

ESTIMAÇÃO DE MODELOS EM TEMPO CONTÍNUO: TAXA DE JUROS DE CURTO PRAZO E SIMULAÇÕES DE MONTE CARLO[†]

ALAN DE GENARO DARIO*
MILTON BAROSSO-FILHO‡

RESUMO

O objetivo deste artigo é estimar parâmetros de modelos de equações diferenciais estocásticas que ajustam o comportamento da taxa de juros de curto prazo brasileira, dada pelo CDI. Baseado em Chan, Karolyi, Longstaff & Sanders (1992), reproduzimos os resultados dos mesmos pelo método GMM. Em seguida, comparamos as propriedades em distribuição, para amostras finitas, das estimativas dos parâmetros dos modelos, pelos métodos GMM e Quasi-Máxima Verossimilhança, através de simulações de Monte Carlo. Finalmente, avaliamos a qualidade do ajuste dos modelos estimados, a partir de informações reais de mercado, pelo critério do Erro Quadrático Médio.

Palavras-chave: Modelos de difusão, método de momentos generalizados, quase-máxima verossimilhança e taxas de juros de curto prazo brasileiras.

ABSTRACT

The aim of this paper is to estimate the parameters of a stochastic differential equation that fits the behavior of Brazilian short run interest rates. Based on Chan, Karolyi, Longstaff & Sanders (1992), we reproduce their results for that Brazilian financial time series using a GMM estimation method. Then, we compare finite sample distribution properties between GMM and Quasi-Maximum Likelihood by Monte Carlo Simulation methods. Finally, estimated models goodness of fit are evaluated by Mean Squared Errors.

Key words: Diffusion Models, GMM method, Quasi-Maximum Likelihood and Brazilian short run interest rates.

JEL Classification: C15, C52 e G12.

[†] Os autores agradecem a Edson Costa Bignotto, Rodrigo De Losso, Marinho Gomes Andrade-Filho e Fernando Postal pelos comentários. Entretanto, erros e omissões são de nossas responsabilidades.

* Mestrando em Economia pelo IPE/USP. Av. Prof. Luciano Gualberto, 908. CEP: 05508-900 São Paulo/SP. E-mail: alandgd@yahoo.com.br

‡ Departamento de Economia, FEA-RP/USP. Av. Dos Bandeirantes, 3900. CEP: 14040-900, Ribeirão Preto/SP. Fone: 55-16-602-3915. E-mail: mbarossi@usp.br

I - Introdução

A estrutura da taxa de juros, ETJ, assim como sua evolução tem sido um tema amplamente estudado e discutido nos últimos 20 anos. A ETJ descreve a relação entre o rendimento de um título livre de risco para diferentes maturidades. Vários modelos têm sido desenvolvidos utilizando-se do arcabouço de processos estocásticos em tempo contínuo. A determinação da estrutura da taxa de juros é, freqüentemente, interpretada como o problema de apreçamento de títulos com cupom zero, a partir de argumentos relacionados à não arbitragem e/ou ao equilíbrio geral. Nos modelos tradicionais admitimos apenas uma variável de estado: a taxa de juros de curto prazo. Assim o processo de difusão é utilizado para representar a dinâmica da mesma. Adicionalmente, o conhecimento da trajetória do preço de um ativo financeiro e/ou das taxas de juros são atributos relevantes ao gerenciamento de risco de mercado.

Ao longo desse mesmo período de tempo, o uso de equações diferenciais estocásticas (SDE) tem se tornado uma prática indispensável ao desenvolvimento de modelos aplicados a finanças e economia. Merton (1990) constitui-se numa referência fundamental a essa afirmação na literatura recente.

Enquanto, teoricamente, pressupõe-se o conhecimento dos parâmetros que caracterizam os processos de difusão, na prática estes parâmetros são desconhecidos e precisam ser estimados. Por sua vez, a maior dificuldade na estimação dos mesmos está no fato de que os modelos são formulados em tempo contínuo, porém as informações acerca das variáveis estão disponíveis em freqüência discreta.

Dessa forma um número considerável de métodos tem sido propostos com o intuito de estimar os parâmetros implícitos da equação diferencial estocástica, a partir de dados amostrados em freqüências discretas. Alguns desses procedimentos estão baseados no método de momentos generalizados, GMM, como propostos por Hansen (1982), Hansen & Scheinkman (1995) e Duffie & Glynn (1997). Outros estão baseados em simulações por inferência indireta, como em Gouriéroux, Monfort & Renault (1993) e, o método dos momentos eficientes EMM de Gallant & Tauchen (1996). A vantagem do método dos momentos generalizados, GMM, sobre os métodos analíticos constitui-se no escopo de modelos a que se aplica.

Métodos não-paramétricos foram desenvolvidos por Aït-Sahalia (2002), Jiang & Knight (1997) e Stanton (1997), entre outros. Recentemente, os métodos bayesianos têm sido aplicados a processos de estimação de parâmetros de equações diferenciais estocásticas (SDE), como em Eraker (2001) e Elerian *et al.* (2002).

Consubstanciado nos aspectos apresentados, o objetivo deste artigo é a estimação de parâmetros de equações diferenciais estocásticas capazes de modelar o comportamento da taxa de juros de curto prazo brasileira, dada pela taxa CDI. Baseado em Chan, Karolyi, Longstaff & Sanders (1992), doravante modelo CKLS, vamos reproduzir os resultados para o caso brasileiro através da estimação dos parâmetros de diferentes modelos de equações diferenciais estocásticas pelo método GMM. Dentre estes, destacamos os modelos: geométrico browniano, de reversão à média, raiz quadrada, entre outros. A estimação dos parâmetros de cada modelo é feita pelo método de momentos generalizados de Hansen (1982).

Em seguida, como objetivo específico, vamos verificar se as estimativas produzidas pelo método GMM são confiáveis, no caso em que a amostra disponível for finita. Para esta finalidade, recorreremos a um estudo de Monte Carlo, em que comparamos a distribuição empírica dos parâmetros estimados, a partir da simulação da trajetória do preço dos ativos, com os resultados obtidos pelos métodos de quase - máxima verossimilhança. Isto é relevante uma vez que a distribuição do

estimador de GMM é, assintoticamente, normal, enquanto que as observações referentes às séries temporais empíricas são finitas.

Empiricamente, a partir da estimação dos parâmetros dos modelos de interesse, comparamos a habilidade de cada um deles em replicar os valores observados, em mercado, da taxa de juros de curto prazo. A aderência desse fato aos dados empíricos é de interesse fundamental ao apreçamento de outros ativos, bem como para estratégias usuais de *hedge*.

Na seção II, discutimos as propriedades do método de estimação de máxima verossimilhança aplicado a parâmetros de modelos de difusão. Comparativamente, a seção III encerra uma discussão acerca do método GMM ao mesmo propósito. Estimativas empíricas, envolvendo a aplicação dos métodos GMM e de Quase-Máxima Verossimilhança são apresentadas na seção IV. Ainda, nesta seção incluímos um exercício de Monte Carlo com o objetivo de avaliar o grau de confiabilidade das estimativas produzidas por ambos os métodos, na estimação de modelos de difusão. Em seguida, na seção V, uma aplicação empírica ao apreçamento de um título de cupom zero no Brasil é apresentada. Finalmente, uma seção conclusiva encerra o artigo.

II - Estimação de máxima verossimilhança para o parâmetro de difusão

A aplicação de equações diferenciais estocásticas, SDE, a modelos de difusão na área de finanças tem sido uma prática extensiva nos últimos anos. Apesar de todo o instrumental analítico que o cálculo estocástico proporciona ao manuseio das equações diferenciais, existe uma dificuldade eminente nos modelos considerados, que é a estimação dos parâmetros da equação diferencial estocástica. Desta maneira os parâmetros, que podem ser unidimensionais ou multidimensionais, devem ser estimados a partir de uma ou de várias realizações do processo, caso uma realização em tempo contínuo estiver disponível, ou ainda, a partir de amostragens discretas, se o processo não for, continuamente, observado, seja devido a imprecisões do instrumento de medição, ou devido à indisponibilidade de observações em todos os instantes de tempo, ou ainda por qualquer outra razão.

