# Testing for Granger Causality in Large Mixed-Frequency VARs*

Thomas Götz† and Alain Hecq‡

Maastricht University, SBE, Department of Quantitative Economics

June 4, 2014

**Abstract**

In this paper we analyze Granger causality testing in a mixed-frequency VAR, originally proposed by Ghysels (2012), where the difference in sampling frequencies of the variables is large. In particular, we investigate whether past information on a low-frequency variable help in forecasting a high-frequency one and vice versa. Given a realistic sample size, the number of high-frequency observations per low-frequency period leads to parameter proliferation problems in case we attempt to estimate the model unrestrictedly. We propose two approaches to solve this problem, reduced rank restrictions and a Bayesian mixed-frequency VAR. For the latter, we extend the approach in Banbura et al. (2010) to a mixed-frequency setup, which presents an alternative to classical Bayesian estimation techniques. We compare these methods to a common aggregated low-frequency model as well as to the unrestricted VAR in terms of their Granger non-causality testing behavior using Monte Carlo simulations. The techniques are illustrated in an empirical application involving daily realized volatility and monthly business cycle fluctuations.

**JEL Codes:** C12, C32
**JEL Keywords:** Granger Causality, Mixed Frequency VAR, Bayesian VAR, Volatility, MIDAS

*We want to particularly thank Eric Ghysels for many fruitful discussions, comments and suggestions on earlier versions of the paper. Moreover, we thank Daniela Osterrieder, Lenard Lieb, Stephan Smeekes and participants of the Workshop on Advances in Quantitative Economics in Maastricht 2013, the CEIS Seminar in Rome 2013, the $33^{rd}$ International Symposium on Forecasting in Seoul 2013, the $24^{th}$ $(EC)^2$ conference in Cyprus 2013 and and the $22^{th}$ SNDE Annual Symposium in New York 2014 for useful suggestions and comments on earlier versions of the paper. The usual disclaimer applies.

†Corresponding author: Thomas Götz, Department of Quantitative Economics, School of Business and Economics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands. Email: t.gotz@maastrichtuniversity.nl

‡Department of Quantitative Economics, School of Business and Economics, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands. Email: a.hecq@maastrichtuniversity.nl

# 1 Introduction

Economic time series are published at various frequencies. While higher frequency variables used to be aggregated (e.g., Silvestrini and Veredas, 2008), it has become more and more popular to consider models that take into account the difference in frequencies of the processes under consideration. As argued extensively in the mixed-frequency literature (e.g., Ghysels et al., 2007), working in a mixed-frequency setup instead of a common low-frequency one is advantageous due to the potential loss of information in the latter scenario and feasibility of the former through MI(xed) DA(ta) S(ampling) regressions (Ghysels et al., 2004), even in the presence of many high-frequency variables compared to the number of observations.

Until recently, the MIDAS literature was limited to the single-equation framework, in which one of the low-frequency variables is chosen as the dependent variable and the high-frequency ones are in the regressors. Since the work of Ghysels (2012) for stationary series and the extension of Götz et al. (2013) and Ghysels and Miller (2013) for the non-stationary and possibly cointegrated case, we can analyze the link between high- and low-frequency series in a VAR system treating all variables as endogenous. Ghysels et al. (2013) define causality in such a mixed-frequency VAR and develop a corresponding test statistic. The authors, however, make an implicit assumption on the variables involved, which does not hold in many economic applications: The number of high-frequency observations within a low-frequency period is rather small, e.g., as in a year/quarter- or quarter/month-example.

In this paper we analyze the finite sample behavior of Granger non-causality tests when the number of high-frequency observations per low-frequency period is large, e.g., as in a month/working day-example (like the application presented in this chapter). To potentially avoid the proliferation of parameters we consider two parameter reduction techniques, reduced rank restrictions and a Bayesian mixed-frequency VAR. With respect to the latter we show how to properly extend the dummy observation approach of Banbura et al. (2010) to a mixed-frequency setting.[1] Importantly, due to stacking the high-frequency variables in the mixed-frequency VAR (Ghysels, 2012), their approach cannot be applied directly such that a properly adapted choice of auxiliary dummy variables corresponding to the prior moments is required. The aforementioned approaches are compared to a common low-frequency VAR, obtained by temporally aggregating the high-frequency variable (**?**), and the unrestricted model, which is expected to suffer from a parameter proliferation problem.

The rest of the paper is organized as follows. In Section 2 notations are introduced, the mixed-frequency VAR (MF-VAR hereafter) for our specific case at hand is presented and Granger (non-)causality is defined formally. Section 3 discusses the approaches to reduce the number of parameters to be estimated, whereby reduced rank restrictions (Section 3.1) as well as Bayesian MF-VARs (Section 3.2) are presented in detail. The finite sample performances of these tests are analyzed via a Monte Carlo experiment in Section 4. An empirical example with U.S. data on the monthly industrial production index and daily volatility in Section 5

[1] Banbura et al. (2010) refer to these variables as dummy observations. To avoid confusion between high- and low-frequency *observations* and auxiliary *variables*, we use the term 'auxiliary dummy variables'.

illustrates the results. Section 6 gives concluding remarks.

# 2 Causality in a Mixed-frequency VAR

## 2.1 Notation

Let us start from a two variable mixed-frequency system, where $y_t$, $t = 1, \ldots, T$, is the low-frequency variable and $x^{(m)}_{t-i/m}$ are the high-frequency variables with $m$ high-frequency observations per low-frequency period $t$. Contrary to Ghysels et al. (2013), we assume $m$ to be rather large as in a year/month- or month/working day-example. We also assume $m$ to be constant rather than varying with $t$.[2] The value of $i$ indicates the specific high-frequency observation under consideration, ranging from the beginning of each $t$-period ($x^{(m)}_{t-(m-1)/m}$) until the end ($x^{(m)}_t$ with $i = 0$). These notational conventions have become standard in the mixed-frequency literature and are similar to the ones in Götz et al. (2014), Clements and Galvão (2008, 2009) or Miller (2012).

Furthermore, let $\underline{W_t} = (W'_{t-1}, W'_{t-2}, \ldots, W'_{t-p})'$, i.e., it stacks the last $p$ low-frequency lags of any process $W$. Finally, $\mathbf{0}_{i\times j}$ ($\mathbf{1}_{i\times j}$) denotes an $(i \times j)$-matrix of zeros (ones), $I_i$ is an identity matrix of dimension $i$, $\otimes$ represents the Kronecker product and *vec* corresponds to the operator stacking the columns of a matrix.[3]

## 2.2 MF-VARs

Considering each high-frequency variable such that

$$X^{(m)}_t = (x^{(m)}_t, x^{(m)}_{t-1/m}, \ldots, x^{(m)}_{t-(m-1)/m})',$$

a dynamic structural equations model for the stationary multivariate process $Z_t = (y_t, X^{(m)\prime}_t)'$ is given by $A_c Z_t = c + A_1 Z_{t-1} + \ldots + A_p Z_{t-p} + \varepsilon_t$,[4] where $\Sigma_\varepsilon$ is a diagonal covariance matrix and $A_c$ pertains to contemporaneous relationships between the series. Note that the parameters in $A_c$ are related to the ones in $A_1$ due to stacking the high-frequency observations $X^{(m)}_t$ in $Z_t$

[2] As long as $m$ is deterministic, even time-varying frequency discrepancies do not pose a problem on a theoretical level. However, the assumption of constant $m$ simplifies the notation greatly (Ghysels, 2012).

[3] Extensions towards representations of higher dimensional multivariate systems as in Ghysels et al. (2013) can be considered, but are left for further research here. Analyzing Granger causality among more than two variables inherently leads to multi-horizon causality (see Lütkepohl, 1993 among others). The latter implies the potential presence of a causal chain: For example, in a trivariate system, $X$ may cause $Y$ through an auxiliary variable $Z$. To abstract from that scenario, Ghysels et al. (2013) often consider cases, in which high- and low-frequency variables are grouped and causality patterns between these groups, viewed as a bivariate system, are analyzed. They study the presence of a causal chain and multi-horizon causality in a Monte Carlo analysis though.

[4] Compared to Ghysels (2012) we put the low-frequency variable first.

(Ghysels, 2012). Explicitly, for a lag length of $p = 1$ the model reads as:

$$\begin{pmatrix} 1 & \beta_1 & \beta_2 & \dots & \beta_m \\ \delta_1 & 1 & -\rho_1 & \dots & -\rho_{m-1} \\ \delta_2 & 0 & 1 & \dots & -\rho_{m-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \delta_m & 0 & 0 & \dots & 1 \end{pmatrix} \begin{pmatrix} y_t \\ x_t^{(m)} \\ x_{t-1/m}^{(m)} \\ \vdots \\ x_{t-(m-1)/m}^{(m)} \end{pmatrix} =$$

$$\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_{m+1} \end{pmatrix} + \begin{pmatrix} \rho_y & \theta_1 & \theta_2 & \dots & \theta_m \\ \psi_2 & \rho_m & \dots & \dots & 0 \\ \psi_3 & \rho_{m-1} & \rho_m & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \psi_{m+1} & \rho_1 & \rho_2 & \dots & \rho_m \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1}^{(m)} \\ x_{t-1-1/m}^{(m)} \\ \vdots \\ x_{t-1-(m-1)/m}^{(m)} \end{pmatrix} + \tag{1}$$

$$\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{(m+1)t} \end{pmatrix}.$$

The matrix $A_c$ links contemporaneous values of $y$ and $x$: $\beta_j \neq 0$ implies that $y_t$ is affected by incoming observations of $X_t^{(m)}$, whereas $\delta_j \neq 0$ implies that the high-frequency observations are influenced by $y_t$ (see Götz and Hecq, 2014). The latter becomes interesting for studying policy analysis, where the high-frequency policy variable(s) may react to current low-frequency conditions (see Ghysels, 2012 for details). Note that an AR($m$) model is assumed for the high-frequency process. If instead it is assumed to follow an AR($q$) model with $q < m$, one should set $\rho_{q+1} = \ldots = \rho_m = 0$ in (1).

Pre-multiplying (1) by $A_c^{-1}$ we get to the mixed-frequency reduced-form VAR($p$), i.e., the aforementioned MF-VAR, model:

$$\begin{aligned} Z_t = &\quad A_c^{-1}c + A_c^{-1}A_1 Z_{t-1} + \ldots + A_c^{-1}A_p Z_{t-p} + A_c^{-1}\varepsilon_t \\ = &\quad \mu + \Gamma_1 Z_{t-1} + \ldots + \Gamma_p Z_{t-p} + u_t \end{aligned}$$

or

$$Z_t = \quad \mu + B'\underline{Z}_t + u_t, \tag{2}$$

where $B = (\Gamma_1, \Gamma_2, \ldots, \Gamma_p)'$. Equation (2) is easy to estimate on small systems. As an example,

let us explicitly write a VAR(1) as

$$\begin{pmatrix} y_t \\ x_t^{(m)} \\ x_{t-1/m}^{(m)} \\ \vdots \\ x_{t-(m-1)/m}^{(m)} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_{m+1} \end{pmatrix} + \underbrace{\begin{pmatrix} \gamma_{1,1} & \gamma_{1,2} & \cdots & \gamma_{1,m+1} \\ \gamma_{2,1} & \gamma_{2,3} & \cdots & \gamma_{2,m+1} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m+1,1} & \gamma_{m+1,2} & \cdots & \gamma_{m+1,m+1} \end{pmatrix}}_{\Gamma_1} \times \begin{pmatrix} y_{t-1} \\ x_{t-1}^{(m)} \\ x_{t-1-1/m}^{(m)} \\ \vdots \\ x_{t-1-(m-1)/m}^{(m)} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \\ \vdots \\ u_{(m+1)t} \end{pmatrix}, \tag{3}$$

$$u_t \sim N(\mathbf{0}_{((m+1)\times 1)}, \Sigma_u), \ \Sigma_u = \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} & \cdots & \sigma_{1,m+1} \\ \sigma_{2,1} & \sigma_{2,2} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m+1,1} & \cdots & \cdots & \sigma_{m+1,m+1} \end{pmatrix}.$$

### 2.3 Granger Causality in MF-VARs

Let the information set generated by the collection of sigma-fields $F_t = \sigma(Z_s, s \leq t), t \geq 0$ be denoted by $\Omega_t$. Furthermore, let $Z_t$ be adapted to that filtration such that $\Omega_t$ represents the set of information available at moment $t$ (**?**). $\Omega_t^W$ denotes the information set containing the information for all stochastic process except $W$ up to moment $t$. $P[X_{t+h}^{(m)}|\Omega_t^W]$ is the best linear forecast of $X_{t+h}^{(m)}$ based on $\Omega_t^W$ and likewise for $P[y_{t+h}|\Omega_t^W]$. Granger non-causality is defined as follows (**?**):

**Definition 1** *$y$ does not Granger cause $X^{(m)}$ if*

$$P[X_{t+1}^{(m)}|\Omega_t^y] = P[X_{t+1}^{(m)}|\Omega_t].$$

*Similarly, $X^{(m)}$ does not Granger cause $y$ if*

$$P[y_{t+1}|\Omega_t^{X^{(m)}}] = P[y_{t+1}|\Omega_t].$$

In other words, $y$ does not Granger cause $X^{(m)}$ if past information of the low-frequency variable do not help in predicting current (or future) values of the high-frequency variables and vice versa (**?**). In terms of (3), testing for Granger non-causality implies the following null (and alternative) hypotheses:

- $y$ does not Granger cause $X^{(m)}$

$$H_0 : \gamma_{2,1} = \gamma_{3,1} = \ldots = \gamma_{m+1,1} = 0,$$
$$H_A : \gamma_{i,1} \neq 0 \text{ for at least one } i = 2, \ldots, m+1.$$

- $X^{(m)}$ does not Granger cause $y$

$$H_0 : \gamma_{1,2} = \gamma_{1,3} = \ldots = \gamma_{1,m+1} = 0,$$
$$H_A : \gamma_{1,i} \neq 0 \text{ for at least one } i = 2, \ldots, m+1.$$

Ghysels et al. (2013) implicitly assume $m$ to be rather small ($m = 3$ or 4 in their case) in order to estimate the MF-VAR and then test for Granger non-causality. With $m$ large (say $m \geq 12$), we would need a much larger sample than is usually applicable for macroeconomic data sets to estimate the parameters and test for causality properly. Additional lags ($p > 1$) would further complicate the issue.

