Previsão macroeconômica para o Brasil utilizando o modelo VAR com dados mistos

Bruna Kasprzak Borges*
Marcelo Savino Portugal†

Resumo

Este trabalho tem como foco a previsão empírica de curto prazo de variáveis macroeconômicas do Brasil aplicando o modelo *mixed-frequency VAR* (MF-VAR) que utiliza dados observados em frequências mistas - mensais e trimestrais. Procuramos estudar se a inclusão de dados mensais aumentam a precisão da previsão do VAR comparativamente ao modelo *quarterly-frequency VAR* (QF-VAR) estimado em duas versões distintas. Os resultados apontam que a utilização das observações mensais dentro do trimestre aumentam os ganhos das previsões de curto prazo, tanto das séries trimestrais como das séries mensais, especialmente para um e dois trimestres a frente.

Palavras-chaves: Previsão macroeconômica. Vetor autoregressivo. Dados em frequência mista. Métodos bayesianos.

Classificação JEL: E37, C32, C53.

Abstract

The aim of this paper is a short-term empirical forecasting exercise to Brazil using the mixed-fequency VAR (MF-VAR) model to macroeconomic data. The VAR we consider uses data at monthly and quarterly frequencies to obtain forecasts of these variables. We evaluate forecasts from the mixed-frequency VAR and compare them to quarterly-frquency VAR (QF-VAR) with two different versions. We find that the switch from a QF-VAR to a MF-VAR improves the forecasting performance for one and two quarters ahead. Overall, the improvements are most pronounced for short-horizont forecasts.

Keywords: Macroeconomic forecasting. Vector autoregression. Mixed-frequency. Bayesian methods.

*SPGG-RS
†Departamento de Economia, PPGE, Universidade Federal do Rio Grande do Sul.

1 Introdução

Um fato inerente ao estudo de séries temporais econômicas é a diferença na frequência de amostragem dos dados. Variáveis macroeconômicas importantes, como o PIB, somente são encontradas em frequência trimestral, enquanto outras variáveis como inflação e taxa de juros são divulgadas mensalmente ou até mesmo em frequência mais elevada. Em aplicações empíricas, os modelos econométricos são geralmente estimados exclusivamente com base em observações mensais ou trimestrais. Nesse caso, considerando que todas as variáveis no modelo precisam estar amostradas na mesma periodicidade, dados mensais, frequentemente, devem ser convertidos em dados trimestrais (Clements & Galvão, 2008). Contudo, esse processo de agregação temporal descarta uma quantidade relevante de informações e pode gerar subespecificação do modelo. A ideia de agregar séries temporais em diferentes frequências foi originalmente proposta por Ghysels *et al.* (2004) com os modelos de regressão MIDAS (*Mixed Data Sampling*). Inicialmente voltada para séries financeiras, a abordagem MIDAS passou, também, a utilizar variáveis macroeconômicas para a previsão. Novos desdobramentos da literatura voltaram-se então para modelos MF-VAR (*mixed-frequency VAR*), isto é, com dados mistos (ver, por exemplo, Kuzin *et al.*, 2011; Ghysels, 2012; Foroni & Marcellino, 2014; Schorfheide & Song, 2015; Mikosch & Neuwirth, 2015)

Nesse contexto, este trabalho tem como foco a previsão empírica de curto prazo de séries macroeconômicas para o Brasil com o uso de dados em frequência mista, isto é, séries mensais e trimestrais utilizando o modelo MF-VAR descrito em Schorfheide & Song (2015). Procuramos estudar se a inclusão de dados mensais divulgados intratrimestre aumentam a precisão da previsão do VAR. A escolha das variáveis macroeconômicas segue a importância na construção empírica de previsões, sendo variáveis trimestrais como o Produto Interno Bruto (PIB) e mensais como taxa de desemprego, taxa de inflação, produção industrial e taxa de juros selecionadas para a aplicação ao caso brasileiro. A literatura aponta que o uso de dados em frequências mistas pode diminuir o viés da agregação temporal e incorporar novas informações à estimação do modelo. Com o intuito de comparação, o modelo VAR na qual todas as séries são agregadas em frequência trimestral - QF-VAR (*quarterly frequency-VAR*) - será implementado também com foco sobre previsão e estimado em duas versões diferentes.

No caso do Brasil, a revisão bibliográfica realizada neste trabalho indica que apenas Alves & Fasolo (2015) utilizam uma abordagem MF-VAR. Os autores desenvolvem um novo algoritmo e o aplicam para dados simulados e empíricos brasileiros. Os resultados indicam que o MF-VAR desenvolvido pelos autores e o VAR bayesiano apresentam performance satisfatória em termos de *nowcasting*, mas não são melhores comparativamente a um sistema univariado com variáveis trimestrais para os outros horizontes de previsão. Por sua vez, a especificação MIDAS foi utilizada por Zuanazzi & Ziegelmann (2014) para previsão macroeconômica e por Junior & Pereira (2011) e dos Santos & Ziegelmann (2012) para a previsão de volatilidade realizada.

Os modelos VAR com dados mistos (MF-VAR) tem o intuito de capturar a dinâmica conjunta entre séries, ao mesmo tempo em que evitam a agregação temporal entre as variáveis. De forma geral, podemos separar os modelos MF-VAR em duas categorias. A primeira utiliza uma abordagem de espaço de estado assumindo cada série de menor frequência como uma série de alta frequência parcialmente latente. A estimação é dada através de filtro de Kalman (Kuzin *et al.*, 2011; Bai *et al.*, 2013; Foroni & Marcellino, 2014) ou métodos bayesianos (Chiu *et al.*, 2012; Schorfheide & Song, 2015). A segunda categoria utiliza uma abordagem "stacking", onde as variáveis com a mesma frequência são empilhadas em um vetor (Francis *et al.*, 2011; Ghysels, 2012; Foroni & Marcellino, 2014; Mikosch & Neuwirth, 2015). Assim, considera-se que os modelos espaço de estado são *parameter-driven*, enquanto os modelos da segunda abordagem VAR são *observation-driven*.

