DISCUSSION PAPER SERIES

No. 9655

TESTING FOR GRANGER CAUSALITY WITH MIXED FREQUENCY DATA

Eric Ghysels, Jonathan B. Hill
and Kaiji Motegi

FINANCIAL ECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP9655.php**

ISSN 0265-8003

TESTING FOR GRANGER CAUSALITY WITH MIXED FREQUENCY DATA

Eric Ghysels, University of North Carolina and CEPR
Jonathan B. Hill, University of North Carolina
Kaiji Motegi, University of North Carolina

Discussion Paper No. 9655
September 2013

Centre for Economic Policy Research
77 Bastwick Street, London EC1V 3PZ, UK
Tel: (44 20) 7183 8801, Fax: (44 20) 7183 8820
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 9655

September 2013

ABSTRACT

Testing for Granger Causality with Mixed Frequency Data*

It is well known that temporal aggregation has adverse effects on Granger causality tests. Time series are often sampled at different frequencies. This is typically ignored, and data are merely aggregated to the common lowest frequency. We develop a set of Granger causality tests that explicitly take advantage of data sampled at different frequencies. We show that taking advantage of mixed frequency data allows us to better recover causal relationships when compared to the conventional common low frequency approach. We also show that the mixed frequency causality tests have higher local asymptotic power as well as more power in finite samples compared to conventional tests.

JEL Classification: C12 and C32
Keywords: granger causality, mixed data sampling (MIDAS), temporal aggression and vector autoregression (VAR)

Eric Ghysels
Department of Economics
University of North Carolina
Gardner Hall CB 3305
Chapel Hill, NC 27599-3305
USA

Email: eghysels@unc.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=135789

Jonathan B. Hill
Department of Economics
Gardner Hall 208B
University of North Carolina
Chapel Hill, NC 27599-3305
USA

Email: jbhill@email.unc.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=177566

Kaiji Motegi
Department of Economics
Gardner Hall 208B
University of North Carolina
Chapel Hill, NC 27599-3305
USA

Email: motegi@email.unc.edu

*The first author acknowledges support of a Marie Curie FP7-PEOPLE-2010-IIF grant and benefited from funding by the Federal Reserve Bank of New York through the Resident Scholar Program as well as the European Central Bank through the Duisenberg Scholar Program.

Submitted 09 September 2013

1 Introduction

It is well known that temporal aggregation may have spurious effects on testing for Granger causality, as noted by Clive Granger himself in a number of papers, see e.g. Granger (1980), Granger (1988), Granger (1995). In this paper we deal with what might be an obvious, yet largely overlooked remedy. Time series processes are often sampled at different frequencies and are typically aggregated to the common lowest frequency to test for Granger causality. The analysis of the present paper pertains to comparing testing for Granger causality with all series aggregated to the common lowest frequency, and testing for Granger causality taking advantage of all the series sampled at whatever frequency they are available. We rely on mixed frequency vector autoregressive models to implement a new class of Granger causality tests.[1]

We show that mixed frequency data Granger causality tests better recover causality patterns in an underlying high frequency process compared to the traditional low frequency approach. We also formally prove that mixed frequency causality tests have higher asymptotic power against local alternatives and show via simulation that this also holds in finite samples involving realistic data generating processes. The simulations indicate that the mixed frequency VAR approach works well for small differences in sampling frequencies - like quarterly/monthly mixtures.

The paper is organized as follows. In Section 2 we first briefly review the Granger causality and MIDAS literatures and then frame mixed frequency VAR models. In Section 3 we develop the mixed frequency data causality tests. Section 4 discusses how we can recover underlying causality using a mixed frequency (henceforth MF) approach compared to a traditional low frequency approach. Section 5 shows that the MF causality tests have higher local asymptotic power than the low frequency ones do. Section 6 reports Monte Carlo simulation results and documents the finite sample power improvements achieved by MF causality test. Section 7 provides some concluding remarks.

2 Mixed Frequency Data Model Specifications

In this section we frame a mixed frequency vector autoregressive (henceforth MF-VAR) model and derive some asymptotic properties. We first provide a short review of the related literature. We then formally present the MF-VAR model. Finally, we establish large sample results for parameter estimators and corresponding Wald statistics.

In the remainder of the paper we will use the following notational conventions. Let $\boldsymbol{A} \in \mathbb{R}^{n \times l}$. The l_2-norm is $|\boldsymbol{A}| := (\sum_{i=1}^{n} \sum_{j=1}^{l} a_{ij}^2)^{1/2} = (\mathrm{tr}[\boldsymbol{A}'\boldsymbol{A}])^{1/2}$; the L_r-norm is $\|\boldsymbol{A}\|_r := (\sum_{i=1}^{n} \sum_{j=1}^{l} E|a_{ij}|^r)^{1/r}$; the determinant is $\det(\boldsymbol{A})$; and the transpose is $\boldsymbol{A}'$. $\mathbf{0}_{n \times l}$ is an $n \times l$ matrix of zeros. $\mathrm{Var}[\boldsymbol{A}]$ is the variance-covariance matrix of a stochastic matrix $\boldsymbol{A}$.

[1] MIDAS, meaning Mi(xed) Da(ta) S(ampling), regression models have been put forward in recent work by Ghysels, Santa-Clara, and Valkanov (2004), Ghysels, Santa-Clara, and Valkanov (2006) and Andreou, Ghysels, and Kourtellos (2010). See Andreou, Ghysels, and Kourtellos (2011) and Armesto, Engemann, and Owyang (2010) for surveys. VAR models for mixed frequency data were independently introduced by Anderson, Deistler, Felsenstein, Funovits, Zadrozny, Eichler, Chen, and Zamani (2012), Ghysels (2012) and McCracken, Owyang, and Sekhposyan (2013). An early example of related ideas appears in Friedman (1962). Foroni, Ghysels, and Marcellino (2013) provide a survey of mixed frequency VAR models and related literature.

2.1 Brief Literature Review

The notion of causality introduced by Granger (1969) is defined in terms of incremental predictive ability, beyond the past observations of a time series process X, by past observations of another time series process Y. Although so-called *Granger causality* has been extended to fairly general settings including nonlinear and random volatility models, it is typically discussed in a linear regression framework, in particular since Sims (1972).

Early contributions by Zellner and Montmarquette (1971) and Amemiya and Wu (1972) pointed out the potentially adverse effects of temporal aggregation on testing for Granger causality. The subject has been extensively researched ever since. See e.g. Granger (1980), Granger (1988), Lütkepohl (1993), Granger (1995), Renault, Sekkat, and Szafarz (1998), Marcellino (1999), Breitung and Swanson (2002), McCrorie and Chambers (2006), Silvestrini and Veredas (2008), among others. It is worth noting that whenever Granger causality and temporal aggregation are discussed, it is typically done in a setting where *all* series are subject to temporal aggregation. In such a setting it is well-known that even the simplest models, like a bivariate VAR(1) with stock (or skipped) sampling, may suffer from spuriously hidden or generated causality, and recovering the original causal pattern is very hard or even impossible in general.

As in the single frequency VAR literature, exploring mixed frequency Granger causality among more than two variables invariably relates to the notion of multi-horizon causality studied by Lütkepohl (1993), Dufour and Renault (1998) and Hill (2007). Of direct interest to us is Dufour and Renault (1998) who generalized the original definition of single-horizon or short run causality to multiple-horizon or long run causality to handle causality chains: in the presence of an auxiliary variable Z, Y may be useful for a multiple-step ahead prediction of X even if it is useless for the one-step ahead prediction. Dufour and Renault (1998) formalize the relationship between VAR coefficients and multiple-horizon causality and Dufour, Pelletier, and Renault (2006) formulate accordingly single step Wald tests of multiple-horizon non-causality. Their framework will be used extensively in our analysis. See Hill (2007) for a sequential method of testing for multiple-horizon non-causality.

In addition to the causality literature, the present paper also draws upon and contributes to the MIDAS literature originated by Ghysels, Santa-Clara, and Valkanov (2004) and Ghysels, Santa-Clara, and Valkanov (2005). A number of papers have linked MIDAS regressions to (latent) high frequency VAR models, such as Foroni, Marcellino, and Schumacher (2013) and Kuzin, Marcellino, and Schumacher (2011), whereas Ghysels (2012) discusses the link between mixed frequency VAR models and MIDAS regressions. None of these papers study in any detail the issue of Granger causality.

2.2 Mixed Frequency VAR Models

We want to characterize three settings which we will refer to as HF, MF and LF - respectively high, mixed and low frequency. We begin by considering a partially latent underlying HF process. Using the notation of Ghysels (2012), the HF process contains $\{\{\boldsymbol{x}_H(\tau_L,k)\}_{k=1}^m\}_{\tau_L}$ and $\{\{\boldsymbol{x}_L(\tau_L,k)\}_{k=1}^m\}_{\tau_L}$, where $\tau_L \in \{0,\ldots,T_L\}$ is the LF time index (e.g. quarter), $k \in \{1,\ldots,m\}$ denotes the HF (e.g. month), and m is the number of HF time periods between LF time indices. In the month versus quarter case, for

example, m equals three since one quarter has three months. Observations $\boldsymbol{x}_H(\tau_L, k) \in \mathbb{R}^{K_H \times 1}$ are called HF variables, whereas $\boldsymbol{x}_L(\tau_L, k) \in \mathbb{R}^{K_L \times 1}$ are latent LF variables because they are not observed at high frequencies - as only some temporal aggregates are available.

Note that two simplifying assumptions have implicitly been made. First, there are assumed to be only two sampling frequencies. Second, it is assumed that m is fixed and does not depend on τ_L. Both assumptions can be relaxed at the cost of much more complex notation and algebra which we avoid for expositional purpose - again see Ghysels (2012).

In reality the analyst's choice is limited to MF and LF cases. Only low frequency variables are aggregated in a MF setting, whereas both high and low frequency variables are aggregated to form a LF process. Following Lütkepohl (1987) we consider only linear aggregation schemes involving weights $\boldsymbol{w} = [w_1, \ldots, w_m]'$ such that:

$$\boldsymbol{x}_H(\tau_L) = \sum_{k=1}^{m} w_k \boldsymbol{x}_H(\tau_L, k) \text{ and } \boldsymbol{x}_L(\tau_L) = \sum_{k=1}^{m} w_k \boldsymbol{x}_L(\tau_L, k). \tag{2.1}$$

Two cases are of special interest given their broad use: (1) *stock* or *skipped* sampling, where $w_k = I(k = m)$; and (2) *flow* sampling, where $w_k = 1$ for $k = 1, \ldots, m$.[2] In summary, we observe:

- all high and low frequency variables $\{\{\boldsymbol{x}_H(\tau_L, j)\}_{j=1}^m\}_{\tau_L}$ and $\{\{\boldsymbol{x}_L(\tau_L, j)\}_{j=1}^m\}_{\tau_L}$ in a HF process;
- all high frequency variables $\{\{\boldsymbol{x}_H(\tau_L, j)\}_{j=1}^m\}_{\tau_L}$ but only aggregated low frequency variables $\{\boldsymbol{x}_L(\tau_L)\}_{\tau_L}$ in a MF process;
- only aggregated high and low frequency variables $\{\boldsymbol{x}_H(\tau_L)\}_{\tau_L}$ and $\{\boldsymbol{x}_L(\tau_L)\}_{\tau_L}$ in a LF process.

A key idea of MF-VAR models is to stack everything observable given a MF process according to their order over time. This results in the following $K = K_L + mK_H$ dimensional vector:

$$\boldsymbol{X}(\tau_L) = [\boldsymbol{x}_H(\tau_L, 1)', \ldots, \boldsymbol{x}_H(\tau_L, m)', \boldsymbol{x}_L(\tau_L)']'. \tag{2.2}$$

Note that $\boldsymbol{x}_L(\tau_L)$ is the last block in the stacked vector - a conventional assumption implying that it is observed after $\boldsymbol{x}_H(\tau_L, m)$. Any other order is conceptually the same, except that it implies a different timing of information about the respective processes. We will work with the specification appearing in (2.2) as it is most convenient.

In order to proceed, we will make a number of standard regulatory assumptions.

Assumption 2.1. The process $\boldsymbol{X}(\tau_L)$ is governed by a VAR(p):

$$\boldsymbol{X}(\tau_L) = \sum_{k=1}^{p} \boldsymbol{A}_k \boldsymbol{X}(\tau_L - k) + \boldsymbol{\epsilon}(\tau_L), \tag{2.3}$$

[2] One can equivalently let $w_k = 1/m$ for $k = 1, \ldots, m$ in flow sampling if the average is preferred to a summation.

where $\boldsymbol{A}_k$ is a $K \times K$ matrix for $k = 1, \ldots, p$, and $\boldsymbol{\epsilon}(\tau_L) = [\epsilon_1(\tau_L), \ldots, \epsilon_K(\tau_L)]'$ is a $K \times 1$ error vector that is assumed to be i.i.d. over τ_L.

Remark: The main results in this paper do not rely on error independence. Indeed, standard asymptotics for our Wald test statistic holds when $\{\boldsymbol{\epsilon}(\tau_L)\}$ is a stationary second order white noise process provided $\{\boldsymbol{X}(\tau_L), \boldsymbol{\epsilon}(\tau_L)\}$ satisfy an additional weak dependence property like strong mixing, along with the stationarity, distribution and moment properties under Assumptions 2.2 - 2.3 below. In this case $\boldsymbol{A}_k$ does not necessarily carry all the usual information about first order causation, where the latter is typically defined in terms of mean-squared-error improvement of an h-step ahead linear forecast error (cfr. Granger (1969), Sims (1972), Dufour and Renault (1998)) (see Section 3 for further detail). This is irrelevant for our purposes, however, because in the tradition of Dufour and Renault (1998) our analysis is primarily about deducing nonlinear restrictions on $\{\boldsymbol{A}_1, ..., \boldsymbol{A}_p\}$ that relate information about predictive ability, and about recovering information on (non-)causation in HF-VAR by using MF- or LF-VAR models. Nevertheless, we impose the i.i.d. assumption in order to simplify some technical arguments. Furthermore, without independence the close relationship between Granger's (1969) and Sims' (1972) notions of causality in terms of linear predictive improvement breaks down, as shown in Florens and Mouchart (1982).

In addition, the following standard assumptions ensure stationarity and geometric strong mixing of the observed time series.[3]

Assumption 2.2. All roots of the polynomial $\det(\boldsymbol{I}_K - \sum_{k=1}^{p} \boldsymbol{A}_k z^k) = 0$ lie outside the unit circle.