Seja (Ω, \mathcal{F}, P) um espaço de probabilidade em que $(\mathcal{F}_t) = \{ \mathcal{F}_t, t > 0 \}$ é uma filtragem. Definindo o processo markoviano $\{Y_t, 0 \leq t \leq T\}$, no espaço de probabilidade acima, como a única solução forte do processo abaixo:

$$dY_t = \mu(Y_t; \theta)dt + \sigma(Y_t)dW_t \quad (1)$$

sendo $\mu : \mathcal{R} \times \Theta \rightarrow \mathcal{R}$, o coeficiente de *drift*, $\sigma : \mathcal{R} \times \Theta \rightarrow \mathcal{R}^+$, o coeficiente de difusão do processo para um conjunto compacto $\Theta \subseteq \mathcal{R}$ e, dW_t um movimento browniano. Seja P_θ uma medida de probabilidade gerada pelo processo $\{Y_t, 0 \leq t \leq T\}$ no espaço $(C[0, T], \mathcal{B}_T)$, em que $C[0, T]$ representa o espaço de funções contínuas dotadas com a norma do supremo e, \mathcal{B}_T a correspondente σ - álgebra de Borel. Defina E_θ , o operador esperança sob a medida P_θ e P_w a medida de probabilidade induzida pelo processo de Wiener.

Na definição do processo $\{Y_t, 0 \leq t \leq T\}$ acima, assumimos, implicitamente, que as condições de Lipschitz e de crescimento linear são satisfeitas. Percebe-se que a função σ não depende do parâmetro θ , desta forma podemos assumir que σ é completamente conhecido ou, se desconhecido, é independente de θ . Sendo σ uma

constante, que pode ser estimada a partir do processo markoviano representado por (1), utilizando o fato de que (Ver Basawa & Prakasa Rao, 1980) :

$$\sum \left[Y_{iT2^{-n}} - Y_{(i-1)T2^{-n}} \right]^2 \xrightarrow[\uparrow_{\infty}]{as} \sigma^2 T \quad (2)$$

Por um instante vamos assumir que σ seja conhecido e igual a um, caso o processo possa ser observado continuamente. O problema de estimação de σ é tratado logo abaixo, quando admitimos o caso em que apenas realizações discretas do processo estocástico em questão estão disponíveis. Supondo que os dados possam ser, continuamente, observados, a função de máxima verossimilhança pode ser definida desde que as seguintes condições sejam satisfeitas:

$$(I) P_{\theta_1} \neq P_{\theta_2} \text{ se } \theta_1 \neq \theta_2 \text{ em } \Theta$$

$$(II) P_{\theta_1} \left(\int_0^T \mu(Y_t, \theta) dt < \infty \right) = 1$$

Caso as condições (I) e (II) acima sejam satisfeitas para todo $\theta \in \Theta$, definimos a função de máxima verossimilhança da seguinte forma :

$$\mathcal{L} = \left\{ \int_0^T \mu(Y_t, \theta) dW_t - \frac{1}{2} \int_0^T \mu(Y_t, \theta) dt \right\} \quad (3)$$

Sendo: $\mathcal{L} := \log \frac{dP_{\theta_1}}{dP_{\theta_2}}$, e $\frac{dP_{\theta_1}}{dP_{\theta_2}}$ a derivada de Radon-Nikodym.

Definição: O estimador de máxima verossimilhança (MLE) $\hat{\theta}(Y_t)$ de θ para uma realização Y_t é definido como um mapeamento mensurável $\hat{\theta} : (C, \mathfrak{B}_T) \rightarrow (\Theta, \tau)$, tal que:

$$\frac{dP_{\hat{\theta}}}{dP_w} = \sup_{\theta \in \Theta} \frac{dP_{\theta}}{dP_w} \quad (4)$$

sendo τ uma σ -álgebra de Borel.

Se o espaço paramétrico Θ for compacto e, a derivada de Radon-Nikodym, $\frac{dP_{\theta}}{dP_w}$, contínua em θ , então pode-se mostrar que existe um estimador mensurável de máxima verossimilhança (MLE). (Ver **LEMA 3.3** em Schmetterer, 1974) Uma vez definido $\Theta \subseteq \mathfrak{R}$, como um conjunto compacto, garantimos a existência do estimador de máxima verossimilhança. Diante disto, todas as propriedades desejáveis e conhecidas de um estimador de MLE, no caso em que as variáveis são independentes e identicamente distribuídas, como eficiência, consistência e normalidade assintótica, são válidas e provadas de maneira análoga, como em Cramér (1946) para o caso de variáveis *iid*¹.

Um exemplo ilustra o método. Seja um processo estocástico exponencial dado pela equação abaixo:

¹ As provas dessas propriedades estão além do escopo deste artigo. A esse respeito, ver Prakasa Rao (1999).

$$dY_t = \theta Y_t dt + \sigma Y_t dW_t$$

em que o objetivo é estimar o parâmetro θ .

A partir de (3), o logaritmo da função verossimilhança é dado por:

$$\mathcal{L} = \left\{ \theta \int_0^T dY_t - \frac{\theta^2}{2} \int_0^T Y_t dt \right\} \quad (5)$$

Desta forma, a estimativa de máxima verossimilhança do parâmetro θ é dada por:

$$\hat{\theta} = \frac{Y(T) - Y(0)}{\int_0^T Y_t dt}$$

A variância aproximada de $\hat{\theta}$ pode ser representada por :

$$\text{var}(\hat{\theta}) = \frac{\sigma^2}{\int_0^T Y_t dt}$$

Sendo que a estimativa de σ^2 é obtida a partir da variação quadrática, expressa na equação (2).

Entretanto, empiricamente, não é possível a observação de realizações em tempo contínuo, portanto devemos trabalhar com observações amostradas em intervalos discretos. Existem, basicamente, duas formas de se proceder. Primeiro, quando a densidade de transição for conhecida em uma forma fechada, permitindo que a função de máxima verossimilhança seja calculada de forma exata. Segundo, na maioria das vezes, a densidade de transição é desconhecida ou de difícil manipulação e alguma aproximação pode ser realizada, permitindo-nos, assim, obter uma pseudo-função de máxima verossimilhança.

No caso em que a densidade de transição $p(y, t; y', t')$ for conhecida para cada ponto, associado ao fato de que Y_t seja um markoviano, então θ pode ser estimado pelo método de máxima verossimilhança, a partir da maximização da seguinte função:

$$\mathcal{L} = \prod_{t=1}^T p(y, t; y', t')$$

O estimador de MLE, $\hat{\theta}$, obtido a partir da maximização da equação acima é consistente e, normalmente, assintótico. A esse respeito, ver Billingsley (1961).

Entretanto, na grande maioria dos casos, a densidade de transição do processo $p(y, t; y', t')$ não é conhecida² e alguma aproximação deve ser realizada. Lo (1988) propõe uma solução na forma da equação de Fokker-Planck-Kolmogorov, em termos numéricos, para cada um dos valores observados, sujeito a uma condição de contorno. Assim, obtemos a função de verossimilhança a partir do fato de que o

² Apenas para exemplificar, os modelos cujas formas são fechadas são: Movimento Browniano, Vasicek e CIR.

modelo, em tempo contínuo, é dotado da propriedade de Markov. A desvantagem deste método está na intensidade computacional requerida.

Santa-Clara (1995) e Pedersen (1995a, 1995b) propõem o método da máxima verossimilhança simulada, em que aproximações consistentes são construídas a partir da discretização da SDE pelo método de Euler-Maruyama e, a função de verossimilhança é obtida, novamente, a partir da propriedade de Markov. Este método também tem a desvantagem de ser intensivo em computação, uma vez que a simulação de um número elevado de aproximações torna-se necessário, até que a convergência seja atingida.

