r b e

ISSN 0034-7140
DOI 10.5935/0034-7140.20160001
Revista Brasileira de Economia, **70**(1), 3–30

Progressividade e Aspectos Distributivos na Previdência Social: Uma Análise com o Emprego dos Microdados dos Registros Administrativos do RGPS*

Luís Eduardo Afonso†

Sumário: **1. Introdução; 2. Fundamentos teóricos e evidências empíricas; 3. Dados e Metodologia; 4. Resultados; 5. Conclusões.**

Palavras-chave: **Previdência Social, Aspectos Distributivos, Progressividade, RGPS, Registros Administrativos.**

Códigos JEL: **H55, J14.**

Este trabalho quantifica os aspectos distributivos das Aposentadorias por Tempo de Contribuição e Aposentadorias por Idade do INSS, com os microdados dos registros administrativos do RGPS. A Taxa de Reposição média foi 82,52%, a Taxa Interna de Retorno média foi 5,32% e o valor médio da Alíquota Necessária foi de 50,53%. Propõe-se também um novo indicador, a Contribuição Adicional e o cálculo da Alíquota Necessária com o emprego da Taxa Interna de Retorno individual. Foram encontradas fortes evidências de progressividade. Valores maiores para todos indicadores foram obtidos para mulheres, indivíduos menos escolarizados, de menor renda e para aposentadorias por idade.

This paper quantifies the distributional aspects and progressivity of the benefits of Length of Contribution Retirement and Age Retirement of INSS, with microdata from administrative records of RGPS. Average Replacement Rate is 82,52%, average Internal Rate of Return is 5,32% and average Necessary Tax Rate is 50,53%. We also propose a new indicator, the Additional Contribution and calculate the Necessary Tax Rate using the individual Internal Rate of Return. Strong evidence of progressivity was found. Higher values for all indicators were obtained for women, less educated individuals and poorer individuals and for the Age Retirement benefit.

*Este artigo é originado da tese de livre-docência do autor, que agradece o financiamento do CNPq, por meio do Edital MCT/CNPq Nº 014/2010–Universal. Foi fundamental o apoio de Rogerio Nagamine Costanzi para a obtenção dos microdados dos registros administrativos do Ministério do Trabalho e Previdência Social. O autor é grato a Cassio Maldonado Turra, Georges Langis, Giselle Datz, João Vinícius de França Carvalho e Samuel de Abreu Pessôa pelos comentários e sugestões feitos a versões anteriores deste trabalho. Finalmente, é importante agradecer o trabalho do assistente de pesquisa Luiz Henrique Cruz e Superti.

†Departamento de Contabilidade e Atuária da FEA/USP. Av. Prof. Luciano Gualberto 908, prédio FEA-3, sala 215. Cidade Universitária, Butantã, São Paulo–SP. CEP 05508-010. E-mail: lafonso@usp.br

1. INTRODUÇÃO

Este trabalho visa quantificar os aspectos distributivos e a progressividade dos benefícios de Aposentadoria por Tempo de Contribuição (ATC) e Aposentadoria por Idade (AI) do Regime Geral de Previdência Social (RGPS) do Instituto Nacional do Seguro Social (INSS). Isto é feito por meio do cálculo de *Indicadores Previdenciários*. São empregados os microdados dos registros administrativos do Ministério da Previdência Social (MPS). Fazem parte do banco de dados 35.000 indivíduos nascidos em 1930, 1935, 1940, 1945, 1950, 1955 e 1960. A hipótese inicial é a existência de progressividade no RGPS. Ou seja, os ganhos oriundos da participação no sistema previdenciário são mais elevados para beneficiários de renda mais baixa, para mulheres e para indivíduos que se aposentam por idade. Inspirado em Pallares-Miralles, Romero & Whitehouse (2012), definem-se os Indicadores Previdenciários como a medida operacional da performance de um sistema previdenciário. Em particular, o foco é a capacidade de a previdência repor renda, reduzir a pobreza e financiar de forma adequada os benefícios.

A relevância do tema fundamenta-se em dois argumentos. O primeiro, de caráter macroeconômico, é o tamanho do dispêndio previdenciário. Em 2014 as despesas com pagamentos de benefícios do RGPS foram de 7,9% do PIB brasileiro, com déficit associado próximo a 1,7% do PIB. O segundo argumento, de natureza microeconômica, é que sistemas previdenciários promovem a realocação intertemporal de recursos dos indivíduos. Este fato pode gerar impactos significativos sobre a distribuição de recursos na economia, tanto de forma intergeracional, quanto de forma intrageracional. Esta realocação pode ocorrer de forma diferenciada para os contribuintes e beneficiários, de acordo com características como coorte, sexo, tipo de benefício, escolaridade (de forma indireta) e nível de renda. Esta redistribuição por vezes pode ser intencional ou não (Gragnolati, Jorgensen, Rocha & Fruttero, 2011), o que é uma evidência da importância de sua quantificação na agenda de estudos sobre temas previdenciários.

Embora a literatura nacional tenha produzido contribuições importantes (Giambiagi & Afonso, 2009; Moura, Jesus Filho, Tafner & Ourives, 2013, dentre outros) existe uma lacuna comparativamente aos trabalhos internacionais no tocante às bases de dados. Grosso modo, os trabalhos sobre previdência podem ser divididos em quatro grupos. No primeiro são empregados *indivíduos representativos*, construídos com base em determinadas características observáveis, nos termos de Leimer (1999). O segundo grupo utiliza microdados em *cross-section*, oriundos dos Censos ou de pesquisas domiciliares amostrais, como a Pesquisa Mensal de Emprego (PME) e a Pesquisa Nacional por Amostra de Domicílios (PNAD). Seu uso apresenta uma restrição: as informações de um determinado momento podem não representar bem a evolução das variáveis de interesse. O terceiro grupo utiliza microdados em painel. Acompanha-se o mesmo conjunto de pessoas por vários anos. O quarto grupo emprega microdados dos registros administrativos das instituições previdenciárias. São informações de fato verificadas de contribuintes e beneficiários, por longos períodos. Como são oriundas de *indivíduos reais*, têm a vantagem de retratar com maior acurácia os sistemas previdenciários.

A literatura sobre aspectos distributivos tem convergido para o emprego de microdados dos registros administrativos. Podem ser citados vários trabalhos nessa linha. Por exemplo, Bertranou & Sánchez (2003), Bucheli, Forteza & Rossi (2010), Reznik, Weaver & Biggs (2009), Shoven & Slavov (2012a), Shoven & Slavov (2012b), e Schröder (2012). No caso brasileiro, a lacuna é o não emprego de registros administrativos, o que limita métodos, força a adoção de premissas e hipóteses por vezes heroicas e limita sobremaneira a generalização das conclusões. As consequências são a reduzida comparabilidade com os trabalhos internacionais e sua menor utilidade como subsídio aos formuladores de políticas públicas.

Uma vez descrito este quadro, considera-se que este trabalho tem a possibilidade de fazer uma contribuição original, por três motivos. O primeiro é o cálculo dos *Indicadores Previdenciários*, de forma similar ao verificado em trabalhos empíricos, com o intuito de analisar a progressividade dos sistemas previdenciários, como Alonso, Hoyo & Tuesta (2015) e Shoven & Slavov (2014). O segundo é o emprego de registros administrativos do RGPS, em consonância com o estado das artes da literatura. Até onde se tem conhecimento, esta é a primeira vez que os microdados do RGPS são usados com esta finalidade. Finalmente, o terceiro é a proposição de algumas modificações na forma de cálculo e interpretação de

alguns dos Indicadores.

Este trabalho tem mais quatro seções, além desta breve introdução. Na seção 2 são apresentados os aspectos mais teóricos mais relevantes da literatura sobre previdência social, com ênfase nos aspectos distributivos. Na sequência faz-se a exposição dos Indicadores Previdenciários e dos principais resultados da literatura empírica relevante. Na seção 3 explica-se a metodologia empregada na construção do banco de dados, tendo como base os registros administrativos do RGPS. A seção 4 traz os resultados do cálculo dos Indicadores Previdenciários. Finalmente, apresentam-se as conclusões da tese.

2. FUNDAMENTOS TEÓRICOS E EVIDÊNCIAS EMPÍRICAS

2.1. Fundamentos e critérios de avaliação de sistemas previdenciários

2.1.1. Fundamentos teóricos

Um sistema previdenciário tem como função básica o pagamento de benefícios de forma continuada às pessoas que deixam o mercado de trabalho, como forma de repor seu padrão de consumo, ainda que de forma parcial, quando de sua aposentadoria. Usualmente o recebimento dos benefícios está condicionado ao pagamento anterior de contribuições previdenciárias, incidentes sobre a renda dos trabalhadores. O RGPS funciona sob o regime de *repartição*. Não há acumulação de recursos. As aposentadorias têm seu valor calculado com base em regras do tipo *benefício definido*. Ou seja, é conhecida a fórmula de cálculo do benefício e, principalmente, como esta está atrelada *ex-ante* a um determinado conjunto de condições. Tendo em vista as características apresentadas e a classificação do RGPS, a partir deste ponto toda a análise terá como foco única e exclusivamente os *regimes de repartição com benefício definido*.

Desde o clássico trabalho de Samuelson (1958), retomado por Aaron (1966), sabe-se que um regime de repartição pode aumentar o bem-estar de cada indivíduo na sociedade. Isto acontecerá se a taxa de crescimento das contribuições (dada pelas taxas de crescimento da renda e da população ativa) superar a taxa de juros da economia. No entanto, esta conclusão, embora fundamental e basilar para parcela expressiva da literatura, não dá conta de várias características importantes dos sistemas previdenciários. Um exemplo são os aspectos distributivos e a progressividade.

A redistribuição feita por um sistema previdenciário pode ser classificada em duas categorias. A primeira é a distribuição intergeracional, ou seja, entre gerações ou coortes diferentes. Este processo é inerente à lógica de um regime PAYG, tendo em vista o financiamento dos benefícios de cada geração pela geração seguinte. Cada coorte pode ser impactada de forma distinta pelo sistema previdenciário. A segunda categoria é a *distribuição intrageracional*. É oriunda da heterogeneidade que existe entre indivíduos de uma mesma coorte, podendo também ser gerada ou acentuada pelo tratamento desigual que o sistema previdenciário dá a trabalhadores da mesma geração. Com base nestas características, define-se que o termo *aspectos distributivos* está ligado à mensuração tanto da distribuição intergeracional, quanto da distribuição intrageracional. O termo *progressividade* é empregado no contexto usual da literatura sobre finanças públicas, conforme discutido, por exemplo, por Musgrave (1985). É de se esperar que a redistribuição gerada pela previdência seja progressiva (Brown & Ip, 2000, p.3).

2.1.2. Critérios de avaliação de sistemas previdenciários

Conforme apontam Brown & Ip (2000), aspectos ligados à adequação, equidade e progressividade são relevantes para a avaliação de sistemas de previdência. Barr & Diamond (2006, 2008, 2009) apontam que um sistema previdenciário deve ter o objetivo de proporcionar segurança econômica na velhice, por meio da suavização do consumo, mitigação da pobreza e redistribuição. O design de um sistema previdenciário deve procurar maximizar essa segurança na velhice, o que inclui o custo de provisão do bem (a previdência social). Estas proposições trazem implícitos alguns critérios de avaliação:

Adequação: capacidade de o sistema repor renda e proporcionar uma forma de seguro contra situações econômicas adversas;

Universalização: proporção de idosos atendida pelo sistema;

Custo de provisão: entendido de forma intertemporal, com base na evolução das variáveis econômico-demográficas que afetam o equilíbrio dos regimes de repartição;

Eficiência ou equidade individual: maior linkage entre contribuições e benefícios;

Incentivo ao crescimento econômico.

Há consenso sobre os pontos mais relevantes do desenho dos sistemas previdenciários. Entretanto, também ficam evidentes as dificuldades operacionais para se encontrar métricas adequadas e, principalmente, dados de boa qualidade, que permitam a mensuração destes pontos. Portanto, com base na discussão apresentada, optou-se por uma escolha que privilegie dois aspectos. O primeiro é a adequação; o segundo é a equidade individual. Esta é uma maneira de analisar e quantificar os aspectos distributivos intrageracionais e intergeracionais, ou seja, verificar a progressividade do RGPS.

2.2. Indicadores Previdenciários

O primeiro indicador é a *Taxa de Reposição* (TR). Corresponde à razão entre o primeiro benefício previdenciário B_{it}, recebido pelo indivíduo i no instante t e a última remuneração anterior à aposentadoria W_{it-1}. Seu cálculo é dado na equação (1). Quando são empregados os valores brutos, este conceito é denominado TR *bruta*. Para o cálculo da TR *líquida*, devem ser usados os valores após a incidência dos tributos. Adotou-se a TR bruta, por quatro motivos. O primeiro é a consistência com relação aos critérios adotados pelo IBGE em suas pesquisas. O segundo é que todos os dados da base original são brutos. O terceiro motivo é que é mais adequado o uso da TR bruta, tendo em vista que o objetivo é analisar características do RGPS. Caso valores líquidos fossem usados, características do sistema tributário afetariam os resultados. O quarto motivo é que a TR bruta é a forma mais adotada na literatura.

$$\mathrm{TR}_i = \frac{B_{it}}{W_{it-1}}. \tag{1}$$

A primeira vantagem da TR é a simplicidade. É possível calcular a TR com duas observações de duas variáveis, em períodos consecutivos. Sendo um indicador "quase contemporâneo", a correção de valores nominais é virtualmente irrelevante. A segunda é a comparabilidade: medidas da TR são comparáveis para benefício de sistemas previdenciários de países distintos. A terceira vantagem é que o cálculo da TR prescinde da escolha da taxa de desconto. O último ponto é a compreensibilidade. Por ser uma medida relativa de poder de compra real, pode ser compreendida até mesmo por leigos.

Os motivos que fazem com que a TR seja tão utilizada, também a fazem incompleta. Como lembram Biggs, Sarney & Tamborini (2009), a TR não permite nenhuma conclusão sobre aspectos *lifetime*, inerentes à lógica dos sistemas previdenciários. Este fato justifica o cálculo de outros indicadores. O segundo indicador é a *Taxa Interna de Retorno* (TIR), apresentada na equação (2). O termo do lado esquerdo é o valor presente das contribuições C_{it} descontadas à taxa TIR, denominado VPC. A somatória tem início quando da primeira contribuição. O termo N corresponde ao último período contributivo. Para os benefícios B_{it}, a explicação é similar. A somatória inicia-se em $N+1$, em que ocorre o recebimento do primeiro benefício, prolongando-se até a idade terminal ω. Este somatório é o *Valor Presente dos Benefícios* (VPB). A TIR, por definição, é a taxa que iguala os fluxos dos dois lados da expressão.

$$\mathrm{VPC}_i = \sum_{t=1}^{N} \frac{C_{it}}{(1+\mathrm{TIR})^t} = \sum_{t=N+1}^{\omega} \frac{B_{it}}{(1+\mathrm{TIR})^t} = \mathrm{VPB}_i\,. \tag{2}$$

A Taxa Interna de Retorno tem características que explicam seu uso tão difundido. A primeira delas é que, assim como a TR, é uma variável que não é expressa em unidades monetárias, o que favorece a compreensão e a comparabilidade entre sistemas diferentes. A segunda, e talvez mais importante, é que

TIR é calculada endogenamente, por meio da equação (2). Desta maneira em seu cálculo não é necessária a escolha de uma taxa de desconto, tendo em vista que esta é exatamente a definição daTIR.

Finalmente, o terceiro indicador é a *Alíquota de Contribuição Necessária* (AliqNec), apresentado na equação (3). Este é um conceito de cunho atuarial, pois representa qual deveria ser a alíquota de contribuição incidente sobre a renda, de tal forma que o valor presente esperado dos benefícios VPB seja igual ao valor presente esperado da renda VPR. Isto pode ser notado pela multiplicação do denominador da equação (3) por AliqNec, o que gera o valor presente das contribuições VPC.

$$\text{AliqNec}_i = \frac{\sum_{t=N+1}^{\omega} \frac{B_{it}}{(1+r)^t}}{\sum_{t=1}^{N} \frac{R_{it}}{(1+r)^t}} = \frac{\text{VPB}}{\text{VPR}}. \tag{3}$$

Dentre todos os indicadores, este merece mais de atenção, devido à discricionariedade, quanto à taxa de desconto r. Quando este cálculo é feito para regimes fundados, em que há acumulação de ativos, a escolha da taxa r é consensual. Sua escolha deve refletir as características dos ativos que irão garantir os benefícios previdenciários. No entanto, para regimes de repartição, tal escolha não é tão óbvia. Uma possível explicação é que o conceito de ativo em um regime de repartição é bem menos claro do que em um regime de capitalização. Neste último, a existência de mercados para os ativos, permite que estes sejam transacionados por ofertantes e demandantes, definindo de maneira mais adequada o seu preço. Já em um regime de repartição, pode ser entendido que o ativo que lastreia os benefícios, denominado *Contribution asset* por Boado-Penas, Valdés-Prieto & Vidal-Meliá (2008) são as contribuições dos trabalhadores ativo, que geram os recursos para pagamento dos benefícios previdenciários a cada instante do tempo.

2.3. Evidências empíricas

Segundo Aldrich (1982), o trabalho pioneiro no emprego da TR foi feito por Horlick (1970). O autor comparou trabalhadores representativos do setor de manufatura de 13 países. O valor médio para a TR está na faixa dos 40%. Este trabalho foi atualizado por Haanes-Olsen (1978), utilizando dados dos mesmos países. O autor encontra resultados mais elevados para a TR, sendo o valor mais baixo de 26% (Reino Unido) e o mais alto de 67% (Suécia). Aldrich (1982) voltou ao tema, analisando os mesmos 12 países, porém com dados do período 1969–1980. Seus resultados reforçam os achados de Haanes-Olsen (1978), que havia encontrado tendência de crescimento da TR, comparativamente a Horlick (1970).

Duas importantes análises para o caso norte-americano são feitas por Leimer (2004, 2007). No primeiro trabalho o autor estuda os aspectos distributivos do OASI para as coortes nascidas até 1922, com o emprego de microdados dos registros administrativos. O autor encontra evidências da existência de um *windfall*[1] para as gerações iniciais. No trabalho de 2007, estas estimativas são atualizadas, incorporando novas premissas demográficas e de evolução do *Trust Fund*. Ratificou-se o resultado de queda da TIR, cujo valor passa de 18,4% para a coorte nascida até 1900, até atingir 2,7% para a coorte nascida em 1945.

Este padrão de resultados é semelhante ao que havia sido reportado por Hurd & Shoven (1985). Neste caso, o objetivo principal é verificar como o início do sistema previdenciário nos EUA afetou a coorte nascida entre 1905 e 1911. Para indivíduos casados, as TIRs reportadas são da ordem de 8,4%, enquanto Leimer (2007) encontra TIRs próximas as 8,6%. Duggan, Gillingham & Greenlees (1993) reforçam os resultados pró-progressividade do OASDI. São analisados os indivíduos de quatro coortes: 1895–1903,

[1] Um regime de repartição baseia-se em transferências das gerações mais novas para as mais antigas. Quando um sistema previdenciário começa, a primeira geração pode receber benefícios, sem que tenha realizado as contribuições necessárias (Feldstein & Liebman, 2002).

1904–1910, 1911–1916 e 1917–1922. A TIR média reportada é de 9,1%. As TRs das mulheres são superiores em 2,5% às TRs masculinas. O mesmo ocorre para os não-brancos e para os trabalhadores de renda mais baixa. Finalmente, são encontradas evidências de *windfall* para as gerações mais antigas.

Um importante avanço nesta literatura ocorre no meio da década de 90 quando Garrett (1995) incorpora a mortalidade diferenciada por faixa de renda. Neste caso, a clara evidência de progressividade no OASDI (por meio de TIRs mais elevadas para os indivíduos de renda mais baixa) é questionada, tendo em vista que as maiores TIRs são obtidas por trabalhadores das faixas intermediárias de renda. No mesmo ano, Duggan, Gillingham & Greenlees (1995) também obtêm resultados menos conclusivos ao empregar taxas de mortalidade por nível de renda. Brown (1998) também incorpora os diferenciais de mortalidade por renda, ao analisar o OASDI e o Canada/Quebec Pension Plans (C/QPP). O autor conclui que ambos os sistemas apresentam progressividade, sendo esta conclusão mais forte para o OASDI.

Uma série importante de contribuições para os países da América Latina é feita por Forteza & Ourens (2009, 2012) e Forteza et al. (2009). No primeiro trabalho são calculadas as TIRs e TRs de 11 países. Os autores encontram progressividade, pois as TIRs são mais elevadas para os contribuintes de renda mais baixa. Obviamente esta conclusão não é válida para os países que têm regimes de capitalização, como Chile, Uruguai e Colômbia. Também de forma genérica, as TRs calculadas são bastante elevadas, atingindo em mais da metade dos países valores superiores a 100%.

Em Forteza et al. (2009) há uma modificação metodológica: os autores empregam registros administrativos dos regimes previdenciários do Chile e do Uruguai. Para o Chile encontram-se TRs muito reduzidas: cerca de 35% para homens e 11% para as mulheres. Quando se incorporam as mudanças legais introduzidas em 2008, os valores aumentam para cerca de 50% e 28%, respectivamente. Forteza & Ourens (2012) voltam a analisar os 11 países do trabalho de 2009, novamente empregando indivíduos representativos, porém desta vez com um modelo de simulação. Mantém-se a conclusão de progressividade, visto que as TIRs são mais elevadas para os indivíduos de renda mais baixa.

No Brasil talvez o primeiro trabalho a tratar deste tema tenha sido Fernandes (1994). Este autor pesquisou a heterogeneidade intergeracional, analisando diversas coortes, a partir de 1930–1935. Por hipótese, no modelo estilizado de previdência utilizado, todos se aposentam aos 60 ou aos 65 anos. Neste segundo caso, a TIR masculina da primeira coorte seria de 1,98%. Para a coorte de 1985–1990, foi encontrada uma taxa negativa: −0,01%. Para as mulheres, os valores seriam de 2,83% e 0,67%, respectivamente. Estes valores enfatizam a hipótese do *windfall* e de progressividade.

Afonso & Fernandes (2005) calculam as TIRs para as coortes nascidas a partir de 1920, com desagregação por nível de educação e região do país, empregando os microdados das PNADs de 1976 a 1999. A TIR média encontrada é de 6,7%. Há evidências de progressividade, pois os valores médios da região nordeste são mais de 1,5 ponto percentual acima das demais regiões. Analogamente, indivíduos com menor nível educacional têm TIR mais de um ponto percentual acima da média. Conclusão similar no tocante às transferências de renda foi obtida por Reis & Turra (2011), que incorporaram as diferenças de composição da população e os diferenciais de mortalidade entre regiões.

Giambiagi & Afonso (2009) concentram seu foco nas ATCs do RGPS, fazendo uso dos microdados da PNAD 2007. Os autores encontram evidências de progressividade: as alíquotas necessárias das mulheres são mais elevadas frente aos homens. Para quase todos os casos analisados os valores são inferiores às alíquotas efetivas (evidência da existência de subsídios cruzados entre os benefícios do RGPS). Quando as TRs são analisadas, o padrão de resultados é um pouco menos claro: os valores obtidos pelos grupos com maior escolaridade são cerca de 12 pontos inferiores em relação aos grupos de menor escolaridade. Conclusões similares foram obtidas também por Penafieri & Afonso (2013).

Os possíveis subsídios cruzados existentes no RGPS são o foco de Caetano (2006). O autor calcula as TIRs de vários grupos de contribuintes, separados por gênero, Espécie de benefício e incidência do fator previdenciário. Para um homem que entrou no mercado de trabalho aos 18 anos e se aposentou por tempo de contribuição aos 53 anos, a TIR é de 1,68%. Para uma mulher que tenha se aposentado após 30 anos de contribuição, a TIR é de 2,51%. Encontram-se evidências de progressividade, sendo que as TIRs mais baixas são encontradas para as ATCs masculinas em que há incidência do fator e as TIRs

mais elevadas são obtidas para as mulheres que se aposentam por idade, recebendo um salário mínimo.

Com base nestes trabalhos é possível chegar a quatro conclusões. É consensual que os sistemas previdenciários têm caráter progressivo. A segunda conclusão é que há fortes evidências de existência de *windfall*, com ganhos para as primeiras gerações com acesso ao sistema previdenciário. A terceira conclusão é que os ganhos são diferentes e decrescentes ao longo do tempo. Por último, verifica-se a existência de transferências intrageracionais, entre indivíduos de uma mesma coorte.

3. DADOS E METODOLOGIA

3.1. Construção do banco inicial a partir dos registros administrativos do RGPS

Os registros são oriundos das bases de dados da Empresa de Tecnologia e Informações da Previdência Social (Dataprev). Os microdados têm periodicidade mensal, compreendendo o período de 27 anos, de janeiro de 1980 a dezembro de 2006. Originalmente foram recebidos dois conjuntos de dados. O primeiro é o *Arquivo de Cadastro e Vínculos e Remunerações*. O segundo conjunto de dados é denominado *Arquivo de Benefícios*, com as informações posteriores ao início do recebimento do benefício previdenciário.

3.1.1. Vínculos e Remunerações

Neste banco há um identificador único para cada indivíduo. A codificação da escolaridade é apresentada na Tabela 1. A variável *Pensão* informa se o benefício é uma pensão ou não. A variável *Competência Vínculo* é uma *dummy* que informa se no mês em questão há registro de remuneração (e contribuição) ao RGPS. A variável tem valor 1 em caso positivo e 0, em caso contrário. Há 324 ocorrências, pois este é o resultado da multiplicação dos 12 meses do ano pelos 27 anos da amostra. Finalmente, a variável *Remuneração-Vínculo* traz o valor da remuneração do segurado, em múltiplos do salário mínimo (SM) nominal vigente na época, multiplicado por 100.

Inicialmente há 5.000 indivíduos de cada coorte, com nascidos nos anos de 1930, 1935, 1940, 1945, 1950, 1955 e 1960, em um total de 34.998 segurados.[2] Há ligeira predominância masculina (53,2%) na amostra. Como o número de categorias é muito elevado e o número de contribuintes por categoria é muito pequeno (considerando os mais de 60% que compõem a categoria 0), decidiu-se fazer uma agregação em quatro *Grupos Educacionais*, da maneira descrita na Tabela 2. Portanto, todos os resultados serão expressos com o emprego destes quatro grupos.

Tabela 1. Distribuição da escolaridade por categoria.

Categoria	Descrição	Número	%
0	Sem informação	21.425	61,22
1	Analfabeto	1.831	5,23
2	Até a 4ª série incompleta do 1º Grau (primário incompleto)	3.438	9,82
3	4ª série completa do 1º Grau (primário completo)	2.752	7,86
4	5ª à 8ª série incompleta do 1º Grau (ginásio incompleto)	1.359	3,88
5	1º grau (ginásio) completo	1.214	3,47
6	2º grau (colegial) incompleto	418	1,19
7	2º grau (colegial) completo	1.399	4,00
8	Superior incompleto	219	0,63
9	Superior completo	943	2,69
Total		34.998	100,00

Fonte: Tabulações do autor, com base nos registros administrativos da Dataprev.

[2]Um problema com dois registros da coorte de 1930 fez com que o banco tivesse na verdade 34.998 indivíduos.

Tabela 2. Grupos educacionais.

Grupo Educacional	Escolaridade (categoria)
0	0
1	1, 2, 3 e 4
2	5, 6 e 7
3	8 e 9

3.1.2. Benefícios

Os bancos do Arquivo de Benefícios têm estrutura similar aos bancos do Arquivo de Cadastro e Vínculos e Remunerações. Há informações sobre a Espécie de benefício, conforme a classificação dos Benefícios de Prestação Continuada do MPS, a data de início do benefício (DIB), a data de cessação do benefício (DCB) e se o benefício é uma *Pensão*. A variável RMI expressa o valor do primeiro pagamento do benefício, expresso em múltiplos do salário mínimo nominal vigente. Porém, de forma distinta do que ocorre para as remunerações, no banco de Benefícios consta apenas o valor do primeiro benefício pago. Por meio do identificador único os componentes deste banco são identificados e atribuídos aos registros do Cadastro de Vínculos e Remunerações. No caso dos benefícios só há registros para indivíduos que no período analisado (1980 a 2006) já tinham a condição de beneficiários. Por este motivo, o tamanho destes bancos é menor do que os bancos de vínculos e remunerações, conforme pode ser visto na Tabela 3. Há 34.998 registros no banco de remunerações, mas no banco de benefícios há apenas 22.916 indivíduos. E deste total, somente 5.270 são aposentadorias por idade e 2.107 são ATCs.

A etapa seguinte consiste na construção de um banco único para contribuições e benefícios para os indivíduos de todas as coortes. Foram mantidas no banco apenas duas espécies de benefícios: Aposentadoria por Tempo de Contribuição (ATC) — Código 42 — e Aposentadoria por Idade (AI) — Código 41. Dois motivos justificam esta escolha. O primeiro é que, dentre todo o rol de benefícios pagos pelo INSS, estas são as espécies que mais claramente podem ser classificadas como previdenciárias. Ou seja, por natureza são benefícios com características programáveis para os quais é razoável pensar na relação entre benefícios e contribuições. O segundo motivo é que a literatura internacional trata basicamente dos benefícios programáveis. Optou-se por excluir do banco de dados todas as pensões. A razão principal é a impossibilidade de identificação destas pensões com base no benefício que lhes deu origem.

Tabela 3. Distribuição dos beneficiários por ano de nascimento.

Ano de nascimento	Número	%
1930	3.319	14,48
1935	4.037	17,62
1940	3.990	17,41
1945	3.928	17,14
1950	3.284	14,33
1955	2.368	10,33
1960	1.990	8,68
Total	22.916	100,00

Fonte: Documento da Dataprev enviado ao autor.

3.2. Cálculo dos valores reais e das contribuições

Uma vez montado o banco inicial e feitos os procedimentos descritos acima, a etapa seguinte consistiu no cálculo das contribuições previdenciárias feitas pelos segurados, pois no banco de dados original não há essa informação. É importante lembrar que os valores da remuneração Rem e do benefício RMI dos indivíduos estão expressos em múltiplos do salário mínimo nominal $\mathrm{SM}_{\mathrm{nom}}$ vigente a cada mês, multiplicado por 100. Desta maneira, faz-se inicialmente o cálculo da Renda Nominal e do Benefício Nominal de cada segurado i a cada mês t.

A seguir é feito o cálculo da contribuição previdenciária de cada trabalhador i no mês t de acordo com as alíquotas de contribuição incidentes sobre o empregado e sobre o empregador. Para que isso fosse possível, foi feita uma extensiva pesquisa na legislação previdenciária para encontrar todas as alíquotas vigentes de janeiro de 1980 a dezembro de 2006. Nestes 27 anos foram contadas 96 alterações de alíquotas, faixas ou limites de contribuição. A maior frequência de mudanças ocorre no período de maior turbulência inflacionária (e de planos econômicos) do país, entre 1986 e 1994.

Após o cálculo dos valores nominais das contribuições, foi feita a atualização dos valores para moeda constante de dezembro de 2006. Foi empregado nos cálculos o mesmo indicador empregado pelo MPS o Índice Nacional de Preços ao Consumidor (INPC). Para valores anteriores à mudança da moeda efetuada em julho de 1994, com o Plano Real, foi adotada a mesma metodologia que o Ipeadata utiliza para a série de salário mínimo nominal. A última etapa do procedimento empírico consistiu na imputação dos valores da 13º parcela anual da renda e dos benefícios, com base no mês de dezembro de cada ano.

Duas hipóteses adotadas nos cálculos devem ser explicitadas. A primeira refere-se ao vínculo empregatício e à situação do contribuinte no mercado de trabalho. Empregados do setor privado de um lado, e trabalhadores por conta própria e empregados domésticos do outro, tradicionalmente vêm sendo tratados de forma diferente pela previdência social no Brasil, com regras contributivas diferenciadas. A alíquota dos conta-própria e dos domésticos sempre foi um pouco inferior a dos assalariados (considerando a soma das alíquotas do empregado e do empregador).

Na base de dados disponibilizada pela Dataprev, não é possível fazer a distinção entre empregados, trabalhadores por conta própria e empregados domésticos. Por este motivo, optou-se por considerar todos os indivíduos da amostra como empregados. Ao se adotar esta hipótese, o resultado é uma possível sobre-estimação das contribuições previdenciárias dos trabalhadores por conta própria e dos trabalhadores domésticos. Parece ser possível considerar que esta hipótese não é tão problemática. Dados do MPS mostram que em setembro de 2006 (data da coleta de dados da PNAD deste ano) havia no país 33,6 milhões de contribuintes. Destes, mais de 79% eram empregados. Ou seja, a possível sobre-estimação aplica-se a cerca de 21% da amostra. Deve ser ressaltado ainda que esta é uma informação do último ano (2006) da amostra. Para anos anteriores, a proporção de contribuintes que não eram empregados assalariados é certamente menor. Quanto mais próximos de 1980 forem os dados, menor será a imprecisão cometida com a hipótese feita.

3.3. Imputação dos valores para períodos não observados

Após o cômputo dos valores reais de todas as variáveis, procede-se à imputação dos dados para os períodos não observados, ou seja, os anos anteriores a 1982 e posteriores a 2006. Os dados iniciais contemplam um período de 25 anos (já contando o descarte dos dados de 1980 e 1981), relativamente reduzido frente ao ciclo de vida, que pode durar de 60 a 70 anos. Dado este fato, é necessário criar anos adicionais, referentes aos períodos pré-1982 e pós-2006, ao banco original. Posteriormente, nesses novos períodos são imputados os valores das variáveis monetárias (renda, contribuições e benefícios) e demais características observáveis. A estratégia adotada dividiu-se em duas partes. Para o período anterior a 1982 criou-se um banco de dados vazio, iniciado em 1944. A este banco foram incorporadas as informações invariáveis de cada indivíduo (identificador, Espécie de benefício, escolaridade). Posteriormente foram incorporadas as variáveis monetárias. As mais importantes são renda e as contribuições e, implicitamente, a *densidade contributiva*. Imputou-se a cada ano t ($t < 1982$) para cada indivíduo i

com um vetor de características X os mesmos valores de renda e contribuições que os indivíduos com o mesmo vetor X de características apresentam para os anos do intervalo 1982–2006. Considerando as variáveis disponíveis e as características fixas de cada indivíduo i da mostra, definiu-se X de acordo com a equação (4).

$$X_i = X_i\,(\text{Idade}, \text{Sexo}, \text{Espécie de Benefício}, \text{Grupo Educacional})\,. \tag{4}$$

Para a segunda parte, o período posterior a 2006, o procedimento é bem mais simples, tendo em vista que todos os trabalhadores já deverão ter obtido o benefício. Desta maneira, mimetizando o procedimento do MPS de manter constante o valor real dos benefícios, replicou-se para os anos posteriores a 2006 o último valor real efetivamente observado.

Um subproduto da manipulação dos dados deste trabalho é a obtenção da *Densidade Contributiva* (DC) para o período observado (1982–2006), conforme apresentado na equação (5). Este é um indicador da relação entre o número de contribuições previdenciárias efetivamente realizadas por cada indivíduo i e o número máximo possível N de contribuições que teriam sido realizadas caso o indivíduo permanecesse o tempo no mercado de trabalho formal ou não tivesse períodos de desemprego. Por construção, para cada segurado o valor de DC deve estar compreendido no intervalo $[0,1]$:

$$\text{DC}_i = \frac{\sum_{t=1}^{N} \text{Nº Contrib}_{it}}{N}. \tag{5}$$

A incorporação da Densidade de Contribuições é uma diferença importante deste artigo em relação a trabalhos que empregam indivíduos representativos ou dados em *cross-section*, como a PNAD. Particularmente neste segundo caso, como não se tem as informações de toda vida ativa dos indivíduos, usualmente opta-se por considerar que o status contributivo foi o mesmo ao longo de toda a vida. Isto significa implicitamente a adoção de uma hipótese de densidade de contribuições igual a 1, o que certamente não ocorre, tendo em vista períodos de desemprego ou informalidade. O valor de DC foi calculado para toda a amostra, no período para o qual estão disponíveis os valores primários de renda (1982–2006). Os dados mostraram que a densidade de contribuições é igual a 0,6975.

A última etapa consistiu na imputação da expectativa de sobrevida. Foram empregadas as tábuas de mortalidade por sexo abreviadas do Brasil, por coorte (com 5 anos) e por quinquênio, disponibilizadas pela Cepal[3]. Estas tábuas foram disponibilizadas para os quinquênios 1960–1965, 1955–1960 e 1950–1955. Como o banco apresenta indivíduos nascidos a partir de 1930, adotou-se um procedimento de extrapolação das expectativas de sobrevida para cada idade, para as coortes mais antigas. Na falta de informações mais precisas adotou-se uma forma de "agravar" as tábuas, com a consequente redução das expectativas de sobrevida. Isto foi feito adotando-se para cada grupo etário uma redução correspondente à média da razão encontrada para as tábuas para as quais se dispõe da informação. Exemplificando: suponha-se que de 1955–1960 para 1960–1965 a expectativa de vida para uma determinada idade passou de 18 para 20 anos; e que de 1950–1955 para 1955–1960 a expectativa de vida para uma determinada idade passou de 16 para 18 anos. Então o aumento médio foi a média entre $2/18$ e $2/16$.

3.4. Cálculo dos Indicadores, exclusão de observações e identificação de *outliers*

O primeiro indicador calculado, pela simplicidade operacional, foi a Taxa de Reposição TR. Seguindo o procedimento de vários autores, o cálculo foi feito, empregando não apenas a última remuneração imediatamente anterior à DIB, mas sim o período de 12 meses anterior à obtenção do benefício. O objetivo é tentar evitar que flutuações de curto prazo afetem os resultados.

Para um número elevado de beneficiários não foi possível efetuar o cálculo da TR, pois o período contributivo esgotava-se muito antes (mais de um ano) antes do início do recebimento da aposentadoria. Aparentemente esse é um *puzzle* inerente às características da base de dados, para o qual não foi

[3]O autor agradece a Guiomar Bay pela disponibilização das tábuas.

possível encontrar uma explicação consistente. Decidiu-se excluir da amostra todos os registros para os quais não foi possível calcular a TR. Com este procedimento, o número de indivíduos caiu de 7.325 para 1.733. Destes foram eliminadas 193 benefícios classificados como pensões, pelos motivos expostos anteriormente. Restaram, portanto apenas 1.540 observações. É inegável que esta é uma drástica redução.

Quando os indicadores TIR e AliqNec foram calculados, verificou-se a existência de alguns *outliers*, claramente incompatíveis com os padrões e resultados esperados para um sistema previdenciário. A identificação destes padrões está ligada aos valores reportados pela literatura empírica, apresentados por exemplo em Afonso & Fernandes (2005), Giambiagi & Afonso (2009), para o caso brasileiro, e Forteza & Ourens (2009). Imagina-se que estes *outliers* possam ter origem em duas fontes. A primeira é algum tipo de problema nos registros administrativos da Dataprev. A segunda são distorções nos dados, em especial nos períodos de alta inflação e sucessivos planos econômicos, entre 1986 e 1994.

Optou-se pela realização de um procedimento de identificação e exclusão de *outliers* com base nos resultados calculados para TR, TIR e AliqNec. Seguindo a recomendação de Weber (2010) utilizou-se a rotina *blocked adaptive computationally efficient outlier nominators* (`bacon`)[4]. O autor fornece evidências de que o seu emprego é vantajoso em relação à rotina `hadimvo`, apresentada por Hadi (1992) e Hadi (1994), visto que consegue identificar um conjunto similar de *outliers*, porém com ganhos computacionais. O emprego da rotina `bacon` implicou a eliminação de 95 dos 1.540 registros que compunham a amostra. É possível considerar como aceitável esta perda de cerca de 6,2%, tendo em vista os objetivos do trabalho e o tipo de microdados utilizados. Os resultados reportados a seguir foram obtidos para os 1.445 registros restantes, que compõem o que se denominou por *Amostra Completa*.

4. RESULTADOS

Uma vez realizado o procedimento de construção do banco de dados, são apresentados os resultados dos *Indicadores Previdenciários*. Todos os valores são apresentados para a *Amostra Completa* e para as categorias de análise: sexo, Espécie de benefício, coorte, grupo educacional e quartil de renda. Antes, porém, são reportadas as estatísticas descritivas mais relevantes. A Tabela 4 apresenta o número de indivíduos da *Amostra Completa*, de acordo com o ano de nascimento e sexo. Os procedimentos de corte no banco explicam porque há apenas 14 indivíduos nascidos em 1960. Em 2006, ano mais recente da amostra, estes indivíduos tinham 46 anos, idade reduzida para a obtenção da ATC. Por outro lado, a

Tabela 4. Número de indivíduos por ano de nascimento e sexo.

	Mulheres		Homens		Total
Ano	Número	%	Número	%	Número
1930	57	64,04	32	35,96	89
1935	137	43,49	178	56,51	315
1940	150	42,02	207	57,98	357
1945	139	42,77	186	57,23	325
1950	66	26,61	182	72,39	248
1955	39	40,21	58	59,79	97
1960	6	42,86	8	57,14	14
Total	594	41,11	851	58,89	1.445

Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

[4]Esta rotina foi desenvolvida por Sylvain Weber e pode ser obtida em `http://www.stata-journal.com/software/sj10-3`.

explicação para o baixo número de nascidos em 1930 (89 pessoas) é oposta. A expectativa de vida das coortes mais antigas é substancialmente menor do que a encontrada para as coortes mais novas.

Com exceção dos anos 1930 e 1950, a proporção de mulheres na amostra é relativamente estável As três coortes mais numerosas na amostra são as de 1940, 1945 e 1950, que representam mais de 64% do total. A Tabela 5 apresenta a tabulação cruzada entre as duas espécies de benefícios e sexo. Na amostra quase 62% dos benefícios são ATCs e o restante, AIs. Há uma diferenciação importante: os homens representam a maioria das ATCs (78,7%) e as mulheres são o grupo majoritário (62,12%) nas AIs. A predominância masculina já era esperada, tendo em vista que no passado a inserção da mulher no mercado de trabalho, particularmente no setor formal era bem menos expressiva. E posteriormente esta reduzida inserção reflete-se em um menor número de benefícios de aposentadoria programáveis.

Tabela 5. Número de beneficiários por Espécie de benefício e sexo.

	Mulheres		Homens		Total	
Espécie	Número	%	Número	%	Número	%
AI (41)	369	62,12	181	37,88	550	38,06
ATC (42)	225	21,27	670	78,73	895	61,94
Total	594	41,11	851	58,89	1.445	100,00

Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

A Tabela 6 mostra que mais de 40% dos beneficiários da Amostra Completa fazia parte do Grupo Educacional 1. Nos grupos 2 e 3, que englobam indivíduos com maior escolaridade, o número de componentes é bem mais baixo. A única diferença digna de nota entre homens e mulheres refere-se à participação nos grupos 0 e 1. Para quase um terço da amostra os microdados não apresentavam o nível educacional. Este problema é mais grave no caso feminino, em que quase 43% dos registros não apresentam essa informação. A informação sobre escolaridade é obtida quando da filiação ao RGPS, ou então em alguma situação que tenha justificado uma atualização das informações cadastrais. Como esta é uma informação acessória, não fundamental para o MPS, é provável que a confiabilidade destes dados não seja muito alta, tendo em vista que esforços adicionais de coleta teriam pouca valia. Esta hipótese é corroborada pela parcela, superior a 32%, de pessoas para as quais não há dados para esta variável. De forma complementar, a Tabela 7 mostra que a qualidade das informações é melhor para as gerações mais novas, tendo em vista o menor número de beneficiários do Grupo 0.

Finalmente a Tabela 8 mostra outra importante diferença entre os sexos, no que se refere ao número de beneficiários por quartil de renda. A presença masculina (feminina) é muito mais expressiva nos quartis de renda mais altos (baixos), evidenciando que a remuneração dos homens é mais elevada que a das mulheres. Esta desigualdade se mostra bastante importante para explicar as diferenças encontradas para os indicadores previdenciários de ambos os sexos.

Tabela 6. Número de beneficiários por sexo e Grupo Educacional.

	Grupo 0		Grupo 1		Grupo 2		Grupo 3		Total
Sexo	Número	%	Número	%	Número	%	Número	%	Número
Mulheres	254	42,76	190	31,99	103	17,34	47	7,91	594
Homens	215	25,26	401	47,12	161	18,92	74	8,70	851
Total	469	32,46	591	40,90	264	18,27	121	8,37	1.445

Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

Tabela 7. Número de beneficiários por ano de nascimento e Grupo Educacional.

	Grupo 0		Grupo 1		Grupo 2		Grupo 3		Total
Ano	Número	%	Número	%	Número	%	Número	%	Número
1930	72	80,90	16	17,98	1	1,12	0	0,00	89
1935	176	55,87	101	32,06	24	7,62	14	4,44	315
1940	121	33,89	157	43,98	58	16,25	21	5,88	357
1945	70	21,54	161	49,54	63	19,38	31	9,54	325
1950	24	9,68	115	46,37	76	30,65	33	13,31	248
1955	6	6,19	34	35,05	37	38,14	20	20,62	97
1960	0	0,00	7	50,00	5	35,71	2	14,29	14
Total	469	32,46	591	40,90	264	18,27	121	8,37	1.445

Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

Tabela 8. Número de beneficiários por sexo e quartil de renda.

	Quartil 1		Quartil 2		Quartil 3		Quartil 4		Total
Sexo	Número	%	Número	%	Número	%	Número	%	Número
Mulheres	232	39,06	183	30,81	99	16,67	80	13,47	594
Homens	96	11,28	186	21,86	268	31,49	301	35,37	851
Total	328	22,70	369	25,54	367	25,40	381	26,37	1.445

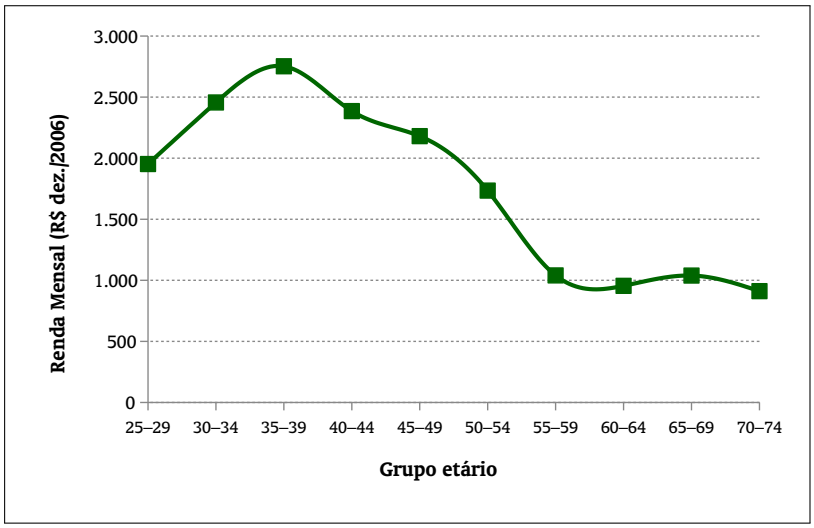
Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

As diferenças de remuneração também podem ser visualizadas por meio da evolução da renda do trabalho ao longo da vida ativa, seja para a Amostra Completa, seja para as categorias de análise previamente apresentadas. Devido ao tamanho reduzido da amostra (problema agravado quando as informações são segregadas por categorias), todos os gráficos são feitos por grupos etário de cinco anos.

A Figura 1 apresenta a renda média ao longo do ciclo de vida, para a Amostra Completa. Há nítida concavidade da curva, com o máximo situando-se antes dos 40 anos (o que é supreendentemente prematuro), decrescendo posteriormente e ficando praticamente estável a partir dos 55 a 59 anos. O formato da curva de renda deste gráfico pode ser mais bem compreendido por meio da comparação com a Figura 2, que traz a renda média por Grupo Educacional. Como era esperado, o Grupo 3 (educação superior) tem um prêmio salarial substantivo em relação aos Grupos 2 (ensino médio) e 1 (ensino fundamental). É razoável inferir que o formato ascendente da curva da Figura 1 para as idades mais baixas é parcialmente explicado pelo comportamento da renda do grupo de maior escolaridade (embora possa se notar que a renda dos demais grupos também tenha concavidade, não tão facilmente visível por causa da escala do gráfico), conforme pode ser visto na Figura 2. O trecho ascendente para as idades mais elevadas não parece ter efeito global significativo, dado o reduzido número de indivíduos que faz parte deste grupo. Desta forma, o efeito dos Grupos 1 e 2 tem certa sobreposição. O comportamento do Grupo 0 (para o qual não se tem informações sobre educação) é bastante similar ao verificado para os Grupos 1 e 2, o que leva a crer que a não-informação do nível educacional é mais expressiva para as duas faixas educacionais inferiores. O perfil de renda é relativamente diferente daquele apresentado por Giambiagi & Afonso (2009, p.162–164), o que é um sinal de que o uso de dados em *cross-section* como a PNAD pode gerar certas discrepâncias frente ao uso de registros administrativos aqui adotados.

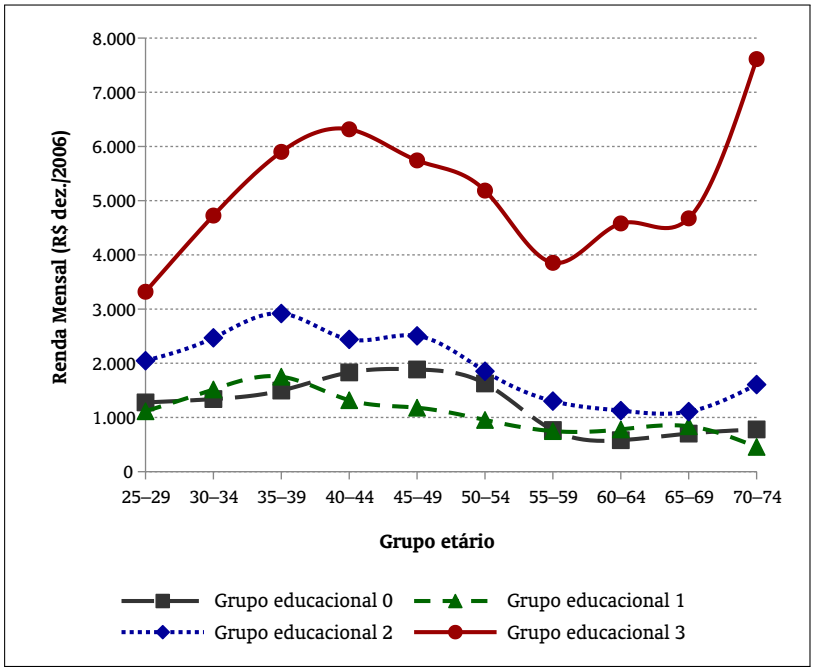
A Figura 3 traz a renda média por idade desagregada por sexo. Para a maioria das idades a renda masculina é mais elevada, o que deve estar ligado às diferenças de escolaridade. Na Figura 4 é mostrada

Figura 1. Renda mensal real (R$ de dez./2006) por Grupo etário – Amostra Completa.



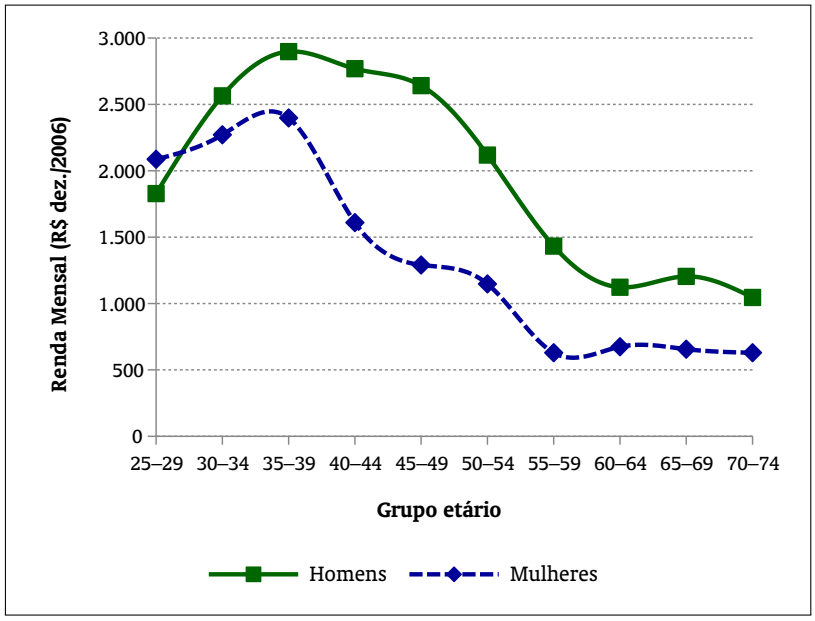
Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

Figura 2. Renda mensal real (R$ de dez./2006) por Grupo etário e Grupo Educacional.