Assumption 2.3. $\boldsymbol{\epsilon}(\tau_L)$ has an absolutely continuous distribution with a bounded joint density $f_\epsilon(\boldsymbol{u}) \leq M$ for all $\boldsymbol{u} \in \mathbb{R}^K$ and some $M \in (0, \infty)$. Further $E[\boldsymbol{\epsilon}(\tau_L)] = \boldsymbol{0}_{K\times 1}$, $||\boldsymbol{\epsilon}(\tau_L)||_{2+\delta} \in (0, \infty)$ for some $\delta > 0$, and $\boldsymbol{\Omega} \equiv E\left[\boldsymbol{\epsilon}(\tau_L)\boldsymbol{\epsilon}(\tau_L)'\right]$ is positive definite. Finally, the sigma-fields induced by the history $\{\boldsymbol{\epsilon}(\tau_L), \boldsymbol{\epsilon}(\tau_L - 1), ...\}$ are strictly increasing.

We assume $\boldsymbol{\epsilon}(\tau_L)$ has a bounded joint density so that, in combination with stationarity, $\{\boldsymbol{X}(\tau_L)\}$ is geometrically strong mixing by classic arguments (cfr. Section 3.2 in Dedecker et al. (2007)). Recall that strong mixing implies mixing (in the ergodic sense) and therefore ergodicity (see Petersen (1983)). The mixing property is useful since the necessary steps for inference involve finite lag functions of $\boldsymbol{\epsilon}(\tau_L)$ and $\boldsymbol{X}(\tau_L)$ which are also mixing and therefore satisfy a standard limit theory (see Dedecker et al. (2007)).[4]

Note also that we do not include a constant term in (2.3) solely to reduce notation, thus $\boldsymbol{X}(\tau_L)$ should be thought of as a de-meaned process. Finally, it is straightforward to allow an infinite order VAR structure, and estimate a truncated finite order VAR model as in Lewis and Reinsel (1985), Lütkepohl and Poskitt (1996), and Saikkonen and Lütkepohl (1996).

[3] Although a large body of literature exists on Granger causality in non-stationary or cointegrated systems (e.g. Yamamoto and Kurozumi (2006)), the generalization is beyond the scope of this paper.

[4] A continuous bounded distribution for $\boldsymbol{\epsilon}(\tau_L)$ can be dispensed with at the cost of additional steps in our limit theory proofs since $\{\boldsymbol{X}(\tau_L)\}$ is geometrically $L_{2+\delta}$-Near Epoch Dependent with respect to an i.i.d. base $\{\boldsymbol{\epsilon}(\tau_L)\}$. All functions of $\boldsymbol{X}(\tau_L)$ and $\boldsymbol{\epsilon}(\tau_L)$ in this paper are also Near Epoch Dependent, and a complete limit theory for such a weak dependence property is available. See Davidson (1994) for a textbook treatment.

2.3 Estimators and Their Large Sample Properties

If the VAR(p) model appearing in (2.3) were standard, then the off-diagonal elements of any matrix $\boldsymbol{A}_k$ would tell us something about causal relationships for some specific horizon. The fact that MF-VAR models involve stacked replicas of high frequency data sampled across different (high frequency) periods implies that potentially multi-horizon causal patterns reside inside any of the matrices $\boldsymbol{A}_k$. It is therefore natural to start with a multi-horizon setting. We do so, at first, focusing on multiple low frequency prediction horizons which we will denote by $h \in \mathbb{N}$.[5]

It is convenient to iterate (2.3) over the desired test horizon in order to deduce simple testable parameter restrictions for non-causality. Recall that under Assumption 2.2 a unique stationary and ergodic solution to (2.3) exists:

$$\boldsymbol{X}(\tau_L) = \sum_{k=0}^{\infty} \boldsymbol{\Psi}_k \boldsymbol{\epsilon}(\tau_L - k), \tag{2.4}$$

where $\boldsymbol{\Psi}_k$ satisfies $\boldsymbol{\Psi}_0 = \boldsymbol{I}_K$, $\boldsymbol{\Psi}_k = \sum_{s=1}^{p} \boldsymbol{A}_s \boldsymbol{\Psi}_{k-s}$ for $k \geq 1$ and $\boldsymbol{\Psi}_k = \boldsymbol{0}_{K\times K}$ for $k < 0$, and $|\boldsymbol{\Psi}_k| = O(\rho^k)$ for some $\rho \in (0,1)$. We then have what Dufour, Pelletier, and Renault (2006) labeled as a (p,h)-autoregression:

$$\boldsymbol{X}(\tau_L + h) = \sum_{k=1}^{p} \boldsymbol{A}_k^{(h)} \boldsymbol{X}(\tau_L + 1 - k) + \sum_{k=0}^{h-1} \boldsymbol{\Psi}_k \boldsymbol{\epsilon}(\tau_L + h - k), \tag{2.5}$$

where

$$\boldsymbol{A}_k^{(1)} = \boldsymbol{A}_k \text{ and } \boldsymbol{A}_k^{(i)} = \boldsymbol{A}_{k+i-1} + \sum_{l=1}^{i-1} \boldsymbol{A}_{i-l} \boldsymbol{A}_k^{(l)} \text{ for } i \geq 2.$$

By convention $\boldsymbol{A}_k = \boldsymbol{0}_{K\times K}$ whenever $k > p$. The MF-VAR causality test exploits Wald statistics based on the OLS estimator of the (p,h)-autoregression parameter set

$$\boldsymbol{B}(h) = \left[\boldsymbol{A}_1^{(h)}, \ldots, \boldsymbol{A}_p^{(h)}\right]' \in \mathbb{R}^{pK\times K}. \tag{2.6}$$

The OLS estimator $\hat{\boldsymbol{B}}(h)$ of $\boldsymbol{B}(h)$ is

$$\hat{\boldsymbol{B}}(h) \equiv \underset{\boldsymbol{B}(h)}{\textit{arg min}} \left\{ \operatorname{vec}\left[\boldsymbol{U}_h(h)\right]' \operatorname{vec}\left[\boldsymbol{U}_h(h)\right] \right\} = \left[\overline{\boldsymbol{W}}_p(h)' \overline{\boldsymbol{W}}_p(h)\right]^{-1} \overline{\boldsymbol{W}}_p(h)' \boldsymbol{W}_h(h),$$

where $\boldsymbol{U}_h(h)$ is a matrix of stacked sums of $\{\boldsymbol{\Psi}_k\}$ and $\{\boldsymbol{\epsilon}(\tau_L)\}$ while $\overline{\boldsymbol{W}}_p(h)$ and $\boldsymbol{W}_h(h)$ are matrices of stacked $\{\boldsymbol{X}(\tau_L)\}$. See Appendix A.1.1 for derivation of $\{\boldsymbol{U}_h(h), \overline{\boldsymbol{W}}_p(h), \boldsymbol{W}_h(h)\}$.

Assumptions 2.1 through 2.3 suffice for $\hat{\boldsymbol{B}}(h)$ to be consistent for $\boldsymbol{B}(h)$ and asymptotically normal. Limits are with respect to $T_L \to \infty$ hence $T_L^* \to \infty$, where $T_L^* = T_L - h + 1$ is the effective sample size for the (p,h)-autoregression.

[5] Another reason for studying multiple horizons is the potential of causality chains when $K_H > 1$ or $K_L > 1$. Note, however, that despite the MF-VAR being by design multi-dimensional there are no causality chains when $K_H = K_L = 1$ since the $m \times 1$ vector of the high frequency observations refers to a single variable.

Theorem 2.1. Under Assumptions 2.1 through 2.3 $\hat{\boldsymbol{B}}(h) \xrightarrow{p} \boldsymbol{B}(h)$ and

$$\sqrt{T_L^*}\text{vec}\left[\hat{\boldsymbol{B}}(h) - \boldsymbol{B}(h)\right] \xrightarrow{d} N\left(\mathbf{0}_{pK^2\times 1}, \boldsymbol{\Sigma}_p(h)\right), \tag{2.7}$$

where $\boldsymbol{\Sigma}_p(h)$ is positive definite.

Remark: See Appendix A.2 for a proof, and see Appendices A.1-A.2 for a complete characterization of $\boldsymbol{\Sigma}_p(h)$.

If all variables were aggregated into a common low frequency and expanded into a (p, h)-autoregression, then h-step ahead non-causality has a simple parametric expression in terms of $\boldsymbol{A}_k^{(h)}$; cfr. Dufour, Pelletier, and Renault (2006). Recall, however, that the MF-VAR has a special structure because of the stacked HF vector. This implies that the Wald-type test for non-causality that we derive is slightly more complicated than those considered by Dufour, Pelletier, and Renault (2006) since in MF-VAR models the restrictions will often deal with linear parametric restrictions across multiple equations. In a generic sense, we show in Section 3 that non-causality between any set of variables in a MF-VAR model can be expressed as linear constraints with respect to $\boldsymbol{B}(h)$. Hence, the null hypothesis of interest is a linear restriction:

$$H_0(h) : \boldsymbol{R}\text{vec}\left[\boldsymbol{B}(h)\right] = \boldsymbol{r}, \tag{2.8}$$

where $\boldsymbol{R}$ is a $q \times pK^2$ selection matrix of full row rank q, and $\boldsymbol{r} \in \mathbb{R}^q$. We leave complete details of the construction of $\boldsymbol{R}$ for Section 3.

With a consistent estimator $\hat{\boldsymbol{\Sigma}}_p(h)$ for the least squares asymptotic variance $\boldsymbol{\Sigma}_p(h)$ which is *almost surely* positive semi-definite for $T_L^* \geq 1$, we can define the Wald statistic

$$W[H_0(h)] \equiv T_L^* \left(\boldsymbol{R}\text{vec}\left[\hat{\boldsymbol{B}}(h)\right] - \boldsymbol{r}\right)' \times \left(\boldsymbol{R}\hat{\boldsymbol{\Sigma}}_p(h)\boldsymbol{R}'\right)^{-1} \times \left(\boldsymbol{R}\text{vec}\left[\hat{\boldsymbol{B}}(h)\right] - \boldsymbol{r}\right). \tag{2.9}$$

Implicitly, of course, $\boldsymbol{R}\hat{\boldsymbol{\Sigma}}_p(h)\boldsymbol{R}'$ must be non-singular for any $\boldsymbol{R} \in \mathbb{R}^{q\times pK^2}$ with full row rank. In view of positive definiteness of $\boldsymbol{\Sigma}_p(h)$ by Theorem 2.1, and the supposition $\hat{\boldsymbol{\Sigma}}_p(h) = \boldsymbol{\Sigma}_p(h) + o_p(1)$, it follows $(\boldsymbol{R}\hat{\boldsymbol{\Sigma}}_p(h)\boldsymbol{R}')^{-1}$ is well defined asymptotically with probability approaching one.

We therefore obtain the following result, which we prove in Appendix A.2.

Theorem 2.2. Let $\hat{\boldsymbol{\Sigma}}_p(h)$ be a consistent estimator for $\boldsymbol{\Sigma}_p(h)$ that is *almost surely* positive semi-definite for any $T_L^* \geq 1$. Given Assumptions 2.1 through 2.3, $W[H_0(h)] \xrightarrow{d} \chi_q^2$ under $H_0(h)$.

Remark: A consistent, *almost surely* positive semi-definite estimator $\hat{\boldsymbol{\Sigma}}_p(h)$ is easily constructed by using Newey and West's (1987) HAC estimator, given the stronger moment assumption $||\boldsymbol{\epsilon}(\tau_L)||^{4+\delta} < \infty$ for some $\delta > 0$.. See Appendix A.1.3 for complete details.

In the remainder of the paper we will provide various tests for Granger causality which are special cases of the generic framework derived so far.

3 Testing Causality with Mixed Frequency Data

In this section we define non-causality when data are sampled at mixed frequencies and describe Wald-type tests associated with it. We first cover some preliminary notions of multiple-horizon causality and extend it to the mixed sampling frequency case. We discuss in detail testing non-causality from one variable to another, and whether they are high or low frequency variables. We also cover non-causality from all high frequency variables to all low frequency variables and vice versa, cases for which we give explicit formulae for the selection matrix $\boldsymbol{R}$ used in the null hypothesis (2.8) and test statistic (2.9).

3.1 Preliminaries

We start with adopting the notion of non-causality to a mixed sampling frequency data filtration setting. Using the notation of Dufour and Renault (1998) we define the relevant information sets for the purpose of characterizing non-causality. In particular, let L^2 be a Hilbert space of covariance stationary real-valued random variables defined on a common probability space, and the covariance as inner product. Moreover, let $\mathcal{I}(\tau_L)$ be a closed increasing subspace of L^2 such that $\mathcal{I}(\tau_L) \subset \mathcal{I}(\tau_L')$ whenever $\tau_L < \tau_L'$, where $\tau_L, \tau_L' \in \mathbb{Z}$.

Furthermore, define the indices i_1, $i_2 \in \{1,\ldots,K_H\}$ and $j_1, j_2 \in \{1,\ldots,K_L\}$, and write $\boldsymbol{x}_{H,i_1} = \{\boldsymbol{x}_{H,i_1}(\tau_L)\}_{\tau_L}$, where $\boldsymbol{x}_{H,i_1}(\tau_L) = [x_{H,i_1}(\tau_L,1),\ \ldots,\ x_{H,i_1}(\tau_L,m)]'$. Note that $\boldsymbol{x}_{H,i_1}(\tau_L)$ is a vector of all m observations of the i_1-th high frequency variable available at period τ_L, whereas $\boldsymbol{x}_{H,i_1}$ is the collection of that vector over all periods. Similarly, we write $x_{L,j_1} = \{x_{L,j_1}(\tau_L)\}_{\tau_L}$ and note that $x_{L,j_1}(\tau_L)$ is a scalar since it is the j_1-th low frequency variable at period τ_L. Finally, let $\boldsymbol{x}_H = \{\boldsymbol{x}_{H,i_1}\}_{i_1}$ and $\boldsymbol{x}_L = \{x_{L,j_1}\}_{j_1}$.

Denote by $\boldsymbol{x}(-\infty,\tau_L]$ the Hilbert space spanned by $\{\boldsymbol{x}(\tau)|\ \tau \leq \tau_L\}$. The information set $\mathcal{I}$ is said to be *conformable* with $\boldsymbol{x}$ if $\boldsymbol{x}(-\infty,\tau_L] \subset \mathcal{I}(\tau_L)$ for all τ_L. We call the information set derived from $\mathcal{I}(\tau_L) = \boldsymbol{X}(-\infty,\tau_L]$, where $\boldsymbol{X}(\tau_L)$ is given in (2.2), as the *MF reference information set in period* τ_L, whereas $\mathcal{I} = \{\mathcal{I}(\tau_L)|\ \tau_L \in \mathbb{Z}\}$ is the *MF reference information set*. Therefore, the only information available up to period τ_L is the high frequency observations of all high frequency variables and the low frequency observations of all low frequency variables. In addition, let $\mathcal{I}_{(H,i_1)}$ denote the MF reference information set except for $\boldsymbol{x}_{H,i_1}$, and let $\mathcal{I}_{(L,j_1)}$ denote the information set except for x_{L,j_1}. Similarly, $\mathcal{I}_{(H)}$ ($\mathcal{I}_{(L)}$) is the MF reference information set except for $\boldsymbol{x}_H$ ($\boldsymbol{x}_L$). Notice that since high and low frequency variables $\boldsymbol{x}_{H,i_1}(\tau_L)$ and $x_{L,j_1}(\tau_L)$ belong to $\boldsymbol{X}(\tau_L)$ for all $i_1 \in \{1,\ldots,K_H\}$ and $j_1 \in \{1,\ldots,K_L\}$, it is clear that the MF reference information set $\mathcal{I} = \{\mathcal{I}(\tau_L)|\ \tau_L \in \mathbb{Z}\}$ is conformable with $\boldsymbol{x}_{H,i_1}(\tau_L)$ and $x_{L,j_1}(\tau_L)$.

