

## Revisão dos Modelos de Vetores Autorregressivos Estatísticos – 2012

---

Os modelos de vetores autorregressivos (VAR) utilizados pelo Banco Central estão divididos em dois grupos: modelos VAR com fundamentação econômica, revisados recentemente, e modelos VAR estatísticos<sup>1</sup>. O objetivo deste boxe é apresentar a revisão dos modelos VAR estatísticos, parte do contínuo esforço de aprimoramento das ferramentas utilizadas pelo Banco Central.

Ao longo das últimas décadas houve considerável avanço no desenvolvimento de técnicas de previsão de séries temporais que exploram o uso de grandes bases de dados, em comparação com métodos tradicionais baseados em reduzido número de séries (*e.g.*, equações multivariadas estimadas com mínimos quadrados ordinários, ou modelos VAR tradicionais). De fato, atualmente, encontra-se disponível um grande conjunto de séries temporais (literalmente, milhares de séries econômicas e financeiras) frequentemente utilizadas por agentes de mercado para projetar agregados macroeconômicos<sup>2</sup>.

Não obstante, o uso de amplo conjunto de variáveis traz desafios de ordem prática, sendo o mais conhecido o elevado número de parâmetros estimados. Nesse caso, um aspecto chave é a imposição de alguma restrição na modelagem tal que o erro de estimação permaneça sob controle (e seja assintoticamente desprezível)<sup>3</sup>. Dentre as técnicas disponíveis na literatura para lidar com essa questão, destacam-se a análise de fatores

---

1/ A revisão dos modelos VAR econômicos está descrita no boxe “Revisão dos Modelos de Vetores Autorregressivos com Fundamentação Econômica – 2012” (Relatório de Inflação de setembro de 2012). Os modelos VAR econômicos e estatísticos anteriores estão descritos no boxe “Modelos de Vetores Autorregressivos” (Relatório de Inflação de junho de 2010).

2/ Segundo Stock e Watson (2006, p. 517): “*Academic work on macroeconomic modeling and economic forecasting historically has focused on models with only a handful of variables. In contrast, economists in business and government, whose job is to track the swings of the economy and to make forecasts that inform decision-makers in real time, have long examined a large number of variables.*”

3/ Segundo Stock e Watson (2006, p. 517): “*...the challenge of many-predictor forecasting is to turn dimensionality from a curse into a blessing.*”

(FA – *Factor Analysis*) e a análise de componentes principais (PCA – *Principal Components Analysis*)<sup>4</sup>.

Em relação aos modelos de fatores, Stock e Watson (2006, p. 524) argumentam que sua aplicação a variáveis macroeconômicas (e.g., Stock e Watson, 2002) sugere que um pequeno número de fatores seria capaz de explicar boa parte da variância dessas variáveis. De fato, em sendo possível obter estimações robustas de tais fatores, a previsão utilizando um grande número de variáveis é simplificada pela utilização de alguns poucos fatores.

Por sua vez, na PCA – técnica utilizada nos modelos VAR estatísticos – as componentes principais se caracterizam como “resumos” das principais fontes de variação temporal (e covariância) de dado grupo de variáveis. Nesse sentido, os modelos VAR estatísticos combinam o uso de componentes principais com a tradicional modelagem VAR. Especificamente, utiliza-se ampla base de séries temporais (Tabela 1), caracterizando diversos setores da economia, que em uma primeira etapa são “resumidas” em algumas (poucas) componentes principais. Numa segunda etapa, tais componentes principais são utilizadas em modelos VAR sem restrição estrutural com o intuito de prever a inflação de preços livres do Índice Nacional de Preços ao Consumidor Amplo (IPCA)<sup>5</sup>.