Modelos espaço de estado envolvem variáveis latentes e, por isso, a necessidade de filtros para lidar com os estados não observáveis, sendo considerados modelos *parameter-driven*. Nessa classe de modelos, o trabalho de Kuzin *et al.* (2011) tem como foco *nowcasting* e previsões dos modelos MIDAS e MF-VAR para o PIB da Zona do Euro. O MF-VAR é estimado com filtro de Kalman permitindo a maximização da verossimilhança associada ao modelo espaço de estado subjacente. Os autores não encontram resultados indicando um dos modelos como superior em termos preditivos. A performance depende do horizonte de previsão e das variáveis incluídas. Bai *et al.* (2013) examinam a relação entre o modelo MIDAS e a abordagem de espaço de estado, comparando-os em termos de previsão. Foroni & Marcellino (2014) comparam a performance de previsão (*nowcasting*) dos *bridge models*, MF-VAR e MIDAS para o PIB trimestral da Zona do Euro utilizando um conjunto amplo de indicadores mensais. O resultado encontrado pelos autores sugere uma melhor performance do modelo MIDAS, especialmente no curto prazo.

Uma primeira abordagem da estimação bayesiana do MF-VAR pode ser encontrada em Chiu *et al.* (2012).[1] Os autores desenvolvem uma alternativa baseada no amostrador de Gibbs para o VAR com dados mistos e em frequência irregular. O foco é a estimação dos parâmetros e funções impulso-resposta. A aplicação empírica confirma o resultado das simulações que indicam ganhos de precisão nas estimativas dos parâmetros comparativamente aos modelos que realizam agregação temporal entre as séries. Na mesma linha de estimação bayesiana do MF-VAR, Schorfheide & Song (2015) tem como objetivo analisar os possíveis ganhos de performance de previsão em *real-time* com a incorporação de dados mensais no VAR. A inferência bayesiana é conduzida através do amostrador de Gibbs conjuntamente a especificação de uma *priori* Minnesota para lidar com a dimensionalidade do VAR. Os autores selecionam um conjunto de 11 variáveis em *real-time* da economia dos Estados Unidos, entre as quais 3 variáveis trimestrais e 8 variáveis mensais. Os resultados indicam que a incorporação de variáveis mensais fornece ganhos de performance expressivos no curto prazo.

Na classe dos modelos *stacking approach* encontram-se aqueles considerados *observation-driven*, isto é, formulados exclusivamente em termos dos dados observáveis e que não envolvem processos latentes. Ghysels (2012) generaliza o modelo MIDAS para o contexto multivariado do VAR. O autor tem como foco as funções impulso resposta baseadas nos dados observáveis (frequências mistas) ao invés de choques gerados por algum processo latente.

Nesse cenário, o principal objetivo deste artigo é comparar as previsões de curto prazo do MF-VAR

[1] A abordagem bayesiana para estimar VAR foi originalmente proposta por Litterman (1980) como solução para o problema de sobreparametrização. Nesse caso, a solução evita que sejam impostas restrições sobre os parâmetros, tal como coeficientes iguais a zero, permitindo sua variação. Assim, a distribuição a *priori* é atualizada por novas informações através da função de verossimilhança, formando a distribuição *a posteriori*.

com o QF-VAR e analisar sua performance de previsão usando um conjunto de variáveis macroeconômicas para o Brasil. A aplicação empírica do MF-VAR segue a especificação de Schorfheide & Song (2015) e os resultados apontam que a utilização das observações mensais dentro do trimestre aumentam a precisão das previsões de curto prazo. Os resultados para PIB, taxa de desemprego, inflação e taxa de juros indicam que o uso de informações mensais leva a uma diminuição considerável da raiz do erro quadrático médio de previsão (RMSE) no curto prazo. No caso das previsões do PIB, a previsão $h = 1$ tem uma redução de 6,87% do RMSE comparativamente ao mesmo horizonte de previsão do QF-VAR. Para as séries mensais da taxa de desemprego, taxa de inflação e taxa de juros, a redução do RMSE gerada pelo MF-VAR para essas séries mensais é mais pronunciada do que para o PIB, a qual é observada em frequência trimestral, em linha com Schorfheide & Song (2015). Para o horizonte $h = 1$, o MF-VAR é capaz de aumentar a precisão das previsões em 32,49% para a taxa de desemprego e 6,80% para a taxa de juros e 1,05% para a taxa de inflação. O uso de informações mensais no modelo de previsão das séries trimestrais mostra-se relevante também para a previsão da série da produção industrial e no caso da previsão $h = 1$, o RMSE do MF-VAR é 13,32% menor em relação ao QF-VAR.

Em resumo, para o grupo das variáveis trimestrais PIB e investimento, os ganhos a partir da inclusão de variáveis mensais é menor comparativamente ao grupo de variáveis mensais não financeiras (taxa de desemprego, taxa de inflação, produção industrial e salário mensal), consistente com os resultados de Schorfheide & Song (2015). Conforme o horizonte de previsão h aumenta, os benefícios de utilizar informações mensais para a previsão, na maioria dos casos, vai diminuindo, mas mantém-se positivo para várias séries. Consistente com a literatura, se o objetivo é gerar previsões para um ou dois trimestres a frente, o modelo MF-VAR torna-se bastante atraente em relação ao QF-VAR. Se o horizonte de previsão for de médio e longo prazo, os resultados indicam que não há incrementos adicionais e o modelo QF-VAR é suficiente para a previsão.

