DEUTSCHE
BUNDESBANK
EUROSYSTEM

Discussion Paper

Deutsche Bundesbank
No 45/2015

Testing for Granger causality in large mixed-frequency VARs

Thomas B. Götz
(Deutsche Bundesbank)

Alain Hecq
(Maastricht University)

Stephan Smeekes
(Maastricht University)

Deutsche Bundesbank, Wilhelm-Epstein-Straße 14, 60431 Frankfurt am Main,
Postfach 10 06 02, 60006 Frankfurt am Main

Tel +49 69 9566-0

Please address all orders in writing to: Deutsche Bundesbank,
Press and Public Relations Division, at the above address or via fax +49 69 9566-3077

Internet http://www.bundesbank.de

ISBN 978–3–95729–217–9 (Printversion)
ISBN 978–3–95729–218–6 (Internetversion)

Non-technical summary

Research Question

When dealing with time series sampled at various frequencies it has become common practice to directly incorporate high-frequency information into the econom(etr)ic model at hand. These specifications were first restricted to the single regression case; with the development of the (stacked) mixed-frequency vector autoregressive (MF-VAR) system (Ghysels, 2015) it is now possible to treat all series similarly and investigate causal effects between them. However, if the difference in frequencies between the series involved is large (as, e.g., in a month/working day scenario), estimation accuracy of the system coefficients is exacerbated, implying the detection of causal effects to be potentially inaccurate. To overcome this issue various parameter reduction techniques are introduced and analyzed. These methods are then evaluated in terms of their ability to detect causality patterns between the series under consideration in the resulting restricted model.

Contribution

Two parameter reduction techniques are discussed in detail: three reduced rank regression (RRR) model variants and a Bayesian MF-VAR. Using a Monte Carlo experiment both approaches are compared in terms of their Granger causality testing behavior with an unrestricted VAR, a (time-aggregated) low-frequency VAR and the *max*-test (Ghysels, Motegi, and Hill, 2015a). To further enhance their finite sample properties we develop and evaluate (whenever possible) two bootstrap variants of these tests. Finally, the methods are applied to U.S. data by investigating channels of causality between the monthly growth rate of the industrial production index (IPI) and daily bipower variation (BV) of the S&P500 index.

Results

We find that, depending on the direction of causality under consideration, a different set of tests results in the best Granger non-causality testing behavior. For the direction from the high- to the low-frequency series, standard testing within the Bayesian MF-VAR, the *max*-test and the restricted bootstrap version of the Wald test in two RRR model versions performs best. For the reverse direction, the unrestricted bootstrap variants of the Bonferroni-corrected Wald tests within the unrestricted VAR and the RRR models dominate. As far as our application is concerned, Granger causality from BV to IPI-growth is clearly supported by the data; evidence for causality in the reverse direction, however, only comes from a subset of tests.

Nichttechnische Zusammenfassung

Fragestellung

Bei der Handhabe unterschiedlich frequenter Zeitreihen ist es nunmehr üblich hochfrequente Informationen direkt in das ökonom(etr)ische Modell enfließen zu lassen. Waren diese Spezifikationen zunächst auf univariate Regressionen beschränkt, so ist es durch die Entwicklung des ("gestapelten") gemischtfrequenten vektor-autoregressiven (MF-VAR) Systems (Ghysels, 2015) nun möglich alle Variablen gleichermaßen zu behandeln und Kausalzusammenhänge zwischen ihnen zu untersuchen. Sollte der Frequenzunterschied jedoch groß sein (wie z.B. in einem Monat/Arbeitstage Szenario), verschlechtert sich die Schätzgenauigkeit der System-Koeffizienten, wodurch die Erfassung kausaler Effekte ungenau werden könnte. Um dieses Problem zu umgehen werden Techniken zur Parameter-Reduzierung vorgestellt und analysiert. Diese Methoden werden dann anhand ihrer Fähigkeit, Kausalzusammenhänge zwischen den entsprechenden Variablen zu erfassen, beurteilt.

Beitrag

Zwei Techniken zur Parameter-Reduzierung werden detailliert diskutiert: Drei Varianten einer Reduzierter-Rang-Regression (RRR) und ein Bayesianisches MF-VAR. Mittels einer Monte Carlo Analyse werden beide Ansätze anhand ihres Granger Kausalitätstestverhalten mit einem unrestriktierten VAR, einem (zeit-aggregierten) niedrigfrequenten VAR und dem *max*-Test (Ghysels et al., 2015a) verglichen. Um ihre Eigenschaften weiter zu verbessern, entwickeln und bewerten wir (wann immer möglich) zwei Bootstrap-Varianten dieser Tests. Als Anwendung untersuchen wir die Kausalzusammenhänge zwischen der Monatswachstumsrate des U.S.-Industrieproduktionsindex (IPI) und der täglichen "bipower" Variation (BV) des S&500 Aktienindex.

Ergebnisse

Wir stellen fest, dass, je nach betrachteter Kausalitäts-Richtung, eine unterschiedliche Gruppe von Tests zum jeweils besten Verhalten führt. Für einen kausalen Effekt von der hoch- zur niedrigfrequenten Reihe erweisen sich Standard-Tests innerhalb des Bayesianischen MF-VARs, der *max*-Test und die restriktierte Bootstrap-Version des Wald Tests in zwei RRR-Varianten als am besten. Für die entgegensetzte Richtung dominieren die unrestriktierten Bootstrap-Varianten der Bonferroni-korrigierten Wald Tests innerhalb des unrestriktierten VARs und der RRR Modelle. In unserer Anwendung wird Granger Kausalität von BV zu IPI-Wachstum klar von den Daten unterstützt; Beweise für Kausalität in die entgegengesetzte Richtung kommen allerdings nur von einer Teilmenge der Tests.

Bundesbank Discussion Paper No 45/2015

Testing for Granger Causality in Large Mixed-Frequency VARs*

Thomas B. Götz

Deutsche Bundesbank

Alain Hecq Stephan Smeekes

Maastricht University

Abstract

We analyze Granger causality testing in a mixed-frequency VAR, where the difference in sampling frequencies of the variables is large. 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 several tests based on reduced rank restrictions, and implement bootstrap versions to account for the uncertainty when estimating factors and to improve the finite sample properties of these tests. We also consider a Bayesian VAR that we carefully extend to the presence of mixed frequencies. We compare these methods to an aggregated model, the *max*-test approach introduced by Ghysels et al. (2015a) as well as to the unrestricted VAR using Monte Carlo simulations. The techniques are illustrated in an empirical application involving daily realized volatility and monthly business cycle fluctuations.

Keywords: Granger Causality, Mixed Frequency VAR, Bayesian VAR, Reduced Rank Model, Bootstrap Test

JEL classification: C11, C12, C32.

*Contact address: Thomas B. Götz, Deutsche Bundesbank, Macroeconomic Analysis and Projection Division, Wilhelm-Epstein-Strasse 14, 60431 Frankfurt am Main. Email: thomas.goetz@bundesbank.de. The authors particularly thank Eric Ghysels for many fruitful discussions, comments and suggestions. Moreover, we thank two anonymous referees, Daniela Osterrieder, Lenard Lieb, Jörg Breitung, Roman Liesenfeld, Jan-Oliver Menz, Klemens Hauzenberger, Martin Mandler and participants of the various conferences and workshops the paper was presented at. This paper supersedes both the mimeo Chauvet, Götz and Hecq (2014), "Realized volatility and business cycle fluctuations: a mixed-frequency VAR approach" and the discussion paper Götz and Hecq (2014), "Testing for Granger causality in large mixed-frequency VARs", Maastricht University, GSBE discussion paper. Discussion Papers represent the authors' personal opinions and do not necessarily reflect the views of the Deutsche Bundesbank or its staff.

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, Sinko, and Valkanov, 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, Santa-Clara, and Valkanov, 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 (2015) for stationary series and the extension of Götz, Hecq, and Urbain (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, Motegi, and Hill (2015b) define causality in such a mixed-frequency VAR and develop a corresponding test statistic. Decent size and power properties of their test, however, are dependent on a relatively small difference in sampling frequencies of the variables involved. Indeed, if the number of high-frequency observations within a low-frequency period is large, size distortions and losses of power may be expected.

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 as, e.g., in a month/working day-example. To avoid the proliferation of parameters we consider two parameter reduction techniques: reduced rank restrictions and a Bayesian mixed-frequency VAR. Both approaches are then compared to (i) temporally aggregating the high-frequency variable (Breitung and Swanson, 2002), (ii) the *max*-test, independently and simultaneously developed by Ghysels et al. (2015a), and (iii) the unrestricted approach.

With respect to reduced rank regressions, the factors are typically not observable and must be estimated. This will obviously affect the distribution of the Wald tests to detect directions of Granger causalities. Consequently, one important contribution of this paper is the introduction of bootstrap versions of these tests (also for the unrestricted VAR), which have correct size even for large VARs and a small sample size.

As far as the Bayesian VAR is concerned, we show how to adapt the analysis to the presence of mixed frequencies. We do so by extending the dummy observation approach of Banbura, Giannone, and Reichlin (2010) to a mixed-frequency setting.[1] Importantly, due to stacking the high-frequency variables in the mixed-frequency VAR (Ghysels, 2015), their approach cannot be applied directly such that a properly adapted choice of auxiliary dummy variables corresponding to the prior moments is required. As these insights transfer beyond Granger causality testing, the adaption of Bayesian methods to mixed-frequency time series marks a significant contribution to the literature by itself.

[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' henceforth.

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. 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 illustrates the results. Section 6 concludes.

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 $x_{t-i/m}^{(m)}$ are the high-frequency variables with m high-frequency observations per low-frequency period t. Throughout this paper 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 ($x_{t-(m-1)/m}^{(m)}$) until the end ($x_t^{(m)}$ with $i = 0$). These notational conventions have become standard in the mixed-frequency literature and are similar to the ones in Götz, Hecq, and Urbain (2014), Clements and Galvão (2008, 2009) or Miller (2012).

Furthermore, let $\underline{W_t} = (W_{t-1}', W_{t-2}', \ldots, W_{t-p}')'$ denote the last p low-frequency lags of any process W stacked. 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.

Remark 1. Extensions towards representations of higher dimensional multivariate systems as in Ghysels et al. (2015b) 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. (2015b) 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.

2.2 MF-VARs

Considering each high-frequency variable such that

$$X_t^{(m)} = (x_t^{(m)}, x_{t-1/m}^{(m)}, \ldots, x_{t-(m-1)/m}^{(m)})', \tag{1}$$

[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, 2015).

a dynamic structural equations model for the stationary multivariate process $Z_t = (y_t, X_t^{(m)\prime})'$ is given by

$$A_c Z_t = c + A_1 Z_{t-1} + \ldots + A_p Z_{t-p} + \varepsilon_t. \tag{2}$$

Note that the parameters in A_c are related to the ones in A_1 due to stacking the high-frequency observations $X_t^{(m)}$ in Z_t (Ghysels, 2015).[3] Explicitly for a lag length of $p = 1$, the model reads as:

$$\begin{pmatrix} 1 & \beta_1 & \beta_2 & \ldots & \beta_m \\ \delta_1 & 1 & -\rho_1 & \ldots & -\rho_{m-1} \\ \delta_2 & 0 & 1 & \ldots & -\rho_{m-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \delta_m & 0 & 0 & \ldots & 1 \end{pmatrix} \begin{pmatrix} y_t \\ x_t^{(m)} \\ x_{t-1/m}^{(m)} \\ \vdots \\ x_{t-(m-1)/m}^{(m)} \end{pmatrix} =$$
$$\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_{m+1} \end{pmatrix} + \begin{pmatrix} \rho_y & \theta_1 & \theta_2 & \ldots & \theta_m \\ \psi_2 & \rho_m & \ldots & \ldots & 0 \\ \psi_3 & \rho_{m-1} & \rho_m & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \psi_{m+1} & \rho_1 & \rho_2 & \ldots & \rho_m \end{pmatrix} \begin{pmatrix} y_{t-1} \\ x_{t-1}^{(m)} \\ x_{t-1-1/m}^{(m)} \\ \vdots \\ x_{t-1-(m-1)/m}^{(m)} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{(m+1)t} \end{pmatrix}. \tag{3}$$

In general, we assume the high-frequency process to follow an AR(q) process with $q < m$ (we set $q = m$ in the equation above; if $q < m$, one should set $\rho_{q+1} = \ldots = \rho_m = 0$). For non-zero values of β_j or δ_j the matrix A_c links contemporaneous values of y and x, a feature referred to as nowcasting causality (Götz and Hecq, 2014).[4]

Pre-multiplying (3) by A_c^{-1} we get to the mixed-frequency reduced-form VAR(p) model:

$$\begin{aligned} Z_t = &\ \mu + \Gamma_1 Z_{t-1} + \ldots + \Gamma_p Z_{t-p} + u_t \\ = &\ \mu + B' \underline{Z}_t + u_t. \end{aligned} \tag{4}$$

Consequently, $\mu = A_c^{-1} c$, $\Gamma_i = A_c^{-1} A_i$ and $u_t = A_c^{-1} \varepsilon_t$. Let $B = (\Gamma_1, \Gamma_2, \ldots, \Gamma_p)'$ and $B(z) = 1 - \sum_{j=1}^{p} \Gamma_j z^j$. We then make the following assumptions on the MF-VAR:

Assumption 1. Z_t is generated by the MF-VAR(p) in (4), for which it holds that (i) the roots of the matrix polynomial $B(z)$ all lie outside the unit circle; (ii) u_t is independent and identically distributed (i.i.d.) with $\mathbb{E}(u_t) = 0$, $\mathbb{E}(u_t u_t') = \Sigma_u$, with Σ_u positive definite, and $\mathbb{E}(||u_t||^4) < \infty$, where $||\cdot||$ is the Frobenius norm.

Assumption 1(i) ensures that the MF-VAR is I(0), while (ii) is a standard assumption to ensure validity of the bootstrap for VAR models, see, e.g., Paparoditis (1996), Kilian (1998) or Cavaliere, Rahbek, and Taylor (2012). For the derivation of the limit distribution of the test statistics (ii) could be weakened (e.g., Ghysels et al., 2015b), but as we rely on

[3] Compared to Ghysels (2015) we simply reverse the mixed-frequency vector Z_t and put the low-frequency variable first.

[4] $\beta_j \neq 0$ 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, 2015 for details). Note that Götz, Hecq, and Lieb (2015) consider a noncausal MF-VAR in which one allows for non-zero elements in the lead coefficients of the A_c matrix.

the aforementioned literature to establish the asymptotic validity of our bootstrap tests proposed in Section 3.1.4, we assume i.i.d. error terms here.