As variáveis utilizadas nos modelos VAR estatísticos são provenientes de seis grupos de indicadores, sendo que os modelos estão divididos em duas classes e são estimados com dados mensais, iniciando-se em janeiro de 2000. Conforme mencionado no box “Modelos de Vetores Autorregressivos” (vide Nota 1), na primeira classe (“Classe 1”) são extraídas componentes principais

---

4/ PCA e FA são abordagens conceitualmente distintas. PCA consiste em transformar matematicamente um conjunto original de variáveis em outro conjunto (de mesma dimensão) de variáveis denominadas “componentes principais”, independentes entre si e estimadas com o objetivo de reter, em ordem de estimação, o máximo de informação em termos da variação total contida nos dados. Cada componente principal é uma combinação linear das variáveis originais, e a primeira componente principal apresenta a maior variação comum dos dados. FA é um nome genérico dado a uma classe de métodos estatísticos multivariados utilizados para descrever a variabilidade entre variáveis observadas correlacionadas, em termos de um número potencialmente menor de variáveis não observadas denominadas “fatores”. As variáveis observadas são modeladas na FA como combinações lineares dos fatores, adicionadas de termos de erro. Vide Johnson e Wichern (1992) para maiores detalhes.

5/ Cabe registrar que a construção dos modelos VAR estatísticos teve como objetivo inicial ser um passo intermediário na construção de modelos do tipo FAVAR (*Factor-Augmented VAR*), originalmente proposto por Bernanke *et al.* (2005). Nestes, os autores modelam conjuntamente, em um VAR estrutural, os fatores e as variáveis macroeconômicas de interesse. Em particular, os autores propõem uma metodologia que combina a análise estrutural dos modelos VAR com técnicas de *Factor Analysis* aplicadas a grandes bases de dados, e que se mostra de grande utilidade na identificação dos mecanismos de transmissão de política monetária.

de variáveis de todos os diferentes grupos de indicadores, totalizando (originalmente) 1536 diferentes modelos. Por exemplo, no grupo de atividade econômica, obtêm-se duas componentes principais de energia (primeira e segunda componentes), extraídas dos três índices de uso de energia elétrica, e duas componentes principais de produção, extraídas da utilização da capacidade instalada, do PIB real e da produção industrial. Na segunda classe (“Classe 2”), os modelos são formados por variáveis selecionadas a partir dos grupos mencionados na Tabela 1 e combinações convexas entre as duas primeiras componentes principais de cada grupo de indicadores utilizado. Nessa classe, são utilizados apenas três grupos de indicadores de cada vez, totalizando (originalmente) 1440 modelos.

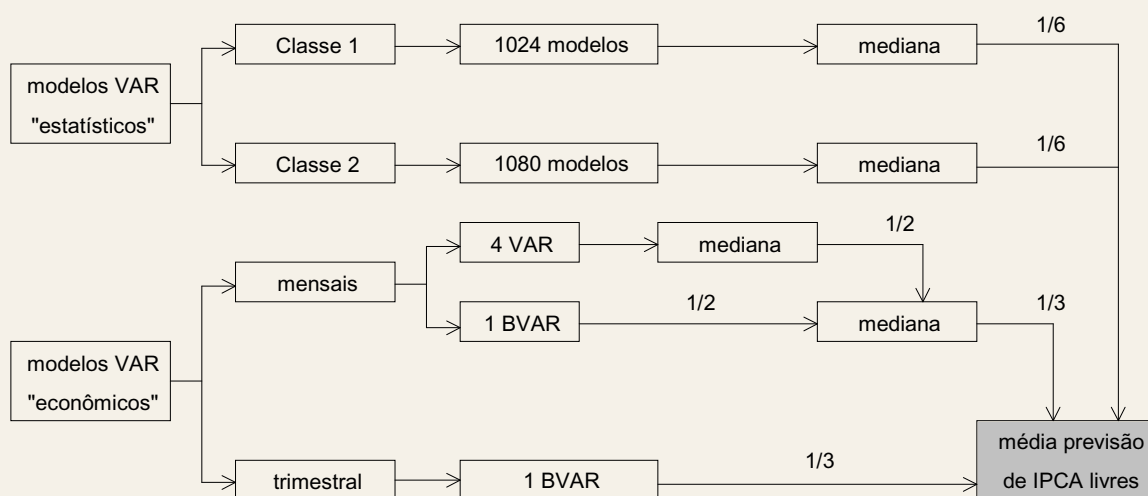
**Tabela 1 – Séries temporais utilizadas nos modelos VAR estatísticos**

