

Previsões Macroeconômicas Baseadas em Modelos TVP-VAR: Evidências Para o Brasil

João F. Caldeira,^{*} Guilherme V. Moura,[†] André A. P. Santos[‡]

Sumário: 1. Introdução; 2. TVP-VARs de grande dimensão; 3. Modelos competidores; 4. Dados e resultados empíricos; 5. Conclusão.

Palavras-chave: VAR Bayesiano, Parâmetros Variando no Tempo, Previsão, Modelo de Estado-Espaço.

Códigos JEL: C53, E43, G17.

Modelos baseados em vetores autoregressivos com parâmetros variantes no tempo e contendo efeitos heterocedásticos (TVP-VAR) propostos por Koop & Korobilis (2013) são utilizados na previsão da inflação (IPCA), da taxa de juros (SELIC) e do indicador mensal do PIB (IBC-Br) para diversos horizontes. Estratégias de previsão baseadas em seleção e combinação dinâmicas entre diferentes especificações também são utilizadas. As previsões são comparadas com as oriundas de modelos VAR bayesianos, modelos VAR aumentados com fatores e outros modelos competidores através da metodologia *model confidence set*. Os resultados indicam que a especificação TVP-VAR é a única que está sempre no conjunto de melhores modelos, independentemente da variável analisada ou do horizonte de previsão escolhido.

Vector autoregressive models with time-varying parameters and heteroskedastic effects, proposed in Koop & Korobilis (2013) and known as TVP-VAR models, are used to predict inflation, the interest rate, and a monthly GDP indicator for several forecast horizons. Dynamic model selection and averaging strategies for the different model specifications are also considered. Forecasts are compared to those from Bayesian VAR models, factor augmented VARs, and other competing models using the model confidence set methodology. The results indicate that TVP-VAR is the only specification always included in the set of best performing models, independently from the variable analyzed, and from the forecast horizon.

^{*}Departamento de Economia da Universidade Federal do Rio Grande do Sul.

[†]Departamento de Economia da Universidade Federal de Santa Catarina.

[‡]Universidade Federal de Santa Catarina, Departamento de Economia. Campus Universitário Trindade, Florianópolis, SC, Brasil. CEP 88049-970. E-mail: andre.portela@ufsc.br



1. INTRODUÇÃO

Koop & Korobilis (2013) desenvolveram métodos de estimação para modelos baseados em vetores autoregressivos com parâmetros variantes no tempo e contendo efeitos heterocedásticos (modelos TVP-VAR) que se aplicam mesmo quando o número de séries de tempo disponível é grande. Em aplicação para dados trimestrais da economia americana e usando até 25 variáveis endógenas, estes autores mostram que modelos TVP-VAR de grande dimensão são não só viáveis, como também úteis para prever a inflação, o produto real e as taxas de juros. O presente trabalho pretende aplicar a metodologia de Koop & Korobilis (2013) para prever dados mensais da economia brasileira. Adicionalmente, incorpora-se o modelo VAR aumentado com fatores (FAVAR) entre os modelos competidores, uma vez que Rossi & Carvalho (2009) já utilizaram esta abordagem para lidar com um grande número de variáveis macroeconômicas para a economia brasileira. Além disso, diferentemente de Koop & Korobilis (2013), a avaliação empírica das previsões dos diversos modelos alternativos é implementada através da abordagem *model confident set* (MCS) de Hansen, Lunde & Nason (2011), o que permite comparar simultaneamente vários modelos quando não se tem um *benchmark* claro.

Previsões acuradas de agregados macroeconômicos são fundamentais para a tomada de decisão de investimento, poupança, alocação de ativos e de política econômica. Recentemente, avanços em tecnologia da informação permitiram que um número muito maior de séries ficasse à disposição, enquanto desenvolvimentos computacionais permitiram a modelagem simultânea de diversas séries macroeconômicas para a previsão de variáveis importantes como inflação, produto agregado e a taxa de juros. As abordagens econométricas mais utilizadas para lidar com tais problemas são os modelos de vetores autorregressivos (VAR) e modelos de fatores dinâmicos (ver, por exemplo, Bańbura, Giannone & Reichlin, 2010; Litterman, 1986; Stock & Watson, 2002a, 2002b).

Modelos VAR têm sido usados para a análise estrutural e previsões em macroeconomia desde o trabalho seminal de Sims (1980). Porém, aplicações empíricas utilizando VARs sofrem da “maldição da dimensionalidade”, uma vez que o número de parâmetros destes modelos cresce exponencialmente a medida que novas variáveis são incluídas, afetando a precisão das estimativas. Dessa forma, análises empíricas são baseadas em um número reduzido de variáveis, geralmente entre três e oito (Bańbura et al., 2010; Giannone & Reichlin, 2006), porém, conforme mostrado em Bernanke, Boivin & Elias (2005), tal prática pode gerar estimativas enviesadas devido a variáveis omitidas. Para contornar as dificuldades relacionadas à estimação de modelos VARs com grande número de variáveis, a literatura recente tem recorrido aos modelos de fatores como forma de limitar o número de parâmetros a ser estimado ou à abordagens bayesianas que utilizam informações a priori para aumentar a precisão das estimativas. Essas abordagens podem ser aplicadas a problemas empíricos que envolvem um grande número de variáveis e vêm obtendo sucesso tanto no contexto de previsão quanto na análise estrutural (ver, por exemplo, Bańbura et al., 2010; Bernanke et al., 2005; Carriero, Kapetanios & Marcellino, 2009, 2012; Stock & Watson, 2005). Mais recentemente, evidências empíricas do bom desempenho preditivo de modelos VAR bayesianos de grande dimensão vêm se acumulando na literatura (ver Carriero et al., 2009; Giannone, Lenza, Momferatou & Onorante, 2014; Koop & Korobilis, 2013, entre outros).

Este artigo se baseia nesta literatura internacional para prever a inflação (IPCA), a taxa de juros (SELIC) e a atividade econômica (IBC-Br) brasileira. Para realizar as previsões conjuntas são usados modelos VAR heterocedásticos de grande dimensão e com parâmetros variando ao longo do tempo (TVP-VAR). Os resultados destes são comparados aos de diversos modelos competidores, incluindo VAR clássico, VAR bayesiano (BVAR), modelos VAR aumentados com fatores dinâmicos (FAVAR), passeio aleatório, entre outros. Adicionalmente, pretende-se analisar a relevância da utilização de um grande número de variáveis, de parâmetros variantes no tempo, bem como de efeitos heterocedásticos para a previsão da média incondicional das três séries macroeconômicas acima.

Para superar as dificuldades inerentes à estimação e à previsão recursivas de modelos TVP-VARs com elevada dimensão, a abordagem baseada em fatores de esquecimento proposta recentemente por Koop & Korobilis (2013) é utilizada, evitando então o uso de métodos de estimação baseados em simulação. Vale

destacar que a abordagem de Koop & Korobilis (2013) necessita da estimação de apenas três parâmetros, independentemente do número de variáveis incluídas no TVP-VAR, o que elimina por completo a “maldição da dimensionalidade”. As estratégias de combinação dinâmica de modelos (DMA) e seleção dinâmica de modelos (DMS), desenvolvidas por Raftery, Kárný, & Ettlér (2010) e aprimoradas por Koop & Korobilis (2013) também são utilizadas nos exercícios de previsão. Mais especificamente, a abordagem DMS é utilizada para seleção dinâmica de um único modelo de previsão em cada ponto do tempo, enquanto a abordagem DMA combina as previsões dos vários modelos com pesos que variam de período a período. O universo de modelos adotados neste artigo contém TVP-VARs com 3, 7 e 20 variáveis, possibilitando selecionar modelos de diferentes dimensões em cada ponto do tempo.

A literatura brasileira sobre previsões macroeconômicas é vasta, o que demonstra o interesse e a importância deste tema. Para previsão de inflação, modelos baseados em curva de Phillips como proposto por Arruda, Ferreira & Castelar (2011) ou modelos puramente estatísticos como o de Carlos & Marçal (2013) e o de Chauvet (2001) têm sido usados com frequência. Para a previsão do PIB, Céspedes, Chauvet & Lima (2006) destaca a importância de parâmetros variantes no tempo para levar em consideração as quebras estruturais e choques sofridos pela economia brasileira, o que motiva o uso de TVP-VARs. R. T. Ferreira, Bierens & Castelar (2005) corrobora a evidência em favor de parâmetros variando no tempo ao usar modelos não lineares de índice de difusão, modelos com efeito *threshold* e com mudança markoviana de regime para prever o PIB. Ademais, estes autores consideram combinações simples de previsões oriundas de diferentes modelos e mostram que tais combinações de previsões geram resultados mais acurados do que modelos individuais. D. Ferreira & Palma (2014) também mostram que a combinação dinâmica de modelos (DMA), bem como a seleção dinâmica de modelos (DMS) melhoram significativamente as previsões de modelos individuais para a inflação brasileira.

Para avaliar a abordagem TVP-VAR empiricamente considerou-se um conjunto de 20 variáveis macroeconômicas brasileiras amostradas mensalmente durante o período de 2003:M1 a 2013:M12 e diversos modelos competidores. As previsões foram avaliadas com base na média dos erro de previsão ao quadrado (MSFE), na soma da log-verossimilhança preditiva, no erro de previsão ao quadrado acumulado e com base na metodologia *model confidence set* (MCS) de Hansen et al. (2011), a qual permite selecionar, dentre um conjunto de modelos, aqueles que podem ser considerados estatisticamente superiores aos demais.

Métodos capazes de promover a avaliação entre modelos incluem, por exemplo, o teste *equal predictive ability* (EPA) de Diebold & Mariano (1995) e o *superior predictive ability* (SPA) de Hansen (2005). Porém, tais métodos possuem algumas limitações, como comparação restrita a pares de modelos e avaliação relativamente a um único modelo *benchmark*, respectivamente. Com o intuito de superar as limitações do EPA e do SPA, Hansen et al. (2011) desenvolveram o *model confidence set* (MCS). Uma vantagem de destaque no MCS é a possibilidade de efetuar comparações múltiplas, isto é, para um conjunto grande de modelos, sem a necessidade da definição de um modelo *benchmark*, como ocorre no SPA. No MCS é possível reduzir um conjunto de modelos para um conjunto menor que contém o(s) melhor(es) modelo(s) dado um nível de significância (Hansen et al., 2011). Essas características do MCS são fundamentais para o exercício empírico implementado neste trabalho, uma vez que são consideradas previsões realizadas por um conjunto grande de modelos e não há consenso sobre qual é o modelo a ser batido, e, portanto, qual deve ser o *benchmark*.

Os resultados encontrados indicam ganhos em termos de acurácia preditiva das variáveis de interesse com uso de TVP-VARs. Apesar dos modelos TVP-VAR com DMA não superarem os demais competidores em todos os casos, estes estão sempre entre os modelos com melhor desempenho, como indicado pela análise *model confidence set*. Em outras palavras, apesar de praticamente todos os modelos alternativos apresentarem bom desempenho na previsão para alguma das variáveis ou para algum horizonte de tempo, somente a abordagem TVP-VAR com DMA está entre os melhores para todas as variáveis e todos os horizontes de previsão analisados.

Os resultados encontrados mostram que o conjunto de modelos que fazem parte do MCS muda consideravelmente entre as variáveis e entre os horizontes de previsão considerados. Especificamente, quando



são analisadas as previsões 1-passo à frente para o IBC-Br, nota-se que o MCS é composto basicamente pelos modelos TVP-VAR com estratégias DMA e DMS, e TVP-VAR de grande dimensão. Para os horizontes de previsão mais longos o MCS é formado por um número maior de modelos. Ao analisar as previsões para o IPCA percebe-se que os modelos VAR, BVAR e FAVAR fazem um bom trabalho em todos os horizontes de previsão e estão presentes no MCS na maioria dos casos, exceto nas previsões mais longas, em que o destaque vai para os modelos TVP-VAR de pequena dimensão. No caso das previsões para a SELIC, há uma preferência pelos modelos TVP-VAR de pequena dimensão e TVP-VAR (DMA e DMS), principalmente para os horizontes de até 6-meses à frente. Cabe destacar, porém, que para todas as variáveis consideradas, os modelos TVP-VAR com estratégias DMA e DMS são os únicos que estão sempre presentes no MCS em todos os horizontes de previsão analisados.

A abordagem TVP-VAR com estratégias DMA e/ou DMS para fazer previsão da atividade econômica (IBC-Br) gera resultados em média 20% melhores que os modelos competidores quando se considera o horizonte de previsão de 1 mês à frente. Quando consideramos horizontes de previsão mais longos, modelos FAVAR apresentam bom desempenho, entretanto, ainda perdem para os TVP-VARS.

