uma série temporal: um *approach* puramente numérico e outro de modelagem estatística. O *approach* numérico difere do segundo por não especificar um modelo estatístico que se suponha que a série obedeça. O *approach* numérico abrange a família dos métodos baseados na minimização de uma soma quadrática propostos por Deaton (1971)<sup>12</sup> – que segue o princípio da preservação de movimentos –, Bassie (1958), e Ginsburgh (1973). O método de modelagem estatística, por sua vez, envolve os modelos baseados em processos ARIMA propostos por Hillmer e Trabelsi (1987), modelos de espaço de

<sup>12</sup> Para uma aplicação deste procedimento veja Fonzo e Mariani (2003).

estado propostos por Durbin e Quenneville (1997) e, os modelos que usam um conjunto de regressões tais como Cholette e Dagum (1994), Mian e Laniel (1993), assim como, as referências citadas por estes autores<sup>13</sup>. Uma aplicação deste procedimento pode ser encontrada em Di Fonzo e Mariani (2003).

Neste artigo fazemos uso de um procedimento de *benchmarking* o qual utiliza modelos estruturais, onde uma série temporal é decomposta em componentes de tendência, sazonalidade e irregular<sup>14</sup>. A seguir alinhavamos os passos seguidos – em ordem – por Durbin e Koopman (2004) e Durbin e Quenneville (1997).

Define-se um *modelo em espaço de estado* (EE) *linear Gaussiano* pelas seguintes equações:

$$\begin{aligned}
 y_t &= Z_t \alpha_t + \varepsilon_t, & \varepsilon_t &\sim N(0, H_t), \\
 \alpha_{t+1} &= T_t \alpha_t + R_t \eta_t, & \eta_t &\sim N(0, Q_t), \\
 \alpha_1 &\sim N(a_1, P_1).
 \end{aligned}
 \tag{3.1}$$

$y_t$  é um vetor de observações  $p \times 1$ ;  $\alpha_t$  é chamado de vetor de *estado*, é não-observável e de dimensão  $m \times 1$ ;  $\varepsilon_t$  e  $\eta_t$  são termos de erro independentes; as matrizes *do sistema*  $Z_t$ ,  $T_t$ ,  $R_t$ ,  $H_t$  e  $Q_t$ , em geral, contém parâmetros desconhecidos, os quais aglutinam-se em um *vetor de parâmetros*  $\psi$ . A estimação do modelo em (3.1) é efetuada pelo *filtro de Kalman* (FK) (vetor de estado) combinado à *máxima verossimilhança* (vetor de parâmetros). O FK é formado por um conjunto de equações que estimam, recursivamente no tempo, a média e a variância condicional do vetor de estado. Detalhes sobre essas equações, sobre suas deduções respectivas e sobre a combinação

<sup>13</sup> Em adição, Chow e Lin (1971) propuseram um modelo multivariado de mínimos quadrados generalizados para interpolação, distribuição e extrapolação de séries temporais. Embora não sendo um método de benchmarking em um sentido estrito, o método proposto está relacionado ao de modelagem estatística, particularmente ao último grupo.

<sup>14</sup> Caminhando em direção oposta, Moauro e Savio (2005) apresentam um procedimento para desagregação de séries temporais observadas em uma dada frequência para outra de frequência maior. O método sugerido é um modelo multivariado SUTSE estimado com filtros de Kalman.

destas com a estimação de máxima verossimilhança se encontram em Harvey (1990) cap. 3 e Durbin e Koopman (2001) cap. 4 e 7.

Por questões de facilidade de exposição, suponhamos que as observações sejam mensais e que elas sejam obtidas em *surveys*, realizados também mensalmente ( $s = 12$ ). Denotemo-las, por  $y_t$  e os verdadeiros valores que se deseja estimar por  $y_t^*$  para  $t = 12(i - 1) + j$ ,  $i = 1, \dots, l$  e  $j = 1, \dots, 12$ , onde  $l$  é o número de anos. Assim, o erro dos *surveys* é  $e_t = y_t - y_t^*$ , denotado por  $\sigma_t^s \xi_t^s$ , onde  $\sigma_t^s$  é o desvio-padrão do erro de *survey* no tempo  $t$ . O erro é modelado como um processo AR(1) com variância unitária. Em princípio processos ARMA de ordem mais alta podem ser usados. Assumamos que os valores de  $\sigma_t^s$  estejam disponíveis a partir da prática dos *surveys* e que os erros sejam livres de viés.

Os valores de *benchmark* são dados por  $x_t = \sum_{j=1}^{12} y_{12(i-1)+j}^*$  para  $i = 1, \dots, l$ . Suponhamos que os valores anuais estejam disponíveis para todos os anos de estudo; assim, na prática, os valores censitários terão em geral uma defasagem de um ano ou dois em relação às observações dos *surveys*. Como modelo da série de observações, usamos

$$y_t = \mu_t + \gamma_t + \varepsilon_t + \sigma_t^s + \xi_t^s \quad t = 1, \dots, 12l, \quad (3.2)$$

onde  $\mu_t$  é a tendência,  $\gamma_t$  é o termo sazonal.

A série é organizada da seguinte forma

$$y_1, \dots, y_{12}, x_1, y_{13}, \dots, y_{24}, x_2, y_{25}, \dots, y_{12l}, x_l.$$

Consideremos que o instante na série em que o *benchmark*  $x_t$  ocorre é  $t = (12i)'$ ; assim, o instante  $t = (12i)'$  ocorre na série entre  $t = 12i$  e  $t = 12i + 1$ . Tomando o modelo de *random walk* integrado para o componente de tendência e o de *quasi-random walk* para a componente sazonal, temos

$$\Delta^2 \mu_t = \zeta_t, \quad \zeta_t \sim \text{NID}(0, \sigma_\zeta^2),$$

$$\gamma_t = -\sum_{j=1}^{11} \gamma_{t-j} + \omega_t, \quad \omega_t \sim \text{NID}(0, \sigma_\omega^2). \quad (3.3)$$

É conveniente colocar os erros das observações – de medida – dentro do vetor de estado, assim temos

$$\alpha_t = (\mu_t, \dots, \mu_{t-11}, \gamma_t, \dots, \gamma_{t-11}, \varepsilon_t, \dots, \varepsilon_{t-11}, \xi_t^s)'$$

Dessa forma,  $y_t = Z_t \alpha_t$ , onde

$$Z_t = (1, 0, \dots, 0, 1, 0, \dots, 0, 1, 0, \dots, 0, \sigma_t^s), \quad t = 1, \dots, n,$$

e  $x_i = Z_t \alpha_t$  onde

$$Z_t = (1, \dots, 1, 0, \dots, 0, 1, \dots, 1, 0) \quad t = (12i)',$$

para  $i = 1, \dots, l$ . Pode-se então escrever a transição de  $\alpha_t$  para  $\alpha_{t+1}$  entre  $t = 12i - 11$  e  $t = 12i - 1$ . De  $t = 12i$  para  $t = (12i)'$  a transição é a identidade. De  $t = (12i)'$  para  $t = 12i + 1$ , a transição é a mesma que de  $t = 12i - 11$  para  $t = 12i - 1$ .

A matriz de transição  $T_t$  é uma matriz bloco-diagonal com elementos

$$\begin{bmatrix} 2 & -1 & 0_{1,10} \\ \mathbf{I}_{11} & 0_{11,1} \end{bmatrix}, \begin{bmatrix} \mathbf{I}'_{11} & 0_{1,1} \\ \mathbf{I}_{11} & 0_{11,1} \end{bmatrix}, \mathbf{I}_k, \begin{bmatrix} 0_{1,12} \\ \mathbf{I}_{11} & 0_{11,1} \end{bmatrix}, \phi \quad \text{para } t = 1, \dots, n,$$

e é igual a  $\mathbf{I}_{37}$  quando  $t = (12i)'$ ,  $i = 1, \dots, l$ . O vetor de erros de estado é  $v_t = [\xi_t, \omega_t, \varepsilon_t, \kappa_t]$ , se  $t = 12i + 2, \dots, 12(i + 1)$  e,  $v_t = 0$  para  $t = (12i)'$ ,  $i = 1, \dots, l$ . Os demais termos do modelo estão descritos acima.

