

Método RAWS/RAW para estimação anual da Matriz de Insumo-Produto na referência 2000 das Contas Nacionais

Resumo

O objetivo do artigo é construir um método, denominado RAWS/RAW, para estimar uma série anual de Matrizes de Insumo-Produto (MIPs) na referência 2000 das contas nacionais. O RAWS/RAW diferencia-se de outros procedimentos aplicados ao Brasil por fundamentar-se em algoritmos que otimizam o uso da informação relevante disponível. Desenvolve-se o algoritmo RAW a partir do algoritmo RAWS apresentado em AUTORX (2013), incorporando-se seus avanços metodológicos e as tabelas auxiliares das MIPs 2000/2005 por ele estimadas. A partir das MIPs completas de 2000 e 2005 obtidas com o RAWS, são interpoladas as MIPs 2001-2004 e projetadas as MIPs 2006-2009 pelo RAW. Para testar o método RAWS/RAW, projeta-se uma versão da MIP 2005 com as informações da MIP 2000, a qual, segundo diferentes medidas de precisão, apresenta desempenho superior ao de Guilhoto e Sesso-Filho (2005, 2010) na comparação com a MIP 2005 do IBGE.

Palavras-chave: insumo-produto; Contas Nacionais; RAS; balanceamento de matrizes.

Abstract

The aim of this paper is to build the so-called RAWS/RAW method to estimate an annual series of Input-Output Tables (IOTs) in 2000 reference of Brazilian national accounts. The RAWS/RAW differs from other procedures applied to Brazil in that it is based on algorithms that optimize the use of relevant information available. We develop the RAW algorithm from the RAWS algorithm presented in AUTORX (2013), by incorporating both its methodological advances and auxiliary tables of the 2000/2005 IOTs estimated there. Using the 2000 and 2005 IOTs obtained from the RAWS algorithm, we both interpolate 2001-2004 IOTs and project 2006-2009 IOTs with the RAW algorithm. To test the RAWS/RAW method, we project a version of the 2005 IOT using information from the 2000 IOT, which, according to various measures of accuracy, outperforms Guilhoto and Sesso-Filho (2005, 2010) in comparison with the 2005 IOT released by IBGE.

Keywords: input-output; National Accounts; RAS; matrix balancing.

JEL: D57, C67, E01

1. Introdução

A riqueza de informações proporcionada pela Matriz de Insumo-Produto (MIP) sobre as relações de comércio inter setoriais tem como custo a dificuldade envolvida na sua construção. Em razão do elevado volume de dados necessário, a MIP tipicamente é divulgada por órgãos de estatísticas oficiais com grande defasagem e em periodicidade superior a um ano. No caso brasileiro, para referência 2000 do Sistema de Contas Nacionais (SCN 2000) a MIP foi divulgada em periodicidade quinquenal, para os anos de 2000 e 2005 (IBGE, 2008).. Ainda que essa periodicidade possa ser suficiente para a compreensão de mudanças a longo prazo na estrutura produtiva, a disponibilidade de dados anuais é desejável para a análise de anos particulares, em especial conforme aumenta a defasagem após a divulgação da MIP oficial mais recente. O objetivo do presente artigo é propor um método para a estimação da MIP anual que, em comparação com metodologias já adotadas no Brasil (Guilhoto e Sesso-Filho, 2005 e 2010; Grijó e Berni, 2006), faça um uso mais abrangente dos dados disponíveis no SCN 2000.

As principais fontes de dados para construir a MIP, que apresenta relações de comércio entre os setores produtivos da economia, são as Tabelas de Recursos e Usos (TRUs), as quais exibem as relações de produção e venda entre os produtos e setores da economia. Na referência 2000 do SCN, o IBGE divulgou as TRUs em periodicidade anual, para 2000-2009. O empecilho para a construção da MIP na mesma periodicidade reside na incompatibilidade entre as fontes de informação da tabela de recursos e da tabela de usos. Pois a tabela de recursos apresenta a oferta dos produtos pelos setores da economia a preços básicos – ou seja, sem margens e impostos indiretos – e somente com a produção nacional. Em contraposição, a tabela de usos registra a demanda dos setores produtivos e dos componentes da demanda final sem distinção entre produção nacional ou importada e com valores a preços de consumidor, que incluem as margens e impostos.

Desta forma, para a construção da MIP faz-se necessário colocar os dados das duas tabelas na mesma valoração, e usualmente os preços básicos são os escolhidos porque a exclusão das margens e impostos faculta maior aproximação com as relações técnicas inter setoriais, que são o objeto de interesse da MIP. Assim, devem ser estimadas tabelas de distribuição dos totais por produto de margens, impostos indiretos e

importações – apresentados junto à tabela de recursos –, que serão extraídos de cada célula da tabela de usos a preços de consumidor. Portanto, o principal passo na estimação da MIP consiste em escolher a técnica adequada para a estimação destas tabelas auxiliares da MIP.

A literatura internacional sobre estimação de MIPs é vasta, pois os procedimentos adotados dependem não só da disponibilidade de dados em cada país, mas também do algoritmo de balanceamento de matrizes escolhido para conciliá-los. O algoritmo mais utilizado para tal fim é o método RAS – em sua forma original ou em uma de suas extensões –, que faz uma distribuição dos totais conhecidos de somas nas linhas e colunas da tabela que se quer estimar conforme as proporções dadas por uma projeção inicial desta matriz. Contudo, conforme as particularidades dos dados censitários disponíveis, torna-se necessário adaptar o RAS, o que dá origem aos chamados métodos híbridos de projeção de matrizes (Miller e Blair, 2009, cap. 7).

Para o Brasil, o método mais conhecido de estimação de MIPs anuais é o de Guilhoto e Sesso-Filho (2005, 2010), que não emprega o RAS. Os autores utilizam somente informações provenientes das TRUs do ano para o qual estão estimando a MIP, distribuindo os totais de margens, impostos e importações por linha (por produto) conforme as proporções da tabela de usos a preços de consumidor. De fato, não há como aplicar o RAS original diretamente, pois não são conhecidos os totais por coluna (por setor) referentes a cada tabela auxiliar.

Já o método de Grijó e Berni (2006) combina as TRUs com projeções da mais recente MIP divulgada pelo IBGE, por meio de um algoritmo RAS modificado. A projeção inicial é dada pelas informações da MIP do ano mais recente e a restrição de soma nas linhas pelas TRUs do ano estimado. A modificação consiste em substituir a restrição referente à soma nas colunas de cada tabela auxiliar pela restrição de que a soma das tabelas estimadas de usos a preços básicos, impostos indiretos, margens e importações deve ser igual à tabela de usos a preços de consumidor.

Entretanto, este método foi desenvolvido para as MIPs brasileiras da década de 1990, na referência 1985 do SCN, para as quais todas as tabelas auxiliares foram apresentadas pelo IBGE (1999) junto às MIPs 1990-1996. Na referência 2000 das contas nacionais, contudo, somente as tabelas de usos a preços básicos e de importações foram divulgadas

com as MIPs de 2000 e 2005. AUTORX (2013) desenvolve o algoritmo RAWS para a estimação de todas as tabelas auxiliares das MIPs de 2000 e 2005, um método híbrido que adapta o RAS levando em consideração diversas particularidades dos dados brasileiros.

No presente artigo, é desenvolvido o algoritmo RAW que, pela combinação das extensões teóricas e tabelas apresentadas em AUTORX (2013) com o procedimento de Grijó e Berni (2006), permite utilizar um conjunto de informações bastante abrangente para a estimação das MIPs anuais de 2000 a 2009. Esse procedimento aprimora a metodologia de Grijó e Berni (2006) em diversos aspectos, na derivação do problema de minimização de perda de informação e tratamento das células negativas, como no RAWS (AUTORX, 2013), e mesmo na estimação dos dados, como nas colunas de variação de estoques, na tabela de imposto de importação e nas linhas de totais de margens de comércio e transporte. O RAW é também adaptado para estimar os anos de 2001 a 2004 por meio da interpolação das tabelas de 2000 e 2005, enquanto as tabelas de 2006 a 2009 são projetadas a partir da MIP 2005. O procedimento completo para estimar as tabelas auxiliares de 2000/2005, apresentado por AUTORX (2013), e as MIPs 2001-2004 e 2006-2009 é denominado *método RAWS/RAW* para estimação anual de MIPs.

O desempenho do método RAWS/RAW é avaliado por meio de uma versão da MIP 2005 projetada pelo RAW a partir apenas das TRUs e da MIP 2000 divulgada pelo IBGE, acrescida das tabelas estimadas pelo RAWS para este ano. Essa projeção é comparada com a MIP 2005 do IBGE e com a estimada por Guilhoto e Sesso-Filho (2010), através de diversas medidas de precisão.

O artigo é composto por sete seções. Após esta introdução, trata-se na segunda seção de metodologias para a projeção de MIPs e na terceira seção dos procedimentos para a construção das estimativas iniciais. Na quarta seção dá-se a derivação completa do algoritmo RAW, desde a formulação e resolução do problema de minimização de perda de informação até o algoritmo para cômputo das soluções. A avaliação do método RAWS/RAW, pela comparação com a MIP 2005 e com o método de Guilhoto e Sesso-Filho (2005, 2010), é discutida na quinta seção. Na sexta seção, são apresentados os

ajustes para a estimação final da série de matrizes de 2000 a 2009, com a interpolação das MIPs 2001-2004 e projeção das MIPs 2006-2009, considerando a desagregação em 55 setores por 110 produtos. Por fim, a sétima seção conclui o texto.¹

2. Metodologias para projeção da MIP

2.1. Algoritmo RAS para projeção de Matrizes de Insumo-Produto.

Em razão dos custos envolvidos em amplos levantamentos estatísticos necessários para a construção de matrizes de insumo-produto, variadas técnicas de atualização ou projeção por métodos não censitários foram desenvolvidas (algumas são apontadas em Miller e Blair, 2009, cap.7 e Lahr e De Mesnard, 2004). A principal é o método RAS, cuja versão mais conhecida é a proposta por Richard Stone (Stone, 1962; Stone e Brown, 1962), embora métodos similares já fossem empregados previamente (Deming e Stephan, 1940). Conforme Lahr e De Mesnard (2004, p. 117), são duas as maiores vantagens do RAS sobre outros métodos: é um algoritmo relativamente simples, que impede a inversão de sinais dos valores e demanda poucos dados.

O RAS é um método biproporcional de ajuste, um algoritmo em que as linhas e colunas de uma matriz de projeção inicial \mathbf{A} sofrem as operações representadas nas matrizes \mathbf{R} e \mathbf{S} para que sejam respeitados os valores, já conhecidos, de somas nas linhas e nas colunas para o ano projetado.² Seja:

- $\mathbf{X} = [x_{ij}]$ a matriz $n \times m$ que se deseja estimar;
- $\mathbf{q}_n = [q_i]$ o vetor $n \times 1$ com os valores já conhecidos de somas nas linhas de \mathbf{X} ;
- $\mathbf{q}_m = [q_j]$ o vetor $m \times 1$ com os valores já conhecidos de somas nas colunas de \mathbf{X} ;
- $\mathbf{A} = [a_{ij}]$ a matriz $n \times m$ que é a projeção inicial da matriz \mathbf{X} .

¹ A versão final da série de MIPs 2000-2009 acompanha o artigo como um apêndice virtual. As rotinas programadas em MATLAB podem ser obtidas mediante contato com o autor.

² Usualmente, para a projeção de MIPs a literatura aplica o RAS à matriz de coeficientes técnicos intersetoriais. Esta matriz geralmente é designada pela letra \mathbf{A} , por isso o \mathbf{A} da sigla RAS. No presente estudo, optou-se pela aplicação do balanceamento a tabelas de transações, mas a letra \mathbf{A} é mantida para designar as tabelas balanceadas para preservar uma notação mais próxima da literatura de métodos biproporcionais.

O primeiro passo do algoritmo é a correção dos erros das somas nas linhas. Forma-se um vetor $\mathbf{r}_1 = [r_{1i}]$ de dimensão $n \times 1$, com $r_{1i} = \frac{q_i}{\sum_{j=1}^m a_{ij}}$. Esse vetor é diagonalizado ($\hat{\mathbf{r}}_1$) e pré-multiplica a matriz \mathbf{A} , resultando em uma matriz $\mathbf{X}_1 = \hat{\mathbf{r}}_1 \cdot \mathbf{A} = [x_{1ij}]$, em que as somas nas linhas igualam o vetor \mathbf{q}_n .

O segundo passo do algoritmo é a correção dos erros nas somas das colunas. Define-se o vetor $\mathbf{s}_1 = [s_{1j}]$ de dimensão $m \times 1$, com $s_{1j} = \frac{q_j}{\sum_{i=1}^n x_{1ij}}$. A matriz \mathbf{X}_1 é pós-multiplicada por esse vetor diagonalizado ($\hat{\mathbf{s}}_1$), o que leva a uma matriz $\mathbf{X}_2 = \mathbf{X}_1 \cdot \hat{\mathbf{s}}_1 = [x_{2ij}]$, em que as somas nas colunas são iguais ao vetor \mathbf{q}_m .

A correção dos erros nas somas das colunas no segundo passo faz com que ressurgam erros nas somas das linhas. Assim, o terceiro passo repete a operação do primeiro, mas com a matriz do passo anterior: $\mathbf{r}_2 = [r_{2i}]$, $r_{2i} = \frac{q_i}{\sum_{j=1}^m x_{2ij}}$, $\mathbf{X}_3 = \hat{\mathbf{r}}_2 \cdot \mathbf{X}_2$. Os passos são repetidos até que os erros nas linhas e colunas cheguem a um nível aceitável. Supondo que isso ocorra em $2k$ passos, teremos:

$$\mathbf{X} = \mathbf{X}_{2k} = \hat{\mathbf{r}}_k \cdot \mathbf{X}_{2k-2} \cdot \hat{\mathbf{s}}_k = \hat{\mathbf{r}}_k \cdot \hat{\mathbf{r}}_{k-1} \cdot \dots \cdot \hat{\mathbf{r}}_2 \cdot \hat{\mathbf{r}}_1 \cdot \mathbf{A} \cdot \hat{\mathbf{s}}_1 \cdot \hat{\mathbf{s}}_2 \cdot \dots \cdot \hat{\mathbf{s}}_{k-1} \cdot \hat{\mathbf{s}}_k$$

Definindo $\mathbf{R} = \hat{\mathbf{r}} = \hat{\mathbf{r}}_k \cdot \hat{\mathbf{r}}_{k-1} \cdot \dots \cdot \hat{\mathbf{r}}_2 \cdot \hat{\mathbf{r}}_1$ e $\mathbf{S} = \hat{\mathbf{s}} = \hat{\mathbf{s}}_1 \cdot \hat{\mathbf{s}}_2 \cdot \dots \cdot \hat{\mathbf{s}}_{k-1} \cdot \hat{\mathbf{s}}_k$, chega-se à expressão que dá nome ao método:

$$\mathbf{X} = \hat{\mathbf{r}} \cdot \mathbf{A} \cdot \hat{\mathbf{s}} = \mathbf{R} \cdot \mathbf{A} \cdot \mathbf{S} \quad (1)$$

Propriedades teóricas do RAS são derivadas em Bacharach (1970). Se a matriz \mathbf{A} e os vetores \mathbf{q}_n e \mathbf{q}_m são não negativos, mostra-se que pelo RAS pode-se calcular iterativamente a solução do problema de minimização de perda de informação de Kullback e Liebler (1951), respeitando propriedades de existência, unicidade e convergência da solução.