Finally, let E and F be two subspaces of L^2, and let $E + F$ denote the Hilbert subspace generated by the elements of E and F. Let $P[\boldsymbol{x}(\tau_L+h)|\ \mathcal{I}(\tau_L)]$ be the best linear forecast of $\boldsymbol{x}(\tau_L+h)$ based on $\mathcal{I}(\tau_L)$ in the sense of a covariance orthogonal projection.

For any generic information set and pair of processes (high or low frequency) the notion of non-causality is defined as follows.

Definition 3.1. (Non-causality at Different Horizons). Suppose that $\mathcal{I}$ is conformable with $\boldsymbol{x}$. (i) $\boldsymbol{y}$ *does*

not cause $\boldsymbol{x}$ at horizon h given $\mathcal{I}$ (denoted by $\boldsymbol{y} \nrightarrow_h \boldsymbol{x} |\, \mathcal{I}$) if:

$$P[\boldsymbol{x}(\tau_L + h)|\mathcal{I}(\tau_L)] = P[\boldsymbol{x}(\tau_L + h)|\mathcal{I}(\tau_L) + \boldsymbol{y}(-\infty, \tau_L]] \qquad \forall \tau_L \in \mathbb{Z}.$$

Moreover, (ii) $\boldsymbol{y}$ *does not cause* $\boldsymbol{x}$ *up to horizon* h *given* $\mathcal{I}$ (denoted by $\boldsymbol{y} \nrightarrow_{(h)} \boldsymbol{x} |\, \mathcal{I}$) if $\boldsymbol{y} \nrightarrow_k \boldsymbol{x} |\, \mathcal{I}$ for all $k \in \{1, \ldots, h\}$.

Definition 3.1 applies to a mixed sampling frequency setting when suitable information set and processes are used.[6] Consider, for example, non-causality from the j_1-th low frequency variable to the i_1-th high frequency variable. Similarly, x_{L,j_1} does not cause $\boldsymbol{x}_{H,i_1}$ at horizon h given $\mathcal{I}$ (denoted by $x_{L,j_1} \nrightarrow_h \boldsymbol{x}_{H,i_1} |\, \mathcal{I}$) if $P[\boldsymbol{x}_{H,i_1}(\tau_L + h) |\, \mathcal{I}_{(L,j_1)}(\tau_L)] = P\left[\boldsymbol{x}_{H,i_1}(\tau_L + h) |\, \mathcal{I}(\tau_L)\right]$ for all $\tau_L \in \mathbb{Z}$. When we consider non-causality between a pair of high frequency series, namely $\boldsymbol{x}_{H,i_1} \nrightarrow_h \boldsymbol{x}_{H,i_2} |\, \mathcal{I}_{(H,i_1)}$ it should be noted that we focus exclusively on low frequency horizons h, or equivalently horizons $h \times m$. Any other horizon, not a multiple of m, are not considered here. They can be handled with the existing same frequency setting of Dufour and Renault (1998).

In a mixed sampling frequency setting, there are six basic cases to consider.

Case 1 (low to low) Non-causality from the j_1-th *low* frequency variable, x_{L,j_1}, to the j_2-th *low* frequency variable, x_{L,j_2}, at horizon h. The null hypothesis can be written as $H_0^1(h) : x_{L,j_1} \nrightarrow_h x_{L,j_2} | \mathcal{I}_{(L,j_1)}$.

Case 2 (high to low) $H_0^2(h) : \boldsymbol{x}_{H,i_1} \nrightarrow_h x_{L,j_1} |\, \mathcal{I}_{(H,i_1)}$.

Case 3 (low to high) $H_0^3(h) : x_{L,j_1} \nrightarrow_h \boldsymbol{x}_{H,i_1} |\, \mathcal{I}_{(L,j_1)}$.

Case 4 (high to high) $H_0^4(h) : \boldsymbol{x}_{H,i_1} \nrightarrow_h \boldsymbol{x}_{H,i_2} |\, \mathcal{I}_{(H,i_1)}$.

Case I (all high to all low) $H_0^I(h) : \boldsymbol{x}_H \nrightarrow_h \boldsymbol{x}_L |\, \mathcal{I}_{(H)}$.

Case II (all low to all high) $H_0^{II}(h) : \boldsymbol{x}_L \nrightarrow_h \boldsymbol{x}_H |\, \mathcal{I}_{(L)}$.

Cases 1 through 4 handle individual variables, while Cases I and II handle entire groups of variables. In the sequel we often consider Cases I and II for simplicity since - viewed as a bivariate system - causality chains can be excluded in both cases since non-causality at one horizon is synonymous to non-causality at all horizons (see Dufour and Renault (1998: Proposition 2.3), cfr. Florens and Mouchart (1982: p. 590)). To avoid tedious matrix notation, we do not treat in detail cases involving non-causation from a subset of all variables to another subset. Our results straightforwardly apply, however, in such cases as well.

3.2 Causality Tests in Mixed Frequency VAR Models

Our next task is to construct the selection matrices $\boldsymbol{R}$ for the various null hypotheses (2.8) associated with the six generic cases. This requires deciphering parameter restrictions for non-causation based on the (p, h)-autoregression appearing in equation (2.5).

[6] Definition 3.1 corresponds to Definition 2.2 in Dufour and Renault (1998) for covariance stationary processes.

Characterizing restrictions on $\boldsymbol{A}_k^{(h)}$ for each case above requires some additional matrix notation. Let $\boldsymbol{N} \in \mathbb{R}^{n\times n}$, and let $a, b, c, d, \iota, \iota' \in \{1, \ldots, n\}$ with $a \leq b$, $c \leq d$, and $(b-a)/\iota$ and $(d-c)/\iota'$ being nonnegative integers. Then we define $\boldsymbol{N}(a : \iota : b, c : \iota' : d)$ as the $(\frac{b-a}{\iota} + 1) \times (\frac{d-c}{\iota'} + 1)$ matrix which consists of the a-th, $(a+\iota)$-th, $(a+2\iota)$-th, ..., b-th rows and c-th, $(c+\iota')$-th, $(c+2\iota')$-th, ..., d-th columns of $\boldsymbol{N}$. To put differently, a signifies the first element to pick, b is the last, and ι is the increment with respect to rows. c, d, and ι' play analogous roles with respect to columns. It is clear that:

$$\boldsymbol{N}(a : \iota : b, c : \iota' : d)' = \boldsymbol{N}'(c : \iota' : d, a : \iota : b). \tag{3.1}$$

A short-hand notation is used when $a = b$: $\boldsymbol{N}(a : \iota : b, c : \iota' : d) = \boldsymbol{N}(a, c : \iota' : d)$. When $\iota = 1$, we write: $\boldsymbol{N}(a : \iota : b, c : \iota' : d) = \boldsymbol{N}(a : b, c : \iota' : d)$. Analogous notations are used when $c = d$ or $\iota' = 1$, respectively.

By Theorem 3.1 in Dufour and Renault (1998) and in view of model (2.5), it follows that $H_0^i(h)$ are equivalent to:

$$\boldsymbol{A}_k^{(h)}(a : \iota : b, c : \iota' : d) = \boldsymbol{0} \text{ for each } k \in \{1, \ldots, p\}, \tag{3.2}$$

where a, ι, b, c, ι', d, and the size of the null vector differ across cases $i = 1, \ldots, 4$ and I and II.[7] In Table 1 we detail the specifics for a, ι, b, c, ι', d in these quantities for each of the six cases.

Table 1: Linear Parametric Restrictions of Non-causality

The null hypotheses of non-causality cases $H_0^i(h)$ for $i = 1, \ldots, 4$ and I and II. can be written as $\boldsymbol{A}_k^{(h)}(a : \iota : b, c : \iota' : d) = \boldsymbol{0}$ for all $k \in \{1, \ldots, p\}$, where a, ι, b, c, ι', d, and the size of the null vector appear as entries to the table.

Cases	a	ι	b	c	ι'	d	$\boldsymbol{0}$
$H_0^1(h)$	$mK_H + j_2$	1	$mK_H + j_2$	$mK_H + j_1$	1	$mK_H + j_1$	1×1
$H_0^2(h)$	$mK_H + j_1$	1	$mK_H + j_1$	i_1	K_H	$i_1 + (m-1)K_H$	$1 \times m$
$H_0^3(h)$	i_1	K_H	$i_1 + (m-1)K_H$	$mK_H + j_1$	1	$mK_H + j_1$	$m \times 1$
$H_0^4(h)$	i_2	K_H	$i_2 + (m-1)K_H$	i_1	K_H	$i_1 + (m-1)K_H$	$m \times m$
$H_0^I(h)$	$mK_H + 1$	1	K	1	1	mK_H	$K_L \times mK_H$
$H_0^{II}(h)$	1	1	mK_H	$mK_H + 1$	1	K	$mK_H \times K_L$

Each case in Table 1 can be interpreted as follows. In Case 1, the $(mK_H + j_2, mK_H + j_1)$-th element of $\boldsymbol{A}_k^{(h)}$ (i.e., the impact of the j_1-th low frequency variable on the j_2-th low frequency variable) is zero if and only if $H_0^1(h)$ is true. Likewise, in Case 2, the $(mK_H + j_1, i_1)$-th, $(mK_H + j_1, i_1 + K_H)$-th, ..., $(mK_H + j_1, i_1 + (m-1)K_H)$-th elements of $\boldsymbol{A}_k^{(h)}$ are all zeros under $H_0^2(h)$. Note that we are testing whether or not all mp coefficients of the i_1-th high frequency variable on the j_1-th low frequency variable are zeros, i.e., the i_1-th high frequency variable has no impact as a whole on the j_1-th low frequency variable at a given horizon h.

When $H_0^3(h)$ holds, all mp coefficients of the j_1-th low frequency variable on the i_1-th high frequency variable are zeros at horizon h. Note that the parameter constraints run across the i_1-th, (i_1+K_H)-th, ..., $(i_1 + (m-1)K_H)$-th rows of $\boldsymbol{A}_k^{(h)}$, not columns. This means that we are testing *simultaneous*

[7] Recall that x_{L,j_1} and $\boldsymbol{x}_{H,i_1}$ belong to $\boldsymbol{X}$ in (2.2) for all $j_1 \in \{1, \ldots, K_L\}$ and $i_1 \in \{1, \ldots, K_H\}$. This is why the non-causality under mixed frequencies is well-defined and Theorem 3.1 in Dufour and Renault (1998) can be applied directly.

linear restrictions *across multiple equations*, unlike Dufour, Pelletier, and Renault (2006) who focus mainly on *simultaneous* linear restrictions *within one equation*, and unlike Hill (2007) who focuses on *sequential* linear restrictions *across multiple equations*.

In Case 4, the i_1-th high frequency variable has no impact on the i_2-th high frequency variable if and only if $H_0^4(h)$ is true. In this case m^2 elements out of $\boldsymbol{A}_k^{(h)}$ are restricted to be zeros for each k, so the total number of zero restrictions is pm^2. Under $H_0^I(h)$, the $K_L \times mK_H$ lower-left block of $\boldsymbol{A}_k^{(h)}$ is a null matrix. Finally, in Case II, the $mK_H \times K_L$ upper-right block of $\boldsymbol{A}_k^{(h)}$ is a null matrix if and only if $H_0^{II}(h)$ is true.

We can now combine the (p,h)-autoregression parameter set $\boldsymbol{B}(h)$ in (2.6) with the matrix construction (3.1), its implication for testable restrictions (3.2), and Table 1, to obtain generic formulae for $\boldsymbol{R}$ and $\boldsymbol{r}$ so that all six cases can be treated as special cases of (2.8).

The above can be summarized as follows:

Theorem 3.1. All hypotheses $H_0^i(h)$ for $i \in \{1,2,3,4,I,II\}$ are special cases of $H_0(h)$ with

$$\boldsymbol{R} = \left[\boldsymbol{\Lambda}(\delta_1)', \boldsymbol{\Lambda}(\delta_2)', \ldots, \boldsymbol{\Lambda}(\delta_{g(a,\iota,b)p})'\right]' \tag{3.3}$$

and

$$\boldsymbol{r} = \boldsymbol{0}_{g(a,\iota,b)g(c,\iota',d)p \times 1}, \tag{3.4}$$

where $g(a,\iota,b) = (b-a)/\iota + 1$, $\delta_1 = pK(a-1) + c$,

$$\delta_l = \delta_{l-1} + K + pK(\iota - 1)I(l - 1 = zp \text{ for some } z \in \mathbb{N}) \tag{3.5}$$

for $l = 2, \ldots, g(a,\iota,b)p$, and $\boldsymbol{\Lambda}(\delta)$ is a $g(c,\iota',d) \times pK^2$ matrix whose $(j,\ \delta + (j-1)\iota')$-th element is 1 for $j \in \{1, \ldots, g(c,\iota',d)\}$ and all other elements are zeros.

Several key points will help us understand (3.3) through (3.5). First, $g(a,\iota,b)$ and $g(c,\iota',d)$ represent how many rows and columns of $\boldsymbol{A}_k^{(h)}$ have zero restrictions for each $k \in \{1,\ldots,p\}$, respectively. The total number of zero restrictions is therefore $q = g(a,\iota,b)g(c,\iota',d)p$ as in (3.4). Second, $\boldsymbol{\Lambda}(\delta)$ has only one nonzero element in each row that is identically 1, signifying which element of $\mathrm{vec}[\boldsymbol{B}(h)]$ is supposed to be zero. The location of 1 is determined by $\delta_1, \ldots, \delta_{g(a,\iota,b)p}$, which are recursively updated according to (3.5). As seen in (3.5), the increment of δ_l is basically K, but an extra increment of $pK(\iota-1)$ is added when $l-1$ is a multiple of p in order to skip some columns of $\boldsymbol{B}(h)$.

Theorem 3.1 provides unified testing for non-causality as summarized below.

Step 1 For a given VAR lag order p and test horizon h, estimate a (p,h)-autoregression.[8]

[8] A potential drawback of our approach as well as Dufour, Pelletier, and Renault (2006) is that the prediction horizon h is fixed at each test and thus the entire set of results for multiple h's may yield a contradiction. See footnote 2 in Hill (2007). Hill (2007) avoids this problem by a sequential multiple-horizon non-causation test, in which a series of individual non-causation tests are performed to deduce causal chains and causation horizon. The present paper takes the Dufour, Pelletier, and Renault (2006) approach because of its simplicity. See Hill (2007) and Salamaliki and Venetis (2013) for a comparison of the two methods.