### 2.4 Remark on Nowcasting Causality

It becomes clear from Section 2.3 that Granger non-causality in both directions is defined in terms of the low frequency index $t$. Given the mixed-frequency nature of the variables under consideration, it may be of interest to analyze whether knowing the values of $x^{(m)}_{t-i/m}, i = 0, \ldots, m-1$, helps to predict $y_t$, which is referred to as instantaneous causality in the common-frequency case (**?**, Lütkepohl, 2005, p.42, Hamilton, 1994, p. 301, Gianetto and Raïssi, 2012).

Götz and Hecq (2014) have introduced the term nowcasting causality as the mixed-frequency analogue to instantaneous causality, because it amounts to predicting $y_t$ using values of the high-frequency variables *within* period $t$. In terms of testing for nowcasting non-causality, Götz and Hecq (2014) argue that $X^{(m)}_t$ is not nowcasting causal for $y_t$ if and only if the corresponding errors of the data generating process of $Z_t$ are uncorrelated (Lütkepohl, 2005). Indeed, testing for nowcasting non-causality corresponds to $\beta_j = \delta_j = 0 \ \forall j$ in (1). In that case, though, $\Sigma_u$ is block diagonal with $\sigma_{1,j} = \sigma_{j,1} = 0$, $j = 2, \ldots, m+1$, because $\Sigma_u = A_c^{-1}\Sigma_\varepsilon A_c^{-1\prime}$. Thus, we can test for nowcasting non-causality by using a Wald test on the $(1,j)$-elements of $\widehat{\Sigma_u}$, i.e., $\widehat{\sigma}_{1,j} = \frac{1}{T}\sum_{t=1}^{T} \widehat{u}_{1t}\widehat{u}_{jt}, j = 2, \ldots, m+1$, where $\widehat{u}_{jt}$ corresponds to the residuals of equation $j$.[5]

## 3 Parameter Reduction

This section presents techniques that we have considered, and evaluated through a Monte Carlo exercise, with the aim to reduce the amount of parameters to be estimated in the MF-VAR model. Two approaches are discussed in detail, reduced rank restrictions and a Bayesian MF-VAR.

[5] In this paper we do not include degree of freedom corrections when computing estimates of covariance matrices due to their invariance asymptotically.

There are many alternative approaches to reduce the number of parameters among which are principal components, Lasso (Tibshirani, 1996) or ridge regressions.[6] The reason not to consider principal components is that it does not necessarily preserve the dynamics of the VAR under the null. To be more precise, nothing prevents, e.g., the first and only principal component to be loading only on $y$ implying that the remaining dynamics enter the error term. In other words, principal components may and will most likely not preserve the autoregressive matrices in (2). This affects, in a possibly drastic manner, the block of parameters we test on for Granger non-causality.

## 3.1 Reduced Rank Restrictions

### 3.1.1 Reduced Rank Regression Model

In order to reduce the number of parameters to estimate in the MF-VAR, we propose the following reduced rank regression model, in which we assume that $rk(B'_{X^{(m)}}) = r < m$ (Carriero et al., 2011) with $B'_{X^{(m)}}$ being obtained from $B'$ by excluding the first columns of $\Gamma_1$, $\Gamma_2$, ..., $\Gamma_p$. The model reads as

$$\begin{aligned} Z_t = &\quad \mu + \gamma_{\cdot,1}\underline{y}_t + \alpha \textstyle\sum_{i=1}^{p} \delta_i' X_{t-i}^{(m)} + \nu_t \\ = &\quad \mu + \gamma_{\cdot,1}\underline{y}_t + \alpha\delta' \underline{X}_t^{(m)} + \nu_t, \end{aligned} \tag{4}$$

where $\gamma_{\cdot,1}$ is the $(m+1) \times p$ matrix containing the first columns of $\Gamma_1, \Gamma_2, \ldots, \Gamma_p$ and $\alpha$ and $\delta = (\delta_1', \ldots, \delta_p')'$ are $(m+1) \times r$ and $pm \times r$ matrices, respectively. Note that (4) can also be written in terms of $\underline{Z}_t$. Let us define $\underline{\Gamma}_i \equiv (\gamma_{\cdot,1}^{(i)}, \alpha\delta_i')$, where $\gamma_{\cdot,1}^{(i)}, i = 1, \ldots, p$, corresponds to the $i^{th}$ column of $\gamma_{\cdot,1}$. Then, (4) is equivalent to

$$Z_t = \mu + \underline{B}'\underline{Z}_t + \nu_t, \tag{5}$$

[6] Cubadda and Guardabascio (2012) analyze a so-called 'medium-$N$' approach arguing that many of the results in the literature favor a number of predictors ($N$) that is considerably larger than in usual small-scale forecasting problems, but not too large for being forced to rely on double ($T$ and $N$) asymptotic methods. The authors find, using Monte Carlo simulations, that, under the so-called Helland and Almoy (1994) condition, both principal components and partial least squares (to be discussed later) provide consistent estimates in such a medium-$N$ framework as only the sample size diverges.

where $\underline{B} = (\underline{\Gamma}_1, \ldots, \underline{\Gamma}_p)'$. For $p = 1$ the models becomes

$$\begin{pmatrix} y_t \\ x^{(m)} \\ x^{(m)}_{t-1/m} \\ \vdots \\ x^{(m)}_{t-(m-1)/m} \end{pmatrix} = \mu + \begin{pmatrix} \gamma_{1,1} \\ \gamma_{2,1} \\ \vdots \\ \gamma_{m+1,1} \end{pmatrix} y_{t-1} + \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{m+1} \end{pmatrix} \delta_1' \begin{pmatrix} x^{(m)}_{t-1} \\ x^{(m)}_{t-1-1/m} \\ \vdots \\ x^{(m)}_{t-1-(m-1)/m} \end{pmatrix} + \nu_t$$

$$= \mu + \underbrace{\left[ \begin{pmatrix} \gamma_{1,1} \\ \gamma_{2,1} \\ \vdots \\ \gamma_{m+1,1} \end{pmatrix}, \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{m+1} \end{pmatrix} \delta_1' \right]}_{\underline{\Gamma}_1} \times \begin{pmatrix} y_{t-1} \\ x^{(m)}_{t-1} \\ x^{(m)}_{t-1-1/m} \\ \vdots \\ x^{(m)}_{t-1-(m-1)/m} \end{pmatrix} + \nu_t,$$

where each $\alpha_i, i = 1, \ldots, m+1$, is of dimension $1 \times r$ and where $\delta_1$ is an $m \times r$ matrix. Hence, we could call $\delta' \underline{X}_t^{(m)}$ a vector of $r$ observable high-frequency factors. Note that $r = m - s$, where $s$ represents the rank reduction we are able to achieve within $\underline{X}_t^{(m)}$.[7]

In terms of parameter reduction, the unrestricted VAR in (2) requires $p(m+1)^2$ coefficients to be estimated in the autoregressive matrices, whereas the VAR under reduced rank restrictions in (4) or (5) needs $p(m+1) + r(m+1) + prm$ parameter estimates. As an example, assume $p = 1$ and $m = 20$. Then, if $r = 1, 2$ or 3, there are, respectively, $62, 103$ or 144 coefficients to be estimated in $\underline{\Gamma}_1$ instead of 441 in $\Gamma_1$ in (3). Note that we do not require $y_{t-1}$ to be included in the same transmission mechanism as the $x$ variables.

### 3.1.2 Obtaining Observable Factors

We use three ways to obtain the observable factors, canonical correlation analysis (CCA hereafter), partial least squares (PLS hereafter) and heterogeneous autoregressive (HAR heferafter) type restrictions.

CCA is based on analyzing the eigenvalues and corresponding eigenvectors of

$$\hat{\Sigma}^{-1}_{\underline{\tilde{X}}^{(m)}\underline{\tilde{X}}^{(m)}} \hat{\Sigma}_{\underline{\tilde{X}}^{(m)}\tilde{Z}} \hat{\Sigma}^{-1}_{\tilde{Z}\tilde{Z}} \hat{\Sigma}_{\tilde{Z}\underline{\tilde{X}}^{(m)}} \tag{6}$$

or, similarly, of the symmetric matrix

$$\hat{\Sigma}^{-1/2}_{\underline{\tilde{X}}^{(m)}\underline{\tilde{X}}^{(m)}} \hat{\Sigma}_{\underline{\tilde{X}}^{(m)}\tilde{Z}} \hat{\Sigma}^{-1}_{\tilde{Z}\tilde{Z}} \hat{\Sigma}_{\tilde{Z}\underline{\tilde{X}}^{(m)}} \hat{\Sigma}^{-1/2}_{\underline{\tilde{X}}^{(m)}\underline{\tilde{X}}^{(m)}}. \tag{7}$$

[7] There are two ways to justify the reduced rank feature of the autoregressive matrix. First, at the model representation level we may assume that $x$ follows an AR($q$) process with $q < m$ and that the last $m - q$ elements of each $X^{(m)}_{t-i}, i = 1, \ldots, p$, have a zero coefficient in the equation for $y_t$. Plugging these restrictions into (2) results in a reduced rank of $B'_{X^{(m)}}$. Second, at the empirical level we can interpret the MF-VAR as an approximation of the VARMA obtained after the block marginalization of a high-frequency VAR for each variable. In this situation, reduced rank matrices may empirically not be rejected by the data.

For a detailed discussion we refer the reader to Anderson (1951) or Vahid and Engle (1993) for the application to common dynamics. Note that $\hat{\Sigma}_{ij}$ represents the empirical covariance matrix of processes $i$ and $j$. Furthermore, $\tilde{Z}$ and $\underline{\tilde{X}}^{(m)}$ indicate $Z_t$ and $\underline{X}_t^{(m)}$, respectively, to be concentrated out by the variables that do not enter in the reduced rank regression, i.e., the intercept and $y_{t-1}$. Denoting by $\hat{V} = (v_1, v_2, \ldots, v_r)$, with $v_i' v_j = 1$ for $i = j$ and 0 otherwise, the eigenvectors corresponding to the $r$ largest eigenvalues of the matrix in (7), we obtain the estimators of $\alpha$ and $\delta$ as:

$$\begin{aligned} \hat{\alpha} = &\ \hat{\Sigma}^{-1}_{\tilde{Z}\tilde{Z}} \hat{\Sigma}_{\tilde{Z}\underline{\tilde{X}}^{(m)}} \hat{\Sigma}^{-1/2}_{\underline{\tilde{X}}^{(m)}\underline{\tilde{X}}^{(m)}} \hat{V} \\ \hat{\delta} = &\ \hat{\Sigma}^{-1/2}_{\underline{\tilde{X}}^{(m)}\underline{\tilde{X}}^{(m)}} \hat{V}. \end{aligned}$$

Note that the estimation of the eigenvectors obtained from the canonical correlation analysis in (6) or (7) may, however, perform poorly with high-dimensional systems, because inversions of the large variance matrices $\hat{\Sigma}^{-1}_{\tilde{Z}\tilde{Z}}$ and $\hat{\Sigma}^{-1}_{\underline{\tilde{X}}^{(m)}\underline{\tilde{X}}^{(m)}}$ are required.

As an alternative to CCA we use a PLS algorithm similar to the one used in Cubadda and Hecq (2011) or Cubadda and Guardabascio (2012). In order to make the solution of this eigenvalue problem invariant to scale changes of individual elements, we compute the eigenvectors associated with the largest eigenvalues of the matrix $\hat{D}^{-1/2}_{\underline{\tilde{X}}^{(m)}\underline{\tilde{X}}^{(m)}} \hat{\Sigma}_{\underline{\tilde{X}}^{(m)}\tilde{Z}} \hat{D}^{-1}_{\tilde{Z}\tilde{Z}} \hat{\Sigma}_{\tilde{Z}\underline{\tilde{X}}^{(m)}} \hat{D}^{-1/2}_{\underline{\tilde{X}}^{(m)}\underline{\tilde{X}}^{(m)}}$ with $\hat{D}_{\underline{\tilde{X}}^{(m)}\underline{\tilde{X}}^{(m)}}$ and $\hat{D}_{\tilde{Z}\tilde{Z}}$ being diagonal matrices having the diagonal elements of, respectively, $\hat{\Sigma}_{\underline{\tilde{X}}^{(m)}\underline{\tilde{X}}^{(m)}}$ and $\hat{\Sigma}_{\tilde{Z}\tilde{Z}}$ as their entries. The computation of $\hat{\alpha}$ and $\hat{\delta}$ works in a similar fashion as with CCA-based factors.