No Brasil, este trabalho é complementar aos estudos de Zuanazzi & Ziegelmann (2014) e Alves & Fasolo (2015) os quais utilizam dados em diferentes frequências em conjunto, respectivamente, às abordagens MIDAS e MF-VAR, com foco sobre previsão macroeconômica. Apesar das diferenças de dados e metodologia, os resultados são similares aos encontrados neste trabalho e sugerem ganhos na previsão utilizando dados em frequências distintas.

O artigo está organizado como segue: na seção 2 é apresentada a metologia, na seção 3 são descritos os dados utilizados bem como a descrição do exercício de previsão, na seção 4 são apresentados os resultados da aplicação empírica e do exercício de previsão e, por fim, a última seção apresenta as considerações finais.

2 Modelo de Vetores Autoregressivos com dados mistos (MF-VAR)

O objetivo desta seção é apresentar o modelo Vetores Autoregressivos com dados mistos (MF-VAR), o qual permite capturar os comovimentos entre séries de frequências mistas. Seguimos o algoritmo de Schorfheide & Song (2015) que utiliza a representação em espaço de estado e métodos bayesianos na estimação do modelo MF-VAR.

Considerando que o MF-VAR aplicado neste trabalho é baseado em um VAR com frequência mensal, as séries trimestrais possuem observações perdidas em dois meses em cada trimestre. Nesse caso, os correspondentes valores mensais são tratados como não observáveis e o modelo é reescrito em espaço de estado.

Em termos gerais, a equação de transição é dada por um VAR em frequência mensal e as equações de medida relacionam as séries observadas às variáveis mensais no vetor de estados.

O modelo assume que a economia evolui em frequência mensal de acordo com o seguinte VAR(p):

$$x_t = \Phi_1 x_{t-1} + ... + \Phi_p x_{t-p} + \Phi_c + u_t, \quad u \sim iidN(0, \Sigma), \tag{1}$$

o vetor $n \times 1$ de variáveis macroeconômicas x_t pode ser decomposto em $x_t = [x'_{m,t}, x'_{q,t}]$, sendo $x_{m,t}$ as variáveis em frequência mensal e $x_{q,t}$ as variáveis mensais não observáveis publicadas em frequência trimestral. Além disso, $z_t = [x'_t, ..., x'_{t-p+1}]'$ e $\Phi = [\Phi_1, ..., \Phi_p, \Phi_c]'$.

Reescrevendo o VAR(p), a equação de transição é descrita por:

$$z_t = F_1(\Phi) z_{t-1} + F_c(\Phi) + v_t, \quad v_{t_m} \sim iidN(0, \Omega(\Sigma)). \tag{2}$$

Denotando $y_{m,t} = x_{m,t}$, na equação de medida quando a variável trimestral está disponível, então, esse dado dado observado é uma média do dado mensal não observável:

$$\tilde{y}_{q,t} = \frac{1}{3}(x_{q,t} + x_{q,t-1} + x_{q,t-2}) = \Lambda_{qz} z_t. \tag{3}$$

Considerando que $y_{q,t}$ é observada somente a cada 3 meses é necessário definir $M_{q,t}$ como a matriz identidade se t corresponde ao último mês do trimestre e zero caso contrário. Essa média de 3 meses da série mensal só é observada em cada terceiro mês. Quando observações de $y_{q,t}$ não estão disponíveis, a linha correspondente de $M_{q,t}$ é zero. Então, definimos:

$$y_{q,t} = M_{q,t} \tilde{y}_{q,t} \quad t = 1, ..., T_b. \tag{4}$$

Seja $y_{m,t}$ o subconjunto de variáveis mensais para as quais no período t as observações são divulgadas após T e $M_{m,t}$ a matriz de seleção:

$$y_{m,t} = M_{m,t} x_{m,t} \quad t = T_b + 1, ..., T. \tag{5}$$

Portanto, a equação de medida pode ser escrita como:

$$y_t = M_t \Lambda_z z_t \quad t = 1, ..., T. \tag{6}$$

Então, M_t é uma sequência de matrizes que selecionam o tempo t quando as variáveis são observadas . Isso influencia a equação de medida que vai mudando ao longo do tempo dependendo se as variáveis trimestrais estão disponíveis ou não. Nota-se que a defasagem do modelo deve ser maior ou igual a 3. Esse é o mínimo *lag* permitido pela estrutura da equação de medida.

2.1 Estimação Bayesiana

A estimação bayesiana do MF-VAR tem início com a definição da distribuição conjunta das variáveis observáveis $Y_{1:T}$, estados latentes $Z_{0:T}$ e parâmetros (Φ, Σ) condicionado a uma pré-amostra $Y_{-p+1:0}$ para inicializar os lags. Neste caso, o amostrador de Gibbs é utilizado para a obtenção dos vetores de estado desconhecidos e implementado através do algoritmo de Carter & Kohn (1994). O amostrador de Gibbs permite gerar amostras a partir da distribuição a posteriori de $(\Phi, \Sigma)|(Z_{0:T}, Y_{-p+1:T})$ e $Z_{0:T}|(\Phi, \Sigma, Y_{-p+1:T})$. A partir dessas amostras é possível simular trajetórias futuras de y_t para caracterizar a distribuição preditiva do modelo.

2.1.1 Distribuição a *priori*

A *priori* é essencial na estimação e previsão em modelos bayesianos e neste caso é utilizada a *priori* Minnesota.[2] Como ressalta Schorfheide & Song (2015), a ideia central da *priori* Minnesota é centrar a distribuição de (Φ) de modo que implique em um passeio aleatório para cada um dos componentes de x_t em (1). A variância da distribuição a *priori* é controlada pelo vetor de hiperparâmetros (λ). É utilizada uma versão atualizada da original proposta por Doan *et al.* (1984). Schorfheide & Song (2015), baseados no trabalho de Sims & Zha (1998), aplicam a distribuição da *priori* na forma da distribuição Normal multivariada Wishart-Inversa (*MNIW*). Essa é uma *priori* conjugada de modo que a *priori MNIW* resulta em uma posteriori que também é *MNIW*.