Aït-Sahalia (1999 e 2002) sugere o uso de expansões analíticas, através da utilização de polinômios de Hermite, em lugar das simulações, como forma de aproximar, por uma forma fechada, mas desconhecida, a densidade de transição. A vantagem deste método está na obtenção de expansões analíticas que aumentam consideravelmente a precisão nas estimações, além de serem menos intensivas em computação. A desvantagem, entretanto, está no fato de que a expansão necessária à convergência do processo de difusão deve ser transformada de tal forma a torná-lo suficientemente gaussiano.

Finalmente, é possível, de uma maneira simples, aproximar a densidade de transição, quando discretizamos a equação diferencial estocástica pelo método de Euler-Maruyama, obtendo-se uma função de pseudo-verossimilhança. A discretização converge, uniformemente, para o processo de difusão original, quando o intervalo da mesma convergir para zero (ver a esse respeito Gikhman & Skorohod, 1965). Adicionalmente, é possível de se verificar que o método de Euler-Maruyama é um caso particular do método de máxima verossimilhança simulada.

Aplicando o método de Euler-Maruyama à equação (1), obtemos a seguinte equação à diferença estocástica:

$$Y_{t+\Delta} - Y_t = \mu(Y_t; \theta)\Delta t + \sigma(Y_t)(W_{t+\Delta} - W_t) \quad (6)$$

Uma vez que o modelo, em tempo contínuo, é um markoviano e, esta propriedade estende-se a qualquer sub-período, amostrado de forma discreta, a partir de uma trajetória em tempo contínuo, a densidade de transição, $p(y_t, t; y_{t+\Delta}, t + \Delta)$, do processo discretizado pode ser expressa como:

$$p(y_t, t; y_{t+\Delta}, t + \Delta) = \frac{1}{\sqrt{2\pi\sigma^2(y_t)\Delta t}} \exp\left\{-\frac{1}{2} \frac{(Y_{t+\Delta} - Y_t - \mu(Y_t; \theta)\Delta t)^2}{\sigma^2(Y_t)\Delta t}\right\}$$

Analogamente ao caso em que a discretização é exata, pode-se calcular a função de probabilidade conjunta a partir do fato que Y_t é um processo de Markov, descrito da seguinte forma:

$$p(y_{(i)}, \dots, y_{(T)}) = \prod_{i=1}^T p(y_{(i)} | y_{(i-1)}) p(y_{(0)})^3$$

Finalmente, os parâmetros do processo de difusão podem ser estimados a partir da seguinte função de pseudo-verossimilhança:

³ Em que: $y_{(i)} = y_{t+\Delta}$.

$$\mathcal{L} = -\frac{1}{2} \sum \left\{ \frac{(Y_{(i)} - Y_{(i-1)} - \mu(Y_{(i-1)}; \theta)\Delta t)^2}{\sigma^2(Y_{(i-1)})\Delta t} + \log(2\pi\sigma^2(Y_{(i-1)})\Delta t) \right\} + \log(p(y_{(0)}))$$

em que: $\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} \mathcal{L}$

Em termos da estimação dos parâmetros do processo de difusão pelo método da pseudo-verossimilhança, um ponto importante a destacar refere-se à consistência do estimador, bem como à distribuição assintótica do mesmo. Segundo Flores-Zmirou (1989), quando a sequência $\Delta_n^3 N$ for de ordem $o(1)$, prova-se que o estimador de máxima verossimilhança, $\hat{\theta}$, obtido a partir da discretização de Euler, é consistente e converge em distribuição para uma distribuição normal, cuja variância é calculada pelo critério de informação de Fischer. Analogamente ao parâmetro de *drift*, o parâmetro de difusão, estimado pelo método da variação quadrática, é consistente.

Para o caso em que $\Delta_n = \Delta$, um valor fixo; isto é, a circunstância em que a frequência em que os dados são observados for constante ao longo de todo período observado, a consistência do estimador, também é objeto de prova em Flores-Zmirou (1989). Baseados neste trabalho destacam-se alguns resultados importantes e, relevantes ao desenvolvimento deste artigo.

TEOREMA 2 (Flores-Zmirou, 1989) ⁴ : Dado o conjunto de condições de regularidade, que são satisfeitas para todo modelo linear, o estimador de mínimo contraste \hat{v}_n apresenta um viés assintótico, como definido a seguir:

$$\Delta \frac{E(\dot{b}A\dot{b})}{2E(\dot{b})} + O(\Delta)$$

Em que $\dot{b} = \frac{\partial b(x; \theta)}{\partial \theta}$ é a derivada da *drift* em relação ao parâmetro do processo de difusão. E A é o gerador infinitesimal do semi-grupo de (X_t) . O Teorema 2 acima apresenta o resultado para uma classe de estimadores conhecida por mínimo contraste, em que o estimador de máxima verossimilhança é um caso particular⁵, quando a função de contraste for: $U(\theta) = -\frac{\operatorname{Log}L(\theta)}{n}$, e o estimador de mínimo contraste for igual a: $U(\theta) = \inf\{U(\theta); \theta \in \Theta\}$. (Ver Dacunha-Castelle & Duflo, 1986)

No caso em que o estimador de contraste for único podemos, a partir do resultado do Teorema 2, reescrevê-lo da seguinte forma:

$$\hat{v}_n = \hat{v}_n - B(\hat{v}_n)$$

Sendo $B(\hat{v}_n)$ o primeiro termo do resultado do Teorema 2.

⁴ Ver Flores-Zmirou (1989) para uma prova do Teorema 2.

⁵ Como condição necessária, o modelo estatístico deve ser identificável isto é, a informação de Kullback é nula, se os dois parâmetros forem iguais, $K(F_\alpha, F_0) = 0$. Isto é, apenas se $\alpha = \theta$, para $\alpha \in \Theta$.

A magnitude do viés assintótico desse estimador é de grandeza $O(\Delta^2)$. Para a finalidade das estimações envolvidas neste artigo, temos que $\Delta = 1/252$, logo o viés assintótico será, no máximo, de grandeza 0,000015.

Definidas as condições sob as quais o estimador de contraste é consistente, podemos obter a distribuição assintótica do estimador. De acordo com Dacunha-Castelle & Duflo (1986), temos o seguinte resultado:

TEOREMA 3 (Dacunha-Castelle & Duflo, 1986): Dado um conjunto de condições de regularidade, se v_Δ for único, então:

$$\sqrt{n}(\hat{v}_n - \hat{v}_0) \xrightarrow{D} N(0, v)$$

Comparado ao estimador de máxima verossimilhança em tempo contínuo, v_0 , a variância do estimador de máxima verossimilhança, baseada em um conjunto de dados discretos, é dada por :

$$\text{var}(\hat{v}_n) = \frac{v_0}{\Delta} + O(1)$$

Adicionalmente, Yoshida (1992) e Florens-Zmirou (1989) mostram que sob algumas condições de regularidade, temos que:

$$\hat{\sigma}^2 = \frac{1}{T\Delta t} \sum_{i=1}^T (Y_{(i)} - Y_{(i-1)})^2$$

em que $\hat{\sigma}^2$ é um estimador consistente de σ^2 .

Como nos casos de variáveis *iid*, o método de máxima verossimilhança é preferido, devido a suas propriedades de consistência, eficiência e normalidade assintótica. Infelizmente, a estimação por máxima verossimilhança de maneira exata é apenas disponível em alguns poucos casos em que a densidade de distribuição é conhecida de maneira exata. Nos demais casos, é comum recorrermos a métodos de aproximação. O método recorrido neste artigo é o da discretização de Euler-Maruyama, por três razões. Primeiro, porque é intuitivo. Segundo, pela facilidade de implementação e, por fim, por ser, computacionalmente, eficiente. Lo (1988) e Melino (1994) argumentam que o estimador de máxima verossimilhança de um processo discretizado por Euler, não é consistente. De acordo com o resultado do Teorema 2, este resultado é, formalmente, provado para todos os estimadores de ML. Entretanto, diante da possibilidade de se medir o viés assintótico, podemos, então, da mesma forma, corrigi-lo a fim de que o mesmo seja dotado da propriedade de consistência, mesmo que não atinja o limite inferior de Cramer-Rao para o caso discretizado.