Finally, we may *impose* the presence of $r = 3p$ factors,[8] inspired by the Corsi HAR-model (Corsi, 2009). For $i = 1, \ldots, p$:

$$\delta_i = \begin{pmatrix} \mathbf{0}_{3(i-1)\times m} \\ \hline 1 \quad \mathbf{0}_{1\times(m-1)} \\ \hline \mathbf{1}_{1\times(\frac{1}{4}m)} \quad \mathbf{0}_{(1\times\frac{3}{4}m)} \\ \hline \mathbf{1}_{1\times m} \\ \hline \mathbf{0}_{3(p-i)\times m} \end{pmatrix}' \Rightarrow \delta' \underline{X}_t^{(m)} = \begin{pmatrix} x_{t-1}^{(m)} \\ \sum_{i=0}^{\frac{1}{4}m-1} x_{t-1-i/m}^{(m)} \\ \sum_{i=0}^{m-1} x_{t-1-i/m}^{(m)} \\ \vdots \\ x_{t-p}^{(m)} \\ \sum_{i=0}^{\frac{1}{4}m-1} x_{t-p-i/m}^{(m)} \\ \sum_{i=0}^{m-1} x_{t-p-i/m}^{(m)} \end{pmatrix}.$$

For $m = 20$, i.e., the month/working day-example mentioned before, and $p = 1$ this corresponds to

$$\delta' \underline{X}_t^{(m)} = \begin{pmatrix} x_{t-1}^{(20)} \\ \sum_{i=0}^{4} x_{t-1-i/20}^{(20)} \\ \sum_{i=0}^{19} x_{t-1-i/20}^{(20)} \end{pmatrix} \equiv \begin{pmatrix} x_{t-1}^{D} \\ x_{t-1}^{W} \\ x_{t-1}^{M} \end{pmatrix},$$

[8] To ensure that $r < m$ we assume hence that $p < \frac{1}{3}m$ in this case.

with $x_t^D, x_t^W$ and $x_t^M$ denoting daily, weekly and monthly measures, respectively. As noted in Ghysels and Valkanov (2012) and Ghysels et al. (2007), these HAR-type restrictions are a special case of MIDAS with step functions introduced in Forsberg and Ghysels (2007). Considering partial sums of regressors $x$ as $X_t(K,m) = \sum_{i=0}^{K} x_{t-i/m}^{(m)}$, a MIDAS regression with $M$ steps reads as

$$y_t = \mu + \sum_{j=1}^{M} \beta_j X_t(K_j, m) + \epsilon_t,$$

where $K_1 < \ldots < K_M$. Alternative to this type of HAR-restrictions, we could use MIDAS restrictions as introduced in Ghysels et al. (2004). Indeed, the difference between HAR and MIDAS models can be small (Ghysels and Valkanov, 2012). However, step functions have the advantage of not requiring non-linear estimation methods, since the distributed lag pattern is approximated by a number of discrete steps. Also, implementing MIDAS restrictions and testing for Granger non-causality implies the well-known Davies (1987) problem, i.e., the parameters determining the MIDAS weights (see Ghysels et al., 2004 for details) are not identified under the null hypothesis.[9]

### 3.1.3 Testing for Granger Non-Causality

Given $r$, we can test for Granger non-causality by defining R as the matrix that picks the set of coefficients we want to do inference on, i.e., R$vec(\widehat{\underline{B}})$, where $\widehat{\underline{B}}$ is the estimator of $\underline{B}$. For a general construction of the matrix R in the presence of several low- and high-frequency variables, we refer the reader to Ghysels et al. (2013). The Wald test is then constructed as

$$\widetilde{\xi}_W = \left[\mathrm{R}vec(\widehat{\underline{B}})\right]' \left(\mathrm{R}\hat{\Omega}\mathrm{R}'\right)^{-1} \left[\mathrm{R}vec(\widehat{\underline{B}})\right], \tag{8}$$

with

$$\hat{\Omega} = (W'W)^{-1} \otimes \hat{\Sigma},$$

where $\hat{\Sigma} = \frac{1}{T}\hat{u}'\hat{u}$ is the empirical covariance matrix of the disturbance terms and $W$ is the regressor set, i.e., an intercept, $y_{t-1}$ and the high-frequency factors $\delta' \underline{X}_t^{(m)}$. As illustrated in Ghysels et al. (2013), $\xi_W$ is asymptotically $\chi^2_{rank(\mathrm{R})}$. A robust version of (8) is also implemented in the empirical section.

[9] To properly test for Granger non-causality in this case, a grid for the weight specifying parameter vector has to be considered and the corresponding Wald tests for each candidate have to be computed. Subsequently, one can calculate the supremum of these tests (Davies, 1987) and obtain an 'asymptotic $p$-value' using bootstrap techniques (see Hansen, 1996 or Ghysels et al., 2007 for details). While this approach is feasible, it is computationally more demanding due to applying bootstrapping within a relatively large (due to large $m$) regression. Admittedly, the HAR-type restrictions provide less flexibility than the MIDAS approach, yet rely on linear estimation of the model, which simplifies the analysis greatly and is therefore very appealing from an applied perspective.

## 3.2 Bayesian MF-VARs

### 3.2.1 Restricted MF-VARs

No restrictions are placed on the coefficients in the autoregressive matrices, i.e., $B$, in (2). While this may seem reasonable for the first equation, i.e., the one for $y_t$, it is less clear for the remaining ones. Ghysels (2012) discusses the issue of parsimony in MF-VAR models by specifying the high-frequency process in such a way as to allow a number of parameters that is independent from $m$. In particular, assuming $x$ to follow an ARX(1) process with only one lag of $y$ in the regressor set yields

$$x^{(m)}_{t-i/m} = \mu_{i+2} + \rho x^{(m)}_{t-(i+1)/m} + \pi y_{t-1} + \upsilon_{(i+2)t} \text{ for } i = 0, \ldots, m-1. \tag{9}$$

Completing the system with the first equation of the unrestricted MF-VAR, i.e.,

$$y_t = \mu_1 + \sum_{k=1}^{p} \left[ \gamma^{(k)}_{1,1} y_{t-k} + \sum_{j=2}^{m+1} \gamma^{(k)}_{1,j} x^{(m)}_{t-k-(m-1)/m} \right] + \upsilon_{1t}, \tag{10}$$

where $\gamma^{(k)}_{i,j}$ corresponds to the $(i,j)$-element of matrix $\Gamma_k$ in (2), leads to the following restricted MF-VAR:

$$\begin{pmatrix} y_t \\ x^{(m)}_t \\ x^{(m)}_{t-1/m} \\ \vdots \\ x^{(m)}_{t-(m-1)/m} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \sum_{i=0}^{m-1} \rho^i \mu_{2+i} \\ \sum_{i=0}^{m-2} \rho^i \mu_{3+i} \\ \vdots \\ \mu_{m+1} \end{pmatrix} + \begin{pmatrix} \gamma^{(1)}_{1,1} & \gamma^{(1)}_{1,2} & \gamma^{(1)}_{1,3} & \cdots & \gamma^{(1)}_{1,m+1} \\ \pi \sum_{i=0}^{m-1} \rho^i & \rho^m & & & \\ \pi \sum_{i=0}^{m-2} \rho^i & \rho^{m-1} & & \mathbf{0}_{m\times(m-1)} & \\ \vdots & \vdots & & & \\ \pi & \rho & & & \end{pmatrix} \times$$
$$\begin{pmatrix} y_{t-1} \\ x^{(m)}_{t-1} \\ x^{(m)}_{t-1-1/m} \\ \vdots \\ x^{(m)}_{t-1-(m-1)/m} \end{pmatrix} + \sum_{k=2}^{p} \begin{pmatrix} \gamma^{(k)}_{1,1} & \gamma^{(k)}_{1,2} & \cdots & \gamma^{(k)}_{1,m+1} \\ & \mathbf{0}_{m\times(m+1)} & & \end{pmatrix}$$
$$\times \begin{pmatrix} y_{t-k} \\ x^{(m)}_{t-k} \\ \vdots \\ x^{(m)}_{t-k-(m-1)/m} \end{pmatrix} + \underbrace{\begin{pmatrix} \upsilon_{1t} \\ \sum_{i=0}^{m-1} \rho^i \upsilon_{(m+1-i)t} \\ \sum_{i=0}^{m-2} \rho^i \upsilon_{(m+1-i)t} \\ \vdots \\ \upsilon_{(m+1)t} \end{pmatrix}}_{\upsilon^*_t}. \tag{11}$$

As for the error terms in (9) and (10), we denote, similarly to Ghysels (2012), that $\mathbb{E}(\upsilon_{it}\upsilon_{it}) = \sigma_{HH}$, $\mathbb{E}(\upsilon_{1t}\upsilon_{it}) = \sigma_{HL}$ for $i = 2, \ldots, m+1$, and that $\mathbb{E}(\upsilon_{1t}\upsilon_{1t}) = \sigma_{LL}$. Furthermore, each

error term is assumed to possess a zero mean and to be normally distributed. Consequently, $\upsilon_t^* \sim N(\mathbf{0}_{(21\times 1)}, \Sigma_{\upsilon^*})$ in (11) with

$$\Sigma_{\upsilon^*} = \begin{pmatrix} \sigma_{LL} & \sigma'_{\cdot L} \\ \sigma_{\cdot L} & \Sigma^H_{\upsilon^*} \end{pmatrix}, \tag{12}$$

where

$$\sigma_{\cdot L} = \sigma_{HL} \left( \sum_{i=0}^{m-1} \rho^i, \sum_{i=0}^{m-2} \rho^i, \ldots, \sum_{i=0}^{1} \rho^i, 1 \right)'$$

and

$$\Sigma^H_{\upsilon^*} = \sigma_{HH} \begin{pmatrix} \sum_{i=0}^{m-1} \rho^{2i} & \sum_{i=0}^{m-2} \rho^{2i+1} & \cdots & \sum_{i=0}^{1} \rho^{2i+m-2} & \rho^{m-1} \\ \sum_{i=0}^{m-2} \rho^{2i+1} & \sum_{i=0}^{m-2} \rho^{2i} & \cdots & \sum_{i=0}^{1} \rho^{2i+m-3} & \rho^{m-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \sum_{i=0}^{1} \rho^{2i+m-2} & \sum_{i=0}^{1} \rho^{2i+m-3} & \cdots & \sum_{i=0}^{1} \rho^{2i} & \rho \\ \rho^{m-1} & \rho^{m-2} & \cdots & \rho & 1 \end{pmatrix}.$$

### 3.2.2 The Auxiliary Dummy Variable Approach for Mixed-Frequency Data

We will rely heavily on the restricted VAR in (11) when formulating prior beliefs for the parameters, which are going to be estimated using Bayesian techniques. Indeed, as pointed out by Carriero et al. (2011), Bayesian methods allow the imposition of such restrictions, while admitting influence of the data as well. Consequently, Bayesian shrinkage has become a standard tool when being faced with large-dimensional estimation problems such as large VARs (e.g., Banbura et al., 2010, Kadiyala and Karlsson, 1997). Within a MF-VAR, Ghysels (2012) describes a way to sample the MIDAS hyperparameters and subsequently formulates prior beliefs for the remaining parameters.[10] Once these hyperparameters are taken care of, the Bayesian analyses of mixed- and common-frequency VAR models are quite similar and, hence, traditional Bayesian VAR techniques (e.g., Kadiyala and Karlsson, 1997, Litterman, 1986) can be applied. This amounts to imposing a normal inverted Wishart prior for the VAR parameters, which retains the principles of the previously specified prior beliefs.

Banbura et al. (2010) show, for the common-frequency case, that adding a set of auxiliary dummy variables to the system is equivalent to imposing the aforementioned normal inverted Wishart prior. In the mixed-frequency scenario this approach can, however, not be applied directly due to the stacked nature of $X_t^{(m)}$ within $Z_t$. In the sequel, we properly extend this auxiliary dummy variable approach to the mixed-frequency setup and, thereby, show precisely

[10] Sekhposyan et al. (2014) use a Sims-Zha shrinkage prior and the algorithm in Waggoner and Zha (2003) to solve the parameter proliferation problem with Bayesian estimation techniques. Bayesian methods within mixed-frequency VARs were also considered by Schorfheide and Song (2012). However, the VAR under consideration differs from the one of Ghysels (2012) in the sense that the high-frequency observations of the low-frequency variables are assumed to be missing. Hence, it is a 'parameter-driven' (Cox et al., 1981) high-frequency VAR with latent variables instead of the 'observation-driven' VAR considered here. We comment on this difference more elaborately at a later stage.

how to choose the moments of the normal inverted Wishart prior when faced with mixed frequencies.

Let us start by considering the MF-VAR in (2) and choosing the moments for the prior distributions of the coefficients in $B$:

$$\begin{aligned}
\mathbb{E}[\gamma_{i,j}^{(k)}] &= \begin{cases} \rho^m & \text{if } i=j=k=1 \\ \rho^{m+j-i} & \text{if } k=1, j=2, i>1 \\ 0 & \text{else} \end{cases}, \\
Var[\gamma_{i,j}^{(k)}] &= \begin{cases} \phi\frac{\lambda^2}{(km+j-2)^2}\mathbb{S}_{LH} & \text{for } i=1, j>1 \\ \phi\frac{\lambda^2}{(km+2-i)^2}\mathbb{S}_{HL} & \text{for } j=1, i>1 \\ \frac{\lambda^2}{(km+j-i)^2} & \text{else} \end{cases},
\end{aligned} \tag{13}$$

where all $\gamma_{i,j}^{(k)}$ are assumed to be a priori independent and normally distributed. Note that the covariance matrix of the residuals is for now assumed diagonal and fixed, i.e., $\Sigma_u = \Sigma = \Sigma_d$ with $\Sigma_d = diag(\sigma_L^2, \sigma_H^2, \ldots, \sigma_H^2)$ of dimension $m+1$. Furthermore, $\mathbb{S}_{ij} = \frac{\sigma_i^2}{\sigma_j^2}, i,j = L,H$. For $\mu$ we take a diffuse prior.