Uma das limitações do RAS original é a condição de que todas as células da matriz negativa devem ser não negativas. Uma solução para este problema foi proposta por Junius e Oosterhaven (2003), o algoritmo *Generalized RAS* (GRAS). Neste algoritmo, a matriz $\mathbf{A} = [a_{ij}]$ é separada em duas, com os elementos positivos em uma e os negativos em outra. As matrizes $\mathbf{A}_+ = [a_{ij}^+]$ e $\mathbf{A}_- = [a_{ij}^-]$ são definidas por:

$$a_{ij}^+ = \begin{cases} 0, & \text{para } a_{ij} \leq 0 \\ a_{ij}, & \text{para } a_{ij} > 0 \end{cases} \quad (2)$$

$$a_{ij}^- = a_{ij} - a_{ij}^+$$

De maneira que $\mathbf{A} = \mathbf{A}_+ + \mathbf{A}_-$ por definição. A solução do problema de minimização de perda de informação é computada numericamente por um algoritmo em que as matrizes \mathbf{A}_+ e \mathbf{A}_- são balanceadas separadamente. Os algoritmos RAWs e RAW derivados nas próximas seções têm o GRAS como ponto de partida, assim por economia de espaço o processo de derivação do GRAS não será apresentado nesta seção. A forma final da matriz estimada pelo GRAS é dada por:

$$\mathbf{X} = \hat{\mathbf{r}} \cdot \mathbf{A}_+ \cdot \hat{\mathbf{s}} + \hat{\mathbf{r}}^{-1} \cdot \mathbf{A}_- \cdot \hat{\mathbf{s}}^{-1} \quad (3)$$

Onde $\hat{\mathbf{r}}^{-1}$ e $\hat{\mathbf{s}}^{-1}$ indicam as inversas das matrizes $\hat{\mathbf{r}}$ e $\hat{\mathbf{s}}$, que são calculadas por um processo similar, mas não idêntico, ao do RAS original.

Evidências empíricas de estudos recentes têm favorecido os métodos GRAS e RAS na comparação com outros. Jackson e Murray (2004), que comparam 10 métodos para dois conjuntos de dados com diferentes medidas de desempenho, constatam que: para dados dos EUA com 23 setores, embora em algumas medidas o RAS seja superado, no geral é o método de melhor desempenho, seguido pelo GRAS; para um exemplo fictício com entradas negativas, o GRAS foi o segundo método de melhor desempenho, apenas ligeiramente inferior a um método não linear, que segundo os autores possivelmente é computacionalmente intratável em problemas de tamanho substancial. Temurshoev, Webb e Yamano (2011) testam oito métodos de estimação de tabelas de recursos e usos, em dados da Holanda e Espanha, e observam que o GRAS não é superado por nenhum dos outros métodos.

Métodos híbridos, que combinam variantes do RAS com dados censitários disponíveis, são promissores para a estimação de matrizes de insumo-produto por unirem as informações disponíveis sobre transações intersetoriais com as propriedades matemáticas interessantes do RAS (Miller e Blair, 2009, cap. 7). Por exemplo, Temurshoev e Timmer (2011) derivam um algoritmo para a estimação conjunta de TRUs, quando apenas dados totais de consumo, valor agregado e demanda final estão disponíveis.

No presente estudo, como foi feito por AUTORX (2013) para estimar as tabelas auxiliares dos anos 2000 e 2005, a partir do GRAS serão derivados métodos híbridos no intuito de utilizar todos os dados disponíveis nas MIPs e TRUs para projetar as MIPs anuais no SCN 2000.

2.2. Estimação da MIP no Brasil

O procedimento de estimação da MIP proposto por Guilhoto e Sesso-Filho (2005) utiliza somente dados da TRU do ano estimado. Os autores propõem os seguintes passos para estimar as tabelas auxiliares:

- i) Eliminar o Dummy Financeiro no processo de cômputo dos coeficientes utilizados para chegar às tabelas auxiliares, o que se aplica apenas aos anos da SCN 1985, já que ele não consta da SCN 2000. Essa etapa deixou de ser necessária, portanto, na revisão do método em Guilhoto e Sesso-Filho (2010);
- ii) Para a alocação de margens e impostos, são calculados coeficientes dividindo cada célula da tabela de usos a preços de consumidor por seu total na linha. Ou seja, tais coeficientes por cada célula, considerando a indexação linha i e coluna j , representam a proporção da produção do produto i que é vendida ao setor ou demanda final j . Os totais por produto das margens de comércio e transporte e dos impostos são multiplicados por esses coeficientes para formar as tabelas auxiliares;
- iii) Para alocar as importações e o imposto de importação, o componente da demanda final “Exportação de Bens e Serviços” recebe o mesmo tratamento conferido à Dummy Financeira, ou seja, ele é excluído da matriz usada para calcular os coeficientes de distribuição. A seguir, o mesmo método adotado para as margens e impostos é aplicado.

O procedimento descrito na etapa (ii) equivale ao primeiro passo do RAS, a correção dos erros nas linhas sem qualquer ajuste nas colunas. Na notação da seção anterior, cada tabela auxiliar X é estimada por um algoritmo do tipo “RA” com uma única etapa, em

que a matriz A é a tabela de usos a preços ao consumidor e a matriz R é o vetor diagonalizado em que cada entrada é igual à relação entre a respectiva margem, imposto ou importação total na linha e a soma na linha de A .

A proposta de Grijó e Berni (2006) para estimar a MIP consiste em projetar as estruturas das tabelas auxiliares do ano mais recente, sujeitas às restrições impostas pela TRU do ano estimado. As tabelas auxiliares da MIP de 2002, ainda sob a referência SCN 1985, são estimadas pela projeção da MIP de 1996, nas seguintes etapas:

- i) Construção de uma matriz de coeficientes para a tabela de usos a preços básicos e de matrizes de *mark-ups* para cada uma das tabelas auxiliares no ano de 1996.³ Para a tabela de usos a preços básicos, a matriz de coeficientes é formada pela divisão elemento a elemento da própria tabela de usos a preços básicos pela tabela de usos a preços ao consumidor. As matrizes de *mark-ups* de cada uma das tabelas auxiliares são formadas pela divisão elemento a elemento da respectiva tabela auxiliar pela tabela de usos a preços básicos;
- ii) Projeção inicial das tabelas do ano 2002. Multiplicando elemento a elemento a matriz de coeficientes de usos a preços básicos de 1996 pela tabela de usos a preços ao consumidor de 2002, tem-se a estimativa inicial da tabela de usos a preços básicos de 2002. Multiplicando esta última tabela elemento a elemento por cada uma das matrizes de *mark-up*, obtém-se as estimativas iniciais das tabelas auxiliares para 2002;
- iii) Intervenções manuais nas projeções iniciais para imputar valores a células que eram nulas em 1996, mas devem ser não nulas em 2002;
- iv) As projeções iniciais são balanceadas por um método RAS adaptado, em que as restrições são: (a) as somas das linhas das tabelas projetadas devem igualar os totais por produto apresentados na TRU de 2002, e (b) a soma das tabelas auxiliares estimadas com a tabela de usos a preços básicos estimada

³ Cabe apontar que os autores agregaram todos os tipos de impostos em uma única tabela e que não separaram as margens e impostos incidentes sobre produtos nacionais e importados, diferentemente do que era feito pelo IBGE nas MIPs da década de 90.

deve ser igual à tabela de usos a preços ao consumidor disponibilizada na TRU 2002.

Pode-se demonstrar que o procedimento de construção das projeções iniciais é na realidade igual para todas as tabelas, tanto as auxiliares quanto a de usos a preços básicos. Basta aplicar a taxa de crescimento entre o ano referência e o estimado da célula v_{ij} da tabela de usos a preços ao consumidor sobre a respectiva célula x_{ij} da tabela auxiliar ou de usos a preços básicos do ano referência.⁴

No balanceamento, como apenas as somas nas linhas de cada uma das matrizes são conhecidas, não é possível construir a matriz \mathbf{S} , que requer o conhecimento prévio dos totais nas colunas. Para contornar esse problema, Grijó e Berni fazem uma projeção inicial da tabela de usos a preços básicos e submetem essa tabela a um balanceamento conjunto com as tabelas auxiliares estimadas.

Note-se que a tabela de usos a preços básicos não carrega qualquer informação nova que já não esteja incorporada às tabelas auxiliares, visto que ela é calculada pelo IBGE como um resíduo da tabela de usos a preços ao consumidor menos a soma das tabelas auxiliares.⁵ Entretanto, com a imposição dessa restrição adicional garante-se que o ajuste das células ao total nas linhas será bem distribuído entre a tabela de usos a preços básicos e as tabelas auxiliares. Sem essa restrição, a tabela de usos a preços básicos é

⁴ Em notação diversa da adotada por Grijó e Berni (2006), seja $\mathbf{V}_0 = [v_{ij0}]$ a tabela de usos a preços ao consumidor do ano referência e $\mathbf{V} = [v_{ij}]$ a mesma tabela para o ano estimado; $\mathbf{X}_0^U = [x_{ij0}^U]$ e $\mathbf{A}^U = [a_{ij}^U]$ são, respectivamente, as tabelas de usos a preços básicos para o ano referência e a estimativa inicial desta tabela no ano estimado; $\mathbf{X}_0^K = [x_{ij0}^K]$ e $\mathbf{A}^K = [a_{ij}^K]$ são as tabelas auxiliares para o ano referência e a estimativa inicial da respectiva tabela no ano estimado, onde K é o indexador da tabela auxiliar. Em Grijó e Berni (2006), os autores derivam as projeções iniciais apresentando como diferentes entre si os procedimentos de estimação das tabelas \mathbf{A}^U e \mathbf{A}^K . As respectivas matrizes de coeficientes e *mark-ups* são $\mathbf{M}^U = [m_{ij}^U]$ e $\mathbf{M}^K = [m_{ij}^K]$, com $m_{ij}^U = \frac{x_{ij0}^U}{v_{ij0}}$ e $m_{ij}^K = \frac{x_{ij0}^K}{x_{ij0}^U}$, e as projeções iniciais para o ano estimado são calculadas por $a_{ij}^U = m_{ij}^U \cdot v_{ij}$ e $a_{ij}^K = m_{ij}^K \cdot a_{ij}^U$. Entretanto, substituindo as definições dos coeficientes e *mark-ups* nas duas últimas equações, chega-se a $a_{ij}^U = x_{ij0}^U \cdot \frac{v_{ij}}{v_{ij0}}$ e $a_{ij}^K = x_{ij0}^K \cdot \frac{v_{ij}}{v_{ij0}}$, ou seja, na realidade o procedimento de cálculo é o mesmo para todas as tabelas e equivale a aplicar, a cada célula das tabelas do ano referência correspondentes às que serão estimadas, a taxa de crescimento observada para a respectiva célula da tabela de usos a preços do consumidor.

⁵ Mesmo a projeção inicial da tabela de usos a preços básicos é por definição idêntica à tabela de usos a preços ao consumidor menos a soma das projeções iniciais das tabelas auxiliares. Tomando as definições da nota de rodapé anterior e dado que $x_{ij0}^U = v_{ij0} - \sum_K x_{ij0}^K$, temos que $a_{ij}^U = x_{ij0}^U \cdot \frac{v_{ij}}{v_{ij0}}$ implica que $a_{ij}^U = (v_{ij0} - \sum_K x_{ij0}^K) \cdot \frac{v_{ij}}{v_{ij0}} \Rightarrow a_{ij}^U = v_{ij0} \cdot \frac{v_{ij}}{v_{ij0}} - \sum_K \frac{v_{ij}}{v_{ij0}} \cdot x_{ij0}^K \Rightarrow a_{ij}^U = v_{ij} - \sum_K a_{ij}^K$.

calculada por resíduo e está sujeita a incongruências, como valores negativos onde apenas se esperariam valores positivos.

No presente artigo, propõe-se um método de projeção anual das MIPs que faz uso dessa restrição de soma das tabelas proposta por Grijó e Berni (2006), mas com melhorias e ajustes adicionais que serão expostos adiante.

3. Definições e projeções iniciais.

3.1. Definições.

Para os anos em que a MIP será projetada, também serão estimadas a tabela de usos a preços básicos (X_U) e todas as tabelas auxiliares: importações (X_{IM}), imposto de importação (X_{TM}), ICMS (X_{TC}), IPI (X_{TP}), outros impostos líquidos de subsídios (X_{TS}), margens de comércio (X_{MC}) e margens de transporte (X_{MT}). Para tanto, será aplicado o método RAW, explicado adiante, sobre uma matriz de estimativas iniciais A , que é composta por um agrupamento das estimativas iniciais A_K de todas as matrizes estimadas, com $K = \{U, IM, TM, TC, TP, TS, MC, MT\}$. Assim são definidas as matrizes A e X como:

$$A = [a_{ij}^k] = \begin{bmatrix} A_U \\ A_{IM} \\ A_{TM} \\ A_{TC} \\ A_{TP} \\ A_{TS} \\ A_{MC} \\ A_{MT} \end{bmatrix}, X = [x_{ij}^k] = \begin{bmatrix} X_U \\ X_{IM} \\ X_{TM} \\ X_{TC} \\ X_{TP} \\ X_{TS} \\ X_{MC} \\ X_{MT} \end{bmatrix}$$

De maneira similar, definem-se os vetores de somas nas linhas e colunas de X :

$$q_n = [q_i^k] = \begin{bmatrix} q_U \\ q_{IM} \\ q_{TM} \\ q_{TC} \\ q_{TP} \\ q_{TS} \\ q_{MC} \\ q_{MT} \end{bmatrix}, q_m = [q_j] = V' \cdot \iota_n$$

Onde:

- $i = \{1, \dots, 8n\}$ é o indexador das linhas;
- $j = \{1, \dots, m\}$ é o indexador das colunas;
- $k \in \mathbf{K} = \{U, IM, TM, TC, TP, TS, MC, MT\}$ indica a respectiva submatriz de \mathbf{A} ou \mathbf{X} , ou subvetor de \mathbf{q}_n ;
- $\mathbf{q}_n = [q_i^k]$ é o vetor de dimensão $8n \times 1$ que reúne os vetores \mathbf{q}_K correspondentes aos totais nas linhas de cada submatriz;
- $\mathbf{q}_m = [q_j]$ é o vetor de dimensão $m \times 1$ correspondente às somas nas colunas da matriz $\mathbf{V} = [v_{ij}]$, a tabela de usos a preços ao consumidor, cuja transposta é designada por \mathbf{V}' ;
- \mathbf{t}_n é um vetor unitário de dimensão $n \times 1$.

3.2. Projeções iniciais: caso geral.

Tome-se $\mathbf{V}_0 = [v_{ij0}]$ como a tabela de usos a preços ao consumidor do ano de referência da projeção, $\mathbf{V} = [v_{ij}]$ a mesma tabela para o ano projetado, $\mathbf{X}_{K0} = [x_{ij0}^k]$ a tabela auxiliar ou de usos a preços básicos do ano referência indexada por $\mathbf{K} = \{U, IM, TM, TC, TP, TS, MC, MT\}$ e $\mathbf{A}_K = [a_{ij}^k]$ como a estimativa inicial da mesma tabela indexada por \mathbf{K} para o ano projetado.