III - Estimação pelo método dos momentos generalizados - GMM

O método de momentos generalizados (GMM) de Hansen (1982) tem sido utilizado para a estimação dos parâmetros tanto em modelos lineares, como em modelos não lineares. A disseminação deste método decorre de algumas vantagens em relação ao método de máxima verossimilhança. A primeira relaciona-se ao fato de que o estimador de GMM prescinde do conhecimento da distribuição do processo, bem como da hipótese de normalidade. A título de validade das propriedades

assintóticas do mesmo, é obrigatório que sua distribuição seja estacionária, ergódica e, ainda, que a existência dos momentos relevantes esteja garantida. A segunda vantagem do método de GMM é uma decorrência do fato de que o estimador e o respectivo erro padrão são consistentes, mesmo na circunstância em que o erro seja, condicionalmente, heterocedástico.

Da aplicação do método de Euler-Maruyama à equação (1) e, tomando a expressão em termos dos dois primeiros momentos, podemos obter o seguinte:

$$E[Y_{t+\Delta t} | Y_t] = Y_t + \mu(Y_t; \theta)\Delta t$$

$$E[Y_{t+\Delta t} - E[Y_{t+\Delta t} | Y_t] | Y_t]^2 = \sigma^2(Y_t)\Delta t$$

Em termos de notação, podemos simplificar escrevendo $E_t[Y_{t+\Delta t}] = E[Y_{t+\Delta t} | Y_t]$ e, de uma forma geral, para uma função $g(\cdot)$ qualquer, podemos também escrever o seguinte: $E_t[g(Y_{t+\Delta t})] = E[g(Y_{t+\Delta t}) | Y_t]$. Logo, os erros de previsão Δt passos à frente podem ser escritos da seguinte forma:

$$\begin{aligned} \varepsilon_{t+\Delta t}^{(1)} &= Y_{t+\Delta t} - (Y_t + \mu(Y_t; \theta)\Delta t) \\ \varepsilon_{t+\Delta t}^{(2)} &= (Y_{t+\Delta t} - E_t(Y_{t+\Delta t}))^2 - \sigma^2(Y_t)\Delta t \end{aligned}$$

Ainda, assumindo a condição de que $\varepsilon_{t+\Delta t}^{(1)}$ e $\varepsilon_{t+\Delta t}^{(2)}$, sejam, por hipótese, ortogonais a um conjunto de variáveis instrumentais, então, daí segue-se que:

$$\begin{aligned} E[\mathbf{u}_{t+\Delta t}(\theta)] &= \mathbf{0} \\ \mathbf{u}_{t+\Delta t}(\theta) &= (\varepsilon_{t+\Delta t}^{(1)}, Y_t \varepsilon_{t+\Delta t}^{(1)}, \varepsilon_{t+\Delta t}^{(2)}, Y_t \varepsilon_{t+\Delta t}^{(2)})' \end{aligned}$$

O estimador de GMM é obtido pela substituição de $E[\mathbf{u}_{t+\Delta t}(\theta)]$ por sua contrapartida amostral, $g_T(\theta)$, tomando um conjunto de T observações amostradas de maneira discreta, em que:

$$g_T(\theta) = \frac{1}{T} \sum_{i=1}^T \mathbf{u}_i(\theta)$$

Em seguida, escolhemos os parâmetros que minimizam a seguinte forma quadrática:

$$\hat{\theta}_T = \underset{\theta \in \Theta}{\operatorname{argmim}} \left\{ J_T(\theta) = g_T(\theta)' W_T(\theta) g_T(\theta) \right\}$$

Em que $W_T(\theta)$ é uma matriz de pesos simétrica e definida positiva, que deve convergir para uma matriz simétrica definida positiva e não estocástica. Desta forma o estimador de GMM depende da escolha de $W_T(\theta)$. Hansen (1982) mostra que a escolha ótima de $W_T(\theta)$, denotada por $W_T^*(\theta)$, é dada pelo inverso da matriz de densidade espectral na frequência zero⁶. Assim, $W_T^*(\theta) = S^{-1}(\theta)$, em que :

⁶ Cochrane (2001) contém uma exposição didática do método GMM e de suas propriedades, bem como formas de estimação.

$$S(\theta) = \sum_{j=-\infty}^{\infty} E \left[\mathbf{u}_i(\theta) \mathbf{u}_{i-j}(\theta)' \right]$$

Quando utilizamos a matriz $W_T^*(\theta)$, isto implica que a distribuição de $\hat{\theta}$ pode ser dada por:

$$\sqrt{T} \hat{\theta} \stackrel{a}{\sim} N \left(\theta, (D' W_T^* D)^{-1} \right) \quad (7)$$

Neste caso, D' é a matriz Jacobiana⁷ de $g_T(\theta)$ em relação a θ .

Dentre as propriedades do estimador de GMM, destacamos que, assintoticamente, não é viesado, sendo consistente, para qualquer matriz de pesos $W_T(\theta)$. Quando a matriz $W_T^*(\theta)$ for utilizada, o estimador de GMM é, assintoticamente, eficiente na classe dos estimadores definidos pela condição de ortogonalidade.

Uma vez estimados os parâmetros que minimizam a forma quadrática, pode-se testar se o ajuste obtido é ou não adequado, tendo por referencial um teste de hipótese, construído a partir da expressão (7). O **LEMA 4.2** de Hansen (1982), permite-nos testar a hipótese nula de que o modelo estimado é adequado, de uma maneira simples e prática, dada pela seguinte expressão:

$$T g_T(\hat{\theta})' (W_T^*(\theta))^{-1} g_T(\hat{\theta}) \rightarrow \chi_{\nu-\lambda}^2 \quad (8)$$

Em que ν e λ são, respectivamente, o número de momentos e o número de parâmetros.

Adicionalmente, considere um modelo definido em tempo discreto. A média condicional de r_t é expressa por:

$$E[r_{t+1} | r_t] = \alpha + (\beta + 1)r_t$$

Desta forma, podemos afirmar que o modelo acima discretizado é semelhante a um modelo AR(1) com erros heterocedásticos condicionados. A variância condicional de r_{t+1} pode, então, ser escrita da seguinte forma⁸:

$$V[r_{t+1}^{(h)} | r_t^{(h)}] = \sigma_{t+1}^{(h)2} = \sigma_{0,h}^2 \left(|r_t^{(h)}|^\gamma + \sigma_{1,h} \right)^2 \quad (9)$$

De acordo com Broze *et. al.* (1995), a não linearidade, induzida pela forma da volatilidade, faz das técnicas tradicionais de séries temporais um ferramental inútil à determinação da estacionaridade de um modelo mais geral. Entretanto, Tweedie (1975) e Tweedie (1983) provam teoremas que solucionam o problema da estacionaridade, da ergodicidade e da existência dos momentos para o caso de

⁷ Ver Cochrane (2001).

⁸ Nos modelos apresentados na Tabela I, assumimos que o termo $\sigma_{1,h}$ seja nulo. Normalmente, a introdução de um termo como tal é motivada pela literatura envolvendo modelos ARCH (Ver Broze *et. al.*, 1995).

modelos não lineares. Os resultados estão baseados na teoria de cadeias de Markov, em espaço de estado contínuo.