No caso das previsões para a inflação, a maioria dos modelos competidores apresenta desempenho levemente superior ao TVP-VAR para previsões um passo à frente. Especificamente, os modelos BVAR alcançam os melhores resultados nas previsões um passo à frente para o IPCA, porém, perdem para diferentes especificações TVP-VAR para previsões três, seis e doze meses à frente.

Analisando as previsões para a taxa de juros (SELIC) percebe-se que o modelo TVP-VAR com apenas três variáveis supera quase todos os demais modelos, principalmente para os horizontes de 1 e 3 meses à frente. Os erros de previsão desta especificação são em média 5% a 14% menores do que os dos competidores. Já para o horizonte mais longo, os modelos BVAR e FAVAR conseguem os melhores resultados. Entretanto, novamente o TVP-VAR com abordagem DMA e principalmente com a abordagem DMS apresentam desempenho próximo aos obtidos pelos melhores modelos em todos os casos, se confirmando como a metodologia mais estável e confiável para a previsão de variáveis macroeconômicas.

Em termos gerais, percebeu-se que, para todas as variáveis e horizontes de previsão considerados, a inclusão de efeitos heterocedásticos são importantes ainda que o foco seja apenas a previsão da média condicional. Já a inclusão de parâmetros variantes no tempo e de uma grande base de dados contendo um grande número de variáveis não parece ter um efeito tão consistente em todos os casos. TVP-VARS, BVARs e modelos FAVAR que utilizam todas as 20 variáveis disponíveis apresentam resultados muito bons para a previsão de inflação. Porém, este padrão não se repete para a previsão do IBC-Br, onde modelos de tamanho médio apresentam os melhores resultados, nem para a previsão da taxa SELIC, situação na qual os modelos com apenas três variáveis são os mais satisfatórios. Modelos que consideram parâmetros variantes no tempo têm desempenho levemente superior para previsões da inflação, mas não parecem oferecer uma vantagem consistente quando se analisa os resultados para todas as três variáveis de interesse e para todos os horizontes de previsão.

Este artigo está estruturado em cinco seções, incluindo esta introdução. Na seção 2 o modelo TVP-VAR, bem como sua estimação e as estratégias de combinação e seleção dinâmicas de modelos são apresentadas. A seção 3 apresenta os modelos competidores, bem como as métricas de avaliação do desempenho utilizadas. A seção 4 apresenta e discute os resultados das previsões obtidas e a seção 5 ressalta as principais conclusões do artigo.

2. TVP-VARS DE GRANDE DIMENSÃO

A estimação dos modelos TVP-VAR segue a abordagem de Koop & Korobilis (2013) e utiliza o filtro de Kalman e fatores de esquecimento. É possível escrever o modelo TVP-VAR no formato de estado espaço como

$$y_t = Z_t \beta_t + \epsilon_t, \quad (1)$$

$$\beta_{t+1} = \beta_t + u_t, \quad (2)$$

onde ϵ_t é i.i.d. $\mathcal{N}(0, \Sigma_t)$ e u_t é i.i.d. $\mathcal{N}(0, Q_t)$; ϵ_t e u_s são independentes um do outro para quaisquer s e t ; y_t é um vetor $M \times 1$ contendo observações do período $t = 1, \dots, T$; Z_t é uma matriz $M \times k$ que contém as p defasagens das variáveis observáveis y e as constantes (logo, $k = (1 + pM) \cdot M$); β_t é um vetor $k \times 1$ de estados que contém os parâmetros variantes no tempo.

Após a seleção da especificação de Σ_t e Q_t , da distribuição a priori das condições iniciais β_0 , Σ_0 e Q_0 e dos outros parâmetros restantes, a inferência bayesiana a respeito dos parâmetros é baseada no filtro de Kalman e em métodos de simulação do tipo MCMC (ver, por exemplo, Cogley & Sargent, 2005). Porém, o custo computacional desta estimação é enorme, principalmente para modelos VAR de grande dimensão. Portanto, Koop & Korobilis (2013) propõem utilizar aproximações simples para Q_t e Σ_t com o intuito de evitar o uso de métodos MCMC.

A aproximação de Q_t é feita através de um fator de esquecimento no algoritmo do filtro de Kalman, como sugerido por Raftery et al. (2010). Mais especificamente, a distribuição dos parâmetros variantes no tempo para o período $t - 1$ baseada na informação disponível até o período $t - 1$ é dada por

$$\beta_{t-1} | y_{t-1} \sim \mathcal{N}(\beta_{t-1|t-1}, V_{t-1|t-1}), \quad (3)$$

onde $y_{t-1} = (y_1, \dots, y_{t-1})'$; $\beta_{t-1|t-1}$ são os estados filtrados do período $t - 1$ e $V_{t-1|t-1}$ sua variância. O filtro de Kalman atualiza a matriz $V_{t-1|t-1}$ usando a expressão $V_{t|t-1} = V_{t-1|t-1} + Q_t$, e este é o único passo onde o conhecimento da matriz Q_t se faz necessário. Portanto, substituindo a expressão anterior por

$$V_{t|t-1} = \frac{1}{\lambda} V_{t-1|t-1}, \quad (4)$$

não há mais necessidade de se estimar ou simular Q_t e o fator de esquecimento $\lambda \in (0, 1]$ irá governar a variabilidade de $V_{t|t-1}$. Note que se $\lambda = 1$, $V_{t|t-1} = V_{t-1|t-1}$ e $Q_t = 0$, o que faz com que o modelo TVP-VAR passe a ser um modelo VAR com parâmetros constantes. O uso dessa restrição será explorada na seção 4.

A abordagem de fatores de esquecimento não é nova e foi desenvolvida justamente para simplificar a estimação de modelos de estado espaço quando as restrições computacionais eram muito maiores do que as atuais (ver, por exemplo, Jazwinski, 1970, para uma discussão mais detalhada sobre fatores de esquecimento e sua motivação). A nomenclatura fator de esquecimento vem da constatação de que nessa abordagem as j observações mais recentes tem peso λ^j e apenas as $\frac{1}{1-\lambda}$ últimas observações terão peso na estimação. Com isso, Jazwinski (1970) argumenta que esquecimento é comparável a uma janela móvel de estimação.

A aproximação para a matriz de covariância na equação de medida (1), Σ_t , é feita através de uma média móvel exponencial (EWMA) (ver, por exemplo, Brockwell & Davis, 2009, seção 1.4):

$$\hat{\Sigma}_t = \kappa \hat{\Sigma}_{t-1} + (1 - \kappa) \hat{\epsilon}_t \hat{\epsilon}_t', \quad (5)$$

onde $\hat{\epsilon}_t = y_t - \beta_t | Z_t$ é o erro de previsão já calculado através das próprias recursões do filtro de Kalman. Seguindo Koop & Korobilis (2013), os parâmetros λ e κ serão escolhidos usando a metodologia de seleção de modelos DMS, descrita na próxima subseção, tendo como base em um conjunto finito de valores. Raftery et al. (2010) simplesmente fixam $\lambda = 0,99$ afirmando que os resultados não são sensíveis a variações entre 0,97 e 0,995, com isso, serão considerados neste artigo valores tal que $\lambda \in \{0,97; 0,98; 0,99; 1\}$, onde $\lambda = 1$ caracteriza o VAR com parâmetros fixos. Para o parâmetro da EWMA, valores tradicionais na literatura são considerados e $\kappa \in \{0,94; 0,96; 0,98\}$.

2.1. Combinação dinâmica de modelos (DMA) e seleção dinâmica de modelos (DMS)

Um dos problemas da previsão em tempo real é a incerteza a respeito de qual modelo fornece a melhor aproximação ao processo gerador dos dados. Muitas vezes, vários modelos competidores estão disponíveis e nem sempre é claro qual a melhor especificação a ser utilizada. Além disso, podem surgir



algumas questões como, por exemplo: i) os coeficientes associados às variáveis previsoras podem mudar ao longo do tempo; ii) o número de potenciais previsores pode ser grande; iii) o melhor modelo para previsão pode mudar ao longo do tempo. Para superar tais questões, este artigo utiliza a estratégia de combinação dinâmica de modelos (DMA) desenvolvida por Raftery et al. (2010).

Koop & Korobilis (2013) propõem o uso dessa mesma abordagem para a seleção dinâmica de modelos (DMS), onde um único (potencialmente diferente) modelo é usado para fazer previsão em cada ponto do tempo. As abordagens DMA ou DMS são interessantes para previsão de variáveis macroeconômicas, uma vez que permitem que o modelo de previsão mude ao longo do tempo e, ao mesmo tempo, permitem que os parâmetros de cada modelo variem ao longo do tempo. Com efeito, D. Ferreira & Palma (2014) mostram que esta estratégia supera diversos modelos concorrentes na previsão da taxa de inflação brasileira.

Neste artigo, o conjunto de modelos considerados para a seleção e combinação dinâmica contém diferentes especificações TVP-VAR com diferentes conjuntos de variáveis. Suponha um conjunto de K modelos e que $L_t \in \{1, 2, \dots, K\}$ determine qual modelo se aplica em cada período do tempo. A partir de previsões oriundas de todos os modelos, a abordagem DMA calcula a probabilidade de que o modelo k é o melhor modelo para prever observações do período t , dadas as informações disponíveis em $t - 1$, ou seja, $\text{Prob}(L_t = k | y_{t-1}) = \pi_{t|t-1,k}$. Uma vez calculado $\pi_{t|t-1,k}$, essas probabilidades podem ser usadas para fazer combinação das previsões dos modelos. A abordagem DMS envolve selecionar o modelo com maior $\pi_{t|t-1,k}$ em cada ponto do tempo e usá-lo para fazer as previsões.

O algoritmo recursivo desenvolvido por Raftery et al. (2010) para o cálculo das probabilidades de cada modelo possui recursões semelhantes às do filtro de Kalman e também conta com um fator de esquecimento, aqui denotado por α , para reduzir o custo computacional. Dado uma probabilidade inicial a priori, $\pi_{0|0,k}$, a equação de transição das probabilidades é dada por

$$\pi_{t|t-1,k} = \frac{\pi_{t-1|t-1,k}^\alpha}{\sum_{l=1}^K \pi_{t-1|t-1,l}^\alpha}, \tag{6}$$

onde $0 < \alpha \leq 1$ é tratado como fixo e uma análise de sensibilidade será apresentada na avaliação empírica. A equação de atualização das probabilidades de cada modelo é dada por

$$\pi_{t|t,k} = \frac{\pi_{t|t-1,k} \cdot p_k(y_t | \underline{y}_{t-1})}{\sum_{l=1}^K \left[\pi_{t|t-1,l} \cdot p_l(y_t | \underline{y}_{t-1}) \right]}, \tag{7}$$

onde $p_k(y_t | \underline{y}_{t-1})$ é a densidade preditiva (verossimilhança preditiva) do modelo k avaliada em y_t , que é uma medida de desempenho de previsão. Esta densidade preditiva é obtida através do filtro de Kalman e possui fórmula fechada (ver, por exemplo, Frühwirth-Schnatter, 2006, p405–410). A abordagem DMA consiste em utilizar uma média ponderada das previsões individuais usando $\pi_{t|t-1,k}$ para o cálculo dos pesos de cada modelo. Já a abordagem DMS envolve selecionar apenas o modelo com a maior probabilidade preditiva, $\pi_{t|t-1,k}$, em cada ponto do tempo, e usá-lo no exercício de previsão.

Dessa forma, a probabilidade preditiva de cada modelo k , no período de tempo t , $\pi_{t|t-1,k}$, pode ser escrita como

$$\pi_{t|t-1,k} \propto \prod_{i=1}^{t-1} \left[p_k(y_{t-i} | y_{t-i-1}) \right]^{\alpha^i}.$$

Logo, o modelo k irá receber mais peso caso tenha apresentado bom desempenho preditivo no passado recente, enquanto o controle do que deve ser considerado passado recente é feito pelo fator de esquecimento, α . Por exemplo, se $\alpha = 0,99$, o desempenho preditivo de dois anos atrás receberá menos do que 80% do peso do desempenho do último período, enquanto $\alpha = 0,95$ determina que as previsões de cinco anos atrás tem apenas 30% do peso das últimas previsões. Já $\alpha = 0,001$ dará pouquíssimo peso

para qualquer previsão, se aproximando do caso tradicional de combinação de previsões usando pesos iguais.

Portanto, essa abordagem requer a seleção dos fatores de esquecimento α e λ , bem como do fator de decaimento κ . Seguindo Koop & Korobilis (2013), $\lambda \in \{0,97; 0,98; 0,99; 1\}$ e $\kappa \in \{0,94; 0,96; 0,98\}$ serão selecionados usando a metodologia DMS. Neste caso, diferentes valores para estes parâmetros definem diferentes modelos e usa-se a seleção dinâmica de modelos, DMS, para escolher entre as diferentes combinações de parâmetros. Com isso, é possível utilizar $\lambda = 1$, o que determina um VAR com parâmetros fixos, em determinado período de tempo, mas também um VAR com parâmetros variando rapidamente ($\lambda = 0,97$) em outro momento. A especificação padrão utiliza a sugestão de Raftery et al. (2010) e fixa $\alpha = 0,99$. Entretanto, serão apresentados resultados para $\alpha \in \{0,001; 0,95; 0,99; 1\}$, permitindo uma análise dos efeitos de diferentes graus de mudanças entre os modelos.