Para cada ano projetado pelo método RAW, o ano de referência é o ano imediatamente anterior, cujos dados são indexados pelo subscrito 0, como em v_{ij0} e x_{ij0}^k .⁶

Seja $\hat{\mathbf{A}}_K = [\hat{a}_{ij}^k]$ o caso geral das estimativas iniciais, definido por:

$$\hat{a}_{ij}^k = \begin{cases} 0 & , \text{para } v_{ij0} = 0 \text{ e } k \neq U \\ v_{ij} & , \text{para } v_{ij0} = 0 \text{ e } k = U \\ x_{ij0}^k \cdot \frac{v_{ij}}{v_{ij0}} & , \text{para } v_{ij0} \neq 0 \end{cases} \quad (4)$$

⁶ Na estimação definitiva dos anos de 2001 a 2004 por interpolação, que será apresentada na subseção 6.1, os anos de referência são fixados em 2000 e 2005, de maneira que o subscrito 0 indicará o ano 2000 e o subscrito 5 o ano 2005.

Para as duas primeiras partes da definição, a célula correspondente da tabela de usos a preços ao consumidor é nula no ano referência ($v_{ij0} = 0$). A distribuição do valor v_{ij} será necessariamente arbitrária entre os α_{ij}^k no balanceamento da restrição W , pois todos os x_{ij0}^k são nulos. Como nestes casos o valor de v_{ij} usualmente é pequeno, optou-se por aloca-lo todo a α_{ij}^U , zerando as células correspondentes das tabelas intermediárias.

O caso geral de fato é o da terceira parte da definição, em que as tabelas auxiliares e a tabela de usos a preços básicos serão estimadas aplicando às células da mesma tabela para o ano referência a taxa de crescimento observada para a célula correspondente na tabela de usos a preços ao consumidor. Conforme demonstrado na seção 2.2, esse procedimento é equivalente à construção de tabelas de *mark-up* proposta por Grijó e Berni (2006).

Algumas intervenções sobre as tabelas estimadas pelo procedimento da equação (4) são necessárias para completar as projeções iniciais, conforme o descrito nas subseções a seguir.

3.3. Coluna “variação de estoque” das tabelas A_U e A_{IM} :

A coluna de variação de estoques é difícil de projetar de maneira consistente, uma vez que suas células são resíduos e os sinais podem variar de ano para ano sem seguir qualquer padrão. Como os métodos de projeção derivados do RAS preservam os sinais das células da matriz de projeção inicial, um sinal invertido que deveria ocorrer apenas no ano base irá se propagar em todos os anos projetados se não forem adotadas intervenções adicionais na projeção inicial. Apenas as tabelas de usos a preços básicos X_U e importações X_{IM} têm entradas não nulas na coluna de variação de estoques, então são essas as tabelas cujas projeções sofrerão ajustes para lidar com esse problema.

As variações de estoques são conhecidas em todos os anos para a tabela de usos a preços de mercado V . A hipótese adotada para a projeção é que em geral os sinais das células da coluna de variação de estoques em X_U e X_{IM} devem ser os mesmos das células equivalentes em V .⁷ Assim, na projeção inicial o ajuste consiste em modificar as

⁷ Dos 110 produtos, no ano 2000 isso não ocorre em apenas três produtos de X_{IM} . No ano de 2005, há um produto de X_U e nove produtos de X_{IM} com sinal na variação de estoque diferente do sinal em V .

células da coluna variação de estoques de X_U e X_{IM} nas quais o sinal for oposto ao sinal da célula equivalente na tabela V .⁸ Nestes casos, o valor absoluto da célula é igualado à unidade e o sinal é invertido.

Ou seja, designando por $\ddot{A}_K = [\ddot{a}_{ij}^k]$ a tabela com as intervenções sobre a variação de estoque, onde $j = E$ designa a coluna “variação de estoque”, teremos:

$$\ddot{a}_{iE}^k = \begin{cases} (-1) \cdot \frac{x_{iE0}^k}{|x_{iE0}^k|}, \text{ para } k \in \{U, IM\} \text{ e } \frac{x_{iE0}^k}{v_{iE0}} < 0 \\ \dot{a}_{iE}^k, \text{ nos demais casos} \end{cases} \quad (5)$$

$$\ddot{a}_{ij}^k = \begin{cases} \ddot{a}_{iE}^k, \text{ para } j = E \\ \dot{a}_{ij}^k, \text{ para } j \neq E \end{cases} \quad (6)$$

3.4. Inversões de sinal ou zeragem na soma da linha.

Intervenções também devem ser feitas quando o total da linha é nulo no ano projetado ($q_i^k = 0$) ou no ano referência ($q_{i0}^k = 0$), ou ainda quando há mudança de sinal no total da linha ($\frac{q_i^k}{q_{i0}^k} < 0$). Se $\dot{A}_K = [\dot{a}_{ij}^k]$ é a tabela com as intervenções relacionadas aos totais nas linhas, define-se:

$$\dot{a}_{ij}^k = \begin{cases} 0, \text{ para } q_i^k = 0 \\ v_{ij} \cdot \frac{q_i^k}{|q_i^k|}, \text{ para } q_{i0}^k = 0 \text{ ou } \frac{q_i^k}{q_{i0}^k} < 0, \text{ exceto se } k = TS \text{ e } j \in J \\ \dot{a}_{ij}^k, \text{ nos demais casos} \end{cases} \quad (7)$$

Onde $J = \{\text{Consumo da administração pública, Consumo das ISFLSF, Variação de estoque}\}$.

Na primeira parte da definição em (7), é estipulado que a célula da projeção inicial deve ser zerada se o total na linha é nulo no ano estimado.

Na segunda parte, quando o total da linha é nulo no ano referência ou sofre inversão de sinal, a referência para a distribuição na linha será a linha correspondente da tabela de usos a preços ao consumidor, $V = [v_{ij}]$. O termo que multiplica v_{ij} é necessário para

⁸ Como será exposto na subseção 4.5, pode ser necessário permitir o sinal contrário para que o algoritmo do RAW convirja. Nesses casos, que serão verificados após a estimação de todas as tabelas, a estimativa inicial da variação de estoques na linha em questão sofrerá nova alteração e as tabelas serão estimadas outra vez.

adequar o sinal da célula, pois todos os elementos da tabela V são não negativos, de maneira que o sinal deve ser invertido quando $q_i^k < 0$.

Quanto à exceção para as três colunas da tabela de outros impostos líquidos de subsídios A_{TS} , ela é necessária porque essas três colunas devem ser mantidas zeradas, em conformidade com as colunas no ano referência. Ela é aplicada somente à tabela A_{TS} porque as outras duas tabelas que também devem ter essas colunas zeradas, A_{TC} e A_{TP} , não tiveram nenhuma linha enquadrada na segunda parte da definição (7) entre 2001 e 2005.

3.5. Tabela de imposto de importação A_{TM} :

Na tabela de imposto de importações, será preservada a proporcionalidade com a tabela de importações, conforme a segunda parte da definição abaixo. Outra regra é usada nas colunas “Exportação de bens e serviços” e “Variação de estoque” da demanda final, que são zeradas para manter a coerência com as tabelas de imposto de importações dos anos de referência estimadas por AUTORX (2013), nas quais estas colunas só têm entradas nulas. O mesmo deve ocorrer nas linhas em que o total de imposto de importação é nulo. Assim, se $\dot{A}_{IM} = [\dot{a}_{ij}^{IM}]$ é a projeção inicial da tabela de importações, para a tabela de imposto de importações a projeção inicial $\ddot{A}_{TM} = [\ddot{a}_{ij}^{TM}]$ será dada por:

$$\ddot{a}_{ij}^{TM} = \begin{cases} 0, & \text{para: } \begin{cases} j = \text{exportações ou variação de estoque} \\ q_i^{TM} = 0 \end{cases} \\ \dot{a}_{ij}^{IM}, & \text{para os demais casos} \end{cases} \quad (8)$$

E a forma geral das tabelas de projeção inicial atualizada com o imposto de importação estimado por (8) será dada por $\ddot{A}_K = [\ddot{a}_{ij}^k]$, onde:

$$\ddot{a}_{ij}^k = \begin{cases} \ddot{a}_{ij}^{TM}, & \text{para } k = TM \\ \dot{a}_{ij}^k, & \text{para } k \neq TM \end{cases} \quad (9)$$

3.6. Valores negativos nas tabelas de margens, A_{MC-} e A_{MT-} :

Após o cálculo conforme a equação (7), as projeções iniciais das tabelas de margens ainda sofrem ajustes nas respectivas linhas de distribuição dos totais nas colunas. Na

tabela de margens de comércio $\mathbf{A}_{MC} = [a_{ij}^{MC}]$, a linha do produto “060101. Comércio” será substituída pela linha com os totais de somas nas colunas da tabela $\mathbf{A}_{MC+} = [a_{ij}^{MC+}]$. O mesmo procedimento é adotado para a linha do produto “070101. Transporte de carga” na tabela de margens de transporte $\mathbf{A}_{MT} = [a_{ij}^{MT}]$, considerando as somas nas colunas de $\mathbf{A}_{MT+} = [a_{ij}^{MT+}]$.

Assim, a partir de (9), $\mathbf{A}_{K+} = [a_{ij}^{k+}]$ e $\mathbf{A}_{K-} = [a_{ij}^{k-}]$ são definidas por:

$$a_{ij}^{k+} = \begin{cases} 0, & \text{para } \ddot{a}_{ij}^k \leq 0 \\ \ddot{a}_{ij}^k, & \text{para } \ddot{a}_{ij}^k > 0 \end{cases} \quad (10)$$

$$a_{ij}^{k-} = \begin{cases} \sum_{i=6n+1}^{7n} -a_{ij}^{MC+}, & \text{para } k = MC \text{ e } i = 060101. \text{ Comércio} \\ \sum_{i=7n+1}^{8n} -a_{ij}^{MT+}, & \text{para } k = MT \text{ e } i = 070101. \text{ Transporte de carga} \\ \ddot{a}_{ij}^k, & \text{para } k \notin \{MC, MT\} \text{ e } \ddot{a}_{ij}^k < 0 \\ 0, & \text{para } \ddot{a}_{ij}^k \geq 0 \end{cases} \quad (11)$$

Pela junção das definições (10) e (11), pode-se enfim definir as projeções iniciais das tabelas $\mathbf{A}_K = [a_{ij}^k]$, com $k \in K = \{U, IM, TC, TP, TS, MC, MT\}$, por:

$$\mathbf{A}_K = \mathbf{A}_{K+} + \mathbf{A}_{K-} \Rightarrow [a_{ij}^k] = [a_{ij}^{k+} + a_{ij}^{k-}] \quad (12)$$

4. Método RAW para projeção anual das MIPs

Nesta seção, apresenta-se o *método RAW* para projeção das MIPs a partir de um ano base em que as tabelas auxiliares estejam disponíveis. A projeção é derivada como resultado de um problema de minimização de perda de informação.

O procedimento aplicado para a projeção é uma ampliação do proposto por Grijó e Berni (2006). Ele será designado por método RAW, ao invés de manter a sigla RAS como fazem Grijó e Berni, uma vez que os valores de soma na coluna são desconhecidos, de maneira que não há uma restrição S no ajuste.

As principais diferenças do método RAW para Grijó e Berni (2006), além do procedimento para construção das projeções iniciais exposto na seção anterior, são: fundamentação como um problema de minimização de perda de informação; generalização do algoritmo para tratar corretamente os valores negativos, conforme o GRAS; tratamento diferenciado para os valores e sinais das colunas de variação de estoques; linhas em que são distribuídos os totais por coluna de margens de comércio e transporte nas respectivas tabelas são atualizadas a cada rodada como somas nas colunas.

4.1. Problema de minimização de perda de informação.

Adota-se a função perda do GRAS com as correções propostas por Lenzen, Wood e Gallego (2007) e Huang, Kobayashi e Tanji (2008), a mesma adotada por Temurshoev e

$$\text{Timmer (2011) e AUTORX (2013): } z_{ij}^k = \begin{cases} \frac{x_{ij}^k}{a_{ij}^k}, & \text{se } a_{ij}^k \neq 0 \\ 1, & \text{se } a_{ij}^k = 0 \end{cases}$$

Para $a_{ij}^k \neq 0$, com $k \in \mathbf{K} = \{U, IM, TM, TC, TP, TS, MC, MT\}$, o problema de minimização de perda de informação é dado por:

$$z_{ij}^k = \underset{z_{ij}^k}{\operatorname{argmin}} \sum_{i=1}^{5n} \sum_{j=1}^m |a_{ij}^k| \cdot (z_{ij}^k \cdot (\ln z_{ij}^k - 1) + 1)$$

sujeito a (13)

$$(I) \quad \sum_{j=1}^m x_{ij}^k = \sum_{j=1}^m a_{ij}^k \cdot z_{ij}^k = q_i^k$$

$$(II) \quad \sum_{k \in \mathbf{K}} x_{ij}^k = \sum_{k \in \mathbf{K}} a_{ij}^k \cdot z_{ij}^k = v_{ij}$$

A este problema de minimização associa-se a seguinte função lagrangeana:

$$\begin{aligned} \mathcal{L} = & \sum_{i=1}^{8n} \sum_{j=1}^m a_{ij}^{k+} \cdot (z_{ij}^{k+} \cdot (\ln z_{ij}^{k+} - 1) + 1) - \sum_{i=1}^{8n} \sum_{j=1}^m a_{ij}^{k-} \cdot (z_{ij}^{k-} \cdot (\ln z_{ij}^{k-} - 1) + 1) \\ & - \sum_{i=1}^{8n} \lambda_i^k \cdot \left(\sum_{j=1}^m a_{ij}^k \cdot z_{ij}^k - q_i^k \right) - \sum_{i=1}^n \sum_{j=1}^m \delta_{ij} \cdot \left(\sum_{k \in K} a_{ij}^k \cdot z_{ij}^k - v_{ij} \right) \end{aligned} \quad (14)$$

As condições de primeira ordem obtidas nas derivações em relação aos multiplicadores λ_i^k e δ_{ij} são as restrições (I) e (II) em (13). As demais condições de primeira ordem são:

i) Para $a_{ij}^k \in A_{K^+}$, com $K^+ = \{U^+, IM^+, TM, TC, TP, TS^+, MC^+, MT^+\}$:

$$\frac{\partial \mathcal{L}}{\partial z_{ij}^{k+}} = 0$$

$$\Rightarrow a_{ij}^{k+} \cdot (\ln z_{ij}^{k+} - 1) + a_{ij}^{k+} - \lambda_i^k \cdot a_{ij}^{k+} - \delta_{ij} \cdot a_{ij}^{k+} = 0$$

$$\Rightarrow a_{ij}^{k+} \cdot (\ln z_{ij}^{k+} - \lambda_i^k - \delta_{ij}) = 0$$

Com $a_{ij}^{k+} \neq 0$, $z_{ij}^{k+} = \frac{x_{ij}^{k+}}{a_{ij}^{k+}}$, $r_i^k = e^{\lambda_i^k}$ e $w_{ij} = e^{\delta_{ij}}$:

$$x_{ij}^{k+} = r_i^k \cdot a_{ij}^{k+} \cdot w_{ij} \quad (15)$$

Note-se que, para TM , TC e TP , o sinal $^+$ não é necessário porque essas tabelas só têm valores não negativos.

ii) Para $a_{ij}^k \in A_{K^-}$, com $K^- = \{U^-, IM^-, TS^-\}$:

$$\frac{\partial \mathcal{L}}{\partial z_{ij}^{k-}} = 0$$

$$\Rightarrow -a_{ij}^{k-} \cdot (\ln z_{ij}^{k-} - 1) - a_{ij}^{k-} - \lambda_i^k \cdot a_{ij}^{k-} - \delta_{ij} \cdot a_{ij}^{k-} = 0$$

$$\Rightarrow a_{ij}^{k-} \cdot (\ln z_{ij}^{k-} + \lambda_i^k + \delta_{ij}) = 0$$

Com $a_{ij}^{k-} \neq 0$, $z_{ij}^{k-} = \frac{x_{ij}^{k-}}{a_{ij}^{k-}}$, $(r_i^k)^{-1} = e^{-\lambda_i^k}$ e $(w_{ij})^{-1} = e^{-\delta_{ij}}$:

$$x_{ij}^{k-} = (r_i^k)^{-1} \cdot a_{ij}^{k-} \cdot (w_{ij})^{-1} \quad (16)$$

Como as tabelas indexadas por TM , TC e TP não têm entradas negativas, não estão incluídas neste caso.