Assim, emprega-se uma forma modificada do modelo de espaço de estado introduzido no começo da seção, no qual a variância do erro da equação de observações é zero para  $t = 1, \dots, n$  e é igual a variância do *benchmark* para  $t = (12i)'$ ,  $i = 1, \dots, l$ . Supõe-se também que a matriz dos erros de *benchmark* seja uma matriz diagonal



porque, de outra forma, o vetor de estado se tornaria muito grande. A equação de transição é então dada por

$$\alpha_t = T_t \alpha_{t-1} + R_t v_t, \quad t = 1, \dots, 12, (12)', 13, \dots \quad (3.4)$$

onde  $R_t$  é a matriz de seleção formada apropriadamente pelas colunas da matriz identidade  $I_{37}$ .

### 3.3 Segunda Etapa: Procedimentos de Estimação em Espaço de Estado<sup>15</sup>

A partir dos resultados obtidos na seção 3.1 os componentes estruturais do PIB trimestral foram extraídos utilizando-se modelos de espaço de estado univariados e multivariados. Foram estimados três modelos estruturais univariados: um com valores omissos, um com processo de *benchmarking* e um terceiro com *missings* e *benchmarking*. Além deles, estimou-se também um modelo estrutural multivariado, SUTSE, uma generalização do modelo estrutural básico.

As estimações foram feitas para o período 1960 a 1996 com as séries em índices nos modelos univariado com valores omissos e multivariado. Neste último usamos as mesmas variáveis do modelo da seção (3.1). Nos modelos com processo de *benchmarking*, utilizamos as séries em quantidades e o PIB foi estimado em valores reais de 1980<sup>16</sup>. Neste último, a soma dos trimestres de cada ano é diferente do PIB anual computado oficialmente; para as demais séries, o total anual é igual à soma das quantidades/valores trimestrais.

Os experimentos preliminares com modelos estruturais básicos univariados e multivariados em frequência trimestral indicaram a presença de uma forte heterocedasticidade, que, se não tratada, poderia comprometer os resultados estimados.

<sup>15</sup> Uma síntese da metodologia é apresentada em Cerqueira (2007).

<sup>16</sup> A série do PIB trimestral em valores constantes de 1980 é calculada multiplicando-se a série de índice pelo valor do PIB médio trimestral de 1980.

A heterocedasticidade no período 1960 a 1996 tem relação com o comportamento da taxa de crescimento da economia: um longo período de crescimento de 1960 a 1980, seguido de um período de taxas mais de crescimento, oscilantes e sem tendência definida. Como também, guarda relação com o comportamento da taxa de inflação<sup>17</sup>. A partir de 1973, a taxa de inflação entra numa rota ascendente, intercalada por períodos de quedas violentas derivadas dos congelamentos de preços. A partir de meados de 1994, com o Plano Real, ela cai abruptamente e passa a seguir uma rota declinante.

A solução para o problema segue o caminho dos mínimos quadrados ponderados. O procedimento é tratar a heterocedasticidade através da normalização das séries com uma série que seja proporcional aos recíprocos dos desvios-padrão dos erros. Uma série adequada seria então aquela que guardasse relação com o PIB e com a inflação. A receita tributária pela teoria, *coeteris paribus*, deveria guardar uma relação de endogeneidade com o PIB e a inflação. No entanto, no caso da economia brasileira, diversas modificações na tributação foram efetuadas durante o período, o que conduziu a um aumento progressivo e exógeno da receita tributária real. A opção como fator de ponderação recaiu sobre o índice<sup>18</sup> da receita tributária real<sup>19</sup>. Como a série está disponível mensalmente, calculou-se a receita do trimestre e, a partir dela, a série de índice. Com isso, a receita tributária real deixou de ser uma variável isolada no modelo e passou a compor combinações com as demais. No que concerne ao PIB, esta série – em valor ou em índice – dividida pelo índice da receita tributária passará a ser a variável a ser estimada – PIB ponderado. O PIB estimado será recuperado multiplicando-se o PIB ponderado pelo índice da receita.

<sup>17</sup> A taxa de inflação é definida como a variação percentual do IGP-DI da FGV.

<sup>18</sup> A razão do uso da variável em índice é que a receita real expressa em “reais” de 1980 assume valores muito baixos devido às conversões monetárias causadas pela inflação, o que faz os valores divididos pela série explodirem, tornando o procedimento estatístico e computacional pouco manejável.

<sup>19</sup> A taxa de inflação seria uma ótima candidata, não fossem os congelamentos de preços ocorridos durante o período, os quais levaram a taxa de inflação próxima à zero. Assim, a divisão de uma série por ela tem como resultado a geração de *outliers* de difícil tratamento.

Experimentos com a estimação de modelos estruturais básicos com a série ponderada do PIB indicaram uma falha no processo de estimação dos componentes estruturais, decorrente do surgimento de *spikes* antes inexistentes na série. A solução foi estimar um modelo estrutural<sup>20</sup> para a série de receita real e extrair os *outliers* identificados na série de resíduos auxiliares da equação de medida e a quebra observada nos resíduos da equação de estado<sup>21</sup>. Desta forma, para corrigir a heterocedasticidade identificada, será utilizada a série de receita tributária ajustada, devidamente tratada com *dummies* de política e de intervenção. As novas séries ajustadas são plotadas no apêndice, figura A2.

Para efeito das estimações, os parâmetros foram tomados com tendo densidade a priori difusa. Assim, não foram supostos valores a priori da matriz de covariância, embora *priors* para os coeficientes das variáveis explicativas tenham sido utilizados nas estimações feitas com valores omissos.

O primeiro modelo estimado (PIB1) considera a presença de valores omissos na série do PIB trimestral, no período que antecede a 1980. O modelo utilizado foi o modelo estrutural básico com tendência, mas sem inclinação, com *dummies* sazonais estocásticas, termo irregular e variáveis explicativas<sup>22</sup>. A variável do lado esquerdo da equação de medida é a série de índice do PIB ponderado ajustado com valores observados de 1980 a 1996, e com valores omissos de 1960 a 1979, os quais desejamos estimar. Se o período anterior a 1980 for inteiramente coberto com valores omissos, o pacote não realiza adequadamente as recursões. Então, da série estimada de PIB selecionou-se aleatoriamente 50% da amostra para assumir valores omissos, para os

---

<sup>20</sup> O modelo utilizado inclui componente estrutural de tendência sem inclinação, ciclo de curta duração, *dummies* sazonais trigonométricas e termo irregular.

<sup>21</sup> O critério para extração foi o resíduo normalizado assumir um valor maior do que 3.5. Foram feitas intervenções nos trimestres 1980.3, 1986.1, 1986.4, 1989.3, 1990.2, 1991.4, 1992.1 no termo irregular e uma intervenção no nível da série em 1989.4.

<sup>22</sup> O modelo foi estimado em *Eviews* 5.1, que usa o procedimento descrito em Koopman, Shephard e Doornik (1999) para estimações com *missings*. Vale a pena ter em conta que o pacote funciona “mais” apropriadamente quando são fornecidos *priors* dos valores iniciais, matriz de covariância e valores dos coeficientes do sistema.

trimestres anteriores a 1980<sup>23</sup>. Por construção, estes valores omissos são substituídos, através do algoritmo de suavização, por valores interpolados, consistentes com a estrutura de componentes do modelo. As variáveis explicativas – veículos, cimento e energia elétrica – devidamente ajustadas entram no lado direito da equação de medida, apropriadamente estacionarizadas<sup>24</sup>, conforme recomendado por DK (2004). O PIB reportado não inclui o componente irregular, por ser este uma estimativa do distúrbio aleatório. Assim, sem esse termo, o PIB estimado fornece uma informação sobre a capacidade dos componentes de tendência e sazonalidade estocástica explicar o seu comportamento.