Grupo de indicadores	Variáveis selecionadas
Atividade econômica	comércio varejista, três indicadores de energia elétrica, produção industrial mensal, PIB real, utilização da capacidade instalada, desemprego.
Externo	VIX, Embi, câmbio, PPI norte-americano ( <i>all commodities</i> ), índice de preços de exportação, índice de preços de importação, índice de <i>quantum</i> de exportação, índice de <i>quantum</i> de importação.
Financeiro	Selic, Selic real calculada de quatro formas (deflacionada pelo IGP e pelo IPCA, com expectativas de 3 e 12 meses), spreads sobre a Selic, calculados para pessoa física, jurídica, total e para créditos com taxa do BNDES.
Preços	Preços administrados, IGP-DI, IPC-BR, IPC-FIPE, preços livres.
Monetário	M1, M2, M3, M4, papel moeda em poder do público, base monetária e depósitos à vista.
Choques	Índice de <i>commodities</i> CRB, energia elétrica, gasolina, IPA-IPC, óleo lubrificante e petróleo.

A previsão da inflação de preços livres para cada classe é dada pela mediana das previsões da própria classe. Posteriormente, as medianas das duas classes são combinadas com as previsões dos modelos VAR com fundamentação econômica para formar uma previsão agregada da inflação de preços livres do IPCA. A Figura 1 ilustra o processo de agregação geral com os respectivos pesos. Por fim, a previsão agregada da inflação de preços livres, combinada com as previsões da inflação de preços administrados, construídas de forma independente, constituem a trajetória projetada da inflação plena<sup>6</sup>.

6/ Os detalhes sobre a projeção dos preços administrados estão descritos no boxe “Preços Administrados por Contratos e Monitorados: atualização dos modelos de projeção e dos coeficientes de repasse cambial” (Relatório de Inflação de junho de 2012).

Figura 1 – Modelos de vetores autorregressivos



Importante destacar que a revisão dos modelos VAR estatísticos envolveu três aspectos: i) alteração das especificações com baixa capacidade preditiva, visando maior acuidade; ii) tratamento mais adequado de observações extremas (*outliers*) da inflação de preços livres do IPCA, visando garantir previsões de longo prazo dos modelos convergindo para a média histórica (média incondicional); e iii) eliminação de modelos com defasagens excessivas, visando obter modelos mais parcimoniosos e, assim, reduzir o número de parâmetros identificados na estimação.

Com relação ao primeiro aspecto, o grau de acuidade das previsões foi aferido de acordo com o desempenho fora da amostra – pelo cálculo dos erros quadráticos médios (EQMs) e dos erros médios, com horizontes de previsão variando entre um e trinta e seis meses. Nesse sentido, para cada um dos 2976 modelos originais foi construído um exercício pseudo *out-of-sample*, em que a amostra de estimação inicia-se em janeiro de 2000 e termina entre dezembro de 2006 e setembro de 2012. A revisão dos modelos resultou em diminuição do erro médio de previsão e do EQM, em grande parte dos horizontes considerados. O ganho de acuidade das previsões mostrou-se estatisticamente significativo, de acordo com os resultados do teste Diebold-Mariano-West (DMW) (vide Diebold e Mariano (1995) e West (1996)). Para verificar a robustez dos resultados, foram conduzidos testes utilizando a estatística modificada por Harvey *et al.* (1997), os quais apresentaram resultados

compatíveis. Por fim, a revisão também implicou redução do número total de modelos de 2976 para 2104, sendo 1024 na Classe 1 e 1080 na Classe 2, conforme mostra a Figura 1.