As tabelas A_{MC-} e A_{MT-} , as quais têm elementos não nulos apenas nas linhas em que há a distribuição das margens totais nas colunas, não se enquadram neste caso porque são ajustadas por um procedimento específico.⁹ Os valores dessas linhas são calculados pela soma nas colunas das tabelas A_{MC+} e A_{MT+} , com o sinal invertido:

iii) Para $a_{ij}^k \in A_{K-}$, com $K^- = \{MC^-, MT^-\}$:

$$x_{ij}^{k-} = \begin{cases} -q_j^{MC+} = \sum_{i=6n+1}^{7n} -x_{ij}^{MC+}, & \text{para } k = MC \text{ e } i = 060101. \text{ Comércio} \\ -q_j^{MT+} = \sum_{i=7n+1}^{8n} -x_{ij}^{MT+}, & \text{para } k = MT \text{ e } i = 070101. \text{ Transporte de carga} \\ 0 & , \text{ para os demais casos} \end{cases} \quad (17)$$

Quanto às condições de segunda ordem, como a função lagrangeana (14) é estritamente convexa, existe uma solução única, que é ponto de mínimo da função.

4.2. Cômputo dos multiplicadores r_i^k e w_{ij} .

Para os multiplicadores r_i^k das linhas, as equações (15) e (16) são substituídas na restrição (I) de (13), o que implica em:

$$\sum_{j=1}^m x_{ij}^k = \sum_{j=1}^m r_i^k \cdot a_{ij}^{k+} \cdot w_{ij} + \sum_{j=1}^m (r_i^k)^{-1} \cdot a_{ij}^{k-} \cdot (w_{ij})^{-1} = q_i^k$$

⁹ Foi testada outra versão do método de projeção em que essas linhas de totais de margens de comércio e transporte são tratadas da mesma forma que as outras células: a projeção inicial construída a partir dos dados do ano anterior, as tabelas A_{MC-} e A_{MT-} balanceadas da mesma forma que as outras A_{K-} e foram adicionadas duas restrições adicionais, de que as somas nas colunas das tabelas de margens de comércio e transporte são nulas. Esse método de projeção teve resultados inferiores ao do RAW na projeção para 2005, então não foi descrito para não aumentar a extensão do texto. Esse resultado é coerente com o fato de que essas linhas de totais de margens são calculadas pelo IBGE pela soma na coluna, de maneira que a estrutura de anos anteriores não é relevante.

$$\Rightarrow (r_i^k)^2 \cdot \sum_{j=1}^m a_{ij}^{k+} \cdot w_{ij} - r_i^k \cdot q_i^k + \sum_{j=1}^m a_{ij}^{k-} \cdot (w_{ij})^{-1} = 0$$

Serão três casos na definição dos multiplicadores das linhas. No primeiro caso, quando $q_i^k = 0$, todas as células da linha são nulas, então se define $r_i^k = 1$. Para as linhas de distribuição dos totais nas colunas de margens de comércio e transporte não haverá balanceamento da linha, então também se define $r_i^{MC} = 1$ e $r_i^{MT} = 1$. No segundo caso, para linhas com $q_i^k > 0$, pode haver elementos positivos e negativos se forem linhas das tabelas de usos a preços básicos e de importações, enquanto nas demais tabelas os elementos são não negativos. A equação de segundo grau é resolvida considerando apenas a raiz positiva porque os multiplicadores r_i^k são positivos. No terceiro caso, com $q_i^k < 0$, o que ocorre apenas na tabela de outros impostos menos subsídios, uma vez que as tabelas A_{MC-} e A_{MT-} não são balanceadas, os elementos da linha são todos não positivos, de tal forma que os termos a_{ij}^{k+} são todos nulos e a expressão para r_i^k se torna diferente do caso anterior. As três partes da definição de r_i^k , correspondentes a cada um dos casos, estão em (18):

$$r_i^k = \begin{cases} 1 & , \text{para: } \begin{cases} q_i^k = 0 \\ k = MC \text{ e } i = 060101. \text{Comércio} \\ k = MT \text{ e } i = 070101. \text{Transporte de carga} \end{cases} \\ \frac{q_i^k + \sqrt{(q_i^k)^2 - 4 \cdot \left(\sum_{j=1}^m a_{ij}^{k+} \cdot w_{ij}\right) \cdot \left(\sum_{j=1}^m \frac{a_{ij}^{k-}}{w_{ij}}\right)}}{2 \cdot \sum_{j=1}^m a_{ij}^{k+} \cdot w_{ij}} & , \text{para } q_i^k > 0 \\ \frac{\sum_{j=1}^m \frac{a_{ij}^{k-}}{w_{ij}}}{q_i^k} & , \text{para } q_i^k < 0 \text{ exceto se } k = MC \text{ ou } k = MT \end{cases} \quad (18)$$

Os multiplicadores w_{ij} são obtidos substituindo as equações (15), (16) e (17) na restrição (II) de (13), onde $K^* = \{U, IM, TS\}$, $K^{**} = \{MC, MT\}$, $K = K^* \cup K^{**} \cup \{TM, TC, TP\}$:

$$\sum_{k \in K} x_{ij}^k = \sum_{k \in K} r_i^k \cdot a_{ij}^{k+} \cdot w_{ij} + \sum_{k \in K^*} (r_i^k)^{-1} \cdot a_{ij}^{k-} \cdot (w_{ij})^{-1} + \sum_{k \in K^{**}} x_{ij}^{k-} = v_{ij}$$

$$\Rightarrow w_{ij}^2 \cdot \sum_{k \in K} r_i^k \cdot a_{ij}^{k+} - w_{ij} \cdot \left(v_{ij} - \sum_{k \in K^{**}} x_{ij}^{k-} \right) + \sum_{k \in K^*} (r_i^k)^{-1} \cdot a_{ij}^{k-} = 0$$

Agrupando os termos entre parênteses:

$$C_{1ij} = \sum_{k \in K} r_i^k \cdot a_{ij}^{k+}$$

$$C_{2ij} = v_{ij} - \sum_{k \in K^{**}} x_{ij}^{k-} = \begin{cases} v_{ij} + q_j^{MC+}, & \text{para } i = 060101. \text{ Comércio} \\ v_{ij} + q_j^{MT+}, & \text{para } i = 070101. \text{ Transporte de carga} \\ v_{ij}, & \text{para os demais casos} \end{cases}$$

$$C_{3ij} = \sum_{k \in K^*} (r_i^k)^{-1} \cdot a_{ij}^{k-}$$

$$\Rightarrow w_{ij}^2 \cdot C_{1ij} - w_{ij} \cdot C_{2ij} + C_{3ij} = 0 \quad (19)$$

São três soluções possíveis, apresentadas na definição (20). No primeiro caso, se $C_{1ij} = 0$ e $v_{ij} = 0$, não são necessários ajustes nas respectivas células das tabelas balanceadas, então se define $w_{ij} = 1$. O segundo caso é a resolução da equação em (19) quando $C_{1ij} = 0$ e $v_{ij} \neq 0$, caso em que somente termos negativos são balanceados. O terceiro caso, por fim, é a solução da equação (19) para $C_{1ij} \neq 0$, com a raiz negativa desconsiderada.

$$w_{ij} = \begin{cases} 1 & , \text{ para } C_{1ij} = 0 \text{ e } v_{ij} = 0 \\ \frac{C_{3ij}}{C_{2ij}} & , \text{ para } C_{1ij} = 0 \text{ e } v_{ij} \neq 0 \\ \frac{C_{2ij} + \sqrt{C_{2ij}^2 - 4 \cdot C_{1ij} \cdot C_{3ij}}}{2 \cdot C_{1ij}} & , \text{ para } C_{1ij} \neq 0 \end{cases} \quad (20)$$

4.3. Forma matricial das soluções.

Para exprimir em forma matricial as soluções do método RAW de projeção, tomam-se as mesmas notações de AUTORX (2013):

- Os vetores e matrizes de multiplicadores são designados por $\mathbf{r}_K = [r_i^k]$ e $\mathbf{W} = [w_{ij}]$.

- Vetores unitários serão designados por ι_n ou ι_m , com dimensão de uma coluna por n ou m linhas, respectivamente.

- Sejam duas matrizes quaisquer $Y_A = [y_{ij}^A]$ e $Y_B = [y_{ij}^B]$ de mesma dimensão, as operações elemento a elemento de matrizes e vetores serão designadas pelas notações:

$Y_A \odot Y_B = [y_{ij}^A \cdot y_{ij}^B]$ é o produto elemento a elemento, ou produto de Hadamard, de Y_A e Y_B .

$Y_A \oslash Y_B = [y_{ij}^A / y_{ij}^B]$ é a divisão elemento a elemento de Y_A e Y_B , onde $y_{ij}^B \neq 0$, para todo i e j .

$(Y_A)^{\circ 2} = [y_{ij}^{A^2}] = [y_{ij}^A \cdot y_{ij}^A]$ é o produto elemento a elemento de Y_A pelo próprio Y_A .

$(Y_A)^{\circ 1/2} = [\sqrt{y_{ij}^A}]$ é a raiz quadrada elemento a elemento de Y_A .

- Como é usual, a diagonalização de um vetor \mathbf{y} é designada por $\hat{\mathbf{y}}$ e a transposição e a inversão de uma matriz são designadas por Y' e Y^{-1} , respectivamente.

Assim, com $K = \{U, IM, TM, TC, TP, TS, MC, MT\}$, a forma matricial da definição dos multiplicadores das linhas em (18) é dada por:

$$r_K = \begin{cases} 1 & , \text{ para: } \begin{cases} q_i^k = 0 \\ k = MC \text{ e } i = 060101. \text{ Comércio} \\ k = MT \text{ e } i = 070101. \text{ Transp.} \end{cases} \\ \mathbf{B}_1 \oslash (2 \cdot ((\mathbf{A}_{k+} \odot \mathbf{W}) \cdot \iota_m)) & , \text{ para } q_i^{TS} > 0 \\ ((\mathbf{A}_{k-} \oslash \mathbf{W}) \cdot \iota_m) \oslash \mathbf{q}_k & , \text{ para } q_i^{TS} < 0 \end{cases} \quad (21)$$

Onde:

$$\mathbf{B}_1 = \mathbf{q}_k + \left((\mathbf{q}_k)^{\circ 2} - 4 \cdot ((\mathbf{A}_{k+} \odot \mathbf{W}) \cdot \iota_m) \odot ((\mathbf{A}_{k-} \oslash \mathbf{W}) \cdot \iota_m) \right)^{\circ 1/2}$$

A forma matricial dos multiplicadores de soma das tabelas em (20), por sua vez, é dada por:

$$\mathbf{W} = \begin{cases} 1 & , \text{ para } C_{1ij} = 0 \text{ e } v_{ij} = 0 \\ \mathbf{C}_3 \oslash \mathbf{C}_2 & , \text{ para } C_{1ij} = 0 \text{ e } v_{ij} \neq 0 \\ \left(\mathbf{C}_2 + ((\mathbf{C}_2)^{\circ 2} - 4 \cdot \mathbf{C}_1 \odot \mathbf{C}_3)^{\circ 1/2} \right) \oslash (2 \cdot \mathbf{C}_1) & , \text{ para } v_{ij} \neq 0 \end{cases} \quad (22)$$

Onde:

$$C_1 = \hat{r}_U \cdot A_{U+} + \hat{r}_{IM} \cdot A_{IM+} + \hat{r}_{TM} \cdot A_{TM} + \hat{r}_{TC} \cdot A_{TC} + \hat{r}_{TP} \cdot A_{TP} + \hat{r}_{TS} \cdot A_{TS+} + \\ + \hat{r}_{MC} \cdot A_{MC+} + \hat{r}_{MT} \cdot A_{MT+}$$

$$C_2 = V - X_{MC-} - X_{MT-}$$

$$C_3 = \hat{r}_U^{-1} \cdot A_{U-} + \hat{r}_{IM}^{-1} \cdot A_{IM-} + \hat{r}_{TS}^{-1} \cdot A_{TS-}$$

Por fim, as estimativas finais das tabelas auxiliares e de usos a preços básicos X_K , coerentes com as equações de (15) a (17), serão dadas por:

$$X_U = \hat{r}_U \cdot A_{U+} \odot W + \hat{r}_U^{-1} \cdot (A_{U-} \oslash W) \quad (23)$$

$$X_{IM} = \hat{r}_{IM} \cdot A_{IM+} \odot W + \hat{r}_{IM}^{-1} \cdot (A_{IM-} \oslash W) \quad (24)$$

$$X_{TM} = \hat{r}_{TM} \cdot A_{TM} \odot W \quad (25)$$

$$X_{TC} = \hat{r}_{TC} \cdot A_{TC} \odot W \quad (26)$$

$$X_{TP} = \hat{r}_{TP} \cdot A_{TP} \odot W \quad (27)$$

$$X_{TS} = \hat{r}_{TS} \cdot A_{TS+} \odot W + \hat{r}_{TS}^{-1} \cdot (A_{TS-} \oslash W) \quad (28)$$

$$X_{MC} = X_{MC+} + X_{MC-} \quad (29)$$

Onde:

$$X_{MC+} = \hat{r}_{MC} \cdot A_{MC+} \odot W$$

$$X_{MC-} = \begin{cases} -\iota_n \cdot X_{MC+}, & \text{na linha referente a } i = 060101. \text{Comércio} \\ 0 & \text{, nas demais células} \end{cases}$$

$$X_{MT} = X_{MT+} + X_{MT-} \quad (30)$$

Onde:

$$X_{MT+} = \hat{r}_{MT} \cdot A_{MT+} \odot W$$

$$X_{MT-} = \begin{cases} -\iota_n \cdot X_{MT+}, & \text{na linha referente a } i = 070101. \text{Transporte de carga} \\ 0 & \text{, nas demais células} \end{cases}$$

4.4. Algoritmo para o cômputo das soluções.

Os valores das soluções X_{MC-} e X_{MT-} e dos multiplicadores r_K e W , os quais são interdependentes, serão calculados pelo seguinte algoritmo, programado em MATLAB:

- *Passo 0*: obtenção das estimativas iniciais A_K das tabelas de passagem e de usos a preços básicos, de acordo com o exposto na seção 3;

- *Passo 1*: a primeira rodada será o balanceamento das linhas, então considera-se $w_{ij} = 1$, para todas as células de $W = [w_{ij}]$. Os multiplicadores r_K serão calculados pela definição (21);

- *Passo 2*: com os r_{MC} e r_{MT} calculados no passo anterior, são computadas estimativas intermediárias das tabelas X_{MC-} e X_{MT-} , conforme as equações (29) e (30). Estas estimativas, assim como os demais multiplicadores r_K calculados no passo anterior, serão empregadas no cálculo de W pela equação (22);

- *Passos de 3 a N*: a cada etapa, os multiplicadores r_K e W e as tabelas X_{MC-} e X_{MT-} são estimados novamente tomando como dadas as estimativas dos passos anteriores. Este processo continua até que o maior elemento da matriz $DIF = W_{(passo N)} - W_{(passo N-1)}$ seja menor que 10^{-6} , que equivale a R\$ 1,00, uma vez que os valores das tabelas estão designados em milhões de Reais;

- *Passo N+1*: as estimativas finais dos multiplicadores r_K e W são substituídas nas equações (23) a (30) para a obtenção das estimativas X_K das tabelas auxiliares;

- *Passo N+2*: conforme o procedimento a ser exposto na próxima subseção, se não houver convergência do algoritmo para alguma das linhas, sinais são invertidos na coluna variação de estoques e toda a estimação até o passo $N+1$ é repetida;

- *Passo N+3*: toma-se o ano estimado como referência para o ano seguinte e os passos de 0 a $N+2$ são refeitos para este ano, o que se repete até que todos os anos estejam estimados.

