TEXTO PARA DISCUSSÃO Nº 449

ESTIMAÇÃO DE HIPERPARÂMETROS EM MODELOS DE PREVISÃO*

Hedibert Freitas Lopes**
Alexandra Mello Schmidt**
Ajax R. Bello Moreira***

Rio de Janeiro, dezembro de 1996

** Do Instituto de Matemática da UFRJ.
*** Da Diretoria de Pesquisa do IPEA.

O IPEA é uma fundação pública vinculada ao Ministério do Planejamento e Orçamento, cujas finalidades são: auxiliar o ministro na elaboração e no acompanhamento da política econômica e prover atividades de pesquisa econômica aplicada nas áreas fiscal, financeira, externa e de desenvolvimento setorial.

TEXTO PARA DISCUSSÃO tem o objetivo de divulgar resultados de estudos desenvolvidos direta ou indiretamente pelo IPEA, bem como trabalhos considerados de relevância para disseminação pelo Instituto, para informar profissionais especializados e colher sugestões.

ISSN 1415-4765

SERVIÇO EDITORIAL

Rio de Janeiro – RJ
Av. Presidente Antônio Carlos, 51 – 14º andar – CEP 20020-010
Telefax: (021) 220-5533
E-mail: editrj@ipea.gov.br

Brasília – DF
SBS Q. 1 Bl. J, Ed. BNDES – 10º andar – CEP 70076-900
Telefax: (061) 315-5314
E-mail: editbsb@ipea.gov.br

SUMÁRIO

RESUMO

Os modelos de previsão têm sido adotados com uma formulação que utiliza hiperparâmetros para tornar os modelos mais flexíveis, como, por exemplo, os fatores de desconto da variância, dos parâmetros do modelo dinâmico bayesiano, ou os coeficientes da Priori de Litterman do modelo auto-regressivo vetorial bayesiano (BVAR) ou os coeficientes não-lineares do modelo de função de transferência. Surge então o problema da sua valoração apropriada para estimação dos parâmetros de interesse.

Os hiperparâmetros introduzem não-linearidades e, em geral, não existe solução analítica para o cálculo do seu valor mais provável, ou da sua distribuição a posteriori, exigindo métodos numéricos para a obtenção dos dois resultados. Este artigo compara os resultados — previsão, parâmetros e capacidade preditiva — condicionais ao valor mais provável do hiperparâmetro com os não-condicionais derivados do levantamento da distribuição a posteriori através do método de amostragem por importância (MCIS), análogo ao método (SIR) de amostragem e reamostragem por importância. Esta comparação é feita para as três diferentes utilizações de hiperparâmetros já mencionadas, em modelos de previsão da balança comercial brasileira.

Depois de estimada a posteriori dos hiperparâmetros do modelo BVAR, utilizamos o Fator de Bayes para comparar a **performance** entre os modelos que não utilizam nenhuma informação a priori e aqueles que utilizam a Priori de Litterman.

ABSTRACT

Forecasting models have been widely used with hyperparameters which turns models more flexible. Examples of hyperparameters are the discount factors, or variances of the state equations in the Dynamic Linear Models, coefficientes of Litterman´s priori in Bayesian Vector Autoregression Models and the estimate coefficients in non-linear transfer function models. One problem that arises is how to make the apropriate choice of the value of the hyperparameter in order to have the estimation of the parameters of interest.

The use of hyperparameters introduce non-linnearity and, in general, there is not an analytical solution to estimate their mode — empirical bayes — or posterior distribution. This paper compare hyperparameters estimation, by numerical methods, using empirical bayes with posterior estimation by Monte Carlo Importance Sampling (MCIS) which is analogue to the Sampling Importance Resampling Method (SIR). We have compared the three different hyperparameters especification indicated above, for trade balance forecast models.

After obtain the posterior of the hyperparameters in the BVAR model the Bayes Factor is used to compare the performance of the model using no information a priori with that which makes use of the Litterman's Priori, so the effect of using priori is discussed for that model.

1 - INTRODUÇÃO

A literatura de modelos de previsão[1] tem adotado freqüentemente uma formulação que utiliza hiperparâmetros como forma de tornar mais flexíveis estes modelos. Por exemplo, os hiperparâmetros são utilizados no modelo dinâmico bayesiano (MDB) [West e Harrison (1989)], para definir os fatores de desconto, ou no modelo estrutural [Harvey (1989)], para estimar a variância das equações de estado, ou no modelo auto-regressivo vetorial bayesiano (BVAR), para definir as variâncias a priori [Litterman (1986)] dos parâmetros auto-regressivos.

Os hiperparâmetros quantificam propriedades dos modelos que são constantes ao longo da amostra e que condicionam e estabelecem relações não-lineares com os demais parâmetros do modelo.[2] Nos modelos clássicos, os hiperparâmetros são estimados utilizando algoritmos de otimização[3] para obter o estimador de máxima verossimilhança (EMV) e a incerteza de seu estimador pode ser medida localmente através do Hessiano no ponto de ótimo ou globalmente levantando a sua distribuição utilizando procedimentos como o **bootstrap** [Efron (1982)].

Nos modelos bayesianos, os hiperparâmetros podem ser valorados de forma independente da amostra para que os modelos tenham certas propriedades desejadas,[4] ou podem ser considerados como "parâmetros" do modelo, e, então, são variáveis aleatórias cuja distribuição a posteriori é estimada utilizando métodos numéricos.[5] Obtida a distribuição a posteriori, é possível derivar as distribuições marginais dos parâmetros do modelo propriamente dito, bem como da previsão e de medidas de capacidade preditiva do modelo. No entanto, estes resultados não-condicionais aos hiperparâmetros podem ser de cálculo muito oneroso.

No caso em que a distribuição a posteriori dos hiperparâmetros tem a forma unimodal e é concentrada em torno do seu máximo, os resultados não-condicionais ao hiperparâmetro são bem aproximados pela priori empírica que maximiza a verossimilhança preditiva — **empirical bayes** —, isto é, podemos usar a aproximação: $P(\theta|X) \approx P(\theta|\lambda^*,X)$ [Bernardo e Smith (1994)]. Como em geral esta distribuição pode não ser unimodal, ou suficientemente concentrada em torno do seu máximo, é uma questão empírica em que situações o EMV utilizado no **empirical bayes** pode substituir o estimador não-condicional derivado do MCIS/SIR.

[1] Modelos em que os parâmetros evoluem no tempo e são estimados utilizando filtro de Kalman.
[2] Portanto, são "parâmetros" de um outro tipo e, por isto, são chamados de hiperparâmetros.
[3] Necessários devido à dificuldade de se obter a expressão analítica do estimador. Além disso, para que estes algoritmos não convirjam para máximos locais, são necessários cuidados especiais.
[4] E aqui os hiperparâmetros são informação a priori e não estimada a partir da amostra.
[5] Também necessários devido à dificuldade da derivação da expressão analítica da distribuição.

Este artigo compara[6] os resultados não condicionais ao hiperparâmetro, obtidos com a metodologia de amostragem por importância,[7] com os condicionais ao valor mais provável — que chamaremos de estimador de máxima verossimilhança (EMV) — para diversos modelos de previsão da balança comercial [ver Moreira e Fiorencio (1996)]. As variáveis utilizadas nestes modelos são as exportações de manufaturados, importações não-petróleo, PIB e a taxa de câmbio real. Utilizando essas variáveis, foram analisados:

a) Um modelo BVAR, para a previsão conjunta das variáveis, estimado por equação e conjuntamente. Nesse caso, os hiperparâmetros são os coeficientes de "aperto" da Priori de Litterman. No caso de estimação conjunta, foi utilizada a propriedade do VAR ser um modelo de componentes comuns e, portanto, foi estimado como um MDB de componentes comuns (MDBCC) [Quintana (1985), Barbosa (1989) e West e Harrison (1989)].

b) MDB de decomposição em tendência e sazonalidade. Nesse caso, os hiperparâmetros são os fatores de desconto, da tendência e da sazonalidade.

c) MDB com função de transferência. Esses modelos admitem que o efeito da taxa de câmbio real sobre as exportações (importações) se dá através de um mecanismo de saturação. Neste caso, o hiperparâmetro é o coeficiente que mede a velocidade de saturação que corresponde a um coeficiente de uma função de transferência de primeira ordem.

O efeito da Priori de Litterman sobre as previsões, como garante a teoria assintótica, diminui com o aumento da amostra. Portanto, é uma questão empírica verificar a sensibilidade do modelo à escolha da priori em modelos de pequenas amostras. Para essa avaliação, é proposto um procedimento que utiliza o Fator de Bayes Acumulado (FBA), que compara o modelo estimado com a Priori de Litterman a um modelo equivalente, porém com priori vaga. Utilizando este esquema foi possível avaliar, ao longo da amostra, o efeito da Priori de Litterman nos modelos BVARs acima referidos.

O artigo está desenvolvido da seguinte maneira: na Seção 2 é sumariada a utilização do método de amostragem por importância para a estimação da distribuição a posteriori dos hiperparâmetros, enquanto na Seção 3 um exercício de simulação é apresentado como uma sugestão ao número de replicações requerido para obtenção da posteriori. As Seções 4 a 6 apresentam os resultados obtidos com o modelo BVAR, com o modelo de decomposição e com o modelo de função de transferência, respectivamente. A Seção 7 conclui o artigo.

[6] Diferentes modelos da abordagem bayesiana e não modelos clássicos com modelos bayesianos. Os comentários sobre a abordagem clássica apenas situam o problema.

[7] Equivalente ao método de reamostragem por importância (SIR — Sampling Importance Re-sampling).

2 - METODOLOGIA

2.1 - Estimativa das Distribuições a Posteriori

Os modelos com hiperparâmetro podem ser colocados numa estrutura hierárquica, onde, no primeiro nível da hierarquia, é especificado o modelo para variáveis **Y,** dados os parâmetros θ; no segundo nível, é especificada a priori para θ, dado λ; e, finalmente, no terceiro nível, temos a priori para os hiperparâmetros λ. Explicitamente:

Modelo	$(Y_t \mid \theta_t)$	$\sim$	$p(y_t \mid \theta_t)$
Priori (estágio I)	$(\theta_t \mid \theta_{t-1},\lambda)$	$\sim$	$p(\theta_t \mid \theta_{t-1},\lambda)$
Priori (estágio II)	(λ)	$\sim$	$p(\lambda)$ (1)

Neste artigo, o primeiro nível da hierarquia corresponde a um MDB tal como descrito no Apêndice e que engloba os três casos que serão estudados, que são o BVAR, o modelo de decomposição e a função de transferência. Por exemplo, no caso dos modelos BVARs, o segundo nível da hierarquia corresponde à Priori de Litterman, que é definida apenas para o período inicial (t=0). Já no caso da especificação dos fatores de desconto no modelo de decomposição ou do coeficiente de saturação da função de transferência, o segundo nível da hierarquia é definido para todos os períodos da amostra. O modelo acima chama a atenção para o fato de que os hiperparâmetros são definidos de forma constante ao longo de toda a amostra e que, portanto, não serão estimados de forma recursiva.

