r b e

ISSN 0034-7140 (print) | ISSN 1806-9134 (online)
DOI 10.5935/0034-7140.20170013
Revista Brasileira de Economia, 71(3), 261–299

Fragilidade Bancária com (e sem) Serviço Sequencial*

Jefferson Bertolai†
Matheus Melo‡

Sumário: **1. Introdução; 2. A escolha de Sistema Bancário; 3. Instabilidade Financeira sem Serviço Sequencial; 4. Instabilidade Financeira com Serviço Sequencial; 5. Observações finais; A. Demonstrações; B. Figuras.**

Palavras-chave: **Serviço sequencial; Contágio; Corridas bancárias.**

Códigos JEL: **E4, E6.**

Este artigo apresenta os modelos e principais resultados de Allen e Gale (2000) e de Bertolai et alii (2016) como casos limites de um mesmo problema de escolha de sistema bancário ótimo para, em seguida, estabelecer resultados complementares a estas referências. Estuda-se formas alternativas de abordar o tema de fragilidade bancária, com base na exigência ou não de serviço sequencial ao escolher o sistema bancário ótimo. A primeira contribuição é complementar ao resultado de contágio estabelecido por Allen e Gale (2000) e mostra que existe uma intervenção no mercado interbancário capaz de eliminar o colapso generalizado do sistema bancário provocado pelo contágio entre bancos. A segunda contribuição é generalizar o resultado de existência de corrida bancária de Bertolai et alii (2016). São estabelecidas as condições sob as quais somente os k últimos depositantes de cada um dos bancos da economia não participam da crise bancária ao manter seus recursos investidos no sistema bancário.

This paper presents the models and main results of Allen e Gale (2000) and Bertolai et alii (2016) as limiting cases of the society's problem in choosing the optimal banking system. Next, complementary results to these references are established. Alternative approaches to banking fragility are studied, highlighting the effects of imposing the sequential service constraint on the choice of the optimal banking system. The first contribution is complementary to the contagion result of Allen e Gale (2000) and shows that there exists an intervention in the interbank market that eliminates the widespread collapse of the banking system caused by the contagion. In the second contribution, the bank-run's existence result of Bertolai et alii (2016) is generalized. In their results, everyone but the last depositor (or the last two depositors) participate in the bank run. For arbitrary k, we establish under which conditions only the last k depositors in each bank do not participate in the bank run by keeping their resources invested in the banking system.

1. INTRODUÇÃO

Uma importante referência na literatura acadêmica que estuda instabilidade financeira é Diamond e Dybvig (1983). Um dos precursores em utilizar a abordagem de Desenho de Mecanismos para estudar fragilidade bancária[1] (possibilidade de corridas bancárias) e o seguro depósito como forma de contornar

*Agradecemos os comentários e sugestões do avaliador anônimo.

†Professor e Pesquisador da Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto da Universidade de São Paulo. E-mail: jbertolai@fearp.usp.br

‡Mestre em Economia pela FEARP/USP. E-mail: mamelo@usp.br

[1]Outra referência neste sentido é Bryant (1980). Para uma referência anterior a respeito do efeito de *moral hazard* do seguro depósito sobre as decisões de investimento dos bancos, mas que utiliza a abordagem de equilíbrio, ver Kareken e Wallace (1978).

Recebido em 13 de maio de 2017
Aceitação Final em 03 de julho de 2017

tal problema, Diamond e Dybvig (1983) mostram que a existência de corrida bancária pode estar associada à busca da sociedade em alocar de forma eficiente a liquidez da economia. Argumentam então que um mecanismo de seguro depósito é capaz de eliminar a fragilidade financeira ao garantir proteção àqueles indivíduos que optam por manter seus recursos investidos no sistema financeiro.

Wallace (1988) argumenta, no entanto, que o seguro depósito proposto por Diamond e Dybvig (1983) não é factível devido a uma importante característica do sistema bancário, o *serviço sequencial*. No sistema bancário, os depositantes são atendidos de forma sequencial: não há a opção de operar o sistema acumulando demandas de saques para somente então escolher o nível dos (e executar os) pagamentos aos depositantes. Wallace (1988) observa que, embora Diamond e Dybvig (1983) tenham reconhecido a importância do serviço sequencial para modelos de Teoria Bancária, os autores não se utilizaram desta característica de forma adequada. O seguro depósito proposto por Diamond e Dybvig (1983) requer conhecer toda a demanda de saques antes de executar os pagamentos ou acionar o seguro depósito.

A partir da contribuição de Wallace (1988), a literatura que estuda fragilidade bancária pode ser dividida entre trabalhos que se utilizam de modelos nos quais o sistema bancário não pode operar sem respeitar o serviço sequencial e estudos nos quais a restrição de serviço sequencial é ignorada. As vantagens de impor serviço sequencial ao modelo já é reconhecida em Diamond e Dybvig (1983). A vantagem em ignorar o serviço sequencial é tratabilidade. Uma importante referência na literatura que se desenvolveu optando pela tratabilidade (ao ignorar a restrição de serviço sequencial) é Allen e Gale (2000). Após o cálculo do sistema bancário socialmente ótimo (escolha da forma eficiente de alocar a liquidez na economia), os autores mostram que algumas estruturas de mercado de depósitos interbancários são capazes de implementar (descentralizar) a alocação ótima.[2] Tais estruturas, no entanto, criam interligações entre os bancos capazes de transmitir choques localizados para toda a economia. É demonstrado que o surgimento inesperado de um choque localizado em um banco pode se propagar para todos os demais bancos (por meio do mercado interbancário) e gerar uma crise generalizada no sistema bancário, em um processo denominado pelos autores de *Contágio Financeiro*.

A complexidade do sistema bancário socialmente ótimo sob a exigência de serviço sequencial é documentada em Green e Lin (2003). Mais importante ainda, estes autores mostram que não há equilíbrio de corrida bancária sob o sistema escolhido pela sociedade se *(i)* os depositantes possuem uma boa estimativa sobre a quantidade de depositantes que visitaram o banco antes deles e *(ii)* os choques de liquidez (necessidade de consumo imediato) são independentes.[3] Tal resultado é central para a literatura em Teoria Bancária, pois impõe o desafio de encontrar um outro modelo capaz de explicar as crises bancárias existentes no mundo real.

Em uma resposta a este desafio, Peck e Shell (2003) propõem um modelo no qual os depositantes sabem pouco sobre a quantidade de depositantes que visitaram o banco antes deles e o choque de liquidez é bastante profundo. Os autores constroem um exemplo numérico no qual o sistema bancário ótimo possui equilíbrio de corrida bancária.[4] Bertolai et alii (2014) estudam a economia proposta por Peck e Shell (2003), mas preservando a intensidade do choque de liquidez utilizado por Green e Lin (2003). É demostrado que o sistema bancário ótimo possui equilíbrio de corrida bancária se a taxa de retorno da economia é baixa e/ou a população da economia é grande. Mais recentemente, Bertolai et alii (2016) demonstram que a alocação ótima de liquidez da economia pode gerar corrida bancária mesmo se os depositantes possuem uma boa estimativa sobre a quantidade de depositantes que visitaram o banco antes deles,[5] um importante contraponto ao resultado de Green e Lin (2003). Enquanto o modelo em Green e Lin (2003) pode ser visto como uma descrição de sistemas bancários nos quais a integração

[2]Para uma referência sobre a dificuldade de coexistência de bancos e mercados, ver Jacklin (1987).

[3]Para a importância da hipótese de independência, ver Andolfatto et alii (2007) e Ennis e Keister (2009).

[4]Ennis e Keister (2009) relaxam a hipótese de independência utilizada por Green e Lin (2003) e propõem um algoritmo capaz de computar exemplos de alocação ótima com equilíbrio de corrida bancária, se os choques de demanda por liquidez são correlacionados e a quantidade de depositantes não é muito grande.

[5]E se os choques forem independentes.

completa entre bancos é factível, o modelo proposto por Bertolai et alii (2016) pode ser visto como o caso no qual a integração é necessariamente parcial. Os bancos são capazes de prover (e obter) liquidez aos (dos) demais bancos no longo prazo, mas no curto prazo eles operam em isolamento físico e informacional.

A contribuição deste artigo reside em apresentar os modelos e principais resultados de Allen e Gale (2000) e de Bertolai et alii (2016) como casos limites de um mesmo *problema de escolha de sistema bancário ótimo* para, em seguida, estabelecer resultados complementares a estas referências. O artigo trata de formas alternativas de estudar fragilidade bancária com base na exigência ou não de serviço sequencial (ao escolher o sistema bancário ótimo). O primeiro resultado é complementar ao resultado de contágio estabelecido por Allen e Gale (2000) e mostra que existe uma intervenção no mercado interbancário capaz de eliminar o colapso generalizado do sistema bancário provocado pelo contágio entre bancos.

O segundo resultado é complementar ao resultado de existência de corrida bancária de Bertolai et alii (2016). São estabelecidas as condições sob as quais somente os k últimos depositantes de cada um dos bancos da economia não participam da crise bancária ao manter seus recursos investidos no sistema bancário.

O artigo está dividido em quatro seções. A primeira seção apresenta os principais objetos e conceitos utilizados nas seções 3 e 4. Em especial, ela apresenta o problema de escolha de sistema bancário, o conceito de instabilidade bancária e o conceito e implicações da restrição de serviço sequencial. As seções 3 e 4 discutem, respectivamente, os modelos de Allen e Gale (2000) e Bertolai et alii (2016). Um objetivo destas seções é apresentar formas de estudar instabilidade financeira na ausência e na presença de serviço sequencial. Outro objetivo é apresentar as contribuições deste artigo. Na subseção 3.3, é proposta uma intervenção no mercado interbancário capaz de eliminar o contágio entre bancos durante uma crise identificado por Allen e Gale (2000). Nas subseções 4.1 e 4.2, generaliza-se o resultado de Bertolai et alii (2016). O apêndice é dedicado a demonstrações e figuras não incluídas no texto.

2. A ESCOLHA DE SISTEMA BANCÁRIO

Considere uma economia com $N = 2$ indivíduos inicialmente idênticos, os quais vivem por dois períodos (data 1 e data 2) e estão considerando estabelecer um contrato entre si para organizar a distribuição de consumo entre datas e entre indivíduos. Este contrato estabelece o conjunto de regras do que será denominado o *sistema bancário* desta economia. A escolha contratual será feita no início da data 1 a fim de maximizar a *utilidade esperada* dos indivíduos e precisa respeitar as restrições de *factibilidade*, de *participação* e de *incentivos*, conforme descrito a seguir.

A utilidade esperada (*ex ante*) é a mesma para os dois indivíduos, uma vez que estes são inicialmente idênticos. Ela depende de choques sob os quais os indivíduos estão sujeitos. Após a assinatura do contrato e em $t = 1$, o indivíduo sofre ou não um choque de preferência, denominado *choque de liquidez*. Este choque ocorre com probabilidade $p \in (0,1)$ e torna o indivíduo ávido por consumo imediato (na data 1). Isto é modelado supondo que ele aufere utilidade somente a partir do consumo na data 1, c_1. Com probabilidade $1-p$, o indivíduo não sofre o choque de liquidez e, portanto, não possui necessidade de consumo imediato inadiável. Neste caso, supõe-se que ele aufere utilidade a partir do consumo agregado (nas duas datas), $c_1 + c_2$. Se a relação de preferência do indivíduo é representada pela função utilidade $u(x) = \frac{x^{1-\delta}}{1-\delta}$, então a utilidade do indivíduo será $u(c_1)$ se sofrer o choque de liquidez e será $u(c_1+c_2)$ se não sofrer o choque. No primeiro caso, o indivíduo é dito ser do tipo 0, ou ainda *impaciente*, e no segundo caso ele é dito ser do tipo 1, ou ainda *paciente*.

O momento exato durante a data 1 no qual cada indivíduo descobre se foi ou não atingido pelo choque de liquidez é incerto. Um dado indivíduo será o primeiro descobrir seu tipo (se sofreu ou não o choque) segundo a probabilidade $1/N = 0.5$. Ele será o segundo a descobrir com probabilidade $1/N = 0.5$. Se $\omega_i \in \{0,1\}$ denota o tipo do i-ésimo indivíduo a descobrir se sofreu ou não o choque, então $\omega \equiv (\omega_1,\omega_2) \in \{0,1\}^2$ denota a realização de choques na economia e é denominado estado da

natureza. O conjunto $\Omega \equiv \{0,1\}^N = \{(0,0),(0,1),(1,0),(1,1)\}$ é chamado de conjunto de estados da natureza.

Como o choque de liquidez é independente entre indivíduos e $\Pr(\omega_i = 0) = p$, então a probabilidade de cada estado $\omega \in \Omega$ é $\Pr(\omega_1,\omega_2) = p^{2-\omega_1-\omega_2}(1-p)^{\omega_1+\omega_2}$. Suponha que o i-ésimo indivíduo a descobrir seu tipo no estado da natureza ω consome $x_i(\omega)$ na data 1 e $y_i(\omega)$ na data 2. Então a utilidade esperada *ex ante* de cada indivíduo é

$$U(x,y) \equiv \sum_{\omega\in\Omega} \Pr(\omega) \sum_{i=1}^{2} \frac{1}{N} u\Big[x_i(\omega) + \omega_i y_i(\omega)\Big].$$

A factibilidade do contrato depende da quantidade de recursos na economia em cada data. Inicialmente, a dotação de recursos de cada indivíduo é $e = 1$ e, portanto, da economia como um todo é $Ne = 2$. Não há dotação de recursos ao longo da data 1 e na data 2. Existe uma tecnologia de armazenamento com retornos constante de escala, a qual gera (para cada unidade de investimento) uma unidade de recurso no curto prazo (data $t = 1$) e $R > 1$ unidades de recurso no longo prazo (data $t = 2$). Considere um dado estado da natureza $\omega \in \Omega$. Sendo $\sum_{i=1}^{N} x_i(\omega)$ o consumo total dos indivíduos na data 1 e este não pode ser maior do que os recursos disponíveis nesta data, então factibilidade do contrato exige que

$$\sum_{i=1}^{N} x_i(\omega) \leq Ne.$$

Os recursos não consumidos na data 1, $Ne - \sum_{i=1}^{N} x_i(\omega)$, são mantidos armazenados e geram $R\left(Ne - \sum_{i=1}^{N} x_i(\omega)\right)$ unidades de recursos na data 2. Sendo $\sum_{i=1}^{N} y_i(\omega)$ o consumo total dos indivíduos na data 2 e este não pode ser maior do que os recursos disponíveis nesta data, então factibilidade do contrato exige que

$$\sum_{i=1}^{N} y_i(\omega) \leq R\left(Ne - \sum_{i=1}^{N} x_i(\omega)\right).$$

Finalmente, o contrato não pode exigir quantidades de consumo negativas e, portanto, $x_i(\omega) \geq 0$ e $y_i(\omega) \geq 0$ para todo $i \in \{1,2\}$.

A participação dos indivíduos no contrato é voluntária e decidida no momento da assinatura do contrato (início da data 1). Os indivíduos estarão dispostos a participar do contrato se a utilidade esperada *ex ante* obtida sob o contrato for maior do que (ou igual) a utilidade esperada *ex ante* obtida fora do contrato. A opção de não participar do contrato é chamada de *autarquia*, uma vez que neste caso o indivíduo decide seu padrão de investimento e de consumo de forma independente dos demais indivíduos. Ele investe toda sua dotação e na tecnologia de armazenagem. Quando sofrer o choque de liquidez, o indivíduo escolherá consumir na data 1 todo recurso investido e consumirá nada na data 2. Se não sofrer o choque de liquidez, o indivíduo escolherá consumir nada na data 1 e todo recurso investido na data 2. Portanto, a utilidade esperada obtida fora do contrato bancário é dada por

$$U_a \equiv pu(e) + (1-p)u(Re).$$

Conclui-se que o contrato induzirá a participação dos indivíduos (atenderá a restrição de participação) se, e somente se, $U(x,y) \geq U_a$.

O problema de incentivo ao escolher o contrato tem origem na incapacidade da sociedade (todos os indivíduos em conjunto) observar todos os choques sofridos por cada um dos indivíduos. Embora o momento no qual o indivíduo sofre o choque é observado pela sociedade,[6] o tipo de choque é *informação*

[6] Wallace (1988) utiliza a analogia com uma máquina que faria pagamentos durante a noite enquanto todos (a sociedade) estão dormindo. O choque de liquidez neste caso é o nível de fome que cada indivíduo acorda durante a noite. A máquina observa o momento em que os indivíduos acordam e vão até a máquina requerer seu consumo $x_i(\omega)$ na data 1, mas não observa o nível de fome do indivíduo.

privada de cada indivíduo. Caso seja *ex ante* atrativo para a sociedade descobrir o tipo de cada indivíduo, o contrato firmado precisa prover incentivo para o indivíduo revelar seu verdadeiro tipo.

Considere a decisão do primeiro indivíduo a descobrir seu tipo, supondo que ele sabe ser o primeiro e acredita que o outro indivíduo irá anunciar ser do tipo 0 com probabilidade $q \in [0,1]$. Suponha que ele sofreu o choque de liquidez e, portanto, sua utilidade é $u(c_1)$. Se revelar a verdade (ser do tipo 0) e o outro indivíduo revelar ser do tipo ω_2, ele obterá níveis de consumo $x_1(0,\omega_2)$ na data 1 e $y_1(0,\omega_2)$ na data 2 e nível de utilidade $u[x_1(0,\omega_2)]$. Com isso, sua utilidade esperada em revelar a verdade é

$$U_T^1(0;q) \equiv qu[x_1(0,0)] + (1-q)u[x_1(0,1)]. \tag{1}$$

Caso minta sobre seu tipo (anuncie ser tipo 1) e o outro indivíduo revelar ser do tipo ω_2, ele obterá o plano de consumo $[x_1(1,\omega_2),y_1(1,\omega_2)]$ e nível de utilidade $u[x_1(1,\omega_2)]$. Logo, sua utilidade esperada em mentir sobre seu tipo é

$$U_L^1(0;q) \equiv qu[x_1(1,0)] + (1-q)u[x_1(1,1)]. \tag{2}$$

Portanto, o primeiro indivíduo a descobrir seu tipo possui incentivo a revelar verdadeiramente seu tipo quando ele é do tipo 0 se e somente se $U_T^1(0;q) \geq U_L^1(0;q)$.

Considere agora o caso em que o primeiro indivíduo não sofreu o choque de liquidez e, portanto, sua utilidade é $u(c_1 + c_2)$. Se revelar a verdade (ser do tipo 1) e o outro indivíduo revelar ser do tipo ω_2, ele obterá níveis de consumo $x_1(1,\omega_2)$ na data 1 e $y_1(1,\omega_2)$ na data 2 e nível de utilidade $u[x_1(1,\omega_2) + y_1(1,\omega_2)]$. Com isso, sua utilidade esperada em revelar a verdade é

$$U_T^1(1;q) \equiv qu[x_1(1,0) + y_1(1,0)] + (1-q)u[x_1(1,1) + y_1(1,1)]. \tag{3}$$

Caso minta sobre seu tipo (anuncie ser tipo 0) e o outro indivíduo revelar ser do tipo ω_2, ele obterá o plano de consumo $[x_1(0,\omega_2),y_1(0,\omega_2)]$ e nível de utilidade $u[x_1(0,\omega_2) + y_1(0,\omega_2)]$. Logo, sua utilidade esperada em mentir sobre seu tipo é

$$U_L^1(1;q) \equiv qu[x_1(0,0) + y_1(0,0)] + (1-q)u[x_1(0,1) + y_1(0,1)]. \tag{4}$$

Portanto, o primeiro indivíduo a descobrir seu tipo possui incentivo a revelar verdadeiramente seu tipo quando ele é do tipo 1 se e somente se $U_T^1(1;q) \geq U_L^1(1;q)$.

Em resumo, o primeiro indivíduo a descobrir seu tipo possui incentivo a revelar verdadeiramente seu tipo, quando ele é do tipo $\theta \in \{0,1\}$ e acredita que o outro indivíduo revelará ser do tipo 0 com probabilidade $q_1 \in [0,1]$, se e somente se $U_T^1(\theta;q_1) \geq U_L^1(\theta;q_1)$.

De forma análoga, o segundo indivíduo a descobrir seu tipo possui incentivo a revelar verdadeiramente seu tipo $\theta \in \{0,1\}$ quando ele acredita que o outro indivíduo revelou ser do tipo 0 com probabilidade $q_2 \in [0,1]$, se e somente se $U_T^2(\theta;q_2) \geq U_L^2(\theta;q_2)$, em que

$$\begin{aligned} U_T^2(\theta;q_2) &= \sum_{\omega_1=0}^{1} q_2^{1-\omega_1}(1-q_2)^{\omega_1} u\Big[x_2(\omega_1,\theta) + \theta y_2(\omega_1,\theta)\Big] \\ U_L^2(\theta;q_2) &= \sum_{\omega_1=0}^{1} q_2^{1-\omega_1}(1-q_2)^{\omega_1} u\Big[x_2(\omega_1,1-\theta) + \theta y_2(\omega_1,1-\theta)\Big]. \end{aligned}$$

Lembrando que a escolha contratual é feita no início da data 1 a fim de maximizar a utilidade esperada dos indivíduos e precisa respeitar as restrições de factibilidade, de participação e de incentivos. Ou seja, o *arranjo bancário* (x,y) é escolhido para resolver o seguinte problema de otimização:

$$\max_{x,y} U(x,y) \quad \text{sujeito a:} \begin{cases} x_i(\omega) \geq 0 \text{ e } y_i(\omega) \geq 0, & \forall i \in \{1,2\}, \forall \omega \in \Omega \\ \sum_{i=1}^{N} x_i(\omega) \leq Ne, & \forall \omega \in \Omega \\ \sum_{i=1}^{N} y_i(\omega) \leq R\left(Ne - \sum_{i=1}^{N} x_i(\omega)\right), & \forall \omega \in \Omega \\ U(x,y) \geq U_a & \\ U_T^i(\theta;p) \geq U_L^i(\theta;p), & \forall i \in \{1,2\}, \forall \theta \in \{0,1\} \end{cases}$$

em que se supõe na última restrição que o indivíduo decide ou não revelar a verdade para a sociedade supondo que o outro indivíduo está revelando a verdade com certeza (probabilidade 1). Ou seja, ele supõe que o outro indivíduo revelerá ser do tipo 0 com probabilidade p.

O arranjo bancário (x,y) escolhido pelos indivíduos como solução do problema acima possui as seguintes propriedades, por construção:

(a) Existe recurso suficiente para fazer todos os pagamentos programados pelo contrato para cada estado da natureza $\omega \in \Omega$.

(b) Os dois indivíduos aceitarão participar do (assinar o) contrato, uma vez que a utilidade sob o contrato é maior do que fora dele: $U(x,y) \geq U_a$.

(c) Existe um equilíbrio (no jogo de anúncios induzido pelo arranjo (x,y)) no qual os dois indivíduos revelam seu verdadeiro tipo. Se um dado indivíduo acredita que o outro está revelando a verdade para a sociedade, sua *melhor resposta* é revelar a verdade, uma vez que neste caso sua utilidade em revelar a verdade é maior do que a utilidade obtida ao mentir: $U_T^i(\theta;p) \geq U_L^i(\theta;p)$.

(d) A quantidade ótima de bancos é indeterminada, tendo em vista a tecnologia de estocagem com retorno de escala constante.[7] O arranjo bancário (x,y) pode ser visto como a descrição de um sistema com $M = 1$ banco com $L = 2$ depositantes, assim como de um sistema com $M = 2$ bancos com $L = 1$ depositante cada um, desde que estes operem em completa integração (física e informacional).[8]

O equilíbrio descrito no item (c) acima é chamado de equilíbrio com *revelação da verdade* e descreve o funcionamento do sistema financeiro em tempos normais, nos quais não há crise.

O sistema bancário (x,y) escolhido pelos indivíduos proporciona uma descrição para eventos de crise bancária se ele possui um outro equilíbrio, chamado de equilíbrio de *corrida bancária*. Neste equilíbrio de instabilidade bancária, indivíduos pacientes (tipo 1) mentem sobre seu tipo quando eles acreditam que outros pacientes estão mentindo e que todos os impacientes (tipo 0) estão revelando a verdade. Ou seja, o sistema bancário definido pelo contrato (x,y) possui um equilíbrio de corrida bancária se e somente se

$$\begin{array}{lcll} U_T^i(1;1) & < & U_L^i(1;1) & \text{para algum } i \in \{1,2\} \\ U_T^i(0;1) & \geq & U_L^i(0;1) & \text{para todo } i \in \{1,2\}. \end{array} \tag{5}$$

A condição (5) determina a extensão da corrida bancária. Se ela é satisfeita para todo i, a corrida é dita *completa*. Se ela é satisfeita para algum i, mas não para todo i, a corrida é dita *parcial*.

[7] Este é um resultado análogo ao que ocorre no modelo de Crescimento Neoclássico com retorno de escala constante, no qual o número de firmas usado para descentralizar a alocação ótima por meio de mercados competitivos é indeterminado.

[8] Pode-se ainda interpretar que (x,y) descreve um sistema com $M \in \mathbb{N}$ bancos com $L = 2$ depositantes cada um, desde que eles operem em completo isolamento – físico e informacional.

Parte da literatura relacionada ao modelo de Diamond e Dybvig (1983) estuda os sistemas bancários escolhidos para resolver o problema descrito acima. Uma importante referência neste sentido é Allen e Gale (2000). A subseção 2.1 apresenta a alocação ótima para o problema acima e apresenta o sistema bancário estudado por Allen e Gale (2000) como o caso limite no qual $N \to \infty$.

Inspirada em Wallace (1988), outra parte da literatura destaca a importância de o contrato entre os indivíduos satisfazer uma restrição adicional, chamada de restrição de *serviço sequencial*. Tal restrição não permite que os pagamentos escolhidos para o primeiro indivíduo dependa do anúncio do segundo indivíduo. Os pagamentos precisam ser imediatos e, portanto, efetuados em sequência. Especificamente, o contrato (x,y) precisa satisfazer

$$x_1(0,1) = x_1(0,0) \qquad \text{e} \qquad x_1(1,0) = x_1(1,1). \tag{6}$$

A subseção 2.2 apresenta a alocação ótima para o problema acima sujeito à restrição de serviço sequencial (6), a qual pode ser vista como a descrição de um sistema de $M \in \mathbb{N}$ bancos operando em isolamento com $N = 2$ depositantes cada um.[9] O sistema bancário estudado por Bertolai et alii (2016) é então apresentado como o caso limite no qual o número de bancos é muito grande ($M \to \infty$) e os bancos operam sob integração completa na data 2.

Propriedades gerais do sistema bancário

Algumas propriedades ótimas do sistema bancário (x,y) são comuns aos casos *sem* serviço sequencial e *com* serviço sequencial. Tais propriedades simplificam sensivelmente o problema de escolha de (x,y). Em especial, os itens (d) até (g) a seguir permitem reduzir a quantidade de variáveis de escolha do problema acima de $16 (= 2N2^N)$ para duas: $x_1(0,0)$ e $x_2(1,0)$.

(a) a restrição de participação $U(x,y) \geq U_a$ é sempre inativa e, portanto, pode ser ignorada na escolha de (x,y).

(b) a restrição de factibilidade $\sum_{i=1}^{N} x_i(\omega) \leq Ne$ é sempre satisfeita e, portanto, pode ser ignorada na escolha de (x,y).

(c) a restrição de incentivos $U_T^i(\theta;p) \geq U_L^i(\theta;p)$ é sempre inativa[10] e, portanto, pode ser ignorada na escolha de (x,y).

(d) não há desperdício de recursos, uma vez que a restrição de factibilidade da data 2, $\sum_{i=1}^{N} y_i(\omega) \leq R(Ne - \sum_{i=1}^{N} x_i(\omega))$ é sempre ativa.

(e) os indivíduos anunciando ser impacientes (tipo 0) não recebem pagamento na data 2:

$$y_1(0,0) = y_1(0,1) = y_2(0,0) = y_2(1,0) = 0.$$

(f) os indivíduos anunciando ser pacientes (tipo 1) não recebem pagamento na data 1:

$$x_1(1,0) = x_1(1,1) = x_2(0,1) = x_2(1,1) = 0.$$

(g) há tratamento igualitário na data 2 entre as pessoas anunciando ser pacientes:

$$y_1(1,1) = y_2(1,1) = Re.$$

[9] Conforme destacado na nota de rodapé 8.

[10] Conforme estabelecido por Green e Lin (2003).

Observe que, independentemente da presença ou não da restrição de serviço sequencial, o sistema bancário ótimo provê seguro completo contra o choque de posicionamento na fila de saques para os indivíduos anunciando ser pacientes (tipo 1).

Com base nas propriedades acima, a escolha do sistema bancário ótimo resolve o seguinte problema:

$$\max_{x,y} \sum_{\omega} P(\omega) \left(\sum_{i=1}^{2} \frac{1-\omega_i}{N} u\Big[x_i(\omega)\Big] + \frac{|\omega|}{N} u\Big[y(\omega)\Big] \right) \tag{7}$$

$$\text{sujeito a} \begin{cases} x_i(\omega) \geq 0 \quad \text{e} \quad y(\omega) \geq 0 & \forall \omega \in \Omega \\ |\omega| y(\omega) = R\left(Ne - \sum_{i=1}^{2}(1-\omega_i)x_i(\omega)\right) & \forall \omega \in \Omega \end{cases}$$

em que $|\omega| = \sum_i \omega_i$ é a quantidade de anúncios de tipo 1 em ω e (ω_{-i},θ) é construído a partir de $\omega \in \Omega$ substituindo a i-ésima coordenada de ω por θ. Ou seja, $(\omega_{-1},\theta) = (\theta,\omega_2)$ e $(\omega_{-2},\theta) = (\omega_1,\theta)$.

2.1. Sistema bancário *sem* serviço sequencial

Na ausência da restrição de serviço sequencial, o sistema bancário é capaz de aguardar todos os anúncios de tipo na data 1 antes de iniciar os pagamentos desta data. Como consequência, o contrato ótimo entre os indivíduos é bastante simples, conforme estabelecido pelo lema 2.1 a seguir.

***Lema* 2.1** (Alocação ótima sem serviço sequencial)**.** *O sistema bancário ótimo sem serviço sequencial*

(i) trata igualmente as pessoas anunciando ser impacientes;

(ii) paga mais àqueles anunciando ser pacientes do que àqueles anunciando ser impacientes, pois para todo $\omega \notin \{(0,0),(1,1)\}$

$$u'[x_i(\omega_{-i},0)] = Ru'[y(\omega)]$$

para todo i *tal que* $\omega_i = 0$.

Demonstração. Seja $\mu_\omega \Pr(\omega)/N$ o multiplicador de Lagrange associado a restrição de factibilidade. O lagrangeano associado ao problema de maximização é

$$\sum_{\omega\in\Omega} \frac{P(\omega)}{N} \left\{ \sum_{i=1}^{2} (1-\omega_i) u[x_i(\omega)] + |\omega| u(y(\omega)) \right\} + \sum_{\omega\in\Omega} \frac{P(\omega)}{N} \mu_\omega \left[R\left(Ne - \sum_{i=1}^{2} (1-\omega_i) x_i(\omega) \right) - |\omega| y(\omega) \right].$$

A condição de primeira ordem em $x_i(\omega_{-i},0)$ é

$$\frac{P(\omega_{-i},0)}{N} \left[(1-\omega_i) u'(x_i(\omega_{-i},0)) - R\mu_{(\omega_{-i},0)}(1-\omega_i) \right] = \frac{P(\omega_{-i},0)}{N} [u'(x_i(\omega)) - R\mu_\omega] = 0.$$

Sabe-se que a razão $\dfrac{P(\omega)}{N} \neq 0$, então $u'(x_i(\omega)) = R\mu_\omega$. Para o caso em que $\omega = (0,0)$, tem-se $u'(x_1(0,0)) = R\mu_\omega = u'(x_2(0,0))$. Como u' é estritamente decrescente (dada a concavidade estrita da função utilidade), tem-se $x_1(0,0) = x_2(0,0)$. A condição de primeira ordem em relação à $y(\omega_{-i},1)$:

$$\frac{P(\omega_{-i},1)}{N} \left[\omega_i u'(y(\omega_{-i},1)) - \mu_{(\omega_{-i},1)} \omega_i \right] = \frac{P(\omega_{-i},1)}{N} \left[u'(y_i(\omega)) - \mu_\omega \right] = 0.$$

Lembrando que $\frac{P(\omega)}{N} \neq 0$, então $u'(y(\omega)) = \mu_\omega$. Logo, para $\omega \notin \{(0,0),(1,1)\}$, tem-se

$$Ru'(y(\omega)) = R\mu_\omega = u'(x_i(\omega)),$$

e, portanto, $u'(y(\omega)) < u'(x_i(\omega))$. Como u' é estritamente descrescente, segue que $y(\omega) > x_i(\omega)$. □

Como consequência do lema 2.1 e da restrição de recursos ativa, quando $\delta = 2$ tem-se $x_1(0,0) = x_2(0,0) = e$, $y(1,0) = y(0,1) = \sqrt{R}x_1(0,1) = \sqrt{R}x_2(1,0) = \sqrt{R}Ne/(1+\sqrt{R})$. Como já tínhamos que $y(1,1) = Re$, o contrato ótimo está completamente descrito.

Na ausência da restrição de serviço sequêncial, o nível de consumo de cada pessoa depende somente da quantidade de anúncios de cada tipo. Em particular, o sistema bancário ótimo provê seguro completo contra o choque de posicionamento na fila de saques também para os indivíduos anunciando ser impacientes (tipo 0).

Considere agora o caso $N > 2$. A generalização da demonstração do lema 2.1 para este caso é direta e, portanto, o tratamento igualitário na data 1 continua válido. Com isso, o problema de maximização sem serviço sequencial pode ser reescrito como:

$$\max_{x,y} \sum_{j=0}^{N} \Pr(|\omega| = j) \left[\frac{N-j}{N} u[x(j)] + \frac{j}{N} u[y(j)] \right] \tag{8}$$

$$\text{sujeito a } \forall j \in \{0,1,\cdots,N\} \begin{cases} x(j) \geq 0 \quad \text{e} \quad y(j) \geq 0 \\ \dfrac{N-j}{N} x(j) + \dfrac{j}{N} y(j) R^{-1} \leq 1, \end{cases} \tag{9}$$

em que $N - |\omega|$ é a quantidade de impacientes no estado $\omega \in \Omega$ e $\Pr(|\omega| = n) = \binom{N}{n} p^{N-n}(1-p)^n$ é a probabilidade de a quantidade de pacientes ser igual a n. O corolário 2.2 a seguir generaliza o lema 2.1 para o caso com um número arbitrário de depositante $N \in \mathbb{N}$.

***Corolário* 2.2** (Sistema bancário ótimo sem serviço sequencial)**.** *Suponha $\delta = 2$ e $N \in \mathbb{N}$ depositantes. O sistema bancário ótimo sem serviço sequencial é dado por (x,y) tal que para cada $\omega \in \Omega$*

$$(x(|\omega|), y(|\omega|)) = \begin{cases} (e,0) & \text{se } 0 = |\omega| < N \\ \alpha\left(\frac{|\omega|}{N}\right)\left(e, \sqrt{R}e\right) & \text{se } 0 < |\omega| < N \\ (0,Re) & \text{se } 0 < |\omega| = N \end{cases}$$

em que $\alpha(q) = \dfrac{\sqrt{R}}{\sqrt{R}(1-q)+q}$ *para* $q \in [0,1]$.

Demonstração. Ver apêndice A. □

A economia estudada em Allen e Gale (2000), na qual há um contínuo de indivíduos, pode ser vista como o caso limite da economia acima quando $N \to \infty$. O problema de maximização apresentado em Allen e Gale (2000) é

$$\max_{x,y} \Big\{ \gamma u(x) + (1-\gamma) u(y) \Big\} \qquad \text{sujeito a} \quad \begin{cases} x \geq 0 \quad \text{e} \quad y \geq 0 \\ \gamma x + (1-\gamma) R^{-1} y \leq 1 \end{cases}$$

em que $\gamma \in [0,1]$. Usando uma Lei dos Grandes Números, a variável aleatória $|\omega|/N$ converge para $(1-p)$ quando $N \to \infty$. Ou seja, para todo $\varepsilon > 0$, tem-se

$$\lim_{N\to\infty} \Pr\left(\left|\frac{|\omega|}{N} - (1-p)\right| > \varepsilon\right) = 0.$$

Com isso, quando $N \to \infty$, a função objetivo em (8) converge para $pu(x) + (1-p)u(y)$ e a única restrição de factibilidade em (9) revelante do ponto de vista probabilístico é $px + (1-p)R^{-1}y \leq 1$.

Tomando $\gamma = p$, o sistema bancário ótimo na economia de Allen e Gale (2000) e no caso limite da economia acima coincidem. Do corolário 2.2 e do fato de que $\alpha(1-p) = \dfrac{\sqrt{R}}{1+p(\sqrt{R}-1)}$, o contrato ótimo converge para

$$(x,y) = \left(\frac{\sqrt{R}}{p(\sqrt{R}-1)+1}e, \frac{R}{p(\sqrt{R}-1)+1}e\right)$$

o qual coincide com a alocação ótima apresentada em Allen e Gale (2000) se $\gamma = p$ e $e = 1$.

2.2. Sistema bancário *com* serviço sequencial

Considere novamente a escolha de sistema bancário com $N = 2$ indivíduos. Conforme discutido acima, com serviço sequencial o contrato ótimo precisa satisfazer a restrição $x_1(0,0) = x_1(0,1)$. Ou seja, o pagamento para o indivíduo na posição $i = 1$ com anúncio ω_1 não pode depender do anúncio do indivíduo na posição posterior, $i = 2$. Portanto, o problema de escolha de sistema bancário é resolver o problema (7) sujeito a restrição de serviço sequencial. Ou seja,

$$\max_{x,y} \sum_{\omega} P(\omega)\left(\sum_{i=1}^{2}\frac{1-\omega_i}{N}u\Big[x_i(\omega_{-i},0)\Big] + \frac{|\omega|}{N}u\Big[y(\omega_{-i},1)\Big]\right) \tag{10}$$

$$\text{sujeito a} \begin{cases} x_i(\omega) \geq 0 \quad \text{e} \quad y(\omega) \geq 0 & \forall \omega \in \Omega \\ |\omega| y(\omega) = R\left(Ne - \sum_{i=1}^{2}(1-\omega_i)x_i(\omega)\right) & \forall \omega \in \Omega \\ x_1(0,0) = x_1(0,1) \end{cases}$$

***Lema* 2.3** (Alocação ótima com serviço sequencial)**.** *O sistema bancário ótimo com serviço sequencial é definido pelas restrições de factibilidade e por*

$$(x_1(0,0), x_2(1,0)) = \left(\frac{1}{1+\sqrt{p+(1-p)/R}}, \frac{1}{1+\sqrt{1/R}}\right)Ne.$$

Como consequência, o contrato ótimo não trata igualmente as pessoas anunciando ser impacientes.

Demonstração. O problema de maximização não é separável em $\omega \in \Omega$ devido a existência da restrição de serviço sequencial (o mesmo pagamento ocorre em histórias diferentes, como exemplo tome o pagamento $x_1(0)$ que ocorre nas histórias $(0,0)$ e $(0,1)$). Desse modo, o problema de maximização acima pode ser reescrito como:

$$p^2\left[\frac{1}{2}u(x_1(0,0))+\frac{1}{2}u(Ne-x_1(0,0))\right]+(1-p)^2\left[\frac{1}{2}u(y_1(1,1))+\frac{1}{2}u(RNe-y_1(1,1))\right]$$
$$+p(1-p)\left[\frac{1}{2}u(x_1(0,0))+\frac{1}{2}u(R(Ne-x_1(0,0)))\right]$$
$$+p(1-p)\left[\frac{1}{2}u(R(Ne-x_2(1,0)))+\frac{1}{2}u(x_2(1,0))\right]$$

em que as restrições de factibilidade e de serviço sequencial foram substituídas na função objetivo. A condição de primeira ordem em $x_1(0,0)$ é

$$\frac{p^2}{N}\left[u'(x_1(0,0))-u'(Ne-x_1(0,0))\right]+\frac{p(1-p)}{N}\left[u'(x_1(0,0))-Ru'[R(Ne-x_1(0,0))]\right]=0$$

Usando $\delta=2$, obtém-se

$$p+(1-p)R^{-1}=\frac{u'(x_1(0,0))}{u'\left[Ne-x_1(0,0)\right]}=\left(\frac{Ne-x_1(0,0)}{x_1(0,0)}\right)^2=\left(\frac{Ne}{x_1(0,0)}-1\right)^2$$

e, portanto, $x_1(0,0)=\frac{1}{1+\sqrt{p+(1-p)/R}}Ne$. A condição de primeira ordem em $x_2(1,0)$ é

$$\frac{p(1-p)}{N}\left[u'(x_2(1,0))-Ru'(R(Ne-x_2(1,0)))\right]=0.$$

Usando $\delta=2$ e $\frac{p(1-p)}{N}\neq 0$, obtém-se $u'(x_2(1,0))=R^{-1}u'(Ne-x_2(1,0))$ e, portanto,

$$R^{-1}=\frac{u'(x_2(1,0))}{u'(Ne-x_2(1,0))}=\left(\frac{Ne-x_2(1,0)}{x_2(1,0)}\right)^2=\left(\frac{Ne}{x_2(1,0)}-1\right)^2$$

de onde se obtém $x_2(1,0)=\frac{1}{1+\sqrt{1/R}}Ne$.

□

Após o anúncio de tipo 0 na primeira posição, o serviço sequencial impede o sistema de aguardar o anúncio da segunda posição. Conforme estabelecido no lema 2.1, quando é possível aguardar o anúncio, o sistema paga para a primeira posição $x_1(0,0)=e$ se o segundo anúncio é tipo 0 e $x_1(0,1)=\alpha(1/N)e=\frac{\sqrt{R}}{\sqrt{R}+1}Ne$ se o segundo anúncio é tipo 1. Aqui, o sistema estima que o segundo anúncio será tipo 0 com probabilidade p e paga $x_1(0,0)=x_1(0,1)$ tais que

$$e<x_1(0,0)=\frac{\sqrt{R}}{\sqrt{R}+\sqrt{pR+(1-p)}}Ne<\alpha(1/N)e.$$

O lema 2.3 mostra que não há seguro completo na data 1 contra o choque de posicionamento na fila de saques, uma vez que $x_2(0,0)=Ne-x_1(0,0)<Ne-e=e$ e, portanto, $x_2(0,0)<e<x_1(0,0)$. Este resultado de seguro incompleto na data 1 decorre da exigência de serviço sequencial, tendo em vista que a única diferença entre os problemas de escolha do sistema bancário ótimo das subseções 2.1 e 2.2 é a presença da restrição de serviço sequencial no problema (10).

O resultado de seguro incompleto se mantém com $N>2$ depositantes em cada um de $M\in\mathbb{N}$ bancos da economia – desde que eles operem sob integração completa. Similarmente, a agenda de

pagamentos descrita no lema 2.3 se mantém com $M > 1$ bancos e $N = 2$ depositantes cada um, desde que não exista integração entre os bancos (eles operem em isolamento).[11,12]

O sistema bancário estudado em Bertolai et alii (2016) pode ser visto como um caso intemediário de integração entre $M \in \mathbb{N}$ bancos da economia. Os bancos operam em isolamento na data 1, mas operam sob integração completa na data 2. Por exemplo, suponha que há $M = 2$ bancos e que, ao final da data 1, os anúncios foram $\omega^{(1)} = (\omega_1^{(1)},\omega_2^{(1)})$ no banco 1 e $\omega^{(2)} = (\omega_1^{(2)},\omega_2^{(2)})$ no banco 2. Os pagamentos na data 1 foram $\left(x_1^{(1)}(\omega^{(1)}),x_2^{(1)}(\omega^{(1)})\right)$ no banco 1 e $\left(x_1^{(2)}(\omega^{(2)}),x_2^{(2)}(\omega^{(2)})\right)$ no banco 2. A integração completa entre os bancos na data 2, permite financiar os pagamentos na data 2 usando a poupança agregada da economia,

$$R\left[Ne - \sum_{i=1}^{2}(1-\omega_i^{(1)})x_i^{(1)}(\omega^{(1)})\right] + R\left[Ne - \sum_{i=1}^{2}(1-\omega_i^{(2)})x_i^{(2)}(\omega^{(2)})\right],$$

enquanto que o isolamento entre os bancos na data 2 limita os pagamentos do banco $m \in \{1,2\}$ na data 2 à sua poupança: $|\omega^{(m)}|y^{(m)}(\omega) \leq R\left(Ne - \sum_{i=1}^{2}(1-\omega_i^{(m)})x_i^{(m)}(\omega^{(m)})\right)$.

Mais geralmente, se $\tilde{\omega} = (\omega^{(1)},\cdots,\omega^{(M)})$ e se $x^{(m)}(\omega^{(m)})$ e $y^{(m)}(\tilde{\omega})$ denotam os pagamentos do banco $m \in \{1,2,\cdots,M\}$ nas datas 1 e 2, respectivamente, então a restrição de factibilidade sob integração completa na data 2 é dada por

$$\sum_{m=1}^{M}|\omega^{(m)}|y^{(m)}(\tilde{\omega}) \leq \sum_{m=1}^{M} R\left[Ne - \sum_{i=1}^{2}(1-\omega_i^{(m)})x_i^{(m)}(\omega^{(m)})\right]. \tag{11}$$

Ou seja, a restrição de factibilidade em (10) é substituída pela sua soma em $\omega^{(m)}$, a restrição (11). A utilidade esperada *ex ante* dos indivíduos é[13]

$$\sum_{m=1}^{M}\frac{1}{M}\sum_{\omega^{(m)}\in\Omega}P\left(\omega^{(m)}\right)\left(\sum_{i=1}^{2}\frac{1-\omega_i^{(m)}}{N}u\left[x_i^{(m)}(\omega_{-i}^{(m)},0)\right] + \frac{|\omega^{(m)}|}{N}u\left[y^{(m)}(\tilde{\omega})\right]\right), \tag{12}$$

a qual pode ser vista como a soma da função objetivo de (10) em $m \in \{1,2,\cdots,M\}$, normalizada por $1/M$. O problema de escolha de sistema bancário (x,y) em Bertolai et alii (2016) consiste em maximizar (12) sujeito a (11) e $\sum_{i=1}^{2}(1-\omega_i^{(m)})x_i^{(m)}(\omega^{(m)}) \leq Ne$ para cada m e cada $\omega^{(m)}$. Quando $M \to \infty$, a utilidade (12) e a restrição (11) convergem[14] para, respectivamente,

$$p^2\left[u(c_1)+u(c_2)\right] + p(1-p)\left[u(c_1)+u(y)\right] + p(1-p)\left[u(y)+u(\bar{c})\right] + (1-p)^2\left[2u(y)\right]$$

e

$$p(1-p)[R(Ne-c_1)-y] + p(1-p)[R(Ne-\bar{c})-y] + (1-p)^2[RNe-2y] = 0, \tag{13}$$

em que para todo $m \in \mathbb{N}$ a variável c_1 denota $x_1^{(m)}(0,0)$, c_2 denota $x_2^{(m)}(0,0)$, $\bar{c}$ denota $x_2^{(m)}(1,0)$ e y denota o pagamento $y^{(m)}(\tilde{\omega})$, o qual deixa de ser aleatório quando $M \to \infty$. As demais restrições são reescritas como $c_1 + c_2 = Ne$, $c_1 \leq Ne$ e $\bar{c} \leq Ne$.

[11] Ver notas de rodapé 7 e 8.

[12] Ennis e Keister (2009) e Bertolai et alii (2014) propõem algoritmos para calcular a alocação ótima no caso geral $N \in \mathbb{N}$ mantendo a hipótese de função utilidade $u(c) = \frac{c^{1-\delta}}{1-\delta}$.

[13] Definida antes da distribuição (aleatória) dos indivíduos entre os bancos.

[14] Conforme demonstrado em Bertolai et alii (2016).

***Lema* 2.4.** *O sistema bancário ótimo com um contínuo de bancos é definido por* $(c_1,\bar{c},y)$ *tal que* $u'(\bar{c}) = Ru'(y)$

$$\begin{aligned} u'(c_1) &= (1-p)Ru'(y) + pu'(Ne - c_1) \\ y &= \frac{R}{N}[p(Ne-c_1) + p(Ne-\bar{c}) + (1-p)Ne]. \end{aligned}$$

Demonstração. Ver apêndice A. □

Observe a relação entre o pagamento na data 2 e o pagamento na data 1 após anúncio de tipo 1 na primeira posição, $u'(\bar{c}) = Ru'(y)$. Bertolai et alii (2016) mostram que tal resultado pode ser generalizado[15] para o caso $N > 2$, de forma que há tratamento igualitário na data 1 logo após algum anúncio de tipo 1 na data 1. Há tratamento desigual na data 1 somente enquanto não surgir um anúncio tipo 1. Trata-se da mesma relação entre o pagamento na data 2 e o pagamento na data 1 escolhida no caso sem serviço sequencial, conforme estabelecido no Lema 2.1 e no Corolário 2.2.

3. INSTABILIDADE FINANCEIRA *SEM* SERVIÇO SEQUENCIAL

Considere o modelo de Allen e Gale (2000), no qual *não há* a exigência de serviço sequencial na escolha de sistema bancário. Conforme discutido, este modelo pode ser visto como um caso limite da economia apresentada na subseção 2.1 quando $N \to \infty$.

As duas datas da seção 2 são divididas em três períodos $t \in \{0,1,2\}$, de forma que o período 0 corresponde ao início da data 1. A tecnologia de armazenamento é apresentada como alternativas de investimento em ativos de curto e de longo prazo. Cada unidade de bem de consumo investida em ativo de curto prazo no período t produz 1 unidade do bem de consumo em $t+1$. Cada unidade de bem de consumo investida em ativo de longo prazo em $t = 0$ produz $R > 1$ unidades de consumo em $t = 2$ e $r \in (0,1)$ de unidades de consumo em $t = 1$.

A economia é dividida em quatro regiões *ex ante* idênticas: regiões A, B, C e D. Cada região possui um contínuo de indivíduos *ex ante* idênticos, de medida $1/4$. Com probabilidade $p \in (0,1)$ o indivíduo sofre o choque de liquidez e se torna impaciente (tipo 0). Com probabilidade $1-p$ ele não sofre o choque e se torna paciente (tipo 1). A probabilidade p varia entre regiões, sendo p_i a probabilidade de um indivíduo se tornar impaciente na região $i \in \{A,B,C,D\}$. Ela é uma variável aleatória com dois valores possíveis, p_H e p_L tais que $0 < p_L < p_H < 1$. Há dois estados possíveis e equiprováveis: S_1 e S_2 com $\Pr(S_1) = \Pr(S_2) = 1/2$. No estado agregado S_1, tem-se $p_A = p_C = p_H$ e $p_B = p_D = p_L$ e no estado agregado S_2, tem-se $p_A = p_C = p_L$ e $p_B = p_D = p_H$. A Tabela 1 organiza tal informação de maneira transparente.

Tabela 1: Choques de liquidez por região

Est/Reg	A	B	C	D
S_1	p_H	p_L	p_H	p_L
S_2	p_L	p_H	p_L	p_H

[15] Desde que N não seja muito grande ou R seja suficientemente baixo.

Como existe um contínuo de indivíduos na região $i \in \{A,B,C,D\}$, a Lei dos Grandes Números garante que a proporção de indivíduos impacientes nesta região será p_i.[16] Logo, a quantidade de impacientes na região i é $p_i/4$. Conclui-se então que a quantidade de impacientes na economia é

$$\gamma \equiv \sum_{i\in\{A,B,C,D\}} \frac{p_i}{4} = \frac{p_H}{4} + \frac{p_L}{4} + \frac{p_H}{4} + \frac{p_L}{4} = \frac{p_H + p_L}{2}$$

nos dois estados da natureza S_1 e S_2. Toda a incerteza é resolvida em $t = 1$ quando o estado da natureza (S_1 ou S_2) é revelado e cada consumidor descobre seu tipo.

Conforme antecipado na subseção 2.1, o problema de escolha de sistema bancário nesta economia é[17]

$$\max_{c_1,c_2}\Big\{\gamma u(c_1) + (1-\gamma)u(c_2)\Big\} \qquad \text{sujeito a} \quad \begin{cases} c_1 \geq 0 \quad \text{e} \quad c_2 \geq 0 \\ \gamma c_1 + (1-\gamma)R^{-1}c_2 \leq 1 \end{cases} \tag{14}$$

em que c_t denota o pagamento do sistema bancário no período $t \in \{1,2\}$. A simplicidade do problema acima decorre de três propriedade da alocação ótima: (i) os pagamentos não dependem do estado da natureza, dado que não existe incerteza agregada; (ii) os pagamentos não dependem da região do indivíduo, dado que os indivíduos são *ex ante* idênticos e a utilidade é estritamente côncava; (iii) o investimento no ativo de curto prazo é igual a quantidade total de pagamentos no períodos 1, γc_1; o investimento no ativo de longo prazo é igual a quantidade total de pagamentos no períodos 2, $(1-\gamma)c_2$.

Seja $a = (s,l,c_1,c_2)$ uma alocação nesta economia, em que s é o investimento no ativo de curto prazo e l é o investimento no ativo de longo prazo. Seja $a^* = (s^*,l^*,c_1^*,c_2^*)$ a solução de (14). Então, $s^* = \gamma c_1^*$ e $l^* = (1-\gamma)c_2^*/R$. Além disso, o Lema 2.1 (ao tomar o limite $N \to \infty$) garante que $u'(c_1^*) = Ru'(c_2^*)$. Dada a concavidade estrita de u e $R > 1$, tem-se $c_1^* < c_2^*$ e, portanto, a restrição de incentivos $u(c_2) \geq u(c_1)$ é atendida por a^*. Os indivíduos pacientes acham fracamente melhor revelar seu verdadeiro tipo ao invés de mentir e consumir em $t = 1$.

3.1. Descentralização (*O Mercado Interbancário de Depósitos*)

Allen e Gale (2000) mostram que a alocação ótima a^* pode ser descentralizada por meio de um sistema bancário competitivo. Ou seja, por meio de um grande número de bancos, cada região paga aos seus indivíduos c_1^* na data 1 e c_2^* na data 2.

O mercado interbancário (integração entre os bancos) é usado para transferir recursos entre as diferentes regiões. Há inicialmente $1/4$ unidades de recursos em cada região e a demanda por recursos no período 1 será $p_H c_1^*/4$ nas regiões com muitos impacientes e $p_L c_1^*/4$ nas regiões com poucos impacientes. Cada região investe em ativo de curto prazo $s^*/4 = \gamma c_1^*/4$ e em ativo de longo prazo e $l^*/4 = (1-\gamma)c_2^*/4R$, de forma que no agregado e economia investe s^* no curto prazo e l^* no longo prazo. Como $p_L < \gamma < p_H$, as regiões com muitos impacientes possuem excesso de demanda por consumo no período 1, dado por $(p_H - \gamma)c_1^*/4$, e excesso de oferta de recursos no período 2, dado por $(1 - p_H - \gamma)c_2^*/4$. Similarmente, as regiões com poucos impacientes possuem excesso de recursos no período 1, dado por $(\gamma - p_L)c_1^*/4$ e excesso de demanda por recursos no período 2, dado por $(1 - p_L - \gamma)c_2^*/4$. O mercado interbancário opera em ambos os períodos para transferir recursos das regiões com excesso de oferta de recursos para as regiões com excesso de demanda por recursos. Tais

[16] Por exemplo, no estado S_1, a região A tem uma proporção p_H de indivíduos impacientes (assim como a região C) e a região B tem uma proporção p_L de indivíduos impacientes (assim como a região D).

[17] A desigualdade $\gamma c_1 + (1-\gamma)R^{-1}c_2 \leq 1$ exige que a soma entre a quantidade de ativos de curto γc_1 e de longo prazo $(1-\gamma)c_2$ não ultrapasse a dotação da economia: 1 unidade de bem de consumo. Como existe um contínuo de agentes e cada agente tem uma unidade do bem, no agregado há uma unidade de recurso.

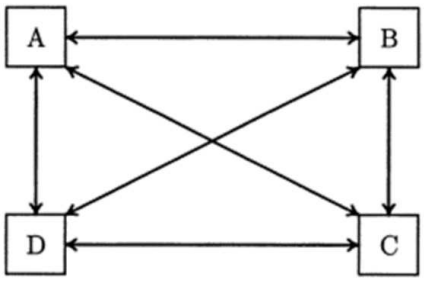
transferências permitem honrar as promessas de pagamento c_1^* e c_2^* em cada uma das regiões sem a necessidade de resgatar o ativo de longo prazo.

A alocação descentralizada é descrita em termos do comportamento de um banco representativo em cada região $i \in \{A,B,C,D\}$. Cada consumidor deposita sua 1 unidade de bem no banco da sua região, o qual investe em um portfólio $(s^i,l^i) \geq 0$ e oferece em troca um contrato (c_1^i,c_2^i) que permite ao consumidor retirar c_1^i unidades de consumo em $t = 1$ ou c_2^i unidades de consumo em $t = 2$. O contrato é suposto invariante aos choques de liquidez das regiões. Para implementar a alocação ótima por meio do setor bancário descentralizado, deve-se ter $a^i = (s^i, l^i, c_1^i, c_2^i) = (s^*/4, l^*/4, c_1^*, c_2^*)$.

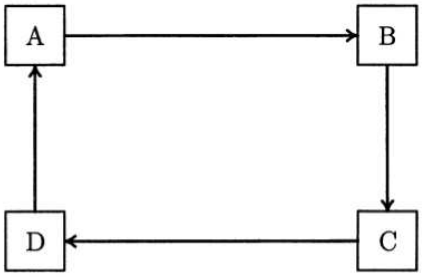
No mercado interbancário, os bancos podem depositar parte de sua dotação inicial em (transferir em $t = 0$ para) bancos de outras regiões. Se o banco receptor está na região i, ele promete pagar ao banco depositante c_1^i se o depósito for resgatado em $t = 1$ e c_2^i se o depósito for resgatado em $t = 2$. Allen e Gale (2000) estudam dois tipos de mercados interbancários, os quais são apresentados na Figura 1. A flecha aponta para o banco que recebe o depósito e tem origem no banco que deposita. O mercado interbancário apresentado na subfigura 1a é chamado de *completo*, pois cada banco pode depositar em todos os demais bancos. O mercado interbancário apresentado na subfigura 1b é chamado de *incompleto*, pois o banco de cada região pode depositar em somente uma região e receber depósito de somente uma outra região. Por exemplo, o banco da região B pode depositar recursos somente no banco da região C e receber recursos somente da região A.

Figura 1: Tipos de mercado interbancário em Allen e Gale (2000).

(a) Integração completa



(b) Integração incompleta



***Lema* 3.1** (Descentralização)**.** *É factível pagar em todas as regiões c_1^* no período $t = 1$ e c_2^* no período $t = 2$, tanto no caso de mercado interbancário completo quanto no caso de mercado interbancário incompleto.*

- *Se o mercado é completo, basta o banco da região $i \in \{A,B,C,D\}$ manter $z^i = (p_H - \gamma)/8$ depósitos em cada região $j \neq i$.*
- *Se o mercado é incompleto, basta o banco da região $i \in \{A,B,C,D\}$ manter $z^i = (p_H - \gamma)/4$ depósitos na região permitida.*

Demonstração. Ver apêndice A. □

Nas duas estruturas de mercado consideradas, as regiões com alta demanda por consumo em $t = 1$ resgatam seus depósitos no mercado interbancário em $t = 1$ e pagam os depósitos obtidos no interbancário em $t = 2$. As regiões com baixa demanda por consumo em $t = 1$ resgatam seus depósitos em $t = 2$ e e pagam os depósitos obtidos em $t = 1$.

3.2. Contágio

Considere agora a existência de um estado da natureza imprevisível $\bar{S}$. Neste estado, a proporção de impaciente é igual a γ nas regiões B, C e D e igual a $\gamma + \epsilon$ na região A, para $\epsilon \in (0{,}1 - \gamma)$, conforme apresentado na Tabela 2.

Tabela 2: Choques de liquidez por região com pertubação

Est/Reg	A	B	C	D
S_1	p_H	p_L	p_H	p_L
S_2	p_L	p_H	p_L	p_H
$\bar{S}$	$\gamma + \epsilon$	γ	γ	γ

A imprevisibilidade de $\bar{S}$ significa que sua realização ocorre com probabilidade zero. Com isso, o sistema bancário ótimo não muda e continua sendo dado por a^*. Nos estados S_1 e S_2 o comportamento de equilíbrio da economia continua o mesmo. No estado $\bar{S}$, contudo, o comportamento será diferente.