Para a projeção de teste do método nos anos de 2001, 2002, 2003, 2004 e 2005, a convergência foi obtida, respectivamente, com $N = 234, 96, 45, 41$ e 86 . Nos anos de 2001 e 2002, a estimação foi repetida para resolver o problema mencionado no passo $N+2$.

4.5. Ajustes nas colunas de variação de estoques.

Um dos problemas do método RAS e seus derivados é que o algoritmo pode não convergir quando as linhas e colunas balanceadas são esparsas, isto é, se possuem poucas células não nulas. Como expõem Miller e Blair (2009, p. 335), nestes casos pode ser impossível que estas células atendam simultaneamente as restrições de somas nas linhas e colunas. No método RAW, esse problema pode surgir se há a combinação de linhas esparsas relacionadas à restrição R com somas esparsas de células relativas à restrição W . Em particular, como foi imposto na projeção inicial que as células não nulas de variação de estoque das tabelas estimadas devem ter os mesmos sinais das células equivalentes na tabela de usos a preços ao consumidor, esse problema pode ocorrer se for necessário que o sinal de alguma dessas células de variação de estoque seja o oposto.

Para as projeções de teste do método RAW de 2001 a 2005 a partir do ano 2000, isso ocorreu somente para o produto “020301. Carvão mineral”, nos anos de 2001 e 2002. Na tabela intermediária de importações do ano 2000, a linha deste produto tem apenas duas células não nulas, ambas com valores positivos: nas compras do setor “0321. Fabricação de aço e derivados” e na coluna variação de estoques. A mesma estrutura é reproduzida na projeção inicial de 2001. Ou seja, o total de importações deste produto deve ser alocado somente nessas duas células, somente com sinais positivos. O problema ocorre porque tais células também estão sujeitas à restrição W , mas a soma dos valores correspondentes a essas duas células na tabela de usos a preços ao consumidor é menor que o total na linha do produto na tabela de importações. Ou seja, mesmo que todo o valor destas duas células na tabela de usos a preços ao consumidor seja alocado somente na tabela de importações para respeitar a restrição W , ainda assim a soma nesta linha da tabela de importações será inferior ao total imposto pela restrição R . Com isso, o algoritmo não converge e os erros são propagados às outras tabelas também, além da tabela de importações.

A solução adotada consistiu em permitir valores com sinais invertidos na coluna variação de estoques, o que torna possível o atendimento simultâneo das restrições R e

W . Uma vez que esse problema pode surgir em diferentes linhas quando o método for aplicado para outros anos ou com tabelas de projeção inicial A alternativas, foi desenvolvido um algoritmo para a identificação e correção automática do problema.

Como dentre as tabelas estimadas apenas a de usos a preços básicos e a de importações têm valores não nulos nas colunas de variação de estoques, o algoritmo é desenhado para intervir no balanceamento destas tabelas, quando necessário. Deve-se ressaltar que, uma vez que as estimações de todas as tabelas estão interligadas pela restrição W , erros na estimação de uma delas irão se propagar pelas outras. Por isso, pode ser necessário que a intervenção para corrigir o problema seja feita em outra tabela, não naquela em que o erro é observado, a princípio. No caso em questão, como é necessário aumentar a soma dos valores alocados nas duas células da linha “020301. Carvão mineral” da tabela de importações, a intervenção ocorrerá na tabela de usos a preços básicos, com a inversão do sinal da variação de estoque de positivo para negativo. Apenas com essa troca de sinal, de maneira automática o algoritmo RAW irá elevar o valor da variação de estoque na tabela de importações no montante necessário para atender a restrição R na linha, enquanto um valor correspondente será lançado com sinal negativo na variação de estoques da tabela de usos a preços básicos para atender a restrição W .

De forma geral, se uma das duas tabelas X_U ou X_{IM} tem um erro de estimação na linha relacionado ao sinal da variação de estoque, irá provocar outro erro na linha equivalente da outra tabela de magnitude semelhante e com sinal oposto. A correção dependerá de qual tabela deve ter a soma na linha elevada e qual deve ter a soma na linha reduzida. Cabe destacar que, em razão das intervenções nas projeções iniciais das colunas de variação de estoque descritas na seção 3, necessariamente serão iguais os sinais das células correspondentes nas colunas de variação de estoque de X_U e X_{IM} . Sofrerão os ajustes as linhas em que o erro para o valor conhecido de soma na linha, em valor absoluto, for maior que 0,5, que é a margem de arredondamento do IGBE. Na linha em que houver o problema, a intervenção dependerá do sinal das células de variação de estoque em X_U e X_{IM} : se ambas forem positivas, será igualada a -1 aquela cuja soma da linha deve ser diminuída; se ambas forem negativas, será igualada a 1 aquela cuja soma precisa ser aumentada. A seguir, todo o balanceamento do RAW é repetido, substituindo as células alteradas nas projeções iniciais. As células de variação de

estoque que sofrem as intervenções são igualadas a I ou $-I$ para que, no ajuste da linha pelo RAW, recebam uma alocação pequena de valores, apenas para tornar o ajuste possível.

O algoritmo que corrige esse problema foi integrado ao exposto na subsecção anterior e é acionado no *passo* $N+2$. Suas etapas são apresentadas a seguir:

- *Passo 0*: após a estimação das tabelas \mathbf{X}_K no *passo* $N+1$ do algoritmo RAW, explicado na subsecção 4.4, são calculados vetores de erros nas linhas para as tabelas de usos a preços básicos e importações, respectivamente $\mathbf{e}_U = [e_i^U]$ e $\mathbf{e}_{IM} = [e_i^{IM}]$, onde $\mathbf{e}_K = \mathbf{X}_K \cdot \mathbf{t}_m - \mathbf{q}_K$ para $K = \{U, IM\}$;

- *Passos de 1 até M*: sejam $\mathbf{x}_E^U = [x_{iE}^U]$ e $\mathbf{x}_E^{IM} = [x_{iE}^{IM}]$ as colunas de variação de estoque das tabelas \mathbf{X}_U e \mathbf{X}_{IM} , respectivamente, da mesma forma que $\mathbf{a}_E^U = [a_{iE}^U]$ e $\mathbf{a}_E^{IM} = [a_{iE}^{IM}]$ são as colunas correspondentes nas projeções iniciais destas tabelas. As projeções iniciais dessas colunas serão alteradas conforme a seguinte regra, aplicada a cada linha $i = \{1, \dots, n\}$:

- Para $e_i^U > 0,5$ e $e_i^{IM} < -0,5$:
 - Se $x_{iE}^U > 0$ e $x_{iE}^{IM} > 0 \Rightarrow$ define-se $a_{iE}^U = -1$;
 - Se $x_{iE}^U < 0$ e $x_{iE}^{IM} < 0 \Rightarrow$ define-se $a_{iE}^{IM} = 1$;

- Para $e_i^U < -0,5$ e $e_i^{IM} > 0,5$:
 - Se $x_{iE}^U > 0$ e $x_{iE}^{IM} > 0 \Rightarrow$ define-se $a_{iE}^{IM} = -1$;
 - Se $x_{iE}^U < 0$ e $x_{iE}^{IM} < 0 \Rightarrow$ define-se $a_{iE}^U = 1$;

- Para demais casos, a_{iE}^U e a_{iE}^{IM} não são alterados.

- *Passo M+1*: todo o processo de projeção pelo algoritmo RAW é repetido, substituindo nas projeções iniciais \mathbf{A}_U e \mathbf{A}_{IM} as células alteradas nos passos de 1 a M .

4.6. Matriz inversa de Leontief.

Com a tabela X_U estimada, é seguido o procedimento das notas técnicas da MIP 2000/2005 (IBGE, 2008) para a obtenção da matriz inversa de Leontief. Devem ser assumidas hipóteses de correspondência entre produtos e setores para a passagem da TRU, que relaciona produtos a setores, para a MIP, que relaciona setores a setores.

Tomem-se as seguintes definições de matrizes e vetores:

- X_R é a tabela de recursos, com os setores nas colunas e produtos nas linhas;
- X_r é a tabela X_R apenas com as colunas dos setores de atividade, de dimensão $n \times n$;
- X_u é a tabela de usos a preços básicos X_U apenas com as colunas dos setores de atividade, sem as colunas da demanda final, de dimensão $n \times n$;
- g é o vetor de produção total dos setores de atividade, de dimensão $n \times 1$, que é igual à transposição da linha de somas nas colunas da tabela X_u ;
- q_U , o vetor de dimensão $n \times 1$ de demanda total de cada produto a preços básicos, dado pela soma nas linhas de X_U , é igual ao vetor de somas nas linhas de X_r , que correspondem à produção total de cada produto;
- I é a matriz identidade de dimensão $n \times n$.

Sumariamente, as etapas para o cômputo da MIP são:

- i) Para a correspondência entre produtos e setores na produção, adota-se a hipótese de tecnologia do setor, segundo a qual as informações disponíveis são sobre a estrutura de insumos das atividades, não dos produtos. Calcula-se então a matriz B dos coeficientes técnicos produto por atividade: $B = X_u \cdot \hat{g}^{-1}$, onde \hat{g}^{-1} é a inversa do vetor g diagonalizado;
- ii) Para a alocação da demanda por cada produto aos setores, segue-se a proporção da participação de cada setor na produção do produto. Isso é feito pelo cálculo da matriz D de *market-share*: $D = X_r' \cdot \hat{q}_U^{-1}$, onde X_r' é a transposta da matriz X_r ;
- iii) A matriz K de coeficientes técnicos intersetoriais é dada por: $K = D \cdot B$;

iv) Por fim, a matriz de insumo-produto ou matriz inversa de Leontief L é calculada por: $L = (I - K)^{-1}$.

5. Desempenho do método de estimação RAWS/RAW.

Nesta seção, o desempenho do método de estimação RAWS/RAW é avaliado com a MIP 2005 projetada pelo RAW a partir da MIP 2000, que inclui as tabelas auxiliares estimadas pelo RAWS em AUTORX (2013). A estimativa pelo método RAWS/RAW é comparada com a MIP 2005 divulgada pelo IBGE e com a estimada por Guilhoto e Sesso-Filho (2010) para o mesmo ano.¹⁰

Seguindo Jensen (1980), Miller e Blair (2009, p. 323-327) classificam em duas formas a precisão na estimação de MIPs: partitiva e holística. A precisão partitiva refere-se a medidas de desempenho na estimação de células individuais, aplicadas à matriz de coeficientes intersetoriais K . A precisão holística diz respeito ao desempenho nos usos mais frequentes da matriz inversa de Leontief L , como o cálculo de multiplicadores de impacto.

Ambos os tipos de medidas são aplicados em um ano em que a matriz é conhecida e pode ser comparada com as estimativas da mesma matriz produzidas pelos métodos sob teste. As chamadas medidas partitivas resultam da aplicação, sobre a matriz conhecida em confronto com a estimada, de procedimentos para calcular a distância entre duas matrizes. Variam conforme os critérios adotados para aferir o erro na estimação de cada célula e ponderar os erros de todas as células. Já as medidas holísticas, que são aplicadas especificamente à matriz inversa de Leontief, comparam multiplicadores e outras métricas de impacto intersetoriais computadas para a MIP conhecida e a estimada. De acordo com Miller e Blair, as medidas partitivas são mais rigorosas, mas um bom desempenho nas medidas holísticas pode ser suficiente para fins práticos.

No presente estudo, serão aplicados os dois tipos de medidas, considerando todas as informações disponíveis para 2005 do conjunto de tabelas que compõem a MIP. Esse conjunto de informações empregado na mensuração de desempenho é mais abrangente

¹⁰ A MIP 2005 calculada pelo método de Guilhoto e Sesso-Filho foi obtida no site do Nereus, <<http://www.usp.br/nereus/?fontes=dados-matrizes>>.

que o considerado em outras estimativas das MIPs brasileiras. Grijó e Berni (2006) não apresentam medidas de avaliação do seu método de projeção. Em Guilhoto e Sesso-Filho (2005, 2010), é verificada a precisão holística, pela comparação de medidas de impacto extraídas da matriz L .

Nesta seção, algumas das medidas de precisão holística adotadas por Guilhoto e Sesso-Filho são calculadas para a matriz L de 2005 estimada pelo método RAWS/RAW. Medidas de precisão partitiva serão aplicadas sobre tabelas e vetores disponibilizados pelo IBGE para 2005. As tabelas são a de usos a preços básicos X_U e a auxiliar de importações X_{IM} . Os vetores são somas de colunas das tabelas auxiliares que podem ser extraídas da tabela H apresentada por AUTORX (2013): soma total dos três tipos de impostos (q_m), margens de comércio (h_{MC}) e margens de transporte mais impostos incidentes sobre o produto “070101. Transporte de carga” (h_{MT}).

5.1. Medidas de precisão holística.

Para aferir a precisão holística do método de estimação RAWS/RAW, foram aplicadas à matriz inversa de Leontief L de 2005 as medidas de ligações intersetoriais que são adotadas em ambos os artigos de Guilhoto e Sesso-Filho (2005, 2010): multiplicadores de produção simples e índices de Rasmussen-Hirschman, para frente e para trás. Para cada uma dessas medidas, resulta um vetor com 55 componentes, correspondente ao número p de setores de produção. Tais vetores calculados para a matriz L estimada são comparados com os correspondentes da matriz L divulgada pelo IBGE, por índices de correlação de Pearson e Spearman. As correlações são confrontadas com as mesmas medidas calculadas para a MIP 2005 de Guilhoto e Sesso-Filho (2010).

i) Multiplicador de produção simples.