This specification of prior beliefs is derived from the Minnesota prior in Litterman (1986) and was extended to the mixed-frequency case by Ghysels (2012). In short, $\lambda$ determines the tightness of the prior distributions around the ARX(1) specification in (9), the influence of low- on high-frequency data and vice versa is controlled by $\phi$ and, finally, $\mathbb{S}_{ij}$ governs the difference in scaling between $y$ and the $x$-variables. Note that $\lambda \approx 0$ results in the posterior coinciding with the prior, whereas $\lambda = \infty$ causes the posterior mean to coincide with the OLS estimate of the unrestricted VAR in (2). From (13) it becomes clear that the prior variances decrease very quickly as $m$ grows. Hence, in contrast to De Mol et al. (2008), we can already assume that, for a given sample size, $\lambda$ needs to *grow* with the size of the system in order for the data to have any influence on the estimates also when $m$ is large.

Define $\underline{Z} = (\underline{Z}_1^{\mu}, \ldots, \underline{Z}_T^{\mu})'$, where $\underline{Z}_t^{\mu} = (\underline{Z}_t', 1)'$, and let us re-write the MF-VAR in (2) in the following way:

$$\underbrace{vec(Z)}_{(m+1)T\times 1} = \underbrace{\underline{Z}^*}_{(m+1)T\times(m+1)n} \underbrace{vec(B^*)}_{(m+1)n\times 1} + \underbrace{vec(E)}_{(m+1)T\times 1}, \tag{14}$$

where $Z = (Z_1, \ldots, Z_T)'$, $\underline{Z}^* = I_{m+1} \otimes \underline{Z}$, $E = (u_1, \ldots, u_T)'$, $B^* = (B', \mu)'$ and $n = (m+1)p+1$. In order to drop the undesirable feature of a fixed and diagonal covariance matrix $\Sigma$, we impose a normal inverted Wishart prior (at the cost of having to set $\phi = 1$) with the following form (Kadiyala and Karlsson, 1997):

$$\begin{aligned}
&vec(B^*)|\Sigma \sim N(vec(B_0^*), [Z_0^{*\prime}(\Sigma^{-1} \otimes I_T)Z_0^*]^{-1}) \text{ and} \\
&\Sigma \sim iW(V_0, v_0),
\end{aligned} \tag{15}$$

where $Z_0^* = Z_0 \otimes I_{m+1}$.[11] $B_0^*$ and $Z_0$ have to be chosen as to let expectations and variances of

[11] We can re-write the MF-VAR in (2) as in Banbura et al. (2010). In that case, the variance of $vec(B^*)$ simplifies to $\Sigma \otimes (Z_0'Z_0)^{-1} = \Sigma \otimes \Omega_0$. However, it will prove useful to express the MF-VAR as done in (14) in our mixed-frequency setup.

the elements in $B^*$ coincide with the moments of the prior in (13). Likewise, $V_0$ and $v_0$ need to be set such that $\mathbb{E}[\Sigma] = \Sigma_d$, i.e., the fixed, diagonal covariance matrix introduced before.

As for the common-frequency case, the priors in (13) ensure that more recent lags provide more reliable information than more distant ones. However, due to the stacked nature of $X_t^{(m)}$, the prior variances $Var[\gamma_{i,j}^{(k)}]$ do not only decrease for larger low-frequency lag lengths $k = 1, \ldots, p$, but also *within* a low-frequency period. Put differently, given $\lambda$, $\phi$ and $\mathbb{S}_{ij}$, the coefficients are not shrunk according to the low-frequency time difference between the corresponding variables, but according to their *high-frequency* time difference. This is the reason why not only $k$, but also $m, j$ and $i$ appear in the denominators of the prior variances. As a consequence, and in severe contrast to the common-frequency case in Banbura et al. (2010), the matrix $Z_0$, which needs to be chosen to match prior variances with the variance of $vec(B^*)$, changes with each column of $B^*$.

Still, we can show that adding $T_d = n^{aux} + (m+1)$, with $n^{aux} = m(2p+1)$, auxiliary dummy variables $Y_d$ and $X_d$ to (14) is equivalent to imposing the normal inverted Wishart prior in (15). To this end, let

$$\begin{aligned} &B_0^{*aux} = (X_d' X_d)^{-1} X_d' Y_d, \\ &Z_0^{aux} = X_d \\ &V_0 = (Y_d - X_d B_0^{*aux})'(Y_d - X_d B_0^{*aux}) \text{ and} \\ &v_0 = m + 3 = T_d - n^{aux} + 2, \end{aligned} \tag{16}$$

and subsequently set

$$\begin{aligned} &vec(B_0^*) = S' vec(B_0^{*aux}) \text{ and} \\ &Var[vec(B^*)] = S'[Z_0^{*aux\prime}(\Sigma^{-1} \otimes I_T) Z_0^{*aux}]^{-1} S, \end{aligned} \tag{17}$$

where $S$ is an $[(m+1)n^{aux} \times (m+1)n]$-dimensional selection matrix filled with ones and zeros, the precise construction of which is given in Appendix A. Intuitively speaking, $Z_0^{aux}$ is an auxiliary matrix constructed as the 'union' of the $Z_0$ matrices corresponding to the different columns of $B^*$. The non-random matrix $S$ then selects, for each column of $B^*$, the corresponding elements of $Z_0^{aux}$ in order to let the variance of each element in $B^*$ match the corresponding prior variance. Likewise, $B_0^{*aux}$ is an auxiliary matrix from which we derive $B_0^*$. The auxiliary dummy variables that imply a matching of the prior moments turn out to be

$$\underbrace{Y_d}_{T_d \times (m+1)} = \begin{pmatrix} \mathbf{0}_{2(m-1)\times 1} & \vdots & \\ \sigma_L m \rho^m / \lambda & \vdots & D_\rho \\ 0 & \vdots & \\ \hdashline \mathbf{0}_{(m(2p-1)-1)\times(m+1)} & & \\ diag(\sigma_L, \sigma_H, \ldots, \sigma_H) & & \\ \mathbf{0}_{1\times(m+1)} & & \end{pmatrix}, \tag{18}$$

$$\underbrace{X_d}_{T_d \times n^{aux}} = \begin{pmatrix} J_p^1 \otimes diag(\sigma_L, \sigma_H)/\lambda & \mathbf{0}_{2pm\times(m-1)} & \mathbf{0}_{2pm\times 1} \\ \mathbf{0}_{(m-1)\times 2pm} & J_p^2 \sigma_H / \lambda & \mathbf{0}_{(m-1)\times 1} \\ \mathbf{0}_{(m+1)\times 2pm} & \mathbf{0}_{(m+1)\times(m-1)} & \mathbf{0}_{(m+1)\times 1} \\ \mathbf{0}_{1\times 2pm} & \mathbf{0}_{1\times(m-1)} & \epsilon \end{pmatrix}, \tag{19}$$

where $D_\rho = diag^a(m\rho^m, (m-1)\rho^{m-1}, \ldots, 2\rho^2, \rho)\frac{\sigma_H}{\lambda} \otimes (0,1)'$, with $diag^a(\cdot)$ denoting an anti-diagonal matrix. Furthermore,

$$J_p^1 = diag(1, 2, \ldots, m, m+1, \ldots, 2m, \ldots, \ldots, pm),$$
$$J_p^2 = diag(pm+1, pm+2, \ldots, pm+m-1).$$

The last line of both, $Y_d$ and $X_d$, corresponds to the diffuse prior for the intercept ($\epsilon$ is a very small number), the block above imposes the prior for $\Sigma$ and the remaining blocks set the priors for the coefficients $\gamma_{i,j}^{(k)}$. As Banbura et al. (2010), we set $\sigma_i^2 = s_i^2, i = L, H$, where $s_i^2$ is the variance of a residual from an AR($p$), respectively an AR($mp$), model for $y_t$, respectively $x_t^{(m)}$.[12]

Let us now augment the model in (14) by the auxiliary dummy variables in (18) and (19). To this end, let $\underline{Z}_j^d = (\underline{Z}', (X_d S_j)')', j = 1, \ldots, m+1$, where $S_j$ is the $j^{th}$ block of the selection matrix $S$ (see Appendix A for details). Here, $S_j$ picks the elements of $Z_0^{aux} = X_d$ corresponding to column $j$ of $B^*$. Then, the augmented system becomes

$$\underbrace{vec(Z_*)}_{(m+1)(T+T_d)\times 1} = \underbrace{\underline{Z}_*^*}_{(m+1)(T+T_d)\times(m+1)n} \underbrace{vec(B^*)}_{(m+1)n\times 1} + \underbrace{vec(E_*)}_{(m+1)(T+T_d)\times 1}, \tag{20}$$

where $Z_* = (Z', Y_d')'$, $E_* = (E', E_d')'$ and $\underline{Z}_*^*$ is block diagonal with $\underline{Z}_*^* = diag\{\underline{Z}_1^d, \underline{Z}_2^d, \ldots, \underline{Z}_{m+1}^d\}$ and off-diagonal elements equal to $\mathbf{0}_{(T+T_d)\times n}$. The posterior then has the form

$$\begin{aligned} &vec(B^*)|\Sigma, Z \sim N(vec(\hat{B}^*), [\underline{Z}_*^{*\prime}(\Sigma^{-1} \otimes I_{T+T_d})\underline{Z}_*^*]^{-1}) \text{ and} \\ &\Sigma|Z \sim iW(\hat{V}, T+m+3), \end{aligned} \tag{21}$$

where

$$vec(\hat{B}^*) = [\underline{Z}_*^{*\prime}(\Sigma^{-1} \otimes I_{T+T_d})\underline{Z}_*^*]^{-1}\underline{Z}_*^{*\prime}(\Sigma^{-1} \otimes I_{T+T_d})vec(Z_*),$$

and where $\hat{V} = \hat{E}_*'\hat{E}_*$ with $vec(\hat{E}_*) = vec(Z_*) - \underline{Z}_*^* vec(\hat{B}^*)$ such that $\hat{E}_*$ is a $(T+T_d)\times(m+1)$ matrix. Note that the posterior mean of the coefficients equals the GLS estimate of a SUR regression of $vec(Z_*)$ on $\underline{Z}_*^*$. As for the common-frequency case, it can be checked that it also coincides with the posterior mean for the prior setup in (13). In practice, we estimate $\Sigma$ in the standard way, i.e., $\hat{\Sigma} = \frac{1}{T+T_d}\hat{E}_*^{ols\prime}\hat{E}_*^{ols}$, where $\hat{E}_*^{ols}$ denotes the OLS residuals of the system in (20).[13]

As a final comment, note that (10) actually corresponds to a U-MIDAS regression (Foroni et al., 2012) without contemporaneous observations of the high-frequency variable. It is worth mentioning that MIDAS restrictions (Ghysels et al., 2004) can be imposed here as an alternative. However, doing so implies leaving the linear framework, which is needed to apply the auxiliary

[12] We propose to choose $\lambda$ in such a way as to control the size of our Granger causality tests, because their analysis is the ultimate purpose of this work. To this end, we also allow $\lambda$ to change with $T$, the sample size. More details on how we choose $\lambda$ are given in the Monte Carlo section.

[13] The $i^{th}$ column of $\hat{E}_*^{ols}$, $i = 1, \ldots, m+1$, corresponds to the residuals of a regression of $(Z_{\cdot,i}', Y_{d,i}')'$ on $\underline{Z}_i^d$, where $Z_{\cdot,i}$ and $Y_{d,i}$ denote the $i^{th}$ columns of $Z$ and $Y_d$, respectively.

dummy variable approach presented above. Furthermore, even after having drawn the MIDAS hyperparameters, one still faces the aforementioned Davies (1987) problem when attempting to test for Granger non-causality from $X^{(m)}$ to $y$ (see the end of Section 3.1.2). For those reasons we leave the corresponding parameters in (10) unrestricted.

### 3.2.3 Testing for Granger Non-Causality

Ultimately, our goal is to compare our parameter reduction techniques with each other, and two benchmark models to be introduced below, in terms of their Granger non-causality testing behavior. To this end, the implementation of the Bayesian MF-VAR via the auxiliary dummy variable approach is preferable to classical Bayesian methods, since we obtain a closed-form solution for the posterior mean of the coefficients of interest. Intuitively speaking, we 'disguise' the Bayesian approach as a frequentist approach by adding our auxiliary dummy variables. As for the parameter estimates, it is desirable to deal with a testing framework that is common to all approaches considered in this paper, i.e., the Wald test.[14] Consequently, to subsequently test for Granger non-causality, the relevant test statistic is computed as

$$\xi_W^* = [Rvec(\hat{B}^*)]' \left\{ R[\underline{Z}_*^{*\prime}(\hat{\Sigma}^{-1} \otimes I_{T+T_d})\underline{Z}_*^*]^{-1} R' \right\}^{-1} [Rvec(\hat{B}^*)], \tag{22}$$

where $R$ is the suitably adjusted matrix picking the set of coefficients to test for Granger non-causality. Note that the auxiliary dummy variables enter $\xi_W^*$ through $vec(\hat{B}^*)$, $\underline{Z}_*^*$ and $\hat{\Sigma}$.