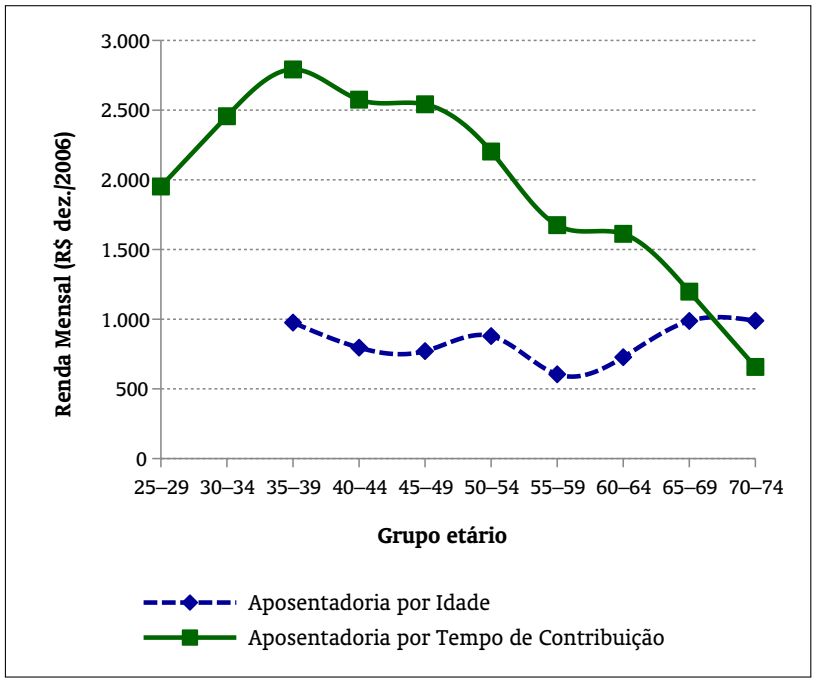
Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

Figura 3. Renda mensal real (R$ de dez./2006) por Grupo etário e sexo.



Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

Figura 4. Renda mensal real (R$ de dez./2006) por Grupo etário e Espécie de benefício.



Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

a renda dos indivíduos que se aposentaram por tempo de contribuição e dos que obtiveram o benefício por idade. Há evidentes diferenças, com maiores ganhos para o primeiro grupo. Este resultado está em consonância com fatos estilizados da previdência social no Brasil, em que indivíduos que obtêm a ATC têm renda mais elevada e se aposentam mais cedo, além de terem valor de benefício mais elevado. Este argumento pode ser visualizado na Tabela 9, na qual são reportados os valores médios da renda e dos benefícios, por quartil de renda e para a Amostra Completa. Também é apresentada a relação entre benefícios e renda, que decresce monotonicamente com o aumento da renda. Este indicador fornece mais uma evidência de que o RGPS, para os benefícios em análise, apresenta progressividade. Análise similar pode ser feita com base na Tabela 10, na qual os valores são segregados pela Espécie de benefício.

Tabela 9. Valor médio dos benefícios e da renda por Quartil de renda (Valores em R$ de dezembro de 2006).

	Quartil 1	Quartil 2	Quartil 3	Quartil 4	Amostra Completa
Benefício médio	382,67	745,51	1.309,16	2.054,92	1.151,55
Renda média	288,61	580,31	1.207,60	4.533,21	1.715,67
Benefício/Renda (%)	132,59	128,47	108,41	45,33	67,12
Número	328	369	367	381	1.445

Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

Tabela 10. Valor médio dos benefícios e da renda por Espécie de benefício (Valores em R$ de dezembro de 2006).

	AI	ATC	Amostra Completa
Benefício médio	640,99	1.465,31	1.151,55
Renda média	697,31	2.341,47	1.715,67
Benefício/Renda (%)	91,92	62,58	67,12
Número	550	895	1.445

Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

4.1. Taxa de Reposição (TR)

Nesta seção são apresentados os resultados obtidos para a Taxa de Reposição TR. São reportados os valores para a Amostra Completa e por categoria de análise. A TR média para os 1.445 indivíduos da Amostra Completa foi de elevados 82,52%. A Tabela 11 reporta esses valores. Parece haver evidências de redução da TR ao longo do tempo para as coortes mais novas. Esta diminuição pode ser entendida como uma evidência da existência de um *windfall* associado à expansão do sistema previdenciário no Brasil. Este fato é similar ao verificado por Leimer (2004, 2007) e Duggan et al. (1993) para os EUA.

A Tabela 12 apresenta os valores desagregados por sexo. A TR feminina é cerca de 10 pontos percentuais superior à masculina. Talvez esta diferença esteja associada às características da remuneração média dos segurados e à existência de um teto para o valor dos benefícios. Esta linha de raciocínio é corroborada pelos resultados da Tabela 13, que mostra que as TRs das AIs são mais de 20 pontos percentuais mais elevadas que as das ATCs. Os resultados da Tabela 14 mostram que com exceção do Grupo 0 os valores são tão mais elevados quanto menor é o nível educacional. Os valores da TR são reportados para os quartis de renda na Tabela 15. Quanto maior a renda, mais baixa é a TR.

Este primeiro conjunto de resultados deve ser analisado em conjunto. Chama a atenção o elevado valor médio encontrado comparativamente a outros trabalhos. Estes valores são bem mais altos que

Tabela 11. Taxa de Reposição (TR) por coorte.

Coorte	*N*	Média (%)	Desvio-padrão	Mínimo	Máximo
1930	89	95,14	32,74	10,42	230,16
1935	315	86,37	36,20	7,16	246,17
1940	357	87,71	33,68	6,25	241,16
1945	325	84,23	36,53	5,86	266,96
1950	248	73,35	37,13	8,64	261,07
1955	97	60,45	30,11	10,98	165,60
1960	14	58,71	26,08	24,51	105,62
Amostra Completa	1.445	82,52	36,12	5,86	266,96

Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

Tabela 12. Taxa de Reposição (TR) por sexo.

Sexo	*N*	Média (%)	Desvio-padrão	Mínimo	Máximo
Feminino	594	88,32	32,54	5,86	246,17
Masculino	851	78,46	37,92	6,18	266,96
Amostra Completa	1.445	82,52	36,12	5,86	266,96

Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

Tabela 13. Taxa de Reposição (TR) por Espécie de benefício.

Espécie de benefício	*N*	Média (%)	Desvio-padrão	Mínimo	Máximo
AI (41)	550	96,27	30,04	7,16	241,16
ATC (42)	895	74,07	36,95	5,86	266,96
Amostra Completa	1.445	82,52	36,12	5,86	266,96

Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

Tabela 14. Taxa de Reposição (TR) por Grupo Educacional.

Grupo Educacional	*N*	Média (%)	Desvio-padrão	Mínimo	Máximo
Grupo Educacional 0	469	86,63	33,63	5,86	231,41
Grupo Educacional 1	591	89,55	35,03	12,40	266,96
Grupo Educacional 2	264	75,57	34,47	6,18	196,64
Grupo Educacional 3	121	47,39	31,49	8,64	166,98
Amostra Completa	1.445	82,52	36,12	5,86	266,96

Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

Tabela 15. *Taxa de Reposição (TR) por quartil de renda.*

Quartil de renda	*N*	Média (%)	Desvio-padrão	Mínimo	Máximo
Quartil 1	328	101,42	27,81	17,72	246,17
Quartil 2	369	89,94	33,96	7,16	241,16
Quartil 3	367	86,54	32,19	12,40	266,96
Quartil 4	381	55,18	32,41	5,86	243,26
Amostra Completa	1.445	82,52	36,12	5,86	266,96

Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

aqueles reportados por Giambiagi & Afonso (2009) e Penafieri & Afonso (2013).

Estas diferenças podem ser um sinal de que o RGPS é generoso, comparativamente aos regimes previdenciários de outros países. Evidencia-se que as primeiras gerações foram beneficiadas. Em outras palavras, estes resultados fornecem um subsídio à tese da existência de transferências intergeracionais e principalmente da existência de um *windfall* no RGPS, nos moldes apontados por Feldstein & Liebman (2002) e Feldstein (2005a, 2005b). O segundo aspecto refere-se à existência de transferências intrageracionais. Foram encontrados valores diferentes de TR para grupos separados por critérios distintos, como gênero, Espécie de benefício, nível educacional e renda. Valores mais altos foram obtidos para as AIs, frente à ATCs, para as mulheres em relação aos homens, para os indivíduos menos educados em relação aos de maior escolaridade e para os primeiros quartis de renda, comparativamente aos últimos quartis. Quanto menor a renda do segurado, mais elevada será sua TR.

4.2. Taxa Interna de Retorno (TIR)

O segundo Indicador Previdenciário é a Taxa Interna de retorno (TIR). O valor médio real obtido foi de 5,32% ao ano para a Amostra Completa, o que também está bastante acima dos padrões encontrados na literatura. Os dados por coorte evidenciam um expressivo aumento dos valores para as coortes mais novas (Tabela 16), de forma distinta do que foi verificado para a TR, em que os valores mais elevados foram encontrados para as coortes mais antigas. Um padrão similar de resultados havia sido encontrado por Afonso & Fernandes (2005), porém sem a desagregação entre trabalhadores dos setores privado e público. Os valores médios aqui encontrados (5,32%) são inferiores àqueles verificados no trabalho citado (6,7%). Porém são bem mais elevados, em média, que os valores calculados por Caetano (2006) e Penafieri & Afonso (2013). De forma similar ao verificado para a TR, a TIR média dos homens é sensivelmente mais baixa que a das mulheres. Este fato pode ser verificado na Tabela 17.

Os resultados separados por Grupo Educacional, reportados na Tabela 18, parecem apontar que os segurados de escolaridade baixa têm TIR bem mais elevada que os de escolaridade média ou superior. Na sequência (Tabela 19), a desagregação por Espécie de benefício mostra que as TIRs calculadas para as AIs são cerca de 4 pontos percentuais mais altas do que aquelas verificadas para as ATCs. Os dados por quartil de renda (Tabela 20) apresentam o mesmo padrão: maiores valores estão associados aos grupos de renda mais baixa, com bastante similaridade entre o segundo e o terceiro quartis, mas com notável diferença entre os extremos da distribuição. Estes elevados resultados reforçam as evidências de progressividade do RGPS.

4.3. Alíquota Necessária (AliqNec)

A Alíquota Necessária representa a alíquota que deveria incidir sobre a renda do segurado para que o valor presente das suas contribuições fosse igual ao valor presente dos benefícios. Os valores encontrados são bastante elevados. O valor médio ultrapassa 50%. O valor de AliqNec cresce a partir da coorte de 1930, até atingir o máximo para a coorte de 1940, em que o valor máximo ultrapassa os 55%. Para as

Tabela 16. *Taxa Interna de Retorno (TIR) por coorte.*

Coorte	*N*	Média (% a.a.)	Desvio-padrão	Mínimo	Máximo
1930	89	4,34	2,67	−3,12	14,76
1935	315	4,75	3,84	−10,56	21,08
1940	357	5,90	4,95	−3,90	21,75
1945	325	5,53	4,77	−1,84	25,97
1950	248	4,82	3,28	−3,99	22,27
1955	97	6,22	3,53	−0,90	15,74
1960	14	7,70	4,83	1,72	16,94
Amostra Completa	1.445	5,32	4,25	−10,56	25,97

Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

Tabela 17. *Taxa Interna de Retorno (TIR) por sexo.*

Sexo	*N*	Média (% a.a.)	Desvio-padrão	Mínimo	Máximo
Feminino	594	6,33	4,64	−10,56	25,97
Masculino	851	4,62	3,79	−4,09	21,54
Amostra Completa	1.445	5,32	4,25	−10,56	25,97

Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

Tabela 18. *Taxa Interna de Retorno (TIR) por Grupo Educacional.*

Grupo Educacional	*N*	Média (% a.a.)	Desvio-padrão	Mínimo	Máximo
Grupo Educacional 0	469	6,54	4,72	−10,56	25,97
Grupo Educacional 1	591	5,65	3,89	−2,68	22,10
Grupo Educacional 2	264	3,95	3,46	−3,99	21,54
Grupo Educacional 3	121	2,01	2,82	−4,09	17,27
Amostra Completa	1.445	5,32	4,25	−10,56	25,97

Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

Tabela 19. *Taxa Interna de Retorno (TIR) por Espécie de benefício.*

Espécie de benefício	*N*	Média (% a.a.)	Desvio-padrão	Mínimo	Máximo
AI (41)	550	7,86	4,47	−10,56	25,97
ATC (42)	895	3,77	3,24	−4,09	22,27
Amostra Completa	1.445	5,32	4,25	−10,56	25,97

Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

Tabela 20. *Taxa Interna de Retorno (TIR) por quartil de renda.*

Quartil de renda	*N*	Média (% a.a.)	Desvio-padrão	Mínimo	Máximo
Quartil 1	328	6,66	5,04	−3,12	22,10
Quartil 2	369	5,59	4,20	−4,09	20,51
Quartil 3	367	5,39	4,07	−3,99	25,97
Quartil 4	381	3,85	3,13	−10,56	18,40
Amostra Completa	1.445	5,32	4,25	−10,56	25,97

Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

gerações seguintes foram encontrados valores mais reduzidos, chegado até os 43% da geração de 1960 (ver Tabela 21). A Tabela 22 mostra que a alíquota feminina deveria ser próxima a 58% para prover equilíbrio entre contribuições e benefícios, valor mais de 12 pontos percentuais superior àquele encontrado para os homens. Da mesma forma verificada para a TIR (e de forma oposta à verificada para a TR), os valores decrescem monotonicamente para os níveis educacionais mais elevados. Há grande diferença entre os valores dos Grupos de educação 1 e 3 (Tabela 23). Para as Espécies de benefícios, verifica-se comportamento similar, com valores bem mais elevados para as AIs comparativamente às ATCs (Tabela 24). Finalmente, quando a dimensão de análise é a renda, os valores mais elevados (quase 53%) são encontrados para indivíduos de renda mais baixa (Tabela 25). À medida que a renda aumenta, o valor da alíquota diminui, até chegar ao valor médio de pouco mais de 40% para o quartil de renda mais elevada.

Os resultados encontrados para as Alíquotas Necessárias merecem alguma reflexão. Um primeiro ponto refere-se à comparação dos valores calculados, em relação aos demais indicadores. Pode-se afirmar que estes valores são bastante consistentes, se analisados à luz dos demais resultados deste trabalho. Valores mais elevados para TIR, AliqNec e TR estão associados à níveis de renda mais baixos, à aposentadoria por idade, às mulheres e aos grupos com menor nível educacional. Como subsídio adicional ao argumento pró-progressividade, apresenta-se a Tabela 26, na qual é feita uma tabulação cruzada entre os valores da TIR e da TR, por quartil. Há clara correlação entre estes dois indicadores: valores mais elevados para a TIR correspondem a valores maiores para a TR e vice-versa.

4.4. Qual a taxa de desconto adequada para um regime de repartição? A proposta do uso da TIR

É possível argumentar que a taxa de juros *de mercado* pode não ser a escolha mais adequada para um regime de repartição, tendo-se em vista a inexistência de ativos neste tipo de regime. Tal escolha eventualmente poderia ser defensável se considerarmos a ótica (ou as finanças) da entidade previdenciária. Suponha-se que por qualquer motivo, as contribuições associadas a um regime de repartição sejam eliminadas. É o caso, por exemplo, de uma mudança imediata para um regime de capitalização. Com a eliminação do fluxo de receitas contributivas, o ente previdenciário deve buscar uma alternativa, tendo em vista que é necessário fazer o *phasing-out* dos benefícios do regime de repartição.

Eliminando-se a possibilidade de emissão de moeda, a única possibilidade restante é a emissão de títulos, com aumento da dívida pública e o pagamento de juros aos compradores dos títulos. Desta forma, o custo de oportunidade para o governo, associado à eliminação das contribuições, está ligado à taxa de juros de longo prazo dos títulos públicos. Para entender o problema de outra forma vale voltar ao conceito de *Contribution Asset*, explorado no texto de Boado-Penas et al. (2008, p.92–93), juntamente com o *Actuarial Balance Sheet* do regime previdenciário. A abordagem dos autores é engenhosa, visto que conseguem reunir em um mesmo trabalho princípios econômicos associados ao regime de repartição e princípios atuariais (a preocupação com a solvência intertemporal da entidade previdenciária). Neste caso, calcular os ativos do sistema significa, de forma tautológica, calcular o valor do passivo previdenciário.

Tabela 21. Alíquota Necessária (AliqNec) por coorte.

coorte	*N*	Média (%)	Desvio-padrão	Mínimo	Máximo
1930	89	49,69	27,72	3,20	136,42
1935	315	52,98	32,66	1,65	184,32
1940	357	55,87	35,84	4,87	198,75
1945	325	48,68	28,66	5,40	166,35
1950	248	45,52	27,61	7,07	204,77
1955	97	43,73	23,23	9,16	124,92
1960	14	43,16	26,24	15,67	96,49
Amostra Completa	1.445	50,53	31,13	1,65	204,77

Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

Tabela 22. Alíquota Necessária (AliqNec) por sexo.

Sexo	*N*	Média (%)	Desvio-padrão	Mínimo	Máximo
Feminino	594	57,59	33,68	1,65	204,77
Masculino	851	45,60	28,21	3,20	198,75
Amostra Completa	1.445	50,53	31,13	1,65	204,77

Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

Tabela 23. Alíquota Necessária (AliqNec) por Grupo Educacional.

Grupo Educacional	*N*	Média (%)	Desvio-padrão	Mínimo	Máximo
Grupo Educacional 0	469	60,37	33,98	1,65	204,77
Grupo Educacional 1	591	53,95	28,79	5,45	177,61
Grupo Educacional 2	264	38,20	24,97	4,87	198,75
Grupo Educacional 3	121	22,53	12,49	3,51	102,92
Amostra Completa	1.445	50,53	31,13	1,65	204,77

Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

Tabela 24. Alíquota Necessária (AliqNec) por Espécie de benefício.

Espécie de benefício	*N*	Média (%)	Desvio-padrão	Mínimo	Máximo
AI (41)	550	69,72	31,93	1,65	198,75
ATC (42)	895	38,74	23,95	3,51	204,77
Amostra Completa	1.445	50,53	31,13	1,65	204,77

Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

Tabela 25. Alíquota Necessária (AliqNec) por quartil de renda.

Quartil de renda	*N*	Média (%)	Desvio-padrão	Mínimo	Máximo
Quartil 1	328	52,89	26,55	4,86	128,35
Quartil 2	369	55,13	35,72	3,51	168,95
Quartil 3	367	54,67	34,11	4,87	204,77
Quartil 4	381	40,05	23,84	1,65	184,32
Amostra Completa	1.445	50,53	31,13	1,65	204,77

Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

Tabela 26. Observações por quartil de TIR e quartil de TR.

Quartil	1		2		3		4		Total
	Número	%	Número	%	Número	%	Número	%	Número
1	131	34,20	98	25,59	81	21,15	73	19,06	383
2	109	28,84	94	24,87	91	24,07	84	22,22	378
3	93	24,73	101	26,86	93	24,73	89	23,67	376
4	41	13,31	81	26,30	92	29,87	94	30,52	308
Total	374	25,88	374	25,88	357	24,71	340	23,53	1.445

Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

Nesta seção propõe-se o uso da TIR para descontar os valores monetários em um regime de repartição. Esta escolha pode ser mais bem compreendida deslocando-se o foco da entidade previdenciária para o beneficiário. Um regime previdenciário é um tipo de contrato social envolvendo o governo e os trabalhadores. O recebimento do benefício está condicionado à realização de contribuições. Em outras palavras, o trabalhador aceita uma redução em sua renda líquida presente para que possa obter renda futura. A relação entre os fluxos de contribuições e de benefícios define a TIR associada à sua participação no sistema previdenciário. Ou, expressando de outra forma, o custo de oportunidade para o trabalhador, associado a não contribuir à previdência, corresponde a não obter a TIR.

Com base neste argumento, nesta seção o cálculo da Alíquota Necessário é refeito, desta vez empregando como taxa de desconto o valor da TIR individual, calculado previamente para cada beneficiário da amostra. O emprego de um valor de TIR para cada indivíduo é justificado pelo fato desta, conforme apontado anteriormente, refletir o custo de oportunidade de cada segurado. De certa forma, é como se fosse empregado como taxa de desconto o valor da *biological interest rate* inerente a um regime de repartição, apresentada pela primeira vez no clássico trabalho de Samuelson (1958), porém com agentes heterogêneos e com a generalização apresentada por Settergren & Mikula (2005, p.124). A heterogeneidade fundamenta-se nos valores diferentes obtidos para os Indicadores Previdenciários, particularmente a TIR, calculados para as categorias de análise aqui empregadas. Já a dedução apresentada por esta dupla, mostra que a TIR pode ser expressa como a soma de três termos:

(i) Mudanças nas contribuições, ligadas às alterações no nível de renda, na participação na PEA e nas alíquotas de contribuição;
(ii) Mudanças no *turnover duration*[5], ligadas a alterações na estrutura demográfica;
(iii) Juros (recebidos ou pagos) sobre os ativos (positivos ou negativos) do sistema.

[5]O *turnover duration* pode ser entendido de forma simplificada como a diferença das idades médias dos ativos e dos inativos do sistema.

O valor médio de AliqNec é bem mais próximo para a média dos indivíduos. Também se nota expressiva redução na dispersão dos valores calculados, pois o coeficiente de variação reduz-se de cerca de 0,62 para 0,44. Quando os resultados por categorias são analisados notam-se expressivas diferenças em relação aos valores originais de AliqNec, calculados com o emprego de uma taxa de desconto de 3% ao ano. Para as gerações mais novas, o valor torna-se pouco superior a 20% (Tabela 27). Outro exemplo ocorre com a alíquota por gênero. No cálculo inicial, o valor médio de AliqNec para mulheres (homens) era de 57,59% (45,60%). Empregando-se a TIR, os valores alteram-se, respectivamente, para 26,67% e 30,36 (Tabela 28). Ou seja, a Alíquota Necessária para as mulheres agora é inferior à masculina. A explicação é que inicialmente os benefícios de ambos os sexos foram trazidos a valor presente à mesma taxa, 3%. Com a TIR média, a taxa empregada para os homens aumenta para 4,62%; já para as mulheres o aumento é ainda maior, pois a TIR média é de 6,33. Este aumento afeta mais os benefícios do que as contribuições, tendo em vista que as aposentadorias encontram-se mais distantes no tempo, o que reduz mais fortemente seu valor presente. De todas as categorias analisadas, a que apresenta a mudança de resultados mais radical é a Espécie de benefício. O novo valor médio de AliqNec para as ATCs (31,61%) é sensivelmente menor que o valor para as AIs (24,34%). Novamente a explicação é que no caso das AIs, o aumento na taxa de desconto é bem mais drástico do que para as ATCs.

Tabela 27. Alíquota Necessária (AliqNec) com o uso da TIR por coorte.

Coorte	*N*	Média (%)	Desvio-padrão	Mínimo	Máximo
1930	89	30,67	12,09	3,45	78,24
1935	315	32,42	16,25	1,01	188,50
1940	357	29,20	14,71	0,99	100,28
1945	325	29,03	9,70	3,36	65,57
1950	248	26,19	7,78	2,11	92,55
1955	97	21,70	4,37	1,37	30,88
1960	14	20,19	2,95	16,09	25,50
Amostra Completa	1.445	28,85	12,68	0,99	188,50

Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

Tabela 28. Alíquota Necessária (AliqNec) com o uso da TIR por sexo.

Sexo	*N*	Média (%)	Desvio-padrão	Mínimo	Máximo
Feminino	594	26,67	14,08	0,99	188,5
Masculino	851	30,36	11,37	1,01	92,55
Amostra Completa	1.445	28,85	12,68	0,99	188,5

Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

5. CONCLUSÕES

Os resultados obtidos fornecem evidências significativas de progressividade no RGPS, para os benefícios de ATC e AI. Esta progressividade é resultado de terem sido encontrados valores mais elevados dos indicadores TR, TIR e AliqNec para as gerações mais antigas, mulheres, beneficiários com menor escolaridade e renda mais baixa. Os resultados são consistentes entre si. Este estudo corrobora para o Brasil os resultados encontrados por outros pesquisadores para vários países, com especial destaque para os EUA. No entanto, com exceção da TR, não há sinais tão claros da existência de *windfall* no RGPS. Também

Tabela 29. Alíquota Necessária (AliqNec) com o uso da TIR por Grupo Educacional.

Grupo Educacional	*N*	Média (%)	Desvio-padrão	Mínimo	Máximo
Grupo Educacional 0	469	26,25	14,33	0,99	188,50
Grupo Educacional 1	591	29,01	10,91	1,20	89,88
Grupo Educacional 2	264	31,63	12,07	1,37	100,28
Grupo Educacional 3	121	32,01	13,28	8,71	76,40
Amostra Completa	1.445	28,85	12,68	0,99	188,50

Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

Tabela 30. Alíquota Necessária (AliqNec) com o uso da TIR por Espécie de benefício.

Espécie de benefício	*N*	Média (%)	Desvio-padrão	Mínimo	Máximo
AI (41)	550	24,34	13,42	0,99	188,50
ATC (42)	895	31,61	11,36	1,37	100,28
Amostra Completa	1.445	28,85	12,68	0,99	188,50

Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

Tabela 31. Alíquota Necessária (AliqNec) com o uso da TIR por quartil de renda.

Quartil de renda	*N*	Média (%)	Desvio-padrão	Mínimo	Máximo
Quartil 1	328	25,48	13,66	0,99	78,24
Quartil 2	369	28,48	12,14	2,23	76,40
Quartil 3	367	30,04	11,86	1,01	100,28
Quartil 4	381	30,95	12,52	2,11	188,50
Amostra Completa	1.445	28,85	12,68	0,99	188,50

Fonte: Elaboração do autor, com base nos registros administrativos da Dataprev.

se pode considerar como um achado a variabilidade verificada para estes indicadores. Aparentemente, há grande heterogeneidade no RGPS que não vinha sendo captada de forma adequada em trabalhos anteriores.

O valor médio encontrado para a TR é de elevados 82,5%, bem acima da média internacional. Para todas as categorias de análise foram encontrados princípios de progressividade. Os valores encontrados também são bem mais altos do que os encontrados por outros autores, como Giambiagi & Afonso (2009) e Penafieri & Afonso (2013). Esta parece ser uma diferença relevante dos achados deste trabalho em relação a trabalhos que empregaram grupos de indivíduos representativos. Conclusões similares são obtidas com base nos resultados da TIR. Porém para este indicador, o valor médio de 5,32% a.a. não é tão mais elevado do que outros estudos encontraram.

O valor médio obtido para AliqNec, superior a 50%, é bastante elevado frente a outras evidências reportadas na literatura. Embora também tenham sido obtidas evidências de progressividade para este indicador, elas parecem ser menos significativas do que as verificadas com a TR e a TIR. Por exemplo, a diferença entre os quartis de renda é de apenas 12 pontos percentuais. Há valores muito altos, que talvez esteja associada a densidades contributivas baixas, o que faz com que o período contributivo efetivo seja baixo frente do período de recebimento do benefício. É possível que estes valores sejam resultado da arbitragem racional que os contribuintes faziam frente a regras de cálculo do valor do benefício e

condições de elegibilidade pouco rigorosas de. Não se pode descartar que, particularmente em épocas mais distantes, tenha havido maior subdeclaração de renda (e contribuições) com relação aos períodos que não faziam parte do cálculo do valor da aposentadoria, o que impactou positivamente AliqNec.

Com base nestes resultados foi feito um exercício adicional, relacionado a uma corrente da literatura cujo foco é a análise da solvência e das medidas de riqueza associadas aos sistemas previdenciários. Parte desta literatura advoga, fundamentada no conceito de *Contribution Asset* (Boado-Penas et al., 2008), que a taxa de desconto empregada para regimes de repartição não deve ser a taxa de juros de ativos sem risco. O correto seria empregar como taxa de desconto o retorno obtido pelos contribuintes por fazerem parte de um regime de repartição (Settergren & Mikula, 2005). Os valores de AliqNec foram recalculados, empregando para cada indivíduo os valores de suas TIRs individuais previamente calculados.

Esta mudança altera bastante os resultados. Para as mulheres, beneficiários da AI, com menor nível de escolaridade e dos primeiros quartis de renda, os valores de AliqNec são bem menores. Isso ocorre principalmente porque os benefícios são descontados com o emprego de taxas bem mais elevadas do que havia sido feito inicialmente. Para os grupos do extremo oposto da distribuição, ocorre fenômeno inverso. A interpretação dos resultados não é imediata. Pode ser entendido que ela é uma medida mais adequada, do ponto de vista individual da existência dos aspectos distributivos, tendo em vista a heterogeneidade verificada nos resultados das seções anteriores.

Quando analisados em conjunto os resultados deste artigo permitem aprofundar o conhecimento sobre aspectos do sistema previdenciário no Brasil, com foco nos dois benefícios programáveis mais importantes do RGPS. Ao contrário do que o senso comum usualmente costuma advogar, tais benefícios não são nem um pouco reduzidos, quando comparados às contribuições correspondentes, por meio dos indicadores calculados. Deve ser citada também o cálculo da densidade contributiva, da ordem de 70%. Esta informação não pode ser obtida com o emprego de dados em *cross-section* ou de indivíduos representativos, como tem sido feito no Brasil. De um lado, a forte evidência de progressividade captada pelos indicadores é um retrato da generosidade das regras previdenciárias. De outro, a reduzida densidade contributiva é outra evidência de que o ônus do financiamento dos benefícios recai sobre parcela reduzida dos trabalhadores. Em conjunto, tais características são o reflexo de um desenho inadequado do RGPS, tendo como resultados déficits de grande magnitude, que oneram a sociedade, impõem custos elevados e prejudicam a adequada alocação de recursos públicos.

Tendo em vista os resultados e os pontos levantados, pode-se considerar que este trabalho fez uma contribuição original à literatura brasileira sobre previdência. Evidências significativas de progressividade no RGPS, para as aposentadorias por tempo de contribuição e por idade são o achado mais relevante. De forma complementar, imagina-se que trabalhos posteriores possam seguir metodologia similar e investigar os benefícios de risco, até o momento pouco abordados no Brasil.

REFERÊNCIAS BIBLIOGRÁFICAS

Aaron, H. (1966). The social insurance paradox. *The Canadian Journal of Economics and Political Science / Revue canadienne d'Economique et de Science politique*, *32*(3), 371–374. Disponível em: http://www.jstor.org/stable/139995

Afonso, L. E., & Fernandes, R. (2005). Uma estimativa dos aspectos distributivos da previdência social no Brasil. *Revista Brasileira de Economia*, *59*(3), 295–334. Disponível em: http://bibliotecadigital.fgv.br/ojs/index.php/rbe/article/view/971

Aldrich, J. (1982). The earnings replacement rate of old-age benefits in 12 countries, 1969–80. *Social Security Bulletin*, *45*(11), 3–11. Disponível em: https://www.socialsecurity.gov/policy/docs/ssb/v45n11/v45n11p3.pdf

Alonso, J., Hoyo, C. & Tuesta, D. (2015). A model for the pension system in Mexico: Diagnosis and recommendations. *Journal of Pension Economics and Finance*, *14*(1), 76–112. doi: 10.1017/S147474721400016X

Barr, N., & Diamond, P. (2006). The economics of pensions. *Oxford Review of Economic Policy*, *22*(1), 15–39 . doi: 10.1093/oxrep/grj002

Barr, N., & Diamond, P. (2008). *Reforming pensions: Principles and policy choices*. Oxford: Oxford Univ. Press.

Barr, N., & Diamond, P. (2009). Reforming pensions: Principles, analytical errors and policy directions. *International Social Security Review*, *62*(2), 5–29. doi: 10.1111/j.1468-246X.2009.01327.x

Bertranou, F. M., & Sánchez, A. P. (2003). Características y determinantes de la densidad de aportes a la Seguridad Social en la Argentina 1994–2001. In F. M. Bertrano, M. Bourquín & H. Pena (Eds.), *Historias laborales en la Seguridad Social* (pp. 37–64). Buenos Aires: Ministerio de Trabajo, Empleo y Seguridad Social. Disponível em: http://www.ilo.org/buenosaires/publicaciones/WCMS_BAI_PUB_101/lang--es/index.htm

Biggs, A. G., Sarney, M. & Tamborini, C. R. (2009, January). *A progressivity index for social security* (Issue Paper Nº 2009-01). Washington, DC: Social Security Administration (SSA). Disponível em: https://www.ssa.gov/policy/docs/issuepapers/ip2009-01.html

Boado-Penas, M. d. C., Valdés-Prieto, S. & Vidal-Meliá, C. (2008). The actuarial balance sheet for pay-as-you-go finance: Solvency indicators for Spain and Sweden. *Fiscal Studies*, *29*(1), 89–134. doi: 10.1111/j.1475-5890.2008.00070.x

Brown, R. L. (1998). Social security: Regressive or progressive? *North American Actuarial Journal*, *2*(2), 1–23. doi: 10.1080/10920277.1998.10595687

Brown, R. L., & Ip, J. (2000). Social security: Adequacy, equity, and progressiveness. *North American Actuarial Journal*, *4*(1), 1–17. doi: 10.1080/10920277.2000.10595866

Bucheli, M., Forteza, A. & Rossi, I. (2010). Work histories and the access to contributory pensions: The case of Uruguay. *Journal of Pension Economics and Finance*, *9*(3), 369–391. doi: 10.1017/S1474747208003776

Caetano, M. A.-R. (2006, agosto). *Subsídios cruzados na previdência social brasileira* (Texto para Discussão Nº 1211). Brasília, DF: Instituto de Pesquisa Econômica Aplicada – IPEA. Disponível em: http://repositorio.ipea.gov.br/handle/11058/2183#

Duggan, J. E., Gillingham, R. & Greenlees, J. S. (1993). Returns paid to early social security cohorts. *Contemporary Economic Policy*, *11*(4), 1–13. doi: 10.1111/j.1465-7287.1993.tb00396.x

Duggan, J. E., Gillingham, R. & Greenlees, J. S. (1995). *Progressive returns to social security? An answer from social security records* (Research Paper Nº 9501). Washington DC: Treasury Department. Disponível em: http://www.treasury.gov/resource-center/economic-policy/Documents/rp9501.pdf

Feldstein, M. (2005a). Rethinking social insurance. *American Economic Review*, *95*(1), 1–24. doi: 10.1257/0002828053828545

Feldstein, M. (2005b). Structural reform of social security. *Journal of Economic Perspectives*, *19*(2), 33–55. doi: 10.1257/0895330054048731

Feldstein, M., & Liebman, J. B. (2002). Social security. In M. Feldstein & A. J. Auerbach (Eds.), *Handbook of public economics* (Vol. 4, pp. 2245–2324). Amsterdam: North-Holland.

Fernandes, F. (1994). Sistema previdenciário e desigualdades inter e intrageracionais no Brasil: O papel da dinâmica demográfica. *Previdência em Dados*, *9*(4), 5–36.

Forteza, A., Apella, I., Fajnzylber, E., Grushka, C., Rossi, I. & Sanroman, G. (2009, December). *Work histories and pension entitlements in Argentina, Chile and Uruguay* (SP Discussion Paper Nº 0926). Washington, DC: World Bank. Disponível em: http://documents.worldbank.org/curated/en/2009/12/11587564/work-histories-pension-entitlements-argentina-chile-uruguay

Forteza, A., & Ourens, G. (2009, August). *How much do Latin American pension programs promise to pay back?* (Social Protection Discussion Paper Nº 0927). Washington, DC: World Bank. Disponível em: http://documents.worldbank.org/curated/en/2009/12/11587833/much-latin-american-pension-programs-promise-pay-back

Forteza, A., & Ourens, G. (2012). Redistribution, insurance and incentives to work in Latin-American pension programs. *Journal of Pension Economics and Finance*, *11*(3), 337–364. doi: 10.1017/S1474747211000564

Garrett, D. M. (1995). The effects of differential mortality rates on the progressivity of social security. *Economic Inquiry*, *33*(3), 457–475. doi: 10.1111/j.1465-7295.1995.tb01874.x

Giambiagi, F., & Afonso, L. E. (2009). Cálculo da alíquota de contribuição previdenciária atuarialmente equilibrada: Uma aplicação ao caso brasileiro. *Revista Brasileira de Economia*, *63*(2), 153–179. Disponível em: http://bibliotecadigital.fgv.br/ojs/index.php/rbe/article/view/1207

Gragnolati, M., Jorgensen, O. H., Rocha, R. & Fruttero, A. (2011). *Growing old in an older Brazil : Implications of population ageing on growth, poverty, public finance, and service delivery* (Directions in Development /

Human Development). Washington DC: World Bank. Disponível em: https://openknowledge.worldbank.org/handle/10986/2351

Haanes-Olsen, L. (1978). Earnings-replacement rate of old-age benefits, 1965–75, selected countries. *Social Security Bulletin*, *41*(1), 3–14. Disponível em: http://www.ssa.gov/policy/docs/ssb/v41n1/v41n1p3.pdf

Hadi, A. S. (1992). Identifying multiple outliers in multivariate data. *Journal of the Royal Statistical Society. Series B (Methodological)*, *54*(3), 761–771. Disponível em: http://www.jstor.org/stable/2345856

Hadi, A. S. (1994). A modification of a method for the detection of outliers in multivariate samples. *Journal of the Royal Statistical Society. Series B (Methodological)*, *56*(2), 393–396. Disponível em: http://www.jstor.org/stable/2345910

Horlick, M. (1970). The earnings replacement rate of old-age benefits: An international comparison. *Social Security Bulletin*, *33*(3), 3–16. Disponível em: http://www.ssa.gov/policy/docs/ssb/v33n3/v33n3p3.pdf

Hurd, M. D., & Shoven, J. B. (1985). The distributional impact of social security [National Bureau of Economic Research Project Report]. In D. A. Wise (Ed.), *Pensions, labor, and individual choice* (pp. 193–222). Boston, MA: University of Chicago Press. Disponível em: http://www.nber.org/chapters/c7134

Leimer, D. R. (1999, April). *Lifetime redistribution under the Social Security Program: A literature synopsis* (ORES Working Paper Series Nº 81). Washington, DC: Social Security Administration (SSA). Disponível em: https://ssa.gov/policy/docs/workingpapers/wp81.pdf

Leimer, D. R. (2004, February). *Historical redistribution under the Social Security Old-Age and Survivors Insurance and Disability Insurance programs* (ORES Working Paper Nº 102). Washington, DC: Social Security Administration (SSA). Disponível em: https://www.ssa.gov/policy/docs/workingpapers/wp102.html

Leimer, D. R. (2007, December). *Cohort-specific measures of lifetime social security taxes and benefits* (ORES Working Paper Nº 110). Washington, DC: Social Security Administration (SSA). Disponível em: https://www.ssa.gov/policy/docs/workingpapers/wp110.html

Moura, R. L. d., Jesus Filho, J. d., Tafner, P. S. B. & Ourives, L. H. d. C. (2013). Social security effects on income distribution: A counterfactual analysis for Brazil. *Applied Economics Letters*, *20*(7), 631–637. doi: 10.1080/13504851.2012.725922

Musgrave, R. (1985). A brief history of fiscal doctrine. In M. Feldstein & A. J. Auerbach (Eds.), *Handbook of public economics* (Vol. 1, pp. 1–59). Amsterdam: North-Holland. doi: 10.1016/S1573-4420(85)80004-5

Pallares-Miralles, M., Romero, C. & Whitehouse, E. (2012, June). *International patterns of pension provision II: A worldwide overview of facts and figures* (Social Protection & Labor Discussion Paper Nº 1211). Washington, DC: The World Bank. Disponível em: https://openknowledge.worldbank.org/bitstream/handle/10986/13560/703190NWP0SPL000Box370035B00PUBLIC0.pdf

Penafieri, A. C., & Afonso, L. E. (2013). O impacto da mudança da regra de cálculo das aposentadorias por tempo de contribuição do INSS: O Fator Previdenciário é atuarialmente justo? *Economia Aplicada*, *17*(4), 667–694. doi: 10.1590/S1413-80502013000400007

Reis, C. d. B., & Turra, C. M. (2011, dezembro). Desigualdade regional na distribuição dos recursos previdenciários: Uma análise das transferências intrageracionais e de período. In *39° Encontro Nacional de Economia da ANPEC*, Foz do Iguaçu – PR. Disponível em: http://anpec.org.br/encontro/2011/inscricao/arquivos/000-014c04d01c28b616ad0a50a96758a87b.pdf

Reznik, G. L., Weaver, D. A. & Biggs, A. G. (2009, April). *Social security and marginal returns to work near retirement* (Issue Paper Nº 2009-02). Washington, DC: Social Security Administration (SSA). Disponível em: https://www.ssa.gov/policy/docs/issuepapers/ip2009-02.html

Samuelson, P. A. (1958). An exact consumption-loan model of interest with or without the social contrivance of money. *Journal of Political Economy*, *66*(6), 467–482. Disponível em: http://www.jstor.org/stable/1826989

Schröder, C. (2012). Profitability of pension contributions: Evidence from real-life employment biographies. *Journal of Pension Economics and Finance*, *11*(3), 311–336. doi: 10.1017/S147474721100031X

Settergren, O., & Mikula, B. D. (2005). The rate of return of pay-as-you-go pension systems: a more exact consumption-loan model of interest. *Journal of Pension Economics and Finance*, *4*(2), 115–138. doi: 10.1017/S1474747205002064

Shoven, J. B., & Slavov, S. N. (2012a, February). *The decision to delay social security benefits: Theory and evidence* (Working Paper Nº 17866). National Bureau of Economic Research. doi: 10.3386/w17866

Shoven, J. B., & Slavov, S. N. (2012b, July). *When does it pay to delay social security? The impact of mortality, interest rates, and program rules* (Working Paper Nº 18210). National Bureau of Economic Research. doi: 10.3386/w18210

Shoven, J. B., & Slavov, S. N. (2014). Does it pay to delay social security? *Journal of Pension Economics and Finance*, *13*(2), 121–144. doi: 10.1017/S1474747213000309

Weber, S. (2010). bacon: An effective way to detect outliers in multivariate data using Stata (and Mata). *Stata Journal*, *10*(3), 331-338. doi: http://www.stata-journal.com/article.html?article=st0197

r b e

ISSN 0034-7140
DOI 10.5935/0034-7140.20160002
Revista Brasileira de Economia, **70**(1), 31–40

Optimal Insider Strategy with Law Penalties*

José Fajardo†

Contents: **1. Introduction; 2. Model; 3. Optimization Problem; 4. Example; 5. Simulations; 6. Optimal Penalties; 7. Conclusions; Appendix. Proof of the Main Theorem.**

Keywords: **Insider Trading, Law Penalties, Social Welfare.**

JEL Code: **G11, K42.**

We study the optimal continuous trading strategy of an insider who is subject to the possibility of law penalties due to her illegal trading activity. Also, we discuss how to obtain the optimal penalty rule in order to maximize a welfare function.

Estudamos a estratégia contínua ótima de um insider que esta sujeito as penalidades da lei devidas a suas acividades de negócios ilegais. Também, analizamos como obter a regra de penalidade ótima que maximize uma funçao de bem-estar.

1. INTRODUCTION

It is very well known and documented that in many markets some traders had more information than others, more exactly they construct their strategies based on the knowledge of future events, this future information is known as the inside information and these traders are known as being the insiders of the market. Of course this kind of trading is illegal and the insiders are exposed to the risk of being penalized by the law.

The evidence of insider trade and possible implications have been analyzed in previous literature for some markets, as for example in the credit default swap and future markets, by Acharya & Johnson (2007) and Hsu & Lee (2014), respectively. Also, the optimal strategy of the insider had been analyzed in many contexts, see for example Pikovsky & Karatzas (1996), Elliott & Jeanblanc (1999), Grorud & Pontier (1998), Corcuera, Imkeller, Kohatsu-Higa, & Nualart (2004), Kohatsu-Higa & Sulem (2006) and Di Nunno, Meyer-Brandis, Øksendal, & Proske (2006), among others. But, in all of them the risk to be penalized by the existence of insider trading law is not taken into account. One way to avoid the possibility to be penalized is to use inconspicuous strategies, as in the equilibrium models of Kyle (1985) and Back (1992), see also Campi & Çetin (2007) and Ly Vath (2007). It requires aggregate demand to be a Brownian motion.[1]

*I would like to thank the comments of an anonymous referee that helps to improve the present version of the paper. Also, I would like to thank CNPq from Brazil for financial support. The usual disclaimer applies.

†Escola Brasileira de Administração Pública e de Empresas – FGV/EBAPE. Praia de Botafogo, 190, 4º e 5º andares, Rio de Janeiro. CEP 22250-900. E-mail: pepefb@gmail.com

[1]It means that when adding the insider's portfolio to the rest of portfolios in the market it remains a noise, that is, the regulator can not observe any abnormal behavior. This kind of strategy is called inconspicuous.

On the other hand, in the default literature the possibility to dishonor commitment is considered but the traders incurs in penalties proportional to the size of the dishonored part, see for example Dubey, Geanakoplos, & Shubik (2005) and Fajardo (2009). We can understand this penalties as a disutility caused by its illegal behavior, as for example loss of credit or being in jail. In the insider trading context, we can also think on penalties due to the use of insider information, in this way we can size the crime by the gains produced by the insider trading or the loss avoided. In the latter model we can observe directly the proportional penalties in the market, while in the former not.

More exactly, in the US law the penalties, both civil and criminal, for insider trading are severe. First, there are private civil remedies, as found in Section 20A of the Exchange Act of 1934. Persons who are harmed by insider trading can bring actions in most circumstances to recover the illegal profits (or avoided losses) enjoyed by wrongful traders in contemporaneous trading. Furthermore, the US Securities and Exchange Commission (SEC) has the authority to impose criminal penalties, civil penalties, and punitive civil awards against wrongful traders. Congress passed the Insider Trading Sanctions Act in 1984 to toughen penalties for illegal traders. The civil penalty in such a suit can include disgorgement of profits and a penalty of up to *three times* the ill-gotten profits.

In this paper we combine the utility penalty approach and the stochastic optimization approach, used in the above literature, to capture a more realistic insider trading behavior. Basically, given a specific kind of insider's information set[2] we solve the optimization problem for a logarithmic utility investor who in face linear utility penalties due to the possibility to incur in law penalties and then compute the additional utility obtained by the insider. More precisely, the optimal solution obtained for the insider will be a function of the probability of being captured by the market regulator and the proportional constant used in the linear penalty function, that can be interpreted as the severity of the law. Obtaining in this way an economically and meaningful strategy for the insider. Of course, in practice more severe penalties can occur, as for example the loss of the license to trade. Finally, we study the penalty rule that maximize a kind of welfare function, finding that full enforcement is the optimal penalty rule.

It is worth noting that DeMarzo, Fishman, & Hagerty (1998), analyzed the optimal enforcement strategy that maximize welfare in a model were the insider also in face penalties due to her illegal behavior but in a very different discrete-time model. Their results implies that regulators must try to minimize the bid-ask spreads observed in the market. In another related result Liu & Zhang (2011) study the impact of public information on trading behavior, finding that the existence of such public information makes the market more efficient.

The paper is organized as follows: In section 2 we introduce our model. In section 3 we obtain the optimal strategy. In section 4 and section 5 we have an example and some simulations, respectively. In section 6 we discuss the optimal penalty rule and section 7 concludes.

2. MODEL

Let $(\Omega,\mathcal{F},\mathbf{P})$ be a probability space with filtration $\mathbb{F} = (\mathcal{F}_t)_{[0,T]}$ satisfying the usual conditions, i.e. the filtration is right-continuous and each $\mathcal{F}_t$ contains all $(\mathcal{F},\mathbb{P})$-null sets. Let $T \in (0,\infty)$ be a fixed time horizon and for simplicity we assume that $\mathcal{F}_0$ is trivial up to $\mathbf{P}$-null sets.

Let G be a $\mathcal{F}$-measurable random variable with values in a Polish space $(\mathcal{U})$.[3] Then we call the filtration $\mathbb{G}$ defined by $\mathcal{G}_t = \mathcal{F}_t \vee \sigma(G)$, $t \in [0,T]$, the initially enlarged filtration of $\mathbb{F}$ and it will describe the insider's information flow.[4]

[2] In our model the insider's information set is a set of information that contains all public information and other additional information, to be specified later.

[3] A Polish space is a separable completely metrizable topological space.

[4] We recommend the reader the excellent text about enlargement of filtrations: `https://www.minet.uni-jena.de/Marie-Curie-ITN/EoF/talks/jeanblanc_introduction.pdf`

In our market there will be two agents trading in a continuous-time setting : The ordinary or honest trader who makes his decisions based on all the $\mathcal{F}_t$-measurable random variables, the insider or dishonest trader who observe the same information of the ordinary trader plus the additional information given by random variable G. We can interpret this additional information as for example G being the price of some asset at time $\tau \geq T$.

The discounted stock price process $S = (S_t)_{t\in[0,\tau]}$ is considered to be a positive continuous $(\mathbf{P},\mathbb{F})$-semimartingale. To be more exactly we consider a fixed continuous local $(\mathbf{P},\mathbb{F})$-martingale M with $M_0 = 0$ and a predictable process α such that $\mathbf{E}\Big[\int_0^T \alpha_t^2 d\langle M\rangle_t\Big] < \infty$, then

$$dS_t = S_t\Big(dM_t + \alpha_t d\langle M\rangle\Big), \qquad \forall t \in [0,T] \text{ and } S_0 > 0,$$

by Itô's formula

$$S_t = S_0 e^{\int_0^t \left(\alpha_s - \frac{1}{2}\right) d\langle M\rangle_s + M_t}.$$

Now by Theorem 2.1 in Jacod (1985) and Lemma 2.8 in Amendinger, Imkeller, & Schweizer (1998)[5] we know that M is a $\mathbb{G}$-semimartingale the following process $\widetilde{M}$ is a local $\mathbb{G}$-martingale:

$$\widetilde{M}_t = M_t - \int_0^t \beta_s^G d\langle M\rangle_s, \qquad t \in [0,T],$$

with $\beta = (\beta_t^G)$ being a $\mathcal{P}(\mathbb{F}) \otimes \mathcal{U}$-measurable process. Then from the insiders point of view the stock price evolution will be

$$dS_t = S_t\left[d\widetilde{M}_t + \left(\beta_t^G + \alpha_t\right) d\langle M\rangle_t\right], \tag{1}$$

by Itô's formula

$$S_t = S_0 e^{\int_0^t \left(\alpha_s + \beta_s^G - \frac{1}{2}\right) d\langle M\rangle_s + \widetilde{M}_t}. \tag{2}$$

We also consider that for the ordinary trader our market is complete in the sense that any bounded random variable $\mathcal{F}_T$-measurable can be replicated by a self-financing strategy, that is can be expressed as the sum of a constant and a stochastic integral respect to S. Also, for him our market is arbitrage free in the sense that there is a probability measure $\mathbf{Q}^{\text{ord}}$ equivalent to $\mathbf{P}$ on $(\Omega,\mathcal{F}_T)$ such that S is a local $(\mathbf{Q}^{\text{ord}},\mathbb{F})$-martingale.

Now in order to place the insider in complete and arbitrage free market[6] is sufficient to impose the following assumption.

Assumption 1. *The regular conditional distributions of G given $\mathcal{F}_t$ are equivalent to the law of G, i.e.:*

$$\mathbf{P}\left[G \in \cdot|\mathcal{F}_t\right](\omega) \sim \mathbf{P}\left[G \in \cdot\right], \quad \forall t \in [0,T] \text{ and } \mathbf{P} \qquad \text{a.s. } \omega \in \Omega.$$

3. OPTIMIZATION PROBLEM

Now in order to set our optimization problem we need to define utility functions, penalty functions and the set of admissible strategies available for the traders.

Definition 1. A function $U : (0,\infty) \to \mathbb{R}$ is an utility function, if

(i) It is strictly increasing, strictly concave and continuously differentiable on $(0,\infty)$

(ii) Satisfy Inada's condition:

$$\lim_{x\downarrow 0} U'(x) = \infty \quad \text{and} \quad \lim_{x\to\infty} U'(x) = 0. \tag{3}$$

[5]Both results assume that there exist a positive measure σ-finite $\xi \in (U,\mathcal{U})$ such that $\mathbf{P}[G \in \cdot|\mathcal{F}_t] \ll \xi(\cdot)$, for all $t > 0$, $\omega \in \Omega$.

[6]See Amendinger (2000) or Jeanblanc, Yor, & Chesney (2009).

Examples of utility functions are $U(x) = x^c$, $0 < c < 1$, and $U(x) = \log(x)$.

Definition 2. We say that $f : (0,\infty) \to \mathbb{R}$ is a penalty function if f is a convex function.