Suponha realizado o estado $\bar{S}$. Indivíduos e bancos irão decidir se retiram seus depósitos em $t = 1$ ou $t = 2$. Os indivíduos impacientes sempre sacam em $t = 1$, uma vez que não auferem utilidade em consumir em $t = 2$. Os consumidores pacientes tem a opção de sacar em $t = 1$ ou em $t = 2$ e vão optar pelo período que lhe fornecer maior quantidade de consumo. Para o comportamente dos bancos em $\bar{S}$, Allen e Gale (2000) fazem as seguintes hipóteses: os bancos são obrigados a pagar c_1^* em $t = 1$, sempre que factível, e os ativos remanescentes em $t = 2$ são liquidados e distribuídos igualmente entre os depositantes (indivíduos e bancos) que decidem retirar em $t = 2$. Se não é possível pagar c_1^* para todos que desejam sacar em $t = 1$, os bancos são obrigados a liquidar todos os seus ativos em $t = 1$ e distribuir os recursos assim obtidos igualmente entre seus depositantes (indivíduos e bancos).

A ordem individualmente ótima na qual os bancos liquidam seus ativos é chamada por Allen e Gale (2000) de *pecking order*. Primeiramente, os bancos liquidam o ativo de curto prazo. Em seguida, liquidam os depósitos interbancários. Por fim, liquidam o ativo de longo prazo. A otimalidade deste ordenamento decorre dos diferentes custos de liquidação (custo de se obter consumo corrente (em $t = 1$) em termos de consumo futuro (em $t = 2$)) de cada ativo:[18]

$$1 < \frac{c_2^*}{c_1^*} < \frac{R}{r}$$

em que $c_1^* < c_2^*$ decorre da condição de otimalidade $u'(c_1^*) = Ru'(c_2^*)$ e r pode ser escolhido suficientemente pequeno para que $R/r > c_2^*/c_1^*$ seja satisfeita, uma vez que a alocação ótima é invariante a r.

Considere daqui em diante a principal estrutura de mercado estudada por Allen e Gale (2000), a estrutura incompleta apresentada na Figura 1b. Suponha que o banco da região i depositou $z^i = (p_H - \gamma)/4$ na única região em que ele é capaz de fazê-lo.[19] Ao se deparar com $(\gamma + \epsilon)/4$ saques em $t = 1$ (cada um deles de valor prometido c_1^*), o banco da região A irá demandar seus depósitos junto ao banco B, uma vez que seus ativos de curto prazo $s^*/4 = \gamma c_1^*/4$ são insuficientes para honrar seus compromissos em $t = 1$. Com isso, o banco da região B demanda seus depósitos junto ao banco C, pois o banco B não é capaz de pagar a demanda dos seus $\gamma/4$ depositantes e do banco A, z^A, utilizando somente sua dotação de ativo de curto prazo, $s^*/4 = \gamma c_1^*/4$. De forma análoga, o banco da região C

[18] Uma unidade de ativo de curto prazo vale 1 unidade de consumo hoje e 1 unidade de consumo amanhã (custo 1). Ao liquidar 1 unidade de depósito o banco desiste de c_2^* unidades de consumo futuro e obtém c_1^* unidades de consumo presente (custo c_2^*/c_1^*). Se o banco liquida 1 unidade de ativo de longo prazo, ele troca R unidades de consumo futuro por r unidades de consumo presente (custo R/r).

[19] A depositou em B, B depositou em C, C depositou em D e D depositou em A.

demanda seus depósitos junto ao banco D e o banco da região D demanda seus depósitos junto ao banco A. Como os depósitos entre os bancos tem o mesmo valor, o mercado interbancário não é capaz de auxiliar o banco A a satisfazer suas obrigações sem recorrer aos ativos de longo prazo. O saque do banco A junto ao banco B é usado para pagar o saque do banco D junto ao banco A.

Do exposto acima, o banco da região A necessita liquidar também ativos de longo prazo para honrar seus compromissos em $t = 1$. Logo, ele não será capaz de honrar o compromisso c_2^* junto aos seus depositantes em $t = 2$. Se o banco A liquidar x unidades de ativo de longo prazo, ele pagará na data 2 somente $c_2^A = 4R[l^*/4 - x]/(1 - \gamma - \epsilon)$. Os indivíduos pacientes continuarão a sacar no período 2 se e somente se $c_2^A \geq c_1^*$, ou seja, se e somente se $x \leq [l^* - (1 - \gamma - \epsilon)c_1^*/R]/4$. Caso contrário, eles optarão por sacar no período 1, o que caracteriza[20] uma *corrida bancária* na região A. Como a região A possui $s^*/4 = \gamma c_1^*/4$ unidades de ativo de curto prazo e $(\gamma + \epsilon)/4$ indivíduos impacientes, então o banco A precisa obter $\epsilon c_1^*/4$ unidades de consumo por meio da liquidação do ativo de longo prazo. Logo, o banco da região A liquida $x = \epsilon \frac{c_1^*}{4}$ unidades de ativo de longo prazo.

***Lema* 3.2.** *Seja $\beta^* = \frac{r}{R-r}\frac{c_2^* - c_1^*}{c_1^*}$ e suponha realizado o estado $\bar{S}$. Então, existe corrida bancária na região A se e somente se*

$$\beta^*(1 - \gamma) < \epsilon. \tag{15}$$

Demonstração. Ao liquidar $x = \epsilon c_1^*/4r$ unidades de ativo de longo prazo, o banco A é capaz de pagar no período 2

$$c_2^A = \frac{4R[l^*/4 - \epsilon c_1^*/4r]}{1 - \gamma - \epsilon} = \frac{(1 - \gamma - \epsilon)c_2^* + \epsilon c_2^* - R\epsilon c_1^*/r}{1 - \gamma - \epsilon} = c_2^* + \frac{\epsilon}{1 - \gamma - \epsilon}\left(\frac{c_2^*}{c_1^*} - \frac{R}{r}\right)c_1^*.$$

Por hipótese, $R/r > c_2^*/c_1^*$. Então $c_2^A < c_2^*$ para todo $\epsilon \in (0{,}1 - \gamma)$ e c_2^A é descrescente em ϵ. Logo, $c_2^A \geq c_1^*$ se e somente se $\frac{c_2^* - c_1^*}{c_1^*} + \frac{\epsilon}{1-\gamma-\epsilon}\left(\frac{c_2^*}{c_1^*} - \frac{R}{r}\right) \geq 0$. Ou seja,

$$\beta^*\frac{R - r}{r} = \frac{c_2^* - c_1^*}{c_1^*} \geq \frac{\epsilon}{1 - \gamma - \epsilon}\left(\frac{R}{r} - \frac{c_2^*}{c_1^*}\right) = \frac{\epsilon}{1 - \gamma - \epsilon}\left(\frac{R}{r} - \beta^*\frac{R - r}{r} - 1\right)$$

$$(1 - \gamma - \epsilon)\beta^*\frac{R - r}{r} \geq \epsilon\,(1 - \beta^*)\,\frac{R - r}{r}$$

o que é equivalente a $\epsilon \leq (1 - \gamma)\beta^*$. □

Se $\epsilon \leq \beta^*(1 - \gamma)$, não há corrida bancária na região A e o banco desta região é capaz de absorver o choque ϵ por conta própria. O choque não tem repercussão sobre bancos em outras regiões. Os indivíduos pacientes da região A, no entanto, estão em pior situação: a prematura liquidação do ativo de longo prazo não permite ao bancos pagar c_2^* em $t = 2$.

Se $\epsilon > \beta^*(1 - \gamma)$, haverá corrida bancária. Todos os $1/4$ indivíduos da região A e o banco D demandarão seus depósitos em $t = 1$. O valor de liquidação dos ativos do banco A é $s^*/4 + rl^*/4 + z^A q^B$, em que q^i é o valor do depósito no banco $i \in \{A,B,C,D\}$ quando há corrida bancária no banco A. Como $q^B \leq c_1^*$, o banco A não é capaz de honrar a promessa $q^A = c_1^*$, pois

$$\frac{s^* + rl^*}{4} + z^A q^B \leq \frac{\gamma c_1^* + r(1 - \gamma)c_2^*/R}{4} + z^D c_1^* < \left(\frac{1}{4} + z^D\right)c_1^*$$

em que a desigualdade estrita decorre de $R/r > c_2^*/c_1^*$. Logo, $q^A = \frac{s^* + rl^* + 4z^A q^B}{1 + 4z^D} < c_1^*$ e todos os indivíduos da região A serão prejudicados pelo choque ϵ, assim como haverá um *spillover effect* negativo para a região D – uma vez que seus depósitos no interbancário perdem valor.

[20] Conforme discutido na seção 2.

O *spillover effect* sobre o banco da região D reduz o valor de seus ativos, de forma que os ativos de curto prazo $s^*/4$ e o depósito no interbancário $z^D q^A$ são insuficientes para honrar os pagamentos em $t=1$, $(\gamma/4+z^C)c_1^*$, pois $q^A < c_1^*$ implica em $s^*/4+z^D q^A < (\gamma/4+z^C)c_1^*$. Logo, o banco da região D também necessita liquidar ativos de longo prazo para honrar seus compromissos em $t=1$ e, portanto, ele não será capaz de honrar o compromisso c_2^* junto aos seus depositantes em $t=2$. Se o banco D liquidar x unidades de ativo de longo prazo, ele pagará na data 2 somente $c_2^D = 4R[l^*/4 - x]/(1-\gamma)$. Os indivíduos pacientes continuarão a sacar no período 2 se e somente se $c_2^D \geq c_1^*$, ou seja, se e somente se $x \leq [l^* - (1-\gamma)c_1^*/R]/4$. Caso contrário, eles optarão por sacar no período 1, o que caracteriza uma *corrida bancária* na região D. Então o banco D precisa obter $(\gamma/4 + z^C)c_1^* - (s^*/4 + z^D q^A) = z^D(c_1^* - q^A) > 0$ unidades de consumo por meio da liquidação do ativo de longo prazo. Logo, o banco da região D liquida $x = z^D(c_1^* - q^A)/r$ unidades de ativo de longo prazo.

Analogamente, o banco da região C liquida $x = z^C(c_1^* - q^D)/r$ unidades de ativo de longo prazo e o banco da região B liquida $x = z^B(c_1^* - q^C)/r$ unidades de ativo de longo prazo.

***Lema* 3.3.** *Seja $z = p_H - \gamma$. A corrida bancária no banco A gera uma corrida bancária via contágio em todas as regiões se $\beta^* < z$, ou seja*

$$\frac{r}{R} < \frac{zc_1^*}{zc_2^* + (1+z)(c_2^* - c_1^*)}. \tag{16}$$

Neste caso, os valores de liquidação são $q^A = q^B = q^C = q^D = s^* + rl^*$.

Demonstração. Seja $Y \equiv c_1^* - \frac{r(1-\gamma)}{Rz}(c_2^* - c_1^*)$. Ao liquidar $x = z^D(c_1^* - q^A)/r$ unidades de ativo de longo prazo, o banco D é capaz de pagar no período 2

$$c_2^D = \frac{4R[l^*/4 - z^D(c_1^* - q^A)/r]}{1-\gamma} = c_2^* - \frac{Rz}{r(1-\gamma)}(c_1^* - q^A),$$

pois $4z^i = z$ para todo $i \in \{A,B,C,D\}$. Então, $c_2^D \geq c_1^*$ se e somente se $c_2^* - \frac{Rz}{r(1-\gamma)}(c_1^* - q^A) - c_1^* \geq 0$. Ou seja, $q^A \geq Y$. De forma análoga, $c_2^C \geq c_1^*$ se e somente se $q^D \geq Y$. Ainda, $c_2^B \geq c_1^*$ se e somente se $q^C \geq Y$. Note que $q^B \leq c_1^*$ e, portanto, $q^A = \frac{s^* + rl^* + 4z^A q^B}{1+4z^A} \leq \frac{s^* + rl^* + zc_1^*}{1+z}$. Logo,

$$\begin{aligned}
Y - q^A &\geq Y - \frac{\gamma c_1^* + (r/R)(1-\gamma)c_2^* + zc_1^*}{1+z} \\
&= c_1^* - \frac{r(1-\gamma)}{Rz}(c_2^* - c_1^*) - \frac{\gamma c_1^* + (r/R)(1-\gamma)c_2^* + zc_1^* \pm c_1^*}{1+z} \\
&= -\frac{r(1-\gamma)}{Rz}(c_2^* - c_1^*) - \frac{(r/R)(1-\gamma)c_2^* - (1-\gamma)c_1^*}{1+z} \\
&= (1-\gamma)\left[\frac{c_1^*}{1+z} - \frac{r}{R}\left(\frac{c_2^* - c_1^*}{z} + \frac{c_2^*}{1+z}\right)\right] \\
&= (1-\gamma)\left[\frac{c_1^*}{1+z} - \frac{r}{R}\left(\frac{(1+z)(c_2^* - c_1^*) + zc_2^*}{z(1+z)}\right)\right] > (1-\gamma)\left[\frac{c_1^*}{1+z} - \frac{c_1^*}{1+z}\right] = 0,
\end{aligned} \tag{17}$$

em que se usou (16) na última desigualdade.

Segue que ocorre corrida bancária na região D. O valor de liquidação no banco D é $q^D = \frac{s^* + rl^* + 4z^D q^A}{1+4z^D} \leq \frac{s^* + rl^* + zc_1^*}{1+z}$, pois $q^A \leq c_1^*$. Álgebra análoga a (17) estabelece que $q^D < Y$ e, portanto, há corrida na região C. O valor de liquidação no banco C é $q^C = \frac{s^* + rl^* + 4z^C q^D}{1+4z^C} \leq \frac{s^* + rl^* + zc_1^*}{1+z}$, pois $q^D \leq c_1^*$. Álgebra análoga a (17) estabelece que $q^C < Y$ e, portanto, há corrida na região B.

Como há corrida bancária em todas as regiões, então $q^i = X + Zq^j$ para todo $(i,j) \in \{(A,B),(B,C),(C,D),(D,A)\}$, em que $X \equiv \frac{s^* + rl^*}{1+z}$ e $Z \equiv \frac{z}{1+z} \in (0,1)$. Logo,

$$\begin{pmatrix} 1 & -Z & 0 & 0 \\ 0 & 1 & -Z & 0 \\ 0 & 0 & 1 & -Z \\ -Z & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} q^A \\ q^B \\ q^C \\ q^D \end{pmatrix} = \begin{pmatrix} X \\ X \\ X \\ X \end{pmatrix}.$$

Observe que $q^A = q^B = q^C = q^D = X/(1-Z) = s^* + rl^*$ é a única solução dos sistema acima, pois o determinante da matriz de coeficientes é $1 - Z^4 > 0$.

□

3.3. Uma intervenção no interbancário para eliminar contágio

Considere uma intervenção no mercado interbancário no estado $\bar{S}$ sob a estrutura de mercado incompleta apresentada na Figura 1b,[21] cujo objetivo é evitar a corrida bancária em todos os bancos. Suponha que (15) e (16) são válidas e, portanto, há corrida bancária em todos os bancos na ausência de intervenção, conforme estalecido pelos Lemas 3.2 e 3.3.

Uma autoridade central impõe o seguinte mecanismo de coordenação aos bancos no estado $\bar{S}$: (i) todos os bancos liquidam somente $(1/4)\epsilon c_1^*/4r$ unidades de ativo de longo prazo; (ii) o banco A liquida somente $(3/4)\epsilon c_1^*/4$ unidades em depósitos junto ao banco B; (iii) o banco B liquida somente $(2/4)\epsilon c_1^*/4$ unidades em depósitos junto ao banco C; (iv) o banco C liquida somente $(1/4)\epsilon c_1^*/4$ unidades em depósitos junto ao banco D; e (iv) o banco D não resgata seus depósitos junto ao banco A em $t = 1$.

O saque do banco i não pode exceder $z^i c_1^*$, ou seja, $\epsilon \leq 16z^i/3 = 4z/3$. Adicionalmente, o banco A será capaz de cumprir sua promessa de pagamento em $t = 1$ se

$$\frac{1}{4}(\gamma + \epsilon)c_1^* + 0 \leq \frac{s^*}{4} + \frac{1}{4}\frac{\epsilon c_1^*}{4} + \frac{3}{4}\frac{\epsilon c_1^*}{4}$$

em que o termo nulo representa o saque realizado pelo banco D e o termo $\epsilon c_1^*/16 = r\epsilon c_1^*/(16r)$ é o montante de recursos obtido pelo banco A por meio da liquidação de ativo de longo prazo. Analogamente, os bancos B, C e D serão capazes de cumprir suas promessas de pagamento em $t = 1$ se, respectivamente,

$$\begin{aligned} \frac{\gamma}{4}c_1^* + \frac{3}{4}\frac{\epsilon c_1^*}{4} &\leq \frac{s^*}{4} + \frac{1}{4}\frac{\epsilon c_1^*}{4} + \frac{2}{4}\frac{\epsilon c_1^*}{4} \\ \frac{\gamma}{4}c_1^* + \frac{2}{4}\frac{\epsilon c_1^*}{4} &\leq \frac{s^*}{4} + \frac{1}{4}\frac{\epsilon c_1^*}{4} + \frac{1}{4}\frac{\epsilon c_1^*}{4} \\ \frac{\gamma}{4}c_1^* + \frac{1}{4}\frac{\epsilon c_1^*}{4} &\leq \frac{s^*}{4} + \frac{1}{4}\frac{\epsilon c_1^*}{4} + 0 \end{aligned}$$

em que o termo nulo representa o saque realizado pelo banco D no banco A. Como $s^* = \gamma c_1^*$, então as quatro condições acima são satisfeitas com igualdade.

Sob a intervenção descrita, todos os bancos conseguem pagar c_1^* aos impacientes e aos demais bancos e o custo do choque ϵ é absorvido pelos indivíduos pacientes de todas as regiões. Os pacientes recebem $c_2^i < c_2^*$ para todo $i \in \{A,B,C,D\}$, conforme estabelecido pelo Lema 3.4 a seguir.

***Lema* 3.4.** *Sob o mecanismo de intervenção, o pagamento aos pacientes na data* 2 *é dado por*

$$c_2^A = c_2^B = c_2^C = c_2^D = c_2^* - \frac{\epsilon c_1^*}{4}\frac{R/r - c_2^*/c_1^*}{1 - \gamma - \epsilon/4} < c_2^*.$$

[21] A análise é restrita a esta estrutura de mercado por ser este o principal caso estudado por Allen e Gale (2000) e ser aquele no qual o contágio possui efeitos mais profundos e disseminados.

Demonstração. Ver apêndice A. □

Os indivíduos pacientes continuarão a sacar no período 2 se e somente se $c_2^i \geq c_1^*$ para todo $i \in \{A,B,C,D\}$. Portanto, para que elimine a corrida bancária, a intervenção precisa satisfazer:

$$c_2^* - c_1^* \geq \frac{\epsilon c_1^*}{4} \frac{R/r - c_2^*/c_1^*}{1-\gamma-\epsilon/4} = c_1^* \frac{R/r - c_2^*/c_1^*}{4(1-\gamma)/\epsilon - 1}$$

$$(4(1-\gamma)/\epsilon - 1)\, \frac{c_2^* - c_1^*}{c_1^*} \geq R/r - c_2^*/c_1^*$$

$$4(1-\gamma)/\epsilon \frac{c_2^* - c_1^*}{c_1^*} \geq \frac{R}{r} - \frac{c_2^*}{c_1^*} + \frac{c_2^* - c_1^*}{c_1^*} = \frac{R-r}{r}$$

$$\epsilon \leq 4\beta^*(1-\gamma),$$

em que se usou $\beta^* = \frac{r}{R-r}\frac{c_2^*-c_1^*}{c_1^*}$. De (15), conclui-se ser necessário que $\epsilon \in (\beta^*(1-\gamma), 4\beta^*(1-\gamma)]$ para que exista corrida bancária no banco A sem intervenção e não ocorra corrida bancária sob a intervenção em banco algum. A proposição (3.5) sumariza tal resultado.

***Proposição* 3.5.** *Existe (ϵ,r) sob o qual há corrida bancária em todas as regiões na ausência de intervenção e não há corrida bancária na presença da intervenção descrita acima. Basta que $\beta^* < z$, $\epsilon \leq \frac{4z}{3}$ e*

$$\beta^*(1-\gamma) < \epsilon \leq 4\beta^*(1-\gamma).$$

4. INSTABILIDADE FINANCEIRA *COM* SERVIÇO SEQUENCIAL

Considere o modelo de Bertolai et alii (2016), no qual *há* a exigência de serviço sequencial na escolha de sistema bancário.[22] Conforme discutido, o caso a ser estudado nesta seção (no qual há um contínuo de bancos) pode ser visto como um caso limite da economia apresentada na subseção 2.2 quando o número de bancos é muito grande ($M \to \infty$) e os bancos se integram parcialmente, compartilhando as reservas bancárias em $t = 2$.

As preferências dos indivíduos são levemente diferentes daquelas apresentadas na seção 2, embora os principais resultados descritos naquela seção continuem válidos. Como em Diamond e Dybvig (1983), as preferências são representadas por uma função utilidade $u(c_1)$ se o indivíduo sofre o choque de preferência e $u(c_1+c_2)$ caso contrário. A função $u : \mathbb{R} \to [-\infty, \infty)$ é duas vezes diferenciável, estritamente crescente, côncava, tal que $\lim_{x\to 0} u'(x) = \infty$ e a aversão ao risco relativa em e é maior que 1:[23] $\delta \equiv -\dfrac{eu''(e)}{u'(e)} > 1$. Como consequência do conceito de integração parcial utilizado, as transferências em $t = 1$ em um dado banco não dependem de anúncios feitos em outros bancos. No entanto, o pagamento em $t = 2$ depende do que acontece nos demais bancos.

Considere inicialmente o caso já apresentado na subseção 2.2, no qual há dois depositantes ($N = 2$) em cada banco. O problema de escolha de sistema bancário (x,y) consiste em maximizar

$$p^2\Big[u(c_1)+u(c_2)\Big] + p(1-p)\Big[u(c_1)+u(y)\Big] + p(1-p)\Big[u(y)+u(\bar{c})\Big] + (1-p)^2\Big[2u(y)\Big]$$

sujeito às restrições $c_1 + c_2 = Ne$ e

$$p(1-p)[R(Ne-c_1)-y] + p(1-p)[R(Ne-\bar{c})-y] + (1-p)^2[RNe-2y] = 0, \qquad (18)$$

[22] A fim de facilitar a comparação com a apresentação em Bertolai et alii (2016), será usada nesta seção a notação original destes autores. Para facilitar a comparação com a apresentação da seção 2, a notação será relacionada sempre que necessário.

[23] Esta função tem como caso particular a função utilizada na seção 2: $u(c) = c^{1-\delta}/(1-\delta)$. Tal função possui aversão ao risco relativa constante e igual a δ em todo domínio, não só no ponto $c = e$.

em que para todo banco $m \in [0,1]$ as variáveis c_1, c_2, $\bar{c}$ e y denotam, respectivamente, os pagamentos $x_1^{(m)}(0,0)$, $x_2^{(m)}(0,0)$, $x_2^{(m)}(1,0)$ e $y^{(m)}(\tilde{\omega})$. Bertolai et alii (2016) generalizam o Lema 2.4 para funções de utilidade mais gerais e mostram que a solução ótima satisfaz as seguintes condições de primeira ordem:

$$u'(\bar{c}) = Ru'(y) \quad (19)$$

$$u'(c_1) = (1-p)Ru'(y) + pu'(Ne - c_1) \quad (20)$$

$$y = \frac{R}{N}[p(Ne - c_1) + p(Ne - \bar{c}) + (1-p)Ne]. \quad (21)$$

A existência de corrida bancária nesta economia depende do incentivo dos indivíduos pacientes em revelar a verdade quando estes acreditam que outros pacientes não estão revelando a verdade. Suponha que todos os indivíduos acreditam que os indivíduos na primeira posição estão anunciando tipo 0 e que todos os indivíduos na segunda posição estão revelando a verdade. Com isso, todos acreditam que p bancos estão recebendo anúncios $\omega = (0,0)$ e os demais $1-p$ bancos estão recebendo $\omega = (0,1)$. Logo, os indivíduos não acreditam que o pagamento em $t = 2$ será determinado pela restrição (18). Ele será dado por

$$(1-p)[R(Ne - c_1) - y] = 0, \quad (22)$$

já que os bancos com história $(0,0)$ exaurem suas reservas em $t = 1$, pois $c_1 + c_2 = Ne$, e os bancos com história $(0,1)$ pagam c_1 em $t = 1$ e poupam $Ne - c_1$ para $t = 2$.

O incentivo para revelar ou não a verdade depende do tipo do agente e de sua posição na fila de saques. Para os indivíduos impacientes a decisão é trivial: eles sempre revelam a verdade (tipo 0), uma vez que não auferem utilidade consumindo na segunda data. Para os indivíduos pacientes, a decisão é mais sofisticada. Considere inicialmente um indivíduo paciente na segunda posição da fila de saques. Ao mentir sobre seu tipo (anunciar tipo 0), ele espera receber $Ne - c_1$, pois acredita que o primeiro indivíduo em sua fila revelou ser do tipo 0 e consumiu c_1. Caso revele a verdade, ele espera receber $R(Ne - c_1)$, conforme implicado por (22). Como $R > 1$, tal indivíduo acha melhor revelar a verdade. Considere agora um indivíduo paciente na primeira posição da fila de saques. Ao mentir sobre seu tipo (anunciar tipo 0), ele espera receber c_1 e, caso revele a verdade, ele espera receber $R(Ne - c_1)$. Logo, mentir sobre seu tipo (se juntar à corrida bancária na primeira posição) é atrativo se

$$F(R) \equiv u(c_1) - u[R(Ne - c_1)] > 0.$$

Note das condições de primeira ordem acima que $(c_1, \bar{c}, y) \to (e,e,e)$ quando $R \to 1$. Logo, o indivíduo paciente na primeira posição fica indiferente entre revelar ou não a verdade durante uma corrida quando $R \to 1$, pois $F(1) = 0$. Se utilizando da abordagem proposta por Bertolai et alii (2014), Bertolai et alii (2016) estudam a existência de corrida por meio da derivada $F'(R)$ na vizinhança de $R = 1$, ao destacar que se $F'(1) \equiv \lim_{R \to 1} F'(R) > 0$, então existe corrida bancária para R suficientemente próximo de (e maior do que) 1. No caso $N = 2$ ela é dada por

$$F'(R) = u'(c_1)\frac{dc_1}{dR} - u'[R(Ne - c_1)]\left(Ne - c_1 - R\frac{dc_1}{dR}\right).$$

Logo, $F'(1) = u'(e)[2c_1' - e]$, em que $c_1' \equiv \lim_{R \to 1} dc_1/dR$. Os autores calculam c_1' diferenciando o sistema de condições de primeira ordem em R e, em seguida, estudando o limite das derivadas obtidas quando $R \to 1$.

***Lema* 4.1.** *Suponha $N = 2$. Existe corrida bancária (com pacientes revelando a verdade somente na segunda posição) para $R \approx 1$ se*

$$\frac{\delta - 1}{\delta} > \frac{1 + 2p}{2(1 - p)},$$

pois $c_1' = \frac{1-p}{1+2p}\frac{\delta-1}{\delta}e$.

Demonstração. Denote $\bar{c}' \equiv \lim_{R\to 1} d\bar{c}/dR$ e $y' \equiv \lim_{R\to 1} dy/dR$. A derivada de (19) é dada por $u''(\bar{c})\frac{d\bar{c}}{dR} = u'(y) + Ru''(y)\frac{dy}{dR}$, cujo limite é $u''(e)\bar{c}' = u'(e) + u''(e)y'$. Usando $\delta = -eu''(e)/u'(e)$, tem-se $\bar{c}' = y' - e/\delta$. O limite da derivada de (20) em R quando $R \to 1$ é dada por $u''(e)c_1' = (1 - p)[u'(e) + u''(e)y'] - pu''(e)c_1'$. Logo, $c_1' = \frac{1-p}{1+p}(y' - e/\delta)$. Finalmente, diferenciando (21) em R e tomando o limite da derivada assim obtida quando $R \to 1$, obtém-se $y' = e - (p/2)(c_1' + \bar{c}')$. Conclui-se com isso que

$$\bar{c}' = e\frac{\delta - 1}{\delta} - \frac{p}{2}(c_1' + \bar{c}') \quad \text{e} \quad c_1' = \frac{1 - p}{1 + p}\bar{c}'.$$

Logo,

$$\frac{2 + p}{2}\bar{c}' = e\frac{\delta - 1}{\delta} - \frac{p}{2}c_1' \qquad \Rightarrow \qquad \bar{c}' = \frac{1}{2 + p}\left(2e\frac{\delta - 1}{\delta} - pc_1'\right)$$

$$(1 + p)c_1' = \frac{(1 - p)}{2 + p}\left(2e\frac{\delta - 1}{\delta} - pc_1'\right)$$

$$[(1 + p)(2 + p) + p(1 - p)]c_1' = 2(1 - p)e\frac{\delta - 1}{\delta}$$

$$[2 + 4p]c_1' = 2(1 - p)e\frac{\delta - 1}{\delta}$$

e, portanto, $c_1' = \frac{1-p}{1+2p}\frac{\delta-1}{\delta}e$. Segue que,

$$F'(1) = u'(e)[2c_1' - e] = u'(e)e\left[2\frac{1 - p}{1 + 2p}\frac{\delta - 1}{\delta} - 1\right]$$

e, portanto, existe corrida se $\frac{\delta-1}{\delta} > \frac{1+2p}{2(1-p)}$. □

Do Lema 4.1, existe corrida bancária parcial (restrita à primeira posição da fila) se p e suficientemente baixo e δ é suficientemente alto. Por exemplo, se $e = 1$, $p = 0.05$ e $\delta = 3$, então $c_1' = 19/33$. Neste caso, $F'(1) = \frac{5}{33}u'(1) > 0$ e existe corrida para R suficientemente baixo.

Para o caso geral com $N \in \mathbb{N}$ depositantes, o estudo de existência de corrida é análogo. Bertolai et alii (2016) mostram que o problema de escolha de sistema bancário (x,y) consiste em maximizar

$$\mathbb{E}\left[\sum_{i=1}^{N}(1 - \omega_i)u[x_i(\omega^i)] + N(1 - p)u(y)\right]$$

sujeito a $x_i(\omega^i) \geq 0$,

$$\sum_{i=1}^{N}(1 - \omega_i)x_i(\omega^i) \leq Ne \tag{23}$$

$$y \leq \frac{R}{N(1 - p)}\mathbb{E}\left[Ne - \sum_{i=1}^{N}x_i(\omega^i)\right]. \tag{24}$$

O lagrangeano associado a este problema é demonstrado ser

$$\hat{\phi}[u(y)-\lambda y]+\bar{\phi}[u(\bar{c})-\lambda R\bar{c}]+\sum_{i=1}^{N}(p^N[u(c_i)-\mu c_i]+\phi_i[u(c_i)-\lambda R c_i])$$

em que μ e λ são os multiplicadores de Lagrange associados, respectivamente, a (23) e (24). Como no caso $N=2$, a variável c_i denota o pagamento $x_i(0,0,\cdots,0)$, a variável $\bar{c}$ denota o pagamento $x_i(\omega)$ para $\omega \neq (0,0,\cdots,0)$ e y é o pagamento em $t=2$. Os coeficientes $\hat{\phi},\bar{\phi}$, e ϕ_i são tais que $\phi_i = p^i - p^N$, $\bar{\phi} = Np-(p+p^2+\cdots+p^N)$ e $\hat{\phi}+\bar{\phi}+\sum_{i=1}^{N-1}\phi_i = N(1-p^N)$. O Lema a seguir é demonstrado por Bertolai et alii (2016) e é a generalização dos Lemas 2.4 e 4.1.

***Lema* 4.2** (Bertolai et alii (2016))**.** *Para $R \approx 1$, o sistema bancário ótimo com um contínuo de bancos e $N \in \mathbb{N}$ é definido por $(c_1,c_2,\cdots,c_{N-1},\bar{c},y)$ tal que*

$$\begin{aligned} u'(\bar{c}) &= Ru'(y) \\ u'(c_i) &= p^{N-i}u'(c_N)+(1-p^{N-i})Ru'(y) \\ (1-p^N)RNe &= \hat{\phi}y+R\bar{\phi}\bar{c}+R\sum_{i=1}^{N-1}\phi_i c_i \end{aligned} \tag{25}$$

e, portanto, $y' = \bar{c}' + e/\delta$ e

$$\begin{aligned} c_i' &= \left(1-\frac{p^{N-i}}{1-p^N}(1-p)N\right)\bar{c}' \\ \bar{c}' &= e\frac{\delta-1}{\delta}\frac{1-p}{1-\frac{p^N}{1-p^N}(1-p)N} > 0. \end{aligned}$$

Demonstração. Ver Bertolai et alii (2016). □

A hipótese de R suficientemente baixo é usada para demonstrar que as restrições de incentivo (revelação da verdade) são atendidas pela solução do problema acima (no qual tais restrições foram ignoradas)[24] e para demostrar que o pagamento em $t=1$ é constante após o primeiro anúncio de tipo 1.

O estudo sobre existência de corrida bancária para o caso geral $N \in \mathbb{N}$ segue a mesma estratégia do caso $N=2$. Contrói-se a função $F(R)$ que determina existência de corrida bancária quando é não negativa. Usando que $F(1)=0$ também no caso geral, estuda-se o sinal da derivada $F'(1) \equiv \lim_{R\to 1}F'(R)$. A condição de corrida será dada por $F'(1)>0$.

Suponha novamente que todos os indivíduos acreditam que os indivíduos nas primeiras $N-1$ posições estão anunciando tipo 0 e que todos os indivíduos na última posição (N) estão revelando a verdade. Com isso, todos acreditam que p bancos estão recebendo anúncios $\omega = (0,\cdots,0,0) \in \{0,1\}^N$ e os demais $1-p$ bancos estão recebendo $\omega = (0,0,\cdots,0,1) \in \{0,1\}^N$. Logo, os indivíduos não acreditam que o pagamento em $t=2$ será determinado pela restrição (25). Ele será dado por

$$(1-p)\left[R\left(Ne-\sum_{i=1}^{N-1}c_i\right)-y\right]=0, \tag{26}$$

[24] O resultado enunciado na seção 2 de que tais restrições são sempre atendidas é estabelecido por Green e Lin (2003) para economias em que os bancos operam em completo isolamento ou sob completa integração. A economia em Bertolai et alii (2016) estuda o caso de integração parcial e, portanto, o resultado de Green e Lin (2003) não é válido necessariamente.

pois os bancos com história $(0,0,\cdots,0,0)$ exaurem suas reservas em $t=1$, já que $\sum_{i=1}^{N} c_i = Ne$, e os bancos com história $(0,0,\cdots,0,1)$ pagam no total $\sum_{i=1}^{N-1} c_i$ em $t=1$ e poupam $Ne - \sum_{i=1}^{N-1} c_i$ para $t=2$.

Como no caso $N=2$, os indivíduos impacientes e os indivíduos pacientes situados na última posição sempre anunciam seus verdadeiros tipos: os impacientes não prezam por consumo em $t=2$ e os pacientes na posição N preferem anunciar a verdade para obter $R\left(Ne - \sum_{i=1}^{N-1} c_i\right)$ do que mentir para obter $Ne - \sum_{i=1}^{N-1} c_i$. Considere então um paciente na posição $i \leq N-1$. Ao mentir sobre seu tipo, o indivíduo receberá c_i com certeza. Se revelar a verdade na posição i, o paciente receberá[25] $y = R\left(Ne - \sum_{i=1}^{N-1} c_i\right)$. Portanto, é atrativo para o paciente da posição i mentir sobre seu tipo se

$$F_i(R) \equiv u(c_i) - u\left[R\left(Ne - \sum_{j=1}^{N-1} c_j\right)\right] > 0.$$

Como $(c_i,\bar{c},y) \to (e,e,e)$ quando $R \to 1$ para todo $i \in \{1,2,\cdots,N\}$, então $F_i(1)=0$. A derivada de $F_i(R)$ é neste caso $F_i'(1) = u'(e)[c_i' - e - c_N']$, pois de $\sum_{j=1}^{N} c_j = Ne$ sabe-se que $c_N' = -\sum_{j=1}^{N-1} c_j'$. Usando o Lema 4.2 para obter c_i' e $\bar{c}'$ e (26) para obter y', Bertolai et alii (2016) calculam as condições sob as quais $F_i'(1) > 0$ para todo $i \leq N-1$. Ou seja, $c_i' > e + c_N'$ para todo $i \leq N-1$.

***Proposição* 4.3** (Bertolai et alii (2016))**.** *Suponha $N \in \mathbb{N}$. Existe corrida bancária (com pacientes revelando a verdade somente na posição N) para $R \approx 1$ se*

$$\frac{\delta - 1}{\delta} f_1(1) > \frac{1 - p^N}{N(1-p)} - p^N$$

em que $f_k(n) = \frac{1-p^k}{k(1-p)} - p^k - (1-p)p^n$

Demonstração. Ver Bertolai et alii (2016). □

Note que a proposição 4.3 tem como caso particular o Lema 4.1 quando $N=2$. De forma similar, é demonstrado existência de corrida bancária na qual os indivíduos pacientes revelam a verdade somente nas duas últimas posições da fila ($N-1$ e N).

Suponha agora que todos os indivíduos acreditam que os indivíduos nas primeiras $N-2$ posições estão anunciando tipo 0 e que todos os indivíduos na penúltima e na última posição ($N-1$ e N) estão revelando a verdade. Com isso, todos acreditam que p^2 bancos estão recebendo anúncios $\omega = (\mathbf{0}^{N-2},0,0) \in \{0,1\}^N$, $p(1-p)$ bancos estão recebendo anúncios $\omega = (\mathbf{0}^{N-2},0,1) \in \{0,1\}^N$, $(1-p)p$ bancos estão recebendo anúncios $\omega = (\mathbf{0}^{N-2},1,0) \in \{0,1\}^N$ e os demais $(1-p)^2$ bancos estão recebendo $\omega = (\mathbf{0}^{N-2},1,1) \in \{0,1\}^N$, em que $\mathbf{0}^n \in \{0,1\}^n$ é a história parcial de anúncios com $n \in \{1,2,\cdots,N\}$ anúncios iguais a zero. Logo, os indivíduos acreditam que o pagamento em $t=2$ será determinado pela seguinte restrição

$$p\left[R\left(Ne - \sum_{i=1}^{N-1} c_i\right) - y\right] + p\left[R\left(Ne - \sum_{i=1}^{N-2} c_i - \bar{c}\right) - y\right] + (1-p)\left[R\left(Ne - \sum_{i=1}^{N-2} c_i\right) - 2y\right] = 0, \qquad (27)$$

pois os bancos com história $(\mathbf{0}^{N-2},0,0)$ exaurem suas reservas em $t=1$, os bancos com história $(\mathbf{0}^{N-2},0,1)$ poupam $Ne - \sum_{i=1}^{N-1} c_i = c_N$ para $t=2$, os bancos com história $(\mathbf{0}^{N-2},1,0)$ poupam

[25] Apesar de este anúncio provocar o surgimento de um banco com história de anúncios $\omega \notin \{(0,0,\cdots,0,0),(0,0,\cdots,0,1)\}$, o efeito da poupança deste banco na poupança agregada é desprezível tendo em vista que existe na economia um contínuo de bancos. Portanto, ao revelar a verdade na posição i, o paciente não afeta seu pagamento na data 2.

$Ne - \sum_{i=1}^{N-2} c_i - \bar{c} = c_N + c_{N-1} - \bar{c}$ para $t = 2$ e os bancos com história $(\mathbf{0}^{N-2},1,1)$ poupam $Ne - \sum_{i=1}^{N-2} c_i = c_N + c_{N-1}$ para $t = 2$.

Novamente os indivíduos impacientes sempre anunciam seus verdadeiros tipos, uma vez que não prezam por consumo em $t = 2$. Considere o incentivo de um indivíduo paciente mentir ou não na posição $i \in \{1,2,\cdots,N\}$. Ao mentir sobre seu tipo, o indivíduo receberá c_i. Se revelar a verdade, o paciente receberá[26] pagamento y que satisfaz (27), ou seja, $y = (R/2)\left(c_N + c_{N-1} - p(\bar{c} - c_N)\right)$. Portanto, será atrativo para o paciente na posição i mentir sobre seu tipo se

$$F_i(R) \equiv u(c_i) - u\left[\frac{R}{2}\left(c_N + c_{N-1} - p(\bar{c} - c_N)\right)\right] > 0.$$

Novamente $F_i(1) = 0$ para todo $i \in \{1,2,\cdots,N\}$. A derivada de $F_i(R)$ é neste caso $F_i'(1) = u'(e)[c_i' - e - \frac{1}{2}(c_N' + c_{N-1}' - p(\bar{c}' - c_N'))]$, pois de $\sum_{j=1}^{N} c_j = Ne$ sabe-se que $c_N' = -\sum_{j=1}^{N-1} c_j'$. Usando o Lema 4.2 para obter c_i' e $\bar{c}'$ e (27) para obter y', Bertolai et alii (2016) calculam as condições sob as quais $F_i'(1) > 0$ para todo $i \leq N-2$ e $F_i'(1) \leq 0$ para $i \in \{N-1,N\}$.

***Proposição* 4.4** (Bertolai et alii (2016))**.** *Suponha $N \in \mathbb{N}$. Existe corrida bancária (com pacientes revelando a verdade somente nas posições $N-1$ e N) para $R \approx 1$ se*

$$\frac{\delta - 1}{\delta} f_2(2) > \frac{1 - p^N}{(1-p)N} - p^N \geq \frac{\delta - 1}{\delta} f_2(1)$$

em que $f_k(n) = \frac{1-p^k}{k(1-p)} - p^k - (1-p)p^n$.

Demonstração. Ver Bertolai et alii (2016). □

Diferentemente do obtido na proposição 4.3, na proposição 4.4 há duas desigualdades. A primeira desigualdade garante que é atrativo para todo indivíduo $i \leq N-2$ participar da corrida e a segunda desigualdade garante que os demais dois (últimos) depositantes não queiram participar da corrida. Na proposição 4.3, há somente uma desigualdade, pois o último depositante da fila sempre prefere revelar a verdade.

4.1. Outros equilíbrios de corrida bancária

Considere agora o caso em que todos os indivíduos acreditam que os indivíduos nas primeiras $N-3$ posições estão anunciando tipo 0 e que todos os indivíduos nas três últimas posições ($N-2$, $N-1$ e N) estão revelando a verdade. Com isso, todos acreditam que a quantidade de bancos recebendo anúncios $\omega = (\mathbf{0}^{N-3},\omega_{N-2},\omega_{N-1},\omega_N) \in \{0,1\}^N$ é dada por

$$\Pr(\omega_{N-2},\omega_{N-1},\omega_N) = p^{3-(\omega_{N-2}+\omega_{N-1}+\omega_N)}(1-p)^{(\omega_{N-2}+\omega_{N-1}+\omega_N)}.$$

Se $s(\omega_{N-2},\omega_{N-1},\omega_N)$ denota a poupança para o período $t = 2$ do banco com história de anúncios $\omega = (\mathbf{0}^{N-3},\omega_{N-2},\omega_{N-1},\omega_N)$ e $A = Ne - \sum_{i=1}^{N-3} c_i$ denota a reserva bancária após $N-3$ anúncios de tipo 0, então

[26] Novamente, o paciente não afeta seu pagamento na data 2 ao revelar a verdade na posição i.

$$s(x) = \begin{cases} A - c_{N-2} - c_{N-1} - c_N & = 0 & \text{se } x = (0,0,0) \\ A - c_{N-2} - c_{N-1} - 0 & = c_N & \text{se } x = (0,0,1) \\ A - c_{N-2} - 0 - \bar{c} & = c_N + c_{N-1} - \bar{c} & \text{se } x = (0,1,0) \\ A - c_{N-2} - 0 - 0 & = c_N + c_{N-1} & \text{se } x = (0,1,1) \\ A - 0 - \bar{c} - \bar{c} & = c_N + c_{N-1} + c_{N-2} - 2\bar{c} & \text{se } x = (1,0,0) \\ A - 0 - \bar{c} - 0 & = c_N + c_{N-1} + c_{N-2} - \bar{c} & \text{se } x = (1,0,1) \\ A - 0 - 0 - \bar{c} & = c_N + c_{N-1} + c_{N-2} - \bar{c} & \text{se } x = (1,1,0) \\ A - 0 - 0 - 0 & = c_N + c_{N-1} + c_{N-2} & \text{se } x = (1,1,1) \end{cases}. \quad (28)$$

Logo, os indivíduos acreditam que o pagamento em $t = 2$ será determinado pela seguinte restrição

$$\sum_{x \in \{0,1\}^3} \Pr(x) \left[Rs(x) - (x_1 + x_2 + x_3)y \right] = 0, \quad (29)$$

a qual generaliza (27). Mais uma vez, os indivíduos impacientes sempre anunciam seus verdadeiros tipos, uma vez que não prezam por consumo em $t = 2$. Considere o incentivo de um indivíduo paciente mentir ou não na posição $i \in \{1,2,\cdots,N\}$. Ao mentir sobre seu tipo, o indivíduo receberá c_i. Se revelar a verdade, o paciente receberá[27] pagamento y que satisfaz (29), ou seja, $y = \frac{R}{3(1-p)} \sum_{x \in \{0,1\}^3} \Pr(x)s(x)$. Portanto, é atrativo para o paciente na posição i mentir sobre seu tipo se

$$F_i(R) \equiv u(c_i) - u\left[\frac{R}{3(1-p)} \sum_{x \in \{0,1\}^3} \Pr(x)s(x) \right] > 0.$$

Novamente $F_i(1) = 0$ para todo $i \in \{1,2,\cdots,N\}$, pois $\sum_{x \in \{0,1\}^3} \Pr(x)s(x) \to 3(1-p)e$ quando $R \to 1$. A derivada de $F_i(R)$ para $R = 1$ é neste caso

$$F_i'(1) = u'(e) \left[c_i' - y' \right],$$

em que $y' = e + \frac{1}{3(1-p)} \sum_{x \in \{0,1\}^3} \Pr(x)s'(x)$ é a derivada do pagamento em $t = 2$ definido por (29), avaliada em $R = 1$, e $s'(x) \equiv \lim_{R \to 1} ds(x)/dR$. A proposição 4.5 a seguir apresenta as condições sobre os parâmetros sob as quais $F_i'(1) > 0$ para todo $i \leq N-3$ e $F_i'(1) \leq 0$ para $i \in \{N-2, N-1, N\}$.

***Proposição* 4.5.** *Suponha $N \in \mathbb{N}$. Existe corrida bancária (com pacientes revelando a verdade somente nas posições $N-2$, $N-1$ e N) para $R \approx 1$ se*

$$\frac{\delta - 1}{\delta} f_3(3) > \frac{1 - p^N}{(1-p)N} p^N - p^N \geq \frac{\delta - 1}{\delta} f_3(2) \quad (30)$$

em que $f_k(n) = \frac{1-p^k}{k(1-p)} - p^k - (1-p)p^n$.

Demonstração. Observe que $F_i'(1) > 0$ se e somente se $c_i' > y'$, em que $y' = e + \sum_{x \in \{0,1\}^3} \frac{\Pr(x)}{3(1-p)} s'(x)$. Usando (28), note também que

[27] Mais uma vez, o paciente não afeta seu pagamento na data 2 ao revelar a verdade na posição i.

$$\begin{aligned}\sum_{x\in\{0,1\}^3}\Pr(x)s'(x) &= (1-p^3)c'_N+\left[1-p^3-p^2(1-p)\right]c'_{N-1}\\ &\quad+\left[1-p^3-2p^2(1-p)-p(1-p)^2\right]c'_{N-2}-p(1-p)(2+p)\bar{c}'\\ &= (1-p)\Big[(1+p+p^2)c'_N+(1+p)c'_{N-1}+c'_{N-2}-p(2+p)\bar{c}'\Big]\end{aligned}$$

em que se usou $1-p^3=(1-p)(1+p+p^2)$ na segunda igualdade. Logo, $y'=e+\frac{1}{3}\left[(1+p+p^2)c'_N+(1+p)c'_{N-1}+c'_{N-2}-(2+p)p\bar{c}'\right]$. Usando c'_i e $\bar{c}'$ definidos no Lema 4.2, obtém-se

$$\begin{aligned}y'-e &= \frac{\bar{c}'}{3}\left[(1+p+p^2)\left(1-\frac{(1-p)N}{1-p^N}\right)+(1+p)\left(1-\frac{p(1-p)N}{1-p^N}\right)\right.\\ &\quad\left.+\left(1-\frac{p^2(1-p)N}{1-p^N}\right)-p(2+p)\right]\\ &= \frac{\bar{c}'}{3}\left[(3+2p+p^2-p(2+p))-[1+p+2p^2+p(1+p)]\frac{(1-p)N}{1-p^N}\right]\\ &= \frac{\bar{c}'}{3}\left[3-[2p^2+(1+p)^2]\frac{(1-p)N}{1-p^N}\right].\end{aligned}$$

Portanto, é atrativo mentir na posição i se e somente se

$$\begin{aligned}\left(1-\frac{p^{N-i}(1-p)N}{1-p^N}\right)\bar{c}'=c'_i &> y'=e+\frac{\bar{c}'}{3}\left[3-[2p^2+(1+p)^2]\frac{(1-p)N}{1-p^N}\right]\\ \left(3-3p^{N-i}\frac{(1-p)N}{1-p^N}\right)\bar{c}' &> 3e+\bar{c}'\left[3-[2p^2+(1+p)^2]\frac{(1-p)N}{1-p^N}\right]\\ 0 &> 3+\frac{\bar{c}'}{e}\left[3p^{N-i}-2p^2-(1+p)^2\right]\frac{(1-p)N}{1-p^N}\\ -3\frac{e}{\bar{c}'}\frac{1-p^N}{(1-p)N} &> 3p^{N-i}-[2p^2+(1+p)^2].\end{aligned}$$

Como o lado direito da desigualdade acima é crescente em i, então basta verificar sob quais condições a desigualdade é satisfeita para $i=N-3$ e violada para $i=N-2$. Ou seja,

$$\frac{3}{1-p}f_3(3)=[2p^2+(1+p)^2]-3p^3>3\frac{e}{\bar{c}'}\frac{1-p^N}{(1-p)N}\geq\left[2p^2+(1+p)^2\right]-3p^2=f_3(2)\frac{3}{1-p}.$$

Do Lema 4.2, tem-se $\left(1-p^N\frac{(1-p)N}{1-p^N}\right)\frac{\bar{c}'}{e}=\frac{\delta-1}{\delta}(1-p)$. Logo,

$$3\frac{e}{\bar{c}'}\frac{1-p^N}{(1-p)N}=\frac{1-p^N}{(1-p)N}\frac{3}{1-p}\left(1-p^N\frac{(1-p)N}{1-p^N}\right)\frac{\delta}{\delta-1}=\frac{\delta}{\delta-1}\frac{3}{1-p}\left(\frac{1-p^N}{(1-p)N}-p^N\right).$$

Portanto, a condição de existência da referida corrida é

$$p^N+f_3(3)\frac{\delta-1}{\delta} > \frac{1-p^N}{(1-p)N}\geq p^N+f_3(2)\frac{\delta-1}{\delta}$$

a qual é equivalente à condição (30).

□

4.2. O caso geral

Considere agora o caso em que todos os indivíduos acreditam que os indivíduos nas primeiras $N-k$ posições estão anunciando tipo 0 e que todos os indivíduos nas k últimas posições ($N-k+1$, $N-k+2$, $\cdots$, $N-1$ e N) estão revelando a verdade. Com isso, todos acreditam que a quantidade de bancos recebendo anúncios $\omega = (\mathbf{0}^{N-k},\omega_{N-k+1},\omega_{N-k+2},\cdots,\omega_{N-1},\omega_N) \in \{0,1\}^N$ é dada por

$$\Pr(\omega_{N-k+1},\omega_{N-k+2},\cdots,\omega_{N-1},\omega_N) = p^{k-|\omega|_k}(1-p)^{|\omega|_k},$$

em que $|\omega|_k = \sum_{i=1}^{k}\omega_{N-k+i}$. Se $s(\omega_{N-k+1},\cdots,\omega_{N-1},\omega_N)$ denota a poupança para o período $t=2$ do banco com história de anúncios $\omega = (\mathbf{0}^{N-k},\omega_{N-k+1},\cdots,\omega_{N-1},\omega_N)$ e $A = Ne - \sum_{i=1}^{N-k} c_i = \sum_{i=1}^{k} c_{N-k+i}$ denota a reserva bancária após $N-k$ anúncios de tipo 0, então

$$\begin{aligned} s(x) &= A - \sum_{i=1}^{k}(1-x_i)\left[c_{N-k+i}\left(I_{[\sum_{j=1}^{i}x_j=0]}\right) + \bar{c}\left(1 - I_{[\sum_{j=1}^{i}x_j=0]}\right)\right] \\ &= A - (k-|x|_k)\bar{c} - \sum_{i=1}^{k}\left(c_{N-k+i}-\bar{c}\right)I_{[|x|_i=0]} \end{aligned} \tag{31}$$

em que $|x|_i = \sum_{j=1}^{i}x_j$ denota a quantidade de pacientes entre as posições $N-k+1$ e $N-k+i$. A função $I_{[expr]}$ é igual a 1 se a sentença $expr$ é verdadeira e igual a 0 caso contrário. Na segunda igualdade também foi usado que $(1-x_i)=1$ sempre que $I_{[|x|_i=0]}=1$. Logo, os indivíduos acreditam que o pagamento em $t=2$ será determinado pela seguinte restrição

$$\sum_{x\in\{0,1\}^k}\Pr(x)\left[Rs(x) - y|x|_k\right] = 0. \tag{32}$$

Mais uma vez, os indivíduos impacientes sempre anunciam seus verdadeiros tipos, uma vez que não prezam por consumo em $t=2$. Considere o incentivo de um indivíduo paciente mentir ou não na posição $i \in \{1,2,\cdots,N\}$. Ao mentir sobre seu tipo, o indivíduo receberá c_i. Se revelar a verdade, o paciente receberá[28] pagamento $\tilde{y}_k$ que satisfaz (32), ou seja, $\tilde{y}_k = \frac{R}{k(1-p)}\sum_{x\in\{0,1\}^k}\Pr(x)s(x)$. Portanto, é atrativo para o paciente na posição i mentir sobre seu tipo se

$$F_i(R) \equiv u(c_i) - u\left(\tilde{y}_k\right) > 0.$$

Novamente $F_i(1)=0$ para todo $i \in \{1,2,\cdots,N\}$, pois $\sum_{x\in\{0,1\}^k}\Pr(x)s(x) \to k(1-p)e$ quando $R\to 1$. A derivada de $F_i(R)$ para $R=1$ é neste caso

$$F_i'(1) = u'(e)\left[c_i' - \tilde{y}_k'\right],$$

em que $\tilde{y}_k' = e + \frac{1}{k(1-p)}\sum_{x\in\{0,1\}^k}\Pr(x)s'(x)$ é a derivada do pagamento em $t=2$ definido por (32), avaliada em $R=1$, e $s'(x) \equiv \lim_{R\to 1} ds(x)/dR$. A proposição 4.6 a seguir apresenta as condições sobre os parâmetros sob as quais $F_i'(1)>0$ para todo $i \le N-k$ e $F_i'(1)\le 0$ para $i > N-k$.

***Proposição* 4.6.** *Suponha $N\in\mathbb{N}$. Existe corrida bancária (com pacientes revelando a verdade somente nas k últimas posições) para $R\approx 1$ se*

$$\frac{\delta-1}{\delta}f_k(k) > \frac{1-p^N}{(1-p)N} - p^N \ge \frac{\delta-1}{\delta}f_k(k-1) \tag{33}$$

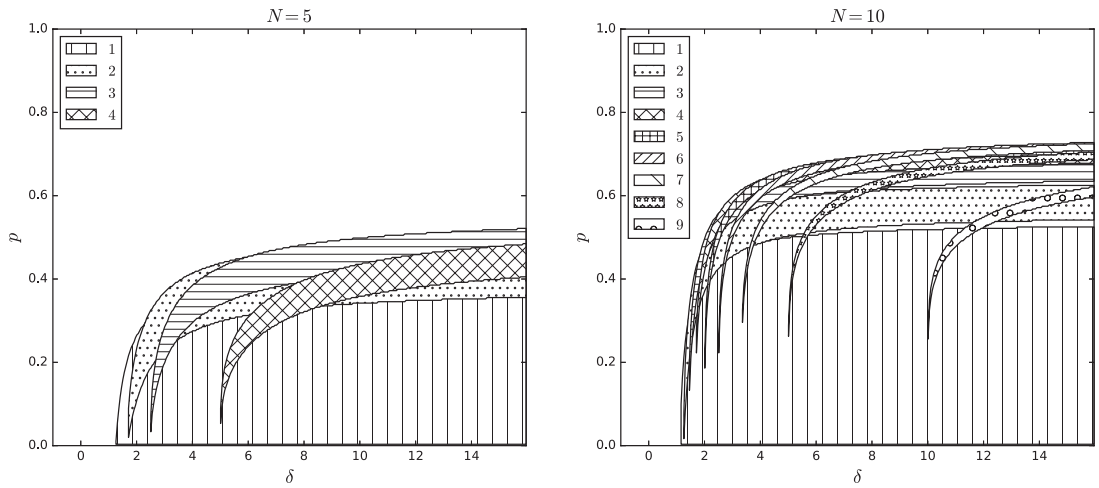
em que $f_k(n) = \frac{1-p^k}{k(1-p)} - p^k - (1-p)p^n$.

[28] Mais uma vez, o paciente não afeta seu pagamento na data 2 ao revelar a verdade na posição i.

Demonstração. Ver apêndice A. □

Note que (33) não é satisfeita para $k = N$, independentemente do valor de (N,p,δ). Tal resultado é consistente com o atendimento das restrições de *truth-telling* quando $R \approx 1$, conforme estabelecido por Bertolai et alii (2016). Quando todos acreditam que todos estão revelando a verdade (revelando verdade nas últimas $k = N$ posições), a melhor opção para cada um dos indivíduos é revelar a verdade. Observe ainda que $f_1(1) = (1-p)^2$, $f_1(0) = 0$ e $\lim_{N\to\infty} \frac{1-p^N}{(1-p)N} - p^N = 0$. Logo, (33) é satisfeita para $k = 1$ quando N é suficientemente grande[29] e quando $N = 2$ e δ é suficientemente grande.[30] A Figura a seguir ilustra o padrão de existência de corrida para $k \in \mathbb{N}$ tal que $1 \leq k < N$.

Figura 2: Economias nas quais há corrida



São apresentados exemplos para $N = 3$ e $N = 10$. Os gráficos apresentam em áreas hachuradas os valores de (δ,p) para os quais há corrida. São estudados valores $p \in (0,1)$ e $\delta \in (1,16)$. Os gráficos sugerem não haver padrão claro em relação ao número de equilíbrios de corrida. Há casos sem corrida bancária, casos com somente um equilíbrio de corrida e casos com mais de um equilíbrio de corrida. Eles mostram que é cada vez menor o conjunto de economias nas quais há equilíbrio de corrida somente nas primeiras posições da fila de saques ($k \approx N$).[31]

5. OBSERVAÇÕES FINAIS

Este artigo propõe uma linguagem sob a qual os modelos de Allen e Gale (2000) e Green e Lin (2003) podem ser vistos como casos particulares. Ambos são baseados em Diamond e Dybvig (1983) e estudam o sistema bancário escolhido pela sociedade para alocar (distribuir) eficientemente a liquidez da economia. Além disso, este artigo complementa dois importantes resultados na literatura de Teoria Bancária: a possibilidade de contágio durante uma crise bancária (estabelecida por Allen e Gale (2000)) e a existência de corrida bancária (estabelecida por Bertolai et alii (2016) em uma economia bastante similar àquela estudada por Green e Lin (2003)). Demonstra-se que *(i)* o contágio existente no modelo de Allen e Gale (2000) pode ser eliminado com uma intervenção bastante parcimoniosa no mercado interbancário;

[29] Conforme já estabelecido pela proposição 4.3.

[30] Conforme já estabelecido pelo Lema 4.1.

[31] Os demais exemplos apresentados no apêndice B corroboram esta afirmação.

e *(ii)* as condições para existência de corrida pode ser generalizadas para o caso em que os pacientes nas últimas $k \in \{1,2,\cdots,N-1\}$ posições na fila de saques não participam da corrida.

BIBLIOGRAFIA

Allen, F. & Gale, D. (2000). Financial contagion. *Journal of Political Economy*, 108(1):1–33.

Andolfatto, D., Nosal, E., & Wallace, N. (2007). The role of independence in the Green–Lin Diamond–Dybvig model. *Journal of Economic Theory*, 137(1):709–715.

Bertolai, J. D., Cavalcanti, R. d. O., & Monteiro, P. K. (2014). Run theorems for low returns and large banks. *Economic Theory*, 57(2):223–252.

Bertolai, J. D., Cavalcanti, R. d. O., & Monteiro, P. K. (2016). Bank runs with many small banks and mutual guarantees at the terminal stage. Mimeo.

Bryant, J. (1980). A model of reserves, bank runs, and deposit insurance. *Journal of Banking & Finance*, 4(4):335–344.

Diamond, D. W. & Dybvig, P. H. (1983). Bank runs, deposit insurance, and liquidity. *Journal of Political Economy*, 91(3):401–419.