Ademais, é necessário definir uma distribuição a priori para os parâmetros iniciais do VAR, β_0 . Seguindo Koop & Korobilis (2013), adota-se uma distribuição a priori Normal para β_0 muito similar à priori de Minnesota desenvolvida por Doan, Litterman & Sims (1984). Como os dados utilizados são estacionários, a média desta a priori é definida como $\mathbb{E}[\beta_0] = 0$. A matriz de covariância da distribuição a priori de Minnesota é usualmente diagonal, dessa forma, denotando $\text{Var}(\beta_0) = \underline{V}$, com V_i designando os elementos da diagonal, a matriz de covariância da distribuição a priori para β_0 é dada por

$$V_i = \begin{cases} \frac{\gamma}{r^2}, & \text{para os coeficientes da defasagem } r, \text{ onde } r = 1, \dots, p \\ a = 10^2, & \text{para os interceptos} \end{cases}, \quad (8)$$

onde γ é o hiperparâmetro que controla o grau de encolhimento nos parâmetros do VAR e $a = 10^2$ caracteriza uma a priori não-informativa a respeito dos interceptos. Exceto onde estiver explícito, os modelos VAR estimados possuem duas defasagens ($p = 2$) e o denominador r^2 em (8) tem o intuito de aumentar o grau de encolhimento nas defasagens mais longas justamente para minimizar os efeitos da sobreparametrização. Para selecionar o valor do hiperparâmetro γ , a estratégia DMS descrita acima também será utilizada com base no seguinte conjunto de valores: $\gamma \in \{10^{-5}; 0,001; 0,005; 0,01; 0,05; 0,1\}$. Diferentes valores de γ definem diferentes distribuições a priori e, consequentemente, diferentes modelos permitindo o uso da abordagem DMS descrita acima. É importante perceber que, independentemente do número de variáveis utilizadas, o modelo TVP-VAR apresentado aqui contém apenas três parâmetros a serem estimados: o fator de esquecimento λ , o parâmetro de suavização κ e o grau de encolhimento da a priori para β_0 .

Diferentemente de estudos anteriores, onde as abordagens DMA e DMS foram usadas para selecionar entre modelos com diferentes variáveis explicativas, estas estratégias serão usadas aqui para selecionar entre modelos de diferentes dimensões: TVP-VAR pequeno, médio e grande. Em resumo, cada modelo é definido pela dimensão do TVP-VAR e um valor para λ , κ e γ . Com três dimensões de TVP-VAR, seis valores para γ e 12 combinações de λ e κ , tem-se um conjunto de 216 modelos diferentes.

3. MODELOS COMPETIDORES

Os principais resultados apresentados neste artigo são baseados em modelos TVP-VAR que utilizam a abordagem DMS para selecionar os únicos três parâmetros do modelo (λ , κ e γ). Têm-se ainda três diferentes conjuntos de variáveis com 3, 7 ou 20 variáveis e estes serão denominados conjuntos pequeno, médio e grande, respectivamente. Resultados individuais para os modelos TVP-VAR pequeno, médio e grande são apresentados separadamente, mas a estratégia de seleção dinâmica de modelos DMS também será usada para selecionar o conjunto de variáveis que mais se adequa à previsão de determinada variável em um dado período de tempo. O principal *benchmark* usado para avaliar todas as diversas especificações e modelos competidores será a estratégia TVP-VAR-DMA, onde previsões usando os três conjuntos de séries serão combinadas dinamicamente usando a metodologia DMA. O fator de esquecimento para as estratégias DMA e DMS é fixado em $\alpha = 0,99$, como sugerido por Raftery et al. (2010),



e casos diferentes destes são apresentados como teste de robustez e são indicados explicitamente nas tabelas 1 a 3.

Além disso, resultados para o caso especial do modelo VAR heteroscedástico com parâmetros fixos são apresentados com base na parametrização $\lambda = 1$ e $\kappa = 0,96$ para cada um dos três conjuntos de variáveis e permitem analisar a importância de parâmetros variantes no tempo. Resultados para modelos VAR homocedástico obtidos fixando-se $\lambda = 1$ e estimando-se Σ através de $\hat{\Sigma} = \frac{1}{t-1} \sum_{t=1}^{t-1} \hat{\epsilon}_t \hat{\epsilon}_t'$ também são apresentados para cada um dos três conjuntos de dados, permitindo uma análise da importância de considerar-se efeitos heterocedásticos.

Como sugerido por Stock & Watson (2002b) e Bernanke & Boivin (2003), uma poderosa alternativa para a previsão macroeconômica quando diversas variáveis observáveis estão à disposição é o uso de fatores comuns. Baseado nesta ideia, Bernanke et al. (2005) desenvolvem um modelo VAR aumentado com fatores comuns (FAVAR), possibilitando incorporar a informação de diversas séries econômicas de forma parcimoniosa em um modelo VAR. Esta metodologia foi utilizada por Rossi & Carvalho (2009) para decompor os choques monetários da economia brasileira e se coloca como competidora à metodologia TVP-VAR. Mais especificamente, o modelo FAVAR estimado pode ser escrito como

$$\begin{bmatrix} F_t \\ Y_t \end{bmatrix} = \Phi(L) \begin{bmatrix} F_{t-1} \\ Y_{t-1} \end{bmatrix} + v_t, \quad (9)$$

onde $\Phi(L)$ é um polinômio de defasagem de ordem p , $v_t \sim \text{i.i.d.}(0, Q_t)$; F_t é um vetor $K \times 1$ de fatores comuns; e Y_t contém o IPCA, o IBC-Br e a taxa SELIC. Os fatores comuns são estimados em uma etapa anterior utilizando o conjunto de 17 variáveis formado por todas as 20 variáveis disponíveis, menos as séries do IBC-Br, IPCA e SELIC contidas em Y_t (ver Rossi & Carvalho, 2009, para mais informações a respeito da estimação de FAVAR). Resultados para $K = 1, 3$ e 5 e para $p = 2$ são apresentados na seção 4.2.

Mais recentemente, Bańbura et al. (2010) mostrou que modelos VAR bayesianos (BVAR) com um grande número de variáveis observáveis podem superar o poder preditivo de modelos FAVAR. A introdução de informação a priori na estimação dos inúmeros parâmetros de um modelo VAR permite aumentar a precisão das estimativas, o que influencia positivamente as previsões baseadas nestes modelos. O desempenho empírico dos BVARs em situações onde várias séries estão disponíveis os coloca como competidores para os TVP-VARs. Com isso, modelos com três, sete e vinte séries são estimados utilizando uma distribuição a priori de Minnesota adaptada para as características dos dados. Uma vez que todas as variáveis são transformadas em séries estacionárias (ver o autorefapp para maiores detalhes a respeito das transformações realizadas), a distribuição a priori utilizada é dada por

$$\beta_{i,j,r} \sim \mathcal{N}(0, \sigma_{i,j,r}^2),$$

onde $\beta_{i,j,r}$ é o coeficiente do VAR referente à r -ésima defasagem da variável j na equação definindo a média condicional da variável i . Seguindo Doan et al. (1984),

$$\sigma_{i,j,r} = \frac{\gamma}{r^\phi} w_{i,j} \begin{pmatrix} \hat{\sigma}_{u,j} \\ \hat{\sigma}_{u,i} \end{pmatrix}, \quad (10)$$

onde $\hat{\sigma}_{u,i}$ é uma estimativa do desvio padrão do coeficiente autoregressivo da série i calculada com base em uma autoregressão univariada de primeira ordem. O hiperparâmetro ϕ estabelece como o desvio padrão da distribuição a priori relativa aos parâmetros das defasagens do VAR varia à medida que o número de defasagens aumenta. Logo, de acordo com (10), a confiança em uma estimativa nula (uma vez que a média da distribuição a priori é dada por $\mathbb{E}[\beta_{i,j,k}] = 0$) aumenta para coeficientes de defasagens muito longas, fazendo com que haja uma preferência por modelos mais parcimoniosos. Para reforçar a idéia de um modelo parcimonioso, faz-se $\phi = 2$, encolhendo fortemente os parâmetros das defasagens maiores para zero. Os pesos $w_{i,j}$ definem a precisão relativa da distribuição a priori para a defasagem da variável j em comparação com a precisão do coeficiente para a própria defasagem e, portanto, são normalizados

de forma que $w_{i,i} = 1, \forall i$. De forma a caracterizar o entendimento a priori de que $\beta_{i,j} = 0$ é mais provável do que $\beta_{i,i} = 0$, seleciona-se $w_{i,j} = 0,4, \forall i \neq j$, o que também privilegia um modelo parcimonioso. Já a precisão da distribuição a priori para a própria defasagem de primeira ordem é definida por $\frac{1}{\gamma^2}$, pois, como fica claro em (10), $\sigma_{i,j,k} = \gamma$ quando $i = j$ e $r = 1$. Como argumentado em Koop & Korobilis (2013), o desempenho de BVARs depende fortemente do grau de encolhimento dos parâmetros. Com o objetivo de encolher os parâmetros para próximo da média da distribuição a priori, selecionou-se $\gamma = 0,2$, como proposto em Litterman (1986).

Por fim, modelos VAR com abordagem clássica são estimados para as três variáveis alvo e resultados são apresentados para as especificações VAR(1), VAR(2) e para o caso especial do passeio aleatório.

3.1. Medidas de desempenho e avaliação das previsões

A estatística básica para comparação de previsões é a média dos erros de previsão ao quadrado (MSFE). Dado um conjunto de T previsões fora da amostra para o horizonte de h -passos à frente, pode-se calcular a média dos erros de previsão ao quadrado (MSFE) para o IBC-Br, IPCA e taxa SELIC, do modelo m como

$$\text{MSFE}_m^i = \frac{1}{T} \sum_{t=1}^T \left(\hat{y}_{t+h|t,m}^i - y_{t+h}^i \right)^2, \quad (11)$$

onde y_{t+h}^i é a i -ésima variável observada em $t+h$, e $\hat{y}_{t+h|t,m}^i$ é a previsão para esta variável feita pelo modelo m para o período de tempo t .

Embora seja frequentemente usado na avaliação de previsões, o MSFE tem a desvantagem de ser uma média para toda a amostra, não fornecendo nenhuma informação sobre em que parte da amostra determinado modelo apresenta seu melhor ou pior desempenho. Para superar essa deficiência, analisa-se graficamente o erro quadrado de previsão acumulado (CSFE), proposto por Welch & Goyal (2008). A série de erro de previsão acumulado mostra claramente quando um determinado modelo supera ou perde para um dado benchmark. O CSFE é definido como

$$\text{CSFE}_{m,T}^i = \sum_{t=1}^T \left[\left(\hat{y}_{t+h|t,\text{bench}}^i - y_{t+h}^i \right)^2 - \left(\hat{y}_{t+h|t,m}^i - y_{t+h}^i \right)^2 \right]. \quad (12)$$

Quando o modelo m supera o *benchmark*, o $\text{CSFE}_{m,T}$ será uma série crescente. Se o *benchmark* produz previsões mais acuradas, então $\text{CSFE}_{m,T}$ será decrescente.

Além de medidas estatísticas tradicionais, testes desenvolvidos especificamente para avaliação de previsão foram utilizados. Ademais, em situações onde um grande número de modelos é avaliado, existe a possibilidade de ao menos um modelo ser melhor do que o *benchmark* e, mesmo assim, perder o teste de previsão (fenômeno também conhecido como “*data snooping*”), uma vez que testes estatísticos estão sujeitos a erros do tipo I ou II. Para contornar o problema de *data snooping* e realizar uma comparação mais adequada entre todos os modelos é aplicada a abordagem proposta por Hansen et al. (2011), conhecida como *model confidence set* (MCS).

Especialmente quando o conjunto de modelos alternativos é grande, muitas aplicações baseadas em MSFE e/ou CSFE não irão detectar apenas um modelo dominante que supera significativamente todos os competidores, pois raramente os dados serão suficientemente informativos para dar uma resposta tão clara, o que dificulta a determinação do melhor modelo. Entretanto, a metodologia *model confidence set* proposta por Hansen et al. (2011) permite reduzir o número de modelos potenciais para um conjunto mais restrito — o conjunto de confiança de modelos — que contém o melhor modelo com determinado grau de confiança. Como são calculadas previsões para três variáveis macroeconômicas e para quatro horizontes de previsão distintos, modelos incluídos no MCS de todos estes casos têm maior chance de ser o melhor modelo de previsão. O MCS possui diversas vantagens em relação a testes de habilidade preditiva superior. Dentre elas está o fato do MCS não necessitar de um modelo *benchmark* específico para



a comparação, o que é útil em uma aplicação com diversos modelos competidores sem um *benchmark* óbvio como a apresentada neste artigo (ver Hansen et al., 2011, seção 4.1 para outras vantagens do MCS em relação a testes de habilidade preditiva superior).