Step 2 Calculate a, ι, b, c, ι', d according to Table 1 for a given case of non-causality relation. Put those quantities into (3.3) and (3.4) to get $\boldsymbol{R}$ and $\boldsymbol{r}$.

Step 3 Use $\boldsymbol{R}$ and $\boldsymbol{r}$ in order to calculate the Wald test statistic $W[H_0(h)]$ in (2.9).

Since Table 1 and Theorem 3.1 are rather abstract, we present a concrete example of how $\boldsymbol{R}$ and $\boldsymbol{r}$ are constructed based in our trivariate simulation. In Section 6.2 we fit a MF-VAR(1) model with prediction horizons $h \in \{1, 2, 3\}$ to two high frequency variables X and Y and one low frequency variable Z with $m = 3$. In this case the mixed frequency vector appearing in (2.2) can be written as:

$$\boldsymbol{W}(\tau_L) = [X(\tau_L, 1), Y(\tau_L, 1), X(\tau_L, 2), Y(\tau_L, 2), X(\tau_L, 3), Y(\tau_L, 3), Z(\tau_L)]'.$$

Notice that $K_H = 2$, $K_L = 1$, and hence $K = 7$ in this example. Although the construction of $\boldsymbol{R}$ and $\boldsymbol{r}$ do not depend on the value of h, consider $h = 1$ for simplicity, and write the parameter matrix:

$$\boldsymbol{A}_1 = \begin{bmatrix} a_{11} & \dots & a_{17} \\ \vdots & \ddots & \vdots \\ a_{71} & \dots & a_{77} \end{bmatrix} \quad \text{or} \quad \boldsymbol{A}_1' = \begin{bmatrix} a_{11} & \dots & a_{71} \\ \vdots & \ddots & \vdots \\ a_{17} & \dots & a_{77} \end{bmatrix}.$$

Since $p = h = 1$, $\boldsymbol{B}(h)$ appearing in (2.6) is simply $\boldsymbol{A}_1'$.

Consider the null hypothesis that Z does not cause X at horizon 1. This null hypothesis is equivalently $a_{17} = a_{37} = a_{57} = 0$ since a_{17}, a_{37}, and a_{57} represent the impact of $Z(\tau_L - 1)$ on $X(\tau_L, 1)$, $X(\tau_L, 2)$, and $X(\tau_L, 3)$, respectively. Note that a_{17}, a_{37}, and a_{57} are respectively the 7th, 21st, and 35th element of vec$[\boldsymbol{B}(h)]$ appearing in (2.8). Hence, the proper choice of $\boldsymbol{R}$ and $\boldsymbol{r}$ is:

$$\boldsymbol{R} = \begin{bmatrix} \mathbf{0}_{1\times 6} & 1 & \mathbf{0}_{1\times 13} & 0 & \mathbf{0}_{1\times 13} & 0 & \mathbf{0}_{1\times 14} \\ \mathbf{0}_{1\times 6} & 0 & \mathbf{0}_{1\times 13} & 1 & \mathbf{0}_{1\times 13} & 0 & \mathbf{0}_{1\times 14} \\ \mathbf{0}_{1\times 6} & 0 & \mathbf{0}_{1\times 13} & 0 & \mathbf{0}_{1\times 13} & 1 & \mathbf{0}_{1\times 14} \end{bmatrix} \quad \text{and} \quad \boldsymbol{r} = \mathbf{0}_{3\times 1}. \tag{3.6}$$

We now confirm that the same $\boldsymbol{R}$ and $\boldsymbol{r}$ can be obtained via Table 1 and Theorem 3.1. Non-causality from Z to X falls in Case 3 with $i_1 = j_1 = 1$ (i.e. non-causality from the first low frequency variable to the first high frequency variable). Using Table 1, we have that $(a, \iota, b, c, \iota', d) = (1, 2, 5, 7, 1, 7)$ and therefore $g(a, \iota, b) = 3$, $g(c, \iota', d) = 1$, and $\{\delta_1, \delta_2, \delta_3\} = \{7, 21, 35\}$ in view of Theorem 3.1. This implies that $\boldsymbol{r} = \mathbf{0}_{3\times 1}$ and $\boldsymbol{R} = [\boldsymbol{\Lambda}(7)', \boldsymbol{\Lambda}(21)', \boldsymbol{\Lambda}(35)']'$, where $\boldsymbol{\Lambda}(\delta)$ is a 1×49 vector whose δ-th element is 1 and all other elements are zeros for $\delta \in \{7, 21, 35\}$. We can therefore confirm that Table 1 and Theorem 3.1 provide correct $\boldsymbol{R}$ and $\boldsymbol{r}$ shown in (3.6).

4 Recovery of High Frequency Causality

The existing literature on Granger causality and temporal aggregation has three key ingredients. Starting with (1) a data generating process (DGP) for HF data, and (2) specifying a (linear) aggregation scheme, one is interested in (3) the relationship between causal patterns - or lack thereof - among the HF series

and the inference obtained from LF data when *all* HF series are aggregated. So far, we refrained from (1) specifying a DGP for HF series and (2) specifying an aggregation scheme. We will proceed along the same path as the existing literature in this section with a different purpose, namely to show that the MF approach recovers more underlying causal patterns than the standard LF approach does. While conducting Granger causality tests with MF does not resolve all HF causal patterns, using MF instead of using exclusively LF series promotes sharper inference.

We first start with a fairly straightforward extension of Lütkepohl (1984), establishing the link between HF-VAR and MF data representations. We then analyze the link between HF, MF and LF causality.

4.1 Temporal Aggregation of VAR Processes

Lütkepohl (1984) provides a comprehensive analysis of temporal aggregation and VAR processes. We extend his analysis to a MF setting. While the extension is straightforward, it provides us with a framework that will be helpful for the analysis in the rest of the paper.

Let $K^* = K_H + K_L$, and define $\overline{\boldsymbol{X}}(\tau_L, k) = [\boldsymbol{x}_H(\tau_L, k)', \boldsymbol{x}_L(\tau_L, k)']' \in \mathbb{R}^{K^*}$ for $k = 1, \ldots, m$. Note that part of the $\overline{\boldsymbol{X}}$ vector process is obviously latent, namely the high frequency observations of the LF process, represented by the $\boldsymbol{x}_L(\tau_L, k)$ elements of the vector process.

To proceed, let $\mathcal{L}_H$ denote the *high frequency* lag operator, in particular

$$\mathcal{L}_H^l \overline{\boldsymbol{X}}(\tau_L, k) = \overline{\boldsymbol{X}}(\tau_L - \iota, \iota')$$

with

$$\iota = \begin{cases} 0 & \text{if } 0 \leq l < k \\ 1 + \lfloor \frac{l-k}{m} \rfloor & \text{if } l \geq k \end{cases} \quad \text{and} \quad \iota' = \begin{cases} k - l & \text{if } 0 \leq l < k \\ \iota m + k - l & \text{if } l \geq k. \end{cases}$$

Note $\lfloor x \rfloor$ is the largest integer not larger than x. For example, $\mathcal{L}_H \overline{\boldsymbol{X}}(\tau_L, 2) = \overline{\boldsymbol{X}}(\tau_L, 1)$ and $\mathcal{L}_H \overline{\boldsymbol{X}}(\tau_L, 1) = \overline{\boldsymbol{X}}(\tau_L - 1, m)$. Letting $\mathcal{L}_L$ be the *low frequency* lag operator, we have that $\mathcal{L}_L \overline{\boldsymbol{X}}(\tau_L, 1) = \mathcal{L}_H^m \overline{\boldsymbol{X}}(\tau_L, 1) = \overline{\boldsymbol{X}}(\tau_L - 1, 1)$.

Assume that $\{\{\overline{\boldsymbol{X}}(\tau_L, k)\}_k\}_{\tau_L}$ follows a VAR(p) process with $p \in \mathbb{N} \cup \{\infty\}$:

$$\overline{\boldsymbol{X}}(\tau_L, k) = \sum_{l=1}^{p} \boldsymbol{\Phi}_l \mathcal{L}_H^l \overline{\boldsymbol{X}}(\tau_L, k) + \overline{\boldsymbol{\eta}}(\tau_L, k), \tag{4.1}$$

where $\overline{\boldsymbol{\eta}}(\tau_L, k) \overset{i.i.d.}{\sim} (\boldsymbol{0}_{K^* \times 1}, \boldsymbol{V})$. The coefficient matrix $\boldsymbol{\Phi}_l$ is partitioned in the following manner:

$$\boldsymbol{\Phi}_l = \begin{bmatrix} \boldsymbol{\Phi}_{HH,l} & \boldsymbol{\Phi}_{HL,l} \\ \boldsymbol{\Phi}_{LH,l} & \boldsymbol{\Phi}_{LL,l} \end{bmatrix},$$

where $\boldsymbol{\Phi}_{yz,l} \in \mathbb{R}^{K_y \times K_z}$ with $y, z \in \{H, L\}$.

A general linear aggregation scheme is considered, appearing in (2.1). By an application of Theorem 1 in Lütkepohl (1984), the mixed frequency vector $\boldsymbol{X}(\tau_L)$ defined in (2.2) and the low frequency vector

defined as

$$\underline{\boldsymbol{X}}(\tau_L) = [\boldsymbol{x}_H(\tau_L)', \boldsymbol{x}_L(\tau_L)']' \in \mathbb{R}^{K^*} \tag{4.2}$$

follow VARMA processes. More specifically, we have the following.

Theorem 4.1. Suppose that an underlying high frequency process follows a VAR(p). Then the corresponding MF process is a VARMA(p_M, q_M), and the corresponding low frequency process is a VARMA(p_L, q_L). Moreover,

$$p_M \leq \text{deg}\left[\det(\overline{\mathbb{A}}(\mathcal{L}_L))\right] \equiv g \text{ and } p_L \leq g,$$

where g is the degree of polynomial of $\det(\overline{\mathbb{A}}(\mathcal{L}_L))$. Furthermore,

$$q_M \leq \max\left\{\text{deg}\left[\overline{\mathbb{A}}_{kl}(\mathcal{L}_L)\right] - g + p_M |\ k,l = 1,\ldots,mK^*\right\},$$

where $\overline{\mathbb{A}}_{kl}(\mathcal{L}_L)$ is the (k,l)-th cofactor of $\overline{\mathbb{A}}(\mathcal{L}_L)$. Similarly,

$$q_L \leq \max\left\{\text{deg}\left[\overline{\mathbb{A}}_{kl}(\mathcal{L}_L)\right] - g + p_L |\ k,l = 1,\ldots,mK^*\right\}.$$

Finally, if the high frequency VAR process is stationary then so are the mixed and low frequency VARMA processes.

Remark: See Appendix B for a proof, and for completeness the construction of $\overline{\mathbb{A}}(\mathcal{L}_L)$.

In general it is impossible to characterize p_M, q_M, p_L, or q_L exactly (cfr. Lütkepohl (1984)). Nevertheless, if the HF process $\{\boldsymbol{X}(\tau_L, k)\}$ is governed by a VAR(p) then the MF and LF processes $\{\boldsymbol{X}(\tau_L)\}$ and $\{\underline{\boldsymbol{X}}(\tau_L)\}$ have VARMA representations, and therefore VAR(∞) representations under the assumption of invertibility. Thus, one can still estimate those invertible VARMA processes by using a finite order approximation as in Lewis and Reinsel (1985), Lütkepohl and Poskitt (1996), and Saikkonen and Lütkepohl (1996). Moreover, the VARMA order can be characterized under certain simple cases such as stock sampling with $p = 1$.

Example : *stock sampling with p = 1:* Suppose that an underlying HF process follows a VAR(1) $\overline{\boldsymbol{X}}(\tau_L, k) = \boldsymbol{\Phi}_1 \mathcal{L}_H^1 \overline{\boldsymbol{X}}(\tau_L, k) + \overline{\boldsymbol{\eta}}(\tau_L, k)$ where $\overline{\boldsymbol{\eta}}(\tau_L, k) \overset{i.i.d.}{\sim} (\boldsymbol{0}_{K^* \times 1}, \boldsymbol{V})$. Then it is easy to show that the corresponding MF process also follows a VAR(1) if we consider stock sampling:

$$\boldsymbol{X}(\tau_L) = \boldsymbol{A}_1 \boldsymbol{X}(\tau_L - 1) + \boldsymbol{\epsilon}(\tau_L). \tag{4.3}$$

The parameter $\boldsymbol{A}_1$ is

$$\boldsymbol{A}_1 = \begin{bmatrix} \boldsymbol{0}_{K_H \times (m-1)K_H} & \boldsymbol{\Phi}_{HH,1}^{[1]} & \boldsymbol{\Phi}_{HL,1}^{[1]} \\ \vdots & \vdots & \vdots \\ \boldsymbol{0}_{K_H \times (m-1)K_H} & \boldsymbol{\Phi}_{HH,1}^{[m]} & \boldsymbol{\Phi}_{HL,1}^{[m]} \\ \boldsymbol{0}_{K_L \times (m-1)K_H} & \boldsymbol{\Phi}_{LH,1}^{[m]} & \boldsymbol{\Phi}_{LL,1}^{[m]} \end{bmatrix}, \tag{4.4}$$

where

$$\boldsymbol{\Phi}_l^k \equiv \begin{bmatrix} \boldsymbol{\Phi}_{HH,l}^{[k]} & \boldsymbol{\Phi}_{HL,l}^{[k]} \\ \boldsymbol{\Phi}_{LH,l}^{[k]} & \boldsymbol{\Phi}_{LL,l}^{[k]} \end{bmatrix}.$$

It follows that $\boldsymbol{\epsilon}(\tau_L) \overset{i.i.d.}{\sim} (\mathbf{0}_{K\times 1}, \boldsymbol{\Omega})$ where $\boldsymbol{\Omega}$ can be explicitly characterized as a function of $\boldsymbol{\Phi}_1$ and $\boldsymbol{V}$. First, observe that

$$\boldsymbol{\epsilon}(\tau_L) = \begin{bmatrix} \sum_{k=1}^{1} \begin{bmatrix} \boldsymbol{\Phi}_{HH,1}^{[1-k]} & \boldsymbol{\Phi}_{HL,1}^{[1-k]} \end{bmatrix} \boldsymbol{\eta}(\tau_L, k) \\ \vdots \\ \sum_{k=1}^{m} \begin{bmatrix} \boldsymbol{\Phi}_{HH,1}^{[m-k]} & \boldsymbol{\Phi}_{HL,1}^{m-k]} \end{bmatrix} \boldsymbol{\eta}(\tau_L, k) \\ \sum_{k=1}^{m} \begin{bmatrix} \boldsymbol{\Phi}_{LH,1}^{[m-k]} & \boldsymbol{\Phi}_{LL,1}^{[m-k]} \end{bmatrix} \boldsymbol{\eta}(\tau_L, k) \end{bmatrix}.$$