A *priori* é implementada com o método de estimação proposto por Litterman (1986), de forma que os dados observados são aumentados (*data augmentation*) com observações *dummy*. O uso de *data augmentation* (adição de variáveis auxiliares ao amostrador) é realizado com o intuito de facilitar a implementação do amostrador de forma que a análise é, então, realizada em um espaço expandido incluindo não somente os parâmetros do modelo, mas, também, variáveis latentes e obervações perdidas.

A *priori* Minnesota é gerada por observações *dummy* (x_*, w_*) que são indexadas por 5×1 vetor de hiperparâmetros λ com elementos λ_i. Usando uma pré-amostra, seja $\underline{x}$ e $\underline{s}$ vetores $n \times 1$ de médias e desvio padrão. Para as séries mensais esse cálculo é direta. Para as séries trimestrais, equiparamos $\underline{x_q}$ e $\underline{s}$ com, respectivamente, a média e o desvio-padrão da pré-amostra dos valores trimestrais observados.

[2] Para Litterman (1986), a escolha da *priori* deve considerar três regularidades estatísticas de séries macroeconômicas: 1) o comportamento da tendência das séries temporais macroeconômicas; 2) valores correntes possuem mais informação que valores passados; 3)valores passados de uma dada variável contém mais informação de seu estado corrente comparativamente aos valores passados de outras variáveis.

2.1.2 Inferência *a posteriori*

A distribuição conjunta dos dados, variáveis latentes e parâmetros condicionada em algumas observações para inicializar os lags é dada por:

$$p(Y_{1:T}, Z_{0:T}, \Phi, \Sigma | Y_{-p+1:0}, \lambda) = p(Y_{1:T}|Z_{0:T}) p(Z_{1:T}|z_0, \Phi, \Sigma) p(z_0|Y_{-p+1:0}) p(\Phi, \Sigma|\lambda) \tag{7}$$

A distribuição de $Y_{1:T}|Z_{1:T}$ é dada pela probabilidade de $Y_{1:T}$ que satisfaz (6). A densidade $p(Z_{1:T}|z_0, \Phi, \Sigma)$ é obtida a partir da regressão linear em (2). A densidade condicional $p(z_0|Y_{-p+1:0})$ é definida como gaussiana. A inicialização de z_{T_-} é dada utilizando as observações atuais, sendo T_- a pré-amostra. No caso das variáveis mensais isso é direto e nas variáveis trimestrais é dado como igual ao valores trimestrais observados, assumindo que durante esses períodos os valores intra-trimestre são dados pela médias observadas durante o trimestre. Então, $p(z_{T_-})$ é uma função de probabilidade. Φ e Σ são definidos como iguais a suas respectivas médias a *priori* e é aplicado o filtro de Kalman para a pré-amostra $Y_{T_-:0}$ para o sistema descrito em (2) e (6). Por sua vez, $p(\Phi, \Sigma|\lambda)$ representa a densidade a *priori* dos parâmetros do VAR. Então, a densidade a posteriori condicional dos parâmetros VAR e estados latentes do MF-VAR é descrita por:

$$\begin{aligned} p(\Phi, \Sigma | z_{0:T}, Y_{-p+1:T}) &\propto p(Z_{1:T}|z_0, \Phi, \Sigma) p(\Phi, \Sigma|\lambda) \\ p(Z_{0:T}|\Phi, \Sigma, Y_{-p+1:T}) &\propto p(Y_{1:T}, Z_{1:T}) p(Z_{1:T}|z_0, \Phi, \Sigma) p(z_0|Y_{-p+1}) \end{aligned} \tag{8}$$

Neste caso, o amostrador de Gibbs que itera as duas distribuições a posteriori condicionais em (8)

$$p(\Phi, \Sigma|z_{0:T}, Y_{-p+1:T}) \quad \text{e} \quad p(Z_{0:T}|\Phi, \Sigma, Y_{-p+1:T})$$

é implementado através do algoritmo de Carter & Kohn (1994).[3] Condicionada a $Z_{0:T}$, a regressão multivariada (2) é linear e gaussiana com *priori* conjugada. Como a *priori* pertence à família MNIW, então o MCMC (*Markov chain Monte Carlo*) pode ser usado diretamente para obter a posteriori. Então, o MF-VAR é um modelo espaço de estado linear gaussiano e a amostragem da posteriori condicional de $p(Z_{0:T}|\Phi, \Sigma, Y_{-p+1:T})$, isto é, a sequência $Z_{0:T}$ condicional aos parâmetros do VAR, é implementada com um suavizador padrão. A inicialização para o passo do filtro de Kalman do suavizador é obtida da distribuição $p(z_0|Y_{-p+1})$.

2.2 Função de verossimilhança marginal e seleção de hiperparâmetros

Na estimação bayesiana, um dos aspecto importante refere-se à seleção dos hiperparâmetros (λ), os quais controlam as regularidades estatísticas de séries macroeconômicas. No MF-VAR são escolhidos os hiperparâmetros que conjuntamente maximizam a densidade marginal dos dados[4] (MDD). De acordo com Schorfheide & Song (2015), a melhor performance para o MF-VAR é atingida para valores de λ entre os

[3] O amostrador de Gibbs segue o algoritmo eficiente de Carter and Kohn (1994) para a amostragem MCMC, onde todos os estados são atualizados conjuntamente em um passo. Esse algoritmo é mais eficiente do que os amostradores com um único estado, os quais realizam a atualização de um vetor de estados de cada vez.