O multiplicador de produção simples - conforme a nomenclatura de Miller e Blair (2009, p. 245), ou multiplicador de produção tipo I, segundo Guilhoto e Sesso-Filho - de cada setor da matriz é dado pela soma de todos os elementos da matriz L na coluna correspondente. Representa o aumento de demanda na economia como um todo decorrente da expansão em uma unidade na demanda final do setor. Se $L = [l_{ij}]$ é a

matriz inversa de Leontief, o multiplicador de produção simples do setor j , com $j = \{1, \dots, p\}$, é dado por:

$$m_j = \sum_{i=1}^p l_{ij} \quad (31)$$

De maneira que $\mathbf{m} = [m_j]$ é o vetor de multiplicadores de produção simples, de dimensão $p \times 1$.

ii) *Índice de ligações intersetoriais de Rasmussen-Hirschman para trás.*

Mede a capacidade de estímulo no resto da economia da demanda final do setor assim como o multiplicador de produção, mas é normalizado pela média dos elementos de \mathbf{L} , de maneira que os setores com índices maiores que a unidade têm capacidade de encadeamento acima da média e aqueles com índices menores que a unidade têm capacidade abaixo da média. Os índices RH para trás são calculados dividindo o multiplicador de produção pelo número de setores e dividindo novamente pela média dos elementos da matriz de Leontief. Ou seja, se $\mathbf{rht} = [rht_j]$ é o vetor com os índices RH para trás correspondentes a todos os setores de atividades $j = \{1, \dots, p\}$, temos:

$$rht_j = \left(\frac{m_j}{p}\right) / \left(\sum_{j=1}^p \frac{m_j}{p^2}\right)$$

Simplificando e substituindo a expressão (31), chega-se a:

$$rht_j = \frac{\sum_{i=1}^p l_{ij}}{\sum_j \sum_i \frac{l_{ij}}{p}} \quad (32)$$

iii) *Índice de ligações intersetoriais de Rasmussen-Hirschman para frente.*

Estão relacionados ao papel estratégico do setor como ofertante de recursos ao resto da economia. Os índices RH para frente são computados apenas substituindo na expressão (100) a soma na coluna pela soma dos elementos na linha da matriz de Leontief. Assim,

se $\mathbf{rhf} = [rhf_i]$ é o vetor com os índices RH para frente correspondentes a todos os setores de atividades $i = \{1, \dots, p\}$, temos:

$$rhf_i = \frac{\sum_{j=1}^p l_{ij}}{\sum_i \sum_j \frac{l_{ij}}{p}} \quad (33)$$

As correspondências entre os multiplicadores de produção e índices RH de ligações intersetoriais calculados para a matriz estimada e os calculados para a matriz do IBGE são avaliadas pelos índices de correlação de Pearson e Spearman, como fazem Guilhoto e Sesso-Filho (2005, 2010). O coeficiente de correlação de Pearson mede a correspondência entre duas séries de valores, enquanto o coeficiente de correlação de Spearman mede a correspondência do ordenamento entre as duas séries de valores.

A Tabela 1 apresenta, para os dois métodos considerados, os índices de correlação de Pearson e Spearman dos índices RH calculados pela MIP do IBGE contra os calculados pelo método em questão. Os índices de correlação dos multiplicadores de produção tipo I não são apresentados porque, por construção, são iguais aos dos índices RH para trás.

Tabela 1: Comparação de diferentes métodos na estimação da matriz inversa de Leontief (L), 2005.

Métodos	RH para trás		RH para frente	
	Pearson	Spearman	Pearson	Spearman
RAWS/RAW	0,992	0,984	0,9997	0,998
Guilhoto e Sesso-Filho	0,957	0,946	0,9948	0,984

Fontes: IBGE (2007, 2008) e Guilhoto e Sesso-Filho (2010). Elaboração própria.

Da tabela nota-se que, de acordo com as quatro medidas consideradas, a precisão holística do método RAWS/RAW é superior à do método de Guilhoto e Sesso-Filho. A diferença é mais acentuada no índice RH para trás, cujas medidas de correlação são as mesmas do multiplicador de produção simples. No índice RH para frente, em que as diferenças dos dois métodos são menores, as correlações para o método RAWS/RAW só não são arredondadas para a unidade com três ou quatro casas decimais.

Apesar das diferenças, pode-se perceber que nas medidas holísticas a diferenciação entre os métodos é pequena, uma vez que a precisão holística já era alta no método de

Guilhoto e Sesso-Filho. Na próxima subseção, com a apresentação dos métodos partitivos, mais elementos de comparação são considerados.

5.2. Medidas de precisão partitiva: tabelas de usos a preços básicos X_U e de importações X_{IM} .

Para o ano de 2005, o IBGE divulga junto à MIP as tabelas X_U e X_{IM} , às quais serão utilizadas para a aferição da precisão partitiva do método de estimação. Serão empregadas medidas de posição e de dispersão para comparar os métodos.

Dos cinco critérios de posição adotados, os três primeiros são apresentados em Miller e Blair (2009, p. 334) como medidas de distância entre duas matrizes. Os dois últimos, que são médias ponderadas em termos percentuais desenvolvidas a partir das medidas anteriores, serão enfatizados nos comentários porque sua interpretação é mais imediata.

Em cada uma das medidas, emprega-se a notação:

- $X_K = [x_{ij}^k]$ é a tabela de dimensão $n \times m$ divulgada pelo IBGE, com $k \in K = \{U, IM\}$;

- $\hat{X}_K = [\hat{x}_{ij}^k]$ é a tabela de dimensão $n \times m$ estimada, com $k \in K = \{U, IM\}$;

As medidas de posição adotadas são:

i) *Desvio absoluto médio (DAM)*.

É a média aritmética dos desvios, em valores absolutos, entre as células correspondentes das tabelas X_K e \hat{X}_K . É calculado por:

$$DAM = \sum_{i=1}^n \sum_{j=1}^m \frac{|\hat{x}_{ij}^k - x_{ij}^k|}{n \cdot m} \quad (34)$$

ii) *Desvio absoluto ponderado (DAP)*.

Essa medida é a média ponderada dos desvios absolutos, em que os fatores de ponderação são iguais à proporção entre o valor absoluto da célula correspondente em \mathbf{X}_K e a soma dos valores absolutos de todas as células desta matriz. Assim:

$$DAP = \sum_{i=n}^n \sum_{j=1}^m \frac{|x_{ij}^k|}{\sum_i \sum_j |x_{ij}^k|} \cdot |\hat{x}_{ij}^k - x_{ij}^k| \quad (35)$$

iii) *Desvio absoluto percentual médio (DAPM).*

Trata-se da média aritmética dos desvios proporcionais, em valores absolutos, das células da tabela $\hat{\mathbf{X}}_K$ em relação às correspondentes em \mathbf{X}_K .¹¹ Sua fórmula de cálculo é:

$$DAPM = \sum_{i=n}^n \sum_{j=1}^m \frac{1}{n \cdot m} \cdot \left| \frac{\hat{x}_{ij}^k - x_{ij}^k}{x_{ij}^k} \right| \quad (36)$$

iv) *Desvio absoluto percentual ponderado (DAPP).*

Definido como a média ponderada dos valores absolutos dos desvios proporcionais, com a ponderação dada pela razão entre o valor absoluto da célula em \mathbf{X}_K e a soma dos valores absolutos de todas as células de \mathbf{X}_K . Ou seja,

$$\sum_{i=n}^n \sum_{j=1}^m \frac{|x_{ij}^k|}{\sum_i \sum_j |x_{ij}^k|} \cdot \left| \frac{\hat{x}_{ij}^k - x_{ij}^k}{x_{ij}^k} \right|$$

Que, simplificada, se torna:

$$DAPP = \sum_{i=n}^n \sum_{j=1}^m \frac{|\hat{x}_{ij}^k - x_{ij}^k|}{\sum_i \sum_j |x_{ij}^k|} \quad (37)$$

Note-se que a mesma medida é obtida dividindo o desvio absoluto médio (DAM) pela média dos valores absolutos das células de \mathbf{X}_K :

$$DAPP = \frac{DAM}{\text{média}(|x_{ij}^k|)} = \left(\sum_{i=n}^n \sum_{j=1}^m \frac{|\hat{x}_{ij}^k - x_{ij}^k|}{n \cdot m} \right) / \left(\sum_{i=n}^n \sum_{j=1}^m \frac{|x_{ij}^k|}{n \cdot m} \right) = \sum_{i=n}^n \sum_{j=1}^m \frac{|\hat{x}_{ij}^k - x_{ij}^k|}{\sum_i \sum_j |x_{ij}^k|}$$

¹¹ Para as células em que $x_{ij}^k = 0$, foi considerado $x_{ij}^k = 1$ no denominador do erro proporcional.

v) *Desvio absoluto ponderado total (DAPT).*

Por fim, esta medida é obtida pela razão da média ponderada dos desvios absolutos, que é a medida DAP, sobre a média ponderada das células da tabela X_R , em valor absoluto.

Ou seja:

$$DAPT = \left(\sum_{i=n}^n \sum_{j=1}^m \frac{|x_{ij}^k|}{\sum_i \sum_j |x_{ij}^k|} \cdot |\hat{x}_{ij}^k - x_{ij}^k| \right) / \left(\sum_{i=n}^n \sum_{j=1}^m \frac{|x_{ij}^k|}{\sum_i \sum_j |x_{ij}^k|} \cdot |x_{ij}^k| \right) \quad (38)$$

A Tabela 2 apresenta as cinco medidas de posição, calculadas para os métodos RAWS/RAW e Guilhoto e Sesso-Filho nas tabelas X_U e X_{IM} .

Tabela 2: Medidas de posição para a comparação de métodos na estimação das tabelas de usos a preços básicos (X_U) e de importações (X_{IM}), 2005.

Tabelas/Métodos	DAM	DAP	DAPM	DAPP	DAPT
<i>Usos a preços básicos</i>					
RAWS/RAW	12	138	5%	2%	0,3%
Guilhoto e Sesso-Filho	50	1772	29%	9%	4%
<i>Importações</i>					
RAWS/RAW	7	357	44%	19%	9%
Guilhoto e Sesso-Filho	22	1046	293%	58%	26%

Fontes: IBGE (2007, 2008) e Guilhoto e Sesso-Filho (2010). Elaboração própria.

Em todas as medidas e para as duas tabelas, os erros são consideravelmente menores no método RAWS/RAS. Contrastando as duas tabelas, X_U e X_{IM} , nota-se que as medidas proporcionais de erros são menores para a tabela de usos a preços básicos. Particularmente relevantes são as duas últimas medidas, por serem erros proporcionais aferidos a partir de médias ponderadas. Para estas duas medidas, o erro médio não é muito elevado na estimação de X_U pelo método de Guilhoto e Sesso-Filho, mas ainda assim foi reduzido substancialmente no método RAWS/RAS. Na tabela X_{IM} , os erros médios nas medidas DAPP e DAPT são altos para o método Guilhoto e Sesso-Filho e mais razoáveis no método RAWS/RAS.

Para avaliar a dispersão das estimativas, serão aplicados dois tipos de medidas de dispersão. A primeira medida é o desvio padrão dos desvios absolutos, designado pela

sigla DP na Tabela 3.¹² A outra é a proporção das células com desvios proporcionais superiores a 10%, 25%, 50% e 100%. Essas medidas de dispersão estão na Tabela 3.

Tabela 3: Medidas de dispersão para a comparação de métodos na estimação das tabelas de usos a preços básicos (X_U) e de importações (X_{IM}), 2005.

Tabelas/Métodos	DP	Nº de setores com erro superior a:				
		10%	25%	50%	100%	Total
<i>Usos a preços básicos</i>						
RAWS/RAW	77	6%	2%	0,9%	0,5%	100%
Guilhoto e Sesso-Filho	446	21%	7%	2%	0,8%	100%
<i>Importações</i>						
RAWS/RAW	64	15%	10%	6%	2%	100%
Guilhoto e Sesso-Filho	161	28%	24%	19%	13%	100%

Fontes: IBGE (2007, 2008) e Guilhoto e Sesso-Filho (2010). Elaboração própria.

Todas as medidas de dispersão dos erros são menores no método RAWS/RAS. Para o desvio padrão, a diferença entre os métodos é maior na tabela X_U . O desvio padrão das tabelas X_U e X_{IM} é semelhante para o método RAWS/RAS, mas para o método de Guilhoto e Sesso-Filho ele é bem maior na tabela X_U .

Na medida de proporção de setores com erro superior ao designado, para o método RAWS/RAS têm erro proporcional superior a 10% apenas 6% das células na tabela X_U e 15% das células na tabela X_{IM} . Conforme se aumenta o erro estipulado, essas proporções de células diminuem, até que para erros superiores a 100% do valor da célula a proporção é quase nula na tabela X_U e de 2% na tabela X_{IM} . Na comparação com o método de Guilhoto e Sesso Filho, nota-se que inicialmente a diferença entre os métodos é maior na tabela X_U , mas conforme se aumenta o erro estipulado a diferença se torna maior para a tabela X_{IM} . Por fim, cabe apontar que em ambos os métodos essas medidas de erro são maiores na tabela X_{IM} em comparação com X_U .

¹² Também foi calculado o desvio padrão dos desvios proporcionais, mas os resultados não são apresentados por serem muito elevados (acima de 1.000%, para as duas tabelas e nos dois métodos), porque não há ponderação e as células com valor nulo ou quase nulo podem ter erros proporcionais muito altos.

5.3. Medidas de precisão partitiva: vetores de soma de impostos q_m , margens de comércio h_{MC} e margens de transporte h_{MT} totais por coluna.

As MIPs 2000 e 2005 divulgadas pelo IBGE não apresentam as tabelas auxiliares de ICMS (X_{TC}), IPI (X_{TP}), outros impostos líquidos de subsídios (X_{TS}), margens de comércio (X_{MC}) e margens de transporte (X_{MT}). Entretanto, os dados divulgados para tais anos trazem informações a respeito das somas nas colunas destas tabelas auxiliares, as quais foram utilizadas no método RAWs de AUTORX (2013) para a estimação das cinco tabelas. Para empregar esta informação no teste do método RAWs/RAW de estimação de MIPs, a MIP 2000 – que inclui as cinco tabelas estimadas pelo RAWs – é projetada para todos os anos até 2005. As somas nas colunas destas tabelas auxiliares estimadas para 2005 serão então confrontadas com a informação extraída da MIP divulgada pelo IBGE para testar a precisão do método na estimação de tais tabelas.

Da MIP 2005 divulgada pelo IBGE são extraídos os vetores q_m , h_{MC} e h_{MT} , conforme os procedimentos descritos por AUTORX (2013). Recapitulando, seja a matriz H definida por $H = V - X_U - X_{IM} - X_{TM}$, com as tabelas de usos a preços ao consumidor (V), usos a preços básicos (X_U) e importações (X_{IM}) dadas pelo IBGE e a tabela de imposto de importações (X_{TM}) calculada a partir da tabela de importações de acordo com o descrito em AUTORX (2013, seção 3.1). O vetor q_m é composto pelas somas nas colunas de H , enquanto os vetores h_{MC} e h_{MT} são, respectivamente, as linhas referentes aos produtos “060101. Comércio” e “070101. Transporte de carga” em H .

Estes três vetores serão empregados para testar a estimação das cinco tabelas auxiliares mencionadas. O vetor q_m é igual às somas nas colunas da tabela de impostos X_T , formada pela junção dos três tipos de impostos, $X_T = X_{TC} + X_{TP} + X_{TS}$. O vetor h_{MC} , com o sinal invertido, corresponde às somas nas colunas da tabela X_{MC+} de margens de comércio sem a linha em que os totais são lançados com os sinais negativos, que é a do produto “060101. Comércio”. Por fim, o vetor h_{MT} , também com o sinal invertido, é a coleção das somas nas colunas X_{MT+} de margens de comércio sem a linha em que os totais são lançados com os sinais negativos, acrescidas dos impostos (ICMS e outros impostos líquidos de subsídios) que incidem sobre o produto “070101. Transporte de carga”.

Esses três vetores são comparados aos seus equivalentes calculados como somas de colunas das tabelas auxiliares estimadas para 2005. Para isso, são empregadas as mesmas medidas partitivas expostas na subseção anterior, com os ajustes necessários por se tratar da comparação de vetores e não de matrizes.