A abordagem MCS começa com um conjunto de modelos candidatos $\mathcal{M}_0 = (1, \dots, m_0)$. O desempenho do modelo i em relação ao modelo j , no tempo t , é dado por

$$d_{ij,t} = L(\xi_{j,t}) - L(\xi_{i,t}), \quad i, j = 1, \dots, m_0 \quad \text{e} \quad t = 1, \dots, T. \quad (13)$$

onde $\xi_{i,t}$ é o erro de previsão do modelo i no tempo t e $L(\cdot)$ pode ser uma função de perda qualquer. Neste artigo a raiz do erro de previsão ao quadrado será utilizada. As previsões são avaliadas em termos da perda esperada:

$$\mathbb{E}[d_{ij}] = \mathbb{E}[L(\xi_{j,t}) - L(\xi_{i,t})].$$

Começando com o conjunto de todos os modelos, $\mathcal{M} = \mathcal{M}_0$, testa-se repetidamente a hipótese nula de igual acurácia preditiva ao nível de significância α :

$$H_0 : \mu_{i,j} = \mathbb{E}[d_{ij,t}] = 0 \quad \text{para todo } i, j \in \mathcal{M}, \quad (14)$$

onde μ_{ij} é a esperança do diferencial de perda entre os modelos i e j . Contanto que a hipótese nula seja rejeitada para algum modelo j , o procedimento MCS elimina de forma sequencial o modelo com pior performance pertencente a \mathcal{M} . Esta eliminação de modelos é repetida até que a hipótese nula (14) não seja mais rejeitada ao nível de confiança $1 - \alpha$. Os modelos remanescentes compõem então o conjunto de confiança $\widehat{\mathcal{M}}_{1-\alpha}^*$.

A hipótese nula na equação (14) é testada por meio da estatística de alcance, T_R , que é baseada na seguinte estatística t :

$$t_{ij} = \frac{\bar{d}_{ij}}{\sqrt{\widehat{\text{Var}}(\bar{d}_{ij})}} \quad \text{para } i, j \in \mathcal{M},$$

onde $\widehat{\text{Var}}(\bar{d}_{ij})$ é uma estimativa de $\text{Var}(\bar{d}_{ij})$ e é obtida através de um procedimento de *bootstrap*. A estatísticas de teste usada é definida por

$$T_R \equiv \max_{i,j \in \mathcal{M}} |t_{ij}|. \quad (15)$$

A aproximação da distribuição desconhecida da estatística de teste (15) usada na implementação do procedimento MCS é feita através do *bootstrap* estacionário proposto por Politis & Romano (1994), sendo que o tamanho do bloco depende do horizonte de previsão¹ e, como mencionado acima, a função de perda, L , considerada é a raiz do erro de previsão ao quadrado.

4. DADOS E RESULTADOS EMPÍRICOS

Nesta seção é apresentada a base de dados utilizada e os resultados empíricos obtidos com as diferentes especificações do modelo TVP-VAR descrito anteriormente e com os modelos competidores considerados.

¹Para previsões 1-mês à frente é utilizado bloco de tamanho 3 meses, para 3-meses à frente usa-se bloco de 6 meses, para 6-meses à frente o bloco é de 12 meses e para 12-meses à frente usa-se bloco de tamanho 20 meses.

4.1. Dados

A base de dados utilizada é composta por 20 séries macroeconômicas brasileiras compreendendo uma ampla gama de categorias, tais como indicador do PIB, produção industrial, utilização da capacidade, emprego e desemprego, preços ao consumidor, preços ao produtor, salários, lançamentos imobiliários, mercado de ações, taxa de juros, taxa de câmbio e agregados monetários. A escolha das variáveis foi baseada em Koop & Korobilis (2013), ajustando para a disponibilidade das séries para o Brasil. Todas as séries foram obtidas do IPEA ou do Banco Central do Brasil e cobrem os meses de 2003:M1 a 2014:M3. Seguindo Stock & Watson (2009) e Koop & Korobilis (2013), todas as séries são transformadas de forma que fiquem estacionárias. Conforme já mencionado, TVP-VARs de diferentes tamanhos são considerados: TVP-VAR de pequena dimensão com três variáveis, TVP-VAR de média dimensão com 7 variáveis e um TVP-VAR de elevada dimensão com 20 variáveis.

O Apêndice apresenta a lista das variáveis, as transformações realizadas e que variáveis são utilizadas nos diferentes tamanhos de TVP-VAR considerados. O desempenho dos modelos é analisado em relação às previsões para o IPCA (inflação), IBC-Br (atividade econômica) e taxa SELIC realizada (taxa de juros). Estas são as únicas três variáveis utilizadas no TVP-VAR de pequena dimensão.

4.2. Resultados

As previsões recursivas foram calculadas com base em uma janela em expansão, sendo que a janela inicial possui 48 observações para cada uma das séries, compreendendo o período de 2003:M1 a 2007:M1. Além das especificações individuais, são consideradas também as estratégias DMA e DMS aplicadas a modelos TVP-VAR estimadas através da abordagem de fatores de esquecimento, permitindo a alternância entre modelos de diferentes dimensões ao longo do tempo.

Com este exercício de previsão, pretende-se compreender melhor se: i) modelos com parâmetros variando ao longo do tempo possuem maior habilidade preditiva; ii) se modelos heterocedásticos melhoram a previsão da média condicional em relação a modelos homocedásticos; e iii) se o uso de um grande número de variáveis melhora a previsão das três séries de interesse.

A Figura 1 mostra os valores selecionados do hiperparâmetro γ da distribuição a priori para as diversas dimensões de TVP-VARs em cada ponto do tempo. Similar ao que foi encontrado por Koop & Korobilis (2013), percebe-se aqui também uma tendência de crescimento do grau de encolhimento à medida que número de variáveis (e consequentemente o número de parâmetros) do TVP-VAR aumenta, o que demonstra a importância crescente de encolhimento à medida que o número de parâmetros a ser estimado aumenta. Já a Figura 2 apresenta os valores selecionados pela metodologia DMS para o parâmetro λ em cada instante do tempo para o TVP-VAR pequeno, médio e grande. Apesar de exibir alguma variação ao longo do tempo, os valores ótimos para o λ geralmente ficam próximos a um, indicando que os coeficientes do VAR exibem pouca variação.

A Figura 3 apresenta evolução das probabilidades posteriores para cada um dos modelos TVP-VAR de diferentes tamanhos usadas nos procedimentos DMS e DMS. Nota-se que o modelo TVP-VAR de pequena dimensão aparentemente é superior aos demais, ao menos no período posterior a 2008, quando a probabilidade posterior do TVP-VAR pequeno fica próxima a 0,50. Dado que no TVP-VAR-DMS o modelo utilizado é aquele com maior probabilidade, o TVP-VAR pequeno é o mais usado nessa abordagem para o período de 2008 em diante. Nos demais pontos do tempo há bastante mudanças entre os três tamanhos de TVP-VAR considerados. Por exemplo, no período anterior à crise financeira (entre 2005 e 2009), a metodologia DMS seleciona o TVP-VAR grande em vários momentos para fazer as previsões.

O exercício de previsão é realizado em pseudo tempo real, isto é, nunca são utilizadas informações que não estavam disponíveis no momento em que a previsão é realizada. Cabe ressaltar que nos casos onde o interesse é prever diversos passos à frente as previsões realizadas são iteradas, o que aumenta o custo computacional uma vez que é necessário fazer simulação preditiva. Isso ocorre pois quando $h > 1$ não existe uma forma analítica para a densidade preditiva. As simulações preditivas são realizadas de duas formas diferentes: na primeira e mais simples, mantêm-se fixos os parâmetros do TVP-VAR obtidos



Figura 1. Valores do parâmetro de encolhimento γ , estimado para cada ponto do tempo e cada dimensão do VAR.

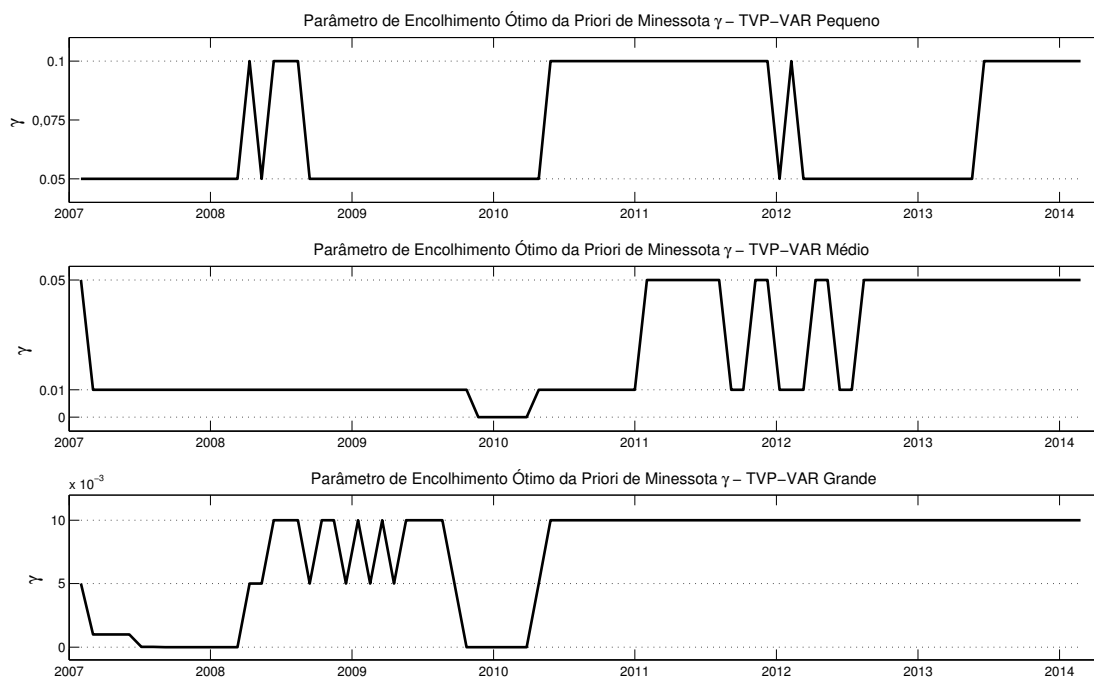


Figura 2. Valores do fator de esquecimento λ , estimado para cada ponto do tempo e cada dimensão do VAR.