[4] Também chamada de verossimilhança marginal.

extremos $\lambda = 0$ e $\lambda \to \infty$. A resolução da MDD envolve integrar os estados latentes. A probabilidade *a posteriori* de λ é proporcional a MDD:

$$p(Y_{1:T}|Y_{-p+1:0}, \lambda) = \int p(y_{1:T}, Z_{0:T}, \Phi, \Sigma|Y_{-p+1:0}, \lambda) d(Z_{0:T}\Phi, \Sigma). \tag{9}$$

Para gerar as previsões do MF-VAR, condicionamos o valor de λ_T que maximiza o logaritmo da MDD:

$$\ln p(Y_{1:T}|Y_{-p+1:0}, \lambda) = \sum_{t=1}^{T} \ln \int p(y_t|Y_{-p+1:t-1}, \Phi, \Sigma) p(\Phi, \Sigma|Y_{-p+1:t-1}) d(\Phi, \Sigma). \tag{10}$$

O logaritmo da densidade marginal dos dados pode ser interpretado como a soma da pontuação preditiva um passo a frente. A escolha de λ baseada na MDD tende a funcionar bem em inferência, assim como, em previsão, de acordo com Schorfheide & Song (2015). No caso da geração das previsões do MF-VAR, cada horizonte de previsão é condicionado sobre o valor $\hat{\lambda_T}$ que maximiza o logaritmo da densidade marginal dos dados.

2.2.1 Aproximação da densidade marginal dos dados

A resolução da MDD a partir de (9) envolve resolver a integral envolvendo estados latentes. Uma forma de estimar a MDD é utilizando o estimador média harmônica de Geweke (1999).

$$\frac{1}{p(Y_{1:T}|Y_{-p+1:0}, \lambda)} = \frac{p(W_{1:T}, w_0|Y_{1:T}, Y_{-p+1:0})}{p(W_{1:T}, Y_{1:T}, w_0|Y_{-p+1:0})} \tag{11}$$

onde $W_{1:T}$ engloba os valores não observáveis de $x_{q,t}$ para o primeiro e o segundo mês de cada trimestre e w_0 contém os valores correspondentes para o período de inicialização $t = -p + 1, ..., 0$. Então, a aproximação da MDD é dada por:

$$\hat{p}(Y_{1:T}|Y_{-p+1:0}, \lambda) = c \left[\frac{1}{N} \sum_{i=1}^{N} \frac{f_0(w_0^{(i)}) f(W_{1:T}^{(i)})}{p(Z_{1:T}^{(i)}|z_0^{(i)}, \lambda) p(z_0^{(i)}|Y_{-p+1:0}, \lambda)} \right]^{-1} \tag{12}$$

A constante c captura o termo Jacobiano associado com a mudança de variável de $(w_0, W_{1:T}, Y_{1:T})$ para $(z_0, Z_{1:T})$. A função f segue Geweke (1999) sendo uma distribuição normal multivariada TRIMMED com média $\hat{\mu}_{W_{1:T}} = \frac{1}{N}\sum_{i=1}^{N} W_{1:T}^{(i)}$ e variância $\hat{\sum}_{W_{1:T}} = \frac{1}{N}\sum_{i=1}^{N} W_{1:T}^{(i)} W_{1:T}^{(i)\prime} - \hat{\mu}_{W_{1:T}} \hat{\mu}'_{W_{1:T}}$. É definido $f_0(w_0^{(i)}) = p(z_0^{(i)}|Y_{-p+1:0}, \lambda)$ de forma que os termos se cancelem. A avaliação do denominador é obtida usando as constantes de normalização para a distribuição MNIW (ver Del Negro & Schorfheide (2011)).

3 Aplicação empírica

Esta seção apresenta a aplicação empírica dos modelos descritos na seção 2 para o caso brasileiro. São apresentadas as variáveis utilizadas, bem como a escolha dos hiperparâmetros utilizados na construção das previsões do MF-VAR e do VAR baseado na agregação temporal em dados trimestrais (QF-VAR). A especificação QF-VAR é estimada de duas maneiras: i) QF-VAR(*priori* informativa), ou seja, estimado de forma semelhante ao MF-VAR e ii) VAR(OLS) estimado por Mínimos Quadrados Ordinários (OLS).[5]

3.1 Variáveis macroeconômicas

São consideradas 9 variáveis, sendo 3 variáveis trimestrais e 6 variáveis mensais. As variáveis trimestrais são o PIB real (PIB), o consumo real do governo (GOV) e a formação bruta de capital fixo real (INV). As variáveis mensais são taxa de desemprego na Região Metropolitana de São Paulo (UNR), taxa de juros SELIC (JUROS), taxa de inflação IPCA (INF), Pesquisa Industrial Mensal Brasil[6](IND), IPCA núcleo médias aparadas com suavização (INFCORE) e o índice do salário real médio na indústria em São Paulo (AWI). As variáveis que entram no VAR em logaritmo são: PIB, GOV, INV, IND e AWI e as demais variáveis são divididas por 100.

Os dados abrangem o período de jan./96 a jun./17. A previsão é realizada com uma abordagem recursiva a qual considera uma sequência crescente de dados $Y_{-p+1:T}$, $T = T_{min}, ..., T_{max}$, para reestimar os parâmetros dos modelos (MF-VAR e QF-VAR) e gerar as previsões para os períodos $T+1, ..., T+H$. O horizonte de previsão máximo é $H = 12$ meses. O período $t = 1$ corresponde a jul./96, T_{min} a jun./12 e T_{max} a jun./16, resultando em 49 amostras de estimação.

Observações mensais para as 3 séries em frequência trimestral são obtidas aplicando o amostrados de Gibbs para período amostral inteiro (jan./96 a jun./17).[7] Como as previsões são avaliadas em médias trimestrais[8], as trajetórias mensais simuladas são agregadas da mesma forma. Considerando que as variáveis trimestrais somente são observadas nessa frequência, avaliamos as previsões das médias trimestrais apesar do modelo ser resolvido na frequência mensal subjacente. Então, $h = 1, ..., H$ é contado em trimestres ao invés de meses.