Ennis, H. M. & Keister, T. (2009). Bank runs and institutions: The perils of intervention. *The American Economic Review*, 99(4):1588–1607.

Green, E. J. & Lin, P. (2003). Implementing efficient allocations in a model of financial intermediation. *Journal of Economic Theory*, 109(1):1–23.

Jacklin, C. J. (1987). Demand Deposits, Trading Restrictions, and Risk Sharing. *Contractual Arrangements for Intertemporal Trade*, 1:26–47.

Kareken, J. & Wallace, N. (1978). Deposit Insurance and Bank Regulation: A Partial Equilibrium Exposition. *Journal of Business*, 51(Jul):413–438.

Peck, J. & Shell, K. (2003). Equilibrium bank runs. *Journal of political Economy*, 111(1):103–123.

Wallace, N. (1988). Another attempt to explain an illiquid banking system: the Diamond and Dybvig model with sequential service taken seriously. *Quarterly Review*, pages 3–16.

A. DEMONSTRAÇÕES

A.1. Corolário 2.2

Demonstração. Seja $\mu_\omega P(\omega)/N$ o multiplicador de Lagrange associado a restrição de factibilidade no estado $\omega \in \Omega$. O lagrangeano do problema de maximização é dado por:

$$\sum_{\omega\in\Omega}\frac{P(\omega)}{N}\left[(N-|\omega|)u(x(\omega))+|\omega|u(y(\omega))\right]+\sum_{\omega\in\Omega}\mu_\omega\frac{P(\omega)}{N}\left[Ne-\left[(N-|\omega|)x(\omega)+|\omega|y(\omega)R^{-1}\right]\right].$$

A condição de primeira ordem em $x(\omega)$ é:

$$\frac{P(\omega)}{N}(N-|\omega|)[u'(x(\omega))-\mu_\omega]=0.$$

Como $P(\omega)/N \neq 0$, para $|\omega| < N$ tem-se $u'(x(\omega)) = \mu_\omega$. A condição de primeira ordem em $y(\omega)$ é:

$$\frac{P(\omega)}{N}|\omega|[u'(y(\omega)) - \mu_\omega R^{-1}] = 0.$$

Dado que $P(\omega)/N \neq 0$, para $|\omega| > 0$ tem-se $Ru'(y(\omega)) = \mu_\omega$. Por fim, derivando o lagrangeano em relação à μ_ω, obtém-se

$$P(\omega)\left[1 - \frac{N-|\omega|}{N}x(\omega) - \frac{|\omega|}{N}R^{-1}y(\omega)\right] = 0$$

e, portanto, $(1-|\omega|/N)x(\omega) + (|\omega|/N)y(\omega)/R = e$, visto que $P(\omega) \neq 0$.

(i) Se $0 < |\omega| < N$, então $u'(x(\omega)) = \mu_\omega = Ru'(y(\omega))$. Como $\delta = 2$, tem-se $y(\omega) = \sqrt{R}x(\omega)$. Usando a restrição de factibilidade, tem-se:

$$\begin{aligned} \frac{N-|\omega|}{N}x(\omega) + \frac{|\omega|}{N}R^{-1}\sqrt{R}x(\omega) &= e \\ \left[\left(1 - \frac{|\omega|}{N}\right)\sqrt{R} + \frac{|\omega|}{N}\right]x(\omega) &= \sqrt{R}e \end{aligned}$$

e, portanto, $x(\omega) = \alpha(|\omega|/N)e$. Consequentemente, $y(\omega) = \alpha(|\omega|/N)\sqrt{R}e$.

(ii) Se $0 = |\omega| < N$, então $u'(x(\omega)) = \mu_\omega$. Usando a condição de primeira ordem em μ_ω, obtém-se $x(\omega) = e$. Portanto, $\mu_\omega = u'(e)$. No cenário em que depositantes pacientes não existem, os depositantes impacientes consomem toda a dotação da economia: $(x(\omega), y(\omega)) = (e,0)$.

(iii) Se $0 < |\omega| = N$, então $Ru'(y(\omega)) = \mu_\omega$. Usando a condição de primeira ordem em μ_ω, obtém-se $y(\omega) = Re$. Portanto, $\mu_\omega = Ru'(Re)$. Já para o cenário com nenhum depositante impaciente, os depositantes pacientes consomem toda a dotação da economia: $(x(\omega),y(\omega)) = (0,Re)$.

□

A.2. Lema 2.4

Demonstração. Seja μ o multiplicador de Lagrange associado a restrição $c_1 + c_2 = Ne$ e λ o multiplicador associado a restrição (13). As demais restrições valem com desigualdade estrita no ótimo. O lagrangeano associado ao problema de maximização é

$$\begin{aligned} p^2(u(c_1)+u(c_2)) + p(1-p)(u(c_1)+u(y)) + p(1-p)(u(y)+u(\bar{c})) + 2(1-p)^2u(y) \\ +\mu(Ne - c_1 - c_2) \\ +\lambda[(1-p^2)RNe - p(1-p)(Rc_1+y) - p(1-p)(R\bar{c}+y) - 2(1-p)^2y]. \end{aligned}$$

As condições de primeira ordem para o ótimo são:

$$\begin{aligned} p^2u'(c_1) + p(1-p)u'(c_1) - \mu - p(1-p)\lambda R &= 0 \\ p^2u'(c_2) - \mu &= 0 \\ p(1-p)u'(\bar{c}) - p(1-p)\lambda R &= 0 \\ (1-p)\left[pu'(y) + pu'(y) + 2(1-p)u'(y) - 2\lambda p - 2\lambda(1-p)\right] &= 0 \\ Ne - c_1 - c_2 &= 0 \\ (1-p)\left[\frac{1-p^2}{1-p}RNe - p(Rc_1+y) - p(R\bar{c}+y) - 2(1-p)y\right] &= 0. \end{aligned}$$

De onde se obtém $\mu = p^2u'(c_2)$, $\lambda = u'(y)$ e $c_1 = Ne - c_2$. Substituindo tais resultados nas equações restantes e usando $\delta = 2$, se obtém os resultados do lema. □

A.3. Lema 3.1

Demonstração. Lembrando que as regiões com alta demanda por consumo em $t = 1$ resgatam seus depósitos no interbancário em $t = 1$ e as regiões com baixa demanda por consumo em $t = 1$ resgatam seus depósitos no interbancário em $t = 2$. Sem perda de generalidade, suponha realizado o estado S_1 e, portanto, há alta demanda por consumo em $t = 1$ nas regiões A e C e baixa demanda por consumo em $t = 1$ nas regiões B e D.

Considere inicialmente o caso de mercado interbancário completo e suponha $z^i = (\omega_H - \gamma)/8$. Os investimentos em $t = 0$ são factíveis, pois as trocas de depósito entre os bancos tem o mesmo valor e, portanto, se cancelam. O pagamento c_1^* será factível sem a liquidação de ativo de longo prazo para os bancos das regiões A e C se

$$\frac{\omega_H}{4}c_1^* + \frac{\omega_H - \gamma}{8}c_1^* \leq \frac{s^*}{4} + 3\frac{\omega_H - \gamma}{8}c_1^*.$$

O banco da região A (C) promete para o estado S_1 pagar $\omega_H/4$ saques (para indivíduos impacientes) em sua própria região e enviar $z^j = (\omega_H - \gamma)/8$ depósitos para o banco da região C (A). Como cada pagamento em $t = 1$ é igual a c_1^*, então a demanda total por pagamentos em $t = 1$ é dada pelo lado esquerdo da restrição acima. Por outro lado, o banco dispõe de $s^*/4$ unidades de ativo de curto prazo e resgata $3z^i = 3(\omega_H - \gamma)/8$ depósitos nas demais regiões, os quais prometem c_1^* unidades de consumo cada um. Como $s^* = \gamma c_1^*$, então a restrição acima é satisfeita com igualdade. Similarmente, o pagamento c_1^* será factível sem a liquidação de ativo de longo prazo para os bancos das regiões B e D se

$$\frac{\omega_L}{4}c_1^* + 2\frac{\omega_H - \gamma}{8}c_1^* \leq \frac{s^*}{4}.$$

O banco da região B (D) promete para o estado S_1 pagar $\omega_L/4$ saques (para indivíduos impacientes) na própria região em $t = 1$ e enviar $2z^j = 2(\omega_H - \gamma)/8$ depósitos para os bancos das regiões A e C. Como cada pagamento em $t = 1$ é igual a c_1^*, então a demanda total por pagamentos em $t = 1$ é dado pelo lado esquerda restrição acima. Por outro lado, o banco possui $s^*/4$ unidades de ativo de curto prazo. Como $\omega_H - \gamma = \gamma - \omega_L$ e $s^* = \gamma c_1^*$, a restrição acima é satisferita com igualdade.

Os bancos das regiões A e C prometem para $t = 2$ pagar $(1 - \omega_H)/4$ saques (para indivíduos pacientes) na própria região e enviar $2(\omega_H - \gamma)/8$ depósitos para as regiões B e D. Como cada pagamento em $t = 2$ é igual a c_2^* e o banco possui $Rl^*/4$ unidades de ativo de longo prazo, então o pagamento c_2^* será factível nas regiões A e C se

$$\frac{1 - \omega_H}{4}c_2^* + 2\frac{\omega_H - \gamma}{8}c_2^* \leq R\frac{l^*}{4}.$$

Já a região B (D) promete para o estado S_1 em $t = 2$ pagar $(1 - \omega_L)/4$ saques (para indivíduos pacientes) na própria região e enviar $(\omega_H - \gamma)/8$ depósitos para a região D (B). Como cada pagamento em $t = 2$ é igual a c_2^*, então o pagamento c_2^* será factível nas regiões B e D se

$$\frac{1 - \omega_L}{4}c_2^* + \frac{\omega_H - \gamma}{8}c_2^* \leq R\frac{l^*}{4} + 3\frac{\omega_H - \gamma}{8}c_2$$

em que $l^*/4$ e a quantidade de ativo de longo prazo disponível e $3(\omega_H - \gamma)/8$ é o recurso obtido ao resgatar os depoósitos nas demais regiões. Usando o fato de que $l^* = (1 - \gamma)c_2^*/R$, conclui-se que as duas restrições acima são satisfeitas com igualdade.

Considere agora o caso de mercado interbancário *incompleto* e suponha $z^i = (\omega_H - \gamma)/4$. Como no caso de mercado completo, os investimentos em $t = 0$ são factíveis, pois as trocas de depósito entre

os bancos tem o mesmo valor e, portanto, se cancelam. O pagamento c_1^* será factível sem a liquidação de ativo de longo prazo para os bancos das regiões A e C e para os bancos das regiões B e D se, respectivamente,

$$\frac{\omega_H}{4}c_1^* \leq \frac{s^*}{4} + z^i c_1^* \qquad \text{e} \qquad \frac{\omega_L}{4}c_1^* + z^i c_1^* \leq \frac{s^*}{4}.$$

Como $s^* = \gamma c_1^*$ e $z^i = (\omega_H - \gamma)/4$, então ambas as restrições são satisfeitas com igualdade. O pagamento c_2^* será factível nas regiões A e C e B e D se, respectivamente,

$$\frac{1-\omega_H}{4}c_2^* + z^i c_2^* \leq R\frac{l^*}{4} \qquad \text{e} \qquad \frac{1-\omega_L}{4}c_2^* \leq R\frac{l^*}{4} + z^i c_2^*.$$

Como $l^* = (1-\gamma)c_2^*/R$ e $z^i = (\omega_H - \gamma)/4$, então ambas as restrições são satisfeitas com igualdade. □

A.4. Lema 3.4

Demonstração. Ao liquidar $\epsilon c_1^*/16r$ unidades de ativo de longo prazo, o banco i pagará na data 2 somente c_2^i tal que

$$\begin{aligned}
\left(\frac{1-\gamma-\epsilon}{4} + z^D\right)c_2^A &= R\left(\frac{l^*}{4} - \frac{\epsilon c_1^*}{16r}\right) + \left(z^A - \frac{3\epsilon}{16}\right)c_2^B \\
\left(\frac{1-\gamma}{4} + z^A - \frac{3\epsilon}{16}\right)c_2^B &= R\left(\frac{l^*}{4} - \frac{\epsilon c_1^*}{16r}\right) + \left(z^B - \frac{2\epsilon}{16}\right)c_2^C \\
\left(\frac{1-\gamma}{4} + z^B - \frac{2\epsilon}{16}\right)c_2^C &= R\left(\frac{l^*}{4} - \frac{\epsilon c_1^*}{16r}\right) + \left(z^C - \frac{\epsilon}{16}\right)c_2^D \\
\left(\frac{1-\gamma}{4} + z^C - \frac{\epsilon}{16}\right)c_2^D &= R\left(\frac{l^*}{4} - \frac{\epsilon c_1^*}{16r}\right) + z^D c_2^A.
\end{aligned}$$

Multiplicando cada uma das equações por 4, o sistema de equações pode ser reorganizado como $AX = M$ tal que $X = (c_2^A, c_2^B, c_2^C, c_2^D)'$, $M = R(l^* - \epsilon c_1^*/4r)[1,1,1,1]'$ e

$$A = \begin{bmatrix}
1-\gamma+z-\epsilon & \frac{3\epsilon}{4} - z & 0 & 0 \\
0 & 1-\gamma+z-\frac{3\epsilon}{4} & \frac{\epsilon}{2} - z & 0 \\
0 & 0 & 1-\gamma+z-\frac{\epsilon}{2} & \frac{\epsilon}{4} - z \\
-z & 0 & 0 & 1-\gamma+z-\frac{\epsilon}{4}
\end{bmatrix}.$$

Seja $\alpha = 1-\gamma+z$. Note que $1-\gamma > \omega_H - \gamma = z$ e $0 < \epsilon < 1-\gamma$. Logo, $\alpha > z$ e $\alpha > \epsilon$. O sistema é possível e determinado, com solução única, pois

$$|A| = (\alpha-\epsilon)(-1)^2 \begin{vmatrix}
\alpha - \frac{3\epsilon}{4} & \frac{\epsilon}{2} - z & 0 \\
0 & \alpha - \frac{\epsilon}{2} & \frac{\epsilon}{4} - z \\
0 & 0 & \alpha - \frac{\epsilon}{4}
\end{vmatrix} + \left(\frac{3\epsilon}{4} - z\right)(-1)^3 \begin{vmatrix}
0 & \frac{\epsilon}{2} - z & 0 \\
0 & \alpha - \frac{\epsilon}{2} & \frac{\epsilon}{4} - z \\
-z & 0 & \alpha - \frac{\epsilon}{4}
\end{vmatrix}$$

e, portanto,

$$\begin{aligned}
|A| &= (\alpha-\epsilon)\left(\alpha-\frac{3\epsilon}{4}\right)\left(\alpha-\frac{\epsilon}{2}\right)\left(\alpha-\frac{\epsilon}{4}\right)-\left(z-\frac{3\epsilon}{4}\right)z\left(z-\frac{\epsilon}{2}\right)\left(z-\frac{\epsilon}{4}\right)\\
&> (\alpha-\epsilon)\left(z-\frac{3\epsilon}{4}\right)\left(z-\frac{\epsilon}{2}\right)\left(z-\frac{\epsilon}{4}\right)-\left(z-\frac{3\epsilon}{4}\right)z\left(z-\frac{\epsilon}{2}\right)\left(z-\frac{\epsilon}{4}\right)\\
&= (\alpha-z-\epsilon)\left(z-\frac{3\epsilon}{4}\right)\left(z-\frac{\epsilon}{2}\right)\left(z-\frac{\epsilon}{4}\right)\\
&= (1-\gamma-\epsilon)\left(z-\frac{3\epsilon}{4}\right)\left(z-\frac{\epsilon}{2}\right)\left(z-\frac{\epsilon}{4}\right)>0.
\end{aligned}$$

Utilizando o palpite de que os consumos em todas as regiões são iguais a uma constante $\bar{c}$, o sistema será dado por $\bar{c}A[1,1,1,1]' = M$. Dado que as linhas da matriz A somam cada um delas $(1-\gamma-\epsilon/4)$, então,

$$\begin{aligned}
\bar{c} &= \frac{R(l^*-\epsilon c_1^*/4r)}{1-\gamma-\epsilon/4}=\frac{R((1-\gamma)c_2^*/R-\epsilon c_1^*/4r)}{1-\gamma-\epsilon/4}=\frac{(1-\gamma\pm\epsilon/4)c_2^*-R\epsilon c_1^*/4r)}{1-\gamma-\epsilon/4}\\
&= c_2^*+\frac{\epsilon c_1^*}{4}\frac{c_2^*/c_1^*-R/r}{1-\gamma-\epsilon/4}=c_2^*-\frac{\epsilon c_1^*}{4}\frac{R/r-c_2^*/c_1^*}{1-\gamma-\epsilon/4}.
\end{aligned}$$

Como $\bar{c}$ satisfaz o sistema e a solução do sistema é única, $\bar{c}$ é a única solução possível para o sistema de equações acima. □

A.5. Proposição 4.6

Demonstração. Observe que $F_i'(1) > 0$ se e somente se $c_i' > \tilde{y}_k'$, em que $\tilde{y}_k' = e + \sum_{x\in\{0,1\}^k}\frac{\Pr(x)}{k(1-p)}s'(x)$. Usando (31), note também que

$$\sum_{x\in\{0,1\}^k}\Pr(x)s(x) = A-kp\bar{c}-\sum_{x\in\{0,1\}^k}\Pr(x)\sum_{i=1}^{k}I_{[|x|_i=0]}\left(c_{N-k+i}-\bar{c}\right) \qquad \text{(A-1)}$$

em que se usou $\sum_{x\in\{0,1\}^k}\Pr(x)(k-|x|_k)=k-k(1-p)$. A seguir, será demonstrado por indução que para todo $l\in\{0,1,\cdots,k-1\}$

$$\sum_{x\in\{0,1\}^k}\Pr(x)\sum_{i=1}^{k}I_{[|x|_i=0]}\left(c_{N-k+i}-\bar{c}\right)=\sum_{\tilde{x}\in\{0,1\}^{k-l}}\Pr(\tilde{x})g(\tilde{x},l) \qquad \text{(A-2)}$$

em que $g(x,0)=\sum_{i=1}^{k}\left(c_{N-k+i}-\bar{c}\right)I_{[|x|_i=0]}$ para todo $x\in\{0,1\}^k$ e

$$g(x,l+1) = pg[(x,0),l]+(1-p)g[(x,1),l]$$

para todo $x\in\{0,1\}^{k-(l+1)}$. A condição (A-2) é trivialmente satisfeita quando $l=0$. Suponha que (A-2) é válida para $l\in\{0,1,\cdots,k-2\}$. Integrando na $(k-l)$-ésima entrada $\tilde{x}$, obtém-se

$$\begin{aligned}\sum_{\tilde{x}\in\{0,1\}^{k-l}} \Pr(\tilde{x})g(\tilde{x},l) &= \sum_{\tilde{x}\in\{0,1\}^{k-l-1}} \Big[\Pr(\tilde{x},0)g[(\tilde{x},0),l] + \Pr(\tilde{x},1)g[(\tilde{x},1),l]\Big] \\ &= \sum_{\tilde{x}\in\{0,1\}^{k-(l+1)}} \Pr(\tilde{x})\Big[pg[(\tilde{x},0),l] + (1-p)g[(\tilde{x},1),l]\Big] \\ &= \sum_{\tilde{x}\in\{0,1\}^{k-(l+1)}} \Pr(\tilde{x})g(\tilde{x},l+1)\end{aligned}$$

em que se usou $\Pr(\tilde{x}) = \Pr(\tilde{x},0)/p = \Pr(\tilde{x},1)/(1-p)$. Portanto, (A-2) se verifica para todo $l \in \{0,1,\cdots,k-1\}$. Agora, será demonstrado que para todo $l \in \{0,1,\cdots,k-1\}$

$$g(x,l) = \sum_{i=1}^{k-l-1} (c_{N-k+i} - \bar{c})\, I_{[|x|_i=0]} + \sum_{i=k-l}^{k} p^{l-k+i}\,(c_{N-k+i} - \bar{c})\, I_{[|x|_{k-l}=0]} \qquad \text{(A-3)}$$

em que $\sum_{i=j}^{l} m(i) = 0$ se $j > l$ para toda função m. A condição (A-3) é trivialmente satisfeita quando $l = 0$. Suponha que (A-3) é válida para $l \in \{0,1,\cdots,k-2\}$. Então,

$$\begin{aligned}g(x,l+1) &= pg[(x,0),l] + (1-p)g[(x,1),l] \\ &= p\left(\sum_{i=1}^{k-l-1} (c_{N-k+i} - \bar{c})\, I_{[|(x,0)|_i=0]} + \sum_{i=k-l}^{k} p^{l-k+i}\,(c_{N-k+i} - \bar{c})\, I_{[|(x,0)|_{k-l}=0]}\right) \\ &\quad +(1-p)\sum_{i=1}^{k-l-1} (c_{N-k+i} - \bar{c})\, I_{[|(x,1)|_i=0]}.\end{aligned}$$

pois $I_{[|(x,1)|_{k-l}=0]} = I_{[\sum_{j=1}^{k-l-1} x_j+1=0]} = 0$. Logo, usando $I_{[|(x,0)|_i=0]} = I_{[|(x,1)|_i=0]} = I_{[|x|_i=0]}$ para $i < k-l$, tem-se

$$g(x,l+1) = \sum_{i=1}^{k-l-1} (c_{N-k+i} - \bar{c})\, I_{[|x|_i=0]} + \sum_{i=k-l}^{k} p^{l+1-k+i}\,(c_{N-k+i} - \bar{c})\, I_{[|(x,0)|_{k-l}=0]}.$$

Logo,

$$\begin{aligned}g(x,l+1) &= \sum_{i=1}^{k-l-2} (c_{N-k+i} - \bar{c})\, I_{[|x|_i=0]} + \sum_{i=k-l-1}^{k-l-1} p^0\,(c_{N-k+i} - \bar{c})\, I_{[|x|_{k-l-1}=0]} \\ &\quad + \sum_{i=k-l}^{k} p^{l+1-k+i}\,(c_{N-k+i} - \bar{c})\, I_{[|x|_{k-l-1}=0]} \\ &= \sum_{i=1}^{k-1-(l+1)} (c_{N-k+i} - \bar{c})\, I_{[|x|_i=0]} + \sum_{i=k-(l+1)}^{k} p^{(l+1)-k+i}\,(c_{N-k+i} - \bar{c})\, I_{[|x|_{k-(l+1)}=0]}.\end{aligned}$$

Portanto, (A-3) se verifica para todo $l \in \{0,1,\cdots,k-1\}$. Em particular, para $l = k-1$ tem-se $g(x,k-1) = \sum_{i=1}^{k} p^{i-1}\,(c_{N-k+i} - \bar{c})\, I_{[x_1=0]}$. Avaliando a condição (A-2) para $l = k-1$, implica

$$\begin{aligned}\sum_{x\in\{0,1\}^k}\Pr(x)\sum_{i=1}^{k}I_{[|x|_i=0]}\left(c_{N-k+i}-\bar{c}\right) &= \sum_{\tilde{x}\in\{0,1\}}\Pr(\tilde{x})g(\tilde{x},k-1)\\ &= pg(0,k-1)+(1-p)g(1,k-1)\\ &= \sum_{i=1}^{k}p^i\left(c_{N-k+i}-\bar{c}\right)\end{aligned}$$

em que se usou $g(1,k-1)=0$. Finalmente, pode-se reescrever (A-1) como

$$\sum_{x\in\{0,1\}^k}\Pr(x)s(x) = A-kp\bar{c}-\sum_{i=1}^{k}p^i\left(c_{N-k+i}-\bar{c}\right).$$

Lembrando que $A'=\sum_{i=1}^{k}c'_{N-k+i}$, segue que

$$\begin{aligned}\tilde{y}'_k &= e+\frac{1}{k(1-p)}\sum_{x\in\{0,1\}^k}\Pr(x)s'(x)\\ &= e+\frac{1}{k(1-p)}\left[\sum_{i=1}^{k}c'_{N-k+i}-kp\bar{c}'-\sum_{i=1}^{k}p^i\left(c_{N-k+i}-\bar{c}\right)\right]\\ &= e+\frac{1}{k(1-p)}\left[\sum_{i=1}^{k}(1-p^i)c'_{N-k+i}+\bar{c}'\sum_{i=1}^{k}p^i-kp\bar{c}\right]\\ &= e+\frac{\bar{c}'}{k(1-p)}\left[\sum_{i=1}^{k}(1-p^i)\frac{c'_{N-k+i}}{\bar{c}'}+p\frac{1-p^k}{1-p}-kp\right].\end{aligned}$$

Usando c'_i e $\bar{c}'$ definidos no Lema 4.2, obtém-se $\sum_{i=1}^{k}(1-p^i)\frac{c'_{N-k+i}}{\bar{c}'} = \sum_{i=1}^{k}(1-p^i)\left(1-p^{k-i}\frac{(1-p)N}{1-p^N}\right)$. Logo,

$$\begin{aligned}\sum_{i=1}^{k}(1-p^i)\frac{c'_{N-k+i}}{\bar{c}'} &= \sum_{i=1}^{k}(1-p^i)-\frac{(1-p)N}{1-p^N}\sum_{i=1}^{k}(p^{k-i}-p^k)\\ &= k+\sum_{i=1}^{k}p^i-\frac{(1-p)N}{1-p^N}\left(\sum_{i=1}^{k}p^{k-i}-kp^k\right)\\ &= k-p\frac{1-p^k}{1-p}-\frac{k(1-p)N}{1-p^N}\left(\frac{1}{k}\sum_{i=0}^{k-1}p^i-p^k\right)\\ &= k-p\frac{1-p^k}{1-p}-\frac{k(1-p)N}{1-p^N}h\left(k\right)\end{aligned}$$

em que $h(k)=\frac{1}{k}\frac{1-p^k}{1-p}-p^k$. Logo,

$$\begin{aligned}\tilde{y}'_k &= e + \frac{\bar{c}'}{k(1-p)}\left[k - p\frac{1-p^k}{1-p} - \frac{k(1-p)N}{1-p^N}h(k) + p\frac{1-p^k}{1-p} - kp\right] \\ &= e + \frac{\bar{c}'}{k(1-p)}\left[(1-p)k - \frac{k(1-p)N}{1-p^N}h(k)\right] \\ &= e + \bar{c}' - \frac{\bar{c}'N}{1-p^N}h(k).\end{aligned}$$

Portanto, é atrativo mentir na posição i se e somente se

$$\begin{aligned}\left(1 - \frac{p^{N-i}(1-p)N}{1-p^N}\right)\bar{c}' = c'_i &> \tilde{y}'_k = e + \bar{c}' - \frac{\bar{c}'N}{1-p^N}h(k) \\ -\frac{p^{N-i}(1-p)N}{1-p^N}\bar{c}' &> e - \frac{\bar{c}'N}{1-p^N}h(k) \\ \frac{\bar{c}'N}{1-p^N}\left(h(k) - p^{N-i}(1-p)\right) &> e \\ f_k(N-i) &> \frac{e}{\bar{c}'}\frac{1-p^N}{N},\end{aligned}$$

pois $f_k(N-i) = h(k) - (1-p)p^{N-i}$. Como $f_k(N-i)$ é decrescente em i, então basta verificar sob quais condições a desigualdade é satisfeita para $i = N-k$ e violada para $i = N-(k-1)$. Ou seja,

$$f_k(k) > \frac{e}{\bar{c}'}\frac{1-p^N}{N} \geq f_k(k-1).$$

Do Lema 4.2, tem-se $\left(1 - p^N\frac{(1-p)N}{1-p^N}\right)\frac{\bar{c}'}{e} = \frac{\delta-1}{\delta}(1-p)$. Logo,

$$\frac{e}{\bar{c}'}\frac{1-p^N}{N} = \frac{1-p^N}{(1-p)N}\left(1 - p^N\frac{(1-p)N}{1-p^N}\right)\frac{\delta}{\delta-1} = \frac{\delta}{\delta-1}\left(\frac{1-p^N}{(1-p)N} - p^N\right).$$

Portanto, a condição de existência da referida corrida é

$$f_k(k) > \left[\frac{1-p^N}{(1-p)N} - p^N\right]\frac{\delta}{\delta-1} \geq f_k(k-1),$$

ou seja, a condição (30). □

B. FIGURAS

Figura B-1: Economias nas quais há corrida: $N \in \{3,6,9\}$.

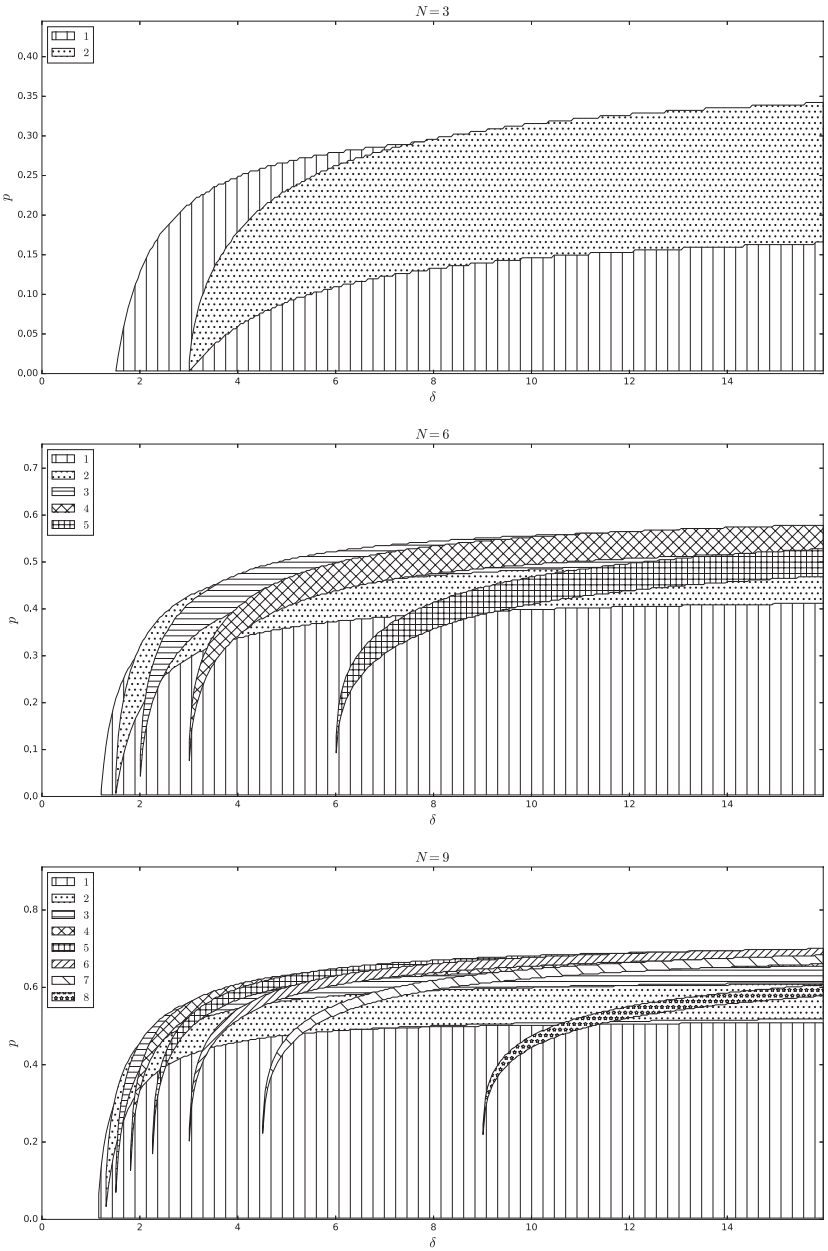
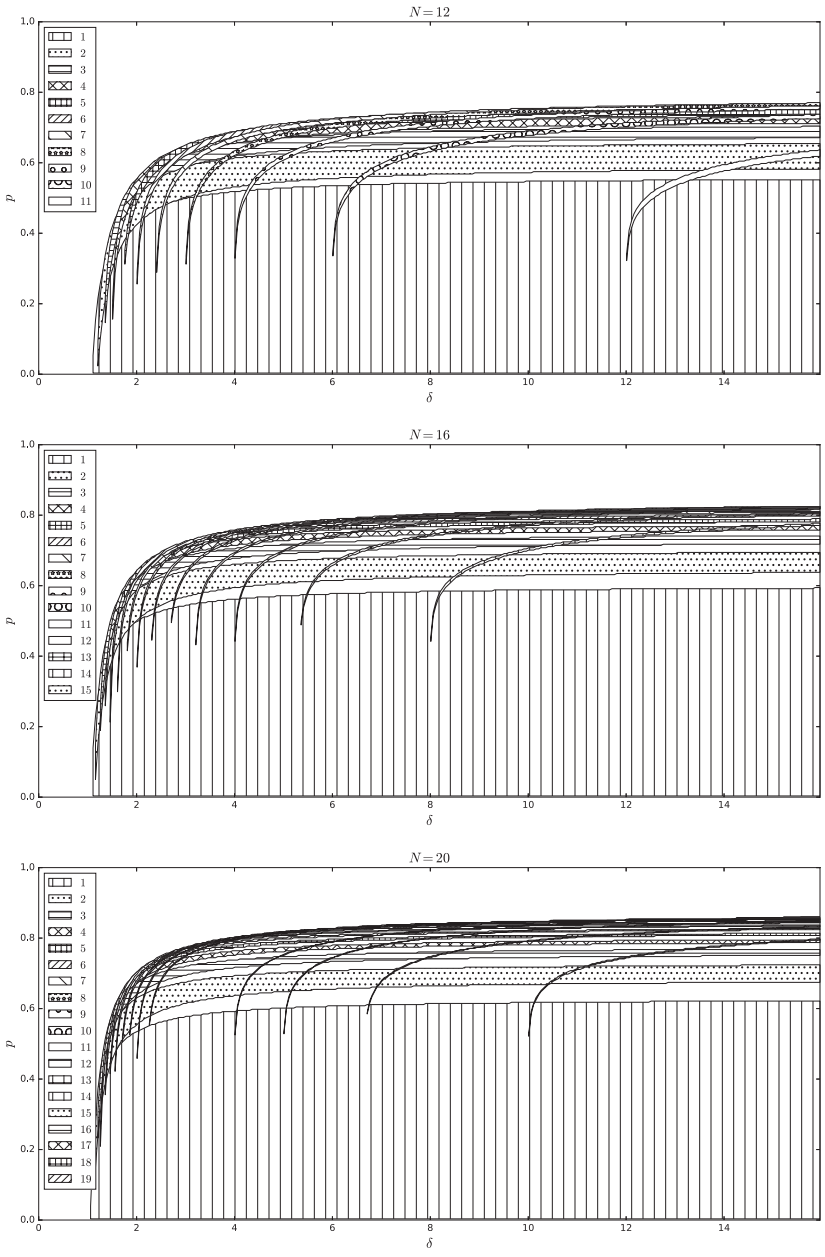


Figura B-2: Economias nas quais há corrida: $N \in \{12,16,20\}$.



r b e

ISSN 0034-7140 (print) | ISSN 1806-9134 (online)
DOI 10.5935/0034-7140.20170014
Revista Brasileira de Economia, 71(3), 301–320

On the Effects of Non-Tariff Measures on Brazilian Exports*

Lucas P. do C. Ferraz†
Marcel Ribeiro‡
Pedro Monasterio§

Contents: **1. Introduction; 2. Related literature; 3. Sample and non-tariff measures database; 4. Empirical strategy; 5. Results; 6. Final remarks.**

Keywords: **Non-tariff measures; SPS; TBT; Extensive margin of exports; Selection model; Firm heterogeneity.**

JEL Code: **C01, F13, L13.**

This article estimates the effects of TBT/SPS measures on Brazilian exports, based on the recent literature of gravity models. A Heckman selection model – theoretically grounded on the seminal Melitz model of heterogeneous firms – highlights the role played by zero trade flows as well as firm heterogeneity for estimations using gravity equations, two factors usually omitted in standard specifications. We show that both extensive and intensive margins of Brazilian exports are, on average, negatively affected by pre-existing TBT and SPS measures. Moreover, we contribute to the empirical literature by providing comprehensive evidence of the sectoral effects of such measures over domestic exports. We find that TBT/SPS measures - though predominantly harmful - can foster exports in several sectors of the Brazilian economy.

Este artigo estima os impactos das medidas TBT/SPS sobre as exportações brasileiras, baseado na literatura recente de modelos gravitacionais. Um modelo de seleção de Heckman é proposto – com fundamentos teóricos no modelo de firmas heterogêneas de Melitz – que destaca a relevância dos fluxos nulos de comércio, assim como da heterogeneidade das firmas, dois fatores usualmente omitidos em especificações convencionais. Nós mostramos que ambas as margens, intensiva e extensiva, das exportações brasileiras são, em média, negativamente afetadas pelas TBTs e SPSs. Ademais, contribuímos para a literatura empírica provendo ampla evidência no nível setorial para as exportações domésticas. Encontramos que medidas TBT/SPS, embora predominantemente prejudiciais, podem estimular as exportações em vários setores da economia Brasileira.

1. INTRODUCTION

An important trend in trade policy in recent decades is the remarkable reduction of tariff barriers imposed on international trade. This pattern is both a result of several trade liberalization rounds at the GATT/ WTO level and, more recently, the consequence of the sharp increase of regional trade agreements worldwide. Just over the last twenty years, more than four hundred preferential trade agreements were notified at the WTO. Over the same period, however, the multilateral trade system has also witnessed a growing number of notifications of non-tariff measures such as TBT (Technical Barriers to Trade) and SPS

*Lucas Ferraz gratefully acknowledges financial support from INMETRO.

†Sao Paulo School of Economics – FGV. E-mail: lucas.ferraz@fgv.br

‡Sao Paulo School of Economics – FGV. E-mail: marcel.ribeiro@fgv.br

§Sao Paulo School of Economics – FGV. E-mail: pedro.monasterio@fgv.br

Received on Feb 25th, 2016
Final Acceptance on May 03th, 2017

(Sanitary and Phytosanitary measures) being submitted by its members, with their legal underpinnings claimed to be based on both the TBT and SPS agreements formalized at the WTO's Uruguay round.

Despite the fact that notifications of TBT and SPS measures are expected to be grounded in pre-existing international standards and scientific evidence, its widespread dissemination among WTO members raised concerns of a new wave of protectionism, now disguised under the umbrella of trade regulation on product standards and safety.

Surprisingly, there is a scarce literature on the broad effects of such regulations on international trade, specially in the case of Brazil. Based on a comprehensive data set and addressing some recent criticisms on the correct specification of gravity models, this paper seeks to fill this gap, bringing new evidence on the impacts of TBT and SPS measures on Brazilian exports. We show that the number of TBT/SPS measures potentially affecting Brazilian exports has increased substantially over the last decades. Results of our econometric analysis suggest that, on average, TBT and SPS measures negatively influence Brazilian exports at both intensive and extensive margins of trade. When extending the analysis to the sectoral level, we find that TBT and SPS measures could foster trade in some sectors. In order to reconcile these apparently conflicting results, we provide a rather intuitive discussion based on trade theory. Our estimations also suggest that information contained in zero trade flows should be take into consideration in order to avoid possible bias on the estimated coefficients, a clear drawback of the traditional OLS approach. When it comes to firm heterogeneity at the intensive margin of trade, we show that Brazilian exports are hardly influenced by this variable.

2. RELATED LITERATURE

According to the World Trade Organization (WTO), Sanitary and Phytosanitary (SPS) measures may be defined as any measures applied: (1) "to protect human or animal life from risks arising from additives, contaminants, toxins or disease-causing organisms in their food; (2) to protect human life from plant - or animal - carried diseases; (3) to protect animal or plant life from pests, diseases, or disease-causing organisms; (4) to prevent or limit other damage to a country from the entry, establishment or spread of pests". By the same token, Technical Barriers to Trade (TBT) "cover all technical regulations, voluntary standards and the procedures to ensure that these are met, from car safety to energy-saving devices, to the shape of food cartons. TBT measures can still cover topics related to human health such as pharmaceutical restrictions or the labeling of cigarettes, nutrition claims and concerns, quality and packaging regulations".

SPS and TBT measures are generally classified as non-tariff measures (NTM) and most of the existing empirical literature about its effects over bilateral trade flows are based on standard gravity equations. Regardless of the real objectives behind the imposition of non tariff measures such as TBT and SPS by importing countries, several studies have provided evidence on their likely negative effects on trade, using rather standard OLS estimations (for a detailed review, see for example, Leamer, 1990, Otsuki et al., 2001, Moenius, 2004, Fontagné et al., 2005, Disdier et al., 2008).

The majority of the empirical studies in the field are restricted to a limited number of products and sectors. For instance, Otsuki et al. (2001) estimate the effect of product-specific regulations on trade for cereals, dried and preserved fruits, and nuts, for a sample of 31 exporting countries and 15 importing countries. They find that aflatoxin B1 standards in importing countries have a negative impact on trade flows of cereals and nuts, while they do not affect trade for dried and preserved fruits. The work by Disdier et al. (2008) concentrates on the effects of TBT/SPS measures on agricultural trade. They show that while developing countries' exports to OECD countries are significantly reduced by the presence of TBT/SPS measures, they do not affect trade among OECD members. Moreover, European imports tend to be more negatively influenced by TBT/SPS measures in comparison to imports of other OECD countries. Fontagné et al. (2005) find that pre-existing TBT/SPS measures reduce bilateral exports in a set of 61 countries, particularly in agricultural products. When evaluating the effects of NTM at the sector level,

both Disdier et al. (2008) and Fontagné et al. (2005) find positive as well as negative NTM effects on trade. However, while the latter differentiate between the effects of TBT and SPS measures on trade, the former treat both measures as the same, using just one dummy variable. To our knowledge, the work by Moenius (2004) is the first to evaluate the effects of all types of standards (environmental as well as technical) on trade for a rather comprehensive range of products (471 SITC industries) over a 16-year period. Using a sample of 12 OECD countries, he shows that the impact of country-specific standards on trade varies substantially across products, with negative effects concentrated in agricultural exports, whereas trade promoting effects concentrated in manufacturing exports. A major drawback of Moenius' work is the non-inclusion of bilateral tariffs in the estimated equations.

Interestingly, a relatively recent debate on the possible existence of misspecifications in standard gravity equations raised serious concerns over the credibility of many previous empirical studies using gravity models and addressing a broad set of issues in trade theory, including the effects of NTM on trade flows. In this regard, the work by Anderson and van Wincoop (2003) highlights the importance of controlling for the unobservable "multilateral resistance" terms in gravity equations, being the first article to formally address this issue, based on a theoretical model of monopolistic competition. In this regard, both Hummels (1999) and Feenstra (2015) advocate the use of directional (exporter and importer) fixed effects in cross-section estimations using gravity equations. For panel data analysis, Olivero and Yotov (2012) demonstrate that the multilateral resistance terms should be accounted for by exporter-time and importer-time fixed effects. Pointing out to another possible misspecification problem in gravity regressions, the work by Silva and Tenreyro (2006) suggests that under heteroskedasticity, the parameters of log-linearized gravity equations estimated by OLS may lead to biased estimates of the true elasticities. More recently, another influential work by Helpman et al. (2008), inspired on previous work by Melitz (2003), raised the issue of firm heterogeneity and also on the correct treatment of zero trade flows in traditional gravity equations. This has prompted a new generation of empirical studies, now addressing specification concerns such as the possible existence of sample selection bias and the influence of heterogeneous firms in gravity estimations. Among this new generation of empirical studies focused on the effects of NTM on trade flows are Disdier and Marette (2010), Xiong and Beghin (2012) and Crivelli and Groeschl (2016). All three studies measure the effects of product-specific regulations on the intensive and extensive margins of trade over a rather narrow set of products and sectors, using the Heckman selection model (Heckman, 1979). For instance, Disdier and Marette (2010) find no significant effect of maximum residue levels of pesticides (MRLs) on market entry (extensive margin) for food and agricultural products imported by OECD countries. However, they find a negative MRL effect on OECD imports of crustaceans (intensive margin). Xiong and Beghin (2012) evaluate the effect of EU aflatoxin standards on trade in groundnuts between the EU-15 and a set of 9 African countries from 1989 to 2006. They find an insignificant effect of the MRL imposed by EU on African exports of groundnuts. The more recent and comprehensive work by Crivelli and Groeschl (2016) evaluate the effects of a rather specific group of SPS measures (Specific trade concerns) [1] on bilateral trade in food and agricultural products involving 114 importer and 124 exporting countries over a time period of 15 years (1996-2010). They find that, on the whole, SPS measures negatively influence the extensive margin of trade (probability to export) but may foster trade at the intensive margin (current volume exported). Moreover, when they separate the SPS measures in two categories, they find that those more prone to generate extra fixed costs to exporters (concerns related to conformity assessments) negatively influence the extensive margin of exporters. By the same token, those SPSs more prone to generate extra variable costs to exporters (concerns related to the characteristic of products) are the ones that foster trade at the intensive margin. Despite the fact this new generation of studies take into consideration the possible role played by zero trade flows in gravity estimations, they keep concentrated in a rather

[1] SPS specific trade concerns reveal particular restrictive product standards that exporters from countries raising a concern face in a given export market.

restricted number of sectors and NTMs. Moreover, they also ignore the potential role played by firm heterogeneity in estimations using gravity equations.

According to the work by Helpman et al. (2008) if the probability to become an exporter is correlated to the decision on how much to export, the estimated impact of NTM on trade flows using standard gravity OLS regressions are likely to be downward biased. Regarding firm heterogeneity, the authors point out that standard gravity equations "confound the effects of trade barriers on firm-level trade with their effects on the proportion of exporting firms". Accordingly, if firm heterogeneity is not somehow included as a explanatory variable in standard gravity equations, its absence may induce an upward bias on the estimated effects of NTM on trade flows.

The issue of sample selection bias in gravity models can be properly addressed through the Heckman's selection model (Heckman, 1979). In this article, we use Heckman's model on its two-stage version. The first stage specifies a Probit model which (ceteris paribus) estimates the impact of a NTM on the probability of a firm to become an exporter. The first stage also estimates the inverse Mills ratio (the so called Heckman's lambda) that must be added as a additional explanatory variable in the (second stage) standard gravity equation. Its noteworthy that the second stage equation estimates the impact of a NTM on bilateral trade flows, conditional on the fact that firms are already exporters.

The issue of firm heterogeneity can be also addressed in the second stage gravity equation through the fraction of exporting firms in each sector. Based on the seminal model of heterogeneous firms by Melitz (2003), Helpman et al. (2008) show that the fraction of exporting firms in each exporting sector and bilateral trade flow can be estimated based on the inverse of the cdf of the unit-normal distribution applied to the estimated first stage probabilities of a firm to become an exporter. The estimated fraction of exporting firms in each sector for each bilateral trade flow can then be added as a regressor in the second stage standard gravity equation.

The current article contributes by providing new and sound evidence on the impacts of TBT and SPS measures over Brazil's exports, considering a rather comprehensive dataset including 4-digit level exports for all partners from 2006 to 2013. Our gravity specification incorporates recent methodological advances in the field, providing more reliable coefficient estimations. To our knowledge, this is the first work to provide evidence on the individual impacts of TBT/SPS measures on Brazilian trade for both extensive and intensive margins as well as for several exporting sectors.

The remainder of the paper is structured as follows. Section 3 provides detailed information on how the database for our panel estimations was built. Section 4 describes our empirical strategy. Section 5 shows the main results and discusses the economic intuition embedded in the reported coefficients. It also dialogs with the previous literature and discusses some policy implications and extensions derived from our results. The last section concludes.

3. SAMPLE AND NON-TARIFF MEASURES DATABASE

3.1. Sample

Brazilian exports (in current dollars) as well as import tariff data were obtained from the World Integrated Trade Solutions (WITS) of the World Bank. The data are annual from 2006 to 2013, according to the four-digit classification of the Harmonized System (HS04). We could have used data at the HS06 disaggregated level, but the notifications are defined mostly in four digits or less. Tariff data used in this work are sectoral simple averages. The advantage of using simple averages - rather than the weighted averages by trade flows - is to circumvent possible endogeneity in the estimation procedure.

Tariff data has many missings which would decrease our sample four fold. In order to fill as many as possible missings with a simple but reasonable rule, we proceeded as follows. When there was missing information on applied tariffs for a product j, importer i at period t, we used the mean tariff over countries for the same product and year. When the missing data was from a country member (not member) from Mercosur, we used the mean from countries members (not members) of Mercosur. One

can think of this procedure as representing the notion of the Most Favored Nation (MFN) clause, which is the usual treatment of tariffs in the GATT (General Agreement on Tariffs and Trade) legal framework. Strictly speaking, the MFN clause would input this importing country to replace its missing import tariff to a value equal to the minimum import tariff among countries which do not have a FTA or PTA with this particular country. Since this would change substantially the distribution of tariffs, we believed it would be more reasonable to input the mean. Most importantly, our results are robust to the exclusion of tariffs from the sample.

GDP data were obtained from the World Bank. We used GDP data in current dollars since the 4-digit export data is only available in current dollars as well. We also used sea transport costs associated with importing from the Trading Across Borders database (Doing business, World Bank). More specifically, we use the number of documents required to import as an excluded variable in our two-step Heckman selection model.

The information on non-tariff measures (TBT and SPS) applied to Brazilian exports was mostly sourced from World Trade Organization (WTO) database. It will be explained in detail in the next subsection.

Our sample consists of 179 importers (countries and territories), 1222 (HS04) products observed over eight years, from 2006 to 2013. The time span was limited by the availability of sea transport costs which starts at 2006 for most countries.

3.2. Building a database on non-tariff measures applied to Brazilian exports

WTO members must notify their non-tariff measures as required by the SPS and TBT agreements. Notifications are multilateral, i.e., they apply to all WTO members. The notifications from 1995 to 2013 are available in the Integrated Trade Intelligence Portal (I-TIP) of WTO. Notification is a document issued by a importing country which describes the requirements imposed on its imports of several products. This includes both the product coverage of the measure and their Harmonized Classification System (HS) code. This classification may be HS02, HS04 or HS06 depending on the details of the requirements. However, only 34.6% of the existing notifications describes its HS code. To circumvent this problem we used additional information available from other sources such as the Brazilian National Institute of Metrology, Quality and Technology (Inmetro) and the Centre for WTO Studies (CWS). While Inmetro provided us product codes for additional TBT notifications, the CWS provided the codes for the additional SPS notifications. Data from Inmetro start in 2001. Product codes were available at the HS04 level.

We defined a non-tariff measure as a specific criteria imposed by an importing country over the exports of its trade partners with respect to a specific product. Therefore, one notification may define several measures. Thus, our measure of NTM was constructed by the following steps. First, we assigned notifications to their respective HS04 products. The following criteria was used: i) notifications with regard to HS02 codes were assigned to all of its breakdowns of HS04 codes; ii) notifications with regard to HS06 codes were discarded since the export data are in HS04 level;[2] iii) the notifications belonging to the European Union were assigned to their respective members taking into account the date of entry for each country.

The literature usually treats the EU as a single country when it comes to member countries notifications of SPS and TBT measures. However, when working with panel data this may be misleading since the number of EU members increases over time in the sample. This is particularly important as the EU is one of the most active issuers of TBT and SPS measures as well as one of Brazil's main trade partners.

The construction of the dummy variable for measures, NTM_{ijt}, followed the criteria below:

$$NTM_{ijt} = \begin{cases} 1 & \text{if } t \geq \tau_{ij} \\ 0 & \text{if } t < \tau_{ij} \end{cases}$$

[2]The notifications limited to their HS06 code descriptions represented 0.32% of the notifications in the database.

where t denotes year and τ_{ij} is the first year when the importing country i notified a given TBT, SPS or both measures on product j. SPS_{ijt} and TBT_{ijt} dummies are defined analogously but only using TBT or SPS measures. Implicitly, its was assumed that, once notified, measures do not expire. For instance, a SPS measure issued in 2002 will not only impose restrictions in that year but over all years afterwards. In principle, countries may withdraw their notifications, but the WTO database do not provide this information. In practice, however, we believe that new measures usually impose more restrictive requirements over pre-existing ones, so that authorities do not bother to withdraw the less restrictive measures.

Figure 1: Adoption of new SPS and TBT by year

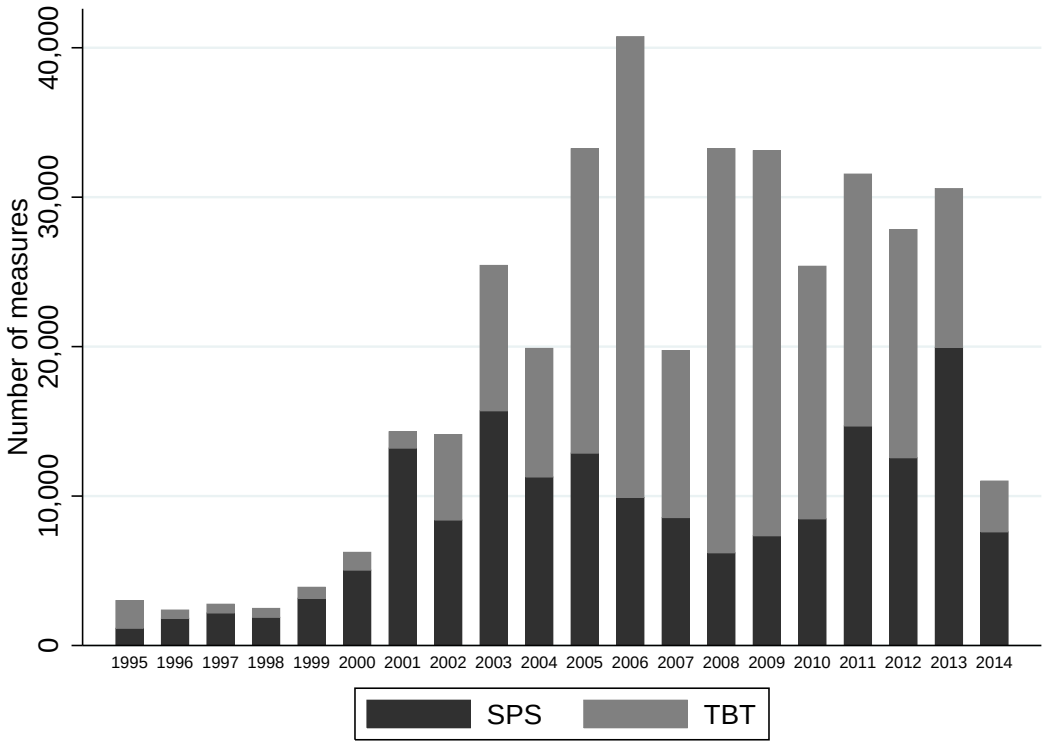


Figure 1 shows the number of new measures adopted from 1995 to 2014 for all countries in the sample, from the perspective of exporting firms in Brazil. It is clear that the number of measures have increased in the 2000's. However, it is important to point out that we had limited information on the notifications of TBT measures before 2001, as the dataset sourced from INMETRO starts at this year. When it comes to SPS measures, our dataset from CWS starts in 1995.

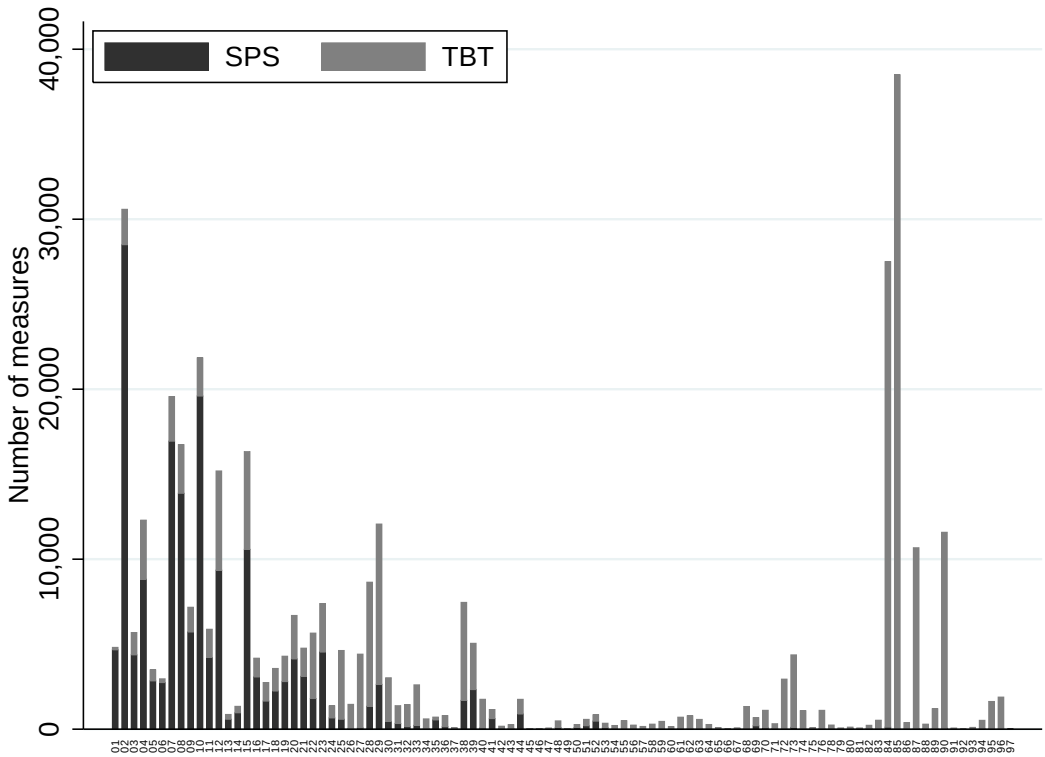
Notwithstanding the fact that our sample starts in 2006, we used information from NTM issued from 1995 up to 2013, as they get accumulated through time.

It is important to acknowledge the distribution of measures among products. Figure 2 reveals this information at the two-digit classification of the harmonized system (HS02). While SPS measures are

widespread over agricultural chapters, TBT measures tend to be concentrated in a few chapters related to manufacturing activities.[3]

Moreover, chapters related to agribusiness are heavily affected by both types of measures. Therefore, it is important to distinguish the effects of these types of measures on producing firms in exporting countries, since they affect different sectors as well as are likely to impose distinct requirements on imports.

Figure 2: Adoption of new SPS and TBT by product



In this study, the dummy variables for TBT and SPS measures treat in a symmetric way the intrinsic potential for each of those measures to affect trade flows. Obviously, this could lead to misleading conclusions about their impacts on trade flows since their effects could be very different in terms of trade restrictiveness. To our knowledge, there is no paper in the literature that was able to classify each of these measures according to their potential to affect trade flows in a satisfactory way. The treatment of each SPS and TBT measures by their potential to affect trade flows would demand, to say the least, a very time consuming and deep technical examination of each existing measure, with quite uncertain and likely questionable results. Therefore, this issue is out of the scope of this work.

[3]The most affected chapter by SPS measures is 02 (Meat). For TBT, the most affected chapters are 84 (Machinery and Mechanical appliances) and 85 (Electrical machinery and equipment).

4. EMPIRICAL STRATEGY

The literature quantifying the impacts of NTM on trade flows is relatively scarce and generally uses linear regression models.[4] One exception, for instance, is Crivelli and Groeschl (2016) which uses the same econometric framework as the one applied in this article but only for a limited number of sectors and for a small set of NTMs related to specific trade concerns.

Helpman et al. (2008) shows that to measure barriers to trade properly it is essential to take into account the extensive margin decision, i.e., whether or not to export. Therefore, empirical works using gravity equations should also take into consideration pre-existing zero bilateral trade flows between countries in the sample, which usually account for a large share of the observations. This point is even more pronounced in our sample since we use bilateral trade data at the HS04 level. Therefore, the probability of a particular product not being traded by a pair of countries is much larger than the probability of that pair not to trade at all. Therefore, the extensive margin is even more critical in our case.

Figure 3 shows the share of observations by year depending on whether bilateral trade flows are zero or positive as well as whether it is affected by measures or not. Some points are noteworthy. First, around 80% of the observations in the sample correspond to zero trade flows, which would be neglected in the case of simple OLS regressions. Second, over 10-15% of the zero trade flows are affected by a NTM. If the absence of trade were induced by the presence of a NTM one would underestimate the effects of the NTM on trade if not taking this into account. Finally, the NTM coverage is broad once the majority of the Brazilian exporting products are affected by at least one NTM.

Existing NTM are likely to impose higher fixed and/or variable costs for exporting firms in order to attend the required standards. This explain why a NTM is usually viewed as a trade barrier. Despite of that, there are alternative channels through which a NTM can actually induce higher exports from affected countries. For instance, it may be the case that compliance with a NTM could increase foreign demand for exporting firms, so long as importers associate complying behavior to higher quality producers. Another important channel that can lead to higher exports after a NTM is imposed is through its general equilibrium effects. Since NTM are multilateral in nature and exporting firms are heterogeneous in productivity, the expected change in relative international production costs due to compliance with a given NTM tend to favor the most productive players in the export market.

Therefore, applied works on the effects of NTM on trade flows should not take for granted that they operate as legal barriers to trade as its is assumed, for instance, in Kee et al. (2009). As a matter of fact, our results show that there are products which are affected positively by both TBT and SPS measures as well as products which are affected negatively by them. Most importantly, NTM may also affect the extensive margin of trade, which standard linear regression gravity models, by its very nature, are not able to capture.