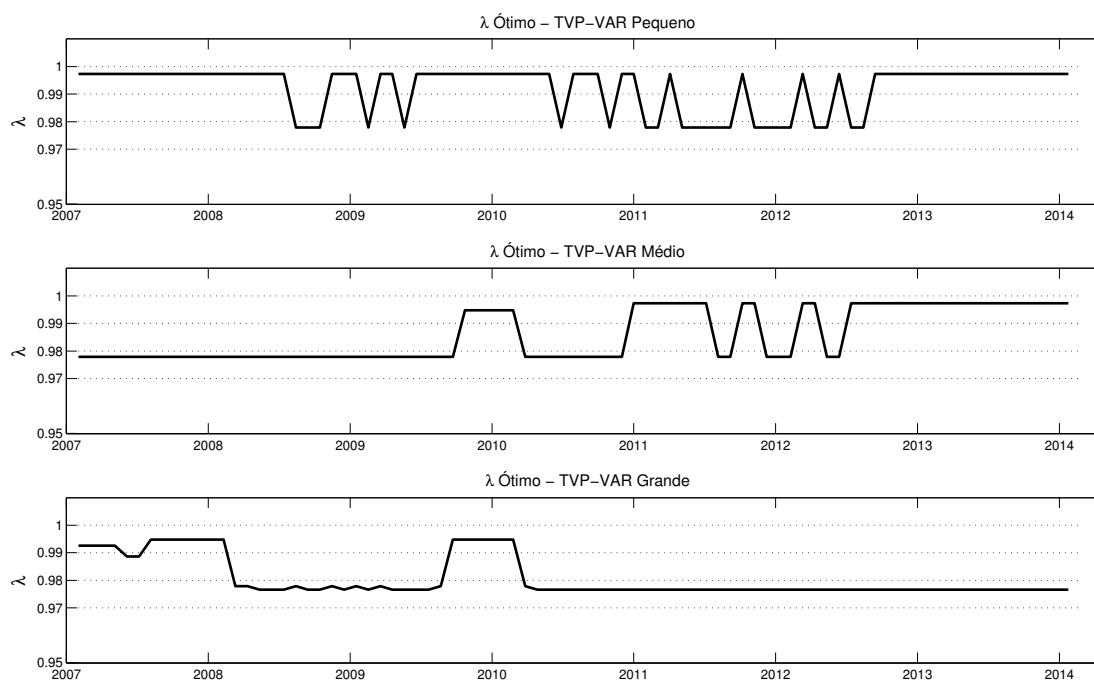
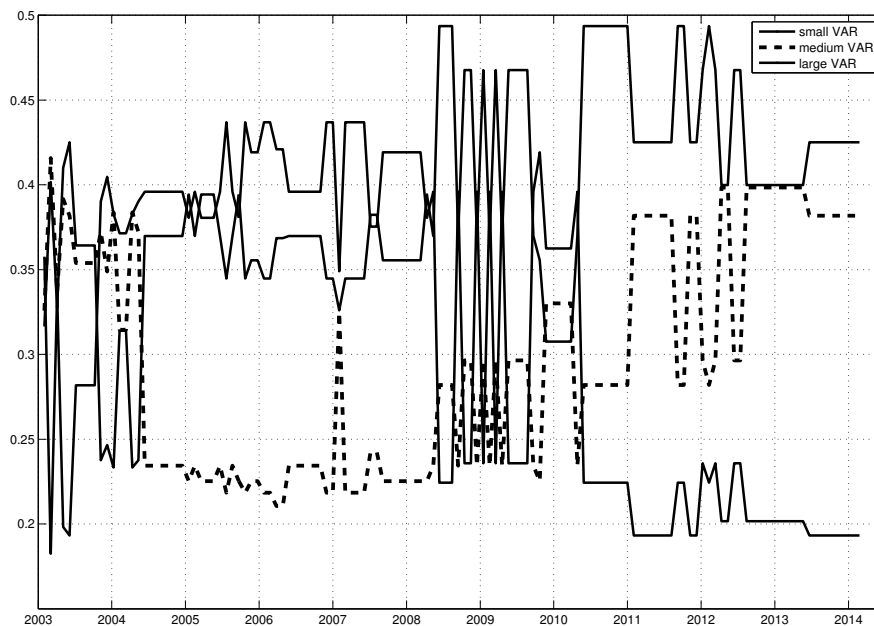


Figura 3. Probabilidades a posteriori dos modelos TVP-VARs pequeno/médio/grande.

no tempo T para prever as variáveis em $T+h$. Na segunda alternativa, chamada de $\beta_{T+h} \sim RW$, permite-se que os parâmetros mudem nas previsões fora da amostra através da simulação a partir da equação de transição dos estados (2) para obter-se uma realização de β_{T+h} .

A acurácia das previsões fora da amostra é avaliada em termos da média dos erros de previsão ao quadrado (MSFE), sendo que são reportados os resultados dos MSFEs relativos ao modelo considerado como *benchmark* (TVP-VAR-DMA):

$$\text{RMSFE}_m^i = \frac{\text{MSFE}_m^i}{\text{MSFE}_m^{\text{bench}}}$$

Logo, um número maior do que 1 implica que o modelo TVP-VAR com estratégia DMA apresenta desempenho melhor do que o modelo considerado.

As tabelas 1 a 3 apresentam os MSFEs de cada modelo individual relativos aos MSFEs resultantes das previsões com o modelo TVP-VAR-DMA para os horizontes de previsão $h = 1, 3, 6$ e 12 meses, sendo que o período de avaliação das previsões se inicia em 2007:M2 e vai até 2014:M3. Para as previsões calculadas através do filtro de Kalman, as tabelas trazem também a soma da log-verossimilhança preditiva, que é uma métrica de previsão que envolve toda a distribuição preditiva (para uma motivação mais completa do uso de verossimilhanças preditivas ver, por exemplo, Geweke & Amisano, 2011). Mais precisamente, é apresentado o resultado da soma da log-verossimilhança preditiva para cada modelo individual menos a soma da log-verossimilhança preditiva para o TVP-VAR-DMA. Portanto, valores negativos indicam que o modelo em questão tem pior desempenho do que o modelo *benchmark*. A abordagem MCS é utilizada nas tabelas para selecionar o conjunto de modelos que apresentam previsões superiores para uma dada função de perda.² Os resultados do MCS são apresentados de forma que um asterisco (*) indica que aquela previsão particular está contida no MCS ao nível de $\alpha = 0,10$.

Ao analisar a parte da Tabela 1 referente aos MSFEs percebe-se que a maior parte dela é preenchida com números maiores do que um, indicando que a abordagem TVP-VAR DMA está prevendo melhor que

²A função de perda, L , considerada aqui é o erro de previsão ao quadrado.



Tabela 1. MSFE e soma da log-verossimilhança preditiva relativa ao TVP-VAR-DMA: IBC-Br.

Modelo	MSFE				Soma da log-verossimilhança preditiva			
	<i>h</i> = 1	<i>h</i> = 3	<i>h</i> = 6	<i>h</i> = 12	<i>h</i> = 1	<i>h</i> = 3	<i>h</i> = 6	<i>h</i> = 12
VAR Pequeno								
TVP-VAR	1,11	0,98*	1,01	1,00*	-2,3	1,8	-0,1	-1,2
TVP-VAR ($\beta_{T+h} \sim RW$)	1,13	0,97*	1,01	1,00*	-3,2	1,3	-0,9	-2,3
TVP-VAR($\lambda = 0,99; \kappa = 0,96; \alpha = 0,99$)	1,12	0,98*	1,02	1,00	-3,2	1,0	-0,5	-1,1
TVP-VAR ($\alpha = 1$)	1,12	0,98*	1,01	1,01	-2,3	1,9	-0,2	-1,0
TVP-VAR ($\alpha = 0,95$)	1,12	0,99*	1,01	1,00*	-2,4	2,4	0,4	-0,4
TVP-VAR ($\alpha = 0,001$)	1,08	0,97*	1,01	0,99*	-2,0	2,0	0,9	0,6
VAR. heterocedastico	1,11	0,98*	0,99*	1,01	-1,6	2,2	0,3	-0,7
VAR. homocedastico	1,13	0,99*	1,00	0,99*	2,7	3,9	4,2	3,2
VAR Médio								
TVP-VAR	0,98	1,00*	0,97*	1,01	5,3	0,8	10,0	1,1
TVP-VAR ($\beta_{T+h} \sim RW$)	0,98*	0,99	0,97*	1,00	5,4	0,5	9,7	0,8
TVP-VAR($\lambda = 0,99; \kappa = 0,96; \alpha = 0,99$)	0,98*	1,00*	0,98*	1,00	4,5	-0,1	8,7	0,5
TVP-VAR ($\alpha = 1$)	0,97*	1,00	0,97	1,00	5,8	0,8	10,1	1,0
TVP-VAR ($\alpha = 0,95$)	0,96*	0,99	0,98	0,98*	6,6	1,5	11,1	1,8
TVP-VAR ($\alpha = 0,001$)	1,01	0,99*	1,01	1,02	0,1	0,0	11,7	-0,7
VAR. heterocedastico	0,98*	1,01	0,97*	1,02	0,8	1,4	10,9	-0,6
VAR. homocedastico	1,02	1,07	0,98*	1,09	-1,7	0,1	10,6	0,1
VAR Grande								
TVP-VAR	1,00	0,99*	0,96*	1,01	0,7	1,2	11,6	-0,8
TVP-VAR ($\beta_{T+h} \sim RW$)	0,98*	1,01	0,99	0,99*	3,5	0,7	12,8	3,0
TVP-VAR($\lambda = 0,99; \kappa = 0,96; \alpha = 0,99$)	0,98*	1,01	0,98*	1,03	5,0	-3,2	10,5	-7,0
TVP-VAR ($\alpha = 1$)	0,99	0,99*	0,96*	1,02	0,4	8,5	11,5	-2,4
TVP-VAR ($\alpha = 0,95$)	0,98*	1,00	0,95*	1,02	1,7	0,2	11,2	-1,1
TVP-VAR ($\alpha = 0,001$)	1,01	1,03	1,03	1,09	-0,5	-4,7	-0,6	-14,3
VAR. heterocedastico	0,97*	1,01*	0,97*	1,06	1,8	-1,5	11,1	-3,0
VAR. homocedastico	0,98*	1,05	0,98*	1,12	0,7	-6,9	10,5	-12,7
Modelos Competidores								
Passo Aleatório	1,85	1,86	2,09	1,93				
VAR(1) MQO pequeno	1,19	1,18	1,02	1,04				
VAR(2) MQO pequeno	0,97*	1,13	1,04	1,06				
BVAR pequeno	1,14	1,04	1,04	1,05				
BVAR médio	1,08	1,03	1,03	1,05				
BVAR grande	1,07	1,02	1,01	1,04				
FAVAR (1-Fator)	1,23	1,05	1,02	1,03				
FAVAR (3-Fatores)	1,20	1,05	1,03	1,03				
FAVAR (5-Fatores)	1,16	1,05	1,03	1,04				
DMA/DMS VAR								
TVP-VAR-DMA	1,00*	1,00*	1,00*	1,00*	0,0	0,0	0,0	0,0
TVP-VAR-DMS	1,00*	0,97*	0,98*	0,98*	0,8	0,9	10,5	11,1
TVP-VAR-DMA ($\beta_{T+h} \sim RW$)	0,97*	0,98*	0,96*	0,99*	-2,8	-4,8	-4,64	-4,45
TVP-VAR-DMS ($\beta_{T+h} \sim RW$)	0,95*	0,97*	1,03	0,98*	2,3	3,0	-1,7	0,67
TVP-VAR-DMA ($\lambda = 0,99; \kappa = 0,96; \alpha = 0,99$)	0,99*	0,98*	1,04	1,00*	0,6	1,0	-2,4	0,4
TVP-VAR-DMS ($\lambda = 0,99; \kappa = 0,96; \alpha = 0,99$)	0,98*	0,97*	0,98*	0,97*	1,0	0,9	1,5	3,0
TVP-VAR-DMA ($\alpha = 1$)	0,99*	0,99*	0,99*	1,00*	-7,2	1,8	-1,0	0,6
TVP-VAR-DMS ($\alpha = 1$)	1,00*	0,98*	1,00*	0,99*	1,0	8,7	1,3	2,8
TVP-VAR-DMA ($\alpha = 0,95$)	0,99*	1,03	1,05	1,00*	0,7	-0,9	1,4	0,6
TVP-VAR-DMS ($\alpha = 0,95$)	1,00*	0,97*	0,99*	0,98*	0,5	8,9	10,5	10,4
TVP-VAR-DMA ($\alpha = 0,001$)	1,02	0,98*	1,04	1,02*	-8,0	-2,5	-1,6	-1,7
TVP-VAR-DMS ($\alpha = 0,001$)	0,99*	0,96*	1,00*	1,01*	-0,2	4,7	2,0	-6,4

Notas: Esta tabela apresenta a média dos erros de previsão ao quadrado (MSFE) de cada modelo específico relativos à média dos erros de previsão ao quadrado do modelo TVP-VAR-DMA. Valores menores que 1 indicam que o modelo em questão supera o TVP-VAR-DMA em termos de previsões. Alternativamente, valores maiores do que 1 indicam que o modelo em questão faz previsões piores que TVP-VAR-DMA por esse critério. As log-verossimilhanças preditivas 1-passo à frente também são apresentadas em relação ao modelo TVP-VAR-DMA, considerado como *benchmark*. Valores negativos (positivos) indicam que o modelo *benchmark* é melhor (pior) que o modelo em questão. Os asteriscos indicam que o modelo faz parte do $\hat{M}_{0,90}^*$, o conjunto dos modelos superiores que contém os melhores modelos com probabilidade não menor que 90%. Valores em negrito indicam o melhor desempenho no horizonte de previsão considerado.

Tabela 2. MSFE e soma da log-verossimilhança preditiva relativa ao TVP-VAR-DMA: IPCA.