Examples of penalty functions are $f(x) = -cx^\gamma$, $0 < \gamma < 1$ and $c > 0$, that we can interpret as a penalty proportional to the crime and $f(x) = c(x - \overline{G})^+$, where $\overline{G}$, can be interpreted as an exogenously fixed maximum average gain level observed in the market and the constant penalty c can be interpreted as the severity of the law. That is, if you have extraordinary gains and insider trade is confirmed, you will have to pay c times the ill-gotten gains, for example three times. As we can see, this kind of linear penalty functions suits very well in our insider model, since we can observe the proportional penalty. Opposite to the default model of Dubey et al. (2005), where the proportional penalties are not directly observed in the market.

Definition 3. Let $x > 0$ be the initial wealth and denote by $\mathbb{E}$ a generic filtration.

(i) We say that an $\mathbb{R}$-valued process $\pi = (\pi_t)_{[0,T]}$ is an $\mathbb{E}$-portfolio if it is an $\mathbb{E}$-adapted predictable process and $\int_0^T \pi_t^2 d\langle M\rangle_t < \infty$, **P** a.s.

(ii) We say that $\mathbb{E}$-portfolio π is self-financing if the discounted wealth process denoted by $V(x,\pi)$ satisfies the following equation:

$$V_t(x,\pi) = x + \int_0^t \pi_s dS_s, \quad \forall t \in [0,T],$$

and we denote by $g_t(\pi) = \int_0^t \pi_s dS_s$ de gain process.

(iii) We say that $\mathbb{E}$-portfolio π is admissible if it is self-financing and $\mathbf{E}[U^-(V_T(x,\pi)) + (-f)^-(g_T(\pi))] < \infty$, where U is an utility function, f a penalty function, and $U^-(x) = \max\{0,-U(x)\}$. The set of admissible portfolios will be denoted by $\mathcal{A}_{\mathbb{E}}(x,T)$.

In order to simplify we will make the following change of variable $\eta_t = \pi_t \frac{S_t}{V_t}$ and we denote the resulting wealth process by $V_t(x,\eta)$. Then

$$dV_t(x,\eta) = \eta_t V_t(x,\eta)\frac{dS_t}{S_t}.$$

By equation (1) and Itô's formula, we have:

$$V_T(x,\eta) = xe^{\int_0^T \eta_t d\widetilde{M}_t + \int_0^T \eta_t(\beta_t^G + \alpha_t)d\langle M\rangle_t - \frac{1}{2}\int_0^T \eta_t^2 d\langle M\rangle_t}. \tag{4}$$

For abuse of notation we will denote the respective set of admissible process by $\mathcal{A}_{\mathbb{E}}(x,T)$. Now the ordinary or honest trader has to solve the following problem

$$\max_{\eta\in\mathcal{A}_{\mathbb{F}}(x,T)} \mathbf{E}\Big[U\Big(V_T(x,\eta)\Big)\Big]$$

and the insider has to solve

$$\max_{\eta\in\mathcal{A}_{\mathbb{G}}(x,T)} \mathbf{E}\Big[U\Big(V_T(x,\eta)\Big) - f\Big(g_T(\eta)\Big)\Big]$$

Now we consider the following case: $U(x) = \log(x)$ and $f(x) = \lambda_2 x$, where $\lambda_2 > 0$ is interpreted as an exogenous proportional constant that represents the severity of the civil or criminal penalty due to the illegal trading. Furthermore, we assume that there is a probability λ_1 to be discovered by regulators and

incur in penalties. Moreover, we assume that this probability is independent of the trading volume.[7] Then, the insider problem turns into

$$\max_{\eta\in\mathcal{A}_{\mathbb{G}}(x,T)} \mathbf{E}\Big[\log\Big(V_T(x,\eta)\Big) - \lambda_1\lambda_2\Big(g_T(\eta)\Big)\Big],$$

where we define $g_T(\eta) \coloneqq g_T\left(\frac{\pi}{V}\right)$.

Our main result is stated as follows:[8]

Theorem 1. *If for all $l \in \mathcal{U}$ the process $\beta^l \in L^1_{\text{loc}}(M,\mathbf{P},\mathbb{F})$, then an optimal strategy for the ordinary trader is given by $\eta^{\text{ord}}(t) = \alpha_t$, $t \in [0,T]$ and the corresponding maximal expected utility is given by*

$$\mathbf{E}\left[\log\left(V_T\left(x,\eta^{\text{ord}}\right)\right)\right] = \log x + \frac{1}{2}\mathbf{E}\left[\int_0^T \alpha_t^2 d\langle M\rangle_t\right].$$

With the optimal strategy for the insider is given by

$$\eta^{\text{ins}} = \Big(1-\lambda_1\lambda_2\Big)\Big(\alpha_t + \beta_t^G\Big), \qquad t\in[0,T] \tag{5}$$

the respective maximal expected utility is

$$\mathbf{E}\Big[\log\Big(V_T(x,\eta^{\text{ins}})\Big) - \lambda_1\lambda_2 g_T\Big(\eta^{\text{ins}}\Big)\Big] = \log x + \frac{(1-\lambda_1\lambda_2)^2}{2}\mathbf{E}\left[\int_0^T \left(\left(\beta_t^G\right)^2 + \alpha_t^2\right)d\langle M\rangle_t\right].$$

Observe that the above result holds also for λ_1 and λ_2 changing deterministically in time, i.e., we can consider $\lambda_1(t)$ and $\lambda_2(t)$.

Remark 1. It is interesting to see the following cases:

- If the probability to be captured is zero ($\lambda_1 = 0$), then we have a *full insider trading* $\eta_t^{\text{ins}} = \beta_t^G + \alpha_t$, $\forall t \in [0,T]$.
- If the probability to be capture is one ($\lambda_1 = 1$), then if we allow the penalty to change in time (more exactly $\lambda_2 = \frac{\beta_t^G}{\alpha_t+\beta_t^G}$), we obtain $\eta^{\text{ins}} = \eta^{\text{ord}}$.
- If $0 < \lambda_1 < 1$, and $0 < \lambda_2 < \infty$,[9] such that $\lambda_1\lambda_2 \gtrless 1$, Then, $\eta_t^{\text{ord}} \gtrless \eta_t^{\text{ins}}$, $\forall t \in [0,T]$.
- The additional utility of the insider is given by

$$\frac{(1-\lambda_1(t)\lambda_2)^2}{2}\mathbf{E}\left[\int_0^T \left((\beta_t^G)^2 + \alpha_t^2\right)d\langle M\rangle_t\right] - \frac{1}{2}\mathbf{E}\left[\int_0^T \alpha_t^2 d\langle M\rangle_t\right].$$

  If $\lambda_1\lambda_2 > 2$, then this additional utility is positive. It give us an explanation of why there exist insider trading even though we have strong penalties i.e. a higher λ_2, since in real markets λ_1 is too small, for example if $\lambda_1 = 0.05$, we need a penalty $\lambda_2 > 40$ in order to discourage insider trading.

[7] We can say that this assumption is quite unrealistic since a high individual trading volume can become suspicious. On the other hand higher trading volumes do not necessarily result in higher gains.

[8] Its proof is presented in an Appendix at the end of this paper.

[9] Remember that by law the penalty is finite.

4. EXAMPLE

Assume that our prices follow the Black and Scholes model, i.e., M_t is given by $\sigma(t)B_t$ and $L = B(T_0)$, $T_0 \geq T \geq t$, where $(\{B_t\})_{\{t\geq 0\}}$ is a Brownian motion and the drift is denoted by $\mu(t)$. Then

$$\pi^{\text{ins}}(t) = (1-\lambda_1\lambda_2)\left[\frac{\mu(t)-r(t)}{\sigma^2(t)} + \frac{1}{\sigma(t)}\frac{B(T_0)-B(t)}{T_0-t}\right].$$

If we choose

$$\lambda_1(t) = \frac{1-(T_0-t)^\gamma}{\lambda_2}, \qquad \gamma > 1 \quad \text{and} \quad \lambda_2 > 1-(T_0-t)^\gamma, \forall t,$$

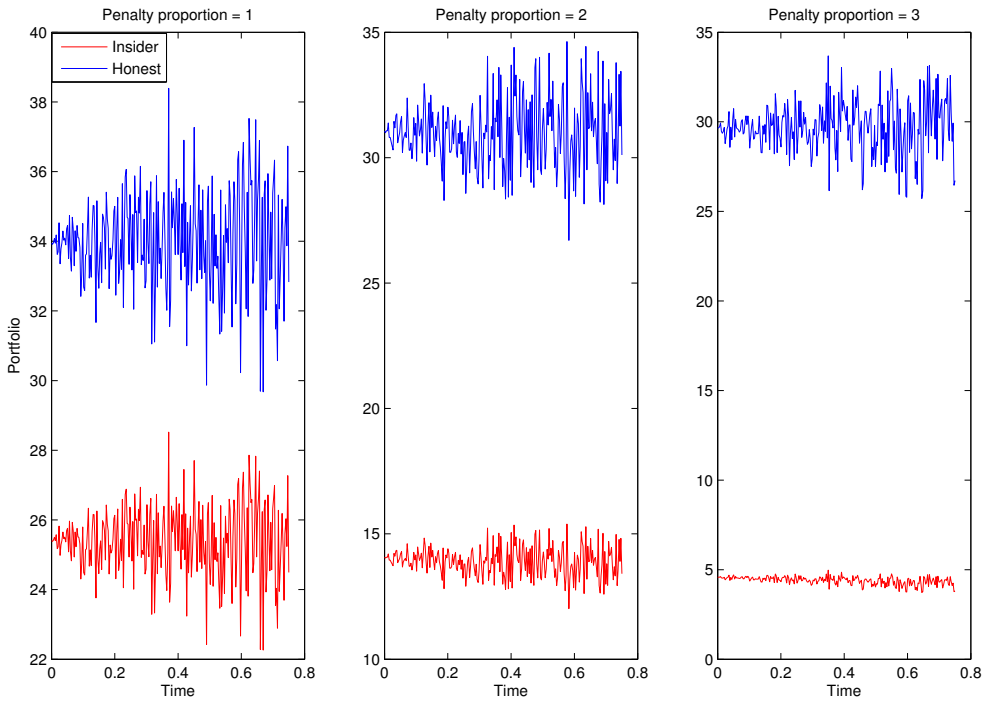
when $T_0 = T < 1$ we obtain a finite utility. This result avoids the unrealistic infinite utility found by Pikovsky & Karatzas (1996).

5. SIMULATIONS

For our simulations we consider, as in the above example, the process $M_t = \sigma_t B_t$ for all $t \leq T$ and $T \leq T_0 \leq 1$. We choose constant market coefficients $\mu(t) = 0.00085$ and $\sigma(t) = 0.035$ for all $t \leq T$, but the simulations could be easily extended to time-varying market coefficients. The annual interest rate is constant $r(t) = 0.01$, for all $t \leq T$.

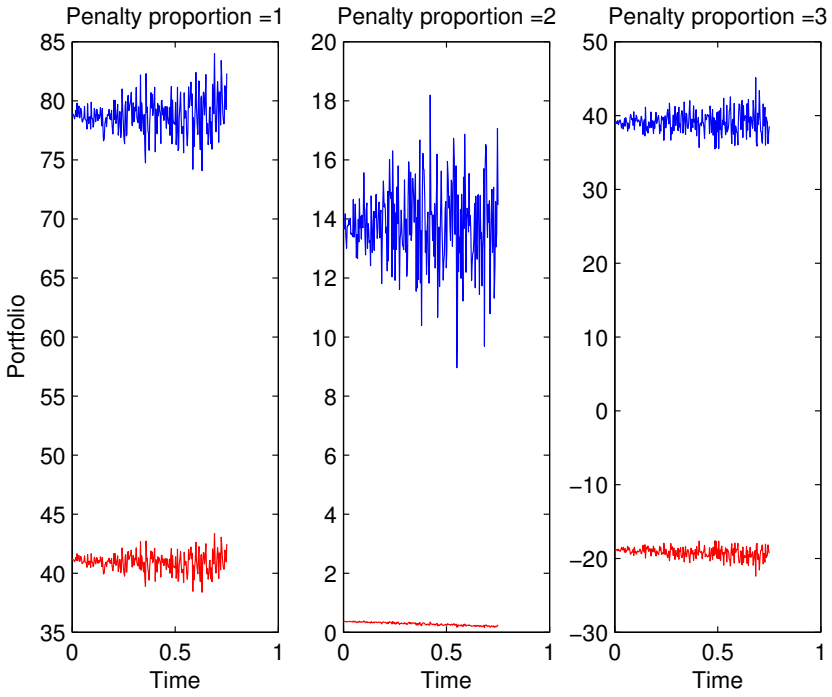
Now we consider a time-varying probability of being captured and punished $\lambda_1(t) = 1-(T_0-t)^\gamma$, $\forall t \in [0,T]$ and $\gamma > 1$, where we can understand γ as the effectiveness of the regulator in detecting and punish the insider, a greater γ means a more effective regulator. On the other hand, λ_2 represents the severity of the law penalty. In Figura 1, we show different portfolios for the ordinary and insider trader, when $\gamma = 1.5$ and $\lambda_2 \in \{1,2,3\}$.

Figure 1. Portfolios for different Penalty levels with $\gamma = 1.5$. Parameters used in the simulations $\mu(t) = 0.00085$, $\sigma(t) = 0.035$ and $r(t) = 0.01$ for all $t \leq T$.



We can see that insider controls more her trading volume when penalty is higher, because the effectiveness of the regulator is quite considerable, with $\lambda_1 \in [0.2845, 0.2885]$. But if we consider a more effective regulator, namely $\gamma = 3$, with $\lambda_1 \in [0.4880, 0.4932]$, even with less penalty the insider would need to control more her trade, as we can see in Figura 2.

Figure 2. Portfolios for different Penalty levels with $\gamma = 3$. Parameters used in the simulations $\mu(t) = 0.00085$ $\sigma(t) = 0.035$ and $r(t) = 0.01$ for all $t \leq T$.



6. OPTIMAL PENALTIES

There is a debate in Law and Economics if we must penalize or not insider trading. This debate exists in both fields during a long time, and the main arguments pass through the main question: *Is insider trading efficient in some sense or not?* Even though we let insider trading to be a crime it could be the case that insider trading be Pareto improving, as proved by Zame (1993) in the case of default penalties.

Then, in our model we can ask: *What is the penalty level that maximize the social welfare?* That is, if we take both ordinary and insider traders and put together in a welfare function, it would be possible to find an optimal penalty level for insider trading, in the sense that penalty not to be too high, since then we can reduce drastically the volume, and not too low such that every body would desire to be an insider resulting in a no trade equilibria.

It is known that insider trading has some benefits since it contributes to more informative stock prices.[10] On the other hand many critics of insider trading have argued that insider trading discourages investments due to information asymmetries, as was pointed out by Ausubel (1990). More recently, Acharya & Johnson (2010) have shown that a greater number of insiders leads to more insider trading.

There are more empirical evidence that try to establish the impact of the enforcement of insider trading law in stock prices and liquidity, see Meulbroek (1992), Bhattacharya & Daouk (2002), Ackerman,

[10] See Carlton & Fischel (1983).

Van Halteren, & Maug (2008), Beny (2007), Fernandes & Ferreira (2009).

In what follows we show a result consistent with the enforcement of insider trading law, that is we find that there is a penalty different from zero that maximize a social welfare function, in that sense we understand our penalty as an enforcement of the insider trading law. More exactly, we find that in our complete market case to forbid insider trading is the optimal enforcement rule.

Theorem 2. *Assume that there are 2 traders: ordinary and insider. Then, given the optimal choices* $(\eta^h)_{h\in\{\text{ord,ins}\}}$ *and a probability* $\lambda_1 > 0$*, the optimal penalty for any* $\gamma \in \mathbb{R}^2_{++}$*,*

$$\lambda_2^* \in \arg\max_{\lambda_2} \sum_{h\in\{\text{ord,ins}\}} \gamma^h \Big(\mathbf{E}\Big[\log\Big(V_T(x,\eta^h)\Big)\Big]\Big) - \lambda_1\lambda_2\gamma^{\text{ins}}\Big(\mathbf{E}\Big[G_T(\eta^{\text{ins}})\Big]\Big),$$

is $\lambda_2^* = \infty$.

Proof. Follows directly from the specific solutions of the optimal strategies for the ordinary and insider traders. □

It is important to notice that we are not assuming that this "welfare allocation" is an equilibrium allocation. To derive the correct equilibrium allocations we need the presence of a market maker.

7. CONCLUSIONS

In this paper we introduce law penalties to optimal insider trading models in a very simple way and show under mild conditions that there is a solution to the insider optimal strategy problem. Then, we analyze the penalty level that maximize a kind of social welfare function. We found that the full enforcement of insider trading law is optimal.

Our results indicates that legislators must try to increase the penalties to its possible maximum in order to discourage insider trading behavior. It is in line with the recommendation of Section 1079A(a)(1)(A) of the Dodd-Frank act that requires the U.S. Sentencing Commission to review its guidelines for security fraud offenses in order that penalties under these guidelines account appropriately for the potential and actual harm to the markets from such offenses.

Some important and interesting extensions to be done are the study of the equilibrium allocations and the analysis of the incomplete markets case.

REFERENCES

Acharya, V. V., & Johnson, T. C. (2007). Insider trading in credit derivatives. *Journal of Financial Economics*, *84*(1), 110–141. doi: 10.1016/j.jfineco.2006.05.003

Acharya, V. V., & Johnson, T. C. (2010). More insiders, more insider trading: Evidence from private-equity buyouts. *Journal of Financial Economics*, *98*(3), 500–523. doi: 10.1016/j.jfineco.2010.08.002

Ackerman, A., Van Halteren, J., & Maug, E. (2008, February). Insider trading legislation and acquisition announcement: Do laws matter? In *Efa 2006 zurich meetings*. doi: 10.2139/ssrn.868708

Amendinger, J. (2000). Martingale representation theorems for initially enlarged filtrations. *Stochastic Processes and their Applications*, *89*(1), 101–116. doi: 10.1016/S0304-4149(00)00015-6

Amendinger, J., Imkeller, P., & Schweizer, M. (1998). Additional logarithmic utility of an insider. *Stochastic Processes and their Applications*, *75*(2), 263–286. doi: 10.1016/S0304-4149(98)00014-3

Ausubel, L. M. (1990). Insider trading in a rational expectations economy. *The American Economic Review*, *80*(5), 1022–1041. Retrieved from http://www.jstor.org/stable/2006759

Back, K. (1992). Insider trading in continuous time. *Review of Financial Studies*, *5*(3), 387–409. doi: 10.1093/rfs/5.3.387

Beny, L. (2007). Insider trading laws and stock markets around the world: An empirical contribution to the theoretical law and economics debate. *Journal of Corporation Law*, *32*(2), 237–300.

Bhattacharya, U., & Daouk, H. (2002). The world price of insider trading. *The Journal of Finance*, *57*(1), 75–108. doi: 10.1111/1540-6261.00416

Campi, L., & Çetin, U. (2007). Insider trading in an equilibrium model with default: A passage from reduced-form to structural modelling. *Finance and Stochastics*, *11*(4), 591–602. doi: 10.1007/s00780-007-0038-4

Carlton, D. W., & Fischel, D. R. (1983). The regulation of insider trading. *Standford Law Review*, *35*, 857–895.

Corcuera, J. M., Imkeller, P., Kohatsu-Higa, A., & Nualart, D. (2004). Additional utility of insiders with imperfect dynamical information. *Finance and Stochastics*, *8*(3), 437–450. doi: 10.1007/s00780-003-0119-y

DeMarzo, P., Fishman, M., & Hagerty, K. (1998). The optimal enforcement of insider trading regulations. *Journal of Political Economy*, *106*(3), 602–632. doi: 10.1086/250023

Di Nunno, G., Meyer-Brandis, T., Øksendal, B., & Proske, F. (2006). Optimal portfolio for an insider in a market driven by Lévy processes. *Quantitative Finance*, *6*(1), 83–94. doi: 10.1080/14697680500467905

Dubey, P., Geanakoplos, J., & Shubik, M. (2005). Default and punishment in general equilibrium. *Econometrica*, *73*(1), 1–37. doi: 10.1111/j.1468-0262.2005.00563.x

Elliott, R. J., & Jeanblanc, M. (1999). Incomplete markets with jumps and informed agents. *Mathematical Methods of Operations Research*, *50*(3), 475–492. doi: 10.1007/s001860050082

Fajardo, J. (2009). Pricing and optimality with default spreads. *The Quarterly Review of Economics and Finance*, *49*(49), 686–692. doi: 10.1016/j.qref.2007.12.002

Fernandes, N., & Ferreira, M. A. (2009). Insider trading laws and stock price informativeness. *Review of Financial Studies*, *22*(5), 1845–1887. doi: 10.1093/rfs/hhn066

Grorud, A., & Pontier, M. (1998). Insider trading in a continuous time market model. *International Journal of Theoretical and Applied Finance*, *1*(3), 331–347. doi: 10.1142/S0219024998000199

Hsu, C.-H., & Lee, H.-C. (2014). Insider trading and information revelation with the introduction of futures markets. *Economic Modelling*, *43*, 173–182. doi: 10.1016/j.econmod.2014.07.037

Jacod, J. (1985). Grossissement initial, hypothèse (H'), et Théorème de Girsanov. In T. Jeulin & M. Yor (Eds.), *Grossissements de filtrations: Exemples et applications – Lecture Notes in Mathematics* (Vol. 1118, pp. 15–35). Berlin: Springer-Verlag.

Jeanblanc, M., Yor, M., & Chesney, M. (2009). *Mathematical methods for financial markets*. Springer-Verlag London. doi: 10.1007/978-1-84628-737-4

Kohatsu-Higa, A., & Sulem, A. (2006). Utility maximization in an insider influenced market. *Mathematical Finance*, *16*(1), 153–179. doi: 10.1111/j.1467-9965.2006.00266.x

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica*, *53*(6), 1315–1335. doi: 10.2307/1913210

Liu, H., & Zhang, Z. (2011). Insider trading with public and shared information. *Economic Modelling*, *28*(4), 1756–1762. doi: 10.1016/j.econmod.2011.03.009

Ly Vath, V. (2007). Competitive market equilibrium under asymmetric information. *Decisions in Economics and Finance*, *30*(2), 79–94. doi: 10.1007/s10203-007-0073-9

Meulbroek, L. K. (1992). An empirical analysis of illegal insider trading. *The Journal of Finance*, *47*(5), 1661–1699. doi: 10.1111/j.1540-6261.1992.tb04679.x

Pikovsky, I., & Karatzas, I. (1996). Anticipative portfolio optimization. *Advances in Applied Probability*, *28*(4), 1095–1122. doi: 10.2307/1428166

Zame, W. R. (1993). Efficiency and the role of default when security markets are incomplete. *The American Economic Review*, *83*(5), 1142–1164. Retrieved from http://www.jstor.org/stable/2117553

APPENDIX. PROOF OF THE MAIN THEOREM

Proof. From equation (4),

$$\log(V_T(x,\eta)) = \log x + \int_0^T \eta_t d\widetilde{M}_t + \int_0^T \eta_t \left(\beta_t^G + \alpha_t\right) d\langle M\rangle_t - \frac{1}{2}\int_0^T \eta_t^2 d\langle M\rangle_t,$$

and as $g_T(\eta) = \int_0^T \eta_t \frac{dS_t}{S_t}$ with equation (1), we have

$$g_T(\eta) = \int_0^T \eta_t d\widetilde{M}_t + \int_0^T \eta_t \left(\beta_t^G + \alpha_t\right) d\langle M\rangle_t.$$

Then

$$\log(V_T(x,\eta)) - \lambda_1\lambda_2 g_T(\eta) = \log x + (1-\lambda_1\lambda_2)\int_0^T \eta_t d\widetilde{M}_t + (1-\lambda_1\lambda_2)\int_0^T \eta_t\left(\beta_t^G + \alpha_t\right) d\langle M\rangle_t - \frac{1}{2}\int_0^T \eta_t^2 d\langle M\rangle_t,$$

and

$$\begin{aligned}\log(V_T(x,\eta)) - \lambda_1\lambda_2 g_T(\eta) = \log x &+ (1-\lambda_1\lambda_2)\int_0^T \eta_t d\widetilde{M}_t \\ &+ \frac{1}{2}\int_0^T \left[(1-\lambda_1\lambda_2)\left(\beta_t^G + \alpha_t\right)\right]^2 d\langle M\rangle_t \\ &- \frac{1}{2}\int_0^T \left[(1-\lambda_1\lambda_2)\left(\beta_t^G + \alpha_t\right) - \eta_t\right]^2 d\langle M\rangle_t.\end{aligned}$$

If $\mathbf{E}\left[\int_0^T \eta_t^2 d\langle M\rangle_t\right] < \infty$, the local $\mathbb{G}$-martingale $\int_0^T \eta_t d\widetilde{M}_t$ would be a martingale, then would have expectation zero,

$$\begin{aligned}\mathbf{E}\Big[\log(V_T(x,\eta)) - \lambda_1\lambda_2 g_T(\eta)\Big] = \log x &+ \frac{1}{2}\mathbf{E}\left[\int_0^T \left[(1-\lambda_1\lambda_2)\left(\beta_t^G + \alpha_t\right)\right]^2 d\langle M\rangle_t\right] \\ &- \frac{1}{2}\mathbf{E}\left[\int_0^T \left[(1-\lambda_1\lambda_2)\left(\beta_t^G + \alpha_t\right) - \eta_t\right]^2 d\langle M\rangle_t\right],\end{aligned}$$

from here the maximal expected utility for the insider is attained at

$$\eta^{\text{ins}} = (1-\lambda_1\lambda_2)\left(\beta_t^G + \alpha_t\right), \qquad \forall t \in [0,T].$$

Finally, it is easy to prove that $\psi^{\text{ins}} \in \mathcal{A}_{\mathbb{G}}$. The case of the ordinary trader is similar, it is enough to set $\lambda_2 = 0$ and $\beta^G \equiv 0$. □

r b e

ISSN 0034-7140
DOI 10.5935/0034-7140.20160003
Revista Brasileira de Economia, **70**(1), 41–59

Estimating Brazilian Monthly GDP: a State-Space Approach*

João Victor Issler,† Hilton Hostalacio Notini‡

Contents: **1. Introduction; 2. State-Space Model and the Kalman Filter for Interpolation; 3. Empirical Results; 4. Conclusions; Appendix. Detailed Interpolation Models.**

Keywords: **GDP Interpolation, State-Space Representation, Kalman Filter, Composite and Leading Indicators, Nowcasting, Forecasting.**

JEL Code: **C32,E32, E37.**

This paper has several contributions. The first is to employ a superior interpolation method that enables to *estimate*, *nowcast* and *forecast* monthly Brazilian GDP for 1980–2012 in an integrated way—see Bernanke, Gertler, & Watson (1997) [*Systematic monetary policy and the effects of oil price shocks* (Brookings Papers in Economic Activity No.1)]. Second, along the spirit of Mariano & Murasawa (2003) [A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics*, 18(4), 427–443], we propose and test a myriad of interpolation models and interpolation auxiliary series—all coincident with GDP from a business-cycle dating point of view. Based on these results, we finally choose the most appropriate monthly indicator for Brazilian GDP. Third, this monthly GDP estimate is compared to an economic activity indicator widely used by practitioners in Brazil—the Brazilian Economic Activity Index (IBC-Br). We found that our monthly GDP tracks economic activity better than IBC-Br. This happens by construction, since our state-space approach imposes the restriction (discipline) that our monthly estimate must add up to the quarterly observed series in any given quarter, which may not hold regarding IBC-Br. Moreover, our method has the advantage to be easily implemented: it only requires conditioning on two observed series for estimation, while estimating IBC-Br requires the availability of hundreds of monthly series. Third, in a nowcasting and forecasting exercise, we illustrate the advantages of our integrated approach. Finally, we compare the chronology of recessions of our monthly estimate with those done elsewhere.

*Esse artigo tem três contribuições originais. A primeira é o emprego de uma técnica superior de interpolação para o PIB brasileiro em bases trimestrais para a frequência mensal, método este que auxilia no nowcast e na previsão fora da amostra do crescimento do PIB de forma integrada Bernanke, Gertler, & Watson (1997) [*Systematic

*We gratefully acknowledge the comments of participants at CIRET 2008 in Santiago, Chile. We also thank Regis Bonelli and Silvia Matos for allowing us to reproduce the results of their nowcast exercise. All errors are ours. We thank Marcia Waleria Machado, Marcia Marcos, and Rafael Burjack for excellent research assistance and thank CNPq-Brazil, FAPERJ and INCT for financial support.

†Escola Brasileira de Economia e Finanças / Fundação Getúlio Vargas (EPGE/FGV). Praia de Botafogo 190, sala 1100, Rio de Janeiro – RJ, Brasil. CEP 22250-900. E-mail: joao.issler@fgv.br

‡Finance Manager at Itaú-Unibanco Holding SA.E-mail: hilton@fgvmail.br

monetary policy and the effects of oil price shocks *(Brookings Papers in Economic Activity No.1)]. A segunda, no espírito de Mariano & Murasawa (2003) [A new coincident index of business cycles based on monthly and quarterly series.* Journal of Applied Econometrics*, 18(4), 427–443], é a de propor e testar uma miríade de modelos de interpolação e de séries auxiliares para interpolar o PIB, sendo as últimas todas coincidentes ao PIB do ponto de vista dos ciclos de negócios da economia brasileira. Baseados nesses resultados, escolhemos o melhor modelo para estimar o PIB brasileiro em bases mensais. A terceira é a comparação entre as dentro e fora da amostra entre a nossa série de PIB mensal e um indicador do PIB mensal popularmente utilizado pelo mercado financeiro — o Índice de Atividade Econômica do Banco Central (IBC-Br). Nessas comparações, notamos que o nosso indicador do PIB mensal não só segue mais de perto as variações do PIB dentro da amostra, como também o prevê de forma mais precisa fora da amostra. O primeiro resultado é esperado, pois impomos a disciplina de que as variações trimestrais do PIB mensal devam ser iguais às do PIB trimestral. Porém, impor essa disciplina ao PIB interpolado também melhora a previsão fora da amostra do modelo que o contém, algo que a técnica de interpolação em si não poderia garantir, mas que observamos na prática. Esse resultado aponta na direção de que talvez a construção da série mensal do IBC-Br não seja muito eficiente, pois requer o armazenamento e acompanhamento de centenas de séries mensais, enquanto a construção do nosso PIB mensal requer apenas o armazenamento e acompanhamento de duas séries auxiliares.*

1. INTRODUCTION

Any modern society is concerned with its current "state" of economic activity and what should be that state in the near future. Entrepreneurs and individuals are interested in the question because their profits and welfare are a function of it. Governments also have an interest in the subject for budgetary and welfare issues.

For the U.S., the most educated estimate of economic-activity turning points is embodied in the binary variable announced by the NBER Business Cycle Dating Committee: recession vs. expansion for any given month. These announcements are based on the consensus of a panel of experts, and they are made some time (usually six months to one year) after a turning point in the business cycle has occurred. So, today, we only possess *old* dating's for the state of the economy. An alternative to infer business-cycle conditions would be to use Gross Domestic Product (GDP). Indeed, Stock & Watson (1989) argue that, if we were to choose one variable to best represent the state of the economy, this variable would be GDP. They claim that "[…] fluctuations in aggregate output are at the core of the business cycle so the cyclical component of real GDP is a useful proxy for the overall business cycle […]".

Unfortunately, GDP is also not timely available to continuously infer what is the current state of the economy. First, as far as we know, all countries compute GDP on a quarterly and/or on an annual frequency, but not on a monthly frequency. Second, there are delays in the release of quarterly GDP. Of course, delays vary across countries. In Brazil the delay in the initial release of quarterly GDP is bigger than three months and can reach up to six months at times. Moreover, this initial release is subject to major revisions in six months' time.

The fact that a given statistic is only available with some delay does not invalidate its later use for inference if we possess a long-span time-series of it. Indeed, referring to the decisions of the NBER Business Cycle Dating Committee, Issler & Vahid (2006) argue that:

> Suppose that we are asked to construct an index of the health status of a patient. Also, suppose that we know that the best indicator of the health of the patient is the results

> of a blood test. However, blood samples cannot be taken too frequently, and test results are only available with a lag, sometimes too long to be useful. Our index therefore must be a function of variables such as blood pressure, pulse rate and body temperature that are readily available at regular frequencies. In order to estimate the best way to combine these variables into an index, would we (i) use the historical data on these variables only, or, (ii) use the historical blood test results as well? The answer is, obviously, the latter.

Here, GDP and NBER dating are the equivalent to the results of the blood test. What should be the equivalent to blood pressure, pulse rate and body temperature? Since the pioneering work of Burns & Mitchell (1946), we have been endowed with coincident and leading indicators of economic activity, which are timely proxies, respectiverly, of the current state of economic activity and what will it be in the near future. This suggests that a superior strategy for having timely and more frequent estimates of economic activity is to combine GDP (or NBER dating) with information on these timely proxies. Issler and Vahid focused on combining NBER dating information with coincident and leading series of economic activity. Our focus here will be to combine the latter data for Brazil with Brazilian GDP.

As argued above, this strategy can improve our knowledge of GDP in two dimensions. The first is that we will possess more frequent business-cycle estimates, i.e., monthly instead of quarterly. This solves an interpolation problem that was proposed by Bernanke, Gertler, & Watson (1997) with an econometric model that encompassed most interpolation models that had been previously proposed in the literature; see also the extension in Mönch & Uhlig (2005). The second is that we will be able to *nowcast* and *forecast* GDP using information on coincident and leading series of economic activity. This solves a *nowcasting* problem, to say the least, and has been proposed by Mariano & Murasawa (2003).

The main contributions of this paper are twofold. First, we propose a joint model for Brazilian GDP and Brazilian coincident series that can be used to interpolate and nowcast GDP. Our model is based on the encompassing methods proposed by Bernanke et al. (1997) and in Mönch & Uhlig (2005). Second, for forecasting, we will propose an alternative model for Brazilian GDP and Brazilian leading series. It is essentially an extension of the model proposed by Mariano & Murasawa (2003)—who are mostly concerned with nowcasting—but will be used for forecasting purposes since we will switch the roles of coincident and leading series in implementing the model. As is clear, our main contribution is empirical, although our motivation to use and combine currently available methods is somewhat original.

In the business-cycle dimension, the ideas and models discussed in this paper are also related to the work of Stock & Watson (1988, 1991, 1993a, 1993b, 2002a, 2002b), Chauvet (1998), Kim & Nelson (1999), Cuche & Hess (2000), Liu & Hall (2001), and Mariano & Murasawa (2003). Regarding the literature on Brazilian business cycles, our paper is related to the work of Contador (1977), Cardoso (1981), Contador & Santos Filho (1987), Nakane (1994), Contador & Ferraz (1999), Spacov (2001), Issler & Spacov (2000), Chauvet (1998, 2001, 2002), Picchetti & Toledo (2002), Duarte, Issler, & Spacov (2004), Hollauer, Issler, & Notini (2009), BCB/Copom (2010), Issler, Notini, & Rodrigues (2013), Issler, Notini, & Rodrigues (2013) and Issler, Notini, Rodrigues, & Soares (2013).

The econometric models used for interpolation (also in nowcasting and forecasting) in this paper employ a state-space approach. They have three advantages over other interpolation methods: first, they allow the estimation of the unobserved monthly GDP with aggregation consistency, i.e., they ensure that the sum of three-months unobserved GDP data in a given quarter is equal to the respective quarterly observed GDP data. Second it encompasses a wide range of models in the literature—allows different ways to treat non-stationarity and different assumptions about the monthly residuals. Third, nowcasting and forecasting are computed imposing the restriction (discipline) that monthly estimates should add up to quarterly observations in-sample, which yields a superior behavior out-of-sample, as our empirical results confirm.

Finally, with the estimated monthly GDP series on hand, we (re)establish a chronology of recessions of the Brazilian economy, comparing our results with that of the recent literature and with those of the Brazilian Business-Cycle Dating Committee (CODACE). Our turning points are determined using a

standard method applied to our interpolated GDP figures—the Bry and Boschan (1971) algorithm.

The paper is organized as follows. In section 2, we review the state-space models used to interpolate quarterly GDP. In section 3 we present our main empirical results including a dating chronology for Brazil business cycles. Section 4 concludes.

2. STATE-SPACE MODEL AND THE KALMAN FILTER FOR INTERPOLATION

2.1. State-Space Representation

In this section, we give a brief review of the use of state-space models and the Kalman filter. More detailed descriptions can be found in Hamilton (1994) and Harvey (1989). The Kalman filter provides an efficient computational (recursive) mean to estimate the state of a stochastic process—usually put in state-space form. The latter encompasses relationships between observable and non-observable series and a key issue is the estimation of non-observables using past information and/or all information available.

The state-vector representation is given by a system of two vector equations. First, the *state or transition equation* describes the dynamics of the $r \times 1$ state vector of possibly unobserved series ξ_t we want to estimate. The second vector equation represents the *observation or measurement equation* linking the state vector to the $n \times 1$ vector containing the observed series (y_t) and possibly some related series (x_t) which are either exogenous or pre-determined.

The *state-space representation* of the dynamics of y_t for $t = 1,\ldots,T$,where T is the number of observations in our sample, is given by the following system of two vector equations:

$$\xi_{t+1} = F\xi_t + v_{t+1}, \tag{1}$$

$$y_t = A'x_t + H'\xi_t + w_t, \tag{2}$$

where F, A', and H', are matrices of parameters of dimensions $r \times r$, $n \times k$, and $n \times r$, respectively. So far, these matrices do not vary across time, but that can be relaxed in a more general setup. Indeed, this will be the case for interpolation, where matrix H will vary within a given year but is held fixed for any given month across different years.

Both vector equations have unpredictable error terms[1] which, for estimation purposes, are assumed to have a multivariate *Normal* distribution as follows:

$$\begin{pmatrix} v_t \\ w_t \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} Q & 0 \\ 0 & R \end{pmatrix}\right). \tag{3}$$

The coefficients matrices F, A', and H', and the two variance-covariance matrices Q and R can be estimated consistently by maximizing the conditional log-likelihood function of the system, given initial conditions $\xi_{1|0}$ and on its variance-covariance matrix, labelled $P_{1|0}$; we discuss that below. For the time being we assume that these parameters are known.

2.2. Filtering and Smoothing

Suppose we observe $y_1,x_1,y_2,x_2,\ldots,y_T,x_T$, but want to estimate the series in the unobserved state-vector ξ_t. Denote by $\mathcal{I}_t$ the information set using the observed series $y_1,x_1,y_2,x_2,\ldots,y_t,x_t$ and the conditional forecast of ξ_{t+1} using information up to t as

$$\xi_{t+1|t} = \mathbb{E}\left[\xi_{t+1} \,|\, \mathcal{I}_t\right].$$

[1] Multivariate white-noise processes.

Notice that we have assumed that (1) does not depend on x_t, which yields

$$\mathbb{E}\left[\xi_t|x_t,\mathcal{I}_{t-1}\right] = \mathbb{E}\left[\xi_t|\mathcal{I}_{t-1}\right] = \xi_{t|t-1}\,. \tag{4}$$

Also, from (1),

$$\xi_{t+1|t} = F\xi_{t|t}\,. \tag{5}$$

To link that with $\xi_{t|t-1}$, we note that the structure in (1) and (2) is linear, and recall the usual structure for updating a linear projections:

$$\xi_{t|t} = \xi_{t|t-1} + \mathbb{E}\left[\left(\xi_t - \xi_{t|t-1}\right)\left(y_t - y_{t|t-1}\right)'\right] \times \left\{\mathbb{E}\left[\left(y_t - y_{t|t-1}\right)\left(y_t - y_{t|t-1}\right)'\right]\right\}^{-1} \times \left(y_t - y_{t|t-1}\right). \tag{6}$$

Combining (1) and (6), and using standard formulas for recursive mean-squared error matrices, we obtain

$$\begin{aligned}\xi_{t+1|t} &= F\xi_{t|t-1} + F\mathbb{E}\left[\left(\xi_t - \xi_{t|t-1}\right)\left(y_t - y_{t|t-1}\right)'\right] \times \left\{\mathbb{E}\left[\left(y_t - y_{t|t-1}\right)\left(y_t - y_{t|t-1}\right)'\right]\right\}^{-1} \\ &\qquad \times \left(y_t - y_{t|t-1}\right) \qquad (7)\\ &= F\xi_{t|t-1} + FP_{t|t-1}H\left(H'P_{t|t-1}H + R\right)^{-1} \times \left(y_t - y_{t|t-1}\right). \qquad (8)\end{aligned}$$

Consider now forecasting y_t using x_t and past information on $\mathcal{I}_{t-1}$:

$$y_{t|t-1} = \mathbb{E}\left[y_t|x_t,\mathcal{I}_{t-1}\right] = A'x_t + H'\xi_{t|t-1}. \tag{9}$$

Combining (8) and (9) we get

$$\xi_{t+1|t} = F\xi_{t|t-1} + FP_{t|t-1}H(H'P_{t|t-1}H + R)^{-1} \times (y_t - A'x_t - H'\xi_{t|t-1}). \tag{10}$$

One can also show that there is a recursion for the variance-covariance matrix $P_{t+1|t}$:

$$P_{t+1|t} = F\left[P_{t|t-1} - P_{t|t-1}H(H'P_{t|t-1}H + R)^{-1}H'P_{t|t-1}\right]F' + Q. \tag{11}$$

The recursive structure in (9), (10) and (11) allows us to state the main results in computing Kalman-filter estimates. First, start the recursion with the unconditional mean and variance-covariance matrix of ξ_1,

$$\xi_{1|0} = \mathbb{E}[\xi_1] \quad \text{and} \quad P_{1|0} = \mathbb{E}\left[\left(\xi_1 - \mathbb{E}[\xi_1]\right)\left(\xi_1 - \mathbb{E}[\xi_1]\right)'\right],$$

and then iterate on (11), (10) and (9) to obtain $P_{t+1|t}$, $\xi_{t+1|t}$ and $y_{t|t-1}$, respectively, for $t = 2,\ldots,T$.

Suppose we are interested in estimating the unobserved state variable, ξ_t. There are two sets of forecasts commonly employed in a Kalman-filter setup: using the full set of observations available $(y_1,x_1,y_2,x_2,\ldots,y_T,x_T)$, which is called the *smoothed* estimate of ξ_t, or, we can forecast ξ_t using only observations available up to period $t-1$, $(y_1,x_1,y_2,x_2,\ldots,y_t,x_t)$, which is called the *filtered* estimate. Both are presented, respectively, below:

$$\xi_{t|T} = \mathbb{E}\left[\xi_t\middle|y_1,x_1,\ldots,y_T,x_T\right], \tag{12}$$

$$\xi_{t|t-1} = \mathbb{E}\left[\xi_t\middle|y_1,x_1,\ldots,y_{t-1},x_{t-1}\right]. \tag{13}$$

2.3. Estimation

Under the assumption that the errors $\begin{pmatrix} v_t \\ w_t \end{pmatrix}$ have a multivariate Normal distribution, i.e., equation (3), consistent and fully-efficient estimation of the parameters in F, A', and H', and the two variance-covariance matrices Q and R, can be performed by maximum likelihood. This assumption implies that the conditional distribution of $y_t \left| x_t, \mathcal{I}_{t-1} \right.$ is Gaussian as well, which can then be used to form the sample log-likelihood function with elements $\ln(f_t)$:

$$\begin{aligned} \sum_{t=1}^{T} \ln(f_t) &= \sum_{t=1}^{T} \ln\left[f_{y_t|x_t,\mathcal{I}_{t-1}}(\cdot) \right] \\ &= \sum_{t=1}^{T} -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\ln\left| H'P_{t|t-1}H + R \right| \\ &\qquad - \frac{1}{2}(y_t - A'x_t - H'\xi_{t|t-1})'(H'P_{t|t-1}H + R)^{-1}(y_t - A'x_t - H'\xi_{t|t-1}), \end{aligned} \tag{14}$$

where $f_{y_t|x_t,\mathcal{I}_{t-1}}(\cdot)$ denotes the (conditional) density function of $y_t|x_t,\mathcal{I}_{t-1}$, and the expressions for $\xi_{t|t-1}$ and $P_{t|t-1}$ are given in (10) and (11), respectively.

2.4. Nowcasting and Forecasting

Suppose we possess observations ranging from $t = 1,2,\ldots,T$, and estimate the parameters in F, A', and H', and the two variance-covariance matrices Q and R, by maximizing the sample log-likelihood (14). As we argued before, there is some delay in making GDP readings available timely and continuously. Thus, when we have the reading for quarterly GDP on T (comprising the months $T-2$, $T-1$, and T), calendar time already ranges somewhere from $T+3$ through $T+6$, and we already possess coincident-series monthly observations somewhere from $T+1$ through $T+4$. So although we refer to this problem as *nowcasting* because of the status of calendar-time, from an econometric point-of-view this is indeed an out-of-sample *forecasting* problem with the added twist that we possess already the realizations of the series in x_t.

Our starting point is the *smoothed* state at the end of the sample: $\xi_{T|T}$. With, that, using (5), we can then produce

$$\xi_{T+1|T} = F\xi_{T|T}. \tag{15}$$

We are now in a position to forecast y_{T+1} using the available information on x_{T+1} and our previous forecast $\xi_{T+1|T}$ given in (15). From (9),

$$\mathbb{E}\left[y_{T+1} \left| x_{T+1}, \mathcal{I}_T \right. \right] = A'x_{T+1} + H\xi_{T+1|T}. \tag{16}$$

This completes the one-step-ahead forecasts $\mathbb{E}[y_{T+1}|x_{T+1},\mathcal{I}_T]$ and $\xi_{T+1|T}$. From then on, we will still have observed x's to condition on, but our information on y's will be kept at $t = T$. The recursive structure in (9), (10) and (11) can then be used to produce nowcasts for $T+2$:

$$\mathbb{E}\left[y_{T+2} \left| x_{T+2}, \mathcal{I}_T \right. \right] = A'x_{T+2} + H\xi_{T+2|T}, \tag{17}$$

$$\begin{aligned} \xi_{T+2|T+1} = F\xi_{T+1|T} &+ FP_{T+1|T}H(H'P_{T+1|T}H + R)^{-1} \\ &\times \left(\mathbb{E}\left[y_{T+2} \left| x_{T+2}, \mathcal{I}_T \right. \right] - A'x_{T+2} - H\xi_{T+2|T} \right), \end{aligned} \tag{18}$$

where it should be noted that $P_{T+1|T}$ can be obtained by using (11):

$$P_{t+1|t} = F\left[P_{t|t-1} - P_{t|t-1}H(H'P_{t|t-1}H + R)^{-1}H'P_{t|t-1} \right]F' + Q, \tag{19}$$

which shows that, given the end-of-sample estimate $P_{T|T-1}$, one obtains immediately $P_{T+1|T}$ since we possess estimates of F, H', Q and R. The procedure for $T+h$, $h=3,4,\ldots$, can replicate that of $T+2$, as long as we possess observations on x up until $T+h$, where h is the number of additional observations we have on x vis-a-vis monthly GDP. This completes the nowcasting problem.

The truly out-of-sample problem starts when we do not possess anymore observations on the x's, which occurs after $T+h$. Here, our proposed strategy is to keep the same state-space representation given in the previous section, switching only the series in x: in the nowcasting exercise these essentially are coincident series of economic activity. In the out-of-sample exercise, these will be leading series of economic activity, i.e., series that lead economic activity and therefore are useful for forecasting.

2.5. The Encompassing Model for Interpolation

First, a word of caution about using the term *interpolation* here. GDP is a *flow* variable for which we possess quarterly observations we want to *distribute* within the months in the quarter. Stock & Watson (2010) cite Harvey (1989) to note that the problem of allocating a quarterly flow as such is referred to as *distribution*, whereas *interpolation* estimates monthly values of *stock* variables from quarterly values. Despite this technical distinction for *flow* and *stock* variables, several authors still refer to *interpolated GDP*, a term that is now ingrained in the literature, being the reason why we employ it here.

Bernanke et al. (1997) proposed a general state-space model that encompassed several competing models used in the literature for interpolation: Chow & Lin (1971), Fernández (1981), and Mitchell & Jones (2005), for example. They assume that unobserved monthly GDP (labelled as y_t^+ here) follows an $\mathrm{AR}(p)$ process explained by regressors x_t and an $\mathrm{AR}(1)$ error term. The variables in x_t are co-variates (pre-determined or exogenous), which should have a high correlation with the series being interpolated: much of the contemporaneous behavior of the interpolated series comes from them. Also in x_t are deterministic series such as a constant and/or seasonal dummies. The model proposed for unobserved monthly GDP y_t^+ is

$$\begin{aligned}\left(1-\phi_1 L-\cdots-\phi_p L^p\right)y_t^+ &= x_t\beta+u_t\\ u_t &= \rho u_{t-1}+\varepsilon_t\,.\end{aligned} \tag{20}$$

Observed quarterly GDP (labelled as y_t here) is the variable being interpolated from quarterly to monthly frequency. It relates to the interpoland series y_t^+ and obeys

$$y_t=\begin{cases}\displaystyle\sum_{i=0}^{2} y_{t-i}^+, & t=3,6,9,12,\ldots,T \qquad (21)\\ 0, & \text{otherwise.} \qquad (22)\end{cases}$$

Hence, quarterly GDP, which we can only observe on months $t=3,6,9,12,\ldots,T$, is the sum of the corresponding monthly GDPs in that quarter.[2] Otherwise, it is just set to a fictional value of zero. Notice that setting $y_t=0$ for the months we do not observe GDP is a way of making quarterly GDP observable at the monthly frequency; see Mönch & Uhlig (2005, Appendix, just above equation (1)). In the Kalman-filter literature for mixed frequency models (see, for example, Giannone, Reichlin, & Small, 2008) a fictional value is usually assumed for missing observations (zero is the most frequent choice). The crucial step is to impose that the fictional observed data has a very large variance, so that the zero value is discounted and overwritten by the Kalman-filter technique. This is exactly how Bernanke et al. (1997) and Mönch & Uhlig proceed.

[2]Note that the aggregation of monthly GDP can also be made averaging the y_t^+'s, i.e., as $y_t=\frac{1}{3}\sum_{i=0}^{2} y_{t-i}^+$. Regardless of how this is done, notice that, up until T, all estimates of monthly unobserved GDP are obtained with the restriction that they add up to quarterly observed GDP data. However, although the parameter estimates embody this restriction, it will not hold for the nowcasting and forecasting period, since we do not possess yet quarterly GDP for this period.

A second issue is how to initialize the values of β, ρ, and the variance of u_t in the Kalman-filter procedure. Mönch & Uhlig (2005) aggregate the covariates in x_t from high (monthly) to low (quarterly) frequency, and run an OLS regression of y_t on its lags and x_t (regression (20)) at the quarterly frequency. This yields estimates of β and the ϕ's, as well as an estimate for the variance of u_t. With the latter, an estimate of ρ is obtained from an OLS regression of u_t on u_{t-1}.

If we assume that the polynomial $(1-\phi_1 L-\cdots-\phi_p L^p)$ is of order one, i.e., $p=1$, with coefficient ϕ, our state-space form is the following:

$$\xi_t = \begin{pmatrix} y_t^+ \\ y_{t-1}^+ \\ y_{t-2}^+ \\ u_t \end{pmatrix} = \begin{pmatrix} \phi & 0 & 0 & \rho \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \rho \end{pmatrix} \begin{pmatrix} y_{t-1}^+ \\ y_{t-2}^+ \\ y_{t-3}^+ \\ u_{t-1} \end{pmatrix} + \begin{pmatrix} x_t\beta \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ 0 \\ 0 \\ \varepsilon_t \end{pmatrix} \tag{23}$$

$$y_t = \mathbf{H}_t'\xi_t\,, \tag{24}$$

where (23) and (24) are respectively the state and the observation equations and the matrix $\mathbf{H}_t'$ is time-varying, with the following format:

$$\mathbf{H}_t' = \begin{cases} \begin{bmatrix} 1 & 1 & 1 & 0 \end{bmatrix}, & t = 3,6,9,12,\ldots,T \\ \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}, & \text{otherwise.} \end{cases} \tag{25}$$

As mentioned before, one interesting characteristic of using the Bernanke et al. (1997) approach is that the model described in (23) and (24) encompasses several data interpolation models. They are summarized in Tabela 1. Also, the Appendix contains some discussion of their features and properties.

2.6. Goodness of Fit Statistics for Interpolated Models

To assess the quality of interpolation, Bernanke et al. (1997) propose the use two R^2 measures of fit. Denoting by $\widehat{y}_{t|T}^+$ the smoothed estimate of monthly GDP, and by $\widehat{u}_{t|T}$ the same estimate of the error term u_t, they consider:

$$R^2_{\text{level}} = \frac{\text{VAR}\left(\widehat{y}_{t|T}^+\right)}{\text{VAR}\left(\widehat{y}_{t|T}^+\right) + \text{VAR}\left(\widehat{u}_{t|T}\right)} \qquad \text{and} \qquad R^2_{\text{diff}} = \frac{\text{VAR}\left(\widehat{\Delta y}_{t|T}^+\right)}{\text{VAR}\left(\widehat{\Delta y}_{t|T}^+\right) + \text{VAR}\left(\widehat{\Delta u}_{t|T}\right)}.$$

Bernanke et al. (1997) claim that it is more informative to report the R^2 in first differences since the same statistic in levels will always be close to unity. Thus, we will compute all models listed in Tabela 1 and compare them regarding their R^2s, picking the one with the best fit.