Para as medidas de posição, as expressões equivalentes às equações (34) a (38) - onde q_j^k representa os elementos dos vetores \mathbf{q}_m , \mathbf{h}_{MC} e \mathbf{h}_{MT} calculados pelos dados do IBGE e \hat{q}_j^k os elementos dos mesmos vetores calculados pelas tabelas auxiliares estimadas - são iguais a:

i) *Desvio absoluto médio (DAM).*

$$DAM = \sum_{j=1}^m \frac{|\hat{q}_j^k - q_j^k|}{m} \quad (39)$$

ii) *Desvio absoluto ponderado (DAP).*

$$DAP = \sum_{j=1}^m \frac{|q_j^k|}{\sum_j |q_j^k|} \cdot |\hat{q}_j^k - q_j^k| \quad (40)$$

iii) *Desvio absoluto percentual médio (DAPM).*

$$DAPM = \sum_{j=1}^m \frac{1}{m} \cdot \left| \frac{\hat{q}_j^k - q_j^k}{q_j^k} \right| \quad (41)$$

iv) *Desvio absoluto percentual ponderado (DAPP).*

$$DAPP = \sum_{j=1}^m \frac{|\hat{q}_j^k - q_j^k|}{\sum_j |q_j^k|} \quad (42)$$

v) *Desvio absoluto ponderado total (DAPT).*

$$DAPT = \left(\sum_{j=1}^m \frac{|q_j^k|}{\sum_j |q_j^k|} \cdot |\hat{q}_j^k - q_j^k| \right) / \left(\sum_{j=1}^m \frac{|q_j^k|}{\sum_j |q_j^k|} \cdot |q_j^k| \right) \quad (43)$$

Os resultados destas medidas, calculadas para os métodos RAWs/RAW e Guilhoto e Sesso-Filho nos três vetores, são apresentados na Tabela 4. Os resultados para os impostos totais são apresentados com e sem o setor exportações da demanda final porque o erro nessa coluna para os impostos é consideravelmente maior que o das outras, principalmente no método de Guilhoto e Sesso-Filho, o que distorce algumas das medidas.

Tabela 4: Medidas de posição para a comparação de métodos na estimação dos vetores de somas nas colunas de impostos totais (q_m), margens de comércio (h_{MC}) e margens de transporte (h_{MT}), 2005.

Tabelas/Métodos	DAM	DAP	DAPM	DAPP	DAPT
<i>Impostos</i>					
RAWS/RAW	223	717	12%	5%	1%
Guilhoto e Sesso-Filho	1.761	15.514	745%	36%	22%
<i>Impostos (sem exportações)</i>					
RAWS/RAW	225	717	10%	5%	1%
Guilhoto e Sesso-Filho	1.456	15.510	59%	29%	22%
<i>Margens de comércio</i>					
RAWS/RAW	64	159	3%	1%	0,2%
Guilhoto e Sesso-Filho	924	13.578	16%	20%	19%
<i>Margens de transporte</i>					
RAWS/RAW	33	52	13%	5%	2%
Guilhoto e Sesso-Filho	206	915	30%	38%	38%

Fontes: IBGE (2007, 2008) e Guilhoto e Sesso-Filho (2010). Elaboração própria.

Novamente, todas as medidas de posição dos erros são menores para o método RAWs/RAW, com diferenças expressivas entre os métodos. Nas medidas de erros ponderados percentuais, DAPP e DAPT, as diferenças vão de 19 a 36 pontos percentuais. Para o método RAWs/RAW, estas duas medidas são particularmente baixas nas margens de comércio, entre 0% e 1%, mas para os impostos e margens de transporte também são pequenas, não superando 5%.

As medidas partitivas de dispersão na estimação dos três vetores são apresentadas na Tabela 5, com duas diferenças em relação às medidas de dispersão da Tabela 3. A primeira é que não é apresentado só o desvio padrão dos erros absolutos, o qual é designado por $|d|$, mas também o desvio padrão dos erros proporcionais, $|d\%|$. A outra

diferença é que, ao invés da proporção de células com erro superior ao designado, é exibida uma contagem do número de setores com erros mais altos, considerando o total de 61 setores, que inclui a demanda final.

Tabela 5: Medidas de dispersão para a comparação de métodos na estimação dos vetores de somas nas colunas de impostos totais (q_m), margens de comércio (h_{MC}) e margens de transporte (h_{MT}), 2005.

Tabelas/Métodos	Desvio Padrão		Nº de setores com erro superior a:				
	d	d%	10%	25%	50%	100%	Total
<i>Impostos</i>							
RAWS/RAW	388	22%	20	6	3	1	61
Guilhoto e Sesso-Filho	4.621	5359%	54	43	28	11	61
<i>Impostos (sem exportações)</i>							
RAWS/RAW	391	13%	19	5	2	0	60
Guilhoto e Sesso-Filho	3.989	55%	53	42	27	10	60
<i>Margens de comércio</i>							
RAWS/RAW	109	4%	4	0	0	0	61
Guilhoto e Sesso-Filho	3.824	23%	29	12	4	1	61
<i>Margens de transporte</i>							
RAWS/RAW	67	30%	20	6	2	1	61
Guilhoto e Sesso-Filho	622	32%	48	26	14	2	61

Fontes: IBGE (2007, 2008) e Guilhoto e Sesso-Filho (2010). Elaboração própria.

Também estas medidas de dispersão são menores para o método RAWS/RAW. Quanto ao desvio padrão, a diferença entre os métodos só é pequena para os erros percentuais das margens de transporte. Na contagem de setores com erros superiores ao estipulado as diferenças entre os métodos são expressivas, principalmente nas primeiras colunas. Note-se que no vetor de margens de comércio estimado pelo RAWS/RAW apenas quatro setores têm erros superiores a 10% e nenhum tem erro superior a 25%.

6. Resultados: Estimativas finais das MIPs 2000-2005.

Nesta seção, são apresentados os procedimentos para a obtenção das estimativas finais da série de MIPs anuais. Entre 2001 e 2004, o método RAWS/RAW será adaptado para interpolar as MIPs de 2000 e 2005. De 2006 a 2009, será aplicado o RAWS/RAW para

a projeção a partir da MIP 2005, com alguns ajustes em razão de especificidades dos dados para esses anos.

6.1. MIPs 2001-2004: interpolação das MIPs 2000 e 2005.

As MIPs projetadas de 2001 a 2005 a partir da MIP 2000 para testar o método de estimação RAWs/RAS serão substituídas por suas versões definitivas. As MIPs 2000 e 2005 serão as divulgadas pelo IBGE, com os ajustes para a revisão da CN 2005/2009 expostos em AUTORX (2013) e as tabelas auxiliares estimadas pelo RAWs.

Já as estimativas definitivas das MIPs 2001 a 2004 serão obtidas a partir de ajustes no método RAW para interpolar as MIPs 2000 e 2005. A interpolação é uma projeção em que a estimativa inicial é dada por uma média ponderada das projeções das tabelas correspondentes para os anos base, no caso 2000 e 2005 (Temurshoev, Webb e Yamano, 2011, p. 94). Para cada ano interpolado, a média ponderada será calculada a partir de duas matrizes de projeção associadas a cada um dos anos base, \mathbf{P}_0 e \mathbf{P}_5 . Sejam $\mathbf{V}_Y = [v_{ijy}]$ as tabelas de usos a preços ao consumidor dos anos base, $\mathbf{V} = [v_{ij}]$ a mesma tabela para cada ano interpolado e $\mathbf{P}_Y = [p_{ijy}]$ as matrizes de projeção, onde $y \in Y = \{0, 5\}$ é o indexador dos anos 2000 e 2005. Define-se:

$$p_{ijy} = \begin{cases} 1 & , \text{para } v_{ijy} = 0 \\ \frac{v_{ij}}{v_{ijy}} & , \text{para } v_{ijy} \neq 0 \end{cases} \quad (44)$$

As matrizes de projeção serão aplicadas sobre as respectivas tabelas dos anos base $\mathbf{X}_{KY} = [x_{ijy}^k]$, onde $k \in K = \{U, IM, TC, TP, TS, MC, MT\}$ é o indexador da tabela e $y \in Y = \{0, 5\}$ o indexador do ano base. Mas as colunas de variação de estoque destas tabelas deverão sofrer ajustes de sinais antes da projeção, pois como foi discutido na seção 3, em geral os sinais das células nas colunas de variação de estoques devem ser os mesmos das células equivalentes na tabela de usos a preços ao consumidor. Assim, se $\dot{\mathbf{X}}_{KY} = [\dot{x}_{ijy}^k]$ indica a tabela com o ajuste na coluna variação de estoque, a qual é designada por $j = E$, teremos:

$$\dot{x}_{ijy}^k = \begin{cases} (-1) \cdot \frac{x_{ijy}^k}{|x_{ijy}^k|} & , \text{para } j = E \text{ e } \frac{x_{iEy}^k}{v_{iEy}} < 0 \\ x_{ijy}^k & , \text{nos demais casos} \end{cases} \quad (45)$$

Com isso, para cada um dos anos interpolados, o caso geral das estimativas iniciais será designado por $\dot{\mathbf{A}}_{\mathbf{K}} = [\dot{a}_{ij}^k]$, com $k \in \mathbf{K} = \{U, IM, TC, TP, TS, MC, MT\}$, e definido por:

$$\dot{a}_{ij}^k = \begin{cases} 0 & , \text{para } v_{ij0} = v_{ij5} = 0 \text{ e } k \neq U \\ v_{ij} & , \text{para } v_{ij0} = v_{ij5} = 0 \text{ e } k = U \\ \left(\frac{5-\gamma}{5}\right) \cdot p_{ij0} \cdot \dot{x}_{ij0}^k + \left(\frac{\gamma}{5}\right) \cdot p_{ij5} \cdot \dot{x}_{ij5}^k & , \text{para } v_{ij0} \neq 0 \text{ ou } v_{ij5} \neq 0 \end{cases} \quad (46)$$

Onde $\gamma \in \{1, 2, 3, 4\}$ é um indexador dos anos de 2001 a 2004, que define ponderações variantes conforme o ano esteja mais próximo de um ou outro ano base, e $\mathbf{X}_{\mathbf{KY}} = [x_{ijy}^k]$ é a tabela correspondente a $\mathbf{A}_{\mathbf{K}}$ nos anos base, com $y \in \mathbf{Y} = \{0, 5\}$.

Nos dois primeiros casos da definição, quando a célula da tabela $\mathbf{V}_{\mathbf{Y}}$ é nula em ambos os anos base e não nula no ano projetado, deve-se adotar um critério de distribuição do valor v_{ij} entre os a_{ij}^k para que a restrição W do balanceamento venha a ser respeitada. Como na seção 3, o valor será todo destinado à célula da tabela $\dot{\mathbf{A}}_{\mathbf{U}}$ e a célula correspondente nas tabelas auxiliares será nula.

O caso geral é o da terceira parte da definição. Quando pelo menos uma das células v_{ijy} não for nula nos anos base, o valor da célula \dot{a}_{ij}^k será determinado por uma média ponderada das projeções dos valores das células correspondentes nos dois anos base.

Assim como na formação das estimativas iniciais para projeção a partir de um ano base, explicada na seção 3, também na interpolação são necessários alguns ajustes à definição mais geral em (46), apresentados nos itens a seguir:

i) Inversões de sinal ou zeragem na soma da linha:

As intervenções referentes a mudanças de sinal e zeragem no total da linha estão na definição (47), onde $\ddot{\mathbf{A}}_{\mathbf{K}} = [\ddot{a}_{ij}^k]$ é a tabela que sofreu essas mudanças:

$$\ddot{a}_{ij}^k = \begin{cases} p_{ij0} \cdot \dot{x}_{ij0}^k, \text{ para } \frac{q_i^k}{q_{i0}^k} > 0 \text{ e } \frac{q_i^k}{q_{i5}^k} < 0 \\ p_{ij5} \cdot \dot{x}_{ij5}^k, \text{ para } \frac{q_i^k}{q_{i0}^k} < 0 \text{ e } \frac{q_i^k}{q_{i5}^k} > 0 \\ v_{ij} \cdot \frac{q_i^k}{|q_i^k|}, \text{ para } q_i^k \neq 0 \text{ e } q_{i0}^k = q_{i5}^k = 0 \\ 0, \text{ para: } \begin{cases} q_i^k = 0 \\ k = TS \text{ e } j \in J \end{cases} \\ \dot{a}_{ij}^k, \text{ nos demais casos} \end{cases} \quad (47)$$

Onde $J = \{\text{Consumo da administração pública, Consumo das ISFLSF, Variação de estoque}\}$.

Nos dois primeiros casos, quando há mudança de sinal no total da linha de 2000 para 2005, somente o ano base cujo total na linha tiver o mesmo sinal da linha no ano interpolado irá ser utilizado na projeção.¹³ No terceiro caso, quando os dois anos base têm total na linha nulo e o ano interpolado é não nulo, considera-se a linha da tabela de usos a preços ao consumidor, com ajuste no sinal. O quarto caso, em que as células são zeradas, ocorre se o total na linha é nulo no ano projetado e, para as colunas especificadas da tabela A_{TS} , para manter nulas células que podem ter sido alteradas indevidamente nos casos anteriores.

ii) Coluna “variação de estoque” das tabelas A_U e A_{IM} :

Para a interpolação, os ajustes nas colunas de variação de estoques são mais complicados que na projeção, pois são várias as combinações possíveis de trocas de sinais e zeragens nos dois anos base e nos anos projetados. Todos os casos são apresentados na definição (49), com $\ddot{A}_K = [\ddot{a}_{ij}^k]$ como a matriz já com essas intervenções na variação de estoques, conforme o definido em (48).

$$\ddot{a}_{ij}^k = \begin{cases} \ddot{a}_{iE}^k, \text{ para } j = E \\ \dot{a}_{ij}^k, \text{ para } j \neq E \end{cases} \quad (48)$$

Onde $j = E$ é a coluna *Variação de estoque*.

¹³ Não ocorrem casos em que o sinal muda em relação a 2000 no ano interpolado e então muda novamente, tornando a ser igual ao sinal de 2000 no ano 2005 $\left(\frac{v_{iE}}{v_{iE0}} < 0 \text{ e } \frac{v_{iE}}{v_{iE5}} < 0\right)$. Se tais casos ocorressem, seriam enquadrados na terceira parte da definição, $\ddot{a}_{ij}^k = v_{ij} \cdot \frac{q_i^k}{|q_i^k|}$.

Apenas as tabelas \mathbf{A}_U e \mathbf{A}_{IM} têm células não nulas de variação de estoque, cujos valores nas projeções iniciais serão estabelecidos com base nos sinais das respectivas células da tabela de usos a preços ao consumidor do mesmo ano interpolado $\mathbf{V} = [v_{ij}]$ e dos anos base $\mathbf{V}_0 = [v_{ij0}]$ e $\mathbf{V}_5 = [v_{ij5}]$.