As previsões são avaliadas com base na raiz do erro quadrático médio (RMSE). A comparação entre o RMSE para as previsões das médias trimestrais do MF-VAR e para as previsões do QF-VAR é realizada através do RMSE relativo:

$$\text{RMSE relativo}(i|h) = 100 \times \frac{RMSE(i|h)}{RMSE_{benchmark}(i|h)} \tag{13}$$

onde *i* denota a variável e *h* refere-se ao horizonte de previsão medido em trimestres. O modelo QF-VAR(*priori* informativa) e o QF-VAR(OLS) são utilizados como *benchmarks*.

Baseado em análises prévias, definimos o número de lags no equação de transição (mensal) do MF-VAR igual a $p_(m) = 6$ e o número de lags no QF-VAR (*priori* informativa) igual a $p_(q) = 2$ e QF-VAR (OLS) igual a $p_(q) = 1$.

Para cada amostra de estimação foram geradas 20.000 amostras a partir da distribuição a posterior dos

[5]Como destacado por Giannone *et al.* (2015), no modelo VAR com *priori flat*, a distribuição a posteriori dos parâmetros é centrada na estimativa OLS.

[6]Os dados anteriores a Dez./01 foram retropolados a partir da série antiga da Pesquisa Industrial Mensal (PIM) para construir uma série com início em 1996.

[7]As 3 séries mensais inferidas são tratadas como dados na avaliação da previsão pontual.

[8]Nesse caso, médias trimestrais referem-se a média para cada 3 meses.

parâmetros do VAR usando o algoritmo MCMC descrito na seção (2.1). São descartadas as primeiras 4.000 amostras e utilizadas as restantes para calcular as aproximações dos momentos a posteriori.

3.2 Seleção dos hiperparâmetros

Como explicado na seção (2.2), os hiperparâmetros são determinados a partir da maximização da log-verossimilhança marginal. Como salientado por Schorfheide & Song (2015), a otimização da verossimilhança marginal ocorre em relação a 5 hiperparâmetros que controlam: o ajuste geral da *priori* (λ_1); a taxa de decaimento da variância conforme aumenta a defasagem dos coeficientes, implicando que quanto mais longa for a defasagem mais a variância se aproximará de zero (λ_2); número de observações utilizadas para obter a *priori* para a matriz de covariância dos erros (λ_3); a soma dos coeficientes dos lags de uma variável em direção a unidade (λ_4); e o parâmetro de ajuste para a co-persistência entre os coeficientes do VAR (λ_5).

A reotimização dos hiperparâmetros do MF-VAR é computacionalmente onerosa. Como espera-se que a escolha ótima dos hiperparâmetros evolua suavemente ao longo do tempo, a reotimização é realizada aproximadamente a cada 2 anos, em linha com Schorfheide & Song (2015). Os resultados da reotimização indicam alguma mudança no valor dos hiperparâmetros. Com isso, optamos por manter os mesmos valores sucessivamente em cada amostra recursiva, realizando a mudança a partir da amostra em que é realizada a reotimização, nesse caso, na amostra recursiva 1, 25 e 49. A tabela (1) mostra a estimativa dos hiperparâmetros para a última reotimização.[9]

Tabela 1: Hiperparâmetros

	λ_1	λ_2	λ_3	λ_4	λ_5
MF-VAR	0,03	3,4	1,00	3,40	3,24
QF-VAR	0,26	3,59	1,00	2,63	1,62

4 Resultados empíricos

Nesta seção apresentamos os resultados do exercício empírico de previsão do MF-VAR para o caso brasileiro.

A tabela (2) apresenta a o RMSE relativo entre o modelo MF-VAR e o QF-VAR(*priori* informativa) como descrito na equação (13). De forma geral, a performance de previsão fora da amostra aponta que a utilização das observações mensais dentro do trimestre aumentam a precisão das previsões de curto prazo.

Os resultados para PIB, taxa de desemprego, inflação e taxa de juros indicam que o uso de informações mensais leva a uma diminuição considerável do RMSE no curto prazo (Figura **??**). No caso das previsões do PIB, a previsão um passo a frente tem uma redução de 6,9% do RMSE comparativamente à previsão do QF-VAR(*priori* informativa). A redução do RMSE permanece até o horizonte $h = 3$. Com o aumento

[9]Os valores dos hiperpâmetros do MF-VAR encontram-se no apêndice A. No caso do QF-VAR (*priori* informativa), os resultados indicaram a manutenção dos mesmos valores para cada amostra recursiva.

do horizonte de previsão para $h = 4$, o QF-VAR(*priori* informativa) alcança o MF-VAR e o diferencial do RMSE entre as previsões dos modelos torna-se negligenciável.

Para as séries mensais da taxa de desemprego, taxa de inflação e taxa de juros, a redução do RMSE gerada pelo MF-VAR para essas séries mensais é mais pronunciada do que para o PIB, a qual é observada em frequência trimestral, em linha com Schorfheide & Song (2015). Para as variáveis divulgadas em frequência mensal, as diferenças entre as previsões do MF-VAR e do QF-VAR(*priori* informativa) tende a ser mais pronunciada do que para o PIB. Para o horizonte $h = 1$, o MF-VAR é capaz de aumentar a precisão das previsões em 32,4% para a taxa de desemprego e 6,8% para a taxa de juros e 1,1% para a taxa de inflação. Para a taxa de desemprego e a inflação os resultados os ganhos tendem a diminuir com o horizonte h de previsão, mas ainda mostram-se substanciais mesmo para $h = 4$. No caso da taxa de juros, o ganho com o uso de informações mensais tende a aumentar conforme aumenta o horizonte h de previsão.