Remark 2. Assumption 1 implies that the data are truly generated at mixed frequencies. Hence, similar to Ghysels et al. (2015a) but unlike Ghysels et al. (2015b), we do not start with a common high-frequency data generating process (DGP hereafter).[5] The latter would imply causality patterns to arise at the high frequency. Consequently, we do not investigate which causal relationships at high frequency get preserved when moving to the mixed- or low-frequency case. Indeed, an extension of our methods along these lines would demand a careful analysis of the mixed- and low-frequency systems corresponding to their latent high-frequency counterpart (see Ghysels et al., 2015b for the unrestricted VAR case).

Equation (4) is easy to estimate for small m, yet becomes a rather large system as the latter grows. For example, in a MF-VAR(1),

$$\begin{pmatrix} y_t \\ x_t^{(m)} \\ x_{t-1/m}^{(m)} \\ \vdots \\ x_{t-(m-1)/m}^{(m)} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_{m+1} \end{pmatrix} + \underbrace{\begin{pmatrix} \gamma_{1,1} & \gamma_{1,2} & \cdots & \gamma_{1,m+1} \\ \gamma_{2,1} & \gamma_{2,3} & \cdots & \gamma_{2,m+1} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{m+1,1} & \gamma_{m+1,2} & \cdots & \gamma_{m+1,m+1} \end{pmatrix}}_{\Gamma_1} \times \begin{pmatrix} y_{t-1} \\ x_{t-1}^{(m)} \\ x_{t-1-1/m}^{(m)} \\ \vdots \\ x_{t-1-(m-1)/m}^{(m)} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \\ \vdots \\ u_{(m+1)t} \end{pmatrix}, \tag{5}$$

$$u_t \overset{i.i.d.}{\sim} (\mathbf{0}_{((m+1)\times 1)}, \Sigma_u), \ \Sigma_u = \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} & \cdots & \sigma_{1,m+1} \\ \sigma_{2,1} & \sigma_{2,2} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m+1,1} & \cdots & \cdots & \sigma_{m+1,m+1} \end{pmatrix}, \tag{6}$$

there are $(m+1)^2$ parameters to estimate in the matrix Γ_1. Additional lags would further complicate the issue.

[5] Given such a situation it would seem 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, Gudmundsson, Lindgren, Bondesson, Harsaae, Laake, Juselius, and Lauritzen, 1981), which contains latent processes, i.e., the high-frequency observations of y. 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, 2014).

2.3 Granger Causality in MF-VARs

Let Ω_t represent the information set available at moment t and let Ω_t^W denote the corresponding set excluding information about the stochastic process W. With $P[X_{t+h}^{(m)}|\Omega]$ being the best linear forecast of $X_{t+h}^{(m)}$ based on Ω, Granger non-causality is defined as follows (Eichler and Didelez, 2009):

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

$$P[X_{t+1}^{(m)}|\Omega_t^y] = P[X_{t+1}^{(m)}|\Omega_t]. \tag{7}$$

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

$$P[y_{t+1}|\Omega_t^{X^{(m)}}] = P[y_{t+1}|\Omega_t]. \tag{8}$$

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 (Granger, 1969). In terms of (5), testing for Granger non-causality implies the following null (and alternative) hypotheses:

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

$$\begin{aligned} &H_0 : \gamma_{2,1} = \gamma_{3,1} = \ldots = \gamma_{m+1,1} = 0, \\ &H_A : \gamma_{i,1} \neq 0 \text{ for at least one } i = 2, \ldots, m+1. \end{aligned} \tag{9}$$

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

$$\begin{aligned} &H_0 : \gamma_{1,2} = \gamma_{1,3} = \ldots = \gamma_{1,m+1} = 0, \\ &H_A : \gamma_{1,i} \neq 0 \text{ for at least one } i = 2, \ldots, m+1. \end{aligned} \tag{10}$$

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.

There are many alternative approaches to reduce the number of parameters among which are principal components, Lasso (Tibshirani, 1996) or ridge regressions (Hoerl and Kennard, 1970, among others). However, using principal components does not necessarily preserve the dynamics of the VAR under the null: nothing prevents, e.g., the first and only principal component to be loading exclusively on y implying that the remaining dynamics enter the error term. In other words, the autoregressive matrices in (4) may and will most likely not be preserved, which naturally affects the block of parameters we test on for Granger non-causality. As for Lasso and ridge regressions, it is well known that they may be interpreted in a Bayesian context. In particular, the latter is equivalent to imposing a normally distributed prior with mean zero on the parameter vector (Vogel, 2002, among others), while the former may be replicated using a zero-mean Laplace prior

distribution (Park and Casella, 2008). Given that our set of models contains a Bayesian VAR for mixed-frequency data, we abstract from its connection to regularized versions of least squares at this stage. Developing a system version of Marsilli (2014) and justifying it from a Bayesian point of view, however, constitutes an interesting avenue for future research.

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 reduce rank regression model, for which we make the following assumption:

Assumption 2. Let $B_{X^{(m)}}$ be the matrix obtained from B' in (4) by excluding the first columns of Γ_1, Γ_2, ..., Γ_p. The rank of this matrix is smaller than the number of high-frequency observation within Z_t, i.e.,

$$rk(B'_{X^{(m)}}) = r < m.$$

The model then reads as follows:

$$\begin{aligned} Z_t = &\ \mu + \gamma_{\cdot,1}\underline{y}_t + \alpha \textstyle\sum_{i=1}^{p} \delta_i^{\prime(m)}{}_{t-i} + \nu_t \\ = &\ \mu + \gamma_{\cdot,1}\underline{y}_t + \alpha\delta' \underline{X}_t^{(m)} + \nu_t, \end{aligned} \tag{11}$$

where $\gamma_{\cdot,1}$ is the $(m+1) \times p$ matrix containing the first columns of $\Gamma_1, \Gamma_2, \ldots, \Gamma_p$, and α and $\delta = (\delta_1', \ldots, \delta_p')'$ are $(m+1) \times r$ and $pm \times r$ matrices, respectively. Note that (11) 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, (11) is equivalent to

$$Z_t = \mu + \underline{B}'\underline{Z}_t + \nu_t, \tag{12}$$

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

$$\begin{aligned} \begin{pmatrix} y_t \\ x^{(m)} \\ x_{t-1/m}^{(m)} \\ \vdots \\ x_{t-(m-1)/m}^{(m)} \end{pmatrix} = &\ \mu + \begin{pmatrix} \gamma_{1,1} \\ \gamma_{2,1} \\ \vdots \\ \gamma_{m+1,1} \end{pmatrix} y_{t-1} + \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{m+1} \end{pmatrix} \delta_1' \begin{pmatrix} x_{t-1}^{(m)} \\ x_{t-1-1/m}^{(m)} \\ \vdots \\ x_{t-1-(m-1)/m}^{(m)} \end{pmatrix} + \nu_t \\ = &\ \mu + \underbrace{\left[\begin{pmatrix} \gamma_{1,1} \\ \gamma_{2,1} \\ \vdots \\ \gamma_{m+1,1} \end{pmatrix}, \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{m+1} \end{pmatrix} \delta_1' \right]}_{\underline{\Gamma}_1} \times \begin{pmatrix} y_{t-1} \\ x_{t-1}^{(m)} \\ x_{t-1-1/m}^{(m)} \\ \vdots \\ x_{t-1-(m-1)/m}^{(m)} \end{pmatrix} + \nu_t, \end{aligned} \tag{13}$$

where each $\alpha_i, i = 1, \ldots, m+1$, is of dimension $1 \times r$ and where δ_1 is an $m \times r$ matrix. Hence, we could call $\delta' \underline{X}_t^{(m)}$ a vector of r high-frequency factors. Note that $r = m - s$,

where s represents the rank reduction we are able to achieve within $\underline{X}_t^{(m)}$. In terms of parameter reduction, the unrestricted VAR in (4) requires $p(m+1)^2$ coefficients to be estimated in the autoregressive matrices, whereas the VAR under reduced rank restrictions in (11) or (12) 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 Γ_1 in. Note that we do not require y_{t-1} to be included in the same transmission mechanism as the x variables.

Remark 3. There are several ways to justify the reduced rank feature of the autoregressive matrix $B_{X^{(m)}}$. First, at the model representation level we may assume that, in the structural model (3), x follows an AR(q) process with $q<m$ and that the last $m-q$ elements of each $X_{t-i}^{(m)}, i=1,\ldots,p$, have a zero coefficient in the equation for y_t. Plugging these restrictions into (4) results in a reduced rank of $B'_{X^{(m)}}$ because matrices $\Gamma_i = A_c^{-1}A_i$ have the rank of A_i. Second, at the empirical level one 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 because of the combinations of many elements. Thus, a small number of dynamic factors can approximate more complicated (possibly nonlinear) dynamics. Finally, the way one typically restricts the MF-VAR in (4), i.e., assuming the high-frequency series to follow an ARX-process (Ghysels, 2015), actually implies a reduced rank representation. Looking slightly ahead, consider Equation (25) or the matrix in (29), which we will use in our Monte Carlo section, for the special case $p=1$. It is obvious that for first order MF-VAR, $Z_t = \Gamma_1 Z_{t-1} + u_t$, with

$$\Gamma_1 = \begin{pmatrix} \gamma_{1,1} & \gamma_{1,2} & \gamma_{1,3} & \cdots & \gamma_{1,21} \\ \gamma^*_{2,1} & \rho^{20} & 0 & \ldots & 0 \\ \gamma^*_{3,1} & \rho^{19} & \vdots & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma^*_{21,1} & \rho & 0 & \ldots & 0 \end{pmatrix}$$

we have, due to the block of zeros, a reduced rank matrix $B_{X^{(m)}}$ with two (restricted coefficient) factors:

$$rk\left\{\begin{pmatrix} \gamma_{1,2} & \gamma_{1,3} & \cdots & \gamma_{1,21} \\ \rho^{20} & 0 & \ldots & 0 \\ \rho^{19} & \vdots & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \rho & 0 & \ldots & 0 \end{pmatrix}\right\} = rk\left\{\begin{pmatrix} 1 & 0 \\ 0 & \rho^{20} \\ 0 & \rho^{19} \\ \vdots & \vdots \\ 0 & \rho \end{pmatrix} AA^{-1} \begin{pmatrix} \gamma_{1,2} & \gamma_{1,3} & \cdots & \gamma_{1,21} \\ 1 & 0 & \ldots & 0 \end{pmatrix}\right\} = 2,$$

where A is a 2×2 full rank rotation matrix, the identity in this particular example. Hence, under H_A, e.g., two factors (although unrestricted in our setting) will capture the reduced rank feature. Note that we could impose (overidentifying) restrictions to retain the large block of zeros in Γ_1. In that sense, the model with $r=2$ would not be misspecified but "suboptimal". On the other hand, with merely one factor we would not be able to impose the block of zeros implying the model to be misspecified. Under the null hypothesis the model is not misspecified with either one or two factors, because the test tries to capture

some dynamics that is not there.

Note that our focus is to compare the performance of different approaches including their bootstrap versions (whenever applicable) and that in this light we view reduced rank regressions as just one particular specification. In particular, next to the Bayesian MF-VAR and the *max*-test, our bootstrap implementation has also made the unrestricted MF-VAR (which is not subject to rank misspecification) a viable option to compare the reduced rank regressions with.

3.1.2 Testing for Granger Non-Causality

Given r and Assumption 2, under the condition that the high-frequency factors $\delta' \underline{X}_t^{(m)}$ are observable, we can estimate (11) by OLS. Letting $\widehat{\underline{\Gamma}} = (\hat{\mu}, \hat{\gamma}_{\cdot,1}, \hat{\alpha})'$ denote the corresponding OLS estimates, 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$vec(\widehat{\underline{\Gamma}})$. For a general construction of the matrix R in the presence of several low- and high-frequency variables,[6] we refer the reader to Ghysels et al. (2015b). The Wald test is constructed as

$$\widetilde{\xi}_W = \left[\mathrm{R}vec(\widehat{\underline{\Gamma}})\right]' (\mathrm{R}\hat{\Omega}\mathrm{R}')^{-1} \left[\mathrm{R}vec(\widehat{\underline{\Gamma}})\right], \tag{14}$$

with

$$\hat{\Omega} = (W'W)^{-1} \otimes \hat{\Sigma}_u, \tag{15}$$

where $\hat{\Sigma}_u = \frac{1}{T}\hat{u}'\hat{u}$ is the empirical covariance matrix of the disturbance terms and $W = (W_1, \ldots, W_T)'$ is the regressor set consisting of $W_t = (1, \underline{y}_t', \delta' \underline{X}_t^{(m)\prime})'$.[7] As illustrated in Ghysels et al. (2015b), ξ_W is asymptotically $\chi^2_{rank(\mathrm{R})}$. A robust version of (14) is also implemented in the empirical section.

Of course, the factors are typically not observable and must be estimated (unless we impose specific factors). Using estimated rather than observed factors in the regression (11) will obviously affect the distribution of $\widetilde{\xi}_W$, certainly in small samples. For this reason we consider bootstrap versions of the tests. Before going into the details on the bootstrap we describe how we estimate or impose the factors.

3.1.3 Estimating the Factors

We use three ways to estimate the factors, canonical correlation analysis (CCA hereafter), partial least squares (PLS hereafter) and heterogeneous autoregressive (HAR hereafter) type restrictions. We briefly present the algorithms that are used to extract those factors.

CCA is based on analyzing the eigenvalues and corresponding eigenvectors of

$$\hat{\Sigma}^{-1}_{\underline{\tilde{X}}^{(m)}\underline{\tilde{X}}^{(m)}} \hat{\Sigma}_{\underline{\tilde{X}}^{(m)}\tilde{Z}} \hat{\Sigma}^{-1}_{\tilde{Z}\tilde{Z}} \hat{\Sigma}_{\tilde{Z}\underline{\tilde{X}}^{(m)}} \tag{16}$$

or, similarly, of the symmetric matrix

$$\hat{\Sigma}^{-1/2}_{\underline{\tilde{X}}^{(m)}\underline{\tilde{X}}^{(m)}} \hat{\Sigma}_{\underline{\tilde{X}}^{(m)}\tilde{Z}} \hat{\Sigma}^{-1}_{\tilde{Z}\tilde{Z}} \hat{\Sigma}_{\tilde{Z}\underline{\tilde{X}}^{(m)}} \hat{\Sigma}^{-1/2}_{\underline{\tilde{X}}^{(m)}\underline{\tilde{X}}^{(m)}}. \tag{17}$$

[6] Within the unrestricted MF-VAR in (4), though.