Table 1. Resulting Model as a Function of ϕ and ρ in (23).

Model	ϕ	ρ
M_1: Static model in levels with IID residuals	0	0
M_2: Static model in levels with AR(1) residuals (Chow & Lin, 1971)	0	free
M_3: Static model in 1st differences with IID residuals (Fernández, 1981)	0	1
M_4: Dynamic model in levels with IID residuals (Mitchell & Jones, 2005)	free	0
M_5: Dynamic model in 1st differences with IID residuals	free	1
M_6: Dynamic model in levels with AR(1) residuals	free	free

Given the best models found in empirical tests, all listed in Tabela 1 above, we can check which of their properties best fit the data: does monthly GDP (y_t) follows an autoregressive (AR) process? What does the structure of the monthly residuals look like? In our interpolation exercise, we estimate the six models summarized in Tabela 1 via Maximum Likelihood using the Kalman filter. For each model, the smoothed estimate of ξ_t serves as our final interpolated monthly GDP sequence.

3. EMPIRICAL RESULTS

3.1. Data

One of the most important factors in the interpolation procedure adopted in this paper is the signal extraction from related series x_t. They represent the main information source for interpolation and must fulfill two requirements:

(i) They have to be available in the desired higher frequency of interpolated GDP (monthly in our case) in a timely fashion.

(ii) They need to have a high correlation with GDP.

The first condition imposes a strong restriction for Brazilian data, since there are only a few time-series that cover the entire estimation period (1980–2012) and are available on a monthly frequency. Following Mariano & Murasawa (2003), natural candidates for series in x_t would be the coincident series used in business-cycle analysis—Industrial Production, Sales, Income, and Employment. Unfortunately, Income and Employment are only available in Brazil after 2002, since a major change in the Monthly Employment Survey (Pesquisa Mensal do Emprego) made it virtually impossible to chain pre- and post-2002 data. Issler, Notini, & Rodrigues (2013) solved this problem by proposing a back-cast algorithm similar to the interpolation method in Bernanke et al. (1997), where pre-2002 data is back-cast on the basis of long-span covariates which are highly correlated with the back-cast series.

Our first step is to get as many series as possible that satisfy the first condition, later checking which of them satisfy the second condition. As a starting point, we considered the first two series that represent monthly economic activity—Industrial Production, labelled *IP*, and Sales, labelled *Sales*. For these variables we have long-span data, and they are believed to have cycles that are concurrent with the latent "business cycles." Since Stock & Watson (1989) argue that, if we were to choose one variable to best represent the state of business cycles, this variable would be the GDP, a natural empirical strategy to interpolate GDP would be to use not only the coincident series used in business-cycle analysis but also additional coincident series as well. The latter can be verified by standard methods—e.g., the Bry and Boschan (1971) algorithm. In order to increase the number of coincident series, we follow Cardoso (1981), from which we have selected additional four candidate series: Energy Demand (labelled *Energy*), Steel Production (labelled *Steel*), Cement Production (labelled *Cement*), and Vehicle Production (labelled *Vehicles*).

Our next step was to test if the potential candidates satisfy some desired statistics requirements (second condition). First, we test if the variables have the same integration order of GDP. Results are summarized in Tabela 2.

Based on the results of the Augmented Dickey-Fuller (ADF) test in Tabela 2, we do not reject the null hypothesis that the series are $I(1)$, which makes us conclude that all analyzed time series have the same integration order of GDP. Next, we apply first differences to all these series and compute the correlation between each first-differenced series and first-differenced GDP, which is reported in Tabela 3.[3] All the analyzed series show a high correlation with the GDP series once both are first-differenced, although

[3] We have to aggregate the monthly series in order to compute correlation coefficients. We also compute correlations using seasonally adjusted (X12) series. The results are very similar and are available upon request.

Table 2. ADF Unit Root Test.

Variables	$H_0 : z$ is $I(1)$	$H_0 : z$ is $I(2)$
GDP*	−0.898	−4.098
	[0.95]	[0.01]
Steel	−0.792	−5.514
	[0.82]	[0.00]
Vehicles	0.522	−6.208
	[0.99]	[0.00]
Cement	−0.220	−3.392
	[0.93]	[0.01]
Energy	−0.208	−26.672
	[0.93]	[0.00]
IP	0.690	−6.270
	[0.99]	[0.00]
Sales	−0.118	−6.156
	[0.95]	[0.00]

Notes: (i) All series are in logs. The specification of the test equation was chosen on the basis of the Schwarz Information Criterion; (ii) the asterisk (*) indicates that a linear trend was included in the test equation; (iii) figures in brackets are p-values.

the correlation of GDP with Industrial Production and Sales is higher than that of other series. This concludes the cyclical analysis of the auxiliary series.[4]

We also investigate the trend pattern between each candidate auxiliary series and GDP. First, we run Johansen's (1991) cointegration test in order to verify whether there is a long-run relationship between GDP and candidate auxiliary series. As expected, we found that all tested series have one common trend with GDP.[5]

Based on the series cycles high correlation and the existence of a cointegration relationship, we select Industrial Production and Sales, and possibly Cement Production as well. It is interesting to note that the traditionally Coincident Series used in business-cycles analysis are the best choices to our interpolation method what confirms their capacity to track very well the economic activity.

Table 3. Correlation Coefficients – GDP and Coincident Series.

Variables	Cement	Vehicles	Steel	Energy	IP	Sales
GDP	0.49	0.33	0.29	0.21	0.66	0.57

3.2. Evaluation Results

In this section we present the results of the Kalman-filter estimation for each model considered. Before estimation, all series have been seasonally adjusted using the X12 procedure. In Tabela 4, we present the estimated parameters and coefficients of the auxiliary variables.

[4]We also compute the correlation coefficient between each candidate series cycles (extracted by Hodrick-Prescott filter) and GDP cycle. Results are very similar to those in Tabela 3.

[5]Results are available upon request.

Table 4. Interpolation Results.

Parameter	M_1	M_2	M_3	M_4	M_5	M_6
ϕ				0.94	−0.67	−0.60
Intercept	11.03	33.51	51.60	−0.029	62.01	45.77
I. Production	0.693	0.198	0.212	0.0480	0.349	0.36
Sales	0.0002	0.0001	0.0001	0.0000	0.0001	0.0001
ρ		1.0041			1.0045	
σ	6.5881	0.8955	0.9439	1.1463	1.2570	1.3259
Log likelihood	205.6269	88.3283	93.7001	109.0752	84.9365	92.7201

Notes: (1) The models are described in the text and in the Appendix. (2) Industrial Production and Sales are used as auxiliary series x_t.

From the results above we can see that the auxiliary coefficients and the estimated parameters change according to the model being estimated. In order to choose the model that best describe our GDP series, we make use of the R^2 measures of fit to gauge the accuracy of interpolation. In Tabela 5, we present for each model, the R^2 measures of fit in first difference for the filtered and smoothed GDP.

With the exception of model M_1, all models have a high R^2, so they show a reasonable interpolation accuracy. The best R^2 measure is obtained for Model M_2, followed by model M_3. Both impose that ρ is zero in (23). They only differ about the stochastic process of the error term: model M_2 imposes that the error is an $\mathrm{AR}(1)$ process, whereas model M_3 imposes that it is an *i.i.d.* process.

Model M_2 is a contemporaneously static model with an $\mathrm{AR}(1)$ error term. Taking first quasi-difference of this model yields a restricted dynamic structure with one lag of dependent explanatory variables—the so called *common-factor* model:

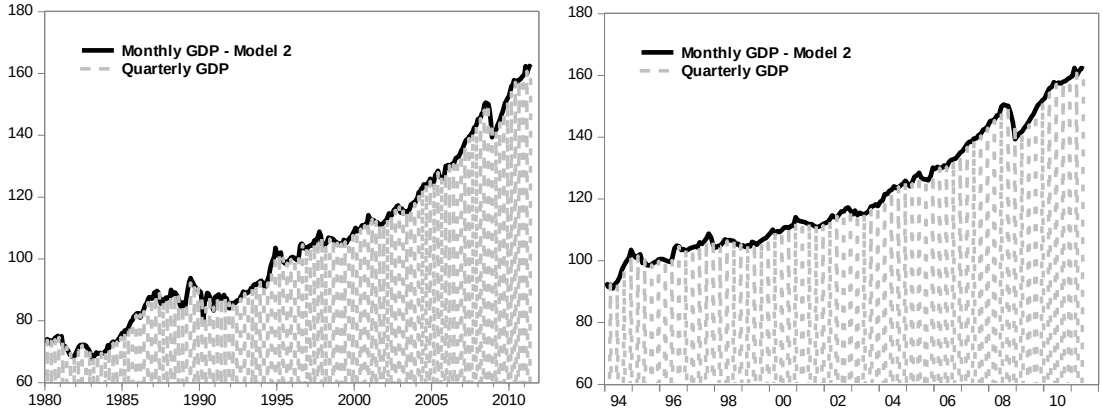
$$\begin{aligned}(1-\rho L)y_t^+ &= (1-\rho L)x_t\beta + (1-\rho L)u_t\\ y_t^+ &= \rho y_{t-1}^+ + x_t\beta - \rho\beta x_{t-1} + \varepsilon_t,\end{aligned}$$

where ε_t is white noise and there is an embedded restriction in the dynamic multipliers due to the existence of a common factor. Figura 1 depicts monthly and quarterly GDP, the former being estimated by Model M_2. The first graph shows the complete sample (1980–2012) while in the second one just the recent period (1994–2012). Both show a smooth interpolated series, which fits closely to the observed quarterly series. Next, we investigate how well interpolated GDP behaves as compared to alternative GDP proxies available in Brazil.

Table 5. R^2_{diffs} Measures of Fit.

	R^2_{diffs}	
Model	GDP_filtered	GDP_smoothed
M_1	0.0273	0.0346
M_2	0.4652	0.8097
M_3	0.4382	0.7923
M_4	0.3609	0.7098
M_5	0.4011	0.5806
M_6	0.3766	0.5736

Figure 1. Monthly (Interpolated) and Quarterly (Observed) GDP.



3.3. Comparing and evaluating the monthly GDP with other economic activity's proxies

Monthly GDP estimated in last section could be widely used by practitioners and academics alike interested in measuring in real time the state of the economic activity. It also could be used to detect turning points in the Brazilian economy. In view of that, this section compares our monthly interpolated GDP with an alternative proxy of the economic activity recently constructed by the Central Bank of Brazil—the Brazilian Economic Activity Index (IBC-Br); see BCB/Copom (2010), for example.[6]

IBC-Br is computed by aggregating monthly time-series from economy-wide supply-side sectors: agriculture, industrial sector and services. In agriculture, the source of information is the Systematic Survey of Agricultural Production released by IBGE. Regarding the industrial sector, IBC-Br monthly index is obtained by averaging the indices for four sub-sectors in our National Accounts, which are available on a monthly basis. They are weighted by value added at basic prices of the Quarterly National Accounts System in the previous year. Finally, the services sector includes time series from the trade activities, transportation, storage and mail services, financial intermediation, insurance, pension funds and related services, real estate and rentals, management, public health and education and social security, and other services.

The first advantage of our monthly GDP over IBC-Br is computational. While putting together IBC-Br requires the aggregation of hundreds of monthly series, our interpolation method just requires the use of two covariate series in estimation. Second, IBC-Br methodology does not impose any restriction (discipline) between estimated levels vis-a-vis that of GDP or between estimated growth rates vis-a-vis that of GDP. In technical language, IBC-Br and GDP need not cointegrate and the cycles in IBC-Br and GDP growth rates need not be synchronized at the quarterly frequency. On the other hand, since our interpolated monthly GDP adds up to quarterly GDP for any given quarter, it *must* display the same short- and long-run behavior of observed GDP by construction, which is the great advantage of imposing discipline at the estimation level. Finally, because our estimate is based on a state-space representation, it can be immediately used to *nowcast* and *forecast* GDP,[7] as discussed in the previous section.

With these points in mind, we now compare these two economic activity proxies. Figura 2 describes both series in levels while in Figura 3 they are depicted in first differences. As can be seen, for the period 2005-12, the two series show a very similar trend pattern. However, this is not the case for their

[6] A detailed methodological description and the complete list of component time series could be found at Brazil's Central Bank homepage: `http://www.bc.gov.br`

[7] We should also mention that the IBC-Br series only begins in 2003, which prevents its use in business-cycle analysis, which requires long-span data.

growth rates. For the latter, we time-aggregate both proxies of GDP at the quarterly frequency, and then compute their respective quarterly growth rates. These results are plotted in Figura 4. As expected, our estimated GDP proxy growth rate tracks that of quarterly GDP perfectly, which is not true regarding IBC-Br. We observe some larger discrepancies, especially in 2003-05 and 2007-08.

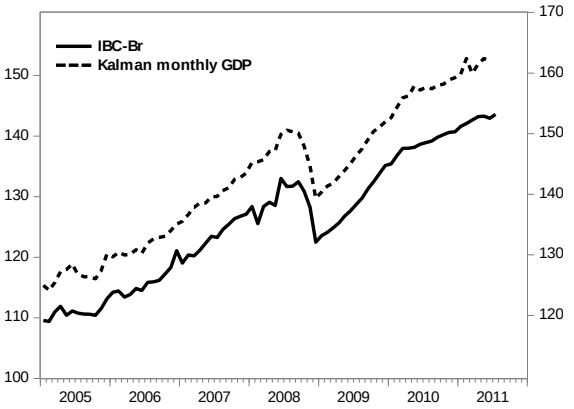


Figure 2. IBC-Br and Kalman-Filter Monthly GDP (levels).

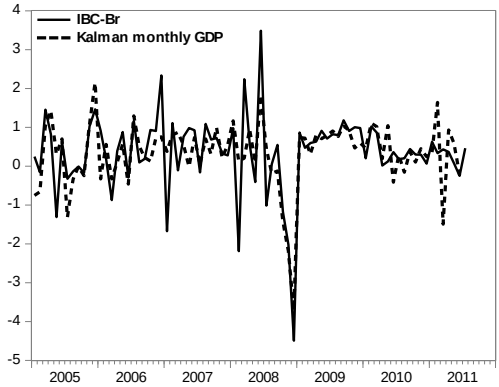
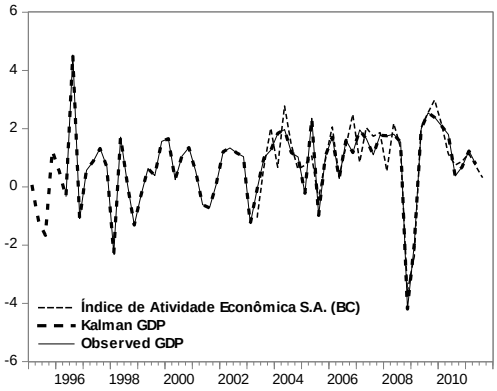


Figure 3. IBC-Br and Kalman monthly GDP in first differences.

Figure 4. Quarterly Growth Rates of GDP, IBC-Br and Kalman-filter monthly GDP.

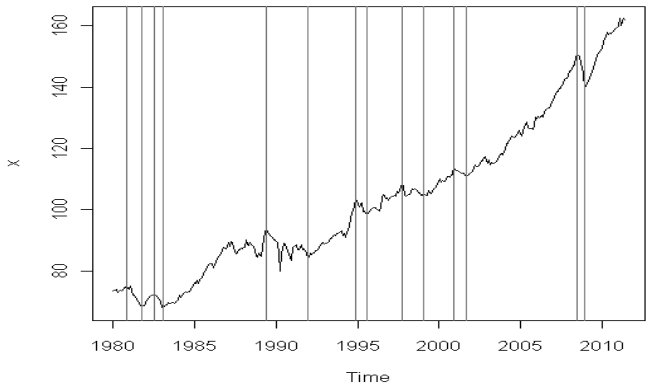


3.4. Detecting Business Cycles Turning Points

In this section, we first compute the chronology of business cycles (recessions vs. expansions) that is consistent with our monthly interpolated GDP, later comparing it with alternative business-cycle dating's, including that of the official Brazilian Business-Cycle Dating Committee (CODACE). Turning points are determined using the Bry and Boschan (1971) algorithm for detecting local minima and maxima of a time series. The dated turning points of Brazilian economic activity according to our interpolated GDP are presented below in Figura 5. In Tabela 6 we make turning-point comparisons regarding alternative datings of Brazilian economic activity.

Results in Tabela 6 show that the dating using our GDP monthly indicator yields results close to that of Duarte et al. (2004), although the latter does not cover such a wide time span as our GDP proxy. Our dating is also close to that of CODACE: on that regard, the most striking differences appear in the dating

Figure 5. Monthly GDP – shaded by Bry Boshan recessions periods.



of the 1987–88 recession and of the 2002–03 recession. Our GDP dating misses these two episodes, while it keeps a very good record vis-a-vis CODACE's datings, if we take into account the fact that the Brazilian Business-Cycle Dating Committee seems to have merged the 1980–81 recession with that of 1982–83.

Finally, we attempt here a historical account of Brazilian recessions according to our monthly GDP proxy. The recession in 1980–81 can be linked to the increase in interest rates by the FED in the early 1980s, which was later responsible for the emerging-market debt crisis (mostly in Latin America). The 1982–83 recession is related to the Latin American debt crisis itself, where international credit to these economies came to a halt after the Mexican moratorium in 1982. In these two recessions, there was an external factor at work.

Our next dated recession (1989–1991) was "home made," which came as the result of the Brazilian government inability to curb high inflation with many unsuccessful economic plans. These all had in common the absence of long-term fiscal discipline, with an initial sudden transitory contraction of money supply.

After the Real Plan, in July 1994, inflation finally gets fairly under control and most recessions are again related to events abroad, which generated capital flight and thus prompted a sudden rise in domestic interest rates as a reaction to keep foreign capital in domestic markets. The Mexican crisis in

Table 6. Turning-Point Comparisons (Peaks and Throughs).

Dated Peaks			**Dated Throughs**		
Brazilian Monthly GDP	Duarte et al.	CODACE	Brazilian Monthly GDP	Duarte et al.	CODACE
1980:11	NA	1980:10	1981:10	NA	
1982:07			1983:02		1983:02
–	1987:04	1987:02	–	1989:02	1988:10
1989:06	1989:08	1989:06	1991:12	1991:12	1991:12
1994:12	1994:12	1994:12	1995:08	1995:09	1995:09
1997:10	1997:10	1997:10	1999:02	1999:02	1999:02
2000:12	2000:12	2000:12	2001:09	2001:09	2001:09
–	NA	2002:10	–	NA	2003:06
2008:07	NA	2008:07	2008:12	NA	2009:01

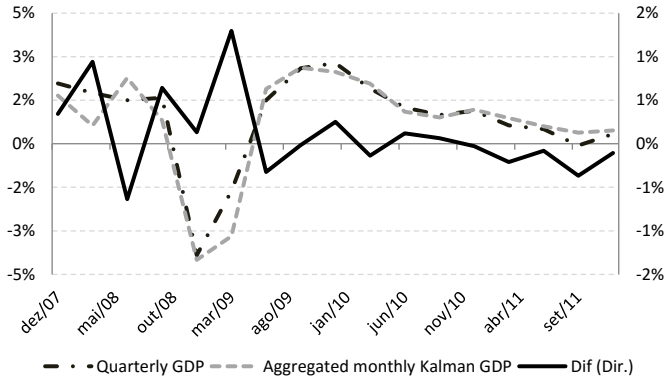
1994 affected Brazil and other emerging markets. In 1997–98, the Asian, the Russian, and the Brazilian crises had similar effects here and in other emerging markets. On these occasions domestic interest rates had risen to very high levels, leading to a reduction in domestic economic activity. In 2000-02 our local energy-supply crisis—the rationing of electrical energy for consumers and industry—was responsible to an economy-wide contraction of economic activity, while in 2008-09 we had the global financial crisis with a short and limited contraction of Brazilian economic activity.

3.5. Nowcasting GDP

In this section we present evidence that one can indeed successfully nowcast Brazilian GDP using our interpolated monthly GDP. Here, we reproduce the results of the exercise in Notini et al. (2012). Instead of resorting to the techniques outlined in section 2.4, which rely on Kalman filtering, they performed a much simpler exercise, where the conditional nowcasting model is estimated by OLS and current observations of the covariates are used to nowcast GDP every month up to one quarter ahead. This quarterly forecast is then compared with the quarterly estimate of IBC-Br—the monthly GDP series put forth by the Central Bank of Brasil—in terms of its accuracy in forecasting actual quarterly GDP.

In their exercise, Notini et al. (2012) set the forecasting horizon to be one quarter. Hence, the GDP model forecasts three months ahead in order to nowcast quarterly GDP. Figura 6 contains the results for this exercise, where the green line is the resulting error series transformed to growth rates. As can be seen below, apart from the beginning of 2009, the monthly GDP model nowcasts GDP with high accuracy.

Figure 6. Nowcast Errors of the Monthly GDP model.



Next, we perform an out-of-sample forecast comparison between the monthly GDP model and IBC-Br. Notice that the latter is widely used by the banking industry to nowcast Brazilian GDP. It is important to note that there is no informational gap between our GDP nowcast and that of IBC-Br. So this is a fair nowcasting competition. Notini et al. (2012) compute the one-quarter ahead Mean Square Error (MSE) to assess the forecast accuracy of both forecasts. This is computed for two different periods: 2007:4 to 2011:4 and 2010:4 to 2011:4. Tabela 7 summarizes the results.

Table 7. Nowcast Mean Squared Error – Monthly GDP Model Versus IBC-Br

Nowcast Period	IBC-Br	Monthly GDP Model
2007.4 – 2011.4	0.46%	0.40%
2010.4 – 2011.4	0.35%	0.16%

As can be seen from Tabela 7, the monthly GDP model outperforms IBC-Br, and its superior performance is even stronger in the recent past, showing its usefulness in nowcasting Brazilian GDP. Furthermore, we should expect the forecasting performance of the monthly GDP model to improve if we use kalman filtered coefficient estimates in nowcasting GDP instead of using OLS estimates.

4. CONCLUSIONS

In this paper, we propose an estimate for real monthly GDP in Brazil for the period 1980–2012, which is an interpolation of the quarterly observed series; see Bernanke et al. (1997) and Mönch & Uhlig (2005). Our monthly GDP proxy is based on a state-space representation which imposes the restriction (discipline) that the monthly proxy adds up to quarterly observed GDP within every single quarter. A key issue is the choice of covariates used in interpolating the quarterly series: we chose to employ Industrial Production and Sales (corrugated paper sales) as covariates in interpolation, since these two series are key in Brazilian business-cycle analysis; see Duarte et al. (2004) and Issler, Notini, & Rodrigues (2013).

The methodology in Bernanke et al. (1997) and Mönch & Uhlig (2005), allows estimating six different interpolation models, some of which have a long tradition in the literature. We are able to assess the fit of different models and related series in order to get the most appropriate monthly real GDP estimate: we evaluate six competing models, and six competing coincident series (Energy Demand, Steel Production, Cement Production, Vehicles Production, Industrial production and Sales), all available at the monthly frequency for the period 1980–2012.

First, we identify the series which behavior is closer to that of GDP, selecting Industrial Production and Sales. Second, we select the state-space model that has the highest goodness-of-fit statistic vis-a-vis GDP. We chose a contemporaneously static model with an $\mathrm{AR}(1)$ error term—which implies a restricted dynamic structure with one lag of dependent an explanatory variables. This best model is then estimated using Industrial Production and Sales as covariates yielding a smoothed estimate of Brazilian monthly GDP which serves as our GDP monthly proxy.

Next, we perform several interesting empirical exercises: we compare our GDP monthly proxy with alternative proxies available for Brazil. Our main comparison is regarding IBC-Br—the monthly GDP proxy made available by the Central Bank of Brazil. We discuss the advantages of our approach vis-a-vis theirs; we (re)establish a chronology of recessions in the recent past of the Brazilian economy using our GDP proxy. Its dating of recessions and expansions is compared with those of Duarte et al. (2004) and to those of the Brazilian Business-Cycle Dating Committee (CODACE), showing a close enough dating; we also present the results of an out-of-sample nowcasting exercise using the monthly GDP model. Finally, its forecast accuracy in forecasting actual quarterly GDP is compared with that of the quarterly estimate of IBC-Br. From 2007:4 to 2011:4, its mean-squared error is 13% smaller than that of IBC-Br, whereas, for the recent past, 2010:4 to 2011:4, it is more than 50% smaller, showing its usefulness in nowcasting Brazilian GDP.

REFERENCES

BCB/Copom (2010). Índice de Atividade Econômica do Banco Central (IBC-Br). In *Relatório de Inflação*, Vol.12 No.1, 24–28. Brasília–DF: Banco Central do Brasil. Retrieved from http://www.bcb.gov.br/?RELINF201003

Bernanke, B. S., Gertler, M., & Watson, M. (1997). *Systematic monetary policy and the effects of oil price shocks* (Brookings Papers in Economic Activity No. 1). Washington, DC: Brookings Institution. Retrieved from http://www.brookings.edu/about/projects/bpea/papers/1997/effects-of-oil-price-shocks-bernanke

Burns, A. F., & Mitchell, W. C. (1946). *Measuring business cycles*. New York: National Bureau of Economic Research (NBER).

Cardoso, E. A. (1981). Uma equação para a demanda de moeda no Brasil. *Pesquisa e Planejamento Econômico*, *11*(3), 617–655.

Chauvet, M. (1998). An econometric characterization of business cycle dynamics with factor structure and regime switching. *International Economic Review*, *39*(4), 969–996. doi: 10.2307/2527348

Chauvet, M. (2001). A monthly indicator of Brazilian GDP. *Brazilian Review of Econometrics*, *21*(1), 1–48. Retrieved from http://bibliotecadigital.fgv.br/ojs/index.php/bre/article/view/3191

Chauvet, M. (2002). The Brazilian business and growth cycles. *Revista Brasileira de Economia*, *56*(1), 75–106. doi: 10.1590/S0034-71402002000100003

Chow, G. C., & Lin, A.-l. (1971). Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. *The Review of Economics and Statistics*, *53*(4), 372–375. doi: 10.2307/1928739

Contador, C. R. (1977). *Ciclos econômicos e indicadores de atividade no Brasil* (Relatório de pesquisa No. 35). Rio de Janeiro: Instituto de Planejamento Econômico e Social / Instituto de Pesquisas (IPEA/INPES).

Contador, C. R., & Ferraz, C. (1999). *Previsão com indicadores antecedentes*. Rio de Janeiro: Silcon.

Contador, C. R., & Santos Filho, W. A. C. d. (1987). Produto Interno Bruto trimestral: bases metodológicas e estimativas. *Pesquisa e Planejamento Econômico*, *17*(3), 711–742.

Cuche, N. A., & Hess, M. K. (2000). Estimating monthly GDP in a general Kalman filter framework: Evidence from Switzerland. *Economic and Financial Modelling*, *7*(4), 153–194.

Duarte, A. J. M., Issler, J. V., & Spacov, A. (2004). Indicadores coincidentes de atividade econômica e uma cronologia de recessões para o Brasil. *Pesquisa e Planejamento Econômico*, *34*(1), 1–37.

Fernández, R. B. (1981). A methodological note on the estimation of time series. *The Review of Economics and Statistics*, *63*(3), 471–476. doi: 10.2307/1924371

Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, *55*(4), 665–676. doi: 10.1016/j.jmoneco.2008.05.010

Hamilton, J. D. (1994). *Time series analysis*. Princeton, NJ: Princeton University Press.

Harvey, A. C. (1989). *Forecasting, structural time series models and the Kalman Filter*. Cambridge University Press.

Hollauer, G., Issler, J. V., & Notini, H. H. (2009). Novo indicador coincidente para a atividade industrial brasileira. *Economia Aplicada*, *13*(1), 5–28. doi: 10.1590/S1413-80502009000100001

Issler, J. V., Notini, H. H., & Rodrigues, C. F. (2013). Constructing coincident and leading indices of economic activity for the Brazilian economy. *Journal of Business Cycle Measurement and Analysis*, *2012*(2), 43–65. doi: 10.1787/19952899

Issler, J. V., Notini, H. H., Rodrigues, C. F., & Soares, A. F. (2013). Constructing coincident indices of economic activity for the Latin American economy. *Revista Brasileira de Economia*, *67*(1), 67–96. doi: 10.1590/S0034-71402013000100004

Issler, J. V., & Spacov, A. D. (2000). *Usando correlações canônicas para identificar indicadores antecedentes e coincidentes da atividade econômica no Brasil* (Relatório de Pesquisa – mimeo). Ministério da Fazenda.

Issler, J. V., & Vahid, F. (2006). The missing link: Using the NBER recession indicator to construct coincident and leading indices of economic activity. *Journal of Econometrics*, *132*(1), 281–303. doi: 10.1016/j.jeconom.2005.01.031

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, *59*(6), 1551–1580. doi: 10.2307/2938278

Kim, C.-J., & Nelson, C. (1999, Sep). *A bayesian approach to testing for Markov switching in univariate and dynamic factor models* (Discussion Paper No. 0035). University of Washington, Department of Economics.

Liu, H., & Hall, S. G. (2001). Creating high-frequency national accounts with state-space modelling: a Monte Carlo experiment. *Journal of Forecasting*, *20*(6), 441–449. doi: 10.1002/for.810

Mariano, R. S., & Murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics*, *18*(4), 427–443. doi: 10.1002/jae.695

Mitchell, T. D., & Jones, P. D. (2005). An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology*, *25*(6), 693–712. doi: 10.1002/joc.1181

Mönch, E., & Uhlig, H. (2005). Towards a monthly business cycle chronology for the Euro Area. *Journal of Business Cycle Measurement and Analysis*, *2005*(1), 43–69. doi: 10.1787/jbcma-2005-5km7v183t48r

Nakane, M. I. (1994). *Testes de exogeneidade fraca e de superexogeneidade para a demanda por moeda no Brasil*. Rio de Janeiro: BNDES.

Notini, H., Issler, J. V., Rodrigues, C., Matos, S., , & Bonelli, R. (2012). Nowcasting Brazilian monthly GDP: A state-space approach [paper presented at CIRET 2012, Vienna].

Picchetti, P., & Toledo, C. (2002). Estimating and interpreting a common stochastic component for the Brazilian Industrial Production Index. *Revista Brasileira de Economia*, *56*(1), 107–120. doi: 10.1590/S0034-71402002000100004

Spacov, A. D. (2001). *Índices antecedentes e coincidentes da atividade econômica brasileira: Uma aplicação da análise de correlação canônica* (Dissertação de Mestrado). Fundação Getúlio Vargas, Escola de Pós-Graduação em Economia (FGV/EPGE), Rio de Janeiro.

Stock, J. H., & Watson, M. W. (1988, November). *A probability model of the coincident economic indicators* (Working Paper No. 2772). National Bureau of Economic Research (NBER). doi: 10.3386/w2772

Stock, J. H., & Watson, M. W. (1989). New indexes of coincident and leading economic indicators. In O. J. Blanchard & S. Fischer (Eds.), *NBER Macroeconomics Annual 1989, volume 4* (pp. 351–409). MIT Press. Retrieved from http://www.nber.org/chapters/c10968

Stock, J. H., & Watson, M. W. (1991). A probability model of the coincident economic indicator. In K. Lahiri & G. H. Moore (Eds.), *Leading economic indicators: New approaches and forecasting records* (pp. 63–89). Cambridge University Press.

Stock, J. H., & Watson, M. W. (Eds.). (1993a). *New research on business cycles, indicators and forecasting*. University of Chicago Press.

Stock, J. H., & Watson, M. W. (1993b). A procedure for predicting recessions with leading indicators: Econometric issues and recent experience. In J. H. Stock & M. W. Watson (Eds.), *New research on business cycles, indicators and forecasting*. University of Chicago Press.

Stock, J. H., & Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, *97*(460), 1167–1179. doi: 10.1198/016214502388618960

Stock, J. H., & Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, *20*(2), 147–162. doi: 10.1198/073500102317351921

Stock, J. H., & Watson, M. W. (2010). *Distribution of quarterly values of GDP/GDI across months within the quarter* [Research Memorandum]. Retrieved from https://www.princeton.edu/~mwatson/mgdp_gdi/Monthly_GDP_GDI_Sept20.pdf

APPENDIX. DETAILED INTERPOLATION MODELS

The literature appoint as the main problem of the models M_1 and M_2 is the fact that they extract signals only from the presumed stochastic process of the original series in a way that no new information is added. We believe that we would be better off enriching the model with additional information contained in the related series.

Model M_1
This is the simpler model which incorporates information from related series. So it can be used as a benchmark when we estimate the more sophisticated models. It can be stated as:

$$y_t = x_t'\beta + \epsilon_t$$

where: x_t' is a vector of related series and ϵ_t a iid error term with distribution $N(0,\sigma^2)$.

Model M_2
This is the Chow & Lin (1971) model. This model is very used in the literature because they are the first to show how to include related series in the interpolation procedure. They suggest a regression model without lagged dependent variables, but autoregressive errors, so we can obtain the Chow & Lin model by fixing $\phi = 0$ and by letting ρ to be estimated freely:

$$\varepsilon_t = \begin{pmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ u_t \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & \rho \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \rho \end{pmatrix} \begin{pmatrix} y_{t-1} \\ y_{t-2} \\ y_{t-3} \\ u_{t-1} \end{pmatrix} + \begin{pmatrix} x_t'\beta \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} \epsilon_t \\ 0 \\ 0 \\ \epsilon_t \end{pmatrix}$$

$$y^+ = h_t'\varepsilon_t$$

It is important to note that in their seminal paper, Chow & Lin do not use the Kalman filter but it can be shown that the Kalman filter above and the Chow & Lin regression yield the same estimates by maximum likelihood (Cuche & Hess, 2000).

Model M_3
This is a variation of the Chow & Lin model, suggested by Fernández (1981), where instead of the regression of y_t in levels, it uses the first differences of y_t in order to account for non-stationarity. This model is obtained by letting regression residuals to be a random walk, i.e. $\phi = 0$ and $\rho = 1$.

Model M_4
This model was suggested by Mitchell & Jones (2005) and it is a dynamic version of model M_1.

Model M_5
A modified version of model M_4 where we impose that $\rho = 1$.

Model M_6
It is the most general version of the Mitchell & Jones (2005) model, where we let ϕ and ρ to be estimated freely.

In our interpolation exercise, we estimate the six models summarized above via Maximum Likelihood. We then perform a set of bilateral likelihood ratio tests to discover whether the imposed restrictions are borne by the data or not and select the most appropriate model accordingly.

r b e

ISSN 0034-7140
DOI 10.5935/0034-7140.20160004
Revista Brasileira de Economia, **70**(1), 61–69

Uma Nota Sobre o Impacto do Preço do Açúcar, do Etanol e da Gasolina na Produção do Setor Sucroalcooleiro

André de Souza Melo,[*] Yony de Sá Barreto Sampaio[†]

Sumário: **1. Introdução; 2. Revisão de Literatura; 3. Metodologia; 4. Resultados; 5. Conclusões.**

Palavras-chave: **Preço da Gasolina, Mercado de Etanol, Mercado de Açúcar, Vetores Autoregressivos.**

Códigos JEL: **L71, Q16, D12.**

Este trabalho relaciona os mercados de açúcar, etanol e gasolina. O açúcar brasileiro se mostra consolidado no mercado internacional. O etanol também ganha destaque com a introdução dos veículos *flex*, surgindo um *trade-off* entre a produção de etanol e açúcar de acordo com os preços das commodities e da gasolina. Utilizando o modelo de Vetores Autorregressivos, analisou-se como as ofertas de etanol e açúcar respondem a choques dos preços do etanol, açúcar e gasolina. Como resultado, o produtor reage mais fortemente a uma mudança no preço do açúcar do que no preço do etanol, demonstrando a preferência em produzir açúcar.

This paper analyses the relationship among sugar, ethanol and gasoline markets. Brazilian sugar is a traditional commodity in international market. Ethanol production has been increasing since the introduction of flex vehicle technology. Thus it emerges a trade-off between ethanol and sugar productions. To develop this analysis, Vector Auroregressions model is used to evaluate the effect of ethanol, sugar and gasoline price shocks on ethanol and sugar supply. As a result, the producer reacts stronger to a change in sugar price than to a change ethanol price and then demonstrates a preference to produce sugar.

1. INTRODUÇÃO

O setor sucroalcooleiro no Brasil apresentou mudanças em sua dinâmica de mercado na última década. Para o etanol, o uso desse combustível como substituto da gasolina renasceu com a introdução dos veículos *flex* em 2003, conferindo flexibilidade de escolha ao consumidor entre os combustíveis. Para o açúcar, o preço internacional dessa commodity aumentou na década de 2000 tendo inclusive aumentado no final da década, em momento em que a maioria das commodities teve seu preço reduzido.

[*]Universidade Federal Rural de Pernambuco / Programa de Pós-Graduação em Administração e Desenvolvimento Rural (UFRPE/PADR). Av. Dom Manoel de Medeiros, s/n, Dois Irmão, Recife–PE. CEP 52171-900. E-mail: andredesouzam@gmail.com

[†]Universidade Federal de Pernambuco / Programa de Pós-Graduação em Economia. Av. dos Economistas, s/n, Recife–PE. CEP 50740-590. E-mail: yonysampaio@gmail.com

Desde o fim da década de 90, o novo contexto de mercados desregulamentados afetou principalmente a forma de comercialização e os preços dos principais produtos finais — o açúcar e o álcool, e da cana-de-açúcar, matéria-prima básica do setor. A avaliação dessas adequações pode ser feita pela observação de mudanças no mix de produção de açúcar e álcool, decorrente de alterações nos preços relativos desses produtos, à medida que eles deixaram de ser estabelecidos pelo governo e passaram a ser determinados sob condições de mercado competitivo (Boff, 2011; Moraes, 2000; Schmidhuber, 2007). Dessa forma, surge um *trade off* entre a produção das duas commodities, pois a indústria sucroalcooleira é caracterizada pela produção dual permitindo a produção de etanol e açúcar em proporções variáveis (Müller, Schmidhuber, Hoogeveen & Steduto, 2007; von Lampe, 2006).

Balcombe & Rapsomanikis (2008) detalham que enquanto o custo de produção de etanol estiver menor que o preço da gasolina, e o valor do produto marginal da cana-de-açúcar no mercado de biocombustíveis exceder o do mercado de açúcar, a cana-de-açúcar será desviada para a produção de etanol, assim aumentando o custo marginal do açúcar. Dessa forma, pode-se observar que a gasolina também é um componente dessa dinâmica do setor sucroalcooleiro. Viegas (2011) destaca que a substitutibilidade entre a gasolina e o etanol criou a oportunidade desse último se destacar no mercado brasileiro de combustíveis. No entanto, com a crise internacional, com a elevação do preço internacional do açúcar, com o aumento do preço da terra, e com a escassez de investimentos em novas usinas influenciou na preferência do produtor em produzir açúcar com relação ao etanol.

Dessa forma, propõe-se realizar um teste econométrico que unifique os dois mercados — gasolina e etanol — com um mercado relacionado com o etanol — o açúcar. Pretende-se, então, analisar como as ofertas de açúcar e etanol reagem a mudanças nos preços do açúcar, etanol e gasolina no Brasil.

2. REVISÃO DE LITERATURA

De acordo com estudo do Centro de Estudos Avançados em Economia Aplicada (CEPEA) (2011) a oferta e a demanda de etanol são determinantes do comportamento do preço de etanol e do açúcar no lado da oferta, então o montante de cana-de-açúcar será alocado em maior quantidade para o mercado mais atrativo e, assim, ocorre uma mudança nos preços. Boff (2011) realiza um estudo semelhante, porém apontando a oferta de etanol e açúcar sendo causadas pelo preço do etanol e açúcar. Os estudos de Moraes (2000), Schmidhuber (2007), Balcombe & Rapsomanikis (2008) e Müller et al. (2007) comentam sobre a flexibilidade na produção entre etanol e açúcar de acordo com o mercado.[1] Eles destacam que o processo de desregulamentação do setor acarretou em uma relação sólida entre os mercados de açúcar e etanol no Brasil.

A literatura também aponta que a entrada da nova tecnologia dos veículos flex ocasionou aumento na demanda de etanol, refletindo na maior produção desta commodity com relação ao açúcar (Cabrini & Marjotta-Maistro, 2007; Giesecke, Horridge & Scaramucci, 2007; Serra, Zilberman & Gil, 2011; Silva & Almeida, 2006). Isso refletiria em um aumento no preço desse último produto (Müller et al., 2007; Tokgoz & Elobeid, 2006). Silva & Almeida (2006) complementam que a relação entre os dois mercados foi modificada pelo persistente aumento do preço do petróleo. O estudo aponta na direção de uma forte ligação dos mercados de açúcar e petróleo que passam pela participação da produção de etanol derivada do açúcar. Schmidhuber (2007), FAO (2008) e Balcombe & Rapsomanikis (2008) comprovam essa interligação entre os três mercados.

Balcombe & Rapsomanikis (2008) e Silva & Almeida (2006) mostram uma hierarquia de causalidade dos mercados de petróleo para açúcar e do açúcar para o etanol. Balcombe & Rapsomanikis (2008) também destacam o papel do preço da gasolina nessa dinâmica. Ou seja, no Brasil o mercado de açúcar exerce maior influência no etanol do que vice versa. Bacchi (2005) comprova essa relação apontando que as variações no preço de açúcar no mercado doméstico, "produto que tem sido considerado carro chefe

[1] Para Smeets et al. (2008), existe uma correlação inversa entre produção de etanol e produção de açúcar. Como muitas usinas produzem açúcar e etanol, a razão da produção de açúcar para o etanol depende do preço relativo desses produtos.

do setor sucroalcooleiro" (Bacchi, 2005), influenciam o preço do etanol.

Para justificar essa preferência do produtor pelo açúcar, Costa, Burnquist & Guilhoto (2006) e Barros (2011) apontam que o mercado de açúcar é mais consolidado do que o mercado de etanol. Segundo eles o mercado de açúcar guarda fortes laços com o setor externo e que os preços domésticos refletem as condições de oferta e demanda global. O mercado de etanol, por sua vez, acompanha as oscilações do preço do açúcar, porém de uma forma mais suave. De acordo Barros (2011), 50% das oscilações do preço do açúcar passam para o etanol. Nos últimos 15 anos a relação de preços entre etanol e açúcar caiu cerca de 60%. Ou seja, o barateamento do etanol demonstra também uma expressiva flexibilidade entre os preços das commodities.

De acordo com Boff (2011) e Viegas (2011) a mistura imposta pelo governo do etanol na gasolina é um fator que gera estabilidade nas produções de etanol e consequentemente de açúcar. O primeiro autor modela o comportamento da oferta de etanol de acordo com mudanças nessa mistura.

Diante do cenário da crise internacional, a oferta de etanol sofreu impactos oriundos de ausência de investimentos em novas usinas e com aumento do preço do açúcar no mercado. Com isso, a demanda interna de etanol não era totalmente suprida.[2] Viegas (2011) defende a intervenção governamental na forma de incentivos para toda cadeia produtiva sucroalcooleira. Entre as intervenções propostas pelo setor é a liberalização do preço da gasolina na refinaria. Segundo Barros (2011) o governo deve ser prudente no estimulo à produção de etanol, pois poderá punir a produção e a exportação de açúcar.

Portanto, a literatura aponta para uma relação de correlação e causalidade entre os mercados de açúcar, etanol e gasolina. A marcante interdependência entre os três mercados sugere uma relação de multicausalidade entre eles. Dessa forma, utiliza-se o modelo de Vetores Autorregressivos (VAR) para captar a dinâmica das variáveis e constatar o sentido da causalidade a qual tem implicações decisivas para a política econômica.

3. METODOLOGIA

No modelo VAR, todas as variáveis devem ser endógenas. Seguindo a revisão, as variáveis escolhidas são a produção de açúcar, a produção de etanol, o preço do açúcar, o preço do etanol e o preço da gasolina.

A estratégia empírica envolve estimar o modelo na forma reduzida e depois recuperar os parâmetros da forma estrutural. Tipicamente, a literatura que investiga os efeitos de choques sobre as variáveis macroeconômicas, dentre outros, segue Sims (1980), e impõe um esquema de identificação recursivo, através da imposição de restrições na matriz de efeitos contemporâneos. Um exemplo de esquema de identificação que estabelece restrições nos parâmetros contemporâneos é a decomposição de Cholesky. A ordem de entrada é mostrada na equação (1). Os choques identificados pelo modelo que fazem parte do escopo do trabalho serão: choque no preço da gasolina, choque no preço do etanol e choque no preço do açúcar. Analisam-se como esses choques afetam a produção de etanol e açúcar.

Então o modelo a ser estimado, de acordo com a identificação de forma recursiva de Cholesky é descrito como

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ a_{21} & 1 & 0 & 0 & 0 \\ a_{31} & a_{32} & 1 & 0 & 0 \\ a_{41} & a_{42} & a_{43} & 1 & 0 \\ a_{51} & a_{52} & a_{53} & a_{54} & 1 \end{bmatrix} \begin{bmatrix} \mathrm{PROD}_t^{\mathrm{AC}} \\ \mathrm{PROD}_t^{\mathrm{ET}} \\ P_t^{\mathrm{AC}} \\ P_t^{\mathrm{ET}} \\ P_t^{\mathrm{GAS}} \end{bmatrix} = \begin{bmatrix} F \end{bmatrix} \begin{bmatrix} \mathrm{PROD}_{t-1}^{\mathrm{AC}} \\ \mathrm{PROD}_{t-1}^{\mathrm{ET}} \\ P_{t-1}^{\mathrm{AC}} \\ P_{t-1}^{\mathrm{ET}} \\ P_{t-1}^{\mathrm{GAS}} \end{bmatrix} + C\xi, \tag{1}$$

onde P_t^{GAS} é o preço médio da gasolina na bomba ao longo do tempo; P_t^{ET} é o preço do etanol hidratado ao produtor; P_t^{AC} representa a o preço de açúcar ao produtor; $\mathrm{PROD}_t^{\mathrm{ET}}$ representa a produção de etanol ao longo do tempo; e $\mathrm{PROD}_t^{\mathrm{AC}}$ a produção de açúcar ao longo do tempo. Além disso, utilizou-se como

[2] Ver `http://www.datagro.com.br`

variáveis exógenas: uma dummy temporal representando a introdução dos veículos flex no Brasil em 2003, uma dummy de intervenção a mistura etanol e gasolina, com valor um para intervenção acima de 24%, e uma dummy temporal representando a crise de 2009.

As variáveis estão em nível, para detectar efeitos de persistência nos choques nas variáveis de interesse. De acordo com Kilian (2010), a vantagem de especificação no nível é que a estimação do VAR é consistente mesmo que as variáveis sejam integradas ou não.

Os dados de produção de etanol anidro e hidratado foram obtidos pela Datagro;[3] os dados de preço de etanol ao produtor (em R$/m³) e do açúcar ao produtor do mercado externo[4] (em R$/t) ao produtor, do Centro de Estudos Avançados em Economia Aplicada (CEPEA); e preço da gasolina (R$/m³) da Agência Nacional de Petróleo e Biocombustível (ANP). Todas as variáveis possuem periodicidade mensal de julho de 2001 até março de 2012. As variáveis foram colocadas em logaritmo para visualizar com maior facilidade as elasticidades.

4. RESULTADOS

Para constatar a estacionariedade das séries, utilizaram-se os testes de raiz unitária de Dickey-Fuller Aumentado (ADF), o teste KPSS (com 5% de significância) e o teste de Phillip Perron de quebra estrutural. Os testes, considerando a presença de intercepto e tendência, mostraram que todas as séries foram estacionárias, exceto o preço do açúcar.

Para analisar a endogeneidade das variáveis em questão utilizou-se o teste de Causalidade de Granger. O resultado do teste indica certo grau de endogeneidade entre as variáveis, destacando que a produção de açúcar não é afetada por nenhuma das variáveis e que os preços da gasolina afeta a produção de etanol com 5% de significância. Em seguida foi determinado o número de defasagens do VAR, para isso utilizou-se o teste de razão de verossimilhança e o teste de Akaike para poucas observações. O número de defasagens indicado pelo teste foi de segunda ordem.

Pela estimação do VAR estrutural extrai-se a função impulso resposta, na qual é identificado como cada variável responde a choques. A Figura 1 mostra que um choque no preço de etanol causa um efeito temporário de queda na produção de açúcar e de aumento na produção de etanol. Com a queda na produção de açúcar, o preço é aumentado. Assim, o açúcar se torna o produto mais atrativo no curto prazo devido ao aumento do preço, então a produção de açúcar torna a crescer.

A Figura 2 ilustra a resposta das produções de açúcar, etanol o preço de etanol quando ocorre um choque no preço do açúcar. Pode-se perceber que, após um choque no preço do açúcar recebido pelo produtor, as produções de açúcar e etanol sofrem um efeito mais prolongado. Pode-se dizer que a atratividade do preço do açúcar recebido ao produtor no mercado externo, possui um maior impacto na dinâmica do setor com relação ao preço doméstico do etanol.

A Figura 3 mostra o impacto do preço da gasolina na dinâmica do setor sucroalcooleiro. Um choque positivo no preço da gasolina causa um aumento na produção de etanol. Como a procura por etanol aumenta devido ao aumento do preço da gasolina, a produção de etanol varia positivamente, porém o preço da commodity também aumenta ao longo do tempo. A mesma dinâmica é vista para o mercado de açúcar, com um choque positivo no preço da gasolina, mais etanol é demandado, menos açúcar é produzido, com isso o preço do açúcar aumenta, aumentando assim a produção de açúcar. Nota-se que esses efeitos são vistos no curto e no longo prazo.

A Tabela 1 mostra como as variáveis de produção se comportam a uma mudança relativa de 1% nos preços do etanol, açúcar e gasolina. Para um choque de 1% no preço de etanol, a maior resposta na produção de açúcar e de etanol é de −0,28% e 0,43% respectivamente. Pelo efeito acumulado, percebe-se apenas um efeito de curto prazo nas produções. Para o choque de 1% no preço do açúcar ocasiona a maior resposta no curto prazo das produções de etanol e açúcar (2 meses após o choque) na ordem de 0,32%

[3] http://www.datagro.com.br

[4] Justifica-se o preço do mercado externo, pois 2/3 da produção do açúcar no Brasil é exportado (http://www.datagro.com.br).

Figura 1. Resposta de um choque no preço de etanol no mercado sucroalcooleiro.

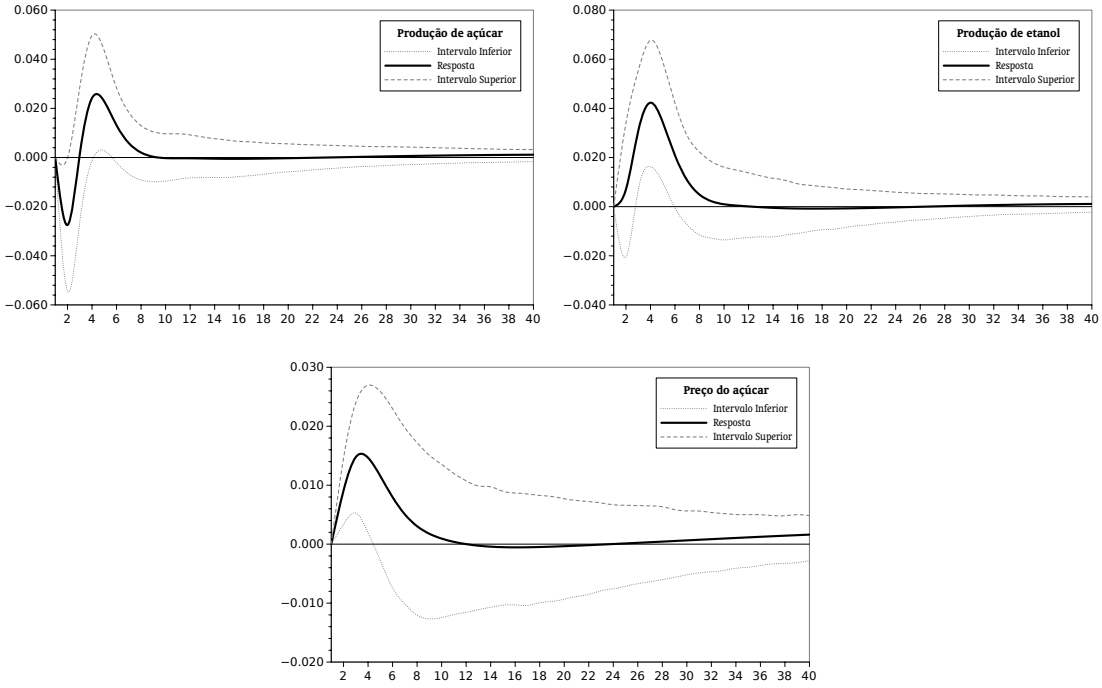


Figura 2. Resposta de um choque no preço do açúcar no mercado sucroalcooleiro.