O segundo modelo univariado é o de *benchmarking*, que torna o PIB trimestral “consistente” com o PIB anual calculado pela IBGE, ao corrigir os erros de medida. Supõe-se que o erro de medida siga um processo AR(1). O modelo foi estimado com o PIB em valores ponderado pelo índice médio do ano da receita tributária<sup>25</sup>. O modelo de estimação foi o gaussiano linear geral<sup>26</sup>, com tendência local, *dummies* sazonais e sem variáveis explicativas. O modelo obedece à estrutura do modelo com processo *benchmarking* tal como descrito na seção 3.2. As matrizes  $Z_t$ ,  $T_t$ , e  $R_t$  construídas para operacionalizar as estimações com *benchmark* estão apresentadas no apêndice B. O primeiro grupo de matrizes refere-se à estimação do modelo univariado reportado na seção 4. Essas matrizes escritas para os dados trimestrais e totais anuais também

---

<sup>23</sup> A seleção foi realizada utilizando 61 pseudonúmeros aleatórios extraídos de uma normal (0,1). Desses números, as duas primeiras casas decimais em módulo eram usadas para escolher uma posição entre 1 e 80, à qual se atribuía um *missing*. Foram necessários 61 números para conseguir selecionar 40, já que alguns eram repetidos e outros superiores a 80. Uma *random seed* foi utilizada.

<sup>24</sup> O teste clássico para a presença de raízes unitárias sazonais, HEGY, em geral rejeita a hipótese nula e, tal como apontado em Maddala e Kim (2002), apresenta várias deficiências. Por esse teste, concluiríamos que não existem raízes unitárias sazonais em nenhuma das variáveis explicativas utilizadas. A solução utilizada para detectar e tratar um componente sazonal estocástico foi pesquisar o comportamento da série através de modelos de espaço de estado. Coincidentemente, o mesmo modelo foi ajustado para todas as séries. Assim, foi estimado um modelo com tendência linear local, sazonalidade estocástica e termo AR(1). Nas séries de veículos e energia elétrica o termo AR(1) estacionário dominou o componente irregular; com isto, este último foi retirado. Nesses casos, as séries estacionarizadas são representadas pelo componente AR(1). No caso da série cimento, a série estacionarizada é representada pelo termo irregular.

<sup>25</sup> O uso da média anual tem como objetivo preservar a relação de *benchmarking* das séries, à exceção do PIB.

<sup>26</sup> A estimação foi feita em *SsfPack3*.

constituem o primeiro passo para a implementação do modelo multivariado com quatro variáveis dependentes, que não chegou a ser estimado. O segundo grupo de matrizes refere-se apenas ao modelo multivariado.

Procedemos também a um exercício de modelo *benchmarking* contendo valores omissos (PIB3) – dado que o modelo trabalha naturalmente com a presença de observações omissas, embora não em grandes proporções da amostra. Estes valores foram concentrados no período anterior a 1980 e escolhidos pela sua atipicidade<sup>27</sup>. As demais características do modelo são como descritas acima. Em ambos os modelos, o PIB reportado não inclui o erro de medida.

Cabe observar que a solução encontrada para tratamento da heterocedasticidade não implicou em não-linearidades nos moldes descritos por Durbin e Queenneville – DQ – (1997). O que fizemos foi transformar variáveis linearmente e estimar o modelo de tal maneira que os valores trimestrais fossem corrigidos para que a soma dos trimestres fosse igual ao total anual. O nosso procedimento de tratamento da heterocedasticidade é uma alternativa ao modelo multiplicativo proposto por DQ (1997), o qual incorre em não-linearidades. Ademais a metodologia proposta é computacionalmente mais eficiente, além de permitir que chequemos através das inovações padronizadas se tratamos corretamente a heterocedasticidade.

O último modelo estimado foi um multivariado em que as quatro variáveis – PIB, veículos, cimento e energia elétrica – estão endogeneizadas. As séries são usadas em índices, divididas previamente pelo índice da receita tributária ajustada de *outliers* (figura A2). A especificação envolve uma tendência local para cada variável, *dummies* sazonais na forma trigonométrica – porque no modelo lidaram melhor com a correlação serial presente –, e variáveis explicativas estacionarizadas, também para tratamento da

---

<sup>27</sup> Foram retirados da amostra e substituídos por valores omissos, os PIBs dos trimestres 1966.1 a 1966.4, 1973.4, 1975.1, 1976.1, 1979.1, 1994.1 e 1994.2. Os dois últimos foram removidos por terem exatamente o mesmo valor.

correlação serial; ver DK (2004). No caso, a escolha recaiu sobre a série de automóveis com seis defasagens. O interesse deste último modelo é exatamente apresentar a extração dos componentes estruturais em um ambiente multivariado e comparar com os resultados dos modelos univariados. O PIB estimado reportado também não inclui o componente irregular, permitindo que os componentes de tendência e sazonalidade estocástica de cada variável expliquem o seu comportamento.

#### 4. Resultados

Nas tabelas 2 e 3 a seguir, são apresentados os testes de diagnósticos dos modelos estimados descritos na seção anterior<sup>28</sup>. Todos os testes foram feitos com a série de resíduos padronizados do modelo, as chamadas inovações; cf. Durbin e Koopman, 2001. O modelo PIB1 refere-se ao modelo com *missings*, o PIB2 com *benchmarking*, e o PIB3 com *missings* e *benchmarking*. O modelo PIB4 é o modelo multivariado estrutural.

TABELA 2: TESTES DE DIAGNÓSTICOS

Modelos	PIB1 <sup>1</sup>	PIB2	PIB3
Estatísticas			
LogL	110.236	229.262	203.655
Iterações	16 (forte)	64 (forte)	61 (forte)
P.E.V.	0.0030	NA	NA
Erro-padrão	0.0549	NA	NA
DH	0.0008 [0.9996] <sup>3</sup>	21.895 [0.0000]	67.648[0.0000] <sup>3</sup>
H(h)	1.2269 [0.2430] <sup>4</sup>	0.7382 [0.8747.] <sup>5</sup>	1.3209[0.1548] <sup>6</sup>
DW	1.9540 [0.7576]	1.8561 [0.3278]	1.7346 [0.0863]
Q(12)	16.505 [0.1234]	12.964 [0.3717]	16.522 [0.168]
CumPer (K-S) <sup>2</sup>	0.0877	0.1544	0.1134
R2	0.9377	NA	NA
RD2	0.7307	NA	NA
AIC	-1.9483	-5.3337	-5.2086
SIC (Bayes)	-1.7985	-5.1729	-5.0406

(1) 107 observações válidas; (2)  $K-S(0.10)=0.1078$ ; (3) O teste utilizado foi o de Bera-Jarque, que também tem distribuição aproximada  $\chi^2_2$ ; (4)  $h = 47$ ; (5)  $h = 58$ ; (6)  $h = 54$ .

<sup>28</sup> Sobre a notação cabem algumas informações. P.E.V. é a predição da variância do erro. DH é o teste de normalidade de Doornik-Hansen com distribuição aproximada  $\chi^2_2$ . H(h) é o teste para presença de heterocedasticidade de Goldfeld-Quandt, com distribuição aproximada F(h,h). Os coeficientes de determinação são reportados para as séries em níveis, diferenças e para diferenças sazonais (RS2). Por fim, são reportados os critérios de informação de Akaike e Schwartz (Bayes).