Modelo	MSFE				Soma da log-verossimilhança preditiva			
	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 1$	$h = 3$	$h = 6$	$h = 12$
VAR Pequeno								
TVP-VAR	1,02	1,01*	0,98	0,90*	1,0	0,7	-1,4	7,0
TVP-VAR ($\beta_{T+h} \sim RW$)	1,05	1,01*	1,05	0,93*	1,8	-0,3	-1,4	3,3
TVP-VAR ($\lambda = 0,99; \kappa = 0,96; \alpha = 0,99$)	1,01*	1,01*	1,09	0,92*	3,7	0,8	2,5	6,6
TVP-VAR ($\alpha = 1$)	1,02	1,00*	1,05	0,92*	1,1	0,9	-1,3	7,3
TVP-VAR ($\alpha = 0,95$)	1,03	1,01*	1,03	0,96*	-0,9	1,5	-0,7	5,2
TVP-VAR ($\alpha = 0,001$)	1,04	1,00	1,06	0,97*	-1,0	-5,3	-0,9	0,5
VAR. heterocedastico	1,03	1,00	1,02	0,95*	0,9	-5,7	-0,3	2,5
VAR. homocedastico	1,03	1,01	1,00*	0,94*	1,5	5,2	2,7	5,2
VAR Médio								
TVP-VAR	1,11	0,97*	0,91*	1,07	-2,8	9,9	22,0	3,7
TVP-VAR ($\beta_{T+h} \sim RW$)	1,09	0,96*	0,92*	1,08	-1,8	10,0	21,0	2,1
TVP-VAR ($\lambda = 0,99; \kappa = 0,96; \alpha = 0,99$)	1,14	0,98*	0,93*	1,08	-4,6	7,9	18,0	2,6
TVP-VAR ($\alpha = 1$)	1,08	0,96*	0,91*	1,07	-1,8	10,0	22,6	3,5
TVP-VAR ($\alpha = 0,95$)	1,27	1,01*	0,92*	1,07	-3,1	8,1	20,0	3,1
TVP-VAR ($\alpha = 0,001$)	1,06	0,94*	0,99*	1,07	1,0	9,5	19,5	-2,7
VAR. heterocedastico	1,14	0,97*	0,90*	1,03*	-2,6	11	23,7	-1,1
VAR. homocedastico	1,29	0,99*	0,88*	1,08	-9,8	6,1	19,9	-1,2
VAR Grande								
TVP-VAR	1,08	0,93*	0,92*	1,07	-3,8	9,8	21,4	0,1
TVP-VAR ($\beta_{T+h} \sim RW$)	1,34	0,99	0,87*	1,09	-8,6	1,0	25,0	4,3
TVP-VAR ($\lambda = 0,99; \kappa = 0,96; \alpha = 0,99$)	1,30	1,03	0,90	1,05	-10,0	6,1	19,7	-2,8
TVP-VAR ($\alpha = 1$)	1,07	0,92*	0,92*	1,06	-3,6	9,6	21,0	-0,2
TVP-VAR ($\alpha = 0,95$)	1,14	0,97*	0,89*	1,08	-4,8	7,9	20,1	0,4
TVP-VAR ($\alpha = 0,001$)	1,12	0,94*	0,93*	1,11	-4,6	8,2	18,7	-8,5
VAR. heterocedastico	1,23	0,98*	0,89*	1,11	-6,4	10,0	21,9	-0,6
VAR. homocedastico	1,34	1,03	0,92*	1,08	-11,0	-5,5	19,1	-2,1
Modelos Competidores								
Passo Aleatório	1,05	1,65	2,14	1,18				
VAR(1) MQO pequeno	1,17	1,05	0,98*	0,99				
VAR(2) MQO pequeno	0,94*	0,96*	0,96*	0,99				
BVAR pequeno	0,95*	0,96*	0,98*	1,00				
BVAR médio	0,93*	0,97*	0,99	1,00				
BVAR grande	0,92*	0,94*	0,95*	1,01				
FAVAR (1-Fator)	0,97*	0,95*	0,98*	1,01				
FAVAR (3-Fatores)	0,96*	0,97*	0,99	1,01				
FAVAR (5-Fatores)	0,95*	0,96*	1,02	1,00				
DMA/DMS VAR								
TVP-VAR-DMA	1,00*	1,00*	1,00*	1,00*	0,0	0,0	0,0	0,0
TVP-VAR-DMS	0,94*	0,99*	1,03	0,98*	2,6	5,9	-6,7	4,7
TVP-VAR-DMA ($\beta_{T+h} \sim RW$)	0,99*	0,99*	0,99*	1,03	7,5	3,5	0,8	-6,5
TVP-VAR-DMS ($\beta_{T+h} \sim RW$)	1,00*	0,99*	1,00	0,98*	0,3	8,1	9,8	7,2
TVP-VAR-DMA ($\lambda = 0,99; \kappa = 0,96; \alpha = 0,99$)	0,99*	0,98*	1,04	0,97*	2,3	3,5	-1,1	3,9
TVP-VAR-DMS ($\lambda = 0,99; \kappa = 0,96; \alpha = 0,99$)	0,94*	0,99*	1,03	0,95*	2,6	4,1	-3,6	5,2
TVP-VAR-DMA ($\alpha = 1$)	1,00*	1,00	0,97*	0,99*	0,0	0,0	1,4	-0,1
TVP-VAR-DMS ($\alpha = 1$)	1,00*	1,01	1,04	0,96*	1,4	4,0	-0,8	2,3
TVP-VAR-DMA ($\alpha = 0,95$)	1,04	1,01	0,99*	1,02	-6,0	-5,9	0,6	-4,6
TVP-VAR-DMS ($\alpha = 0,95$)	0,99*	0,99*	1,01	1,01	0,8	4,0	-1,7	-2,9
TVP-VAR-DMA ($\alpha = 0,001$)	1,07	1,03	1,06	1,01*	-6,1	-12,0	-1,6	-10,0
TVP-VAR-DMS ($\alpha = 0,001$)	1,05	1,00	1,07	0,96*	-1,1	1,5	-12,0	0,6

Notas: Esta tabela apresenta a média dos erros de previsão ao quadrado (MSFE) de cada modelo específico relativos à média dos erros de previsão ao quadrado do modelo TVP-VAR-DMA. Valores menores que 1 indicam que o modelo em questão supera o TVP-VAR-DMA em termos de previsões. Alternativamente, valores maiores do que 1 indicam que o modelo em questão faz previsões piores que TVP-VAR-DMA por esse critério. As log-verossimilhanças preditivas 1-passo à frente também são apresentadas em relação ao modelo TVP-VAR-DMA, considerado como *benchmark*. Valores negativos (positivos) indicam que o modelo *benchmark* é melhor (pior) que o modelo em questão. Os asteriscos indicam que o modelo faz parte do $\hat{M}_{0,90}^*$, o conjunto dos modelos superiores que contém os melhores modelos com probabilidade não menor que 90%. Valores em negrito indicam o melhor desempenho no horizonte de previsão considerado.



Tabela 3. MSFE e soma da log-verossimilhança preditiva relativa ao TVP-VAR-DMA: SELIC.

Modelo	MSFE				Soma da log-verossimilhança preditiva			
	<i>h</i> = 1	<i>h</i> = 3	<i>h</i> = 6	<i>h</i> = 12	<i>h</i> = 1	<i>h</i> = 3	<i>h</i> = 6	<i>h</i> = 12
VAR Pequeno								
TVP-VAR	0,93*	0,86*	0,97*	1,05	5,5	0,3	4,4	0,6
TVP-VAR ($\beta_{T+h} \sim RW$)	0,98*	0,85*	0,99*	1,01	5,0	5,2	3,8	-3,2
TVP-VAR ($\lambda = 0,99; \kappa = 0,96; \alpha = 0,99$)	0,91*	0,83*	0,97*	1,06	5,9	3,8	3,7	-4,5
TVP-VAR ($\alpha = 1$)	0,94*	0,84*	0,97*	1,05	9,1	4,4	8,1	-4,0
TVP-VAR ($\alpha = 0,95$)	0,94*	0,87*	0,97*	1,00*	4,9	4,8	4,6	-0,6
TVP-VAR ($\alpha = 0,001$)	1,05	0,87*	0,99*	1,05	1,5	-0,5	4,6	-3,9
VAR. heterocedastico	0,94*	0,87*	0,97*	1,05	5,1	-0,1	4,6	-4,1
VAR. homocedastico	0,98*	0,88*	1,01	1,02	-5,9	-1,1	-2,9	-5,0
VAR Médio								
TVP-VAR	1,22	1,19	1,06	0,95*	-6,9	-10,0	3,6	15,0
TVP-VAR ($\beta_{T+h} \sim RW$)	0,91*	0,92*	0,90*	0,91*	12,0	3,7	8,8	24,0
TVP-VAR ($\lambda = 0,99; \kappa = 0,96; \alpha = 0,99$)	1,22	1,17	1,06	0,97*	-7,5	-11,0	-2,6	14,0
TVP-VAR ($\alpha = 1$)	1,20	1,17	1,05	0,95*	-6,4	-9,9	-3,49	15,0
TVP-VAR ($\alpha = 0,95$)	1,20	1,22	1,06	0,95*	-6,3	-10,0	-3,54	16,0
TVP-VAR ($\alpha = 0,001$)	1,22	1,10	1,15	1,10	-7,3	-10,0	-1,9	-1,3
VAR. heterocedastico	1,21	1,20	1,06	0,94*	-6,1	-9,5	-3,73	16,0
VAR. homocedastico	1,36	1,28	1,11	0,97*	-19,0	-21,0	-5,1	9,5
VAR Grande								
TVP-VAR	1,53	1,43	1,11	0,98*	-13,0	-15,0	-3,4	1,4
TVP-VAR ($\beta_{T+h} \sim RW$)	1,19	1,28	0,98*	0,92*	-2,9	-7,8	6,32	25,0
TVP-VAR ($\lambda = 0,99; \kappa = 0,96; \alpha = 0,99$)	1,48	1,37	1,07	0,94*	-12,0	-17,0	1,96	15,0
TVP-VAR ($\alpha = 1$)	1,53	1,43	1,10	0,99*	-13,0	-15,0	-0,67	1,8
TVP-VAR ($\alpha = 0,95$)	1,44	1,35	1,13	1,03	-10,0	-12,3	-1,48	-1,2
TVP-VAR ($\alpha = 0,001$)	1,23	1,30	1,12	1,09	-8,3	-14,0	-3,2	-8,2
VAR. heterocedastico	1,42	1,38	1,13	1,03*	-9,6	-13,0	-2,59	-2,0
VAR. homocedastico	1,32	1,26	1,02	0,95*	-18,0	-20,0	-3,1	10,0
Modelos Competidores								
Passeio Aleatório	1,49	1,70	1,05	1,31				
VAR(1) MQO pequeno	0,96*	0,99	0,98*	0,98*				
VAR(2) MQO pequeno	1,17	1,06	1,01	0,96*				
BVAR pequeno	1,11	1,09	1,01	0,97*				
BVAR médio	1,14	1,05	1,00	0,99				
BVAR grande	1,13	1,02	0,98*	1,09				
FAVAR (1-Fator)	1,18	1,02	0,95*	0,90*				
FAVAR (3-Fatores)	1,18	1,02	0,95*	0,90*				
FAVAR (5-Fatores)	1,16	1,00	0,93*	0,88*				
DMA/DMS VAR								
TVP-VAR-DMA	1,00*	1,00*	1,00*	1,00*	0,0	0,0	0,0	0,0
TVP-VAR-DMS	1,00*	0,88*	0,91*	0,99*	1,0	12,0	5,34	2,8
TVP-VAR-DMA ($\beta_{T+h} \sim RW$)	1,02	1,01*	0,99*	0,99*	-1,2	0,6	1,01	4,3
TVP-VAR-DMS ($\beta_{T+h} \sim RW$)	0,98*	0,92*	0,92*	0,99*	2,2	-2,9	3,87	11,2
TVP-VAR-DMA ($\lambda = 0,99; \kappa = 0,96; \alpha = 0,99$)	0,98*	0,97*	1,05	1,06	8,9	7,7	-1,6	-4,1
TVP-VAR-DMS ($\lambda = 0,99; \kappa = 0,96; \alpha = 0,99$)	0,90*	0,81*	0,91*	0,99*	4,8	3,4	5,3	1,2
TVP-VAR-DMA ($\alpha = 1$)	1,00*	1,00	0,99*	0,99	-1,4	0,6	-1,8	1,4
TVP-VAR-DMS ($\alpha = 1$)	0,96*	0,88*	0,93*	1,00	3,7	8,6	5,3	-2,5
TVP-VAR-DMA ($\alpha = 0,95$)	1,04	1,04	1,11	1,10	-5,0	-3,4	-0,2	-1,2
TVP-VAR-DMS ($\alpha = 0,95$)	1,01*	0,93*	0,96*	1,03*	1,9	5,6	3,8	-14,0
TVP-VAR-DMA ($\alpha = 0,001$)	1,18	1,06	1,26	1,14	-4,0	0,6	-9,7	-1,4
TVP-VAR-DMS ($\alpha = 0,001$)	1,05	0,87*	1,12	1,13	-0,4	2,9	-1,1	-9,5

Notas: Esta tabela apresenta a média dos erros de previsão ao quadrado (MSFE) de cada modelo específico relativos à média dos erros de previsão ao quadrado do modelo TVP-VAR-DMA. Valores menores que 1 indicam que o modelo em questão supera o TVP-VAR-DMA em termos de previsões. Alternativamente, valores maiores do que 1 indicam que o modelo em questão faz previsões piores que TVP-VAR-DMA por esse critério. As log-verossimilhanças preditivas 1-passo à frente também são apresentadas em relação ao modelo TVP-VAR-DMA, considerado como *benchmark*. Valores negativos (positivos) indicam que o modelo *benchmark* é melhor (pior) que o modelo em questão. Os asteriscos indicam que o modelo faz parte do $\widehat{M}_{0,90}^*$, o conjunto dos modelos superiores que contém os melhores modelos com probabilidade não menor que 90%. Valores em negrito indicam o melhor desempenho no horizonte de previsão considerado.

os demais modelos. Porém, existem alguns horizontes para os quais os modelos TVP-VAR com dimensão fixa apresentam bom desempenho. Enquanto os TVP-VAR pequenos têm desempenho muito bom para o horizonte $h = 3$, para o horizonte $h = 1$ isto acontece apenas para os modelos com dimensão média e grande. É interessante perceber que, para o IBC-Br, as especificações TVP-VAR-DMA ($\beta_{T+h} \sim RW$) e TVP-VAR-DMS ($\lambda = 0,99$; $\kappa = 0,96$; $\alpha = 0,99$) são as únicas que superam o *benchmark* em todos os horizontes de previsão considerados, indicando a boa qualidade das previsões obtidas através dos TVP-VARs que utilizam as abordagens DMA e DMS. Além disso, nota-se que os modelos competidores tem MSFE quase sempre maior do que um, reforçando a qualidade das previsões do *benchmark* TVP-VAR-DMA.