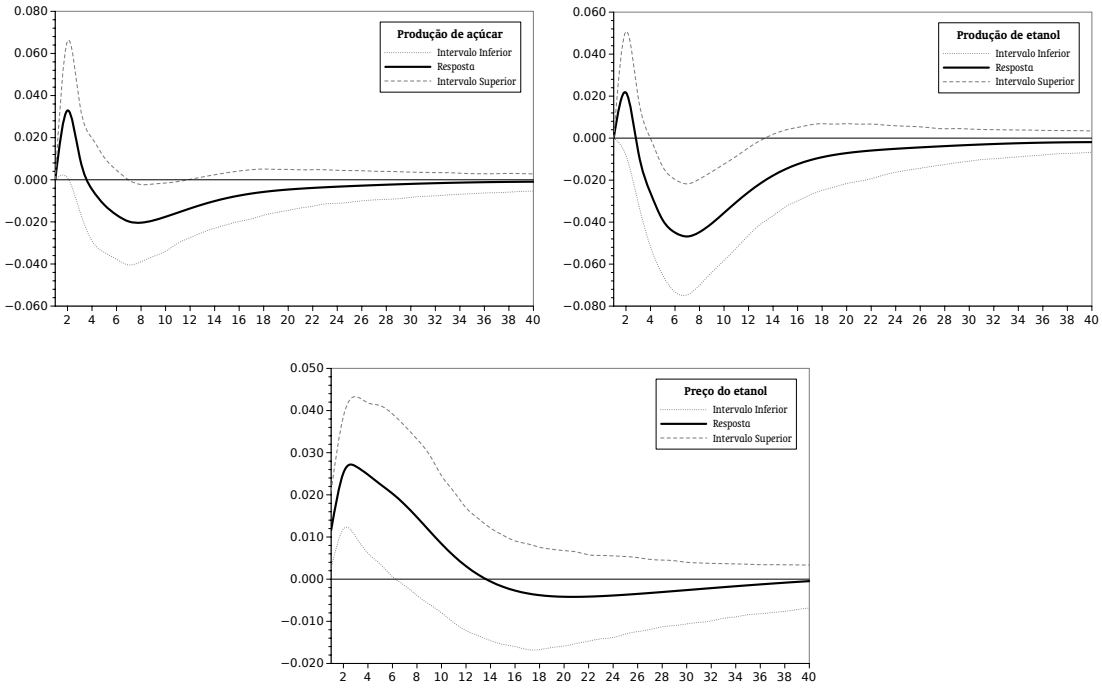


Figura 3. Resposta de um choque no preço da gasolina no mercado sucroalcooleiro.

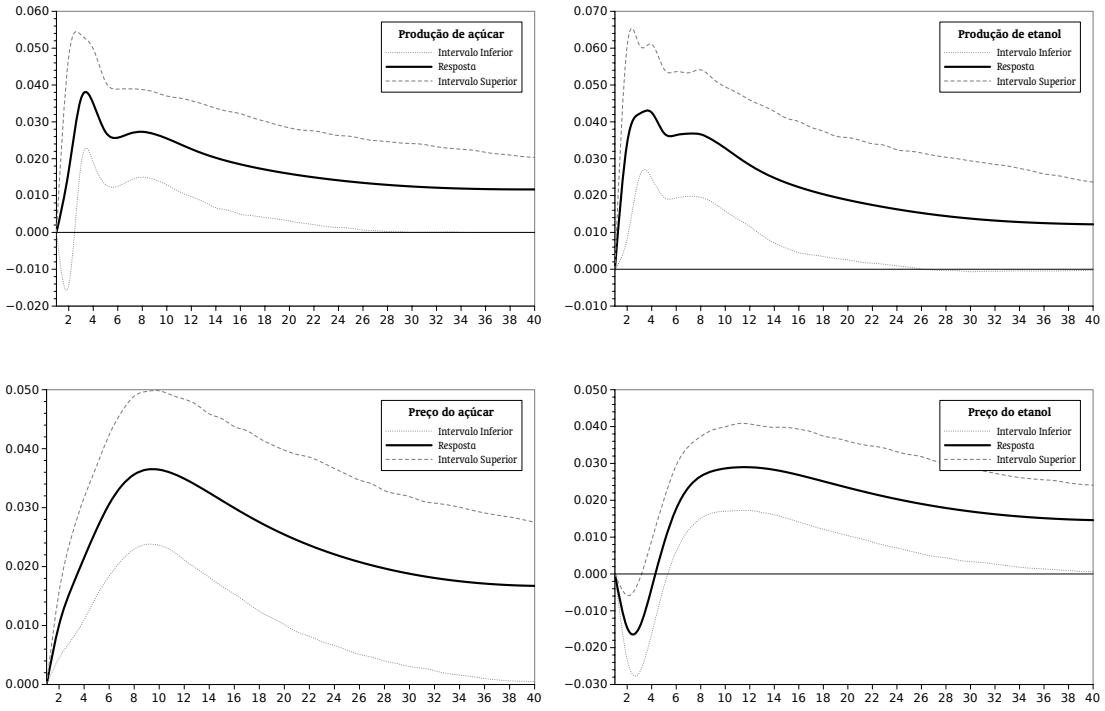


Tabela 1. Respostas Acumuladas (em valores relativos) das variáveis selecionadas a um choque de 1% nos preços do etanol, gasolina e açúcar (%).

Choque			Resposta Máxima	Resposta Acumulada Máxima	Resposta Acumulada até o 10º período	Resposta Acumulada até o 20º período	Resposta Acumulada até o 30º período	Resposta Acumulada até o 40º período
Etanol	→	Produção de Etanol	0,43	1,61	1,61	1,61	1,58	1,56
Etanol	→	Produção de Açúcar	−0,28	0,47	0,47	0,45	0,42	0,40
Açúcar	→	Produção de Etanol	0,32	−6,45	−3,89	−5,78	−6,17	−6,45
Açúcar	→	Produção de Açúcar	0,50	−2,70	−0,96	−2,12	−2,47	−2,70
Gasolina	→	Produção de Etanol	2,26	11,75	15,78	28,88	35,57	39,59
Gasolina	→	Produção de Açúcar	1,91	13,65	11,58	22,10	27,47	30,68

e 0,5%, respectivamente. No longo prazo, o efeito acumulado das respostas indica uma persistência na queda da produção de etanol e certa estabilidade da produção de açúcar. O resultado sugere que o produtor do setor sucroalcooleiro modifica estoques para beneficiar a produção do açúcar no longo prazo.

Com relação ao choque de 1% no preço da gasolina, a maior resposta relativa ocorre com a produção de etanol, quatro meses após o choque, indicando que a troca da gasolina pelo etanol faz a oferta desse produto aumentar no curto prazo, o que corrobora com os resultados de Melo & Sampaio (2014) que mostra um aumento da demanda de etanol quando ocorre um choque no preço da gasolina. O efeito acumulado nas produções de etanol e açúcar mostra um aumento persistente até o quadragésimo mês. Esse resultado demonstra que o mercado sucroalcooleiro é suscetível a mudanças no preço da gasolina.

Na Tabela 2 é mostrada a decomposição de variância. Observa-se que tanto a produção de etanol, como a produção de açúcar possui uma estrutura basicamente autoregressiva. No entanto, destaca-se que a produção de açúcar possui uma representatividade na variância da produção de etanol no curto e no longo prazo. Esse resultado sugere que a produção de etanol tem se destacado de forma residual com relação ao açúcar. A produção de açúcar, por sua vez, possui uma forte estrutura autorregressiva no curto prazo e também no longo prazo. O que se pode sugerir é a consolidação do mercado internacional do açúcar e a tradição exportadora do produto, pode influenciar na decisão do produtor em produzir açúcar em detrimento ao etanol.

Tabela 2. Decomposição da Variância das Produções de Etanol e Açúcar (%).

Variável	Mês	Produção Açúcar	Produção Etanol	Preço Açúcar	Preço Etanol	Preço Gasolina
Produção de Etanol	1	42,6350	57,3650	0,0000	0,0000	0,0000
	20	32,2260	46,0110	8,3360	2,9140	10,5130
	40	31,7870	45,4290	8,2500	2,8630	11,6720
Produção de Açúcar	1	100,0000	0,0000	0,0000	0,0000	0,0000
	20	87,5610	2,5860	2,2960	1,3560	6,2010
	40	86,6250	2,7080	2,3170	1,3410	7,0080

As variáveis dummy da introdução dos veículos flex, crise internacional e intervenção na mistura etanol/gasolina não afetaram as produções de etanol e açúcar, com 5% de significância. Dessa forma, as produções de etanol e açúcar são basicamente determinadas pelas variáveis inerentes ao mercado (preço) e ao mercado do combustível fóssil (preço da gasolina).

Para testar a robustez do modelo apresentado, optou-se por trocar a ordem de entrada das variáveis. Colocando que os preços afetam de forma contemporânea as produções de açúcar e etanol respectivamente os efeitos do impulso resposta e decomposição de variância foram modificados. Segundo Brooks (2008), essa modificação traz algumas implicações. Primeiro, quando a ordem de entrada modifica os resultados é indício que os resíduos da equação são fracamente correlacionados. Segundo, os resultados mostram que as produções de açúcar e etanol apresentaram uma memória praticamente autorregressiva, o que indica que a variação da produção depende praticamente dela mesma e influencia os preços. Dessa forma, optou-se por colocar as produções como variáveis que afetam contemporaneamente as outras, mas não é afetada por nenhuma.

5. CONCLUSÕES

Este trabalho mostrou que a produção de açúcar e do etanol responde a choques de preços de maneira significativa. Porém, as produções respondem em uma menor medida a mudanças no preço do

etanol, sendo considerado apenas o impacto no curto prazo. Ou seja, de acordo com a literatura, a crescente demanda de etanol com a introdução dos veículos flex, que refletiu no preço do biocombustível, influenciou as ofertas de etanol e açúcar apenas no curto prazo, ressaltando assim, que o preço de etanol não gera impactos prolongados nas produções de açúcar e de etanol.

Um choque no preço do açúcar gera impacto, em uma maior medida, nas produções do setor sucroalcooleiro. Esse fenômeno é visto pela consolidação do açúcar no mercado externo. No entanto, grande parte da variação da oferta de açúcar é inerente a fatores ligados a produção de açúcar, como estoques. O modelo sugere que produtor do setor sucroalcooleiro reage mais fortemente a uma mudança no preço do açúcar com relação a uma mudança no preço do etanol. Ou seja, ele revela a preferência em produzir açúcar para o mercado externo a produzir etanol para o mercado interno.

O preço da gasolina também é um fator que impacta na dinâmica do setor sucroalcooleiro. O resultado mostra que o produtor do setor sucroalcooleiro tem um benefício com o aumento da produção e o aumento do preço. O preço da gasolina, por sua vez, é controlado pelo governo, embora tenha sido nominalmente desregulamentado após 2002. Essa, porém, é uma reivindicação dos produtores do setor, que o preço da gasolina atinja a paridade com o preço internacional para que o setor obtenha ganhos com o aumento de preço e produção. Porém, a liberalização total do preço do combustível fóssil pode gerar instabilidade no mercado interno.

Este trabalho possui algumas limitações, como por exemplo, não explicitar no modelo a variação de estoques que é mais estável e apresenta menor grau de sazonalidade. No entanto essa variável não é encontrada em dados mensais. Outro fator de limitação é que não foi possível tornar endógena a variável veículos flex, pois a frota nacional está disponível apenas em dados anuais e também não foi possível contabilizar a dinâmica do mercado quando ocorre um choque estrutural e de políticas.

REFERÊNCIAS BIBLIOGRÁFICAS

Bacchi, M. R. P. (2005). Formação de preços no setor sucroalcooleiro da Região Centro-Sul do Brasil: Relação com o mercado de combustível fóssil. In *XXXIII Encontro Nacional da ANPEC, 6 a 9 de dezembro de 2005*, Natal, RN. Disponível em: http://www.anpec.org.br/encontro2005/artigos/A05A143.pdf

Balcombe, K., & Rapsomanikis, G. (2008). Bayesian estimation and selection of nonlinear vector error correction models: The case of the sugar-ethanol-oil nexus in Brazil. *American Journal of Agricultural Economics*, *90*(3), 658–668. doi: 10.1111/j.1467-8276.2008.01136.x

Barros, G. (2011, 2 de julho) Políticas focadas no curto prazo podem prejudicar os investimentos no etanol. *Folha de São Paulo*.

FAO. (2008). *Soaring food prices: Facts, perspectives, impacts and actions required* (High-Level Conference on World Food Security: The Challenges of Climate Change and Bioenergy – HLC/08/INF/1). Rome, 3–5 June 2008: Food and Agriculture Organization (FAO). Disponível em: http://www.fao.org/fileadmin/user_upload/foodclimate/HLCdocs/HLC08-inf-1-E.pdf

Boff, H. P. (2011). Modeling the Brazilian ethanol market: How flex-fuel vehicles are shaping the long run equilibrium. *China-USA Business Review*, *10*(4), 245–264.

Brooks, C. (2008). *Introductory econometrics for finance* (2ª ed.). Cambridge UK: Cambridge University Press.

Cabrini, M. F., & Marjotta-Maistro, M. C. (2007). Mercado internacional de álcool: Os recentes programas de uso do produto como combustível. *Agroanalysis*, *7*, 36–36.

Costa, C. C. d., Burnquist, H. L. & Guilhoto, J. J. M. (2006). Impacto de alterações nas exportações de açúcar e álcool nas regiões Centro-Sul e Norte-Nordeste sobre a economia do Brasil. *Revista de Economia e Sociologia Rural*, *44*(4), 609–627. doi: 10.1590/S0103-20032006000400001

Giesecke, J. A., Horridge, J. M. & Scaramucci, J. A. (2007, dez). *The downside of domestic substitution of oil with biofuels: Will Brazil catch the Dutch disease?* (Working Paper Nº G-169). Clayton Vic, Australia: Victoria University, Centre of Policy Studies/IMPACT Centre. Disponível em: http://www.copsmodels.com/ftp/workpapr/g-169.pdf

Kilian, L. (2010). Explaining fluctuations in gasoline prices: A joint model of the global crude oil market and the U.S. retail gasoline market. *The Energy Journal*, *31*(2), 103–128. doi: 10.5547/ISSN0195-6574-EJ-Vol31-No2-4

Melo, A. d. S., & Sampaio, Y. d. S. B. (2014). Impacts of gasoline and ethanol prices on Brazilian ethanol demand. *Revista de Economia Contemporânea*, *18*(1), 56–83. doi: 10.1590/141598481813

Moraes, M. A. F. D. d. (2000). *Desregulamentação do setor sucroalcooleiro do Brasil*. São Paulo: Caminho Editorial.

Müller, A., Schmidhuber, J., Hoogeveen, J. & Steduto, P. (2007, Jan 28–31). Some insights in the effect of growing bio-energy demand on global food security and natural resources. In *Linkages between energy and water management for agriculture in developing countries*, Hyderabad, India. Disponível em: http://www.globalbioenergy.org/uploads/media/0701_FAO_Mueller__-_Some_insights_in_the_effect_of_growing_bioenergy_demand_on_global_food_security_and_natural_resources_01.pdf

Schmidhuber, J. (2007, May). *Biofuels: An emerging threat to Europe's food security? Impact of an increased biomass use on agricultural markets, prices and food security: A longer-term perspective* (Policy Paper N^{0} 27). Jacques Delors Institute–Notre Europe. Disponível em: http://www.delorsinstitute.eu/011-1317-Biofuels-An-Emerging-Threat-to-Europe-s-Food-Security.html

Serra, T., Zilberman, D. & Gil, J. (2011). Price volatility in ethanol markets. *European Review of Agricultural Economics*, *38*(2), 259–280. doi: 10.1093/erae/jbq046

Silva, C. M. S., & Almeida, E. L. F. (2006, ago). Formação de um mercado internacional de etanol e suas inter-relações com os mercados de petróleo e açúcar. In *CBE, 16–18 de agosto de 2006*, Rio de Janeiro.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, *48*(1), 1–48. doi: 10.2307/1912017

Smeets, E., Junginger, M., Faaij, A., Walter, A., Dolzan, P. & Turkenburg, W. (2008). The sustainability of Brazilian ethanol: An assessment of the possibilities of certified production. *Biomass and Bioenergy*, *32*(8), 781–813. doi: 10.1016/j.biombioe.2008.01.005

Tokgoz, S., & Elobeid, A. E. (2006, November). *An analysis of the link between ethanol, energy, and crop markets* (CARD Working Paper N^{0} 444). Ames, Iowa: Center for Agricultural and Rural Development / Iowa State University. Disponível em: http://lib.dr.iastate.edu/card_workingpapers/444/

Viegas, T. (2011, junho 20). *A solução para a crise do etanol: Incentivos, subsídios, regulação ou defesa da concorrência?* [webpage]. Blog Infopetro. Acessado em setembro de 2011: https://infopetro.wordpress.com/2011/06/20/a-solucao-para-a-crise-do-etanol-incentivos-subsidios-regulacao-ou-defesa-da-concorrencia/

von Lampe, M. (2006, Feb). *Agricultural market impacts of future growth in the production of biofuels* (Working Paper on Agricultural Policies and Markets). Organisation for Economic Co-operation and Development (OCDE).

r b e

ISSN 0034-7140
DOI 10.5935/0034-7140.20160005
Revista Brasileira de Economia, **70**(1), 71–98

Regional Labor Market Differences in Brazil and Search Frictions: Some Structural Estimates*

Paulo Felipe de Oliveira,† José Raimundo Carvalho‡

Contents: **1. Introduction; 2. Regional Differences in Brazilian Labor Markets; 3. Empirical Implementation of Job Search Models; 4. Empirical Analysis; 5. Results; 6. Concluding Remarks.**

Keywords: **Search Theory, Regional Differences, Structural Econometrics.**

JEL Code: **J64, C14, C41.**

We estimate an equilibrium job search model for six metropolitan areas located in different regions of Brazil. Two mechanisms of wage determination are considered: wage posting by monopsony firms and Nash bilateral bargaining. In order to estimate the model, we use the non-parametric method developed by Bontemps, Robin, & van den Berg [2000. Equilibrium search with continuous productivity dispersion: Theory and nonparametric estimation. *International Economic Review*, 41(2), 305–358]. There is significant heterogeneity among the estimated structural parameters for these regions. We succeed in rationalizing some well-known regional differences in wages, unemployment rates and productivity prevalent in Brazilian labor markets, thus offering new interpretations. Metropolitan regions in the Northeast have much lower λ_0 (arrival rate of wage offers for unemployed workers) and λ_1 (arrival rate of wage offers for employed workers) *vis a vis* areas in the South or Southeast regions. This is a new, and much more precise, result worth considering on the regional inequality debate in Brazil. Regional inequality in wages, besides being an outcome of its regional human capital distribution, can be rationalized as inequality of search frictions brought by differences in λ_1. We found a key (indirect) role for search frictions when analysing productivity differences as well. Since search frictions impact simultaneously on monopsony power as well as on productivity, in order to understand better regional productivity differences we must deepen our analysis on how these structural parameters are differentiated by regions. Labor market frictions add important insights into the regional debate, something not captured by more traditional econometric reduced form approaches.

*This paper has been circulating under the title "Estimation of an Equilibrium Search Model with Productivity Dispersion: a Regional Analysis for Brazil". We thank participants from the 33rd Meeting of the Brazilian Econometric Society, 2011, and the 6th CAEN-EPGE Meeting of Public Policy and Economic Growth, 2013, for valuable comments and suggestions. The usual disclaimer applies.

†Ministério do Desenvolvimento, Indústria e Comércio Exterior. I would like to thank CNPq, Brazil.

‡CAEN/UFC and LECO/CAEN/UFC. E-mail: josecarv@ufc.br

Estimamos um modelo de procura de emprego de equilíbrio para seis regiões metropolitanas localizadas em diferentes regiões do Brasil. Dois mecanismos de determinação dos salários são considerados: postagem de salários por empresas monopsônicas e negociação bilateral a la Nash. Para estimar o modelo, usamos o método não-paramétrico desenvolvido por Bontemps, Robin, & van den Berg [2000. Equilibrium search with continuous productivity dispersion: Theory and nonparametric estimation. International Economic Review, *41(2), 305–358]. Existe heterogeneidade significativa entre os parâmetros estruturais estimados para essas regiões. Conseguimos racionalizar diferenças regionais já bem conhecidas nos salários, taxas de desemprego e produtividade prevalentes nos mercados de trabalho brasileiros, oferecendo assim novas interpretações. Regiões metropolitanas do Nordeste possuem λ_0 (taxa de oferta de salários para trabalhadores empregados) e λ_1 (taxa de oferta de salários para trabalhadores empregados) muito menores* vis a vis *aquelas encontradas nas regiões Sul ou Sudeste. Este é um resultado novo e muito mais preciso que vale a pena ser considerado no debate regional da desigualdade no Brasil. A desigualdade regional nos salários, além de ser resultado da distribuição regional de capital humano, pode ser racionalizada através da desigualdade de fricções trazidas por diferenças entre λ_1's. Evidenciamos também um papel (indireto) importante das fricções de busca quando se analisam as diferenças de produtividade. Como os atritos de busca impactam simultaneamente sobre o poder de monopsônio, bem como sobre a produtividade, a fim de entender melhor as diferenças de produtividade regionais, devemos aprofundar nossa análise sobre como esses parâmetros estruturais são diferenciadas por regiões. Fricções no mercado de trabalho adicionam importantes* insights *ao debate, algo que não é capturado por abordagens econométricas de forma "reduzidas" mais tradicionais.*

1. INTRODUCTION

Job search theory considers the labor market an environment of imperfect information. This implies that transactions in this market need time and other resources to be held, since agents are not fully informed about the opportunities and characteristics which are relevant to the transaction (see Eckstein & van den Berg, 2007 and Rogerson, Shimer, & Wright, 2005). Hence, these models serve as a substitute for the analysis of labor markets outside the neoclassical paradigm of labor supply, which precludes the possibility of involuntary unemployment.

In an environment of imperfect information, firms can exploit the fact that workers are not fully informed about all opportunities and thus offer wages which are lower than the value of labor productivity, even in a situation where firms and workers are homogeneous (Burdett & Mortensen, 1998). Thus, the prevailing wage may be smaller than in an environment of perfect competition, where wages would be equal to the marginal labor productivity. Moreover, from a structural estimation perspective, it is possible to analyze jointly, in a context of market equilibrium, labor market issues such as arrival rates of job offers for employed and unemployed workers, the separation rate of employment relations, firm's productivity, and monopsony power. Hence, one can make an attempt to infer the factors that would account for differences in unemployment rates between regions or would influence the level of frictions present in the labor market on wages, for example.

In this paper, parameters of an equilibrium job search model are estimated, where firms may differ in terms of labor productivity, based on six metropolitan areas located in different regions of Brazil. These metropolitan regions serve as a benchmark to analyse regional labor market differences. Although the methodology employed is not new, we believe our main contribution is to show the feasibility of analysing regional labor market differences by means of a structural model. In fact, our results are in

accordance with stylized facts about regional inequality of Brazilian markets (see Reis & Barros, 1990, Savedoff, 1990, Savedoff, 1995, Azzoni & Servo, 2002, Queiroz & Golgher, 2008, and Freguglia & Menezes-Filho, 2012). It also present new insights and hints on why such inequalities might exist.

The productivity debate in Brazil is also object of our analysis. In order to do that we confront two possible wage determination mechanisms: wage posting *ex ante* by monopsonist firms that set wages unilaterally, based in Bontemps, Robin, & van den Berg (2000), and Nash bilateral bargaining *ex post*, based in Mortensen (2003). The methodology used to estimate the model is the nonparametric method of Bontemps et al. (2000), which enables us to estimate the model without assuming any parametric probability distribution of firms' productivity. Our results show the superiority of the Nash bilateral bargaining model, both in terms of fit as well as in terms of replicating stylised facts about regional productivity differences.

Although empirical analysis based on structural estimation of search models are already quite developed internationally,[1] in Brazil, the literature is virtually nonexistent, and Carvalho (2012) and Oliveira (2011) are rare examples of structural analysis of a search model using Brazilian data (Survey of Living Standards–PPV). However, our paper not only develops and estimates a search model but also suggests that this methodology can shed important insights into the debate over regional labor market differences.

The empirical analysis in this study confirms and shows significant differences between regions. Some regions have a higher expected unemployment duration, as the metropolitan areas of Salvador and Rio de Janeiro, where the expected completed durations of unemployment are equal to approximately eleven months, while for the metropolitan region of Belo Horizonte this term is near four months. In the metropolitan area of Recife, employed workers have a low transition rate to jobs that pay higher wages, while regions such as São Paulo and Porto Alegre have a relatively high mobility. Despite these results, our methodology shows its strength when we are able to rationalize some well-known regional differences in Brazilian labor markets by means of deep structural parameters. Clearly, the role of labor market frictions, internal labor market dynamics and market organization (monopsony level) add important insights into the debate.

In terms of productivities, depending on the type assumed in the wage determination, we arrive at different conclusions. The model with wage posting inferred high-productivity levels, a result also found by Mortensen (2003) and Shimer (2006), and the bilateral bargaining has more plausible and theoretically admissible values that do not violate any theoretical restriction. In the case of bargaining, the metropolitan region of São Paulo has the highest level of productivity and Recife has the lowest level. When compared to Recife, the distribution of firms' productivity in São Paulo is also more dispersed. We believe that our estimates have the potential to contribute to the long-standing debate about regional differences in Brazilian labor markets (see, among others, Reis & Barros, 1990, Savedoff, 1990, Savedoff, 1995, Azzoni & Servo, 2002, Queiroz & Golgher, 2008, and Freguglia & Menezes-Filho, 2012).

Besides this introductory section, the paper is organized as follows: Section 2 presents a brief description of the recent debate about regional differences in Brazilian labor markets, focusing on three main labor market outcomes: wages, unemployment rate and productivity. Section 3 reviews the theoretical and empirical literature on structural estimation of search models in the labor market. Section 4 addresses the empirical part of the paper and it is split into three subsections: Theoretical Background, Data Set and Structural Estimation. Section 5 presents and discusses the main structural estimates and, finally, section 6 draws some final considerations.

2. REGIONAL DIFFERENCES IN BRAZILIAN LABOR MARKETS

According to Azzoni & Servo (2002), among others, Brazil has a high level of regional inequality and the prospects of any improvement from the present to the medium run are at best disappointing. Although

[1] See Eckstein & van den Berg (2007) for a literature review.

this regional inequality can be discussed from many perspectives, income inequality (more specifically labor income) is the subject of most scrutiny in the literature. Azzoni & Servo assert that

> The Northeast region of Brazil is home to 28% of the Brazilian population in 2000; it produced only 13% of the Brazilian GDP in 1998. The rich Southeast region represented 43% of the total population and produced 58% of GDP. Per capita income in the Northeast was 54% below the national average while in the Southeast it was 36% above that level. The poorest state, Piaui, in the Northeast region, had a per capita income level 5.6 times lower than the richest state, Sao Paulo, in the Southeast region.

Hence, it is a well-established fact that Brazilian labor markets are very heterogenous.

A less developed argument is that, besides earnings differences, regional inequality in Brazilian labor markets can be found, at least indirectly, when one looks at other indexes such as unemployment rates, unemployment duration, productivity differences, among other economic variables. In fact, labor market differences in Brazil is by no means a recent trend. As a matter of fact, much of our regional disparities can be traced back to colonial times, where the North-South (more precisely, Northeast-Southeast) disparities started up, enlarging during the last decades of development (Araújo & Lima, 2010). In that sense, we think that a brief literature review on regional differences in Brazilian labor markets that considers those aspects of heterogeneity as well as the recent path of regional inequality is important for a better understanding of our arguments.

In order to start understanding regional differences in Brazilian labor markets during the first decade of the 21st century, we need to understand the macroeconomic environment surrounding that period. The Brazilian economy started to recover from a period of stagnation during the 1980s and mid-1990s, initiating a period of modest growth from 1995 on. But it was only during the period from 2004 up to 2008 that we observe a real acceleration in its growth rate. Basically, this acceleration took place simultaneously with

> [...] the influence from the growth conditions of the world economy and domestic demand driven by real increases in the minimum wage, the expansion of personal credit and higher volume of federal cash transfers resources, and the multiplier effects of export activity benefited from major expansion of foreign demand for industrial and agricultural commodities. (Araújo & Lima, 2010)

The rest of this section discusses three aspects of regional labor market differences in Brazi: wages, unemployment rate and productivity.

2.1. Wage Differences

Although the debate about income inequality is well established among Brazilian scholars (see, among others, Barros & Mendonça, 1995, Langoni, 2005, Barros, Foguel, & Ulyssea, 2006), the debate about income inequality with an emphasis on regional differences on earnings has received much less attention. Few papers like Reis & Barros (1990), Savedoff (1990), Savedoff (1995), Azzoni & Servo (2002), Queiroz & Golgher (2008), and Freguglia & Menezes-Filho (2012) treat in an explicitly manner the stark earnings differences among Brazilian regions, pointing out the important role that human capital unequal distribution has on this inequality as well as other observed and unobserved individual heterogeneities.[2] Tabela 1 shows the evolution of the average monthly earnings between regions in Brazil.

Tabela 1 and Figura 1 show clearly that the Southeast region was the leader in labor productivity until 2007 (only surpassed by the Center-West mainly due to agricultural productiviy) and the Northeast region remains at the last place. Also, the gap between Southeast and Northeast remains almost unchanged.

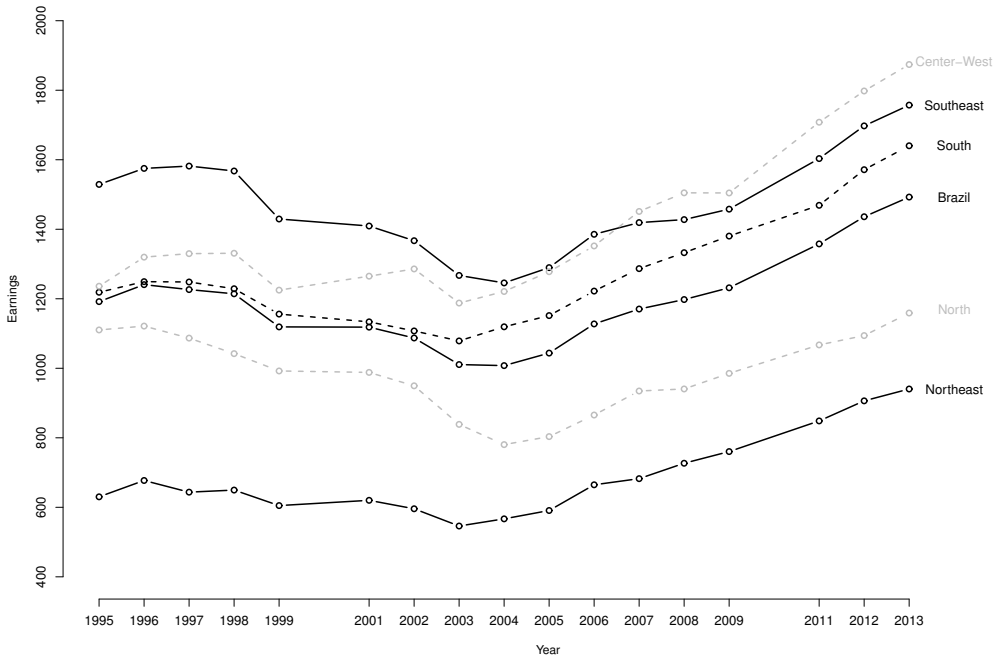
[2]However, some evidence point out to the fact that these differences are considerably reduced once we control for unobserved heterogeneity at the individual level, accordingly to Freguglia & Menezes-Filho (2012).

Table 1. Average Monthly Earnings at main job of occupied population 16 years or older.

Region	2004	2005	2006	2007	2008	2009	2011
Brazil	1,007.68	1,043.71	1,127.58	1,170.46	1,197.75	1,231.56	1,357.84
North	780.41	803.42	865.49	934.61	940.49	985.38	1,067.22
Northeast	566.77	590.78	664.76	682.43	726.74	760.33	848.52
Southeast	1,245.6	1,289.6	1,385.48	1,419.14	1,427.75	1,457.84	1,603.35
South	1,119.33	1,151.6	1,222.12	1,286.72	1,332.61	1,380.37	1,468.88
Center-West	1,220.96	1,277.92	1,352.02	1,451.23	1,504.83	1,504.43	1,707.97

Note: R$ of 09/2013 – INPC.
Source: IBGE/PNAD.

Figure 1. Mean Monthly Earnings at main job of occupied population 16 years or older.



Many Brazilian authors suggest inequality of human capital distribution as the main explanation for that situation (see, among others, Reis & Barros, 1990, Barros & Mendonça, 1995, Duarte, Ferreira, & Salvato, 2004, and Barros et al., 2006). Tabela 2 presents the mean number of years of education by region. Overall, these figures support the human capital explanation for wage income inequality among regions. However, as many others have argued convincingly other explanations like unionization (Arbache, 1999), quality of education (Behrman & Birdsall, 1983), bargaining and mark-up power (Amadeo, 1994), and individual unobserved heterogeneity (Freguglia & Menezes-Filho, 2012) are important to understand wages differences.

2.2. Unemployment Rate Differences

Although the question of regional labor income inequality was the interest of Brazilian scholars, the understanding of regional unemployment differences has received much less attention (see, Corseuil, Gonzaga, & Issler, 1997, Oliveira & Cruz, 2000, and Oliveira & Carneiro, 2001). The seminal paper by

Table 2. *Mean Years of Education 16 years or older.*

	2004	2005	2006	2007	2008	2009	2011
Brazil	6.89	7.01	7.19	7.31	7.47	7.60	8.73
North	6.28	6.47	6.65	6.82	6.97	7.11	7.13
Northeast	5.52	5.67	5.86	6.03	6.24	6.38	6.53
Southeast	7.60	7.71	7.89	8.01	8.11	8.23	8.39
South	7.39	7.50	7.64	7.68	7.90	8.00	8.04
Center-West	7.13	7.29	7.46	7.55	7.76	7.90	8.17

Source: PNAD/IBGE.

Corseuil et al. (1997) presents an analysis of local and national unemployment rates. The authors attempt to measure how aggregate shocks in the Brazilian economy impact on regional labor markets. Corseuil et al. (1997) depart from three basic premises: the sensitivity of regional unemployment with respect to shocks at the national employment; decomposition of regional unemployment into aggregate, regional and industry factors; and fluctuations in regional unemployment.

They draw to main conclusions: i) aggregate shocks are more important than structural regional shocks to explain unemployment dynamics; and ii) it is not easy to define in a clear and unambiguous way whose region has greater natural unemployment rate. So, although Corseuil et al. (1997) are a benchmark in the national literature about regional differences in unemployment dynamics, it falls short over achieving a clear regional explanation for unemployment dynamics. In light of this fact, we resort to more simple empirical evidence in order to have a glimpse on some sort of regional unemployment differences. Tabela 3 and Figura 2 give us some hints.

A first clear trend is the constant decline on unemployment rate over all regions under analysis. A second, and in a sense, dissapointing pattern, is the fact that for Bahia, Pernambuco (both at the Northeast region) and São Paulo (Southeast region), unemployment rates stay higher than the national level all over the years from 2004 up to 2010. So, this signals that any sort of regional disparity in unemployment rate would be a little hard to back up. Notwithstanding that, we could tentatively try to support a view that Southwestern states like Rio de Janeiro and Minas Gerais (with the exception of São Paulo) have lower unemployment rates than those states in the Northeast.

2.3. Productivity Differences

The 1990s brought an unprecedent period of structural change in Brazil's economy. The process of openness of Brazilian economy prompted many changes in the industrial sector, bringing the necessary productive modernization and technological gains in order to be competitive at a global level (see, Feijó & Carvalho, 2002, Galeano & Wanderley, 2013a, 2013b). In terms of labor productivity, at first we saw considerable gains overall. However, these initial gains faded away by the beginning of the 2000s. According to Galeano & Wanderley (2013a), the decrease in labor productivity that started around 1998–2000 was due mainly to the considerable growth on the number of occupied workers.

The first decade of the 21st century witnesses, from 2002 on, first a fall in labor productivity driven by a decrease on the value of industrial transformation coupled with a increment on the number of occupied workers. From 2002 up to 2005, the growth rate of labor productivity was zero, although we saw a little recovery by the end of the decade. By 2007–2008, labor productivity reached the same value as of 1996. Such evidence of a falling labor productivity prompted some authors to question the very process of openness of Brazilian economy as a beneficial one (see, among others, Galeano & Wanderley, 2013a, 2013b, Schettini & Azzoni, 2013, and Cavalcante & De Negri, 2014).

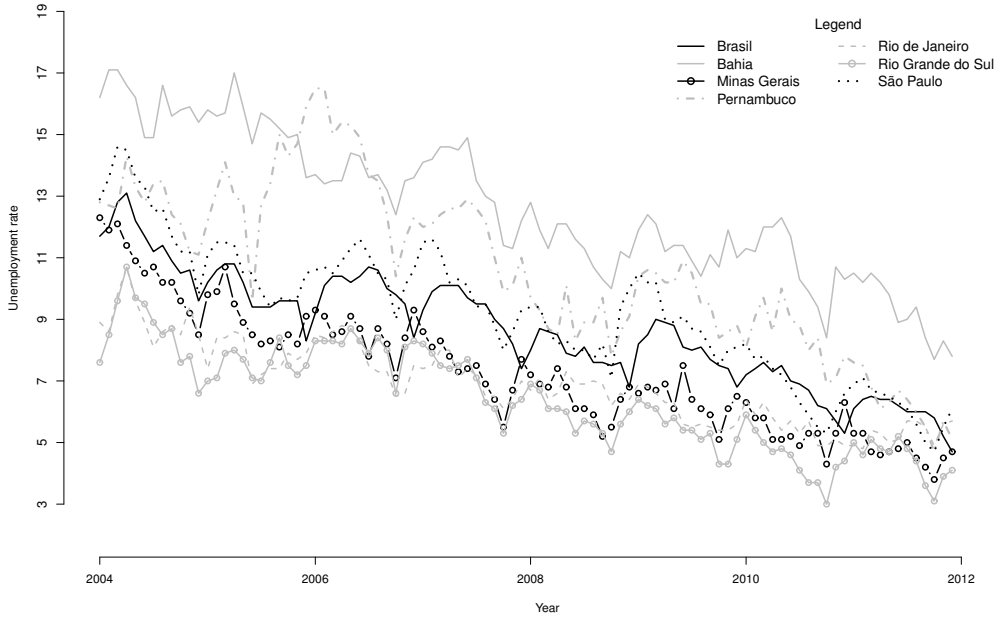
Despite the controversial impact of trade openness in labor productivity growth, there is some consensus among Brazilian scholars (Galeano & Wanderley, 2013a, 2013b; Schettini & Azzoni, 2013) over the

Table 3. Unemployment Rate.

Metropolitan Area	2004	2005	2006	2007	2008	2009	2010	2011
Salvador – Bahia	16.03	15.38	13.57	13.43	11.36	11.37	10.79	9.23
Belo Horizonte – Minas Gerais	10.63	8.98	8.57	7.35	6.43	6.39	5.39	4.68
Recife – Pernambuco	12.68	13.41	13.96	11.78	8.97	9.84	8.46	6.20
Rio de Janeiro – Rio de Janeiro	9.03	7.81	7.83	7.09	6.75	5.96	5.53	5.30
Porto Alegre – Rio Grande do Sul	8.64	7.50	8.12	6.97	5.83	5.42	4.46	4.44
Sao Paulo – Sao Paulo	12.63	10.44	10.60	9.88	8.53	8.91	6.73	6.09
Brazil	11.48	9.83	9.98	9.29	7.89	8.08	6.74	5.98

Source: IpeaData.

Figure 2. Unemployment Rate.



regional impacts, say, spatial reallocation of industrial activity among regions of Brazil. The main effect was a deconcentration of industrial activity from the South and Southeast regions towards North and, mainly, Northeast regions. Actually, the growth rate of labor productivity from the Southeast region was lower than those of all other regions, contributing to slow down the process of regional inequality.

Notwithstanding that increase in labor productivity of all regions *vis a vis* the more developed Southeast region, with the exception of the North region (whose industry is driven mainly by a massive program of fiscal incentives and "Zona Franca de Manaus" in the state of Amazonas) there is an important gap on labor productivity (see Tabela 4). Of course, a gap in labor productivity is well known for having deep impacts on growth and development and for posing an obstacle to regional development, since at least Kaldor (1970).

Hence, the question of regional labor inequalities is both pervasive and last longing in Brazil, with deep historical roots but driven by internal and external macroeconomic shocks, as any modern and complex economy. It presents itself by inequalities in wages, unemployment rate and labor productivity, although a key aspect of our paper is to show that such picture can be much more nuanced as long

Table 4. Industrial Labor Productivity in 2010.

	Extractive Industry	Transformation Industry
Brazil	145.64	26.97
North	448.44	37.21
Northeast	57.78	19.69
Southeast	156.84	30.81
South	18.15	21.87
Center-West	86.38	24.59

Notes: 1,000 R$ of 1996.
Source: Galeano & Wanderley (2013b).

as our econometric estimation exercises shed light on the deep relationships brought by the modern search theory. Next section briefly digresses over the empirical implementation of equilibrium search models.

3. EMPIRICAL IMPLEMENTATION OF JOB SEARCH MODELS

We start by summarizing the two main different approaches that lead to wage dispersion in job search models as a response to Diamond (1971)'s degeneracy of the wage offer distribution. The first approach is represented by the model by Albrecht & Axell (1984). The wage dispersion occurs at Albrecht & Axell (1984) due to the fact that workers are heterogeneous with respect to the value attributed to leisure (or the opportunity cost of labor). Thus, workers differ in the value of reservation wage. By offering a higher wage, the firm increases the level of employment in steady state, but reduces the profit per worker. It is, then, possible that firms which offer different wages have equal profits, enabling the existence of an equilibrium with wage dispersion. Therefore, in this model, the support of the wage offer distribution is equal to a subset of the reservation wages of workers.

The second approach is based mainly on Burdett & Mortensen (1998). The main result of Burdett & Mortensen is an equilibrium with wage dispersion even if workers and firms are identical. The dispersion is produced by the possibility of on-the-job search. Firms can offer wages which are higher than the reservation wage because it would also attract more workers from other employers. That is, the labor supply curve for the firm is upward sloping. Thus, using the isoprofit assumption for firms that offer different wages, it is possible to get an equilibrium where the wage distribution is not degenerate on the common reservation wage, avoiding the Diamond (1971) paradox. Finally, the model by Burdett & Mortensen (1998) is characterized as a model of wage posting. In other words, firms unilaterally decide the wage offer and bargaining between workers and employers will not occur.

As to the empirical evidence, the work by Eckstein & Wolpin (1990) is based on Albrecht & Axell (1984) model and the study by van den Berg & Ridder (1998) is one of the main empirical analysises made from Burdett & Mortensen (1998) model. In Eckstein & Wolpin (1990), workers are homogeneous in terms of productivity and heterogeneous in leisure value and firms differ in efficiency in terms of labor productivity. The authors estimate the unrestricted and restricted model. The restricted model incorporates the impositions generated by the equilibrium job search model. Eckstein & Wolpin (1990) observe that the restricted model fails to explain wage dispersion, which was the main objective of the model. The wage dispersion is almost entirely attributed to measurement error in wages. However, the paper has great appeal because it is the first to develop a methodology to estimate an equilibrium search model of the labor market. Another important feature of the model is that the arrival rate of wage offers is endogenous, determined by the number of active firms. The approach based on the model developed by Burdett & Mortensen (1998) assumes that this rate is exogenous.

The study by van den Berg & Ridder (1998) follows Burdett & Mortensen (1998) model. The authors consider that the labor market is segmented in relation to some characteristics of workers and, in turn, each segment can be considered a labor market. Thus, it is considered that in each segment workers and firms are homogeneous, but there is heterogeneity between markets. This heterogeneity is important to adjust the model to data. The analysis is performed from a panel data of employed and unemployed workers in the Netherlands. From the estimates of the theoretical model's parameters, the authors estimate the degree of monopsony in firms, effects of minimum wages variations and the factors that determine the wage dispersion. The degree of monopsony is estimated between 10% and 17%. Employed workers have a high arrival rate of wage offers that influences the competition by firms, reducing their power. Moreover, the results indicate that only 22% of the observed wage dispersion is explained by the friction in the labor market.

Bontemps et al. (2000) extend the Burdett & Mortensen (1998) model by assuming that firms are heterogeneous in terms of labor productivity and introduce a continuous distribution of firms' productivities. These authors argue that in reality firms use different technologies and that the assumption of identical firms generates a theoretical distribution of wages with an increasing density, which would be as opposed to data which shows that the highest wages occur at a low frequency. Moreover, the authors do not allow the workers to be heterogeneous due to the fact that, from the evidence of Eckstein & Wolpin (1990), this type of heterogeneity explains a small fraction of the variation in wages. Then, the wage dispersion is caused by search friction in the labor market and the difference in the firms' productivity.

The model is estimated by a nonparametric method consisting of three stages and has the advantage of the frictional parameters being estimated consistently independent from the strategy that firms adopt. The model is estimated for some sectors of the French economy (transportation, food, equipment, etc.) using microdata from workers and confronting the results with data from employers. The results indicate that firms exploit friction in the labor market, implying a high monopsony power of firms (20% to 100%).

Sulis (2008) estimates the model developed in Bontemps et al. (2000) using data from workers in Italy. The results indicate that the arrival rate of wage offers are higher for unemployed workers. Monopsony power is estimated to be considerably high, ranging between 50% to 100%. Kyyrä (2007) estimates various specifications of Burdett and Mortensen's model from Finland's microdata. The author estimates the pure homogeneity model, also considering the possibility of measurement error in wages, and the model with heterogeneity in firms' productivity. In the latter case, the author considers the case of discrete and continuous dispersion in type of firms. The results indicate that the model without heterogeneity fits the data for wages only after the introduction of measurement error in wages.

Despite the advance of equilibrium models with wage posting, Mortensen (2003) finds evidence that rejects the model by Bontemps et al. (2000) to Denmark. The productivity levels are overestimated, generating a high level of monopsony for firms. Moreover, the theoretical restriction that the relationship between wages and productivity should be increasing is not satisfied, in other words, that more productive firms offer higher wages. Thus, Mortensen (2003) investigates the possibility that wages are formed through a process of bilateral bargaining. The author suggests this possibility because Denmark has a long history of collective bargaining. The results for this case are admissible for all wages and the estimated levels of productivity are plausible.

For Brazil, analysises of job search models are almost nonexistent. In terms of structural estimation, we can cite the work by Carvalho (2012). The analysis is performed using the microdata from the Survey on Living Standards (PPV), held in 1996 and 1997. Carvalho estimated a search model based on the study of van den Berg & Ridder (1998), but using retrospective duration data. The author uses the results to analyze differences between labor markets in the Northeast and Southeast regions. There is a low arrival rate of wage offers for employed workers in both regions. In terms of productivity, a lower level of productivity for the Northeast region is estimated.

4. EMPIRICAL ANALYSIS

We use empirical data from six metropolitan areas that are covered by the Monthly Employment Survey (PME). From this data, the parameters of the theoretical models are estimated. Although our paper is an empirical one, it is necessary to summarize the theoretical support for the estimable econometric model. Our approach is based on Bontemps et al. (2000), with the possibility of wage determination presented by Mortensen (2003). The model is an extension of Burdett & Mortensen (1998) model, where workers and firms are homogeneous.

4.1. Theoretical Background

Our approach is based on the model developed by Bontemps et al. (2000), also incorporating the wage determination possibility described by Mortensen (2003). The model is an extension of Burdett & Mortensen (1998) model, where workers and firms are homogeneous. Bontemps et al. (2000) incorporate the possibility of heterogeneity in the firms' productivity as an attempt to obtain a better fit for the wage distribution. The assumptions of the model are:

A1. Workers and active firms are represented by a continuum of mass 1 and m, respectively; they are homogeneous with respect to the ability and the firms are heterogeneous in terms of productivity;
A2. Unemployed worker has a utility flow (discounted search costs) equal to b and workers receive wage offers from a distribution $F(w)$ (known) at a rate λ_0 and λ_1 when unemployed and employed, respectively;
A3. The wage offer distribution, $F(\cdot)$, is independent of the worker's state (employed or unemployed) and the support of F is denoted by $\text{supp}(F)$, where $\underline{w} = \inf(\text{supp}(F))$ and $\overline{w} = \text{supp}(\text{supp}(F))$;
A4. An employed worker leaves the job to unemployment at rate δ, which is called the rate of employment separation (quit rate) and workers and firms discount the future at a rate ρ.

Once hypotheses are presented, the next step is to define the strategies used by workers and firms, and define the equilibrium of the model that serves as a basis for empirical analysis.

4.1.1. Workers

From the worker's maximization problem and considering the previous assumptions, the value function for an employed worker, $V^e(w)$, can be written as

$$\rho V^e(w) = w + \lambda_1 \int_w^{\overline{w}} \Big(V^e(\tilde{w}) - V^e(w)\Big) dF(\tilde{w}) + \delta\Big(V^u - V^e(w)\Big), \tag{1}$$

where V^u is the value to be unemployed. Basically, this equation relates with the value of the employment state to the wage that the worker is receiving, plus the additional expected value of a wage offer that exceeds the value of the current wage plus the expected value associated with the return to the state of unemployment, which is negative. Furthermore, we have:

$$\frac{\partial V^e(w)}{\partial w} = \frac{1}{\rho + \delta + \lambda_1(1 - F(w))} > 0, \tag{2}$$

i.e., $V^e(w)$ is increasing in w.

The value flow for an unemployed worker is

$$\rho V^u = b + \lambda_0 \int_0^{\overline{w}} \max\left\{V^e(\tilde{w}) - V^u, 0\right\} \mathrm{d}F(\tilde{w}). \tag{3}$$

Since $V^e(w)$ is increasing, there is a reservation wage, w^r, making $V^e(w^r) = V^u$. That is, there is a determined wage that makes the worker indifferent between accepting a job and remaining unemployed.[3] Thus, from (1) and (3) we arrive at

$$w^r = b + (k_0 - k_1)\int_{w^r}^{\overline{w}} \frac{\bar{F}(\tilde{w})}{\beta + 1 + k_1\bar{F}(\tilde{w})}\,\mathrm{d}\tilde{w}, \tag{4}$$

where $k_0 = \lambda_0/\delta$, $k_1 = \lambda_1/\delta$, $\beta = \rho/\delta$, $\bar{F}(x) \equiv 1 - F(x)$. Note that in the definition of the reservation wage, the worker takes into account not only the flow of utility to be unemployed, but the arrival rates of job offers, the separation rate and the wage distribution in the economy. Before presenting the firms' strategy, it is necessary to comment on worker flows in steady state.

4.1.2. Steady-State Flows

Steady-state worker flows into and out of unemployment must be equal. Thus, we have that at any time $\lambda_0 u$ workers leave the unemployment state[4] and $(1-u)\delta$ fall in unemployment, where u is the measure of unemployed workers. The dynamics of unemployment is given by

$$\dot{u} = (1-u)\delta - u\lambda_0. \tag{5}$$

At steady state, $\dot{u} = 0$. So

$$u\lambda_0 = (1-u)\delta \quad \Leftrightarrow \quad u = \frac{1}{1+k_0}. \tag{6}$$

Therefore, this equation relates the unemployment rate with the duration (λ_0) and incidence (δ).

Moreover, the wage distribution in the stock of employed workers is denoted by $G(\cdot)$, where $E(w) = (1-u)G(w)$ is the fraction of individuals receiving a wage less than or equal to w in the stock of employed worker. The dynamics of $E(w)$ is

$$\dot{E}(w) = \lambda_0 F(w)u - \delta E(w) - \lambda_1(1 - F(w))E(w), \tag{7}$$

where $\lambda_0 F(w)u$ is the fraction of unemployed workers who find a job and receive a wage less than or equal to w, $\delta E(w)$ is the fraction of employed workers receiving a wage less than or equal to w and entering the unemployment state and $\lambda_1(1-F(w))E(w)$ is the fraction of employed workers who receive a wage less than or equal to w and find another job that pays more than w.

Again, at steady-state $\dot{E}(w) = 0$, implying that

$$G(w) = \frac{F(w)}{1 + k_1(1 - F(w))}. \tag{8}$$

Equation (8) establishes the structural relationship between the wages distribution earned by the stock of employed workers or earnings distribution, $G(\cdot)$, and the wage offer distribution, $F(\cdot)$.

Based on these relations, we can develop the analysis for the firm's behavior. We will analyze two types of wage determination: i) wage posting, where the firm unilaterally sets the wage; and ii) Nash bargaining, where, after workers and firms meet in the market, they bargain over the wages to be paid by the firm.