PROPOSIÇÃO 1: Se $\gamma \neq 1$, o processo r_{t+1} é ergódico e estacionário de segunda ordem, se e somente se, $|\beta+1| < 1$ e $0 \leq \gamma < 1$. Se $\gamma = 1$, a condição suficiente à estacionaridade de segunda-ordem torna-se $(\beta+1)^2 + \sigma^2 < 1$. A condição suficiente à ergodicidade é dada por:

$$E \left[\log |(\beta+1)^2 + W_t| \right] < 0$$

Notamos que $\gamma \leq 1$ é uma condição necessária à validade das propriedades destacadas. No caso específico em que $\gamma = 1$, a estacionaridade fraca implica em ergodicidade.

A tabela I abaixo é ilustrativa acerca da propriedade de estacionaridade dos modelos aplicados à taxa de juros de curto prazo, assim como as restrições impostas sobre os parâmetros dos mesmos.

Tabela I - Modelos Para as Taxas de Juros de Curto Prazo: Tipos, Estacionaridade e Restrições Impostas aos Parâmetros

$$dr = (\alpha + \beta r)dt + \sigma r^\gamma dZ$$

Modelo	α	β	σ^2	γ	Estacionaridade
Merton		0		0	Não
Vasicek				0	Sim
CIR SR				1/2	Sim
Dothan	0	0		1	Não
GBM	0			1	Sim
Brennan-Schwartz				1	Sim
CIR VR	0	0		3/2	Não
CEV	0				Sim

Fonte: CKLS (1992) e Sundaresan(2000).

IV - Estimativas empíricas

Inicialmente o método de GMM foi aplicado à série de taxa de juros de curto prazo brasileira, com o intuito de replicar o artigo de CKLS. Posteriormente, aplicamos o método da máxima verossimilhança a três dos modelos mais conhecidos na literatura de SDE aplicadas à finanças: movimento browniano geométrico, modelo de Vasicek e o modelo CIR, com o intuito de comparar as estimativas dos parâmetros produzidas na aplicação de cada um desses métodos. A série utilizada refere-se à taxa CDI diária no período de 04/01/2000 a 28/12/2001. De forma distinta de CKLS, que utilizam observações mensais, optamos por utilizar taxas diárias para evitar a incorrência do viés de discretização, bem como para validar os resultados referentes à convergência, apresentados na seção II.

A despeito da não estacionaridade dos modelos: Merton, CEV e Dothan, decidiu-se por “estimá-los” pelo método de GMM, com intuito de replicar os

resultados de CKLS, embora a aplicação do método de GMM a representações não estacionárias seja um fato já conhecido.

A Tabela III contém os resultados referentes às estimativas dos parâmetros dos modelos de difusão obtidos pelo método GMM, além das estimativas dos parâmetros do modelo irrestrito. Estes resultados indicam que é possível de se verificar um processo de reversão à média na série de taxa de juros de curto prazo no Brasil, contrapondo-se ao resultado encontrado por CKLS no caso dos EUA. Entretanto, pode-se confirmar o padrão encontrado por Parisi (1998) para a série de taxa de juros do Chile.

Adicionalmente, na Tabela II reportamos o valor de equilíbrio estimado para os modelos de reversão à média e a velocidade de ajustamento dos desequilíbrios da taxa de juros de curto prazo. Por exemplo, no caso do modelo CIR SR, o valor ao qual a taxa de juros tende a reverter é da ordem de 15,95% em termos anuais, a uma velocidade de 0,1142. Os valores de corte para os testes de especificação incorretas dos modelos estão referenciados pela estatística J.

Pode-se perceber pelos resultados que nenhum modelo foi rejeitado, um resultado, no mínimo, estranho. Entretanto, Zhou (2000), realiza uma simulação de Monte Carlo, com o intuito de verificar o poder do teste J para um modelo CIR, obtendo-se o resultado de que o poder do mesmo tende a zero, quando o tamanho da amostra cresce. Diante destes resultados, é possível de se admitir que a estatística J como reportada pode não ser um critério de escolha factível aos modelos.

Como é possível de se verificar na seção 4.1, a seguir, o estimador de GMM tende a ser viesado para pequenas amostras, enquanto que o mesmo não ocorre com o estimador de máxima verossimilhança. Desta forma, aplicamos o método de quase - máxima verossimilhança aos dois modelos de reversão à media, Vasicek e CIR SR que demonstraram significância estatística a fim de se verificar a diferença entre as estimativas.

Tabela II - Resultados: Estimação por Quase-Máxima Verossimilhança

MODELO	κ	θ	σ^2	Logaritmo da Verossimilhança
<i>Vasicek</i>	0,2066**	0,1755**	0,03553**	2.713
SE	(0,046)	(0,083)	(0,012)	
<i>CIR</i>	0,1446**	0,168**	0,01936**	683,16
SE	(0,024)	(0,078)	(0,019)	
<i>GBM</i>	0,01	0	0,01869**	1.845
SE	(0,027)		(0,006)	

*Significante em 5%.

Na Tabela II acima reportamos os parâmetros de reversão à média ($\theta = 0,1755$) e a velocidade ($\kappa = 0,2066$) para o modelo de Vasicek. Desta forma, podemos comparar, diretamente, estas estimativas com aquelas obtidas pelo método GMM. Verifica-se a existência de uma diferença na magnitude das mesmas, embora os sinais sejam esperados, não importando o método utilizado. Adicionalmente, todos os parâmetros estimados para os modelos de Vasicek e CIR tornam-se significativos.

TABELA III
Resultados: Estimação de Modelos Alternativos para a Taxa de Juros de Curto Prazo
 $dr = (\alpha + \beta r)dt + \sigma r^\gamma dZ$

MODELO	α	β	σ^2	γ	J-Stat	θ	κ	σ^2	γ
<i>Full</i>	0,2843	-0,1609	0,1226	0,9814	0	0,1767	0,1909	0,1226	0,9814
<i>SE</i>	(0,3911)	(0,2133)	(0,1652)	(1,252)					
<i>Merton</i>	-0,0177	0	0,0392**	0	0,9661	0	0	0,0392	0
<i>SE</i>	(0,0316)		(0,0117)						
<i>Vasicek</i>	0,2198**	-0,1316	0,0413**	0	0,6053	0,166952	0,1316	0,0413	0
<i>SE</i>	(0,0379)	(0,2088)	(0,0122)						
<i>CIR SR</i>	0,028733**	-0,4207*	0,0657**	0,5	1,2855	0,159512	0,1142	0,0228	0,5
<i>SE</i>	(0,01386)	(0,2429)	(0,01677)						
<i>Dothan</i>	0	0	0,0127	1	2,9979	0	0	0,0127	1
<i>SE</i>			(0,0037)**						
<i>GBM</i>	0	-0,0158	0,012	1	2,151	0	0,0158	0,012	1
<i>SE</i>		(0,0171)	(0,0038)**						
<i>Brennan-Schwarz</i>	0,1466	-0,0975	0,0123**	1	2,0803	0,150413	0,0975	0,0123	1
<i>SE</i>	(0,3736)	(0,206)	(0,0039)						
<i>CIR VR</i>	0	0	0,068**	1,5	4,0032	0	0	0,0068	1,5
<i>SE</i>			(0,0021)						
<i>CEV</i>	0	-0,0064	0,0969	0,8307	2,347	0	0,0064	0,0969	0,8307
<i>SE</i>		(0,0187)	(0,1369)	(1,3203)					