[7] A sample size correction, i.e., using $\hat{\Sigma}_u = \frac{1}{T-K_W}\hat{u}'\hat{u}$, where K_W is the amount of elements in W, may alleviate size distortions in finite samples.

For a detailed discussion we refer the reader to Anderson (1951) or, for the application to common dynamics, to Vahid and Engle (1993). 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 (17), we obtain the estimators of α and δ as:

$$\begin{aligned} \hat{\alpha} = &\ \hat{\Sigma}^{-1}_{\tilde{Z}\tilde{Z}} \hat{\Sigma}_{\tilde{Z}\underline{\tilde{X}}^{(m)}} \hat{\Sigma}^{-1/2}_{\underline{\tilde{X}}^{(m)}\underline{\tilde{X}}^{(m)}} \hat{V} \\ \hat{\delta} = &\ \hat{\Sigma}^{-1/2}_{\underline{\tilde{X}}^{(m)}\underline{\tilde{X}}^{(m)}} \hat{V}. \end{aligned} \tag{18}$$

Note that the estimation of the eigenvectors obtained from the canonical correlation analysis in (16) or (17) 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 $i = 1, \ldots, p$:

$$\delta_i = \begin{pmatrix} \mathbf{0}_{3(i-1)\times m} \\ \hdashline 1 \quad \mathbf{0}_{1\times(m-1)} \\ \hdashline \mathbf{1}_{1\times(\frac{1}{4}m)} \quad \mathbf{0}_{(1\times\frac{3}{4}m)} \\ \mathbf{1}_{1\times m} \\ \mathbf{0}_{3(p-i)\times m} \end{pmatrix}' \Rightarrow \delta' \underline{X}_t^{(m)} = \begin{pmatrix} x_{t-1}^{(m)} \\ \sum_{i=0}^{\frac{1}{4}m-1} x_{t-1-i/m}^{(m)} \\ \sum_{i=0}^{m-1} x_{t-1-i/m}^{(m)} \\ \vdots \\ x_{t-p}^{(m)} \\ \sum_{i=0}^{\frac{1}{4}m-1} x_{t-p-i/m}^{(m)} \\ \sum_{i=0}^{m-1} x_{t-p-i/m}^{(m)} \end{pmatrix}. \tag{19}$$

For $p = 1$ and $m = 20$ this corresponds to

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

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

with x_t^D, x_t^W and x_t^M denoting daily, weekly and monthly measures, respectively.

Remark 4. 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 $K_1 < \ldots < K_M$. Alternatively, we could use MIDAS restrictions as introduced in Ghysels et al. (2004), for the difference between the latter and HAR-type restrictions has been shown to 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, thereby simplifying the analysis. Furthermore, and more crucially in the context of this paper, 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]

3.1.4 Bootstrap Tests

Despite the dimensionality reduction achieved by imposing the factor structure, a considerable number of parameters remains to be estimated for conducting the Wald tests such that the tests may still be subject to considerable small sample size distortions. Moreover, the estimation of the factors will have a major effect on the distribution of the Wald test statistic. We therefore also consider a bootstrap implementation of these tests to improve the properties of the tests in finite samples.

Let $\underline{\widehat{B}} = \underline{\widehat{\Gamma}}\widehat{\delta}'$ denote the estimates of $\underline{B}$ obtained by one of the reduced rank methods, and assume that δ is normalized such that its upper $r \times r$ block is equal to the identity matrix.[10] The first bootstrap method we consider is a standard "unrestricted" bootstrap, that may have superior power properties in certain cases (cf. Paparoditis and Politis, 2005). Its algorithm looks as follows.

1. Obtain the residuals from the MF-VAR(p)

$$\tilde{\nu}_t = Z_t - \hat{\mu} + \underline{\widehat{B}}' \underline{Z}_t, \qquad t = p+1, \ldots, T. \tag{21}$$

2. Draw the bootstrap errors $\nu_1^*, \ldots, \nu_T^*$ with replacement from $\tilde{\nu}_{y,p+1}, \ldots, \tilde{\nu}_{y,T}$.

3. Letting $\underline{Z}_t^* = (Z_{t-1}^{*\prime}, \ldots, Z_{t-p}^{*\prime})'$, construct the bootstrap sample Z_t^* recursively as

$$Z_t^* = \hat{\mu} + \underline{\widehat{B}}' \underline{Z}_t^* + \nu_t^*, \qquad t = 1, \ldots, T, \qquad Z_0^* = 0. \tag{22}$$

[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. Admittedly, HAR-type restrictions provide less flexibility than the MIDAS approach, yet their simplicity makes them very appealing from an applied perspective.

[10] This ensures that the rotation of the factors is the same for the original sample and the bootstrap sample, which ensures for the first bootstrap method that the correct bootstrap null hypothesis is imposed. In the simulations we experimented with different normalizations which didn't change any results.

Note that the null hypothesis in the bootstrap is not imposed which has consequences for the next step.

4. Estimate the bootstrap equivalent of (11), where the factors are estimated using the same method as for the original sample, and obtain the bootstrap Wald test statistic $\widetilde{\xi}_W^*$. As the "true" bootstrap parameters governing the Granger causality are different from zero, we need to adapt the bootstrap Wald test such that it tests the correct null hypothesis. Observing that those parameters form a subset of the estimates in $\widehat{\underline{\Gamma}}$ used in (22), the appropriate Wald test statistic is

$$\widetilde{\xi}_W^* = \left[\mathrm{R}vec(\widehat{\underline{\Gamma}}^*) - \mathrm{R}vec(\widehat{\underline{\Gamma}})\right]' (\mathrm{R}\hat{\Omega}^*\mathrm{R}')^{-1} \left[\mathrm{R}vec(\widehat{\underline{\Gamma}}^*) - \mathrm{R}vec(\widehat{\underline{\Gamma}})\right], \tag{23}$$

where all quantities with a superscript '*' are calculated analogously to their sample counterparts but then using the bootstrap sample.

5. Repeat steps 2 to 4 B times, and calculate the bootstrap p-value as the proportion of bootstrap samples for which $\widetilde{\xi}_W^* > \widetilde{\xi}_W$.

While this bootstrap method has the advantage that it can be used for testing Granger causality in both directions (as only the matrix R needs to change), it also has some drawbacks. In particular, even with the dimensionality reduction provided by the reduced rank estimation, it still relies on the generation of a bootstrap sample based on an $(m+1)$-dimensional MF-VAR(p) that requires $p(m+1)+r(m+1)+prm$ parameters to estimate. This may make the bootstrap unstable and prone to generate outlying samples, with potential size distortions or a loss of power as a result.

Therefore we consider a second bootstrap method that imposes the null hypothesis of no Granger causality and in doing so achieves a further significant reduction of dimensionality. A downside of this method is that it requires separate bootstrap algorithms for testing in the two opposite directions. We describe the procedure here for the null hypothesis that $X^{(m)}$ does not Granger cause y; we comment on the test in the opposite direction below.

1. Estimate an AR(p) model for y and obtain the residuals

$$\tilde{u}_{y,t} = y_t - \hat{\mu} - \sum_{j=1}^{p} \hat{\rho}_{y,j} y_{t-j}, \qquad t = p+1, \ldots, T.$$

2. Draw the bootstrap errors $u_1^*, \ldots, u_T^*$ with replacement from $\tilde{u}_{y,p+1}, \ldots, \tilde{u}_{y,T}$.

3. Construct the bootstrap sample y_t^* recursively as

$$y_t^* = \sum_{j=1}^{p} \hat{\rho}_{y,j} y_{t-1}^* + u_t^*, \qquad t = 1, \ldots, T, \qquad y_0^* = 0,$$

thus imposing the null of no Granger causality from $X^{(m)}$ to y.

4. Letting $X_t^{(m)*} = X_t^{(m)}$, use the bootstrap sample to estimate the MF-VAR using the same estimation method as for the original estimator, and obtain the corresponding Wald test statistic $\widetilde{\xi}_W^*$. As the null hypothesis of no Granger causality has been imposed in the bootstrap, $\widetilde{\xi}_W^*$ now has the same form as $\widetilde{\xi}_W$ in (14).

Note that only p parameters need to be estimated in order to generate the bootstrap sample, most likely making the procedure more stable than the unrestricted bootstrap above. A similar bootstrap procedure can be implemented for the reverse direction of causality by only resampling $X_t^{(m)}$ while keeping y_t fixed. As one can resample $X_t^{(m)}$ independently of y_t, it can be done in the original high-frequency as a univariate process. Said differently, it only requires fitting an AR(q) process to $x_t^{(m)}$, again achieving a large reduction of dimensionality.

It is also possible in step 2 of both algorithms to use the wild bootstrap to be robust against heteroskedasticity. In this case the bootstrap errors are generated as $u_t^* = \xi_t^* \tilde{u}_t$ where we generate $\xi_1^*, \ldots, \xi_T^*$ as independent standard normal random variables. This would constitute the so-called "recursive wild bootstrap" scheme that Brüggemann, Jentsch, and Trenkler (2014) prove to be asymptotically valid under conditionally heteroskedastic errors. In particular, employing the wild bootstrap then allows us to replace the i.i.d. assumption in Assumption 1 with the martingale difference sequence assumption in Assumption 2.1 of Brüggemann et al. (2014). We implement the wild bootstrap version in the empirical application in Section 5.

Remark 5. Kilian (1998) and Paparoditis (1996) prove the asymptotic validity of the unrestricted VAR bootstrap under Assumption 1. The validity of the restricted bootstrap can be derived from the results for AR(p) processes in Bose (1988), and the observation that conditioning is a valid approach in the absence of Granger causality. Van Giersbergen and Kiviet (1996) provide a detailed examination of the two different types of bootstraps in the context of ADL models and demonstrate that the restricted bootstrap even works well in the absence of strong exogeneity.[11]

Remark 6. Gonçalves and Perron (2014) consider factor-augmented regression models, with the factors estimated by principal components. In this setting the asymptotic impact of factor estimation depends on which asymptotic framework is assumed. Under the asymptotic framework, in which factor estimation has a non-negligible effect on the limit distribution of the regression estimators, the bootstrap correctly mimics this effect and is therefore asymptotically valid. Consequently, the bootstrap provides a much more accurate approximation to the finite sample distribution of the regression estimators than the asymptotic approximation, that assumes a negligible effect of factor estimation. We expect the bootstrap to have similar properties in our setting where the factors are estimated using CCA or PLS.

3.2 Bayesian MF-VARs

3.2.1 Restricted MF-VARs

Ghysels (2015) discusses the issue of parsimony in MF-VAR models by specifying the high-

[11] If the test statistic (14) is asymptotically pivotal, we may expect the bootstrap to provide asymptotic refinements as well and, hence, reduce small sample size distortions (cf. Bose, 1988; Horowitz, 2001).

frequency process in such a way as to allow a number of parameters that is independent from m. Indeed, while it seems reasonable to leave the equation for y_t unrestricted,[12] it is less clear for the remaining ones. Hence, let us make the following assumption (see Ghysels, 2015 or the numerical examples in Ghysels et al., 2015a).

Assumption 3. The high-frequency process $x_t^{(m)}$ follows an AR(1) model with one lag of the low-frequency variable in the regressor set

$$x_{t-i/m}^{(m)} = \mu_{i+2} + \rho x_{t-(i+1)/m}^{(m)} + \pi y_{t-1} + \upsilon_{(i+2)t} \tag{24}$$

for $i = 0, \ldots, m-1$.

Completing the system with the equation for y_t leads to the following restricted MF-VAR:

$$\begin{aligned}
\begin{pmatrix} y_t \\ x_t^{(m)} \\ x_{t-1/m}^{(m)} \\ \vdots \\ x_{t-(m-1)/m}^{(m)} \end{pmatrix} = &\begin{pmatrix} \mu_1 \\ \sum_{i=0}^{m-1} \rho^i \mu_{2+i} \\ \sum_{i=0}^{m-2} \rho^i \mu_{3+i} \\ \vdots \\ \mu_{m+1} \end{pmatrix} + \begin{pmatrix} \gamma_{1,1}^{(1)} & \gamma_{1,2}^{(1)} & \gamma_{1,3}^{(1)} & \cdots & \gamma_{1,m+1}^{(1)} \\ \pi \sum_{i=0}^{m-1} \rho^i & \rho^m & & & \\ \pi \sum_{i=0}^{m-2} \rho^i & \rho^{m-1} & & \mathbf{0}_{m\times(m-1)} & \\ \vdots & \vdots & & & \\ \pi & \rho & & & \end{pmatrix} \times \\
&\begin{pmatrix} y_{t-1} \\ x_{t-1}^{(m)} \\ x_{t-1-1/m}^{(m)} \\ \vdots \\ x_{t-1-(m-1)/m}^{(m)} \end{pmatrix} + \sum_{k=2}^{p} \begin{pmatrix} \gamma_{1,1}^{(k)} & \gamma_{1,2}^{(k)} & \cdots & \gamma_{1,m+1}^{(k)} \\ & \mathbf{0}_{m\times(m+1)} & & \end{pmatrix} \\
&\times \begin{pmatrix} y_{t-k} \\ x_{t-k}^{(m)} \\ \vdots \\ x_{t-k-(m-1)/m}^{(m)} \end{pmatrix} + \underbrace{\begin{pmatrix} \upsilon_{1t} \\ \sum_{i=0}^{m-1} \rho^i \upsilon_{(m+1-i)t} \\ \sum_{i=0}^{m-2} \rho^i \upsilon_{(m+1-i)t} \\ \vdots \\ \upsilon_{(m+1)t} \end{pmatrix}}_{\upsilon_t^*},
\end{aligned} \tag{25}$$

where $\gamma_{i,j}^{(k)}$ corresponds to the (i,j)-element of matrix Γ_k in (4). As for the error terms $\upsilon_{it}, i = 1, \ldots, m+1$, we set $\mathbb{E}(\upsilon_{it}\upsilon_{it}) = \sigma_{HH}$, $\mathbb{E}(\upsilon_{1t}\upsilon_{it}) = \sigma_{HL}$ for $i = 2, \ldots, m+1$, and $\mathbb{E}(\upsilon_{1t}\upsilon_{1t}) = \sigma_{LL}$. Furthermore, each error term is assumed to possess a zero mean and to be normally distributed. Consequently, $\upsilon_t^* \sim N(\mathbf{0}_{((m+1)\times 1)}, \Sigma_{\upsilon^*})$, whereby we refer the reader to Ghysels (2015) for the exact composition of Σ_{υ^*}.