Um exemplo simples, mas bastante complexo para ser tratado analiticamente, é:

Modelo	$(y_t \mid \theta, y_{t-1}) \sim N(\theta y_{t-1}, \sigma^2)$	
Priori (estágio I)	$\theta \mid \sigma^2 \sim N(0, \sigma^2\lambda)$	
	$\sigma^2 \sim GI(\alpha,\beta)$	
Priori (estágio II)	$\lambda \sim Beta(\xi,\zeta)$	(2)

Um exemplo similar pode ser encontrado em Berger (1985, Seção 4.6.2), onde é apresentado o núcleo da posteriori para λ. Entretanto, para casos mais gerais é necessária a utilização de métodos aproximados para obtenção de características a posteriori.

Utilizando um teorema proposto por Smith e Gelfand (1992), amostras da posteriori para λ em (1), $p(\lambda|D_T)$, podem ser obtidas tomando amostras de $p(\lambda)$ e calculando para cada λ_i sua verossimilhança $L(\lambda_i,D_T)$. Onde $D_T = \{\mathbf{y},D_0\}$, $y=\{y_1,...,y_T\}$ é a amostra e D_0 é o conjunto de informação a priori.Esquematizando, temos que:

(1) Se $\lambda_1,...,\lambda_n$ é uma amostra de tamanho n de $p(\lambda)$.

(2) Se λ^* é uma observação gerada da distribuição discreta$\{\lambda_1, \lambda_2..., \lambda_n\}$, onde:

$$p(\lambda^* = \lambda_j) = q_j = L(\lambda_j, D_T) \,/\, \sum_{k=1}^{n} L(\lambda_k, D_T) \qquad j=1,2,...,n \qquad (3)$$

(3) Então, $p(\lambda^* < a) \xrightarrow{n\to\infty} \int_{-\infty}^{a} p(\lambda|D_T)d\lambda$. Ou, de outra forma, λ^* é aproximadamente uma observação gerada da posteriori $p(\lambda|D_T)$.

Dessa forma, se for gerada $\lambda_1^*,\ldots,\lambda_m^*$, uma amostra aleatória de tamanho m da distribuição discreta, teremos, por exemplo:

$$\tilde{\lambda} = \tfrac{1}{m}\sum_{k=1}^{m} \lambda_k^* \xrightarrow{n\to\infty} \int_{-\infty}^{+\infty} \lambda p(\lambda|D_T)d\lambda = E(\lambda|D_T)$$

$$\tfrac{1}{m}\sum_{k=1}^{m} (\lambda_k^* - \tilde{\lambda})^2 \xrightarrow{n\to\infty} V(\lambda|D_T)$$

A verossimilhança $L(\lambda,D_T)$ corresponde à densidade preditiva de $\mathbf{y}=\{y_1,...y_T\}$, dado λ,D_0, e que nos MDB ou nos MDBCC tem expressão analítica (Apêndice 1), onde $D_T=\{\mathbf{y},D_0\}$:

$$L(\lambda,D_T)=L(\lambda,\mathbf{y},D_0) = p(\mathbf{y}\,|\lambda,D_0) = \int_{\Theta} p(y|\theta,D_0)p(\theta|\lambda)d\theta = E_{\theta|\lambda}[L(\theta,y|D_0)] \qquad (4)$$

A distribuição a posteriori do hiperparâmetro é dada pela expressão abaixo:

$$p(\lambda \mid D_T) = L(\lambda,D_T)p(\lambda) \,/\, \int_{\lambda} L(\lambda,D_T)p(\lambda)d\lambda \qquad (5)$$

Se a priori $p(\lambda)$ é vaga, $p(\lambda\,|D_T)$ é simplesmente a verossimilhança padronizada para se tornar uma função de probabilidade:

$$p(\lambda \mid D_t) = L(\lambda,D_t) \,/\, \int_{\lambda} L(\lambda,D_t)\, d\lambda \qquad (5')$$

Obtida a distribuição a posteriori $p(\lambda \mid D_t)$, é possível calcular as características do modelo marginal. Por exemplo, o valor esperado dos parâmetros do modelo $E(\theta|D_t)$ pode ser calculado da seguinte forma:

$$\begin{aligned} E(\theta|D_T) &= \int_{\theta} \theta p(\theta|D_T)d\theta = \int_{\theta}\int_{\lambda} \theta\, p(\theta|D_T,\lambda)p(\lambda|D_T)d\lambda d\theta \\ &= \int_{\lambda}\{\int_{\theta} \theta\, p(\theta|D_T,\lambda)d\theta\}\, p(\lambda|D_T)d\lambda \\ &= \int_{\lambda} E(\theta|\lambda,D_T)\, p(\lambda|D_T)\, d\lambda \end{aligned} \qquad (6)$$

Nos MDB é comum denotar a média a posteriori dos parâmetros θ, no instante t=T, por $E(\theta|\lambda,D_T)= m_T(\lambda)$. Assim, uma aproximação para esses momentos seria:

$$\sum_{i=1}^{n} m_T(\lambda_i) q_i^* = \sum_{i=1}^{n} E(\theta|\lambda_i, D_T) q_i^* \xrightarrow{n\to\infty} E(\theta|D_T)^{8} \quad (7)$$

No caso da amostragem por importância (MCIS), q_i^* é dado por (3), no caso da reamostragem (SIR), é a proporção de vezes que λ_i é reamostrado.[9] Vale ressaltar que, no caso da utilização do SIR, vários λ_is não serão reamostrados e, portanto, seus respectivos pesos q_i^* serão nulos; por exemplo, se λ_1 for reamostrado três vezes numa re-amostra com m elementos, então q_i^*=3/m. Essa nomenclatura permite incluir nesse esquema de amostragem o procedimento Integração por Monte Carlo com função de importância (MCIS), onde a função de importância seria $p(\lambda)$. Além disso, em (7) o limite só é válido para o procedimento SIR, se $m\to\infty$ também.[10] Maiores detalhes sobre Monte Carlo com função de importância podem ser encontrados em Van Dijk e Kloek (1980,1986) e Efron (1982).

De maneira análoga, a esperança da previsão a um passo, $E(y_{t+1}|\mathbf{y})$, pode ser calculada da seguinte forma:

$$E(y_{t+1}|D_T) = \int_\lambda E(y_{t+1}|\lambda, D_T)\, p(\lambda/D_T)\, d\lambda \quad (8)$$

Logo, uma aproximação direta seria:

$$\sum_{i=1}^{n} E(y_{t+1}|\lambda_i, D_T) q_i^* \xrightarrow{n,m\to\infty} E(y_{t+1}|D_T) \quad (9)$$

As equações (8) e (9) podem ser utilizadas para realizar projeções para mais de um período quando as variáveis explicativas não são estocásticas. Nos modelos VARs, entretanto, os regressores são estocásticos. Neste caso, um procedimento **heurístico** habitual é usado [Lutkepohl (1991, p.85)]. Por exemplo, num modelo VAR(p), a previsão para mais de um período pode ser (dado λ):

$$\hat{y}_{t+h}(\lambda) = \sum_{k=1}^{p} E(\theta_{t,k}|D_T, \lambda)\hat{y}_{t+h-k}(\lambda) \quad (10)$$

onde $E(\theta_{t,k}|D_T)$ é uma matriz de estimada no período t e relativo à defasagem k. O valor esperado da previsão incondicional a λ seria então dado por:

$$\sum_{i}^{n} \hat{y}_{t+h}(\lambda_i) q_i^* \xrightarrow{n,m\to\infty} \int_\lambda \hat{y}_{t+h}(\lambda) p(\lambda|D_T) d\lambda = E[\hat{y}_{t+h}|D_T] \quad (11)$$

[8] Naturalmente, a distribuição a posteriori dos parâmetros está definida apenas para o período final da amostra, uma vez que os hiperparâmetros foram obtidos utilizando toda a amostra. Os parâmetros dos períodos intermediários também devem ser definidos da mesma forma $p(\theta_t | D_T,\lambda)$.

[9] A seção seguinte compara os dois procedimentos — amostragem por importância (MCIS) ou re-amostragem por importância (SIR).

[10] Ou seja, quando os resultados do SIR convergem para o MCIS.

Para medir a capacidade preditiva dos modelos, seja a previsão condicional (12) — e no caso dos modelos BVARs a aproximação (12a), equivalente a (9) — e o erro de previsão condicional (12b):

$$\tilde{y}_{t+1}(\lambda) = E(y_{t+1}|D_t,\lambda) \quad (12)$$

$$\tilde{y}_{t+h}(\lambda) = \sum_{k=1}^{p} E(\theta_{t,k}|D_t,\lambda)\tilde{y}_{t+h-k}(\lambda) \quad (12a)$$

$$\tilde{e}_{t+h}(\lambda) = y_{t+h} - \tilde{y}_{t+h}(\lambda) \quad (12b)$$

Utilizando os erros de previsão condicional pode-se construir uma aproximação para o erro quadrático médio do modelo não-condicional, como em (13) abaixo, ou para qualquer outra medida de capacidade preditiva como o desvio absoluto médio (DAM) ou o índice de Theil-U:[11]

$$\sum_{i=1}^{n} \tilde{e}_{t+h}^2(\lambda_i).q_i^* \xrightarrow{n,m\to\infty} \int_{\lambda} \tilde{e}_{t+h}^2(\lambda)p(\lambda|D_T)d\lambda = E[\tilde{e}_{t+h}^2] \quad (13)$$

2.2 - Comparação dos Métodos de Aproximação

Os métodos de amostragem por importância, MCIS, e o de amostragem e re-amostragem por importância, SIR, mencionados anteriormente, são assintoticamente comparáveis. Vimos na seção anterior que Smith e Gelfand (1992) apresentam um resultado que garante, por exemplo, a seguinte convergência:

$$\frac{1}{m}\sum_{j=1}^{m} \lambda_j^* \xrightarrow{n,m\to\infty} E(\lambda|D_T)$$

Do outro lado, os métodos MCISs garantem que, por exemplo:

$$\tilde{\lambda} = \frac{\sum_{i=1}^{n} \lambda_i L(\lambda_i, D_T)}{\sum_{i=1}^{n} L(\lambda_i, D_T)} \xrightarrow{n\to\infty} \frac{\int \lambda p(y|\lambda, D_0)p(\lambda)d\lambda}{\int p(y|\lambda, D_0)p(\lambda)d\lambda} = E(\lambda|D_T)$$

Ou seja, quando n,m→∞, os dois procedimentos SIR e MCIS produzem resultados similares. Entretanto, o SIR evidencia claramente a estrutura bayesiana de estimação, além de obter boas aproximações mesmo quando M << n, por exemplo, M=0.1n [Rubin (1988), Smith e Gelfand (1992)], o que pode ser interessante, por exemplo, no caso do cálculo de estimativas dos parâmetros suavizados.

[11] Estas estatísticas foram construídas para ser comparáveis com as equivalentes do modelo condicional. No entanto, contêm a impropriedade de considerar como D_t o conjunto relevante para estimar as previsões e D_T para estimar a distribuição a posteriori do hiperparâmetro.

Tanto o método SIR quanto o método MCIS produzem, com priori vaga, uma distribuição a posteriori, $p(\lambda|D_T)$, cuja moda corresponde ao EMV, como pode ser visto na equação (5´), o que explicita a relação entre os estimadores clássicos e bayesianos. Como o estimador do hiperparâmetro não é, necessariamente, normalmente distribuído, a moda pode ser diferente da média e também os momentos de segunda ordem do estimador, medidos localmente no ponto de máximo através do Hessiano, podem ser diferentes de uma medida global $V(\lambda|D_T)$.[12] Além disto, a reparametrização freqüentemente requerida para o algoritmo de busca dificulta a estimativa da incerteza dos hiperparâmetros.