TABELA 3: TESTES DE DIAGNÓSTICOS

Mod. Multivar.	PIB4			
	PIB	CIM	EE	AUTO
Estatísticas				
LogL	1457.73	1457.73	1457.73	1457.73
Iterações	38 (Very Strong)	38	38	38
P.E.V.	0.0088	0.0068	0.0103	0.0117
Erro-padrão	0.0938	0.0825	0.1017	0.1079
DH	2.8403 [0.2417]	4.2242 [0.1210]	0.6834 [0.7105]	0.2926 [0.8639]
H(46)	0.5234 [0.9848]	0.6467 [0.9285]	1.0551[0.4282]	1.3326 [0.1668]
DW	1.8763 [0.4611]	1.9279 [0.6675]	1.8781 [0.4677]	1.8263 [0.3007]
Q(10)	13.667 [0.0909]	10.423 [0.2366]	13.986 [0.0821]	9.2378 [0.3226]
CumPer (K-S)	0.1991	0.1181	0.2270	0.1537
R2	0.8112	0.6041	0.7272	0.6424
RD2	0.4631	0.4855	0.4515	0.5275
RS2	0.3384	0.3584	0.3269	0.4242
AIC	-3.7610	-4.0187	-3.6007	-3.4809
SIC (Bayes)	-2.3247	-2.5824	-2.1644	-2.0446

O modelo PIB1, sem benchmarking, apresenta diagnósticos adequados. É importante frisar a inexistência de quaisquer indicações de quebras no termo de nível da série, nem da presença de *outliers* em 1980, um ano crítico, dado que nele ocorre o encadeamento da série estimada e a série publicada pelo IBGE<sup>29</sup>. Podemos concluir que o modelo estimado capturou adequadamente os componentes estruturais da série, o que provavelmente decorreu de a sazonalidade ter sido especificada estocasticamente.

As avaliações das predições um passo a frente, feitas para o período 1980 a 1996, estão reportadas na tabela 4. Esse é o período para o qual estão disponíveis as estimativas oficiais do IBGE. Apenas ressaltamos que as estatísticas são bem claras quanto a acurácia das estimativas realizadas pelo modelo PIB1. O gráfico do PIB1 – estimado pela soma dos componentes de tendência e sazonais – convertido para valores reais de 1980 aparece na figura 3. Na quinta coluna, as estatísticas se referem ao PIB em valores estimado com os coeficientes da equação de longo prazo do procedimento de Engle-Granger sem encadeamento com a série do IBGE; ver seção 3.1.

<sup>29</sup> Para corroborar esta afirmativa, fizemos uma regressão da série PIB1, em níveis e em diferenças, contra termos determinísticos (constante, tendência e *dummies* sazonais), estimada por mínimos quadrados recursivos. Os coeficientes, assim como seus intervalos de confiança, não apresentam mudanças de comportamento em 1980, apesar de serem bastante instáveis. Os resultados estão disponíveis aos interessados.

TABELA 4: AVALIAÇÃO DAS PREDIÇÕES: 1980 A 1996

Modelo	PIB1	PIB2	PIB3	PIB4	YQ6096
Estatísticas					
Raiz Quadrada do EQM	0.0089	0.0112	0.0158	0.0245	0.0427
EQM (%)	0.0069	0.0078	0.0121	0.0196	0.0372
Erro Absoluto Médio	0.0070	0.0073	0.0091	0.0199	0.0356
Erro Absoluto Médio (%)	0.0054	0.0052	0.0066	0.0158	0.0312
Coef. Desigualdade de Theil	0.0034	0.0043	0.0060	0.0094	0.0185
Proporção do Viés	0.0008	0.2847	0.1428	0.0021	0.0398
Proporção da Variância	0.0008	0.2685	0.0947	0.0050	0.0263
Proporção da Covariância	0.9984	0.4469	0.7625	0.9929	0.9339
Pseudo R2	0.9969	0.9979	0.9928	0.9770	0.9198

O modelo PIB2, corrige os valores trimestrais do PIB de tal maneira que a soma dos PIBs trimestrais estimados iguala-se ao PIB oficial do IBGE. A heterocedasticidade foi inteiramente tratada. Não há presença de correlação serial nas inovações padronizadas, mas o teste K-S é rejeitado a 1%. A normalidade das inovações é rejeitada, o que não chega a ser um problema, dado que não pretendemos fazer inferência estatística sobre parâmetros desconhecidos, mas sim extrair componentes estruturais. Contudo, é uma indicação de que o comportamento dos valores observados não está sendo representado apropriadamente pelo modelo em consideração. Frisa-se também que não foram encontrados quaisquer indícios de quebras no termo de nível da série resultante nem de presença de *outliers* na vizinhança de 1980.

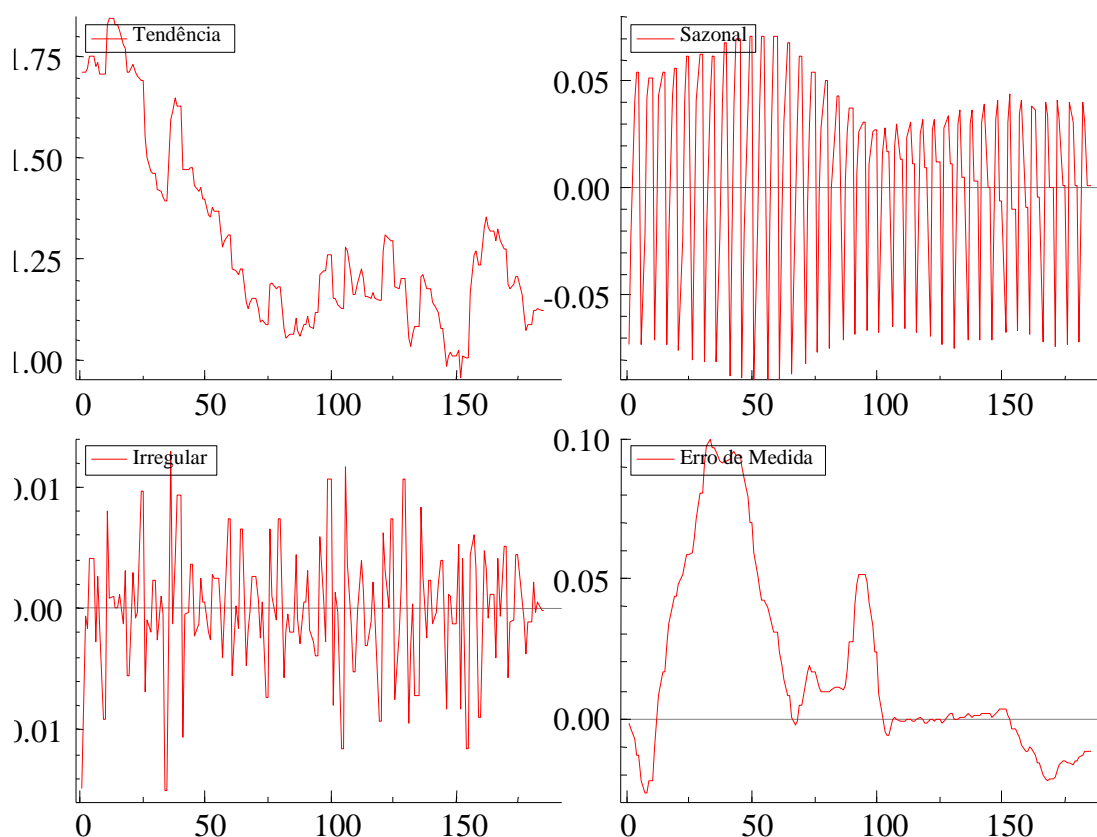
Entre as limitações dos resultados desse modelo, ressaltamos a ausência de *dummies* de intervenção, o que foi detectado através de experimentos realizados com o modelo com estrutura de *benchmarking* – rodados em ambiente de modelo estrutural com a especificação incluindo variáveis explicativas. Por outro lado, o erro de medida estimado possui um coeficiente de correlação serial de primeira ordem muito próximo de um  $(0.9727)^{30}$ , o que significa que este termo absorveu ou refletiu a correlação serial que, de outra forma, estaria contida nas inovações. Interpreta-se essa forte persistência

<sup>30</sup> O erro de medida estimado  $\xi_t$  sem os totais anuais tem ordem de integração próxima de um. Os testes de raiz unitária com polinômio de defasagens apropriadamente especificado aceitam em conjunto a hipótese nula de raiz unitária;  $\xi_t$  segue assim um processo ARIMA(4,1,0). A hipótese do modelo é que  $\xi_t$  é AR(1)  $\sim$  (0,1).



como sendo “herdada” de estruturas que, potencialmente, deveriam estar embutidas nos demais componentes do vetor de estado<sup>31</sup>, conforme se visualiza no quarto painel da figura 2. Julgamos esse problema superável com a mudança da especificação do modelo<sup>32</sup>.