## 3.3 Common Low-Frequency VARs

Before the introduction of MIDAS regression models, high-frequency variables were usually aggregated to the low frequency in order to obtain a common frequency for all variables appearing in a regression (Silvestrini and Veredas, 2008 or Marcellino, 1999). Likewise for systems, the observations of a high-frequency, say monthly, variable were usually aggregated to a low frequency, say quarterly, such that the VAR could be estimated in the resulting common frequency.

Temporally aggregating the high-frequency variable also leads to a great reduction in parameters which need to be estimated. After all, each set of $m$ high-frequency variables per $t$-period is aggregated into one low-frequency observation. In terms of the $(p = 1)$-example in (3), instead of $(m+1)^2$ parameters to estimate in $\Gamma_1$, a common low-frequency VAR as in

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \underbrace{\begin{pmatrix} \gamma_{1,1}^{LF} & \gamma_{1,2}^{LF} \\ \gamma_{2,1}^{LF} & \gamma_{2,2}^{LF} \end{pmatrix}}_{\Gamma_1^{LF}} \begin{pmatrix} y_{t-1} \\ x_{t-1} \end{pmatrix} + \begin{pmatrix} u_{1,t} \\ u_{2,t} \end{pmatrix} \tag{23}$$

[14] Similar to the relationship between the posterior mean of the coefficients to the GLS estimate of an adequately constructed SUR regression, there is large sample correspondence between classical Wald and Bayesian posterior odds tests (**?**). For certain choices of the prior distribution, the posterior odds ratio is approximately equal to the Wald statistic. **?** shows that for any significance level $\alpha$ there exist priors such that the aforementioned correspondence holds, and vice versa.

requires only 4 parameter estimates in $\Gamma_1^{LF}$.[15] Note that $x_t = W(L^{1/m})x_t^{(m)}$, where $W(L^{1/m})$ denotes a high-frequency lag polynomial of order $A$, i.e., $W(L^{1/m})x_t^{(m)} = \sum_{i=0}^{A} w_i x_{t-i/m}^{(m)}$ (Silvestrini and Veredas, 2008). This generic specification nests the two dominating aggregation schemes in the literature, Point-in-Time ($A = 0, w_0 = 1$) and Average sampling ($A = m-1, w_i = 1/m \ \forall i$), where the former is usually applied to stock and the latter to flow variables. In view of the high-frequency variable we consider in our empirical application, the natural logarithm of bipower variation, we focus on Average sampling in this paper.

Of course, the decrease in parameters to estimate comes at the cost of disregarding information embedded in the high-frequency process. As argued in Miller (2011), if the aggregation scheme used to compute $x_t$ is different from the true underlying aggregation scheme, potentially crucial high-frequency information will be forfeited. Additionally, aggregating a high-frequency variable may lead to 'spurious' (non-)causality in the common low-frequency setup (**?**), since causality is a property which is not invariant to temporal aggregation (Marcellino, 1999 or Sims, 1971).

As far as testing for Granger non-causality is concerned, we can rely on the Wald statistic in (8), where the set of regressors $W$, the matrix $R$ and the coefficient matrix $B$ are suitably adjusted.

### 3.4 Unrestricted VARs

Finally, we can attempt to estimate the full MF-VAR in (2) ignoring the fact that, given the sample sizes usually available, the amount of parameters may be too large to estimate them properly or to test for any causality pattern adequately. Still, it may serve as benchmark case. To this end we estimate the MF-VAR using ordinary least squares disregarding the parameter proliferation problem we inevitably run into.[16] Similar to the previous section, we can test for Granger non-causality using the Wald statistic in (8).[17]

## 4 Monte Carlo Simulations

In order to assess the finite sample performance of our different parameter reduction techniques, we conduct a Monte Carlo experiment. Unlike Ghysels et al. (2013), we do not start with a common high-frequency data generating process (DGP hereafter), but rather a mixed-frequency one. The former implies the assumption that the low-frequency variable has high-frequency observations that are missing. Given such a situation it seems natural to cast the model in state space form and estimate the parameters using the Kalman filter. However, this amounts to a 'parameter-driven' model (Cox et al., 1981), which contains latent processes, by construction.

[15] Of course, a similar argument holds for $p > 1$: Instead of $p(m+1)^2$ parameters to estimate in the autoregressive matrices $\Gamma_1, \ldots, \Gamma_p$ in (2), the common low-frequency VAR($p$) requires only $4p$ parameter estimates in $\Gamma_1^{LF}, \ldots, \Gamma_p^{LF}$.

[16] In this sense the comparison is related to the one of U-MIDAS (Foroni et al., 2012) and MIDAS regression models.

[17] Again, $W$, $R$ and $B$ have to be adjusted adequately.

The latter is a feature we try to avoid in our MF-VAR. Say we are interested in the impact of shocks to one or several variables on the whole system. Using a high-frequency DGP with missing observations implies that shocks to these latent processes are also latent and unobservable. This is undesirable given that, e.g., policy shocks are, of course, observable (Foroni and Marcellino, 2013).

Recall from Section 2.1 that $Z_t = (y_t,\, x_t^{(m)},\, \ldots,\, x_{t-(m-1)/m}^{(m)})' = (y_t,\, X_t^{(m)\prime})'$. In light of our empirical investigation (Section 5) we set $m = 20$, i.e., a month/ working day-example. Furthermore, we start by investigating the case where $p = 1$ and keep the analysis of higher lag orders for future research.

As far as investigating the size of our Granger non-causality tests is concerned, we assume that the data are generated as a mixed-frequency white noise process, i.e.,

$$Z_t = u_t. \tag{24}$$

To analyze the power of the tests, we generate two different DGPs that are closely related to the restricted VAR in (11):

$$Z_t = \Gamma_P Z_{t-1} + u_t \tag{25}$$

with

$$\Gamma_P = \begin{pmatrix} \gamma_{1,1} & \gamma_{1,2} & \gamma_{1,3} & \cdots & \gamma_{1,21} \\ \gamma_{2,1}^* & \rho^{20} & 0 & \ldots & 0 \\ \gamma_{3,1}^* & \rho^{19} & \vdots & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma_{21,1}^* & \rho & 0 & \ldots & 0 \end{pmatrix},$$

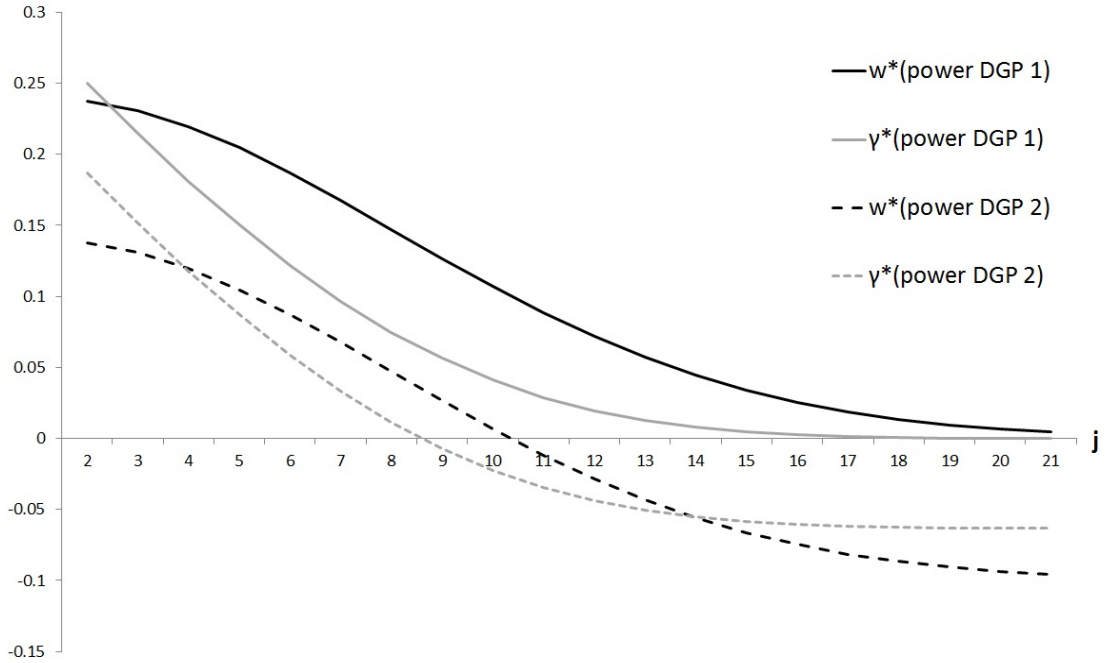
where $\gamma_{1,1} = 0.5$, $\rho = 0.6$ and $\gamma_{1,j} = 2w_{j-1}^*(-0.01)$ for $j > 1$.[18] The parameter values for $w_j^*$ and $\gamma_{j,1}^*$, $j = 2, \ldots, 21$, are depicted in Figure 1.[19]

The reason to consider two different DGPs for power is that due to the zero-mean feature of $w_{j-1}^*(\psi)$ and $\gamma_{j,1}^*, j = 2, \ldots, 21$, in the second power DGP, we expect the presence of Granger causality to be 'hidden' when Average sampling the high-frequency variable as discussed in Section 3.3. If this expectation can be confirmed, the situation underlying the second power

[18] Additionally, we have considered three alternative DGPs for size, all based on the restricted VAR in (11): (i) $\gamma_{1,i} = \gamma_{i,1}^* = 0$, (ii) $\gamma_{1,i} = 0$ and $\gamma_{i,1}^* \neq 0$ or (iii) $\gamma_{i,1}^* = 0$ and $\gamma_{1,i} \neq 0\ \forall i = 2, \ldots, 21$. Results are available upon request.

[19] The parameter values have been chosen to mimic part of the structure of the restricted VAR in (11) and to ensure stability of the system. The first DGP we consider for investigating power is characterized by positive, but decreasing coefficients $w_{j-1}^*(\psi)$ and $\gamma_{j,1}^*$. In other words, the later the observation of $x$ in $t-1$, the larger its impact on $y_t$. Furthermore, positive values of $y$ in the preceding period will increase $x$ towards the end of the current period, but hardly have an impact on it at the beginning. One can think of a delayed impact of $y$ on $x$ here. For the second power DGP the general pattern of the coefficients $w_{j-1}^*(\psi)$ and $\gamma_{j,1}^*$ does not change, except that they now possess a mean of zero. This implies that large values of $x$ early in $t-1$ have a negative effect on $y_t$, whereas large $x$-values late in $t-1$ have a positive effect. Similarly, positive values of $y$ in period $t-1$ have a negative impact on early $x$-values in period $t$, but a positive impact on $x$-values towards the end of the period.

Figure 1: Parameter values for $w^*$ and $\gamma^*_{j,1}$




Note: This figure graphs the parameter values for $w^*_{j-1}(\psi)$ and $\gamma_{j,1}, j = 2, \ldots, 21$. In the first power DGP $w^*_i(\psi) = w_i(\psi)$ with $w_i(\psi) = \frac{exp(\psi i^2)}{\sum_{i=1}^{20} exp(\psi i^2)}$, whereas in the second power DGP $w^*_i(\psi) = w_i(\psi) - \overline{w(\psi)}$, where the horizontal bar symbolizes the arithmetic mean. This specification actually corresponds to the two-dimensional exponential Almon lag polynomial with the first parameter set to zero (Ghysels et al., 2007). Likewise, $\gamma^*_{j,1} = \gamma_{j,1}$ (first power DGP) and $\gamma^*_{j,1} = \gamma_{j,1} - \overline{\gamma_{\cdot,1}}$ (second power DGP) for $j = 2, \ldots, 21$ with $\gamma_{2,1} = 0.25$ and $\gamma_{j,1} = (\gamma_{j-1,1})^{1.11}$ for $j = 3, \ldots, 21$. Note that $\overline{\gamma_{\cdot,1}} = \frac{1}{20} \sum_{j=2}^{21} \gamma_{j,1}$.

DGP would be one, in which classical temporal aggregation fails (Marcellino, 1999) and an alternative way of dealing with high-frequency observations would be called for. The first power DGP serves as a benchmark in the sense that we do not a priori expect one approach to be better or worse than the others.

For the methods in Section 3 we analyze size and size-adjusted power for $T = 50, 250$ and 500, corresponding to roughly 4, 21 and 42 years of monthly data. Note that an additional 100 monthly observations are used to initialize the process. As far as the error term is concerned, we assume $u_t \sim N(\mathbf{0}_{(21\times1)}, \Sigma_u)$, where $\Sigma_u$ has the same structure as $\Sigma_{\upsilon^*}$ in (12) with $\sigma_{LL} = 0.5$, $\sigma_{HH} = 1$ and $\sigma_{HL} = -0.1$.[20]

[20]We choose these values since they are close to the data we use in the empirical section to come. In particular, the sample variance of the low-frequency variable, an estimator of $\sigma_{LL}$, turns out to be 0.55, the mean of the sample variances of $x^{(20)}_{t-i/20}, i = 0, \ldots, 19$ is equal to 1.09 and approximates $\sigma_{HH}$ and, finally, the mean of the covariances between $y$ and $x^{(20)}_{t-i/20}, i = 0, \ldots, 19$ estimating $\sigma_{HL}$ is $-0.22$. We deviate from the exact values to achieve positive-definiteness of $\Sigma_u$. Note also that since $\sigma_{HL} \neq 0$ nowcasting causality is present (see Section

Similar to the lag length, we restrict the amount of factors (when considering CCA and PLS) to one in this paper and leave the analysis of higher factor dimensions for further research. The figures in the tables below represent the percentage amount of rejections at the 5% level.[21] All figures are based on 2,500 replications and are computed using GAUSS12.