Para uma compreensão mais intuitiva da definição (49), consideremos o valor nulo como um terceiro sinal, ou seja, cada célula pode ter três sinais possíveis: $\{+, -, 0\}$. Pode-se então dizer que, conforme os dois primeiros casos de (49), quando o sinal de v_{ij} é igual ao de um dos anos base e diferente do outro, considera-se apenas a projeção do ano base com o mesmo sinal. No terceiro caso da definição, quando o sinal de v_{ij} é diferente dos dois anos base, iguala-se a variação de estoque a 1 ou -1 , sem mudar o sinal. Por fim, no quarto caso a projeção não é alterada, o que ocorre se o sinal de v_{ij} for igual ao de v_{ij0} e v_{ij5} .

$$\dot{a}_{iE}^k = \begin{cases} p_{iE0} \cdot \dot{x}_{iE0}^k, \text{ para: } \begin{cases} \frac{v_{iE}}{v_{iE0}} > 0 \text{ e } \frac{v_{iE}}{v_{iE5}} < 0 \\ \frac{v_{iE}}{v_{iE0}} > 0 \text{ e } v_{iE5} = 0 \\ v_{iE} = v_{iE0} = 0 \text{ e } v_{iE5} \neq 0 \end{cases} \\ p_{iE5} \cdot \dot{x}_{iE5}^k, \text{ para: } \begin{cases} \frac{v_{iE}}{v_{iE0}} < 0 \text{ e } \frac{v_{iE}}{v_{iE5}} > 0 \\ v_{iE0} = 0 \text{ e } \frac{v_{iE}}{v_{iE5}} > 0 \\ v_{iE0} \neq 0 \text{ e } v_{iE} = v_{iE5} = 0 \end{cases} \\ \frac{\ddot{a}_{iE}^k}{|\ddot{a}_{iE}^k|}, \text{ para: } \begin{cases} \frac{v_{iE}}{v_{iE0}} < 0 \text{ e } \frac{v_{iE}}{v_{iE5}} < 0 \\ v_{iE} \neq 0 \text{ e } v_{iE0} = v_{iE5} = 0 \\ v_{iE0} = 0 \text{ e } \frac{v_{iE}}{v_{iE5}} < 0 \\ \frac{v_{iE}}{v_{iE0}} < 0 \text{ e } v_{iE5} = 0 \end{cases} \\ \ddot{a}_{iE}^k, \text{ para: } \begin{cases} \frac{v_{iE}}{v_{iE0}} > 0 \text{ e } \frac{v_{iE}}{v_{iE5}} > 0 \\ v_{iE} = v_{iE0} = v_{iE5} = 0 \end{cases} \end{cases} \quad (49)$$

iii) Tabela de imposto de importação \mathbf{A}_{TM} :

Para a tabela de imposto de importação, o processo é igual ao descrito na seção 3. No caso geral, as proporções serão as mesmas da tabela de importações. As células serão

nulas quando o total da linha for nulo e nas colunas de exportações e variação de estoque. Portanto, a tabela $\ddot{\mathbf{A}}_{TM} = [\ddot{a}_{ij}^{TM}]$ é dada por:

$$\ddot{a}_{ij}^{TM} = \begin{cases} 0, & \text{para: } \begin{cases} j = \text{exportações ou variação de estoque} \\ q_i^{TM} = 0 \end{cases} \\ \ddot{a}_{ij}^{IM}, & \text{para os demais casos} \end{cases} \quad (50)$$

De maneira que $\ddot{\mathbf{A}}_{\mathbf{K}} = [\ddot{a}_{ij}^k]$ é igual a:

$$\ddot{a}_{ij}^k = \begin{cases} \ddot{a}_{ij}^{TM}, & \text{para } k = TM \\ \ddot{a}_{ij}^k, & \text{para } k \neq TM \end{cases} \quad (51)$$

iv) Valores negativos nas tabelas de margens, \mathbf{A}_{MC-} e \mathbf{A}_{MT-} :

As tabelas \mathbf{A}_{MC-} e \mathbf{A}_{MT-} também seguirão o exposto na seção 3. Seja $\mathbf{A}_{\mathbf{K}+} = [a_{ij}^{k+}]$ definida por:

$$a_{ij}^{k+} = \begin{cases} 0, & \text{para } \ddot{a}_{ij}^k \leq 0 \\ \ddot{a}_{ij}^k, & \text{para } \ddot{a}_{ij}^k > 0 \end{cases} \quad (52)$$

As tabelas $\mathbf{A}_{\mathbf{K}-} = [a_{ij}^{k-}]$, com $k \in \mathbf{K} = \{MC, MT\}$, serão definidas pelas somas das margens nas colunas nas linhas de totais e anuladas nas outras linhas. A definição geral de $\mathbf{A}_{\mathbf{K}-} = [a_{ij}^{k-}]$ é dada por:

$$a_{ij}^{k-} = \begin{cases} \sum_{i=6n+1}^{7n} -a_{ij}^{MC+}, & \text{para } k = MC \text{ e } i = 060101. \text{ Comércio} \\ \sum_{i=7n+1}^{8n} -a_{ij}^{MT+}, & \text{para } k = MT \text{ e } i = 070101. \text{ Transporte de carga} \\ \ddot{a}_{ij}^k, & \text{para } k \notin \{MC, MT\} \text{ e } \ddot{a}_{ij}^k < 0 \\ 0, & \text{para } \ddot{a}_{ij}^k \geq 0 \end{cases} \quad (53)$$

Finalmente, unindo as expressões (52) e (53), as projeções iniciais das tabelas $\mathbf{A}_{\mathbf{K}} = [a_{ij}^k]$, com $k \in \mathbf{K} = \{U, IM, TC, TP, TS, MC, MT\}$, serão definidas por:

$$\mathbf{A}_{\mathbf{K}} = \mathbf{A}_{\mathbf{K}+} + \mathbf{A}_{\mathbf{K}-} \Rightarrow [a_{ij}^k] = [a_{ij}^{k+} + a_{ij}^{k-}] \quad (54)$$

Dadas as projeções iniciais em (54), as estimativas finais dos anos interpolados serão obtidas pelo algoritmo RAW conforme os procedimentos descritos na seção 4. Tomando N como o número de iterações do algoritmo apresentado na subseção 4.5.1, para os anos de 2001, 2002, 2003 e 2004 a convergência foi obtida, respectivamente, com $N = 149, 58, 57$ e 46 .

6.2. MIPs 2006-2009: projeção da MIP 2005.

Para estimar as MIPs de 2006 a 2009, será empregado o método RAW, exposto na seção 4. Ao invés da MIP 2000, a MIP 2005 será o ano base, incluindo as tabelas auxiliares estimadas pelo RAWS (AUTORX, 2013) e os ajustes da revisão CN 2005/2009.

A única diferença no procedimento será na construção das projeções iniciais, especificamente nas tabelas A_{TC} e A_{TP} . Para os anos de 2006 a 2009, ocorrem nestas duas tabelas situações em que o total na linha é nulo no ano referência ($q_i^k = 0$), que então não oferece parâmetros para a distribuição do total na linha do ano projetado. Nestes casos, adota-se o procedimento exposto na seção 3, ou seja, a linha correspondente da tabela de usos a preços ao consumidor $V = [v_{ij}]$ do ano projetado é usada como referência. Entretanto, deve-se respeitar a estrutura das tabelas de ICMS e IPI discutida na seção 3.3 de AUTORX (2013): elementos nulos nas colunas de exportações e de setores industriais, exceto “0301. Alimentos e bebidas”, além dos procedimentos diferenciados para o produto “030702. Papel e papelão, embalagens e artefatos”. Ademais, como na A_{TS} , também em A_{TC} e A_{TP} há três colunas que devem ser mantidas nulas, em conformidade com o ano referência. Assim, a definição (7) será substituída por:

$$a_{ij}^k = \begin{cases} 0 & , \text{para } q_i^k = 0 \\ v_{ij} \cdot \frac{q_i^k}{|q_i^k|} & , \text{para } q_{i0}^k = 0 \text{ ou } \frac{q_i^k}{q_{i0}^k} < 0, \text{ exceto se: } \begin{cases} k \in \{TC, TP, TS\} \text{ e } j \in J^* \\ k \in \{TC, TP\} \text{ e } j \in J^{**} \\ k \in \{TC, TP\} \text{ e } i, j = P, J \end{cases} \\ \ddot{a}_{ij}^k & \text{definida por (74), nos demais casos} \end{cases} \quad (7^*)$$

Onde: $J^* = \{\text{Consumo da administração pública, Consumo das ISFLSF, Variação de estoque}\}$; $J^{**} = \{\text{setores da indústria, exceto a coluna } j = 0301. \text{ Alimentos e bebidas e a linha } i = 030702. \text{ Papel e papelão, embalagens e artefatos}\}$; $P = 030702. \text{ Papel e papelão, embalagens e artefatos}$ e $J = 030801. \text{ Jornais, revistas, discos e outros produtos gravados}$.

Para os anos 2006, 2007, 2008 e 2009, a convergência do algoritmo RAW (seção 4.4) foi obtida, respectivamente, com $N = 157, 74, 59$ e 68 .

7. Conclusão

Apresentou-se neste texto o algoritmo RAW para projeção de MIPs, o qual aprimora o método de Grijó e Berni (2006), adotando avanços recentes da literatura de projeção de matrizes para fazer o melhor uso da informação disponível nas MIPs e TRUs da década de 2000. Combinado com o algoritmo RAWS de AUTORX (2013), dá origem ao método RAWS/RAW para a estimação anual das matrizes de insumo-produto na referência 2000 das contas nacionais. Os dois algoritmos podem ser apresentados como um único método porque o RAW é uma extensão natural do RAWS, adotando seus desenvolvimentos metodológicos e as tabelas estimadas em AUTORX (2013). A diferença é o conjunto de dados considerado: enquanto o RAWS estima as tabelas auxiliares da MIP para 2000 e 2005, o RAW projeta as MIPs completas de 2001-2004 e 2006-2009.

Para testar o método, as tabelas de 2005 foram estimadas apenas com as informações da MIP 2000 e as TRUs, aplicando o RAW à MIP 2000 completada pelas tabelas auxiliares estimadas com o RAWS em AUTORX (2013). Estas estimativas de teste do método RAWS/RAW foram comparadas com as mesmas tabelas estimadas para 2005 por Guilhoto e Sesso-Filho (2010) e a MIP 2005 original do IBGE. Em todas as medidas – holísticas e partitivas, de posição e dispersão – o desempenho do RAWS/RAW foi melhor. Nas medidas partitivas, que fazem a comparação célula a célula das tabelas e vetores, as diferenças entre métodos são mais marcantes que nas medidas holísticas, que comparam métricas de impacto intersetoriais extraídas da matriz inversa de Leontief.

Conforme as medidas partitivas, o ganho de precisão com a estimação pelo RAWS/RAW foi considerável na tabela de usos a preços básicos, mas foi ainda maior nas tabelas auxiliares. Esse é um resultado importante, pois a disponibilidade de estimativas precisas das tabelas auxiliares é relevante para variadas aplicações, como efeitos de choques sobre importações (Dietzenbacher, Albino e Kutzt, 2005) e efeitos de impostos na estrutura produtiva (Siqueira, Nogueira e Souza, 2001).

Por fim, cabe destacar que a base conceitual do método RAWS/RAW poderá ser aplicada a outros problemas em desenvolvimentos posteriores, como a estimação de Matrizes de Contabilidade Social e a compatibilização com MIPs divulgadas em outros períodos. Em particular, porque as TRUs do ano de 2009 foram as últimas da referência 2000 das contas nacionais, de maneira que o método precisará ser adaptado para dar sequência à série de matrizes quando o IBGE passar a divulgar os dados da nova referência.

Referências

AUTORX. **Estimação das tabelas auxiliares de impostos e margens da Matriz de Insumo-Produto com mínima perda de informação: algoritmo RAWS**. [Texto não publicado, disponibilizado como anexo do presente artigo].

BACHARACH, M. **Biproportional Matrices and Input–Output Change**. Cambridge: Cambridge University Press, 1970.

DEMING, W. E.; STEPHAN, F. F. **On a least-squares adjustment of a sampled frequency table when the expected marginal totals are known**. *The Annals of Mathematical Statistics*, v. 11, n. 4, p. 427–444, dez. 1940.

DIETZENBACHER, E.; ALBINO, V.; KUHTZ, S. “**The fallacy of using US-type Input-Output tables**”. *In: INTERNATIONAL CONFERENCE ON INPUT-OUTPUT TECHNIQUES*, 15. Beijing: 2005.

GUILHOTO, J. J. M.; SESSO FILHO, U. A. **Estimação da matriz insumo-produto a partir de dados preliminares das contas nacionais**. *Economia Aplicada*, São Paulo, v. 9, n. 2, p. 277-299, abr./jun. 2005.

_____. **Estimação da matriz insumo-produto utilizando dados preliminares das contas nacionais: aplicação e análise de indicadores econômicos para o Brasil em 2005.** *Economia & Tecnologia*, ano 6, v.23, p. 53-62, out./dez. 2010.

GRIJÓ, E.; BERNI, D. A. **Metodologia completa para a estimativa de matrizes de insumo-produto.** *Teoria e Evidência Econômica, Passo Fundo*, v. 14, n. 26, p. 9-42, mai. 2006.

HUANG, W.; KOBAYASHI, S.; TANJI, H. **Updating an input-output matrix with sign-preservation:** some improved objective functions and their solutions. *Economic Systems Research*, v. 20, n. 1, p. 111–123, mar. 2008.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Matriz de Insumo-Produto:** Brasil 1996. Rio de Janeiro: IBGE, 1999.

_____. **Sistema de Contas Nacionais:** Brasil 2004-2005. Rio de Janeiro: IBGE, 2007. (Contas Nacionais, n. 20).

_____. **Matriz de Insumo-Produto:** Brasil 2000-2005. Rio de Janeiro: IBGE, 2008. (Contas Nacionais, n. 23).

_____. **Sistema de Contas Nacionais:** Brasil 2005-2009. Rio de Janeiro: IBGE, 2011. (Contas Nacionais, n. 34).

JACKSON, R. W.; MURRAY, A. T. **Alternative input-output matrix updating formulations.** *Economic Systems Research*, v. 16, n. 2, p. 135–148, jun. 2004.

JENSEN, R. C. **The concept of accuracy in regional input-output models.** *International Regional Science Review*, v. 5, n. 2, p. 139-154, 1980.

JUNIUS, T.; OOSTERHAVEN, J. **The solution of updating or regionalizing a matrix with both positive and negative entries.** *Economic Systems Research*, v.15, n. 1, p. 87–96, mai. 2003.

KULLBACK, S.; LIEBLER, R. A. **On information and sufficiency.** *The Annals of Mathematical Statistics*, v. 22, n. 1, p. 79–86, mar. 1951.

LAHR, M. L.; DE MESNARD, L. **Biproportional techniques in input-output analysis: table updating and structural analysis.** Economic Systems Research, v. 16, n. 2, p. 115–134, jun. 2004.

LENZEN, M.; WOOD, R.; GALLEGO, B. **Some comments on the GRAS method.** Economic Systems Research, v. 19, n. 4, p. 461–465, dez. 2007.

MILLER, R. E.; BLAIR, P. D. **Input-output analysis: foundations and extensions.** Englewood Cliffs: Prentice-Hall, 2. ed., 2009.

SIQUEIRA, R.; NOGUEIRA, J.; SOUZA, E. **A incidência final dos impostos indiretos no Brasil: Efeitos da tributação de insumos.** Revista Brasileira de Economia, v. 55, n. 4, p. 513–544, out./dez. 2001.

STONE, R. **Multiple classifications in social accounting.** Bulletin de l'Institut International de Statistique, v. 39, n. 3, p. 215–233, 1962.

STONE, R.; BROWN, A. **A Computable Model of Economic Growth.** London: Chapman & Hall, 1962.

TEMURSHOEV, U.; TIMMER, M. P. **Joint estimation of supply and use tables.** Papers in Regional Science. Oxford, v. 90, n. 4, p. 863-882, nov. 2011.

TEMURSHOEV, U.; WEBB, C.; YAMANO, N. **Projection of supply and use tables: methods and their empirical assessment.** Economic Systems Research, v. 23, n. 1, p. 91-123, mar. 2011.