Therefore, we follow the specification by Helpman et al. (2008) where the extensive margin decision of a exporting firm is represented by the following Probit model (first stage equation in Heckman's two-stage model):

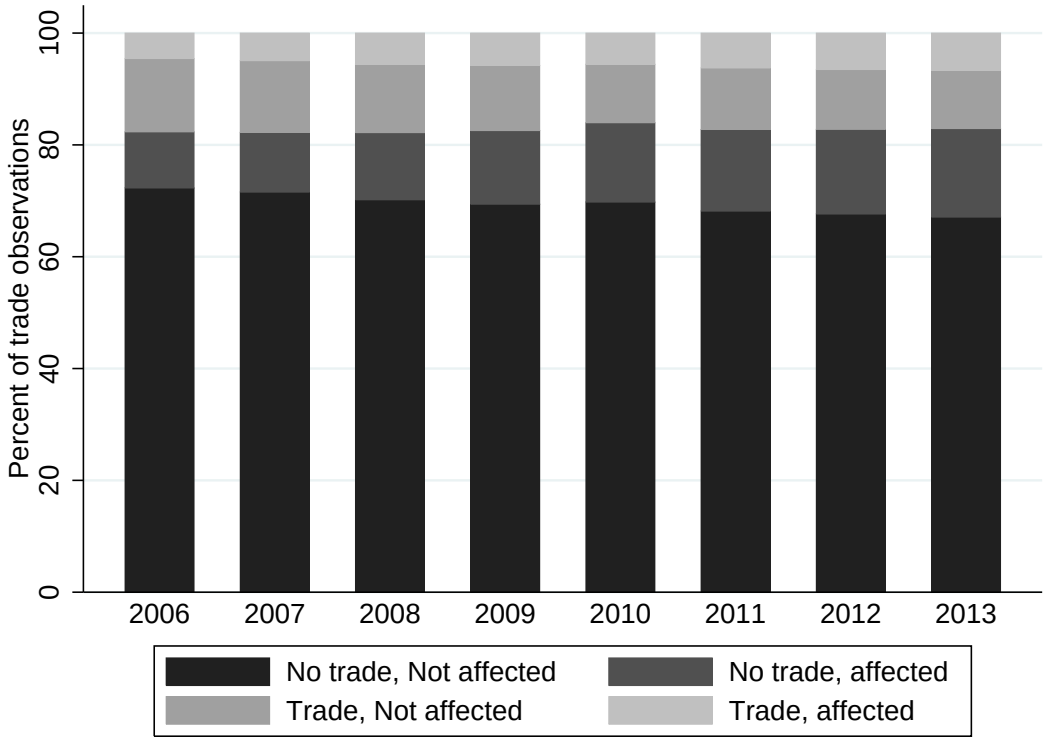
$$\rho_{ijt} \equiv Pr\left(T_{ijt} = 1 | W^*\right) = \Phi\left(\beta^* NTM_{ijt} + Z_{ijt}\theta^* + \alpha_i^* + \gamma_j^* + \eta_t^*\right) \quad (1)$$

where T_{ijt} is an indicator variable which is equal to one if there are positive Brazilian exports of product j to country i in period t. $W^* = \left(\begin{smallmatrix} NTM_{ijt} & Z_{ijt} & \alpha_i^* & \gamma_j^* & \eta_t^* \end{smallmatrix}\right)$ denote the vector of all observables. The intensive margin decision is given originally by the following specification (second stage equation):

$$y_{ijt} = \beta NTM_{ijt} + \beta_\lambda \hat{\lambda}_{ijt} + log(\beta_z exp(\hat{z}_{ijt} + \hat{\lambda}_{ijt}) - 1) + X_{ijt}\theta + \alpha_i + \gamma_j + \eta_t + \varepsilon_{ijt} \quad (2)$$

[4] See, for instance, Disdier et al. (2008) and Fontagné et al. (2005).

Figure 3: Trade and Non-tariff Measures



Note: Author's calculations based on 179 importing countries and 1222 HS04 products.

with $\varepsilon_{ijt} \sim \mathcal{N}(0,\sigma^2)$ that has correlation ρ with the structural error from first step. y_{ist} denote the (FOB) value of exports of a HS04 product j from Brazil, in period t, to destination i. The parameters $\alpha_i(\alpha_i^*)$, $\gamma_j(\gamma_j^*)$ and $\eta_t(\eta_t^*)$ are dummies of importers, products and years, respectively. X_{ijt} is a vector for a standard set of gravity control variables and Z_{ijt} is the vector which includes X_{ijt} as well as the excluding variable from the first stage equation.

The variables $\hat{z}_{ijt} = \Phi^{-1}(\hat{\rho}_{ijt})$ and $\hat{\lambda}_{ijt} \equiv \lambda(\hat{z}_{ijt}) = \frac{\phi(\hat{z}_{ijt})}{\Phi(\hat{z}_{ijt})}$ denote firm heterogeneity and the inverse Mills ratio (non-selection hazard) which are constructed based on the extensive margin estimation. The inverse Mills ratio control for the extensive margin decision from the first stage equation (possible selection bias due to existing zero trade flows). Following the specification by Helpman et al. (2008) and given that our data is disaggregated at the product level, the variable $\hat{z}_{ijt}$ represents the within-product Brazilian firm heterogeneity.

Equation (2) is an extended gravity equation that can be derived from the Melitz model of heterogeneous firms (Melitz, 2003), as described in detail by Helpman et al. (2008). However, as discussed by Helpman et al. (2008), their findings on the relevance of sample selection bias and firm heterogeneity in gravity equations are robust to any other formulations, including linear versions of equation (2). Due to the large dimension of our sample, the non-linear gravity specification proved extremely time con-

suming, requiring several days for a single simulation. Therefore, we opted for the linear specification of equation (2), replacing the term in logs (on its right-hand side) by the single variable $\hat{z}_{ijt}$.[5]

Control variables includes importer's GDP and applied tariffs. Since we are estimating the effects of non-tariff measures on Brazilian exports it is essential to include tariffs in the gravity equation, once NTM might be used as protectionist instruments replacing standard tariff barriers. Since Brazil is the only exporter in the sample, the importer country fixed effects controls for all time-invariant conventional gravity variables such as distance, common religion, language and legal origin, land border, colonial ties and so on.

For an appropriate identification of Heckman's second stage equation, we selected variables to be excluded from the second stage equation (where they must have no explanatory power) but to be included in the first stage. In this regard, Helpman et al. (2008) use information on the regulation costs of firm entry at the country pair level. More specifically, they build a country pair index on the number of documents and time required to start a business. This article uses regulation costs that are trade-specific and so we believe this must be a more appropriate instrument. Therefore, the number of documents required to import a standardized cargo by sea at the destination country was selected as the excluded variable in the first stage.[6] This information was sourced from the Trading Across Borders database (Doing Business, World Bank).

Even though we controlled for product fixed effects, there are roughly two reasons for the effects of NTM to be different across sectors. First, Brazilian sectors differs in terms of competitiveness. Second, a NTM can specify very different requirements depending on the sector it applies. Therefore, in order to capture sector-specific effects of NTM on trade flows, a extended gravity specification was built where the dummy variables for SPS_{ijt} and TBT_{ijt} measures were allowed to interact with sectorial dummies. Since there were 1222 sectors at the HS04 level in the original sample, we took a more parsimonious approach and introduced a new set of sectorial dummy variables at the 2-digit classification level covering 96 chapters. In the final specification, therefore, we added $182(=96\times2)$ more regressors into the model, instead of 2444 as it would be the case if we decided to stick to the original specification. Last, the HS02 specific coefficients can capture variation over its HS04 sub-sectors which would not be possible with HS04 specific coefficients instead. Hence, we believe this final extended specification may allow more reliable estimates when it comes to sector-specific effects of NTM on trade.

5. RESULTS

Table 1 reports the estimated NTM effects on Brazilian exports applying rather traditional econometric approaches with gravity equations, for the aggregated version of the gravity model (when the sector-level effects of NTM are not taken into consideration). The idea is to show how misleading traditional approaches may be, when some of the issues raised in the recent literature on gravity models are ignored. While columns 1, 2 and 3 reports the effects of NTM (TBT or SPS) on Brazilian exports, columns 4, 5 and 6 split the NTM dummy variable into two distinct dummies: one to capture the effects of TBT measures and another one to capture the effects of SPS measures. As shown in column 1 for a pooled OLS regression, when no fixed effects are taken into consideration, the impacts of NTM on trade flows are positive and statistically significant at 1%. When both country-specific and product-specific fixed effects are added to the model (column 2), it turns out that the estimated coefficient on the NTM dummy variable becomes negative and statistically not different from zero. Column 3 shows the results now

[5] This same linear version of equation (2) is also tested in Helpman et al. (2008) showing outstanding robustness. In order to avoid collinearity, the variables $\hat{z}_{ijt} = \Phi^{-1}(\hat{\rho}_{ijt})$ and $\hat{\lambda}_{ijt} \equiv \lambda(\hat{z}_{ijt}) = \frac{\phi(\hat{z}_{ijt})}{\Phi(\hat{z}_{ijt})}$ will be tested separately.

[6] Since Helpman et al. (2008) had limited information on country pair regulatory costs, they decided to work with "religion" as the excluded variable, despite its lack of a clear economic intuition (Crivelli and Groeschl, 2016, followed the same approach). In our case, we also tested "delays at customs" in the importing countries as an excluded variable but it performed worse in comparison to the number of documents required to import.

using a fixed panel specification where the estimated coefficient on the NTM dummy variable becomes positive but remains statistically not different from zero. When the NTM dummy variable is split into TBT and SPS dummies, the results are qualitatively similar, exception made to the coefficient on the SPS measure, that now becomes statistically different from zero through columns 4-6.

Results like the ones reported in Table 1 should be viewed with caution. As will be seen in Table 2, if either the probability to become and exporter is somehow correlated with the decision on how much to export or if the proportion of exporting firms varies substantially throughout Brazilian sectors, the coefficients reported in Table 1 are likely to be biased (see Helpman et al., 2008). Standard OLS approaches like the ones reported in Table 2 were widely used – with some small variations – in the previous empirical literature on the NTM effects on trade, such as in the works of Otsuki et al. (2001), Moenius (2004), Fontagné et al. (2005), Disdier et al. (2008).

Table 1: Aggregate effect of NTMs on Brazilian exports: OLS and Fixed effects

	OLS	OLS	FE	OLS	OLS	FE
	(1)	(2)	(3)	(4)	(5)	(6)
NTM	0.659***	-0.0301	0.0211			
	(0.0703)	(0.0328)	(0.0289)			
TBT				0.414***	-0.00257	0.0179
				(0.0622)	(0.0330)	(0.0303)
SPS				0.901***	-0.203**	0.128**
				(0.161)	(0.0828)	(0.0502)
Tariff	0.0157	-0.00477**	0.00321***	0.0111	-0.00452*	0.00320***
	(0.00965)	(0.00239)	(0.00108)	(0.00920)	(0.00236)	(0.00108)
log(GDP)	0.356***	0.710***	0.878***	0.350***	0.712***	0.878***
	(0.0161)	(0.0415)	(0.0390)	(0.0163)	(0.0415)	(0.0390)
Mercosur	2.375***	-0.228***	-0.374***	2.350***	-0.229***	-0.372***
	(0.0759)	(0.0660)	(0.0610)	(0.0758)	(0.0660)	(0.0611)
FE: Country and Product	No	Yes	No	No	Yes	No
FE: Country-Product	No	No	Yes	No	No	Yes
Observations	318896	318896	318896	318896	318896	318896
Adj. R^2	0.087	0.481	-	0.089	0.481	-

Note: Product cluster-robust standard errors in parenthesis. All regressions includes time fixed effects.

The results for the two-stage Heckman selection model are shown in Table 2. Similarly to Table 1, the first columns in the table (columns 1-4) show the results for the effects of NTM on Brazilian exports, while the last columns (columns 5-8) show the results for the individual effects of both TBT and SPS measures.

The two-stage Heckman selection model explains both the domestic firm decision on whether or not to become an exporter to a given destination (first stage probit regressions in columns 1 and 5) as well as on how much to export (second stage OLS regression in columns 3, 4, 7 and 8). More precisely, the first-stage equation is devoted to explain the sector-level extensive margin of Brazilian exports while the second stage equation explains the sector-level intensive margin decision of existing exporters. Results in Table 2 suggest that a NTM can affect both the sector-level extensive and intensive margins decisions.

Table 2: Aggregate effect of NTMs on Brazilian exports: Heckman and HMR(2008)

	Probit	OLS	Heckman	HMR	Probit	OLS	Heckman	HMR
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
NTM	-0.0558***	-0.0302	-0.147***	-0.0251				
	(0.0135)	(0.0328)	(0.0321)	(0.0514)				
TBT					-0.0452***	-0.00258	-0.116***	0.00167
					(0.0125)	(0.0330)	(0.0320)	(0.0456)
SPS					-0.150***	-0.203**	-0.397***	-0.189
					(0.0353)	(0.0828)	(0.0844)	(0.131)
Tariff	0.00682***	-0.00477**	0.00993***	-0.00539	0.00703***	-0.00452*	0.0105***	-0.00518
	(0.00179)	(0.00239)	(0.00259)	(0.00527)	(0.00176)	(0.00236)	(0.00252)	(0.00530)
log(GDP)	0.321***	0.710***	1.146***	0.681***	0.323***	0.712***	1.150***	0.682***
	(0.0159)	(0.0416)	(0.0461)	(0.224)	(0.0159)	(0.0416)	(0.0457)	(0.220)
Mercosur	-0.157***	-0.228***	-0.393***	-0.214*	-0.158***	-0.229***	-0.396***	-0.214*
	(0.0289)	(0.0661)	(0.0655)	(0.126)	(0.0289)	(0.0661)	(0.0655)	(0.125)
log(Doc.)	-0.0429***	-0.00390			-0.0438***	-0.00412		
	(0.0126)	(0.0298)			(0.0126)	(0.0298)		
Mills			2.829***				2.836***	
			(0.125)				(0.121)	
Heterog.				0.0908				0.0941
				(0.694)				(0.680)
Obs.	1638279	318896	318873	318896	1638302	318896	318873	318896
Adj. R^2	0.526	0.481	0.502	0.481	0.526	0.481	0.502	0.481

Note: Product cluster-robust standard errors in parenthesis. All regressions includes country, product and time fixed effects.

A qualitative assessment of the results reported in the first and second columns of Table 2 suggest that most of the variables that help to explain the extensive margin decision of Brazilian exporters (Probit estimation) also explain their intensive margin decisions (pooled OLS). The resulting correlation between residuals from the two equations highlights the importance of zero trade flows to be taken into account in gravity estimations, as it points out to the existence of sample selection bias. As a matter of fact, when the Mills ratio estimated in the first-stage is added to the second stage equations (columns 3 and 7) the results clearly suggest that ignoring sample selection bias in traditional gravity models (columns 2 and 6) may lead to a significant underestimation of the effects of NTM on trade flows (downward bias). When it comes to the effects of sector-specific firm heterogeneity on Brazilian exports, results in columns 4 and 8 suggest that they are not statistically different from zero. However, controlling for sector-specific firm heterogeneity does not seem to alter significantly the coefficients from traditional gravity estimations using OLS (columns 2 and 6), even though it tends to slightly understate their effects on trade flows, as predicted by Helpman et al. (2008).

Results in Table 2 provide evidence that controlling for sample selection bias is relatively more important than controlling for firm heterogeneity when it comes to explaining the impacts of NTM on Brazilian exports. This result is apparently at odds with Helpman et al. (2008) predictions about the relative magnitude of these two effects, suggesting that the higher relevance of firm heterogeneity in explaining trade flows, as obtained by them using country-level bilateral trade flows, may be dependent on the existing level of aggregation in the database. Arguably, the more disaggregated the database, the higher the incidence of zero trade flows and the lower the variation in the proportion of exporting firms at the sectoral level, as it seems to be the case in our study.

Overall, the departure from a simple OLS gravity approach to a more suitable two-stage Heckman selection model reveals that, on the whole, TBT and SPS measures negatively influence Brazilian exports

at both intensive and extensive margins of trade. Therefore, it is likely that the TBT/SPS measures imposed against Brazilian exports imply both substantial fixed (affecting the extensive margin) and variable costs (affecting the intensive margin) for exporting firms in Brazil. Moreover, ignoring the possibility of sample selection bias brought about by the presence of zero trade flows in the dataset may lead to substantial underestimation of NTM effects on trade flows, as suggested by Helpman et al. (2008). This discussion will be resumed in subsection 5.1, when we evaluate sectoral results.

Table 3 reports the results for the extended model, when 96 (HS02) sector dummy variables are allowed to interact with both TBT and SPS dummy variables, resulting in 192 additional regressors. Results in columns 1 to 4 are the same as the ones reported in columns 5 to 7 of Table 2 for the group of controls and are represented again for the sake of comparison. As can be seen from columns 5 to 8 in Table 3, extending the model to capture the sector-level impacts of TBT and SPS measures slightly improves the fit without significantly changing the coefficients for the group of control variables, attesting the robustness of our results.

Table 3: Disaggregate effect vs aggregate specification

	Aggregate NTM effects				Disaggregate NTM effects by HS02			
	Probit	OLS	Heckman	HMR	Probit	OLS	Heckman	HMR
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
log(GDP)	0.323***	0.712***	1.150***	0.682***	0.351***	0.738***	1.200***	0.635***
	(0.0159)	(0.0416)	(0.0457)	(0.220)	(0.0158)	(0.0418)	(0.0455)	(0.169)
Mercosur	-0.158***	-0.229***	-0.396***	-0.214*	-0.151***	-0.213***	-0.357***	-0.169*
	(0.0289)	(0.0661)	(0.0655)	(0.125)	(0.0291)	(0.0662)	(0.0660)	(0.0970)
log(Doc.)	-0.0438***	-0.00412			-0.0510***	-0.00747		
	(0.0126)	(0.0298)			(0.0128)	(0.0298)		
Mills			2.836***				2.777***	
			(0.121)				(0.111)	
Heterog.				0.0941				0.294
				(0.680)				(0.475)
Obs.	1638302	318896	318873	318896	1638279	318896	318896	318896
Adj. R^2	0.526	0.481	0.502	0.481	0.532	0.488	0.508	0.488

Note: Product cluster-robust standard errors in parenthesis. All regressions includes country, product and time fixed effects.

It is important to notice that the estimated coefficients from the two-stage Heckman selection model do not coincide with the marginal effects. For instance, since our results suggest that firm heterogeneity does not play a role in explaining Brazilian sectoral exports, the marginal effects for the second stage equation can be calculated as: [7]

$$me = \beta + \beta_\lambda \left[\lambda(\beta^* NTM_{ijt} + Z_{ijt}\theta^* + \alpha_i^* + \gamma_j^* + \eta_t^*) - \lambda(Z_{ijt}\theta^* + \alpha_i^* + \gamma_j^* + \eta_t^*)\right] \quad (3)$$

where, formally, $me = \mathbb{E}[y_{ijt}|W^*, NTM_{ijt} = 1] - \mathbb{E}[y_{ijt}|W^*, NTM_{ijt} = 0]$.

5.1. Sectoral results, extensions and policy implications

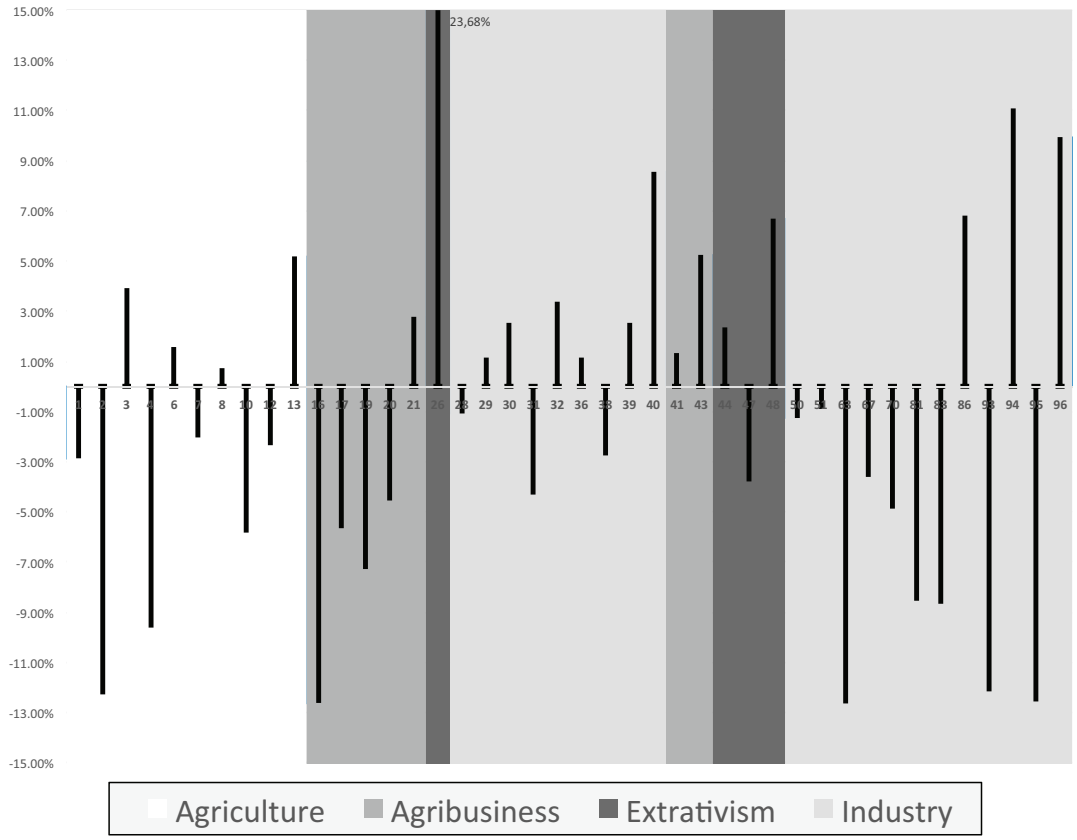
Results from the two-stage Heckman selection model shown in Table 2 suggested that, on average, a NTM works as a trade barrier for both extensive and intensive margins of Brazilian exports. Therefore,

[7] In fact, we are calculating the discrete change effects once we are estimating the effects of dummy variables.

TBT/SPS measures negatively affect the trade performance of both potential and current exporters in the country. As new production requirements are imposed to exporting firms, higher production costs are an obvious channel through which a NTM may affect exporters.

However, a sectoral level analysis on the effects of NTM over Brazilian exports reveals that its effects on trade are not unambiguously negative, as previously assumed in empirical works such as, for example, Kee et al. (2009). Figures 4 and 5 show the sectoral level marginal effects of SPS and TBT measures on the extensive margin of Brazil's exports, respectively. These marginal effects stem from the first-stage equation of the Heckman selection model (equation 1), where the TBT and SPS dummy variables are allowed to interact with sectoral dummies at the two-digit level of the harmonized system (Table 3). As for the intensive margin of Brazil's exports, Figures 6 and 7 show the sectoral level marginal effects of SPS and TBT measures, respectively, stemming from the second-stage equation of the Heckman selection model (linearized version of equation 2), computed by equation 3. In both cases, extending the analysis to the sectoral level shows that, though harmful on average, TBT and SPS measures can foster trade in some sectors of the domestic economy, a result already detected in previous empirical studies (see, for instance, Fontagné et al., 2005, Disdier et al., 2008).

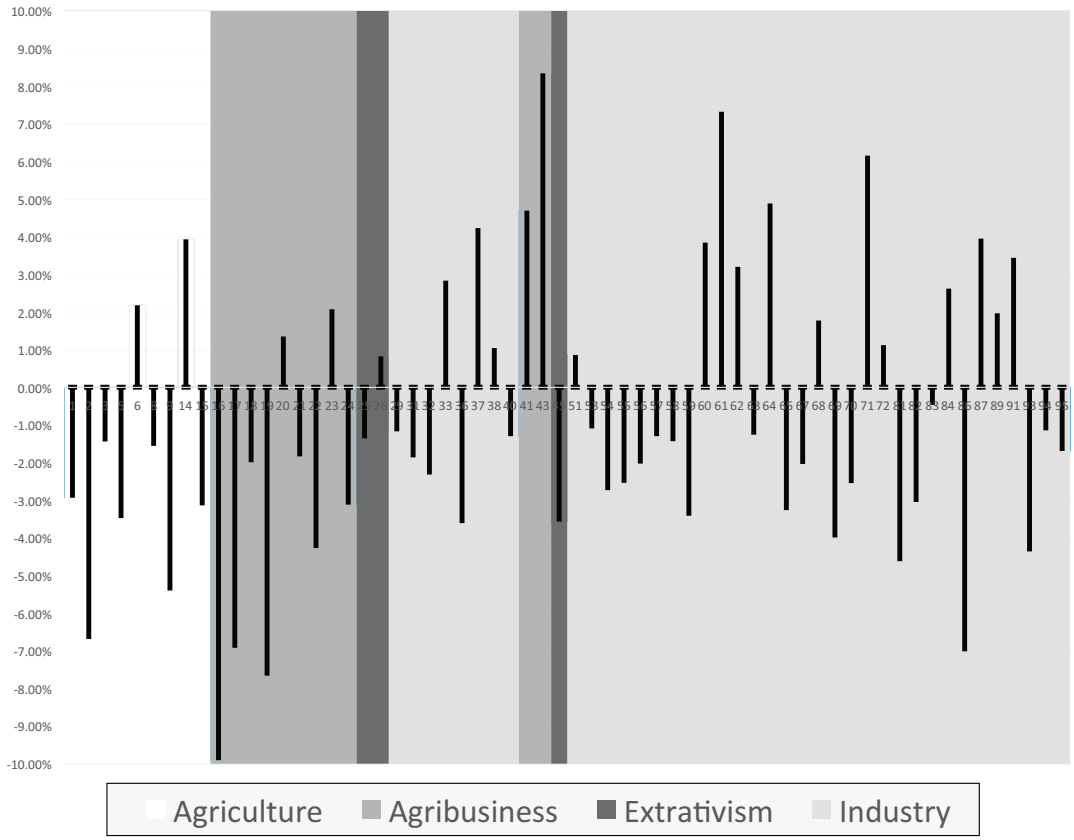
Figure 4: Marginal effect of SPS on sectoral Brazilian exports: Probit



Note: Marginal effects of products that have statistically significant coefficients at 5%.

Therefore, pre-existing SPS and TBT measures may actually foster Brazilian exports in several sectors of the domestic economy, though these measures may be predominantly harmful for the economy as

Figure 5: Marginal effect of TBT on sectoral Brazilian exports: Probit

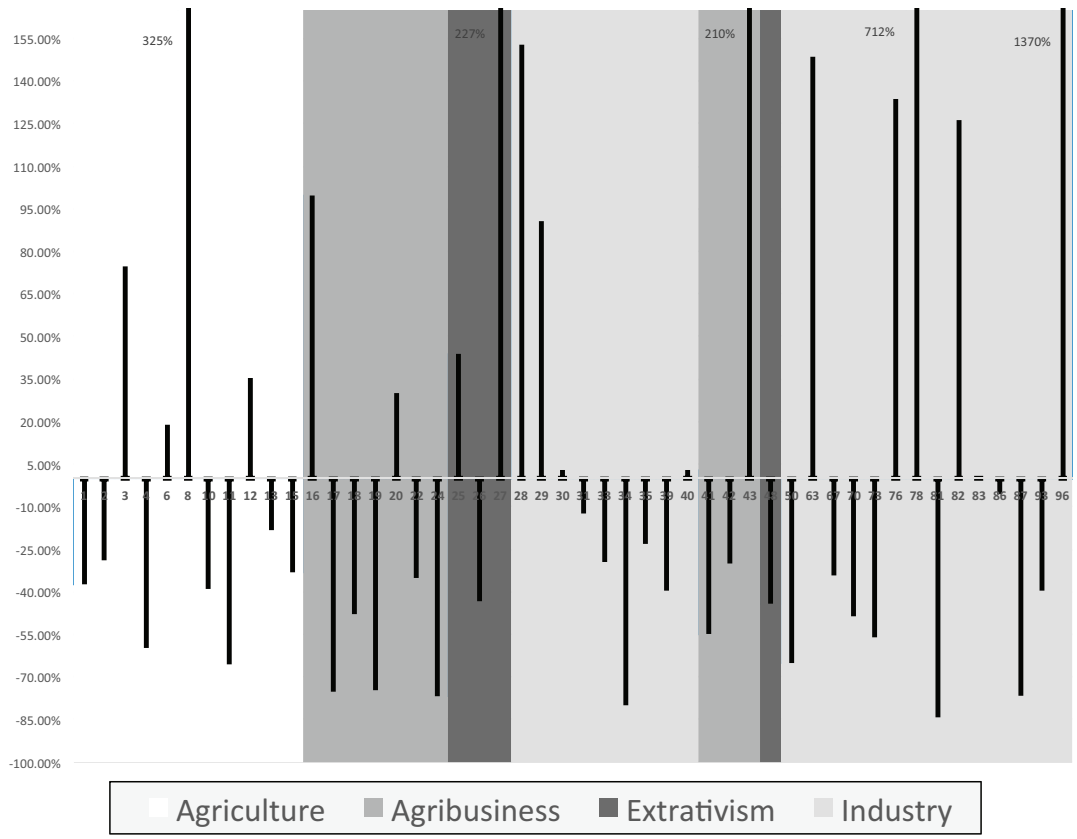


Note: Marginal effects of products that have statistically significant coefficients at 5%.

a whole, as results in Table 2 have suggested. In order to reconcile this apparently conflicting results, it is important to keep in mind that a NTM may potentially affect exports through, at least, four main channels, according to standard trade theory. First, as it imposes more strict requirements on production process by firms, it may lead to higher marginal costs to export. Second, more strict requirements on production may entail new investments in technology and equipments by firms, increasing fixed costs to export. Third, compliance with a NTM may positively affect consumer preferences in importing countries, shifting import demand and consequently bursting exports. Fourth, more efficient exporters tend to be more resilient to increasing costs of production then less efficient ones. Therefore, higher production costs are expected to be relatively more harmful to the exports of less efficient suppliers. In this case, depending on the magnitude of relative effects, more efficient suppliers may be better off with the imposition of a NTM. A more sophisticated theoretical framework representing the four channels described above can be found in the seminal work by Melitz (2003), where both extensive and intensive margins of trade are integrated in a monopolistic competition model (Krugman, 1980) with heterogeneous firms.

Since the influence of TBT and SPS measures on trade flows involves unobservable demand and supply factors, their final effects on trade are hard to anticipate. This is why the direction and the magnitude of TBT/SPS measures on trade flows is essentially an empirical issue.

Figure 6: Marginal effect of SPS on sectoral Brazilian exports: Heckman



Note: Marginal effects of products that have statistically significant coefficients at 5%.

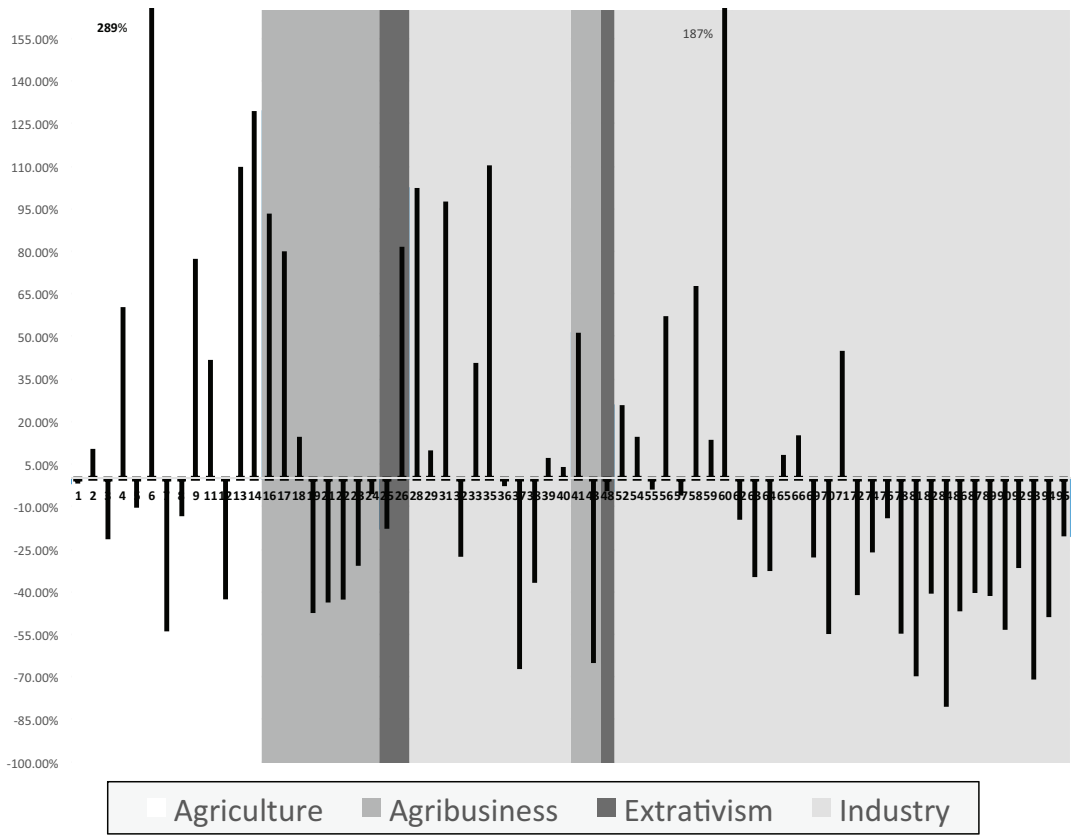
Sectoral level results for the extensive margin decision, represented in Figures 4 and 5, suggest that both SPS and TBT measures can negatively affect the probability of a Brazilian firm to become an exporter through the imposition of unaffordable fixed costs. This sectoral information has relevant implications for exporting promoting policies in the country, particularly the ones devoted to stimulate the participation of SMEs (Small and Medium sized Enterprises) in the international market. Taking into account the sector specific additional costs demanded for compliance with existing TBT and SPS measures in importing countries may significantly undermine the expected benefits of those policies.

When it comes to the intensive margin decisions, Figures 6 and 7 show that, though predominantly negative, existing TBT and SPS measures can again contribute to burst exports in several sectors of the Brazilian economy. Comparing their effects across sectors, however, while the negative effects of SPS measures tend to be evenly distributed among sectors in Brazil, the negative effects of TBT measures tend to be relatively more concentrated in capital intensive sectors. Given the well known lack of comparative advantages of manufacturing exports in Brazil, this result is hardly unexpected.

From the two-stage Heckman selection model used in this work it is straightforward to estimate the ad valorem equivalents[8] of TBT and SPS measures for any sectoral bilateral trade flows, based on their

[8]Diference between world and domestic prices caused by existing TBT/SPS measures at the country border.

Figure 7: Marginal effect of TBT on sectoral Brazilian exports: Heckman



Note: Marginal effects of products that have statistically significant coefficients at 5%.

previously estimated marginal effects on exports (see, for example Kee et al., 2009). In this regard, a relatively new trend in the empirical literature of NTM is to estimate the general equilibrium effects of preferential trade and regulatory agreements using the pre-estimated ad valorem equivalents of NTM as inputs into computable general equilibrium models (see, for example Harrison and Pearson, 1994, Andriamananjara et al., 2003, 2004, Francois et al., 2005, Fugazza and Maur, 2008). In most of the cases, the idea is to estimate the effects of trade agreements involving regulatory clauses related to mutual recognition and/or harmonization of standards. For instance, Harrison and Pearson (1994) simulated the effects of regulations harmonizations in the EU in the post-Maastricht era. Their results suggested that the impact of harmonization of standards among EU countries could reach an impressive 2.4% of EU's GDP. However, as already discussed in previous paragraphs and also argued in Baldwin et al. (2000), the notification of TBTs and SPSs by importing countries are likely to generate extra fixed as well as variable costs for exporting firms. Therefore, when working in conjunction with the ad valorem equivalents of those notifications, CGE models should somehow accommodate an imperfect competition market structure able to represent export-specific fixed costs due to the existence of NTM, which was hardly the case in most of the previous empirical studies in the field. To our knowledge, the first attempt to represent those fixed costs can be found in Zhai (2008) and more recently in Akgul et al. (2014).

It is noteworthy that the ad valorem equivalents estimated by the two-stage Heckman selection model are compatible with CGE models that do not incorporate the extensive margin decision, which are still predominant in the empirical literature. Therefore, in those models the ad valorem equivalents of TBTs and SPSs must represent estimations of pure extra variable costs and shall not be influenced by any kind of fixed costs. This is the case as the marginal effect from equation (3) is conditional on the decision of entry. This must be true since fixed costs must only influence entrepreneur's decision to become an exporter (Heckman's first stage equation representing the effects of NTM on the extensive margin of trade) but not his decision on how much to export, given that he is already an exporter (Heckman's second stage equation representing the effects of NTM on the intensive margin of trade). Again, this reasoning is fully backed by the theoretical foundations found in Melitz' seminal model of heterogeneous firms.

The common view of TBT and SPS measures as real barriers to trade should not lead one to underestimate the potential social benefits they might provide. Unlike tariffs, the optimal policy for TBT/SPS measures will often not be their full elimination but rather their rationalization to the social utility maximizing level, a point already made in Fugazza and Maur (2008). In this regard, it seems advisable that Brazilian trade authorities concentrate more efforts on the formalization of regulatory agreements with Brazil's main trade partners, in order to minimize the cost-benefit ratio of existing regulations and, whenever possible, to avoid duplication costs and unnecessary red tape.

6. FINAL REMARKS

The number of TBT and SPS notifications at the WTO has grown significantly over the last decades, following the explosion of preferential trade agreements and the decline in global import tariffs. While recognized as legal instruments devoted to protect the welfare of importing countries, its widespread use has raised suspicions of a new wave of protectionism among WTO's members.

This paper provides sounding and comprehensive empirical evidence of the effects of TBT/SPS measures over Brazilian exports, detailed at the 2-digit level of the harmonized system, when most of previous works in the literature have focused on particular sectors and specific measures. While not formally addressing the potential opportunistic use of NTM against Brazilian exports, this paper has shown that TBT and SPS measures can indeed work as significant trade barriers for exporting firms in Brazil, depending on the sector where they are applied. In more extreme sector-level cases, existing SPS and TBT measures can substantially limit the export potential of Brazilian firms, as the magnitudes of their estimated Probit effects have suggested.

This paper has also touched upon another set of relevant issues on the impacts of NTM on trade flows. Starting with a new econometric specification based on the previous work by Helpman et al. (2008), it has shown how previous gravity approaches on the effects of NTM over trade flows might be misleading. Moreover, we discussed the main channels through which a NTM may affect trade flows, based on the insights from the Melitz model, in order to better clarify the net effects of NTM on both the extensive and intensive margins decisions of exporting firms. Accordingly, the determination of both direction and magnitude of the effects of TBT/SPS measures on trade flows will depend crucially on the interaction of key supply and demand factors operating in exporting and importing countries, respectively, being essentially an empirical issue.

BIBLIOGRAPHY

Akgul, Z., Villoria, N. B., & Hertel, T. (2014). Introducing Firm Heterogeneity into the GTAP Model with an Illustration in the Context of the Trans-Pacific Partnership Agreement. Technical report, Department of Agricultural Economics, Purdue University, West Lafayette, IN.

Anderson, J. E. & van Wincoop, E. (2003). Gravity with Gravitas: A Solution to the Border Puzzle. *American Economic Review*, 93(1):170–192.

Andriamananjara, S., Dean, J., Feinberg, R., Ferrantino, M. J., Ludema, R., & Tsigas, M. E. (2004). The Effects of Non-Tariff Measures on Prices, Trade, and Welfare: CGE Implementation of Policy-Based Price Comparisons. Technical Report 15863.

Andriamananjara, S., Ferrantino, M. J., & Tsigas, M. E. (2003). Alternative Approaches in Estimating the Economic Effects of Non-Tariff Measures: Results from Newly Quantified Measures. Technical Report 15872.

Baldwin, R. E., McLaren, J., & Panagariya, A. (2000). Regulatory Protectionism, Developing Nations, and a Two-Tier World Trade System. *Brookings Trade Forum*, pages 237–293.

Crivelli, P. & Groeschl, J. (2016). The Impact of Sanitary and Phytosanitary Measures on Market Entry and Trade Flows. *The World Economy*, 39(3):444–473.

Disdier, A.-C., Fontagné, L., & Mimouni, M. (2008). The Impact of Regulations on Agricultural Trade: Evidence from the SPS and TBT Agreements. *American Journal of Agricultural Economics*, 90(2):336–350.

Disdier, A.-C. & Marette, S. (2010). The combination of gravity and welfare approaches for evaluating nontariff measures. *American Journal of Agricultural Economics*, 92(3):713–726.

Feenstra, R. C. (2015). *Advanced international trade: theory and evidence*. Princeton University Press.

Fontagné, L., Mimouni, M., & Pasteels, J.-M. (2005). Estimating the Impact of Environmental SPS and TBT on International Trade. *Integration and Trade Journal*, 22:7–37.

Francois, J., Van Meijl, H., & Van Tongeren, F. (2005). Trade liberalization in the Doha Development Round. *Economic Policy*, 20(42):350–391.

Fugazza, M. & Maur, J.-C. (2008). Non-Tariff Barriers in Computable General Equilibrium Modelling. Technical Report 38.

Harrison, W. J. & Pearson, K. R. (1994). Multiregional and Intertemporal AGE Modelling via GEMPACK. Technical report.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the Econometric Society*, pages 153–161.

Helpman, E., Melitz, M., & Rubinstein, Y. (2008). Estimating Trade Flows: Trading Partners and Trading Volumes. *The Quarterly Journal of Economics*, 123(2):441–487.

Hummels, D. (1999). Toward a Geography of Trade Cost. *GTAP Working Papers*.

Kee, H. L., Nicita, A., & Olarreaga, M. (2009). Estimating Trade Restrictiveness Indices. *The Economic Journal*, 119(534):172–199.

Krugman, P. (1980). Scale economies, product differentiation, and the pattern of trade. *The American Economic Review*, 70(5):950–959.

Leamer, E. E. (1990). Latin America as a target of trade barriers erected by the major developed countries in 1983. *Journal of Development Economics*, 32(2):337–368.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, 71(6):1695–1725.

Moenius, J. (2004). Information versus product adaptation: The role of standards in trade. Technical report, Working Paper, International Business & Markets Research Center.

Olivero, M. & Yotov, Y. (2012). Dynamic gravity: Theory and empirical implications. *Canadian Journal of Economics*, 45(1):64–92.

Otsuki, T., Wilson, J. S., & Sewadeh, M. (2001). What price precaution? European harmonisation of aflatoxin regulations and African groundnut exports. *European Review of Agricultural Economics*, 28(3):263–284.

Silva, J. S. & Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics*, 88(4):641–658.

Xiong, B. & Beghin, J. (2012). Does European aflatoxin regulation hurt groundnut exporters from Africa? *European Review of Agricultural Economics*, 39(4):589–609.

Zhai, F. (2008). Armington Meets Melitz: Introducing Firm Heterogeneity in a Global CGE Model of Trade. *Journal of Economic Integration*, 23(3):575–604.

r b e

ISSN 0034-7140 (print) | ISSN 1806-9134 (online)
DOI 10.5935/0034-7140.20170015
Revista Brasileira de Economia, 71(3), 321–345

Multiplicador dos Gastos do Governo em Períodos de Expansão e Recessão: Evidências Empíricas para o Brasil

Vanessa Grudtner*
Edilean Kleber da Silva Bejarano Aragon†

Sumário: **1. Introdução; 2. Estratégia empírica; 3. Resultados; 4. Considerações finais; A. Estimação do VAR não linear; B. Estimação das funções de impulso-respostas generalizadas; C. Impulso-respostas generalizadas estimadas.**

Palavras-chave: **VAR não linear; Multiplicador dos gastos do governo; Ciclo econômico.**

Códigos JEL: **C32, E32, E62, H30.**

Neste trabalho, nós estimamos um vetor autorregressivo de transição não-linear com a finalidade de verificar se o multiplicador dos gastos do governo brasileiro são dependentes do ciclo econômico. O período de análise vai do primeiro trimestre de 1999 ao quarto de 2015. Além dos gastos consolidados do governo, calculamos o multiplicador de consumo, investimento e salários pagos pelo governo aos servidores públicos. Os resultados mostram que os multiplicadores dos gastos do governo se comportam da mesma maneira em períodos de recessão e expansão. Esse resultado é robusto a choques nos gastos do governo dados em períodos de expansão forte e recessão profunda, ao multiplicador de gastos do governo controlado pelo investimento, à razão dívida/PIB e ao grau de abertura econômica e taxa de câmbio.

In this paper, we estimate a nonlinear transition autoregressive vector with a purpose of verification if the Brazilian government expenditure multiplier is dependent of the economic cycle. The analysis period runs from the first quarter of 1999 to the fourth quarter of 2015. In addition to the government's consolidated costs, were calculated the multiplier for government's consumption, government investment and wages paid by government to public employees. The results show that government spending multipliers behave in the same way in periods of recession and expansion. This result is robust to shocks in government spending given in periods of strong expansion and deep recession, the multiplier of government spending driven by investment, the debt/GDP ratio, and the degree of economic openness and exchange rate.

1. INTRODUÇÃO

O interesse dos economistas sobre os efeitos macroeconômicos da política fiscal foi reaceso em 2008, com os primeiros estudos para avaliar o impacto dos pacotes fiscais, como o *American Recovery and Reinvestment Act*, lançados pelo governo dos EUA para impulsionar o produto e o emprego da economia durante a crise.[1] Até então, a política fiscal havia ocupado papel central na macroeconomia apenas no período pós-Grande Depressão, guiada pelo keynesianismo Blanchard e Perotti (2002). Nos anos seguintes, o ceticismo com relação à eficácia da política fiscal foi crescendo, principalmente após o trabalho de Barro (1974), baseado na equivalência ricardiana. Paralelamente a isso, o desenvolvimento do mercado

*Aluna de mestrado do PPGE – UFPB. E-mail: vgrudtner@gmail.com

†Professor do PPGE – UFPB. E-mail: edilean@hotmail.com

[1] Ver Romer e Bernstein (2009), Eggertsson (2009) e Taylor (2011), por exemplo.

Recebido em 23 de fevereiro de 2017
Aceitação Final em 14 de julho de 2017

financeiro aumentou a efetividade da política monetária, que foi suficiente para diminuir as flutuações cíclicas do produto norte-americano a partir dos anos 1980 até 2008, quando a demanda agregada norte-americana entrou em colapso e os Banco Central dos EUA baixaram sua taxa de juros para um valor próximo de zero.

Com o passar do tempo, a discussão sobre o impacto de políticas fiscais se refletiu no debate sobre o tamanho do multiplicador fiscal e sob quais situações econômicas o multiplicador seria maior ou menor.[2] Auerbach e Gorodnichenko (2012) encontraram multiplicadores de gastos em infraestrutura e outros tipos de investimento público nos EUA maiores que os multiplicadores de consumo do governo. Chahrour et alii (2012) e Leeper et alii (2011) encontraram multiplicadores de gastos do governo maiores do que multiplicadores de impostos. Christiano et alii (2011) mostraram que os multiplicadores são maiores quando o aumento dos gastos do governo são feitos em um período onde a taxa nominal de juros está no seu limite inferior.

No que se refere à possibilidade de que o multiplicador variar de acordo com o estado da economia, Pereira e Lopes (2010) e Kirchner et alii (2010) usaram um VAR estrutural com parâmetros variando no tempo e simulações Bayesiana e chegaram à conclusão que os multiplicadores de gastos do governo não são muito diferentes em expansões e recessões. Por sua vez, Auerbach e Gorodnichenko (2012) estimaram um modelo STVAR e encontraram um multiplicador de pico para os gastos do governo nos EUA de 2,5 em recessões e 0,6 em expansões. Fazzari et alii (2013) estimaram um SVAR com *threshold* e encontraram que os efeitos de um choque nos gastos do governo sobre a produção são significativamente maiores e mais persistentes quando a economia está com um alto grau de recursos subutilizados, o que se verifica em períodos de recessão. Ramey e Zubairy (2014), utilizando o método de Jordà (2005), não encontraram indícios de que os multiplicadores são maiores em recessões, nem quando foram utilizados os mesmos períodos de amostragem, dados, definições das variáveis, definição de estado da economia e choques utilizados por Auerbach e Gorodnichenko (2012). Para os autores, as diferenças encontradas se deram pelos pressupostos, muitas vezes irrealistas, que Auerbach e Gorodnichenko (2012) usaram para calcular as suas funções de impulso-resposta tais como: $i)$ a taxa de crescimento médio do produto ser exógeno ao modelo e dado pela média móvel de sete trimestres; $ii)$ a economia permanecer em recessão ou expansão pelo menos por vinte trimestres e; $iii)$ as mudanças nos gastos do governo não impactarem o estado da economia. Caggiano et alii (2015) estimaram um VAR não linear que se aproxima do estimado por Auerbach e Gorodnichenko (2012), mas levaram em consideração as críticas feitas por Ramey e Zubairy (2014). Esses autores constataram que o multiplicador dos gastos dos governo fiscal é maior quando a economia se encontra em recessão profunda.

Teoricamente, os efeitos reais assimétricos da política de gastos governamentais em diferentes fases do ciclo econômico podem surgir em um modelo Keynesiano tradicional (IS-LM-OA-DA) com uma curva de oferta agregada (OA) convexa. Nesse caso, em períodos de produto baixo (quando a economia está em recessão), a curva OA é pouco inclinada e o multiplicador fiscal é relativamente grande. Por outro lado, quando a economia está em expansão, a curva OA é muito inclinada e o multiplicador fiscal é relativamente pequeno. Baseados em modelos DSGE propostos para estudar os efeitos de política fiscal, Eggertsson (2009) e Christiano et alii (2011) verificam que o multiplicador dos gastos do governo é maior quando o limite inferior da taxa nominal de juros está ativo (e a economia está recessão). A razão para isso é que, quando a taxa nominal de juros está próxima de zero, uma política fiscal expansionista não altera essa variável e, consequentemente, não gera um descolamento dos gastos privados (consumo e investimento). Recentemente, Canzoneri et alii (2016) apresentam um modelo teórico com fricções no mercado financeiro em que a variação cíclica nos custos de intermediação bancária gera um comportamento anticíclico para o *spread* entre a taxa de depósitos e a taxa de empréstimos bancários, e cria

[2]De forma simplificada, o multiplicador fiscal estima o efeito que uma variação em um instrumento de política fiscal tem sobre o produto. Esse multiplicador pode ser de "pico",que informa o maior efeito sobre o produto em qualquer trimestre após a mudança de política, "de impacto", que traz o efeito imediato da mudança de política ou "acumulativos", que representa o somatório dos efeito de política fiscal sobre o produto durante um determinado período.

um acelerador financeiro que é muito mais forte em recessões do que em expansões. Nesse caso, uma política fiscal expansionista teria maiores efeitos sobre o produto agregado em períodos de recessão.

Posto isto, o objetivo deste trabalho é verificar se o multiplicador dos gastos do governo no Brasil depende da fase do ciclo de negócios. Para isso, estima-se um vetor autorregressivo de transição não-linear (STVAR) modelado para discriminar as respostas dinâmicas do produto a choques fiscais em períodos de recessão e expansão econômica. A análise é baseada nas funções de impulso-respostas generalizadas (GIRFs) estimadas de forma a captar a endogeneidade da transição de um estado da economia para outro, após um choque fiscal. O estudo é feito com dados trimestrais para o período de 1999 a 2015.

A análise do tamanho do multiplicador fiscal em tempos bons e ruins é particularmente importante para a economia brasileira. Por mais que o país não tenha atingido o limite inferior da taxa nominal de juros, a política monetária sempre teve como principal objetivo o controle da inflação, ficando em segundo plano quando se trata de política econômica anticíclica. Em 2008, o governo brasileiro utilizou vários instrumentos de política fiscal para combater a crise, aumentando seus gastos e cortando impostos.[3] Mesmo com a política fiscal expansionista sendo feita de maneira discricionária, sem se saber ao certo seu efeito sobre a economia, em 2010 a economia brasileira já havia superado a crise [Pastore et alii (2004)]. A partir de 2012, de acordo com Gobetti e Orair (2015), o multiplicador dos gastos do governo brasileiros mostrou certa ineficiência em produzir os efeitos econômicos esperados. De 2010 ao primeiro semestre de 2014, a economia brasileira se manteve relativamente estável, mesmo o governo mantendo a taxa real de expansão das despesas acima de 4%. Apesar de não se saber ao certo quais as medidas adotadas tiraram o Brasil da crise 2008 e aprofundaram a de 2015, há indícios de que o multiplicador fiscal para o Brasil é dependente do estado em que se encontra a economia. Isso porque as mesmas medidas de aumento de gastos surtiram efeitos contrários sobre o produto em períodos diferentes de implementação: em 2008 a economia estava em recessão e em 2014 não.

No Brasil, a maioria dos trabalhos sobre multiplicador fiscal procura estimar o valor desse parâmetro. Por exemplo, Peres e Ellery Junior (2009) estimaram um modelo VAR e concluíram que os multiplicadores dos gastos são positivos em torno de 0,3 e da receita situam-se em torno de -0,2. Cavalcanti e Silva (2010) incluíram a dívida pública em um VAR e observaram que a omissão da dívida pública pode superestimar os efeitos de choques fiscais. Pires (2012) também estimou um modelo VAR e concluiu que o multiplicador fiscal no Brasil tende a ser maior que um, sendo o multiplicador dos investimentos mais elevado que o do consumo do governo.[4] Ainda dentro dessa literatura, é importante ressaltar que são poucos os trabalhos que relacionam o multiplicador fiscal com a evolução do ciclo econômico para o Brasil. Pires (2014) e Samuel et alii (2016) utilizam modelos Markov-Switching e não encontram diferenças entre os multiplicadores em períodos de recessão e expansão. O trabalho que mais se assemelha ao que estamos propendo é o de Orair et alii (2016). Esses autores estimaram multiplicadores de diferentes tipos de gasto público através de um modelo STVAR, seguindo a metodologia de Auerbach e Gorodnichenko (2012). Os autores foram os únicos que encontraram estado-dependência nos multiplicadores fiscais para o Brasil. Seus resultados mostram que, durante depressões econômicas severas, o efeito de alguns tipos de gasto público, como investimentos, benefícios sociais e gastos com pessoal, são maiores do que em tempos normais. Da mesma forma, as respostas do produto aos impulsos fiscais são insignificantes ou muito baixos em situações de expansão econômica.

O presente trabalho busca contribuir com a literatura nacional em dois sentidos: o primeiro é estimar o multiplicador dos gastos do governo através de um modelo calibrado de forma que as recessões definidas pelo modelo coincidam com as recessões definidas pelo Comitê de Datação de Ciclos (CODACE),

[3] Os cortes no impostos se deram, principalmente, com a redução do IPI sobre linha brando e carros. Além disso, com a finalidade de incentivar a demanda agregada ainda mais, algumas outras medidas foram implantadas, tais como a manutenção do preço da energia elétrica e gasolina.

[4] Outros trabalhos que tratam de estimar o multiplicador fiscal para o Brasil utilizando abordagens diferentes são Moura (2015) e Cavalcanti e Vereda (2015).

já que as recessões identificadas pelos modelos markovianos propostos por Pires (2014) e Samuel et alii (2016) podem não condizer perfeitamente com as datações históricas das recessões brasileiras; o segundo é estimar um STVAR seguindo Caggiano et alii (2015), que incorpora as críticas de Ramey e Zubairy (2014) ao método de estimação de Auerbach e Gorodnichenko (2012). Além disso, o multiplicador dos gastos do governo é desagregado em consumo do governo, investimento do governo e massa salarial paga pelo governo aos funcionários públicos. Por fim, nós verificamos a robustez dos resultados a choques nos gastos do governo dados em períodos de expansão forte e recessão profunda, e controlamos as estimações do multiplicador pelo investimento público, razão dívida/PIB, grau de abertura econômica e taxa de câmbio.

Em suma, os resultados deste trabalho mostram que o multiplicador dos gastos do governo não é estatisticamente diferente nos períodos de recessão e expansão, assim como os multiplicadores de consumo, do investimento e dos gastos do governo com salários aos servidores públicos. As GIRFs indicam que em períodos de recessão o choque nos gastos do governo leva ao aumento do produto e dos gastos do governo. As receitas do governo aumentam com o choque, mas se estabilizam em um patamar mais baixo que o verificado antes. O choque nos gastos do governo em períodos de expansão não tem efeitos sobre os gastos do governo, receitas do governo e produto.

Além desta introdução, este artigo conta com mais três seções. Na segunda seção, o modelo econométrico e a base de dados são apresentados. A terceira seção traz os resultados do trabalho, bem como os testes de robustez. Por fim, a última seção apresenta as conclusões da pesquisa.

2. ESTRATÉGIA EMPÍRICA

Nesta seção, nós apresentamos o modelo empírico utilizado para estimar o multiplicador dos gastos do governo nas diferentes fases do ciclo de econômico, bem como os dados utilizados nas estimações.

2.1. O modelo econométrico

O estado-dependência do multiplicador dos gastos governamentais é estimado através de um vetor autorregressivo de transição não-linear (STVAR), dado por:

$$x_t = F(z_{t-1})\Pi_R(L)x_{t-1} + [1 - F(z_{t-1})]\Pi_E(L)x_{t-1} + \varepsilon_t \quad (1)$$

$$\varepsilon_t \sim N(0,\Omega_t) \quad (2)$$

$$\Omega_t = F(z_{t-1})\Omega_R + [1 - F(z_{t-1})]\Omega_E \quad (3)$$

$$F(z_t) = exp(-\gamma z_t)/[1 + exp(-\gamma z_t)] \quad (4)$$

onde $\gamma > 0$, $z_t \sim N(0,1)$, $x_t = (g_t,t_t,y_t)'$ é o vetor de variáveis endógenas a serem modeladas, g_t denota os gastos reais do governo, t_t são as receitas reais do governo, y_t é o produto agregado, $F(z_{t-1})$ é a função de transição que captura a probabilidade da economia estar em recessão, γ regula a suavidade de transição entre os estados, z_t é o indicador de transição, Π_R e Π_E são os coeficientes do VAR que capturam a dinâmica do sistema durante recessões e expansões (respectivamente), ε_t é o vetor de resíduos das formas reduzidas, com média zero e estado-dependência da matriz de variância-covariância dado por Ω_t. As estruturas de covariância dos resíduos em recessões e expansões são representadas respectivamente por Ω_R e Ω_E. Assume-se no modelo que as variáveis podem ser descritas por uma combinação de dois VAR lineares, um adequado para descrever a economia durante recessões e o outro durante expansões. A transição de um estado para outro é regulado pela variável padronizada de transição z_t. O parâmetro de suavização γ afeta a probabilidade de se estar em uma recessão, $F(z_t)$, ou seja, quanto maior γ, mais rápida é a transição de um estado para outro. Percebe-se que o modelo (1) – (4) permite a não-linearidade surgir tanto das relações contemporâneas quanto das relações dinâmicas do sistema econômico.

O modelo é estimado por Máxima Verossimilhança e Monte Carlo via Cadeias de Markov (MCMC).[5] Tanto o VAR linear quanto o não-linear serão estimados com três *lags*. A escolha de seleção foi feita com base no critério de informação de *Schwarz* aplicado ao modelo linear estimado para toda a amostra. Esse critério, de acordo com, Ventzislav e Lutz (2005), é o critério mais acurado para se definir o número de *lags* de um VAR com dados trimestrais para até 120 observações.

Seguindo Koop et alii (1996), a função de impulso-resposta generalizada é programada para captar a interação entre a evolução das variáveis no vetor x_t e a variável de transição, que é diretamente influenciada pelo crescimento do produto, ou seja, é modelado o *feedback* da evolução do produto no vetor x_t para o indicador de transição z_t e, consequentemente, a probabilidade $F(z_{t-1})$. Com isso, a probabilidade $F(z_t)$ é endogeneizada.

Para calcular as GIRFs, nós seguimos Koop et alii (1996) e definimos os valores iniciais para os *lags* do VAR, a variável indicadora de transição z_t e, por (4), $F(z_t)$. Em seguida, são simulados dois cenários, um onde todos os choques são identificados pela decomposição de Cholesky da matriz de variância e covariância dada em (3), e outro com os mesmos choques mais um $\delta > 0$ referente à primeira realização do choque. As GIRFs dos choques δ são dadas pela diferença entre os dois cenários. Para cada condição inicial, z, é salva a realização mediana de 500 realizações estocásticas diferentes. Essas etapas são repetidas até que 500 condições iniciais, associadas a recessões (expansões), sejam consideradas. Então, as GIRFs são construídas considerando estas 500 realizações medianas e normalizando os choques.[6]

2.2. Dados

A base da análise diz respeito ao vetor $x_t = (g_t,t_t,y_t)'$, onde g_t é o logaritmo natural dos gastos reais do governo para os três entes da federação (governos federal, estaduais e municipais), t_t é o logaritmo natural das receitas reais do governo (líquidas de transferências de assistência social, previdência e subsídios) e y_t é o logaritmo natural do produto interno bruto (PIB) real.

