Preços-Sombra no Sistema de Pagamentos: Uma Abordagem Dual para a Política Monetária Intradiária*

Rodrigo Andrés de Souza Peñaloza**

Palavras-chave: Banco Central; risco sistêmico; políticas monetárias; sistema de pagamentos; dualidade; preços-sombra.

Códigos JEL: C61; E51; E52; E58.

O funcionamento do sistema de liqüidação pelo valor bruto em tempo real é modelado como uma otimização na qual enfileiramento e fracionamento de pagamentos e acordos de recompra surgem como soluções primais. O problema dual associado à maximização do fluxo de pagamentos é então usado para a determinação dos preços-sombra dos bancos no sistema de pagamentos. Esses preços-sombra podem ser usados para a personalização das políticas monetárias intradiárias tais como requerimentos de reserva inicial, acordos de recompra, extensão bilateral de crédito no mercado interbancário intradiário, etc. de modo a tornar eficiente o uso da liqüidez sistêmica.

We model the functioning of real-time gross settlement systems for large-value transfers as a linear programming problem in which queueing arrangements, splitting of payments, and Lombard loans arise as primal solutions. Then we use the dual programming problem associated with the maximization of the total flow of payments in order to determine the shadow-prices of banks in the payment system. We use these shadow-prices to set personalized intraday monetary policies such as reserve requirements, availability of Central Bank credit to temporarilly illiquid banks, extension of

*Artigo recebido em nov. 2004 e aprovado em jun. 2005. Agradeço a Joe Ostroy (UCLA) pelas longas e frutíferas conversas que tivemos durante os anos de minha estadia na UCLA, bem como aos seguintes membros da UCLA Economic Theory Center: David Levine, Bill Zame, Alberto Bennardo e David Rahman. Uma versão anterior deste trabalho circulou com o título "On shadow-prices of banks in real-time gross settlement systems", texto para discussão #71, Banco Central do Brasil, e também nos Anais do XXV Encontro da Sociedade Brasileira de Econometria, EBE 2003.

**Universidade de Brasília, Departamento de Economia - FACE. E-mail: penaloza@unb.br

intraday interbank credit exposures, etc., so as to make the use of systemic liquidity more efficient.

1. Introdução

Neste artigo apresentamos um modelo de determinação de preços-sombra para os parâmetros de política monetária intradiária no sistema de pagamentos. Chamamos de políticas monetárias intradiárias os seguintes parâmetros: reservas iniciais no Banco Central, acordos de recompra (também chamados de empréstimos Lombard), extensão do *haircut* sobre esses empréstimos, extensão da exposição bilateral em um mercado interbancário no intradia[1] e possibilidades de *overdraft,* como, por exemplo, os *net debit caps* do sistema FEDWIRE americano.

As compensações dos pagamentos interbancários mediante transferências monetárias de grandes valores das contas-reserva dos bancos no Banco Central podem ocorrer, *grosso modo*, de acordo com dois arranjos institucionais distintos: sistema de liqüidação defasada pelo valor líqüido (sistema LDL) ou pelo sistema de liqüidação bruta em tempo real (sistema LBTR). No sistema LDL todas as transferências vão sendo anotadas e no final do dia os débitos líqüidos são pagos. No sistema LBTR as transferências são compensadas pelo seu valor bruto no momento em que chegam. O sistema LDL é mais econômico em termos de liqüidez para as compensações interbancárias, pois os bancos pagam apenas os débitos líqüidos, mas é claramente mais propenso ao risco de liqüidez devido à defasagem de tempo entre o envio da ordem de pagamento e sua liqüidação. Dadas as características de rede do sistema de pagamentos, o risco sistêmico inerente ao sistema LDL pode ser bastante elevado.[2] Já o sistema LBTR pode reduzir consideravelmente a exposição ao risco sistêmico, pois os pagamentos ocorrem em tempo real. Porém, é obviamente mais custoso em termos de liqüidez. Com efeito, de acordo com o *Lamfalussy Report* (Bank for International Settlements (1997:9)), o tamanho e a duração da exposição aos riscos de liqüidez e de crédito são os fatores básicos para o potencial aumento do risco sistêmico.

Apesar desse evidente *trade-off* entre os dois arranjos institucionais, a substituição do sistema LDL pelo LBTR tem sido uma tendência mundial, incluindo-se aí os países europeus membros e não-membros da União Européia, vários países

[1]Uso indistintamente os termos "intradiário" e "no intradia", por julgar que essa distinção é irrelevante, da mesma forma como é irrelevante, por exemplo, a distinção entre "marginal" e "na margem".

[2]A exposição ao risco no sistema LDL pode ser reduzido pela utilização de *clearings* que operem com garantias (agradeço a um parecerista anônimo por notar a relevância desse fato).

asiáticos e latino-americanos. No Brasil, o sistema LBTR no Sistema de Transferência de Reservas (STR) – conhecido como o *Novo SPB* – é vigente desde de maio de 2002. Essa tendência é conseqüência das recomendações do Banco de Compensações Internacionais (BIS) na decada de 1990, após quase vinte anos de estudos sobre as boas práticas de segurança nos sistemas de pagamentos.

É economicamente relevante, portanto, saber em que medida o sistema LBTR pode se tornar mais econômico em termos de liqüidez sem, contudo, perder suas características definidoras. Dentre as soluções encontradas estão os sistemas de enfileiramento de mensagens de pagamentos para os quais inexiste saldo suficiente, os acordos de recompra intradiários, ou seja, empréstimos do Banco Central mediante títulos que sirvam de garantia, e a criação de um mercado interbancário intradiário, sob cuja vigência o Banco Central abster-se-ia de sua função de emprestador de última instância em prol de uma participação ativa dos próprios bancos na provisão de liqüidez intradiária.

A literatura sobre o tema concentra-se nos efeitos do desenho do sistema de pagamentos sobre o comportamento estratégico dos bancos (ver DeBandt e Hartmann (2000) para um *survey*). Para citar dois efeitos importantes, há o problema do *free-rider,* em que um banco adia seus pagamentos esperando receber liqüidez de outros bancos, o que é socialmente ineficiente (Bech e Garrat, 2003), e existe também o problema de saber qual dois sistemas é socialmente eficiente na presença de informação assimétrica por parte dos bancos (Freixas e Parigi, 1998).

Nosso modelo faz o caminho inverso. Dado o padrão de pagamentos interbancários intradiários e dado que o sistema deve ser do tipo LBTR, qual o melhor desenho que o Banco Central pode impor ao sistema de pagamentos? Para podermos responder a essa pergunta usamos os preços-sombra. O sistema de pagamentos é encarado como uma grande rede na qual pagamentos fluem ao longo do dia. A linearidade surge naturalmente das características de rede do sistema. Nossa formulação primal do sistema LBTR é original, inexiste na literatura. O desenho ótimo é, assim, a solução primal. Entretanto, a novidade maior de nosso modelo é o uso da solução dual para apreçamento das políticas monetárias intradiárias. É nesse aspecto que reside a segunda contribuição teórica do modelo. Com ele é possível encarar a gestão de liqüidez intradiária por parte da autoridade monetária como um problema microecômico. O sistema primal-dual do modelo permite-nos responder a outras perguntas importantes para o Banco Central, algumas datando inclusive de Bagehot (1873):

- Quando é que um aumento do requerimento de reservas melhora o fluxo de pagamentos?
- Será uma boa idéia extender crédito intradiário grátis para bancos ilíqüidos?
- Vale a pena permitir empréstimos *overnight* entre bancos?
- Empréstimos Lombard realmente melhoram o fluxo de pagamentos?
- Empréstimos Lombard podem ser alocados otimamente – e com racionamento – ou será que o Banco Central deve estendê-los sempre que sejam requisitados?
- Existe um mecanismo de enfileiramento ótimo que minimiza as necessidades de liqüidez sistêmica?
- Pode um mercado interbancário intradiário substituir o Banco Central em seu papel de único provedor de liqüidez intradiária?
- Como a falência de um banco pode afetar o fluxo de pagamentos?

