Nowcasting Brasil

Pedro G. Costa Ferreira
FGV IBRE

Daiane Marcolino de Mattos
FGV IBRE

Guilherme Branco Gomes
FGV EPGE

Tiago dos Guaranys Martins
FGV EPGE

Resumo

Inserir resumo aqui

Palavras-chave: nowcasting, PIB, dynamic factors model.

1 Introdução

Decisões de política monetária e investimento são tomadas com base nas condições presentes e futuras da economia mesmo quando as variáveis usadas para descrever esse estado não estão acessíveis. falar da defasagem do PIB do Brasil. Por conta disso, a previsão em tempo real, ou simplesmente *nowcasting*, se torna um tema de relevância.

Os métodos de previsão em tempo real, que se desenvolveram nas últimas décadas, são baseados em modelos de fatores dinâmicos (DFM) e dependem de algum poder computacional para lidar com uma grande quantidade de dados. Veja Stock & Watson (2006) para uma revisão a respeito dessa literatura.

O termo nowcasting é cunhado em Giannone et al. (2008) e definido como a previsão do presente, do passado recente ou do futuro próximo. Neste artigo, os autores mostram como reduzir em apenas dois fatores dinâmicos a informação contida em dezenas de séries temporais mensais com o intuito de explicar o PIB (dos Estados Unidos) de curto prazo dos trimestres cuja informação ainda não está disponível. Após esse estudo, muitos outros continuaram explorando o uso de DFM na previsão em tempo real, como por exemplo Bańbura & Rünstler (2011) e Banbura et al. (2011). No primeiro os autores fazem uma análise mostrando como a divulgação de certas variáveis influenciam na atualização da previsão do PIB. Além disso, os autores também propõem uma medida mensal para a variável trimestral, uma vez que os fatores extraídos são mensais. No segundo, os autores propõem estimar o modelo via outro método (Expectation Maximization) e não mais por dois estágios como se fazia em Giannone et al. (2008).

O objetivo do artigo é encontrar um modelo de previsão para o PIB do Brasil segundo as propostas desenvolvidas nos artigos do parágrafo anterior. Para disseminar o uso da metodologia e reproduzir esse

Somos gratos a Domenico Giannone por disponibilizar os códigos em Matlab, assim como comentários relevantes sobre a bibliografia.

trabalho, desenvolveu-se o pacote `nowcasting` no R com tais métodos e algumas outras ferramentas que facilitam o tratamento de séries temporais para tal uso e que permitem analisar a importância de cada variável numa previsão. As funções de estimação foram apenas traduzidas para a linguagem, uma vez que Giannone et al. (2008) e Banbura et al. (2011) forneceram os códigos em Matlab.

A estrutura do artigo é a seguinte: na seção 2 são apresentados o arcabouço teórico sobre modelos de fatores dinâmicos, os métodos de estimação e a base de dados; na seção 3 tem-se os resultados da aplicação metodológica, a análise dos resultados e previsão extendida para o PIB do Brasil no ano de 2018. Por fim, na seção 4, têm-se as considerações finais. Embora todo o contexto apresentado aqui seja referente ao PIB, a metodologia pode ser aplicada a outras séries temporais.

2 Metodologia

2.1 Modelo de Fatores Dinâmicos

Seja $x_t = (x_{1,t}, x_{2,t}, ..., x_{N,t})^{'}$ o vetor que representa as N séries temporais mensais transformadas para satisfazerem a suposição de estacionariedade. A especificação geral do modelo de fator dinâmico (DFM em inglês) é dada por:

$$x_t = \mu + \Lambda f_t + \varepsilon_t \tag{1}$$

$$f_t = \sum_{i=1}^{p} A_i f_{t-i} + B u_t, \quad u_t \sim NID(0, I_q) \tag{2}$$

em que μ é o vetor de médias incondicionais, f_t é o vetor de fatores comuns (não observados) de dimensão $r \times 1$ modelados por um processo VAR(p) em que as matrizes A_i de dimensão $r \times r$ representam os coeficientes do processo autorregressivo e B de dimensão $r \times q$, Λ é uma matriz de constantes de dimensão $N \times r$ e ε_t é um vetor de componentes idiossincráticos, tal que $\Psi = E[\varepsilon_t \varepsilon_t^{'}]$. Assume-se ainda que $E[\varepsilon_t u_{t-k}'] = 0$ para qualquer k.

No chamado *modelo de fatores dinâmicos exato* assume-se que os componentes de erro são mutualmente não correlacionados em todos os lags, isto é, $E[\varepsilon_{i,t}\varepsilon_{j,s}] = 0$ para $i \neq j$. No entanto, é possível que a componente idiossincrática siga um processo AR(p) tal como se mostra na equação (3). Tal procedimento é encontrado com mais detalhes em Banbura et al. (2011).

$$\varepsilon_{i,t} = \sum_{i=1}^{p} \alpha_i \varepsilon_{i,t-i} + e_{i,t}, \quad e_{i,t} \sim NID(0, \sigma_i^2) \tag{3}$$

em que $E[e_{i,t}e_{j,s}] = 0$ para $i \neq j$.