FIGURA 2: Componentes Extraídos do Modelo PIB2



Todas as observações feitas para o modelo PIB2 se aplicam ao modelo PIB3 estimado contendo alguns valores omissos; ver seção anterior. Em especial, destacamos três resultados dos modelos de *benchmarking*. O primeiro – não reportado nesta versão

<sup>31</sup> Também é importante ter em conta que, devido às inúmeras dificuldades computacionais para estimação desse modelo, esta análise é incompleta, posto não ter sido possível estimar os resíduos auxiliares.

<sup>32</sup> Para testar a validade do procedimento de *benchmarking* empregado podemos adotar o seguinte procedimento: (i) tomar uma série observada ou simulada com frequência trimestral e cujos totais anuais sejam iguais à soma dos trimestres e (ii) introduzir um erro seguindo um processo AR(1) nos dados reais; (iii) em seguida, aplicar o procedimento de *benchmarking* e comparar os resultados para observar se a série foi corrigida do erro de medida, tal como proposto pelo método.

do artigo –, é que o modelo PIB2 estimado sem tratamento de heterocedasticidade produz praticamente a mesma série, não havendo diferença material entre elas. Assim, para esse modelo de estado de espaço – *benchmarking* – o filtro de Kalman é robusto à heterocedasticidade. O segundo é que a presença de dados omissos por sua vez afeta as estimativas, mas não de modo expressivo. O terceiro é um resultado evidenciado pela análise da avaliação das predições para o período 1980 a 1996, apresentada na tabela 4. O aparente viés das estimativas vem do fato de que a partir de 1990.4 o PIB estimado sobreestima sistematicamente as estimativas do PIB feitas pelo IBGE (ver figura 3). Em especial o PIB3 não acompanha os movimentos bruscos do PIB estimado pelo IBGE ocorridos em 1991, subestimando-os, o que sugere que o nosso procedimento está efetivamente corrigindo as estimativas trimestrais oficiais, de modo que a soma dos trimestres seja consistente com – igual a – o PIB anual. Nesse caso, o viés detectado é uma característica do PIB trimestral oficial estimado e, que, portanto, são as estimativas do IBGE que não apresentam consistência.

O modelo multivariado associado ao PIB4 estimado envolveu as séries de índice do PIB, de produção de cimento, de veículos automotores e consumo de energia elétrica no Rio de Janeiro e em São Paulo. Todas as séries foram divididas pelo índice da receita tributária real, extraíndo-se previamente os *outliers*. A série de produção de automóveis em defasagens, apropriadamente estacionarizada, entrou como variável explicativa com o intuito de controlar a correlação serial. Os testes dos resíduos padronizados indicam que os pressupostos básicos do modelo estão sendo respeitados. Por outro lado, não há nos resíduos auxiliares das variáveis indicações da presença de *outliers* ou de quebras, o que indica que o modelo captura adequadamente os componentes estruturais da

variável<sup>33</sup>. O PIB4 é estimado como a soma dos componentes de tendência e sazonal a menos do termo irregular. A sua aderência aos dados do IBGE é evidente.

Vale ressaltar, que os componentes sazonais extraídos dos modelos estimados, (ver figura 2 para o modelo PIB2; os demais não foram reportados), não apresentam quebras significativas em torno do ano de 1980 (observação 29). O que se observa é a sazonalidade oscilando ao longo de todo o período. O que permite deduzir, que o encadeamento, **per si**, da série estimada na primeira etapa com a do IBGE não implicou em alterações materiais do padrão sazonal.

Na figura 3 estão plotados os quatro PIBs estimados pela metodologia de espaço de estado e o PIBPROXY - estimado pela estática de Engle-Granger e encadeado com a série do IBGE. Olhados individualmente, cada um dos resultados estimados apresenta boas propriedades e todos representam um avanço em relação às estimativas do PIB com os coeficientes da estática de Engle-Granger. Por sua vez, os procedimentos de estimação utilizando filtro de Kalman com valores omissos e com processo de *benchmarking* são inovadores na literatura brasileira, ao que se juntam o procedimento de tratamento da heterocedasticidade e o emprego de variáveis explicativas em modelos univariados e multivariados.

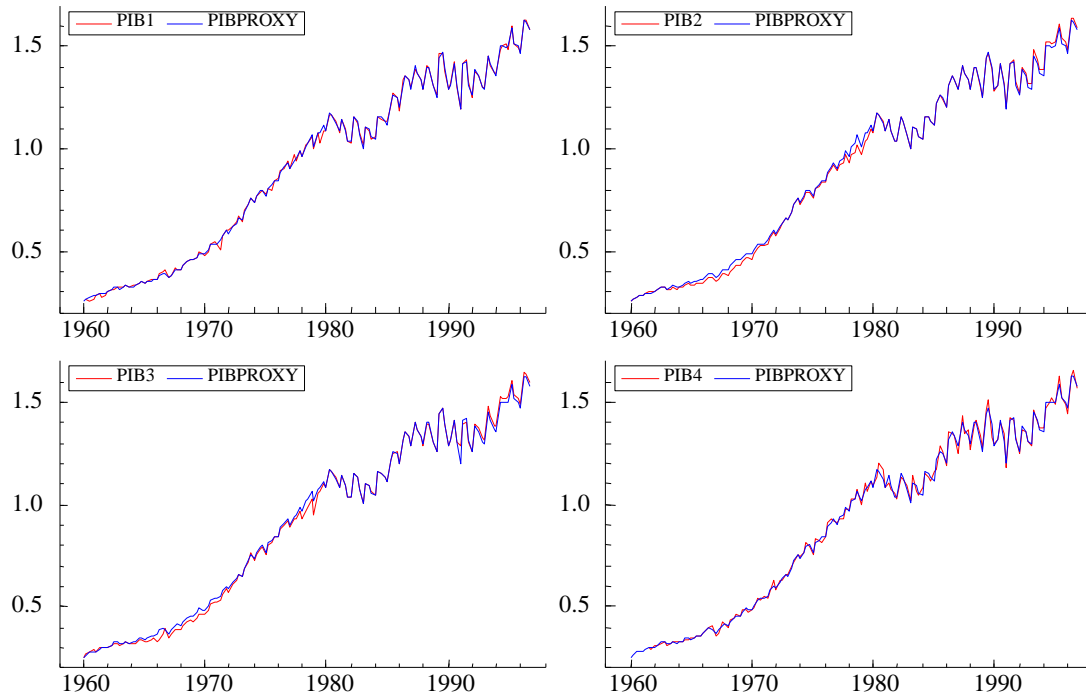
As estatísticas dos testes de diagnóstico e da avaliação da capacidade preditiva dos modelos não indicam claramente qual modelo escolher, entre o univariado com valores omissos (PIB1) e o multivariado (PIB4), para serem usados para recuperação do PIB trimestral no período 1960 a 1979 e, como substituto do PIB do IBGE nos anos de 1980 a 1996. Os modelos com *benchmarking* (PIB2 e PIB3) por corrigirem o erro de medida existente nos dados trimestrais, harmonizam os dados recuperados com os

---

<sup>33</sup> Os componentes extraídos da equação do PIB – não reportados – indicam que o componente de tendência se assemelha ao do modelo PIB1, embora a sazonalidade apresente um padrão mais instável e de maior amplitude, provavelmente devido a presença de quatro variáveis dentro do modelo. O termo irregular, tal como no modelo do PIB1, é um termo auto-regressivo estável, no caso um AR(8).

valores anuais conhecidos. E constituem um aperfeiçoamento em relação aos próprios dados trimestrais da fonte oficial, IBGE, no período 1980 a 1996.

FIGURA 4: PIBs Estimados



## 5. Conclusão

Neste artigo apresentamos uma metodologia para a estimação do PIB trimestral para o período 1960 a 1996 e reportamos as estimativas realizadas. Embora o objetivo inicial fosse estimá-lo apenas para o período 1960 a 1979, a necessidade de uma amostra maior nos levou a fazer estimativas para um período mais longo. A necessidade de uma *proxy* judiciosa para o PIB para períodos anteriores a 1980 se relaciona diretamente à realização de estudos empíricos que cubram um período mais longo da história econômica brasileira.