Respostas a essas perguntas têm implicação imediata na condução eficiente da política monetária intradiária. O Banco Central estipula, por exemplo, que uma fração dos depósitos à vista sejam retidos no Banco Central sob a forma de reservas. A função dessas reservas é dupla: primeiro, é fonte natural de liqüidez para o fluxo suave dos pagamentos interbancários; segundo, tem o óbvio objetivo de ajustar a liqüidez do mercado para a estabilização do nível de preços na economia no médio e longo prazos. Sob esta última função jazem algumas ineficiências. Embora seja ótimo do ponto de vista macroeconômico o ajustamento da fração dos depósitos retida no Banco Central, pode ser sub-ótimo do ponto de vista da gestão intradiária de liqüidez exigir que todos os bancos cumpram as mesmas exigências, como, por exemplo, a mudança do compulsório de 45% para 60% dos depósitos à vista e alguns meses depois de volta aos 45%. Por que não exigir que um banco retenha 45% de seus depósitos e que outro retenha 20%? A teoria econômica reza que os agentes econômicos (neste caso, os bancos) devem ser tratados de forma personalizada, de modo que cada agente internalize as externalidades de rede que sua presença causa no mercado. Em outras palavras, cada um é responsável por seu produto marginal. A despersonalização da política monetária intradiária (ou seja, a não internalização das externalidades) pode gerar um mau uso da liqüidez do sistema, mau uso este se refletindo na existência de liqüidez que acaba por não ser usada para finalização dos pagamentos. O desenho do sistema de pagamentos pode ser rígido o suficiente para que certos pagamentos não possam ser liqüidados, mesmo que exista no sistema o montante monetário necessário para tal. Com efeito, a simples determinação de regras de enfileiramento como *first-in-first-out* com graus de prioridade, por mais que permita grande fluidez dos pagamentos na prática, é "potencialmente" restritiva. O custo de oportunidade desse dinheiro inativo representa o custo privado arcado pelos bancos dentro do sistema de pagamentos, o que chamaremos de custo de liqüidez sistêmica. Com essa filosofia em mente, um dos subprodutos de nosso modelo é a determinação das contribuições marginais de cada banco (para cada período do dia e para cada restrição do sistema LBTR) para o custo mínimo de liqüidez sistêmica. Os preços-sombra podem assim ser usados para personalizar as políticas monetárias intradiárias de modo a zerar o custo mínimo de liqüidez sistêmica.

Para darmos uma visão intuitiva do *framework* de nosso modelo, considere os seguintes problemas de programação linear:

$$(\mathcal{P})\left\{\begin{array}{ll}\max & \mathbf{c}'\mathbf{y}\\ s.a & \mathbf{A}\mathbf{y}\leqq\mathbf{b}\\ & \mathbf{y}\geqq\mathbf{0}\end{array}\right. \qquad (\mathcal{D})\left\{\begin{array}{ll}\min & \mathbf{b}'\sigma\\ s.a & \mathbf{A}'\sigma\geqq\mathbf{c}\\ & \sigma\geqq\mathbf{0}\end{array}\right.$$

O problema $(\mathcal{D})$ é o dual de $(\mathcal{P})$, que é dito, por sua vez, o primal. A solução σ^* para $(\mathcal{D})$ é o vetor de preços-sombra dos parâmetros $\mathbf{b}$ do primal. O que eles medem? Suponha que a função-objetivo primal é o fluxo de pagamentos no STR e que $\mathbf{y}^*$ é a solução do primal. Sob certas condições, $\mathbf{c}'\mathbf{y}^*=\mathbf{b}'\sigma^*$. Uma das coordenadas de $\mathbf{b}$ pode ser o valor que o Banco Central atribui a um certo título possuído por um banco e custodiado no SELIC. Se o correspondente preço-sombra é igual a 2, por exemplo, então se o Banco Central aumentar essa valoração em \$1 – reduzindo assim o *haircut* –, o fluxo máximo de pagamentos aumentará de \$2. Essa é a contribuição marginal desse parâmetro. Se o Banco Central estimar esses preços-sombra, então será capaz de saber se o *haircut* de cada título é excessivo ou não, se deve reduzi-lo ou subi-lo e ordenar essas variações.

A seção 2 apresenta o modelo descritor do sistema LBTR. A seção 3 apresenta o problema primal do Banco Central, especificando sua função-objetivo e as restrições a serem atendidas. A seção 4 descreve o problema dual, mostra em que sentido o valor econômico de um pagamento interbancário difere de seu valor contábil e apresenta a equação para eficiência de liqüidez sistêmica. A seção 5 mostra como usar os preços-sombra para a determinação da política monetária intradiária. Na seção 6 apresentamos exemplos que ilustram a aplicabilidade de nosso modelo. Finalmente, a seção 7 conclui o artigo.

2. O Sistema de Liqüidação Bruta em Tempo Real

Nesta seção apresentamos o modelo de compensações ótimas de pagamentos interbancários em um sistema de liqüidação bruta em tempo real (LBTR).

As instituições que participam do sistema LBTR são ditas *bancos*. Denotamos por $\mathbf{B} = \{1, ..., n\}$ o conjunto dos bancos participantes do LBTR. Cada banco mantém uma conta reserva no Banco Central que será usada para as transferências de fundos ao longo do dia. Denotamos por B_o^i o saldo inicial de reservas do banco i.

O dia será dividido em um número finito, $K+1$, de períodos, $\mathbf{T} = \{t_o < t_1 < \cdots < t_K\}$, em que t_o denota o começo do dia e $t_K \equiv T$ o período de encerramento. Cada período $t \in \mathbf{T}$ representa o instante em que as compensações ocorrem, ou seja, o instante em que uma transferência de fundos é efetivada. O intervalo $\Delta t_k = (t_{k-1}, t_k]$ é interpretado como o intervalo de tempo dentro do qual uma mensagem de transferência de fundos foi enviada e processada. Assim, quando dissermos que uma compensação ocorreu no instante $t = t_k$, estaremos dizendo que a mensagem que lhe deu origem e o seu processamento (verificação sintática, criptográfica, etc.) deram-se no intervalo $\Delta t_k = (t_{k-1}, t_k]$.

O valor monetário da transferência de fundos do banco i para o banco j no período τ é denotado por $x_{ij}(\tau)$. Como um banco não transfere fundos para si mesmo em qualquer instante, temos que $x_{ii}(\tau) \equiv 0$, $\forall i \in \mathbf{B}$, $\forall t \in \mathbf{T}$.

Uma das variáveis de escolha do sistema LBTR é a fração $\upsilon_{ij}(\tau, t)$ do pagamento $x_{ij}(\tau)$ que será compensada no período $t \geqslant \tau$, a qual será dita simplesmente uma compensação. De fato, uma mensagem de pagamento pode ser enviada em um dado momento, mas sua compensação pode ocorrer em um momento futuro do dia. É assim, por exemplo, quando uma mensagem é enfileirada por insuficiência de fundos. Podemos impor as seguintes restrições fundamentais sobre as compensações, $\forall i, j \in \mathbf{B}$:

(a) $\sum_{t \geqslant \tau} \upsilon_{ij}(\tau, t) \geqslant 0$, $\forall \tau \in \mathbf{T}$

(b) $\upsilon_{ij}(\tau, t) \equiv 0$, $\forall \tau > t$, $\forall t \in \mathbf{T}$

(c) $\sum_{t \geqslant \tau} \upsilon_{ij}(\tau, t) \leqslant 1$, $\forall \tau \in \mathbf{T}$

A condição (a) diz que pelo menos alguma fração do pagamento $x_{ij}(\tau)$ tem que ser compensada até o final do dia. A condição (b) diz que um pagamento que ainda não foi enviado não pode ser compensado. A condição (c) diz que a transferência interbancária de fundos associada ao pagamento $x_{ij}(\tau)$ não pode ser maior do que o próprio valor do pagamento.

Temos ainda restrições que não são fundamentais, no sentido de que podem variar conforme as características legais ou operacionais do banco central que opera as transferências de fundos:

(d) $\upsilon_{ij}(\tau,t) \equiv 0$, $\forall \tau < t$, $\forall t \in \mathbf{T}$

(e) $\upsilon_{ij}(\tau,t) \in \{0,1\}$

(f) $\upsilon_{ij}(\tau,t) \in [0,1]$

(g) $\upsilon_{ij}(\tau,t) \in [a_{ij}, b_{ij}]$, onde $a_{ij} < 0$ e $b_{ij} > 1$

(h) $\upsilon_{ij}(\tau,t) \in \mathbb{R}$

(i) $\sum_{t=\tau}^{t^*} \upsilon_{ij}(\tau,t) \leqslant 1$, $t^* < t_K$

A condição (d) diz que a mensagem de pagamento não pode ser enfileirada. A compensação deve ocorrer no instante $\tau = t$ em que a mensagem é enviada. A condição (e) diz que uma mensagem pode ser enfileirada, mas a transferência de fundos não pode ser fracionada. A condição (f) diz que uma mensagem pode ser enfileirada e que a transferência de fundos pode ser fracionada. A condição (g) descreve o mercado interbancário intradiário. Suponha, por exemplo, que $x_{ij}(\tau) = \$100$ e que $\upsilon_{ij}(\tau,t) = 1,1$. Então o banco i ordena uma transferência de fundos para o banco j no periodo τ para compensação no período $t \geqslant \tau$, mas a transferência efetiva envolve um montante adicional de \$10, interpretado como um empréstimo intradiário do banco i para o banco j. Note que as condições (a) e (c) juntas garantem que o empréstimo descrito pela condição (g) é quitado ainda dentro do mesmo dia. Os limites $a_{ij} < 0$ e $b_{ij} > 1$ dizem que a exposição de crédito interbancário intradiário é limitada. Esses limites estão sob controle do Banco Central e, como tais, podem ser apreçados por seus preços-sombra. A condição (h) também descreve a possibilidade de mercado interbancário intradiário, porém sem quaisquer limites à exposição bilateral de crédito. A condição (i) diz que o pagamento $x_{ij}(\tau)$ enviado no período τ é temporalmente crítico, isto é, sua compensação total deve dar-se com finalidade até o periodo t^*, antes do fim do dia.