As séries utilizadas na análise são as de produto interno bruto, consumo do governo, investimento do governo, impostos arrecadados pelo governo, massa salarial paga pelo governo aos funcionários públicos, razão dívida/PIB, transações correntes em porcentagem do PIB, taxa de câmbio e $F(z_t)$. O período da análise vai do primeiro trimestre de 1999 ao quarto trimestre de 2015, 1999T1 – 2015T4. Todas as séries utilizadas apresentam as 68 observação disponíveis, com exceção da série de razão Dívida/PIB, disponível a partir de 2001T4 e das informações sobre o rendimento pagos pelo governo aos funcionários públicos, disponíveis a partir de 2002T2, utilizadas para controle e análise de robustez dos resultados. Todas as séries nominais foram deflacionadas pelo deflator do PIB trimestral com base em 1996T1[7], logaritimizadas e dessazonalizadas, quando necessário, pelo método X-12-ARIMA.

As séries de PIB e de consumo do governo foram retiradas das Contas Nacionais Trimestrais, disponibilizadas pelo Instituto Brasileiro de Geografia e Estatística (IBGE). As séries de investimento e receita do governo não estão disponíveis para todo o período de análise em base trimestral. A Secretaria de Tesouro Nacional passou a disponibilizá-las trimestralmente a partir de 2010. No entanto, várias pesquisas feitas pelo Instituto de Pesquisa Econômica Aplicada (IPEA) levaram a estimações dessas séries. Santos et alii (2012) estimaram as séries de formação bruta de capital fixo (investimento) do governo e Santos et alii (2010) as séries de carga tributária líquida (impostos líquidos de transferências de assistência sociais, de previdência e de subsídios) do governo.[8]

[5] Para detalhes da estimação do modelo, ver Apêndice A.

[6] Os procedimentos utilizados para obtenção das GIRFs estão apresentados com mais detalhes no Apêndice B.

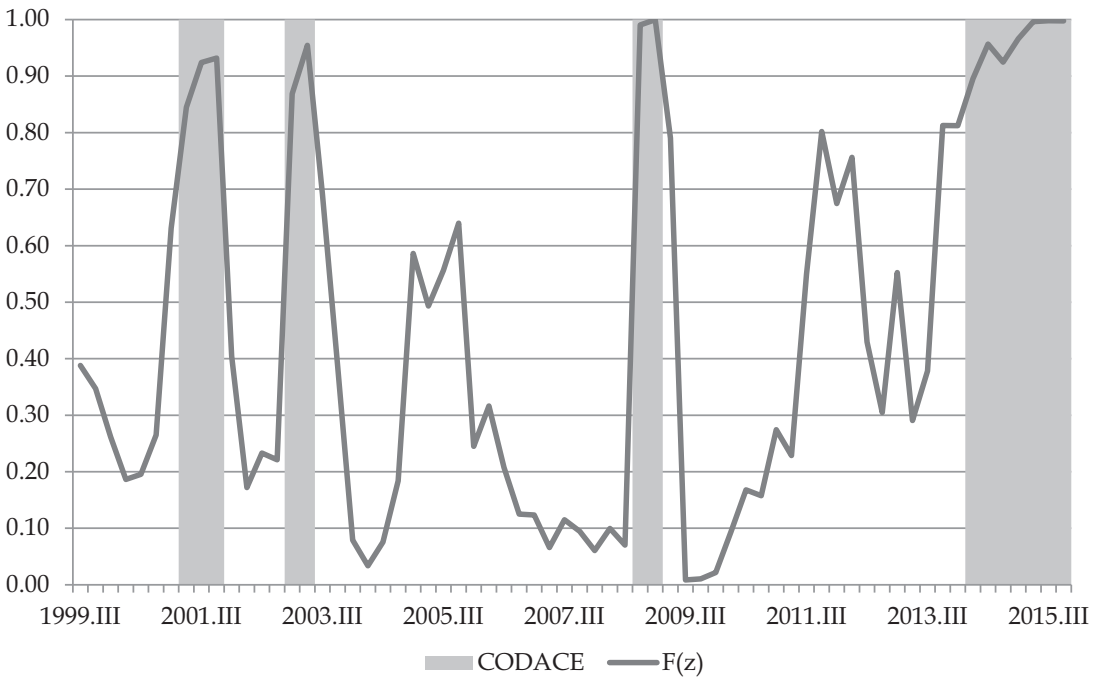
[7] O deflator do PIB trimestral foi construído com base nas informações de PIB real e nominal encadeados a preço de 1995 disponibilizadas pelo IBGE nas Contas Nacionais Trimestrais.

[8] Apesar das estimações feitas nesses artigos não abrangerem todo o período em que vai ser estimado o multiplicador fiscal, os autores seguiram atualizando esses dados e disponibilizaram as séries por e-mail.

A série de massa salarial paga pelo governos aos funcionários públicos foi construída multiplicando as informações da população empregada no setor público pelo Rendimento médio real efetivo dos empregados no setor público disponível na Pesquisa Mensal de Emprego do IBGE (PME). Lembrando que a PME é uma amostra coletada apenas nas regiões metropolitanas de Recife, Salvador, Belo Horizonte, Rio de Janeiro, São Paulo e Porto Alegre. Mesmo assim, é uma *proxy* útil para se verificar a covariância do que o governo gasta com salários e as demais variáveis utilizadas no VAR. A série de razão dívida/PIB, foi retirada do Ipeadata, assim como a série de Transações correntes dos últimos 12 meses como % PIB, e a série de Índice da taxa de câmbio real (IPCA), necessária para chegar à série de variação de taxa de câmbio.

A última série necessária para a análise é $F(z_t)$. De acordo com a Equação (4), para a construção da $F(z_t)$ é necessário definir a variável de transição z_t e o valor de calibração do parâmetro de suavização γ, de forma que, por definição, a função $F(z_t)$ assuma valores maiores do que $0{,}85$, em períodos de recessão. A frequência e os períodos das recessões históricas brasileiras são identificadas pelo CODACE.

Figura 1: Probabilidade de estar em uma fase de recessão



De acordo com o CODACE, a economia brasileira esteve em recessão em $20{,}6\%$ dos trimestres do período de 1999T1 a 2015T4. A variável de transição z_t é dada pela média móvel padronizada da taxa de crescimento percentual do PIB real, trimestre a trimestre. A quantidade de trimestres incluídas na média móvel que define z_t e o valor de γ devem satisfazer: $Pr[F(z_t) \geq 0{,}85] \approx 20{,}6\%$. Com isso, optou-se pelo $\gamma = 2{,}65$ e pela média móvel de dois períodos.[9] A Figura (1) mostra o comportamento de $F(z_t)$ ao longo do período de análise. Pode-se verificar que os maiores valores de $F(z_t)$ estão claramente associados aos períodos de recessão identificados pelo CODACE.

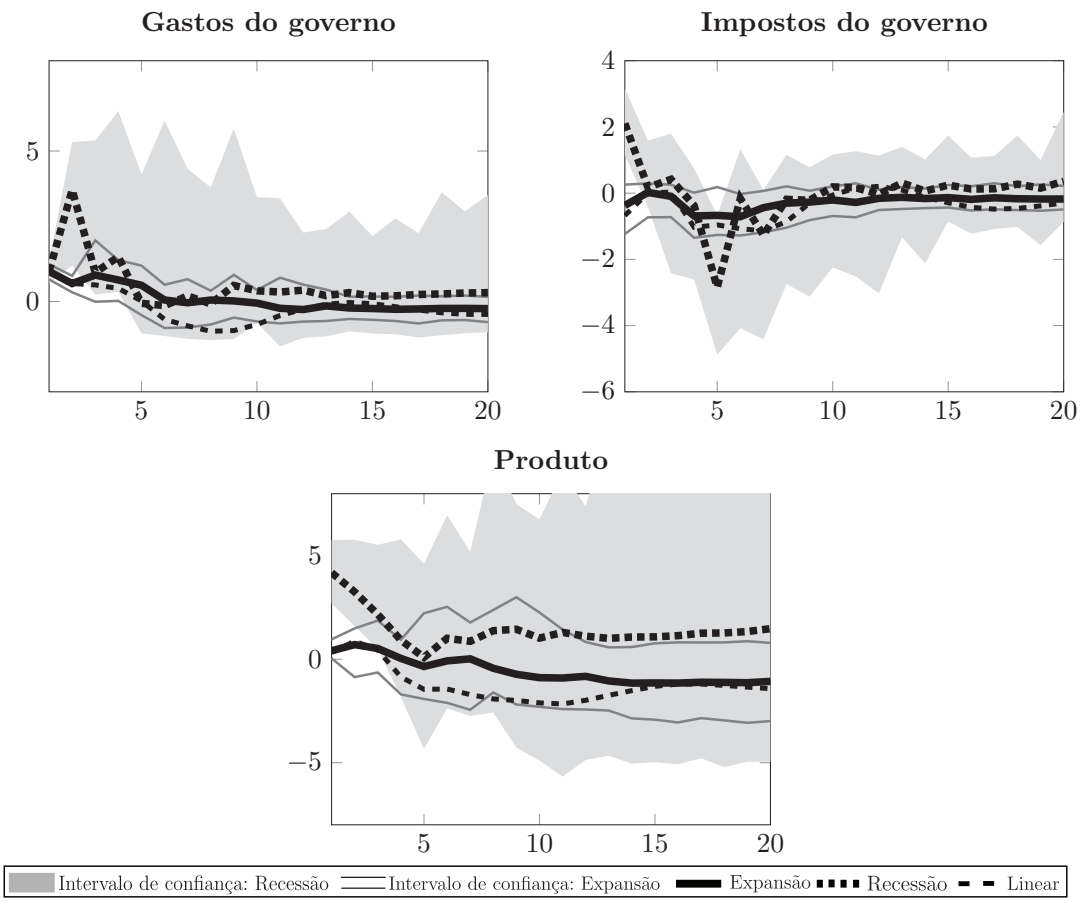
[9]Pela Equação (4), é possível ainda calcular o *threshold* $\overline{z}$ que separa as observações ocorridas em períodos expansivos das observações ocorridas em períodos recessivos para a estimação do modelo. Neste trabalho, dados $\gamma = 2{,}65$ e $F(z_t) = 0{,}85$, o valor de $\overline{z}$ é -0,65%.

3. RESULTADOS

A estimação do modelo (1)-(4) foi feita de acordo com os procedimentos apresentados na seção anterior. Em todas as estimações realizadas, os critérios de informação Akaike e Schwarz indicaram que a estimação do VAR não linear traz ganhos com relação à especificação linear.[10]

A Figura 2 traz o impacto do choque nos gastos do governo no VAR linear e não-linear, bem como os intervalos de confiança de 90% estimados para cada estado da economia. Os gastos do governo são dados pela soma do consumo e investimento do governo. Para o VAR linear, observa-se que um choque positivo nos gastos do governo leva a uma redução dos gastos do governo por três anos. Após três anos, os gastos do governo voltam à sua trajetória pré-choque. As receitas do governo caem com o choque nos gastos e só voltam à sua trajetória pré-choque após dez trimestres. O produto, por sua vez, sofre um leve aumento com o choque, mas atinge uma trajetória de queda suave até o décimo primeiro trimestre, voltando a subir lentamente, sem alcançar o valor pré-choque durante o horizonte de cinco anos.

Figura 2: Impulso Resposta Generalizada para choque nos gastos do governo: Modelo Linear, Recessões e Expansões



[10]Os resultados podem ser disponibilizados após requisitados aos autores.

A estimação do VAR não-linear mostra que, em períodos de recessão, o choque nos gastos do governo aumenta o produto e os gastos por um ano. A partir de então, o produto e os gastos sofrem uma queda até assumir sua nova trajetória, mais elevada que as verificadas no período pré-choque. As receitas do governo no estado de recessão aumentam com o choque, mas logo caem até alcançar valores menores do que os verificados anteriormente. Após dois anos, as receitas do governo voltam ao seu patamar pré-choque. Já o choque sobre os gastos do governo em períodos de expansão não tem efeito sobre os gastos, receitas do governo e produto. Os gastos do governo, voltam para sua trajetória lentamente durante os vinte trimestres analisados. As receitas do governo se recuperam em dois trimestres da queda sofrida pelo choque nos gastos e se mantém em sua trajetória com leves oscilações. As GIRFs estimadas sugerem ainda que o produto não chega a sofrer efeitos do choque nos gastos governamentais em períodos de expansão.

Em relação à assimetria nos efeitos dos choques nos gastos do governo sobre o produto, pode-se verificar que os intervalos de confiança se sobrepõem substancialmente, principalmente depois de um ano. Essa sobreposição sugere que a reação do produto ao choque nos gastos não é necessariamente diferente quando a economia está em recessão ou expansão.

O próximo passo é calcular o multiplicador dos gastos do governo e os intervalos de confiança associados. O multiplicador é dado pela integral do produto dividido pela integral da resposta do choque, ou seja, $\sum_{h-1}^{H} y_h / \sum_{h-1}^{H} g_h$, onde H é o horizonte de tempo escolhido. As variações percentuais são convertidas em reais, redimensionando a razão pela taxa média do nível de produto sobre os gastos públicos. Dessa forma, leva-se em conta a persistência dos choques e uma possível inversão de trajetória durante o tempo.

As estimativas obtidas para o multiplicador dos gastos governamentais são apresentadas na Tabela 1 considerando horizontes de até cinco anos. Pode-se observar que a magnitude do multiplicador é maior nos períodos de recessão apenas nos horizontes de quatro e oito trimestres. Mesmo as magnitudes sendo diferentes, o intervalo de confiança já dá indícios de que os efeitos da política fiscal são, em geral, simétricos. As únicas possíveis assimetrias podem ser constatadas no multiplicador estimado para os horizontes de quatro e doze trimestres. Por exemplo, para o horizonte de quatro trimestres, o multiplicador de períodos expansivos é estatisticamente igual a zero, enquanto o multiplicador recessivo não. Já no horizonte de doze trimestres, o multiplicador de períodos de recessão é estatisticamente igual a zero, enquanto o multiplicador de períodos de expansão é maior que zero. Por fim, é possível destacar que as estimativas do multiplicador fiscal são estatisticamente maiores que 1 apenas nos horizontes de 16 e 20 trimestres. Isso sugere que os maiores efeitos reais dos gastos do governo são verificados no médio prazo.

Para testar se o multiplicador dos gastos do governo é estatisticamente diferente entre os estados da economia, nós construímos um teste baseado na diferença do multiplicador estimado para os períodos de recessão e expansão. Esse teste computa a diferença do multiplicador de recessão e expansão condicionados ao mesmo conjunto de acontecimentos extraídos aleatoriamente dos elementos do modelo. A densidade empírica da diferença entre o multiplicador de recessão e expansão é baseada nas 500 realizações de cada horizonte de interesse.

As densidades empíricas da diferença dos multiplicadores para os horizontes de análise da função impulso-resposta com as bandas de 90% do intervalo de confiança está apresentada na Figura 3. O zero estando dentro do intervalo de confiança significa que não há diferença entre o multiplicador de gastos em recessão e expansão, ou seja, os efeitos da política de gastos governamentais sobre o produto são simétricos entre as diferentes fases do ciclo econômico.

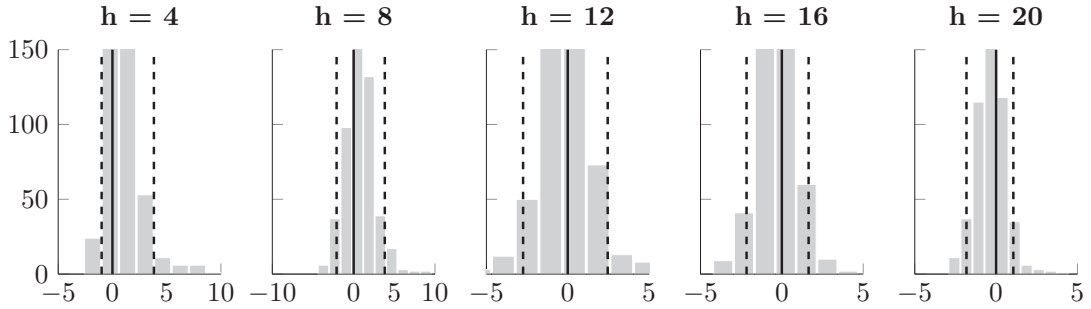
Os resultados encontrados indicam que o multiplicador dos gastos do governo não é estatisticamente diferente em períodos de recessão e expansão. Contudo, ainda há a possibilidade de algum dos componentes dos gastos do governo apresentar um multiplicador dependente do estado da economia. Para tanto, são estimadas três especificações do modelo (1) a (4) com a variável g_t sendo representada pelo consumo do governo, investimento do governo e gastos do governo com salários de servidores

Tabela 1: Multiplicadores dos gastos do governo

Horizonte/Estado	Expansão	Recessão
4	0,45	1,11
	[-2,17; 1,39]	[-0,07; 2,38]
8	0,94	1,75
	[-1,04; 2,73]	[-0,54; 3,66]
12	1,84	1,80
	[0,06; 3,29]	[-0,40; 3,40]
16	2,50	2,21
	[1,55; 3,59]	[1,02; 3,89]
20	2,94	2,59
	[2,13; 4,23]	[1,77; 4,21]

Notas: Tabela elaboração com base no VAR estimado. Variações percentuais convertidas em reais. Multiplicadores acumulativos.

Figura 3: Diferença entre os multiplicadores de gastos do governo em períodos de recessão e expansão

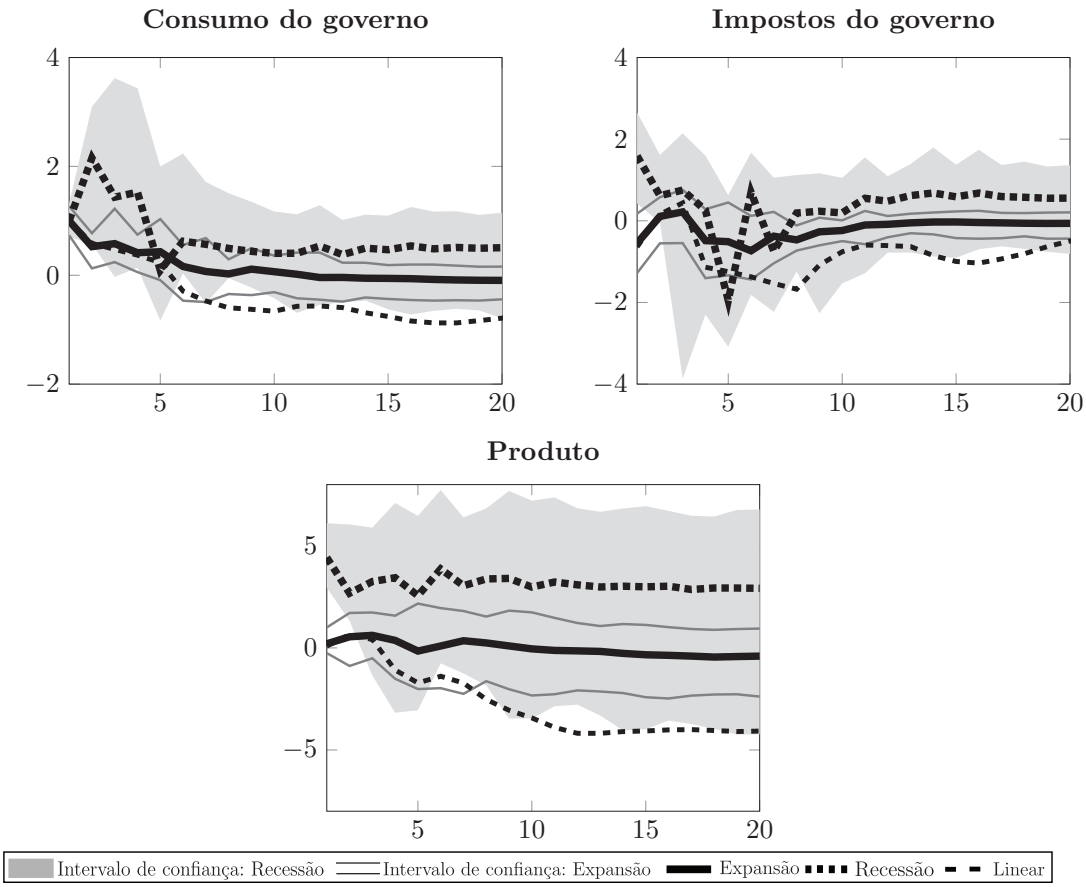


públicos. A finalidade é verificar se o governo teria incentivos a concentrar gastos em algumas dessas áreas como medida anticíclica.

Na Figura 4, é possível observar os efeitos do choque no consumo do governo para o VAR linear e não-linear. As GIRFs mostram que um choque no consumo do governo dado em períodos recessivos leva o consumo do governo e o produto para patamares mais elevados do que antes do choque. Além disso, o choque no consumo do governo em períodos recessivos leva a um aumento na arrecadação do governo no primeiro ano. Após uma queda no quinto trimestre, o choque no consumo do governo estabiliza as receitas do governo em uma trajetória mais elevada do que se encontrava antes do choque. Quando o choque se dá em períodos expansivos, o consumo do governo, as receitas dos governo e o produto voltam às suas trajetórias pré choque. No curto prazo, o choque no consumo do governo em períodos expansivos faz com que as receitas governamentais caiam.

A Figura 5 traz os impactos do choque nos investimentos do governo no VAR linear e não-linear. As GIRFs mostram que o choque nos investimentos dado em períodos expansivos faz com que essa variável volte à sua trajetória pré choque depois de dez trimestres. Já o choque dado em períodos recessivos leva o investimento do governo para um patamar mais elevado do que o verificado antes do choque. Os efeitos que o choque nos investimentos apresentam no VAR linear e não linear são semelhantes: o choque nos investimentos leva ao aumento das receitas do governo, seguida de uma queda. Em dois

Figura 4: Impulso Resposta Generalizada para choque no consumo do governo: Modelo Linear, Recessões e Expansões

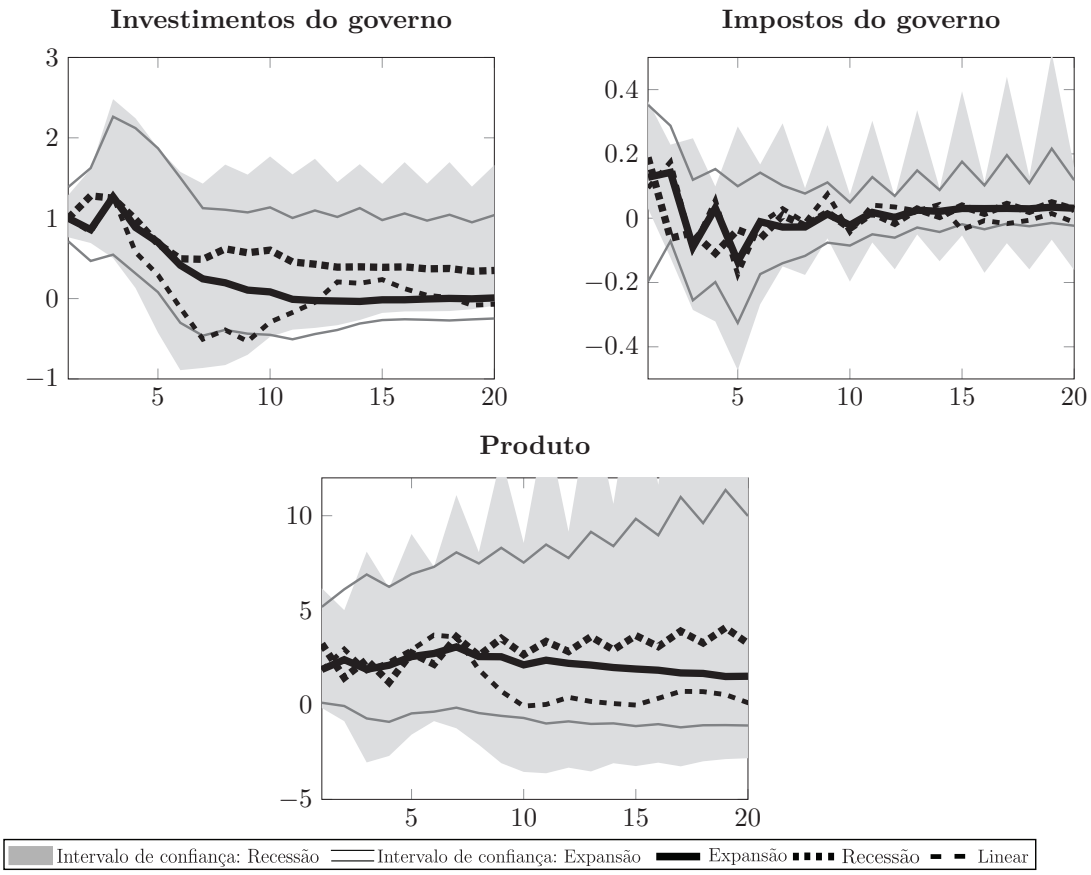


anos as receitas do governo voltam a sua trajetória pré choque. O produto aumenta com o choque nos investimentos e se mantém em uma trajetória mais elevada do que a verificada antes do choque.

Os efeitos do choque nos gastos do governo com salários pagos a servidores públicos no VAR linear e não-linear são apresentados na Figura 6. Como se pode observar, em períodos recessivos, o montante salarial pago pelo governo se estabiliza em patamares mais elevados do que os verificados antes do choque. Já quando o choque acontece em períodos expansivos, a massa salarial volta aos patamares pré choque. Nota-se também que, no curtíssimo prazo, o choque nos salários pagos pelo governo aumenta a receita do governo quando o choque é dado em períodos expansivos e diminui a arrecadação do governo quando o choque é dado em períodos recessivos. Depois de um ano as receitas do governo voltam ao seu patamar pré choque. O produto aumenta quando o choque nos salários é dado em períodos recessivos, mas diminui quando o choque é dado em períodos expansivos.

A Tabela 2 traz os multiplicadores acumulativos do consumo, investimento e massa salarial paga pelo governo estimados para horizontes de um a cinco anos. Inicialmente, é possível notar que o multiplicador do consumo parece ser maior nos períodos recessivos em comparação com os períodos expansivos. Em contraposição, observa-se que o multiplicador do investimento é, em geral, maior em períodos de recessão. Já o multiplicador de salário pago pelo governo apresenta valores maiores em períodos de recessão apenas para o horizonte de um e dois anos. Apesar dessas constatações, é

Figura 5: Impulso Resposta Generalizada para choque nos investimentos do governo: Modelo Linear, Recessões e Expansões



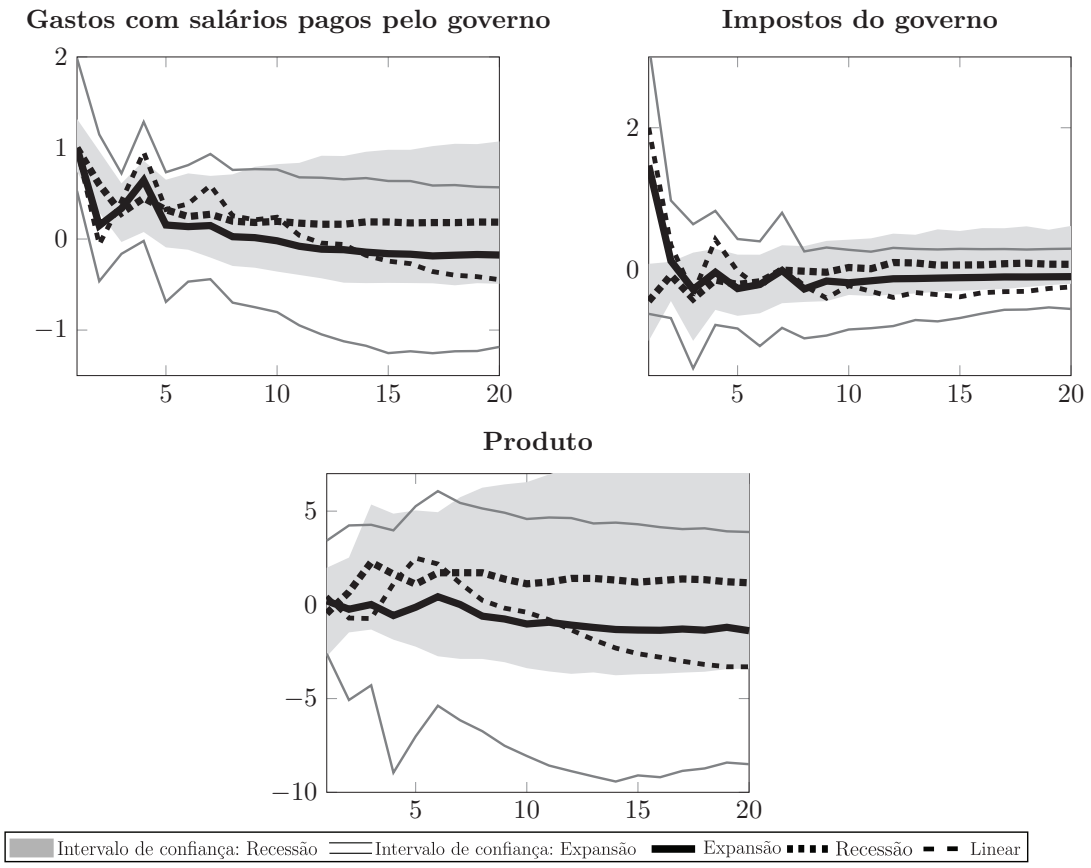
válido ressaltar que, em geral, não há indícios de que os multiplicadores são diferentes nos períodos de recessão e expansão: todos os multiplicadores de períodos de expansão estão contidos no intervalo de confiança dos multiplicadores de períodos recessão e vice-versa. Adicionalmente, nota-se que, no horizonte de um ano, não se pode afirmar que os multiplicadores são estatisticamente diferentes de zero.

As densidades empíricas das diferenças dos multiplicadores do consumo, investimento e massa salarial entre períodos de recessão e expansão são apresentadas nas Figuras 7, 8 e 9. Como se pode observar, não há evidências de que os multiplicadores analisados são dependentes do estado da economia, ou seja, são diferentes em períodos de recessão e expansão para todos os horizontes de análise. Esse resultado também foi verificado quando considerou-se os gastos agregados do governo.

3.1. Testes de robustez dos resultados

Como vimos anteriormente, as evidências empíricas mostram que o multiplicador dos gastos do governo não é dependente do estado da economia brasileira. Nessa seção são feitas algumas estimações alternativas para verificar a robustez desse resultado. Em específico, verifica-se se a simetria no multiplicador se mantém quando: $i)$ são consideradas apenas recessões profundas e expansões fortes; $ii)$ o multiplicador de consumo do governo é controlado para os investimentos; $iii)$ o multiplicador é con-

Figura 6: Impulso Resposta Generalizada para choque na massa salarial paga pelo governo: Modelo Linear, Recessões e Expansões



trolado pela trajetória da dívida pública e; iv) o multiplicador é controlado por variáveis de abertura econômica e taxa de câmbio.

3.1.1. Eventos extremos

Caggiano et alii (2015) não encontraram evidências de que o multiplicador fiscal depende do estado da economia norte-americana. Diante disso, esses autores levantaram a hipótese de que recessões e expansões moderadas são intrínsecas à economia e, portanto, o multiplicador não tem um efeito anticíclico tão atenuante. Por isso, reestimaram um modelo STVAR dividindo a amostra não só em recessões e expansões, mas também em eventos extremos (expansões fortes e recessões profundas).

Neste trabalho, nós seguimos Caggiano et alii (2015) e estimamos o modelo não-linear (1)-(4) considerando eventos extremos.[11] A identificação desses eventos foi baseada nos valores da variável de transição (z). Como essa variável é normalizada, cada realização de z acima de dois foi associada à expansão forte e cada realização abaixo de menos dois foi associada à recessão profunda.

Os resultados das estimações são sumarizadas por meio das densidades empíricas da diferença do multiplicador dos gastos do governo entre períodos de fortes expansões e profundas recessões, mostra-

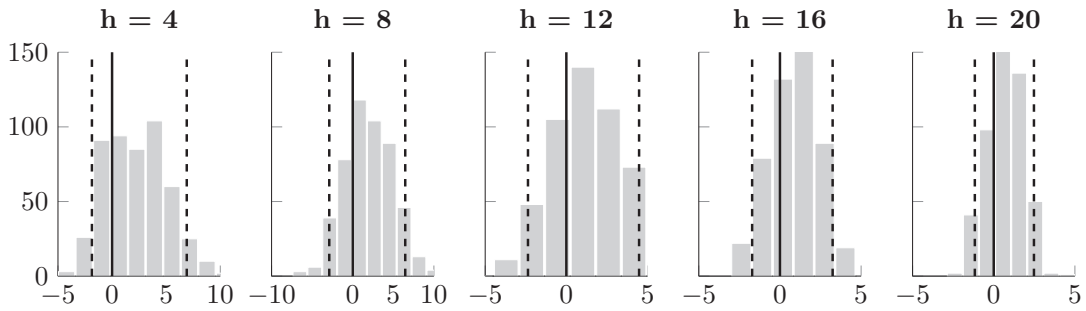
[11] As GIRFs do VAR linear e não-linear associadas aos eventos médios e extremos para a economia brasileira são apresentadas na Figura C-1, em anexo.

Tabela 2: Multiplicadores do consumo, investimento e massa salarial do governo

	Consumo		Investimento		Salário	
h	Expansão	Recessão	Expansão	Recessão	Expansão	Recessão
4	0,80	2,43	2,54	2,08	1,84	3,65
	[-3,24; 2,19]	[-0,58; 5,91]	[-0,50; 4,55]	[-1,80; 4,10]	[-13,85; 22,01]	[-7,13; 8,86]
8	1,68	3,08	4,22	3,68	4,66	5,35
	[-1,39; 3,90]	[-0,51; 7,35]	[0,07; 7,74]	[1,09; 5,63]	[-10,94; 16,78]	[-1,93; 13,03]
12	2,92	4,02	5,16	4,84	7,59	6,68
	[1,68; 5,07]	[1,31; 7,03]	[1,17; 7,97]	[2,33; 7,65]	[0,81; 17,67]	[3,74; 13,78]
16	3,75	4,77	5,84	5,76	8,77	7,67
	[2,47; 5,57]	[2,60; 6,66]	[2,55; 8,68]	[3,30; 9,15]	[4,16; 13,46]	[5,46; 11,78]
20	4,31	5,13	6,35	6,54	8,65	7,90
	[3,04; 5,69]	[3,43; 6,45]	[3,20; 9,70]	[3,89; 10,31]	[5,84; 11,19]	[6,42; 10,18]

Notas: Tabela elaboração com base no VAR estimado. Variações percentuais convertidas em reais. Multiplicadores acumulativos.

Figura 7: Diferença entre os multiplicadores de consumo do governo em períodos de recessão e expansão



das na Figura 10. Mais uma vez, as evidências indicam que o multiplicador não é dependente do estado da economia, mesmo em eventos extremos.

3.1.2. Consumo e investimento

Um outro teste de robustez realizado foi estimar o modelo (1)-(4) para verificar o efeito de um choque no consumo do governo, mas controlando para investimento. Nessa caso, $x_t = (c_t, i_t, t_t, y_t)'$, onde c_t é o logaritmo natural do consumo do governo e i_t é o logaritmo natural do investimento do governo.[12]

A Figura 11 mostra que, conforme se verificou para o multiplicador de consumo sem controle, o multiplicador de consumo controlado para investimentos não apresenta assimetria relacionada à fase do ciclo de negócios.

[12]As GIRFs estão apresentadas na Figura C-2, em anexo.

Figura 8: Diferença entre os multiplicadores de investimento do governo em períodos de recessão e expansão

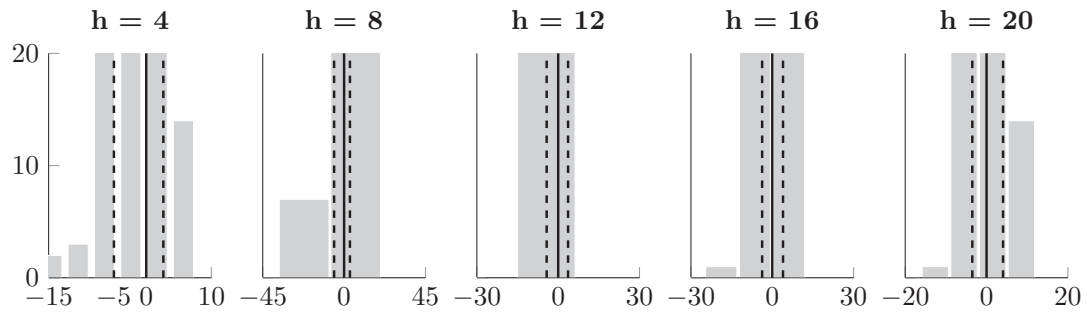


Figura 9: Diferença entre os multiplicadores de massa salarial paga pelo governo em períodos de recessão e expansão

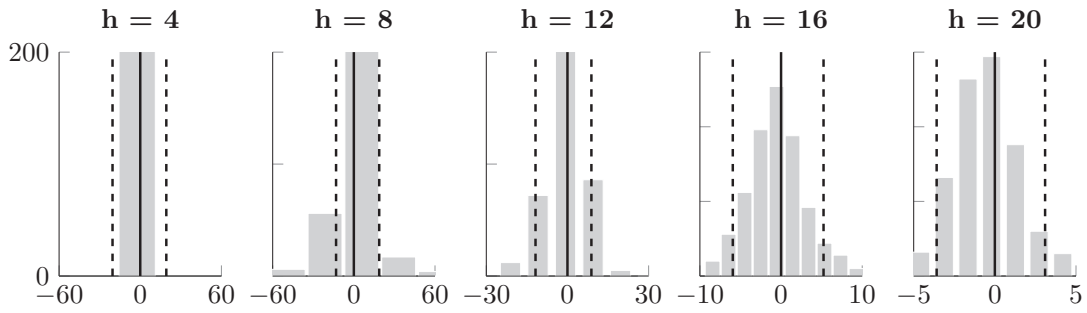
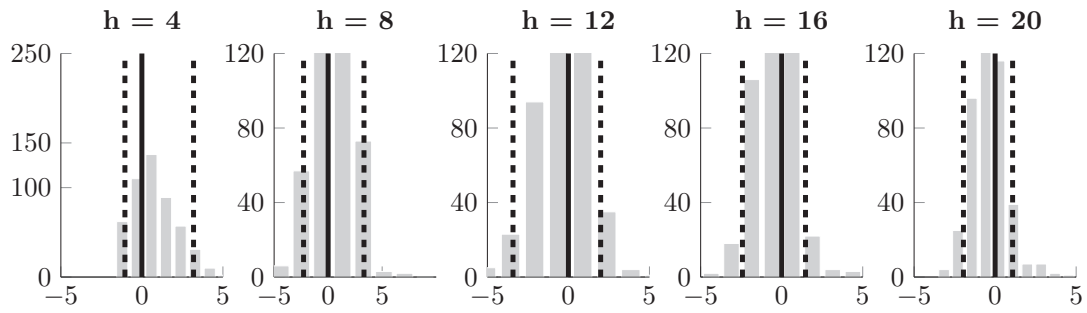


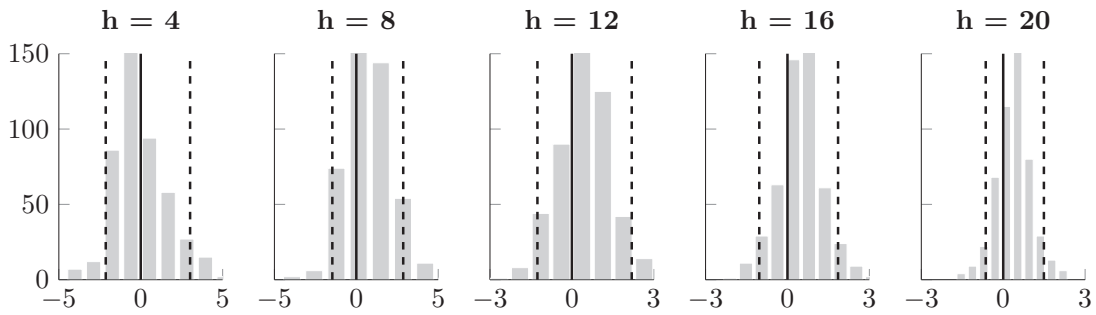
Figura 10: Diferença entre os multiplicadores de gastos do governo em períodos de recessão e expansão: eventos extremos



3.1.3. Razão dívida/PIB

Outra variável de controle importante é a razão dívida/PIB. Uma limitação da maioria dos trabalhos que estimam multiplicadores fiscais é não considerar adequadamente o papel da dívida pública na determinação da política fiscal. De acordo com Favero e Giavazzi (2007), se um VAR omite a dívida como variável explicativa e não leva em consideração o impacto de choques nos gastos e receitas públicas sobre a evolução da dívida, as funções de impulso-resposta estimadas são viesadas. A ideia é que um choque fiscal expansionista tende a aumentar a dívida pública. Esse aumento da dívida tem que gerar

Figura 11: Diferença entre os multiplicadores de consumo do governo controlado para investimento em períodos de recessão e expansão

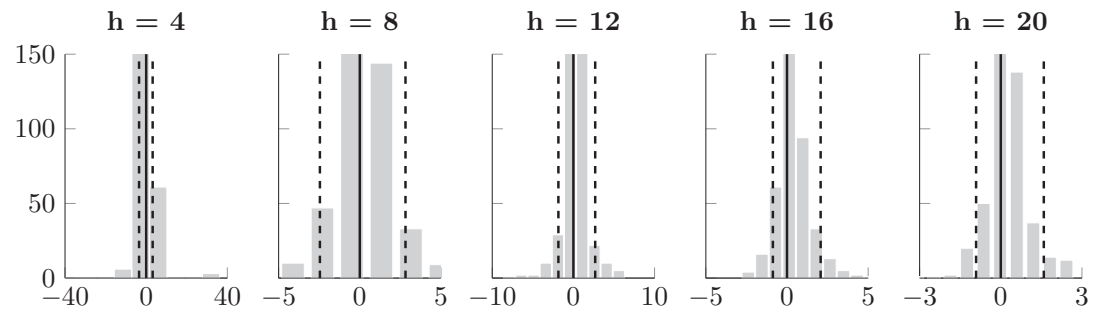


um movimento fiscal compensatório no longo prazo, visto que a razão dívida/PIB não pode crescer indefinidamente.

Em conformidade com isso, Ilzetzki et alii (2013) fazem um estudo com dados em painel dividindo os países entre os que apresentam razão dívida/PIB elevada e os países com razão baixa. Os resultados encontrados por esses autores mostram que países com elevados níveis de dívida pública apresentam multiplicadores menores do que países com níveis baixos de dívida. Cavalcanti e Silva (2010) mostram, para o Brasil, que a exclusão da dívida pública superestima os efeitos de choques fiscais.

Para analisar se a dívida está influenciando a não-linearidade do VAR, nós estimamos o modelo (1)-(4) incluindo a razão dívida/PIB como variável endógena. Nesse caso, $x_t = (g_t, t_t, y_t, d_t)'$, onde d_t é o logaritmo natural da razão dívida/PIB do Brasil. Os resultados desse modelo são apresentados na Figura 12, que mostra as densidades empíricas das diferenças do multiplicador entre períodos de recessão e expansão. Como se pode verificar, mesmo controlando o modelo para a dívida pública, o multiplicador dos gastos do governo não é estado-dependente.

Figura 12: Diferença entre os multiplicadores de gastos do governo controlado para dívida pública em períodos de recessão e expansão



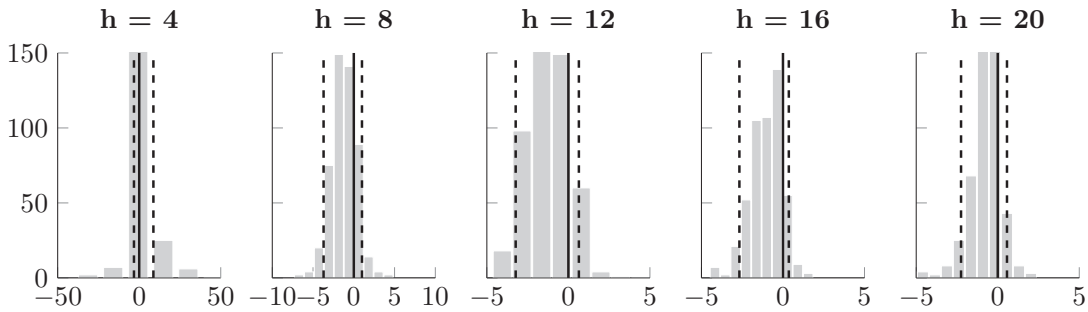
3.1.4. Abertura econômica e taxa de câmbio

O último controle feito para verificar a robustez dos resultados foi adicionar o grau de abertura econômica e taxa de câmbio ao modelo. Ilzetzki et alii (2013) verificam que a flexibilidade da taxa de câmbio afeta os multiplicadores, assim como o grau de abertura ao comércio. Para controlar o VAR para essas

variáveis, nós acrescentamos as variáveis transações correntes e a variação do logaritmo natural da taxa de câmbio real ao vetor x do modelo base.[13]

A Figura 13 apresenta as densidades empíricas das diferenças do multiplicador entre as fases do ciclo econômico. Mais uma vez, os resultados sugerem que os efeitos reais do choque nos gastos governamentais não são diferentes em recessão e expansão.

Figura 13: Diferença entre os multiplicadores de gastos do governo controlado para o grau de abertura econômica e taxa de câmbio em períodos de recessão e expansão



3.2. Discussão dos resultados

A ideia de que a política de gastos governamentais expansionistas tem efeitos assimétricos na economia segue a teoria keynesiana, a qual defende que um aumento nos gastos do governo só tem efeito positivo na produção e no consumo quando a economia está com capacidade ociosa: capacidade industrial subutilizada e desemprego elevado. Nesse contexto, um aumento nos gastos deveria impulsionar a demanda agregada, aumentar a produção e o emprego na economia. Por outro lado, um aumento nos gastos do governo quando a economia está em plena capacidade tende a gerar apenas inflação, se a política monetária acomodar a taxa de juros.

Entretanto, as evidências empíricas verificadas nesse trabalho sugerem que o multiplicador dos gastos do governo não depende do estado da economia. Esse resultado também foi encontrado por Pereira e Lopes (2010) e Ramey e Zubairy (2014) para os Estados Unidos, por Kirchner et alii (2010) para a zona do Euro e, por Pires (2014) e Samuel et alii (2016) para a economia brasileira.

Como a teoria do multiplicador fiscal está intimamente ligada à propensão marginal a consumir, dizer que o multiplicador é o mesmo durante as recessões e expansões é dizer que a propensão marginal não se alteram durante os ciclos econômicos curtos da economia. Talvez apenas se alterem nos ciclos econômicos longos, que reflitam mudanças de hábito dos agentes. O que é diferente do que aconteceu com os períodos de taxa de juros zero dos EUA, onde os agentes foram obrigados a reformular suas decisões de poupança e investimento, uma vez que o consumo presente passou a ser mais atrativo que o futuro.

Orair et alii (2016), por sua vez, encontram dependência de estado no multiplicador dos gastos do governo brasileiro. Entretanto, encontraram esses resultados seguindo o método de estimação de Auerbach e Gorodnichenko (2012), que assumiu alguns pressupostos irrealistas para a construção das funções de impulso-respostas, como a taxa de crescimento do produto ser dada exogenamente, a economia permanecer em recessão ou expansão por um período pré determinado e, principalmente, não permitir que choques nos gastos do governo levem a economia para outro estado.

[13] As GIRFs estão apresentadas na Figura C-4, em anexo.

Além disso, os autores estimam dois tipos de multiplicadores para períodos de recessão e expansão: o multiplicador de pico e o acumulativo de 1 a 48 períodos. O problema é que em um horizonte de análise de cinco anos, por exemplo, ter um multiplicador de períodos recessivos que assume um máximo estatisticamente diferente do multiplicador de períodos expansivos no terceiro ano, não diz nada para efeitos de política anticíclica. O mesmo acontece com o multiplicador acumulativo de 48 períodos. Como afirmam Auerbach e Gorodnichenko (2012) os modelos não lineares têm a propriedade de tornar os multiplicadores dependentes do tempo e da trajetória. Logo, faz mais sentido comparar o multiplicador durante vários horizontes de tempo, como feito nesse trabalho.

Apesar de tudo, Orair et alii (2016) só encontram dependência de estado no multiplicador para eventos de recessão e expansão extremos. O que, dadas as críticas, é mais um indício de que o multiplicador dos gastos do governo não é dependente do ciclo econômico no Brasil.

As possíveis explicações para a simetria do multiplicador dos gastos do governo no Brasil, especificamente, pode estar ligado a algumas questões: $i)$ a curva de oferta agregada da economia brasileira ser linear, e não convexa, nesse caso, em períodos de produto baixo a curva OA teria a mesma inclinação do que teria em períodos de produto alto, gerando o mesmo multiplicador; $ii)$ a taxa de juros brasileira responder mais a aumentos nos gastos do governo do que se pensa, não atuando da mesma forma que atuaria caso a economia se encontrasse no limite inferior da taxa de juros; e $iii)$ o mercado financeiro brasileiro ser pouco desenvolvido e uma fração baixa da população ter acesso a empréstimos e depósitos, não gerando o acelerador financeiro tão mais forte em recessões, como proposto por Canzoneri et alii (2016).

4. CONSIDERAÇÕES FINAIS

Neste trabalho nós estimamos o multiplicador dos gastos do governo brasileiro dependente do estado da economia, através de um vetor autorregressivo de transição não-linear (STVAR), modelado para discriminar respostas dinâmicas a choques nos gastos em tempos bons e ruins, ou seja, em períodos de recessões e expansões.

As funções de impulso-respostas generalizadas, calculadas de forma a captar a endogeneidade da transição de um estado da economia para outro, mostram que o choque nos gastos do governo leva ao aumento desses gastos e do produto em períodos de recessão. As receitas do governo aumentam com o choque, mas se estabilizam em um patamar mais baixo que o verificado antes do choque. Por sua vez, o choque nos gastos do governo em períodos de expansão não tem efeitos expressivos sobre os gastos do governo, receitas do governo e produto agregado.

Além do multiplicador de gastos do governo, foram estimados os multiplicadores de consumo, investimentos e massa salarial do governo. Os testes baseados nas diferenças dos multiplicadores estimados para os períodos de recessão e expansão indicaram que os efeitos reais da política fiscal não dependem da fase do ciclo econômico. Adicionalmente, nós verificamos que esses resultados são robustos a choques nos gastos do governo dados em períodos de expansão forte e recessão profunda, e às estimativas do multiplicador controladas pelo investimento, razão dívida/PIB, grau de abertura econômica e taxa de câmbio.

Uma importante limitação deste trabalho é a não inserção das expectativas dos gastos governamentais no VAR em razão da inexistência de uma base de dados com essas informações. Muitos autores constatam que as expectativas de gastos formuladas pelo agentes impacta de maneira significante a dinâmica dos instrumentos fiscais [ver, por exemplo, Ramey (2011)]. Posto isso, uma sugestão para pesquisa futura seria estimar as expectativas dos agentes econômicos em relação aos gastos governamentais e estender a análise para verificar se o multiplicador é dependente do estado da economia, dados choques antecipados e não antecipados nesses gastos.

BIBLIOGRAFIA

Auerbach, A. J. & Gorodnichenko, Y. (2012). Measuring the Output Responses to Fiscal Policy. *American Economic Journal: Economic Policy*, 4(2):1–27.

Barro, R. J. (1974). Are Government Bonds Net Wealth? *Journal of Political Economy*, 82(6):1095–1117.

Blanchard, O. & Perotti, R. (2002). An Empirical Characterization of the Dynamic Effects of Changes in Government Spending and Taxes on Output. *The Quarterly Journal of Economics*, 117(4):1329–1368.

Caggiano, G., Castelnuovo, E., Colombo, V., & Nodari, G. (2015). Estimating Fiscal Multipliers: News From A Non-linear World. *Economic Journal*, 0(584):746–776.

Canzoneri, M., Collard, F., Dellas, H., & Diba, B. (2016). Fiscal Multipliers in Recessions. *Economic Journal*, 126(590):75–108.

Cavalcanti, A. F. H. C. & Vereda, L. (2015). Fiscal Policy Multipliers in a DSGE Model for Brazil. *Brazilian Review of Econometrics*, SAMBA(1):197–233.

Cavalcanti, M. A. & Silva, N. L. (2010). Dívida pública, política fiscal e nível de atividade: Uma abordagem VAR para o Brasil no período 1995-2008. *Economia Aplicada*, 14(4):391–418.

Chahrour, R., Schmitt-Grohé, S., & Uribe, M. (2012). A Model-Based Evaluation of the Debate on the Size of the Tax Multiplier. *American Economic Journal: Economic Policy*, 4(2):28–45.

Christiano, L., Eichenbaum, M., & Rebelo, S. (2011). When Is the Government Spending Multiplier Large? *Journal of Political Economy*, 119(1):78–121.

Eggertsson, G. B. (2009). What fiscal policy is effective at zero interest rates? Staff Reports 402, Federal Reserve Bank of New York.

Favero, C. & Giavazzi, F. (2007). Debt and the effects of fiscal policy. Working Papers 07-4, Federal Reserve Bank of Boston.

Fazzari, S., Morley, J., & Panovska, I. (2013). State-Dependent Effects of Fiscal Policy. Discussion Papers 2012-27A, School of Economics, The University of New South Wales.

Gobetti, S. W. & Orair, R. O. (2015). Política Fiscal em Perspectiva: o ciclo de 16 anos (1999-2014). *Revista de Economia Contemporânea*, 19:417–447.

Ilzetzki, E., Mendoza, E. G., & Végh, C. A. (2013). How big (small?) are fiscal multipliers? *Journal of Monetary Economics*, 60(2):239–254.

Jordà, Ò. (2005). Estimation and Inference of Impulse Responses by Local Projections. *American Economic Review*, 95(1):161–182.

Kirchner, M., Cimadomo, J., & Hauptmeier, S. (2010). Transmission of government spending shocks in the Euro area: Time variation and driving forces. Working Paper Series 1219, European Central Bank.

Koop, G., Pesaran, M. H., & Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics*, 74(1):119–147.

Leeper, E. M., Traum, N., & Walker, T. B. (2011). Clearing Up the Fiscal Multiplier Morass. NBER Working Papers 17444, National Bureau of Economic Research, Inc.

Moura, G. V. (2015). Multiplicadores Fiscais e Investimento em Infraestrutura. *Revista Brasileira de Economia*, 69(1):53–73.

Orair, R. O., Siqueira, F. d. F., & Gobetti, S. W. (2016). Política Fiscal e Ciclo Econômico: uma análise baseada em multiplicadores do gasto público.

Pastore, A. C., Gazzano, M., & Pinotti, M. C. (2004). *A crise internacional e a mudança na política econômica*, chapter 8, pages 241–284.

Pereira, M. C. & Lopes, A. S. (2010). Time varying fiscal policy in the U.S. CEMAPRE Working Papers 1004, Centre for Applied Mathematics and Economics (CEMAPRE), School of Economics and Management (ISEG), Technical University of Lisbon.

Peres, M. A. F. & Ellery Junior, R. d. G. (2009). Efeitos dinâmicos dos choques fiscais do governo central no PIB.

Pires, M. C. d. C. (2012). Controvérsias recentes sobre multiplicadores fiscais. *Banco Interamericano de Desenvolvimento (BID). Multiplicadores Fiscais no Brasil, Brasília, BID*, pages 9–27.

Pires, M. C. d. C. (2014). Política fiscal e ciclos econômicos no Brasil. *Economia Aplicada*, 18:69–90.

Ramey, V. A. (2011). Identifying Government Spending Shocks: It's all in the Timing. *The Quarterly Journal of Economics*, 126(1):1–50.

Ramey, V. A. & Zubairy, S. (2014). Government Spending Multipliers in Good Times and in Bad: Evidence from U.S. Historical Data. NBER Working Papers 20719, National Bureau of Economic Research, Inc.

Romer, C. & Bernstein, J. (2009). The job impact of the American Recovery and Reinvestment Plan. Staff reports, American Council of Economic Advisers.

Samuel, M. A. C. B., Lima, E. C. R., & de Paula, L. F. (2016). Mudanças de Regime e Multiplicadores Fiscais no Brasil entre 1999-2012: uma avaliação empírica. ANPEC - Associação Nacional dos Centros de Pósgraduação em Economia [Brazilian Association of Graduate Programs in Economics], Anais do XLIII Encontro Nacional de Economia [Proceedings of the 43rd Brazilian Economics Meeting].

Santos, C., Orair, R., Gobetti, S., Ferreira, A., Rocha, W., Silva, H., & Britto, J. (2012). Estimativas mensais da formação bruta de capital fixo pública no Brasil (2002-2010). *Economia Aplicada*, 16(3):445–473.

Santos, C. H. M. d., Silva, A. C. M., & Ribeiro, M. B. (2010). Uma metodologia de estimação da carga tributária líquida brasileira trimestral no período 1995-2009. *Revista de Economia Contemporânea*, 14:209–235.

Taylor, J. B. (2011). An Empirical Analysis of the Revival of Fiscal Activism in the 2000s. *Journal of Economic Literature*, 49(3):686–702.

Ventzislav, I. & Lutz, K. (2005). A Practitioner's Guide to Lag Order Selection For VAR Impulse Response Analysis. *Studies in Nonlinear Dynamics & Econometrics*, 9(1):1–36.

A. ESTIMAÇÃO DO VAR NÃO LINEAR

Seguindo Auerbach e Gorodnichenko (2012) e Caggiano et alii (2015), a função de Máxima Verossimilhança do modelo (1)-(4) pode ser escrita como:

$$\ln L = \alpha + \frac{1}{2}\sum_{t=1}^{T} \ln |\Omega_t| - \frac{1}{2}\sum_{t=1}^{T} \varepsilon_t^{'} \Omega_t^{-1} \varepsilon_t \qquad \text{(A-1)}$$

já que, pela Equação 1, $\varepsilon_t = x_t - F(z_{t-1})\Pi_R(L)x_{t-1} + [1 - F(z_{t-1})]\Pi_E(L)x_{t-1}$. O objetivo é estimar os parâmetros $\Psi = \{\gamma, \Omega_R, \Omega_E, \Pi_R(L), \Pi_E(L)\}$, onde $\Pi_{R,E}(L) = [\Pi_{R,E,1} ... \Pi_{R,E,gl}]$. Como o

modelo é linear em $\{\Pi_R(L),\Pi_E(L)\}$, para um dado $\{\gamma,\Omega_R,\Omega_E\}$, os coeficientes $\{\Pi_R(L),\Pi_E(L)\}$ podem ser estimados minimizando $\frac{1}{2}\sum_{t=1}^{T}\varepsilon_t^{'}\Omega_t^{-1}\varepsilon_t$. Chamando $w_t = F(z_{t-1})x_{t-1} + [1 - F(z_{t-1})]x_{t-1}...F(z_{t-1})x_{t-gl}+[1-F(z_{t-1})]x_{t-gl}$, pode-se reescrever $\varepsilon_t = x_t - \Pi w_t^{'}$, com $\Pi_{R,E}(L) = [\Pi_{R,E,1}...\Pi_{R,E,gl}]$. Assim, a função objetivo se torna:

$$\frac{1}{2}\sum_{t=1}^{T}(x_t - \Pi w_t^{'})\Omega_t^{-1}(x_t - \Pi w_t^{'}) =$$

$$trao\left[\frac{1}{2}\sum_{t=1}^{T}(x_t - \Pi w_t^{'})\Omega_t^{-1}(x_t - \Pi w_t^{'})\right] =$$

$$\frac{1}{2}\sum_{t=1}^{T}trao\left[(x_t - \Pi w_t^{'})\Omega_t^{-1}(x_t - \Pi w_t^{'})\right]$$

Dado que a condição de primeira ordem de Π é $\sum_{t=1}^{T}(w_t^{'}x_t\Omega_t^{-1} - w_t^{'}w_t\Pi^{'}\Omega_t^{-1}) = 0$, aplicando o operador de vetorização, tem-se:

$$vec\left(\sum_{t=1}^{T}w_t^{'}x_t\Omega_t^{-1}\right) = vec\left(\sum_{t=1}^{T}w_t^{'}w_t\Pi^{'}\Omega_t^{-1}\right) = \sum_{t=1}^{T}vec\left(w_t^{'}w_t\Pi^{'}\Omega_t^{-1}\right) =$$

$$\sum_{t=1}^{T}\left(vec\Pi^{'}\right)\left(\Omega_t^{-1}\otimes w_t^{'}w_t\right) = vec\Pi^{'}\sum_{t=1}^{T}\left(\Omega_t^{-1}\otimes w_t^{'}w_t\right)$$

o que gera

$$vec\Pi^{'} = \left[\sum_{t=1}^{T}\left(\Omega_t^{-1}\otimes w_t^{'}w_t\right)\right]^{-1} vec\left[\sum_{t=1}^{T}w_t^{'}x_t\Omega_t^{-1}\right] \quad \text{(A-2)}$$

Esse processo é repetido para diferentes valores de $\{\gamma,\Omega_R,\Omega_E\}$. Mas, para que os diferentes valores $\{\gamma,\Omega_R,\Omega_E\}$ gerem matrizes Ω_R e Ω_E positivas definidas, trabalha-se com um vetor alternativo, $\Psi = \{\gamma,chol\Omega_R,chol\Omega_E,\Pi_R(L),\Pi_E(L)\}$, onde $chol$ representa a decomposição de Cholesky. Assim, para cada conjunto de valores de Ω_R e Ω_E estimados, Π é obtido e $logL$ é computado em (A-1).