Inicialmente, realizamos um experimento com um modelo simples, porém rigoroso, dado seus estimadores serem superconsistentes, que foi o de estimar os coeficientes da equação estática do teste de cointegração de dois estágios de Engle-

Granger realizado com frequência anual. As séries utilizadas foram o PIB, a produção de cimento, veículos automotores, consumo industrial de energia elétrica na região Rio de Janeiro e São Paulo e receita tributária do tesouro real. O vetor de cointegração estimado foi, então, aplicado sobre os dados trimestrais das séries para gerar a estimativa do PIB trimestral. Os resultados indicaram que a melhor conduta seria estender os estudos com uma metodologia mais robusta, o que nos conduziu a estimar modelos que utilizam métodos de espaço de estado.

Continuamos a pesquisa com diferentes modelos estruturais, dadas as estimativas realizadas com o modelo acima. Estimamos um modelo univariado com valores omissos no período 1960 a 1979, utilizando variáveis explicativas. Um modelo com processo de *benchmarking*, que corrige o erro de medida dos valores dos trimestres de modo que a soma dos PIBs trimestrais totalize o PIB anual respectivo. Realizamos também um experimento com modelo de *benchmarking* e valores omissos. Finalmente, estimamos um modelo estrutural multivariado a quatro séries e com variáveis explicativas. Todos os modelos foram estimados com correção de heterocedasticidade. Estimamos, portanto, quatro modelos concorrentes e quatro séries de PIB trimestral. À exceção do modelo de *benchmarking*, nenhum dos PIBs recuperados inclui o componente irregular. A lógica do *benchmarking* considera a componente irregular porque o componente de erro de medida é retirado da sua composição.

Os resultados foram bastante satisfatórios no que concerne aos diagnósticos dos resíduos, os quais respeitam as hipóteses do modelo gaussiano linear. Da mesma forma, os PIBs estimados apresentam boa aderência aos dados e previram adequadamente o PIB oficial no período pós-1980. O PIB estimado com *benchmarking* corrige ao máximo os erros de medida e, inequivocamente mantém consistência com o PIB anual, aperfeiçoando os próprios dados oficiais. A nível global, o modelo estimado com

*missing values* e o modelo SUTSE apresentaram as melhores performances. Em conjunto, as estatísticas reportadas não indicam claramente qual a melhor estimativa. Certamente, todas as estimativas representaram uma melhora em relação ao PIB estimado com a estática de Engle-Granger. A escolha para prever o PIB no período 1960 a 1996 deve ficar recair sobre uma delas.

Como recomendação de estudos futuros indicamos três direções. A primeira é utilizando os modelos discutidos neste artigo pesquisar por outros períodos, preferencialmente que se estendam até o presente e que permitam a inclusão de outras variáveis relacionadas com o PIB. Como consequência natural o caminho para realização de extrapolações fica aberto. Uma segunda complementar é a de estender a metodologia para modelos multivariados de benchmarking, o que traria maior precisão e confiabilidade às séries finais do PIB trimestral. Finalmente, sugerimos estender o procedimento para a recuperação e correção de erros de medida de dados mensais, a partir de dados trimestrais previamente corrigidos.

A principal contribuição do artigo foi apresentar uma metodologia consistente e original para a recuperação e estimação do PIB trimestral do período 1960 a 1996, e que ademais, é capaz de corrigir os dados da pesquisa trimestral do IBGE. Por fim, nossa metodologia oferece bases concretas para o desenvolvimento de um procedimento de previsão/extrapolação do PIB trimestral.

## 6. Referências Bibliográficas

Banco Central do Brasil. “Boletim Mensal”, maio de 1994.

Banco Central do Brasil. <http://bcb.gov.br/>, acessado em 21 de junho de 2004.

Bassi, V.L., (1958). “Appendix A,” in *Economic Forecasting*, ed. by V.L. Bassi, New York: McGraw-Hill.

Bertsekas, D.P. (1976). *Dynamic Programming and Stochastic Control*. New York: Academic Press.

Caporale, G. M and Pittis, N. (2002). “Unit Root versus Other Types of Time Heterogeneity, Parameter Time Dependence and Superexogeneity”, *Journal of Forecasting*, 21, 207-223.

Cardoso, E. (1981). “Uma Equação para a Demanda de Moeda no Brasil”, *Pesq. Plan. Econ.*, 11(3), 617-655.

Cerqueira, L.F. (1993). “Demanda por Moeda, Regressões Espúrias e Níveis de Instabilidade Hiperinflacionários”. *Dissertação de mestrado não publicada, Departamento de Economia, Pontifícia Universidade Católica, RJ.*

Cerqueira, L.F. (2007). *Metodologia de Estimação do PIB Trimestral Utilizando Procedimentos de Cointegração e Filtros de Kalman*. Textos Para Discussão 216, Departamento de Economia, UFF, Niterói, Junho 2007.

Chow, G., C., and Lin, A. (1971), “Best Linear Unbiased Interpolation, Distribution and Extrapolation of Time Series by Related Series,” *The Review of Economics and Statistics*, 53, 372-375.

Cholette P.A. and E.B. Dagum (1994), Benchmarking Time Series with Autocorrelated Survey Errors, *International Statistical Review* 62(3), 365-377.

Contador, C.R. e Santos Filho, W.A.C. (1987). “Produto Interno Bruto Trimestral: Bases Metodológicas e Estimativas”, *Pesq. Plan. Econ.*, 17(3), 711-742.

Denton, F. T. (1971), “Adjustment of Monthly or Quarterly Series to Annual Totals: An Approach Based on Quadratic Minimization,” *Journal of the American Statistical Association*, 66, 99-102.

Di Fonzo, T. and Marini, M. (2003) “Benchmarking Systems of Seasonally Adjusted Time Series According to Denton's Movement Preservation Principle”, Dipartimento di Scienze Statistiche, Università di Padova, Working Paper n. 2003-09.

*Durbin, J. and Koopman, S.J. (2004). Time Series Analysis by State Space Methods. Oxford : Oxford University Press.*

*Durbin, J. and Quenneville, B. (1997). “Benchmarking by State Space Models”, International Statistical Review, 65(1), 21-48.*

*Engle, R.F. and Granger, C. (1987). “Cointegration and Error Correction: Representation, Estimation, and Testing”, in Engle, R. and Granger, C.W.J. (eds.), Long-Run Economic Relationships: Readings in Cointegration. Oxford: Oxford Univ. Press, (1991).*

*Engle, R.F. and Granger, C.W.J. (1991). “Introduction” to Long-Run Economic Relationships: Readings in Cointegration, op. cit., (1991).*

*Elliot, G., Rothenberg, T.J. & Stock, J.H. (1996). “Efficient Tests for an Autoregressive Unit Root”, Econometrica, 64, 813-836.*

*FGV, “Conjuntura Econômica”, abri de 1977 e agosto de 1979.*

*Ginsburgh, V.A. (1973). “A Further Note on the Derivation of Quarterly Figures Consistent with Annual Data,” Applied Statistics, 22(3), 368–374.*

*Harvey, A.C. (1990). Forecasting, Structural Time Series and the Kalman Filter. Cambridge: Cambridge University Press.*

*Harvey, A.C and Koopman, S.J. (1992). “Diagnostic Checking of Unobserved Time Series Models”, J. Business and Economic Statistics, 10, 377-389.*

*Harvey, A.C and Koopman, S.J. (1997). “Multivariate Structural Time Series Models, In Heji, C., Schumacher, H., Hanzon, B. and Praagman, C. (eds.) Systematic Dynamics in Economics and Financial Models, pp. 269-298. Chichester: John Wiley and Sons.*

*Harvey, A.C and Koopman, S.J. (2000). “Signal Extraction and the Formulation of Unobserved Components Models”, Econometrics Journal, 3, 84-107.*

*Hillmer S.C. and Trabelsi, A. (1987), Benchmarking of Economic Time Series, Journal of the American Statistical Association, 82(400), 1064-1071.*



IBGE, “Anuário Estatístico do Brasil”, 1990 e 1994.