### 4.1 Size

Table 1 contains the size results for the the common low-frequency VAR (LF-VAR hereafter), the unrestricted VAR (U-VAR hereafter), the VARs after reduced rank restrictions have been imposed and the Bayesian MF-VAR (B-MF-VAR hereafter) implemented via the auxiliary dummy variable approach. With respect to reduced rank restrictions, the factors are obtained via CCA, PLS or HAR-type restrictions.

Table 1: Size of Granger Non-Causality Tests

| S'le Size | LF-VAR | U-VAR | CCA | PLS | HAR | B-MF-VAR | Direction |
|---|---|---|---|---|---|---|---|
| T=50 | 0.062 | 0.13 | 0.202 | 0.346 | 0.067 | 0.05 | |
| T=250 | 0.059 | 0.067 | 0.294 | 0.349 | 0.052 | 0.058 | $X^{(m)}$ to $y$ |
| T=500 | 0.056 | 0.061 | 0.3 | 0.36 | 0.057 | 0.054 | |
| T=50 | 0.052 | 0.926 | 0.858 | 0.634 | 0.604 | 0.713 | |
| T=250 | 0.054 | 0.108 | 0.185 | 0.139 | 0.097 | 0.271 | $y$ to $X^{(m)}$ |
| T=500 | 0.045 | 0.066 | 0.124 | 0.105 | 0.063 | 0.146 | |
| T=50 | | 0.086 | 0.113 | 0.131 | 0.065 | 0.435 | $y$ to $X^{(m)}$ |
| T=250 | | 0.057 | 0.082 | 0.095 | 0.056 | 0.097 | |
| T=500 | | 0.041 | 0.066 | 0.082 | 0.04 | 0.061 | (Bonferroni) |

Note: The figures represent the percentage of rejections of the Wald test statistic in (22) for B-MF-VAR or the one in (8), properly adapted if necessary, for the remaining methods at the 5% level. Prior to estimation, the initial VAR has been transformed into a common low-frequency VAR as discussed in Section 3.3, left unrestricted as discussed in Section 3.4, reduced rank restrictions, with CCA-, PLS- or HAR-based factors, have been imposed as discussed in Section 3.1 or the VAR has been augmented to a Bayesian MF-VAR through the auxiliary dummy variable approach in Section 3.2. The number of factors (for CCA and PLS) and the lag length of the estimated VARs are equal to 1. The underlying DGP is found in (24), where the variance-covariance matrix of the error term is equal to $\Sigma_u$.

Let us focus on the top two blocks of Table 1. While the low-frequency VAR has size close to the nominal one, the unrestricted approach suffers from parameter proliferation for small $T$. Note that the former result is not surprising because the flat aggregation scheme underlying the low-frequency VAR is correct in this mixed-frequency white noise DGP. For large sample sizes the two approaches perform almost equally well. As far as reduced rank restrictions and testing the direction from $X^{(m)}$ to $y$ is concerned, computing the factors by CCA or PLS

2.4). If we set it equal to zero instead, nowcasting causality would have been absent. The results are, however, similar in both cases.

[21] Outcomes for the 10% and 1% levels are available upon request.

delivers size distortions, whereas the use of HAR-type factors leads to very good size results irrespective of $T$. With respect to testing for causality in the reverse direction, the tests are severely oversized for small $T$, whereas they perform more satisfactory (especially for HAR) as $T$ grows. Finally, the Bayesian MF-VAR yields optimal size results for the direction from $X^{(m)}$ to $y$ by construction as we choose $\lambda$ as to control the size of Granger non-causality tests in this direction (see below). For the reverse direction, however, a similar picture as for CCA or PLS arises, i.e., size distortions also for large $T$.

When testing for Granger non-causality from $y$ to $X^{(m)}$ in all approaches except the low-frequency VAR is the test oversized for small $T$, but yields better size results for a larger sample size. Note that we compute a joint test on $20p$ parameters in this case, which distorts our size results, especially for small $T$. In order to address this issue, we re-compute the rejection frequencies using the Bonferroni correction (Dunn, 1961) and report the results in the lowest block of Table 1. It turns out that the results improve for all sample sizes, yet only for the reduced rank restrictions with HAR-type factors (and the unrestricted VAR to a lesser degree) are the outcomes satisfactory also for small $T$.

As far as B-MF-VAR and the choice of $\lambda$ is concerned, closer inspection of the prior variances of the parameters of interest reveals that $Var[\gamma_{1,j}^{(k)}] < Var[\gamma_{i,1}^{(k)}]$ for $i, j = 2, \ldots, m+1$ and $k = 1, \ldots, p$ (see (13)). Consequently, for a given $m$ and $\lambda$ the parameters corresponding to Granger (non-)causality from $X^{(m)}$ to $y$ get shrunk more than the ones corresponding to the direction from $y$ to $X^{(m)}$. As the Bonferroni correction presents a way to lower size distortions in the latter case, we opted to set $\lambda$ as to control size in the former direction. Denoting the value of $\lambda$ that equates the corresponding actual and nominal size for a certain $T$ and $m$ by $\lambda(T, m)$, we obtain $\lambda(50, 20) \approx 3.12$, $\lambda(250, 20) \approx 4.8$ and $\lambda(500, 20) \approx 5.6$.[22]

### 4.2 Power

Table 2 reports size-adjusted power of the Granger non-causality tests for the various parameter reduction techniques and the unrestricted VAR. The top half of the table contains the outcomes for the first power DGP ($w_{j-1}^*(\psi)$ and $\gamma_{j,1}^*, j = 2, \ldots, 21$, possess non-zero means), while the bottom half does so for the second power DGP ($w_{j-1}^*(\psi)$ and $\gamma_{j,1}^*, j = 2, \ldots, 21$, possess means of zero). Within each half, the top two blocks correspond to the usual Wald statistics and the bottom block reports the results after using the Bonferroni correction.

Let us start by considering the first power DGP, for which no method is a priori expected to perform better or worse than the others. When testing for Granger non-causality from $X^{(m)}$ to $y$, all parameter reduction approaches lead to very good power results. For the reverse direction we obtain a similar pattern as for size, i.e., size-adjusted power is low for small $T$, yet improves quickly as $T$ grows. Although the results improve slightly when using the Bonferroni correction

[22] For $m = 4, 5, 7$ and 10 we obtain $\lambda(i, 4) = 1.3, 1.45, 1.6$, $\lambda(i, 5) = 1.55, 2, 2.8$, $\lambda(i, 7) = 1.8, 2.4, 3.7$ and $\lambda(i, 10) = 2.45, 2.6, 4.5$ for $i = 50, 250$ and 500, respectively. A detailed analysis of the relationship between $m, T$ and $\lambda$ as well as the potential introduction of an equation-dependent $\lambda$, say $\lambda_L$ and $\lambda_H$, to address the different evolvement of prior variances of the Granger (non)-causality determining parameters, are of great interest, but beyond the scope of this paper and thus left for future work.

Table 2: Size-adjusted Power of Granger Non-Causality Tests

| S'le Size | LF-VAR | U-VAR | CCA | PLS | HAR | B-MF-VAR | Direction |
|---|---|---|---|---|---|---|---|
| T=50 | 0.992 | 1 | 0.909 | 0.975 | 1 | 1 | |
| T=250 | 1 | 1 | 1 | 1 | 1 | 1 | $X^{(m)}$ to $y$ |
| T=500 | 1 | 1 | 1 | 1 | 1 | 1 | |
| T=50 | 0.182 | 0.063 | 0.004 | 0.098 | 0.116 | 0.172 | |
| T=250 | 0.866 | 0.88 | 0.857 | 0.908 | 0.935 | 0.931 | $y$ to $X^{(m)}$ |
| T=500 | 0.995 | 0.999 | 0.999 | 0.999 | 0.999 | 0.999 | |
| T=50 | | 0.125 | 0.181 | 0.181 | 0.25 | 0.094 | $y$ to $X^{(m)}$ |
| T=250 | | 0.97 | 0.976 | 0.969 | 0.982 | 0.98 | |
| T=500 | | 1 | 1 | 1 | 1 | 1 | (Bonferroni) |
| T=50 | 0.046 | 0.963 | 0.636 | 0.865 | 0.999 | 1 | |
| T=250 | 0.054 | 1 | 1 | 1 | 1 | 1 | $X^{(m)}$ to $y$ |
| T=500 | 0.064 | 1 | 1 | 1 | 1 | 1 | |
| T=50 | 0.058 | 0.054 | 0.016 | 0.057 | 0.065 | 0.144 | |
| T=250 | 0.049 | 0.35 | 0.304 | 0.42 | 0.411 | 0.376 | $y$ to $X^{(m)}$ |
| T=500 | 0.058 | 0.779 | 0.829 | 0.862 | 0.826 | 0.766 | |
| T=50 | | 0.075 | 0.071 | 0.063 | 0.098 | 0.076 | $y$ to $X^{(m)}$ |
| T=250 | | 0.472 | 0.528 | 0.489 | 0.514 | 0.519 | |
| T=500 | | 0.874 | 0.928 | 0.906 | 0.896 | 0.888 | (Bonferroni) |

Note: The underlying DGP is found in (25), where the top half of the Table reports the outcomes for the first power DGP, in which $w^*_{j-1}(\psi)$ and $\gamma^*_{j,1}, j = 2, \ldots, 21$, contain a non-zero mean, and where the bottom half gives the results for the second power DGP, in which they possess a mean of zero. For the rest see Table 1.

(except for B-MF-VAR), the general pattern remains the same.

Let us now turn to the second power DGP. Most importantly, the results for the low-frequency VAR show that aggregation of the high-frequency variable by averaging annihilates the causality features between the series for all sample sizes (size-adjusted power actually almost coincides with the nominal size of 5%). Put differently, the outcomes confirm that Granger causality is not invariant to temporal aggregation (Marcellino, 1999) if the causal effects get averaged out as in this case. For the other approaches a similar pattern as for the first power DGP emerges: Granger non-causality from $X^{(m)}$ to $y$ is usually rejected, whereby power is slightly lower for small $T$ in case of CCA and, to a lesser extent, PLS.[23] For the reverse direction, power is very small for $T = 50$ (whereby B-MF-VAR provides the least bad outcome and CCA the worst) and is satisfactory for $T = 500$. For $T = 250$ size-adjusted power equals about 30-40% for all approaches (except LF-VAR, of course) implying much smaller values than

[23]PLS has the disadvantage that the loadings corresponding to the first factor are bound to be larger than zero. With several negative coefficients in the underlying DGP, this could explain losses of power for the direction from $X^{(m)}$ to $y$. This problem would most likely vanish for a larger amount of factors, since negative loadings are possible from the second factor onwards.

in the top half of the table. Overall, it seems that reduced rank restrictions with HAR-type factors results in best size-adjusted power (except maybe for $T = 50$ and testing the direction from $y$ to $X^{(m)}$). Application of the Bonferroni correction yields similar outcomes.

A final comment on the results of the Monte Carlo study. Note that these results are specific to one particular example, in which $m = 20$, the lag length and number of factors are one and which is based on DGPs that are closely related to the restricted VAR in (11). It is well possible that the outcomes change for different situations, i.e., reduced rank restrictions using factors based on CCA and PLS or B-MF-VAR may perform better in other scenarios. In any case, both approaches are valuable contributions to the set of parameter reduction techniques that could be considered in practice as the next section shows.

# 5 Application

We apply the parameter reduction approaches described in Section 3 to a MF-VAR consisting of the monthly growth rate of the industrial production index (*ipi* hereafter), a measure of business cycle fluctuations, and the logarithm of daily bipower variation (*bv* hereafter), a measure of volatility. While the degree to which macroeconomic variables can help to predict volatility movements has been investigated widely in the literature (see Schwert, 1989b, Hamilton and Gang, 1996, or Engle and Rangel, 2008, among others), the reverse, i.e., whether the future path of the economy can be forecasted using return volatility, has been granted comparably few attention (examples are Schwert, 1989a, Mele, 2007, and Andreou et al., 2000). Instead of using an aggregate measure of volatility measure (Chauvet et al., 2013) such as, e.g., realized volatility or GARCH, we use daily bipower variation computed on 5-minute returns.[24] With the *bv*-series being available at a higher frequency than most indicators of business cycle fluctuations, we obtain the mixed-frequency framework analyzed in this paper.