The covariance matrix $\boldsymbol{\Omega} \equiv \boldsymbol{E}[\boldsymbol{\epsilon}(\tau_L)\boldsymbol{\epsilon}(\tau_L)']$ has a block representation

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Omega}_{1,1} & \dots & \boldsymbol{\Omega}_{1,m} & \boldsymbol{\Omega}_{1,m+1} \\ \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{\Omega}_{1,m}' & \dots & \boldsymbol{\Omega}_{m,m} & \boldsymbol{\Omega}_{m,m+1} \\ \boldsymbol{\Omega}_{1,m+1}' & \dots & \boldsymbol{\Omega}_{m,m+1}' & \boldsymbol{\Omega}_{m+1,m+1} \end{bmatrix} \in \mathbb{R}^{K\times K}, \tag{4.5}$$

with components

$$\boldsymbol{\Omega}_{i,j} = \sum_{k=1}^{i} \begin{bmatrix} \boldsymbol{\Phi}_{HH,1}^{[i-k]} & \boldsymbol{\Phi}_{HL,1}^{[i-k]} \end{bmatrix} \boldsymbol{V} \begin{bmatrix} \boldsymbol{\Phi}_{HH,1}^{[j-k]'} \\ \boldsymbol{\Phi}_{HL,1}^{[j-k]'} \end{bmatrix} \text{ for } i,j \in \{1,\dots,m\} \text{ and } i \le j, \tag{4.6}$$

$$\boldsymbol{\Omega}_{i,m+1} = \sum_{k=1}^{i} \begin{bmatrix} \boldsymbol{\Phi}_{HH,1}^{[i-k]} & \boldsymbol{\Phi}_{HL,1}^{[i-k]} \end{bmatrix} \boldsymbol{V} \begin{bmatrix} \boldsymbol{\Phi}_{LH,1}^{[m-k]'} \\ \boldsymbol{\Phi}_{LL,1}^{[m-k]'} \end{bmatrix} \text{ for } i \in \{1,\dots,m\}$$

and

$$\boldsymbol{\Omega}_{m+1,m+1} = \sum_{k=1}^{m} \begin{bmatrix} \boldsymbol{\Phi}_{LH,1}^{[m-k]} & \boldsymbol{\Phi}_{LL,1}^{[m-k]} \end{bmatrix} \boldsymbol{V} \begin{bmatrix} \boldsymbol{\Phi}_{LH,1}^{[m-k]'} \\ \boldsymbol{\Phi}_{LL,1}^{[m-k]'} \end{bmatrix}. \tag{4.7}$$

Similarly, the LF process follows a VAR(1):

$$\underline{\boldsymbol{X}}(\tau_L) = \underline{\boldsymbol{A}}_1 \underline{\boldsymbol{X}}(\tau_L - 1) + \underline{\boldsymbol{\epsilon}}(\tau_L), \tag{4.8}$$

where

$$\underline{\boldsymbol{A}}_1 = \boldsymbol{\Phi}_1^m, \tag{4.9}$$

and $\underline{\boldsymbol{\epsilon}}(\tau_L) \overset{i.i.d.}{\sim} (\mathbf{0}_{K^*\times 1}, \underline{\boldsymbol{\Omega}})$. The covariance matrix follows by noting $\underline{\boldsymbol{\epsilon}}(\tau_L) = \sum_{k=1}^{m} \boldsymbol{\Phi}_1^{m-k} \boldsymbol{\eta}(\tau_L, k)$ hence

$$\underline{\boldsymbol{\Omega}} = \sum_{k=1}^{m} \boldsymbol{\Phi}_1^{m-k} \boldsymbol{V} (\boldsymbol{\Phi}_1^{m-k})' \in \mathbb{R}^{K^*\times K^*}. \tag{4.10}$$

4.2 Causality and Temporal Aggregation

Felsenstein et al. (2013) explore conditions for identifying a HF process based on MF data. When their conditions are satisfied, recovery of HF causality is trivially feasible by looking at off-diagonal elements of the identified HF-VAR coefficients. The conditions for identification are stringent, however, and one may therefore wonder what happens if they are not satisfied. In this subsection we fill some of the gap by focusing on testing for causality since this does not require full identification of the entire HF process.

Since Granger causality is based on information sets, we need to define reference information sets for HF- and LF-VAR processes. To this end, we rewrite a HF-VAR(p) process in (4.1) with a single time index t: $\boldsymbol{Y}_t = \sum_{l=1}^{p} \boldsymbol{\Phi}_l \boldsymbol{Y}_{t-l} + \boldsymbol{\xi}_t$, where $\boldsymbol{Y}_t \in \mathbb{R}^{K^*}$ is simply a single-index version of $\boldsymbol{X}(\tau_L, k)$. One way of mapping (τ_L, k) to t is to let $t = m(\tau_L - 1) + k$ so that $\boldsymbol{Y}_1$ corresponds to $\boldsymbol{X}(1,1)$. The same mapping is used between $\boldsymbol{\xi}_t$ and $\boldsymbol{\eta}(\tau_L, k)$. Recall from Section 3.1 that $\mathcal{I}(\tau_L)$ is the MF *reference information set in period* τ_L, while $\mathcal{I} = \{\mathcal{I}(\tau_L) |\ \tau_L \in \mathbb{Z}\}$ is the MF *reference information set*. We now introduce HF and LF versions of the information set. The *HF reference information set at time* t is defined as $\overline{\mathcal{I}}(t) = \boldsymbol{Y}(-\infty, t]$. The *HF reference information set* is defined as $\overline{\mathcal{I}} = \{\overline{\mathcal{I}}(t) \,|\, t \in \mathbb{Z}\}$. The prediction horizon for non-causality given $\overline{\mathcal{I}}$ is in terms of the high frequency, denoted by $\overline{h} \in \mathbb{Z}$. For example, non-causality from all high frequency variables to all low frequency variables at high frequency horizon $\overline{h}$ given $\overline{\mathcal{I}}$ is written as $\boldsymbol{x}_H \nrightarrow_{\overline{h}} \boldsymbol{x}_L \,|\, \overline{\mathcal{I}}$. Similarly, the *LF reference information set at time* τ_L is defined as $\underline{\mathcal{I}}(\tau_L) = \underline{\boldsymbol{X}}(-\infty, \tau_L]$, where $\underline{\boldsymbol{X}}(\tau_L)$ is given in (4.2). The *LF reference information set* is defined as $\underline{\mathcal{I}} = \{\underline{\mathcal{I}}(\tau_L) \,|\, \tau_L \in \mathbb{Z}\}$. Whether (non-)causality is preserved under temporal aggregation depends mainly on three conditions: an aggregation scheme, VAR lag order p, and the presence of an auxiliary variable and therefore the possibility of causality chains. The existing literature has found that temporal aggregation may hide or generate causality even in very simple cases. We show that the MF approach recovers underlying causality patterns better than the traditional LF approach.

Theorem 4.2. Consider the linear aggregation scheme appearing in (2.1) and assume a HF-VAR(p) with $p \in \mathbb{N} \cup \{\infty\}$. Then, the following two properties hold when applied respectively to all low and all high frequency processes: (i) If $\boldsymbol{x}_H \nrightarrow \boldsymbol{x}_L \,|\, \mathcal{I}$, then $\boldsymbol{x}_H \nrightarrow \boldsymbol{x}_L \,|\, \underline{\mathcal{I}}$. (ii) If $\boldsymbol{x}_L \nrightarrow \boldsymbol{x}_H \,|\, \overline{\mathcal{I}}$, then $\boldsymbol{x}_L \nrightarrow \boldsymbol{x}_H \,|\, \mathcal{I}$.

Proof: See Appendix C.

Note that the prediction horizon in Theorem 4.2 is arbitrary since there are no auxiliary variables involved. This follows since we only examine the relationship between all low and all high frequency processes respectively.[9]

Theorem 4.2 part (i) states that non-causality from all high frequency variables to all low frequency variables is preserved between MF and LF processes, while part (ii) states that non-causality from all low frequency variables to all high frequency variables is preserved between HF and MF processes. One might incorrectly guess from Theorem 4.2 part (ii) that $\boldsymbol{x}_L \nrightarrow \boldsymbol{x}_H \,|\, \mathcal{I} \Rightarrow \boldsymbol{x}_L \nrightarrow \boldsymbol{x}_H \,|\, \underline{\mathcal{I}}$. This statement does not hold in general. A simple counter-example is a HF-VAR(2) process with stock sampling, $m = 2$, $K_H = K_L = 1$,

[9] We can deduce virtually no theoretical results in the presence of auxiliary variables since potential causal chains complicate causality patterns substantially.

$$\mathbf{\Phi}_1 = \begin{bmatrix} \phi_{HH,1} & 0 \\ \phi_{LH} & 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{\Phi}_2 = \begin{bmatrix} \phi_{HH,2} & 0 \\ 0 & 0 \end{bmatrix}.$$

Assume that $\phi_{HH,1}$, $\phi_{HH,2}$, and ϕ_{LH} are all nonzero. Note that, given $\overline{\mathcal{I}}$, x_L does not cause x_H while x_H does cause x_L. In this particular case, we can derive the corresponding MF- and LF-VAR(1) processes. The MF coefficient is

$$\boldsymbol{A}_1 = \begin{bmatrix} \phi_{HH,2} & \phi_{HH,1} & 0 \\ \phi_{HH,1}\phi_{HH,2} & \phi_{HH,1}^2 + \phi_{HH,2} & 0 \\ \phi_{LH}\phi_{HH,2} & \phi_{LH}\phi_{HH,1} & 0 \end{bmatrix}, \tag{4.11}$$

while the LF coefficient is

$$\underline{\boldsymbol{A}}_1 = \begin{bmatrix} \phi_{HH,1}^2 + \phi_{HH,2} & \phi_{HH,1}\phi_{HH,2}/\phi_{LH} \\ \phi_{LH}\phi_{HH,1} & \phi_{HH,2} \end{bmatrix}. \tag{4.12}$$

Equations (4.11) and (4.12) indicate that x_L does not cause x_H given $\mathcal{I}$, but x_L does cause x_H given $\underline{\mathcal{I}}$. Thus, we confirm that non-causality from all low frequency variables to all high frequency variables is *not necessarily* preserved between MF and LF processes.

Summarizing Theorem 4.2 and the counter-example above, a crucial condition for non-causality preservation is that the information for the "right-hand side" variables (i.e. $\boldsymbol{x}_L$ for (i) and $\boldsymbol{x}_H$ for (ii)) is not lost by temporal aggregation. In this sense, the MF approach yields more implications on hidden causality patterns than the LF approach, which switches directly from a HF process by aggregating all variables.

To conclude the subsection we again focus on stock sampling with p =1 as this particular case yields much sharper results.

Example: *stock sampling with p = 1:* When p = 1 and stock sampling is of interest, the exact functional form for the MF and LF processes is known and appear in (4.3) and (4.8). Equation (4.4) highlights what kind of causality information gets lost by switching from a HF- to MF-VAR. Similarly, (4.9) reveals the information loss when moving from a MF- to LF-VAR. This brings us to the following theorem.

Theorem 4.3. Consider stock sampling with $p = 1$. Then, the corresponding MF-VAR and LF-VAR processes are also of order 1. Furthermore, non-causation among the HF-, MF-, and LF-VAR processes is related as follows.

i. *In Case 1 (low $\nrightarrow$ low) and Case 2 (high $\nrightarrow$ low), non-causation up to HF horizon m given the HF information set $\overline{\mathcal{I}}$ implies non-causation at horizon* 1 *given the MF information set $\mathcal{I}$, which is necessary and sufficient for non-causation at horizon* 1 *given the LF information set $\underline{\mathcal{I}}$.*

ii. *In Case 3 (low $\nrightarrow$ high) and Case 4 (high $\nrightarrow$ high), non-causation up to HF horizon m given $\overline{\mathcal{I}}$ is necessary and sufficient for non-causation at horizon* 1 *given $\mathcal{I}$, which implies non-causation at horizon* 1 *given $\underline{\mathcal{I}}$.*

iii. *In Case I (all high $\nrightarrow$ all low), non-causation at HF horizon* 1 *given* $\overline{\mathcal{I}}$ *implies non-causation at horizon* 1 *given* $\mathcal{I}$*, which is necessary and sufficient for non-causation at horizon* 1 *given* $\underline{\mathcal{I}}$*.*

iv. *In Case II (all low $\nrightarrow$ all high), non-causation at HF horizon* 1 *given* $\overline{\mathcal{I}}$ *is necessary and sufficient for non-causation at horizon* 1 *given* $\mathcal{I}$*, which implies non-causation at horizon* 1 *given* $\underline{\mathcal{I}}$*.*

Proof: See Appendix D.

Although Theorem 4.3 is much sharper than Theorem 4.2 due to much stronger assumptions, they share an interesting feature that causality tends to be contaminated more when temporal aggregation discards information for "right-hand side" variables. For example, item 2 shows that no relevant information for testing low-to-high or high-to-high causality is lost when moving from $\overline{\mathcal{I}}$ to $\mathcal{I}$ (i.e., when aggregating low frequency variables), while some information is lost when moving from $\mathcal{I}$ to $\underline{\mathcal{I}}$ (i.e., when aggregating high frequency variables).

Theorem 4.3 suggests that the MF causality test should never perform worse than the low frequency causality test, and the former should be more powerful than the latter especially when Cases 3, 4, and II are of interest. Sections 5 and 6 verify this point by a local asymptotic power analysis and a Monte Carlo simulation, respectively.

5 Local Asymptotic Power Analysis

The goal of this section is to show that the MF causality tests have higher local asymptotic power compared to the LF causality test. We need to constrain our attention to analytically tractable DGPs, which is why we consider a bivariate HF-VAR(1) process with stock sampling. As shown in the previous section, for the bivariate HF-VAR(1) one can derive analytically the corresponding MF- and LF-VAR(1) processes. Recall that Case I considers unidirectional causality from the high frequency variable to the low frequency variable, while Case II considers unidirectional causality from the low frequency variable to the high frequency variable. We first compute the local asymptotic power functions for both cases, and then plot them in a numerical exercise.

Case I: High-to-Low Causality In order to characterize local asymptotic power, assume that the high frequency DGP is given by:

$$\boldsymbol{X}(\tau_L, k) = \boldsymbol{\Phi}(\nu/\sqrt{T})\mathcal{L}_H \boldsymbol{X}(\tau_L, k) + \boldsymbol{\eta}(\tau_L, k), \tag{5.1}$$

where

$$\boldsymbol{\Phi}(\nu/\sqrt{T}) = \begin{bmatrix} \rho_H & 0 \\ \nu/\sqrt{T} & \rho_L \end{bmatrix}$$

with $\rho_H, \rho_L \in (-1, 1)$, where $\nu \in \mathbb{R}$ is the usual Pitman drift parameter. Assume for computational simplicity that $\boldsymbol{\eta}(\tau_L, k) \overset{i.i.d.}{\sim} (\mathbf{0}_{2\times 1}, \boldsymbol{I}_2)$, hence the errors are uncorrelated and $\boldsymbol{X}(\tau_L, k)$ has a strictly stationary solution. In the true DGP, the low frequency variable does not cause the high frequency

variable, while for $\nu \neq 0$ the high frequency variable causes the low frequency variable with a marginal impact of $\nu/\sqrt{T}$ which vanishes as $T \to \infty$. First note we have $p = h = 1$. We will therefore simplify notation, namely denote the least squares asymptotic covariance matrix $\mathbf{\Sigma}_p(h)$ as $\mathbf{\Sigma}_1$.