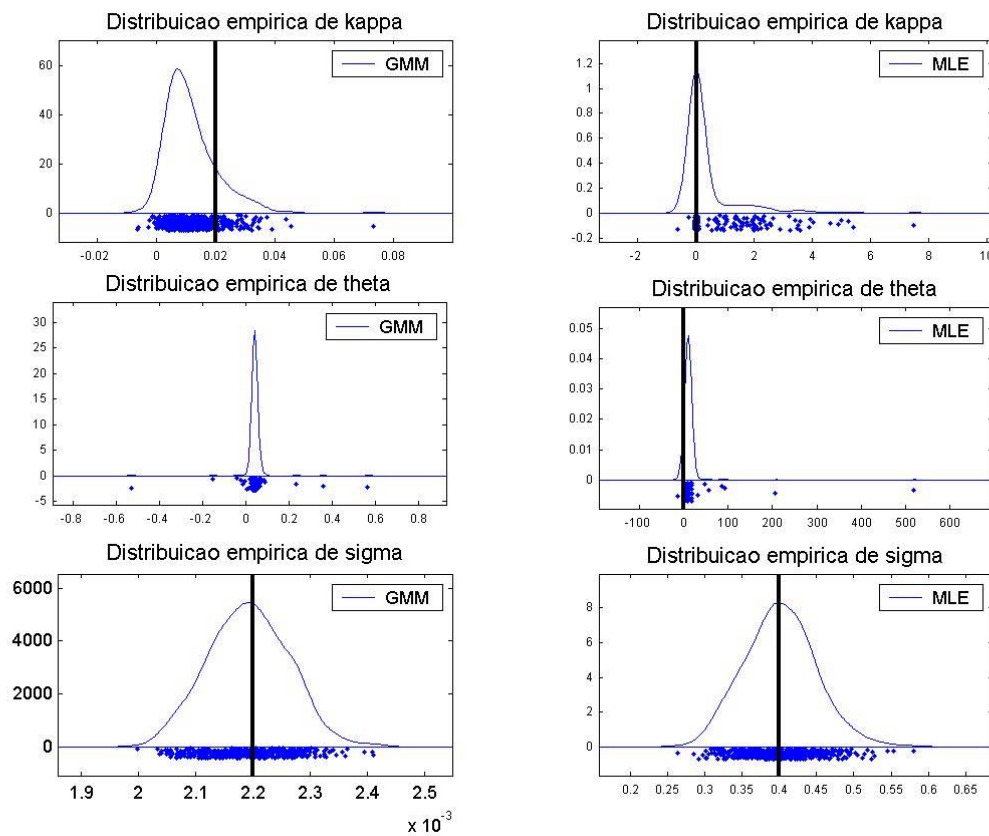
** Significante em 5%.

* Significante em 10%.

4.1 - Um exercício de Monte Carlo

A literatura internacional pertinente à estimação de parâmetros de modelos de difusão demonstra uma tendência à utilização do método de GMM, em detrimento do método de quase - máxima verossimilhança. Nos gráficos a seguir, apresentamos os resultados de um exercício de Monte Carlo para os modelos de Cox-Ingersoll-Ross, CIR, Vasicek e Browniano Geométrico, que se mostraram significativos quando das estimações de seus parâmetros nas seções anteriores. Para tanto, construímos uma série com 300 observações, para cada um dos modelos, e analisamos a distribuição empírica dos estimadores de QML e de GMM para o caso em que o número de replicações é igual a 500. Nos gráficos I, II e III, a seguir, a barra vertical escura denota o parâmetro gerador dos dados e, os eixos, os valores: estimados e verdadeiro para os parâmetros dos modelos. Os resultados para o modelo CIR são apresentados no Gráfico I, logo abaixo.

Gráfico I - Modelo Cox-Ingersoll-Ross

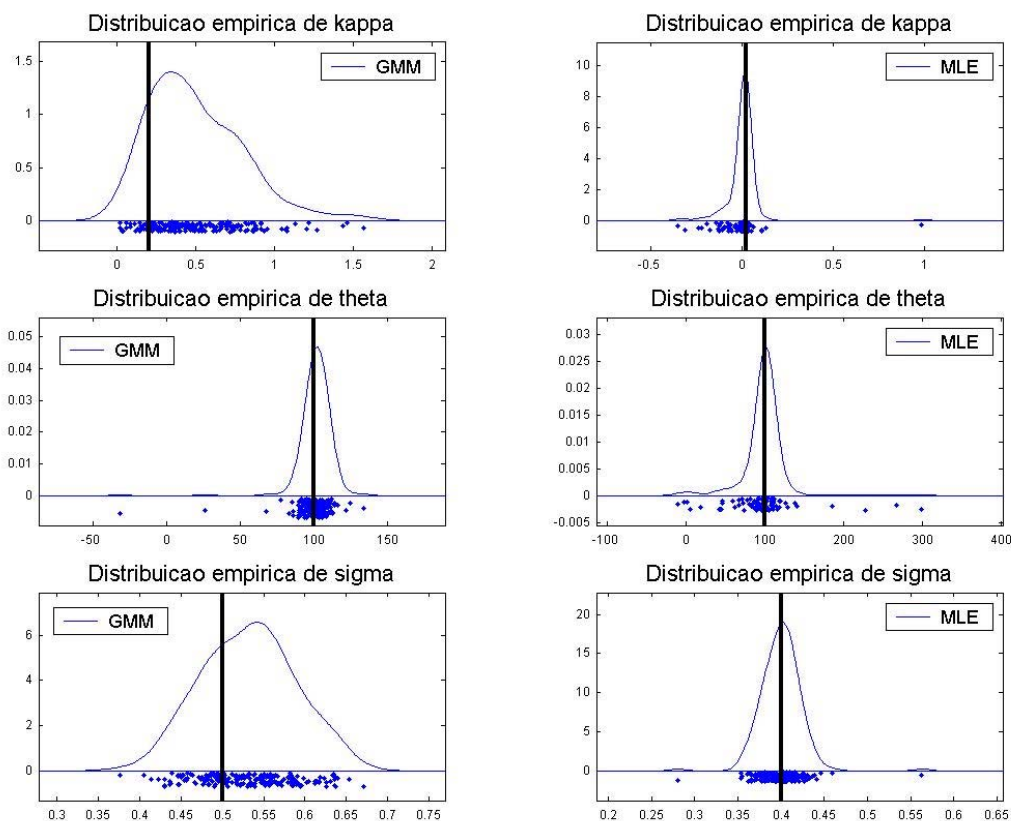


Os resultados sugerem que os estimadores de quase - máxima verossimilhança não são viesados para pequenas amostras e, são, aparentemente, distribuídos como uma normal. Entretanto, os estimadores de GMM são, para alguns dos parâmetros, viesados para a circunstância em que as amostras são pequenas. Os valores calculados para os Erros Quadráticos Médios comparativos, entre os parâmetros de mesma natureza, estimados pelos três modelos admitidos, estão na Tabela IV.

O Gráfico II abaixo contém os resultados referentes às simulações para o modelo de Vasicek. Da observação do mesmo, bem como dos resultados da Tabela IV, podemos

verificar que o estimador de quase - máxima verossimilhança mantém uma superioridade relativa, quando comparado ao estimador de GMM.