[3] It is assumed that the worker accepts a job when he/she is indifferent.

[4] In equilibrium, no firm posts a wage below the reservation wage of the worker, which implies that $\bar{F}(w^r) = 0$. Therefore, the risk of leaving the state of unemployment is equal to λ_0.

4.1.3. Wage Posting

In the wage posting setting, firms determine wages and, given these wages, each firm faces a labor supply curve, $l(w)$. The employment level of a firm that offers a wage w is

$$\begin{aligned} l(w) &= \lim_{\epsilon \to 0} \frac{1-u}{m} \frac{G(w) - G(w-\epsilon)}{F(w) - F(w-\epsilon)} = \frac{1-u}{m} \frac{\mathrm{d}G(w)}{\mathrm{d}F(w)} \\ &= \frac{1-u}{m} \frac{1+k_1}{\left[1 + k_1(1-F(w))\right]^2}, \end{aligned} \tag{9}$$

where $(1-u)(G(w) - G(w-\epsilon))$ is the fraction of employed workers who are receiving a wage in the range $[w-\epsilon, w]$, and $m(F(w) - F(w-\epsilon))$ is the fraction of firms offering a wage in the range $[w-\epsilon, w]$. This equation says that the fraction of workers who receive a wage w are uniformly distributed among firms that offer these wages. When $k_1 = 0$ (no on-the-job search), all firms have the same work force in equilibrium, which equals $(1-u)/m$.

Firms may differ in terms of labor productivity p, and p's distribution is denoted by $\Gamma_0(p)$ with $\underline{p}_0 \geq 0$ being the infimum of its support and $\bar{p}$ the supreme. Assume that $E_{\Gamma_0}(p) < \infty$.

It is considered that a worker generates a revenue flow equal to p and is independent of the number of workers in the firm. Thus, p is the labor productivity at the firm and the firm is of type p. The firm's objective is to maximize the profit flow at steady state:

$$\pi(p, w) = (p - w) l(w). \tag{10}$$

Bontemps et al. (2000) show that in the case of a continuous distribution of firms' productivities, there is a function K that maps $\mathrm{supp}(\Gamma)$ on $\mathrm{supp}(F)$ such that the set K_p is represented by a single point $K(p)$. So the first order condition for the profit maximization problem of the firm is

$$-\left[1 + k_1(1 - F(w))\right] + 2k_1 f(w)(p - w) = 0, \tag{11}$$

under the restriction that $w \geq \max\{w^r, w_{min}\}$, where $w = K(p)$. Firms with the lowest possible level of productivity will offer a wage $\underline{w}$. The second order condition is

$$f'(w)\left[1 + k_1(1 - F(w))\right] - k_1 f(w)^2 < 0, \tag{12}$$

which is equivalent to saying that $f(w)[1 + k_1(1 - F(w))]$ is decreasing. This implies that the theory can be tested. For the model to be admissible, a second order condition must be satisfied for all wages in the sample.

Bontemps et al. (2000) derive the following expression for $K(p)$:

$$K(p) = p - \left(1 + k_1 \bar{\Gamma}(p)\right)^2 \int_{\underline{w}}^{p} \frac{\mathrm{d}x}{\left(1 + k_1 \bar{\Gamma}(x)\right)^2}. \tag{13}$$

This is the fundamental equation of the model, because it defines the firms' strategy. That is, the wage offer is a function that depends on the firm's productivity, the level of friction in the labor market (k_1) and the distribution of active firms' productivity. Using the structural relation between $G(\cdot)$ and $F(\cdot)$, we can rewrite (13) as

$$K^{-1}(w) = w + \frac{1 + k_1 G(w)}{2k_1 g(w)}. \tag{14}$$

Therefore, the structural parameters of the model are necessary to infer the productivity level associated with a given wage. Given the model solution for the case of monopsonistic firms, we proceed now to describe the solution to the case of bilateral bargaining between firms and workers.

4.1.4. Nash Bargaining

Mortensen (2003) suggests that the hypothesis that wages are unilaterally determined by firms may not be admissible. Another feature of the wage posting model is to infer high levels of firms' productivities (Shimer, 2006). So another alternative for wage determination is the possibility that it might be the result of a Nash bargaining process between workers and firms. This process may occur as a result of the presence of unions and classes in the labor market, which is not a very unrealistic assumption for Brazilian labor markets. For example, Mortensen (2003) finds empirical evidence that the appropriate model for Danish data is the bargain. Obviously, this result does not necessarily apply to all economies.

The analysis is somewhat different from the previous problem. First, we can write the value function of a type p firm that pays a wage equal to w as

$$\rho J(p,w) = p - w - \left(\delta + \lambda_1 \bar{F}(w)\right) J(p,w). \tag{15}$$

Note that it is assumed that the value of a vacant position not occupied is zero (free entry condition). Rewriting (15), we have

$$J(p,w) = \frac{p-w}{\rho + \delta + \lambda_1 \bar{F}(w)}. \tag{16}$$

The value functions for employed and unemployed workers remain the same as those in section 4.1.1. After workers and firms meet, the wage is defined as the Nash solution of the bilateral bargaining process in respect of the surplus value of $V^e(w) - V^u$, for workers and $J(p,w)$, for firms, since the value of staying with the position not occupied is zero for the firm. I.e.,

$$\mathcal{W}(p) = \underset{w \geq w^r}{\arg\max} \left(V^e(w) - V^u\right)^{\alpha} J(p,w)^{1-\alpha}, \tag{17}$$

where $\alpha \in (0,1)$ represents the bargaining power of workers. Thus, the first order condition for an interior solution is

$$\alpha \frac{V^{e'}(w)}{V^e(w) - V^u} - (1-\alpha)\left(\frac{1}{p-w} - \frac{\lambda_1 f(w)}{\rho + \delta + \lambda_1 \bar{F}(w)}\right) = 0, \tag{18}$$

where $w \equiv \mathcal{W}(p)$. Again, $\underline{p} = \underline{w}$, which implies that $\mathcal{W}(\underline{p}) = \underline{w}$ since this is the only viable wage for the firm that could be accepted by the worker. Note that the "external option" for the worker is V^u, independent of the individual being unemployed or employed. This is because the bargaining process is *ex post* and as soon as the worker accepted the job, the only external option is unemployment. In other words, workers cannot return to previous employment. Moreover, it is assumed that workers observe the firm's productivity level when they find a job and thus can infer the resulting wages if they accept the job. Thus, if $\mathcal{W}(p)$ is increasing, employed workers only change to more productive firms.

The inverse function, $\mathcal{W}^{-1}(w)$, obtained from the wage offer function, $\mathcal{W}(p)$ can be derived from (18) as

$$p \equiv \mathcal{W}^{-1}(w) = w + \frac{1}{\frac{\alpha}{1-\alpha}\frac{V^{e'}(w)}{V^e(w) - V^u} + \frac{\lambda_1 f(w)}{\rho + \delta + \lambda_1 \bar{F}(w)}}. \tag{19}$$

This function relates with the productivity level p associated with a given wage w generated from (17).

Moreover, using the fact that $V^e(w^r) = V^u$, it follows that from (2) the surplus that the worker obtains from the matching is

$$V^e(w) - V^u = \int_{w^r}^{w} \frac{1}{\rho + \delta + \lambda_1 \bar{F}(w')} \, \mathrm{d}w'. \tag{20}$$

Substituting (2) and (20) in (19), we arrive at

$$\mathcal{W}^{-1}(w) = w + \frac{(1-\alpha)(\rho+\delta+\lambda_1\bar{F}(w))\displaystyle\int_{w^r}^{w}\frac{1}{\rho+\delta+\lambda_1\bar{F}(w')}\,\mathrm{d}w'}{\alpha+\lambda_1 f(w)(1-\alpha)\displaystyle\int_{w^r}^{w}\frac{1}{\rho+\delta+\lambda_1\bar{F}(w')}\,\mathrm{d}w'}. \tag{21}$$

Again, the model can be tested. As Mortensen (2003) and Shimer (2006) asserted[5] the model is admissible if $\frac{\partial \mathcal{W}^{-1}(w)}{\partial w} > 0$.

So, after describing the two different mechanisms of wage determination, the following subsection details how the empirical analysis was performed. It is divided into the database and econometric analysis.

4.2. Data Set

Our analysis is carried out based on a longitudinal microdata from the Monthly Employment Survey (PME) of 2009,[6] which is a database collected by the Brazilian Institute of Geography and Statistics (IBGE). This survey is conducted in six major metropolitan areas of Brazil, namely: Salvador, Recife, Belo Horizonte, São Paulo, Rio de Janeiro and Porto Alegre. The workers who were interviewed answer several questions related to the labor market and demographic characteristics. Some questions are fundamental to the analysis, as job search duration of an unemployed worker, employment duration, wages, labor market position (employed or unemployed).

The sub-sample is obtained after the initial selection of workers who answered all four consecutive interviews in 2009. Individuals who were out of the labor force were excluded because the model allows only two states, employment and unemployment. Individuals were selected from 16 to 55 years old. All workers who were working on some interviews in the public sector, as self-employed, employer or unpaid, or who were in jobs with working hours shorter than 30 hours were excluded from the subsample. Finally, in order to eliminate potential outliers, we excluded wages below R$ 300.00 (808 observations)[7] and 1% higher wages (370 observations) from the subsample. After this cut the final subsample totaled 46,367 workers.

With respect to the generated variables, it was possible to observe that for unemployed workers we obtain the elapsed time of job search until the date of the first interview,[8] t_{0b}, and calculate the residual time in which the individual remained unemployed for the other three remaining interviews, t_{0f}. If the worker leaves the state of unemployment to employment in this period, we observe the accepted wage, w_0, which is a realization of the wage offer distribution, $F(w)$. Furthermore, for individuals who responded that they were seeking a job for 5 years or more,[9] the unemployment durations were treated as left censored, $d_{0b} = 1$ and for those who remained unemployed in remaining interviews the unemployment duration was considered as right censored, $d_{0f} = 1$. Thus, we have $d_{0b} = 0$ on the date of the first interview if the worker is unemployed for less than 60 months and $d_{0f} = 0$ if the unemployed worker leaves the state of unemployment in the three months after the date of the first interview.

For employed workers, we observe the job durations on the date of the first interview, t_{1b}, and wages, w_1, which is a realization of $G(w)$. Likewise, the time that the worker remained employed in

[5] Note that the model by Shimer (2006) differs from Mortensen (2003) model, but produces similar results.

[6] Menezes-Filho & Picchetti (2000) and Penido & Machado (2002) also use only one year to perform a duration analysis.

[7] As of January, 2009, the minimum wage was equal to R$ 415.00. After the adjustment in February, 2009, the minimum wage rose to R$ 465.00.

[8] For the interested reader we can provide a diagram showing a detailed account for all possible types of labor market dynamics sampled.

[9] Due to the wording of the questionnaire, the job search duration reported by the respondent is limited to 60 months.

relation to other interviews, t_{1f}, is computed. The worker can leave the current job to unemployment, $v = 1$, or to another job, $v = 0$. Both alternatives are considered and, if the worker stays unemployed less than 1 month before going into another job, it is considered as a job-to-job transition, as it is done in Sulis (2008). Moreover, if the worker remains in the same job during the remaining interviews, the employment duration is right censored, $d_{1f} = 1$.

Tabela 5 provides descriptive statistics regarding the final subsample. Included are some statistics related to demographic characteristics in order to provide an overview of the composition of each metropolitan market. All metropolitan areas have a male proportion between 53–54%. In terms of educational level, the structure is similar across regions, except for some differences. São Paulo has the largest proportion of workers with higher education, 12%, while in the metropolitan area of Recife, that proportion is only 5%. The mean age of workers is 33 years, and there is not a significant difference between regions.

Table 5. Descriptive statistics of subsample.

	Total	Recife	Salvador	Belo Horizonte	Rio de Janeiro	São Paulo	Porto Alegre
Number of workers	46,367	3,272	5,380	9,333	8,377	12,564	7,441
Employed (%)	87	83	79	90	88	86	91
Unemployed (%)	13	17	21	10	12	14	9
Men (%)	53.81	53.55	53.75	54.61	53.67	53.38	53.81
Education							
Uneducated (%)	1	1	1	1	1	1	0
Literacy (%)	23	21	20	25	24	19	26
Elementary education (%)	20	18	18	22	21	18	24
High School (%)	48	55	54	45	44	50	44
University (%)	9	5	7	7	10	12	6
Age	32.93	32.79	32.87	32.28	34.48	32.35	33.09
(std. dev.)	(10.12)	(9.66)	(9.59)	(10.14)	(2.10)	(10.23)	(10.23)
Unemployed							
Left censored (%)	0.78	0.52	1.45	0.00	0.28	1.75	0.94
Right censored (%)	74.21	72.03	86.27	58.18	86.91	72.74	64.59
Incomplete duration; mean	8.00	6.26	25.09	4.66	11.77	7.59	7.09
(std. dev.)	(8.79)	(6.97)	(10.91)	(4.5)	(11.52)	(6.99)	(8.22)
Employed							
Left Censored (%)	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Right censored (%)	87.93	74.81	89.33	83.95	94.42	90.13	86.59
Incomplete duration; mean	55.95	57.54	59.9	50.43	62.85	54.77	53.95
(std. dev.)	(67.28)	(65.74)	(71.74)	(63.95)	(72.89)	(65.39)	(64.78)
Transitions to Unemployment	3,364	582	297	898	270	680	637
Transitions to another job	1,504	1,998	159	443	143	386	275
Earnings Distribution $G(w)$							
Minimum	300	300	300	300	300	300	300
P_{10}	465	415	415	465	465	480	465
Q_1	500	465	465	470	500	600	550
Q_2	700	505	550	600	700	800	700
Q_3	1,000	700	800	915	1,000	1,200	1,000
P_{90}	1,680	1,000	1,400	1,500	1,700	2,000	1,600
P_{90}/P_{10}	3.61	2.41	3.37	3.23	4.17	3.44	3.66
Mean	926.48	658.51	786.52	867.61	927.54	1,087.44	936.29
(std. dev.)	(751.37)	(406.27)	(647.12)	(699.6)	(765.88)	(879.76)	(681.47)

The mean of unemployment duration ($t_{0b} + t_{0f}$), also considering the incomplete duration, is 8 months in Brazil. Because there are workers who remained unemployed, it is likely to understate the actual mean of unemployment durations. The metropolitan area of Belo Horizonte has the lowest average,

4.66 months, and Rio de Janeiro has the largest, 11.77. For employed workers on the date of the first interview, the mean of employment duration ($t_{1b} + t_{1f}$) is 55.95 months. Rio de Janeiro has the highest mean, 62.85, while Belo Horizonte has the lowest, 50.43 months.

Regarding the wage distribution of employed workers on the date of the first interview, we observe a considerable difference between the metropolitan areas studied. The metropolitan region of São Paulo has the highest mean, R$ 1,087.44, and Recife has the lowest average, R$ 658.51, which represents only 60% of the state of São Paulo. Observing the ratio between the ninetieth and tenth percentiles, Recife also has a low-wage dispersion when compared to other metropolitan areas.

Finally, Figura 3 shows kernel estimates of the density functions of the accepted wages distribution by workers who left the unemployment state during the survey, $f(w)$ (wage offer distribution), and the wages distribution earned by workers who were employed at the date of the first interview, $g(w)$ (earnings distribution). Theory predicts that the wage offer distribution is dominated by the earnings distribution, due to the fact that the employed workers can migrate to jobs that pay higher wages. Referring to Figura 3, there is little difference between $f(w)$ and $g(w)$ for the metropolitan areas of Recife and Salvador. However, for other regions, especially for São Paulo and Porto Alegre, there is a considerable difference between these distributions, which may be showing greater mobility of workers to jobs that pay better wages.

4.3. Structural Estimation

Job search models are usually estimated by maximum likelihood. In the case of the model in question, the wage offers distribution $F(w)$ does not have an explicit form. Thus, Bontemps et al. (2000) proposed a three-stage procedure to estimate the nonparametric model.

First, we need to specify the likelihood function. The model predicts that the unemployment duration is exponentially distributed with parameter λ_0, due to the fact that the time between two events of a Poisson process is exponentially distributed. The exponential distribution has the characteristic of being memoryless. Thus, this distribution has the property that the elapsed unemployment duration until the date of interview, t_{0b}, and the residual unemployment duration(referring to remaining interviews), t_{0f}, are independent and exponentially distributed[10] with parameter λ_0. The probability of sampling an unemployed worker at the date of the first interview is equal to $1/(1+k_0)$ and if he/she receives a wage offer, w_0, this wage offer will be a realization of $F(\cdot)$. Thus, we can write the likelihood function for an unemployed worker as

$$\mathcal{L}_d = \frac{\lambda_0^{2-d_{0b}-d_{0f}}}{1+k_0} \exp[-\lambda_0(t_{0b}+t_{0f})] f(w_0)^{(1-d_{0f})}, \tag{22}$$

where d_{0b} is equal to 1 if t_{0b} is left censored, and d_{0f} is equal to 1 if t_{0f} is right censored.

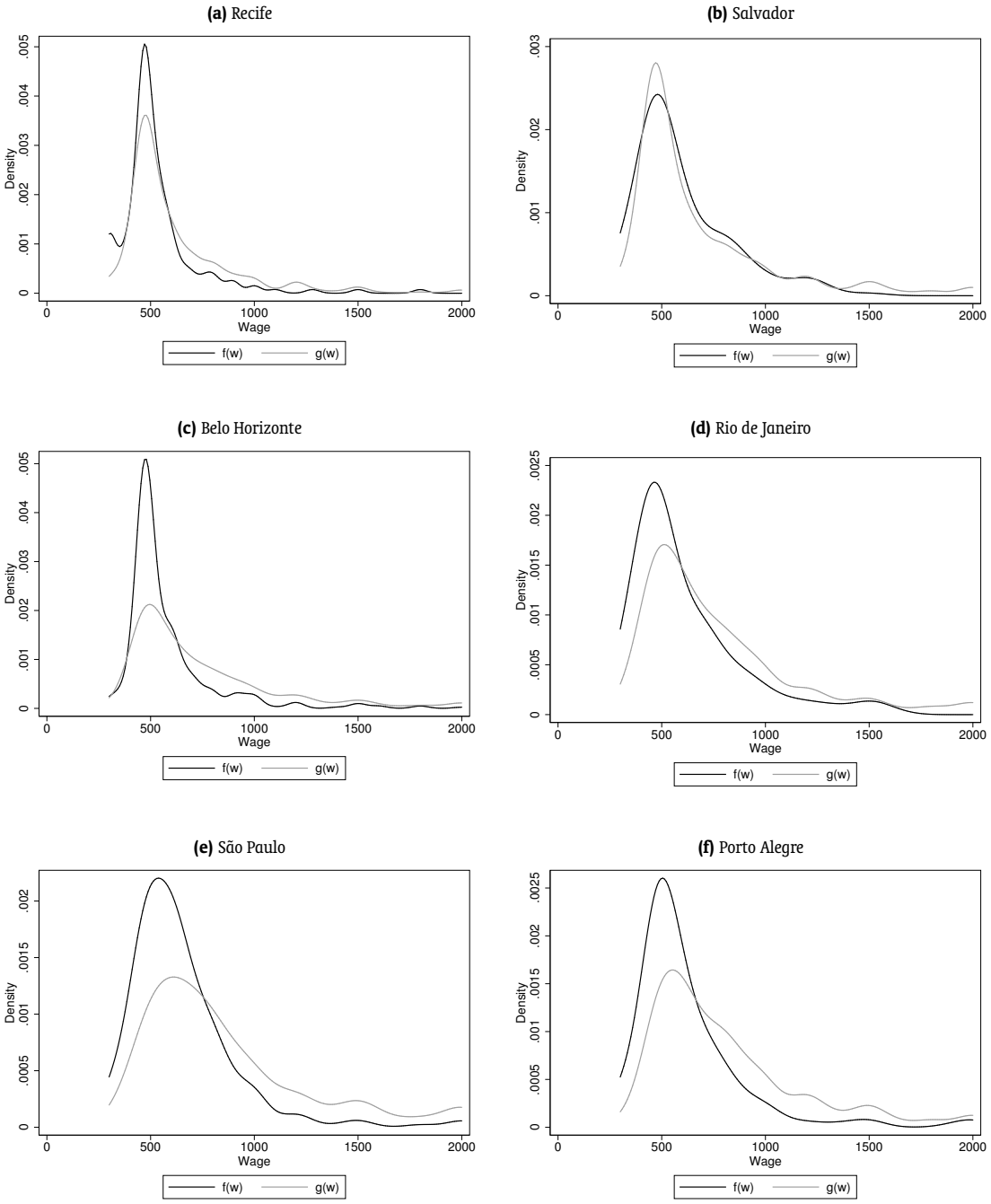
The probability that an employed worker is observed is equal to $k_0/(1+k_0)$ and the wage received w_1 is a realization of the earnings distribution, $G(\cdot)$. Job duration is exponentially distributed with parameter θ, where $\theta = \delta + \lambda_1 \bar{F}(w_1)$, given w_1. The probability that a worker leaves the current job to unemployment or to another job are equal to $\delta/(\delta + \lambda_1 \bar{F}(w_1))$ and $\lambda_1 \bar{F}(w_1)/(\delta + \lambda_1 \bar{F}(w_1))$, respectively. Again, t_{1b} and t_{1f} are the elapsed and residual job duration, respectively, and are assumed to be independent. The likelihood function for an employed worker is

$$\mathcal{L}_e = \frac{k_0}{1+k_0} g(w_1) \Big(\delta + \lambda_1 \bar{F}(w_1)\Big)^{1-d_{1b}} \exp\Big[-\Big(\delta + \lambda_1 \bar{F}(w_1)\Big)(t_{1b}+t_{1f})\Big] \times \Big[\delta^{v} \Big(\lambda_1 \bar{F}(w_1)\Big)^{1-v}\Big]^{1-d_{1f}}, \tag{23}$$

where d_{1b} is equal to 1 if t_{1b} is left censored, d_{1f} is equal to 1 if t_{1f} is right censored, and v is equal to 1 if the transition is to unemployment and 0 if it is to another job. Therefore, one can write the likelihood

[10] For more details see Lancaster (1990) and Cameron & Trivedi (2005).

Figure 3. Kernel estimates of wage distributions.



function for a sample of size N as

$$\mathcal{L} = \prod_{i=1}^{N} \mathcal{L}_{di}^{x} \mathcal{L}_{ei}^{(1-x)}, \tag{24}$$

where x is equal to 1 if the worker is unemployed at the time of the first interview and 0 if he/she is employed.

Moreover, using the relation between $F(\cdot)$ and $\Gamma(\cdot)$, we can generate the density function of firms' productivities, $\gamma(\cdot)$. Thus, considering that the wage policy function[11] is $w(p)$, the density function of firms' productivity is

$$\frac{\mathrm{d}\Gamma(p)}{\mathrm{d}p} \equiv \gamma(p) = f(w)w'(p). \tag{25}$$

Using the inverse relationship with $w(p)$, we can rewrite (25) as

$$\gamma(p) = \frac{f(w)}{(w^{-1})'(w)}, \tag{26}$$

where $(w^{-1})'(w) = \frac{\partial w^{-1}(w)}{\partial w}$. Therefore, it becomes possible to obtain an expression for $\gamma(p)$ for the two wage determination cases.

As already mentioned, $F(w)$ does not have an explicit form, which makes the use of the likelihood function alone impossible. Thus, we adopt the procedure proposed by Bontemps et al. (2000). To estimate the bilateral bargaining model, we set $\rho = 0$. As it is also assumed in Burdett & Mortensen (1998) in the wage posting model, due to the fact that this parameter cannot be identified from the data. We also set $\alpha = 0.5$ and $\widehat{w}^r = \min\{w_i\}_{i=1}^{N}$.

Note that by this procedure, the frictional parameters estimation (λ_0, λ_1 and δ), which is held in the first two steps, is based only on worker's behavior. Thus, it is expected that the estimatives of these parameters are consistent with different forms of firms' behavior (Bontemps et al., 2000). Another important point is the fact that for the model not to be rejected by the data it must have $\gamma(p) > 0$. Then, from equation (26) wages and productivities must be positively related, which is a restriction generated by the theoretical model. For the estimation of $g(w)$, we use the Gaussian kernel function.

Finally, standard errors or confidence intervals of $\hat{\lambda}_0$, $\hat{\lambda}_1$ and $\hat{\delta}$, are obtained using bootstrap replications procedures, including the first stage. The bootstrap method is also useful to generate the confidence interval of $k_1 = \lambda_1/\delta$, serving as an alternative to the delta method. The results follow.

5. RESULTS

This section presents the results of the structural estimation of the job search model. The results are divided into two sections. The first presents estimation results for the frictional parameters that are independent of the any wage setting behavior of firms. The second section presents the results for the estimated productivities distribution considering the two possibilities of wage determination.

5.1. Frictional Parameters

The frictional parameters are a measure of the search frictions present in the labor market that represent the difficulty of workers and employers to establish working relationships. Tabela 6 presents the estimatives for the six metropolitan areas that constitutes the PME.

First, the arrival rate of wage offers by unemployed workers, λ_0, shows considerable heterogeneity among the metropolitan areas studied. This rate in the metropolitan area of Belo Horizonte is 22.77%,

[11] In the case of wage posting models, $w(p) \equiv K(p)$, and to the bilateral bargain $w(p) = \mathcal{W}(p)$.

Table 6. Estimates of frictional parameters.

	λ_0 unemployed arrival rate offer	λ_1 employed arrival rate offer	δ separation rate	k_1 level of friction
Recife	0.1497 [0.1361;0.1618]	0.0089 [0.007;0.0106]	0.0196 [0.0186;0.0205]	0.4515 [0.3467;0.5571]
Salvador	0.0898 [0.0849;0.0942]	0.0133 [0.0111;0.0159]	0.0153 [0.0146;0.016]	0.8670 [0.7283;1.0824]
Belo Horizonte	0.2277 [0.2191;0.2397]	0.0307 [0.0275;0.0343]	0.0164 [0.0158;0.0171]	1.8683 [1.6828;2.1442]
Rio de Janeiro	0.0915 [0.086;0.096]	0.0209 [0.0183;0.024]	0.0115 [0.0111;0.012]	1.8182 [1.536;2.1304]
São Paulo	0.1284 [0.1234;0.1329]	0.0306 [0.0271;0.0337]	0.0141 [0.0136;0.0145]	2.1697 [1.8963;2.4559]
Porto Alegre	0.1615 [0.1523;0.1713]	0.0311 [0.0271;0.0356]	0.0141 [0.0135;0.0146]	2.2078 [1.8941;2.6102]

Notes: 2.5% and 97.5% percentiles of bootstrap distribution. 100 replications. Time unit: months.

while in Salvador this rate is only 8.98%, which implies an expected completed unemployment duration[12] of approximately 11 months. This rate has a direct influence on unemployment rates, because the faster unemployed workers find jobs, the lower is the unemployment rate in the economy, given a level of destruction of employment relations (separation rate).

Compared to other studies, the estimated values of λ_0 in our paper are relatively higher. In Bontemps et al. (2000) this rate is on average 0.07 and in Sulis (2008) this rate ranges between 0.04 to 0.07, but the author highlights the fact that the measure of unemployment duration used in the study is the time that the worker remains outside the administrative records used, which may be caused by unemployment, employment in public service, self-employment and inactivity. On the other hand, according to Bontemps et al. (2000) the parameters estimated by Kiefer & Neumann (1993) are two times higher than those estimated by the first authors, which would be closer to the values we have found here.

Regarding the arrival rate of wage offers for employed workers, λ_1, we find that this rate is considerably smaller than λ_0. This is in accordance with the international literature. Compared to other metropolitan areas, Recife has a considerably low level of λ_1, only 0.89%. São Paulo and Porto Alegre have a rate approximately three times greater than Recife's. This fact is already an indication that workers have greater mobility in these metropolitan areas, which implies higher competition among employers.

The separation rate is more homogeneous among the metropolitan areas. On average, approximately 1.5% of employed workers become unemployed per month. Rio de Janeiro has the lowest separation rate among the regions analyzed, indicating that the labor market in this region has a lower turnover. The results coming from the international literature are distinct. Van den Berg & Ridder (1998) and Bontemps et al. (2000) estimate a rate on average of 0.005 and 0.0061 for the Netherlands and France, respectively. On the other hand, Sulis (2008) estimated a rate of 0.0128 for Italy and Bunzel et al. (2001) estimate δ between 0.01 to 0.02 for Denmark, which are closer to those estimated by us. Kyyrä (2007) found high values of δ for Finland, estimatives of δ in his work range from 0.05 to 0.01.

It is about the right place to deepen our analysis and try to connect our estimates with the prevalent regional inequalities in Brazilian labor markets described in section 2. We start by rationalizing regional

[12] As the duration of unemployment is assumed distributed exponentially with parameter λ_0, the expected full length is simply equal to $1/\lambda_0$.

inequalities on unemployment rates and the estimated deep parameters of our model, say, λ_0, λ_1 and δ on Tabela 6. Since δ is roughly homogenous, we focus on both λ_0 and λ_1. For instance, let us start with Bahia, which presented in 2009, accordingly to Tabela 3, an unemployment rate of 11.37 and compare it to Porto Alegre (with a 5.42 unemployment rate). Do our structural estimates have a bite on that regional inequality? We think the answer is yes. While Salvador has the lowest rate of wage offer while unemployed (0.0898) and one of the lowest rate of wage offer while employed (0.013), ahead of Recife only, Porto Alegre has almost twice Salvador's wage offer while unemployed (0.1615) and almost three times its rate of wage offer while employed (0.0311). So, no wonder Salvador and Porto Alegre occupy the extremes in terms of unemployment inequality in Brazil. More specifically, from an efficiency point of view any policy aiming improvements on Salvador's unemployment rate must act on its very low arrival of wage offer while employed, something that has to do with its internal labor market.

If we want to draw a more regional explanation here, definitely we would look at the values of λ_1, the arrival rate of wage offer while employed. The second column in 6 brings us an important hint: metropolitan areas belonging to the South and Southeast regions have at least twice, sometimes three times, the corresponding λ_1 of Recife and Salvador, both areas in the Northeast. In other words, the scanty debate about regional differences in unemployment rate can be both backed up and, more importantly, can be deepened with our results, by noting that the inefficiency of internal labor markets in the Northeast is a driving force of that phenomena.

From λ_1 and δ one can obtain k_1, which is a parameter of great importance in the model because it is a measure of the level of friction in the labor market (van den Berg & van Vuuren, 2003). This is because k_1 measures the number of expected job offers to be received by a worker during an episode of employment, reflecting the level of competition among firms in the market. Hence, in a market that employed workers receive alternative offers at a higher rate, employers have incentives to offer better wages to reduce the outflow of workers. Besides this effect, of course, workers move more quickly into jobs that pay better wages, which implies that the earnings distribution, $g(w)$, tends to dominate the wage offer distribution, $f(w)$. Now, let us step into the regional inequality in wages appearing in Tabela 1.

The effect of k_1 on the distributions is evident at Figura 3 (page 87). It is observed that for the metropolitan areas that have lower k_1 values, the distributions of $f(w)$ and $g(w)$ are closer, which is the case of the metropolitan areas of Recife and Salvador. However, São Paulo and Porto Alegre have values of k_1, approximately four times higher than that of Recife and the effect is that $g(w)$ moves away from $f(w)$, i.e., workers have a faster wage growth in these regions. Thus, we got some evidence that the higher the level of friction in the market (lower k_1) is, the more concentrated the wage distribution is because workers have a low transition rate to jobs that pay better wages. Hence, our structural estimates corroborate past finds and, again, shed light on the debate: regional inequality in wages, besides being an outcome of its regional human capital distribution, can be rationalized as inequality of labor market search frictions.

How reasonable are our estimates? Tabela 7 gives us some hints: it shows a comparison between our estimatives[13] and those estimated in other studies. The estimated value of λ_0 in our paper is considerably higher than those on other studies, but Carvalho (2012) has a value close to that estimated in this study. As explained earlier, some authors make use of nonemployment durations, which include individuals who remained out of the workforce instead of using strictly unemployment duration data. Therefore, it is expected that the nonemployment duration mean be higher than the unemployment duration mean, which is calculated based on those workers who are actively seeking employment. There is also a chance of encountering a memory bias due to the fact that respondents underestimate the actual time they are searching for a job, which contribute to the estimation of high values of λ_0. As to λ_1, our estimatives are only slightly less than the estimated in van den Berg & Ridder (1998).

[13]The values in Tabela 7 are related to the estimation of the model for the total sample.

Table 7. Comparison of estimated values of frictional parameters.

	Country	λ_0	λ_1	δ	k_1
Our results	Brazil – PME	0.129	0.022	0.015	1.467
Van den Berg & Ridder (1998)	Netherlands	0.033	0.047	0.005	9.400
Bontemps et al. (2000)	France	0.063	0.008	0.006	1.333
Bunzel et al. (2001)	Denmark	0.028	0.010	0.015	0.235
Sulis (2008)	Italy	0.043	0.006	0.013	0.462
Carvalho (2012)	Brazil – PPV	0.137	0.005	0.005	1.000

By comparing k_1, the value found here is close to the estimate for France by Bontemps et al. (2000). However, in van den Berg & Ridder (1998), the value of k_1 is estimated at about 9.40, which is considerably high and the authors find a low level of monopsony power by firms. On the other hand, Sulis (2008) estimated a low k_1 which means high levels of monopsony power.

In the model with homogeneous firms, such as Burdett & Mortensen (1998), wage dispersion is caused only by search frictions. However, when we include heterogeneity in the firms' productivity, the dispersion is also caused by differences of firms. The next section investigates the productivity distribution from the two possibilities of firms' behavior assumed in this work and trace a parallel between that and regional differences in labor productivity.

5.2. Productivities Distribution

As we saw in section 2.3, the first decade of the 21st century witnesses a falling labor productivity all over Brazil. Here, we will try to give alternative explanations not for this decline but for the still prevalent regional differences in productivity. The productivity distribution is obtained from the third step of the estimation process (Bontemps et al., 2000). Such step is related to finding the productivity levels associated with observed wages in the sample, exploiting the first order conditions, given the frictional parameters estimated in previous steps. In addition, for each productivity level, we estimate the corresponding value of the density function, $\gamma(p)$ associated to $\Gamma(\cdot)$ which is one of the model's primitive.

This step is performed for each possible wage determination analyzed: bilateral bargaining and wage posting. One of the restrictions shared by both wage determination mechanisms is to observe an increasing relationship between wages and productivities. This condition is satisfied for the bilateral bargaining case, but not wage posting. Mortensen (2003) finds the same result for the labor market in Denmark. This result[14] can be seen in Figura 4. In the wage posting case, the theory is rejected at lowest observed wages, where the relationship is decreasing, implying that those wages cannot be profits maximizing for firms. That is, the model does not explain the left tail of the wage distribution. The shape of the relationship between wages and productivity is similar to that found by Mortensen (2003, p.103).

Besides the fact of not observing an increasing relationship between wages and productivity, the wage posting model generates implausible values for the firms' productivity levels. Tabela 8 shows the estimated productivity levels for the stock of employed workers on the date of the first interview. As the largest and most developed metropolis, it is expected that firms in São Paulo have the highest average productivity level. However, the metropolitan area of Recife has the highest average productivity level and after Recife, Salvador has the second highest. This may be due to the fact that estimated productivity

[14]It represents the relationship between wages and productivity for the metropolitan region of São Paulo. Other metropolitan areas have an analogous pattern.

Figure 4. Relationship between wages e productivities.

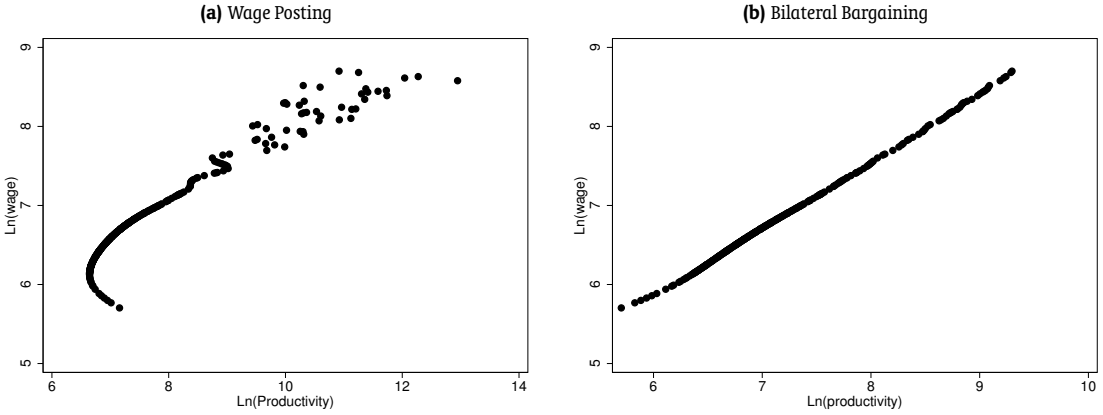


Table 8. Estimated productivities distribution: wage posting.

	Minimum	P_{10}	Q_1	Q_2	Q_3	P_{90}	P_{90}/P_{10}	Mean
Recife	849.94	851.63	851.63	1,149.04	2,812.47	6,298.05	7.40	5,903.52
Salvador	730.30	748.52	748.52	1,109.06	2,308.59	8,190.00	10.94	5,350.19
Belo Horizonte	642.59	661.69	676.81	941.36	1,993.42	6,159.22	9.31	4,264.10
Rio de Janeiro	694.92	712.04	747.33	1,174.25	2,327.15	8,801.40	12.36	4,424.64
São Paulo	768.79	781.69	883.96	1,268.75	3,139.06	8,156.23	10.43	4,147.55
Porto Alegre	694.14	710.52	788.53	1,068.96	2,070.59	6,424.08	9.04	4,142.61

Note: P_{10}, P_{90}, Q_1, Q_2, Q_3 are percentiles e quartiles.

levels are too high for firms that pay the highest wages and are in the right tail of the distribution. Therefore, as the probability mass at the end of the distribution tends to zero, the estimated productivity level tends to be extremely high, inflating the mean. In Figura 3 (on page 87), we find that the density function of wages to Recife is closer to zero for higher wages than to São Paulo.

Shimer (2006) and Mortensen (2003) noted that the model with wage posting tends to estimate implausible productivity levels. Sulis (2008) also found high values for the estimated productivity and restricts its analysis to qualitative questions. Bontemps et al. (2000) argue that large differences in productivity and wages for some firms are due to the presence of the employer's large capital stock, not treated in the model, which would generate a positive effect on labor productivity. Hence, from now on we stick to the bilateral bargaining model both on empirical grounds and on the Brazilian union structure described, for instance, in Amadeo (1994).

Tabela 9 shows the estimated productivity distribution for employed workers, considering the case in which wages are determined as a result of a bilateral bargaining process. In this case, the metropolitan area of Recife has the lowest average productivity level, R$ 956.46. Now, São Paulo has, on average, the highest estimated labor productivity level, R$ 1,617.22, approximately 69% higher than Recife's average. Note that the values are much smaller than those presented in Tabela 8. In terms of dispersion and considering the ratio between the ninetieth and tenth percentiles, São Paulo and Rio de Janeiro have the greatest productivity dispersion.

An important result is obtained when comparing the metropolitan areas of Rio de Janeiro and Porto Alegre. According to Tabela 5, the mean wage of an employed worker is R$ 927.54 and R$ 936.39 for Rio de Janeiro and Porto Alegre, respectively. However, Porto Alegre has a lower average productivity

Table 9. Estimated productivities distribution: bilateral bargaining.

	Minimum	P_{10}	Q_1	Q_2	Q_3	P_{90}	P_{90}/P_{10}	Mean
Recife	300.00	515.61	593.03	656.90	1,018.52	1,591.67	3.09	956.46
Salvador	300.00	507.43	581.71	717.42	1,166.16	2,317.39	4.57	1,174.23
Belo Horizonte	300.00	568.27	574.95	763.98	1,289.42	2,373.75	4.18	1,265.98
Rio de Janeiro	300.00	576.47	625.32	926.82	1,434.59	2,775.68	4.81	1,372.77
São Paulo	300.00	605.46	767.02	1,056.84	1,757.31	3,164.49	5.23	1,617.22
Porto Alegre	300.00	575.30	686.68	904.65	1,398.28	2,504.62	4.35	1,352.76

Note: P_{10}, P_{90}, Q_1, Q_2, Q_3 are percentiles e quartiles.

level than that of Rio de Janeiro. Thus, it was expected that workers were better off in the region with more productive firms. However, from Tabela 6, we have that the level of search frictions in the labor market is higher in Rio de Janeiro than in Porto Alegre. Therefore, workers in this region have a higher transition rate towards more productive firms that pay higher wages. This result is important because it reflects the inefficiency problem due to imperfect information in the labor market, where less productive firms retain workers with low wages even if there are jobs associated with higher levels of productivity and wages. In essence, Tabela 8 reflects those figures present on Tabela 4 (page 78). However, our methodology allows us to go further: let us move to the density functions.

The estimated distributions have similar shapes, with a concentration of firms at a low productivity level and a long right tail.[15] However, for example, it may be noted that the firms' productivities distribution is less dispersed in Recife than it is in São Paulo. And this is crucial to understand the link between firms' productivities, wages and labor market frictions, bringing a new host of possible explanations to regional inequality.

To analyze the difference between the productivity levels and wages received by workers, it is interesting to observe what is called the "monopsony power index". This index is defined by:

$$\mu(p) = \frac{p - \mathcal{W}(p)}{p}, \tag{27}$$

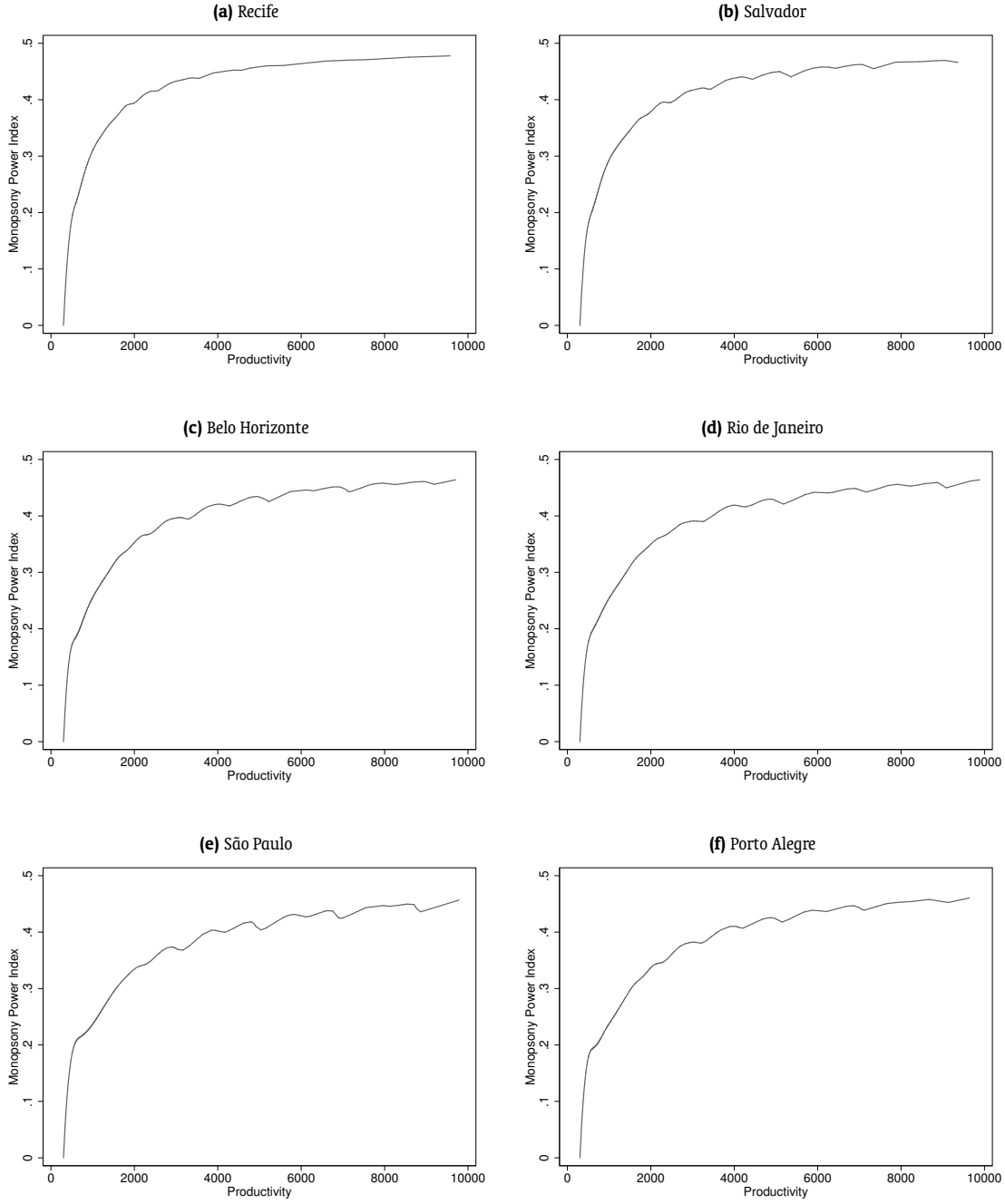
where $\mathcal{W}(p) \equiv w$. The analysis is performed only for the bargaining model, due to the fact that the wage posting model does not generate acceptable results from the theoretical point of view. This index provides the share of labor productivity that is appropriated by the firm. The index is equal to 0 in the case of wages equal to productivity, and is equal to 1, in which case the firm appropriates all gains of productivity.

Thus, from the results obtained in the case of wage determination via bilateral bargaining one can infer a lower monopsony power compared to the case where firms post wages, as evidenced by in Bontemps et al. (2000) and Sulis (2008). Part of this difference is explained by the fact that if the firms post wages, they have the full power on the wage determination and the only factor influencing the reduction in their power would be the competition between firms. In the case of bilateral bargaining, apart from the competition with other firms, workers have some power over the wage determination.

Figura 5 shows the monopsony power index for the six metropolitan regions at each productivity level. For lower productivity levels, all regions provide similar values for its monopsony index. In addition to these levels, we find that the resulting wages are closer to the labor productivity value. This is probably because less productive firms experiencing a higher rate of workers outflow, so they react by increasing the wages of workers in order to reduce this outflow. This index grows rapidly up to the productivity level around R$2,000. At this productivity level, the index is approximately 40%

[15] The interested reader can obtain the graph for estimated productivities from the authors.

Figure 5. Monopsony Power Index.



and 30% in Recife and São Paulo, respectively. This result reveals that employed workers in Recife are in a worse situation than those in São Paulo. This is because, for the same productivity level, workers in Recife receive relatively lower wages. What can be partly explained by the higher level of friction present in the labor market in this region.

So, what does the trio "monopsony power", productivity and wages can explain in terms of regional labor market inequalities? We think that, in a sense, we could interpret our evidence by noting that Figura 5 seems to back up a pattern where Recife and Salvador (both in the Northeastern region), specially between a productivity range of 1,000–3,000, present higher monopsony power due to either existing higher labor market frictions (see Tabela 7 on page 91) or *ceteris paribus* the role of unions on these regions, the latter point having some support on papers by Amadeo (1994), Carneiro (1998), and Menezes-Filho, Chahad, Zylberstajn, & Pazello (2008). Although we offer new initial hints, we think we still lack a complete structural explanation on the regional productivity inequality in Brazil.

6. CONCLUDING REMARKS

By considering the presence of search frictions in the labor market, job search theory is able to generate interesting theoretical results, such as involuntary unemployment, monopsony power and wage dispersion. However, although there is already a literature on empirical studies based on models of job search which are well developed internationally,[16] this literature is almost nonexistent in Brazil. Thus, this study sought to develop an empirical analysis for the Brazilian labor market based on the job search theory and, more importantly, we use the deep estimated structural parameters to shed new light over the important debate about labor regional inequality in Brazil.

To this end, the paper developed a comparison between metropolitan areas that constitute the PME based on a estimated equilibrium job search model with continuous productivity dispersion. Two forms of wage determination are considered: bilateral bargaining and wage posting. We tried to get evidence to determine which way is best suited to the Brazilian labor market, and found favorable evidence to the bilateral bargaining solution. The work follows the paradigm of structural econometrics by incorporating theoretical constraints upon model estimation. The analysis is performed using the nonparametric method developed in Bontemps et al. (2000). From the estimation of frictional parameters and the type of wage determination made, the method allows to recover the productivity distribution (unobserved) associated with the observed wages distribution.

Regarding the results, there are differences in the values of estimated frictional parameters for the regions. The unemployment duration is considerably lower in the metropolitan area of Belo Horizonte, 4.39 months, while for Salvador, the expected duration of a complete episode of unemployment is approximately 11 months. Metropolitan areas like Recife and Salvador have a higher level of search friction, which has direct impact on competition between firms. With respect to the type of wage determination, the determination hypothesis via a bilateral bargaining process seems more appropriate than the assumption of wage posting. In the case of wage posting, we are not able to find the (theoretically expected) strictly increasing relationship between wages and productivity, which is a violation of the theoretical restriction that wage policy function must be increasing in relation to firm's productivity. This condition is satisfied if we assume the case of bilateral bargaining. This result is the same found by Mortensen (2003) to Denmark.

From the estimated productivity values, we calculated the proportion of productivity that remains within the firm, i.e., it is not appropriated in the form of wage by the worker. Comparing the metropolitan areas of Recife and São Paulo, one realizes that the latter contains lower levels of monopsony power than the former. For the same level of productivity, a worker in the metropolitan region of São Paulo has a higher wage. This is possibly a result of a higher level of competition between firms in this region caused by a lower level of search friction in the labor market. In terms of public policy, it is evident

[16] See Eckstein & van den Berg (2007).

that the information flow among workers and employers have a key role in the labor market. Thus, employment agencies could possibly reduce the information cost, reducing the inefficiencies generated by search friction in the market. Moreover, more flexible labor laws can reduce the costs associated with changes between jobs by workers decreasing regional disparities on earnings.

Overall, we are able to rationalize Brazilian regional inequalities on unemployment rates and the estimated deep parameters of our model, say, λ_0, λ_1 and δ. In fact, our results strongly suggest that an important share of regional unemployment rate inequality can be explained by search frictions, more specifically the arrival rate of job offers while employed (λ_1). Metropolitan regions in the Northeast have much lower λ_0 and λ_1 *vis a vis* areas in the South or Southeast regions. This is a new, and much more precise, result worth considering on the regional inequality debate in Brazil.

Also, regional inequality in wages, besides being an outcome of its regional human capital distribution, can be rationalized as inequality of labor market search frictions brought by differences in λ_1. We found a key (indirect) role for search frictions when analysing productivity differences. Since search frictions impact simultaneously on monopsony power as well as on productivity, in order to start to understand regional productivity differences we must deepen our analysis on how these structural parameters are differentiated by regions.

Some extensions of the work become interesting. The analysis can be extended to various demographic groups defined by gender, age, education, etc. Another extension would be a study based on a model that incorporates the presence of formal and informal sectors, which seems to be relevant to the analysis of the Brazilian labor market.

REFERENCES

Albrecht, J. W., & Axell, B. (1984). An equilibrium model of search unemployment. *Journal of Political Economy*, *92*(5), 824-40. doi: 10.1086/261260

Amadeo, E. J. (1994). Bargaining power, mark-up power, and wage differentials in Brazil. *Cambridge Journal of Economics*, *18*(3), 313–322. Retrieved from http://cje.oxfordjournals.org/content/18/3/313.short

Araújo, T. P. d., & Lima, R. A. d. (2010). Aspectos estruturais do mercado de trabalho em contexto recente da economia brasileira: Contraponto Nordeste–Sudeste. In A. Moretto, J. D. Krein, M. Pochmann, & J. Macambira (Eds.), *Economia, desenvolvimento regional e mercado de trabalho do Brasil* (pp. 105–138). Fortaleza: Instituto de Desenvolvimento do Trabalho; Banco do Nordeste do Brasil; Centro de Estudos Sindicais e de Economia do Trabalho.

Arbache, J. S. (1999). Do unions always decrease wage dispersion? The case of Brazilian manufacturing. *Journal of Labor Research*, *20*(3), 425–436. doi: 10.1007/s12122-999-1009-7

Azzoni, C. R., & Servo, L. (2002). Education, cost of living and regional wage inequality in Brazil. *Papers in Regional Science*, *81*(2), 157–175. doi: 10.1111/j.1435-5597.2002.tb01228.x

Barros, R. P. d., Foguel, M. N., & Ulyssea, G. (Eds.). (2006). *Desigualdade de renda no Brasil: Uma análise da queda recente* (Vol. 1). Brasília DF: Ipea. Retrieved from http://repositorio.ipea.gov.br/bitstream/11058/3249/1/Desigualdade%20de%20renda%20no%20Brasil%20-%20v.%201.pdf

Barros, R. P. d., & Mendonça, R. S. P. d. (1995, July). *Os determinantes da desigualdade no Brasil* (Texto para Discussão No. 377). Rio de Janeiro: Instituto de Pesquisa Econômica Aplicada (IPEA). Retrieved from http://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=3490

Behrman, J. R., & Birdsall, N. (1983). The quality of schooling: Quantity alone is misleading. *The American Economic Review*, 928–946.