Os EMVs foram calculados, para os exercícios do artigo, utilizando o algoritmo de busca de Davidon-Flecher-Powell [Brian (1984)], que utiliza o gradiente da função de verossimilhança, computados numericamente, para calcular, a cada iteração, o Hessiano. Este algoritmo é uma generalização do método de Newton-Raphson que utiliza direções conjugadas para acelerar a busca. Como alguns dos hiperparâmetros são definidos apenas em alguns segmentos da reta, foram adotadas reparametrizações destes hiperparâmetros que permitem utilizar algoritmos não-restritos para resolver problemas com este tipo de restrição.[13]

2.3 - Fator de Bayes

Assintoticamente o efeito da priori é desprezível, portanto a Priori de Litterman seria desnecessária assintoticamente e os modelos BVARs seriam simplesmente modelos VARs. Como as amostras são finitas, é interessante verificar, em cada período do tempo, se a priori continua trazendo alguma informação ou se já existe informação suficiente no conjunto de dados. Para isto ser verificado será utilizado o Fator de Bayes Acumulado (FBA), como definido em West e Harrison (1989). O FBA permite comparar dois modelos através de suas densidades preditivas. O FBA será utilizado para comparar o modelo BVAR, onde λ é fixado em $E(\lambda|\mathbf{y})$, indicado por M_0, com um modelo VAR clássico equivalente (M_1). Assim, o logaritmo do FBA, que compara os modelos M_0 e M_1 ao longo de k observações, é dado por:

$$\log(H_t(k)) = \log(p(Y_{t-k+1},....Y_t|D_0,M_0)) - \log(p(Y_{t-k+1},.....Y_t|M_1)) \qquad (14)$$

[12]Estas duas medidas são semelhantes quando a concavidade da distribuição é aproximadamente constante.

[13] No caso dos coeficientes de Litterman que determinam uma variância, foi feita uma reparametrização — tomando o quadrado da variável utilizada no algoritmo — que garante que a busca é feita apenas no quadrante positivo. No caso da volatilidade proporcional utilizada no modelo dinâmico bayesiano, foi feita uma reparametrização que garante que os valores estejam num particular segmento da reta [0,7-1].

3 - AVALIAÇÃO DO MÉTODO SIR/MCIS

A estimativa da distribuição a posteriori utilizando o método SIR está fundamentada no teorema indicado na seção anterior que garante que, se o número de amostras tomados da priori for muito grande, os estimadores adotados convergem em distribuição para a posteriori. Entretanto, o teorema não indica a velocidade de convergência e não se tem como avaliar o tamanho necessário da amostra. Para calcular este número, foram construídos dois experimentos em que se conhecia analiticamente a posteriori. Nos dois exemplos o modelo é normal e escolhem-se prioris conjugadas, de forma que as posterioris são bem-definidas. Num exemplo reamostra-se a média da normal para uma dada variância, enquanto no outro reamostra-se a variância da normal para uma dada média.

Estes métodos tendem a fornecer resultados ruins quando a distribuição a priori e a função de verossimilhança estão afastadas entre si. Naturalmente, se as duas funções estiverem próximas, o tamanho da amostra da priori deve ser menor do que se as duas distribuições estiverem muito distanciadas. Por esse motivo, o exercício foi repetido com diferentes tamanhos de amostra e diferentes graus de distanciamento. Esse último foi medido pela probabilidade de ser sorteado da priori um valor menor ou igual ao verdadeiro valor do parâmetro.

3.1 - Modelo Normal com Variância Conhecida

O primeiro exercício considera a verossimilhança normal com média θ, desconhecida, e variância σ^2 =1. Como priori para θ usamos a distribuição conjugada normal com média μ e variância τ^2. Assim se $x_1,x_2,...,x_T$ são independentes e identicamente distribuídos, com $x_i|\theta \sim N(\theta,\sigma^2)$, onde σ^2 =1, e se $\theta \sim N(\mu,\tau^2)$, pode-se demonstrar que a distribuição a posteriori de θ é $P(\theta|X) \sim N(\mu_1, \tau_1^2)$, onde :

$$\mu_1 = \frac{T\sigma^{-2}\bar{x} + \tau^{-2}\mu}{T\sigma^{-2} + \tau^{-2}} \qquad \text{e} \qquad \tau_1^{-2} = T\sigma^{-2} + \tau^{-2}$$

O método SIR foi utilizado para estimar a distribuição a posteriori utilizando uma mesma amostra de T=50 observações normais com média, θ=0, e variância, σ^2 =1. O grau de distanciamento entre as distribuições foi medido por $P(|\theta|\leq 3)$ na priori. Assim, quanto maior o valor desta probabilidade maior será a área coincidente entre verossimilhança e distribuição a priori.

A qualidade da estimativa da distribuição a posteriori foi avaliada através da comparação com a distribuição teórica, quanto à média, à variância e ao formato da distribuição. A igualdade entre distribuições teórica e estimada foi verificada utilizando o teste de Kolmogorov.[14]

[14] A estatística de teste é $D = \max\left|F - F^*\right|$, o máximo da diferença entre a verdadeira distribuição acumulada e a distribuição acumulada aproximada.

A Tabela 1 apresenta os resultados em que cada linha descreve um exercício. Em cada caso foi sorteada uma amostra de tamanho n da distribuição a priori $\theta \sim N(\mu,\tau^2)$, para $\tau^2=1$, com a qual foram calculadas a média, a variância e o teste de Kolmogorov. Na Tabela 1 estão também apresentados a média teórica e o valor crítico do teste de Kolmogorov ao nível de 1%.[15] Os resultados completos estão no Apêndice 3.

Tabela 1
Estimativa de média dados σ^2, tamanho da amostra e distanciamento

μ	$P(\lvert\theta\rvert\leq 3)$	Média Teórica	M	Média Estimada	Variância Estimada	Teste de Kolmogorov	V.Crít. (1%)
0	0.9773	-0.0283	100	-0.0207	0.0205	0.0936	0.196
			1.000	-0.0271	0.0205	0.0500	0.062
1.2	0.7257	-0.0048	100	-0.0134	0.0163	0.0928	0.196
			1.000	0.0142	0.0175	0.0957+	0.062
1.5	0.5000	0.0011	100	0.1285	0.0482	0.2688+	0.196
			1.000	-0.0098	0.0212	0.0761+	0.062

Obs.: Nesse exercício a variância teórica é $\tau_1^2=0.0196$. + Rejeita-se H_0: distribuições iguais.

Os exercícios consideram desde uma situação favorável em que a média da priori coincide com a média da verossimilhança até uma situação em que a probabilidade de ser sorteado um valor menor do que o verdadeiro valor é de apenas 27%. Para cada caso foram consideradas amostras da priori de três tamanhos, M=(100,500,1.000), da priori. Os resultados mostram que o procedimento SIR funciona melhor se a priori engloba pelo menos 70% da região relevante e se M=1.000 pelo menos.

3.2 - Modelo Normal com Variância Desconhecida

O segundo experimento foi feito para uma distribuição normal com média conhecida e variância, σ^2, desconhecida. A distribuição para a precisão, $\phi=1/\sigma^2$, foi a distribuição conjugada gama. Assim se $x_1,x_2,...,x_T$ são independentes e identicamente distribuídos, com $x_i|\theta \sim N(\theta,\phi^{-1})$, e se $\phi \sim G(1/\sigma_0^2,1)$, pode-se demonstrar que a distribuição a posteriori de ϕ é uma gama com parâmetros $T/2+1/\sigma_0^2$ e $1+Ts^2/2$, onde:

$$s^2 = \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})^2$$

O método SIR foi utilizado a partir de uma mesma amostra de tamanho T=100 da normal padrão, isto é, $\theta=0$ e $\phi^{-1}=1$. Em cada exercício foi sorteada uma amostra de tamanho M de uma distribuição $G(\sigma_0^{-2},1)$, com a qual foram calculadas a média, a variância, e a estatísitica do teste de Kolmogorov. Na tabela, são também apresentados a média e variância teórica e o valor crítico do teste de Kolmogorov. Os resultados também mostram que a média e a variância da distribuição a posteriori são estimadas mesmo em situações muito desfavoráveis, mas que, à

[15] Algoritmo de geração de variáveis aleatórias pode ser encontrado em Ripley (1987) e Thisted (1988).

medida que aumenta o distanciamento entre as distribuições, o teste de Kolmogorov rejeita a hipótese de igualdade entre as distribuições, teóricas e estimadas, da posteriori.

Tabela 2
Estimativa de σ^2 dados a média, tamanho da amostra e distanciamento

σ_0^{-2}	p(ϕ<1)	Média Teórica	Variância Teórica	M	Média Estimada	Variância Estimada	Teste de Kolmogorov	V.Crít. (1%)
1.2	0.54	1.1099	0.0241	100	1.1060	0.0259	0.200	0.196
				1.000	1.1148	0.0214	0.056	0.062
2.8	0.10	1.1445	0.0248	100	1.1495	0.0234	0.119	0.196
				1.000	1.1362	0.0254	0.060	0.062
3.3	0.05	1.1554	0.0250	100	1.2039	0.0230	0.154	0.196
				1.000	1.1637	0.0238	0.045	0.062

Obs.: + Rejeita-se H_0: distribuições iguais.

Analisando a Tabela 2, podemos concluir que os momentos da variância a posteriori são bem calculados para qualquer distanciamento do verdadeiro valor e para qualquer tamanho da amostragem-reamostragem. Entretanto, quanto mais distante da verdadeira distribuição, mais fortemente o teste de Kolmogorov rejeita a hipótese nula de distribuições iguais. Os resultados completos estão no Apêndice 3.

Para finalizar, vale dizer que nos exercícios empíricos que analisaremos mais adiante as propriedades dos hiperparâmetros nos darão informação suficiente para garantir que, com baixa probabilidade, estejamos nas situações desfavoráveis.

4 - ANÁLISE DE UM MODELO BVAR

O coeficiente de aperto da Priori de Litterman corresponde a um modelo VAR com priori, isto é, um VAR bayesiano. Para tanto, utilizamos como exemplo o modelo apresentado em Moreira e Fiorencio (1996) e que tem as seguintes

variáveis:[16] exportações (desagregadas no conjunto dos produtos manufaturados, exceto suco de laranja, café e açúcar industrializados, e no conjunto dos demais produtos), importações (excluída a importação de petróleo), PIB e a taxa de câmbio (deflacionada pelo INPC e multiplicada pelo IPA dos Estados Unidos).[17]
Litterman (1986) apresenta uma forma de especificação a priori para os parâmetros do VAR que tem gerado bons resultados do ponto de vista preditivo em modelos empiricamente estudados. No nosso caso, y_t=(exp.,imp.,PIB,câmb.) é um vetor de dimensão quatro e a especificação VAR(p) seria:

[16] As variáveis foram medidas em logaritmos e consideradas espúrias as observações relativas a 86.4, 90.3, e 94.4 e o modelo foi estimado com dados trimestrais para o período 1975.I a 1995.IV.
[17] Esta particular desagregação pretende isolar do modelo o efeito da política de produção do petróleo nacional e as parcelas das exportações de manufaturados mais semelhantes às **commodities**.

$$y_t = \mu + \Phi D_t + \sum_{l=1}^{p} \Gamma_l y_{t-l} + \varepsilon_t$$

as distribuições a priori para $(\mu_i, \Phi_i, \Gamma_{1,i}, \cdots, \Gamma_{p,i})$ são obtidas da seguinte forma: em geral a priori para (μ_i, Φ_i) é vaga. Já os coeficientes auto-regressivos Γ_{lij} (coeficiente auto-regressivo de ordem *l* da j-ésima variável na equação da i-ésima variável do VAR) têm médias 0 e variâncias definidas por V_{ij}^{l} (Priori de Litterman):

$$V_{ij}^{l} = \left(\frac{\lambda_j}{l^d} \right)^2 \frac{\hat{\sigma}_j^2}{\hat{\sigma}_i^2}$$

onde $\hat{\sigma}_i^2$ é uma estimativa de σ^2, d é um fator de decaimento das variâncias a priori (em geral, d=1); quanto mais longa for a defasagem mais a variância se aproximará de zero e mais certeza, a priori, se terá que $\Gamma_{lij}=0$; λ_j são os coeficientes de Litterman. Se $(\lambda_j \approx 0)$, então a probabilidade, a priori, da variável afetar (y_i) é pequena; por outro lado, se $(\lambda_j \rightarrow \infty)$, esta probabilidade é alta, o que corresponde ao caso de priori não-informativa. Serão esses os coeficientes considerados como hiperparâmetros.[18]

A Priori de Litterman é definida para cada equação e por bloco de variável explicativa endógena. Neste caso, temos quatro variáveis endógenas e, portanto, um total de quatro hiperparâmetros por equação ou 16 hipeparâmetros para o conjunto das equações. Estes coeficientes podem ser especificados com restrição, por exemplo, impondo um mesmo hiperparâmetro para o mesmo bloco em todas as equações, ou o mesmo hiperparâmetro para todos os blocos de cada equação.