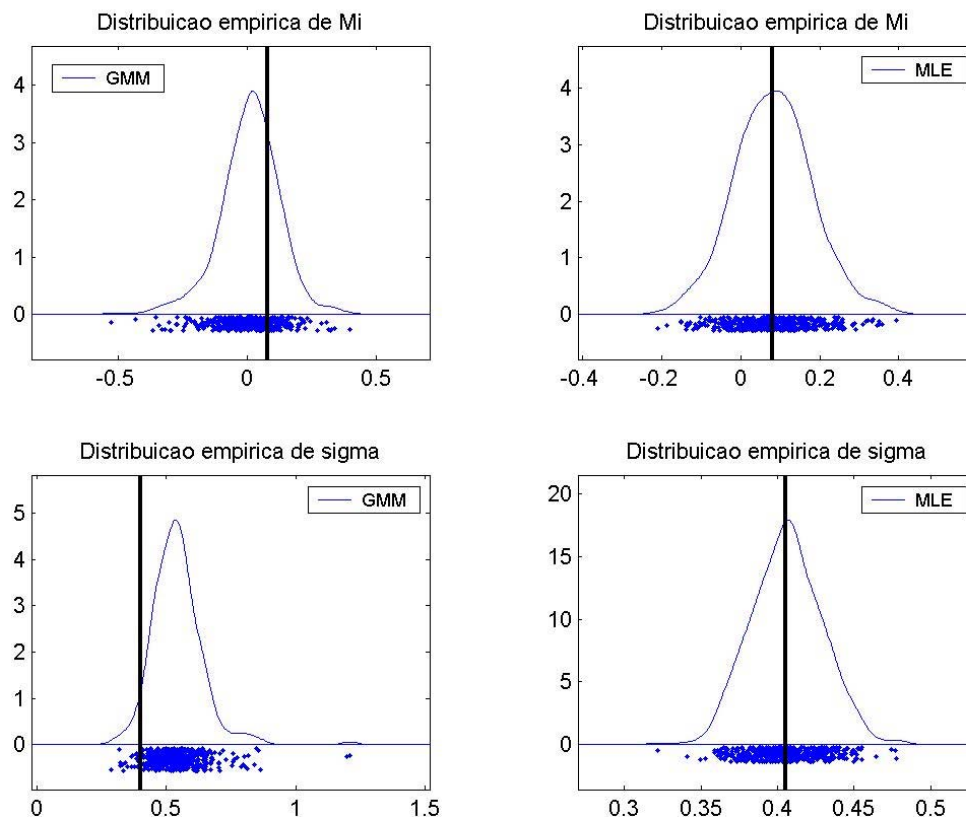
Gráfico II - Modelo de Vasicek



Por último, apresentamos no gráfico III os resultados para o modelo de Black & Scholes. Os resultados estão na mesma direção dos anteriores, isto é, os estimadores de quase - máxima verossimilhança tendem a ser não viesados para pequenas amostras e, aproximadamente, distribuídos como uma normal.

Podemos, a partir dos resultados das simulações de Monte Carlo, afirmar que o método de GMM não é o mais indicado à estimação de parâmetros de processos de difusão. Além da manifestação de um viés pronunciado, a dispersão das estimativas de GMM para os parâmetros dos três modelos de difusão, escolhidos e produzidos pela simulação de Monte Carlo, é mais elevada do que a dispersão apresentada pelas estimativas por Quase-Máxima Verossimilhança, QML. Portanto, o método QML apresenta propriedades desejáveis a um estimador, como por exemplo, a ausência de viés, a normalidade no caso de pequenas amostras e consistência.

Gráfico III - Modelo *Browniano Geométrico*



A Tabela IV abaixo resume os valores calculados para os EQMs, que confirmam as conclusões acima relatadas.

Tabela IV - Erros Quadráticos Médios das Estimativas dos Parâmetros Obtidas a Partir da Simulação de Monte Carlo

Parâmetros	CIR - SR		VASICEK		BROWNIANO	
	QML	GMM	QML	GMM	QML	GMM
<i>Teta</i>	0,0402	0,3657	1,0512	2,6150	-	-
<i>Kapa</i>	0,0123	0,3260	0,0925	0,4130	0,074	0,137
<i>Sigma</i>	0,000069	0,0488	0,1037	0,1420	0,0227	0,1711

Fonte: Elaboração dos autores.

V - Apreçamento de um título com cupom zero: uma aplicação ao Brasil

Na seção IV estimamos alguns dos modelos existentes e capazes de caracterizar o comportamento da taxa de juros de curto prazo. Dentre estes, os modelos CIR SR e Vasicek são estatisticamente significantes. Entretanto não utilizamos nenhum critério de escolha quanto ao grau de aderência dos mesmos à descrição do comportamento da taxa de juros de curto prazo no Brasil. Na seção IV apresentamos um exercício de Monte Carlo, a partir do qual é possível demonstrar que o estimador de quase - máxima verossimilhança produz estimativas mais apropriadas no que diz respeito ao viés e à normalidade para pequenas amostras, comparando-o ao método de GMM.

O objetivo desta seção é utilizar os resultados da seção IV para apreçar um título de cupom zero pelos modelos: Vasicek e CIR SR. Além disso, recorreremos ao critério do erro quadrático médio, como forma de concluirmos qual(is) dentre esses modelos produz(em) estimativas mais próximas das observações reais em mercado financeiro brasileiro.

De acordo com as características intrínsecas a um título de cupom zero comparáveis através das estimativas produzidas pelos modelos Vasicek e CIR SR, optamos pela Letra do Tesouro Nacional, LTN. Duas datas⁹ em que os vencimentos estavam disponíveis, segundo informações da ANDIMA, foram escolhidas na amostra, a partir das quais calculamos o preço teórico sugerido em cada modelo e comparamos esses valores com o valor de fato transacionado¹⁰ em mercado. Em seguida, calculamos a raiz do erro quadrático médio e a utilizamos como critério comparativo de escolha do modelo que melhor se ajusta, no período observado, à descrição da trajetória da taxa de juros de curto prazo no Brasil.

Como podemos observar, a partir da tabela V, o modelo de Vasicek apresenta um erro quadrático médio inferior ao do modelo CIR para diferentes maturidades. Desta forma, diante do fato que ambos os modelos são estatisticamente significantes e, da observação de que o modelo de Vasicek é dotado do menor EQM, então podemos escolhê-lo como o mais apropriado à descrição do comportamento da taxa de juros de curto prazo brasileira, no período de tempo analisado.

Destacamos, como ponto negativo ao modelo de Vasicek, a possibilidade de se obter taxas de juros negativas. Como estamos utilizando apenas taxas de juros nominais, isto é um problema. Entretanto, dado o patamar em que as taxas de juros oscilam no Brasil, a probabilidade de ocorrência de uma taxa de juros negativa é dada por um conjunto de pontos, cuja medida é nula.

TABELA V - Comparação: Resultados Estimados¹¹ e PU Observado

	VENCIMENTO	Vasicek	PU	CIR SR
12/09/01	03/10/01	98,79776	98,85043	99,91642
	07/11/01	96,9506	97,0015	99,45787
	09/01/02	93,71981	93,39199	97,6778
	06/02/02	92,32177	91,6852	96,52045
	03/04/02	89,5936	88,38745	93,59977
	03/07/02	85,34675	83,09401	87,39817
08/12/00	03/01/01	98,83902	98,91987	99,90808
	31/01/01	97,56346	97,72963	99,59421
	28/02/01	96,30301	96,54369	99,06456
	04/04/01	94,74855	95,04004	98,11218
	02/05/01	93,52176	93,83411	97,12988
	06/06/01	92,00909	92,34381	95,64409
	04/07/01	90,81549	91,15529	94,26321
	01/08/01	89,63647	89,97536	92,72459
	03/10/01	87,03643	87,34892	88,75173
	05/12/01	84,50836	84,75441	84,19197
	09/01/02	83,13448	83,32898	81,45967
	EQM	2,803688		12,6336

⁹ Os mesmos resultados verificam-se para um número maior de pontos no tempo ou datas, e estão disponíveis, sob requisição.

¹⁰ Os autores agradecem à ANDIMA por disponibilizar, prontamente, os dados.

¹¹ Os parâmetros necessários ao cálculo do preço do título, para cada um dos modelos, foram obtidos a partir dos valores da tabela II.

VI - Conclusões

A preocupação em reproduzir a estrutura da taxa de juros, ETJ, bem como a evolução da mesma tem sido um tema amplamente estudado e discutido nos últimos 20 anos. Vários modelos têm sido desenvolvidos utilizando-se do arcabouço de processos estocásticos em tempo contínuo a essa finalidade. Nos modelos tradicionais admitimos apenas uma variável de estado: a taxa de juros de curto prazo. Assim o processo de difusão é utilizado para representar a dinâmica da mesma.

Com base neste problema teórico, estimamos parâmetros de equações diferenciais estocásticas com o intuito de modelar o comportamento da taxa de juros de curto prazo brasileira, dada pela taxa CDI. Baseado em Chan, Karolyi, Longstaff & Sanders (1992), reproduzimos seus resultados para a série temporal brasileira através da estimação dos parâmetros de diferentes modelos pelo método GMM¹², evidenciando uma reversão à média.