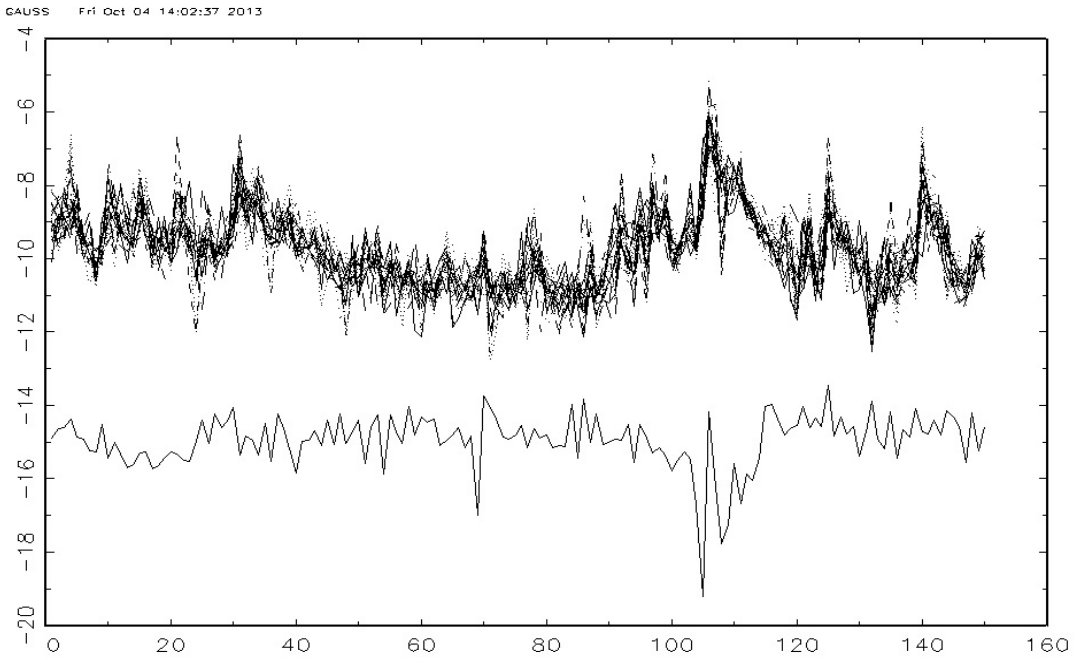
The sample covers the period from January 2000 to June 2012, yielding a sample size of $T = 150$, and $m = 20$. Figure 2 plots the data. We take $m = 20$ as it is the maximum amount of working days that is available in *every* month throughout the sample we deal with. Consider the following table illustrating the notation for our month/working day-example, where $t = 2012M06$, i.e., June 2012.

In this particular case, it means that although there are 21 open days in June 2012, we do not consider the first day, i.e., June 1. For May 2012 we do not consider the first three days. An alternative (balanced) strategy would have been to take the maximum number of days in a *particular* month (i.e., 23, usually in July, August or October) and to create additional values for non-existing days in other months whenever necessary. As far as the treatment of daily data is concerned we have also taken $bv_t^{(20)} = bv_{t-1/20}^{(20)}$ when there are no quotations for $bv_t^{(20)}$. Because $m = 20$ we have $BV_t^{(20)} = (bv_t^{(m)}, bv_{t-1/m}^{(m)}, \ldots, bv_{t-(m-1)/m}^{(m)})'$.

[24] Data are obtained from the database in Heber, Gerd, Asger Lunde, Neil Shephard and Kevin Sheppard (2009), Oxford-Man Institute's realized library (version 0.2), Oxford-Man Institute, University of Oxford. Note that *bv* is considered as a measure that is less sensitive to jumps compared to realized volatility.

| Notation | Meaning |
|---|---|
| $ipi_t$ | $ipi$ in June, 2012 |
| No working day | $bv$ on June 30, 2012 |
| $bv_t^{(20)}$ | $bv$ on June 29, 2012 |
| $bv_{t-1/20}^{(20)}$ | $bv$ on June 28, 2012 |
| $\vdots$ | $\vdots$ |
| $bv_{t-19/20}^{(20)}$ | $bv$ on June 4, 2012 |
| $ipi_{t-1}$ | $ipi$ in May, 2012 |
| $bv_{t-1}^{(20)}$ | $bv$ on May 31, 2012 |
| $bv_{t-1-1/20}^{(20)}$ | $bv$ on May 30, 2012 |

Figure 2: Growth Rate of Industrial Production Index and the logarithm of Bipower Variation




Note: This figure shows the monthly growth rate of the industrial production index (lower line), i.e., $ipi$, and the logarithm of daily bipower variation (top lines), i.e., $BV^{(20)}$, for the time period from January 2000 to June 2012

Table 3 contains the outcomes of Granger non-causality tests for all approaches discussed in Section 3. Note that a length of $p = 1$ and 2 is considered and that the figures represent $p$-values. For all approaches, except the one described in Section 3.2, we also consider a heteroscedasticity consistent variant of (8) by computing a robust estimator of $\Omega$ (see Ravikumar et al., 2000) to

account for the potential presence of a time varying multivariate process:

$$\hat{\Omega}_R = T((W'W)^{-1} \otimes I_{m+1})\hat{S}_0((W'W)^{-1} \otimes I_{m+1}), \tag{26}$$

where

$$\hat{S}_0 = \frac{1}{T}\sum_{t=1}^{T}(W_t \otimes \hat{u}_t)'(W_t \otimes \hat{u}_t).$$

Table 3: Testing for Granger Non-Causality between $\Delta lnIPI$ and $lnBV$

| | Factors | $p=1$ | | $p=2$ | | Direction |
|---|---|---|---|---|---|---|
| | | Wald | White | Wald | White | |
| LF-VAR | | $< 0.001$ | 0.002 | $< 0.001$ | 0.021 | *bv* to *ipi* |
| | | 0.001 | 0.015 | 0.004 | 0.05 | *ipi* to *bv* |
| U-VAR | | 0.128 | 0.069 | 0.028 | 0.005 | *bv* to *ipi* |
| | | 0.127 | 0.005 | 0.04 | $< 0.001$ | *ipi* to *bv* |
| CCA | $r=1$ | 0.02 | 0.039 | 0.064 | 0.118 | |
| | $r=2$ | 0.05 | 0.09 | 0.004 | 0.031 | *bv* to *ipi* |
| | $r=3$ | 0.003 | 0.009 | 0.01 | 0.061 | |
| | $r=1$ | 0.09 | $< 0.001$ | 0.011 | $< 0.001$ | |
| | $r=2$ | 0.135 | 0.006 | 0.002 | $< 0.001$ | *ipi* to *bv* |
| | $r=3$ | 0.213 | 0.012 | 0.002 | $< 0.001$ | |
| PLS | $r=1$ | $< 0.001$ | 0.002 | 0.001 | 0.013 | |
| | $r=2$ | 0.002 | 0.006 | $< 0.001$ | 0.011 | *bv* to *ipi* |
| | $r=3$ | $< 0.01$ | 0.003 | 0.001 | 0.028 | |
| | $r=1$ | 0.057 | $< 0.001$ | 0.015 | $< 0.001$ | |
| | $r=2$ | 0.063 | $< 0.001$ | 0.011 | $< 0.001$ | *ipi* to *bv* |
| | $r=3$ | 0.055 | $< 0.001$ | 0.011 | $< 0.001$ | |
| HAR | $r=3$ | 0.007 | 0.019 | 0.011 | 0.072 | *bv* to *ipi* |
| | $r=3$ | 0.081 | $< 0.001$ | 0.016 | $< 0.001$ | *ipi* to *bv* |
| B-MF-VAR | | 0.006 | | 0.001 | | *bv* to *ipi* |
| | | 0.022 | | 0.003 | | *ipi* to *bv* |

Note: The figures represent the $p$-values associated with the Wald test statistic in (22) for B-MF-VAR, the one in (8), properly adapted if necessary, for the remaining methods, labeled 'Wald', and the one derived from (26), labeled 'White'. The models considered are the common low-frequency VAR as discussed in Section 3.3, the unrestricted VAR as discussed in Section 3.4, the VAR with reduced rank restrictions, using CCA-, PLS- or HAR-based factors, as discussed in Section 3.1 and the B-MF-VAR implemented via the auxiliary dummy variable approach as discussed in Section 3.2. The lag order $p$ is equal to 1 or 2 and the number of factors (for CCA and PLS) is equal to $1, 2$ or 3.

The top part of the table shows the outcomes after the initial MF-VAR has been transformed into a common low-frequency VAR by temporally aggregating the high-frequency variable: $bv_t^{(aggr)} = ln(\frac{1}{20}\sum_{i=0}^{19} exp(bv_{t-i/20}^{(20)}))$. Clearly, the figures suggest bi-directional Granger causality.

The second block from the top presents the outcomes when leaving the MF-VAR unrestricted. The outcomes partly differ if we consider the standard Wald test or its robust version. Based on the latter we usually reject Granger non-causality in both directions. Based on the former, however, we only do so for $p = 2$; for $p = 1$ no Granger causality is found in any direction. However, although not displayed here,[25] the Bonferroni correction points towards Granger causality from *ipi* to $BV^{(20)}$. Hence, apart from the Wald test for $p = 1$, we conclude again bi-directional Granger causality.

As far as parameter reduction by reduced rank restrictions is concerned, canonical correlation tests[26] favor $r = 2$ factors whereas the approach of Cubadda and Hecq (2011) selects $r = 1$. Consequently, we present the results for $1, 2$ as well as 3 factors, i.e., $r = 1, 2, 3$. Of course, for the HAR-type factors, the number of factors is $r = 3p$. The third block of Table 3 displays the results when the factors are computed using CCA, whereas the subsequent two blocks show the outcomes when they are based on PLS and HAR-type restrictions, respectively.

For all three approaches based on reduced rank restrictions, Granger causality from $BV^{(20)}$ to *ipi* is detected for almost all combinations of $p$, $f$, 'Wald' and 'White'. For the reverse direction and $p = 1$, 'White' also leads to a rejection of Granger non-causality, whereas 'Wald' either suggests a rejection of Granger non-causality at roughly 6% (PLS) and 9% (HAR or CCA with $r = 1$) or the absence of Granger causality altogether (CCA with $r > 1$). For a lag length of 2, both tests imply a rejection of the null hypothesis in almost cases. These results are generally confirmed by the use of the Bonferroni correction.

Finally, the bottom part of Table 3 contains the results for the Bayesian MF-VAR implemented through the auxiliary dummy variable approach.[27] Here, the figures clearly suggest bi-directional Granger causality between the two variables of interest. Again, this conclusion is confirmed by the use of the Bonferroni correction.

Overall, the outcomes point towards bi-directional Granger causality between uncertainty in financial markets, as measured by bipower variation, and business cycle fluctuations, as measured by growth in the industrial production index.

[25] Results available upon request.

[26] For *i.i.d.* normally distributed disturbances, the likelihood ratio test statistic for the null hypothesis that there exist at most $r = 20-s$ common feature combinations within $X_t^{(m)}$ is given by $\zeta_{LR(s)} = -T\sum_{j=1}^{s} \ln(1-\hat{\lambda}_j)$, where $\hat{\lambda}_j$ denotes the $s = 1, \ldots, 20$, smallest eigenvalues associated with $\hat{\Sigma}_{\tilde{Z}\tilde{Z}}^{-1/2}\hat{\Sigma}_{\tilde{Z}\underline{\tilde{X}}^{(m)}}\hat{\Sigma}^{-1}_{\underline{\tilde{X}}^{(m)}\underline{\tilde{X}}^{(m)}}\hat{\Sigma}_{\underline{\tilde{X}}^{(m)}\tilde{Z}}\hat{\Sigma}_{\tilde{Z}\tilde{Z}}^{-1/2}$, i.e., the matrix in (7) with the roles of $\tilde{Z}$ and $\underline{\tilde{X}}^{(m)}$ interchanged.

[27] Note that we do not compute a heteroscedasticity consistent variant of (22) for this case.

# 6 Conclusion

Using the example of analyzing the link between uncertainty in financial markets and economic fluctuations, we investigated Granger non-causality testing in a MF-VAR, where the mismatch between the sampling frequencies of the variables under consideration is large. Indeed, the observable MF-VAR by Ghysels (2012) quickly becomes very large when the mismatch between sampling frequencies grows, causing estimation and inference to be potentially problematic due to parameter proliferation.

To potentially avoid this issue we discussed two parameter reduction techniques in detail, reduced rank restrictions and a Bayesian MF-VAR approach. The former achieves a reduction in parameters by looking for a reduced rank structure within the coefficient matrix pertaining to the high-frequency variables (Carriero et al., 2011). The corresponding observable high-frequency factors were computed either by CCA or PLS, or imposed as HAR-type factors. The Bayesian MF-VAR was implemented through the mixed-frequency extension of the approach in Banbura et al. (2010). Using a Minnesota prior for the coefficients (Ghysels, 2012), it was shown that augmenting the unrestricted MF-VAR by a set of auxiliary dummy variables and estimating the resulting system by GLS is equivalent to considering the mean of the corresponding posterior distribution. Finally, we also transformed the MF-VAR into a common low-frequency one by a simple aggregation scheme. Subsequently, we compared these approaches with each other and an unrestricted VAR in terms of their Granger non-causality testing behavior using Monte Carlo simulations.

In our particular case we found that while the low-frequency VAR may lead to very poor results, the unrestricted VAR resulted in size distortions for small sample sizes due to parameter proliferation. Reduced rank restrictions based on CCA and PLS yielded relatively poor size and power outcomes, whereas the ones based on HAR-type factors lead to very good size and the overall best power results. Despite power being nearly as high as for the latter approach, the Bayesian MF-VAR approach lead to size distortions when testing for causality in the direction from low- to high-frequency variables. Size for the reverse direction was nearly optimal by construction. Reduced rank restrictions and Bayesian MF-VARs are, in any case, valuable contributions to the available set of techniques to address the proliferation of parameters in large MF-VAR models.

An application investigating the presence of a causal link between business cycle fluctuations and volatility illustrated the practical usefulness of these approaches. In most cases, bi-directional Granger causality between uncertainty in financial markets and the growth rate of the industrial production index was detected.

A number of avenues should be pursued in future work: A formal comparison of the canonical correlation approach with alternative options; An analysis of the relationship between the tightness of the prior distribution around the restricted MF-VAR, the sample size and the frequency discrepancy between the variables ($\lambda$, $T$ and $m$); A potential generalization of the the paper's findings to different data generating processes, especially a common high-frequency

DGP;[28] The consideration of higher lag orders and a larger number of factors for CCA and PLS; Reduced-rank Bayesian VARs; The assessment of the various models' forecasting performances; An extension of the Monte Carlo study to investigate the presence of causal chains and their impact on Granger causality tests.