Em seguida, dado os valores iniciais de Ψ^0, utiliza-se o método de MCMC para construir cadeias com o comprimento dos parâmetros N do modelo. Para isso, traça-se o vetor de parâmetros $\Theta^n = \Psi^n + \psi^n$, para as $n+1$ cadeias. Ψ^n representa o estado vigente e ψ^n o vetor de choques dado por $N(0,\Omega_\Psi)$. O valor inicial de Θ^0 é dado pela aproximação de segunda ordem de Taylor do modelo (1)-(4).

O modelo, então, é escrito regredindo x_t contra as *lags*. Os resíduos dessa regressão são empregados para ajustar a matriz de variância-covariância do VAR variável no tempo e então estimar Ω_R e Ω_E usando máxima verossimilhança. Dadas essas estimativas e o valor calibrado de γ, pode-se construir Ω_t e, então, os valores de $\Pi_R(L),\Pi_E(L)$ por (A-2).

A matriz diagonal inicial, Ω_Ψ, é calibrada para 1% dos valores dos parâmetros e ajustada para que as 20.000 repetições gerem uma taxa de aceitação próxima de 0,3. São feitas N = 50.000 repetições, onde as últimas 20% são retidas para inferência, como feito por Caggiano et alii (2015).

B. ESTIMAÇÃO DAS FUNÇÕES DE IMPULSO-RESPOSTAS GENERALIZADAS

Feita as estimativas apresentadas no Apêndice A, as funções de impulso-respostas generalizadas (GIRFs) do STVAR são estimadas seguindo a abordagem proposta por Koop et al. (1996). Nessa abordagem, a função de impulso-respostas generalizada é programada para captar a interação entre a evolução das

variáveis no vetor x_t e a variável de transição, que é diretamente influenciada pelo crescimento do produto, ou seja, é modelado o *feedback* da evolução do produto no vetor x_t para o indicador de transição z_t e, consequentemente, a probabilidade $F(z_{t-1})$. Com isso, a probabilidade $F(z_t)$ é endogeneizada.

Para calcular as GIRFs, de acordo com Koop et alii (1996), em primeiro lugar define-se dentro de todas as observações disponíveis o conjunto de todos os acontecimentos, $\lambda_i \in \Lambda$ e, então, se separa os acontecimentos de períodos recessivos dos acontecimentos períodos expansivos. Para cada acontecimento λ_i, calcula-se a variável de transição z_{λ_i}. $z_{\lambda_i} \leq \bar{z}$, $\lambda_i \in \Lambda^R$, onde Λ^R é o conjunto de todos os acontecimentos de períodos recessivos. Se $z_{\lambda_i} > -\bar{z}$, $\lambda_i \in \Lambda^E$, o conjunto de todos os acontecimentos de períodos recessivos.

Logo, seleciona-se um acontecimento aleatório, λ_i de cada conjunto, Λ^E e Λ^R. De cada λ_i, tem-se $\hat{\Omega}_{\lambda_i}$, dada por:

$$\hat{\Omega}_{\lambda_i} = F(z_{\lambda_i})\hat{\Omega}_R + [1 - F(z_{\lambda_i})]\hat{\Omega}_E \tag{B-3}$$

A decomposição de Cholesky da matriz de variância-covariância $\hat{\Omega}_{\lambda_i}$ é dada por:

$$\hat{\Omega}_{\lambda_i} = \hat{C}_{\lambda_i}\hat{C}'_{\lambda_i} \tag{B-4}$$

Ortogonalizando os resíduos se obtém o choque estrutural:

$$e^{(R,E)}_{\lambda_i} = \hat{C}^{-1}_{\lambda_i}\hat{\varepsilon} \tag{B-5}$$

O termo $e^{(R,E)}_{\lambda_i}$ é um vetor de choques, com horizonte h, definido em 20, nesse trabalho, tal qual:

$$e^{(R,E)*}_{\lambda_i} = \left\{e^*_{\lambda_{i,t}}, e^*_{\lambda_{i,t+1}},..., e^*_{\lambda_{i,t+h}}\right\} \tag{B-6}$$

Pode-se definir um outro conjunto de choques, $e^{(R,E)\delta}_{\lambda_i}$, semelhante ao (B-6), exceto pelo k-ésimo choque em $e^{(R,E)*}_{\lambda_i}$, que sobre uma perturbação $\delta > 0$ (em G, no caso desse trabalho). Assim, transformando $e^{(R,E)*}_{\lambda_i}$ e $e^{(R,E)\delta}_{\lambda_i}$ em

$$\hat{\varepsilon}^{(R,E)*}_{\lambda_i} = \hat{C}^{-1}_{\lambda_i} e^{(R,E)*}_{\lambda_i} \tag{B-7}$$

e

$$\hat{\varepsilon}^{(R,E)\delta}_{\lambda_i} = \hat{C}^{-1}_{\lambda_i} e^{(R,E)\delta}_{\lambda_i} \tag{B-8}$$

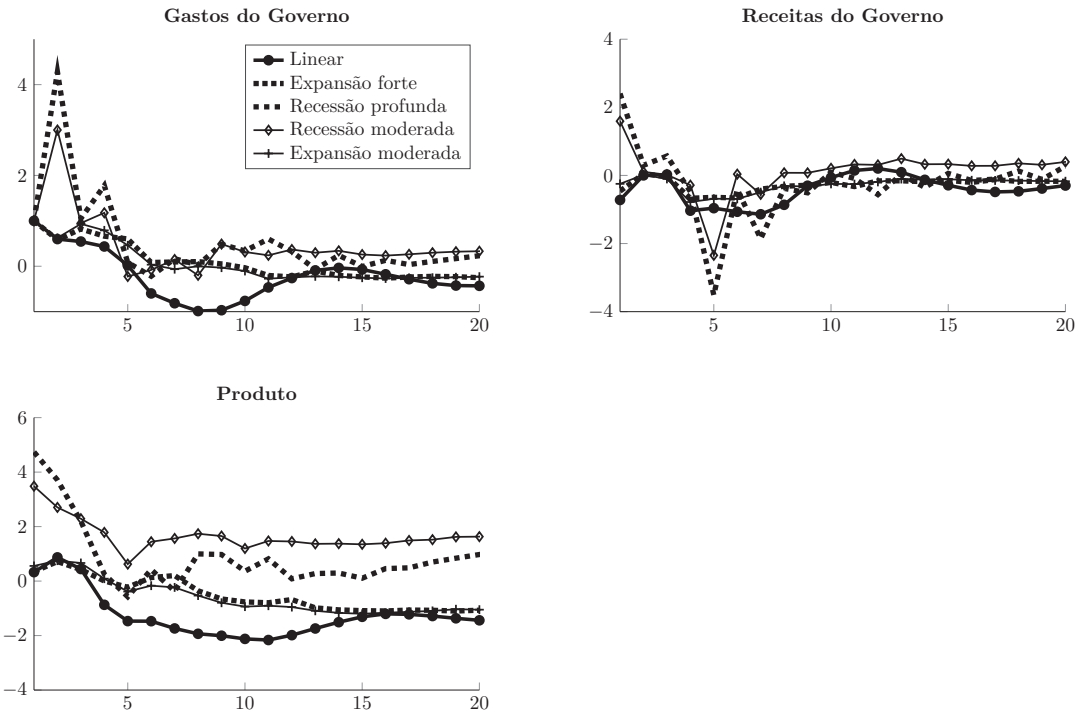
pode-se gerar as duas sequências de $x^{(R,E)*}_{\lambda_i}$ e $x^{(R,E)\delta}_{\lambda_i}$ e a $GIRF^j(h,\delta,\lambda_i)$. Condicionada ao acontecimento λ_i, são calculadas $GIRF$ de $j = 1,...,500$, ou seja, $GIRF^1(h,\delta,\lambda_i)$, $GIRF^2(h,\delta,\lambda_i),...,GIRF^{500}(h,\delta,\lambda_i)$.

Além disso, os passos pra se obter a $GIRF$ são repetidos para os $i = 500$ acontecimentos aleatórios recessivos e expansivos. Então, toma-se a média para se obter $GIRF^R(h,\delta,\Lambda_R)$ e $GIRF^E(h,\delta,\Lambda_E)$. O intervalo de confiança de 90% da impulso-resposta é construído, pegando de cada horizonte de cada estado o 5º e o 95º percentis das densidades de $GIRF^{(1:500),R}$ e $GIRF^{(1:500),E}$.

Em resumo, são simulados dois cenários, um onde todos os choques são identificados pela decomposição de Cholesky da matriz de variância e covariância dada em (3), e outro com os mesmos choques mais um $\delta > 0$ referente à primeira realização do choque. As GIRFs dos choques δ são dadas pela diferença entre os dois cenários. Para cada condição inicial, z, é salva a realização mediana de 500 realizações estocásticas diferentes. Essas etapas são repetidas até que 500 condições iniciais, associadas a recessões (expansões), sejam consideradas. Então, as GIRFs são construídas considerando estas 500 realizações medianas e normalizando os choques.

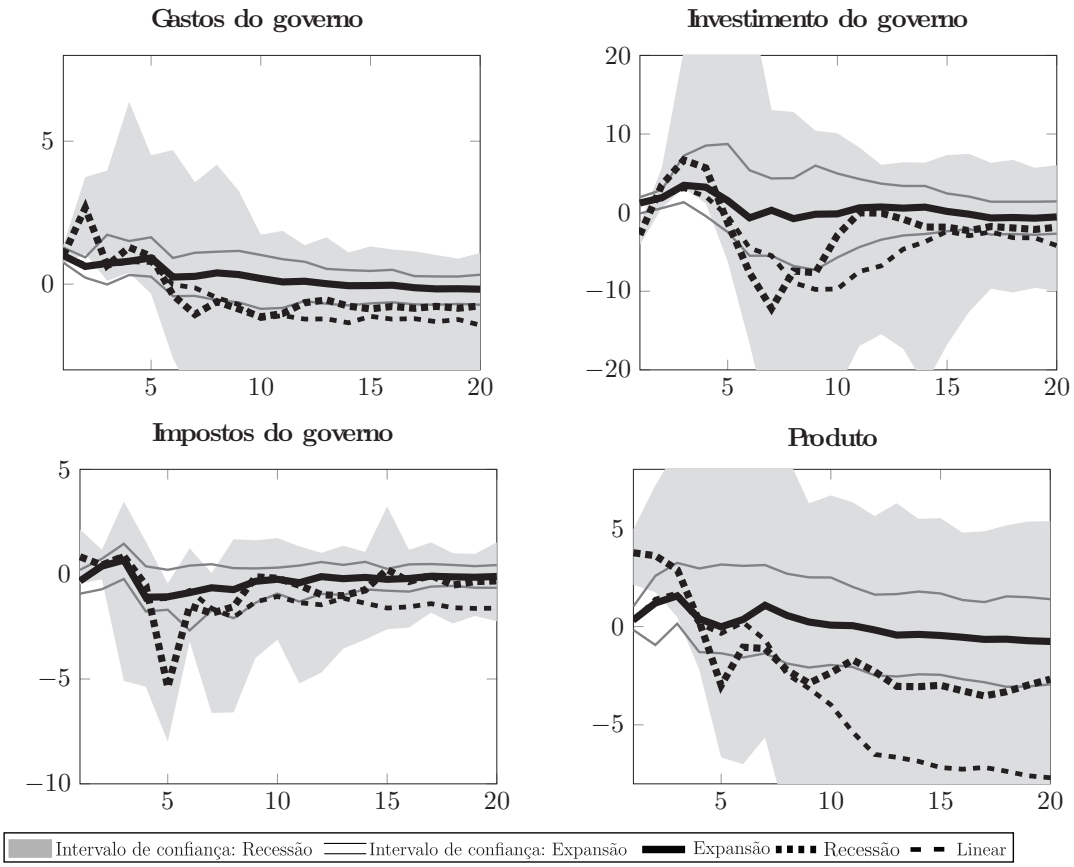
C. IMPULSO-RESPOSTAS GENERALIZADAS ESTIMADAS

Figura C-1: Impulso Resposta Generalizada para choque nos gastos do governo: Modelo Linear, Recessões e Expansões Extremas



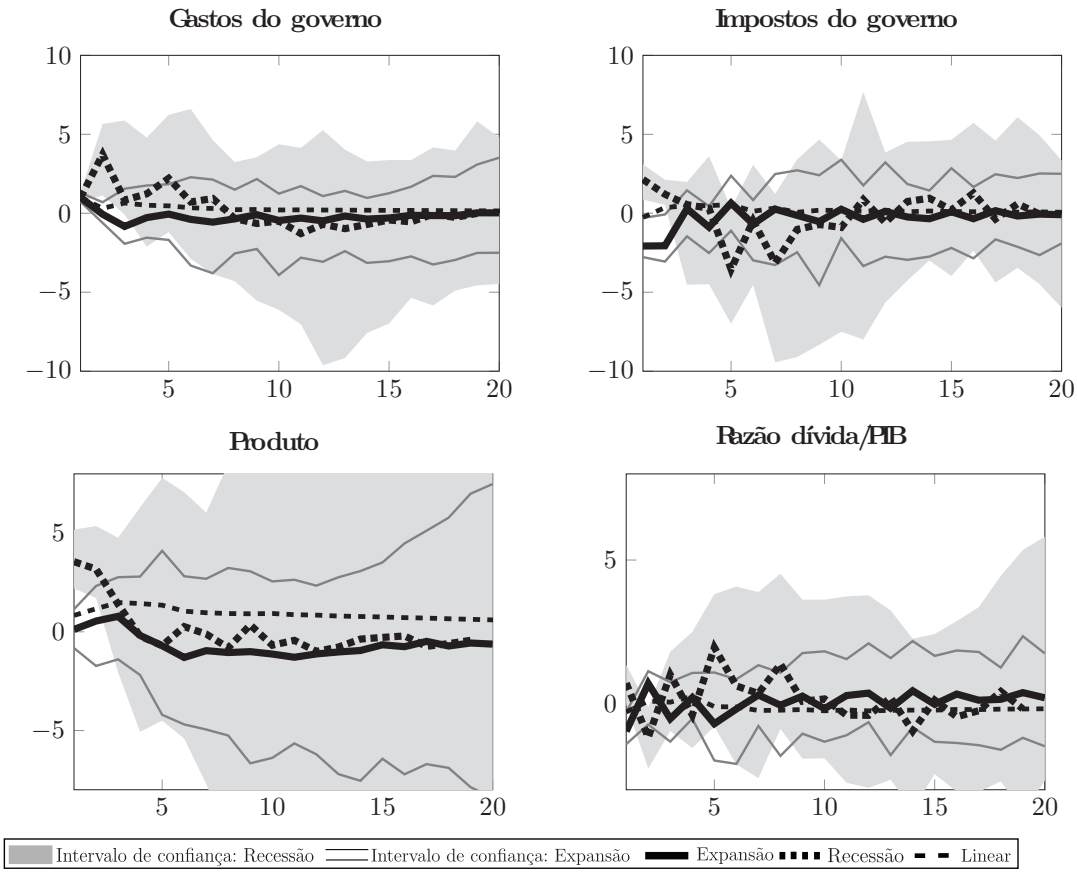
Notas: Recessões profundas e expansões fortes associadas a acontecimentos a dois desvios-padrões da média da variável de transição. Respostas médias a um choque nos gastos do governo normalizado para um. Var estimado com uma contante e três *lags* com dados de 1999T4 a 2015T4.

Figura C-2: Impulso Resposta Generalizada para choque no consumo do governo controlado pelos gastos com investimento: Modelo Linear, Recessões e Expansões



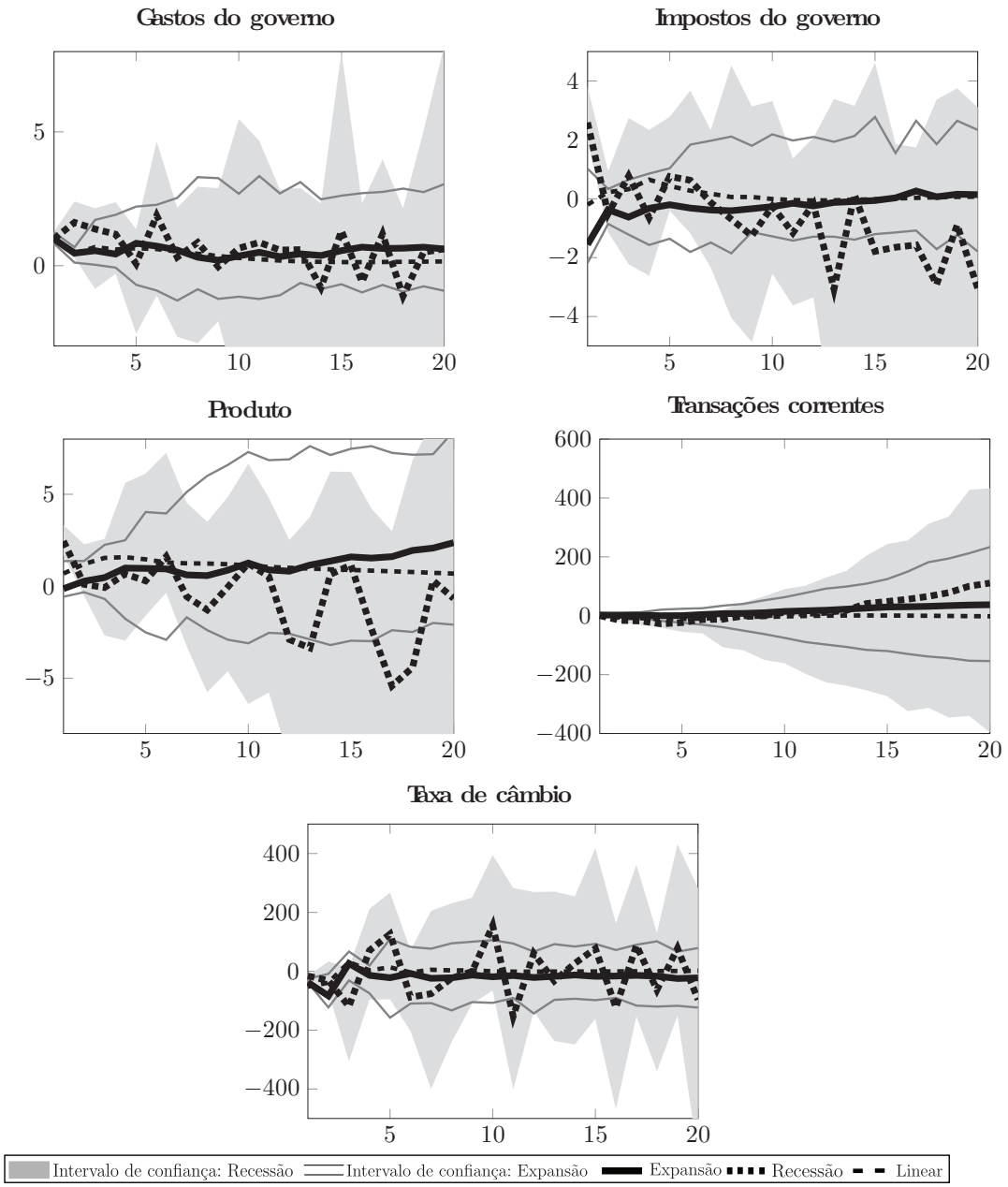
Notas: Respostas médias a um choque no consumo do governo normalizado para um. Var estimado com uma contante e três *lags* com dados de 1999T4 a 2015T4.

Figura C-3: Impulso Resposta Generalizada para choque nos gastos do governo controlado pela razão dívida/PIB: Modelo Linear, Recessões e Expansões



Notas: Respostas médias a um choque nos gastos do governo normalizado para um. Var estimado com uma contante e três *lags* com dados de 2002T3 a 2015T4.

Figura C-4: Impulso Resposta Generalizada para choque no consumo do governo controlado pelo grau de abertura econômica e taxa de câmbio: Modelo Linear, Recessões e Expansões



Notas: Respostas médias a um choque nos gastos do governo normalizado para um. Var estimado com uma contante e três *lags* com dados de 1999T4 a 2015T4.

r b e

ISSN 0034-7140 (print) | ISSN 1806-9134 (online)
DOI 10.5935/0034-7140.20170016
Revista Brasileira de Economia, 71(3), 347–359

What Drives Inequality of Brazilian Cross-State Household Credit?*

Paulo Matos†
Joaquim Correia‡

Contents: **1. Introduction; 2. Brazilian credit market; 3. Methodology; 4. Empirical exercise; 5. Conclusion.**

Keywords: **Household credit; Panel data; Poverty; Per capita income; Gini index; Years of schooling.**

JEL Code: **O54, O16, G18.**

According to Matos et al. (2013) credit policy in Brazil has been discriminatory and strongly characterized by a regional bias. We address this issue by aiming to identify Brazilian cross-state credit drivers. Methodologically we follow Matos (2017) by proposing a panel model to estimate relationships between real per capita Brazilian household credit and a set of relevant social, economic and financial variables. Our main findings considering all federal subnational entities during the period from 2004 to 2013 suggest that demand for credit plays a more relevant role than the supply thereof. Our evidence may be useful in writing theoretical models dealing with credit patterns in Brazil.

Segundo Matos et al. (2013), a política creditícia no Brasil tem sido discriminatória e fortemente caracterizada por um viés regional. Este artigo aborda este tema, visando identificar os determinantes da concessão de crédito nos estados brasileiros. Metodologicamente, o exercício empírico segue Matos (2017) ao se basear fazer uso de um painel na estimação das relações entre o crédito real per capita obtido pelas famílias e um conjunto de variáveis sociais, econômicas e financeiras. Considerando todas as unidades da federação ao longo do período de 2004 a 2013, os resultados obtidos sugerem que os determinantes pelo lado da demanda sejam mais influentes que os determinantes pelo lado da oferta. Essa evidência empírica pode ser útil no desenho de novos arcabouços teóricos sobre crédito no país.

1. INTRODUCTION

When observing the literature discussing the economic role played by credit markets – from seminal papers such as Schumpeter (1934) to more recent contributions such as Rajan and Zingales (1998) – most of the related research diverges into two different but related approaches: a) the effects of credit on social and economic variables[1] and b) the analysis of credit drivers.

Here, we add to the discussion about credit drivers, following one of the most promising lines of research developed by Cottarelli et al. (2005), Kiss et al. (2006), Hansen and Sulla (2013) and Matos (2017).

*First draft December 2015. The financial support from CNPQ-Brazil is gratefully acknowledged. We are thankful to anonymous referees and the editor for their insightful comments. We thank Ricardo Brito, Arnaldo Santos and seminar participants at Federal University of Ceará for their comments and suggestions. The usual disclaimer applies.

†CAEN-UFC, Graduate Program in Economics at Federal University of Ceará, Fortaleza – Ceará, Brazil.E-mail: paulomatos@caen.ufc.br

‡CAEN/UFC.E-mail: joaquimfcneto@gmail.com

[1]Concerning the role of financial intermediation on poverty, development and income inequality, we can mention some recent World Bank's contributions, as Ravallion (2001) and Beck et al. (2006).

Received on May 24th, 2017
Final Acceptance on June 01th, 2017

More specifically, we are the first to enter this debate aiming to identify significant drivers able to explain the heterogeneity of household credit across states of Brazil (including the federal district). This country is a very interesting case because the Brazilian context is even more cumbersome and depressed than in other emerging economies. Following this, we list features that motivate our country choice.

First, Bernanke and Gertler (1995) suggest theoretically that the volume of credit is pro-cyclical, but this is not necessarily true for Brazil. The austerity established within this economy in the '90s is held in check by the deepest economic and political crisis in its recent democratic history. In a counterintuitive way, credit volume provided to Brazilian families is reaching high levels even for a scenario characterized by elevated household loan interest rates. Under the argument that Brazil has a low credit to GDP ratio (54% in mid-2015) the government has stimulated the debt increase among the population, irrespective of the level of human capital, profile of default or employment status. Therefore, for the first time in decades, household debt in relation to cumulative income over the last twelve months within the National Financial System is higher than 46%, even with annual interest rates close to 35% for individuals considering all modalities of lending, according to the Central Bank of Brazil. A decade ago, this debt ratio was less than 19%.

Second, one can see a recent, intuitive but uncomfortable evidence reported in Matos et al. (2015) about Brazilian household loan delinquency. According to the main findings of this paper, poverty and unemployment play a significant role in household decisions to honor a financial commitment or not.

Third, Brazilian economists used to say: "there are two nations in the same territory"; this usually refers to income. However according to Matos et al. (2013), based on an application of the Phillips and Sul (2007) framework to Brazilian states, it is also possible to prove the existence of a discriminatory credit policy. These authors are able to illustrate the formation of two clubs strongly characterized by a regional bias with, a representative presence of states located in the Northeast and North regions included in the second group.

Fourth, unfortunately this heterogeneity is broader because one can see many other disparities related to the financial system or human capital. For instance, on average in the Northeast and North regions, families and companies may have difficulty gaining access to the banking sector because there were 5.5 agencies per 100,000 inhabitants; whereas in the South, there were 14 agencies per 100,000 inhabitants (according to the Central Bank of Brazil, during the period of 2004 to 2013). On the demand side, while the Northeast has the lowest values for years of schooling (ranging from 4.9 to 6.1) the average value in the South is higher than 7.4 years.

Fifth, Brazilian and other Latin American economies have experimented with a recent financial liberalization starting in the 1990s, after moving toward an open, market-based development model in place of one that relied on the state and was closed to foreign trade and capital flows. Stallings (2006), Edwards et al. (2007) and Hira and Gaillard (2011) are relevant sources about the consequences of this experience. To summarize, this evidence suggests the following key ideas: i) there exists a discriminatory credit policy across Brazilian states, and ii) there exists a robust, persistent and broader heterogeneity between those states based on many social and economic variables. This unusual scenario has motivated us to study Brazilian credit drivers through panel data, covering the period from 2004 to 2013, applying this method to all federal subnational states.

We are convinced that it is timely and opportune to ask how human capital, employment and income, combined with the development of the Brazilian financial system can influence heterogeneous levels of Brazilian household credit volume.

Methodologically, we follow the recent related literature, such as Kiss et al. (2006), Hansen and Sulla (2013) and Matos (2017), to propose a contemporaneous panel model; in other words, we estimate the relationship between the level of real per capita household and a set of economic variables rather than trying to identify causality in the model. However, because credit may display autoregressive behavior, our approach can be useful in a dynamic version when one intends to measure the effects of timing and causality in this market.

The key issue to be addressed here is the choice of a representative, feasible, parsimonious and observable set of exogenous variables from among a wide range of intuitive variables that may drive the evolution of credit in Brazilian states. Although we have looked to extract these series from many different data sources, our empirical exercise lacks cultural, historical and political variables. In our final specification, we work only with macroeconomic, microeconomic, labor market, social and financial variables.

Our main findings suggest that demand for credit plays a more relevant role than supply thereof. More specifically, households' levels of human capital and real income are significantly related to improving credit; reducing the level of inequality across families is also important. We also show that poverty can motivate household credit through microcredit programs developed by public banks.

This article is structured into five sections including this introduction. Section 2 gives an account of an overview about credit evolution in Brazil, before presenting our methodology in the third section. We discuss our empirical analysis in the fourth section. Concluding remarks are offered in the fifth section.

2. BRAZILIAN CREDIT MARKET

According to the World Bank dataset, Brazil's credit to GDP growth rate during the period from 2004 to 2011 was one of the highest worldwide at 11.4%. According to Matos (2017) the average value of credit to GDP ratio for Brazil during this period is approximately 39%, the third ranked value in a sample containing 14 Latin American economies. Only Panama and Honduras, with 77% and 44%, respectively, show higher ratios than Brazil.

On the one hand, those numbers characterize the Brazilian credit market as a highlight among emerging countries. In this sense, we must mention that the strong growth of household credit in this period is due to the growth of payroll loans and incentives to mortgage, according to Doctor (2015).

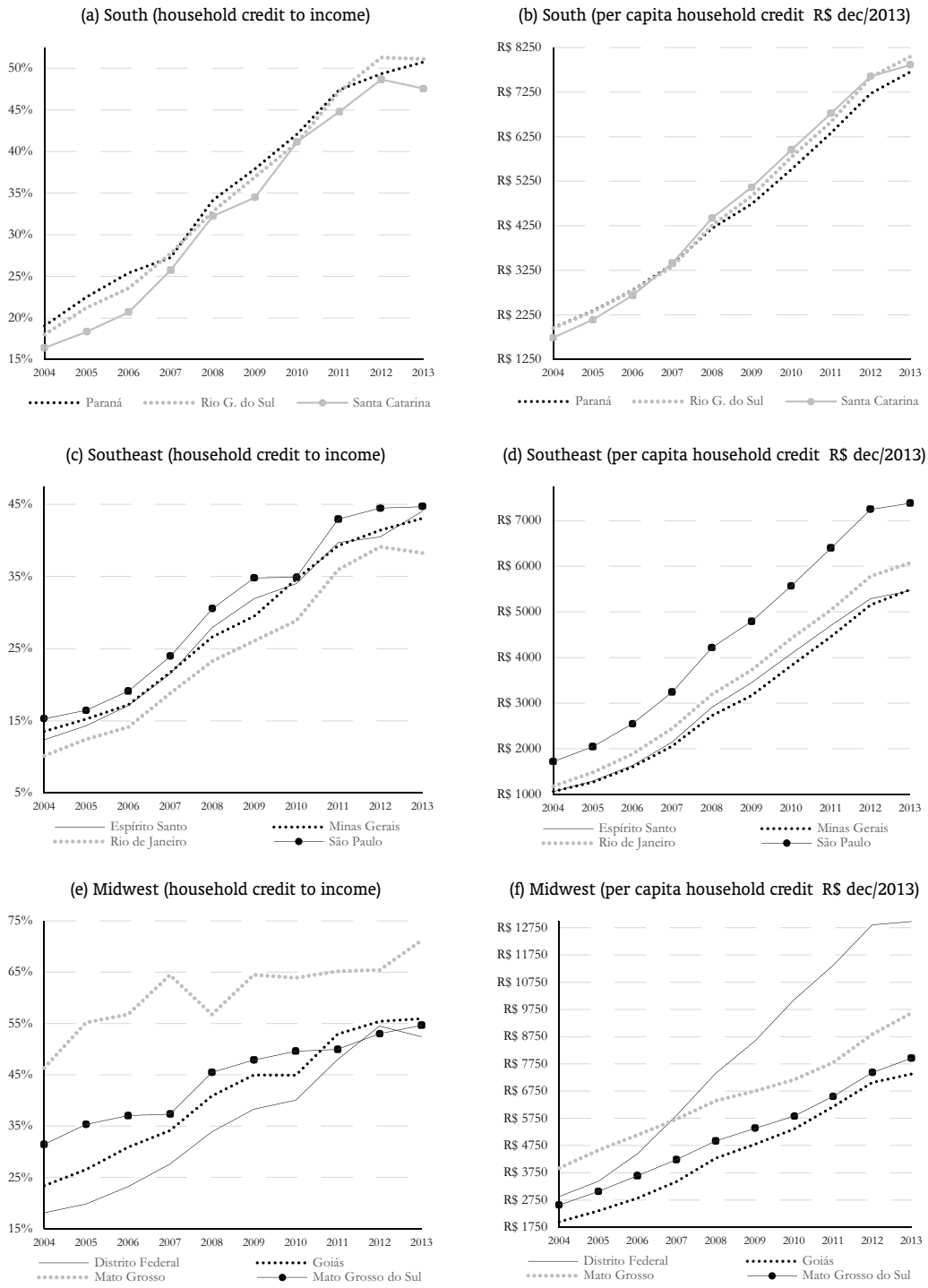
However, it is necessary to disaggregate credit in the Brazilian states so that inequality can be illustrated in that country. In Figure 1, we plot the evolution from 2004 to 2013 of our endogenous variable, real per capita household credit for each Brazilian federal subnational entity, considering all types of credit to families.

The evolution of the aggregate volume of household credit shown in Figure 1 shows a visible growth trend in all time series, both in absolute per capita terms or as a ratio of per capita income. Except for entities located in the Midwest region, the pattern of evolution is such that Brazilian entities begin with a credit to income ratio ranging from 5% to 20%, and reach ratios ranging from 35% to 55% at the end of the period (2013). The exceptions are Mato Grosso with 70% and Amazonas and Pará, with less than 35%.

When we address real per capita series, the divergences are much more worrying because the regions with lower per capita income also display lower levels of per capita household credit, as we can see in the North and Northeast regions. Once more, in Amazonas, Pará and Sergipe, we see per capita credit levels less than R$ 2750 in 2013, while in the Southeast region all entities reach values above R$ 5000, and in the Southern region, the level is close to R$ 8000. In accordance with previous publications such as Matos et al. (2013), we see in this figure apparent similarities and convergence patterns among entities located in a same region, especially in the regions of South, Southeast and Northeast. In the Midwest region we find some extreme series when we consider credit to income ratio, such as Mato Grosso, and Distrito Federal when we observe real per capita credit.

This heterogeneous cross-state context has motivated us to identify possible credit drivers.

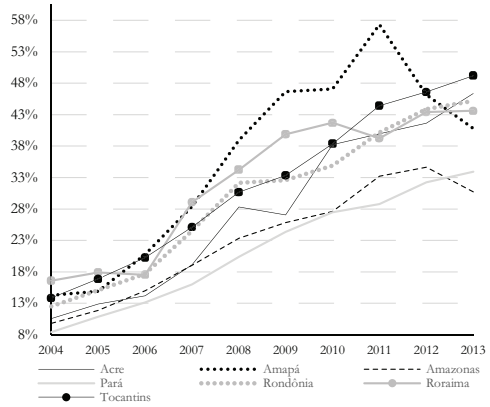
Figure 1: Brazilian household credit (per capita and as a ratio of household income), over the period from 2004 to 2013, for all Brazilian federal subnational entities.



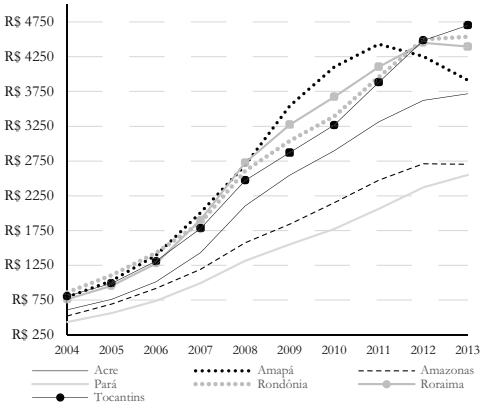
Source: Central Bank of Brazil and IPEADATA.

Figure 1: Continued from previous page.

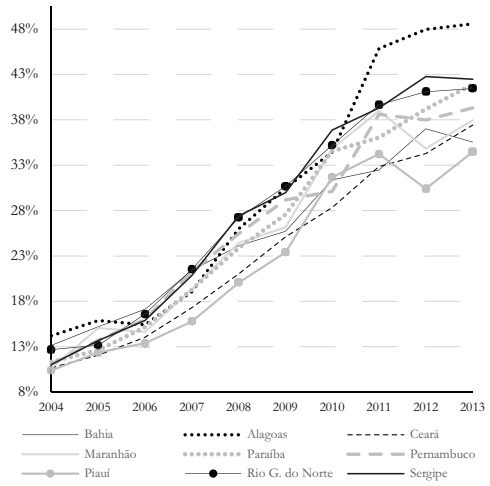
(g) North (household credit to income)



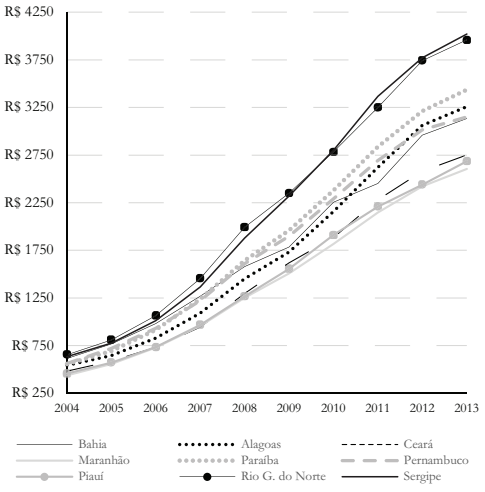
(h) North (per capita household credit R$ dec/2013)



(i) Northeast (household credit to income)



(j) Northeast (per capita household credit R$ dec/2013)



Source: Central Bank of Brazil and IPEADATA.

3. METHODOLOGY

Some recent empirical contributions to the discussion of credit drivers are Cottarelli et al. (2005), Kiss et al. (2006), Hansen and Sulla (2013) and Matos (2017). The main difference between the existing literature and our work concerns cross-section data because these studies have used cross-country panel data, while our purpose is modelling the cross-state heterogeneity in Brazil. As we have already emphasized, the key issue to be addressed here (associated with the estimation methodology) is the choice of a representative, feasible, parsimonious and observable set of exogenous variables from among a wide range of intuitive variables which may drive credit evolution in Brazilian entities.

Regarding the supply side, we can see in Brock and Suarez (2000) some possible impacts of household delinquency on credit market issues in Latin American economies, as this phenomenon spreads

and increases in volume, motivating us to use this variable in an explanatory role. According to Holmstrom and Tirole (1997), the level lending banks engage in depends positively on the quantity of bank capital. According to Bernanke and Blinder (1992), an exogenous reduction in insured deposits is accompanied by a decreased in bank lending. To measure the effect of the presence of banks and financial institutions we followed Nguyen (2014), who shows that branch closures (even in crowded markets in the United States) can have large effects on local credit supply. Finally, from the perspective defended by Alesina and Perotti (1996), highly unequal societies would create incentives for individuals to engage in activities outside the formal market, as well as discouraging socio-political instability in the accumulation of capital in view of the increased uncertainty affecting the credit market.

The demand side seems to be more complex, since some intuitive variables are latent for developing countries. According to Cottarelli et al. (2005), growth in the credit to private sector ratio was essentially driven by domestic saving flows, which suggests to using variables that capture this saving profile of families and companies. Becker (1975) states that the poor tend to save by investing in human capital, while the rich save by accumulating physical and financial assets, while according to Coudert and Pouvelle (2010) a higher level of income can be associated with a higher level of credit. In the same direction, richer families and bigger companies tend to offer better collateral sources. For the same reasons, in general, societies with more years of schooling should be able to demand more credit. Regarding the labor market, most credit programs (even in developing economies) require a minimum level of formal employment, except for some programs such as some temporary microcredit lines.

Given our purpose, we believe that the class of linear models is not adequate in which the error in each time period is assumed to be uncorrelated with the explanatory variables in the same time period because this may be a strong and unnecessary assumption. We can confront this issue and also account for the omitted variables problem by applying a balanced panel data framework, if we assume that the unobserved effect – a random variable, drawn from the population – can be treated as a time constant. Our panel model in its full and unrestricted version can be written as:

$$\begin{aligned} HC_{i,t} = \varphi_i + \varphi_H \cdot HD_{i,t} + \varphi_D \cdot BD_{i,t} + \varphi_N \cdot NB_{i,t} + \varphi_G \cdot GI_{i,t} \\ + \varphi_S \cdot SC_{i,t} + \varphi_U \cdot UN_{i,t} + \varphi_P \cdot PO_{i,t} + \varphi_I \cdot IN_{i,t} + \varepsilon_{i,t} \end{aligned} \tag{1}$$

where the subscript i refers to each Brazilian entity, and t to each year of our sample. Here, our dependent variable is real per capita household loan volume, denoted by HC.

On the supply side, household delinquency is indicated by HD, BD refers to real per capita bank deposits, number of bank branches per 100,000 adults is given by NB and GI is Gini income inequality index. On the demand side, SC measures years of schooling, UN corresponds to unemployment rate as a percent of total labor force, PO meaning the rate of poor people, and household real per capita income is given by IN.

The nature of this unobserved effect (φ_i) and whether it will be treated as a random or fixed effect is discussed based on respective tests, whose results are considered to define our final panel estimation setup.

Regarding the properties of estimators, we do not have to care about the cases where T is of the same order as N, or where T is much larger than N because our N could be seen as sufficiently larger than T, so we can assume rough independence in the cross-section dimension. Finally, while some other econometric techniques require strong assumptions about distributions, under the panel model, we only need the stationarity of the underlying process.

4. EMPIRICAL EXERCISE

4.1. Dataset and summary statistics

In principle, all applied research working with credit, economic, social and financial variables in developing economies has to address the trade-off between T and N. For the Brazilian economy, our first

main limitation concerns the time-series, even when extracting the series from the most important economic data source – the Central Bank of Brazil and IPEADATA. We aim to have a balanced panel for all twenty-seven Brazilian federal subnational entities, so we work with main variables on a yearly frequency during the period from 2004 to 2013 (the most recent year of macroeconomic series). With regard to the endogenous variable, we have already discussed its evolution in each Brazilian entity, plotted in Figure 1. In Table 1, we observe its average values, considering the period from 2004 to 2013. It is usually believed that Brazil is a continental country and that the heterogeneity of its per capita income is not good. However, Matos et al. (2013) show that for the period from 2004 to 2009, we also see a kind of abyss in terms of credit policy created by a huge dispersion, and by the formation of two groups strongly characterized by regionalism. This regionalism showed a representative presence of states located in the Northeast and North regions that was present in the second group. This trend remains true until 2013, according to the first column in Table 1.

In the Northeast region, credit values range from approximately R$ 1,440.00 to R$ 2,210.00; this upper bound is less than half the volume observed in Paraná – the lower bound in the South region. We can also show that the wealthiest regions (the South and Southeast) are upwards of three orders of magnitude larger than the values presented by regions with lower per capita credit, such as the Northeast. In the North, values are higher than R$ 2,000.00 for all states, with the exception of Pará and Amazonas. In the Midwest and Southeast, there is an intraregional homogeneity with some outlier behaviors due to the presence of São Paulo and Distrito Federal.

Concerning all exogenous variables used here, we report their average values in Table 1. Since we aim to model the evolution of household credit in Brazil, our explanatory variables should be able to measure its oscillation among Brazilian entities and along the time series, considering idiosyncratic Brazilian issues related to the financial system, human capital, the labor market, poverty and income distribution.

We observe that, in addition to having higher values for per capita credit, the South and Southeast regions also display better or more comfortable values for most exogenous variables, corroborating what Brazilian economists used to say: "there are two nations in the same territory". On the supply side, we can see that in the second column in Table 1, in accordance with previous publications of the Central Bank of Brazil, there are apparent similarities and convergence patterns among entities located in the same region. However, in the South, loan delinquency ranges between 3% and 5%, in the Northeast (with the exceptions of Sergipe, Rio Grande do Norte and Paraíba) loan delinquency is greater than 6.8%. In the Southeast, Rio de Janeiro is dispersed across the other ones, whose patterns are very similar to each other.

With regard to the ratio of the number of branches per 100,000 habitants, the disparity again takes place in a regional context, where Maranhão acts as the lower bound in terms of perception of the banking system, with a rate almost four times less than that of São Paulo. In the Northeast and North regions families and companies may have difficulty accessing the banking sector because there are 5.5 agencies per 100,000 inhabitants, while in other regions, this ratio is closer to or above 10 per 100,000 inhabitants. In the South, this variable has an average close to 14 per 100,000 inhabitants.

We also take into account the total value of demand, time and savings deposits at domestic money deposit banks – commercial banks and other financial institutions that accept transferable deposits, such as demand deposits. This variable ranges from about R$ 300 in Piauí to more than R$ 22,260 in Distrito Federal.

Heterogeneity remains even if we limit ourselves to an intraregional analysis, with São Paulo showing a volume of per capita deposits close to four times that of Minas Gerais.

Finally, looking at the levels of Gini index we can see a cross regional oscillation, so that the Northeastern and Northern states, as well as Distrito Federal, have higher values, but with a lower dispersion if we compare them with poverty levels for all entities.

On the demand side, while states in regions of the Midwest, Southeast and South have rates of poverty ranging between 5.6% and 11.7%; other regions present astounding numbers such as 40.6% in

Table 1: Average values of the variables over the period from 2004 to 2013, for all Brazilian federal subnational entities.

	Endogenous variable	Supply side					Demand side		
	Household real per capita loan volume (R$ dec/2013)	Household loan delinquency rate (%)	Number of Bank Branches per 100,000 habitants	Real per capita bank deposit (R$ dec/2013)	Gini inequality index	Years of schooling	Unemployment rate as a percent of total labor force (%)	Poverty (%)	Household real per capita income (R$ dec/2013)
	Source: Central Bank of Brazil	Source: Central Bank of Brazil	Source: Central Bank of Brazil	Source: Central Bank of Brazil	Source: IPEADATA	Source: IPEADATA	Source: IPEADATA	Source: IPEADATA	Source: IPEADATA
Northeast region									
Alagoas	1,739.21	6.812	4.565	532.02	0.564	4.939	10.565	39.898	5,643.36
Bahia	1,779.73	7.282	5.874	865.93	0.554	5.691	10.585	32.596	6,783.93
Ceará	1,515.86	6.925	4.755	1,786.78	0.547	5.667	7.666	33.149	6,265.49
Maranhão	1,442.51	7.966	4.056	325.07	0.560	5.261	7.990	40.596	5,433.31
Paraíba	1,887.03	5.849	5.180	519.05	0.566	5.528	9.124	33.051	6,953.80
Pernambuco	1,809.52	7.671	5.862	1,089.94	0.555	6.059	11.330	35.062	6,670.69
Piauí	1,479.59	7.122	4.138	302.23	0.563	5.054	5.945	36.406	6,345.89
Rio Grande do Norte	2,206.08	5.880	5.179	759.77	0.559	6.014	10.199	29.580	7,577.43
Sergipe	2,194.71	5.963	8.373	1,306.01	0.556	6.135	10.266	29.462	7,492.07
North region									
Acre	2,201.63	5.029	5.791	690.54	0.572	6.254	8.217	30.164	7,740.09
Amapá	2,815.85	6.033	5.104	402.96	0.515	7.686	12.861	26.056	7,725.02
Amazonas	1,677.00	6.752	4.716	801.90	0.529	7.230	10.757	28.591	7,095.44
Pará	1,434.96	6.117	4.500	566.40	0.516	6.126	8.961	30.110	6,520.81
Rondônia	2,729.49	4.809	6.655	825.90	0.511	5.193	6.930	20.345	8,893.13
Roraima	2,751.49	7.032	5.270	660.77	0.541	7.318	10.382	27.293	8,083.83
Tocantins	2,656.44	5.269	6.123	526.20	0.533	6.340	6.899	25.023	7,971.95
Midwest region									
Distrito Federal	7,993.85	4.632	13.277	22,260.47	0.605	9.462	10.795	10.660	21,521.44
Goiás	4,558.85	4.942	10.329	955.69	0.510	6.938	7.105	10.043	10,699.47
Mato Grosso	6,593.70	4.821	9.988	1,175.50	0.515	6.810	6.967	10.789	10,661.08
Mato Grosso do Sul	5,157.57	3.718	9.242	995.21	0.523	6.964	6.542	9.729	11,351.56
Southest region									
Espírito Santo	3,202.58	4.495	10.848	2,341.72	0.523	7.153	7.869	10.272	10,755.52
Minas Gerais	3,080.62	4.496	9.864	1,951.89	0.514	6.768	7.707	10.438	10,347.20
Rio de Janeiro	3,522.25	6.568	11.384	3,928.09	0.543	8.223	9.684	11.653	13,780.25
São Paulo	4,515.72	4.921	15.059	8,501.05	0.505	8.121	8.880	9.092	14,176.02
South region									
Paraná	4,619.82	3.860	12.756	4,084.51	0.504	7.346	5.906	10.582	12,524.44
Rio Grande do Sul	4,757.91	3.590	14.046	3,170.53	0.500	7.326	6.070	10.703	13,052.96
Santa Catarina	4,770.15	3.736	14.347	2,052.12	0.452	7.610	4.046	5.660	13,884.43

Maranhão or 39.9% in Alagoas. Following Barro and Lee (2013) suggestion for human capital, we use years of schooling, and according to Table 1, the Northeast region shows us the lowest values, ranging from 4.9 to 6.1. However, the average value for the South region is higher than 7.4 years.

With few exceptions (such as Piauí), unemployment levels in the Northeast region are quite higher than those registered in the South region. Observing the Midwest region's behavior as suggested by other variables, we can characterize it as a type of intermediate region geographically as well as statistically. The exception again is Distrito Federal, which has a very high unemployment rate comparable to the Northeast region. Regarding per capita income, except for São Paulo in the Southeast region and Distrito Federal in the Midwest region we observe an intraregional homogeneity among the average values. As seen with credit, the values displayed by the South and Southeast regions for per capita income range from 2 to 3 times the values registered in the North and Northeast regions.

4.2. Main results and discussion

First, as a type of preliminary test we can see that all variables are stationary at 1%, except for the number of bank branches and bank deposits, according to Table 2. To address this issue, we work with these series in level and also in first difference (stationary series) to see the robustness of our results.

Table 2: Unit Root Panel Test^{a,b}.

Supply side			
Household loan delinquency rate	Number of Bank Branches per 100,000 adults	Real per capita bank deposit	Gini inequality index
-13.0776**	1.2576	-0.4136	-19.6156**
[0.0000]	[0.8957]	[0.3396]	[0.0000]
Demand side			
Years of schooling	Unemployment rate as a percent of total labor force	Poverty	Household real per capita income
-6.7488**	-3.8082**	-8.5309**	-9.6995**
[0.0000]	[0.0001]	[0.0000]	[0.0000]
Endogenous			
Household real per capita loan volume			
-2.5147**			
[0.0060]			

Note: ^{a} Levin et al. (2002) panel unit common root test with intercept, over the period from 2004 to 2013. ^{b}Respective p-values are reported in the box brackets.Indicates the rejection of nonstationary hypothesis at 5% level.*Indicates the rejection of nonstationary hypothesis at 1% level.

Concerning the setup, according to Table 3 we must treat unobserved effect φ_i as a fixed effect based on the rejection of the Hausman (1978) test, and because we are working with all Brazilian entities. To

be robust to heteroskedasticity, we conduct our statistical inferences using the White cross-section covariance coefficients method. See Wooldridge (2002) for a further discussion about our estimation setup.

Our main results are also reported in Table 3. We adopt an empirical strategy aiming to exploit the lack of a microfundamented model by performing a type of backward procedure, estimating from an unrestricted (full) version to a better-specified one. In detail, this backward procedure suggests that the best fitting version is the one with the highest value for the power of explanation considering all possible combinations of all variables (i.e. all eight variables, then combinations considering only seven variables, then combinations considering only six variables, and so on). We also estimate restricted versions, taking into account only one side, demand or supply, to explain the levels of equilibrium of Brazilian household credit.

The results of the unrestricted specification in the first column of Table 3 show some instances of individual insignificance in supply side variables and unemployment. The results of one-sided versions (second and third columns of Table 3) show that observing only demand or supply variables does not seem to be an adequate solution because some individual insignificance or even explanatory power, mainly in the supply-side model.

We provide our final conclusive evidence based on the better-specified version (fourth column of Table 3), where all parameters are individually significant at 1%. It is nontrivial that financial system variables do not seem to be relevant in this final specification. As one could expect theoretically, for the supply side our evidence corroborates Alesina and Perotti (1996), since highly unequal societies could affect the credit market. Regarding the demand side, we are aligned with Coudert and Pouvelle (2010) when we show that a higher level of income can be associated with a higher level of credit.

We can also show a positive and significant relation between household credit and human capital measured by years of schooling; at the same time, our results suggest that poorer individuals tend to obtain more loans in Brazil. These results are qualitatively robust, using the number of bank branches and bank deposits in level or in first difference. Both variables are not relevant in our better specified version.

Our final conclusive evidence is based on the F-statistic and our adjusted R^2, which is higher than 95%. This leads us to believe that we have offered an evidence about the overall fit of the model proposed; thus we are able to suggest some feasible policies.

Table 3: Full Model Panel Estimation[a,b,c,d].

	Unrestriced model	Restricted models		Better specified model
		Supply side	Demand side	
Constant	-5.132e-03*	1.212e-05**	-1.318e-05**	-4.777e-04*
	[0.036]	[0.001]	[0.000]	[0.030]
Supply side				
Household loan delinquency rate	0.777	-45.671		
	[0.975]	[0.514]		
Number of Bank Branches per 100,000 adults	108.861	413.565*		
	[0.058]	[0.026]		
Real per capita bank deposit	0.116		0.657**	
	[0.259]	[0.000]		
Gini inequality index	-1.573e-04**	-2.540e-05**		-1.633e-05**
	[0.000]	[0.000]		[0.000]
Demand side				
Years of schooling	803.728**		1.064 e-03**	866.906**
	[0.006]		[0.000]	[0.003]
Unemployment rate as a percent of total labor force	6.142		-18.195	
	[0.844]		[0.581]	
Poverty	86.883**		54.496**	90.995**
	[0.000]		[0.000]	[0.000]
Household real per capita income	0.862**		0.860**	0.931**
	[0.000]		[0.000]	[0.000]
Complementary results				
Adjusted R^2	95.646%	79.622%	93.765%	95.324%
F-statistic	174.819	36.034	135.852	183.772
	[0.000]	[0.000]	[0.000]	[0.000]
Hausmant test	144.035	66.899	139.417	144.900
	[0.000]	[0.000]	[0.000]	[0.000]

Note: [a]Panel estimation results based on estimation over the period from 2004 to 2013, for all twenty-seven Brazilian federal subnational entities. [b]We specify the White cross-section covariance coefficients method, to get robustness to heteroskedasticity. [c]We take into account for fixed effects, as suggested by the rejection of the H0 of the Hausman (1978) test and because we use all Brazilian entities. [d]Respective p-values are reported in the box brackets. *Indicates the rejection of individually insignificant parameter hypothesis at 5% level. **Indicates the rejection of individually insignificant parameter hypothesis at 1% level.

5. CONCLUSION

In practice, our evidence indicates that household characteristics, mainly their levels of human capital and income, are relevant by providing society more equal access to the financial system as growth catalyzer. The evidence as to whether poverty in the states motivates acquisition of household credit through programs such as Microcredit (developed by Bank of Northeast) shows the relevance of this type of initiative, as does the rapidly growing market share held by such types of credit and payroll loans among households.

Theoretically, we know that a definite answer must be the consequence of a microfundamented model, which has yet to be written or tested – and we have certainly not found it. However, our evidence can work as a starting point for this consumption-based model because the omission of relevant variables can be a source of standard models' inability to address credit patterns in Brazil. We recognize that our approach requires further generality, in the sense that this exercise could be performed on a longer spanning time-series, based on more disaggregated data, as cross-city panel data, or taking into account cross-effects due to contagion in the region, as evidenced by Viale et al. (2014) for Mercosur. We also recognize that further progress should require use of primary source data based on questionnaires. Another possible extension would be to examine household credit by type of credit.

BIBLIOGRAPHY

Alesina, A. & Perotti, R. (1996). Income distribution, political instability, and investment. *European Economic Review*, 40(6):1203–1228.

Barro, R. J. & Lee, J. W. (2013). A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics*, 104:184–198.

Beck, T., Demirguc-Kunt, A., & Levine, R. (2006). Finance, inequality, and poverty: cross-country evidence. World Bank Policy Research Working Paper. World Bank, Washington D.C.

Becker, G. S. (1975). Human capital. *Chicago Press*.

Bernanke, B. S. & Blinder, A. S. (1992). The federal funds rate and the channels of monetary transmission. *American Economic Review*, 82:901–921.

Bernanke, B. S. & Gertler, M. (1995). Inside the black box: the credit channel of monetary policy transmission. Technical report, National Bureau of Economic Research.

Brock, P. L. & Suarez, L. R. (2000). Understanding the behavior of bank spreads in Latin America. *Journal of Development Economics*, 63(1):113–134.

Cottarelli, C., Dell'Ariccia, G., & Vladkova-Hollar, I. (2005). Early birds, late risers, and sleeping beauties: Bank credit growth to the private sector in Central and Eastern Europe and in the Balkans. *Journal of Banking & Finance*, 29(1):83–104.

Coudert, V. & Pouvelle, C. (2010). Assessing the sustainability of credit growth: The case of Central and Eastern European countries. *The European Journal of Comparative Economics*, 7(1):87.

Doctor, M. (2015). Assessing the changing roles of the Brazilian Development Bank. *Bulletin of Latin American Research*, 34(2):197–213.

Edwards, S., Esquivel, G., & Márquez, G. (2007). Introduction to "The Decline of Latin American Economies: Growth, Institutions, and Crises". In *The Decline of Latin American Economies: Growth, Institutions, and Crises*, pages 1–14. University of Chicago Press.

Hansen, N.-J. H. & Sulla, M. O. (2013). *Credit Growth in Latin America: Financial Development or Credit Boom?* Number 13–106 in Working Paper. International Monetary Fund.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the Econometric Society*, 46:1251–1271.

Hira, A. & Gaillard, N. (2011). The bottom line: The fundamental weaknesses of Latin American finance. *Bulletin of Latin American Research*, 30:163–182.

Holmstrom, B. & Tirole, J. (1997). Financial intermediation, loanable funds, and the real sector. *Quarterly Journal of Economics*, 112(3):663–691.

Kiss, G., Nagy, M., & Vonnák, B. (2006). Credit growth in Central and Eastern Europe: convergence or boom? Technical report, Magyar Nemzeti Bank Working Paper. (Budapest: Central Bank of Hungary).

Levin, A., Lin, C.-F., & Chu, C.-S. J. (2002). Unit root tests in panel data: asymptotic and finite-sample properties. *Journal of Econometrics*, 108(1):1–24.

Matos, P., Sampaio, R., Moura, L., et al. (2015). On the key drivers of Brazilian household loan delinquency. *International Economics Letters*, 4(2):80–90.

Matos, P., Vasconcelos, J., & Penna, C. (2013). Política creditícia no Brasil: o sertão vai virar mar? *Economia (Anpec)*, 14:1–129.

Matos, P. R. F. (2017). On the Latin American Credit Drivers. *Emerging Markets Finance and Trade*, 53(2):306–320.

Nguyen, H.-L. Q. (2014). Do bank branches still matter? The effect of closings on local economic outcomes. *Massachusetts Institute of Technology, Working Paper*.

Phillips, P. C. & Sul, D. (2007). Transition modeling and econometric convergence tests. *Econometrica*, 75(6):1771–1855.

Rajan, R. G. & Zingales, L. (1998). Financial Dependence and Growth. *American Economic Review*, 88:559–586.

Ravallion, M. (2001). Growth, inequality and poverty: looking beyond averages. *World Bank Research Working Paper*. World Bank, Washington, D.C.

Schumpeter, J. A. (1934). The theory of economic development. *Harvard University Press*. Cambridge.

Stallings, B. (2006). *Finance for development: Latin America in comparative perspective*. Brookings Institution Press.

Viale, A. M., Bessler, D. A., & Kolari, J. W. (2014). On the structure of financial contagion: Econometric tests and Mercosur evidence. *Journal of Applied Economics*, 17(2):373–400.

Wooldridge, J. M. (2002). *Econometric analysis of cross section and panel data*. MIT press.

r b e

ISSN 0034-7140 (print) | ISSN 1806-9134 (online)
DOI 10.5935/0034-7140.20170017
Revista Brasileira de Economia, 71(3), 361–377

Casamentos Seletivos e Desigualdade de Renda no Brasil

Luciene Pereira*
Cezar Santos†

Sumário: **1. Introdução; 2. Crescimento dos casamentos seletivos; 3. Casamentos seletivos e desigualdade de renda; 4. Conclusão; A. Dados; B. Curva de Lorenz e Índice de Gini; C. Algoritmo Sinkhorn-Knopp (1967).**

Palavras-chave: **Desigualdade de renda; Casamentos; Pareamento.**

Códigos JEL: **D13, D31, J12, O11.**

O objetivo deste artigo é investigar a evolução do mercado de casamento e o seu impacto na distribuição de renda no Brasil. A partir da análise de dados dos censos demográficos de 1970 a 2010, encontramos evidências de que as pessoas estão cada vez mais se casando com parceiros de características semelhantes, ao longo dos anos. O aumento do número de casamentos seletivos, a princípio, não é capaz de afetar negativamente a desigualdade de renda. No entanto, quando realizamos alguns exercícios contrafactuais, é possível perceber que a melhora na distribuição de renda, ocorrida nesse período, poderia ter sido ainda maior caso essa tendência no mercado de casamento não tivesse ocorrido.