Assuming stock sampling and general $m \in \mathbb{N}$, the corresponding MF-VAR(1) process of dimension $K = m + 1$ (since $K_H = K_L = 1$) is as follows:

$$\boldsymbol{X}(\tau_L) = \boldsymbol{A}(\nu/\sqrt{T})\boldsymbol{X}(\tau_L - 1) + \boldsymbol{\epsilon}(\tau_L), \tag{5.2}$$

where

$$\boldsymbol{A}(\nu/\sqrt{T}) = \begin{bmatrix} \mathbf{0}_{1\times(m-1)} & \rho_H & 0 \\ \vdots & \vdots & \vdots \\ \mathbf{0}_{1\times(m-1)} & \rho_H^m & 0 \\ \mathbf{0}_{1\times(m-1)} & \sum_{k=1}^{m} \rho_H^{k-1}\rho_L^{m-k}(\nu/\sqrt{T}) & \rho_L^m \end{bmatrix} \tag{5.3}$$

and $\boldsymbol{\epsilon}(\tau_L) \overset{i.i.d.}{\sim} (\mathbf{0}_{K\times 1}, \mathbf{\Omega})$ See (4.5)-(4.7) in Section 4.1 for a characterization of $\mathbf{\Omega}$. The MF-VAR(1) being estimated is:

$$\boldsymbol{X}(\tau_L) = \boldsymbol{A} \times \boldsymbol{X}(\tau_L - 1) + \boldsymbol{\epsilon}(\tau_L)$$

with coefficient matrix $\boldsymbol{A} = \boldsymbol{A}(\nu/\sqrt{T})$. Table 1 and Theorem 3.1 provide us the Case I selection matrix $\boldsymbol{R}$ to formulate the null hypothesis of high-to-low non-causality:

$$H_0^I : \boldsymbol{R}\text{vec}\left[\boldsymbol{A}'\right] = \mathbf{0}_{m\times 1} \text{ where } \boldsymbol{R} \in \mathbb{R}^{m\times K^2}.$$

Thus, the corresponding local alternatives $H_A^{I,L}$ are written as

$$H_A^{I,L} : \boldsymbol{R}\text{vec}\left[\boldsymbol{A}'\right] = (\nu/\sqrt{T})\boldsymbol{a},$$

where by (5.3) it follows $\boldsymbol{a}$ is the $m \times 1$ vector $[0, ..., 0, \sum_{k=1}^{m} \rho_H^{k-1}\rho_L^{m-k}]'$. Now let $\hat{\boldsymbol{A}}$ be the least squares estimator of $\boldsymbol{A}$. Theorem 2.2 implies that $W[H_0^I] \overset{d}{\to} \chi_m^2$ as $T \to \infty$ under H_0^I. Similarly, by classic arguments it is easy to verify under $H_A^{I,L}$ that $W[H_A^{I,L}] \overset{d}{\to} \chi_m^2(\kappa_{MF})$, where $\chi_m^2(\kappa_{MF})$ is the non-central chi-squared distribution with m degrees of freedom and non-centrality parameter κ_{MF} :

$$\kappa_{MF} = \nu^2 \boldsymbol{a}' \left[\boldsymbol{R}\mathbf{\Sigma}_1\boldsymbol{R}'\right]^{-1} \boldsymbol{a}, \tag{5.4}$$

where $\mathbf{\Sigma}_1$ is the asymptotic variance of $\hat{\boldsymbol{A}}$, in particular

$$\mathbf{\Sigma}_1 = \mathbf{\Omega} \otimes \mathbf{\Upsilon}_0^{-1} \quad \text{with} \quad \mathbf{\Upsilon}_0 = \sum_{i=0}^{\infty} \boldsymbol{A}^i \mathbf{\Omega} \boldsymbol{A}^{i'} \text{ where } \boldsymbol{A} \equiv \lim_{T\to\infty} \boldsymbol{A}(\nu/\sqrt{T}). \tag{5.5}$$

Equation (5.5) can be obtained from non-local least squares asymptotics with $\boldsymbol{A} \equiv \lim_{T\to\infty} \boldsymbol{A}(\nu/\sqrt{T})$. See Appendices A.1.1 and A.1.2, in particular the example in Appendix A.1.2, for details on deriving

$\boldsymbol{\Sigma}_1$ in (5.5). Using the discrete Lyapunov equation, $\boldsymbol{\Upsilon}_0$ can be characterized by:

$$\text{vec}[\boldsymbol{\Upsilon}_0] = (\boldsymbol{I}_{K^2} - \boldsymbol{A} \otimes \boldsymbol{A})^{-1}\text{vec}[\boldsymbol{\Omega}].$$

Let $F_0 : \mathbb{R} \to [0,1]$ be the cumulative distribution function (c.d.f.) of the null distribution, χ^2_m. Similarly, let $F_1 : \mathbb{R} \to [0,1]$ be the c.d.f. of the alternative distribution, $\chi^2_m(\kappa_{MF})$. The local asymptotic power of the MF high-to-low causality test, $\mathcal{P}$, is given by:

$$\mathcal{P} = 1 - F_1\left[F_0^{-1}(1-\alpha)\right], \tag{5.6}$$

where $\alpha \in [0,1]$ is a nominal size.

We now derive the local asymptotic power of the LF high-to-low causality test. First, the LF-VAR(1) process corresponding to (5.1) is given by:

$$\underline{\boldsymbol{X}}(\tau_L) = \underline{\boldsymbol{A}}(\nu/\sqrt{T})\underline{\boldsymbol{X}}(\tau_L - 1) + \underline{\boldsymbol{\epsilon}}(\tau_L), \tag{5.7}$$

where

$$\underline{\boldsymbol{A}}(\nu/\sqrt{T}) = \begin{bmatrix} \rho_H^m & 0 \\ \sum_{k=1}^m \rho_H^{k-1}\rho_L^{m-k}(\nu/\sqrt{T}) & \rho_L^m \end{bmatrix} \tag{5.8}$$

and $\underline{\boldsymbol{\epsilon}}(\tau_L) \overset{i.i.d.}{\sim} (\boldsymbol{0}_{2\times 1}, \underline{\boldsymbol{\Omega}})$. Note that $\underline{\boldsymbol{\Omega}}$ is characterized in (4.10).

Suppose that we fit a LF-VAR(1) model with coefficient matrix $\underline{\boldsymbol{A}} \in \mathbb{R}^{2\times 2}$, that is $\underline{\boldsymbol{X}}(\tau_L) = \underline{\boldsymbol{A}}\underline{\boldsymbol{X}}(\tau_L - 1) + \underline{\boldsymbol{\epsilon}}(\tau_L)$. The null hypothesis of high-to-low non-causality is that the lower-left element of $\underline{\boldsymbol{A}}$ is zero:

$$H_0^I : \boldsymbol{R}\text{vec}\left[\underline{\boldsymbol{A}}'\right] = 0,$$

where $\boldsymbol{R} = [0,0,1,0]$. The corresponding local alternative hypothesis is:

$$H_A^{I,L} : \boldsymbol{R}\text{vec}\left[\underline{\boldsymbol{A}}'\right] = \sum_{k=1}^m \rho_H^{k-1}\rho_L^{m-k}(\nu/\sqrt{T}).$$

Let $\hat{\underline{\boldsymbol{A}}}$ be the least squares estimator of $\underline{\boldsymbol{A}}$. We have that $W[H_0^I] \xrightarrow{d} \chi^2_1$ as $T \to \infty$ under H_0^I, while $W[H_A^{I,L}] \xrightarrow{d} \chi^2_1(\kappa_{LF})$ under $H_A^{I,L}$ with κ_{LF} given by:

$$\kappa_{LF} = \frac{\left(\nu \sum_{k=1}^m \rho_H^{k-1}\rho_L^{m-k}\right)^2}{\boldsymbol{R}\underline{\boldsymbol{\Sigma}}_{\boldsymbol{1}}\boldsymbol{R}'},$$

where $\underline{\boldsymbol{\Sigma}}_{\boldsymbol{1}}$ is the asymptotic variance of $\hat{\underline{\boldsymbol{A}}} \equiv \lim_{T\to\infty}\{\underline{\boldsymbol{A}}(\nu/\sqrt{T})\}$, in particular as in (5.5) it can be shown $\underline{\boldsymbol{\Sigma}}_1 = \underline{\boldsymbol{\Omega}} \otimes \underline{\boldsymbol{\Upsilon}}_0^{-1}$ with $\underline{\boldsymbol{\Upsilon}}_0 = \sum_{i=0}^\infty \underline{\boldsymbol{A}}^i\underline{\boldsymbol{\Omega}}\underline{\boldsymbol{A}}^{i'}$. The local asymptotic power of the LF high-to-low causality test is given by (5.6), where F_0 is the c.d.f. of χ^2_1 and F_1 is the c.d.f. of $\chi^2_1(\kappa_{LF})$.

Case II: Low-to-High Causality Assume that the true DGP is given by (5.1) with

$$\mathbf{\Phi}(\nu/\sqrt{T}) = \begin{bmatrix} \rho_H & \nu/\sqrt{T} \\ 0 & \rho_L \end{bmatrix}$$

with ρ_H, $\rho_L \in (-1, 1)$. Assume again that $\boldsymbol{\eta}(\tau_L, k) \overset{i.i.d.}{\sim} (\mathbf{0}_{2\times 1}, \boldsymbol{I}_2)$. In the true DGP, the high frequency variable does not cause the low frequency variable, while the low frequency variable causes the high frequency variable, a relationship which vanishes as $T \to \infty$.

Assuming stock sampling and general $m \in \mathbb{N}$, the corresponding MF-VAR(1) process is given by (5.2) with

$$\boldsymbol{A}(\nu/\sqrt{T}) = \begin{bmatrix} \mathbf{0}_{1\times(m-1)} & \rho_H & \sum_{k=1}^{1} \rho_H^{k-1}\rho_L^{1-k}(\nu/\sqrt{T}) \\ \vdots & \vdots & \vdots \\ \mathbf{0}_{1\times(m-1)} & \rho_H^m & \sum_{k=1}^{m} \rho_H^{k-1}\rho_L^{m-k}(\nu/\sqrt{T}) \\ \mathbf{0}_{1\times(m-1)} & 0 & \rho_L^m \end{bmatrix}. \tag{5.9}$$

Our model is again a MF-VAR(1) model, so the local asymptotic power of the MF low-to-high causality test can be computed exactly as in Case I with only two changes. First, $\boldsymbol{a}$ in (5.4) has different elements here: $\boldsymbol{a} = [\sum_{k=1}^{1} \rho_H^{k-1}\rho_L^{1-k}, \ldots, \sum_{k=1}^{m} \rho_H^{k-1}\rho_L^{m-k}]'$. Second, the selection matrix $\boldsymbol{R}$ is specified according to Case II in Section 3.2. These differences will produce an interesting asymmetry between the MF high-to-low causality test and the MF low-to-high causality test.

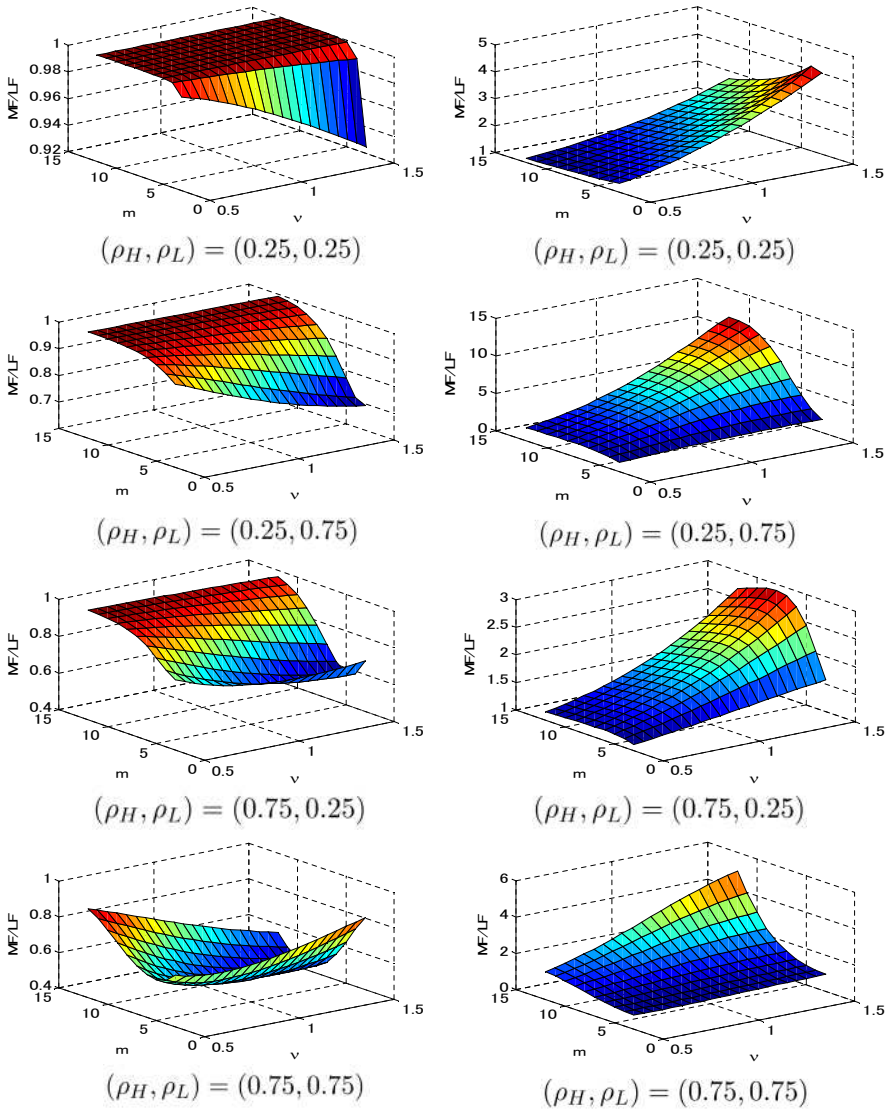
We now consider the LF low-to-high causality test. The LF-VAR(1) process is given by:

$$\underline{\boldsymbol{A}}(\nu/\sqrt{T}) = \begin{bmatrix} \rho_H^m & \sum_{k=1}^{m} \rho_H^{k-1}\rho_L^{m-k}(\nu/\sqrt{T}) \\ 0 & \rho_L^m \end{bmatrix}. \tag{5.10}$$

The local asymptotic power of the LF low-to-high causality test can again be computed exactly as in Case I with the only difference being that $\boldsymbol{R} = [0, 1, 0, 0]$ here, so there is no asymmetry between the LF high-to-low causality test and the LF low-to-high causality test.