Os hiperparâmetros foram estimados por máxima verossimilhança e suas distribuições a posteriori foram obtidas utilizando o método MCIS/SIR. Foram sorteadas 1.000 amostras de uma distribuição a priori dos hiperparâmetros [lognormal com $E(\lambda)=6.14$ e $Prob(\lambda>40)=5\%$]. A utilização da distribuição lognormal garante que os valores sejam positivos, como convém a um desvio padrão, e os parâmetros desta distribuição foram definidos considerando que o hiperparâmetro é um coeficiente de Litterman, que, portanto, um valor de 40 implica um valor suficientemente não-informativo, e que os valores informativos do coeficiente estão centrados em torno de 6.

Foram estimadas 25 distribuições a posteriori dos hiperparâmetros (λ). Para cada uma das quatro equações, quatro por bloco e também um para todos os blocos. Na estimação conjunta, um para cada um dos quatro blocos e também um para todo o

[18]Outras informações sobre o procedimento bayesiano de estimação aplicado ao VAR para outras especificações de priori e outros tipos de abordagem podem ser encontradas em Kadiyala e Karlsson (1993), Koop (1992), Lima, Migon e Lopes (1993) e Lopes (1994).

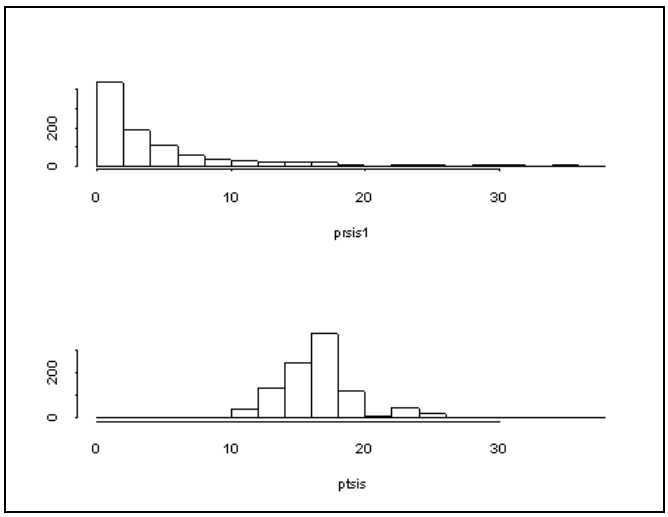
sistema. A capacidade preditiva foi avaliada com o modelo condicionado ao λ mais provável; à moda no caso de distribuição discreta; à média a posteriori de λ, $E(\lambda|D_T)$; ao estimador de máxima verossimilhança λ. Para comparação, são apresentados, também, os resultados de um modelo equivalente sem hiperparâmetros, ou seja, na forma VAR clássica.[19]

O primeiro histograma abaixo mostra para o caso de um único hiperparâmetro para todas as equações e blocos, a sua distribuição a priori e a distribuição a posteriori obtidas. A comparação dos dois ilustra o ganho de informação obtido. Os resultados para os demais casos estão sinteticamente caracterizados pela média e o intervalo de máxima densidade a posteriori — que é descrito pelos percentis 3 e 97% — apresentados nas Tabelas 3 e 4.

Tabela 3
Hiperparâmetros : média e intervalo de máxima densidade a posteriori

Equação/ Regressores	Sistema	PIB	Exportação	Importação	Câmbio
Bloco:PIB	17.7(13.6,39.0)	15.0(5.8,12.8)	5.7(0.7,30.0)	6.8(0.4,28.6)	1.3(0.1,5.5)
Bloco:Exp.	9.7(5.9,20.9)	0.3(0.2,1.1)	5.1(1.0,25.9)	3.6(0.2,10.6)	2.7(0.1,9.2)
Bloco:Imp.	12.7(7.5,20.9)	6.7(0.3,6.6)	1.1(.0,3.6)	1.6(0.1,6.2)	2.5(0.1,10.0)
Bloco:Câmbio	16.0(3.6,38.7)	3.0(0.1,2.1)	8.3(0.3,30.7)	2.6(0.0,14.3)	6.3(0.3,38.7)
Todos os Blocos	16.6(15.7,27.7)	7.2(7.1,19.0)	1.7(0.3,4.3)	0.7(0.1,2.3)	0.8(0.1,2.0)

Gráfico 1
Histograma da priori e posterioi de um coeficiente de Litterman



A Tabela 3 mostra o valor médio e respectivo intervalo de máxima densidade a posteriori.[20] Em geral, os hiperparâmetros são estimados com muita incerteza —

[19] Por economia de espaço não foram apresentados os resultados para as restrições do tipo um hiperparâmetro para o bloco da dependente defasada e outro para os demais blocos.
[20] Medidos nos percentis 3 e 97%.

intervalos de máxima densidade a posteriori muito grandes [21] — e que a hipótese de igualdade entre os hiperparâmetros estimados sem restrição e com restrição é, em muitos casos, rejeitada. Os hiperparâmetros relativos a um certo bloco, por exemplo o do PIB, no sistema e em cada uma das equações, podem ser comparados ao longo da linha. De outro lado, os hiperparâmetros dos diferentes blocos de cada uma das equações são comparados ao longo das colunas.

As tabelas a seguir utilizam a notação (M1) para indicar que a equação ou o conjunto de equações foi estimado com quatro hiperparâmetros, (M2) para indicar que a equação ou o conjunto foi estimado com 1 hiperparâmetro e finalmente (M0) para indicar o modelo estimado na forma clássica — sem utilizar a Priori de Litterman.

A Tabela 4 mostra que os logaritmos da verossimilhança preditiva (LVP) do modelo condicionado a diferentes valores do hiperparâmetro — média a posteriori, moda a posteriori ou EMV — são muito semelhantes entre si, seja considerando cada uma das equações ou para o sistema como um todo. Entretanto, as verossimilhanças são diferentes quando comparamos modelos estimados com restrições, por exemplo comparando o modelo M2 com o M1, resultado que está de acordo com a comparação dos intervalos de máxima densidade a posteriori dos hiperparâmetros. Esta tabela também mostra que o modelo estimado de forma clássica é significativamente pior do que o estimado com alguma informação a priori.

Tabela 4
LVP no VAR e BVAR com diferentes restrições [22]

Modelo	Condicional	Sistema	PIB	Exportação	Importação	Câmbio
M1	Média	-2976.8	-2780.9	-2850.3	-2849.3	-2823.1
M1	Moda	-2977.1	-2780.6	-2848.4	-2846.0	-2820.2
M1	EMV	-2977.2	-2781.6	-2848.9	-2845.5	-2820.2
M2	Média	-2976.9	-2782.8	-2850.4	-2847.7	-2821.0
M2	Moda	-2976.6	-2782.8	-2850.3	-2846.3	-2820.9
M2	EMV	-2976.9	-2782.5	-2850.4	-2846.3	-2820.9
M0	-	-2991.1	-2795.5	-2859.7	-2861.2	-2835.9

A Tabela 5 mostra que a introdução da restrição de que o hiperparâmetro de um certo bloco seja o mesmo para todas as equações reduz a verossimilhança de forma significativa.[23] Esta tabela mostra o LVP de cada equação estimada separadamente e em conjunto — caso em que foi imposta a restrição de um mesmo hiperparâmetro em cada bloco em todas as equações.

[21] E com uma distribuição a posteriori assimétrica como atesta a comparação do ponto médio do intervalo com a média. Esta assimetria já invalida a utilização das medidas habituais de incerteza de estimadores.
[22] As verossimilhanças são comparáveis por coluna.
[23] Consideramos que alterações de até uma unidade do logaritmo da verossimilhança não são significativas.

Tabela 5
LVP no BVAR condicionado à média

Modelo	Estimação	PIB	Exportação	Importação	Câmbio
M1	Conjunta	-2783.0	-2851.5	-2850.9	-2827.2
M1	por equação	-2780.9	-2850.3	-2849.3	-2823.1
M2	Conjunta	-2784.0	-2852.5	-2852.1	-2888.4
M2	por equação	-2782.8	-2850.4	-2847.7	-2821.0
M0	-	-2795.5	-2859.7	-2861.2	-2835.9

Utilizando o Theil-U[24] do erro de previsão a um período,[25] é possível comparar o modelo marginal — integrado para todos os valores do hiperparâmetro — com o modelo condicional ao valor médio do hiperparâmetro, seja na estimativa por equação ou conjuntamente.

Os resultados apresentados na Tabela 6 mostram que o efeito das restrições sobre os hiperparâmetros fica menos visível, que o modelo marginal não é substancialmente diferente do modelo condicional à média do hiperparâmetro. Aqui também (M0) é significativamente pior. As previsões apresentadas na Tabela 6 e o Theil-U do erro de previsão a três períodos mostrado na Tabela 7 têm a mesma característica.

Tabela 6
Outros resultados do modelo estimado por equação e conjuntamente

	Modelos	Condic.	Theil-U				$E(y_{t+1}/t)$			
			PIB	Exp.	Imp.	Câmb.	PIB	Exp.	Imp.	Câmb.
Equação	M1	Marginal	0.552	0.837	0.774	1.003	4.946	2.763	3.299	1.521
	M1	Média	0.541	0.841	0.797	1.055	4.942	2.774	3.297	1.521
	M2	Marginal	0.559	0.845	0.759	1.020	4.941	2.765	3.297	1.527
	M2	Média	0.556	0.838	0.764	1.017	4.936	2.766	3.297	1.525
Sistema	M1	Marginal	0.561	0.857	0.813	1.118	4.934	2.801	3.287	1.509
	M1	Média	0.558	0.857	0.815	1.120	4.932	2.805	3.287	1.508
	M2	Marginal	0.536	0.870	0.839	1.136	4.932	2.811	3.282	1.502
	M2	Média	0.552	0.870	0.840	1.137	4.931	2.813	3.281	1.502
	M0	-	0.780	1.200	1.156	1.355	4.922	2.836	3.270	1.487

[24] Theil-U é uma medida da comparação do erro de previsão (e) com o obtido de um modelo ingênuo $E(y_t/t\text{-}1)=y_{t-1}$, que faz sentido para séries integradas. A medida é $TU=(\sum e_t^2/\sum (y_t-y_{t-1})^2)^{1/2}$.