This paper aims to study the evolution of marriage markets and its impact on income inequality in Brazil. Using data from Brazilian censuses from 1970 to 2010, we find evidence that the degree of assortative mating has increased in Brazil. This higher degree of sorting does not seem at first to influence income inequality, since this has decreased over time. However, by computing counterfactual analyses, we find that income inequality would have decreased even further were it not for the higher degree of assortative mating in the marriage market.

1. INTRODUÇÃO

Decisões individuais (tais como casamento, divórcio, fecundidade e educação) afetam diretamente a dinâmica demográfica de um país e estão relacionadas a variáveis macroeconômicas (como renda *per capita* e desigualdade social). O interesse em entender como mudanças demográficas e desigualdade de renda se relacionam tem sido crescente na literatura. Em especial para o Brasil, Barros et alii (2000) apresenta evidências do impacto de fatores demográficos (o tamanho e a composição etária da população) sobre a distribuição de renda entre os anos de 1976 e 1996. Ele chega à conclusão, a partir de microssimulações com dados da PNAD, que o efeito da mudança demográfica nesse período sobre a redução da pobreza corresponde a aproximadamente 15% do efeito que o crescimento da renda teve sobre tal redução.[1] Estudar a relação entre mudanças demográficas e distribuição de renda se mostra particularmente interessante para o Brasil, pois o país possui elevados índices de desigualdade. Segundo dados do Banco Mundial, cerca de 90% das economias apresentam um grau de desigualdade de renda inferior à brasileira. Além disso, o país apresenta um fato notório, o seu grau de desigualdade

*Bahia Asset Management.E-mail: lucienetmp@gmail.com

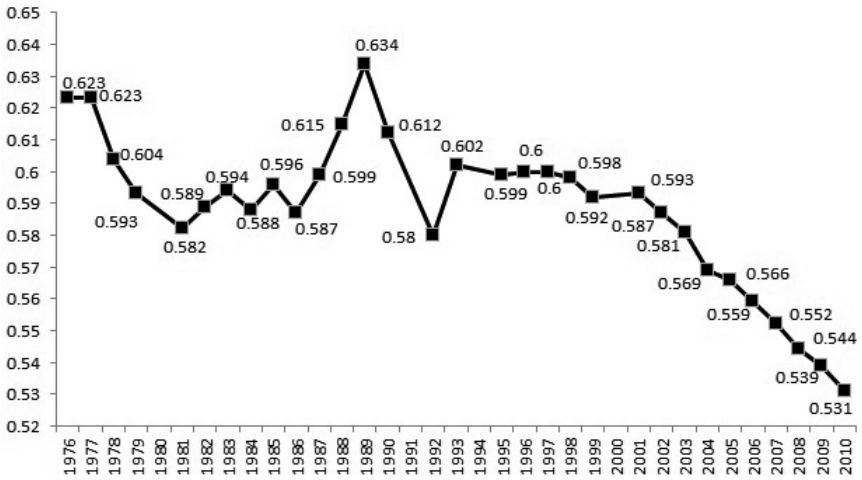
†FGV/EPGE - Escola Brasileira de Economia e Finanças.E-mail: cezar.santos@fgv.br

[1]Entre os anos de 1976 a 1996, a renda per capita cresceu em média 3% ao ano.

Recebido em 19 de abril de 2016
Aceitação Final em 07 de julho de 2017

vem apresentando uma queda contínua e acentuada desde 2001, como podemos observar na Figura 1 abaixo.[2]

Figura 1: Evolução do Índice de Gini, de 1976 a 2010



Nota: Dados do Banco Mundial.

Neste artigo, estudamos a evolução do mercado de casamento no Brasil nos últimos quarenta anos e como isso tem afetado a distribuição de renda da economia. A partir da análise de dados dos censos demográficos de 1970 a 2010, verificamos que as pessoas estão cada vez mais propensas a se casar com outras que possuem características socioeconômicas semelhantes (aqui representadas por 4 níveis educacionais distintos). Esse fenômeno social será chamado neste artigo de crescimento do casamento seletivo (ou também de queda na diversidade conjugal).

O crescente fenômeno de casamentos entre classes socioeconômicas semelhantes já vem sendo largamente documentado em países desenvolvidos por sociólogos e economistas.[3] Para o Brasil, Neri (2005) parece ser o único a se debruçar sobre essa questão. O estudo analisa de forma geral as relações entre casamento e economia, distinguindo os padrões de mulheres e homens. E, mais especificamente, mensura o quanto das relações conjugais acontecem entre pessoas do mesmo nível educacional e como isso evoluiu entre 1970 e 2000. E a partir da análise do percentual de casamentos ocorridos entre os mesmos grupos educacionais (56,7% em 1970 e 49,6% em 2000), conclui que houve um aumento na diversidade conjugal (ou queda no casamento seletivo). O nosso artigo apresenta conclusões opostas a Neri (2005) e reafirma, para o Brasil, a tendência apontada na literatura internacional. Isso ocorre pois Neri (2005) não leva em consideração o fato de que as distribuições marginais de homens e mulheres entre os diferentes níveis educacionais mudaram ao longo dos anos. Quando esse fato é incorporado às análises, somos capazes de concluir que houve um aumento nos casamentos seletivos no Brasil.

Além de estudar essa tendência, também analisamos como ela é capaz de afetar a distribuição de renda do país. A literatura mostra que, para países desenvolvidos, o crescente número de casamentos seletivos tem contribuído para o aumento da desigualdade de renda entre as famílias. No entanto, para o Brasil verificamos em um primeiro momento o oposto, isto é, a desigualdade de renda melhorou entre 1970 e 2010, apesar de os brasileiros estarem cada vez mais se casando com parceiros de características semelhantes. Contudo, a partir de exercícios contrafactuais, mostramos que a distribuição de renda

[2]Barros et alii (2006) analisa as causas e consequências dessa queda entre 2001 e 2004.

[3]Ver Schwartz e Mare (2005), Blossfeld e Timm (2003), Greenwood et alii (2014), Greenwood et alii (2016), Eika et alii (2014).

poderia ter melhorado ainda mais caso essa tendência no mercado de casamentos não estivesse presente. No primeiro experimento, medimos o efeito sobre o Índice de Gini caso os casamentos tivessem ocorrido de forma aleatória e observamos que, tanto para 1970 quanto para 2010, há uma melhora nesse índice. No segundo, trocamos o percentual de solteiros entre os anos de 1970 e 2010 e mais uma vez verificamos uma melhora na desigualdade de renda, embora o efeito se mostre pequeno. Esses exercícios contrafactuais só fazem sentido e produzem algum efeito na distribuição de renda caso a participação de mulheres no mercado de trabalho seja positiva e crescente, o que se mostra verdadeiro quando analisamos os dados. A partir disso, realizamos o exercício de assumir que todos os casamentos ocorreram de forma aleatória em 1970 e 2010 mas trocando a participação das mulheres casadas no mercado de trabalho entre os anos de 1970 e 2010. E comprovamos que a inserção de mulheres casadas afeta a distribuição de renda.

Investigar o impacto de um maior número de casamentos seletivos sobre a distribuição de renda requer um certo cuidado. Durantes os últimos quarenta anos, observamos, pela Figura 3, que o nível educacional da população brasileira como um todo tem aumentado e que as mulheres a partir de 1990 passaram a ter em média mais anos de estudo que os homens. Portanto a distribuição marginal de homens e mulheres entre os diferentes níveis educacionais mudou nesse período. Assim, precisamos distinguir os efeitos sobre a distibuição de renda provocados pelas mudanças dos níveis de educação de homens e mulheres e pela maior tendência de casamentos seletivos. Para isso, no exercício contrafactual de trocar a distribuição de casamentos entre os anos de 1970 e 2010, utilizamos a tabela de contingência padronizada.[4] E mais uma vez observamos uma melhora na desigualdade de renda.

Deve-se notar também que há interações entre algumas das margens acima que não são exploradas no presente artigo. Por exemplo, há vários estudos que relacionam interações entre casamento e participação feminina no mercado de trabalho: ver, por exemplo, o artigo clássico de Lundberg (1985). Goldin (2006) relaciona mudanças na família, educação e oferta de trabalho feminina por um período de mais de 100 anos. Há ainda estudos sobre a relação entre os retornos da educação no mercado de casamento: ver, por exemplo, Chiappori et alii (2009) e Greenwood et alii (2016).

O artigo está organizado da seguinte maneira. Na próxima seção, analisamos a evolução da diversidade conjugal no mercado de casamentos brasileiro. A partir disso, verificamos como essa tendência pode afetar a distribuição de renda e realizamos alguns experimentos contrafactuais na seção 3. Por fim, concluímos na seção 4.

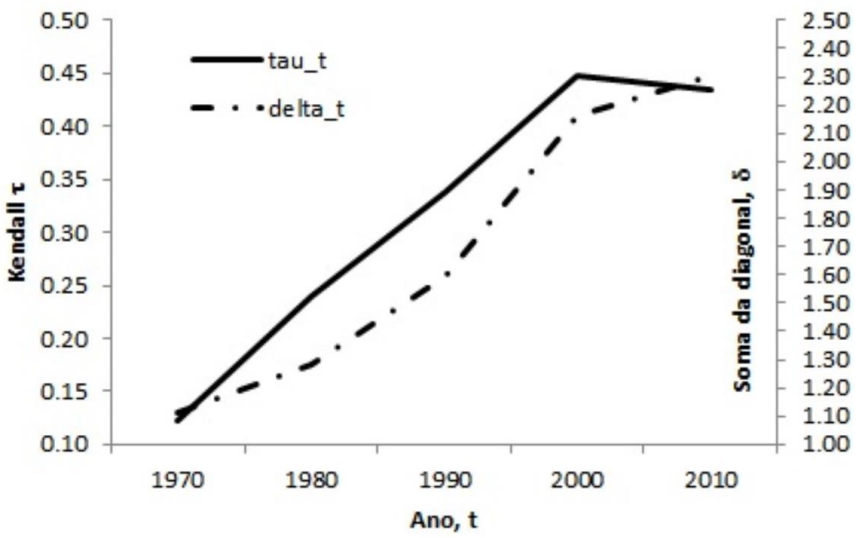
2. CRESCIMENTO DOS CASAMENTOS SELETIVOS

Para analisar o crescente fenômeno de casamentos entre pessoas de características semelhantes, dividimos as classes socioeconômicas brasileiras em quatro níveis educacionais: educação primária incompleta (PC-), primária (PC), secundária (SC) e universidade (C). Primeiramente, estimamos o coeficiente de correlação de Kendall τ entre os níveis educacionais do marido e da esposa para cada ano da nossa amostra (1970, 1980, 1991, 2000 e 2010). A evolução desse coeficiente ao longo do tempo pode ser observada na Figura 2 a seguir. Note que, apesar da série não ser monotonicamente crescente, o coeficiente τ é claramente maior em 2010 (0.43516) do que em 1970 (0.12160).

Em segundo, computamos a tabela de contingência dos casamentos (os anos de 1970 e 2010 podem ser vistos na Tabela 1). Cada célula representa o percentual de casamentos com aquele específico pareamento educacional entre marido e esposa. Vale mencionar que a moda entre todas as combinações educacionais de marido e esposa em 1970 era entre pessoas de baixa escolaridade (ensino primário incompleto) e passou para casais com ensino secundário em 2010, ou seja, houve uma melhora dos níveis

[4] Método proposto por Mosteller (1968) e descrito com maiores detalhes na seção 3.

Figura 2: Crescimento de casamentos seletivos



Nota: As variáveis τ_t and δ_t são medidas da tendência de casamentos seletivos para os anos $t = 1970,..,2010$. Quanto maior o valor de uma dessas variáveis, maior é o grau de pareamento seletivo entre os casais. Veja o texto para maiores detalhes e o Apêndice para a descrição dos dados usados.

educacionais na população brasileira (que pode ser facilmente verificada ao observarmos a evolução do número médio de anos de estudo para mulheres e homens na Figura 3 abaixo).[5]

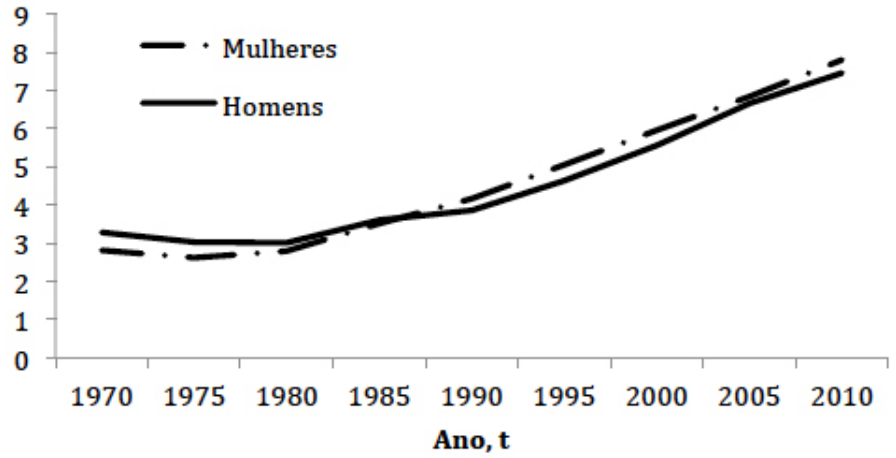
Computamos também as tabelas de contigência quando a hipótese de casamento aleatório é assumida (em itálico na Tabela 1). A partir dessas duas categorias de tabela de contigência, podemos apresentar outra estatística que indica a queda na diversidade conjugal ao longo dos anos. Essa estatística é a razão entre a soma dos traços (diagonal principal) das tabelas de contigência real e aleatória. Note que os elementos da diagonal principal representam os casamentos entre pessoas de mesmo nível educacional, assim quanto mais alto forem esses números, maior é a fração na população de casamentos entre indivíduos com características semelhantes. Como δ é a razão da soma da diagonal principal das tabelas de contingência real e aleatória, quanto mais alto for essa razão, maior é o grau de casamento seletivo. Observando ainda a Figura 2, podemos notar que essa estatística é sempre maior do que um (o que indica que o pareamento entre marido e mulher de mesmo nível educacional é sempre superior ao caso aleatório) e que ela cresceu mais do que 100% entre 1970 e 2010. Portanto, o grau de casamento seletivo cresceu de forma considerável na sociedade brasileira.

3. CASAMENTOS SELETIVOS E DESIGUALDADE DE RENDA

A literatura tem apontado que um crescimento na tendência de casamentos seletivos vem se traduzindo em uma maior desigualdade de renda para países desenvolvidos (em especial os Estados Unidos). O interesse nessa relação causal pode ser observado tanto em termos teóricos quanto em termos pu-

[5]Repare que as mulheres apresentaram um crescimento mais acelerado do que os homens de acumulação de anos de estudo até os anos 1990. A partir desse ano, ambos apresentaram tendência semelhante, embora as mulheres com um maior nível de educação.

Figura 3: Evolução dos anos médios de estudo



Nota: Dados de Barro e Lee (2010), população brasileira de 25 ou mais anos de idade.

ramente empíricos.[6] Seguindo a metodologia de Greenwood et alii (2014) de um arcabouço empírico que independe de um modelo teórico, estudaremos essa relação para o Brasil. Em um primeiro momento, quando apenas olharmos para a melhora do Índice de Gini entre 1970 e 2010 e o crescimento dos casamentos seletivos nesse mesmo período, chegaremos à conclusão de que o Brasil se diferencia dos demais desenvolvidos. No entanto, quando realizamos exercícios contrafactuais, podemos perceber que a melhora na desigualdade de renda brasileira poderia ter sido ainda maior caso o aumento na tendência de casamentos seletivos não tivesse ocorrido.

A primeira análise consiste em olhar para as estatísticas de renda de pessoas casadas nos diferentes níveis educacionais. Na Tabela 1, cada célula nos fornece a renda daquele determinado tipo de casal em relação à renda média da população inteira, casados e solteiros. Quando analisamos de forma geral todas as células, podemos perceber que, com exceção dos casamentos entre pessoas com escolaridade primária incompleta, em todos os tipos de pareamento, a renda do casal se aproximou à renda média da população entre 1970 e 2010, em outras palavras, a renda da sociedade como um todo se tornou mais homogênea. Além disso, quando olhamos para a diagonal principal das matrizes (ou seja, para os casamentos em que os parceiros têm o mesmo nível educacional), as rendas dos casais de níveis educacionais secundário e universidade foram as que mais se aproximaram da renda média.[7] E vale lembrar que o casamento entre pessoas de nível secundário se tornou o mais comum na sociedade em 2010. Assim, o tipo de casamento de maior frequência foi um dos que mais teve a sua renda se aproximando à renda média da economia. Portanto, apenas por essa análise inicial, temos indícios de que a distribuição da renda brasileira tenha melhorado no período em questão. Mas para confirmar, precisamos computar a Curva de Lorenz e o Índice de Gini.

Para isso, defina $f_{i,j}$ como a fração das famílias que são do tipo i e estão no percentil j de renda e $r_{i,j}$ como a renda desse tipo de família em relação à renda média. O índice percentual j é expresso em termos de fração (ou seja, 0.10 no lugar de 10). Os tipos são classificados da seguinte maneira: casados ou solteiros. No caso dos casados, cada membro da família é identificado também pelo nível

[6] Para análises com modelos teóricos, ver Fernàndez e Rogerson (2001) e Greenwood et alii (2016). Para trabalhos empíricos, ver Lam (1997), Cancian e Reed (1998, 1999), Schwartz (2010) e Greenwood et alii (2014).

[7] Deve-se notar que o retorno à educação no Brasil deve ter um papel importante também. Para este assunto, ver por exemplo Barbosa Filho e Pessôa (2008).

Tabela 1: Tabela de Contingência: Casamento Seletivo e Renda por Nível Educacional

Casamento Seletivo por Nível Educacional								
1970								
Marido	Esposa							
	PC-		PC		SC		C	
PC-	**0.850**[1]	**0.790**[2]	0.018	*0.046*	0.009	*0.035*	0.000	*0.006*
PC	0.036	*0.058*	**0.018**	**0.003**	0.009	*0.003*	0.001	*0.000*
SC	0.011	*0.032*	0.011	*0.002*	**0.012**	**0.001**	0.001	*0.000*
C	0.003	*0.022*	0.007	*0.001*	0.011	*0.001*	**0.004**	**0.000**
Marginal	0.900		0.054		0.041		0.006	
2010								
PC-	**0.166**	**0.068**	0.078	*0.081*	0.039	*0.095*	0.007	*0.047*
PC	0.049	*0.068*	**0.148**	**0.081**	0.077	*0.095*	0.017	*0.047*
SC	0.016	*0.069*	0.047	*0.082*	**0.180**	**0.097**	0.052	*0.048*
C	0.001	*0.029*	0.006	*0.035*	0.031	*0.041*	**0.086**	**0.020**
Marginal	0.232		0.279		0.327		0.162	
Renda Familiar por Nível Educacional								
1970								
PC-	0.565[3]	*0.029*[4]	1.583	*0.118*	2.424	*0.326*	4.172	*0.425*
PC	1.81	*0.037*	2.785	*0.075*	3.469	*0.228*	5.653	*0.349*
SC	2.852	*0.033*	3.928	*0.053*	4.580	*0.174*	6.158	*0.301*
C	5.160	*0.025*	6.324	*0.037*	7.042	*0.091*	8.690	*0.177*
2010								
PC-	0.357	*0.348*	0.428	*0.329*	0.610	*0.364*	1.062	*0.536*
PC	0.459	*0.296*	0.603	*0.309*	0.729	*0.334*	1.408	*0.481*
SC	0.629	*0.262*	0.724	*0.256*	1.034	*0.318*	1.976	*0.457*
C	1.420	*0.188*	1.684	*0.203*	2.120	*0.227*	4.102	*0.362*

Note: Cada célula na tabela de contingência possui duas entradas. No painel superior, elas se referem: 1) ao pareamento observado entre marido e esposa e 2) ao que aconteceria se os casamentos fossem aleatórios. No painel inferior, elas denotam: 3) a renda da família relativa à renda média total e 4) a fração da renda feminina na renda total da família. A linha Marginal nos mostra a fração de mulheres em cada nível educacional, ou seja, a distribuição marginal de educação para as mulheres.

educacional. A esposa recebe uma classificação de acordo com a participação no mercado de trabalho (trabalha ou não). Os solteiros são apenas identificados pelo seu nível educacional. E ainda há dez percentis (decis) de renda, de maneira que $j \in \{0.1,0.2,...,1.0\}$. São gerados, assim, 360 (i,j) possíveis combinações de família para cada ano. A fração da renda agregada que cabe ao percentil j é dada por $s_j = \sum_i f_{i,j} r_{i,j}$. A fração de renda acumulada no percentil p é $l_p \equiv \sum_j^p s_j = \sum_j^p \sum_i f_{i,j} r_{i,j}$. A Curva de Lorenz pode ser representada pelo gráfico de l_p contra $p = \sum_j^p \sum_i f_{i,j}$. E o Índice de Gini, g, é duas vezes a área entre a Curva de Lorenz e a reta de 45°.[8] Se p se mover continuamente, então o coeficiente de Gini passa a ser definido por $g = 2\int_0^1 |l_p - p|\, dp$, onde $0 \leq g \leq 1$. Quanto maior for g, maior será

[8]Veja o Apêndice para maiores detalhes de como essas curvas são construídas.

a desigualdade de renda. A Curva de Lorenz e o Índice de Gini são funções de $f_{i,j}$ e $r_{i,j}$, para todo i e j. Assim, podemos escrever $l_p = \text{LORENZ}_p(\{f'_{ij}\},\{r_{ij}\})$ e $g = \text{GINI}(\{f'_{ij}\},\{r_{ij}\})$.

As Curvas de Lorenz de 1970 e 2010 nos mostram uma queda na desigualdade de renda, com o Índice de Gini indo de 0.609 para 0.572.[9] A Tabela 2 nos mostra a renda relativa de cada percentil. Note que, entre 1970 e 2010, os percentis mais baixos (0.1 e 0.2) apresentaram uma perda acentuada em sua renda relativa, em especial o percentil 0.1 com uma queda de 82% (indo de 0.03258 para 0.00572). No entanto, todos os percentis entre 0.3 e 0.7 viram as suas rendas relativas crescerem, com destaque para os percentis 0.4 e 0.5 que tiveram um crescimento de 13% e 10%, respectivamente. Já os percentis 0.8 e 0.9 apresentaram uma pequena queda, na ordem de 4% e 2% respectivamente. E por fim, o percentil 1 verificou um aumento na renda relativa de apenas 1,7%. A partir dessa análise, podemos perceber que, em linhas gerais, a renda relativa de praticamente todos os percentis (com exceção para os dois mais baixos) tem se aproximado da renda média, o que corrobora a melhora da distribuição de renda brasileira. Quando comparamos essa distribuição de renda brasileira com a norte-americana entre 1960 e 2005, podemos perceber o quanto a evolução foi diferente entre essas duas economias.[10] Nos Estados Unidos, houve um crescimento da renda relativa dos percentis mais altos (0.8, 0.9 e 1), enquanto os mais baixos (0.1 até 0.7) viram a sua renda cair. E essa tendência, naturalmente, se traduziu em uma piora da desigualdade (e aumento do Índice de Gini).

Tabela 2: Desigualdade de Renda

	Renda familiar relativa à média	
Percentil de renda	1970	2010
0.1	0.03258	0.00572
0.2	0.14659	0.11823
0.3	0.22434	0.23933
0.4	0.29699	0.33641
0.5	0.38578	0.42601
0.6	0.50822	0.54831
0.7	0.70218	0.71148
0.8	1.00548	0.98516
0.9	1.66101	1.59004
1	5.14919	5.23769

Embora para o Brasil não observemos uma piora na desigualdade de renda, apesar do aumento de casamentos seletivos, contrariando assim o que a literatura vem apontando nos países desenvolvidos, surge uma pergunta: será que a distribuição de renda brasileira teria melhorado ainda mais caso o crescimento na tendência de casamentos seletivos não tivesse ocorrido? Para respondê-la, realizamos o seguinte experimento: substituímos o pareamento observado na tabela de contingência presente na Tabela 1 pelo padrão da tabela de contingência aleatória (representada em itálico também na Tabela 1). Vejamos como isso é realizado.

Primeiro, defina $\mathcal{M}$ como o conjunto de índice para os casados e $\mathcal{S}$, o conjunto para os solteiros. O conjunto $\mathcal{M}$ é definido particularmente da seguinte forma. Considere uma distribuição qualquer $\{f_{ij}\}$ de famílias. Vale lembrar que as famílias compostas por casais são indexadas pelo nível educacional do marido e da esposa, a participação de cada um no mercado de trabalho e o número de filhos. Defina

[9] Ver Tabela 2.

[10] Ver Figure 2 em Greenwood et alii (2014).

$\mathcal{M}_{E_H}$ como os conjuntos que contêm os índices de todas as famílias de casais cujo marido possui nível educacional $E_H \in \{PC-,PC,SC,C\}$. De forma similar, os conjuntos $\mathcal{M}_{E_W}$ contêm todas as famílias de casais cuja esposa possui nível educacional E_W. Além disso, $\mathcal{M}_{LFP_H}$ ($\mathcal{M}_{LFP_W}$) contêm todas as famílias de casados cujo (a) marido (esposa) possui a participação no mercado de trabalho $LFP_{H(W)} \in \{TRABALHA_{H(W)}, \tilde{}TRABALHA_{H(W)}\}$. E por fim, $\mathcal{M}_{KIDS}$ contém todas as famílias de casados com um particular número de filhos $KIDS \in \{0,1,2,2+\}$. O conjunto de todas as famílias de casados com um particular pareamento de nível educacional $\mathcal{M}_{E_H,E_W}$ é dado por

$$\mathcal{M}_{E_H,E_W} = \mathcal{M}_{E_H} \cap \mathcal{M}_{E_W}.$$

Assim, o conjunto que contém todas as diferentes famílias é:

$$\mathcal{M} = \bigcup\nolimits_{E_H,E_W,LFP_H,LFP_W,KIDS} (\mathcal{M}_{E_H} \cap \mathcal{M}_{E_W} \cap \mathcal{M}_{LFP_H} \cap \mathcal{M}_{LFP_W} \cap \mathcal{M}_{KIDS}),$$

onde o termo dentro do parênteses representa todos os casados do tipo $(E_H,E_W,\ LFP_H,\ LFP_W,KIDS)$.

Com os conjuntos $\mathcal{M}$ e $\mathcal{S}$ devidamente definidos, agora mostraremos como de fato o experimento de casamento aleatório é realizado. Ele se resume a substituir os $\{f_{i,j}\}$'s observados para $(i,j) \in \mathcal{M}$ pelo conjunto obtido caso os casamentos tivessem ocorrido de forma aleatória, aqui representado por $\{\widetilde{f}_{ij}\}$ para $(i,j) \in \mathcal{M}$. O Índice de Gini é então dado por GINI($\{f'_{ij}\},\{r_{ij}\}$), onde $\{f'_{ij}\} \equiv \{\widetilde{f}_{ij}\}_{\mathcal{M}} \cup \{f_{ij}\}_{\mathcal{S}}$. Assim, considere o primeiro elemento da tabela de contingência de 1970 presente na Tabela 1, que representa os casamentos entre pessoas com nível primário de educação incompleto. Em 1970, essa fração era de 0.85 e pode ser representada da seguinte forma

$$\frac{\sum_{i \in \mathcal{M}_{PC-,PC-}} \sum_{j=0.1}^{1} f_{ij}^{1970}}{\sum_{i \in \mathcal{M}} \sum_{j=0.1}^{1} f_{ij}^{1970}} = 0.85.$$

Agora, impondo que os casamentos sejam aleatórios, a fração desse tipo de casamento passa a ser 0.79. Chame essa nova distribuição de $\widetilde{f}_{ij}^{1970}$ e de forma similar à expressão acima, temos

$$\frac{\sum_{i \in \mathcal{M}_{PC-,PC-}} \sum_{j=0.1}^{1} \widetilde{f}_{ij}^{1970}}{\sum_{i \in \mathcal{M}} \sum_{j=0.1}^{1} f_{ij}^{1970}} = 0.79.$$

Assim, a razão entre o número total de casamentos aleatórios entre pessoas com nível primário incompleto e o total de casamentos desse tipo observados nos dados é

$$\frac{\sum_{i \in \mathcal{M}_{HS-,HS-}} \sum_{j=0.1}^{1} \widetilde{f}_{ij}^{1970}}{\sum_{i \in \mathcal{M}_{HS-,HS-}} \sum_{j=0.1}^{1} f_{ij}^{1970}} = \frac{0.79}{0.85} = 0.93.$$

Portanto, sob a hipótese de casamentos aleatórios, o número de casamentos do tipo $(PC-,PC-)$ é reduzido por um fator de $0.93 = 0.79/0.85$. Assuma que essa redução possa ser generalizada para todos os percentis de renda, isto é, para todos os j's. Assim, quando realizarmos o exercício de casamentos aleatórios, a fração $\widetilde{f}_{ij}^{1970}$ deve ser construída da seguinte forma

$$\widetilde{f}_{ij}^{1970} = \frac{0.79}{0.85} f_{ij}^{1970}, \text{ para } i \in \mathcal{M}_{PC-,PC-} \text{ e todo } j.$$

Esse procedimento é realizado para todos os demais 15 tipos de casamentos. E para os indivíduos solteiros e divorciados, mantemos as frações originais (provenientes dos dados), ou seja, $\widetilde{f}_{ij}^{1970} = f_{ij}^{1970}$.

Na Tabela 3, podemos observar os resultados desse exercício contrafactual. Note que, tanto para 1970 quanto para 2010, há uma queda no Índice de Gini se comparamos com os dados (indo de 0.609 para 0.587 em 1970 e de 0.572 para 0.543 em 2010). A partir desses números, podemos concluir que o crescimento na tendência de casamentos seletivos impacta a distribuição de renda em uma economia e que a queda ligeiramente maior para o ano de 2010 mostra um nível de casamento seletivo mais elevado. Assim, a resposta à pergunta anterior fica evidente: a desigualdade de renda brasileira teria melhorado ainda mais caso a diversidade conjugal não tivesse caído ao longo dos anos.

Tabela 3: Índice de Gini, Dados e Experimentos

Base para o Índice de Gini	1970	2010
Dados	0.609	0.572
Pareamento Aleatório	0.587	0.543
Aleatório + 2010 PFFT	0.562	
Aleatório + 1970 PFFT		0.585
Tabela Padronizada	0.603	0.561
Tabela Padronizada + 2010 PFFT	0.591	
Tabela Padronizada + 1970 PFFT		0.600
Mais solteiras (nível 2010)	0.614	
Menos solteiras (nível 1970)		0.568
Mais mulheres trabalhando (nível 2010)	0.596	
Menos mulheres trabalhando (nível 1970)		0.603

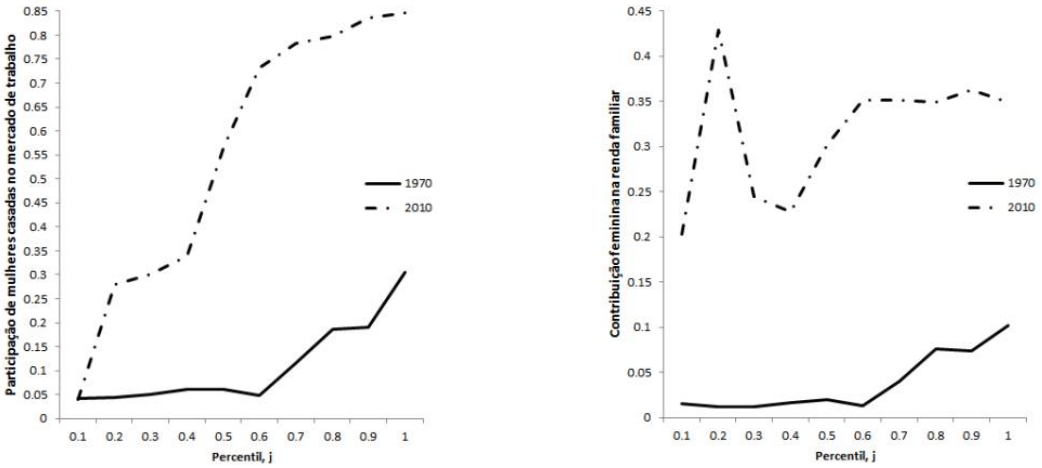
Outro experimento interessante a ser feito é verificar o que ocorreria com a distribuição de renda em 1970 e 2010, caso trocássemos o percentual de solteiros entre os anos, ou seja, caso assumíssemos que o nível de solteiros de 1970 fosse o mesmo de 2010 e o de 2010 fosse igual ao de 1970. Note pela Tabela 3 que o Índice de Gini subiria de 0.609 para 0.614 (diferença de 0.005) em 1970 e cairia de 0.572 para 0.568 (diferença de 0.004) em 2010. Esses resultados nos levam a conclusão de que um maior número de casamentos ajuda a melhorar a distribuição de renda brasileira, embora esse efeito não seja muito expressivo.

Os dois experimentos acima só causam algum impacto na distribuição de renda se pelo menos algumas mulheres casadas participarem do mercado de trabalho. Olhando para a Figura 4, que nos mostra a participação de mulheres casadas nesse mercado por percentil de renda, podemos perceber que essa participação é muito maior em 2010 do que em 1970, para todos os percentis acima de 0.1. Além disso, houve um crescimento bastante elevado dessa participação a partir do (incluindo) percentil 0.5, com destaque para o 0.6 (o percentual de mulheres casadas trabalhando cresceu de 5% para 73%). Ainda na Figura 4, também podemos observar a contribuição da renda da mulher na renda da família por percentil. Note que, novamente, esse tipo de participação é muito superior em 2010 se comparada a 1970. Além disso, para todos os percentis houve um crescimento da renda da mulher na renda total da família da ordem de 20 pontos percentuais (p.p.).[11] Assim, temos fortes indícios de que a oferta de mão de obra das mulheres casadas no mercado de trabalho impacta a distribuição de renda.

Para comprovar o ponto anterior, realizaremos o seguinte exercício contrafactual: primeiro, assuma que todos os casamentos ocorreram de forma aleatória em 1970 e 2010 e depois, que as mulheres casadas em 1970 participaram no mercado de trabalho com o mesmo nível de 2010 e que em 2010 elas trabalharam ao nível de 1970. O algoritmo para obter as tabelas de contingência assumindo casamentos aleatórios já foi descrito anteriormente. Suponha agora que, além disso, impomos que as mulheres

[11] O percentil 0.2 viu sua participação crescer quase 42 p.p., o que pode ser uma anomalia dos dados.

Figura 4: Mulheres casadas no mercado de trabalho, 1970 e 2010



Nota: O lado direito mostra a participação de mulheres casadas no mercado de trabalho por percentil de renda para 1970 e 2010. O lado esquerdo ilustra a fração da renda feminima na renda total da família.

casadas participem do mercado de trabalho em 1970 com os níveis de 2010.[12] A participação de mulheres casadas na força de trabalho em 1970, quando assumimos a hipótese de casamentos aleatórios, é

$$\frac{\sum_{i\in\mathcal{M}_{TRABALHA_W}}\sum_{j=0.1}^{1}\widetilde{f}_{ij}^{1970}}{\sum_{i\in\mathcal{M}}\sum_{j=0.1}^{1}\widetilde{f}_{ij}^{1970}}=0.13,$$

enquanto essa participação em 2010 é

$$\frac{\sum_{i\in\mathcal{M}_{TRABALHA_W}}\sum_{j=0.1}^{1}f_{ij}^{2010}}{\sum_{i\in\mathcal{M}}\sum_{j=0.1}^{1}f_{ij}^{2010}}=0.64.$$

Chamemos essa nova distribuição contrafactual de famílias de casados em 1970 de $\widehat{f}_{ij}^{1970}$, para $i\in\mathcal{M}$ e para todo j. Essa distribuição para 1970 deve nos dar a participação feminina no mercado de trabalho ao nível de 2010, ou seja,

$$\frac{\sum_{i\in\mathcal{M}_{TRABALHA_W}}\sum_{j=0.1}^{1}\widehat{f}_{ij}^{1970}}{\sum_{i\in\mathcal{M}}\sum_{j=0.1}^{1}\widehat{f}_{ij}^{1970}}=0.64.$$

É válido ter em mente que a fração de pessoas casadas em 1970 não muda nesses experimentos contrafactuais, isto é,

$$\sum_{i\in\mathcal{M}}\sum_{j=0.1}^{1}f_{ij}^{1970}=\sum_{i\in\mathcal{M}}\sum_{j=0.1}^{1}\widetilde{f}_{ij}^{1970}=\sum_{i\in\mathcal{M}}\sum_{j=0.1}^{1}\widehat{f}_{ij}^{1970}.$$

Portanto,

$$\frac{\sum_{i\in\mathcal{M}_{TRABALHA_W}}\sum_{j=0.1}^{1}\widehat{f}_{ij}^{1970}}{\sum_{i\in\mathcal{M}}\sum_{j=0.1}^{1}\widetilde{f}_{ij}^{1970}}=\frac{0.64}{0.13}\frac{\sum_{i\in\mathcal{M}_{TRABALHA_W}}\sum_{j=0.1}^{1}\widetilde{f}_{ij}^{1970}}{\sum_{i\in\mathcal{M}}\sum_{j=0.1}^{1}\widetilde{f}_{ij}^{1970}}=0.64.$$

[12] O procedimento é o mesmo para o caso em que as mulheres em 2010 trabalharam ao nível de 1970.

Impor a taxa de participação feminina no mercado de trabalho de 2010 em 1970 implica em escalonar para cima todas as combinações (i,j) de famílias de casados em que a esposa trabalha. Por outro lado, as famílias de casados em que a esposa não trabalha devem ser escalonadas para baixo de tal maneira que a fração total de famílias de casados não se altere.

Portanto, a distribuição contrafactual $\{\widehat{f}_{ij}^{1970}\}$ deve ser construída da seguinte maneira:

$$\widehat{f}_{ij}^{1970} = \frac{0.64}{0.13}\widetilde{f}_{ij}^{1970}, \text{ para } i \in \mathcal{M}_{TRABALHA_W} \text{ e para todo } j,$$

e

$$\widehat{f}_{ij}^{1970} = \frac{1-0.64}{1-0.13}\widetilde{f}_{ij}^{1970}, \text{ para } i \in \mathcal{M}_{\text{-}TRABALHA_W} \text{ e para todo } j.$$

Logo, a fração total de famílias de casados permanece constante,

$$\begin{aligned}
\sum_{i\in\mathcal{M}}\sum_{j=0.1}^{1}\widehat{f}_{ij}^{1970} &= \sum_{i\in\mathcal{M}_{TRABALHA_W}}\sum_{j=0.1}^{1}\widehat{f}_{ij}^{1970} + \sum_{i\in\mathcal{M}_{\text{-}TRABALHA_W}}\sum_{j=0.1}^{1}\widehat{f}_{ij}^{1970} \\
&= \frac{0.64}{0.13}\sum_{i\in\mathcal{M}_{TRABALHA_W}}\sum_{j=0.1}^{1}\widetilde{f}_{ij}^{1970} + \frac{1-0.64}{1-0.13}\sum_{i\in\mathcal{M}_{\text{-}TRABALHA_W}}\sum_{j=0.1}^{1}\widetilde{f}_{ij}^{1970} \\
&= 0.64\sum_{i\in\mathcal{M}}\sum_{j=0.1}^{1}\widetilde{f}_{ij}^{1970} + (1-0.64)\sum_{i\in\mathcal{M}}\sum_{j=0.1}^{1}\widetilde{f}_{ij}^{1970} \\
&= \sum_{i\in\mathcal{M}}\sum_{j=0.1}^{1}\widetilde{f}_{ij}^{1970}.
\end{aligned}$$

Assim como a distribuição contrafactual do experimento anterior, a distribuição para solteiros e divorciados permanece inalterada, $\widehat{f}_{ij}^{1970} = f_{ij}^{1970}$.

Nesse experimento, os Índices de Gini passam a ser 0.562 para 1970 e 0.585 para 2010. A hipótese de casamentos aleatórios faz com que as rendas fiquem mais equalizadas entre as famílias, pois acaba diversificando a renda entre maridos e esposas de diferentes níveis educacionais. Mas esse efeito só se torna real à medida em que as mulheres casadas participarem do mercado de trabalho. Assim, comparando o Índice de Gini do caso de casamentos aleatórios com o caso de casamentos aleatórios mais a troca entre os anos da participação das mulheres casadas no mercado de trabalho, observamos que essa estatística caiu de 0.587 para 0.562 em 1970 e cresceu de 0.543 para 0.585 em 2010. Portanto, o fato das mulheres casadas se inserirem mais ou menos no mercado afeta a desigualdade de renda brasileira.

Outro importante experimento a se pensar seria medir a distribuição de renda caso os casamentos em 2010 tivessem ocorrido da mesma maneira que em 1970, ou seja, caso a tabela de contingência de 2010 fosse substituída pela de 1970. No entanto, esse exercício contrafactual apresenta um grande problema. Em 2010, as pessoas (homens e principalmente mulheres) tinham níveis educacionais mais altos do que em 1970. Com isso, a distribuição marginal de maridos e esposas em cada nível educacional mudou bastante entre 1970 e 2010, inviabilizando em um primeiro momento esse novo experimento. Assim, caso realizássemos esse experimento sem qualquer correção, não seria possível distinguir os efeitos dos casamentos seletivos dos efeitos do aumento do nível educacional da população sobre a distribuição de renda.[13] Para contornar essa questão, podemos utilizar o procedimento proposto por Mosteller (1968) e construir a chamada tabela de contingência padronizada para 1970, usando as distribuições marginais de níveis educacionais para homens e mulheres de 2010, e para 2010 usando as distribuições de 1970.

Mosteller (1968) sugere que, quando comparamos duas tabelas de contingência, elas devem primeiro ser padronizadas de maneira que ambas possuam as mesmas distribuições marginais associadas às colunas e às linhas. Considere uma tabela 4×4. Ela pode ser padronizada tal que cada elemento das duas

[13] Isto é, a mudança das distribuições marginais de homens e mulheres entre os diferentes níveis educacionais.

distribuições marginais seja $1/4$. Isso pode ser feito usando o algoritmo de Sinkhorn e Knopp (1967), que escalona cada linha e cada coluna de forma iterativa.[14] A padronização preserva o padrão central de associação em uma tabela de contingência. Por exemplo, Tan et alii (2004) notam que tal padronização não altera as razões de possibilidade (medida típica usada para avaliar o padrão de associação entre variáveis) em uma tabela de contingência.

As tabelas padronizadas para 1970 e 2010 estão ilustradas na Tabela 4. A soma dos elementos da diagonal principal em 2010 é maior do que em 1970, o que nos leva a concluir que houve um crescimento na tendência de casamentos seletivos.

Tabela 4: Tabela de Contingência Padronizada: Casamento Seletivo por Nível Educacional

Casamento Seletivo por Nível Educacional				
1970				
Distribuição Marginal $= (1/4, \cdots, 1/4)$				
Marido	Esposa			
	PC-	PC	SC	C
PC-	**0.201**	0.028	0.017	0.004
PC	0.034	**0.114**	0.067	0.035
SC	0.012	0.074	**0.101**	0.063
C	0.003	0.034	0.065	**0.148**
Marginal, Esposas	$1/4$	$1/4$	$1/4$	$1/4$
2010				
Distribuição Marginal $= (1/4, \cdots, 1/4)$				
PC-	**0.166**	0.058	0.021	0.005
PC	0.058	**0.130**	0.048	0.014
SC	0.022	0.048	**0.130**	0.049
C	0.004	0.013	0.050	**0.182**
Marginal, Esposas	$1/4$	$1/4$	$1/4$	$1/4$

Nota: O painel superior mostra a tabela de contingência para 1970 quando ela é normalizada usando o algoritmo de Sinkhorn-Knopp tal que a distribuição marginal de homens e mulheres pelos níveis educacionais seja igual $1/4$. O painel inferior mostra o mesmo para 2010.

Note que cada distribuição marginal das tabelas padronizadas não precisa necessariamente ser $1/4$. A padronização da tabela de 1970 pode ser tal que cada distribuição marginal coincida com aquelas provenientes dos dados de 2010, ou vice-versa. Esses resultados são mostrados na Tabela 5.

A partir dessas tabelas de contingência padronizadas, podemos computar a desigualdade de renda para o experimento que estávamos pensando anteriormente. Os Índices de Gini associados são 0.603 para 1970 e 0.561 para 2010. Assim, caso as pessoas em 2010 se casassem como em 1970 (seguindo a tabela de contingência padronizada), o coeficiente de Gini teria caído de 0.572 para 0.561, ou seja, teríamos uma melhora na distribuição de renda brasileira.

Além do experimento com essas tabelas de contingência padronizadas, também podemos realizar outro exercício adicionando a hipótese de troca da participação das mulheres casadas no mercado de trabalho entre os anos. Assim, assumindo para 1970 a distribuição proveniente da tabela de contingência padronizada e a participação das mulheres casadas ao nível de 2010, o Índice de Gini vai de 0.609

[14] Ver Apêndice para maiores detalhes sobre esse algoritmo.

Tabela 5: Tabela de Contingência Padronizada: Casamento Seletivo por Nível Educacional

Casamento Seletivo por Nível Educacional				
1970				
Usando as Distribuições Marginais de 2010				
Marido	Esposa			
	PC-	PC	SC	C
PC-	**0.201**	0.047	0.036	0.005
PC	0.024	**0.131**	0.101	0.034
SC	0.008	0.082	**0.146**	0.059
C	0.001	0.017	0.043	**0.063**
Marginal, Esposas	0.234	0.277	0.326	0.161
2010				
Usando as Distribuições Marginais de 1970				
PC-	**0.830**	0.034	0.013	0.000
PC	0.047	**0.012**	0.005	0.000
SC	0.018	0.004	**0.012**	0.001
C	0.007	0.003	0.010	**0.005**
Marginal, Esposas	0.902	0.053	0.040	0.006

Nota: O painel superior mostra a tabela de contingência para 1970 quando ela foi normalizada usando o algoritmo de Sinkhorn-Knopp tal que a distribuição marginal de homens e mulheres pelos níveis educacionais seja igual aos dados de 2010. O painel inferior mostra a tabela de contingência para 2010 quando ela foi normalizada tal que a distribuição marginal de homens e mulheres pelos níveis educacionais seja igual aos dados de 1970.

(dados) para 0.591. Fazendo o mesmo para 2010 e diminuindo a participação das mulheres casadas ao nível de 1970, essa estatística cresce de 0.572 (dados) para 0.6, aumentando assim a desigualdade renda. Mais uma vez mostramos que a participação das mulheres no mercado de trabalho é crucial para entender a distribuição de renda.

4. CONCLUSÃO

Neste artigo, a partir de análises dos censos demográficos brasileiros e de exercícios puramente contábeis, mostramos que ao longo do tempo os brasileiros estão se casando cada vez mais com parceiros que possuem características socioeconômicas semelhantes. Esse fenômeno já vem sendo observado para alguns países desenvolvidos e está largamente documentado por sociólogos e economistas. Além disso, a literatura tem mostrado que a tendência crescente de casamentos seletivos impacta de forma negativa a desigualdade de renda. Quando nos debruçamos sobre essa questão para o Brasil, na verdade observamos, em um primeiro momento, o contrário: a desigualdade de renda melhorou de 1970 para 2010, embora a frequência de casamentos seletivos tenha aumentado. No entanto, quando realizamos alguns exercícios contrafactuais, podemos observar que a distribuição de renda brasileira teria melhorado ainda mais, caso a tendência de casamentos seletivos não tivesse ocorrido.

BIBLIOGRAFIA

Barbosa Filho, F. d. H. & Pessôa, S. (2008). Retorno da Educação no Brasil. *Pesquisa e Planejamento Econômico*, 38(1):97–126.

Barro, R. & Lee, J.-W. (2010). A New Data Set of Educational Attainment in the World, 1950-2010. *Journal of Development Economics*, 104:184–198.

Barros, R. P. d., de Carvalho, M., Franco, S., & Mendonça, R. (2006). Consequências e Causas Imediatas da Queda Recente da Desigualdade de Renda Brasileira. Technical report, Discussion Papers 1201, IPEA.

Barros, R. P. d., Firpo, S., Guedes, R., & Leite, P. (2000). Demographic Changes and Poverty in Brazil. Technical report, Discussion Papers 0096, IPEA.

Blossfeld, H.-P. & Timm, A. (2003). *Who Marries Whom? Educational Systems as Marriage Markets in Modern Societies*. Kluwer Academiic, Dordrecht.

Cancian, M. & Reed, D. (1998). Assessing the Effects of Wives'Earnings on Family Income Inequality. *The Review of Economics and Statistics*, 80(1):73–79.

Cancian, M. & Reed, D. (1999). The Impact of Wives' Earnings on Income Inequality: Issues and Estimates. *Demography*, 36(2):173–84.

Chiappori, P.-A., Iyigun, M., & Weiss, Y. (2009). Investment in Schooling and the Marriage Market. *American Economic Review*, 99(5):1689–1713.

Eika, L., Mogstad, M., & Zafar, B. (2014). Educational Assortative Mating and Household Income Inequality. NBER Working Paper 20271.

Fernàndez, R. & Rogerson, R. (2001). Sorting and Long-Run Income Inequality. *Quarterly Journal of Economics*, 116(4):1305–41.

Goldin, C. (2006). The Quiet Revolution that Transformed Women's Employment, Education, and Family. NBER Working Paper 11953.

Greenwood, J., Guner, N., Kocharkov, G., & Santos, C. (2014). Marry your Like: Assortative Mating and Income Inequality. *American Economic Review Papers & Proceedings*, 104(5):348–353.

Greenwood, J., Guner, N., Kocharkov, G., & Santos, C. (2016). Technology and the Changing Family: A Unified Model of Marriage, Divorce, Educational Attainment and Married Female Labor-Force Participation. *American Economic Journal: Macroeconomics*, 8(1):1–41.

Lam, D. (1997). Demographic Variables and Income Inequality. In Rosenzweig, M. R. & Stark, O., editors, *Handbook of Population and Family Economics*, pages 1015–1059. Elsevier North Holland.

Lundberg, S. (1985). The Added Worker Effect. *Journal of Labor Economics*, 3(1):11–37.

Mosteller, F. (1968). Association and Estimation in Contingency Tables. *Journal of the American Statistical Association*, 63(321):1–28.

Neri, M. (2005). *Sexo, Casamento e Economia*. FGV/IBRE, CPS, Rio de Janeiro.

Rao, V. M. (1969). Two Decompositions of Concentration Ratio. *Journal of the Royal Statistical Society, Series A (General)*, 132(3):418–425.

Schwartz, C. R. (2010). Earnings Inequality and the Changing Association between Spouses' Earnings. *American Journal of Sociology*, 115(5):1524–1557.

Schwartz, C. R. & Mare, R. D. (2005). Trends in Educational Assortative Marriage from 1940 to 2003. *Demography*, 42(4):621–646.

Sinkhorn, R. & Knopp, P. (1967). Concerning Nonnegative Matrices and Doubly Stochastic Matrices. *Pacific Journal of Mathematics*, 21(2):343–348.

Tan, P.-N., Kumar, V., & Srivastava, J. (2004). Selecting the Right Objective Measure for Association Analysis. *Information Systems*, 29(4):293–313.

A. DADOS

A base de dados usada neste estudo é o censo demográfico brasileiro dos anos de 1970, 1980, 1991, 2000 e 2010 e pode ser facilmente obtida através da página do IPUMS (Integrated Public Use Microdata Series) na internet. As variáveis incluídas na nossa base de dados são: ano do censo (nome da variável: year), código identificador da esposa (sploc), número de membros em uma família (famsize), número de filhos (nchild), idade (age), sexo (sex), estado civil (marst), nível educacional (educ), relação de emprego (empstat) e renda individual (inc). No nosso estudo só consideramos casais e solteiros entre 25 e 54 anos. Os adultos vivem por conta própria ou com seus filhos, que têm idade inferior a 19 anos. Famílias em que existem outros membros tais como avós, tios/tias, ou quaisquer outros indivíduos estão excluídas da base de dados. Assim como famílias com subfamílias de qualquer tipo. Além disso, viúvos/viúvas e indivíduos casados cujas as esposas estão ausentes também são excluídos. As rendas são estritamente positivas.

Na nossa análise, há 384 tipos diferentes de famílias e elas são classificadas da seguinte maneira:

1. Estado civil: casados, homens nunca casados, mulheres nunca casadas, homens divorciados e mulheres divorciadas;
2. Educação: primária incompleta, primária, secundária e universidade;
3. Mercado de trabalho: trabalha, não trabalha;
4. Filhos: nenhum filho, um, dois, ou mais de dois filhos.

Por fim, as famílias são divididas em dez percentis. Assim, para cada ano, temos 3840 (i,j) combinações de famílias para todos os tipos e decis.

B. CURVA DE LORENZ E ÍNDICE DE GINI

Suponha uma amostra de diferentes tipos de famílias, $i \in \{1,2,...,m\}$, situadas em diferentes percentis, $j \in \mathcal{J}$, da distribuição de renda. Defina f_{ij} como a fração de famílias que são do tipo i e estão no percentil j e r_{ij} como a renda y_{ij} relativa à renda média y. A renda de cada família é ajustada para a base equivalente de um adulto de acordo com a metodologia da OCDE, que considera o primeiro adulto como 1, o segundo como 0.5 e cada criança como 0.3. As rendas equalizadas das famílias são então divididas pela renda média da população.

A fração de renda do percentil j é

$$s_j = \sum_i f_{ij} r_{ij}.$$

A Curva de Lorenz é derivada plotando as frações acumuladas da população indexadas por p,

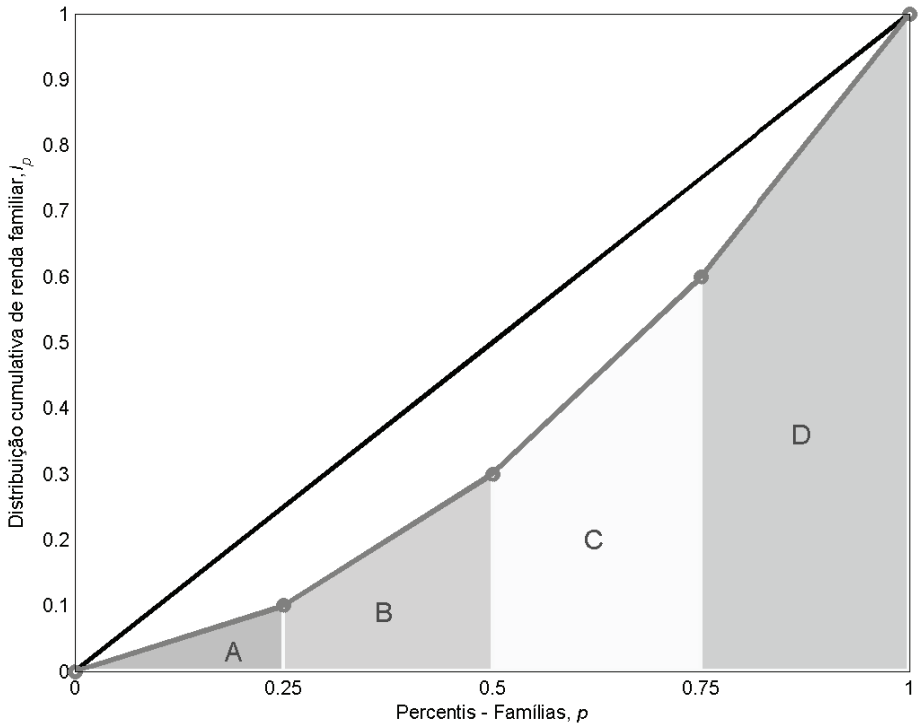
$$p = \sum_j^p \sum_i^m f_{ij}$$

no eixo x contra a fração de renda acumulada pelo percentil p,

$$l_p = \sum_{j}^{p} s_j,$$

no eixo y. Suponha que o intervalo unitário seja dividido em n segmentos iguais. Assim, $j \in \mathcal{J} = \{1/n, \cdots, 1-1/n, 1\}$.

Imagine o exemplo em que $n = 4$ (quartil). A Curva de Lorenz descrita acima está plotada na Figura 1. O Índice de Gini associado a essa Curva de Lorenz é duas vezes a área entre a curva e a linha de 45 graus. O índice também pode ser calculado como $1 - 2\Delta$, onde Δ é a área abaixo da Curva de Lorenz. No caso em que $n = 4$, o índice é a soma das áreas do triângulo A e dos trapezóides B, C, e D. As coordenadas no eixo x são dadas por 0, $p_1 = 0.25$, $p_2 = 0.5$, $p_3 = 0.75$, e 1.0. As coordenadas no eixo y são 0, $l_{0.25}$, $l_{0.5}$, $l_{0.75}$, e 1.0.



Nota: A figura mostra a construção de uma Curva de Lorenz quando há quatro percentis (quartis). O Índice de Gini é duas vezes a área entre a linha de 45 graus e a Curva de Lorenz.

Assim, utilizando as fórmulas para as áreas geométricas A, B, C, e D, o Índice de Gini, g, pode ser derivado como

$$g = 1 - 2\left(\underbrace{\frac{p_1 l_1}{2}}_{\text{Área } A} + \underbrace{\frac{(l_1 + l_2)(p_2 - p_1)}{2}}_{\text{Área } B} + \underbrace{\frac{(l_2 + l_3)(p_3 - p_2)}{2}}_{\text{Área } C} + \underbrace{\frac{(l_3 + 1)(1 - p_3)}{2}}_{\text{Área } D}\right).$$

Após rearranjar a expressão acima, o Índice de Gini é dado por

$$g = (p_1 l_2 - p_2 l_1) + (p_2 l_3 - p_3 l_2) + (p_3 - l_3).$$

As frações acumuladas da população, os p's, são baseadas nos quartis; i.e., $p_1 = 1/4$,$p_2 = 2/4, \cdots$. Portanto, a expressão anterior pode ser escrita como

$$g = \frac{1}{4}\left[(l_2 - 2l_1) + (2l_3 - 3l_2) + (3 - 4l_3)\right].$$

No caso geral de n percentis, o Índice de Gini é

$$g = \sum_{p=1/n}^{1-1/n} [pl_{p+1} - (p + 1/n)l_p].$$

A versão dessa fórmula para um número arbitrário de grupos de renda e de sub-populações (tipos) é apresentada em Rao (1969).

C. ALGORITMO SINKHORN-KNOPP (1967)

1. Comece uma iteração com uma tabela de contingência.
2. Essa tabela de contingência possui uma distribuição marginal associada a cada linha (marido) obtida a partir da soma de cada linha ao longo de suas colunas para obter o total daquela linha. Divida cada linha por 4 vezes o seu total. A distribuição marginal asscoiada às linhas passa a ser $(1/4,1/4,1/4,1/4)$.
3. Compute a distribuição marginal associada a cada coluna (esposa) somando cada coluna ao longo de suas linhas para obter o total daquela coluna. Divida cada coluna por 4 vezes o seu total.
4. Recompute a distribuição marginal associada a cada linha Verifique se ela mudou após as duas etapas anteriores, checando a distância até a distribuição marginal desejada $(1/4,1/4,1/4,1/4)$. Se atingiu o nível de proximidade desejado, então pare. Caso contrário, volte ao passo 1.

r b e

ISSN 0034-7140 (print) | ISSN 1806-9134 (online)
DOI 10.5935/0034-7140.20170018
Revista Brasileira de Economia, 71(3), 379–394

Multicointegração e Sustentabilidade da Política Fiscal no Brasil com Regime de Quebras Estruturais (1997-2015)

Divanildo Triches*
Luís Antônio Sleimann Bertussi†

Sumário: **1. Introdução; 2. Solvência e endividamento do setor público; 3. Condução da política fiscal; 4. Aspectos metodológico de multicointegração com quebra estrutural; 5. Descrição e análise dos resultados; 6. Conclusões; A. Síntese dos estudos empíricos e modelos aplicados.**

Palavras-chave: **Multicointegração; Política fiscal; Déficits orçamentários; Sustentabilidade.**

Códigos JEL: **H30, H62, H63.**

Este artigo tem como objetivo investigar a sustentabilidade da política fiscal do Governo Central do Brasil no período de 1997 a 2015, levando em conta os impactos dos aumentos do gasto público sobre a dinâmica da sustentabilidade da dívida. Para isso, trata-se da dinâmica fiscal pré-crise e pós-crise, e as principais mudanças na evolução das receitas e despesas públicas. Para testar a sustentabilidade da política fiscal um modelo de multicointegração com quebras estruturais é empregado com séries trimestrais. O déficit nominal teve um crescimento acentuado, passando de R$ 343,9 bilhões em 2014, para R$ 613,0 bilhões em 2015, resultado, sobretudo, do impacto das despesas com juros nominais sobre a dívida pública que crescerem 6,64% para 8,50% do produto da economia no período. Os resultados dos testes multicointegração com quebras estruturais indicam que a hipótese da existência de uma relação de equilíbrio de longo prazo as variáveis de fluxo receita, a despesa e o estoque de dívida do setor governamental não pode ser aceita. Isso evidência uma dinâmica de sustentabilidade da política fiscal fraca, pois na média o crescimento das receitas acumuladas é significativamente menor do que as despesas acumuladas.