O Banco Central possui uma linha de crédito para empréstimos Lombard aos bancos. Seja $M_i \geqslant 0$ o montante de crédito que o Banco Central torna disponível para o banco i, no caso de ele requerer empréstimos Lombard durante o dia. Em caso de necessidade de liqüidez intradiária no período t, o banco i requer um empréstimo $\pi_i(t)$ do Banco Central oferecendo títulos como garantia e com o compromisso de recompra até o fim do dia. O Banco Central determina o *haircut* sobre o valor dos títulos, o que equivale a uma taxa de juros sobre o empréstimo Lombard. Denotamos o preço do empréstimo Lombard feito no período t por $r_i(t)$. Como essa taxa de juros é um parâmetro do modelo, podemos determinar seu preço-sombra. Os títulos adquiridos pelo Banco Central são ativos financeiros que serão, então, apreçados pelos preços-sombra. A manutenção de reservas no Banco Central é custosa por duas razões: (a) elas não são remuneradas e, (b) para evitar iliqüidez, os bancos têm que contar com empréstimos Lombard que são apreçados por meio de *haircuts* sobre os valores das garantias.

3. Alocação Ótima de Liqüidez no Sistema de Pagamentos

3.1 Custo de liqüidez sistêmica

De acordo com o BIS (ver Bank for International Settlements (1997)), a liqüidez líqüida agregada intradiária em qualquer instante do dia é dada pelas reservas totais menos a soma de todos os pagamentos que têm de ser liqüidados naquele instante. Em nosso modelo, essa definição é refinada de modo a incorporar tanto o papel do enfileiramento como o fluxo acumulado de pagamentos. A isso chamamos *custo de liqüidez sistêmica*. Ela mede o montante de dinheiro que permanece "parado" no sistema. Em outras palavras, é o montante monetário que está no sistema mas não está sendo usado para fins de compensação. O custo de oportunidade da liqüidez sistêmica é o que o setor bancário perde por participar do sistema LBTR.

O que queremos é determinar o esquema de compensações e o perfil de empréstimos Lombard ao longo do dia para cada banco de modo a minimizar o custo de liqüidez sistêmica:

$$\Lambda = \underbrace{\sum_{i=1}^{n} B_o^i}_{reservas\ totais\ iniciais} - \underbrace{\frac{1}{T}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=t_o}^{t_K}\sum_{\tau=t_o}^{t} x_{ij}(\tau)v_{ij}(\tau,t)}_{fluxo\ total\ acumulado\ (sobre\ os\ bancos)\ médio\ (sobre\ o\ tempo)}$$

Como as reservas iniciais são dadas no início do dia, a função objetivo é o fluxo total de pagamentos que podem ser compensados de acordo com a regra de compensação existente:

$$\Phi = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=t_o}^{t_K}\sum_{\tau=t_o}^{t} x_{ij}(\tau)\upsilon_{ij}(\tau,t)$$

3.2 Restrições do sistema LBTR

Há três conjuntos de restrições.

LIQ *Restrições de liqüidez.* No instante t, pagamentos que saem do banco i não podem exceder seu balanço corrente, isto é, seu balanço inicial mais as transferências líqüidas até o instante imediatamente anterior:

$$\begin{aligned}\sum_{j}\sum_{\tau\leqslant t} x_{ij}(\tau)\upsilon_{ij}(\tau,t) \leqslant B_o^i + \alpha_t\pi_i(t) + \\ +\sum_{s<t}\sum_{\tau\leqslant s}\sum_{j} x_{ji}(\tau)\upsilon_{ji}(\tau,s) - \sum_{s<t}\sum_{\tau\leqslant s}\sum_{j} x_{ij}(\tau)\upsilon_{ij}(\tau,s)\end{aligned}$$

em que:

$$\alpha_t = \begin{cases} +1 & \text{if } t<T \\ -1 & \text{if } t=T \end{cases}$$

CRE *Restrições de crédito.* O Banco Central pode emprestar até M_i ao banco i em troca de garantias. Em cada período o banco i pode requerer um empréstimo Lombard sob alguma taxa de juros. Tais linhas de crédito reduzem-se à medida em que empréstimos vão sendo tomados. No fim do dia, os empréstimos devem ser pagos (o banco i recompra suas garantias):

$$\begin{aligned}\pi_i(t_o) &\leqslant M_i \\ \pi_i(t_k) &\leqslant M_i - \sum_{\ell=0}^{k-1}\pi_i(t_\ell) \\ \pi_i(t_K) &\geqslant \sum_{\ell=0}^{K-1}(1+r_i(t_\ell))\pi_i(t_\ell)\end{aligned}$$

CON *Restrições de consistência.* Qualquer que seja a forma como um pagamento $x_{ij}(\tau)$ é enfileirado, parte dele tem que ser compensada até o fim do dia:

$$0 \leqslant \sum_{t \geqslant \tau} \upsilon_{ij}(\tau, t) \leqslant 1$$

Além disso, $\upsilon_{ij}(\tau, t) \equiv 0$, $\forall \tau > t$, o que significa que um pagamento que ainda não foi enviado não pode ser enfileirado. Note que as restrições de consistência são apenas as restrições fundamentais (a), (b) e (c) sobre as compensações.

3.3 O problema primal do Banco Central

As variáveis de escolha do Banco Central no sistema LBTR são (i) o mecanismo de enfileiramento representado pelas compensações, ou seja, $\{\upsilon_{ij}(\tau, t)\}$, e (ii) o perfil de empréstimos Lombard $\{\pi_i(t)\}$.

O Banco Central quer encontrar o mecanismo de enfileiramento e a seqüência de empréstimos Lombard de modo a minimizar o custo de liqüidez sistêmica sujeito às restrições de liqüidez, de crédito e de consistência:

$$\begin{cases} \text{MIN} & \text{custo de liqüidez sistêmica } (\Lambda) \\ \text{sujeito a} & \text{LIQ, CRE e CON} \end{cases}$$

Alternativamente, o Banco Central quer maximizar o fluxo total de pagamentos sujeito às mesmas restrições:

$$\begin{cases} \text{MAX} & \text{fluxo total de pagamentos } (\Phi) \\ \text{sujeito a} & \text{LIQ, CRE e CON} \end{cases}$$

Portanto, o problema primal do Banco Central é:

$$\left\{\begin{array}{lll}
\max_{\{\upsilon_{ij}(t,\tau),\pi_i(t)\}} & & \sum_{i\in\mathbf{B}}\sum_{j\in\mathbf{B}}\sum_{t\in\mathbf{T}}\sum_{\tau\leqslant t} x_{ij}(\tau)\upsilon_{ij}(\tau,t) \\
\quad \textit{sujeito a} & (i) & \sum_j x_{ij}(t_o)\upsilon_{ij}(t_o,t_o) \leqslant B_o^i \\
& (ii) & \sum_j x_{ij}(t)\upsilon_{ij}(t,t) + \sum_{\tau<t}\sum_j x_{ij}(\tau)\upsilon_{ij}(\tau,t) \leqslant B_o^i \\
& & +\sum_{\tau<t}\sum_{s\leqslant\tau}[\sum_j x_{ji}(s)\upsilon_{ji}(s,\tau) \\
& & -\sum_j x_{ij}(s)\upsilon_{ij}(s,\tau)] + \alpha_t\pi_i(t),\ t\in\mathbf{T}\backslash\{t_o\} \\
& (iii) & 0 \leqslant \sum_{\tau\geqslant t}\upsilon_{ij}(t,\tau) \leqslant 1, \forall t\in\mathbf{T}, \forall i,j\in\mathbf{B} \\
& (iv) & 0 \leqslant \upsilon_{ij}(\tau,t) \leqslant 1 \\
& (v) & 0 \leqslant \pi_i(t_o) \leqslant M_i \\
& (vi) & \pi_i(t_k) \leqslant M_i - \sum_{\ell=0}^{k-1}\pi_i(t_\ell), \quad 1\leqslant k\leqslant K-1 \\
& (vii) & \pi_i(t_K) \geqslant \sum_{\ell=0}^{K-1}(1+r_i(t_\ell))\pi_i(t_\ell) \\
& (viii) & \pi_i(t) \geqslant 0, \forall i\in\mathbf{B}, \forall t\in\mathbf{T}
\end{array}\right.$$

A solução primal $\{\upsilon_{ij}^*(\tau,t), \pi_i^*(\tau) : t \geqslant \tau,\ \tau = t_o, \ldots, t_K\}$ é uma regra de compensação e uma alocação seqüencial de empréstimos Lombard que resolvem o problema acima.

Após a introdução de algumas notações vetoriais, será possível reescrevermos o problema primal de maneira compacta e simples.