IBGE. “Sistema IBGE de Recuperação Automática-SIDRA”, <http://www.sidra.ibge.gov.br/>, acessado em 21/06/2004.

IPEA. “Ipeadata”, <http://www.ipeadata.gov.br/>, acessado em 21/06/2004.

Johansen, S. (1995). Likelihood-Based Inference in Cointegrated Vector Auto-Regressive Models. New York: Oxford University Press.

Koopman, S. J. and Shephard, H. (1992). “Exact Score for Time Series Models in State Space Form”, *Biometrika*, 79, 823-26.

Koopman, S. J., Shephard, H. and Doornik, J.A. (1999). “Statistical Algorithms for Models in State Space Using SsfPack 2.2”, *Econometrics Journal*, 2, 107-160.

Koopman, S.J., Harvey, A.C., Doornik, J.A. and Shephard, N. (2000). Stamp 6.3: Structural Time Series Analyzer, Modeller, and Predictor”. London: Timberlake Consultants.

Maddala, G.S. and Kim, In-Moo (2002). Unit Roots, Cointegration, and Structural Changes. Cambridge: Cambridge University Press.

Mian, I.U.H., and N. Laniel, (1993). “Maximum Likelihood Estimation of Constant Multiplicative Bias Benchmark Model with Application,” *Survey Methodology*, 19, (December), 165–72.

Moauero, F and Savio, G. (2005). “Temporal Disaggregating Using Multivariate Structural Time Series Models”, *Econometrics Journal*, 8(2), 214-234.

Nakane, M.I. (1994). *Testes de Exogeneidade Fraca e de Superexogeneidade para a Demanda por Moeda no Brasil*. Rio de Janeiro: BNDES.

Pastore, A.C. (1991). “Déficit, Dívida Pública, Moeda e Inflação: Uma Resenha”. Unpublished.

Ng, S. and Perron, P. (1997). “Useful Modifications to Some Unit Root Tests with Dependent Errors and Their Local Asymptotic Properties”, *Review of Economic Studies*, 63, 435-465.

Pinotti, M.C. (1991). “Inovações Financeiras e Imposto Inflacionário”. Tese de Doutorado, FEA, USP, SP.

*Phillips, P.C.B. and Hansen, B.E. (1990). “Statistical Inference in Instrumental Variables Regression with I(1) Processes”, Review of Economic Studies, 58, 407-436.*

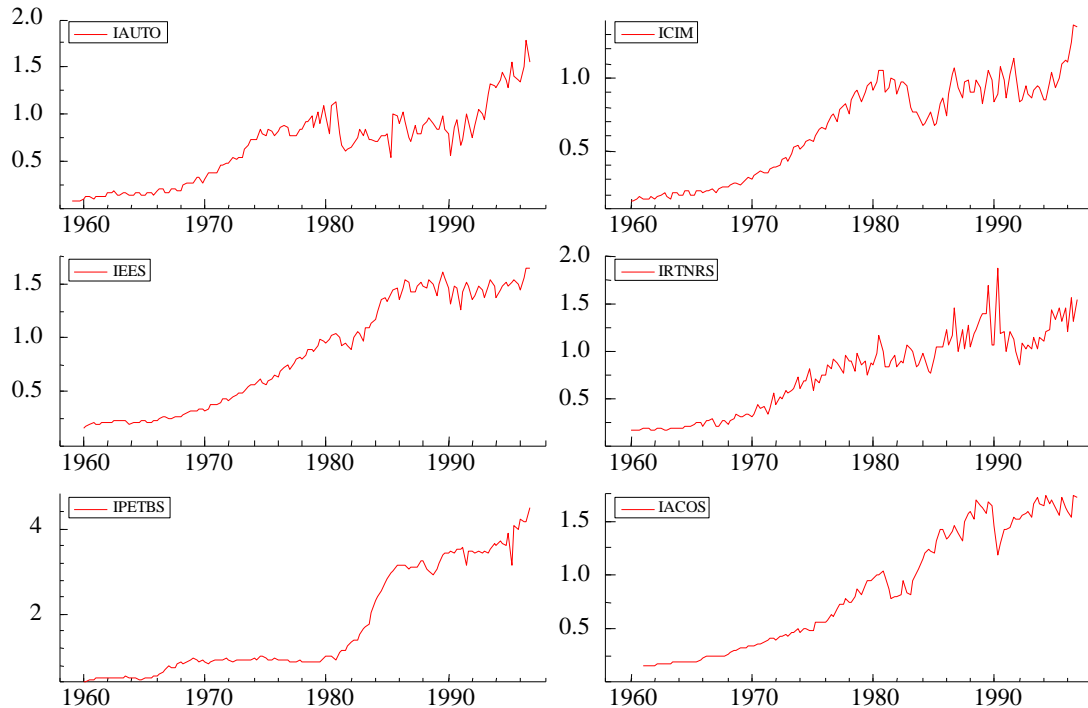
*QMS (2004). Eviews 5. Irvine, CA: Quantitative Micro Software, LLC.*

*Rossi, J.M. (1988). “A Demanda de Moeda no Brasil: O Que Ocorreu a Partir de 1980?”, PPE, 18(1), 37-53.*

## Apêndice A: Dados Utilizados

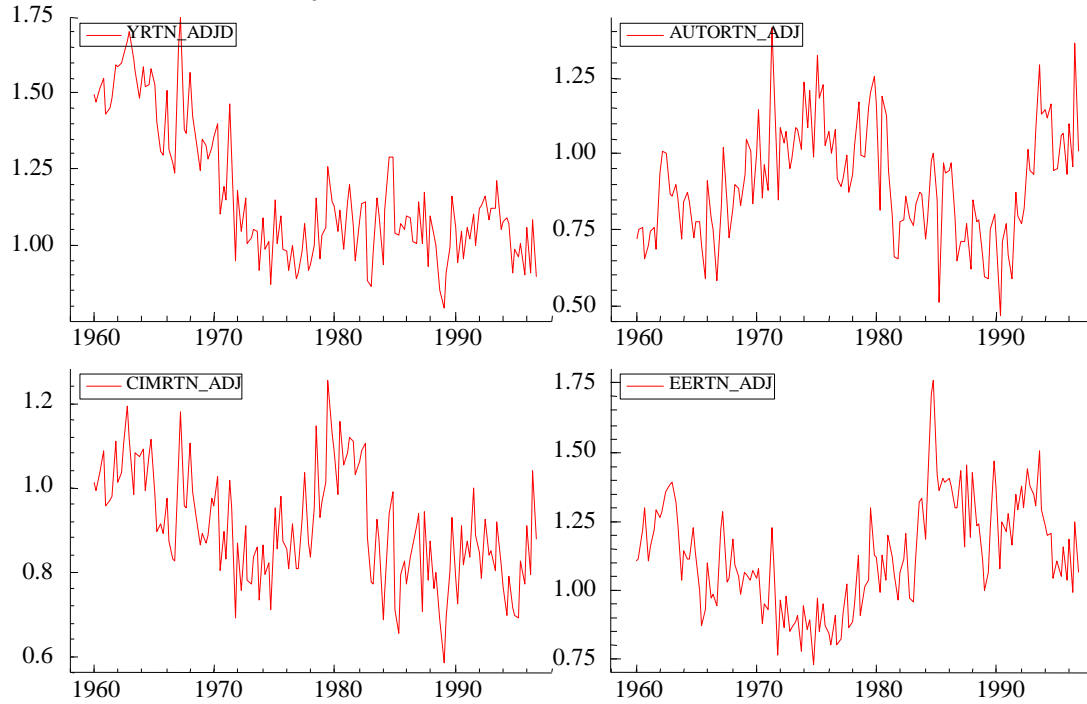
Na figura A1, estão os dados em índices trimestrais encadeados com as observações interpoladas ou projetadas, respectivamente, da produção de automóveis (IAUTO), cimento (ICIM), consumo industrial de energia elétrica (IEES), receita tributária real (IRTNRS), produção de petróleo e LGN (IPETBS) e aço em lingotes (IACOS)<sup>34</sup>. Na figura A2 aparecem, as séries do PIB estimado, YRTN, as séries observadas da produção de automóveis, cimento e consumo industrial de energia elétrica, todas divididas pela série da receita tributária real ajustada e expressa em índices (ver texto).