Tabela 2: RMSE relativo do MF-VAR *versus* QF-VAR (*priori* informativa)

h	UNR	JUROS	INF	IND	INFcore	AWI	PIB	INV	GOV
1	-32,5	-6,8	-1,1	-13,3	12,0	-34,4	-6,9	-4,3	20,4
2	-12,5	-15,5	-10,8	-5,6	14,1	-15,2	-3,8	-6,6	7,0
3	-7,1	-22,0	-8,6	-2,3	1,0	-9,0	-1,0	-6,2	4,0
4	-4,1	-23,9	-7,9	-1,6	3,6	-5,8	0,3	-3,1	3,1

Nota: O RMSE para UNR, JUROS, INF, INFCORE refere-se à previsão do nível das séries. Para as variáveis restantes, o RMSE refere-se ao logaritmo da série.

O uso de informações mensais no modelo de previsão das séries trimestrais mostra-se relevante também para a previsão da série da produção industrial. No horizonte $h = 1$, o RMSE da previsão do MF-VAR é 13,3% menor em relação ao QF-VAR(*priori* informativa). Para os horizontes $h = 2, 3$ e $h = 4$ o incremento decresce, mas o diferencial mantém-se positivo. Considerando que a produção industrial é comumente utilizada como um dos indicadores antecedentes do PIB, o resultado satisfatório da previsão dessa variável pode colaborar para a definição mais precisa de estimativas para a prévia do PIB.

No caso da previsão do salário mensal, os horizontes $h = 1$ e $h = 2$ indicam ganhos expressivos na redução do RMSE do MF-VAR frente ao QF-VAR(*priori* informativa). Conforme h aumenta, o RMSE diminui, mas ainda assim mostra-se relevante mesmo com o aumento do horizonte de previsão. Para as variáveis INFCORE e GOV, o modelo QF-VAR(*priori* informativa) ultrapassa o MF-VAR e os diferenciais tornam-se favoráveis ao modelo que emprega as séries agregadas na frequência trimestral.

Em resumo, para os grupo das variáveis trimestrais PIB e INV, os ganhos a partir da inclusão de variáveis mensais é menor comparativamente ao grupo de variáveis mensais não financeiras (UNR, INF, IND e AWI), consistente com os resultados de Schorfheide & Song (2015). Conforme o horizonte de previsão h aumenta, os benefícios em utilizar informações mensais para a previsão, na maioria dos casos, vai diminuindo, mas mantém-se positivo para várias séries.

A tabela (3) apresenta a o RMSE relativo entre o modelo MF-VAR e o QF-VAR(OLS). De forma geral, a

Tabela 3: RMSE relativo do MF-VAR *versus* QF-VAR (OLS)

h	UNR	JUROS	INF	IND	INFcore	AWI	PIB	INV	GOV
1	-59,2	-73,5	-18,3	-68,0	-33,1	-62,3	-49,7	-53,3	-27,7
2	-4,6	-60,4	0,0	-42,0	-19,2	-30,3	-5,5	-21,6	-3,2
3	55,0	-52,2	5,4	-11,9	-17,0	3,7	44,3	21,1	36,6
4	117,0	-42,6	12,9	16,8	-8,4	40,0	92,2	64,5	79,0

Nota: O RMSE para UNR, JUROS, INF, INFcore refere-se à previsão do nível das séries. Para as variáveis restantes, o RMSE refere-se ao logaritmo da série.

performance de previsão fora da amostra aponta que a utilização das observações mensais dentro do trimestre aumentam a precisão das previsões para $h = 1$ e $h = 2$ passos a frente para todas as variáveis.

Os resultados para PIB, taxa de desemprego, inflação e taxa de juros indicam a diminuição do RMSE para $h = 1$ e $h = 2$. No caso do PIB, a previsão um passo a frente tem uma redução considerável de 49,7% do RMSE comparativamente à previsão do QF-VAR(OLS).

Nesse caso, para os grupo das variáveis trimestrais os ganhos a partir da inclusão de variáveis mensais foram similares ao grupo de variáveis mensais não financeiras (UNR, INF, IND e AWI), diferentemente do caso QF-VAR(*priori* informativa). Na maioria das situações, para o horizonte de previsão superior a $h = 3$, os benefícios em utilizar informações mensais para a previsão desaparece.

Em ambos os casos, se o objetivo é gerar previsões para $h = 1$ e $h = 2$ trimestres a frente, o MF-VAR indica ganhos de performance mais expressivos tanto em relação ao QF-VAR(*priori* informativa) como, também, ao QF-VAR(OLS). Para horizontes superiores a $h = 3$, o incremento diminuiu ou desaparece em relação a ambas as versões do QF-VAR.

5 Considerações finais

Este trabalho teve como foco a previsão empírica de curto prazo de séries macroeconômicas do Brasil utilizando o modelo *mixed-frequency VAR* (MF-VAR), o qual utiliza dados que são observados em frequências mistas - mensais e trimestrais. A especificação do MF-VAR (Schorfheide & Song (2015)) utiliza uma abordagem bayesiana com o intuito de lidar com a dimensionalidade do modelo, bem como, os hiperparâmetros são determinados pelos dados (*data-driven*) através da maximização do logaritmo da densidade marginal dos dados.

Os resultados indicaram ganhos na precisão das previsões do MF-VAR comparativamente ao modelo QF-VAR (*quarterly-frequency VAR*). No caso da versão do QF-VAR isso indica que a utilização de dados mensais melhora a previsão de curto prazo, especialmente a previsão para $h = 1$ e $h = 2$ passos a frente. Com o horizonte de previsão aproximando-se de um ano, o incremento entre as previsões do MF-VAR e do QF-VAR(*priori* informativa) vão diminuindo. Em relação ao QF-VAR(OLS), os ganhos do MF-VAR ocorrerem até $h = 2$ trimestres a frente, pois após esse período os ganhos desaparecem.

De forma geral, os resultados indicaram que a especificação do VAR com observações em diferentes frequências - mensais e trimestrais - é interessante para a previsão das principais variáveis macroeconômicas brasileiras. O resultado principal indicou que incluir informação mensal melhora a performance de previsão de curto prazo, especialmente com o foco em $h = 1$ e $h = 2$ passos a frente.