Seja $\mathbf{0}_n = (0, ..., 0)$ o vetor nulo de $\mathbb{R}^n$. Para cada banco $i \in \mathbf{B}$, defina a matriz:

$$\mathbf{X}_i(t) = \begin{bmatrix} \mathbf{0}_n' \\ \vdots \\ \mathbf{0}_n' \\ x_i(t) \\ \mathbf{0}_n' \\ \vdots \\ \mathbf{0}_n' \end{bmatrix} = \begin{bmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \\ x_{i1}(t) & \cdots & x_{in}(t) \\ 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{bmatrix}_{n\times n} \qquad \longleftarrow i^{\underline{a}}\ \textit{linha}$$

Defina a matriz particionada $n \times n^2$:

$$\mathbf{X}(t) = \left[\ \mathbf{X}_1(t) \mid \quad \cdots \quad \mid \mathbf{X}_n(t)\ \right]_{n\times n^2}$$

Seja:

$$\mathbf{Y}_i(t) = \begin{bmatrix} x_{i1}(t) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & x_{in}(t) \end{bmatrix}$$

Agora defina a matriz particionada $n \times n^2$:

$$\mathbf{Y}(t) = \left[\begin{array}{c|c|c} \mathbf{Y}_1(t) & \cdots & \mathbf{Y}_n(t) \end{array} \right]$$

A matriz dos coeficientes das restrições de liqüidez é dada pela matriz:

$$\mathbf{Q} = \begin{bmatrix} \mathbf{X}(t_o) & \mathbf{0}_n & \mathbf{0}_n & \cdots & \mathbf{0}_n & \cdots & \mathbf{0}_n \\ \mathbf{X}(t_o) - \mathbf{Y}(t_o) & \mathbf{X}(t_o) & \mathbf{X}(t_1) & \cdots & \mathbf{0}_n & \cdots & \mathbf{0}_n \\ \mathbf{X}(t_o) - \mathbf{Y}(t_o) & \mathbf{X}(t_o) - \mathbf{Y}(t_o) & \mathbf{X}(t_1) - \mathbf{Y}(t_1) & \cdots & \mathbf{0}_n & \cdots & \mathbf{0}_n \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \mathbf{X}(t_o) - \mathbf{Y}(t_o) & \mathbf{X}(t_o) - \mathbf{Y}(t_o) & \mathbf{X}(t_1) - \mathbf{Y}(t_1) & \cdots & \mathbf{0}_n & \cdots & \mathbf{0}_n \\ \mathbf{X}(t_o) - \mathbf{Y}(t_o) & \mathbf{X}(t_o) - \mathbf{Y}(t_o) & \mathbf{X}(t_1) - \mathbf{Y}(t_1) & \cdots & \mathbf{X}(t_o) & \cdots & \mathbf{X}(t_K) \end{bmatrix}$$

de dimensão $n(K+1) \times n^2 \frac{1}{2} (K+2)(K+1)$.

Para cada banco $i \in \mathbf{B}$, seja $\upsilon_i(\tau, t) = (\upsilon_{i1}(\tau, t), ..., \upsilon_{in}(\tau, t))$, $\forall t_o \leqslant \tau \leqslant t, \forall t \in \mathbf{T}$. Defina:

$$\upsilon(t_k) = ((\upsilon_i(t_o, t_k))_{1 \leqslant i \leqslant n}, (\upsilon_i(t_1, t_k))_{1 \leqslant i \leqslant n}, ..., (\upsilon_i(t_k, t_k))_{1 \leqslant i \leqslant n}) \in \mathbb{R}^{n+2n+3n+...+(k+1)n}$$

Considere o vetor de $n^2 \frac{1}{2} (K+2)(K+1)$ componentes:

$$\upsilon = (\upsilon(t_o), ..., \upsilon(t_K))$$

Defina $x(t) = (x_1(t), ..., x_n(t))$, de n^2 componentes e seja:

$$\overleftarrow{x}(t_k) = (x(t_o), ..., x(t_k)) = (x(t_\ell))_{t_o \leqslant \ell \leqslant t_k} \in \mathbb{R}^{(k+1)n^2}, 1 \leqslant k \leqslant K$$

Defina:

$$\mathbf{x} = (\overleftarrow{x}(t_o), \overleftarrow{x}(t_1), ..., \overleftarrow{x}(t_K)) \in \mathbb{R}^{n^2 \frac{1}{2}(K+2)(K+1)}$$

Seja $\mathbf{B}_o = (B_o^1, ..., B_o^n)$ o vetor de reservas iniciais e defina:

$$\mathbf{b} = \begin{bmatrix} \mathbf{B}_o \\ \vdots \\ \mathbf{B}_o \end{bmatrix}$$

de dimensão $n(K+1)$.

Podemos ver facilmente que a função-objetivo é:

$$\mathbf{x}'\upsilon = \sum_{i\in\mathbf{B}}\sum_{j\in\mathbf{B}}\sum_{t\in\mathbf{T}}\sum_{\tau\leqslant t} x_{ij}(\tau)\upsilon_{ij}(\tau,t)$$

e, caso não houvesse empréstimos Lombard, as restrições de liqüidez poderiam ser escritas como $\mathbf{Q}\upsilon \leqq \mathbf{b}$, onde $\leqq$ denota a desigualdade componente a componente entre vetores de mesma dimensão.

A matriz dos coeficientes das restrições de consistência é dada pela matriz $n^2(K+1) \times n^2\frac{1}{2}(K+2)(K+1)$:

$$\mathbf{J} = \begin{bmatrix} \mathbf{I}_{n^2} & \mathbf{I}_{n^2} & \mathbf{0} & \mathbf{I}_{n^2} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_{n^2} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_{n^2} & \mathbf{0} & \mathbf{I}_{n^2} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{I}_{n^2} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{n^2} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{I}_{n^2} & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{n^2} \end{bmatrix}$$

onde $\mathbf{I}_{n^2}$ é a matriz identidade de ordem n^2.

Denote por $\mathbf{1}$ o vector $n^2\frac{1}{2}(K+2)(K+1)$-dimensional de 1's. Portanto, as restrições de consistência são descritas pelo sistema $\mathbf{J}\upsilon \leqq \mathbf{1}$.

Defina a matriz $n(K+1) \times n(K+1)$:

$$\mathbf{C} = \begin{bmatrix} -\mathbf{I}_n & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I}_n & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & -\mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{I}_n \end{bmatrix}$$

onde $\mathbf{I}_n$ é a matriz identidade de ordem n e $\mathbf{0}$ é a matriz nula $n \times n$.

Dado o vetor $\mathbf{M} = (M_1, \ldots, M_n)$, considere:

$$\mathbf{m}^o = \underbrace{(\mathbf{M}, \ldots, \mathbf{M})}_{K\ vezes} \in \mathbb{R}^{nK}$$

e defina $\mathbf{m} = (\mathbf{m}^o, \mathbf{0}_n) \in \mathbb{R}^{n(K+1)}$. Seja $\mathbf{R}(t)$ a n-matriz diagonal:

$$\mathbf{R}(t) = \begin{bmatrix} r_1(t) & 0 & \cdots & 0 \\ 0 & r_2(t) & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & r_n(t) \end{bmatrix}$$

Designando por $\mathbf{0}$ a matriz nula $n \times n$, defina a matriz $n(K+1) \times n(K+1)$:

$$\mathbf{R} = \begin{bmatrix} \mathbf{I}_n & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{I}_n & \mathbf{I}_n & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{I}_n & \mathbf{I}_n & \cdots & \mathbf{I}_n & \mathbf{0} \\ \mathbf{I}_n + \mathbf{R}(t_o) & \mathbf{I}_n + \mathbf{R}(t_1) & \cdots & \mathbf{I}_n + \mathbf{R}(t_{K-1}) & -\mathbf{I}_n \end{bmatrix}$$

Então as restrições de crédito são escritas como:

$$\mathbf{R}\pi \leqq \mathbf{m}$$

Assim, o problema primal é reescrito como:

$$(\mathcal{P}) \begin{cases} \max_{(\upsilon,\pi)} & \mathbf{x}'\upsilon \\ \textit{sujeito a} & \begin{bmatrix} \mathbf{Q} & \mathbf{C} \\ \mathbf{J} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix} \begin{bmatrix} \upsilon \\ \pi \end{bmatrix} \leqq \begin{bmatrix} \mathbf{b} \\ \mathbf{1} \\ \mathbf{m} \end{bmatrix} \\ & (\upsilon,\pi) \geqq \mathbf{0} \end{cases}$$

Aqui, obviamente, as matrizes nulas $\mathbf{0}$ têm as dimensões apropriadas para que a divisão em blocos acima seja consistente, o mesmo valendo para o vetor nulo na restrição $(\upsilon,\pi) \geqq \mathbf{0}$.