Veja um exemplo da representação matricial da equação 2 do modelo apresentado considerando as ordens $r = 2$, $p = 2$ e $q = 2$.

$$\begin{bmatrix} f_{1,t} \\ f_{2,t} \\ f_{1,t-1} \\ f_{2,t-1} \end{bmatrix} = \begin{bmatrix} a_{1,1}^1 & a_{1,2}^1 & a_{1,1}^2 & a_{1,2}^2 \\ a_{2,1}^1 & a_{2,2}^1 & a_{2,1}^2 & a_{2,2}^2 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} f_{1,t-1} \\ f_{2,t-1} \\ f_{1,t-2} \\ f_{2,t-2} \end{bmatrix} + \begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u_{1,t} \\ u_{2,t} \end{bmatrix} \tag{4}$$

$$F_t = \begin{bmatrix} A_1 & A_2 \\ I_2 & 0 \end{bmatrix} F_{t-1} + Bu_t \tag{5}$$

2.2 Variáveis trimestrais e mensais

Para que a modelagem apresentada suporte frequências mensais e trimestrais, é utilizada a transformação proposta em Mariano & Murasawa (2003). Assim é possível que variáveis trimestrais como PIB sejam explicadas por outras variáveis de frequências mensais (fatores) ao se obter representações trimestrais para a variável mensal.

Seja Y_t^M o nível de uma variável mensal que representa o fluxo. A representação trimestral dessa variável Y_t^Q é dada por:

$$Y_t^Q = Y_t^M + Y_{t-1}^M + Y_{t-2}^M, \quad t = 3, 6, 9, ... \tag{6}$$

Defina $y_t = Y_t^M - Y_{t-1}^M$ e $y_t^Q = Y_t^Q - Y_{t-3}^Q$. Desse modo pode-se escrever y_t^Q em função de y_t de acordo com o seguinte filtro:

$$y_t^Q = y_t + 2y_{t-1} + 3y_{t-2} + 2y_{t-3} + y_{t-4}, \quad t = 3, 6, 9, ... \tag{7}$$

Assim é possível transformar diferenças mensais em diferenças trimestrais. Além disso, se a variável de interesse for uma taxa de variação, tal filtro é uma possível aproximação:

$$log(y_t^Q) \approx log(y_t) + 2log(y_{t-1}) + 3log(y_{t-2}) + 2log(y_{t-3}) + log(y_{t-4}) \tag{8}$$

2.3 Estimação

Serão descritas duas metodologias para estimação de fatores dinâmicos: *Two Stages* e *Expectation & Maximization*. No primeiro método considera-se que as variáveis explicativas em x_t sejam de mesma periodicidade (mensal) e a variável resposta de periodicidade trimestral. No segundo método, x_t e y_t são de periodicidade mensal.

1. *Two Stages*: Essa abordagem é apresentada no trabalho seminal de Giannone et al. (2008). Considerando o modelo de fatores dinâmicos exato, a estimação dos fatores dinâmicos é feita em duas etapas.

   No primeiro estágio, utilizando um painel ($\bar{X}_t$) balanceado e padronizado, são estimados os parâmetros das matrizes Λ e f_t via PCA (Principal Components Analysis). Por balanceado entende-se as variáveis em x_t sem missings e outliers. A padronização é importante pois a estimação via PCA não é invariante a escala. Os estimadores $\hat{\Lambda}$ e $\hat{f}_t$ podem ser obtidos resolvendo o problema de otimização em (9).

$$\min_{f_1,\ldots,f_T,\Lambda} \frac{1}{NT}\sum_{t=1}^{T}(\bar{X}_t - \Lambda f_t)'(\bar{X}_t - \Lambda f_t) \quad s.t. \quad N^{-1}\Lambda'\Lambda = I_r \tag{9}$$

Em seguida, o estimador da matriz de variância e covariância de ε_t é dado por

$$\hat{\Psi} = diag\left(\frac{1}{T}\sum_{t=1}^{T}(\bar{X}_t - \hat{\Lambda}\hat{f}_t)(\bar{X}_t - \hat{\Lambda}\hat{f}_t)'\right) \tag{10}$$

De acordo com Stock & Watson (2011), a solução de (9) é tal que $\hat{\Lambda}$ são os autovetores da matriz de variância e covariância de $\bar{X}_t$ associados aos r maiores autovalores e $\hat{f}_t$ são as r primeiras componentes principais de $\bar{X}_t$. Os coeficientes das matrizes A_i, $i = 1, 2, ..., p$ são estimados por OLS ao regredir f_t em $f_{t-1}, ..., f_{t-p}$. Por fim BB' é estimado como sendo a matriz de covariância dos resíduos dessa regressão.

No segundo estágio, utiliza-se Kalman smoothing para reestimar os fatores para o painel não balanceado x_t considerando os parâmetros obtidos na etapa anterior.