This paper aims to investigate the sustainability of the fiscal policy of the Brazilian Central Government from 1997 to 2015. It considers the impact of public spending increases on the dynamics of public accounts. For this, the pre-crisis and post-crisis fiscal dynamics and major changes in the evolution of public revenues and expenditure are analyzed. To test the sustainability of fiscal policy, a model of multicointegration with structural breaks is employed with quarterly data. The nominal deficit had a sharp increase, from R$ 343.9 billion in 2014 to R$ 613.0 billion in 2015, a result of the impact of expenses with nominal interest on the public debt to grow from 6.64% to 8.50% of GDP in the period. The Multicointegration test pointed out that the hypothesis of the existence of a long-run equilibrium relation between the flow variables revenue, expenditure and government debt stock cannot be accepted. This evidences a sustainability dynamic of weak fiscal policy, since on average accumulated revenue growth is significantly lower than accumulated expenditure.

*Doutor em Economia pela Universidade Federal do Rio Grande do Sul. Pesquisador e professor no Programa de Pós-Graduação em Economia da Universidade do Vale do Rio dos Sinos, PPGE/Unisinos. E-mail: divanildot@unisinos.br

†Doutorando em Economia pela Universidade do Vale do Rio dos Sinos e Professor na Universidade de Passo Fundo. E-mail: luisbertussi@upf.br

Recebido em 23 de fevereiro de 2017
Aceitação Final em 14 de julho de 2017

1. INTRODUÇÃO

As políticas fiscais expansionistas têm caracterizado a gestão dos agentes políticos no Brasil nas últimas décadas, em especial, pelo uso das despesas de custeio e investimentos como instrumento para interferir na dinâmica do produto da economia. No entanto, em diversos episódios, seja para acelerar o crescimento ou para enfrentar recessões econômicas o uso da política fiscal têm causado descontrole nas contas públicas, aumento do endividamento e pressões inflacionárias. Tais fatos decorrem porque os déficits públicos são financiados por meio de colocação de títulos junto ao setor privado que implica em aumento da dívida governamental e/ou por emissão de moeda. Há um consenso, na análise econômica, da necessidade em manter a relação do nível endividamento público e o produto interno sustentável. No entanto, para Camarero et alii (2015) a atenção dada pelos formuladores da política fiscal tem se concentrado, prioritariamente, na proporção entre déficit e o produto e, na estabilidade e crescimento.

A sustentabilidade fiscal requer que setor governamental seja solvente ou demonstre capacidade de honrar a sua dívida no futuro. Nesse sentido, a literatura empírica tem empregado modelos de cointegração para avaliar a dinâmica de ajustamento das variáveis de estoque e de fluxos, ou seja, saldo fiscal e dívida pública com Afonso (2005), Bajo-Rubio et alii (2010), Claeys (2007), Nikolov et alii (2010), Trachanas e Katrakilidis (2013) para os países europeus e Hatemi e Hatemi-J e Zanella (2013), Luporini (2012), Simonassi et alii (2014) para o caso brasileiro. Outros estudos testaram a sustentabilidade fiscal por meio do modelo Multicointegração baseado em Engsted et alii (1997) como Leachman e Francis (2000), Leachman et alii (2005), Bertussi e Triches (2013). Mas recentemente, a preferência recaiu para modelo que levem em conta a quebra estrutural como Berenguer-Rico e Carrion-i Silvestre (2011), Camarero et alii (2015), Tronzano (2012).[1]

No Brasil, no período pós-crise econômica financeira de 2008, iniciada no mercado *subprime*, os mecanismos de incentivo à atividade econômica por vias fiscais foram amplamente empregados por meio do crescimento nos dispêndios diretos, renúncia tributária, empréstimos subsidiados entre outras medidas. Num primeiro momento, a economia brasileira reagiu bem aos impactos negativos que a crise teve sobre o produto e o emprego. Contudo, passada a crise, a manutenção dos incentivos e a consequente ampliação dos dispêndios com o intuito de fomento à atividade econômica mostrarem-se nocivos para as contas do setor público, bem como para a dinâmica do estoque da dívida.

Nesse contexto, o objetivo do estudo é investigar a sustentabilidade da política fiscal do Governo Central do Brasil, no período de 1997 a 2015, levando em conta os impactos que as ampliações do gasto público tiveram sobre a dinâmica das contas públicas. Para isso utiliza-se metodologia baseada no modelo de multicointegração com quebras estruturais propostos por Berenguer-Rico e Carrion-i Silvestre (2011) e Tronzano (2012).

O tema de pesquisa justifica-se em função da relevância da política fiscal para a sociedade brasileira uma vez que a dinâmica fiscal tem efeitos diretos no equilíbrio macroeconômico do país. Por outro lado, a falta de responsabilidade fiscal pode ter custos elevadíssimos para toda a economia. No Brasil, ao longo da década de 80, o desequilíbrio orçamentário gerou altos custos para a sociedade, como aceleração da taxa inflação. Assim o desempenho da política fiscal para equilíbrio macroeconômico tem motivado a mensuração de sua sustentabilidade. Além disso, pretende-se contribuir com a literatura, no sentido que há ainda poucos estudos para economia brasileira que utiliza esse tipo de metodologia. Claramente, as evidências encontradas mostram que entre 1997 a 2015, a dinâmica de sustentabilidade da política fiscal foi fraca, pois, em termos médios, o crescimento das receitas acumuladas foi significativamente menor do que as despesas.

O artigo está composto por cinco seções, além dessa introdução. Na primeira, faz-se revisão da abordagem sobre a solvência e endividamento do setor público. Na segunda, apresenta-se uma análise da condução da política fiscal. A terceira seção traz a descrição do modelo de multicointegração com

[1] No apêndice, encontra-se um quatro sintético dos estudos empíricos selecionados.

quebra estrutural e a definição das variáveis. Nas quartas e quintas seções são analisados os resultados e apresentadas as conclusões, respectivamente.

2. SOLVÊNCIA E ENDIVIDAMENTO DO SETOR PÚBLICO

A dinâmica da sustentabilidade da dívida pública de um país é considerada um dos norteadores das expectativas de solvência do setor governamental a qual depende da magnitude dos déficits orçamentários. Os déficits orçamentários podem ser financiados de quatro formas: (i) pelo aumento da carga tributária (ii) pela emissão monetária; (iii) tomando emprestado do setor privado interno ou externo na forma de contratos; e (iv) pela emissão de títulos da dívida pública. Outra solução para o desequilíbrio nas contas do governo, como discutido por Sachs e Larrian (2000) e Triches e Morais (2005), poderia se dar por meio de escolha de políticas que combinassem a redução de gastos e aumento de impostos. Contudo, esse tipo de política é de difícil de implementação porque exige apoio da maioria do poder legislativo. A dificuldade aumenta nos países em que os governos são fracos no ponto de vista político e/ou formados por coalizão que sofre forte pressão dos grupos de interesse organizados com discorre Triches (2005).

O equilíbrio intertemporal do orçamento público é fundamental para credibilidade das políticas fiscais e solvência do endividamento governamental. Desse modo, a restrição orçamentária do setor público seguindo Walsh (1998), pode ser definida pela equação (1)

$$G_t + i_{t-1}B_{t-1}^T = T_t + \left(B_t^T - B_{t-1}^T\right) + RBC_t \tag{1}$$

onde o lado esquerdo da equação, consiste dos gastos do setor público em bens, serviços e transferências (G_t), mais o pagamento de juros sobre a dívida pública interna e externa, $\left(i_{t-1}B_{t-1}^T\right)$. O lado direito representa as fontes de recursos, provenientes da arrecadação fiscal (T_t), novas emissões de títulos, $\left(B^T - B_{t-1}^T\right)$, mais as receitas do Banco Central, (RBC_t).

A autoridade monetária, representada pelo Banco Central, também tem uma restrição orçamentária, sendo que esta deve exibir a identidade entre os seus ativos e passivos, dada pela relação (2)

$$\left(B_t^M - B_{t-1}^M\right) + RBC_t = i_{t-1}B_{t-1}^M + \left(H_t - H_{t-1}\right) \tag{2}$$

onde $\left(B_t^M - B_{t-1}^M\right)$ representa montante de títulos do Tesouro em poder do Banco Central, $\left(i_{t-1}B_{t-1}^M\right)$ são juros recebidos do Tesouro e, por fim, $\left(H_t - H_{t-1}\right)$ é a variação da base monetária.[2] A restrição orçamentária do setor governamental consolidada, isto é, o orçamento público e o orçamento do Banco Central passa ser expresso pela equação (3)

$$G_t + i_{t-1}B_{t-1} = T_t + \left(B_t - B_{t-1}\right) + \left(H_t - H_{t-1}\right) \tag{3}$$

As variáveis do lado direito da equação (3) mostram os recursos que o governo pode obter, ou seja, os impostos (T_t) (não está incluída a receita inflacionária), empréstimos do setor privado via emissão de títulos $\left(B_t - B_{t-1}\right)$ e, ainda, a emissão de moeda $\left(H_t - H_{t-1}\right)$.

A sustentabilidade da dívida pública somente ocorre quando for satisfeita restrição orçamentária governamental intertemporal, isto é, valor presente descontado de seus gastos primários (inclusive, encargos financeiros) correntes e futuros não são maiores que o valor presente descontado de sua renda corrente e futura, líquida de qualquer endividamento inicial, como discorre Blanchard et alii (1990) e Bertussi e Triches (2013). Desse modo, as variáveis tais como taxa de crescimento real da economia,

[2]Essa é constituída pelo estoque de moeda mantido pelo público não financeiro mais as reservas bancárias.

relação base monetária e produto, e taxa de juros real caracterizam como determinantes essenciais que afetam essa restrição orçamentária. Tomando a equação (3) em proporção ao produto, chega-se a dinâmica da dívida, conforme equação (4):

$$\Delta b_t = d_t + \big(r_{t-1} - \rho_{t-1}\big)b_t - \lambda m_t \tag{4}$$

onde b representa a razão entre a dívida e o produto; d, o déficit primário também como proporção ao produto; r, a taxa real de juros da dívida; ρ, a taxa real de crescimento do produto; m, a razão base monetária e produto e λ é a variação da base monetária nominal, a qual se pressupõe ser igual à taxa de variação do produto nominal. É notório, pela equação (4) se a taxa real de juros da dívida exceder, persistentemente, a taxa real de crescimento da economia, a dinâmica da dívida passa a ser insustentável ou seguirá uma trajetória explosiva como argumentam Bertussi e Triches (2012).

A definição de sustentabilidade, para IMF (2012), diz que uma instituição está numa situação sustentável se ela satisfaz sua restrição orçamentária, sem a necessidade de uma grande correção nas receitas e despesas, tendo em conta os custos de seu financiamento. Nesse caso, a definição de sustentabilidade agrega os conceitos de solvência e liquidez dos ativos colaterais da dívida. Ainda, segundo IMF (2012), existem algumas condições que tornariam insustentável a dinâmica da dívida, dentre as quais se destacam as seguintes: a) dívida atinge tal magnitude que seja necessária sua reestruturação; b) país acumule dívidas em velocidade superior aos recursos que consegue auferir para pagamento dos serviços da dívida; c) o nível de endividamento da nação cresça acima de sua capacidade, mesmo que será necessário um expressivo ajuste na economia para honrar os serviços da dívida. Partindo de tais pressupostos, os custos do financiamento são cruciais para a sustentabilidade da dívida pública, que a afeta a restrição orçamentária intertemporal como discute Bertussi e Triches (2012).

O estudo pioneiro que trata dos efeitos dos déficits e a sustentabilidade da dívida pública foi o de Domar (1944) em que vincula a dinâmica da dívida ao crescimento da economia. Ele mostra que quando a trajetória taxa de juros nominal cresce acima da taxa de crescimento da economia leva uma dinâmica insustentável da dívida. Para a economia brasileira, os estudos mostraram que a senhoriagem se constituiu num mecanismo de financiamento governamental nos períodos de alta inflação, como abordam Pastore (1995) e Rocha (1997). Para os autores, nas décadas de 80 e 90, o orçamento do governo foi equilibrado e respeita o valor presente. As receitas auferidas com senhoriagem foram cruciais para manutenção da solvência devido à constância dos déficits orçamentários.

Os estudos posteriores que abrange o período de estabilização monetária como de Silva e Gamboa (2011) mostraram que com a inclusão da senhoriagem, o teste de cointegração, ao longo do período de 1986 a 2006, indicou forte sustentabilidade fiscal. Ao passo que sem receitas de senhoriagem, o teste mostrou que a sustentabilidade não pode ser definida como forte nem fraca. A avaliação sustentabilidade fiscal para países da América Latina, partindo de um modelo de multicointegração, foi desenvolvida por Bertussi e Triches (2012). O resultado foi de que não foi possível verificar a existência de relação entre receitas e despesas no longo prazo. Isso ocorreu indiferente do cenário, seja ele de recessão ou de crescimento econômico O Brasil foi um dos países que respeitou restrição orçamentária, no período analisado de 1997 até 2007, embora as despesas tenham apresentado um crescimento maior do que as receitas.

A sustentabilidade fiscal, após a estabilização monetária, teve papel de destaque nas questões finanças públicas, impondo limites ao endividamento e superávit primário, que foram implementados no período pós 1994. Assim, Simonassi e Arraes (2007) avaliou sustentabilidade da política fiscal, no período de 1991 a 2006, utilizando técnicas baseadas nas reações do governo à acumulação de dívida. O autor mostra que o governo brasileiro apresentou dificuldades em vender títulos públicos e passou a usar as receitas de senhoriagem para financiamento os déficits orçamentários e que a política fiscal se mostrou sustentável.

Leachman et alii (2005) utilizaram a metodologia de multicointegração para criar um conjunto de critérios para testar a sustentabilidade fiscal do processo orçamentário em qualquer cenário econômico.

Os autores testaram a sustentabilidade fiscal das receitas e despesas para 15 países industrializados com tais mecanismos. O resultado do teste demonstrou que somente a Noruega e o Reino Unido possuíam políticas fiscais sustentáveis com base no modelo aplicado.

Escario et alii (2012) analisaram a sustentabilidade da política fiscal da Espanha no longo prazo utilizando a metodologia de multicointegração. Eles avaliaram a possibilidade de equilíbrio na relação entre variáveis fluxo e estoque. Os resultados revelaram que a senhoriagem foi essencial para garantir a solvência intertemporal da Espanha. No entanto, houve uma mudança na forma de financiamento do gasto público quando aderiu á moeda comum europeia em 1998, quando perdeu a capacidade de monetização da dívida. Os autores demonstram que o país necessitava de mudanças estruturais para consolidação de um regime fiscal que mantivesse uma relação de equilíbrio de longo prazo sem monetização dos gastos.

Tronzano (2012) utiliza a técnica de multicointegração com quebras estruturais para avaliar a sustentabilidade fiscal, na Índia para o período de 1950 a 2010. O Autor usou uma série de equações de multicointegração com quebras estruturais que rejeitaram a existência de um equilíbrio de longo prazo para as séries de receita e despesa. Tal fato veio a demostrar a insustentabilidade da política fiscal num ambiente estocástico. Por fim, ele conclui que o governo deveria estabelecer regras de equilíbrio entre as receitas as despesas e o estoque de dívida.

Camarero et alii (2015) unificam as metodologias para testar a sustentabilidade da política fiscal com variáveis fluxo e estoque. Eles aplicam os modelos de Berenguer-Rico e Carrion-i Silvestre (2011) para testar a hipótese nula de não cointegração em variáveis I(2) e estendem o modelo para permitir a possibilidade da presença de duas quebras estruturais. Essa metodologia foi aplicada para um grupo de 17 países da OCDE para os quais evidenciam sustentabilidade fiscal fraca e a existência de cointegração entre o déficit e o estoque de dívida. Os resultados demonstram a existência de uma relação de equilíbrio de longo prazo para 6 dos 17 países analisados. Quatro países, como Canadá, Itália, Japão e Reino Unido, demonstram resultados similares de sustentabilidade da política fiscal e dois casos, como os Estados Unidos e a Grécia, apresentaram insustentabilidade da política fiscal a partir de 1983 para a Grécia e para o período entre 1987 e 1999 para os Estados Unidos.

3. CONDUÇÃO DA POLÍTICA FISCAL

3.1. Política econômica no pós-crise 2008

A crise econômica iniciada nos EUA, em setembro 2008, teve impacto significativo nas taxas de crescimento da economia mundial. Em resposta a isso e como medida de estímulo à atividade econômica, diversas nações, incluindo o Brasil, abriram mão de políticas nas quais o equilíbrio orçamentário predominava, para adotar postura de afrouxamento orçamentário com políticas de expansão do gasto público e redução de impostos.

No final de 2008 e início de 2009, o governo brasileiro adotou várias medidas anticíclicas para diminuir os seus efeitos sobre o nível da atividade econômica. De fato, a economia brasileira conseguiu se manter relativamente estável frente ao agravamento das condições econômicas mundiais. Em 2009, a economia brasileira apresentou queda de 0,20 ponto percentual no PIB. As medidas adotadas estimularam a economia que passou a apresentar sinais fortes de recuperação com um crescimento de 7,6% em 2010. Em contrapartida, a dinâmica crescimento média da economia mundial foi de queda de 2,1% e uma elevação 4,1%, nos dois períodos, ou seja 2009 e 2010, respectivamente como ilustra a Tabela 1. Tal desempenho foi ainda mais fraco quando se observa os países que integram a OCDE. Em termos de crescimento médio ao longo de 2004 a 2014, a economia brasileira também se destacou com 3,6% ao ano, enquanto economia mundial cresceu 2,6% e os países da OCDE, apenas 1,5% ao ano.

A utilização de um tripé de políticas macroeconômicas previsíveis com câmbio flutuante, superávit primário de 3% do PIB e sistema de metas de inflações e o consequente acúmulo de reservas internacionais em virtude da entrada excessiva de moeda estrangeira amenizaram os impactos da crise nos

Tabela 1: Evolução do crescimento do produto comparado entre o Brasil e a economia mundial entre 2004 e 2014 (%).

Regiões	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	Média
Mundo	4,1	3,6	4,1	3,9	1,5	-2,1	4,1	2,8	2,2	2,4	2,5	2,6
Brasil	5,7	3,1	4	6	5	-0,2	7,6	3,9	1,8	2,7	0,1	3,6
Países OECD	3,2	2,6	3	2,6	0,2	-3,6	2,9	1,7	1,2	1,4	1,7	1,5

Fonte: Banco Mundial.

anos de 2009 a 2011. Em especial, a possibilidade de utilização de políticas fiscais expansionistas com aumento de despesas e redução de impostos. Isso permitiu que governo brasileiro adotasse políticas macroeconômicas anticíclicas que, por sua vez, possibilitaram uma rápida retomada do crescimento. Contudo, vários indicadores sinalizavam a desaceleração da economia como retração do setor industrial, com 7,4%, dos níveis de consumo e o investimento privado com queda, na ordem de 2% e 7% no último trimestre de 2008.

Diante desse cenário, as políticas adotadas para amenizar a retração econômica foram baseadas em instrumentos monetários, visando uma elevação na liquidez na economia, ou seja; a) redução de compulsórios, b) antecipação de linhas de crédito, c) abertura de novas linhas de crédito e d) intervenção no mercado de câmbio para segurar a variação cambial. No lado da política fiscal, foram adotadas reduções de impostos em diversos seguimentos da economia. Num primeiro momento, no mês de outubro de 2008, foram contempladas com a redução de imposto sobre operações financeiras (IOF) as operações no mercado de câmbio. Posteriormente, passou-se a reduzir a alíquota de IOF, os financiamentos de motocicletas, motonetas e ciclomotores, e as alíquotas de IOF sobre operações de crédito para pessoas físicas foram reduzidas no mês seguinte.

No mês de dezembro 2008, os instrumentos fiscais tiveram predominância, foram reduzidas as alíquotas de imposto sobre produtos industrializados (IPI), incidentes sobre a venda de veículos e caminhões. Tal medida que teria caráter provisório, com vigência de um ano, mas constantemente, foi prorrogada perdurando até dezembro de 2014. As alíquotas de imposto de renda sobre pessoa física (IRPF) foram alteradas, criando-se faixas intermediárias com alíquotas reduzidas. O governo adotou ainda uma política industrial agressiva, para amenizar os impactos no setor produtivo, como repasses de recursos do Banco Nacional de Desenvolvimento Social.

Embora houvesse a recuperação econômica no ano seguinte à crise de 2008, ao longo do primeiro mandato da Presidente Dilma Rousseff, outras políticas fiscais com foco no consumo e investimento foram adotadas. Em agosto de 2011, foi lançado o Plano Brasil Maior que pode ser traduzido como uma série de políticas governamentais expansionistas. O seu intuito era fomentar a atividade econômica, as quais tiveram dois estágios de aplicação 2011 e 2012, conforme MDIC (2011). Outros estímulos também foram adotados com a edição da medida provisória 540, e após convertida em lei que trata da alteração de tributos incidentes sobre a folha de pagamento de algumas categorias. No seu período de maior abrangência, englobou 56 setores, sendo a renúncia fiscal estimada de R$ 22 bilhões, em 2014, conforme RFB (2014).[3]

Apesar de todas essas medidas, a resposta dos indicadores macroeconômicos foi aquém das expectativas. Logo, o governo passou para a segunda fase do Plano Brasil Maior com ampliação dos dispêndios

[3]No seu período de maior abrangência, englobou 56 setores, sendo a renúncia fiscal estimada de R$ 22 bilhões, em 2014. Foram instituídas ainda medidas de impacto nas contas públicas como: i) ampliação da abrangência do simples nacional, ii) crescimento dos repasses pelo BNDES, iii) ampliação do programa microempreendedor individual (MEI) como aponta MDIC (2011).

do BNDES, ampliação das desonerações da folha de pagamento, redução do IPI sobre bens de consumo, entre outras. Em 2013, a redução do PIS e Cofins vigentes sobre os produtos da cesta básica foi anunciada, com impacto nas contas públicas de R$ 7,5 bilhões, no ano de 2014 (Lei N. 12.839, de 9 de julho de 2013), conforme MDIC (2012). Por outro lado, não foram somente as renúncias tributárias e dispêndios ao BNDES os norteadores da ação fiscal de estímulo à economia. Durante o período de 2007 até 2014, o governo procurava aquecer as atividades econômicas com intervenções diretas feitas através de programas como: Minha Casa Minha Vida, PAC (Programa de Aceleração do Crescimento) que teve três etapas com a primeira iniciando em 2007, programa de compras governamentais que foi lançado em 2012 os quais abrangeram até compras militares, entre outros dispêndios.

3.2. Visão geral da situação fiscal do Governo Central

A condução da política fiscal no Brasil, de forma geral, esteve condicionada pelo ciclo político e pela necessidade permanente de equilibrar as contas públicas. Nesse sentido, foram identificados, ao longo de 1997 a 2015, cinco ciclos de geração resultados primário e nominal pelo Governo Central, como ilustra a Figura 1. O primeiro inicia em 1997 e termina em 2008, com a geração de superávits primários que permitiram uma redução do déficit nominal do governo central em virtude de um maior controle das contas públicas. Tal fato foi motivado por questões econômicas, como ciclo de alta da demanda e dos preços das commodities no mercado internacional e por questões legais e políticas, como o acordo com FMI após crise das eleições de 2002. O resultado primário do setor público como proporção do PIB foi superavitário, em média de 3,43% entre 2004 até 2008 como mostra a Tabela 2. Já o pagamento médio de juros sobre a dívida pública foi 6,37% do PIB o que resultou um déficit na necessidade de financiamento do setor público no conceito o nominal, em média, de 2,94%, representando cerca de R$ 56,3 bilhões.

Tabela 2: Resultado primário, nominal e juros nominais de 2004 a 2015 (%PIB).

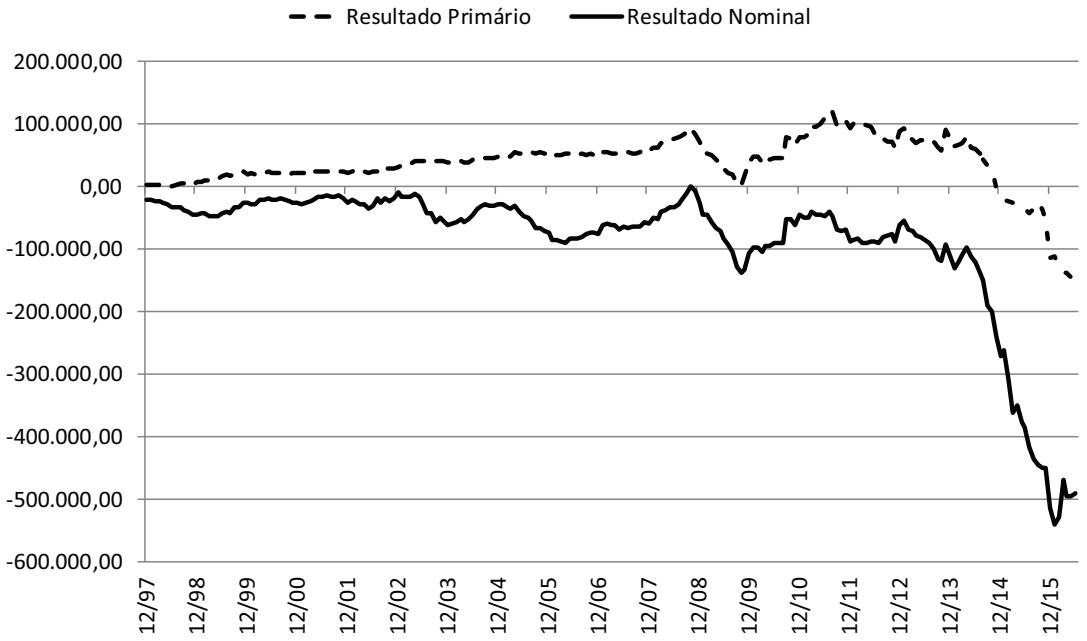
	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Primário	3,69	3,74	3,15	3,24	3,33	1,95	2,62	2,94	2,23	1,77	-0,6	-1,9
Juros Nominais	6,56	7,28	6,72	5,98	5,33	5,14	5,03	5,41	4,54	4,83	5,46	8,5
Nominal	2,87	3,54	3,57	2,74	1,99	3,19	2,41	2,47	2,31	3,05	6,23	10,3

Fonte: BCB.

O segundo ciclo, mais curto, ocorre entre outubro de 2008 a outubro de 2009. Os resultados nominal e primário apresentaram significativa piora em relação ao período anterior motivado pelas políticas anticíclicas praticadas pelo governo, pós crise subprime. Assim, a economia feita pelo governo para pagamento dos juros da dívida teve redução, partindo de um superávit primário de 3,33%, em 2008, para superávit de 1,95% do PIB, em 2009. Em valores absolutos, o resultado foi de R$ 103,5 bilhões, em 2008, para R$ 64,7 bilhões, em 2009, sendo que os dividendos pagos pelas empresas estatais à União, praticamente, dobraram, no ano de 2009, em relação a 2008, totalizando R$ 26,6 bilhões ou 0,80% do PIB. O pagamento de juros sobre o endividamento público permaneceu em torno dos 5% que implicou no déficit nominal de 1,99%, em 2008, para 3,19%, em 2009, em termos de valores correntes, representa um crescimento de R$ 61,9 bilhões para R$ 106,2 bilhões.

O intervalo de outubro de 2009 a julho de 2011 caracteriza-se como o terceiro ciclo em que se observa uma tentativa da retomada da atividade econômica, superávits mais expressivos puderam ser obtidos, em 2010 e 2011, com 2,62% e 2,94%, respectivamente. Esse resultado se deve a uma elevação significativa na arrecadação tributária. Em 2009, a receita total do tesouro nacional foi da ordem de

Figura 1: Resultados primário e nominal do governo entre dezembro de 1997 a abril de 2016 (R$ milhões, acumulados em 12 meses).



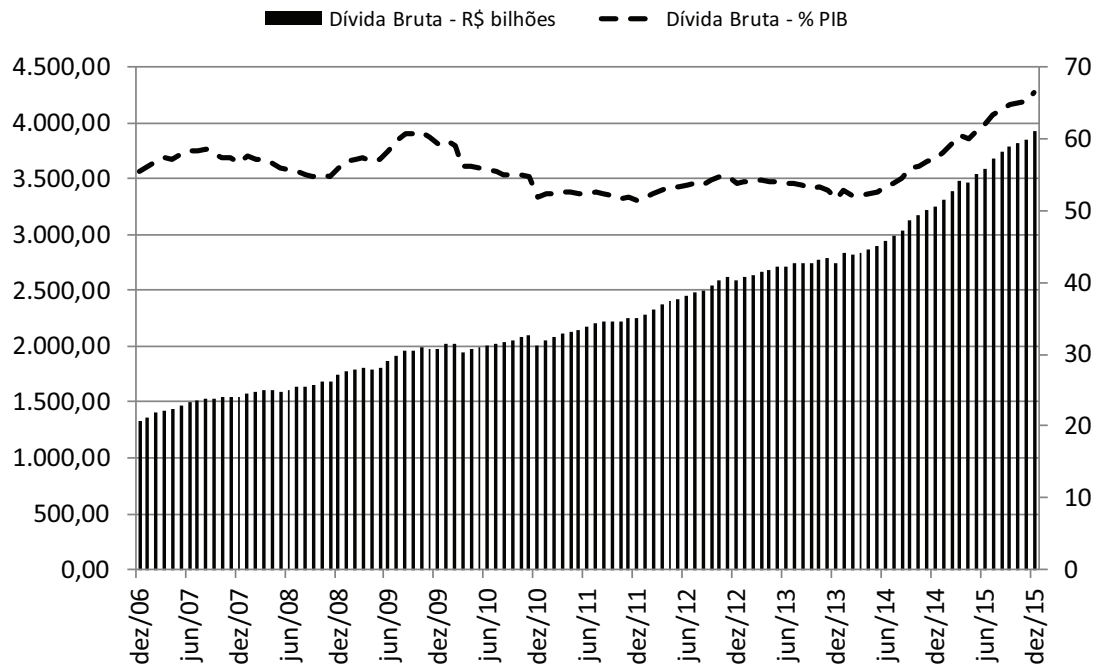
Fonte: STN (2015) e (2016).

R$ 739,3 bilhões ou 22,2% do PIB, uma queda de 1,1% como proporção do produto da economia quando comparada ao ano anterior. Em compensação, em 2010 receita teve um expressivo crescimento, atingindo 23,7% do PIB ou R$ 919,7 bilhões em valores correntes. Sendo assim, a melhora na atividade econômica impactou diretamente no déficit nominal que foi de 2,41% do produto, em 2010, ou R$ 93,6 bilhões.

Contudo, com o declínio das condições econômicas internas e a adoção da nova matriz macroeconômica do governo entre 2011 a 2014, inicia-se o quarto ciclo, no qual as condições fiscais apresentam significativa piora com a geração de déficits primário e nominal em virtude da utilização de políticas fiscais expansionistas. A deterioração do quadro fiscal, em 2012, só não foi maior devido à contração do pagamento dos juros nominais que passaram de 5,41% para 4,54% do PIB, ou seja, uma redução nominal de R$ 22,8 bilhões. A situação fiscal foi se agravando e, no ano de 2013, registrou-se um resultado primário do setor público pior do que no ano da crise econômica mundial. Isso implicou um déficit nominal de 3,05% do PIB ou R$ 157,5 bilhões, com um crescimento nominal de R$ 48,3 bilhões em relação a 2012. Os juros reais mais que dobraram, entre 2012 a 2013, de R$ 61,9 bilhões para R$ 133,1 bilhões ou 2,58% do PIB.

O pior dos cinco ciclos de déficits fiscais inicia, em abril de 2014, o qual foi marcado por uma grave deterioração da situação fiscal do país em que foi registrado um déficit primário de 0,59% ou R$ 32,5 bilhões e, no ano seguinte, cresceu ainda mais, passando para 1,9%. Aliado por uma forte elevação de custo da dívida que passou de 5,5%, em 2014, para 8,5% do PIB em 2015, resultando uma elevação do déficit nominal de 6,0% para a 10,4% em relação ao produto, ainda conforme tabela 2. Em termos absolutos, isso representa, na ordem, de um montante de R$ 343,9 bilhões para R$ 613,0 bilhões. A deterioração das finanças públicas foi também magnificada pelas políticas de contenção a desvalorização

Figura 2: Revolução da dívida bruta do governo geral brasileiro entre dezembro de 2006 a junho de 2016 (R$ milhões).



Fonte: STN (2015) e (2016). *Metodologia utilizada a partir de 2008.

cambial que foi usada por meio de lançamentos de swaps cambiais.[4] O Banco Central realizou perdas em 2014, nessa rubrica, que somaram R$ 17,3 bilhões que acabou impactando no resultado nominal do setor público.

O resultado dessa política fiscal expansionista do governo apresentou uma elevação crescente da dívida bruta geral, passando de em torno de R$ 1,4 trilhão em dezembro de 2006, para cerca de R$ 4,2 trilhões em junho de 2016. Isso representa o crescimento médio anual de aproximadamente 12,1%, conforme pode se notar na Figura 2.

Salienta-se que o ritmo do crescimento da dívida se intensificou a partir do final de 2013. Esse fato é confirmado, como abordado anteriormente pela elevação do déficit nominal, e portanto, pela relação entre a dívida total e o produto interno a qual vinha se mantendo relativamente estável oscilando entre 50% a 60% do um pouco mais em após a crise de 2008. Contudo, a partir de 2013, a razão entre a dívida e o PIB apresenta um crescimento quase exponencial, chegando a 70% em dezembro de 2015. Ao persistir tal tendência, a situação fiscal torna-se insustentabilidade tendo em vista à queda da trajetória da receita provocada pela recessão econômica e manutenção do nível das despesas.

[4]São instrumentos derivativos feitos para trocar a rentabilidade de dois ativos. No caso de swaps cambiais feitos pelo BCB, a troca é feita através de leilões em oferta pública. O BC faz a troca da variação cambial pela taxa de juros, ou seja, o BC assume o risco da variação cambial enquanto o comprador do contrato de swapassume a variação da taxa de juros doméstica. BCB (2013).

4. ASPECTOS METODOLÓGICO DE MULTICOINTEGRAÇÃO COM QUEBRA ESTRUTURAL

O modelo de multicointegração caracteriza-se por testar a relação de equilíbrio entre variáveis de fluxo e de estoque para um determinado período. Portanto, essa metodologia é usada para identificar uma relação de sustentabilidade entre o padrão de comportamento das variáveis de fluxo (receita e despesa) e seus efeitos no equilíbrio com uma variável de estoque (dívida). Assim, a multicointegração é um avanço em relação aos modelos tradicionais de cointegração, pois permite avaliar a uma relação mais profunda, por exemplo, entre as receitas, as despesas e o estoque de dívida, mesmo com possibilidade de existência de quebras estruturais nessas séries temporais.

O modelo com uma única equação de multicointegração foi desenvolvido por Engsted et alii (1997). Eles incorporaram as variáveis integradas na ordem um e dois, ou seja, I(1) e I(2), as quais mantinham uma relação de equilíbrio de longo prazo, entre elas ou uma relação estável. Mais recentemente, Tronzano (2012) e Berenguer-Rico e Carrion-i Silvestre (2011) revisaram esse pressuposto para atender as mudanças estruturais das séries temporais. Tronzano (2012) testou essa metodologia para avaliar a sustentabilidade da política fiscal que consiste:

$$\begin{aligned} Y_t &= \beta_{00} + \beta_{01}t + \beta_{02}1(t > T_b) + \beta_{03}(t - T_b)1(t > T_b) + \alpha t^2 + \beta_{11}\Delta X_t \\ &+ \beta_{12}\Delta X_t 1(t > T_b) + \beta_{21}X_t + \beta_{22}X_t 1(t > T_b) + \mu_t \end{aligned} \tag{5}$$

Essa abordagem expande os modelos padrões de multicointegração para que incorporem quebras estruturais que afetam a relação de equilíbrio dos regressores determinísticos e estocásticos. Assim os coeficientes da equação (5) equivalem as suas definições originais.[5] O termo 1(.) indica o argumento da função, T_b [6] representa o ponto de quebra, os coeficientes β_{02} e β_{03} capturam, respectivamente, uma mudança nas condições de equilíbrio no nível e na tendência. Por sua vez, β_{12} e β_{22} representam, na ordem, uma mudança do efeito da variável fluxo (ΔX_t) e da variável estoque (X_t) sobre os gastos acumulados do governo (Y_t) e μ_t representa os resíduos.

O sistema de estoque e fluxo da equação (5) permite várias possibilidades de cointegração. Se μ_t for integrado na ordem dois, ou I(2) não haverá uma relação de equilíbrio no longo prazo. Se μ_t for I(1), as variáveis forem I(2), (X_t) e (Y_t) cointegram num processo de ordem I(1). Então esse processo é denominado por (Y_t), $(X_t) \sim CI(2,1)$, o qual implica que o governo mantém o fluxo de receita e despesa sob controle, porém, sem se preocupar com o estoque de dívida. Nesse caso, o estoque de dívida seria I(1).

Finalmente, se μ_t for I(0), então, prevalece uma relação de multicointegração na qual as variáveis cointegram num processo estacionário, denominado (Y_t), $(X_t) \sim CI(2,2)$. Nesse caso o governo manteria equilibrado as variáveis fluxo de receita e despesa e, ainda, interviria nas variáveis de fluxo, quando os choques externos tendem a mudar a trajetória do estoque de dívida para níveis fora de controle ou explosivo. O estoque de dívida passa a ser I(0). Além disso, passam a existir duas relações de cointegração em que a sustentabilidade fiscal é alcançada com o controle das duas variáveis de fluxos, receita e despesas e uma de estoque que a dívida.

Assim, seguindo Berenguer-Rico e Carrion-i Silvestre (2011) e Tronzano (2012), para cada quebra estrutural identificada na série é aplicado o teste ADF para os resíduos. Os modelos estimados estão reportados no Tabela 3, em que as especificações A e B não assumem quebra estrutural nos regressores estocásticos. O primeiro inclui uma constante e a tendência linear como regressores determinísticos e o segundo inclui a tendência quadrática.

Os demais modelos, C, D e E, permitem quebras estruturais nos regressores determinísticos e estocásticos. Objetivando a análise dos regressores estocásticos, o modelo com troca de regimes envolve

[5] Para especificações e valores críticos dos testes, veja-se Berenguer-Rico e Carrion-i Silvestre (2011).

[6] Aplicado teste de Bai (1997) e Bai e Perron (2003).

Tabela 3: Modelos estimados para cada troca de regimes.

Modelo	Parte Determinística	Parte Estocástica
Modelo $[A]$	$\alpha = 0$	$\beta_{12} = \beta_{22} = 0$
Modelo $[B]$	$\alpha \neq 0$	$\beta_{12} = \beta_{22} = 0$
Modelo $[C]$	$\alpha = 0$	$\beta_{12} = 0; \beta_{22} \neq 0$
Modelo $[D]$	$\alpha = 0$	$\beta_{12} \neq 0; \beta_{22} = 0$
Modelo $[E]$	$\alpha = 0$	$\beta_{12} \neq 0; \beta_{22} \neq 0$

Nota: Os coeficientes se referem à Equação (1). A troca de regime é observada na Tabela 5.

variáveis integradas na segunda ordem ou I(2), (X_t) no caso dos modelos C.E. e I(1) no modelo D, com (ΔX_t). O modelo E representa a especificação geral, sem restrições nos parâmetros e quebras estruturais, com as variáveis I(1) e I(2).

As series estatísticas utilizadas foram as receitas e as despesas e o estoque da dívida do governo central brasileiro cobrindo o período do primeiro trimestre de 1997 ao quarto trimestre de 2015, ajustada sazonalmente. Todas as séries foram coletadas junto a Secretaria do Tesouro Nacional, Ministério da Fazenda e ao Instituto Brasileiro de Geografia e Estatísticas – IBGE.

Para a realização dos testes econométricos foram utilizadas as séries de receita e despesa nominal em percentual do produto nominal, conforme Tronzano (2012). O emprego das variáveis fiscais expressas em percentual do produto nominal é sustentado pelo comportamento explosivo das séries em valores nominais ou reais. Goyal et alii (2004), Payne et alii (2008), Leachman e Francis (2000) e Leachman et alii (2005) aplicaram esse mesmo tratamento metodológico, demonstrando que um procedimento usual nesse tipo temática.

5. DESCRIÇÃO E ANÁLISE DOS RESULTADOS

As especificações do modelo de multicointegração com quebra estrutural, nas variáveis fiscais, para o Governo Central do Brasil, evidenciam que t_δ^*[7] não excede aos valores críticos para todos os modelos estimados como ilustra a Tabela 4. Esse resultado tende a ser robusto em relação a todos os modelos, desde o mais simples, como modelo A, ao mais completo ou Modelo E.

Tabela 4: Testes de multicointegração com quebra estrutural.

Modelo	$t_\delta^* = \inf t_\delta(\lambda)$- ADF	$t_\delta^* = \inf t_\delta(\lambda)$- Ng-Perron	5%*	10%*
Modelo $[A]$	-1,93	-2,08	-6,20	-5,93
Modelo $[B]$	-3,52	-5,80	-6,61	-6,32
Modelo $[C]$	-2,15	-2,22	-6,48	-6,20
Modelo $[D]$	-3,10	-3,21	-6,45	-6,16
Modelo $[E]$	-2,99	-3,07	-6,68	-6,38

*Refere-se aos valores críticos.

[7] $t_\delta^* = \inf t\delta(\lambda)$: é a estatística do teste de multicointegração desenvolvida por Berenguer-Rico e Carrion-i Silvestre (2011). C.V. significam valores críticos, conforme Berenguer-Rico e Carrion-i Silvestre (2011).

Os testes apontam que a hipótese nula dos resíduos serem I(1) não pode ser rejeitada para os dados fiscais para o Brasil. Isso implica que a variável de estoque é integrada de segunda ordem ou I(2) e apresenta uma relação de cointegração de primeira ordem (Y_t), $(X_t) \sim CI(2,1)$. No entanto, não há uma cointegração completa na qual as receitas (Y_t) e as despesas (X_t), ou seja, não são $CI(2,2)$. Nesse caso, o resultado demonstra a existência de uma relação simples de cointegração entre as variáveis fluxo e exclui a existência de uma relação mais profunda de equilíbrio entre as variáveis fluxo e estoque ou uma multicointegração com a presença de quebras estruturais nas séries.

As estimativas dos coeficientes para cada modelo de regressão de multicointegração com a presença de quebras estruturais estão reportadas na Tabela 5. As duas primeiras colunas apresentam os resultados dos regressores determinísticos e estocásticos. A última coluna relaciona a soma dos quadrados dos resíduos (SSR), o período utilizado para quebra (T_b) e os dois critérios de análise o de AIC e SBC.

Tabela 5: Estimativas dos modelos de muticointegração com quebra estrutural.

	Regressores determinísticos					Regressores Estocásticos							
Modelos	α	β_{00}	β_{01}	β_{02}	β_{03}	β_{11}	β_{12}	β_{21}	β_{22}	SSR	Tb	AIC	SBC
Modelo	-	0,14	0,01	0,04	-0,01	-0,87	-	0,82	-	0,09	2010/	-3,75	-3,57
$[A]$		(3,62)	(4,87)	(1,83)	(-5,43)	(-4,10)	(51,18)				Q3		
Modelo	-0,00	0,11	0,02	0,03	0,00	-0,79	-	0,74	-	0,07	2010/	-3,99	-3,77
$[B]$	(-4,54)	(3,09)	(6,31)	(1,54)	(0,75)	(-4,21)		(32,84)			Q3		
Modelo	-	0,14	0,01	0,05	-0,01	-0,84	-	0,81	0,00	0,09	2010/	-3,75	-3,54
$[C]$		(3,56)	(4,90)	(2,11)	(-5,45)	(3,96)		(45,55)	(-1,38)		Q3		
Modelo	-	0,06	0,01	0,37	-0,00	-0,44	-1,68	0,83	-	0,07	2010/	-3,94	-3,73
$[D]$		(1,46)	(4,74)	(-1,95)	(-2,00)	(4,08)	(56,75)				Q3		
Modelo	-	-0,06	0,01	0,37	-0,01	-0,43	-1,64	0,82	-0,00	0,07	2010/	-3,93	-3,69
$[E]$		(1,46)	(4,63)	(4,43)	(-2,00)	(-1,95)	(-3,96)	(50,28)	(-1,14)		Q3		

Nota: SSR é a soma dos quadrados dos resíduos, Tb, período utilizado para a quebra, AIC e SBS são critérios de informação Akaike e Schwarz, respectivamente.

A questão relevante a salientar é que o coeficiente β_{21} é altamente significativo e em torno de 0,82 para todos os modelos estimados, exceto para o Modelo B que é de 0,74. Isso significa que na média o crescimento das receitas acumuladas é significativamente menor do que as despesas acumuladas. Esse resultado evidencia a existência de uma sustentabilidade fiscal fraca. Nos modelos D e E, os coeficientes β_{01} e β_{02} evidenciam uma mudança na tendência da série após a quebra estrutural (terceiro trimestre de 2010), ou seja, a série apresenta uma mudança de nível de equilíbrio da tendência linear.

Os coeficientes estimados β_{11} e β_{12} ainda nos modelos D e E, suportam a hipótese de uma potencial troca de regime na variável I(1), ΔX_t, de coeficientes significativos e negativos em torno de -0,44 para -1,68. Este resultado é equivalente ao coeficiente K_1 do modelo padrão de multicointegração, no qual a sustentabilidade fiscal é assegurada somente se K_1 for negativo e significativo, demonstrando que as despesas excedem, em média, as receitas. A soma negativa dos coeficientes estimados ($\beta_{11} + \beta_{12}$), após o terceiro quadrimestre de 2010, revela uma significativa piora da situação fiscal do país, com uma possível troca de regime, com despesas aumentando em média acima da receita. Além disso, como a estatística t^*_δ é, em geral, menor do que os critérios de informações de Akaike e Schwarz, o que confirma uma sustentabilidade fiscal do governo central fraca, ao longo de 1997 a 2015.

O modelo de multicointegração com quebras estruturais demonstra, portanto, a ausência de estacionariedade dos resíduos ou que são integrados de ordem um, I(1). Isso significa dizer que há uma cointegração de primeira ordem entre receita e despesa acumuladas. Já a hipótese da existência de multicointegração com a presença de quebras estruturais não pode ser aceita. Por consequência, tal resultado evidencia a inexistência de uma relação de equilíbrio de longo prazo entre as variáveis fiscais de fluxo e de estoque para o Brasil. Esse fato revela que, período estudado, não ocorreu uma relação

de equilíbrio entre os déficits orçamentários e a evolução do estoque de dívida pública, corroborando, novamente, a presença de uma sustentabilidade fiscal bastante fraca para o governo central brasileiro.

Em síntese, o modelo de multicointegração aplicado para testar a sustentabilidade da política fiscal demonstra a inexistência de uma relação de equilíbrio entre receitas, despesas e estoque de dívida para o Brasil. Tal fato vem demostrar que o país enfrenta um profundo desequilíbrio das contas públicas, mantendo essa tendência, o setor governamental poderá entrar num processo *default* no longo prazo em virtude da prática de políticas fiscais com fraca sustentabilidade e a não existência de multicointegração entre as variáveis citadas. Portanto, medidas de equilíbrio das contas públicas no curto prazo não poderão suficientes para alcançar a sustentabilidade da política fiscal no longo prazo.

6. CONCLUSÕES

A dinâmica da sustentabilidade da política fiscal de uma nação tende a ser uma questão crucial condução da política econômica. O equilíbrio da relação entre as receitas, as despesas e o estoque de dívida bruta representa um indicador de confiança e de formação de expectativas para os agentes econômicos. Contudo, com adoção de políticas anticíclicas e os próprios ciclos políticos tendem a magnificar os gastos e, portanto, o comprometimento o equilíbrio intertemporal das contas públicas.

O governo brasileiro implementou várias medidas fiscais, pós-crise de 2008, que gerou mudanças na evolução das receitas e despesas públicas. A consequência foi uma deterioração nos déficits orçamentários. Mais recente, a necessidade de financiamento do setor público no conceito nominal passou de um déficit de R$ 343,9 bilhões, em 2014, para R$ 613,0 bilhões em 2015, tendo como um dos principais componentes os encargos sobre a dívida pública. Em proporção ao produto da economia brasileira que está em retração, o déficit nominal cresceu para 10,3%, pagamento dos juros, 8,5% e a dívida em torno de 70%.

O modelo de multicointegração com quebras estruturais reforça essa conclusão. Portanto a hipótese da existência de uma relação de equilíbrio de longo prazo as variáveis de fluxo e de estoque, ou seja, entre a receita, a despesa e o estoque de dívida do setor governamental não pode ser aceita. Esse resultado vem evidenciar que o governo não estaria adotando medidas corretivas adequadas dos gastos públicos em relação às receitas em um cenário de aumento acelerado do estoque de dívida. Em todos os modelos estimados o coeficiente β_{21} é altamente significativo, situando-se no intervalo de 0,82 a 0,74. Isso significa que na média o crescimento das receitas acumuladas é significativamente menor do que as despesas acumuladas. Essa conclusão é corroborada pela soma negativa dos coeficientes estimados ($\beta_{11} + \beta_{12}$), após o terceiro quadrimestre de 2010, revela uma significativa piora da situação fiscal do país, com uma possível troca de regime o que evidencia a existência de uma sustentabilidade fiscal fraca.

Por fim, o modelo de multicointegração aplicado para testar a sustentabilidade da política fiscal vem demostrar que o país está enfrentando um profundo desequilíbrio das contas públicas. Em caso de persistir essa tendência de aceleração no crescimento do endividamento, o setor governamental brasileiro poderá entrar num processo *default* num prazo mais longo.

BIBLIOGRAFIA

Afonso, A. (2005). Fiscal sustainability: unpleasant European case. *FinanzArchiv*, 61:19–44.

Arghyrou, M. G. & Luintel, K. B. (2007). Government solvency: revisiting some EMU countries. *Journal of Macroeconomics*, 29(2):387–410.

Bai, J. (1997). Estimation of a change point in multiple regression models. *The Review of Economics and Statistics*, 79(4):551–563.

Bai, J. & Perron, P. (2003). Critical values for multiple structural change tests. *The Econometrics Journal*, 6(1):72–78.

Bajo-Rubio, O., Díaz-Roldán, C., & Esteve, V. (2010). On the sustainability of government deficits: some long-term evidence for Spain, 1850–2000. *Journal of Applied Economics*, 13(2):263–281.

Berenguer-Rico, V. & Carrion-i Silvestre, J. L. (2011). Regime shifts in stock–flow I (2)–I (1) systems: the case of US fiscal sustainability. *Journal of Applied Econometrics*, 26(2):298–321.

Bertussi, L. A. S. & Triches, D. (2012). Multicointegração e Políticas Fiscais: Uma Avaliação de Sustentabilidade Fiscal para Argentina, Brasil, México, Peru, Uruguai e Venezuela. *Revista Economia*.

Bertussi, L. A. S. & Triches, D. (2013). Uma revisão da dinâmica macroeconômica da dívida pública e dos testes de sustentabilidade da política fiscal. *Revista Teoria e Evidência Econômica*, 19(40).

Blanchard, O. J., Chouraqui, J.-C., Hagemann, R., & Sartor, N. (1990). The sustainability of fiscal policy: New answers to an old question.

Camarero, M., Carrion-i Silvestre, J. L., & Tamarit, C. (2015). The relationship between debt level and fiscal sustainability in organization for economic cooperation and development countries. *Economic Inquiry*, 53(1):129–149.

Claeys, P. (2007). Sustainability of EU fiscal policies: A panel test. *Journal of Economic Integration*, pages 112–127.

Domar, E. D. (1944). The burden of the debt and the national income. *The American Economic Review*, 34(4):798–827.

Engsted, T., Gonzalo, J., & Haldrup, N. (1997). Testing for multicointegration. *Economics Letters*, 56(3):259–266.

Escario, R., Gadea, M. D., & Sabaté, M. (2012). Multicointegration, seigniorage and fiscal sustainability. Spain 1857–2000. *Journal of Policy Modeling*, 34(2):270–283.

Garcia, M. & Rigobon, R. (2004). A risk management approach to emerging market's sovereign debt sustainability with an application to Brazilian data. Technical report, National Bureau of Economic Research.

Goyal, R., Khundrakpam, J., & Ray, P. (2004). Is India's public finance unsustainable? Or, are the claims exaggerated? *Journal of Policy Modeling*, 26(3):401–420.

Greiner, A., Koeller, U., & Semmler, W. (2006). Testing the sustainability of German fiscal policy: evidence for the period 1960–2003. *Empirica*, 33(2):127–140.

Hatemi-J, A. & Zanella, F. (2013). Testing for the government's intertemporal budget restriction in Brazil during 1823–1889. *Applied Economics*, 45(12):1533–1540.

Leachman, L., Bester, A., Rosas, G., & Lange, P. (2005). Multicointegration and sustainability of fiscal practices. *Economic Inquiry*, 43(2):454–466.

Leachman, L. L. & Francis, B. B. (2000). Multicointegration analysis of the sustainability of foreign debt. *Journal of Macroeconomics*, 22(2):207–227.

Legrenzi, G. & Milas, C. (2012). Nonlinearities and the sustainability of the government's intertemporal budget constraint. *Economic Inquiry*, 50(4):988–999.

Luporini, V. (2012). *Sustainability of Brazilian fiscal policy, once again: corrective policy response over time*. XL Encontro Nacional de Economia, ANPEC, Porto de Galinhas, Pernambuco.

Nikolov, I. A., Green, C. J., & Vasilev, D. (2010). Fiscal Sustainability Convergence: how the crisis caught 'old' Europe Unprepared.

Pastore, A. C. (1995). Déficit Público, a Sustentabilidade do Crescimento das Dívidas Interna e Externa, Senhoriagem e Inflação: Uma Análise do Regime Monetário Brasileiro. *Brazilian Review of Econometrics*, 14(2):177–234.

Payne, J. E., Mohammadi, H., & Cak, M. (2008). Turkish budget deficit sustainability and the revenue-expenditure nexus. *Applied Economics*, 40(7):823–830.

Rocha, F. (1997). Long-run limits on the Brazilian government debt. *Revista Brasileira de Economia*, 51(4):447–470.

Sachs, J. D. & Larrian, F. B. (2000). *Macroeconomia*. Preason, São Paulo.

Silva, R. d. & Gamboa, U. R. d. (2011). Regime fiscal e sustentabilidade da dívida pública brasileira–1986 a 2006. *Revista Gestão & Políticas Públicas*, 1(1).

Simonassi, A. G. & Arraes, R. (2007). Função de resposta fiscal, múltiplas quebras estruturais e a sustentabilidade da dívida pública no Brasil. *Anais do Encontro Nacional de Economia da ANPEC, Recife-PE*.

Simonassi, A. G., Arraes, R. A., & de Sena, A. M. C. (2014). Fiscal reaction under endogenous structural changes in Brazil. *Economia*, 15(1):68–81.

Trachanas, E. & Katrakilidis, C. (2013). Fiscal deficits under financial pressure and insolvency: Evidence for Italy, Greece and Spain. *Journal of Policy Modeling*, 35(5):730–749.

Trachanas, E. & Katrakilidis, C. (2014). Is the Greek budget deficit sustainable after all? Empirical evidence accounting for regime shifts. *Applied Economics*, 46(20):2387–2397.

Triches, D. (2005). Uma análise de economia política e das atitudes dos grupos de interesse no Mercosul. pages 157–176, Caxias do Sul, RS, Educs. Economia regional e integração internacional.

Triches, D. & Morais, I. A. (2005). Déficit público e taxa de inflação: teste de raiz unitária e causalidade para o Brasil de 1991-1999. pages 196–210, Caxias do Sul, RS, Educs. Economia regional e integração internacional.

Tronzano, M. (2012). *Multicointegration and Fiscal Sustainability in India: Evidence from Standard and Regime Shifts Models*. Universitá degli Studi di Genova, Facoltà di Economia, Dipartimento di Economia e Metodi Quantitativi.

Uctum, M., Thurston, T., & Uctum, R. (2006). Public Debt, the Unit Root Hypothesis and Structural Breaks: A Multi-Country Analysis. *Economica*, 73(289):129–156.

Walsh, C. E. (1998). *Monetary theory and policy*. The MIT press, Cambridge, Massachusetts.

Westerlund, J. & Prohl, S. (2008). Panel Cointegration Tests of the Sustainability Hypothesis in Rich OECD Countries. *Applied Economics*, pages 1–10.

A. SÍNTESE DOS ESTUDOS EMPÍRICOS E MODELOS APLICADOS

Referência	Período/Países	Modelo	Frequência	Resultado
Afonso (2005)	1968-1997 (EU-15)	Testes raiz unitária Cointegração	Dados anuais	Testes indicaram ausência de sustentabilidade com exceção da Alemanha, Áustria e Holanda.
Arghyrou e Luintel (2007)	1970:1-1998:3 (Grécia) 1962:2-1997:4 (Itália) 1957:1-1998:4 (Irlanda)	Testes raiz unitária e Cointegração com e sem quebra estrutural	Dados trimestrais	Testes indicaram sustentabilidade da política fiscal para diferentes horizontes do tempo e os desequilíbrios fiscais se ajuntam de forma não linear.
Bajo-Rubio et alii (2010)	1850-2000 (Espanha)	Teste raiz unitária Cointegração	Dados anuais	Déficit fiscal tem sido fortemente sustentável no longo prazo. Não há evidência de significativas quebras estrutural ao longo do período. A sustentabilidade fiscal foi mantida pelo comportamento não linear do governo que agia quando o déficit excedia a 4,5% do PIB.
Claeys (2007)	1970-2001 (EU-15 sem Luxemburgo)	Testes raiz unitária e Cointegração	Dados anuais	Testes indicaram sustentabilidade para o painel e sem sustentabilidade para alguns países individuais.
Garcia e Rigobon (2004)	1994-2003 (Brasil)	VAR e Simulação Monte Carlos	Dados mensais	O déficit público poderia ser considerável sustentável na ausência de risco. Contudo ele apresenta uma trajetória é claramente insustentável. A dinâmica da dívida está proximamente relacionada com *spread* do EMBI+ Brasil.
Greiner et alii (2006)	1964-2001 (Alemanha)	Semi-paramétrico	Dados trimestrais	Dívida pública é sustentável. A sustentabilidade da política fiscal está na dependência de como o governo reage ao aumento da razão dívida/PIB no futuro.
Hatemi-J e Zanella (2013)	1823-1889 (Brasil)	Testes de Cointegração com quebra estrutural desconhecida	Dados anuais	As despesas e receitas do governo são cointegradas. Levando-se em conta as duas quebras estruturais desconhecidas o governo foi solvente no longo prazo particularmente no período imperial.
Legrenzi e Milas (2012)	1960-2009 (Itália)	Modelo não linear com correção de erro	Dados mensais	Testes indicaram um ajustamento não linear mas sustentável no longo prazo.
Luporini (2012)	1991-2011 (Brasil)	Testes raiz unitária e Cointegração	Dados mensais	O parâmetro da reação fiscal e a taxa crescimento da relação dívida/PIB indicam que a estabilidade dessa relação é altamente dependente da taxa de crescimento da economia dada a meta do governo com o superávit primário.
Nikolov et alii (2010)	1990-2008 (EU-14)	Cointegração	Dados anuais	Testes indicaram ausência sustentabilidade da política fiscal.
Simonassi et alii (2014)	1991-2008 (Brasil)	Cointegração	Dados mensais	Testes indicaram quebras estruturais em maio de 1994 e em fevereiro de 2003. A solvência da situação fiscal somente ocorre a partir de maio de 1994.
Trachanas e Katrakilidis (2013)	1970-2010 (Espanha, Grécia e Itália)	Cointegração com quebra e com efeito linear não linear	Dados anuais	Testes indicaram sustentabilidade fiscal fraca no longo prazo para os três países. O processo de ajustamento orçamentário é assimétrico na Itália e na Espanha.
Trachanas e Katrakilidis (2014)	1960-2011 (Grêcia)	Raiz unitária e cointegração com quebra e com efeito linear não linear	Dados anuais	Testes indicaram a presença de quebra estrutural e o déficit orçamentário grego é intentável.
Tronzano (2012)	1960-2011 (Índia)	Multicointegração com mudança de regime	Dados anuais	Testes rejeita a multicointegração entre a receitas e despesas a política fiscal do governo indiano não sustentável num ambiente estocástico.
Uctum et alii (2006)	1970-2002 (G7 e países selecionados da América Latina e da Ásia)	Teste de raiz unitária	Dados anuais	Testes de raiz unitária não capta as ações corretivas tomadas pelos governos. Com controle das quebras estruturais altera os resultados de não estacionariedade. A função reação do governo com quebra estruturais fornece evidência de os governos de G7 são mais propensos de controlar a dívida do que as demais regiões.
Westerlund e Prohl (2008)	(Oito países da OCDE)	Cointegração com quebra	Dados trimestrais	Testes indicaram sustentabilidade da política fiscal.

Nota: Nikolov (2010) traz outros estudos sobre a sustentabilidade fiscal dos países europeus.