4. Preços-sombra da Política Monetária Intradiária

4.1 Problema dual do Banco Central

O problema dual é um instrumento que nos permite fazer análise de sensibilidade. Cada restrição é apreçada de acordo com sua contribuição marginal para a função objetivo otimizada. Esses preços – ditos preços-sombra – são as variáveis de escolha duais:

Restrição Primal	Parâmetro	Preço-sombra
LIQ	B_o^i	$\lambda_i(t)$
CONS	1	$\mu_{ij}(t)$
CRED	M_i	$\xi_i(t)$

Defina $\lambda_i = (\lambda_i(t))_{t=t_o,\ldots,t_K}$ como o vetor de preços-sombra das restrições de liqüidez para o banco i, $\mu_i(t) = (\mu_{ij}(t))_{j=1,\ldots,n}$, $\mu_i = (\mu_i(t))_{t=t_o,\ldots,t_K}$ como o vetor de preços-sombra das restrições de consistência para o banco i, e, finalmente,

$\xi_i = (\xi_i(t))_{t=t_o,\dots,t_K}$ como o vetor de preços-sombra das restrições de crédito para o banco i. Os vetores de preços-sombra das restrições de liqüidez, de consistência e de crédito são denotados, respectivamente, por λ, μ e ξ.

O problema dual é dado por:

$$(\mathcal{D}) \begin{cases} \min_{(\lambda,\mu,\xi)} & \mathbf{b}'\lambda + \mathbf{1}'\mu + \mathbf{m}'\xi \\ \textit{sujeito a} & \begin{bmatrix} \mathbf{Q}' & \mathbf{J}' & \mathbf{0}' \\ \mathbf{C}' & \mathbf{0}' & \mathbf{R}' \end{bmatrix} \begin{bmatrix} \lambda \\ \mu \\ \xi \end{bmatrix} \geqq \begin{bmatrix} \mathbf{x} \\ \mathbf{0} \end{bmatrix} \\ & (\lambda, \mu, \xi) \geqq \mathbf{0} \end{cases}$$

onde o superescrito $'$ denota a transposta.

A função objetivo dual é o valor econômico dos recursos escassos (balanço inicial, crédito intradiário do Banco Central e restrições de consistência) caso eles tivessem de ser comprados:

$$\sum_t \sum_i B_o^i \lambda_i(t) + \sum_{t<T} \sum_i M_i \xi_i(t) + \sum_t \sum_i \sum_j \mu_{ij}(t)$$

4.2 Valor dual *versus* valor de face das transferências interbancárias e dos empréstimos Lombard

Há dois conjuntos de restrições duais. A primeira diz que, em um momento $t \geqslant \tau$, o valor econômico de uma transferência interbancária de grande valor enviada em τ não é seu valor de face $x_{ij}(\tau)$, mas o preço-sombra $\mu_{ij}(\tau)$ de seu enfileiramento mais seu valor de face ajustado por um coeficiente dado pelo preço-sombra corrente $\lambda_i(t)$ do balanço inicial do banco pagador mais o preço-sombra líqüido bilateral (banco pagador e recebedor) futuro (acumulado do instante do envio até o fim do dia) $\sum_{\theta=t+1}^{T}[\lambda_i(\theta) - \lambda_j(\theta)]$, ou seja:

$$\begin{aligned} x_{ij}(\tau)\{\lambda_i(t) + \sum_{\theta=t+1}^{T}[\lambda_i(\theta) - \lambda_j(\theta)]\} + \mu_{ij}(\tau) &\geqslant x_{ij}(\tau) \\ x_{ij}(\tau)\lambda_i(T) + \mu_{ij}(\tau) &\geqslant x_{ij}(\tau),\ \forall \tau \in \mathbf{T} \end{aligned}$$

Assim, o valor econômico ou dual de um pagamento é uma transformação afim de seu valor contábil. A restrição dual diz que o valor contábil ou de face de um pagamento não pode exceder seu valor econômico. Note que o valor econômico de um pagamento varia com o tempo, tendo como referências o final do dia e o período em que o valor do pagamento é avaliado.

O segundo conjunto de restrições duais refere-se aos empréstimos Lombard. Em qualquer período, os preços-sombra acumulados da seqüência futura de empréstimos Lombard devem ser tão altos quanto o preço-sombra corrente de sua reserva inicial:

$$\begin{aligned} -\lambda_i(t) + \sum_{\theta=t}^{T-1} \xi_i(\theta) + (1 + r_i(t))\xi_i(T) &\geqslant 0, \ \forall t < T \\ \lambda_i(T) - \xi_i(T) &\geqslant 0 \end{aligned}$$

ou ainda:

$$\begin{aligned} \sum_{\theta=t}^{T} \xi_i(\theta) + r_i(t)\xi_i(T) &\geqslant \lambda_i(t), \ \forall t < T \\ \xi_i(T) &\leqslant \lambda_i(T) \end{aligned}$$

Assim, um empréstimo Lombard ou acordo de recompra só vale a pena do ponto de vista econômico se seu valor dual jamais é inferior ao valor dual das reservas iniciais. Isso condiz com a boa prática econômica: um empréstimo não pode ser concedido para a realização de uma ação se esta pode ser executada sem o empréstimo, caso contrário teríamos desperdício de liqüidez. No final do dia a desigualdade é reversa justamente pelo fato de ser um empréstimo negativo, ou seja, é a hora do pagamento dos empréstimos feitos ao longo do dia.

4.3 Equação de eficiência de liqüidez sistêmica

O Teorema Fundamental da Programação Linear diz que, sob condições gerais, o valor máximo da função objetivo primal coincide com o valor mínimo da função objetivo dual. Em outras palavras, o montante ótimo de pagamentos compensados até o fim do dia coincide com o valor econômico dos recursos escassos do Banco Central durante o dia. Isso implica a seguinte equação:

$$\Lambda = \sum_{i \in \mathbf{B}} B_o^i(\bar{\lambda}_i - 1) + \sum_{i \in \mathbf{B}} \sum_{j \in \mathbf{B}} \bar{\mu}_{ij} + \sum_{i \in \mathbf{B}} M_i(\bar{\xi}_i - \frac{1}{T}\xi_i(T))$$

em que $\bar{\lambda}_i = \frac{1}{T}\sum_{t \in \mathbf{T}} \lambda_i(t)$, $\bar{\mu}_{ij} = \frac{1}{T}\sum_{t \in \mathbf{T}} \mu_{ij}(t)$ e $\bar{\xi}_i = \frac{1}{T}\sum_{t \in \mathbf{T}} \xi_i(t)$ são os preços-sombra médios (ao longo do dia) do banco i.

A moeda do Banco Central será integralmente usada para compensações sempre que o custo de liqüidez sistêmica mínimo for nulo, isto é, se:

$$\sum_{i\in\mathbf{B}} B_o^i(\bar{\lambda}_i - 1) + \sum_{i\in\mathbf{B}}\sum_{j\in\mathbf{B}} \bar{\mu}_{ij} + \sum_{i\in\mathbf{B}} M_i(\bar{\xi}_i - \frac{1}{T}\xi_i(T)) = 0$$

ou ainda:

$$\sum_{i\in\mathbf{B}} B_o^i\bar{\lambda}_i + \sum_{i\in\mathbf{B}} M_i\bar{\xi}_i + \sum_{i\in\mathbf{B}}\sum_{j\in\mathbf{B}} \bar{\mu}_{ij} = \sum_{i\in\mathbf{B}} B_o^i + \frac{1}{T}\left(\sum_{i\in\mathbf{B}} M_i\right)\xi_i(T)$$

Essa é a equação fundamental para a eficiência de liqüidez. Ela mostra como os requerimentos de reserva bancária, os montantes de crédito intradiário para os bancos individuais, a extensão da exposição interbancária podem ser usados para atingir a eficiência de liqüidez por meio dos preços-sombras. Note que a taxa de juros sobre empréstimos Lombard (*haircuts*) é também uma variável de controle, embora não apareça explicitamente na função-valor dual. Ela está escondida nas restrições duais, mas pode certamente também ser usada pelo Banco Central para atingir a eficiência de liqüidez.

5. Aplicações à Política Monetária Intradiária

Se o Banco Central conhecesse os preços-sombra, ele seria capaz de ditar a política monetária intradiária de modo a fazer $\Lambda = 0$. Sempre que $\Lambda > 0$, algum montante de moeda do Banco Central está parado. O custo de oportunidade de Λ é o que os bancos perdem por participarem do sistema LBTR. Assim, o valor primal $\Lambda \geqslant 0$ nos informa o montante monetário que, apesar de estar dentro do sistema de pagamentos para fins de compensação interbancária, não está sendo utilizado, devido à rigidez do desenho do sistema LBTR ou ainda porque há reservas iniciais em excesso. Como o valor primal é mínimo, Λ serve como uma quota inferior para o custo de liqüidez sistêmica. O custo social do sistema LBTR é o custo de oportunidade de Λ.

Estamos especificamente interessados no uso dos preços-sombra para apreçar os parâmetros de política monetária intradiária. Assim fazendo, o Banco Central será capaz de fazer com que os bancos participantes do sistema LBTR internalizem as externalidades de rede que causam no sistema.