Nesse contexto, duas opções podem ser consideradas ao estimar os fatores:

- *quarterly factors*: as variáveis explicativas mensais podem ser transformadas para representarem quantidades trimestrais seguindo o procedimento visto na seção **??**. Portanto, os fatores embora mensais também representarão quantidades trimestrais e, consequentemente, poderão ser transformados em variáveis de periodicidade trimestral, e assim a equação (11) pode ser estimada para a previsão da variável resposta.
- *monthly factors*: outra opção é estimar os fatores sobre as variáveis originais e ao final aplicar a transformação vista na seção **??** aos fatores para que representem quantidades trimestrais. Em seguida cria-se a variável de periodicidade trimestral que será usada para a previsão da variável resposta em (11).

$$y_t = \beta_0 + \beta'\hat{f}_t + e_t \tag{11}$$

Os parâmetros da equação (11) são estimados por OLS e a previsão de y_{t+h} é obtida como

$$\hat{y}_{t+h} = \hat{\beta}_0 + \hat{\beta}'\hat{f}_{t+h} \tag{12}$$

2. Expectation & Maximization: Banbura et al. (2011) apresentam novidades metodológicas, tanto na especificação do modelo quanto na estimação.

   Primeiramente são impostas restrições com relação aos fatores. Os fatores são divididos em três grupos: global, real e nominal, de modo que a equação (1) pode ser reescrita como:

$$x_t = \mu + \begin{pmatrix} \Lambda_{N,G} & \Lambda_{N,N} & 0 \\ \Lambda_{R,G} & 0 & \Lambda_{R,R} \end{pmatrix} \begin{pmatrix} f_t^G \\ f_t^N \\ f_t^R \end{pmatrix} + \varepsilon_t \tag{13}$$

em que

$$\Lambda = \begin{pmatrix} \Lambda_{N,G} & \Lambda_{N,N} & 0 \\ \Lambda_{R,G} & 0 & \Lambda_{R,R} \end{pmatrix} \tag{14}$$

$$f_t = \begin{pmatrix} f_t^G \\ f_t^N \\ f_t^R \end{pmatrix} \tag{15}$$

O fator global é estimado considerando todas as variáveis explicativas. Já os fatores nominal e real são estimados considerando apenas, respectivamente, as variáveis classificadas em nominais e reais. O parâmetro μ é um vetor de dimensão N de constantes.

Banbura et al. (2011) utilizam uma especificação diferente para o DFM que não é na forma exata, permitindo autocorrelação serial nos componentes idiossincráticos, no qual cada um segue isoladamente um $AR(1)$:

$$\varepsilon_{i,t} = \alpha_i \epsilon_{i,t-1} + e_{i,t}, \quad e_{i,t} \sim i.i.d.N(0, \sigma_i^2) \tag{16}$$

com $E[e_{i,t}e_{j,s}] = 0$ para $i \neq j$.

Uma outra novidade no modelo é que o vetor de séries x_t é composto por variáveis mensais e por uma medida mensal não observada do PIB. O que torna obsoleta a necessidade da equação (11) uma vez que séries de diferentes frequências são incorparadas conjuntamente no processo gerador.

Neste modelo os parâmetros, os fatores comuns e o PIB mensal não são observados, porém todos são estimados conjuntamente utilizando o *Expected Maximization algorithm.* Esse algoritmo utiliza a seguinte estrutura recursiva:

- E-step: A esperança da função de verossimilhança condicional é calculada utilizando as estimativas dos parâmetros estáticos (θ) da iteração anterior, θ_j;
- M-step: Os novos parâmetros, θ_{j+1} são estimados pela maximização da função de verossimilhança do passo anterior, com respeito a θ.

O algoritmo então converge até que os parâmetros sejam iguais nas iterações. O processo recursivo se inicia com as estimativas de PCA apresentadas em Giannone et al. (2008) (primeiro estágio do método *Two Stages*).

2.4 Base de dados

inserir base de dados

3 Nowcasting do PIB brasileiro

apresentar a ordem de cada modelo e quais variáveis
análise pseudo fora da amostra para avaliar a previsão do PIB
avaliar como a divulgação das variáveis impactam na previsão do PIB
previsão do PIB 2018
mostrar link do shiny com a atualização da previsão atual, realtime

4 Considerações finais

concluir o trabalho e apresentar trabalhos futuros

Referências

Banbura, M., Giannone, D., & Reichlin, L. (2011). Nowcasting. *Oxford Handbook on Economic Forecasting*.

Bańbura, M., & Rünstler, G. (2011). A look into the factor model black box: publication lags and the role of hard and soft data in forecasting gdp. *International Journal of Forecasting*, *27*(2), 333–346.

Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, *55*(4), 665–676.

Mariano, R. S., & Murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of applied Econometrics*, *18*(4), 427–443.

Stock, J. H., & Watson, M. (2011). Dynamic factor models. *Oxford Handbook on Economic Forecasting*.

Stock, J. H., & Watson, M. W. (2006). Forecasting with many predictors. *Handbook of economic forecasting*, *1*, 515–554.