Nevertheless, the approaches presented in this paper indicate that many empirical research questions, where the sampling frequencies of the variables involved differ considerably and where a link between these variables is of interest, can be addressed in the future. More importantly, these problems can be tackled without the need to temporally aggregate the high-frequency variable, thus avoiding the risk to create non-existing or conceal existing causality patterns.

[28] It would be useful to explore whether causality patterns in a common high-frequency VAR can be preserved in the MF-VARs after the various parameter reduction approaches have been applied (Ghysels et al., 2013).

## References


Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *The Annals of Mathematical Statistics*, 22(3):327–351.

Andreou, E., Osborn, D. R., and Sensier, M. (2000). A comparison of the statistical properties of financial variables in the usa, uk and germany over the business cycle. *Manchester School*, 68(4):396–418.

Banbura, M., Giannone, D., and Reichlin, L. (2010). Large bayesian vector auto regressions. *Journal of Applied Econometrics*, 25(1):71–92.

Carriero, A., Kapetanios, G., and Marcellino, M. (2011). Forecasting large datasets with bayesian reduced rank multivariate models. *Journal of Applied Econometrics*, 26(5):735–761.

Chauvet, M., Senyuz, Z., and Yoldas, E. (2013). What does financial volatility tell us about macroeconomic fluctuations? Technical report.

Clements, M. and Galvão, A. B. (2008). Macroeconomic forecasting with mixed-frequency data. *Journal of Business & Economic Statistics*, 26:546–554.

Clements, M. and Galvão, A. B. (2009). Forecasting u.s. output growth using leading indicators: An appraisal using midas models. *Journal of Applied Econometrics*, 24(7):1187–1206.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Cox, D. R., Gudmundsson, G., Lindgren, G., Bondesson, L., Harsaae, E., Laake, P., Juselius, K., and Lauritzen, S. L. (1981). Statistical analysis of time series: Some recent developments [with discussion and reply]. *Scandinavian Journal of Statistics*, 8(2):93–115.

Cubadda, G. and Guardabascio, B. (2012). A medium-n approach to macroeconomic forecasting. *Economic Modelling*, 29(4):1099–1105.

Cubadda, G. and Hecq, A. (2011). Testing for common autocorrelation in data rich environments. *Journal of Forecasting*, 30(3):325–335.

Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternatives. *Biometrika*, 74(1):33–43.

De Mol, C., Giannone, D., and Reichlin, L. (2008). Forecasting using a large number of predictors: Is bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics*, 146(2):318–328.

Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association*, 56(293):52–64.

Engle, R. F. and Rangel, J. G. (2008). The spline-garch model for low-frequency volatility and its global macroeconomic causes. *Review of Financial Studies*, 21(3):1187–1222.

Foroni, C. and Marcellino, M. (2013). Mixed frequency structural models: estimation, and policy analysis. Technical report.

Foroni, C., Marcellino, M., and Schumacher, C. (2012). U-midas: Midas regressions with unrestricted lag polynomials. Technical report.

Forsberg, L. and Ghysels, E. (2007). Why do absolute returns predict volatility so well? *Journal of Financial Econometrics*, 5(1):31–67.

Ghysels, E. (2012). Macroeconomics and the reality of mixed frequency data. Technical report.

Ghysels, E. and Miller, J. I. (2013). Testing for Cointegration with Temporally Aggregated and Mixed-frequency Time Series. Technical report.

Ghysels, E., Motegi, K., and Hill, J. (2013). Granger causality tests with mixed data frequencies. Discussion Paper.

Ghysels, E., Santa-Clara, P., and Valkanov, R. (2004). The midas touch: Mixed data sampling regression models. CIRANO Working Papers 2004s-20, CIRANO.

Ghysels, E., Sinko, A., and Valkanov, R. (2007). Midas regressions: Further results and new directions. *Econometric Reviews*, 26(1):53–90.

Ghysels, E. and Valkanov, R. (2012). *Forecasting volatility with MIDAS*, chapter 16, pages 383–401. Handbook of Volatility Models and Their Applications. John Wiley & Sons, Inc., Hoboken, NJ, USA.

Gianetto, Q. G. and Raïssi, H. (2012). Testing instantaneous causality in presence of non constant unconditional variance. Technical report.

Götz, T. B. and Hecq, A. (2014). Nowcasting causality in mixed frequency vector autoregressive models. *Economics Letters*, 122(1):74–78.

Götz, T. B., Hecq, A., and Urbain, J.-P. (2013). *Testing for common cycles in non-stationary VARs with varied frecquency data*, volume 32, pages 361–393. Emerald Group Publishing Limited.

Götz, T. B., Hecq, A., and Urbain, J.-P. (2014). Forecasting mixed frequency time series with ecm-midas models. *Journal of Forecasting*, 33:198–213.

Hamilton, J. D. (1994). *Time series analysis*. Princeton Univ. Press, Princeton, NJ.

Hamilton, J. D. and Gang, L. (1996). Stock market volatility and the business cycle. *Journal of Applied Econometrics*, 11(5):573–93.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica*, 64(2):413–30.

Helland, I. S. and Almoy, T. (1994). Comparison of prediction methods when only a few components are relevant. *Journal of the American Statistical Association*, 89(426):583–591.

Kadiyala, K. R. and Karlsson, S. (1997). Numerical methods for estimation and inference in bayeisan var-models. *Journal of Applied Econometrics*, 12(2):99–132.

Litterman, R. B. (1986). Forecasting with bayesian vector autoregressions - five years of experience. *Journal of Business & Economic Statistics*, 4(1):25–38.

Lütkepohl, H. (1993). *Testing for causation between two variables in higher dimensional VAR models*, pages 75–91. Studies in Applied Econometrics. Springer-Verlag, Heidelberg.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Publishing Company, Incorporated.

Marcellino, M. (1999). Some consequences of temporal aggregation in empirical analysis. *Journal of Business & Economic Statistics*, 17(1):129–136.

Mele, A. (2007). Asymmetric stock market volatility and the cyclical behavior of expected returns. *Journal of Financial Economics*, 86(2):446–478.

Miller, J. I. (2011). Conditionally efficient estimation of long-run relationships using mixed-frequency time series. Working Papers 1103, Department of Economics, University of Missouri.

Miller, J. I. (2012). Mixed-frequency cointegrating regressions with parsimonious distributed lag structures. Technical report.

Ravikumar, B., Ray, S., and Savin, N. E. (2000). Robust wald tests in sur systems with adding-up restrictions. *Econometrica*, 68(3):715–720.

Schorfheide, F. and Song, D. (2012). Real-time forecasting with a mixed-frequency VAR. Technical report.

Schwert, G. W. (1989a). Business cycles, financial crises, and stock volatility. *Carnegie-Rochester Conference Series on Public Policy*, 31(1):83–125.

Schwert, G. W. (1989b). Why does stock market volatility change over time? *Journal of Finance*, 44(5):1115–53.

Sekhposyan, T., McCracken, M. W., and Owyang, M. T. (2014). Real-time forecasting with a large, mixed frequency, bayesian var. Working Paper.

Silvestrini, A. and Veredas, D. (2008). Temporal aggregation of univariate and multivariate time series models: A survey. *Journal of Economic Surveys*, 22(3):458–497.

Sims, C. A. (1971). Discrete approximations to continuous time distributed lags in econometrics. *Econometrica*, 39(3):545–563.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society (Series B)*, 58:267–288.

Vahid, F. and Engle, R. F. (1993). Common trends and common cycles. *Journal of Applied Econometrics*, 8(4):341–360.

Waggoner, D. F. and Zha, T. (2003). A Gibbs sampler for structural vector autoregressions. *Journal of Economic Dynamics and Control*, 28(2):349–366.

# A Construction of the selection matrix $S$

Let us investigate the variance of $vec(B^*)|\Sigma$ more closely. Noting that we have to choose $Z_0$, or $Z_0^*$ in (15), in such a way as to let the variances of the corresponding coefficients coincide with the prior variances in (13), it turns out that we need to set

$$Var[vec(B^*)|\Sigma] = \begin{pmatrix} \sigma_L^2\Omega_0^{(2)} & \mathbf{0}_{n\times n} & \dots & \dots & \mathbf{0}_{n\times n} \\ \mathbf{0}_{n\times n} & \sigma_H^2\Omega_0^{(2)} & \mathbf{0}_{n\times n} & \dots & \mathbf{0}_{n\times n} \\ \vdots & \mathbf{0}_{n\times n} & \sigma_H^2\Omega_0^{(3)} & \dots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \mathbf{0}_{n\times n} \\ \mathbf{0}_{n\times n} & \mathbf{0}_{n\times n} & \dots & \mathbf{0}_{n\times n} & \sigma_H^2\Omega_0^{(m+1)} \end{pmatrix},$$

where

$$\Omega_0^{(i)} = diag( \quad \tfrac{\lambda^2}{(m+2-i)^2\sigma_L^2}, \tfrac{\lambda^2}{(m+2-i)^2\sigma_H^2}, \tfrac{\lambda^2}{(m+3-i)^2\sigma_H^2}, \dots, \tfrac{\lambda^2}{(m+m+1-i)^2\sigma_H^2}, \dots, \\ \dots, \tfrac{\lambda^2}{(pm+2-i)^2\sigma_L^2}, \tfrac{\lambda^2}{(pm+2-i)^2\sigma_H^2}, \dots, \tfrac{\lambda^2}{(pm+m+1-i)^2\sigma_H^2}, \tfrac{1}{\epsilon^2})$$

for $i = 2, \dots, m+1$.

Hence, unlike in the common-frequency case, where $\Omega_0^{(i)} = \Omega_0 \ \forall i$ (Banbura et al., 2010), the set of variances changes due to the stacked nature of the vector $Z_t$ and the specific lag structure for each coefficient (see the variances in (13)). Let us form an auxiliary matrix $\Omega_0^{aux}$, which contains the union of all elements in $\Omega_0^{(2)}, \dots, \Omega_0^{(m+1)}$. Each matrix $\Omega_0^{(i)}$ contains $p+1$ new elements compared to $\Omega_0^{(i-1)}$ for $i > 2$. As there are $n$ elements in $\Omega_0^{(2)}$, we end up with a dimension of $m(2p+1) = n^{aux}$ for the square matrix $\Omega_0^{aux}$:

$$\Omega_0^{aux} = diag( \quad \tfrac{\lambda^2}{1^2\sigma_L^2}, \tfrac{\lambda^2}{1^2\sigma_H^2}, \dots, \tfrac{\lambda^2}{m^2\sigma_L^2}, \tfrac{\lambda^2}{m^2\sigma_H^2}, \tfrac{\lambda^2}{(m+1)^2\sigma_L^2}, \tfrac{\lambda^2}{(m+1)^2\sigma_H^2}, \dots, \tfrac{\lambda^2}{(2m)^2\sigma_H^2}, \\ \dots, \dots, \tfrac{\lambda^2}{(pm)^2\sigma_L^2}, \tfrac{\lambda^2}{(pm)^2\sigma_H^2}, \tfrac{\lambda^2}{(pm+1)^2\sigma_H^2}, \dots, \tfrac{\lambda^2}{(pm+m-1)^2\sigma_H^2}, \tfrac{1}{\epsilon^2}).$$

All that remains is to define a $[(m+1)n^{aux} \times (m+1)n]$-dimensional selection matrix $S$ such that $S'(\Sigma_d \otimes \Omega_0^{aux})S = Var[vec(B^*)|\Sigma]$ . Note that from $\Omega_0^{aux}$ we can then derive $Z_0^{aux} = X_d$ by using $\Omega_0^{aux} = Z_0^{aux'} Z_0^{aux}$.

Let us denote by $\underline{1}_i$ an $n^{aux}$-dimensional column vector with a one in row $i$ and zeros elsewhere. It turns out that $S$ is a block-diagonal matrix, i.e., $S = diag(S_1, S_2, \dots, S_{m+1})$, where each off-diagonal block is $\mathbf{0}_{n^{aux}\times n}$. Each $S_j, j = 1, \dots, m+1$, can be described as follows:

$$S_j = (S_j^1, S_j^2, \underline{1}_{n^{aux}}) \text{ for } j = 2, \dots, m+1,$$

where

$$S_j^1 = ( \quad \underline{1}_{2m-(2j-3)}, \underline{1}_{2m-(2j-4)}, \underline{1}_{2m-(2j-6)}, \dots, \underline{1}_{4m-(2j-2)}, \underline{1}_{4m-(2j-3)}, \dots, \\ \underline{1}_{6m-(2j-2)}, \dots, \dots, \underline{1}_{2pm-(2j-3)}, \underline{1}_{2pm-(2j-4)}, \dots, \underline{1}_{2pm}),$$

$$S_j^2 = \begin{cases} \emptyset & \text{for } j = m+1 \\ (\underline{1}_{2pm+1}, \underline{1}_{2pm+2}, \ldots, \underline{1}_{2pm+m-(j-1)}) & \text{else} \end{cases}$$

and

$$S_1 = S_2.$$

Each of the indices in $S_j$ reveals which row element of the $j^{th}$ column of $B_0^{aux}$ gets selected by $S$. Put differently, the indices that are missing in $S_j$ correspond to elements of $B_0^{aux}$ (in column $j$) that will not get chosen by $S$. This implies that the values of these elements do not play a role for the computation of $vec(B_0^*)$. Consequently, when constructing $B_0^{aux}$, and subsequently also $Y_d$, we assign these elements a value of zero for simplicity.