Suponha, por exemplo, que a única fonte de liqüidez intradiária é o montante de reservas iniciais, além, é claro, das transferências líqüidas. Se o custo de liqüidez sistêmica é positivo, então $\Lambda = \sum_{i\in\mathbf{B}} B_o^i(1-\bar{\lambda}_i) - \sum_{i\in\mathbf{B}}\sum_{j\in\mathbf{B}}\bar{\mu}_{ij} > 0$. Isso significa que existe liqüidez no sistema que não está sendo usada para compensação de pagamentos. Como o Banco central pode ajustar as reservas bancárias para tornar o sistema eficiente? A partir da equação de eficiência $\sum_{i\in\mathbf{B}} B_o^{i*}(1-\bar{\lambda}_i) - \sum_{i\in\mathbf{B}}\sum_{j\in\mathbf{B}}\bar{\mu}_{ij} = 0$, uma solução é:

$$B_o^{i*} = \begin{cases} \frac{1}{\#\mathbf{B}_1(\bar{\lambda}_i-1)}\sum_{k\in\mathbf{B}_o}\{B_o^k(1-\bar{\lambda}_k) - \sum_{j\in\mathbf{B}}\bar{\mu}_{kj}\} - \frac{1}{\bar{\lambda}_i-1}\sum_{j\in\mathbf{B}}\bar{\mu}_{ij} & se\ i\in\mathbf{B}_1 \\ B_o^i & se\ i\in\mathbf{B}_o \end{cases}$$

Aqui, $\mathbf{B}_1 = \{i : \bar{\lambda}_i > 1\}$ é o conjunto de bancos com preço-sombra de liqüidez médio acima da unidade e $\mathbf{B}_o = \{i : 0 \leqslant \bar{\lambda}_i \leqslant 1\} = \mathbf{B}\backslash\mathbf{B}_1$ é o conjunto dos demais bancos. Na solução acima, bancos com baixos preços-sombra mantêm seus históricos de reservas, ao passo que bancos com alto preço-sombra mudam suas reservas para B_o^{i*}.

Em alguns casos, como no FEDWIRE norte-americano, a existência de *overdrafts* é permitida, mesmo sendo um sistema LBTR. A extensão do *overdraft*, conhecida como *net debit cap* (NDC), é obviamente condicionada. Sendo esse o caso, como o Banco Central poderia usar os preços-sombra para determinar o NDC ótimo ou mesmo NDC's personalizados? Seja $D_i(t)$ a extensão de NDC que o Banco Central outorga ao banco i no período t. Claramente podemos interpretar a cessão desse benefício como um montante monetário adicionado aos requerimentos de reserva inicial durante o dia. A equação fundamental para eficiência de liqüidez sistêmica torna-se:

$$\sum_{i\in\mathbf{B}} B_o^i(\bar{\lambda}_i - 1) + \frac{1}{T}\sum_{i\in\mathbf{B}}\sum_{t\in\mathbf{T}} D_i(t)\lambda_i(t) + \sum_{i\in\mathbf{B}}\sum_{j\in\mathbf{B}}\bar{\mu}_{ij} + \sum_{i\in\mathbf{B}} M_i(\bar{\xi}_i - \frac{1}{T}\xi_i(T)) = 0$$

Seja $\Lambda = \sum_{i\in\mathbf{B}} B_o^i(1-\bar{\lambda}_i) - \sum_{i\in\mathbf{B}}\sum_{j\in\mathbf{B}}\bar{\mu}_{ij} > 0$ o custo mínimo de liqüidez sistêmica na ausência de NDC's e empréstimos Lombard, isto é, acordos de recompra. Se os bancos tivessem acesso a um NDC constante, digamos, $\bar{D}_i(t) = D > 0$, então:

$$\sum_{i\in\mathbf{B}} B_o^i(1-\bar{\lambda}_i) - \sum_{i\in\mathbf{B}}\sum_{j\in\mathbf{B}}\bar{\mu}_{ij} - D\sum_{i\in\mathbf{B}}\bar{\lambda}_i = 0$$

ou seja:

$$D^* = \max\{\frac{\sum_{i\in\mathbf{B}} B_o^i(1-\bar{\lambda}_i) - \sum_{i\in\mathbf{B}}\sum_{j\in\mathbf{B}}\bar{\mu}_{ij}}{\sum_{i\in\mathbf{B}}\bar{\lambda}_i}, 0\}$$

isso porque a equação acima poderia admitir solução negativa, o que significaria um imposto uniforme sobre os bancos. Uma solução personalizada é dada por:

$$D_i^* = \frac{B_o^i(1-\bar{\lambda}_i) - \sum_{j\in\mathbf{B}} \bar{\mu}_{ij}}{\bar{\lambda}_i}$$

Conforme D_i^* seja positivo ou negativo, o banco i receberia um NDC ou seria taxado. Assim, NDC's seriam financiados pelos próprios bancos dentro do sistema de pagamentos mediante redistribuição de liqüidez. De um ponto de vista pragmático, uma solução mais recomendável seria beneficiar alguns bancos com NDC's personalizados sem taxação de outros. Obviamente, bancos com alto preço-sombra arcariam com o custo proveniente do não-pagamento de taxas por parte de outros bancos. Se $\mathbf{B}_1 = \{i : \bar{\lambda}_i > 1\}$ é o conjunto de bancos com preço-sombra de liqüidez médio acima da unidade e $\mathbf{B}_o = \{i : 0 \leqslant \bar{\lambda}_i \leqslant 1\} = \mathbf{B}\backslash\mathbf{B}_1$ é o conjunto dos demais bancos, então:

$$D_i^* = \begin{cases} \frac{B_o^i(1-\bar{\lambda}_i) - \sum_{j\in\mathbf{B}} \bar{\mu}_{ij}}{\bar{\lambda}_i} + \frac{1}{\#\mathbf{B}_1\bar{\lambda}_i}\sum_{k\in\mathbf{B}_o}\{B_o^k(1-\bar{\lambda}_k) - \sum_{j\in\mathbf{B}}\bar{\mu}_{kj}\} & se\ i \in \mathbf{B}_1 \\ 0 & se\ i \in \mathbf{B}_o \end{cases}$$

Comparando com a solução anterior, vemos que NDC's são reduzidos pelo montante dado pelo segundo termo na expressão acima. O montante da redução é exatamente igual ao montante das taxas não pagas pelos bancos com baixo preço-sombra dividido igualmente entre os bancos com alto preço-sombra e ponderado pela inversa do preço-sombra médio de liqüidez. Quanto maior o preço-sombra, menor a redução de seu NDC em relação à solução anterior.

6. Exemplos

Nesta seção consideramos alguns exemplos que ilustram o modo como nosso modelo pode responder às questões levantadas na introdução mediante o uso de preços-sombra em um sistema LBTR. Ressaltamos que as conclusões dos exemplos não sugerem regras gerais de conduta da política monetária intradiária.

Suponha que há dois bancos, $\mathbf{B} = \{1, 2\}$, e que o dia está dividido em três períodos, $\mathbf{T} = \{t_o, t_1, t_2\}$, que chamaremos de *manhã*, *tarde* e *noite* (o fim do dia), respectivamente. As transferências a serem feitas ao longo do dia são resumidas nas matrizes abaixo:

$$x(t_o) = \begin{bmatrix} 0 & 80 \\ 120 & 0 \end{bmatrix} \qquad x(t_1) = \begin{bmatrix} 0 & 180 \\ 120 & 0 \end{bmatrix} \qquad x(t_2) = \begin{bmatrix} 0 & 100 \\ 120 & 0 \end{bmatrix}$$

Cada entrada $x_{ij}(t)$ deve ser lida como "o banco i transfere $x_{ij}(t)$ unidades monetárias para o banco j no período t". Os balanços iniciais (as reservas iniciais requeridas pelo Banco Central) são dados por $B_o^1 = \$100$ e $B_o^2 = \$120$. O Banco Central possui um montante monetário, digamos, \$130, para emprestar temporariamente a bancos ilíqüidos. Suponha que ele possui $M_1 = \$50$ disponíveis para empréstimos ao banco 1 e $M_2 = \$80$ para empréstimos ao banco 2.