[25] Para o modelo marginal foi calculado o Theil-U, mas não foi calculada a verossimilhança preditiva.

Tabela 7
Theil-U para o erro de previsão a três trimestres no sistema

Mod.	Condicional	PIB	Exportação	Importação	T.Câmbio
M1	Marginal	0.799	1.075	1.160	1.173
M1	Média	0.810	1.080	1.170	1.181
M2	Marginal	0.831	1.077	1.188	1.201
M2	Média	0.833	1.079	1.191	1.204
M0	-	0.936	1.129	1.362	1.302

4.1 - Fator de Bayes para Comparação VAR e BVAR

Os resultados da seção anterior deixam claro que os modelos que utilizam BVAR são sistemática e significativamente melhores do que os modelos VAR, quando considerada toda a amostra. Mas pode ocorrer que o efeito dos hiperparâmetros diminua ao longo da amostra. Os gráficos a seguir apresentam o FBA descrito na Seção 2.2 com uma janela de 12 períodos. O conjunto de gráficos refere-se à comparação dos modelos condicionados ao valor esperado do hiperparâmetro com um modelo clássico. O primeiro refere-se ao sistema e outros aos modelos do PIB, das exportações de manufaturados, das importações não-petróleo e da taxa de câmbio. Em todos os casos em que o gráfico do FBA estiver no intervalo entre as linhas (-2,2), o modelo BVAR é considerado semelhante ao VAR; caso contrário, será melhor se estiver acima de 2 e pior se estiver abaixo de -2.

Gráfico 2
FBA(12) - Sistema

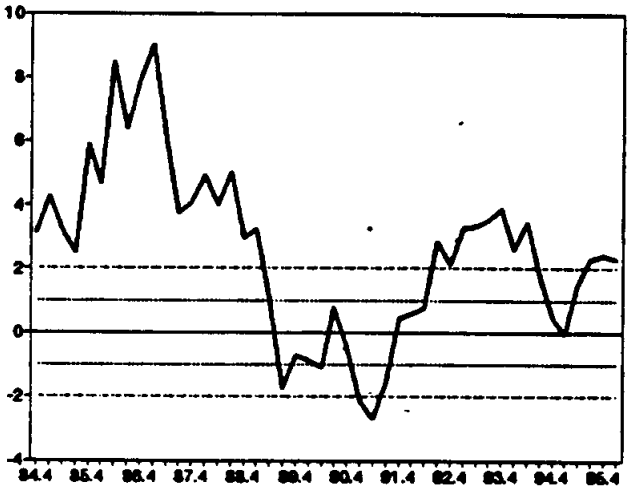
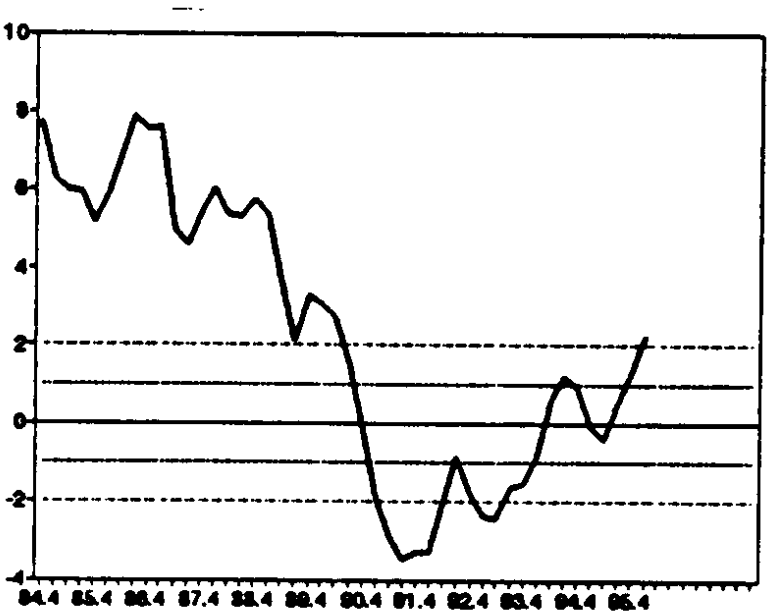


Gráfico 3
FBA(12) - PIB



No início da amostra o modelo BVAR é significativamente melhor, mas, na medida em que a amostra cresce, o modelo BVAR é tão bom quanto o modelo VAR. Nesta circunstância, o FBA muda sua inclinação por volta da década de 90 em todos os casos, sugerindo a ocorrência de uma mudança estrutural não considerada. De fato, o modelo VAR equivale a um BVAR estimado com variância a priori muito grande e, portanto, a variância dos seus parâmetros é maior possibilitando maior facilidade de ajustamento. Neste caso, parece recomendável, do ponto de vista bayesiano, utilizar os procedimentos de monitoramento e intervenção.

Gráfico 4
FBA(12) - Importação

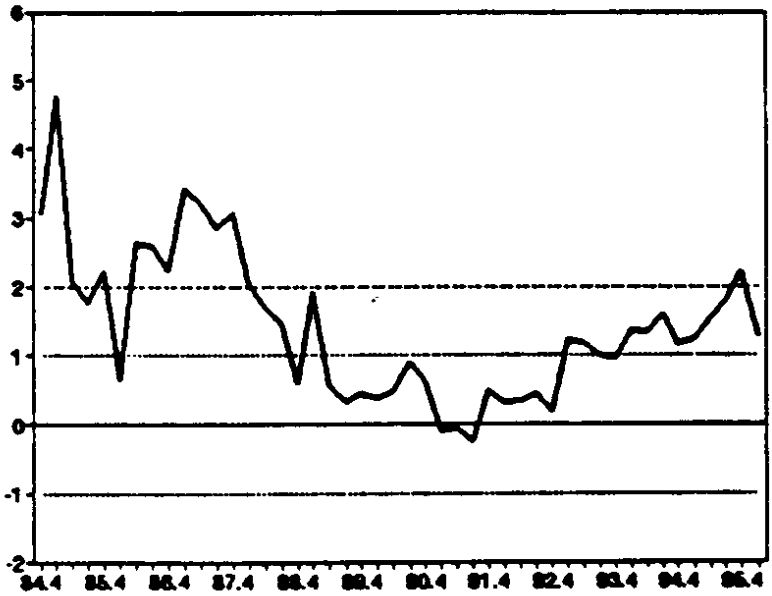
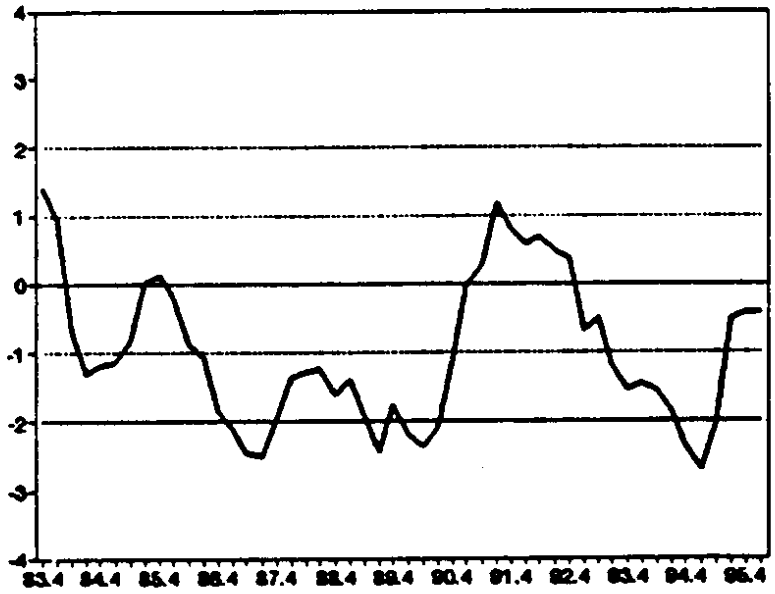


Gráfico 5
FBA(12) - Taxa de Câmbio



5 - ANÁLISE DOS MODELOS DE TENDÊNCIA E SAZONALIDADE

Nos MDB pode ser especificado o fator de desconto que determina a volatilidade de cada parâmetro.[26] Em geral, não existe um critério para definir o fator de desconto, mas no caso[27] do modelo de decomposição tendência e sazonalidade estes fatores podem ser arbitrados.[28] Naturalmente, esse fator pode ser tratado como um hiperparâmetro e estimado com o método MCIS/SIR ou por máxima verossimilhança.

O modelo de decomposição é dado por:

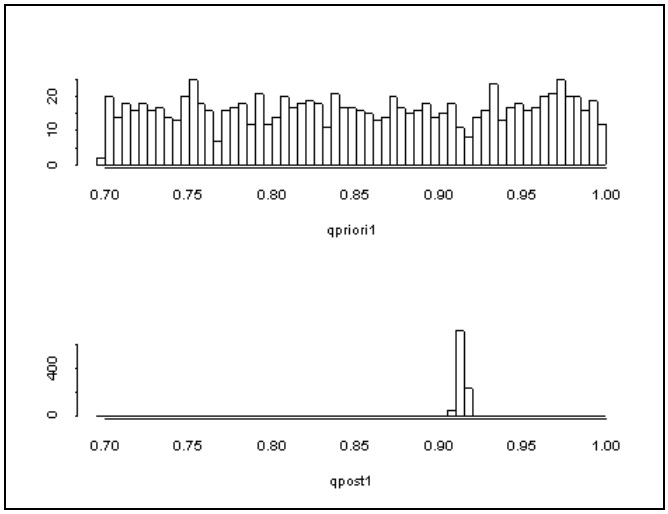
$$
\begin{aligned}
y_t &= \mu_t + S^1_t + e_t & e_t &\approx N(0,V_t)\\
\mu_t &= \mu_{t-1} + \beta_{t-1} + \xi^1_t & (\xi^1_t,\xi^2_t)' &\approx N(0,W_t)\\
\beta_t &= \beta_{t-1} + \xi^2_t & &\\
S_t &= \phi\, S_{t-1} + \xi^s_t & \xi^s_t &\approx N(0,W^s_t)
\end{aligned}
$$

onde μ é a tendência estocástica, β é a inclinação, S é um vetor de (s-1) componentes sazonais — onde a primeira componente é uma parcela da equação de observação — ϕ é uma matriz de rotação de periodicidade (s), $(\xi^1,\xi^2)'$ são os choques estocásticos sobre o bloco da tendência, ξ^s, é o vetor de choques sobre as componentes sazonais. Além disso, V é a variância do modelo, C_t e C^s_t são, respectivamente, as matrizes de covariância a posteriori do bloco de tendência e sazonalidade, $W_t=C_t/\lambda^1$, $W^s_t=C^s_t/\lambda^2$. Dados os hiperparâmetros $(\lambda^1\lambda^2)$ que correspondem aos fatores de descontos, este modelo é um MDB que é estimado segundo o algoritmo do Apêndice 1. Para estimar os hiperparâmetros, foi adotada a metodologia da Seção 2, que foi aplicada tomando 1.000 amostras da distribuição a priori, suposta uniforme [.7,1].[29]

Este modelo é utilizado para prever as mesmas componentes da balança comercial da seção anterior.[30] Para ilustrar um resultado, o Gráfico 6 apresenta os histogramas das distribuições a priori e da posteriori do fator de desconto do bloco de tendência do modelo do PIB.