Pesquisas futuras podem avançar na análise da previsão entre modelos alternativos que utilizam dados com frequência mista. Além disso, podem ser testados conjuntos diferentes de variáveis de acordo o objetivo da previsão.

Referências

ALVES, SERGIO, & FASOLO, ANGELO. 2015. *Not Just Another Mixed Frequency Paper*. Tech. rept. Central Bank of Brazil, Research Department.

BAI, JENNIE, GHYSELS, ERIC, & WRIGHT, JONATHAN H. 2013. State space models and MIDAS regressions. *Econometric Reviews*, **32**(7), 779–813.

CARTER, CHRIS K, & KOHN, ROBERT. 1994. On Gibbs sampling for state space models. *Biometrika*, **81**(3), 541–553.

CHIU, CHING WAI JEREMY, ERAKER, BJØRN, FOERSTER, ANDREW, KIM, TAE BONG, & SEOANE, HERNÁN D. 2012. *Estimating VAR's sampled at mixed or irregular spaced frequencies: a Bayesian approach*. Tech. rept. Federal Reserve Bank of Kansas City.

CLEMENTS, MICHAEL P, & GALVÃO, ANA BEATRIZ. 2008. Macroeconomic forecasting with mixed-frequency data: Forecasting output growth in the United States. *Journal of Business & Economic Statistics*, **26**(4), 546–554.

DEL NEGRO, MARCO, & SCHORFHEIDE, FRANK. 2011. Bayesian macroeconometrics. *The Oxford handbook of Bayesian econometrics*, **293**, 389.

DOAN, THOMAS, LITTERMAN, ROBERT, & SIMS, CHRISTOPHER. 1984. Forecasting and conditional projection using realistic prior distributions. *Econometric reviews*, **3**(1), 1–100.

DOS SANTOS, DOUGLAS GOMES, & ZIEGELMANN, FLÁVIO AUGUSTO. 2012. Multi-Period Volatility Predictions: A Comparative Study Using MIDAS Regressions. *In: 34º Meeting of the Brazilian Econometric Society*.

FORONI, CLAUDIA, & MARCELLINO, MASSIMILIANO. 2014. A comparison of mixed frequency approaches for nowcasting Euro area macroeconomic aggregates. *International Journal of Forecasting*, **30**(3), 554–568.

FRANCIS, NEVILLE, GHYSELS, ERIC, & OWYANG, MICHAEL T. 2011. *The Low-Frequency Impact of Daily Monetary Policy Shocks*. Tech. rept. 2011-009. Federal Reserve Bank of St. Louis Working Paper Series.

GEWEKE, JOHN. 1999. Using simulation methods for Bayesian econometric models: inference, development, and communication. *Econometric Reviews*, **18**(1), 1–73.

GHYSELS, ERIC. 2012. *Macroeconomics and the Reality of Mixed Frequency Data*. Tech. rept. UNC.

GHYSELS, ERIC, SANTA-CLARA, PEDRO, & VALKANOV, ROSSEN. 2004. *The MIDAS Touch: Mixed Data Sampling Regression Models*. Tech. rept. UNC and UCLA.

GIANNONE, DOMENICO, LENZA, MICHELE, & PRIMICERI, GIORGIO E. 2015. Prior selection for vector autoregressions. *Review of Economics and Statistics*, **97**(2), 436–451.

JUNIOR, MARCOS VINÍCIO WINK, & PEREIRA, PEDRO LUIZ VALLS. 2011. Modeling and Forecasting of Realized Volatility: Evidence from Brazil. *Brazilian Review of Econometrics*, **31**(2), 315–337.

KUZIN, VLADIMIR, MARCELLINO, MASSIMILIANO, & SCHUMACHER, CHRISTIAN. 2011. MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the euro area. *International Journal of Forecasting*, **27**(2), 529–542.

LITTERMAN, ROBERT B. 1980. *Techniques for forecasting with Vector Autoregressions*. PhD Thesis, University of Minnesota, Minnesota.

LITTERMAN, ROBERT B. 1986. Forecasting With Bayesian Vector Autoregressions. *Journal of Business & Economic Statistics*, **4**(1), 25–38.

MIKOSCH, HEINER, & NEUWIRTH, STEFAN. 2015. *Real-time forecasting with a MIDAS VAR*. Tech. rept. BOFIT Discussion Papers.

SCHORFHEIDE, FRANK, & SONG, DONGHO. 2015. Real-time forecasting with a mixed-frequency VAR. *Journal of Business & Economic Statistics*, **33**(3), 366–380.

SIMS, CHRISTOPHER A, & ZHA, TAO. 1998. Bayesian methods for dynamic multivariate models. *International Economic Review*, **39**(4), 949–68.

ZUANAZZI, PEDRO TONON, & ZIEGELMANN, FLÁVIO AUGUSTO. 2014. Previsões para o crescimento do PIB trimestral brasileiro com séries financeiras e econômicas mensais: uma aplicação de MIDAS. *Economia Aplicada*, **18**(2), 295–318.

Apêndice A

Hiperparâmetros

Tabela 4: Hiperparâmetros

	λ_1	λ_2	λ_3	λ_4	λ_5
MF-VAR	0,05	3,40	1,00	3,40	2,19
QF-VAR	0,26	3,59	1,00	2,63	1,62

Nota: Refere-se aos hiperparâmetros obtidos a partir da amostra recursiva 1.

Tabela 5: Hiperparâmetros

	λ_1	λ_2	λ_3	λ_4	λ_5
MF-VAR	0,04	3,4	1,00	3,40	3,24
QF-VAR	0,26	3,59	1,00	2,63	1,62

Nota: Refere-se aos hiperparâmetros obtidos a partir da amostra recursiva 25.