Exemplo 1: *Travamento, regra FIFO e ausências de empréstimos Lombard e de mercado interbancário.* Se o padrão de pagamentos fosse como o especificado acima em um sistema LBTR com enfileiramento dado pela regra FIFO (*first-in first-out*) e sem fracionamento de pagamentos, então, no período da manhã, os bancos fariam suas transferências normalmente. O banco 1 ficaria com um saldo de \$140 e o banco 2 com um saldo de \$80. Já no período da tarde, nenhum dos dois bancos poderia fazer as transferências, pois não têm saldos suficientes. As duas mensagens de pagamento são então enfileiradas para o período seguinte. À noite as mensagens enfileiradas da tarde ainda não podem ser pagas, o que impede o pagamento das mensagens noturnas. O dia termina com um total de \$520 não compensados. ■

Exemplo 2: *Regra FAFO e ausência de empréstimos Lombard e mercado interbancário.* Considere a mesma situação anterior, mas com a diferença de que a regra FIFO é substituída pela regra FAFO (*first-available first-out*). É no período da noite que aparecem as diferenças entre essas duas regras. À noite, apenas o banco 1 pode fazer a transferência do período, \$100, mas não pode cumprir a transferência enfileirada, ficando com um saldo final de \$40. O banco 2 termina o dia com um saldo final de \$180. O banco 1 compensou apenas \$180 de um total de \$360 e o banco 2 compensou apenas \$120 de um total de \$360. Faltou serem compensados \$420. A ineficiência se revela no fato de o banco 1 terminar o dia com \$180 e não poder pagar \$180 que deve. Se o banco 2 recebesse esses \$180, ficaria com saldo de \$220 e poderia pagar os \$120 que deve. A isso se soma outra ineficiência. Mesmo que o arranjo acima fosse possível, o banco 2 ainda terminaria o dia com um excesso de reservas no valor de \$100, arcando com seu custo de oportunidade. Nos exemplos seguintes, analisamos como diferentes desenhos do LBTR podem reduzir a ineficiência. ■

Exemplo 3: *Fracionamento e ausências de empréstimos Lombard e de mercado interbancário.* Vamos mostrar agora como nosso modelo pode reduzir o custo de liqüidez sistêmica. A variável de escolha é $v_{ij}(s,t)$, que representa a porção do pagamento $x_{ij}(s)$ do banco i para o banco j no período s e que será compensada no período $t \geqslant s$. A solução primal é dada por:

$$\begin{cases} \upsilon^*_{12}(t_o,t_o) = \upsilon^*_{21}(t_o,t_o) = 1 \\ \upsilon^*_{12}(t_o,t_1) = \upsilon^*_{21}(t_o,t_1) = \upsilon^*_{12}(t_o,t_2) = \upsilon^*_{21}(t_o,t_2) = 0 \\ \upsilon^*_{12}(t_1,t_1) = 7/9 \\ \upsilon^*_{21}(t_1,t_1) = 2/3 \\ \upsilon^*_{12}(t_1,t_2) = 0 \\ \upsilon^*_{21}(t_1,t_2) = 1/3 \\ \upsilon^*_{12}(t_2,t_2) = 4/5 \\ \upsilon^*_{21}(t_2,t_2) = 5/6 \end{cases}$$

Assim, $\upsilon^*_{12}(t_o,t_o) = 1$ significa que o pagamento matutino $x_{12}(t_o) = \$80$ do banco 1 para o banco 2 é compensado por inteiro. Já $\upsilon^*_{21}(t_1,t_1) = \frac{2}{3}$ e $\upsilon^*_{21}(t_1,t_2) = \frac{1}{3}$ significam que $\frac{2}{3}$ do pagamento verspertino $x_{21}(t_1) = \$120$ do banco 2 para o banco 1 (ou seja, \$80) são compensados imediatamente, enquanto que $\frac{1}{3}$ (isto é, \$40) é compensado à noite. O valor ótimo do fluxo total de pagamentos é \$640. Portanto, o custo de liquidez sistêmica é $\Lambda = \$6,67$. Esse é o montante monetário que existe no sistema mas não é usado para finalizar os pagamentos. Uma outra solução é obtida fazendo-se $\upsilon^*_{12}(t_1,t_2) = \frac{2}{9}$ e $\upsilon^*_{12}(t_2,t_2) = \frac{2}{5}$, mantendo-se inalteradas as demais variáveis.[3] Claramente o valor ótimo é o mesmo, \$640. Para os fins deste exemplo, consideraremos apenas a primeira solução acima. Os preços-sombra são dados por:

$$\begin{cases} \lambda^*_1(t_o) = 0 \\ \lambda^*_2(t_o) = \lambda^*_1(t_1) = \lambda^*_2(t_1) = \lambda^*_1(t_2) = \lambda^*_2(t_2) = 1 \\ \mu^*_{12}(t_o) = 80 \\ \mu^*_{21}(t_o) = \mu^*_{21}(t_1) = \mu^*_{21}(t_1) = \mu^*_{12}(t_2) = \mu^*_{21}(t_2) = 0 \end{cases}$$

O valor dual é 640, de modo que não existe defasagem de dualidade. O pagamento vespertino do banco 1 é parcialmente liqüidado. O montante não compensado é $\frac{2}{9} \times \$180 = \$40$. O pagamento noturno do banco 1 também é parcialmente compensado e o montante não compensado é $\frac{1}{5} \times \$100 = \$20$. O montante não compensado do pagamento noturno do banco 2 é $\frac{1}{6} \times \$120 = \$20$. Em suma, o banco 1 possui \$60 a serem compensados e o banco 2 possui \$20. O banco 1 teria que ter seu requerimento de reserva aumentado de $\$y$ de modo a satisfazer a equação $y \times (\lambda^*_1(t_o) + \lambda^*_1(t_1) + \lambda^*_1(t_2)) = \$60$, ou seja, \$30. Por outro

[3] Agradeço a um parecerista anônimo pela solução alternativa. Uma explicação para essa multiplicidade é que pontos ótimos podem estar sobre uma face do poliedro de restrições, e não apenas no conjunto de pontos extremos.

lado, o banco 2 deveria ter suas reservas aumentadas de $\$z$ de modo a termos $z \times (\lambda_2^*(t_o) + \lambda_2^*(t_1) + \lambda_2^*(t_2)) = \$20$, ou seja, \$10. Portanto, se a reserva inicial do banco 1 fosse \$130 em vez de \$100 e se a reserva inicial do banco 2 fosse \$130 em vez de \$120, o sistema teria sido capaz de liqüidar todos os pagamentos intradiários mediante enfileiramento com fracionamento. Em outras palavras, com \$40 a mais no sistema compensam-se \$80 a mais dos pagamentos. ■

Exemplo 4: *Empréstimos Lombard e ausências de enfileiramento e de mercado interbancário.* Recorde que o Banco Central possui $M_1 = \$50$ e $M_2 = \$80$ disponíveis para empréstimo aos bancos 1 e 2, respectivamente. Esses montantes têm que ser alocados otimamente durante o dia. A solução primal é:

$$\begin{cases} \upsilon_{12}^*(t_o) = \upsilon_{21}^*(t_o) = \upsilon_{21}^*(t_2) = 1 \\ \upsilon_{12}^*(t_1) = 7/9 \\ \upsilon_{21}^*(t_1) = 3/4 \\ \upsilon_{12}^*(t_2) = 9/10 \\ \pi_1^*(t_o) = \pi_2^*(t_o) = \pi_1^*(t_1) = \pi_1^*(t_2) = 0 \\ \pi_2^*(t_1) = \pi_2^*(t_2) = 10 \end{cases}$$

Note que o banco 2 faz um acordo de recompra com o Banco Central à tarde, recebendo um empréstimo Lombard de \$10 e pagando-o à noite. O fluxo máximo de pagamentos é \$640, de modo que o custo de liqüidez sistêmica é $\Lambda = \$6,67$. Este exemplo mostra que substituir o arranjo institucional do exemplo 4 por um que inclua empréstimos Lombard não melhora necessariamente o fluxo de pagamentos, porquanto o custo mínimo de liqüidez sistêmica permanece o mesmo. Os preços-sombra são:

$$\begin{cases} \lambda_1^*(t_o) = 0 \\ \lambda_2^*(t_o) = \lambda_1^*(t_1) = \lambda_2^*(t_1) = \lambda_1^*(t_2) = \lambda_2^*(t_2) = 1 \\ \mu_{12}^*(t_o) = 80 \\ \mu_{21}^*(t_o) = \mu_{12}^*(t_1) = \mu_{21}^*(t_1) = \mu_{12}^*(t_2) = \mu_{21}^*(t_2) = 0 \\ \xi_1^*(t_o) = \xi_2^*(t_o) = \xi_1^*(t_1) = \xi_2^*(t_1) = 0 \\ \xi_1^*(t_2) = 80 \\ \xi_2^*(t_2) = 1 \end{cases}$$

Para o Banco Central eliminar o custo de liqüidez sistêmica, a política monetária intradiária deve ser capaz de reduzir esse custo em \$6,67, ou seja, o fluxo máximo de pagamentos deve aumentar em \$20. O montante de crédito disponível para o banco 1 poderia crescer em $\$w$, onde w resolve $w \times \xi_1^*(t_2) = \$20$, ou seja, $w = \$0,25$. Em outras palavras, se o Banco Central emprestasse \$0,25 a mais na forma de acordo de recompra ao banco 1, todo custo de liqüidez sistêmica seria eliminado. ■

Exemplo 5: *Fracionamento com mercado interbancário intradiário.* Suponha agora que os pagamentos podem ser fracionados (e, portanto, enfileirados) e que não existem empréstimos Lombard. Entretanto, existe um mercado interbancário intradiário livre. Por simplicidade, supomos que a taxa de juros desse mercado é nula. A variável de escolha $\upsilon_{ij}(s,t)$ é livre, mas a restrição de consistência $0 \leqslant \sum_{t \geqslant s} \upsilon_{ij}(s,t) \leqslant 1$, $\forall s$, deve ser atendida. A solução primal é:

$$\begin{cases} \upsilon_{12}^*(t_o,t_o) = 5/4 \\ \upsilon_{21}^*(t_o,t_o) = 1 \\ \upsilon_{12}^*(t_o,t_1) = \upsilon_{21}^*(t_o,t_1) = \upsilon_{21}^*(t_o,t_2) = 0 \\ \upsilon_{12}^*(t_o,t_2) = -1/4 \\ \upsilon_{12}^*(t_1,t_1) = 2/3 \\ \upsilon_{21}^*(t_1,t_1) = 5/6 \\ \upsilon_{12}^*(t_1,t_2) = 1/3 \\ \upsilon_{21}^*(t_1,t_2) = 1/6 \\ \upsilon_{12}^*(t_2,t_2) = 3/5 \\ \upsilon_{21}^*(t_2,t_2) = 5/6 \end{cases}$$