Bontemps, C., Robin, J.-M., & van den Berg, G. J. (2000). Equilibrium search with continuous productivity dispersion: Theory and nonparametric estimation. *International Economic Review*, *41*(2), 305–358. Retrieved from http://www.jstor.org/stable/2648885

Bunzel, H., Christensen, B. J., Jensen, P., Kiefer, N. M., Korsholm, L., Muus, L., … Rosholm, M. (2001, January). Specification and estimation of equilibrium search models. *Review of Economic Dynamics*, *4*(1), 90–126. doi: 10.1006/redy.1999.0092

Burdett, K., & Mortensen, D. T. (1998). Wage differentials, employer size, and unemployment. *International Economic Review*, *39*(2), 257–273. doi: 10.2307/2527292

Cameron, A. C., & Trivedi, P. K. (2005). *Microeconometrics: Methods and applications*. Cambridge University Press.

Carneiro, F. G. (1998). Productivity effects in Brazilian wage determination. *World Development*, *26*(1), 139–153. doi: 10.1016/S0305-750X(97)10013-4

Carvalho, J. R. (2012). Desigualdades salariais entre Nordeste e Sudeste: Explicações "estruturais" através de um modelo de busca por emprego com dados retrospectivos. *Revista Econômica do Nordeste*, *43*(4), 119–140.

Cavalcante, L. R., & De Negri, F. (2014). Evolução recente dos indicadores de produtividade no Brasil. In F. De Negri & L. R. Cavalcante (Eds.), *Produtividade no Brasil: Desempenho e determinantes* (Vol. I, pp. 143–172). Brasília DF: Ipea; ABDI.

Corseuil, C. H., Gonzaga, G., & Issler, J. V. (1997, Jan). *Desemprego regional no Brasil: Uma abordagem empírica* (Texto para Discussão No. 475). Instituto de Pesquisa Econômica Aplicada (Ipea). Retrieved from http://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=3681

Diamond, P. A. (1971). A model of price adjustment. *Journal of Economic Theory*, *3*(2), 156–168. doi: 10.1016/0022-0531(71)90013-5

Duarte, A. J. M., Ferreira, P. C., & Salvato, M. A. (2004, March). *Regional or educational disparities? A counterfactual exercise* (Ensaios Econômicos No. 532). Fundação Getulio Vargas (FGV). Retrieved from http://bibliotecadigital.fgv.br/dspace/handle/10438/1013

Eckstein, Z., & van den Berg, G. J. (2007). Empirical labor search: A survey. *Journal of Econometrics*, *136*(2), 531–564. doi: 10.1016/j.jeconom.2005.11.006

Eckstein, Z., & Wolpin, K. I. (1990). Estimating a market equilibrium search model from panel data on individuals. *Econometrica*, *58*(4), 783–808. doi: 10.2307/2938350

Feijó, C. A., & Carvalho, P. G. M. d. (2002). Uma interpretação sobre a evolução da produtividade industrial no Brasil nos anos noventa e as "leis" de Kaldor. *Nova Economia*, *12*(2), 57–78. Retrieved from http://revistas.face.ufmg.br/index.php/novaeconomia/article/view/401

Freguglia, R. S., & Menezes-Filho, N. A. (2012). Inter-regional wage differentials with individual heterogeneity: Evidence from Brazil. *The Annals of Regional Science*, *49*(1), 17–34. doi: 10.1007/s00168-010-0432-2

Galeano, E. A. V., & Wanderley, L. A. (2013a). Produtividade industrial do trabalho e intensidade tecnológica nas regiões do Brasil: Uma análise regional e setorial para o período 1996–2007. *Planejamento e Políticas Públicas*, *40*. Retrieved from http://www.ipea.gov.br/ppp/index.php/PPP/article/view/370

Galeano, E. V., & Wanderley, L. A. (2013b). Um estudo sobre o comportamento da produtividade industrial do trabalho nas regiões do Brasil no período de 1996 a 2010. *Geografares*, *15*, 139–180. doi: 10.7147/GEO15.4839

Kaldor, N. (1970). The case for regional policies. *Scottish Journal of Political Economy*, *17*(3), 337–348. doi: 10.1111/j.1467-9485.1970.tb00712.x

Kiefer, N. M., & Neumann, G. R. (1993). Wage dispersion with homogeneity: The empirical equilibrium search model. In H. Bunzel, P. Jensen, & N. Westergård-Nielsen (Eds.), *Panel data and labor market dynamics* (pp. 57–74). New York: North-Holland.

Kyyrä, T. (2007). Estimating equilibrium search models from Finnish data. *Finnish Economic Papers*, *20*(2). Retrieved from http://www.taloustieteellinenyhdistys.fi/images/stories/fep/fep22007_kyyra.pdf

Lancaster, T. (1990). *The econometric analysis of transition data*. Cambridge University Press.

Langoni, C. G. (2005). *Distribuição da renda e desenvolvimento econômico do Brasil*. FGV Editora.

Menezes-Filho, N. A., Chahad, J. P., Zylberstajn, H., & Pazello, E. T. (2008). Trade unions and the economic performance of Brazilian establishments. *Estudos Econômicos*, *38*(1), 55–72. doi: 10.1590/S0101-41612008000100003

Menezes-Filho, N. A., & Picchetti, P. (2000). Os determinantes da duração do desemprego em São Paulo. *Pesquisa e Planejamento Econômico*, *30*(1). Retrieved from http://ppe.ipea.gov.br/index.php/ppe/article/view/175

Mortensen, D. (2003). *Wage dispersion: Why are similar workers paid differently*. MIT Press.

Oliveira, C. W. d. A., & Carneiro, F. G. (2001). Flutuações de longo prazo do emprego no Brasil: Uma análise alternativa de co-integração. *Revista Brasileira de Economia*, *55*(4), 493–512. doi: 10.1590/S0034-

71402001000400003

Oliveira, C. W. d. A., & Cruz, B. d. O. (2000, Feb). *Desigualdades regionais e elasticidade de longo prazo do emprego nos estados do Nordeste com relação ao emprego nacional* (Texto para Discussão No. 704). Brasília, DF: Instituto de Pesquisa Econômica Aplicada (IPEA). Retrieved from http://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=4126

Oliveira, P. F. A. d. (2011). *Estimação estrutural de um modelo de busca por emprego com dispersão de produtividade: Uma anáise para o Brasil* (thesis). Universidade Federal do Ceará–CAEN, Fortaleza CE.

Penido, M., & Machado, A. F. (2002). *Desemprego: Evidências da duração no Brasil metropolitano* (Texto para Discussão No. 176). Belo Horizonte: UFMG/Cedeplar. Retrieved from http://www.cedeplar.ufmg.br/pesquisas/td/TD%20176.pdf

Queiroz, B. L., & Golgher, A. B. (2008). Human capital differentials across municipalities and states in Brazil. *Population Review*, *47*(2). doi: 10.1353/prv.0.0008

Reis, J. G. A., & Barros, R. P. d. (1990). Desigualdade salarial e distribuição de educação: A evolução das diferenças regionais no Brasil. *Pesquisa e Planejamento Econômico*, *20*(3), 415–478.

Rogerson, R., Shimer, R., & Wright, R. (2005, December). Search-theoretic models of the labor market. *Journal of Economic Literature*, *43*, 959-988. doi: 10.1257/002205105775362014

Savedoff, W. D. (1990). Os diferenciais regionais de salários no Brasil: Segmentação versus dinamismo da demanda. *Pesquisa e Planejamento Econômico*, *20*(3), 521–556.

Savedoff, W. D. (1995). *Wages, labour and regional development in Brazil*. Avebury.

Schettini, D., & Azzoni, C. R. (2013). *Recent productivity growth and regional inequality in Brazil* (Texto para Discussão No. 10-2013). São Paulo: USP/Nereus. Retrieved from http://www.usp.br/nereus/?txtdiscussao=recent-productivity-growth-and-regional-inequality-in-brazil

Shimer, R. (2006). On-the-job search and strategic bargaining. *European Economic Review*, *50*(4), 811–830. doi: 10.1016/j.euroecorev.2006.01.008

Sulis, G. (2008). Wage dispersion and equilibrium search models: Some evidence from Italy. *LABOUR*, *22*(4), 593–627. doi: 10.1111/j.1467-9914.2008.00426.x

van den Berg, G. J., & Ridder, G. (1998). An empirical equilibrium search model of the labor market. *Econometrica*, *66*(5), 1183–1222. doi: 10.2307/2999634

van den Berg, G. J., & van Vuuren, A. (2003, July). *The effect of search frictions on wages* (Discussion Paper No. 3979). London: Centre for Economic Policy Research (CEPR).

r b e

ISSN 0034-7140
DOI 10.5935/0034-7140.20160006
Revista Brasileira de Economia, **70**(1), 99–125

Medidas Recentes de Desoneração Tributária no Brasil: Uma Análise de Equilíbrio Geral Computável

Kênia Barreiro de Souza,[*] Débora Freire Cardoso,[†] Edson Paulo Domingues[‡]

Sumário: **1. Introdução; 2. Revisão de Literatura; 3. Modelo e Base de Dados; 4. Simulações; 5. Resultados; 6. Comentários Finais; Apêndice A. Modelo EGC: Base de Dados e Estrutura Teórica; Apêndice B. Alocação de Investimento e Acumulação de Capital; Apêndice C. Análise de Sensibilidade.**

Palavras-chave: **Desoneração, Folha de Pagamentos, Equilíbrio Geral Computável.**

Códigos JEL: **H25, J38, C68, O25.**

O objetivo deste trabalho é avaliar os impactos econômicos da desoneração da folha de pagamentos conforme a política instituída pelo Governo Federal. Para tanto, utiliza-se um modelo de equilíbrio geral computável desenhado para simulações de estática comparativa. Os resultados mostram efeitos positivos para o crescimento de longo prazo da economia, com elevação do emprego, do PIB e do consumo das famílias. Contudo, verifica-se efeito negativo sobre as exportações. Conclui-se que a escolha dos setores abrangidos pela política pode representar um ponto crucial para sua eficácia.

The objective of this study is to assess the economic impacts of release of payroll as proposed by Federal Government policy. For this purpose, it has been used a computable general equilibrium model designed for comparative static simulations. The results show positive effects for the long-term growth of the economy, increasing employment, GDP and household consumption. However, there is a negative effect on exports. We conclude that the choice of the sectors covered by the policy may represent a crucial point for its effectiveness.

[*]Faculdade de Ciências Econômicas da Universidade Federal de Minas Gerais (FACE/UFMG). Av. Antônio Carlos, 6627, Pampulha, Belo Horizonte MG, Brasil. CEP 31270-901. E-mail: keniadesouza@gmail.com

[†]Centro de Desenvolvimento e Planejamento Regional da Universidade Federal de Minas Gerais (Cedeplar/UFMG). Av. Antônio Carlos, 6627, Pampulha, Belo Horizonte MG, Brasil. CEP 31270-901. E-mail: dfreire@cedeplar.ufmg.br

[‡]Centro de Desenvolvimento e Planejamento Regional da Universidade Federal de Minas Gerais (Cedeplar/UFMG). Bolsista de Produtividade em Pesquisa (CNPq) e do Programa Pesquisador Mineiro (Fapemig). E-mail: epdomin@cedeplar.ufmg.br

1. INTRODUÇÃO

A mitigação dos entraves ao crescimento econômico brasileiro por meio de instrumentos de incentivo à competitividade da produção nacional, sobretudo frente à conjuntura econômica do país e aos efeitos remanescentes da crise mundial, tem sido foco de ampla discussão nas esferas política e acadêmica. Entre os principais obstáculos que impedem a maior competitividade são frequentemente citadas as distorções causadas pelo sistema tributário, especialmente quando se trata do excessivo peso dos encargos trabalhistas sobre a folha de pagamentos.

Os encargos sociais que incidem sobre o mercado de trabalho formal elevam o custo do fator trabalho, levando a uma piora nas condições de competitividade das empresas, ao estímulo à informalidade e a baixa cobertura da previdência social (DIEESE, 2008). De acordo com Varsano et al. (1998), as elevadas contribuições sobre a folha de salários criam uma grande cunha entre o custo do trabalhador para as empresas e o salário efetivamente recebido, estimulando as relações trabalhistas informais que, por sua vez, reduzem a própria base de arrecadação. Assim sendo, pode-se dizer que a onerosa tributação sobre o trabalho eleva o custo da mão de obra e reduz as possibilidades de abertura de novos postos de trabalho, contribuindo para manter o que se chama de "custo Brasil" em patamares mais elevados, o que coloca os produtos nacionais em uma posição desfavorável frente ao comércio internacional.

Em meio a esta discussão que envolve acadêmicos, a classe política, as entidades patronais e os sindicatos, o Governo Federal lançou a proposta de desoneração da folha de pagamentos de alguns setores como parte da política industrial, tecnológica e de comércio exterior do Governo Dilma Rousseff.[1] A política elimina a atual contribuição previdenciária sobre a folha (20%) e adota uma nova contribuição sobre a receita bruta das empresas para setores selecionados.

Neste contexto, o presente estudo busca revisitar a questão dos encargos sociais e avaliar os impactos econômicos de uma desoneração da folha de pagamentos. Torna-se imprescindível que se avaliem os efeitos desta ação, no intuito de apontar seus impactos sobre o emprego, PIB, consumo e fluxo comercial, de modo a ponderar se a política realmente atenua as distorções causadas pelo peso dos encargos trabalhistas sobre a competitividade nacional. Ademais, tendo em vista que a política é direcionada a atividades específicas, é preciso conhecer seus impactos sobre os setores beneficiados, bem como sobre o restante da economia, considerando que os setores e mercados são interligados e o efeito entre eles é sistêmico.

Com vistas a acessar esses resultados, adota-se um modelo de equilíbrio geral aplicado, que se mostra uma ferramenta adequada ao tipo de análise pretendida, ao tomar explicitamente a relação entre os diferentes setores e mercados. O componente inovador deste estudo se encontra na análise de uma desoneração da folha de pagamentos nos moldes da política governamental, que ainda não foi investigada em outros trabalhos.

O artigo está organizado em mais cinco seções, além desta: i) a próxima seção apresenta uma revisão de literatura e as medidas de desoneração da folha de pagamentos, realizando ainda uma breve discussão acerca dos trabalhos sobre modificações tributárias que utilizam modelos de equilíbrio geral como ferramenta analítica; ii) a seção 3 apresenta o modelo e a base de dados; iii) a seção 4 discute as simulações; iv) a seção 5 os resultados obtidos; e v) a seção 6 tece as considerações finais.

2. REVISÃO DE LITERATURA

A discussão em torno dos encargos sociais é bastante polêmica, seja quanto à mensuração de seu peso sobre a folha, ou quanto a seus impactos sobre a economia. Autores como **?** e grande parte do empresariado defendem que os encargos sociais no Brasil equivalem a mais de 100% dos salários pagos. Segundo esse conceito, o empregador despende mais que o dobro do salário do empregado em encargos sociais. Uma segunda interpretação, adotada pelo DIEESE (2011) e por pesquisadores da Unicamp, conclui que o peso dos encargos sociais é de 27,8% sobre a remuneração total do trabalhador (DIEESE, 2011).

[1] Lei nº 12.546 de 14 de dezembro de 2011 e Medida Provisória n° 563 de 2012.

A discrepância entre as duas concepções se deve ao conceito de salário e encargos sociais adotados. O primeiro grupo considera uma noção mais restrita de salário: apenas a remuneração pelo tempo efetivamente trabalhado, ou seja, não inclui, por exemplo, o décimo terceiro salário, férias remuneradas, aviso prévio e despesas de rescisão contratual. Na concepção desses autores, todos esses itens são considerados como encargos sociais, juntamente com as obrigações recolhidas ao INSS para custeio da Previdência Social, salário educação, Sistema S,[2] entre outros. Já o segundo grupo, considera como salário a remuneração total recebida integral e diretamente pelo trabalhador, que inclui, além do salário recebido mensalmente (inclusive nas férias), os ganhos de décimo terceiro e 1/3 de férias e o FGTS e outras verbas rescisórias. De acordo com esse conceito, a diferença entre o montante que a empresa desembolsa e a remuneração do trabalhador constitui os encargos sociais que são recolhidos ao governo, sendo alguns deles repassados a entidades patronais de assistência e formação profissional (DIEESE, 2011).

Seguindo a interpretação adotada pelo DIEESE, a Tabela 1 mostra o conjunto de encargos sociais que incidem sobre a folha de pagamentos. Como se observa, trata-se das contribuições sociais pagas pelas empresas como parte do custo total do trabalho, que não são revertidas em benefício direto e integral ao trabalhador. O FGTS, por exemplo, não está incluído, já que configura um patrimônio individual do trabalhador.

Quanto aos impactos econômicos dos encargos que incidem sobre a folha, também não existe um consenso. Conforme ressalta o DIEESE (2011), existem os que defendem que fatores inibidores do crescimento do emprego estariam muito mais relacionados às condições macroeconômicas que dificultam o investimento e a demanda interna, como altas taxas de juros, arrocho monetário, arrocho fiscal, ausência de políticas setoriais consistentes e ambiente de incerteza econômica, do que com o peso dos encargos sociais. Deste modo, medidas que estimulem a criação de um ambiente propício ao investimento produtivo seriam mais eficazes do que a redução ou eliminação de encargos sociais.

Por outro lado, autores como Amadeo (1992), Pastore (1994a, 1994b) e Chahad & Fernandes (2002) concluem que a redução desses encargos geraria impacto positivo sobre a criação de empregos, formalização, e, portanto, sobre a competitividade da economia brasileira. Entretanto Chahad & Fernandes (2002) ressaltam que as políticas que visem aumentar a atratividade do emprego formal serão mais eficazes se seu foco se estabelecer sobre os encargos de caráter tributário, de modo que os benefícios diretos ao trabalhador não sejam lesados.

Mais recentemente, estudos de abordagem quantitativa têm sido realizados no intuito de mensurar os efeitos econômicos da desoneração dos encargos trabalhistas, como é o caso de Paes (2011), Domingues et al. (2012) e Bitencourt & Teixeira (2008). Paes (2011) simula uma reforma tributária mais ampla

Tabela 1. *Alíquotas de encargos sociais incidentes sobre a folha de pagamentos média mensal das empresas.*

Tipo de encargo	Porcentagem sobre a folha média mensal
INSS	20,00
Seguro contra acidentes do trabalho (média)	2,00
Salário-Educação	2,50
Incra	0,20
Sesi ou Sesc ou Sest	1,50
Senai ou Senac ou Senat	1,00
Sebrae	0,60
Total	27,8

Notas: Guia do INSS. Elaboração: DIEESE (2011).

[2]Sistema S: Senai, Senac, Sesi, Sesc, Senat, Senar, Sebrai.

(nos moldes da Proposta de Emenda à Constituição PEC-233/2008), com modificações de alguns tributos, em que a desoneração da folha de pagamentos está inclusa. Já Domingues et al. (2012), simula uma desoneração de 50,0% sobre os encargos da folha para a indústria, considerando um ajuste por parte do consumo do governo. Em ambos, observam-se efeitos positivos sobre o PIB, emprego, produção, consumo e investimentos.

No entanto, para Bitencourt & Teixeira (2008), apenas a desoneração da folha não seria suficiente para garantir a expansão do emprego e o ganho de competitividade, sendo necessário implementar políticas complementares, à medida que, a necessidade de reajuste dos gastos do governo diante da depreciação de sua receita tornaria os resultados econômicos piores. Seguindo uma linha semelhante, para o DIEESE (2011), ainda, não é razoável que se reduzam as alíquotas da Previdência Social ou de programas educacionais sem que se definam fontes alternativas para o seu financiamento, uma vez que é imprescindível preservar a receita da Previdência.

2.1. Medidas recentes de desoneração da folha de pagamentos

Em meio a essas discussões, o Governo Federal lançou em 2001 uma série de medidas que estabelecem uma nova política industrial, tecnológica e de comércio exterior para o período de 2011 a 2014, visando fortalecer a indústria nacional que está sendo prejudicada pela valorização do real, decorrentes das incertezas do cenário internacional. Tais mudanças aprimoram os avanços obtidos com a Política industrial, Tecnológica e de Comércio Exterior (PITCE 2003–2007) e com a Política de Desenvolvimento Produtivo (PDP 2008–2010), tendo como objetivos centrais acelerar o crescimento do investimento produtivo e de inovação tecnológica das empresas nacionais, e aumentar a competitividade dos bens e serviços nacionais.

As medidas preveem uma série de ações visando desonerações tributárias, financiamento ao investimento e a inovação e desoneração das exportações. Entre suas principais diretrizes está a desoneração da folha de pagamentos,[3] que é consolidada por três ações: i) elimina-se a atual contribuição previdenciária sobre a folha (20%); ii) adota-se uma nova contribuição previdenciária sobre a receita bruta das empresas, cuja alíquota varia de 1 a 2% do total do faturamento, dependendo do setor; e iii) impõe-se um acréscimo igual à alíquota sobre a receita bruta na Cofins sobre as importações, de modo a reduzir as assimetrias na tributação entre o produto nacional e o importado.

Os setores selecionados para a desoneração da folha de pagamentos são:[4] i) Têxtil; ii) Confecções; iii) Couro e calçados; iv) Plásticos; v) Material elétrico; vi) Bens de capital – mecânico; vii) Ônibus; viii) Autopeças; iv) Naval; x) Aéreo; xi) Móveis; xii) TI & TIC;[5] xiii) Hotéis; xiv) Call Centers; e xv) Design Houses (chips). Para os primeiros onze setores listados a alíquota sobre faturamento é de 1%, enquanto nos quatro últimos situa-se em 2%.

De acordo com o Ministério da Fazenda 2012, essa mudança de base da contribuição contempla ainda uma redução da carga tributária dos setores beneficiados, porque a alíquota sobre a receita bruta foi fixada em um patamar inferior àquela alíquota que manteria inalterada a arrecadação (alíquota neutra). A mudança na base de contribuição ocorre para as empresas que se enquadram nas atividades econômicas ou que fabricam produtos industriais listados na Medida. Nestes casos, a empresa obrigatoriamente passa a pagar sua contribuição previdenciária sobre a receita bruta oriunda da venda daqueles produtos.

Com a desoneração da folha o Governo procura aumentar a competitividade da indústria nacional por meio da redução dos custos laborais, estimular as exportações e ampliar a formalização no mercado de trabalho, uma vez que a contribuição previdenciária dependerá da receita e não mais da folha de salários. Objetiva-se ainda, tornar o produto nacional mais competitivo internamente frente ao importado, dada

[3] Para maiores informações consultar Brasil (2012).

[4] Até o momento de execução deste estudo, estes foram os setores selecionados pela medida, contudo, o Governo Federal pretende ampliar o número de setores abrangidos.

[5] Tecnologia da informação (TI) e Tecnologia da informação e da comunicação (TIC).

o aumento da Cofins-Importação para os produtos elegíveis pela política (Brasil, 2012).

2.2. Estudos sobre modificações tributárias

Vários estudos sobre modificações tributárias no Brasil e seus impactos na economia têm sido produzidos. Os modelos de equilíbrio geral são as ferramentas mais empregadas para essa análise, dadas as vantagens de considerarem diversos setores e mercados, e da relação entre eles ser tomada explicitamente. Como modificações de tributos afetam diversos setores de forma diferenciada, e o efeito entre estes é sistêmico, análises em equilíbrio parcial são provavelmente insuficientes no estudo de questões tributárias amplas. Nessa linha metodológica de equilíbrio geral computável (EGC), estudos recentes sobre modificações de tributos na economia brasileira podem ser vistos em Domingues & Haddad (2003), Silva, Tourinho & Alves (2004), Salami & Fochezatto (2009), Ferreira Filho, Santos & Lima (2010), Paes & Bugarin (2006), Paes (2009), Tourinho, Alves & Silva (2010), Cavalcanti & Silva (2010), Cury & Coelho (2010) e Paes (2011). Dos trabalhos citados, alguns estão mais próximos do tema deste artigo e são comentados brevemente.

Tourinho et al. (2010) analisaram três medidas fiscais por meio de um modelo de EGC para medir os impactos de alterações na estrutura tributária nacional. As alterações consideradas nesse trabalho foram a transformação parcial da Cofins em uma contribuição sobre o valor adicionado, a adoção do PIS/Pasep e da Cofins sobre as importações, e a extinção da CPMF que foi adotada em 2007. Seus resultados mostram que para a transformação da Cofins em uma contribuição sobre o valor adicionado os efeitos quase anulam o ganho inicial de arrecadação. Para a adoção do PIS/Pasep e da Cofins sobre as importações os efeitos de equilíbrio geral reduzem o ganho pela metade, e no caso do CPMF eles compensam metade da perda de arrecadação. Por fim, a extinção da CPMF beneficia os trabalhadores de baixa qualificação não formais, cujos salários aumentam em 5,5%. E em geral, a renda das famílias aumenta em cerca de 2%, indistintamente ao longo das classes.

Cury & Coelho (2010) avaliam os impactos econômicos da reforma tributária do PIS/PASEP e do COFINS, que passaram a ser arrecadados por dois regimes (cumulativo e não cumulativo) nos fluxos domésticos e impostos sobre as importações. A avaliação foi realizada por meio de um modelo de equilíbrio geral computável (ano base 2003) adaptado para as novas características do sistema tributário. Mais recentemente, Paes (2011) desenvolveu um modelo de equilíbrio geral (calibrado para o ano de 2005) para verificar os possíveis efeitos macroeconômicos e setoriais da nova reforma tributária — PEC 233/2008 — que unifica alguns tributos federais do consumo no IVA-F (Imposto sobre o Valor Agregado – Federal).[6] Os principais resultados sugerem que pelo lado agregado ocorrerá uma modesta expansão do produto, consumo, emprego e investimento, com pequena perda de arrecadação e aumento de bem-estar. Em relação ao produto setorial, observa-se uma tendência de aumento dos setores industrial e agropecuário em detrimento dos serviços.

3. MODELO E BASE DE DADOS

O modelo de equilíbrio geral computável (EGC) utilizado neste trabalho está calibrado para o ano de 2005,[7] utilizando as informações mais atuais das matrizes de insumo-produto disponibilizadas pelo

[6] A proposta da PEC 233/2008 é unificar o PIS, a COFINS e a CIDE que incidem sobre a receita das empresas e serão então, cobrados de forma unificada.

[7] Vale ressaltar que a utilização da matriz de 2005 decorre da inexistência de dados oficiais disponíveis para o ano de 2001 (ano da política). Essa limitação nos dados pode afetar os resultados obtidos de duas formas: i) pelo efeito escala, pois a cada ano os valores base aumentam para toda a economia e, consequente, o choque é aplicado sobre uma base menor; ii) pelo efeito de mudanças na estrutura produtiva, que pode alterar a composição dos resultados em termos setoriais e na composição dos agregados macroeconômicos. No que tange o primeiro efeito, assumindo um crescimento balanceado da economia brasileira entre 2005 e 2011, o uso de variações percentuais para os resultados minimiza o efeito escala, muito embora os valores absolutos não possam ser diretamente comparados aos valores correntes de 2011. Por sua vez, o segundo efeito decorre de mudanças estruturais, que tendem a ser lentas, portanto, espera-se que não afetem significativamente os resultados.

IBGE. Para este trabalho, o modelo foi configurado para 60 setores, 117 produtos, cinco componentes da demanda final (consumo das famílias, consumo do governo, investimento, exportações e estoques), dois elementos de fatores primários (capital e trabalho) e dois setores de margens (comércio e transportes). Algumas características da base de dados do modelo são importantes para sua utilização no tema de modificações de tributos. Primeiramente, cinco tipos de tributos (IPI, ICMS, COFINS, INSS e FGTS) estão especificados individualmente, incidindo sobre fluxos de compras (IPI, ICMS), remuneração do trabalho (INSS e FGTS) e produção (COFINS). Esta especificação da base de dados e da estrutura teórica do modelo permite projetar políticas de desoneração de impostos por tributo e setor. A simulação empregada neste trabalho, conforme será apresentado mais adiante, está desenhada para exercícios de estática comparativa.

A especificação teórica do modelo se baseia na tradição australiana em modelos EGC, principalmente nos modelos ORANI (Dixon, Parmenter, Sutton & Vincent, 1982) e MONASH (Dixon & Rimmer, 2002). Em diversos aspectos, a especificação teórica segue o padrão em modelos EGC. Os setores produtivos minimizam custos de produção sujeitos a uma tecnologia de retornos constantes de escala, em que a combinação de insumos intermediários e fator primário (agregado) é determinada por coeficientes fixos (Leontief). Na composição dos insumos há substituição via preços entre produto doméstico e importado, por meio de funções de elasticidade de substituição constante (CES). Na composição dos fatores primários também há substituição via preço entre capital e trabalho por funções CES. A demanda é especificada a partir de funções de utilidade não-homotéticas Stone-Geary (Peter, Horridge, Meagher, Naqvi & Parmenter, 1996). Essa especificação divide o consumo dos bens e serviços em parcelas de "luxo" e "subsistência", reservando uma parcela fixa do gasto em subsistência e uma parcela residual em "gasto de luxo", o que permite que modificações na renda causem modificações diferenciadas no consumo dos produtos, daí seu caráter não-homotético. Na composição do consumo do produto entre doméstico e importado, utilizam-se funções de elasticidade de substituição constante (CES).

As exportações setoriais respondem a curvas de demanda negativamente associadas aos custos domésticos de produção e positivamente afetadas pela expansão exógena da renda internacional, adotando-se a hipótese de país pequeno no comércio internacional. O consumo do governo é tipicamente exógeno, podendo estar associado ou não ao consumo das famílias ou à arrecadação de impostos. Os estoques se acumulam de acordo com a variação da produção.

O investimento e o estoque de capital seguem mecanismos de deslocamento intersetorial e de acumulação a partir de regras pré-estabelecidas, associadas a taxas esperadas de retorno e de depreciação do estoque de capital.[8] Assim, setores com elevação na taxa esperada de retorno, calculada endogenamente, atraem investimento. O fator trabalho também segue mecanismos de deslocamento intersetorial a partir do diferencial de rendimentos. O Apêndice A apresenta a estrutura teórica do modelo.

Alguns indicadores da base de dados do modelo são importantes para a análise da incidência tributária da economia e para as simulações que serão feitas com o modelo EGC. A Tabela 2 resume a estrutura tributária presente no modelo. A arrecadação de impostos representa 33% do PIB, cerca de 700 milhões de reais. Os tributos indiretos correspondem a 12,80% do PIB; e 8,32% do PIB são tributos sobre a folha de salários. O restante (9,03% do PIB) são outros tributos e subsídios à produção.

Como o foco deste trabalho são os impostos sobre a folha de salários e a mudança da base de recolhimento da contribuição previdenciária para a receita bruta em alguns setores, cabe detalhar este aspecto da base de dados do modelo. A Tabela 3 mostra a distribuição dos tributos que incidem sobre a folha de salários (INSS e FGTS), os tributos diretos sobre a produção (COFINS e Outros) e a tributação sobre as importações entre os setores da economia brasileira para o ano de 2005. Compete ressaltar que, a fim de simplificar a exposição, reportam-se os dados agregados para os setores Agropecuária, Indústria extrativa, Indústria e Serviços, desagregando os dados apenas para os setores alvo da política.

A incidência de tributos sobre a folha de salários da economia brasileira em 2005 foi de R$ 178,6 bilhões. O setor de serviços deteve a maior participação na arrecadação total sobre a folha de salários, e

[8] O Apêndice B detalha da regra de alocação do investimento e deslocamento intersetorial do capital.

Tabela 2. Resumo da estrutura tributária na base de dados do modelo EGC 2005.

Tributos	Incidência	Total (R$ milhões)	% PIB
Indiretos (IPI, ICMS e Outras taxas)	Consumo intermediário	107.908	5,03
	Investimentos	23.816	1,11
	Consumo das famílias	164.346	7,66
Diretos	Cofins e outros	28.067	1,31
Importações	Importações	9.172	0,43
Trabalhistas	INSS e FGTS	178.564	8,32
Outros tributos e subsídios	Outros custos de produção	193.728	9,03
Total		705.601	32,88

Tabela 3. Tributação sobre salários, produção e importação (2005)

	Folha de salários		**Produção**		**Importação**	
	R$ 2,1 milhões	Taxa sobre salários	R$ 2,1 milhões	Taxa sobre produção	R$ 2,1 milhões	Taxa sobre importação
Agropecuária	9.174	20,3	1.131	0,6	70	1,5
Indústria extrativa	3.331	32,3	619	0,6	9	0,0
Indústria	38.022	24,0	10.822	0,8	9.093	5,6
Serviços	128.462	20,0	15.495	0,7	–	–
Setores alvo da política						
Serviços de alojamento e alimentação	1.434	11,7	297	0,4	–	–
Serviços de Informação	3.651	16,2	2.545	1,8	–	–
Serviços Prestados à empresas	7.581	17,0	986	0,7	–	–
Móveis e Indústrias diversas	900	16,3	238	0,7	218	17,2
Outros equipamentos de transporte	962	24,4	235	1,0	61	1,0
Peças e aces. para veíc. automotores	2.258	24,3	558	1,0	573	6,7
Caminhões e Ônibus	610	30,3	159	0,8	18	2,9
Ap. médico-hosp., medida e óptico	416	22,5	90	0,8	436	5,2
Máquinas e Equipamentos	2.759	23,3	642	1,1	1.430	7,6
Máquinas, ap. e materiais elétricos	1.263	24,0	289	0,9	626	9,0
Mat. eletrônico e equip. de comunic.	985	25,6	296	0,8	731	3,9
Artigos de Borracha e Plástico	1.906	23,5	474	1,0	242	9,6
Artefatos de couro e calçados	1.050	19,9	249	1,1	133	17,1
Artigos do vestuário e acessórios	849	14,6	233	0,8	120	20,7
Têxteis	1.169	22,2	273	0,8	297	13,0

Notas: Elaboração própria a partir da base de dados do modelo.

a indústria representou 24,0% da arrecadação desses impostos. A taxa média de incidência do imposto sobre a folha é em torno de 22,0%, mas variaram significativamente entre os setores. Por exemplo, na indústria extrativa, a taxa implícita chegou a 32,3% da folha de salários, enquanto no setor de serviços e na agropecuária esta taxa fica em torno de 20,0%. Ao se considerar os setores que serão desonerados pela política, a taxa sobre a folha no setor de Caminhões e ônibus chega a 30,3%, enquanto em Serviços de alojamento e alimentação é de 11,7%. Estas diferenças se explicam pela diferente composição da mão de obra em cada setor e pelas regras de tributação do FGTS.[9]

No que se refere aos tributos diretos sobre a produção (COFINS e Outros), a arrecadação em 2005 foi de R$ 28,1 bilhões, sendo que a agropecuária e o setor extrativo participaram com, respectivamente, 4,02% e 2,20% deste montante, cabendo aos setores industriais e de serviços as participações mais representativas de 38,56% e 55,21% da arrecadação total. Comparando este imposto com a sua base de incidência (produção) a alíquota média é de cerca de 0,7%, com alguma variação entre os setores alvo da política (uma taxa implícita de 1,8% em Serviços da Informação e 0,4% em Serviços de alojamento e alimentação).

A Receita de tributos sobre a importação foi de R$ 9,2 bilhões em 2005, sendo que o setor industrial deteve a maior participação na arrecadação desses impostos. Na base de dados do modelo não há incidência de imposto de importação sobre o setor de serviços. Entre os setores elegíveis pela política, há bastante variabilidade quanto à taxa desse imposto em relação à sua base, o valor importado. O setor Artigos de Vestuário e acessórios, por exemplo, apresenta uma taxa de 20,7%, enquanto que para Outros equipamentos de transporte esta taxa é de 1,0%.

4. SIMULAÇÕES

O objetivo deste estudo consiste em aferir os impactos econômicos potenciais da desoneração da folha de pagamentos, nos moldes da política proposta pelo Governo Federal. Desse modo, são realizadas três simulações. As duas primeiras implementam as modificações tributárias exatamente nos setores selecionados pelo plano Brasil Maior, em uma perspectiva de curto prazo (Simulação 1) e longo prazo (Simulação 2). Por sua vez, na terceira simulação todos os setores são desonerados em um horizonte de longo prazo (Simulação 3), com a finalidade de verificar qual seria o desenho de política com os maiores benefícios para a economia brasileira como um todo.

A operacionalização de modelos EGC exige que sejam estabelecidas hipóteses de simulação, ou o conjunto de variáveis exógenas e endógenas, definindo o chamado fechamento do modelo, e permitindo choques exógenos específicos, para determinadas variáveis, que desencadeiam alterações nas variáveis endógenas, por meio de mudanças nos preços relativos e quantidades reais. A partir do fechamento são definidas as hipóteses de operacionalização do modelo, associadas ao horizonte de tempo hipotético das simulações, ou seja, ao período de tempo necessário para que a economia se mova rumo ao novo equilíbrio. Neste trabalho, foram utilizados dois fechamentos diferentes, de curto e de longo prazo para os quais foram feitas as hipóteses descritas no Quadro 1.

Desta forma, pode-se dizer que os resultados representam as alterações no comportamento econômico do país a partir de mudanças exógenas, no caso deste trabalho, desencadeadas por meio das variações na tributação dos setores selecionados pelo programa Brasil Maior.

Assim sendo, para a implementação das simulações 1 e 2 foram necessárias a identificação dos setores-alvo pela política no conjunto de setores do modelo BRIDGE. Para tanto, foram utilizadas as descrições e setores especificados na Lei nº 12546, de 2001, e Medida Provisória nº 563, de 2012, os quais foram compatibilizados com os setores do modelo, conforme pode ser observado no Quadro 2. Conforme as determinações da medida governamental a simulação é composta basicamente de três choques sobre

[9] Como a alíquota do FGTS é fixa em 8% sobre os salários, essas diferenças se explicam por variações na alíquota do INSS por faixa de rendimento do trabalhador e, portanto, dependem da composição da mão de obra em cada setor, dado que a incidência do INSS é reduzida para salários menores.

Quadro 1. Hipóteses de fechamento do modelo.

	Curto Prazo	**Longo prazo**
Fator capital	É fixo no curto prazo, com a taxa de retorno do capital endógena e variando entre os setores	É variável entre os setores, e sua taxa de retorno é fixa (exógena)
Fator trabalho	O salário real é exógeno, sendo o emprego e o salário nominal endógenos	
Consumo real das famílias	Exógeno e não recebe choque	Endógeno e segue as as variações na renda real
Consumo do governo	Exógeno e não recebe choque	Endógeno e segue o consumo das famílias
Investimento	O investimento agregado é exógeno e não recebe choque	O investimento agregado é endógeno e responde às variações na taxa de retorno do capital setorial
Setor Externo	Saldo comercial externo como proporção do PIB é endógeno*. O preço em moeda estrangeira das importações é fixo e o das exportações depende dos ajustes de oferta e demanda no mercado interno.	

*A hipótese de endogeneidade desta variável significa que não há restrição sobre a balança comercial, ou seja, déficits ou superávits contínuos podem ocorrer mesmo no longo prazo.

Quadro 2. Correspondência entre os setores do modelo BRIDGE e os setores beneficiados pelo programa Brasil Maior.

Brasil Maior	**Modelo BRIDGE**
Hotéis	Serviços de Alojamento e Alimentação*
TI & TIC	Serviços de Informação*
Call Center	Serviços Prestados às Empresas*
Móveis	Móveis e indústrias diversas*
Navios	Outros equipamentos de transporte*
Aéreo	Outros equipamentos de transporte*
Autopeças	Peças e acessórios para veículos automotores
Ônibus	Caminhões e Ônibus
Bens de Capital	Aparelhos/instrumentos médico-hospitalar, medida e óptico
Bens de Capital	Máquinas e Equipamentos
Material Elétrico	Máquinas, aparelhos e materiais elétricos
Design House	Material eletrônico e equipamentos de comunicações
Plástico	Artigos de Borracha e Plástico*
Calçados e Couro	Artefatos de couro e calçados
Vestuário	Artigos do vestuário e acessórios
Têxteis	Têxteis

*Choque parcial, segundo a participação da atividade descrita no Brasil Maior no setor correspondente do modelo BRIDGE.

cada um dos setores selecionados: i) isenção de INSS; ii) criação de um imposto sobre a receita bruta; e iii) criação de um imposto sobre importações.[10]

Na Simulação 3, o objetivo é aplicar os mesmos choques sobre todos os setores econômicos, no entanto, considerando os objetivos da política de fortalecimento do setor industrial, foram incluídas as simulações de desonerações para todos os setores industrias, da agropecuária e serviços relacionados. Desta forma, não foram incluídos setores de serviços prestados às famílias, e os setores públicos.

Conforme proposto por Domingues & Haddad (2003), para avaliar alterações de tributos sobre a receita do governo é importante separar quais são os efeitos de primeira ordem (de equilíbrio parcial) de efeito total de primeira e segunda ordem (ou de equilíbrio geral). Assim, sendo BAS a base tributária, TAX a receita dos tributos diretos e t a alíquota tributária, temos que $\mathrm{TAX} = t * \mathrm{BAS}$. Logo, uma variação na tributação implica em dois efeitos:

$$\Delta \mathrm{TAX} = \underbrace{BAS * \Delta t}_{\substack{\text{efeitos de} \\ \text{primeira} \\ \text{ordem}}} + \underbrace{BAS * t}_{\substack{\text{efeitos de} \\ \text{segunda} \\ \text{ordem}}}$$

A Tabela 4 apresenta o efeito de primeira ordem da política de desoneração sobre cada um dos setores selecionados pelo programa. Já na Tabela 5, calcula-se esse efeito para a extensão da modificação tributária em todos os setores (Simulação 3). Em ambas, os valores foram calculados a partir dos dados do modelo BRIDGE, e desagregados por meio da Pesquisa Industrial Anual (PIA) e Pesquisa Anual dos Serviços (PAS), ambas disponibilizadas pelo IBGE (2012).

Conforme a Tabela 4, a variação estimada nos cofres públicos mediante a modificação tributária nos setores selecionados pelo programa é uma redução de cerca de R$ 5,602 milhões na arrecadação de

Tabela 4. *Variação estimada na arrecadação de impostos decorrente da modificação tributária nos setores alvo da política, com base nos dados do modelo BRIDGE (em milhões de reais).*

Setores	Isenção de INSS	Imposto sobre a receita bruta	Imposto de importação	Efeito de primeira ordem na arrecadação pública
Hotéis	−285,38	320,63	161,32	196,57
TI & TIC	−1.360,90	609,57	107,86	−643,46
Call Center	−1.571,36	1.117,15	226,14	−228,07
Móveis	−385,15	246,54	14,81	−123,80
Naval e Aéreo	−494,22	135,18	63,43	−295,62
Autopeças	−1.675,99	558,12	91,65	−1.026,22
Ônibus	−435,52	196,94	6,22	−232,36
Bens de Capital	−2.324,06	711,78	291,65	−1.320,63
Material Elétrico	−940,14	304,87	76,12	−559,15
Design House	−727,11	732,78	387,74	393,41
Plástico	−937,10	332,93	27,57	−576,60
Calçados e Couro	−721,01	235,17	9,12	−476,72
Vestuário	−508,85	280,20	7,00	−221,65
Têxteis	−851,59	337,97	25,79	−487,83
Total	−13.218,38	6.119,82	1.496,42	−5.602,14

[10] O imposto sobre a receita bruta e sobre as importações é de 2% para TI & TIC, hotéis, Call Center, e Design House; e de 1% para os demais setores.

Tabela 5. Variação estimada na arrecadação de impostos decorrente da extensão da modificação tributária para todos os setores, com base nos dados do modelo BRIDGE (em milhões de reais).

Descrição	Isenção de INSS	Imposto de 1% sobre a receita bruta	Imposto de 1% sobre importações	efeito de primeira ordem
Serviços prestados às empresas	−5.556,30	1.396,13	0,00	−4.160,17
Transporte, armazenagem e correio	−4.969,45	1.808,98	0,00	−3.160,47
Construção	−4.138,49	1.676,72	0,00	−2.461,77
Agric., silvicultura, exploração florestal	−3.594,31	1.226,81	44,80	−2.322,70
Pecuária e pesca	−2.899,23	717,96	1,83	−2.179,44
Alimentos e Bebidas	−4.643,69	2.572,96	54,05	−2.016,68
Serviços de informação	−2.682,17	1.402,69	0,00	−1.279,48
Máq. e equipamentos, inc. manut. e reparos	−2.018,34	605,68	203,02	−1.209,64
Peças e acessórios para veículos automotores	−1.668,96	558,12	91,65	−1.019,19
Petróleo e gás natural	−1.892,65	711,96	207,42	−973,27
Eletric. e gás, água, esgoto e limp. urbana	−2.218,30	1.326,35	0,00	−891,95
Artigos de borracha e plástico	−1.392,48	481,86	55,18	−855,44
Prod. de metal – ex. máquinas e equipamentos	−1.280,33	528,60	35,59	−716,14
Fabricação de aço e derivados	−1.376,27	734,82	33,77	−607,68
Celulose e produtos de papel	−992,93	375,06	23,41	−594,46
Máquinas, aparelhos e materiais elétricos	−940,14	304,87	76,12	−559,15
Jornais, revistas, discos	−855,52	293,57	3,78	−558,17
Outros produtos de minerais não-metálicos	−801,96	255,24	16,42	−530,30
Têxteis	−851,59	337,97	25,79	−487,83
Artefatos de couro e calçados	−721,01	235,17	9,12	−476,72
Outros equipamentos de transporte	−720,80	245,59	63,43	−411,78
Produtos químicos	−1.065,25	610,41	162,51	−292,33
Móveis e produtos das indústrias diversas	−594,79	329,19	14,94	−250,66
Caminhões e ônibus	−435,52	196,94	6,22	−232,36
Artigos do vestuário e acessórios	−508,85	280,20	7,00	−221,65
Metalurgia de metais não-ferrosos	−513,55	252,64	47,50	−213,41
Produtos de madeira – exclusive móveis	−405,90	196,20	2,33	−207,37
Produtos e preparados químicos diversos	−365,52	130,20	32,59	−202,73
Tintas, vernizes, esmaltes e lacas	−260,51	84,18	5,81	−170,52
Material eletrônico e equip. de comunicações	−727,11	366,39	193,87	−166,85
Perfumaria, higiene e limpeza	−376,56	175,21	37,64	−163,71
Eletrodomésticos	−256,09	95,60	4,93	−155,56
Ap./instrumentos médico-hosp, medida e óptico	−305,73	106,10	88,63	−111,00
Outros da indústria extrativa	−307,58	121,42	88,14	−98,02
Minério de ferro	−328,50	235,72	0,00	−92,78
Defensivos agrícolas	−225,68	117,19	22,36	−86,13
Álcool	−202,78	123,08	0,01	−79,69
Produtos do fumo	−191,08	95,20	20,11	−75,77
Fabricação de resina e elastômeros	−380,14	240,01	71,56	−68,57
Cimento	−123,82	66,75	0,60	−56,47
Máq. para escritório e equip. de informática	−211,04	118,74	62,68	−29,62
Serviços de manutenção e reparação	−143,30	254,54	0,00	111,24
Automóveis, camionetas e utilitários	−148,05	530,86	46,23	429,04
Refino de petróleo e coque	−823,12	1.227,07	111,16	515,11
Total	−55.115,39	23.750,95	1.972,20	−29.392,24

Nota: Elaboração própria a partir da base de dados do modelo BRIDGE.

impostos (aumento da arrecadação sobre a receita bruta, mais o aumento dos impostos sobre importação, menos o INSS dos setores selecionados).

Com a extensão da modificação tributária para todos os setores (Simulação 3), a redução estimada nos cofres públicos fica em torno de R$ 29,392, conforme reporta a Tabela 5. Os setores estão ordenados segundo as maiores perdas de arrecadação, permitindo observar que as atividades Serviços prestados às empresas; Transporte, armazenagem e correio e Construção seriam as maiores responsáveis pela redução total nos cofres públicos.

Não obstante, o efeito de primeira ordem da política represente uma redução na arrecadação de impostos para a maioria dos setores selecionados, o efeito total de primeira e segunda ordem, depende, ainda, dos efeitos multiplicadores desses incentivos. Desta forma, seguindo o arcabouço de equilíbrio geral, as alterações nos impostos são um choque exógeno que provocará alterações a partir de um equilíbrio inicial, fazendo com que a economia se adapte e atinja um novo equilíbrio, que incorpora os efeitos desencadeados pela política pública. A seção de resultados a seguir analisa esses impactos.

5. RESULTADOS

5.1. Simulações de política: setores selecionados pelo Programa

Para os setores selecionados pelo governo foram analisados os impactos da política sobre a economia brasileira e setores específicos no curto e longo prazo. A Tabela 6 permite observar que, no curto prazo, com investimento e consumo fixos, a impossibilidade de ajuste do capital e a manutenção do consumo das famílias, fazem com que o efeito negativo sobre a receita bruta supere o efeito positivo do corte no INSS. Deste modo, a política exerceria impactos negativos no emprego total (−0,11%), no PIB (−0,06%), Exportações (−0,40%) e Importações (−0,05%).

Tabela 6. *Resultados agregados das simulações de desoneração da folha de pagamentos para setores selecionados pelo governo no curto prazo.*

Variáveis	Variação % Total	Contribuição por tipo de choque		
		Corte do INSS	Imposto sobre a receita bruta	Impostos sobre importações
Emprego	−0,11	0,16	−0,26	−0,01
PIB	−0,06	0,08	−0,14	−0,01
Exportações	−0,40	0,43	−0,69	−0,14
Importações	−0,05	−0,12	0,20	−0,13
Investimento	0,00	0,00	0,00	0,00
Consumo	0,00	0,00	0,00	0,00

Nota: Elaboração própria a partir dos resultados das simulações.

Se a política é mantida, no longo prazo os resultados são positivos, como pode ser observado na Tabela 7. Os setores se ajustam frente a nova relação do custo capital-trabalho, distorcida pela alteração nos impostos, a demanda responde aos novos preços com elevação do consumo, incentivando o investimento, de modo que a economia encontra um equilíbrio com maior nível de produção e emprego. Deste modo, pode-se dizer que o resultado agregado seria condizente aos objetivos do governo de promoção do crescimento econômico, à medida que podem gerar um aumento de 0,27% do emprego total, acompanhado por uma elevação de 0,10% no PIB, 0,25% no consumo das famílias e 0,14% no investimento.

Tabela 7. *Resultados agregados de longo prazo das simulações de desoneração da folha de pagamentos para setores selecionados pelo governo.*

Variáveis	Variação % Total	Contribuição por tipo de choque		
		Corte do INSS	Imposto sobre a receita bruta	Impostos sobre importações
Emprego	0,27	2,44	−1,98	−0,19
PIB	0,10	2,25	−1,96	−0,20
Exportações	−0,65	2,21	−2,41	−0,45
Importações	0,29	1,35	−0,86	−0,21
Investimento	0,14	2,36	−1,99	−0,24
Consumo	0,25	2,11	−1,71	−0,15

Nota: Elaboração própria a partir dos resultados das simulações.

Em relação ao setor externo, o impacto se mantém negativo sobre as exportações (redução de 0,65%), com crescimento das importações (0,29%),[11] que reflete o aumento do emprego e consequentemente da renda e do consumo, que não apenas aumentam o consumo nacional dos bens produzidos localmente, como também pressionam os preços nacionais, acirrando a concorrência com os produtos importados. De fato, ao se analisar o impacto das medidas sobre as importações setoriais, verifica-se que, embora haja redução nas compras externas dos setores alvo da política, devido à nova tarifa imposta, todos os demais setores aumentam a parcela de importados, em decorrência da elevação generalizada nos preços internos da economia, que leva à perda de competitividade perante às importações (cujos preços estão fixos, dada a hipótese de país pequeno no comércio internacional).

A elasticidade de substituição entre bens domésticos e importados é outro fator importante na condução dos resultados. Para algumas atividades com alto grau de substituição, pequenas alterações na razão de preços domésticos/ importados promovem sensíveis variações na parcela de importados demandados pelo setor, como é o caso de algumas commodities da Indústria Alimentícia e da Indústria de Bebidas.[12] Esses setores, por não serem alvo da política, registraram as maiores variações percentuais positivas nas importações frente ao aumento dos preços internos. Portanto, pode-se dizer que a medida estimula a produção doméstica dos setores em que ocorre a elevação da tarifa de importação, já que os bens comprados no exterior ficam relativamente mais caros frente aos nacionais, não obstante desestimula os setores não desonerados. Neste desenho de política, o aumento das importações nos setores excluídos das modificações tributárias supera a redução nos setores elegidos, promovendo o crescimento agregado das importações no longo prazo, o que contribui negativamente para os resultados sobre o PIB.