[12] In fact, viewed as single equation, it boils down to an unrestricted MIDAS model (Foroni, Marcellino, and Schumacher, 2015) without contemporaneous observations of the high-frequency variable. 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 dummy variable approach presented below. 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.

3.2.2 The Auxiliary Dummy Variable Approach for Mixed-Frequency Data

As pointed out by Carriero, Kapetanios, and Marcellino (2011), Bayesian methods allow the imposition of restrictions such as the ones in Assumption 3, while also admitting influence of the data. 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). As for MF-VARs, Ghysels (2015) describes a way to sample the MIDAS hyperparameters and subsequently formulates prior beliefs for the remaining parameters.[13] 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.

We follow the approach of Banbura et al. (2010), which in turn is built on the work of Sims and Zha (1998), showing that adding a set of auxiliary dummy variables to the system is equivalent to imposing a normal inverted Wishart prior. The specification of prior beliefs is derived from the Minnesota prior in Litterman (1986), whereby we center the prior distributions of the coefficients in B around the restricted MF-VAR in (25):

$$\begin{aligned} \mathbb{E}[\gamma_{i,j}^{(k)}] &= \begin{cases} \rho^m & \text{if } i=j=k=1 \\ \rho^{m+j-i} & \text{if } k=1, j=2, i>1 \\ 0 & \text{else} \end{cases}, \\ Var[\gamma_{i,j}^{(k)}] &= \begin{cases} \phi\frac{\lambda^2}{k^2}\mathbb{S}_{LH} & \text{for } i=1, j>1 \\ \phi\frac{\lambda^2}{k^2}\mathbb{S}_{HL} & \text{for } j=1, i>1 \\ \frac{\lambda^2}{k^2} & \text{else} \end{cases}, \end{aligned} \tag{26}$$

where all $\gamma_{i,j}^{(k)}$ are assumed to be a priori independent and normally distributed. 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 $m+1$. The tightness of the prior distributions around the AR(1) specification in (24) is determined by λ,[14] the influence of low- on high-frequency data and vice versa is controlled by ϕ and, finally, $\mathbb{S}_{ij} = \frac{\sigma_i^2}{\sigma_j^2}, i,j = L,H$, governs the difference in scaling between y and the x-variables. For μ we take a diffuse prior.

Remark 7. As for the common-frequency case, the expressions for $Var[\gamma_{i,j}^{(k)}]$ in (26) imply that more recent (low-frequency) lags provide more reliable information than more distant ones. However, due to the stacked nature of $X_t^{(m)}$, the coefficients could be shrunk according to their high-frequency time difference instead. This implies specifying the lag associated with each coefficient in fractions of the low-frequency time index: y_t and $x_{t-1-s/m}^{(m)}$, e.g., are $1+i/m$ low-frequency time periods apart such that the denominator in the corresponding coefficient's prior variance would equal $(1+i/m)^2$.

[13] McCracken, Owyang, and Sekhposyan (2015) 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 (2015). However, they use a latent high-frequency VAR instead of the mixed-frequency system à la Ghysels (2015).

[14] $\lambda \approx 0$ 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 (4).

However, in the context of Granger causality testing such a specification is problematic as the coefficients to test on are shrunk in "opposite directions": y_t and $x^{(m)}_{t-1-1/2}$ are 1.5 t-periods apart, whereas $x^{(m)}_{t-1-1/2}$ and y_{t-1} are separated by only 0.5 low-frequency periods. Consequently, for a given λ, the coefficients are shrunk much more when testing for Granger causality from $X^{(m)}$ to y, especially for large m. This makes it difficult to control the size of the respective tests in one or the other direction. The common-frequency handling of the prior variances in (26) mitigates this issue, although some losses in power have to be assumed.

Note that a treatment of the mixed-frequency nature of the variables along the lines outlined above may well be advantageous in other circumstances (e.g., forecasting) such that we present this approach in detail in Appendix A.[15]

Given the prior beliefs specified before, the analysis is very similar to the one in Banbura et al. (2010). In short, let us write the MF-VAR as

$$Z = \underline{Z}B^* + E,$$

where $\underline{Z} = (\underline{Z}^\mu_1, \dots, \underline{Z}^\mu_T)'$ with $\underline{Z}^\mu_t = (\underline{Z}'_t, 1)'$, $Z = (Z_1, \dots, Z_T)'$, $E = (u_1, \dots, u_T)'$ and $B^* = (B', \mu)'$. Then, one can show that augmenting the model by two dummy variables, Y_d and X_d, is equivalent to imposing a normal inverted Wishart prior that satisfies the moments in (26). Finally, estimating the augmented model by ordinary least squares gives us the posterior mean of the coefficients, on which we can do inference as outlined in the next subsection.

Note that X_d is, in fact, identical to the matrix for the common-frequency VAR (Banbura et al., 2010); Y_d, however, is slightly different due to the prior means being centered around the restricted VAR in (25):

$$\underbrace{Y_d}_{((m+1)(p+1)+1)\times(m+1)} = \begin{pmatrix} \rho^m\sigma_L/\lambda & 0 & 0 & \dots & 0 \\ 0 & \rho^m\sigma_H/\lambda & \rho^{m-1}\sigma_H/\lambda & \dots & \rho\sigma_H/\lambda \\ & & \mathbf{0}_{((m+1)p-2)\times(m+1)} & & \\ & & diag(\sigma_L, \sigma_H, \dots, \sigma_H) & & \\ & & \mathbf{0}_{1\times(m+1)} & & \end{pmatrix},$$

Augmenting the model is then achieved by setting $Z_* = (Z', Y'_d)'$ and $\underline{Z}_* = (\underline{Z}', X'_d)'$.

3.2.3 Testing for Granger Non-Causality

In terms of analyzing the Granger non-causality testing behavior within the Bayesian MF-VAR, the auxiliary dummy variable approach is quite appealing, as it provides a closed-form solution for the posterior mean of the coefficients. It is thus straightforward to compare the testing behavior with the ones from alternative approaches using the Wald

[15] When extending the approach we stick to the standard structure of the Minnesota prior in Litterman (1986). It is, however, also feasible to address the aforementioned issue by introducing, say, three hyperparameters: λ_1 governing the prior tightness for the coefficients determining Granger causality from y to $X^{(m)}_t$, λ_2 for parameters controlling causality in the reverse direction and λ_3 taking care of the remaining coefficients. This strategy is, however, a rather straightforward extension of the theory presented in Appendix A such that we do not present it here.

test.[16] The latter is derived analogously to the one in (14), and is, given Assumption 3, also asymptotically $\chi^2_{rank(\mathrm{R})}$-distributed.

Remark 8. Because the parameter vector $vec(B^*)$ is not assumed fixed but random, construction of the test statistic, and especially its interpretation, need to be treated with special care. We consider Bayesian confidence intervals, in particular highest posterior density (HPD) confidence intervals (Bauwens, Lubrano, and Richard, 2000), the tightness of which is expressed by α, not coincidentally the same letter that denotes the significance level in the frequentist's framework. Here, it is interpreted such that $(1-\alpha)\%$ of the probability mass falls within the respective interval. In other words, the probability that a model parameter falls within the bounds of the interval is equal to $(1-\alpha)\%$. The interval centered at the modal value for unimodal symmetric posterior densities is then called the HPD (Zellner, 1996 or Bauwens et al., 2000). The connection to Wald tests is now immediate using the equivalence between confidence intervals and respective test statistics, whereby similar care is demanded when interpreting results.

3.3 Benchmark Models

3.3.1 Low-Frequency VAR

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, a monthly variable, for example, was usually aggregated to, say, the quarterly frequency such that a VAR could be estimated in the resulting common low frequency. Formally, $x_t = W(L^{1/m})x_t^{(m)}$, 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).[17] As far as testing for Granger non-causality is concerned, we can rely on the Wald statistic in (14), where the set of regressors, the matrix R and the coefficient matrix are suitably adjusted.

Remark 9. Naturally, such temporal aggregation leads to a great reduction in parameters. After all, each set of m high-frequency variables per t-period is aggregated into one low-frequency observation. Of course, this decrease in parameters comes at the cost of disregarding information embedded in the high-frequency process. As argued in Miller (2011), if the aggregation scheme employed is different from the true one underlying the DGP, 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 (Breitung and Swanson, 2002), since causality is a property which is not invariant to temporal aggregation (Marcellino, 1999 or Sims, 1971).

[16] There is a large sample correspondence between classical Wald and Bayesian posterior odds tests (Andrews, 1994). For certain choices of the prior distribution, the posterior odds ratio is approximately equal to the Wald statistic. Andrews (1994) shows that for any significance level α there exist priors such that the aforementioned correspondence holds, and vice versa.

[17] 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.

3.3.2 The *max*-test

Independently and simultaneously to this work, Ghysels et al. (2015a) have developed a new Granger non-causality testing framework, whose parsimonious structure makes it very appealing in a situation, where m is large relative to the sample size. In short, the idea is to focus on the first line of the MF-VAR in (4), but rather than estimating that (U)-MIDAS equation (Foroni et al., 2015) and test for Granger non-causality in the direction from $X^{(m)}$ to y, the authors propose to compute the OLS estimates of β_j in the following h separate regression models:

$$y_t = \mu + \sum_{k=1}^{q} \alpha_{k,j} y_{t-k} + \beta_{j+1} x_{t-1-j/m}^{(m)} + v_j,\ j = 0, \dots, h-1,$$

whereby h needs to be set "sufficiently large" (to achieve $h > pm$). The corresponding *max*-test statistic is then the properly scaled and weighted maximum of $\{\hat{\beta}_1^2, \dots, \hat{\beta}_h^2\}$. Although it has a non-standard limit distribution under H_0, an asymptotic p-value may be obtained in a similar way as when overcoming the Davies (1987) problem (see Remark 4). Testing for Granger non-causality in the reverse direction works analogously in the following regression model:

$$y_t = \mu + \sum_{k=1}^{q} \alpha_{k,j} y_{t-k} + \sum_{k=1}^{h} \beta_{k,j} x_{t-1-k/m}^{(m)} + \gamma_j x_{t+j/m}^{(m)} + v_j,\ j = 1, \dots, l.$$

Again, the corresponding *max*-test statistic is the maximum of $\{\hat{\gamma}_1^2, \dots, \hat{\gamma}_l^2\}$ scaled and weighted properly.[18]

3.3.3 Unrestricted VARs

Finally, we can attempt to estimate the full MF-VAR in (4) ignoring the possibility that the amount of parameters may be too large to perform accurate estimations or adequate Granger non-causality tests. To this end we estimate the MF-VAR using ordinary least squares disregarding the potential parameter proliferation problem. In this sense the comparison is related to the one of U-MIDAS (Foroni et al., 2015) and MIDAS regression models for large m. We can test for Granger non-causality using the Wald statistic in (14). Again, W, R and B have to be adjusted adequately. We also consider a bootstrap version of the unrestricted MF-VAR, which we expect to alleviate size distortions, but which cannot solve power issues due to the parameter proliferation.

[18] A MIDAS polynomial (Ghysels et al., 2004) on $\sum_{k=1}^{h} \beta_{k,j} x_{t-1-k/m}^{(m)}$ may be imposed to increase parsimony. Note that the aforementioned Davies problem does not occur due to testing for Granger non-causality by looking at the OLS estimates of γ_j. Furthermore, Ghysels et al. (2015a) argue that m contemporaneous high-frequency observations of x should be included as instruments in order to handle simultaneity between y and x.

4 Monte Carlo Simulations

In order to assess the finite sample performance of our different parameter reduction techniques, we conduct a Monte Carlo experiment. In light of our empirical investigation we set $m = 20$, i.e., as in a month/ working day-example.[19] 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.,

$$Z_t = u_t. \tag{27}$$

Remark 10. Additionally, we have considered three alternative DGPs for size, all based on the restricted VAR in (25): (i) $\gamma_{1,i} = \gamma^*_{i,1} = 0$ ('diagonal'), (ii) $\gamma_{1,i} = 0$ and $\gamma^*_{i,1} \neq 0$ ('only Granger non-causality from $X^{(m)}$ to y') or (iii) $\gamma^*_{i,1} = 0$ and $\gamma_{1,i} \neq 0$ ('only Granger non-causality from y to $X^{(m)}$') $\forall i = 2, \ldots, 21$, where the parameters $\gamma_{1,i}$ and $\gamma^*_{i,1}$ refer to (29). However, with the respect to the respective testing direction, the outcomes do not differ qualitatively and quantitative differences are very small. Results are available upon request.

To analyze power we generate two different DGPs that are closely related to the restricted VAR in (25):

$$Z_t = \Gamma_P Z_{t-1} + u_t \tag{28}$$

with

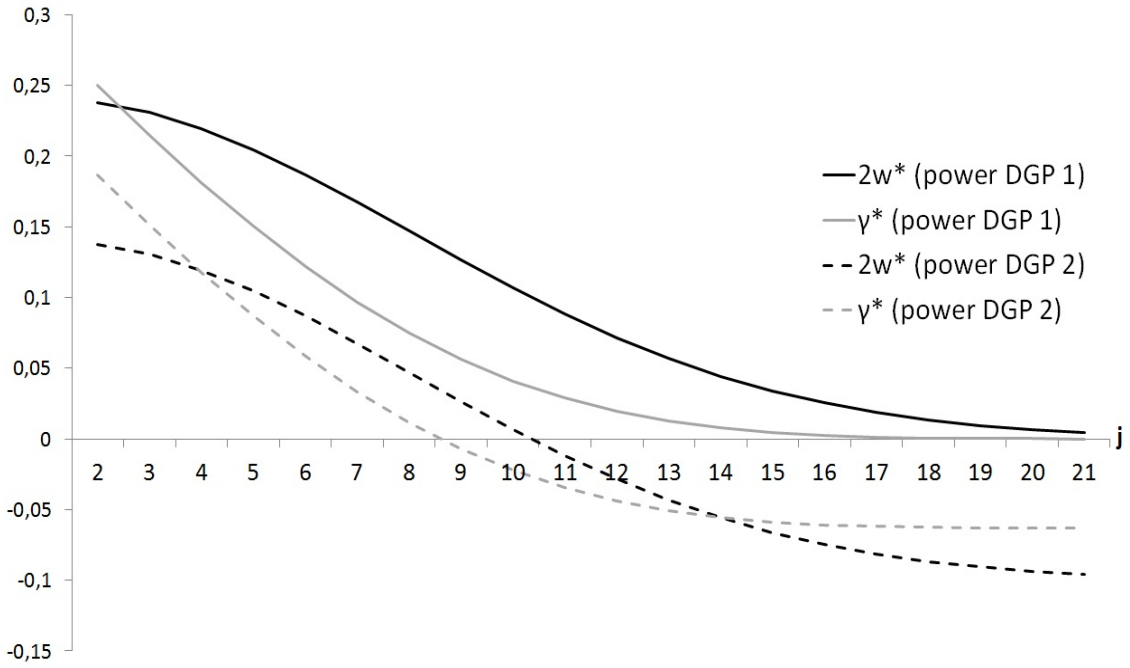
$$\Gamma_P = \begin{pmatrix} \gamma_{1,1} & \gamma_{1,2} & \gamma_{1,3} & \cdots & \gamma_{1,21} \\ \gamma^*_{2,1} & \rho^{20} & 0 & \ldots & 0 \\ \gamma^*_{3,1} & \rho^{19} & \vdots & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma^*_{21,1} & \rho & 0 & \ldots & 0 \end{pmatrix}, \tag{29}$$

where $\gamma_{1,1} = 0.5$, $\rho = 0.6$ and $\gamma_{1,j} = \beta w^*_{j-1}(-0.01)$ for $j > 1$. Note that $w_i(\psi) = \frac{exp(\psi i^2)}{\sum_{i=1}^{20} exp(\psi i^2)}$ corresponds to the two-dimensional exponential Almon lag polynomial with the first parameter set to zero (Ghysels et al., 2007). Now, for the first power DGP we simply set $w^*_i(\psi) = w_i(\psi)$, whereas in the second power DGP $w^*_i(\psi) = w_i(\psi) - \overline{w(\psi)}$ with the horizontal bar symbolizing the arithmetic mean. As far as $\gamma^*_{j,1}$ is concerned we take $\gamma^*_{j,1} = \gamma_{j,1}$ (first power DGP) and $\gamma^*_{j,1} = \gamma_{j,1} - \overline{\gamma_{\cdot,1}}$ (second power DGP) for $j = 2, \ldots, 21$, whereby $\gamma_{j,1} = (\gamma_{j-1,1})^{1.11}$ for $j = 3, \ldots, 21$.[20] As an example, Figure 1 plots βw^*_{j-1} and $\gamma^*_{j,1}$ for $\beta = 2$ and $\gamma_{2,1} = 0.25$.