O banco 1 compensa com finalidade seu pagamento matutino, $x_{12}(t_o) = \$80$, e empresta 25% desse valor, ou seja, $\frac{1}{4} \times 80 = \$20$, ao banco 2 ainda pela manhã. O banco 2 paga à noite os \$20 que tomou emprestado do banco 1 pela manhã, como pode ser deduzido da solução $\upsilon_{12}(t_o,t_2) = -\frac{1}{4}$, que implica $\upsilon_{12}(t_o,t_2)x_{12}(t_o) = -\frac{1}{4} \times 80 = -\$20$. De $\upsilon_{12}(t_1,t_1) = \frac{2}{3}$ e $\upsilon_{12}(t_1,t_2) = \frac{1}{3}$, conclui-se que o banco 1 compensa $\frac{2}{3}$ de seu pagamento vespertino e enfileira $\frac{1}{3}$ para compensar à noite. O fluxo máximo é \$660, de modo que o custo mínimo de liqüidez sistêmica é $\Lambda = \$0$. Portanto, dado o total de reservas $\sum_i B_o^i = \$220$, o arranjo ótimo de fracionamento/enfileiramento e de mercado interbancário intradiário livre levam o sistema a operar eficientemente com capacidade máxima. Este exemplo mostra que o mercado às vezes pode exercer de forma mais eficiente a função usual do Banco Central de emprestador de última instância. Note que o total de pagamentos é de \$720, mas apenas \$660 forma compensados. Isso é o melhor que o sistema de pagamentos com os arranjos institucionais especificados pode fazer a partir de um total de reservas iniciais de \$220. O problema aqui é que as reservas iniciais são insuficientes. Conhecidos os preços-sombra das restrições de liqüidez de cada banco, o Banco Central poderia determinar um aumento de reservas personalizado. Quanto maior o preço-sombra do banco, menor o aumento necessário.[4] ■

Exemplo 6: *Net debit caps.* Dados os preços-sombra do exemplo 4, os preços-sombra médios das restrições de liqüidez para cada banco são $\bar{\lambda}_1^* = \frac{2}{3}$ e $\bar{\lambda}_2^* = 1$ e os de consistência são $\bar{\mu}_{12}^* = \frac{80}{3}$ e $\bar{\mu}_{21}^* = 0$. Se cada banco pudesse ficar descoberto até um certo montante, então seria eficiente estabelecer os *net debit caps* como $D_1^* = \frac{B_o^1(1-\bar{\lambda}_1)-\bar{\mu}_{12}}{\bar{\lambda}_1} = \frac{100(1-\frac{2}{3})-\frac{80}{3}}{\frac{2}{3}} = \$10$ e, similarmente, $D_2^* = \frac{B_o^2(1-\bar{\lambda}_2)-\bar{\mu}_{21}}{\bar{\lambda}_2} = \frac{120(1-1)-0}{1} = \$0$, ou seja, o banco 1 é beneficiado com a possibilidade de ficar descoberto (*overdraft*) até \$10, enquanto que o banco 2 não teria esse benefício. Este exemplo mostra, assim, como a personalização da política monetária intradiária torna o sistema mais eficiente. ■

7. Conclusão

A modelagem do problema primal do Banco Central relativamente à determinação da política monetária intradiária ótima como uma programação linear não é uma simplificação. É, ao contrário, bastante adequada e consistente com a natureza de rede do sistema de pagamentos, em que o fluxo de dinheiro circula entre seus nós. Partimos do pressuposto de que o Banco Central conhece o vetor de pagamentos intradiários. Evidentemente não é assim na realidade. É impossível saber-se antecipadamente qual será o vetor de pagamentos, mas sua distribuição ao longo do dia é certamente conhecida (ver McAndrews e Rajan (2000)) para a distribuição dos pagamentos no sistema FEDWIRE]. O modelo mais adequado é a versão estocástica do que foi aqui apresentado, como em Peñaloza (2004). Entretanto, a principal contribuição econômica de nosso modelo em sua versão determinística é a mesma da estocástica: o poder dos preços-sombra e do tratamento personalizado dos bancos para a eficiência dos pagamentos. As mais importantes

[4] Uma experiência desse tipo ocorreu com sucesso na Suíça durante os anos 1990. Até outubro de 1999, o *Swiss National Bank* (SNB) não fornecia empréstimos intradiários com garantias. A situação mudou porque a autoridade monetária suíça previu um potencial aumento do número e do valor de pagamentos *time-critical* quando da introdução do sistema CLS (Continuous Link Settlement System) (Ver Heller et alii (2000)).

lições que tiraríamos da versão mais geral são também obtidas pela versão determinística. Com efeito, na versão estocástica caímos em uma programação linear em dimensão infinita, em cujo caso os pagamentos e as compensações são ligados por relações de dualidade entre espaços de Banach. Mas as restrições duais possuem o mesmo formato genérico. Por essa razão preferimos abrir mão de tecnicismos desnecessários em favor de uma apresentação mais simples, porém com maior teor intuitivo e similar poder de explicação teórica.

O modelo também é de fácil aplicabilidade empírica a qualquer sistema de pagamentos do tipo LBTR. O sistema de cada país evidentemente requer certas adaptações, mas a estrutura do modelo é geral o suficiente para comportar as especificidades cabíveis. No Brasil, por exemplo, existe enfileiramento com três graus de prioridade e os pagamentos não podem, por lei, ser fracionados. Já em alguns países nórdicos o fracionamento é permitido. Essas diferenças se traduzem em pequenas modificações nas restrições e no domínio de definição da função de compensações.

O Banco Central possui os dados com todos os pagamentos intradiários. A partir desses dados podemos calcular quais teriam sido os preços-sombra caso o sistema funcionasse otimamente. Teríamos assim uma série de preços-sombra para cada política intradiária e para cada banco. As tendências históricas dos preços-sombra podem ajudar na condução da política monetária intradiária ou mesmo na estimação de efeitos sistêmicos marginais de cada banco, estes sendo definidos como a soma dos seus preços-sombra.

A mensagem principal do modelo é propagar o poder dos preços-sombra como critério genuinamente econômico para a boa gestão do sistema de pagamentos. Evidentemente, características políticas e culturais representam uma forte barreira à personalização da política monetária intradiária. Entretanto, isso não impede que o Banco Central, baseado nos preços-sombra, possa fazer com que essa política caminhe sob os auspícios da teoria microeconômica. Sem a pretensão de querer determinar a conduta da gestão intradiária, os preços-sombra podem indicar condutas mais adequadas para objetivos específicos. É nesse sentido que os preços-sombra podem ajudar na gestão do sistema de pagamentos.

Uma extensão natural em nosso projeto de pesquisa é a construção de um jogo bayesiano em que a escolha das políticas monetárias intradiárias do Banco Central afeta o comportamento estratégico dos bancos, juntando assim a literatura existente na qual os bancos são o objeto de estudo com nosso modelo, em que o Banco Central é o agente relevante. A idéia é saber se o desenho do sistema de pagamentos pode ser um equilíbrio bayesiano de Nash. Nessa extensão, o *framework* adequado é o estocástico. Outro caminho a ser seguido é a introdução de recursividade por vários dias consecutivos, o que pode dar um papel maior aos preços-sombra dos empréstimos *overnight*.

Referências

Bagehot, W. (1873). *Lombard Street: A Description of the Money Market*. H. S. King, London.

Bank for International Settlements (1997). Real-time gross settlement systems. CPSS, March, Basle, Switzerland.

Bech, M. & Garrat, R. (2003). The intraday liquidity game. *Journal of Economic Theory*, 109:198–219.

DeBandt, O. & Hartmann, P. (2000). Systemic risk: A survey. European Central Bank, WP #35.

Freixas, X. & Parigi, B. (1998). Contagion and efficiency in gross and net interbank payment systems. *Journal of Financial Intermediation*, 7:3–31.

Heller, D., Nellen, T., & Sturm, A. (2000). The Swiss interbank clearing system. Swiss National Bank, Payment System Subsection.

McAndrews, J. & Rajan, S. (2000). The timing and funding of FEDWIRE funds transfers. FRBNY Economic Policy Review, July. Federal Reserve Bank of New York.

Peñaloza, R. (2004). The pricing of intraday monetary policies. Anais do XXVI Encontro da Sociedade Brasileira de Econometria, EBE 2004, João Pessoa, PB.