[26] Como definido no Apêndice 1 para o caso da variância proporcional.
[27] Quando as matrizes F da equação de observação e G definidas no Apêndice forem constantes no tempo.
[28] Considerando a vida média da informação (n). Neste caso o fator de desconto é igual a 3n-1/3n+1.
[29] Que corresponde ao intervalo de variação relevante para um fator de desconto.
[30] Exclusive a taxa de câmbio que não apresenta um padrão sazonal.

Gráfico 6
Histograma da priori e posteriori do fator de desconto da tendência do PIB



A Tabela 8 apresenta de forma sintética a distribuição a posteriori dos hiperparâmetros dos modelos de todas as variáveis. Em todos os casos, a média, mediana, e a moda da posteriori são aproximadamente iguais entre si e ao estimador de máxima verossimilhança. O intervalo de máxima densidade a posteriori é aproximadamente igual aos limites do conjunto M dos hiperparâmetros em que o LVP é no máximo duas unidades menor do que o LVP modal, sugerindo que este critério de comparação de LVP é razoável empiricamente.Além disso, com aproximadamente 95% de probabilidade, os fatores de desconto estão no intervalo [0.87;1].

Tabela 8
Hiperparâmetros

Modelo	Bloco	Média	Moda	EMV	Perc. 0.03	Perc. 0.5	Perc. 0.97	Mín. em M	Máx. em M
PIB	Tend	0.903	0.903	0.918	0.891	0.903	0.911	0.891	0.911
	Saz	0.984	0.998	0.999	0.997	0.998	0.998	0.997	0.999
Importação	Tend	0.921	0.931	0.914	0.885	0.911	0.935	0.885	0.935
	Saz	0.997	0.998	1.000	0.994	0.997	0.998	0.994	0.998
Exportação	Tend	0.938	0.930	1.000	0.870	0.930	0.993	0.887	0.993
	Saz	0.997	0.999	0.999	0.993	0.998	0.9998	0.999	0.999

Obs.: M ={k∈ amostra tal que |LogVero(k)-LogVero(moda)|<2.

O modelo marginal — somente possível de ser obtido por método numérico — é mais representativo do que os modelos condicionais a algum valor dos hiperparâmetros porque leva em consideração a distribuição de probabilidade a posteriori dos hiperparâmetros e, portanto, os modelos condicionados a todos os possíveis valores dos hiperparâmetros. Alternativamente, num enfoque clássico,

seria escolhido o modelo condicionado ao EMV dos fatores de desconto. Portanto, uma pergunta interessante seria: estas duas abordagens geram resultados substancialmente diferentes do ponto de vista preditivo? A Tabela 9 compara a **performance** preditiva de alguns desses modelos, utilizando como critério medidas do erro de previsão a um período, o Theil-U e o desvio absoluto médio (DAM).

Tabela 9
Resultados dos modelos

Variável	Modelo	T-U	DAM	LVP	Prev.(1)	Prev.(3)
PIB	Marginal	0.918	0.031	-	4.795	4.923
	Média	0.917	0.047	-37232.9	4.796	4.923
	Moda	0.917	0.047	-37232.9	4.796	4.923
	EMV	0.871	0.045	37233.2	4.797	4.925
Importação	Marginal	0.931	0.119	-	3.418	3.584
	Média	0.926	0.119	-39374.4	3.424	3.590
	Moda	0.943	0.120	-39374.6	3.420	3.590
	EMV	0.914	0.117	-39372.4	3.417	3.589
Exportação	Marginal	1.102	0.124	-	2.734	2.813
	Média	1.103	0.124	-38901.5	2.737	2.804
	Moda	1.080	0.121	-38896.8	2.733	2.814
	EMV	1.199	0.174	-38879.5	2.715	2.796

Os modelos condicionais e marginal são bem parecidos no que diz respeito a capacidade preditiva, pelo menos nesse exemplo empírico. Estes resultados indicam que os EMVs podem ser utilizados para estimar os momentos de primeira ordem, entretanto, utilizando o método MCIS/SIR, podemos fazer afirmativas acerca da variabilidade dos hiperparâmetros a posteriori.

A Tabela 9 indica que, no caso das importações e exportações, o EMV obtém resultados significativamente menores do que a moda obtida através do MCIS/SIR. Isto possivelmente é conseqüência de que a distribuição a posteriori é muito mais concentrada do que a priori — especialmente no caso do hiperparâmetro referente ao bloco sazonal — , e que, portanto, não foi amostrado um número suficiente de pontos no domínio de variação relevante do par de fatores de desconto utilizado. Esta característica, que está evidente nos resultados, indica que o procedimento MCIS/SIR deveria ser repetido redefinindo a priori do hiperparâmetro do bloco sazonal.

6 - MODELOS DE FUNÇÃO DE TRANSFERÊNCIA

Os hiperparâmetros também podem ser utilizados para considerar modelos não-lineares como modelos lineares condicionados ao hiperparâmetro que representa o aspecto não-linear. Como um exemplo, vamos considerar um modelo de função

de transferência de primeira ordem que representa um mecanismo de saturação e que é dado por:

$$y_t = \mu_t + E_t + S^1_t + e_t \qquad e_t \approx N(0,V_t)$$
$$\mu_t = \mu_{t-1} + \xi^1_t \qquad \xi^1_t \approx N(0,W_t)$$
$$E_t = \lambda E_{t-1} + \gamma_{t-1} X_t$$
$$\gamma_t = \gamma_{t-1}$$
$$S_t = \phi S_{t-1}$$

onde μ representa a tendência estocástica — admitida sem inclinação — S representa as componentes sazonais e E o termo da função de transferência do regressor X para y, λ mede a velocidade da convergência e γ mede a magnitude do efeito. Neste modelo, admite-se que apenas o termo de tendência tenha uma componente estocástica.

Resolvendo a equação de diferenças, temos $E_t=\gamma_t(1+\lambda L+\lambda^2 L^2+...)X_t$ mostrando que o efeito de X sobre y se dá de forma cumulativa e que o termo λ mede a velocidade com que o efeito converge para o seu valor limite $\gamma_t/(1-\lambda)$.

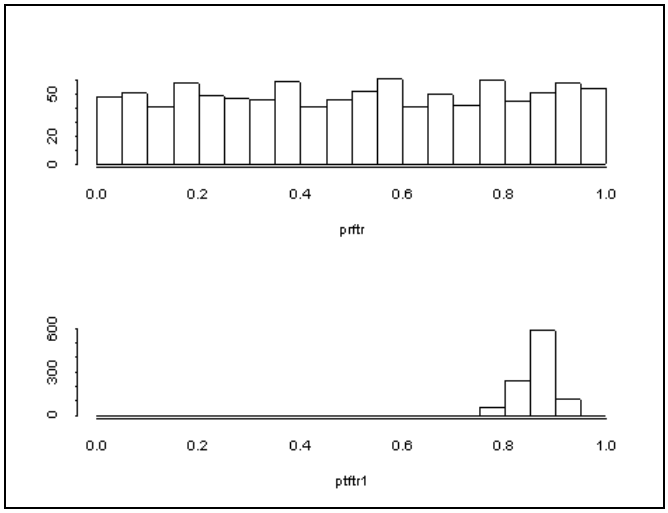
Conhecido o valor de λ, este modelo é um MDB, portanto λ pode ser considerado como um hiperparâmetro e estimado por EMV ou pelo método MCIS/SIR. Neste caso, o hiperparâmetro pode ser transformado em mais um parâmetro do modelo, transformando este num modelo não-linear, que tem mais um parâmetro (λ_t) que será estimado seqüencialmente — ao contrário do hiperparâmetro que é suposto fixo para toda a amostra.

Como um exemplo, consideremos que o efeito da taxa de câmbio sobre as importações e as exportações se dá de forma cumulativa e que este processo pode ser representado por uma função de transferência de primeira ordem do tipo acima definido. Portanto, serão considerados modelos deste tipo para as exportações de manufaturados e as importações não-petróleo estimando a medida (λ) do efeito da taxa de câmbio por EMV e pelo método MCIS/SIR. No caso das exportações de manufaturados, é apresentada também a estimativa de um modelo não-linear em que a medida de saturação foi transformada em mais um parâmetro, utilizando a aproximação proposta por Migon e Harrisson (1985).[31]

O coeficiente de saturação só faz sentido no domínio [0,1], portanto a priori foi definida como uma uniforme neste mesmo intervalo. Também aqui foram extraídas 1.000 amostras da priori. Para ilustrar os resultados, o Gráfico 7 apresenta o histograma da distribuição a priori e à posteriori do hiperparâmetro relativo a equação de exportações.

[31] A não-linearidade afeta apenas o cálculo da distribuição a priori. Os termos não-lineares são então aproximados pelo termo linear da expanção em série de Taylor em torno do valor da posteriori do período anterior. Como esta transformação da posteriori para a priori é suave, esta é uma boa aproximação.

Gráfico 7
Histograma da priori e posteriori da saturação



A Tabela 10 apresenta de forma sintética a distribuição a posteriori dos hiperparâmetros dos dois modelos. Neste caso, também o EMV é semelhante à moda e o intervalo de máxima densidade é semelhante aos limites do conjunto M. A capacidade preditiva do modelo marginal e dos modelos condicionais ao EMV, ou à média, moda da posteriori são muito semelhantes entre si e aos resultados do modelo não-linear que incorpora a velocidade de saturação como mais um parâmetro do modelo.

Tabela 10
Hiperparâmetros

Modelo	Média	Moda	EMV	Perc. 0.03	Perc. 0.5	Perc. 0.97	Mín. em M	Máx. em M	Não-Linear
Importação	0.894	0.904	0.906	0.841	0.896	0.936	0.843	0.946	-
Exportação	0.861	0.861	0.876	0.875	0.783	0.867	0.918	0.795	0.716

Obs.: M ={k∈ amostra tal que |LogVero(k)-LogVero(moda)|<2.

Tabela 11
Resultados dos modelos

Modelo	Condicionamento	T-U	DAM	LVP	Prev.(1)
Importação	Marginal	0.909	0.119	-	3.412
	Média	0.910	0.119	-38409.2	3.412
	Moda	0.908	0.119	-38409.2	3.409
	EMV	0.908	0.119	-38409.2	3.408
Exportação	Marginal	0.971	0.116	-	2.564
	Média	0.972	0.116	-37926.6	2.565
	Moda	0.972	0.116	-37926.5	2.565
	EMV.	0.970	0.116	-37926.6	2.565
	Não-Linear	0.968	0.114	-37457.2	2.556

7 - CONCLUSÃO

A introdução dos hiperparâmetros torna os MDBs mais flexíveis, mas os resultados do modelo são condicionais aos valores escolhidos dos hiperparâmetros. Uma alternativa é utilizar o MBD condicional ao EMV do hiperparâmetro. Uma outra alternativa, numa abordagem bayesiana, é levantar a distribuição a posteriori do hiperparâmetro, para obter resultados não condicionais aos hiperparâmetros.

Como é uma questão empírica avaliar se o EMV é uma boa aproximação para o MCIS/SIR, este artigo faz uma comparação empírica entre estas duas abordagens utilizando como tema modelos de previsão da balança comercial brasileira.