O segundo aspecto considerado no processo de revisão foi a previsão de longo prazo dos modelos. Nos modelos VAR, a previsão de uma variável estacionária tende, no longo prazo, para a sua média incondicional (Hamilton, 1994). Dois fatores relevantes contribuíram para aproximar a previsão de longo prazo dos modelos da média histórica: i) exclusão das observações extremas do IPCA de preços livres no período out/2002 a fev/2003; e ii) consideração de dados de inflação que refletem a nova estrutura de padrão de consumo medida pela Pesquisa de Orçamentos Familiares (POF) 2008-2009 do Instituto Brasileiro de Geografia e Estatística (IBGE)<sup>7</sup>. A exclusão das observações extremas e a consideração da nova POF contribuem para que a média de longo prazo dos modelos selecionados fique próxima da média histórica da inflação a partir de 2004, período a partir do qual se observa uma maior estabilidade do IPCA<sup>8</sup>.

Por fim, considerou-se também o critério da parcimônia no processo de seleção dos modelos. Em particular, identificou-se que as especificações com defasagens mais longas (até 6 na Classe 1 e até 12 na Classe 2) estavam entre as que apresentavam desempenho inferior. Possivelmente, o excesso de parâmetros estimados (*overfitting*) estava contribuindo para reduzir a acuidade das previsões. A eliminação de modelos com defasagens muito longas não apenas reduziu o número de parâmetros estimados, mas também contribuiu para aumentar a acuidade da previsão agregada dos modelos.

Em suma, este boxe apresenta a revisão dos modelos VAR estatísticos. Essa iniciativa faz parte do contínuo esforço de aprimoramento das ferramentas utilizadas pelo Banco Central, bem como de dar transparência a suas ações. Cabe notar que o novo conjunto de modelos é composto por um menor número de especificações econométricas, mas que apresentam parâmetros estimados com maior precisão, melhor capacidade preditiva e propriedades

---

7/ Na estimação, a trajetória do IPCA a partir de 2008 reflete as mudanças introduzidas pela nova POF.

8/ A média incondicional dos modelos é a previsão média (entre a mediana da Classe 1 e a mediana da Classe 2) para a taxa de inflação de preços livres do IPCA (anualizada) para dezembro de 2016, ano que é considerado, para fins práticos, como sendo o “longo prazo”.

de longo prazo mais consistentes com o comportamento histórico do IPCA a partir de 2004.

## Referências

BERNANKE, B., BOIVIN, J., ELIASZ, 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.

DIEBOLD, F.X. e MARIANO, R. (1995). “Comparing Predictive Accuracy”, *Journal of Business and Economic Statistics*, 13, 253-265.

HAMILTON, J.D. (1994). “*Times Series Analysis*”, Princeton University Press, New Jersey, USA.

HARVEY, D., LEYBOURNE, S., NEWBOLD, P. (1997). “Testing the equality of prediction mean squared errors”, *International Journal of Forecasting*, 13(2), 281–291.

JOHNSON, R., WICHERN, D. (1992). “*Applied Multivariate Statistical Analysis*”, Prentice-Hall, New Jersey, USA.

STOCK, J.H., WATSON, M.W. (2002). “Forecasting using principal components from a large number of predictors”, *Journal of the American Statistical Association*, 97 (460), 1167-1179.

\_\_\_\_\_. (2006). “Forecasting with Many Predictors”, In: Elliott, G., Granger, C., Timmermann, A. (Eds.), *Handbook of Economic Forecasting*, Volume 1, 515–54. Elsevier, Amsterdam, North-Holland.

WEST, K.D. (1996). “Asymptotic Inference About Predictive Ability”, *Econometrica*, 64, 1067-1084.