Numerical Exercises To study the local asymptotic power analysis more directly, we rely on some numerical calculations. In Figure 1 we plot the ratio of the local asymptotic power of the MF causality test to that of the LF causality test, which we call the *power ratio* hereafter. We assume a nominal size $\alpha = 0.05$. Panel A focuses on high-to-low causality, while Panel B focuses on low-to-high causality. Each panel has four figures depending on $\rho_H, \rho_L \in \{0.25, 0.75\}$. The x-axis of each figure has $\nu \in [0.5, 1.5]$, while the y-axis has $m \in \{3, \ldots, 12\}$. The case that $m = 3$ can be thought of as the month versus quarter case, while the case that $m = 12$ can be thought of as the month versus year case. Note that the scale of each z-axis is different.

In Panel A, the power ratio varies within $[0.5, 1]$, hence the MF causality test is as powerful as, or is in fact *less* powerful than, the LF causality test. This is reasonable since a MF process contains the same information about high-to-low causality test as the corresponding LF process does (cfr. (5.3), (5.8), and Theorem 4.3) and the former has more parameters: recall that $\boldsymbol{A}$ is $(m+1)\times(m+1)$ while $\underline{\boldsymbol{A}}$ is 2×2.



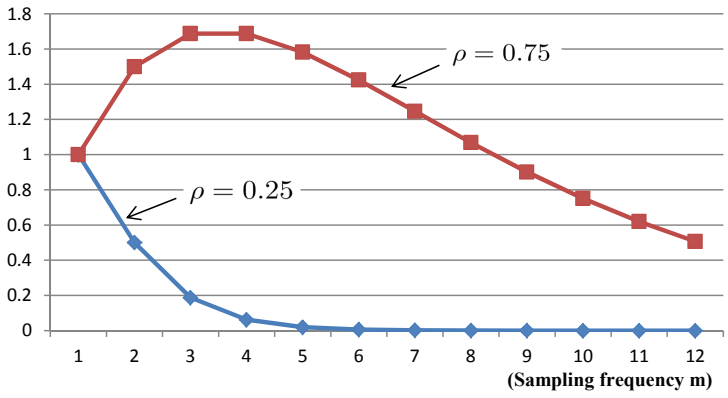
Note: The z-axis of each figure has the power ratio (i.e. the ratio of the local asymptotic power of the MF causality test to that of the low frequency causality test). Note that the scale of each z-axis is different. The x-axis has $\nu \in [0.5, 1.5]$, while the y-axis has $m \in \{3, \ldots, 12\}$.

Figure 1: Local Asymptotic Power of Mixed and Low Frequency Causality Tests

The power ratio tends to be low in the bottom figures of Panel A, where $\rho_H = 0.75$. This result is also understandable since the information loss caused by aggregating a high frequency variable is less severe when it is more persistent.

Panel B highlights the advantage of the MF approach over the LF approach. Note that the power ratio always exceeds one and the largest value of the z-axis is 5, 15, 3, or 6 when $(\rho_H, \rho_L) = (0.25, 0.25)$, $(0.25, 0.75)$, $(0.75, 0.25)$, or $(0.75, 0.75)$, respectively. This result is consistent with (5.9), (5.10), and Theorem 4.3, where we show that a MF process contains more information about low-to-high causality test than the corresponding LF process does. Given the same ρ_L, the power ratio tends to be low when the high frequency variable is more persistent. The reason for this result is again that aggregating a high frequency variable produces less severe information loss when it is more persistent.

Another interesting finding from Panel B is that the power ratio is decreasing in m for $(\rho_H, \rho_L) = (0.25, 0.25)$ and increasing in m for $(\rho_H, \rho_L) = (0.75, 0.75)$. To interpret this fact, let $\rho_H = \rho_L = \rho$ and consider a key quantity in the upper-right block of $\boldsymbol{A}$, $\sum_{k=1}^{m} \rho_H^{k-1} \rho_L^{m-k} = m\rho^{m-1} \equiv f(m)$. Given m, the upper-right block of $\boldsymbol{A}$ has $f(1), \ldots, f(m)$ while that of $\underline{\boldsymbol{A}}$ has $f(m)$ only, therefore it is $\{f(1), \ldots, f(m-1)\}$ that determines the power ratio. Hence, whether the power ratio increases or decreases by switching from m to $m+1$ depends on the magnitude of $f(m)$. If $f(m)$ is close to zero, then the power ratio decreases due to more parameters in a MF-VAR model and negligible informational gain from $f(m)$. If $f(m)$ is away from zero, then the power ratio increases since such a large coefficient helps us reject the incorrect null hypothesis of low-to-high non-causality. Figure 2 plots $f(m)$ for $\rho \in \{0.25, 0.75\}$. It shows that $f(m)$ converges to zero quickly as m grows when $\rho = 0.25$, while it does much more slowly when $\rho = 0.75$. Thus, the power ratio is decreasing in m for $\rho = 0.25$ and increasing in m for $\rho = 0.75$.



Note: The horizontal axis has $m \in \{1, \ldots, 12\}$, while the vertical axis has $m\rho^{m-1}$ for $\rho \in \{0.25, 0.75\}$.

Figure 2: Plot of the Function $m\rho^{m-1}$ - Driver of Local Asymptotic Power Ratios

In summary, the local asymptotic power of the MF low-to-high causality test is higher than that of the LF counterpart. The ratio of the former to the latter increases as a high frequency variable gets less persistent, given the persistence of a low frequency variable. Moreover, the power ratio increases in m

for persistent series, while it decreases in m for transitory series.

6 Power Improvements in Finite Samples

This section conducts Monte Carlo simulations for a bivariate case and a trivariate case to evaluate the finite sample performance of the mixed frequency causality test. In bivariate cases with stock sampling, we know how causality is transferred among HF-, MF-, and LF-VAR processes and hence we can compare the finite sample power of MF and LF causality tests. In trivariate cases we have little theoretical results on how causality is transferred because of potential causality chains, so we just evaluate the performance of the MF causality test itself by checking empirical size and power based on a MF-VAR model. All tests in this section are performed at the 5% level.

6.1 Bivariate Case

This section considers a bivariate HF-VAR(1) with stock sampling as in Section 5 so that the corresponding MF- and LF-VAR models are known and correctly specified. One drawback of this experimental design is that we cannot easily study flow sampling since the corresponding MF and LF processes only have VARMA representations of unknown order, and therefore may not have a finite order VAR representation, by Theorem 4.1.[10]

6.1.1 Simulation Design

We draw J independent samples from a HF-VAR(1) process $\{\{\boldsymbol{X}(\tau_L,k)\}\}$ according to (4.1) with $\boldsymbol{\Phi}_1$ partitioned in two possible ways:

$$(a)\quad \begin{bmatrix} \phi_{HH,1} & \phi_{HL,1} \\ \phi_{LH,1} & \phi_{LL,1} \end{bmatrix} = \begin{bmatrix} 0.4 & 0.0 \\ 0.2 & 0.4 \end{bmatrix} \text{ and } (b)\quad \begin{bmatrix} \phi_{HH,1} & \phi_{HL,1} \\ \phi_{LH,1} & \phi_{LL,1} \end{bmatrix} = \begin{bmatrix} 0.4 & 0.2 \\ 0.0 & 0.4 \end{bmatrix}.$$

Thus we have in (a) unidirectional causality from the high frequency variable to the low frequency variable and in (b) unidirectional causality from the low frequency variable to the high frequency variable. Since we assume stock sampling here, these causal patterns carry over to the corresponding MF- and LF-VAR processes under this parametrization. The innovations are mutually and serially uncorrelated standard normal: $\boldsymbol{\eta}(\tau_L,k) \overset{i.i.d.}{\sim} N(\mathbf{0}_{2\times 1}, \boldsymbol{I}_2)$. The low frequency sample size is $T_L \in \{50, 100, 500\}$. The sampling frequency is taken from $m \in \{2,3\}$, so the high frequency sample size is $T = mT_L \in \{100, 150, 200, 300, 1000, 1500\}$. The case that $(m, T_L) = (3, 100)$ can be thought of as a month versus quarter case covering 25 years.

We extract the MF process $\{\boldsymbol{X}(\tau_L)\}_1^{T_L}$ and the LF process $\{\underline{\boldsymbol{X}}(\tau_L)\}_1^{T_L}$ according to (2.2) and (4.2) respectively, and fit MF-VAR(1) and LF-VAR(1), which are correctly specified. We then compute Wald

[10] In simulations not reported here we explored Lütkepohl and Poskitt's (1996) finite-order approximation for VAR(∞). The resulting test exhibited large empirical size distortions and was therefore not considered in this paper.

statistics for two separate null hypotheses of high-to-low non-causality $H_{H \nrightarrow L}$: $x_H \nrightarrow x_L$ and low-to-high non-causality $H_{L \nrightarrow H}$: $x_L \nrightarrow x_H$, each for horizon $h = 1$.[11] The Wald statistic shown in (2.9) is computed by OLS with two covariance matrix estimators. The first one is based on the Bartlett kernel HAC estimator discussed in Appendix A.1.3. We use a bandwidth of the form $n_{T_L^*} \equiv \max\{1, \lambda(T_L^*)^{1/3}\}$ since this optimizes the estimator's rate of convergence (Newey and West (1994)), while λ is determined by Newey and West's (1994) automatic bandwidth selection. This so-called *HAC case* corresponds to a situation where the researcher merely uses one robust covariance estimation technique irrespective of theory results.[12] The second covariance matrix is the true analytical matrix, and is therefore called the *benchmark case*. This case corresponds to a complete-information situation where the researcher knows the true parameters. The benchmark covariance matrix for the MF-VAR model can be computed according to (5.5). In the LF-VAR model, $\boldsymbol{A}$ and $\boldsymbol{\Omega}$ in that expression should be replaced with $\underline{\boldsymbol{A}}$ and $\underline{\boldsymbol{\Omega}}$, respectively (see (4.4), (4.5), (4.9), and (4.10)).

To circumvent size distortions for small samples $T_L \in \{50, 100\}$, we employ a parametric bootstrap as in Dufour, Pelletier, and Renault (2006, p. 351).[13] Although $p = h = 1$ in this section, we present the procedure with general p and h for completeness. We present the concrete procedure with respect to $H_0^2(h) : \boldsymbol{x}_{H,i_1} \nrightarrow_h x_{L,j_1} | \, \mathcal{I}_{(H,i_1)}$, but all other cases can be treated analogously.

Step 1 Fit an unrestricted MF-VAR(p) model for prediction horizon one to get $\hat{\boldsymbol{B}}(1)$ and $\hat{\boldsymbol{\Omega}}$ (cfr. (2.3) and (2.6)). Also fit an unrestricted MF-VAR(p) model for prediction horizon h to get $\hat{\boldsymbol{B}}(h)$ (cfr. (2.5)).

Step 2 Using (2.9), compute the Wald test statistic based on the actual data, $W[H_0^2(h)]$.

Step 3 Simulate N samples from (2.5) using $\boldsymbol{B}(h) = \hat{\boldsymbol{B}}(h)$ and $\boldsymbol{\Omega} = \hat{\boldsymbol{\Omega}}$ and the hypothesis that $\boldsymbol{\epsilon}(\tau_L)$ is jointly standard normal, where we impose parametric constraints corresponding to $H_0^2(h)$, found in (3.2) and Table 1. Estimates of the impulse response coefficients $\boldsymbol{\Psi}_k$ can be obtained using $\hat{\boldsymbol{B}}(1)$ and (2.4). We denote by $W_i[H_0^2(h)]$ the Wald test statistic based on the i-th simulated sample, where $i \in \{1, \ldots, N\}$.

Step 4 Compute the resulting p-value $\hat{p}_N(W[H_0^2(h)])$, defined as

$$\hat{p}_N(W[H_0^2(h)]) \equiv \frac{1}{N+1}\left(1 + \sum_{i=1}^{N} I(W_i[H_0^2(h)] \geq W[H_0^2(h)])\right).$$

The null hypothesis $H_0^2(h)$ is rejected at level α if $\hat{p}_N(W[H_0^2(h)]) \leq \alpha$.

[11] Notice from (4.3) and (4.8) that $H_{H \nrightarrow L}$ corresponds to $\boldsymbol{A}_1(m+1, 1:m) = \boldsymbol{0}_{1\times m}$ in the MF-VAR and to $\underline{\boldsymbol{A}}_1(2,1) = 0$ in the LF-VAR models, while $H_{L \nrightarrow H}$ corresponds to $\boldsymbol{A}_1(1:m, m+1) = \boldsymbol{0}_{m\times 1}$ in the MF-VAR and to $\underline{\boldsymbol{A}}_1(1,2) = 0$ in the LF-VAR models.

[12] In the special case when $h = 1$, a consistent and *almost surely* positive definite least squares asymptotic variance estimator is easily computed without a long-run variance HAC estimator (see Appendix A.1). Based on this insight, we also tried a sufficiently small λ instead of Newey and West's (1994) automatic selection. The results were similar to those of the HAC case, so are not reported here.

[13] Chauvet, Götz, and Hecq (2013) explore an alternative approach of parameter reductions based on reduced rank conditions, the imposition of an ARX(1) structure on the high frequency variables, and the transformation of MF-VAR into LF-VAR models.

For small sample sizes $T_L \in \{50, 100\}$, we draw $J = 1,000$ samples with $N = 100$ replications in bootstrap. For the larger sample size $T_L = 500$, we draw $J = 100,000$ samples without bootstrap since size distortions do not occur.

We expect the following two results based on Theorem 4.3 and Section 5. First, the MF high-to-low causality test should have the same or lower power than the LF high-to-low causality test does since they contain the same amount of causal information and the former entails more parameters. Second, the MF low-to-high causality test should have higher power than the LF low-to-high causality test does since the former contains more causal information than the latter.

6.1.2 Simulation Results

In Table 2 we report rejection frequencies for the simulation design described in the previous subsection. Notice that, in case (a), size is computed with respect to low-to-high causality while power is computed with respect to high-to-low causality. In case (b), size is computed with respect to high-to-low causality, while power is computed with respect to low-to-high causality. Values in parentheses are the benchmark rejection frequencies based on the analytical covariance matrix, and values not in parentheses concern the HAC case.

Empirical size varies within $[0.035, 0.069]$, so there are no serious size distortions in any case. Focusing on power, the results are consistent with the two conjectures above. First, the gap between rejection rates for MF and LF causality tests for $H_{H \nrightarrow L}$ is not large (see case (a) in Table 2). For example, when $(m, T_L) = (2, 50)$ and the HAC covariance matrix is used, power for the MF high-to-low causality test is 0.125 while power for the LF high-to-low causality test is 0.177. Second, the MF low-to-high causality test has clearly higher power than the LF counterpart (see case (b)). This difference is most prominent for the largest m and T_L, where the rejection frequencies in the HAC case are 0.997 and 0.556 for the MF- and LF-VAR models, respectively. These results indicate that the MF causality test never performs worse than the LF causality test in terms of size and power when sample size is fairly large, and the former is much more powerful than the latter when low-to-high causality is of interest.