Nos casos estudados neste artigo, as duas abordagens fornecem resultados muito semelhantes, seja no caso do BVAR, ou como fator de desconto ou quando é um parâmetro do modelo que por conveniência foi transformado em hiperparâmetro. Naturalmente, este resultado não é generalizável para outros modelos e também outras formas de utilização dos hiperparâmetros no MDB. Para finalizar, podemos listar o que consideramos como principais vantagens e desvantagens de cada uma das abordagens.

O EMV é menos oneroso computacionalmente do que o SIR/MCIS e esta é uma vantagem sensível na atual geração de computadores,[32] por outro lado, seus resultados são ótimos locais,[33] e, devido à reparametrização necessária a maioria dos casos,[34] é difícil estimar o intervalo de confiança e a correlação entre os hiperparâmetros.

O método do MCIS/SIR fornece resultados globais, que dependem da qualidade de informação utilizada na especificação da distribuição a priori. O método não

[32] Processadores Pentium, com 90Mhz.
[33] A menos que se utilizem algoritmos que partam de diversos pontos iniciais e possam ser muito mais onerosos computacionalmente.
[34] Reparametrização que garante que a busca só se dá no conjunto de valores relevante do hiperparâmetro.

funciona bem quando a interseção da priori e da posteriori é muito pequena, seja porque a priori foi mal-especificada, seja porque um dos hiperparâmetros tem uma posteriori muito concentrada.

Resumindo, quando o objeto de interesse são apenas os parâmetros do modelo, ou estatísticas deles derivadas, o EMV pode aproximar razoavelmentge bem o MCIS/SIR — o que é uma questão empírica que se verificou em todos os casos estudados neste artigo —, mas, quando os hiperparâmetros e os seus momentos de segunda ordem são os objetos de interesse, o EMV não pode substituir o MCIS/SIR, seja devido à reparametrização dos hiperparâmetros eventualmente utilizada no procedimento de busca do EMV, seja porque a distribuição a posteriori dos hiperparâmetros não é aproximadamente normal, como foi o caso de todas as situações discutidas neste artigo.

APÊNDICE

A.1 - MDB de Componentes Comuns

O modelo condicional ao hiperparâmetro mencionado na Seção 2 é o modelo dinâmico bayesiano [West e Harrison (1989)], foi generalizado para múltiplas equações com componentes comuns por Quintana (1985) e Barbosa (1989). Este modelo tanto pode ser univariado como multivariado desde seja um modelo de componentes comuns ou que todas as equações do modelo tenham os mesmos regressores. Um exemplo importante é o modelo VAR — onde todas as equações são explicadas pelas mesmas defasagens das variáveis endógenas.

Suponhamos que existam q séries temporais univariadas Y_{tj} (j=1...q), cada uma seguindo um modelo linear dinâmico (MLD) bayesiano padrão, isto é, cada uma com a seguinte especificação:$\{F_t, G_t, V_t\sigma^2_j, W_t\sigma^2_j\}$com F_t (nx1) vetores regressores e a quádrupla conhecida, para todo t, a menos de uma escala σ^2_j (j=1...q). Assim, temos as seguintes equações das observações e de transição do sistema:

$$Y_{tj} = F_t' \theta_{tj} + \nu_{tj} \qquad \nu_{tj} \sim N(0, V_t\sigma^2_j) \qquad \text{(a-1)}$$
$$\theta_{tj} = G_t \theta_{t-1j} + \omega_{tj} \qquad \omega_{tj} \sim N(0, W_t\sigma^2_j) \qquad \text{(a-2)}$$

Note-se que F_t, G_t, V_t,[35] W_t são comuns a todos as séries e cada série tem o seu próprio vetor de estados θ_{tj}. O modelo acima, quando escrito na forma matricial, torna-se, para cada t em {1...T:

$$\mathbf{Y}_t = F_t'\boldsymbol{\theta}_t + \boldsymbol{\nu}_t \qquad \boldsymbol{\nu}_t \sim N(0, V_t\Sigma) \qquad \text{(a-3)}$$
$$\boldsymbol{\theta}_t = G_t\boldsymbol{\theta}_{t-1} + \boldsymbol{\omega}_t \qquad \boldsymbol{\omega}_t \sim N(0, W_t\Sigma) \qquad \text{(a-4)}$$

onde $\mathbf{Y}_t=(Y_{t1},...Y_{tq})$, $\boldsymbol{\nu}_t=(\nu_{t1},...\nu_{tq})$, $\boldsymbol{\theta}_t=(\theta_{tj},...\theta_{tq})$, $\boldsymbol{\omega}_t=(\omega_{tj},...\omega_{tq})$, e $\Sigma=\{\sigma^2_{ji}\}$, $\sigma^2_{jj} = \sigma^2_j$.

Suponhamos que a distribuição abaixo contenha toda a informação a priori sobre as séries estudadas conjuntamente:

$$(\boldsymbol{\theta}_0,\Sigma|D) \sim N\text{-}WI(\mathbf{m}_0,C_0,\mathbf{S}_0,n_0) \qquad \text{(a-5)}$$

ou de outra forma: $(\boldsymbol{\theta}_0|\Sigma,D_0) \sim N(\mathbf{m}_0,C_0,\Sigma)$, $(\Sigma|D_0) \sim WI(\mathbf{S}_0,n_0)$ (a-6)

implicando: $(\boldsymbol{\theta}_0|D_0) \sim T_{n0}(\mathbf{m}_0,C_0,\mathbf{S}_0)$ (a-7)

[35] No caso de modelos homocedásticos, $V_t=1$.

O teorema abaixo resume os passos da evolução e atualização — cálculo das distribuições a priori e a posteriori — dos parâmetros do modelo dinâmico bayesiano de componentes comuns, bem como as distribuições preditivas.

Teorema: As distribuições a posteriori e preditiva para cada instante de tempo t para o modelo (a-3) e (a-4) são obtidas como a seguir:

a) posteriori em t-1: suponhamos que a distribuição a posteriori em t-1 para os parâmetros do modelo $(\boldsymbol{\theta},\boldsymbol{\Sigma})$ seja caracterizada por $(\mathbf{m}_{t-1},C_{t-1},\mathbf{S}_{t-1},n_{t-1})$, ou seja:

$$(\boldsymbol{\theta}_{t-1}|\boldsymbol{\Sigma},D_{t-1}) \sim N(\mathbf{m}_{t-1},C_{t-1},\boldsymbol{\Sigma}), \quad (\boldsymbol{\Sigma}|D_{t-1}) \sim WI(\mathbf{S}_{t-1},n_{t-1}) \tag{a-8}$$

$$(\boldsymbol{\theta}_{t-1}|D_{t-1}) \sim T_{nt-1}(m_{t-1},C_{t-1},\mathbf{S}_{t-1}) \tag{a-9}$$

b) então a distribuição a priori em t é dada por:

$$(\boldsymbol{\theta}_t|\boldsymbol{\Sigma},D_{t-1}) \sim N(\mathbf{a}_{t-1},R_{t-1}), \quad (\boldsymbol{\Sigma}|D_{t-1}) \sim WI(\mathbf{S}_{t-1},n_{t-1}) \tag{a-10}$$

$$(\boldsymbol{\theta}_t|D_{t-1}) \sim T_{nt-1}\ (\mathbf{a}_{t-1},R_{t-1},\mathbf{S}_{t-1}) \tag{a-11}$$

onde: $\mathbf{a}_t=G_t\mathbf{m}_{t-1}$ e $R_t =G_tC_{t-1}G_t' + W_t$ (a-12)

c) a distribuição preditiva em t é dada por:

$$(\mathbf{Y}_t|\boldsymbol{\Sigma},D_{t-1}) \sim N(\mathbf{f}_t,Q_t\boldsymbol{\Sigma}) \tag{a-13}$$

$$(\mathbf{Y}_t|D_{t-1}) \sim T_{nt}\ (\mathbf{f}_t,Q_t\mathbf{S}_{t-1}) \tag{a-14}$$

onde: $\mathbf{f}_t = F_t'\mathbf{a}_t$ e $Q_t = F_t'R_tF_t + V_t$ (a-15)

d) e a distribuição a posteriori em t é dada pelas distribuições (a-16) e (a-17) que completam a caracterização do cálculo recursivo das distribuições do modelo:

$$(\boldsymbol{\theta}_t|\boldsymbol{\Sigma},D_t) \sim N(\mathbf{m}_t,C_t,\boldsymbol{\Sigma}) \quad (\boldsymbol{\Sigma}|D_t) \sim WI(\mathbf{S}_t,n_t) \tag{a-16}$$

$$(\boldsymbol{\theta}_t|D_t) \sim T_{nt}\ (\mathbf{m}_t,C_t,\mathbf{S}_t) \tag{a-17}$$

onde: $A_t = R_t\,F_t/\,Q_t$ $\mathbf{e_t = Y_t - f_t}$ (a-18)

$\mathbf{m}_t = \mathbf{a}_t + A_t\,\mathbf{e}_t'$ $C_t = R_t + A_t\,A_t'\,Q_t$ (a-19)

$\mathbf{S}_t = (n_{t-1}\mathbf{S}_{t-1} + \mathbf{e_t}\,\mathbf{e_t}'/\,Q_t)$, $n_t = n_{t-1} +1$ (a-20)

No caso de $\sum$ desconhecido,[36] a verossimilhança preditiva $p(\mathbf{Y_1,....,Y}_T|D_0)$, o teorema acima nos diz que:

$$p(\mathbf{Y}_t|D_{t-1}) \sim T_{nt}(\mathbf{f}_t, Q_t\mathbf{S}_{t-1}) = G(n_{t-1},q)\,|Q_t\mathbf{S}_{t-1}|^{-1/2}\{n_{t-1} + \mathbf{e}_t'(Q_t\mathbf{S}_{t-1})^{-1}\,\mathbf{e}_t\}^{-(nt-1+q)/2} \quad \text{(a-21)}$$

$$\text{onde:} \quad G(n_{t-1},q) = n_{t-1}^{n_{t-1}/2}\;\Gamma((n_{t-1}+q)/2) \,/\, \{\Gamma(n_{t-1}/2)\,\pi^q\} \quad \text{(a-22)}$$

portanto, o logaritmo da verossimilhança preditiva é dado por:

$$\log(p(\mathbf{Y_1,.Y}_T|D_0)) = \sum_{t=1}^{T} p(\mathbf{Y}_t|D_{t-1}) \quad \text{(a-23)}$$

$$\log(p(\mathbf{Y_1,.Y}_T|D_0)) = k\text{-}0.5^* \sum_{t=1}^{T}\{\log(|Q_tS_{t-1}|)+(n_{t-1}+q)\log(n_{t-1}+\mathbf{e}_t'(Q_tS_{t-1})^{-1}\mathbf{e}_t)\} \quad \text{(a-24)}$$

A matriz W_t de covariância de ω_t, que é diagonal por bloco de equações — o que garante os choques ω de blocos diferentes não são correlacionados —, pode ser definida como proporcional à covariância C_t entre as variáveis de estado. Então, $W_t = \beta C_t \beta'$ para os parâmetros de um certo bloco e 0 em caso contrário. O vetor de coeficientes β determina a volatilidade dos parâmetros do modelo e pode ser definida arbitrariamente ou estimado como um hiperparâmetro.