Ainda na Tabela 7, observando pela decomposição do choque é possível verificar que, para as exportações, os efeitos negativos do aumento de tributos sobre a receita bruta e importações são mais intensos e superam o incentivo da desoneração de encargos trabalhistas, fazendo com que a política deixe de cumprir seu objetivo de aumentar a competitividade dos produtos nacionais e com isto a capacidade de exportação. É preciso lembrar ainda, que a pauta de exportação brasileira é basicamente primário-exportadora, setores esses que não estão incluídos na política. Deste modo, o custo relativo do fator trabalho aumenta para essas atividades, causando desestímulo à sua produção, o que reflete

[11] Vale ressaltar que os resultados são influenciados pela a ausência de restrição à balança comercial no fechamento do modelo, isto é, a possibilidade de despoupança ou poupança externa contínua. Assumimos que a política pode alterar permanentemente a balança comercial. A adoção dessa pressuposição é justificada pelo objetivo proposto no artigo de investigar os efeitos da política sobre o comércio externo. A exogeneidade dessa variável impossibilitaria acessar os efeitos de longo prazo da política sobre o saldo comercial.

[12] Tratam-se exatamente de setores classificados no estudo de Tourinho, Kume & Pedroso (2007) como de "muito alta elasticidade de Armington".

diretamente nas exportações agregadas.

A partir desses resultados, pode-se concluir que a política poderia gerar efeitos positivos apenas no longo prazo, que, ainda assim, merecem algumas ressalvas quanto a sua análise. O efeito de redução nos encargos trabalhistas mostra-se positivo, impulsionando a economia, enquanto o aumento dos impostos sobre a receita bruta compensa parte da perda de arrecadação do governo. Porém, o efeito do aumento dos impostos sobre as importações atua no sentido contrário ao esperado, reduzindo a produção e o emprego.

Essa relação se torna mais clara a partir da decomposição setorial dos resultados[13] de longo prazo (Tabela 8). Para alguns setores, o efeito composto dos três choques (INSS, receita e importações) é positivo em termos da economia como um todo, ou seja, o efeito da desoneração da folha de pagamentos supera o efeito negativo do aumento de impostos. Esse é o caso, por exemplo, de bens de capital, cujos incentivos não são apenas benéficos para o próprio setor, mas geram resultados positivos sobre o PIB, o emprego e o consumo, para toda a economia. Comportamento semelhante ocorre ainda para outros setores como a produção de plásticos, e a indústria têxtil, que devido à redução de seus custos beneficiam os setores para os quais são vendidos como insumos de produção.

Por outro lado, para setores como Call Centers e produtores de material elétrico, embora o próprio setor se beneficie com a redução nos custos de produção, registrando aumento no nível de atividade, o efeito sobre a economia é negativo. Esses dois setores têm os custos do trabalho representando cerca de 50% do valor adicionado, montante um pouco abaixo da média dos demais setores incluídos na política (56%). Para Call Centers, acrescenta-se o fato de que a base de desoneração é menor, já que o setor representa 40% do valor bruto da produção da atividade de serviços prestados às empresas, sendo que em algumas atividades a desoneração abrangeu 100% do setor. Dessa forma, em termos comparativos, são atividades menos favorecidas que as demais, ou seja, que perdem vantagem comparativa. Outro fator importante são as ligações para frente do setor. Para Call Centers, por exemplo, as vendas ocor-

Tabela 8. *Resultados agregados de longo prazo das simulações, decompostos pela contribuição de cada setor no efeito total das políticas de desoneração da folha de pagamentos.*

	PIB	Emprego	Export.	Import.	Invest.	Consumo
Variação total	0,10	0,27	−0,65	0,29	0,14	0,25
Bens de Capital	0,26	0,31	0,20	0,05	0,28	0,23
Plástico	0,08	0,10	0,03	0,04	0,06	0,08
Têxteis	0,06	0,08	−0,05	0,03	0,05	0,08
Design House	0,06	0,05	0,03	0,07	0,09	0,06
Autopeças	0,03	0,04	−0,01	0,09	0,06	0,04
Naval e Aéreo	0,01	0,02	0,03	0,07	0,02	0,01
Ônibus	0,01	0,00	0,02	0,02	0,04	0,01
Vestuário	0,01	0,01	−0,06	0,04	0,02	0,03
Couro e Calçados	−0,01	0,00	−0,05	0,04	0,00	0,01
Móveis	−0,03	−0,01	−0,04	−0,01	−0,05	−0,02
TI & TIC	−0,05	0,02	−0,12	−0,04	−0,12	−0,03
Hotéis	−0,05	−0,05	−0,08	−0,03	−0,05	−0,04
Call Centers	−0,11	−0,11	−0,25	−0,02	−0,10	−0,07
Material Elétrico	−0,18	−0,18	−0,29	−0,06	−0,18	−0,14

Nota: Elaboração própria a partir dos resultados das simulações.

[13] O efeito da simulação sobre o setor i corresponde à soma dos efeitos de corte do INSS, aumento no imposto sobre a receita bruta e aumento no imposto sobre exportações. Pode-se observar que o resultado agregado é a soma dos resultados setoriais.

rem principalmente para o setor de comércio, que, por não estar incluído nas modificações tributárias, registrou redução em seu nível de atividade e consequentemente reduziu suas compras intermediárias. Esses resultados realçam a necessidade da identificação das interdependências setoriais e seus efeitos sistêmicos em uma economia para a formulação eficaz de políticas públicas.

As variações setoriais mostram que a política é eficaz, em um horizonte de longo prazo, para os setores diretamente beneficiados, no entanto, a interação no mercado de fatores, e principalmente, a concorrência por mão de obra, reduz o emprego e a produção dos setores não incluídos na política (para os quais o custo relativo da mão de obra aumenta), como é o caso da agropecuária e a indústria de alimentos (Tabela 9).

As Tabela 10 e 11 apresentam os resultados com maior nível de desagregação setorial. O setor de máquinas e equipamentos foi o principal responsável pelo aumento na produção de bens de capital, com 7,63% de crescimento do emprego e 1,19% de aumento na produção, fundamentalmente decorrente do efeito da desoneração da folha de pagamentos, que supera o efeito negativo da elevação dos tributos.

A simulação mostra ainda que todos os setores-alvo da política obtiveram aumentos de produção e alguns desencadearam resultados positivos em outros setores com os quais possuem elevado comércio intermediário, como Serviços de manutenção e reparação; Serviços imobiliários e de aluguel; e Construção com crescimento da produção respectivamente de 0,35%; 0,33% e 0,13%. Por outro lado, os setores que obtiveram redução na produção e no emprego são aqueles que não receberam incentivos fiscais do governo e para os quais a elevação do preço relativo ao fator trabalho superou os benefícios indiretos gerados pelo crescimento no consumo, emprego e produção de outros setores.

Esses resultados mostram que a escolha das atividades e produtos beneficiados é essencial na determinação de políticas públicas deste tipo, principalmente devido aos efeitos sistêmicos encadeados em outros setores, que podem gerar benefícios ou perdas indiretas para o restante da economia. Os setores diretamente beneficiados são favorecidos em termos de produção, emprego e tornam-se se mais competitivos frente às importações, o que configura importante estímulo à indústria nacional. No longo

Tabela 9. *Resultados setoriais de longo prazo sobre o emprego e a produção em setores agregados decorrentes das políticas de desoneração da folha de pagamentos.*

Setores	Variação % Total	Contribuição por tipo de choque		
		Corte do INSS	Imposto sobre a receita bruta	Impostos sobre importações
Emprego				
Agropecuária	−0,34	1,52	−1,66	−0,20
Indústria Extrativa	−1,63	0,91	−2,24	−0,29
Alimentos	−0,78	0,87	−1,50	−0,15
Bens de Consumo	−0,25	2,41	−2,35	−0,31
Duráveis	1,91	4,21	−2,15	−0,15
Intermediários	0,18	2,76	−2,32	−0,25
Bens de Capital	4,54	7,70	−2,92	−0,24
Serviços	0,11	2,21	−1,92	−0,18
Produção				
Agropecuária	−0,25	1,64	−1,68	−0,21
Indústria Extrativa	−0,36	2,13	−2,16	−0,33
Alimentos	−0,17	1,48	−1,48	−0,17
Bens de Consumo	0,11	2,14	−1,83	−0,19
Duráveis	0,38	2,68	−2,10	−0,20
Intermediários	−0,03	2,51	−2,26	−0,28
Bens de Capital	0,74	3,81	−2,81	−0,26
Serviços	0,12	2,30	−1,97	−0,20

Nota: Elaboração própria a partir dos resultados das simulações.

Tabela 10. Resultados setoriais sobre o emprego no longo prazo decorrentes das políticas de desoneração da folha de pagamentos.

Setores	Variação % Total	Contribuição por tipo de choque		
		Corte do INSS	Imposto sobre a receita bruta	Impostos sobre importações
10 setores mais beneficiados				
Máquinas e equipamentos	7,63	11,32	−3,56	−0,14
Ap. instr. médico-hosp., medida e óptico	6,44	9,11	−2,45	−0,22
Têxteis	4,73	7,65	−2,84	−0,09
Outros equipamentos de transporte	4,15	7,19	−2,62	−0,41
Artigos de borracha e plástico	4,13	7,23	−2,87	−0,24
Peças e acessórios para v. automotores	3,61	6,62	−2,73	−0,28
Móveis e produtos das indústrias diversas	3,42	5,94	−2,34	−0,18
Serviços de informação	2,71	5,10	−2,23	−0,16
Produtos farmacêuticos	2,55	8,16	−4,56	−1,05
Máquinas, aparelhos e materiais elétricos	2,10	4,91	−2,57	−0,24
10 setores menos beneficiados				
Petróleo e gás natural	−1,79	0,75	−2,26	−0,27
Produtos de madeira – exclusive móveis	−1,65	0,88	−2,23	−0,31
Minério de ferro	−1,58	0,70	−1,95	−0,33
Celulose e produtos de papel	−1,21	1,14	−2,12	−0,24
Outros da indústria extrativa	−1,09	1,62	−2,36	−0,34
Produtos do fumo	−1,07	0,75	−1,60	−0,22
Jornais, revistas, discos	−1,03	1,32	−2,16	−0,18
Refino de petróleo e coque	−0,93	1,47	−2,16	−0,24
Metalurgia de metais não ferrosos	−0,91	1,52	−2,16	−0,27
Álcool	−0,90	1,73	−2,34	−0,28

Nota: Elaboração própria a partir dos resultados das simulações.

prazo, trazem efeitos positivos sobre a renda, emprego e investimento. No entanto, vale lembrar que a distorção de recursos causada por esse tipo de política, ao trazer perda de vantagem competitiva para as atividades não incluídas, podem desfavorecer setores importantes, em termos de produção de bens essenciais e mesmo de exportações, bem como a economia como um todo. Um resultado importante encontrado a partir desse desenho de política, foi o desestímulo à produção da indústria alimentícia, importante segmento exportador da indústria brasileira, e mesmo para bens intermediários. Diante desses efeitos, a deterioração da balança comercial, com aumento das importações e queda nas exportações, deve ser um ponto a ser considerado pelos policymakers. A próxima seção apresenta os resultados da simulação alternativa, na qual todos os setores são desonerados, que tem por objetivo apontar qual a escolha de setores que maximiza os ganhos em termos de crescimento econômico para o país.

5.2. Simulação alternativa: desoneração em todos os setores

O objetivo desta seção consiste em apresentar os resultados de longo prazo da simulação de mudança tributária[14] proposta pelo programa Brasil Maior em todos os setores. O interesse é apontar a escolha de setores que traria maior retorno econômico. A Tabela 12 mostra o ranking dos setores em termos do impacto sobre a economia brasileira em termos da variação percentual do PIB provocada pela desoneração tributária de cada um dos setores (1ª coluna), reporta ainda o impacto na produção do próprio setor (2ª coluna).

[14]Todos os setores são desonerados (à exclusão dos setores de serviços prestados às famílias, e os setores públicos) considerando-se a desoneração na folha de pagamentos correspondente ao valor do INSS, compensado por um imposto de 1% sobre o Valor Bruto da Produção e 1% sobre as importações setoriais.

Tabela 11. Resultados setoriais sobre a produção decorrentes das políticas de desoneração da folha de pagamentos.

Setores	Variação % Total	Contribuição por tipo de choque		
		Corte do INSS	Imposto sobre a receita bruta	Impostos sobre importações
10 setores mais beneficiados				
Outros equipamentos de transporte	1,74	4,75	−2,59	−0,42
Máquinas e equipamentos	1,19	4,79	−3,45	−0,15
Têxteis	0,75	3,64	−2,78	−0,10
Ap. Instr. médico-hosp., medida e óptico	0,70	3,31	−2,38	−0,24
Peças e acessórios para v. automotores	0,67	3,65	−2,69	−0,29
Artefatos de couro e calçados	0,66	2,59	−1,77	−0,17
Máquinas, aparelhos e materiais elétricos	0,52	3,31	−2,55	−0,24
Artigos do vestuário e acessórios	0,46	2,57	−1,98	−0,14
Artigos de borracha e plástico	0,43	3,48	−2,80	−0,25
Serviços de manutenção e reparação	0,36	2,59	−2,04	−0,20
10 setores menos beneficiados				
Produtos de madeira – exclusive móveis	−0,63	1,91	−2,19	−0,34
Minério de ferro	−0,60	1,68	−1,92	−0,36
Outros da indústria extrativa	−0,45	2,25	−2,34	−0,36
Produtos do fumo	−0,44	1,36	−1,57	−0,23
Agr., silvicultura, exploração florestal	−0,31	1,71	−1,79	−0,24
Comércio	−0,28	1,86	−1,94	−0,20
Petróleo e gás natural	−0,26	2,27	−2,22	−0,31
Defensivos agrícolas	−0,23	2,20	−2,14	−0,29
Metalurgia de metais não ferrosos	−0,19	2,24	−2,14	−0,29
Serviços de alojamento e alimentação	−0,17	2,18	−2,16	−0,19

Nota: Elaboração própria a partir dos resultados das simulações.

Como mostra a segunda coluna, para alguns setores, embora a política em geral beneficie o próprio setor, incentivando o aumento da produção setorial, os efeitos encadeados entre os demais setores são potencialmente negativos, de modo que a modificação tributária nesses setores impactaria negativamente a economia como um todo. São exemplos as atividades Outros equipamentos de transporte, Produtos de madeira exclusive móveis, Celulose e produtos de papel, Metalurgia de metais não ferrosos, Petróleo e gás natural, Fabricação de aço e derivados, entre outros. Em geral, tratam-se de setores que produzem bens utilizados como insumos por outras atividades, ou, dito de outra forma, com fortes encadeamentos para frente na economia. A desoneração dos salários, ao pressionar a demanda por trabalho, termina por encarecer seu preço, o que reflete no preço final do bem produzido por esses setores, que, por serem utilizados como insumos, elevam o custo de produção de uma série de outras atividades que demandam seus produtos. Vale destacar ainda, que para três setores o impacto das medidas de desoneração sobre sua própria atividade é negativo, são eles Serviços de manutenção e reparação; Automóveis, camionetas e utilitários; e Refino de petróleo e coque.

A seleção de setores com impactos positivos na economia (de 1 a 18 na Tabela 12) teria o potencial de gerar um crescimento de 2,13% no PIB da economia brasileira de longo prazo. Esse crescimento seria acompanhado pelo aumento de 2,27% no emprego e 2,46% nos investimentos. Vale ressaltar, que neste caso, o efeito positivo na atividade econômica poderia compensar a arrecadação do governo via outros impostos, ou via aumento do montante total sobre o qual incidem os impostos da política.

6. COMENTÁRIOS FINAIS

A elevada carga tributária no Brasil e as discussões acerca da reforma do sistema tributário indicam a necessidade de reavaliação das incidências de impostos no país, principalmente quando a excessiva

Tabela 12. *Simulação alternativa: impactos de longo prazo da desoneração da folha de pagamentos para todos os setores.*

	Descrição	Impacto na economia brasileira (variação no PIB, %)	Impacto na produção do próprio setor (variação, %)
1	Transporte, armazenagem e correio	0,33	0,82
2	Pecuária e pesca	0,32	1,09
3	Construção	0,30	0,44
4	Máq. e equipamentos, inc. man, e reparos	0,27	2,13
5	Agricultura, silvicultura, exploração florestal	0,24	0,81
6	Serviços prestados às empresas	0,17	0,78
7	Artigos de borracha e plástico	0,11	0,84
8	Prod. de metal – ex. máq. e equipamentos	0,08	0,44
9	Outros produtos de minerais não-metálicos	0,07	0,57
10	Têxteis	0,06	0,98
11	Máquinas, aparelhos e materiais elétricos	0,06	0,78
12	Peças e acessórios para veículos automotores	0,03	1,11
13	Eletrodomésticos	0,03	0,64
14	Tintas, vernizes, esmaltes e lacas	0,02	0,47
15	Outros da indústria extrativa	0,01	0,77
16	Artigos do vestuário e acessórios	0,01	0,24
17	Caminhões e ônibus	0,01	0,64
18	Cimento	0,00	0,06
19	Produtos e preparados químicos diversos	0,00	0,69
20	Produtos do fumo	0,00	0,35
21	Defensivos agrícolas	0,00	0,27
22	Artefatos de couro e calçados	0,00	0,82
23	Perfumaria, higiene e limpeza	−0,01	0,32
24	Outros equipamentos de transporte	−0,01	1,87
25	Álcool	−0,01	0,25
26	Produtos de madeira – exclusive móveis	−0,01	0,62
27	Celulose e produtos de papel	−0,01	0,61
28	Máq. para escritório e equip. de informática	−0,02	0,13
29	Jornais, revistas, discos	−0,02	0,27
30	Ap./inst. médico-hospitalar, medida e óptico	−0,02	0,56
31	Móveis e produtos das indústrias diversas	−0,03	0,22
32	Petróleo e gás natural	−0,03	0,44
33	Metalurgia de metais não-ferrosos	−0,03	0,48
34	Fabricação de resina e elastômeros	−0,04	0,22
35	Material eletrônico e equip. de comunicações	−0,04	0,22
36	Minério de ferro	−0,04	0,11
37	Serviços de manutenção e reparação	−0,05	−0,26
38	Fabricação de aço e derivados	−0,06	0,23
39	Produtos químicos	−0,07	0,17
40	Elet. e gás, água, esgoto e limpeza urbana	−0,08	0,04
41	Automóveis, camionetas e utilitários	−0,15	−0,48
42	Serviços de informação	−0,16	0,01
43	Alimentos e bebidas	−0,18	0,14
44	Refino de petróleo e coque	−0,29	−0,43

tributação impõe entraves ao crescimento econômico como, por exemplo, ao reduzir a competitividade externa da produção nacional, ao desestimular o investimento e dificultar a absorção de mão de obra. Estes obstáculos ao crescimento são apontados por muitos como devidos, em parte, ao excessivo peso dos encargos trabalhistas sobre a folha de pagamentos, que elevam demasiadamente o custo laboral no mercado de trabalho formal.

Nesse contexto, o Governo Federal lançou em 2001 e ampliou ainda em 2012 uma série de medidas tributárias, cujo objetivo primordial é criar um ambiente propício para o investimento e a inovação, e aumentar a competitividade dos bens e serviços nacionais. Uma das principais medidas tomadas é a desoneração da folha de pagamentos, que ocorre por meio da mudança da base de arrecadação da contribuição previdenciária da folha salarial para a receita bruta. Para os setores elegíveis pela política a contribuição ao INSS (20,0%) incidente sobre a folha será zerada, passando a ser tributada em alíquotas que variam de 1% a 2% sobre sua receita bruta . Soma-se um acréscimo igual à alíquota sobre a receita bruta na importação desses bens.

O presente trabalho avaliou as consequências da desoneração dos impostos sobre a folha de salários e discutiu seus impactos macroeconômicos e setoriais. Para tanto, utilizou-se um modelo de equilíbrio geral computável (BRIDGE), com detalhamento específico da estrutura de impostos diretos e indiretos incidentes sobre 60 setores de atividades e 116 produtos.

Conforme as determinações da medida governamental e seguindo o arcabouço dos modelos de equilíbrio geral foram realizadas três simulações. As duas primeiras foram compostas basicamente por três choques sobre cada um dos setores-alvo da política: i) isenção de INSS; ii) criação de um imposto sobre a receita bruta; e iii) criação de um imposto sobre importações no curto e longo prazo. A terceira teve por objetivo indicar uma seleção alternativa de setores que traria o maior impacto sobre crescimento econômico no longo prazo. Para tanto, os mesmos choques foram implementados, contudo, abrangendo todos os setores.

Os resultados da simulação da política proposta pelo governo federal mostram que, em um horizonte de curto prazo, como consumo e investimento não se ajustam, os impactos econômicos das medidas são negativos. No longo prazo, contudo, a redução na carga tributária gera efeitos positivos para o crescimento da economia, induzindo o aumento no emprego de 0,27%, acompanhado por uma elevação de 0,10% no PIB e 0,25% no consumo das famílias. Ainda, setores não tradicionais da indústria brasileira são incentivados, como bens de capital. Desta forma, pode-se dizer que em termos gerais a redução nos encargos trabalhistas pode auxiliar na redução das ineficiências causadas pela estrutura do sistema tributário brasileiro, uma vez que incentiva a produção ao reduzir os custos de produção.

No entanto ao menos três ressalvas são necessárias:

(i) o aumento dos impostos sobre as importações, cujo objetivo é proteger a indústria nacional, incentiva os setores elegidos pela política, os quais reduzem a parcela de importados, contudo, leva a impactos negativos sobre a economia brasileira, pois onera a importação de insumos produtivos;

(ii) os setores não elegidos pela política perdem vantagem competitiva, o que aumenta as importações e reduz as exportações, já que passam a se defrontar com maior custo do trabalho e maiores preços para insumos domésticos, pois a política ao mesmo tempo em que desonera a produção, pressiona os preços devido ao aumento na demanda interna para consumo e mão de obra; e

(iii) para alguns setores o resultado sobre a economia como um todo é negativo, embora o próprio setor se beneficie, tornando a escolha dos setores ponto crucial para a eficácia da política.

A simulação das modificações tributárias, nos moldes da política, para todos os setores econômicos, estabeleceu um ranking dos setores com maior capacidade de impactar o crescimento econômico, ou, dito de outro modo, apontou a seleção de setores que teriam o potencial de maximizar o crescimento do PIB. Essa seleção teria o potencial de gerar um crescimento de longo prazo de 2,13% no PIB da economia brasileira.

A seleção de setores vencedores, no entanto, é bastante controversa na literatura e remete à política de campões nacionais, já esgotada pelo Governo Federal. Muito se discute a respeito das políticas de caráter vertical, isto é, que seleciona setores ou grupos, em contraposição às políticas de caráter mais horizontal, na qual todos os setores são incentivados, seja por medidas macroeconômicas ou mesmo por políticas multisetoriais. A grande crítica às políticas verticais ocorre pela distorção na alocação de recursos, causada quando apenas alguns segmentos são incentivados, e mesmo pelo ambiente propício à formação de lobbys empresariais. As políticas horizontais, justamente por não fazerem a escolha de setores a serem incentivados, contam, geralmente, com maior apoio da classe empresarial. Os resultados desse artigo mostram que a distorção na alocação de recursos causada por uma política vertical não deve ser ignorada no processo de tomada de decisão da política industrial. Os efeitos sistêmicos de uma política que escolhe setores podem levar a impactos indiretos que se contrapõem ao objetivo inicial, como ocorreu com o desincentivo à alguns setores exportadores e incentivo às importações dos setores não elegidos pela política. É preciso lembrar, no entanto, que na opção por uma política mais transversal, existe o risco de se perpetuar a estrutura industrial exatamente como ela é, no caso brasileiro, como destacado por Almeida (2009), uma estrutura diversificada e especializada na exportação de produtos de baixo conteúdo tecnológico. Vale realçar, ainda, outro resultado encontrado nesse estudo: mesmo no cenário em que todos os setores são desonerados, há distorção na alocação de recursos, de modo que alguns setores ganham e outros perdem.

Por último, vale mencionar que, conforme o próprio Governo Federal ressalta haverá perda de arrecadação para a previdência pública com a política, o que deverá ser compensado com redução nos gastos ou aumento de outro tributo, o que não foi anunciado pelo governo e por esse motivo não foi contemplado neste estudo. Sendo assim, uma possível extensão do trabalho seria a simulação do efeito de redução dos gastos públicos no montante de redução na arrecadação de tributos ou elevação de outros tributos que compensem a perda de receita.

REFERÊNCIAS BIBLIOGRÁFICAS

Almeida, M. (2009, dezembro). *Desafios da real política industrial brasileira do século XXI* (Texto para Discussão Nº 1452). Brasília·DF: Instituto de Pesquisa Econômica Aplicada (IPEA). Disponível em: http://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=4988

Amadeo, E. J. (1992). *Mercado de trabalho, relações industriais e competitividade*. Rio de Janeiro: BNDES.

Bitencourt, M. B., & Teixeira, E. C. (2008). Impactos dos encargos sociais na economia brasileira. *Nova Economia*, *18*(1), 3–86. Disponível em: http://revistas.face.ufmg.br/index.php/novaeconomia/article/view/498

Brasil · Ministério da Fazenda. (2012). *Desoneração da folha de pagamentos* [Cartilha]. Disponível em: http://www1.fazenda.gov.br/portugues/documentos/2012/cartilhadesoneracao.pdf

Cavalcanti, M. A. F. d. H., & Silva, N. L. C. d. (2010). Impactos de políticas de desoneração do setor produtivo: Uma avaliação a partir de um modelo de gerações superpostas. *Estudos Econômicos*, *40*(4), 943–966. doi: 10.1590/S0101-41612010000400008

Chahad, J. P. Z., & Fernandes, R. (Eds.). (2002). *O mercado de trabalho no Brasil: Políticas, resultados e desafios*. São Paulo: LTr.

Cury, S., & Coelho, A. M. (2010). From revenue to value added taxes: Welfare and fiscal efficiency effects in Brazil. *Revista Brasileira de Economia*, *64*(4), 373–392. doi: 10.1590/S0034-71402010000400002

DeVuyst, E. A., & Preckel, P. V. (1997). Sensitivity analysis revisited: A quadrature-based approach. *Journal of Policy Modeling*, *19*(2), 175–185. doi: 10.1016/0161-8938(95)00145-X

DIEESE. (2008, junho). *A proposta de reforma tributária do governo* (Nota Técnica Nº 69). São Paulo·SP: Departamento Intersindical de Estatística e Estudos Socioeconômicos (DIEESE). Disponível em: http://www.dieese.org.br/notatecnica/2008/notatec69propostareformatributariadogoverno.pdf

DIEESE. (2011, julho). *Encargos sociais e a desoneração da folha de pagamentos: Revisitando uma antiga polêmica* (Nota Técnica Nº 101). São Paulo·SP: Departamento Intersindical de Estatística e Estudos Socioeconômicos (DIEESE). Disponível em: https://www.dieese.org.br/notatecnica/2011/notatec101Desoneracao.pdf

Dixon, P. B., & Parmenter, B. (1996). Computable general equilibrium modelling for policy analysis and forecasting. In *Handbook of computational economics* (Vol. 1, pp. 3–85). Elsevier. doi: 10.1016/S1574-0021(96)01003-9

Dixon, P. B., Parmenter, B. R., Sutton, J. & Vincent, D. P. (1982). *ORANI: A multisectoral model of the Australian economy*. Amsterdam: North-Holland.

Dixon, P. B., & Rimmer, M. (2002). *Dynamic general equilibrium modelling for forecasting and policy: A practical guide and documentation of MONASH*. Amsterdan: Elsevier.

Dixon, P. B., & Rimmer, M. T. (1998, Jun). *Forecasting and policy analysis with a dynamic CGE model of Australia* (Working Paper Nº op-90). Victoria University, Centre of Policy Studies/IMPACT Centre. Disponível em: http://www.copsmodels.com/elecpapr/op-90.htm

Domingues, E. P., Betarelli Jr., A. A., Magalhaes, A. S., Carvalho, T. S. & Santiago, F. S. (2010, dezembro). Repercussões setoriais e regionais da crise econômica de 2009 no Brasil: Simulações em um modelo de equilíbrio geral computável de dinâmica recursiva. In *38° Encontro Nacional de Economia da ANPEC*, Salvador.

Domingues, E. P., Cardoso, D. F., Souza, K. d., Motta, G. P. d., Carvalho, T. S., Santiago, F. S., … Junior, A. A. B. (2012, fevereiro). *Crescimento, emprego e produção setorial: Efeitos da desoneração de tributos sobre a folha de salários no Brasil* (Texto para Discussão Nº 456). Belo Horizonte·MG: UFMG/CEDEPLAR. Disponível em: http://www.cedeplar.ufmg.br/pesquisas/td/TD%20456.pdf

Domingues, E. P., & Haddad, E. A. (2003). Política tributária e re-localização. *Revista Brasileira de Economia*, *57*(4), 849–871. doi: 10.1590/S0034-71402003000400009

Domingues, E. P., Magalhães, A. S. & Faria, W. R. (2009). Infraestrutura, crescimento e desigualdade regional: Uma projeção dos impactos dos investimentos do Programa de Aceleração do Crescimento (PAC) em Minas Gerais. *Pesquisa e Planejamento Econômico*, *39*(1), 121–158. doi: http://ppe.ipea.gov.br/index.php/ppe/article/view/1148

Ferreira Filho, J. B. d. S., Santos, C. V. d. & Lima, S. M. d. P. (2010). Tax reform, income distribution and poverty in Brazil: An applied general equilibrium analysis. *International Journal of Microsimulation*, *3*(1), 114–117. doi: http://microsimulation.org/IJM/V3_1/IJM_33.pdf

Frisch, R. (1959). A complete scheme for computing all direct and cross demand elasticities in a model with many sectors. *Econometrica*, *27*(2), 177–196. doi: 10.2307/1909441

Harrison, W. J., & Pearson, K. R. (2002). GEMPACK document GPD-1 [Software]. Melbourne, Australia: Centre of Policy Studies and Impact Project/Monash University.

Hoffmann, R. (2007). Elasticidades-renda das despesas e do consumo de alimentos no Brasil em 2002–2003. In F. G. Silveira, L. M. Servo, T. Menezes & S. F. Piola (Eds.), *Gasto e consumo das famílias brasileiras contemporâneas* (Vol. 2, pp. 463–484). Brasília: IPEA.

IBGE – Instituto Brasileiro de Geografia e Estatística. (2012). *Sistema de Contas Nacionais – Brasil 2004–2008*. Disponível em: http://www.ibge.gov.br/home/estatistica/economia/contasnacionais/2008/

Paes, N. L. (2009). Implantação do Princípio do Destino na cobrança do ICMS e suas implicações dinâmicas sobre os estados. *Revista Brasileira de Economia*, *63*(3), 233–248. doi: 10.1590/S0034-71402009000300003

Paes, N. L. (2011). A reforma tributária: Os efeitos macroeconômicos e setoriais da PEC 233/2008. *Estudos Econômicos*, *41*(2), 487–512. doi: 10.1590/S0101-41612011000200011

Paes, N. L., & Bugarin, M. N. S. (2006). Reforma tributária: Impactos distributivos, sobre o bem-estar e a progressividade. *Revista Brasileira de Economia*, *60*(1), 33–56. doi: 10.1590/S0034-71402006000100003

Pastore, J. (1994a). *Encargos sociais no Brasil e no exterior: Uma avaliação crítica*. Sebrae.

Pastore, J. (1994b). *Flexibilização e contratação coletiva*. LTR.

Peter, M. W., Horridge, M., Meagher, G. A., Naqvi, F. & Parmenter, B. R. (1996, April). *The theoretical structure of MONASH-MRF* (Preliminary Working Paper Nº OP-85). Clayton Vic: Centre of Policy Studies (COPS). Disponível em: http://www.copsmodels.com/ftp/workpapr/op-85.pdf

Salami, C. R., & Fochezatto, A. (2009). Avaliando os impactos de políticas tributárias sobre a economia brasileira com base em um modelo de equilíbrio geral de gerações sobrepostas. *Revista Brasileira de Economia*, *63*(3), 299–314. doi: 10.1590/S0034-71402009000300006

Silva, N. L. C. d., Tourinho, O. A. F. & Alves, Y. L. B. (2004, novembro). *O impacto da reforma tributária na economia brasileira: Uma análise com o modelo CGE* (Texto para Discussão Nº 1056). Rio de Janeiro: Instituto de

Pesquisa Econômica Aplicada (IPEA). Disponível em: http://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=4653

Tourinho, O. A. F., Alves, Y. L. B. & Silva, N. L. C. d. (2010). Implicações econômicas da reforma tributária: Análise com um modelo CGE. *Revista Brasileira de Economia*, *64*(3), 307–340. doi: 10.1590/S0034-71402010000300006

Tourinho, O. A. F., Kume, H. & Pedroso, A. C. d. S. (2007). Elasticidades de Armington para o Brasil: 1986-2002. *Revista Brasileira de Economia*, *61*(2), 245–267. doi: 10.1590/S0034-71402007000200006

Varsano, R., Pessoa, E. d. P., Silva, N. L. C. d., Afonso, J. R. R., Araujo, E. A. & Ramundo, J. C. M. (1998, agosto). *Uma análise da carga tributária no Brasil* (Texto para Discussão Nº 0583). Rio de Janeiro·RJ: Instituto de Pesquisa Econômica Aplicada (IPEA). Disponível em: http://www.ipea.gov.br/portal/index.php?option=com_content&view=article&id=3859

APÊNDICE A. MODELO EGC: BASE DE DADOS E ESTRUTURA TEÓRICA

O modelo EGC deste trabalho é um desenvolvimento do modelo BRIDGE (Domingues, Betarelli Jr., Magalhaes, Carvalho & Santiago, 2010), que por sua vez parte de elementos da especificação dos modelos MONASH e ORANI (Dixon et al., 1982; Dixon & Rimmer, 1998). O modelo segue a tradição australiana de modelagem, do tipo Johansen, nos quais a estrutura matemática é representada por um conjunto de equações linearizadas, as soluções são apresentadas como taxas de crescimento (elasticidades) e variados tipos de fechamento são permitidos. Além disso, assume pressupostos neoclássicos: firmas minimizadoras de custos, famílias maximizadoras de utilidade e *market clearing*.

A operacionalização de um modelo EGC é composta por duas partes. A primeira é a especificação, que consiste em determinar as formas funcionais, baseadas na teoria microeconômica tradicional consolidada. A segunda parte é denominada de calibragem, e consiste na determinação de uma solução inicial. Para a execução dessas duas etapas são necessários dois tipos de dados: os provenientes da matriz de absorção (núcleo da base de dados do modelo), as quais retratam os fluxos da economia, e ainda os parâmetros comportamentais relativos às formas funcionais adotadas (como por exemplo, as elasticidades de exportação, elasticidades de substituição).

A estrutura do núcleo da base de dados do modelo é representada na Figura A-1. As linhas mostram a estrutura de compras realizadas por cada agente representado nas colunas. Cada um dos 120 bens identificados pode ser obtido no mercado nacional ou internacional. Os bens são consumidos pelos 60 setores, para investimento, consumo, consumo do governo, variação de estoques e exportações. Sobre o consumo de bens pelos agentes incidem margens (m) e impostos (t). São utilizados 3 fatores de primários pelos 60 setores: trabalho, capital e terra. Sobre a produção dos setores incidem impostos (`V1PTX`), assim como sobre a folha de salários (`VLTAX`). A matriz de produção (`MAKE`) representa a produção dos c bens pelos i setores, e o imposto sobre as importações é representado pelo vetor `V0TAR`.

O modelo requer equações de demanda para todos agentes e fluxos representados na economia: demanda por produtos domésticos e importados, demanda por fatores de produção, equações de preço dos bens e fatores, equações de equilíbrio de mercado, e equações de agregados econômicos.

Os coeficientes de impostos, por sua vez, tratam a soma dos valores de IPI, ICMS e "Outros impostos menos subsídios" para todos os usuários (menos estoque). Apesar de o modelo permitir o tratamento de incidência dos impostos sobre o fluxo das exportações, no caso brasileiro, tais exportações estão desoneradas por lei. Assim, a calibragem dos coeficientes representativos aos impostos sobre os fluxos básicos (e.g. `V1TAX`) teve um procedimento semelhante aos coeficientes das margens. Os vetores de IPI, ICMS e Outros Impostos menos subsídios também foram distribuídos por indústria, particularmente, para `V1TAX` e `V2TAX` e, nos demais coeficientes, apenas por origem (doméstico e importado), a partir das tarifas calculadas sobre o total do valor básico. A exceção a esse procedimento se concentrou no Governo, nas Exportações e nas Variações de Estoque, os quais não sofrem incidência de impostos.

Além de dados provenientes das matrizes insumo-produto, os modelos de EGC empregam estimativas de elasticidades e parâmetros. Tais valores são geralmente extraídos da literatura, o que é justificado

Figura A-1. Núcleo da base de dados do modelo EGC.

		Agentes					
		Produtores	Investidores	Famílias	Exportações	Governo	Estoques
	Dimensões	i	1	f = 1, ... , 10	1	1	1
Fluxos básicos	c*s	V1BAS	V2BAS	V3BAS	V4BAS	V5BAS	V6BAS
Margens	c*s*m	V1MAR	V2MAR	V3MAR	V4MAR	V5MAR	V6MAR
Impostos	c*s*t	V1TAX	V2TAX	V3TAX	V4TAX	V5TAX	V6TAX
Trabalho	c	V1LAB					
Impostos sobre o trabalho	c	V1TAX					
Capital	c	V1CAP					
Terra	c	V1LND					
Impostos sobre a produção	c	V1PTX					
Outros Custos	c	V1OCT					

c - bens (1, ..., 116)
i - setores (1, ..., 60)
t - impostos indiretos (ICMS, IPI, Outros)
f - famílias
s - origem (doméstico, importado)
m - margens (comércio, transporte)

	Matriz de Produção	Tarifas de Importação
Dimensão	i	1
c	MAKE	V0TAR

pela escassez de dados para a estimação. São utilizados principalmente parâmetros relativos a elasticidades de substituição entre fatores primários, elasticidades de substituição do tipo Armington (entre as variedades domésticas e importadas dos bens) e elasticidade-gasto consumo das famílias.

Muitos desses parâmetros, como a elasticidade de substituição entre fatores primários e a elasticidade de substituição entre bens domésticos e importados, obtidas das estimativas econométricas contidas no Modelo de Equilíbrio Geral Computável Multi-Regional TERM-CEDEPLAR (Domingues, Magalhães & Faria, 2009). A elasticidade de Armington é definida por produto e é idêntica tanto para bens intermediários (SIGMA1), quanto para bens de investimento (SIGMA2) e para a demanda das famílias (SIGMA3). Tais parâmetros foram retirados de Tourinho et al. (2007). Quando necessário, as estimativas foram compatibilizadas aos setores do modelo.

A estrutura de demanda das famílias incorpora também em sua formulação o Parâmetro de Frisch (Frisch, 1959), que é um parâmetro de substituição que mede a sensibilidade da utilidade marginal da renda.[15] Ele é estimado com um valor negativo e é maior, em módulo, quanto mais pobre for a população em análise. Em outras palavras, quanto maior este parâmetro, em módulo, menor o grau de consumo de "luxo" e maior o grau de consumo de "subsistência". O BRIDGE-TAX emprega o valor −2,48, que foi o primeiro a ser estimado para a economia brasileira no modelo já citado (i.e. Multi-Regional TERM-CEDEPLAR). Os resultados apontam uma estimativa de −2,48, que coloca o Brasil, de acordo com a definição original (Frisch, 1959) no extrato de consumo de renda média Domingues et al. (2009). Com-

[15]Para maiores detalhes ver Dixon et al. (1982); Dixon & Rimmer (1998).

plementando a estrutura de demanda das famílias, utiliza-se um parâmetro que mede a elasticidade gasto dispêndio (EPS) das mesmas. Esses valores foram retirados de Hoffmann (2007). Hoffmann calcula a elasticidade-renda de diversos tipos de despesas (e.g. alimentação, habitação, vestuário, transporte, dentre outras), utilizando os dados da POF 2002/2003 (IBGE, 2012).

A apresentação esquemática das soluções de Johansen para modelos EGC é padrão na literatura. A seguir é apresentado um resumo desse procedimento de forma a se obter uma visão de como o modelo é operacionalizado. Maiores detalhes podem ser encontrados em Dixon et al. (1982) e Dixon & Parmenter (1996). O sistema de equações do modelo pode ser escrito como

$$F(V) = 0, \tag{A-1}$$

onde V é um vetor de equilíbrio de dimensão n (número de variáveis), e F é uma função-vetor não linear de dimensão m (número de equações). Supõe-se que F seja diferenciável, e que o número de variáveis seja maior que o número de equações no sistema ($n > m$). Dessa forma, $(n - m)$ variáveis devem ser determinadas exogenamente. Uma solução inicial de equilíbrio, V^{**}, deve ser determinada para fins de calibragem, ou seja, supõe-se que

$$\exists V = V^* \quad \text{tal que} \quad F(V^*) = 0. \tag{A-2}$$

Dada a solução inicial, V^*, um novo conjunto de soluções pode ser obtido para um dado vetor de alterações nas variáveis exógenas. Dessa forma, para se computar o novo conjunto de soluções para o modelo é necessária a partição do vetor V em dois grupos de variáveis, endógenas e exógenas. Seja Y o vetor contendo as m variáveis endógenas e X o vetor contendo as $(n - m)$ variáveis exógenas. A equação (A-2) pode ser reescrita como

$$F(Y,X) = 0. \tag{A-3}$$

O problema é computar Y para um dado X, sendo que não é possível, normalmente, escrever Y como uma função explícita de X. Várias técnicas podem ser empregadas para computar Y. A solução por linearização assume que já se possui alguma solução para o sistema (Y^0,X^0), tal que:

$$F(Y^0,X^0) = 0. \tag{A-4}$$

Em geral a solução inicial é obtida dos dados do período base em que o modelo é calibrado, ou seja, assume-se que o sistema é válido para algum ponto no passado. Para pequenas alterações em $\mathrm{d}X$ e $\mathrm{d}Y$ e sob hipóteses convencionais para o comportamento da função F, a diferenciação total de (A-4) permite obter[16]

$$F_Y(Y,X)\,\mathrm{d}Y + F_X(Y,X)\,\mathrm{d}X = 0, \tag{A-5}$$

onde F_Y e F_X são matrizes de derivadas parciais de F avaliadas em (Y^0,X^0). Para que as soluções sejam obtidas na forma de taxas de variação é conveniente expressar $\mathrm{d}Y$ e $\mathrm{d}X$ como pequenas variações percentuais x e y. Portanto, um típico elemento dos vetores x e y são definidos como

$$y = 100\frac{\mathrm{d}Y}{Y} \qquad x = 100\frac{\mathrm{d}X}{X}. \tag{A-6}$$

Da mesma forma podem ser definidas as seguintes funções:

$$G_Y(Y,X) = F_Y(Y,X)\widehat{Y} \qquad G_X(Y,X) = F_X(Y,X)\widehat{X}, \tag{A-7}$$

onde $\widehat{Y}$ e $\widehat{X}$ são matrizes diagonais. Assim o sistema linearizado torna-se

$$G_Y(Y,X) + G_X(Y,X) = 0. \tag{A-8}$$

[16] As hipóteses usuais são de continuidade e diferenciabilidade de F e solução única. Em geral estas características são atendidas em sistemas walrasianos de equilíbrio geral expressos na forma de equações linearizadas.

Soluções para sistemas como (A-8) podem ser facilmente computadas por meio de técnicas de álgebra linear. Estas representam, entretanto, apenas uma solução acurada para pequenas mudanças em X e Y. Caso contrário, erros de linearização podem ocorrer. A aproximação linear, ou em primeira ordem, da solução de (A-8) é dada por

$$y = -G_Y(Y,X)^{-1}G_X(Y,X)x. \tag{A-9}$$

A equação (A-9) leva à solução de Johansen, Y^J, uma aproximação da solução verdadeira, Yexata. Essa observação leva à ideia de particionar as alterações em X em pequenos passos, e para cada subalteração em X usar a aproximação linear para derivar a sub-resposta de Y. A partir dos novos valores de X e Y as matrizes de coeficientes G_Y e G_X são recomputadas e o processo é repetido para cada passo. Esta técnica é conhecida como método de Euler, uma das mais simples opções em técnicas de integração numérica — processo de uso de equações diferenciais para se mover de uma solução para outra. Para uma aproximação em 3 passos o erro de linearização é sensivelmente menor, aproximando-se da solução exata. Quanto maior o número de passos, melhor a aproximação.

Os modelos EGC são operacionalizados no programa GEMPACK.[17] Além do método de Euler, é possível obter as soluções por dois métodos alternativos, Gragg e Midpoint. Estes métodos são variações do método de Euler e podem, em alguns casos, produzir resultados mais precisos para um mesmo número de passos. No caso do modelo deste trabalho, um procedimento de cálculo em vários estágios (Johansen/Euler) é utilizado, em que o vetor de choques exógenos é dividido em p partes iguais. Desta forma, pode-se definir uma sequência de cálculos, de modo que as matrizes G_Y e G_X sejam reestimadas em cada estágio.

A questão teórica que se coloca diz respeito à maneira em que a solução do modelo converge para a solução verdadeira à medida que p cresça e, caso convirja, ao número de estágios (tamanho de p) necessário para a obtenção de uma solução precisa do modelo. Dixon et al. (1982) demonstram que quando p tender a infinito, o método de vários estágios de Johansen/Euler produzirá uma solução exata do modelo.

APÊNDICE B. ALOCAÇÃO DE INVESTIMENTO E ACUMULAÇÃO DE CAPITAL

A regra de investimento e acumulação de capital segue Dixon et al. (1982), ou regra DPSV, segundo a qual, o investimento realizado no período pode afetar o capital do próprio período (desde que o fechamento do modelo assim o permita, como o fechamento de longo prazo utilizado neste trabalho). A regra é desenhada para um modelo de estática comparativa, com ano base $t = 0$ e com choque que desloque a economia por 1 período. A regra não considera a trajetória do capital do início do período ($t = 0$) até o final do período de análise ($t = 1$), mas permite que o capital seja móvel entre os setores, de tal forma que o estoque de capital se modifique ao longo do período de simulação. Formalmente, a taxa de retorno do capital é dada por

$$R_J(0) = \frac{PK_j}{PI_j} - d, \tag{B-10}$$

em que $R_j(0)$ é a taxa de retorno do capital no início do período para a indústria j, dada pela relação entre o preço do capital na indústria j e o preço do investimento na indústria j menos a taxa de depreciação, d. Ou seja, a rentabilidade do capital aumenta (diminui) se o preço do capital aumenta (diminui) mais (menos) do que o custo de investimento para cada indústria j.

Considerando que o capital demora um período para ser instalado, os investidores são cautelosos, e suas expectativas sobre a taxa de retorno seguem a seguinte regra:

$$R_j(1) = R_j(0)\left(\frac{K_j(1)}{K_j(0)}\right)^{-\beta_j}, \tag{B-11}$$

[17] O modelo foi utilizado na versão 11.0 do programa GEMPACK. Sobre sua utilização na implementação de modelos EGC ver Harrison & Pearson (2002).

em que $R_j(1)$ é a taxa de retorno esperado no final do período, e β_j é um parâmetro positivo. Ou seja, conforme mostra a Figura B-2, quando não há alteração no estoque de capital do setor j, a taxa de retorno esperada do capital é igual a taxa no início do período ($R_j(1) = R_j(0)$), porém, quando o estoque de capital do setor aumenta durante o período, como no ponto A, o retorno do capital reduz para B. De forma similar, reduções no estoque de capital, aumentam o retorno esperado do capital.

Para que o capital seja alocado entre as indústrias, assume-se que existe uma taxa de retorno esperada única para toda a economia (Ω) definida exogenamente, de tal forma, que os investimentos se alocam entre os setores para equalizar o retorno do capital entre as indústrias. Assim, a equação (B-11) é substituída por:

$$R_j(0)\left(\frac{K_j(1)}{K_j(0)}\right)^{\beta_j} = \Omega. \tag{B-12}$$

Finalmente, o investimento I_j realizado no início do período passa a compor o estoque de capital do setor, somando-se ao estoque de capital no início do período $K_j(0)$ descontado da taxa de depreciação d. Assim, no final do período o estoque de capital se torna $K_j(1)$, dado por

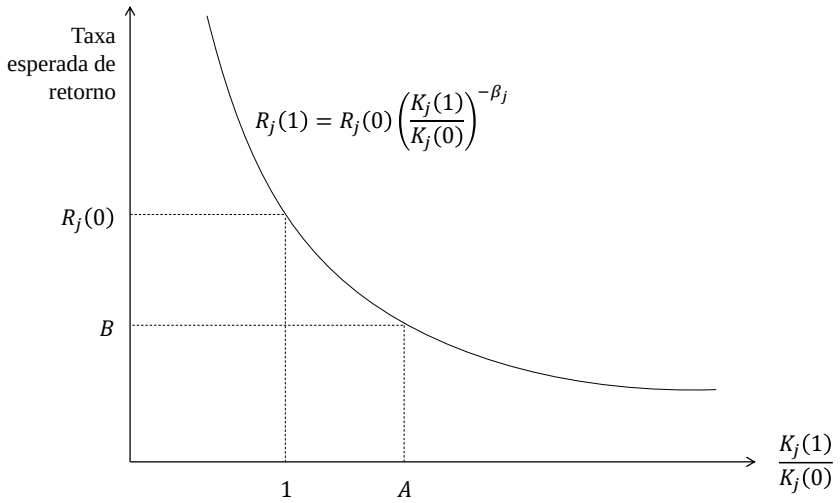
$$K_j(1) = K_j(0)(1-d) + I_j. \tag{B-13}$$

Com o estoque de capital no final do período definido em (B-12), a equação (B-10) define o nível de investimento do setor j. De tal modo, que para produzir novas unidades de capital, cada setor combina commodities em uma estrutura aninhada: no primeiro nível, o investidor representativo do setor i combina bens de capital c, de origem doméstica ou importada, minimizando o custo total de investimento, sujeito a uma estrutura Leontief. No segundo nível, para cada commodity c, o investidor minimiza o custo combinando as variedades doméstica e importada em uma função do tipo CES (hipótese de Armington).

APÊNDICE C. ANÁLISE DE SENSIBILIDADE

Uma série de parâmetros comportamentais são utilizados para a implementação do modelo. Conforme descrito no Apêndice 1, alguns desses parâmetros são baseados em estimativas econométricas e

Figura B-2. *Taxa esperada de retorno do capital para o setor j.*



para outros são atribuídos valores com base em conhecimentos econômicos e suposições. Essa dependência em relação a parâmetros não estimados diretamente para o modelo pode ser avaliada por meio da análise de sensibilidade.

A análise de sensibilidade sistemática empregada neste trabalho segue a metodologia de quadratura gaussiana proposta por DeVuyst & Preckel (1997). Nessa abordagem, o modelo EGC é tratado como um problema de integração numérica cuja solução (resultado das variáveis endógenas) pode ser obtida de forma pontual, e os dois primeiros momentos (média e variância), podem ser obtidos a partir da variação dos parâmetros e ou choques. Estas informações representam dados qualitativos a respeito da sensibilidade dos resultados do modelo a parâmetros específicos, e podem sugerir os elementos para onde a atenção do pesquisador deve estar focada.

Dadas as considerações acima, a análise de sensibilidade sistemática consistiu em testar a robustez dos parâmetros do modelo, a partir de sua variação em um intervalo de 50% para mais ou para menos, com distribuição uniforme em torno do parâmetro utilizado para cada uma das simulações. Os testes mostraram que os resultados agregados e setoriais se mostraram muito pouco sensíveis à variação dos parâmetros. A Tabela C-1, mostra os resultados da análise de sensibilidade para as variáveis macroeconômicas nas simulações de política no curto e longo prazo. Para a simulação 3 os resultados são semelhantes aos apresentados na Tabela, com pequenas variações na média após a segunda ou terceira casa decimal.[18]

Tabela C-1. *Análise de sensibilidade para as simulações da política no curto e longo prazo.*

	Curto prazo			**Longo prazo**		
	Simulação	Média	Desvio padrão	Simulação	Média	Desvio padrão
Emprego	−0,114	−0,111	0,013	0,267	0,273	0,047
PIB	−0,062	−0,060	0,007	0,096	0,102	0,056
Exportações	−0,404	−0,394	0,043	−0,648	−0,642	0,047
Importações	−0,047	−0,047	0,035	0,288	0,294	0,054
Investimento	0,000	0,000	0,000	0,135	0,142	0,079
Consumo	0,000	0,000	0,000	0,245	0,251	0,056

[18] Esses resultados não foram reportados pois na terceira simulação as variações agregadas não possuem um significado específico, uma vez que têm como objetivo único a seleção dos setores. Caso seja de interesse, testes de sensibilidade para setores específicos podem ser obtidos diretamente com os autores.