[19] See Section 5 for a justification of the time-invariance of m within this setup.

[20] The parameter values have been chosen to mimic part of the structure of the restricted VAR in (25) and to ensure stability of the system. In the first DGP late x_{t-1}-observations have a larger impact on y_t, and positive values of y_{t-1} increase x towards the end of the current period, but hardly have an impact at the beginning. In the second power DGP a zero-mean feature is imposed on the coefficients $w^*_{j-1}(\psi)$ and $\gamma^*_{j,1}$ without changing the general pattern of their evolvement.

Figure 1: Parameter values for $2w^*_{j-1}(-0.01)$ and $\gamma^*_{j,1}$ ($\gamma_{2,1} = 0.25$)



Remark 11. 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 (see Section 3.3.1). It is thus a situation, in which classical temporal aggregation is expected not to preserve the causality patterns in the data (Marcellino, 1999). 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 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 Σ_u has the same structure as the covariance matrix of the restricted VAR in (25) with $\sigma_{LL} = 0.5$ and $\sigma_{HH} = 1$.[21]

For CCA and PLS we consider $r = 1, 2$ and leave the analysis of higher factor dimensions for further research. Recall from Remark 3 that under H_0 the model is not misspecified with either amount of factors. Under H_A, however, the model is misspecified for $r = 1$, but not for $r = 2$. In that sense, we implicitly analyze the impact of misspecification, stemming from choosing r too small, on the behavior of the respective test statistics. As far as the Bayesian approach is concerned, we experimented with various values for the hyperparameter and found $\lambda = 0.175$ to be a good choice. Furthermore, we approximate ρ by running a corresponding univariate regression. For the *max*-test we take $q = 1$ and $h = l = m = 20$.

We consider different variants of our size and power DGPs by varying the values of σ_{HL} and β. To be more precise, we choose $\sigma_{HL} = 0, -0.05, -0.1$ when analyzing size properties of our tests in both directions. A value of zero implies the absence of nowcasting causality, whereas the other two values imply some degree of correlation between the low-

[21] These values resemble the data in the empirical section, where $\hat{\sigma}_{LL} = 0.55$ and $\hat{\sigma}_{HH} = 1.09$.

and high-frequency series. For both power DGPs we fix $\sigma_{HL} = -0.05$, though. When investigating Granger causality testing from $X^{(m)}$ to y in the first power DGP we consider $\beta = 0.5, 0.8, 2$; in the second power DGP we only take the $\beta = 0.5$. For causality in the reverse direction we take $\gamma_{2,1} = 0.2$ in both power DGPs, because the results do not seem sensitive in this respect.

The figures in the tables below represent the percentage amount of rejections at $\alpha = 5\%$.[22] All figures are based on 2,500 replications and are computed using GAUSS12. For the bootstrap versions we use $B = 499$.

4.1 Size

Tables 1 and 2 contain the size results for Granger non-causality tests within the following approaches: the low-frequeny VAR (LF), the unrestricted VAR (UNR), reduced rank restrictions using canonical correlations analysis (CCA) or partial least squares (PLS), and with three imposed factors using the HAR model (HAR), the *max*-test (*max*) and the Bayesian mixed-frequency VAR (BMF). With respect to CCA and PLS, the outcomes for $r = 2$ turn out to be qualitatively very similar to the $r = 1$ case, showing that the methods are apparently not affected by the presence of misspecification. Consequently, we only show the outcomes for $r = 1$ to save on space. Furthermore, aside from the results of the standard Wald tests, we only display the outcomes for the best performing bootstrap variant, which are denoted with a superscript '$*$' (e.g., CCA* is the bootstrap CCA-based Wald test). In the case of testing causality in the direction from the high- to the low-frequency series this turns out to be the bootstrap that imposes the null hypothesis, whereas the unrestricted variant dominates for the reverse direction.[23] For both methods we set the lag length within the bootstrap equal to $p = 1$.

When testing causality from y to $X^{(m)}$ size distortions may occur due to computing a joint test on mp parameters from m different equations in the system. In order to address this issue, we take the maximum of the Wald statistics computed equation by equation. We denote these tests by adding a subscript 'b', e.g., CCA$_b$. For the tests with asymptotic critical values we apply a Bonferroni correction to control the size under multiple testing (Dunn, 1961). In similar spirit as in White (2000), the bootstrap implementations of these tests automatically provide an implicit Bonferroni-type correction for multiple testing, as we calculate the corresponding maximum of the single-equation Wald tests within each bootstrap iteration and compare that to the maximum of the original tests. We will refer to all these tests as 'Bonferroni-type' tests to avoid confusion with the *max*-test, and present the corresponding outcomes in Table 3.

Let us start by analyzing the testing behavior from $X^{(m)}$ to y. First, LF has size close to the nominal one, which is not surprising given that the flat aggregation scheme is correct in this white noise DGP (no matter the value of σ_{HL}). The unrestricted VAR, however, incurs some size distortions for small T due to parameter proliferation. Reduced rank restrictions based on CCA and PLS yield considerable size distortions, whereby the imposed HAR-type factor structure delivers good results (despite being slightly oversized

[22] Outcomes for $\alpha = 10\%$ and $\alpha = 1\%$ are available upon request.

[23] The unrestricted bootstrap is seriously oversized testing causality from $X^{(m)}$ to y, while the other way around the bootstrap under the null is mildly oversized for small samples. All outcomes not shown here are available upon request.

Table 1: Size of Granger Non-Causality Tests from $X^{(m)}$ to y

$X^{(m)}$ to y	T	LF	UNR	CCA	PLS	HAR	max	BMF	UNR*	CCA*	PLS*	HAR*
	50	5.4	13.4	18.8	11.4	6.9	4.3	5.7	5.0	4.6	4.9	5.2
$\sigma_{HL} = 0$	250	4.7	6.4	28.6	12.4	5.4	4.4	5.2	4.9	4.7	5.3	5.3
	500	4.9	5.6	29.2	12.8	5.2	4.1	5.3	5.1	5.2	6.2	5.3
	50	5.3	12.6	20.9	16.7	6.6	4.6	5.6	4.7	5.6	8.2	5.0
$\sigma_{HL} = -0.05$	250	4.8	5.8	28.3	17.1	5.5	4.4	5.2	4.7	4.8	9.2	5.4
	500	4.8	6.2	30.0	16.9	4.4	4.2	5.3	5.6	5.2	8.5	4.5
	50	5.0	12.5	20.0	32.6	6.3	4.6	5.0	4.8	4.8	19.6	4.8
$\sigma_{HL} = -0.1$	250	4.6	7.2	28.8	35.7	5.5	4.8	5.1	5.8	5.8	23.2	5.2
	500	4.7	5.4	28.9	36.5	5.3	4.5	5.2	4.8	6.3	23.0	5.2

Note: For testing Granger non-causality from $X^{(m)}$ to y the figures represent percentages of rejections of the asymptotic Wald test statistics and the restricted bootstrap variants (indicated with a superscript '*') of the following approaches: The low-frequeny VAR (LF), the unrestricted approach (UNR), reduced rank restrictions with one factor using canonical correlations analysis (CCA) or partial least squares (PLS), and with three imposed factors using the HAR model (HAR), the max-test (max) and the Bayesian mixed-frequency approach (BMF). The lag length of the estimated VARs is equal to one. The underlying DGP is found in (27), where the variance-covariance matrix of the error term is equal to Σ_u with $\sigma_{HL} = 0, -0.05, -0.1$.

for $T = 50$). Finally, *max* and BMF show almost perfect size, whereby the former is marginally under- and the latter marginally oversized.

Turning to the outcomes of the bootstrap versions we observe that, independent of the value of σ_{HL}, even for $T = 50$ the actual size of the Wald tests within the unrestricted VAR and reduced rank restriction models with CCA- or HAR-type factors is very close to the nominal one, such that the bootstrap tests clearly dominate their asymptotic counterparts. For PLS-based factors this conclusion only holds in the absence of nowcasting causality, and size distortions arise quickly as we increase the contemporaneous correlation between $X^{(m)}$ and y.

Many of the aforementioned statements carry over to testing causality in the reverse direction: LF delivering nearly perfect size (as it is the correct model under the null), *max* performing very well (being only marginally oversized for small T) and BMF being a bit oversized (by a slightly larger degree than in Table 1).[24] The size distortions incurred by UNR, CCA, PLS and HAR are, however, amplified. The use of the Bonferroni-corrected UNR_b, CCA_b, PLS_b and HAR_b can only mitigate this effect, yet not fully eradicate it. The bootstrap version does eliminate the size distrotions, though: actual size of UNR*, PLS* and HAR* becomes oftentimes very close to 5%; only CCA* remains a bit oversized for $T = 50$. The Bonferroni-type tests UNR^*_b, CCA^*_b, PLS^*_b and HAR^*_b all have size very close to the nominal one for all sample sizes.

To sum up, the Granger non-causality tests with size close to the nominal one are LF, *max* and BMF to some degree. Furthermore, these are UNR*, CCA*, PLS* (with at most "mild" nowcasting causality) and HAR* when testing the direction from $X^{(m)}$ to y, and UNR*, CCA*, PLS* and HAR* as well as their Bonferroni-type counterparts when testing the reverse direction. Consequently, these are the cases we focus on when analyzing power (and when dealing with real data in Section 5).

4.2 Power

Tables 4 and 5 contain the corresponding outcomes under the alternative hypothesis. Recall that for the direction from $X^{(m)}$ to y we consider three different values of β within the first power DGP, $\beta = 0.5, 0.8$ and 2. For the second power DGP we fix $\beta = 0.5$. For the reverse direction we always keep $\gamma_{2,1} = 0.2$.

Let us again start with the direction from the high- to the low-frequency variable and first focus on the outcomes for the first power DGP, i.e., the top three blocks of Table 4. Observe (i) how power reaches one asymptotically for all approaches, and (ii) how larger values of β imply an increase in the rejection frequencies for a given T. Naturally, for a large enough β all tests have power equal to one, irrespective of the sample size (as nearly happens for $\beta = 2$). So, in order to compare our different tests let us focus on $\beta = 0.5$.

The asymptotic Wald test within the low-frequency VAR still performs very well, with the highest rejection frequency for $T = 50$. Note, however, that the Granger causality feature in the data does not get averaged out by temporal aggregation in the first power DGP. The bootstrapped version of the Wald test after HAR-type restrictions have been imposed in a reduced rank regression perform almost as well as LF. *max*, BMF and PLS* appear to fall a bit short, but catch up quickly as either T or β grows. A similar

[24] Strangely, BMF seems to be rejected less and less for growing σ_{HL}, with the effect being strongest for small T.

Table 2: Size of Granger Non-Causality Tests from y to $X^{(m)}$

y to $X^{(m)}$	T	LF	UNR	CCA	PLS	HAR	max	BMF	UNR*	CCA*	PLS*	HAR*
	50	5.7	91.1	82.0	58.6	59.8	6.1	7.9	4.6	8.8	5.0	5.2
$\sigma_{HL} = 0$	250	5.5	11.5	11.8	10.6	10.2	4.6	10.4	5.2	5.3	5.0	5.3
	500	5.2	6.7	7.4	6.8	6.6	5.9	7.8	4.6	4.8	4.6	4.8
	50	5.6	92.4	83.8	62.0	60.6	5.7	4.8	5.6	9.4	4.9	5.6
$\sigma_{HL} = -0.05$	250	5.4	10.9	13.2	11.9	10.3	4.8	8.9	4.7	4.5	4.4	4.8
	500	5.0	7.4	9.2	8.7	7.3	5.5	7.3	5.4	5.4	5.3	5.2
	50	5.7	92.1	87.1	64.2	58.5	5.7	0.5	5.5	9.2	4.0	4.4
$\sigma_{HL} = -0.1$	250	4.7	11.0	20.5	14.8	10.5	5.0	5.4	5.2	6.3	5.1	5.3
	500	5.3	7.4	13.8	10.7	7.2	5.5	5.0	5.3	5.7	4.2	5.2

Note: For testing Granger non-causality from y to $X^{(m)}$ the figures represent percentages of rejections of the asymptotic Wald test statistics and the unrestricted bootstrap variants (indicated with a superscript '*') of the following approaches: The low-frequency VAR (LF), the unrestricted approach (UNR), reduced rank restrictions with one factor using canonical correlations analysis (CCA) or partial least squares (PLS), and with three imposed factors using the HAR model (HAR), the max-test (max) and the Bayesian mixed-frequency approach (BMF). The lag length of the estimated VARs is equal to one. The underlying DGP is found in (27), where the variance-covariance matrix of the error term is equal to Σ_u with $\sigma_{HL} = 0, -0.05, -0.1$.