Já no caso específico das previsões para a inflação (IPCA) apresentadas na Tabela 2, o BVAR grande prevê melhor para o horizonte mais curto ($h = 1$), mas apesar de apresentarem bom desempenho nas previsões para 3 e 6 meses à frente, os BVAR são superados pelo TVP-VAR grande. Nas previsões de mais longo prazo (12-meses à frente), o TVP-VAR pequeno faz o melhor trabalho. Em geral, os MSFEs indicam que os TVP-VARs com DMA e DMS são as abordagens de previsão mais consistentes, apresentando bom desempenho em todos os horizontes considerados. Mais especificamente, somente o TVP-VAR-DMA e o VAR heterocedástico estão no MCS para todos os horizontes de previsão, sendo que o VAR heterocedástico apresenta resultados de MSFE pior do que os obtidos pelo TVP-VAR-DMA para os horizontes $h = 1, 3$ e 6. Além disso, o fato do MCS ser formado por um conjunto bem maior de modelos nesse caso indica a dificuldade em se escolher um modelo superior no caso das previsões para o IPCA.

Na Tabela 3, que traz as previsões para a taxa de juros (SELIC), mais uma vez os TVP-VARs com DMA e DMS obtêm excelentes resultados. Os modelos competidores apresentam resultados inferiores para todos os horizontes de previsão, exceto para $h = 12$, onde o FAVAR com cinco fatores apresenta o melhor resultado. Novamente, somente especificações de TVP-VAR estão presentes no MCS para todos os horizontes de previsão. Além disso, os resultados dos TVP-VARs são muito superiores em horizontes de previsão mais curtos, como $h = 1$ e 3.

Na maioria dos casos, as log-verossimilhanças preditivas indicam que a abordagem DMA prevê ligeiramente melhor do que a DMS, embora esse resultado não se reflita nos MSFEs, onde a abordagem DMS exibe desempenho melhor em muitos casos. Uma possível explicação para este resultado é que as estratégias DMA e DMS possuem diferentes graus de encolhimento e este pode afetar de forma diferente as duas funções perda consideradas. A abordagem DMS atribui peso 1 ao melhor modelo e peso zero aos demais, reduzindo a zero a relevância dos modelos não selecionados. Koop & Korobilis (2013) argumentam que esse encolhimento adicional pode fornecer alguns benefícios de previsão à estratégia DMS em relação à DMA, pois ignora completamente modelos com desempenho inferior. Além disso, diante de mudanças rápidas, a metodologia DMS tenderá a se ajustar mais rapidamente do que a DMA, uma vez que pode selecionar um modelo inteiramente novo, ao invés de ajustar os pesos de todos os modelos.

Analisando especificamente os modelos TVP-VAR com dimensão fixa, o primeiro ponto que chama a atenção é que, para a maioria dos casos, as diferentes especificações consideradas levam a resultados parecidos em termos de MSFE. Diferentes tratamentos para o fator de esquecimento α usado no procedimento DMS que escolhe λ , κ e γ não afeta os resultados de forma considerável. O modelo com pequena dimensão tende a prever melhor a taxa de juros que os modelos médios e grandes, embora existam algumas exceções. Já no caso do IBC-Br, o TVP-VAR de pequena dimensão apresenta desempenho inferior aos modelos médios e grandes. Para a inflação, o TVP-VAR grande é superior, exceto quando se considera as previsões de longo prazo, $h = 12$, caso no qual o TVP-VAR pequeno supera todos os demais modelos. Tais resultados indicam um certo grau de incerteza em relação a quais variáveis incluir no TVP-VAR, o que reforça a importância das abordagens DMA e DMS.

Como já indicado na Figura 2, os parâmetros variantes no tempo não parecem contribuir muito para a previsão, uma vez que os resultados obtidos por TVP-VARs de diferentes tamanhos, apesar de levemente superiores, são similares aos obtidos pelos modelos VAR heterocedásticos (estimados exatamente como o TVP-VAR, mas fixando-se $\lambda = 1$). Por outro lado, a importância de se considerar erros heterocedásticos fica evidente quando o resultado dos modelos VAR homocedásticos e heterocedásticos são comparados. Considerando-se as diferentes variáveis analisadas, tamanhos de modelos e também horizontes de previ-



são, o modelo VAR heterocedástico obtém resultados superiores em relação à sua versão homocedástica na ampla maioria dos casos. Entretanto, exatamente como ocorrido com os modelos TVP-VAR, os resultados para modelos VAR com parâmetros fixos não permitem discriminar entre os diferentes tamanhos, uma vez que o desempenho de cada um deles muda de acordo com o horizonte de previsão e com a variável a ser prevista. Por exemplo, para $h = 6$ o modelo VAR médio exibe melhores resultados, mas para os horizontes mais curtos, o modelo pequeno tem melhor desempenho. Assim, mesmo quando se está trabalhando com os modelos VAR com parâmetros fixos, pode ser interessante o uso das abordagens DMA e DMS, possibilitando melhores previsões nos casos onde há incerteza a respeito da dimensão do VAR.

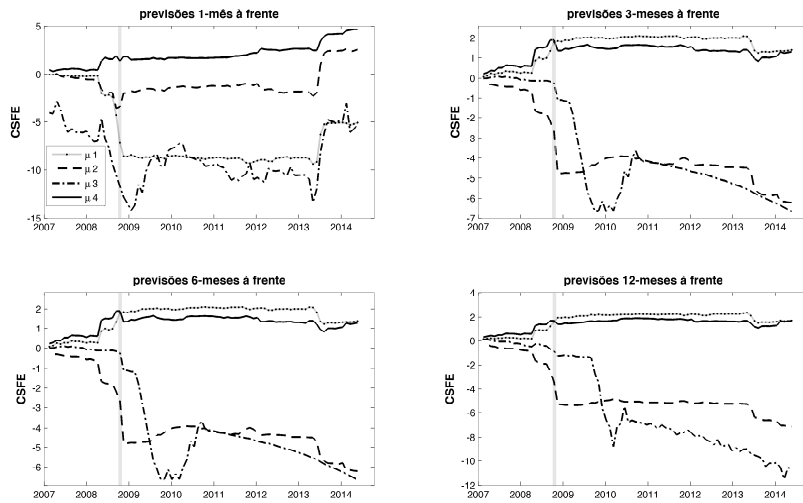
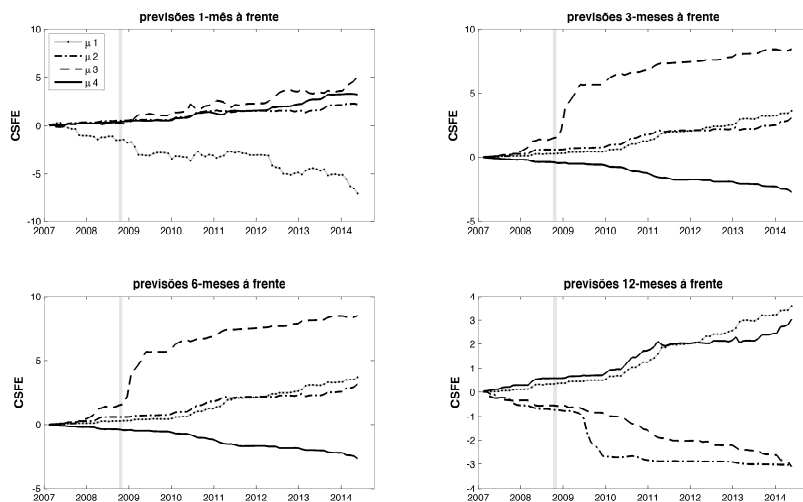
Conforme pode ser observado nas tabelas, os resultados relativos ao MCS indicam que o modelo *benchmark* TVP-VAR-DMA é o único que pertence ao conjunto que contém o melhor modelo com probabilidade não inferior a 90%, $\hat{M}_{0,90}^*$, para todas as variáveis e todos os horizontes de previsão considerado. Ou seja, o modelo TVP-VAR-DMA é o único que apresenta desempenho consistente em todos os 12 problemas de previsão analisados (4 horizontes de previsão e 3 variáveis diferentes).

Em linha com Koop & Korobilis (2013), os resultados encontrados neste trabalho indicam que apesar das abordagens TVP-VAR-DMA e TVP-VAR-DMS nem sempre produzem as melhores previsões, estas estão consistentemente entre os melhores modelos. Entretanto, não é possível identificar uma única especificação que domine todas as demais em todos os horizontes de previsão e variáveis de interesse. Na literatura brasileira, Carlos & Marçal (2013) analisaram diferentes metodologias para a previsão da inflação (IPCA) e também encontram evidências de que não existe um modelo que supere todos os demais em todos os horizontes de previsão. Autores como R. T. Ferreira et al. (2005) e Céspedes et al. (2006) realizaram previsões da taxa de crescimento do PIB real do Brasil usando diferentes abordagens e também não identificaram uma abordagem que domine todas as demais. É justamente essa dificuldade em identificar um único modelo de previsão que sugere a utilização da abordagem MCS.

Segundo Welch & Goyal (2008), para permitir a avaliação dos diferentes modelos em diferentes períodos de tempo, são apresentados gráficos com a diferença entre o erro quadrado de previsão acumulado (CSFE), calculado como em (12), de alguns modelos considerados e do TVP-VAR-DMA ao longo de toda a avaliação fora da amostra. As figuras 4 a 5 mostram o resultado para cada variável e para cada horizonte de previsão. Para facilitar a visualização, o gráfico apresenta apenas os modelos que se mostraram como fortes competidores ao TVP-VAR-DMA em algum momento. São eles: o modelo VAR heterocedástico pequeno (μ_1), o modelo BVAR grande (μ_2), o modelo FAVAR com 5 fatores (μ_3) e o modelo TVP-VAR-DMS com $\lambda = 0,99$, $\kappa = 0,96$ e $\alpha = 0,99$ (μ_4). Cada linha do gráfico representa um modelo diferente e mostra o desempenho individual de cada um em relação ao TVP-VAR-DMA, em particular, um CSFE crescente indica um desempenho superior ao TVP-VAR-DMA, ao passo que CSFE decrescente indica pior desempenho relativo.

A linha vertical nas figuras indica o início da crise financeira internacional.³ Após esse período, as taxas de juros da economia brasileira inicialmente subiram e, em seguida, foram reduzidas drasticamente, de cerca de 14% para 8,5%. Os gráficos do CSFE permitem examinar em detalhes o desempenho mês a mês dos diferentes modelos a partir do início da crise. O CSFE mostra que, para as previsões do IBC-Br, a maioria dos modelos apresentam uma queda no desempenho relativo ao TVP-VAR-DMA no período da crise. O único modelo que não é afetado em nenhum dos horizontes é o TVP-VAR-DMS com $\lambda = 0,99$, $\kappa = 0,96$ e $\alpha = 0,99$, mostrando que a seleção de modelos foi importante para um bom desempenho. Este modelo também apresentou um maior estabilidade durante todo o período analisado, evidenciando a menor volatilidade do erro de previsão ao quadrado obtido pela seleção de modelos. No caso do IPCA, para os horizontes de previsão intermediários, a crise também parece ter favorecido o desempenho relativo dos modelos BVAR. Note que após 2010, a inclinação da linha do modelo BVAR é muito similar às dos modelos VAR heterocedástico e FAVAR, mas o desempenho extremamente superior em uma pequena

³ O período de início da crise utilizado é setembro de 2008, mês da bancarrota do banco de investimentos norte americano Lehman Brothers.