A.2 - Fator de Bayes no Modelo de Componentes Comuns

Um modelo M_0 e um modelo alternativo M_1 podem ser comparados através da verossimilhança preditiva, utilizando o Fator de Bayes definido como a razão entre as verossimilhanças dos dois modelos, ou seja:

$$H_t = \log(p(\mathbf{Y_1,.Y}_t|D_0,M_0)) - \log(p(\mathbf{Y_1,.Y}_t|D_0,M_1)) \quad \text{(a-25)}$$

A avaliação da estabilidade do processo gerador das séries pode ser realizada comparando o modelo $M_0 \equiv (\mathbf{Y}_t \mid D_{t-1}) \sim T_{nt}(\mathbf{f}_t,\ Q_t\mathbf{S}_{t-1})$ como um modelo alternativo, semelhante a M_0, exceto por admitir um maior grau de incerteza (k), ou seja:

$$M_1 \equiv (\mathbf{Y}_t \mid D_{t-1}) \sim T_{nt}(\mathbf{f}_t,\ kQ_t\mathbf{S}_{t-1}) \quad \text{(a-26)}$$

[36] No caso de Σ conhecido a densidade preditiva e o logaritmo da verossimilhaça preditiva são dados por:
$p(Y_t|D_{t-1},\Sigma) \approx N(f_t,\ Q_t\,\Sigma) = \{(2\pi)^q\,|Q_t\Sigma|^{-1/2}\}^{-1/2}\exp\{(-1/2)e_t'(Q_t\Sigma)^{-1}\}$ e
$\log(p(Y_1,...Y_T \mid \Sigma,D_0)) = cte\ -0.5^* \sum_{t=1}^{T}\ \{\log(|Q_t\Sigma|)+ (e_t'(Q_t\Sigma)^{-1}\,e_t)\}$

Utilizando a definição da verossimilhança preditiva e um critério de escolha do melhor modelo, pode-se derivar um procedimento para identificar o modelo mais provável.[37]

A.3 - Teste de Hipótese em Modelos de Componentes Comuns

Seja θ_k a matriz dos k primeiros parâmetros das q equações de um modelo de componentes comuns, e θ_r a matriz dos últimos r parâmetros das q equações do mesmo modelo, e n= k+r o número total de parâmetros em cada equação. Definindo:

$$\theta = (\theta_1,...., \theta_q) = \begin{pmatrix} \theta_{11},...\theta_{1q} \\ ..... \\ \theta_{n1},...\theta_{nq} \end{pmatrix} = \begin{pmatrix} \theta_k \\ \theta_r \end{pmatrix} \qquad \text{(a-27)}$$

Se estivermos interessados em verificar se as r últimas componentes do vetor F_t são significativas conjuntamente às q equações, teremos: H_0:θ_r=0 (ou H_0:θ_r'= 0), onde a segunda é mais conveniente. Fazendo a transposição e vetorização, temos:

$$\theta' = (\theta_k'\theta_r') \text{ e } vec(\theta') = \begin{pmatrix} vec(\theta_k') \\ vec(\theta_r') \end{pmatrix} \qquad \text{(a-28)}$$

adicionando o fato de que: $vec(\theta') \sim T_n (vec(m'), C \otimes S)$ (a-29)

temos que a distribuição de : $vec(\theta_r') \sim T_n (vec(m_r'), C_{22} \otimes S)$ (a-30)

onde: $vec(m') = \begin{pmatrix} vec(m_k') \\ vec(m_r') \end{pmatrix}$ e $C = \begin{pmatrix} C_{11}C_{12} \\ C_{21}C_{22} \end{pmatrix}$

Assim o nível descritivo do teste será $P(F_{rq,n} >= (1/rq)\, m_r (C_{22}^{-1} \otimes S_t^{-1})\, m_r')$.

[37]O Fator de Bayes também pode ser calculado dado o Σ conhecido — por exemplo, utilizando o seu estimador de máxima verossimilhança. Neste caso, temos:
$H_t = \log(p(\mathbf{Y_1,.Y}_t|\Sigma,D_0, M_0)) - \log(p(\mathbf{Y_1,.Y}_t|\Sigma,D_0,M_1))$
$M_0 \equiv (\mathbf{Y}_t \mid \Sigma, D_{t-1}) \approx N(\mathbf{f}_t, Q_t\Sigma)$ e $M_1 \equiv (\mathbf{Y}_t \mid \Sigma, D_{t-1}) \approx N(\mathbf{f}_t, k\, Q_t\Sigma)$

A.4 - Resultados Complementares das Subseções 3.1 e 3.2

Tabela 12
Estimativa de média dados σ^2, tamanho da amostra e distanciamento

μ	$P(\|\theta\|\leq 3)$	μ_1	M	Média Estimada	Variância Estimada	Teste de Kolmogorov	Valor Crítico (1%)
0	0.9773	-0.0283	100	-0.0207	0.0205	0.0936	0.196
			500	-0.0367	0.0204	0.0495	0.088
			1000	-0.0271	0.0205	0.0500	0.062
0.8	0.9192	-0.0126	100	-0.0405	0.0167	0.1806	0.196
			500	-0.0141	0.0162	0.0832	0.088
			1000	-0.0171	0.0189	0.0499	0.062
1.2	0.7257	-0.0048	100	-0.0134	0.0163	0.0928	0.196
			500	0.0003	0.0217	0.0758	0.088
			1000	0.0142	0.0175	0.0957+	0.062
1.5	0.5000	0.0011	100	0.1285	0.0482	0.2688+	0.196
			500	0.0227	0.0194	0.1159+	0.088
			1000	-0.0098	0.0212	0.0761+	0.062
1.8	0.2746	0.007	100	0.0094	0.0099	0.3153+	0.196
			500	0.0057	0.0248	0.1204+	0.088
			1000	0.0064	0.0160	0.1178+	0.062

Obs.: Nesse exercício, a variância teórica é τ_1^2=0.0196 ← + Rejeita-se H_0: distribuições iguais.

Tabela 13
Estimativa de σ^2 dados a média, tamanho da amostra e distanciamento

σ_0^{-2}	$p(\phi<1)$	Média Teórica	Variância Teórica	M	Média Estimada	Variância Estimada	Kolmogorov	Valor Crítico (1%)
1.2	0.54	1.1099	0.0241	100	1.1060	0.0259	0.200	0.196
				500	1.1119	0.0233	0.041	0.088
				1000	1.1148	0.0214	0.056	0.062
2.2	0.21	1.1315	0.0245	100	1.1149	0.0135	0.189	0.196
				500	1.1439	0.0233	0.062	0.088
				1000	1.1379	0.0235	0.051	0.062
2.8	0.10	1.1445	0.0248	100	1.1495	0.0234	0.119	0.196
				500	1.1348	0.0263	0.080	0.088
				1000	1.1362	0.0254	0.060	0.062
3.3	0.05	1.1554	0.0250	100	1.2039	0.0230	0.154	0.196
				500	1.1415	0.0228	0.085	0.088
				1000	1.1637	0.0238	0.045	0.062
5.0	0.003	1.1922	0.0258	100	1.2556	0.0090	0.344+	0.196
				500	1.1972	0.0227	0.119+	0.088
				1000	1.1799	0.0244	0.127+	0.062

BIBLIOGRAFIA

BARBOSA, E.P. **Dynamic bayesian models for vector time series analysis and forecasting**. University of Warwick, UK, 1989 (Ph. D. Thesis).

BERGER, J. **Statistical decision theory and bayesian analysis.** Springer-Verlag, 1985.

BERNARDO, J.M., SMITH, A.F.M. **Bayesian theory**. New York, John Wiley, 1994.

BRIAN, B. **Basic optimization methods.** University of Bradford/School of Mathematical Science, 1984.

EFRON, B. **The bootstrap, jacknife and other resampling plans**. Philadelfia: Society of Industrial and Appled Mathematics, 1982.

GAMERMAN, D., MIGON, H. S. **Inferência estatística: uma abordagem integrada**. Rio de Janeiro, IM-UFRJ, 1993 (Textos de Métodos Matemáticos, 27).

HARVEY, A.C. **Forecasting structural time series models and the Kalman Filter.** Cambridge University Press, UK, 1989.

KADIYALA, K.R., KARLSSON, S. Forecasting with generalized bayesian vector autoregressions. **Journal of Forecasting**, v.12, p.365-378, 1993.

KOOP, G. Aggregate shocks and macroeconomic fluctuations: a bayesian approach. **Journal of Applied Econometrics**, v.7, p.395-411, 1992.

LIMA, E.C.R., MIGON, H.S., LOPES, H. F. Efeitos dinâmicos dos choques de oferta e demanda agregadas sobre o nível de atividade econômica do Brasil. **Revista Brasileira de Economia**, v.47, n.2, p.177-204, 1993.

LITTERMAN, R. Forecasting with bayesian vector autoregressions - five years of experience. **Journal of Business and Economic Statistics**, v.4, p.1, p.25-38, 1986.

LOPES, H.F. **Aplicações de modelos autoregressivos vetoriais**. Rio de Janeiro, IM-UFRJ: Departamento de Métodos Estatísticos, 1994 (Dissertação de Mestrado).

LUTKEPOHL, H. **Introduction to multiple time series.** Berlin: Springer-Verlag, 1991.

MIGON, H.S., HARRISON, P.J. An aplication of non-linear bayesian forecasting to television advertising. In: BERNARDO, J.M., DeGROOT, M.H.,

LINDLEY, D.V. (eds.). **Bayesian Statistics**, n. 2, North-Holland: Amsterdan and Valencia University Press, 1985.

MOREIRA, A.R.B., FIORENCIO, A. **Boletim Conjuntural,** IPEA, abr. 1996.

RIPLEY, B.D. **Stochastic simulation.** New York: John Wiley & Sons, 1986.

QUINTANA, J.M. **A dynamic linear matrix-variable regression model**. Dept. of Statistics/University of Warwick, UK, 1985 (Research Report, 83).

RUBIN, D.B. Using the SIR algorithm to simulate posterior distributions. In: BERNARDO, J.M. **et alii** (eds.). **Bayesian Statistics**, n. 3. Oxford, U.K., Oxford University Press, p. 395-402, 1988.

SMITH, A.F.M., GELFAND, A.E. Bayesian statistic without tears, a sampling-resampling perspective. **The American Statistician**, v.46, n.2, 1992.

THISTED, R.A. **Elements of statistical computing.** Chapman and Hall, 1988.

VAN DIJK, H.K., KLOEK, T. Futher experience in bayesian analysis using Monte Carlo integration. **Journal of Econometrics**, v.14, p.307-328, 1980.

_____. Experiments with some alternatives for simple importance sampling in Monte Carlo integration. In: BERNARDO, J.M. **et alii** (eds.). **Bayesian Statistical**, n. 2, Elsevier Science Publishers B.V., p.511-530, 1986.

Livros Grátis

(http://www.livrosgratis.com.br)

Milhares de Livros para Download:

Baixar livros de Literatura
Baixar livros de Literatura de Cordel
Baixar livros de Literatura Infantil
Baixar livros de Matemática
Baixar livros de Medicina
Baixar livros de Medicina Veterinária
Baixar livros de Meio Ambiente
Baixar livros de Meteorologia
Baixar Monografias e TCC
Baixar livros Multidisciplinar
Baixar livros de Música
Baixar livros de Psicologia
Baixar livros de Química
Baixar livros de Saúde Coletiva
Baixar livros de Serviço Social
Baixar livros de Sociologia
Baixar livros de Teologia
Baixar livros de Trabalho
Baixar livros de Turismo