Table 3: Size of Granger Non-Causality Tests from y to $X^{(m)}$ (Bonferroni)

y to $X^{(m)}$	T	UNR_b	CCA_b	PLS_b	HAR_b	BMF_b	UNR_b^*	CCA_b^*	PLS_b^*	HAR_b^*
	50	16.9	17.3	16.4	13.9	2.9	3.4	4.6	3.9	3.7
$\sigma_{HL} = 0$	250	12.7	13.6	13.1	12.7	5.1	4.6	5.0	5.0	4.9
	500	12.4	12.4	12.1	12.2	4.4	5.0	5.0	4.6	4.9
	50	17.9	18.9	18.3	15.2	2.1	3.8	4.8	4.0	4.0
$\sigma_{HL} = -0.05$	250	13.4	14.5	14.4	12.8	4.8	5.4	5.2	5.4	5.4
	500	12.3	13.7	14.8	12.5	4.0	5.0	5.2	5.0	4.6
	50	17.2	20.2	21.6	15.4	0.4	3.3	5.0	4.3	4.4
$\sigma_{HL} = -0.1$	250	12.0	15.4	17.0	12.2	3.3	4.9	5.3	4.8	4.7
	500	11.9	14.9	17.0	11.7	3.3	5.4	5.2	5.1	5.2

Note: For testing Granger non-causality from y to $X^{(m)}$ the figures represent percentages of rejections of the Bonferroni-type tests (indicated with a subscript 'b') for both the asymptotic Wald test statistics and the unrestricted bootstrap variants (indicated with a superscript '*') of the following approaches: The unrestricted approach (UNR), reduced rank restrictions with one factor using canonical correlations analysis (CCA) or partial least squares (PLS), and with three imposed factors using the HAR model (HAR) and the Bayesian mixed-frequency approach (BMF). For the low-frequency VAR (LF) and the *max*-test (*max*) the Bonferroni-type test is not applicable. The lag length of the estimated VARs is equal to one. The underlying DGP is found in (27), where the variance-covariance matrix of the error term is equal to Σ_u with $\sigma_{HL} = 0, -0.05, -0.1$.

Table 4: Power of Granger Non-Causality Tests from $X^{(m)}$ to y

$X^{(m)}$ to y	T	LF	max	BMF	UNR*	CCA*	PLS*	HAR*
Power-DGP 1, $\beta = 0.5$	50	64.1	39.6	52.8	23.1	12.7	54.0	62.6
	250	99.9	100.0	99.8	99.8	96.0	98.7	100.0
	500	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Power-DGP 1, $\beta = 0.8$	50	93.6	79.4	92.2	54.6	26.6	77.9	93.0
	250	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	500	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Power-DGP 1, $\beta = 2$	50	100.0	100.0	100.0	100.0	85.3	98.4	100.0
	250	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	500	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Power-DGP 2, $\beta = 0.5$	50	5.8	14.5	21.9	9.7	8.0	25.1	22.3
	250	5.0	75.9	79.2	71.2	58.8	79.7	89.4
	500	4.7	98.6	99.0	98.3	87.6	97.3	99.8

Note: For testing Granger non-causality from $X^{(m)}$ to y the figures represent percentages of rejections of the asymptotic Wald test statistics of the low-frequeny VAR (LF), the max-test (max) and the Bayesian mixed-frequency approach (BMF). Furthermore, it contains percentages of rejections of the restricted bootstrap variants (indicated with a superscript '*') of the unrestricted approach (UNR), reduced rank restrictions with one factor using canonical correlations analysis (CCA) or partial least squares (PLS), and with three imposed factors using the HAR model (HAR). The lag length of the estimated VARs is equal to one. The underlying DGP is found in (28) or (29), where the variance-covariance matrix of the error term is equal to Σ_u with $\sigma_{HL} = -0.05$. In power-DGP 1 the Granger causality determining coefficients do not sum up to zero, whereas they do so in power-DGP 2. See Section 4 for details. $\beta = 0.5, 0.8, 2$ in power-DGP 1 and $\beta = 0.5$ in power-DGP 2.

observation holds for UNR* and CCA*, whereby they seem to be markedly less powerful for small T.

These conclusions generally carry over to the second power DGP, with two exceptions: First and foremost, the outcomes of LF show that Average sampling of $X^{(m)}$ annihilates the causality feature between the series for all sample sizes (power actually almost coincides with the nominal size of 5%). Second, the rejection frequencies are lower than they were in the corresponding setting of the first power DGP.

Turning to the direction from y to $X^{(m)}$, it becomes obvious that the outcomes for the first and second power DGP are, once again, similar, with the same two exceptions mentioned before. This being said and leaving aside the ones for LF due Granger causality not being invariant to temporal aggregation (Marcellino, 1999), it suffices to analyze the results of the first power DGP.

Both *max* and BMF have higher power than the bootstrapped Wald tests corresponding to the reduced rank restrictions approaches. However, recall that BMF and *max* were somewhat oversized, especially for small T, which may inflate their power. However, so far we did not consider the Bonferroni-type tests based on the maximum of the individual Wald tests. Given the way in which *max* is constructed though, it seems more natural to compare it to the Bonferroni-type counterparts of the bootstrap tests. Indeed, both approaches rely on a statistic that is computed as the maximum of a set of test statistics.[25] Here it turns out that, for a given T, almost all approaches are more powerful than *max*. The fact that UNR$_b^*$, PLS$_b^*$ and HAR$_b^*$ are all slightly undersized for small T, while *max* is marginally oversized, even strengthens the aforementioned conclusions. Also note that the bootstrap tests based on the reduced rank restrictions seem to be slightly more powerful than UNR$_b^*$, confirming again (though less pronounced than in the other direction) that while the bootstrap can correct size of the unrestricted MF-VAR approach well, the parameter proliferation problem continues to have a negative effect on power.

Combining the power and size outcomes, it seems that BMF, HAR*, *max* (for large enough β when T is small) and PLS* (in the absence of nowcasting causality) are the dominant Granger non-causality testing approaches as far as the direction from $X^{(m)}$ to y is concerned. For the reverse direction of causality, CCA$_b^*$, PLS$_b^*$ and HAR$_b^*$ as well as UNR$_b^*$ and *max* (though both to a lesser degree) appear superior.

5 Application

We apply the approaches described in Section 3 to a MF-VAR consisting of the monthly growth rate of the U.S. industrial production index (*ipi* hereafter), a measure of business cycle fluctuations, and the logarithm of daily bipower variation (*bv* hereafter) of the S&P500 stock index, a realized volatility measure more robust to jumps than the realized variance. 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 predicted using return volatility, has been granted

[25] There is a difference in terms of the underlying regression to obtain the various test statistics, of course. Ghysels et al. (2015a) consider univariate MIDAS regressions with leads, whereas the Bonferroni-type tests are based on estimated coefficients from different equations, i.e., the ones for $X^{(m)}$, of the corresponding system regression.

Table 5: Power of Granger Non-Causality Tests from y to $X^{(m)}$

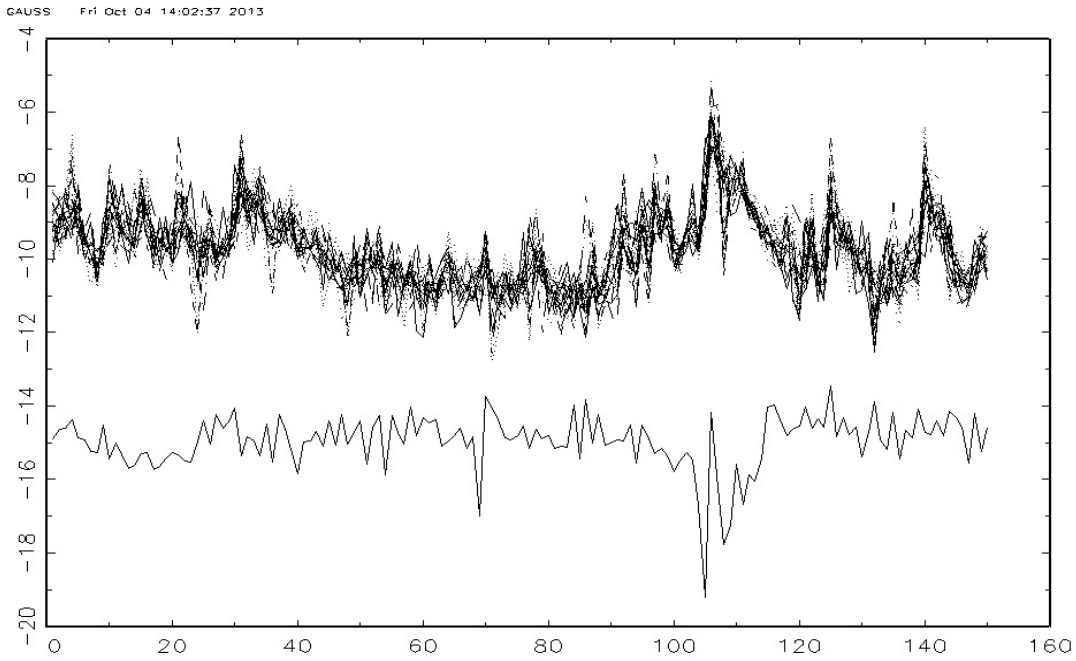
y to $X^{(m)}$	T	LF	max	BMF	UNR*	CCA*	PLS*	HAR*	UNR$^*_b$	CCA$^*_b$	PLS$^*_b$	HAR$^*_b$
	50	6.9	8.0	15.8	5.6	9.9	6.2	6.2	6.8	11.4	10.7	10.8
Power-DGP 1	250	19.9	24.2	31.0	17.7	20.0	20.3	19.3	51.0	54.3	54.7	53.8
	500	36.4	53.2	49.1	39.3	44.7	43.9	42.4	83.1	85.2	85.2	84.6
	50	4.7	7.9	18.4	5.7	10.6	5.7	5.3	5.9	10.0	8.9	8.0
Power-DGP 2	250	5.0	22.9	25.1	13.5	14.6	14.3	14.3	32.2	34.7	34.2	32.6
	500	5.2	49.9	33.7	26.6	29.9	28.4	27.7	57.3	62.2	62.2	59.3

Note: For testing Granger non-causality from y to $X^{(m)}$ the figures represent percentages of rejections of the asymptotic Wald test statistics of the low-frequeny VAR (LF), the max-test (max) and the Bayesian mixed-frequency approach (BMF). Furthermore, it contains percentages of rejections of the unrestricted bootstrap variants (indicated with a superscript '*') of the unrestricted approach (UNR), reduced rank restrictions with one factor using canonical correlations analysis (CCA) or partial least squares (PLS), and with three imposed factors using the HAR model (HAR). For the latter set of approaches, it also shows the outcomes for the Bonferroni-type tests (indicated with a subscript 'b'). The lag length of the estimated VARs is equal to one. The underlying DGP is found in (28) or (29), where the variance-covariance matrix of the error term is equal to Σ_u with $\sigma_{HL} = -0.05$. In power-DGP 1 the Granger causality determining coefficients do not sum up to zero, whereas they do so in power-DGP 2. See Section 4 for details. $\gamma_{2,1} = 0.2$ in both power-DGPs.

comparably few attention (examples are Schwert, 1989a, Mele, 2007, and Andreou, Osborn, and Sensier, 2000). Instead of using an aggregate measure of volatility such as, e.g., monthly realized volatilities (Chauvet, Senyuz, and Yoldas, 2015) or monthly GARCH estimated variances, we use daily bipower variation computed on 5-minute returns obtained from the database in Heber, Lunde, Shephard, and Sheppard (2009). 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 $T = 150$. 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.[26] Consequently, we have $BV_t^{(20)} = (bv_t^{(20)}, bv_{t-1/20}^{(20)}, \ldots, bv_{t-19/20}^{(20)})'$. Figure 2 plots the data.

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



Note: This figure shows the monthly growth rate of the U.S. industrial production index (lower line), i.e., *ipi*, and the logarithm of daily bipower variation of the S&P500 stock index (top lines), i.e., $BV^{(20)}$, for the time period from January 2000 to June 2012. The graph for the former is shifted downwards to enable a visual separation of *ipi* from the *bv*-lines.

Table 6 contains the outcomes of Granger non-causality tests for all approaches discussed in Section 3. As mentioned before, we disregard the cases that showed considerable

[26] Whenever a month contains more than 20 working days we disregard the corresponding amount of days at the beginning of the month. In June 2012, e.g., there are 21 working days such that we do not consider June 1. For May 2012 we disregard 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 $bv_t^{(20)}$.

size distortions in the Monte Carlo analysis. Note that a lag length of $p = 1$ and 2 is considered (the same for the bootstrap) and that the numbers represent p-values (in percentages). For reduced rank restrictions using CCA and PLS we consider one up to three factors, whereby we only show the outcomes for $r = 1$ and $r = 2$ for representational ease; $r = 3$ gives similar results. For the non-bootstrap tests, except the *max*-test and the Wald-test within the Bayesian mixed-frequency VAR,[27] we also consider a heteroscedasticity consistent variant of (14) by computing a robust estimator of Ω (see Ravikumar, Ray, and Savin, 2000) to account for the potential presence of a time varying multivariate process:

$$\hat{\Omega}_R = T((W'W)^{-1} \otimes I_{m+1})\hat{S}_0((W'W)^{-1} \otimes I_{m+1}), \tag{30}$$

where

$$\hat{S}_0 = \frac{1}{T}\sum_{t=1}^{T}(W_t \otimes \hat{u}_t)'(W_t \otimes \hat{u}_t). \tag{31}$$

For the bootstrap tests we achieve the robustness to heteroskedasticity by implementing the wild bootstrap version as described in Section 3.1.4.

If we consider the test statistics that perform best in our Monte Carlo experiment, the outcomes above clearly point towards Granger-causality from $BV^{(20)}$ to *ipi*. This result supports Andreou et al. (2000) concluding that "[...] volatilities may also be useful [...] indicators for both the growth and volatility of industrial production" (p. 15). The situation is less clear cut when looking at Granger-causality from the low-frequency variable to the high-frequency volatility measures. Indeed, *max* and most Bonferroni-type tests seem to reject the null of no Granger-causality, whereas the joint tests do not. However, the higher power obtained on the maximum of the individual Wald-type tests would favor the presence of Granger-causality in the direction from business cycle movements to financial uncertainties as well. A more careful investigation, out of the scope of this paper, could be done on different subsamples in order to analyze whether, e.g., periods before, during or after the financial crises lead to similar conclusions.