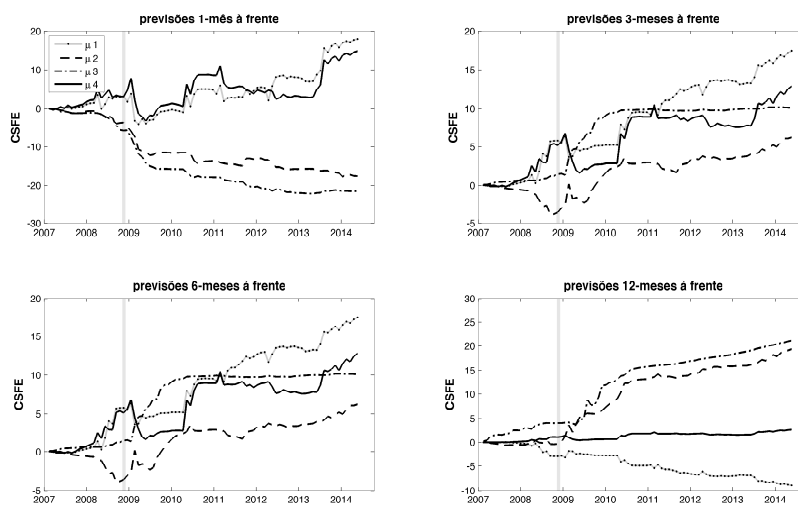
Figura 4. Erros de previsão cumulativos ao quadrado para o IBC-Br e Inflação – CSFE.**(a)** CSFE para o IBC-Br**(b)** CSFE para o IPCA

Notas: As figuras (a) e (b) apresentam os erros de previsão ao quadrado cumulativos (CSFE) em relação ao random walk para as previsões do IBC-Br no Painel (a) e da taxa de inflação (IPCA) no Painel (b). As figuras mostram os CSFEs para os modelos com melhor desempenho nos horizontes de previsão considerados. As barras cinzas indicam período de recessão. As seguintes abreviações são usadas nos gráficos: μ_1 refere-se ao modelo VAR heterocedástico pequeno, μ_2 representa o modelo BVAR de grande dimensão, μ_3 refere-se ao FAVAR com 5 fatores e μ_4 refere-se ao modelo TVP-VAR-DMS $\lambda = 0,99$, $\kappa = 0,96$ e $\alpha = 0,99$).



Figura 5. Erros de previsão cumulativos ao quadrado para a taxa de juros – CSFE.

(a) CSFE para SELIC



Notas: Esta figura apresenta os erros de previsão ao quadrado cumulativos (CSFE) em relação ao random walk para as previsões da taxa de juros. A figura mostra os CSFEs para os modelos com melhor desempenho nos horizontes de previsão considerados. As barras cinzas indicam período de recessão. As seguintes abreviações são usadas nos gráficos: $\mu 1$ refere-se ao modelo VAR heterocedástico pequeno, $\mu 2$ representa o modelo BVAR de grande dimensão, $\mu 3$ refere-se ao FAVAR com 5 fatores e $\mu 4$ refere-se ao modelo TVP-VAR-DMS ($\lambda = 0,99$; $\kappa = 0,96$; $\alpha = 0,99$).

parte da amostra fez com que esse modelo obtivesse um resultado final muito melhor.

5. CONCLUSÃO

Neste artigo investigou-se o uso de diversas especificações de modelos VAR com parâmetros variando no tempo e erros heterocedásticos (TVP-VAR) para a previsão do IBC-Br, inflação e taxa de juros no Brasil. Analisou-se também a estratégia de combinação dinâmica (DMA) de diferentes especificações de TVP-VAR, bem como a estratégia de seleção dinâmica da melhor especificação (DMS). Os resultados das previsões geradas pelos TVP-VAR foram comparadas com modelos VAR bayesianos (BVAR) e modelos VAR aumentados por fatores (FAVAR). As estratégias DMA e DMS ampliam a abordagem TVP-VAR ao permitir não apenas que os parâmetros dos modelos variem ao longo do tempo, mas também permitindo mudanças entre os diferentes modelos. A estimação dos modelos TVP-VAR é facilitada pelo uso de fatores de esquecimento sendo que o modelo final exige apenas a definição de 3 parâmetros, independentemente do número de variáveis utilizadas no VAR, acabando com a “maldição da dimensionalidade” que tanto restringe as aplicações empíricas de modelos VAR. Portanto, este trabalho está diretamente relacionado com o exercício de previsão implementado por Koop & Korobilis (2013) para a economia norte-americana.

As evidências empíricas encontradas indicam os benefícios das abordagens DMA e DMS. Em particular, observa-se que a estratégia TVP-VAR-DMA apresenta previsões satisfatórias para todas as variáveis analisadas e em todos os quatro horizontes de previsão considerados. Mais especificamente, a avaliação de previsões feita através da metodologia de *model confidence set* indicou que o TVP-VAR-DMA é a única das 44 metodologias consideradas que está sempre incluída no conjunto que contém o melhor modelo de previsão com 90% de confiança. Com isso, os resultados mostram que modelos TVP-VAR são ferramentas adequadas para previsão de variáveis macroeconômicas quando se dispõe de um grande painel

de dados e quando há incerteza em relação à dimensão do modelo a ser utilizado.

Adicionalmente, apresenta-se também evidências de que a inclusão de efeitos heterocedásticos são importantes mesmo quando o foco é apenas a previsão da média condicional. Já a inclusão de parâmetros variantes no tempo e de uma grande base de dados contendo várias variáveis não parece ter um efeito tão consistente em todos os casos analisados. TVP-VARS, BVARs e modelos FAVAR que utilizam todas as 20 variáveis disponíveis apresentam resultados muito bons para a previsão de inflação, porém, este padrão não se repete para a previsão do IBC-Br, nem para a previsão da taxa SELIC.

REFERÊNCIAS BIBLIOGRÁFICAS

- Arruda, E. F., Ferreira, R. T. & Castelar, I. (2011). Modelos lineares e não lineares da curva de Phillips para previsão da taxa de inflação no Brasil. *Revista Brasileira de Economia*, 65(3), 237–252. Disponível em: <http://bibliotecadigital.fgv.br/ojs/index.php/rbe/article/view/1523>
- Bañbura, M., Giannone, D. & Reichlin, L. (2010). Large bayesian vector auto regressions. *Journal of Applied Econometrics*, 25(1), 71–92. doi: 10.1002/jae.1137
- Bernanke, B. S., & Boivin, J. (2003). Monetary policy in a data-rich environment. *Journal of Monetary Economics*, 50(3), 525–546. doi: 10.1016/S0304-3932(03)00024-2
- Bernanke, B. S., Boivin, J. & Elias, P. S. (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (favar) approach. *The Quarterly Journal of Economics*, 120(1), 387–422. doi: 10.1162/0033553053327452
- Brockwell, P. J., & Davis, R. A. (2009). *Time series: Theory and methods* [reprint of the 1991 edition]. Springer.
- Carlos, T. C., & Marçal, E. F. (2013, August 28). Forecasting Brazilian inflation by its aggregate and disaggregated data: A test of predictive power by forecast horizon. In S. Henzel (Presid.), *European Economic Association & Econometric Society 2013 Parallel Meetings*, Gothenburg, Sweden. Disponível em: <http://www.eea-esem.com/eea-esem/2013/prog/viewpaper.asp?pid=553>
- Carriero, A., Kapetanios, G. & Marcellino, M. (2009). Forecasting exchange rates with a large Bayesian VAR. *International Journal of Forecasting*, 25(2), 400–417. doi: 10.1016/j.ijforecast.2009.01.007
- Carriero, A., Kapetanios, G. & Marcellino, M. (2012). Forecasting government bond yields with large Bayesian vector autoregressions. *Journal of Banking & Finance*, 36(1), 2026–2047. doi: 10.1016/j.jbankfin.2012.03.008
- Céspedes, B. J. V., Chauvet, M. & Lima, E. C. R. (2006). Forecasting Brazilian output and its turning points in the presence of breaks: A comparison of linear and nonlinear models. *Estudos Econômicos*, 36(1), 5–46. Disponível em: <http://www.revistas.usp.br/ee/article/view/35863>
- Chauvet, M. (2001). Leading indicators of inflation for Brazil. *Pesquisa e Planejamento Econômico*, 31(1), 43–74. Disponível em: <http://ppe.ipea.gov.br/index.php/ppe/article/view/160/95>
- Cogley, T., & Sargent, T. J. (2005). Drift and volatilities: Monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, 8(2), 262–302. doi: 10.1016/j.red.2004.10.009
- Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics*, 13(3), 253–263. doi: 10.1080/07350015.1995.10524599
- Doan, T., Litterman, R. B. & Sims, C. A. (1984). Forecasting and conditional projection using realistic prior distribution. *Econometric Reviews*, 3(1), 1–144. doi: 10.1080/07474938408800053
- Ferreira, D., & Palma, A. (2014). Forecasting inflation with the Phillips curve: A dynamic model averaging approach for Brazil. In *Anais das XXIX Jornada Anuales de Economía*. Banco Central do Uruguay.
- Ferreira, R. T., Bierens, H. & Castelar, I. (2005). Forecasting quarterly Brazilian GDP growth rate with linear and nonlinear diffusion index models. *Economia* (revista da ANPEC), 6(3), 261–292. Disponível em: http://www.anpec.org.br/revista/vol6/vol6n3p261_292.pdf
- Frühwirth-Schnatter, S. (2006). *Finite mixture and markov switching models*. New York, NY: Springer.
- Geweke, J., & Amisano, G. (2011). Hierarchical Markov normal mixture models with applications to financial asset returns. *Journal of Applied Econometrics*, 26(1), 1–29. doi: 10.1002/jae.1119
- Giannone, D., Lenza, M., Momferatou, D. & Onorante, L. (2014). Short-term inflation projections: A Bayesian vector autoregressive approach. *International Journal of Forecasting*, 30(3), 635–644. doi:



- 10.1016/j.ijforecast.2013.01.012
- Giannone, D., & Reichlin, L. (2006). Does information help recovering structural shocks from past observations? *Journal of the European Economic Association*, 4(2-3), 455–465. doi: 10.1162/jeea.2006.4.2-3.455
- Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics*, 23(4), 365–380. doi: 10.1198/073500105000000063
- Hansen, P. R., Lunde, A. & Nason, J. M. (2011). The model confidence set. *Econometrica*, 79(2), 453–497. doi: 10.3982/ECTA5771
- Jazwinski, A. H. (1970). *Stochastic processes and filtering theory*. Academic Press.
- Koop, G., & Korobilis, D. (2013). Large time-varying parameter VARs. *Journal of Econometrics*, 177(2), 185–198. doi: 10.1016/j.jeconom.2013.04.007
- Litterman, R. B. (1986). Forecasting with bayesian vector autoregressions: Five years of experience. *Journal of Business & Economic Statistics*, 4(1), 25–38. doi: 10.1080/07350015.1986.10509491
- Politis, D. N., & Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical Association*, 89(428), 1303–1313.
- Raftery, A. E., Kárný, M. & Ettler, P. (2010). Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. *Technometrics*, 52(1), 52–66. doi: 10.1198/TECH.2009.08104
- Rossi, J. L., Jr., & Carvalho, M. D. d. (2009). Identification of monetary policy shocks and its effects: Favara methodology for the brazilian economy. *Brazilian Review of Econometrics*, 29(2), 285–313. Disponível em: <http://bibliotecadigital.fgv.br/ojs/index.php/bre/article/view/3444>
- Sims, C. A. (1980). Macroeconomics and reality. *Econometrica*, 48(1), 1–48. doi: 10.2307/1912017
- Stock, J. H., & Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97(460), 1167–1179. doi: 10.1198/016214502388618960
- Stock, J. H., & Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2), 147–162. doi: 10.1198/073500102317351921
- Stock, J. H., & Watson, M. W. (2005, June). *Implications of dynamic factor models for VAR analysis* (Working Paper N° 11467). National Bureau of Economic Research (NBER). doi: 10.3386/w11467
- Stock, J. H., & Watson, M. W. (2009). Forecasting in dynamic factor models subject to structural instability. In J. Castle & N. Shephard (Eds.), *The methodology and practice of econometrics: Festschrift in honor of d. f. hendry* (pp. 1–57). Oxford University Press. doi: 10.1093/acprof:oso/9780199237197.003.0007
- Welch, I., & Goyal, A. (2008). A comprehensive look at the empirical performance of equity premium prediction. *The Review of Financial Studies*, 21(4), 1455–1508. doi: 10.1093/rfs/hhm014