FIGURA A1: Variáveis Intervenientes no PIB – Dados Trimestrais



<sup>34</sup> Para as séries que possuem apenas frequência anual para parte do período – produção de petróleo, aço e receita tributária nacional em valores correntes – construiu-se um teste ADF de raiz unitária para um período “específico”. Com a especificação do teste, deduz-se o processo ARIMA da série, *i.e.*, um *random walk* puro, um *random walk* com um processo auto-regressivo de ordem 1 e o coeficiente de auto-regressão associado, ou um processo integrado de segunda ordem. Com essas informações, realiza-se a interpolação descrita em Bertsekas (1976) e encadeia-se a série interpolada com a série trimestral observada a partir do período em que os dados não estão disponíveis. Para a série de produção de petróleo, escolheu-se interpolar no período 1955 a 1969, em função da homogeneidade do crescimento nesse período, ainda não muito pronunciado. A maior parte dos testes indicou a série como sendo I(2), o que nos levou à escolha desta especificação. A interpolação da produção de aço foi feita para o período 1961 a 1997 a partir da constatação de que ela segue um processo do tipo *random walk* puro. A série de receita tributária nacional foi interpolada no período 1952 a 1973, supondo que a série no período era I(2). A série foi convertida para reais e encadeada com a série observada em 1965.3.

FIGURA A2: Variáveis Ajustadas – Dados Trimestrais



## Apêndice B: Matrizes do Modelo de *Benchmarking*

Abaixo apresentamos o grupo de matrizes referente a um modelo de *benchmarking* univariado com nível, *dummies* sazonais e termo de erro de medida, juntamente com os vetores de estados  $\alpha_t$  e de choques de estado  $\tilde{\eta}_t$ , associados às matrizes. Por conveniência, o vetor de observações está inserido dentro do vetor de estado. Lembramos que o modelo espaço de estado é dado por

$$y_t = Z_t \alpha_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, H_t),$$

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t, \quad \eta_t \sim N(0, Q_t), \quad t = 1, \dots, n, \quad \alpha_1 \sim N(a_1, P_1), \quad (3.10)$$

onde no modelo univariado  $y_t$  tem uma variável e no modelo multivariado quatro variáveis empilhadas.

Para os  $Y$ 's (trimestres), as matrizes são

$$Z_t = [1 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0 \ 1],$$

$$T_t = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \phi \end{bmatrix},$$

$$R_t = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Para os  $X$ 's (totais anuais), as matrizes são  $z_t = [1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1]$ ,  $T_t =$

Identidade 12 x 12,

$$R_t = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Os vetores de estados e de choques de estados são respectivamente os seguintes

$$\alpha_t' = \left[ \mu_t, \mu_{t-1}, \mu_{t-2}, \mu_{t-3}, \gamma_t, \gamma_{t-1}, \gamma_{t-2}, \gamma_{t-3}, \varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \varepsilon_{t-3}, \xi_t^s \right]$$

$$e \quad \tilde{\eta}'_t = [\eta_t, \omega_t, \varepsilon_t^*, \kappa_t].$$

As matrizes abaixo se referem a um modelo de *benchmarking* com quatro variáveis dependentes (vetor de medidas com dimensão 4), uma delas sendo o PIB. A especificação de cada variável é a mesma do modelo univariado. As matrizes  $Z_t$ ,  $T_t$ , e  $R_t$  possuem uma dimensão a menos que as matrizes do modelo univariado, porque delas foi retirado o erro de medida  $\xi_t$ , o qual é recolocado posteriormente.

O procedimento de construção das matrizes segue os seguintes passos: (i) reconstruir as matrizes do modelo univariado sem considerar o erro de medida, para os  $Y$ 's e para os  $X$ 's, as quais não são rerepresentadas; (ii) construir as matrizes  $\tilde{Z}_t = Z_t \otimes I_4$ ,  $\tilde{T}_t = T_t \otimes I_4$ , e  $\tilde{R}_t = R_t \otimes I_4$ , nas quais  $\otimes$  é o produto de Kronecker; (iii) construir as matrizes finais, que aparecem a seguir, montadas a partir dos componentes descritos, seguidas das matrizes bloco-diagonal de covariância  $Q_t$  e  $H_t$ .

Para os  $Y$ 's (trimestres) as matrizes são

$$\bar{Z}_t = \begin{pmatrix} \left( \begin{array}{c} 1 \\ 0 \\ 0 \\ 0 \end{array} \right) \\ (\tilde{Z}_t) \end{pmatrix}, \quad \bar{T}_t = \begin{pmatrix} (\tilde{T}_t) & \begin{pmatrix} 0 \\ 48 \times 1 \end{pmatrix} \\ \begin{pmatrix} 0 \\ 1 \times 48 \end{pmatrix} & (\phi) \end{pmatrix}, \quad \bar{R}_t = \begin{pmatrix} (\tilde{R}_t) & \begin{pmatrix} 0 \\ 48 \times 1 \end{pmatrix} \\ \begin{pmatrix} 0 \\ 1 \times 12 \end{pmatrix} & 1 \end{pmatrix},$$

$$Q = \text{diagcat}(\Sigma_\eta, \Sigma_\omega, \Sigma_\varepsilon, \sigma_\xi).$$

Para os  $X$ 's (totais anuais), as matrizes são

$$\bar{Z}_t = \begin{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \\ (\tilde{Z}_t) \end{pmatrix}, \quad \bar{T}_t = \begin{pmatrix} (\tilde{T}_t) & \begin{pmatrix} 0 \\ 48 \times 1 \end{pmatrix} \\ \begin{pmatrix} 0 \\ 1 \times 48 \end{pmatrix} & (1) \end{pmatrix}, \quad \bar{R}_t = \begin{pmatrix} (\tilde{R}_t) & \begin{pmatrix} 0 \\ 48 \times 1 \end{pmatrix} \\ \begin{pmatrix} 0 \\ 1 \times 12 \end{pmatrix} & \begin{pmatrix} 0 \\ 0 \end{pmatrix} \end{pmatrix},$$

$$Q = \text{diagcat}(\Sigma_\eta, \Sigma_\omega, \Sigma_\varepsilon, \sigma_\xi).$$

Observe-se que as matrizes do passo (ii) devem ter as dimensões  $48 \times 4$  ( $\tilde{Z}_t$ ),  $48 \times 48$  ( $\tilde{T}_t$ ) e  $8 \times 12$  ( $\tilde{R}_t$ ), tanto para os  $Y$ 's (trimestres) quanto para os  $X$ 's (totais anuais), e que os blocos de  $Q_t$  são “cheios” porque as componentes de mesma interpretação são correlacionadas. A matriz  $\sigma_\xi$  ( $4 \times 1$ ) é univariada pois existe um erro de medida, e este é associado ao PIB. O vetor de espaço do modelo multivariado não é apresentado, em função da sua dimensão ( $37 \times 1$ ). Porém ele nada mais é do que o vetor de estado do modelo univariado com cada grupo de componentes repetido quatro vezes devidamente indexados pelo número da variável dependente (quatro) correspondente. O vetor de choques de estado é dado por

$$\bar{\eta}'_t = \left[ \eta_{t1}, \eta_{t2}, \eta_{t3}, \eta_{t4}, \omega_{t1}, \omega_{t2}, \omega_{t3}, \omega_{t4}, \varepsilon_{t1}^*, \varepsilon_{t2}^*, \varepsilon_{t3}^*, \varepsilon_{t4}^*, \kappa_t \right].$$

Os  $\eta$ 's estão associados aos níveis; os  $\omega$ 's às sazonalidades; os  $\varepsilon^*$ 's às componentes irregulares; e  $\kappa_t$  é o choque do erro de medida do PIB; lembrando que este erro de medida segue um AR(1), como evidenciado pela matriz  $\bar{T}$  relativa aos  $Y$ 's (trimestres).