6 Conclusion

We investigate Granger non-causality testing in a mixed-frequency VAR, where the mismatch between the sampling frequencies of the variables under consideration is large, causing estimation and inference to be potentially problematic. To avoid this issue we discuss two parameter reduction techniques in detail, reduced rank restrictions and a Bayesian MF-VAR approach, and compare them to (i) a common low-frequency VAR, (ii) the *max*-test approach and (iii) the unrestricted VAR in terms of their Granger non-causality testing behavior. To further improve their finite sample test properties we also consider two bootstrap variants for the reduced rank regression approaches (and the unrestricted VAR).

[27] Note that, as discussed in Section 3.2.2, there is a one-to-one correspondence between the OLS estimator of the augmented model and the prior setting we consider in our Bayesian framework. As we did not discuss setting that corresponds to a robust version of the Wald test we disregard from that test variant here.

Table 6: Testing for Granger Causality between $BV^{(20)}$ and *ipi*

$BV^{(20)}$ to *ipi*	$p=1$		$p=2$	
	Wald	Robust	Wald	Robust
LF	0.0	0.2	0.1	2.1
max	0.1	./.	0.0	./.
BMF	14.4	./.	21.8	./.
UNR*	15.0	17.4	7.3	8.7
CCA1*	2.9	7.9	12.4	20.4
CCA2*	14.5	16.2	19.2	25.1
PLS1*	0.1	0.7	0.2	1.9
PLS2*	0.6	2.8	0.1	1.4
HAR*	1.2	2.2	1.5	2.4
ipi to $BV^{(20)}$	$p=1$		$p=2$	
	Wald	Robust	Wald	Robust
LF	0.1	1.5	0.4	5.0
max	4.4	./.	1.1	./.
BMF	23.5	./.	42.5	./.
UNR*	28.9	34.2	25.8	26.4
CCA1*	25.8	33.5	20.5	38.4
CCA2*	35.9	46.1	17.6	34.3
PLS1*	15.8	24.6	11.4	24.4
PLS2*	18.3	26.9	10.3	18.0
HAR*	19.2	27.0	11.4	18.6
UNR_b^*	5.4	8.9	20.0	22.9
$CCA1_b^*$	4.1	10.5	7.8	16.9
$CCA2_b^*$	5.0	10.8	7.7	15.1
$PLS1_b^*$	1.5	4.3	1.5	9.5
$PLS2_b^*$	1.2	5.2	1.1	4.0
HAR_b^*	2.3	6.5	6.2	14.1

Note: For testing Granger non-causality between $BV^{(20)}$ and *ipi* the figures represent *p*-values (in percentages) of the asymptotic Wald test statistics, the restricted bootstrap variants for the direction from $BV^{(20)}$ to *ipi* and the unrestricted bootstrap variants for the direction from *ipi* to $BV^{(20)}$ (bootstrap variants are indicated with a superscript '*', the Bonferroni-type tests with a subscript 'b'). For all approaches but the *max*-test (*max*) and the Bayesian MF-VAR (BMF) a robustified version is computed as well. The lag length of the estimated VARs is equal to one or two. For CCA and PLS the results for one and two factors are displayed.

For both directions of causality we find a different set of tests to result in the best Granger non-causality testing behavior. For the direction from the high- to the low-frequency series, standard testing within the Bayesian mixed-frequency VAR, the *max*-test of Ghysels et al. (2015a), and the restricted bootstrap version of the Wald test in a reduced rank regression after HAR-type factors have been imposed or PLS-type factors have been computed, the latter of which being restricted to nowcasting non-causality (Götz and Hecq, 2014), perform best. For the reverse direction, the unrestricted bootstrap variants of the Bonferroni-corrected Wald tests within the following models dominate: the unrestricted VAR and reduced rank regressions with CCA-, PLS- or HAR-based factors.

An application investigating the presence of a causal link between business cycle fluctuations and uncertainty in financial markets illustrates the practical usefulness of these approaches. While Granger causality from uncertainty in financial markets to business cycle fluctuations was clearly supported by the data, evidence for causality in the reverse direction only comes from a subset of the tests, yet the more powerful ones according to our simulation results.

A Bayesian VAR estimation with mixed-frequency prior variances

Properly accounting for the high-frequency time difference between the variables involved implies the following prior variances:

$$Var[\gamma_{i,j}^{(k)}] = \begin{cases} \phi \frac{\lambda^2}{(k+(j-2)/m)^2} \mathbb{S}_{LH} & \text{for } i = 1, j > 1 \\ \phi \frac{\lambda^2}{(k+(2-i)/m)^2} \mathbb{S}_{HL} & \text{for } j = 1, i > 1 \\ \frac{\lambda^2}{(k+(j-i)/m)^2} & \text{else} \end{cases}, \tag{32}$$

Now, 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 (4) in the following way:

$$\underbrace{vec(Z)}_{(m+1)T\times 1} = \underbrace{\underline{Z}^*}_{(m+1)T\times(m+1)n} \underbrace{vec(B^*)}_{(m+1)n\times 1} + \underbrace{vec(E)}_{(m+1)T\times 1}, \tag{33}$$

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 Σ, we impose a normal inverted Wishart prior (at the cost of having to set $\phi = 1$) with the following form (Kadiyala and Karlsson, 1997):

$$vec(B^*)|\Sigma \sim N(vec(B_0^*), [Z_0^{*\prime}(\Sigma^{-1} \otimes I_T)Z_0^*]^{-1}) \text{ and } \Sigma \sim iW(V_0, v_0), \tag{34}$$

where $Z_0^* = I_{m+1} \otimes Z_0$ with Z_0 being a $(T \times n)$-matrix. B_0^* and Z_0 have to be chosen as to let expectations and variances of the elements in B^* coincide with the moments in (26). Likewise, V_0 and v_0 need to be set such that $\mathbb{E}[\Sigma] = \Sigma_d$.

Similar to Banbura et al. (2010), we can show that adding auxiliary dummy variables Y_d and X_d, the precise composition of which is given in Appendix B, to (33) is equivalent to imposing the normal inverted Wishart prior in (34). To this end, let

$$\begin{aligned} &B_0^{*aux} = (X_d'X_d)^{-1}X_d'Y_d, \\ &Z_0^{aux} = X_d \\ &V_0 = (Y_d - X_dB_0^{*aux})'(Y_d - X_dB_0^{*aux}) \text{ and} \\ &v_0 = m + 3 = T_d - n^{aux} + 2, \end{aligned} \tag{35}$$

where $n^{aux} = m(2p+1)$ and $T_d = n^{aux} + (m+1)$, and subsequently set

$$\begin{aligned} &vec(B_0^*) = S'vec(B_0^{*aux}) \text{ and} \\ &Var[vec(B^*)] = S'[Z_0^{*aux\prime}(\Sigma^{-1} \otimes I_{T_d})Z_0^{*aux}]^{-1}S, \end{aligned} \tag{36}$$

with $Z_0^{*aux} = I_{m+1} \otimes Z_0^{aux}$ and where S is an $[(m+1)n^{aux} \times (m+1)n]$-dimensional selection matrix, the precise construction of which is given in Appendix C.

Remark 12. 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 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^*.

In order to augment the model in (33), let us define $\underline{Z}_j^d = (\underline{Z}', (X_dS_j)')'$, where S_j is

the j^{th} block of the selection matrix S (see Appendix C for details), $j = 1, \ldots, m+1$.[28] Then, the augmented system becomes

$$\underbrace{vec(Z_*)}_{(m+1)(T+T_d)\times 1} = \underbrace{\underline{Z}^*_*}_{(m+1)(T+T_d)\times(m+1)n} \underbrace{vec(B^*)}_{(m+1)n\times 1} + \underbrace{vec(E_*)}_{(m+1)(T+T_d)\times 1}, \tag{37}$$

where $Z_* = (Z', Y_d')'$, $E_* = (E', E_d')'$ and $\underline{Z}^*_*$ is block diagonal with $\underline{Z}^*_* = diag\{\underline{Z}^d_1, \underline{Z}^d_2, \ldots, \underline{Z}^d_{m+1}\}$. The posterior then has the form

$$vec(B^*)|\Sigma, Z \sim N(vec(\hat{B}^*), [\underline{Z}^{*\prime}_*(\Sigma^{-1} \otimes I_{T+T_d})\underline{Z}^*_*]^{-1}) \text{ and } \Sigma|Z \sim iW(\hat{V}, T+m+3), \tag{38}$$

where $\hat{V} = \hat{E}_*'\hat{E}_*$ with $vec(\hat{E}_*) = vec(Z_*) - \underline{Z}^*_* vec(\hat{B}^*)$.[29] Note that the posterior mean of the coefficients boils down to

$$vec(\hat{B}^*) = [\underline{Z}^{*\prime}_*(\Sigma^{-1} \otimes I_{T+T_d})\underline{Z}^*_*]^{-1}\underline{Z}^{*\prime}_*(\Sigma^{-1} \otimes I_{T+T_d})vec(Z_*), \tag{39}$$

i.e., 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 (26). An example with $p = 1$ and $m = 2$ illustrates the auxiliary dummy variables approach and is provided in Appendix D.

B The auxiliary dummy variables with mixed-frequency prior variances

The auxiliary dummy variables that imply a matching of the prior moments turn out to be

$$\underbrace{Y_d}_{T_d\times(m+1)} = \begin{pmatrix} \mathbf{0}_{2(m-1)\times 1} & \\ \sigma_L\rho^m/\lambda & D_\rho \otimes (0, \sigma_H/\lambda)' \\ 0 & \\ \hline \mathbf{0}_{(m(2p-1)-1)\times(m+1)} & \\ diag(\sigma_L, \sigma_H, \ldots, \sigma_H) & \\ \mathbf{0}_{1\times(m+1)} & \end{pmatrix}, \tag{40}$$

$$\underbrace{X_d}_{T_d\times n^{aux}} = \begin{pmatrix} J^1_p \otimes diag(\sigma_L, \sigma_H)/\lambda & \mathbf{0}_{2pm\times(m-1)} & \mathbf{0}_{2pm\times 1} \\ \mathbf{0}_{(m-1)\times 2pm} & J^2_p\sigma_H/\lambda & \mathbf{0}_{(m-1)\times 1} \\ \mathbf{0}_{(m+1)\times 2pm} & \mathbf{0}_{(m+1)\times(m-1)} & \mathbf{0}_{(m+1)\times 1} \\ \mathbf{0}_{1\times 2pm} & \mathbf{0}_{1\times(m-1)} & \epsilon \end{pmatrix}, \tag{41}$$

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

$$\begin{aligned} J^1_p &= diag(\tfrac{1}{m}, \tfrac{2}{m}, \ldots, 1, \tfrac{m+1}{m}, \ldots, 2, \ldots, \ldots, p), \\ J^2_p &= diag(p + \tfrac{1}{m}, p + \tfrac{2}{m}, \ldots, p + \tfrac{m-1}{m}). \end{aligned} \tag{42}$$

[28] Here, S_j picks the elements of $Z^{aux}_0 = X_d$ corresponding to column j of B^*.

[29] In practice, we estimate Σ 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 (37). Again, a sample size correction alleviates potential size distortions (see Section 3.1.2). 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 $\underline{Z}^d_i$, where $Z_{\cdot,i}$ and $Y_{d,i}$ denote the i^{th} columns of Z and Y_d, respectively.

The last line of both, Y_d and X_d, corresponds to the diffuse prior for the intercept (ϵ is a very small number), the block above imposes the prior for Σ and the remaining blocks set the priors for the coefficients $\gamma_{i,j}^{(k)}$. As in 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 for $x_t^{(m)}$.

C 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 (34), in such a way as to let the variances of the corresponding coefficients coincide with the prior variances in (26), it turns out that we need to set

$$Var[vec(B^*)|\Sigma] = \begin{pmatrix} \sigma_L^2\Omega_0^{(2)} & \mathbf{0}_{n\times n} & \dots & \dots & \mathbf{0}_{n\times n} \\ \mathbf{0}_{n\times n} & \sigma_H^2\Omega_0^{(2)} & \mathbf{0}_{n\times n} & \dots & \mathbf{0}_{n\times n} \\ \vdots & \mathbf{0}_{n\times n} & \sigma_H^2\Omega_0^{(3)} & \dots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \mathbf{0}_{n\times n} \\ \mathbf{0}_{n\times n} & \mathbf{0}_{n\times n} & \dots & \mathbf{0}_{n\times n} & \sigma_H^2\Omega_0^{(m+1)} \end{pmatrix}, \tag{43}$$

where

$$\Omega_0^{(i)} = diag(\begin{array}{l} \frac{\lambda^2}{(1+(2-i)/m)^2\sigma_L^2}, \frac{\lambda^2}{(1+(2-i)/m)^2\sigma_H^2}, \frac{\lambda^2}{(1+(3-i)/m)^2\sigma_H^2}, \dots, \frac{\lambda^2}{(1+1+(1-i)/m)^2\sigma_H^2}, \dots, \\ \dots, \frac{\lambda^2}{(p+(2-i)/m)^2\sigma_L^2}, \frac{\lambda^2}{(p+(2-i)/m)^2\sigma_H^2}, \dots, \frac{\lambda^2}{(p+1+(1-i)/m)^2\sigma_H^2}, \frac{1}{\epsilon^2}) \end{array} \tag{44}$$

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 (26)). Let us form an auxiliary matrix Ω_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 Ω_0^{aux}:

$$\Omega_0^{aux} = diag(\begin{array}{l} \frac{\lambda^2}{(1/m)^2\sigma_L^2}, \frac{\lambda^2}{(1/m)^2\sigma_H^2}, \dots, \frac{\lambda^2}{1^2\sigma_L^2}, \frac{\lambda^2}{1^2\sigma_H^2}, \frac{\lambda^2}{(1+1/m)^2\sigma_L^2}, \frac{\lambda^2}{(1+1/m)^2\sigma_H^2}, \dots, \frac{\lambda^2}{2^2\sigma_H^2}, \\ \dots, \dots, \frac{\lambda^2}{p^2\sigma_L^2}, \frac{\lambda^2}{p^2\sigma_H^2}, \frac{\lambda^2}{(p+1/m)^2\sigma_H^2}, \dots, \frac{\lambda^2}{(p+1-1/m)^2\sigma_H^2}, \frac{1}{\epsilon^2}). \end{array} \tag{45}$$

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 Ω_0^{aux} we can then derive $Z_0^{aux} = X_d$ by using $\Omega_0^{aux} = (Z_0^{aux'} Z_0^{aux})^{-1}$.