Em seguida, verificamos a confiabilidade das estimativas produzidas pelo método GMM para amostras finitas. Para esta finalidade, realizamos um exercício de Monte Carlo, comparando a distribuição empírica dos parâmetros estimados, a partir da simulação da trajetória do preço dos ativos, com os resultados obtidos pelos métodos GMM e QML. Os resultados sugerem que os estimadores QML não são viesados para pequenas amostras e, são, aparentemente, distribuídos como uma normal. Entretanto, os estimadores de GMM são, para alguns dos parâmetros, viesados para a circunstância em que as amostras são pequenas¹³. Estes resultados, até certo ponto, são encontrados em Dell'Aquila *et.al.* (2003).

Diante das conclusões obtidas da simulação de Monte Carlo, estimamos os parâmetros dos modelos CIR SR e Vasicek, para os dados de CDI diários, pelo método QML. Novamente, encontramos evidências de reversão à média.

Finalmente, os modelos Vasicek e CIR SR são comparados por meio do Erro Quadrático Médio, EQM, quando o intuito é a escolha do modelo que seja capaz de apreçar, teoricamente, um título com cupom zero mais próximo ao seu valor real de mercado: neste caso, uma LTN. Como podemos observar, a partir da tabela V, o modelo de Vasicek apresenta um EQM inferior ao do modelo CIR para diferentes maturidades. Desta forma, diante do fato que ambos os modelos são, estatisticamente, significantes e, da observação de que o modelo de Vasicek é dotado do menor EQM, então podemos escolhê-lo como o mais apropriado à descrição do comportamento da taxa de juros de curto prazo brasileira, no período de tempo analisado.

¹² Ver Tabelas II e III.

¹³ Ver Tabela IV.

REFERÊNCIAS

- AÏT-SAHALIA, Yacine (1999). Transition Densities for Interest Rate and Other Nonlinear Diffusions. *The Journal of Finance*, v. LIV: 1361-1395.
- AÏT-SAHALIA, Yacine (2002). Maximum-Likelihood Estimation of Discretely-Sampled Diffusions: A Closed-Form Approximation Approach. *Econometrica*, v. 70: 223-262.
- BASAWA, I.V. & PRAKASA RAO, B.L.S. (1980). *Statistical Inference for Stochastic Processes*. London: Academic Press.
- BILLINGSLEY, Patrick (1961). *Statistical Inference for Markov Processes*. Chicago: Chicago University Press.
- BLACK, Fischer & SCHOLES, Myron (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, v. 81: 637-659.
- BROZE, Laurence, SCAILLET, Olivier & ZAKOÏAN, Jean-Michel (1995). Testing for Continuous-Time Models of the Short-Term Interest Rate. *Journal of Empirical Finance*, v. 2: 199-223.
- CHAN, K; KAROLYI, G; LONGSTAFF, Francis & SANDERS, Anthony (1992). An Empirical Comparison of Alternative Models of Short-Term Interest Rate. *Journal of Finance*, v. XLVII: 1209-1227.
- COCHRANE, John (2001). *Asset Pricing*. Princeton : Princeton University Press.
- CRAMÉR, H (1946). *Mathematical Methods of Statistics*. Princeton: Princeton University Press.
- DACUNHA-CASTELLE, Didier & DUFLO, Marie (1986). **NOME ??** V. II. New York: Springer-Verlag.
- DELL' AQUILLA, Rosario, RONCHETTI, Elvezio & TROJANI, Fabio (2003). Robust GMM Analysis of Models for the Short Term Interest Rates. *Journal of Empirical Finance*, v. 10: 373-397.
- DUFFIE, Darrel & GLYNN, Peter (1997). Estimation of Continuous-Time Markov Process Sampled as Random Time Intervals. *Mimeo*. Stanford Universty, Departamento de Economia.
- ELERIAN, Ola; SIDDHARTHA, Chib & SHEPHARD Neil (2002). Likelihood Inference for Discretely Observed Non-Linear Diffusions. *Econometrica*, v. 69: 959-993.
- ERAKER, Bjørn (2001). MCMC Analysis of Diffusion Models with Applications in Finance. *Journal of Business & Economics Statistics*, v. 17: 177-191.
- FLORES-ZMIROU, D. (1989). Approximate Discrete-Time Schemes for Statistics of Diffusion Processes. *Statistics*, v. 20: 547-557.
- GALLANT, Roy A. & TAUCHEN, George (1996). Which Moments to Match? *Econometric Theory*, v. 12: 657-681.
- GIHKMAM, I & SKOROHOD, A. V (1965). *Introduction to Theory of Random Process*. New York: W.B. Saunders Company.
- GOURIÉROUX, Christian & MONFORT, Alain (1996). *Simulation-Based Econometric Methods*. Core Lectures, Oxford: Oxford University Press.
- GOURIÉROUX, Christian; MONFORT, Alain & RENAULT, Eric (1993). Indirect Inference. *Journal of Applied Econometrics*, v. 8: 85-118.
- HANSEN, Lars Peter (1982). Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica*, v. 50 : 1029-1054.
- HANSEN, Lars Peter & Scheinkman, José A. (1995) Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes. *Econometrica*, v. 63: 767-804.
- JIANG, G. L. & KNIGHT, J. L.(1997). A Nonparametric Approach to the Estimation of Diffusion Process, with Application to a Short-Term Interest Rate Model. *Econometric Theory*, v. 13: 615 - 645.

- LO, Andrew W. (1988). Maximum Likelihood Estimation of Generalized Itô Processes with Discretely Sampled Data. *Econometric Theory*, v. 4: 231-247.
- MERTON, Robert (1990) *Continuous-Time Finance*, 2^a. ed., Malden, MA: Blackwell Publishers.
- MELINO, Angelo (1994) Estimation of Continuous-Time Stochastic Processes in Finance, *Advances in Econometrics, Sixth World Congress, Vol II*, C. Sims (ed.), Cambridge: Cambridge University Press.
- PARISI, Franco (1998). Tasas de Interés Nominal de Corto Plazo em Chile: Una Comparación Empírica de sus Modelos. *Cuadernos de Economía*, v. 35: 161-182.
- PEDERSEN, A. (1995a). A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations. *Scandinavian Journal of Statistics*, v. 22: 55-71.
- PEDERSEN, A. (1995b). Consistency and Asymptotic Normality of an Approximate Maximum Likelihood Estimator for Discretely Observed Diffusion Processes. *Bernoulli*, v. 1: 257-279.
- PRAKASA RAO, B.L.S. (1999). *Semimartingales and Their Statistical Inference*. Boca Raton: CRC Press.
- PRAKASA RAO, B.L.S. (1999). *Statistical Inference for Diffusion Type Process*. London: Arnold.
- SANTA-CLARA, P. (1995). Simulated Likelihood Estimation of Diffusions with an Application to the Short Term Interest Rate. *PhD. Dissertation*, INSEAD.
- SCHMETTERER, L. (1974). *Introduction to Mathematical Statistics*. Berlin: Springer-Verlag.
- STANTON, Richard (1997). A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk. *Journal of Finance*, v. 52: 1973-2002.
- SUNDARESAN, Suresh M. (2000). Continuous-Time Methods in Finance: A Review and an Assessment. *The Journal of Finance*, v. LV: 1569-1622.
- TWEEDIE, R.L. (1975). Sufficient Conditions for Ergodicity and Recurrence of Markov Chains on a General State-Space. *Stochastic Processes and Their Application*, v. 3: 385-403.
- TWEEDIE, R.L. (1983). The Existence of Moments for Stationary Markov Chains. *Journal of Applied Probability*, v. 20: 191-196.
- YOSHIDA, Nakahiro (1992). Estimation for Diffusion Process from Discrete Observation. *Journal of Multivariate Analysis*, v.41: 220-242.
- ZHOU, Hao (2000). A Study of the Finite Sample Properties of EMM, GMM, QMLE, and MLE for the Squared-Root Interest Rate Diffusion Model. *Working Paper FED Washington*.