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:

$$S_j = (S_j^1, S_j^2, \underline{1}_{n^{aux}}) \text{ for } j = 2, \dots, m+1, \tag{46}$$

where

$$S_j^1 = (\quad \underline{1}_{2m-(2j-3)}, \underline{1}_{2m-(2j-4)}, \underline{1}_{2m-(2j-6)}, \ldots, \underline{1}_{4m-(2j-2)}, \underline{1}_{4m-(2j-3)}, \ldots, \underline{1}_{6m-(2j-2)}, \ldots, \ldots, \underline{1}_{2pm-(2j-3)}, \underline{1}_{2pm-(2j-4)}, \ldots, \underline{1}_{2pm}), \tag{47}$$

$$S_j^2 = \begin{cases} \emptyset & \text{for } j = m+1 \\ (\underline{1}_{2pm+1}, \underline{1}_{2pm+2}, \ldots, \underline{1}_{2pm+m-(j-1)}) & \text{else} \end{cases} \tag{48}$$

and

$$S_1 = S_2. \tag{49}$$

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.

D Example of the auxiliary dummy variables approach with mixed-frequency prior variances

Let $p = 1$ and $m = 2$. The prior beliefs in (26) corresponding to this setup, and written in terms of $vec(B^*) = vec((B', \mu)')$, are given by

$$\mathbb{E}(vec(B^*)) = vec\begin{pmatrix} \rho^2 & 0 & 0 \\ 0 & \rho^2 & \rho \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and

$$Var(vec(B^*)) = \lambda^2 \begin{pmatrix} 1 & 0 & & & & & \ldots & & & & & 0 \\ 0 & \mathbb{S}_{LH} & 0 & & & & & \ldots & & & & 0 \\ 0 & 0 & \frac{1}{(3/2)^2}\mathbb{S}_{LH} & 0 & & & & & \ldots & & & 0 \\ 0 & \ldots & 0 & \frac{\sigma_L^2}{\epsilon^2\lambda^2} & 0 & & & & & \ldots & & 0 \\ 0 & & \ldots & 0 & \mathbb{S}_{HL} & 0 & & & & & \ldots & 0 \\ 0 & & & \ldots & 0 & 1 & 0 & & & & & 0 \\ 0 & & & & \ldots & 0 & \frac{1}{(3/2)^2} & 0 & & & \ldots & 0 \\ 0 & & & & & \ldots & 0 & \frac{\sigma_H^2}{\epsilon^2\lambda^2} & 0 & & \ldots & 0 \\ 0 & & & & & & \ldots & 0 & \frac{1}{(1/2)^2}\mathbb{S}_{HL} & 0 & \ldots & 0 \\ 0 & & & & & & & \ldots & 0 & \frac{1}{(1/2)^2} & 0 & 0 \\ 0 & & & & & & & & \ldots & 0 & 1 & 0 \\ 0 & & & & & & & & & \ldots & 0 & \frac{\sigma_H^2}{\epsilon^2\lambda^2} \end{pmatrix}. \tag{50}$$

With $T_d = 9$ and $n^{aux} = 6$ the auxiliary dummy variables look as follows:

$$Y_d = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \frac{\frac{1}{2}\rho\sigma_H}{\lambda} \\ \frac{\rho^2\sigma_L}{\lambda} & 0 & 0 \\ 0 & \frac{\rho^2\sigma_H}{\lambda} & 0 \\ 0 & 0 & 0 \\ \sigma_L & 0 & 0 \\ 0 & \sigma_H & 0 \\ 0 & 0 & \sigma_H \\ 0 & 0 & 0 \end{pmatrix}, X_d = \begin{pmatrix} \frac{\frac{1}{2}\sigma_L}{\lambda} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{\frac{1}{2}\sigma_H}{\lambda} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{\sigma_L}{\lambda} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{\sigma_H}{\lambda} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{\frac{3}{2}\sigma_H}{\lambda} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \epsilon \end{pmatrix}.$$

Let us demonstrate that these auxiliary dummy variables truly imply a matching of the prior moments in (34) and the prior beliefs given above. Simple algebra gives us $Z_0^{aux} = X_d$, $v_0 = 5$ as well as

$$B_0^{*aux} = (X_d'X_d)^{-1}X_d'Y_d = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & \rho \\ \rho^2 & 0 & 0 \\ 0 & \rho^2 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

$$V_0 = (Y_d - X_dB_0^{*aux})'(Y_d - X_dB_0^{*aux}) = \begin{pmatrix} \sigma_L^2 & 0 & \\ 0 & \sigma_H^2 & 0 \\ 0 & 0 & \sigma_H^2 \end{pmatrix}.$$

To obtain $vec(B_0^*)$ we require the selection matrix S. Following the guidelines in Appendix C yields $S = diag\{S_1, S_2, S_3\}$, where the off-diagonal blocks are of dimension 6×4 and

$$S_1 = S_2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, S_3 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Now it is easy to show that, indeed,

$$vec(B_0^*) = S'vec(B_0^{*aux}) = vec\begin{pmatrix} \rho^2 & 0 & 0 \\ 0 & \rho^2 & \rho \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \mathbb{E}(vec(B^*))$$

and $S'[\Sigma \otimes (Z_0^{aux\prime}Z_0^{aux})^{-1}]S = Var(vec(B^*))$ as in equation (50).

Having shown that Y_d and X_d capture the prior moments, we need to augment the MF-VAR. Let us start by re-writing it into the format presented in equation (33):

$$vec\begin{pmatrix} y_2 & x_2^{(2)} & x_{2-1/2}^{(2)} \\ \vdots & \vdots & \vdots \\ y_T & x_T^{(2)} & x_{T-1/2}^{(2)} \end{pmatrix} =$$
$$\left[I_3 \otimes \begin{pmatrix} y_1 & x_1^{(2)} & x_{1/2}^{(2)} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ y_{T-1} & x_{T-1}^{(2)} & x_{T-1-1/2}^{(2)} & 1 \end{pmatrix}\right] \underbrace{vec\begin{pmatrix} \gamma_{1,1} & \gamma_{2,1} & \gamma_{3,1} \\ \gamma_{1,2} & \gamma_{2,2} & \gamma_{3,2} \\ \gamma_{1,3} & \gamma_{2,3} & \gamma_{3,3} \\ \mu_1 & \mu_2 & \mu_3 \end{pmatrix}}_{vec(B^*)} + vec(E).$$

Following the steps outlined at the end of Section 3.2.2 gives us the augmented system:

$$vec\begin{pmatrix} y_2 & x_2^{(2)} & x_{2-1/2}^{(2)} \\ \vdots & \vdots & \vdots \\ y_T & x_T^{(2)} & x_{T-1/2}^{(2)} \\ 0 & 0 & 0 \\ 0 & 0 & \frac{\frac{1}{2}\rho\sigma_H}{\lambda} \\ \frac{\rho^2\sigma_L}{\lambda} & 0 & 0 \\ 0 & \frac{\rho^2\sigma_H}{\lambda} & 0 \\ 0 & 0 & 0 \\ \sigma_L & 0 & 0 \\ 0 & \sigma_H & 0 \\ 0 & 0 & \sigma_H \\ 0 & 0 & 0 \end{pmatrix} = \underline{Z}_*^* vec(B^*) + vec(E_*).$$

with

$$\underline{Z}^*_* = \begin{pmatrix} \begin{matrix} y_1 & x_1^{(2)} & x_{1/2}^{(2)} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ y_{T-1} & x_{T-1}^{(2)} & x_{T-1-1/2}^{(2)} & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{\sigma_L}{\lambda} & 0 & 0 & 0 \\ 0 & \frac{\sigma_H}{\lambda} & 0 & 0 \\ 0 & 0 & \frac{\frac{3}{2}\sigma_H}{\lambda} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \epsilon \end{matrix} & \mathbf{0}_{(T+9)\times 4} & \mathbf{0}_{(T+9)\times 4} \\ \mathbf{0}_{(T+9)\times 4} & \begin{matrix} y_1 & x_1^{(2)} & x_{1/2}^{(2)} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ y_{T-1} & x_{T-1}^{(2)} & x_{T-1-1/2}^{(2)} & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \frac{\sigma_L}{\lambda} & 0 & 0 & 0 \\ 0 & \frac{\sigma_H}{\lambda} & 0 & 0 \\ 0 & 0 & \frac{\frac{3}{2}\sigma_H}{\lambda} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \epsilon \end{matrix} & \mathbf{0}_{(T+9)\times 4} \\ \mathbf{0}_{(T+9)\times 4} & \mathbf{0}_{(T+9)\times 4} & \begin{matrix} y_1 & x_1^{(2)} & x_{1/2}^{(2)} & 1 \\ \vdots & \vdots & \vdots & \vdots \\ y_{T-1} & x_{T-1}^{(2)} & x_{T-1-1/2}^{(2)} & 1 \\ \frac{\frac{1}{2}\sigma_L}{\lambda} & 0 & 0 & 0 \\ 0 & \frac{\frac{1}{2}\sigma_H}{\lambda} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{\sigma_H}{\lambda} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \epsilon \end{matrix} \end{pmatrix}.$$

The GLS estimate of $vec(B^*)$ in the regression above is then a closed-form solution for the posterior mean of the coefficients.

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., D. R. Osborn, and M. Sensier (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.

Andrews, D. W. K. (1994). The large sample correspondence between classical hypothesis tests and Bayesian posterior odds tests. *Econometrica 62*(5), 1207–1232.

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

Bauwens, L., M. Lubrano, and J.-F. Richard (2000). *Bayesian inference in dynamic econometric models*. Oxford University Press.

Bose, A. (1988). Edgeworth correction by bootstrap in autoregressions. *Annals of Statistics 16*, 1709–1722.

Breitung, J. and N. R. Swanson (2002). Temporal aggregation and spurious instantaneous causality in multiple time series models. *Journal of Time Series Analysis 23*, 651–666.

Brüggemann, R., C. Jentsch, and C. Trenkler (2014). Inference in VARs with conditional heteroskedasticity of unknown form. Technical report, Department of Economics, University of Konstanz.

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

Cavaliere, G., A. Rahbek, and A. M. R. Taylor (2012). Bootstrap determination of the co-integration rank in vector autoregressive models. *Econometrica 80*, 1721–1740.

Chauvet, M., Z. Senyuz, and E. Yoldas (2015). What does financial volatility tell us about macroeconomic fluctuations? *Journal of Economic Dynamics and Control 52*(C), 340–360.

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

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

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

Cubadda, G. and A. Hecq (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.

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

Eichler, M. and V. Didelez (2009). On Granger-causality and the effect of interventions in time series. Technical report.

Engle, R. F. and J. G. Rangel (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 M. Marcellino (2014, November). Mixed Frequency Structural Models: Identification, Estimation, And Policy Analysis. *Journal of Applied Econometrics 29*(7), 1118–1144.

Foroni, C., M. Marcellino, and C. Schumacher (2015, 01). Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society Series A 178*(1), 57–82.

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

Ghysels, E. (2015). Macroeconomics and the reality of mixed frequency data. Technical report.

Ghysels, E. and J. I. Miller (2013, June). Testing for Cointegration with Temporally Aggregated and Mixed-frequency Time Series. Working Papers 1307, Department of Economics, University of Missouri.

Ghysels, E., K. Motegi, and J. Hill (2015a). Simple granger causality tests for mixed frequency data. Discussion Paper.

Ghysels, E., K. Motegi, and J. Hill (2015b). Testing for granger causality with mixed frequency data. Discussion Paper.

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

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

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

Gonçalves, S. and B. Perron (2014). Bootstrapping factor-augmented regression models. *Journal of Econometrics 182*, 156–173.

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

Götz, T. B., A. Hecq, and L. Lieb (2015). Real-time mixed-frequency vars: Nowcasting, backcasting and granger causality. Discussion Paper.

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

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

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica 37*(3), 424–438.

Hamilton, J. D. and L. Gang (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.

Heber, G., A. Lunde, N. Shephard, and K. Sheppard (2009). Oxford-man institute's realized library (version 0.2). Oxford-Man Institute, University of Oxford.

Hoerl, A. E. and R. W. Kennard (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics 12*(1), 55–67.

Horowitz, J. L. (2001). The bootstrap. In J. J. Heckman and E. E. Leamer (Eds.), *Handbook of Econometrics*, Volume 5, Chapter 52, pp. 3159–3228. Amsterdam: North Holland Publishing.

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

Kilian, L. (1998). Small-sample confidence intervals for impulse response functions. *Review of Economics and Statistics 80*, 218–230.

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*, pp. 75–91. Studies in Applied Econometrics. Springer-Verlag, Heidelberg.

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

Marsilli, C. (2014). Variable selection in predictive midas models. Working papers 520, Banque du France.

McCracken, M. W., M. T. Owyang, and T. Sekhposyan (2015, October). Real-Time Forecasting with a Large, Mixed Frequency, Bayesian VAR. Working Papers 2015-30, Federal Reserve Bank of St. Louis.

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, August). Mixed-frequency Cointegrating Regressions with Parsimonious Distributed Lag Structures. Working Papers 1211, Department of Economics, University of Missouri.

Paparoditis, E. (1996). Bootstrapping autoregressive and moving average parameter estimates of infinite order vector autoregressive processes. *Journal of Multivariate Analysis 57*, 277–296.

Paparoditis, E. and D. N. Politis (2005). Bootstrap hypothesis testing in regression models. *Statistics & Probability Letters 74*, 356–365.

Park, T. and G. Casella (2008). The Bayesian Lasso. *Journal of American Statistical Association 103*(482), 681–686.

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

Schorfheide, F. and D. Song (2015). Real-time forecasting with a mixed-frequency var. *Journal of Business and Economic Statistics 33*, 366–380. ¡p¿Second version: December 2013; also available as NBER Working Paper 19712¡br /¿ First version: August, 2012; available as FRB Minneapolis Working Paper 701¡/p¿.

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.

Silvestrini, A. and D. Veredas (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.

Sims, C. A. and T. Zha (1998, November). Bayesian Methods for Dynamic Multivariate Models. *International Economic Review 39*(4), 949–68.

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

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

Van Giersbergen, N. P. A. and J. F. Kiviet (1996). Bootstrapping a stable AD model: weak vs strong exogeneity. *Oxford Bulletin of Economics and Statistics 58*, 631–656.

Vogel, C. R. (2002). *Computational methods for inverse problems*. Number 0898715504. Philadelphia: Society for Industrial and Applied Mathematics.

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

White, H. (2000). A reality check for data snooping. *Econometrica 68*.

Zellner, A. (1996). *An introduction to bayesian inference in econometrics*. John Wiley and Sons, Inc.