6.2 Trivariate Case

We now focus on a trivariate MF-VAR model with multiple prediction horizons in order to see if the mixed frequency causality test can capture causality chains properly. Since there is no clear theory on how causality is linked between MF- and LF-VAR processes in the presence of causality chains, we do not consider LF-VAR models, and therefore the aggregation scheme does not matter here.

6.2.1 Simulation Design

Suppose that there are two high frequency variables X and Y and one low frequency variable Z with sampling frequency $m = 3$ so that $K_H = 2$, $K_L = 1$, and $K = 7$. The low frequency sample size is

Table 2: Size and Power for Causality Tests Based on VAR (Stock Sampling)

Rejection frequencies at the 5% level for mixed and low frequency causality tests at horizon $h = 1$. The two cases are (a) $\phi_{HL,1} = 0$ and $\phi_{LH,1} = 0.2$ (unidirectional high-to-low causality) and (b) $\phi_{HL,1} = 0.2$ and $\phi_{LH,1} = 0$ (unidirectional low-to-high causality). In case (a), size is computed with respect to low-to-high causality, while power is computed with respect to high-to-low causality. In case (b), size is computed with respect to high-to-low causality, while power is computed with respect to low-to-high causality. Entries in parentheses are based on the benchmark analytical covariance matrix, and entries not in parentheses are based on the HAC estimator. A parametric bootstrap is employed for $T_L \in \{50, 100\}$ to avoid size distortions. m is the sampling frequency and T_L is the sample size in terms of low frequency.

Sample Size $T_L = 50$

	Case (a)			Case (b)	
	m=2	m=3		m=2	m=3
Size	MF: $0.065(0.051)$ LF: $0.048(0.048)$	MF: $0.046(0.046)$ LF: $0.064(0.055)$		MF: $0.035(0.053)$ LF: $0.037(0.044)$	MF: $0.046(0.049)$ LF: $0.050(0.041)$
Power	MF: $0.125(0.143)$ LF: $0.177(0.203)$	MF: $0.057(0.071)$ LF: $0.072(0.092)$		MF: $0.232(0.261)$ LF: $0.178(0.200)$	MF: $0.186(0.228)$ LF: $0.104(0.109)$

Sample Size $T_L = 100$

	Case (a)			Case (b)	
	m=2	m=3		m=2	m=3
Size	MF: $0.057(0.059)$ LF: $0.047(0.053)$	MF: $0.048(0.041)$ LF: $0.040(0.047)$		MF: $0.054(0.049)$ LF: $0.042(0.045)$	MF: $0.036(0.069)$ LF: $0.046(0.052)$
Power	MF: $0.229(0.253)$ LF: $0.301(0.330)$	MF: $0.098(0.115)$ LF: $0.133(0.140)$		MF: $0.455(0.487)$ LF: $0.317(0.329)$	MF: $0.383(0.427)$ LF: $0.136(0.168)$

Sample Size $T_L = 500$

	Case (a)			Case (b)	
	m=2	m=3		m=2	m=3
Size	MF: $0.059(0.051)$ LF: $0.056(0.052)$	MF: $0.064(0.051)$ LF: $0.055(0.051)$		MF: $0.060(0.052)$ LF: $0.056(0.050)$	MF: $0.066(0.052)$ LF: $0.056(0.053)$
Power	MF: $0.900(0.898)$ LF: $0.943(0.944)$	MF: $0.414(0.390)$ LF: $0.557(0.551)$		MF: $0.998(0.998)$ LF: $0.943(0.944)$	MF: $0.997(0.997)$ LF: $0.556(0.550)$

$T_L = 100$. Define a mixed frequency vector:

$$\boldsymbol{W}(\tau_L) = [X(\tau_L, 1), Y(\tau_L, 1), X(\tau_L, 2), Y(\tau_L, 2), X(\tau_L, 3), Y(\tau_L, 3), Z(\tau_L)]'.$$

Our true DGP is MF-VAR(1):

$$\boldsymbol{W}(\tau_L) = \boldsymbol{A}\boldsymbol{W}(\tau_L - 1) + \boldsymbol{\epsilon}(\tau_L), \quad \boldsymbol{\epsilon}(\tau_L) \overset{i.i.d.}{\sim} (\mathbf{0}_{7\times 1}, \boldsymbol{I}_7)$$

with

$$\boldsymbol{A} = \begin{bmatrix} 0.2 & 0 & -0.3 & 0 & 0.6 & 0 & 0 \\ \boxed{0.3} & 0.3 & \boxed{0.3} & -0.4 & \boxed{0.4} & 0.5 & 0 \\ 0 & 0 & -0.2 & 0 & 0.4 & 0 & 0 \\ \boxed{0} & 0 & \boxed{0.2} & 0.2 & \boxed{0.2} & 0.4 & 0 \\ 0 & 0 & 0 & 0 & 0.3 & 0 & 0 \\ \boxed{0} & 0 & \boxed{0} & 0 & \boxed{0.3} & 0.3 & 0 \\ \underline{0} & \textcircled{0.3} & \underline{0} & \textcircled{0.3} & \underline{0} & \textcircled{0.4} & 0.6 \end{bmatrix}, \tag{6.1}$$

where the nine elements in rectangles represent the impact of X on Y, the three underlined elements represent the impact of X on Z, and the three boxed elements represent the impact of Y on Z. All other non-zero elements are autoregressive coefficients, so not directly relevant for causal patterns. Equation (6.1) thus implies that there are only two channels of causality at $h = 1$: $X \to_1 Y \,|\, \mathcal{I}$ and $Y \to_1 Z \,|\, \mathcal{I}$. In particular, note that X does not cause Z at $h = 1$. For $h \geq 2$, we have three channels of causality because of a causal chain from X to Z via Y: $X \to_h Y \,|\, \mathcal{I}$, $Y \to_h Z \,|\, \mathcal{I}$, and $X \to_h Z \,|\, \mathcal{I}$. This point is verified by observing $\boldsymbol{A}^2$ and $\boldsymbol{A}^3$:

$$\boldsymbol{A}^2 = \begin{bmatrix} 0.04 & 0 & 0 & 0 & 0.18 & 0 & 0 \\ \boxed{0.15} & 0.09 & \boxed{\text{-}0.14} & -0.04 & \boxed{0.61} & 0.14 & 0 \\ 0 & 0 & 0.04 & 0 & 0.04 & 0 & 0 \\ \boxed{0} & 0 & \boxed{\text{-}0.08} & 0.04 & \boxed{0.22} & 0.04 & 0 \\ 0 & 0 & 0 & 0 & 0.09 & 0 & 0 \\ \boxed{0} & 0 & \boxed{0} & 0 & \boxed{0.18} & 0.09 & 0 \\ \underline{0.09} & \textcircled{0.27} & \underline{0.15} & \textcircled{0} & \underline{0.30} & \textcircled{0.63} & 0.36 \end{bmatrix} \tag{6.2}$$

and

$$\boldsymbol{A}^3 = \begin{bmatrix} 0.01 & 0 & -0.01 & 0 & 0.08 & 0 & 0 \\ \boxed{0.06} & 0.30 & \boxed{0.00} & -0.03 & \boxed{0.29} & 0.07 & 0 \\ 0 & 0 & -0.01 & 0 & 0.03 & 0 & 0 \\ \boxed{0} & 0 & \boxed{0.02} & -0.01 & \boxed{0.05} & 0.03 & 0 \\ 0 & 0 & 0 & 0 & 0.03 & 0 & 0 \\ \boxed{0} & 0 & \boxed{0} & 0 & \boxed{0.08} & 0.03 & 0 \\ \underline{0.10} & \textcircled{0.19} & \underline{0.02} & \textcircled{\text{-}0.00} & \underline{0.50} & \textcircled{0.47} & 0.22 \end{bmatrix}. \tag{6.3}$$

We fit (p, h)-autoregression with $p = 1$ and $h \in \{1, 2, 3\}$ to implement the mixed frequency causality

test from an individual variable to another. We are particularly interested in whether we can find *non-causality* from X to Z at $h = 1$ and *causality* from X to Z at $h = 2, 3$. We draw $J = 1,000$ samples and $N = 100$ parametric bootstrap replications to avoid size distortions. The HAC covariance estimator with Newey and West's (1994) automatic bandwidth selection is used as in the bivariate simulation.

6.2.2 Simulation Results

Table 3 reports the rejection frequencies (empirical size is accurate in view of the parametric bootstrap).

Empirical power for the test of $X \nrightarrow_h Y$ is 0.997, 0.815, and 0.161 for horizons 1, 2, and 3, respectively. Diminishing power is reasonable given the diminishing impact of X on Y; see the elements in rectangles in (6.1), (6.2), and (6.3).

Power for the test of $Y \nrightarrow_h Z$ vanishes more slowly as h increases: 1.000, 0.996, and 0.700 for horizons 1, 2, and 3, respectively. In fact the boxed elements of $\boldsymbol{A}^2$ and $\boldsymbol{A}^3$ contain relatively large loadings 0.63 and 0.47, respectively. The intuitive reason for this slower decay is that Y has a more persistent impact on Z than X does on Y; see the upper triangular structure of the rectangles in (6.1).

Finally, the rejection frequency for $X \nrightarrow_h Z$ is 0.047, 0.555, and 0.631 for horizons 1, 2, and 3, respectively. At horizon 1 we get the desired result of non-causality from X to Z, while we have relatively high power for $h = 2, 3$ due to the indirect impact of X on Z via Y (see the underlined elements in (6.1)-(6.3)). Thus, our mixed frequency causality test performs well even in the presence of a causality chain.

Table 3: Rejection Frequency in Trivariate Simulation

This table lists rejection frequencies based on (p, h)-autoregression with $p = 1$ and $h \in \{1, 2, 3\}$. The upper right triangular matrices have empirical size for $Y \nrightarrow_h X$, $Z \nrightarrow_h X$, and $Z \nrightarrow_h Y$. Each test deals with the null hypothesis of non-causality from an individual variable to another at horizon h. We draw $J = 1,000$ samples and $N = 100$ bootstrap replications. The HAC covariance estimator with Newey and West's (1994) automatic bandwidth selection is used. The nominal size is 5%.

Null Hypothesis			$h=1$			$h=2$			$h=3$		
$-$	$Y \nrightarrow_h X$	$Z \nrightarrow_h X$	$-$	0.050	0.052	$-$	0.055	0.047	$-$	0.052	0.062
$X \nrightarrow_h Y$	$-$	$Z \nrightarrow_h Y$	0.997	$-$	0.052	0.815	$-$	0.047	0.161	$-$	0.062
$X \nrightarrow_h Z$	$Y \nrightarrow_h Z$	$-$	0.047	1.000	$-$	0.555	0.996	$-$	0.631	0.700	$-$

7 Concluding Remarks

Time series processes are often sampled at different frequencies and are typically aggregated to the common lowest frequency to test for Granger causality. This paper compares testing for Granger causality with all series aggregated to the common lowest frequency, and testing for Granger causality taking advantage of all the series sampled at whatever frequency they are available. We rely on mixed frequency vector autoregressive models to implement the new class of Granger causality tests.

We show that mixed frequency causality tests better recover causality patterns in an underlying high frequency process compared to the traditional low frequency approach. Moreover, we show formally that mixed frequency causality tests have higher asymptotic power against local alternatives and show via simulation that this also holds in finite sample involving realistic data generating processes. The simulations indicate that the mixed frequency VAR approach works well for small differences in sampling frequencies (denoted by m) like month versus quarter. Current work in progress (Ghysels, Hill, and Motegi (2013)) considers MIDAS regression-based causality tests inspired by Sims (1972) in order to handle relatively large m like month versus year.

References

AMEMIYA, T., AND R. Y. WU (1972): "The Effect of Aggregation on Prediction in the Autoregressive Model," *Journal of the American Statistical Association*, 67, 628–632.

ANDERSON, B. D. O., M. DEISTLER, E. FELSENSTEIN, B. FUNOVITS, P. ZADROZNY, M. EICHLER, W. CHEN, AND M. ZAMANI (2012): "Identifiability of Regular and Singular Multivariate Autoregressive Models from Mixed Frequency Data," in *51st Conference on Decision and Control*, pp. 184–189, Maui, HI. IEEE Control Systems Society.

ANDREOU, E., E. GHYSELS, AND A. KOURTELLOS (2010): "Regression Models with Mixed Sampling Frequencies," *Journal of Econometrics*, 158, 246–261.

——— (2011): "Forecasting with Mixed-Frequency Data," in *Oxford Handbook of Economic Forecasting*, ed. by M. Clements, and D. Hendry, pp. 225–245.

ANDREWS, D. W. K. (1991): "Heteroscedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–85.

ARMESTO, M., K. ENGEMANN, AND M. OWYANG (2010): "Forecasting with Mixed Frequencies," *Federal Reserve Bank of St. Louis Review*, 92, 521–536.

BREITUNG, J., AND N. SWANSON (2002): "Temporal Aggregation and Spurious Instantaneous Causality in Multiple Time Series Models," *Journal of Time Series Analysis*, 23, 651–665.

CHAUVET, M., T. GÖTZ, AND A. HECQ (2013): "Realized Volatility and Business Cycle Fluctuations: A Mixed-Frequency VAR Approach," Working paper, University of California Riverside and Maastricht University.

DAVIDSON, J. (1994): *Stochastic Limit Theory*. New York: Oxford University Press.

DE JONG, R., AND J. DAVIDSON (2000): "Consistency of Kernel Estimators of Heteroscedastic and Autocorrelated Covariance Matrices," *Econometrica*, 68, 407–423.

DEDECKER, J., P. DOUKHAN, G. LANG, J. LEON, S. LOUHICHI, AND C. PRIEUR (2007): *Weak Dependence: With Examples and Applications*. New York: Springer Science+Business Media.

DUFOUR, J., D. PELLETIER, AND E. RENAULT (2006): "Short Run and Long Run Causality in Time Series: Inference," *Journal of Econometrics*, 132, 337–362.

DUFOUR, J., AND E. RENAULT (1998): "Short Run and Long Run Causality in Time Series: Theory," *Econometrica*, 66, 1099–1125.

FELSENSTEIN, E., B. FUNOVITS, M. DEISTLER, B. ANDERSON, M. ZAMANI, AND W. CHEN (2013): "Regular and Singular Multivariate Autoregressive Models and Mixed Frequency Data: Identifiability and Consistent Estimation," Working paper, Vienna University of Technology.

FLORENS, J. P., AND M. MOUCHART (1982): "A Note on Noncausality," *Econometrica*, 50, 583–591.

FORONI, C., E. GHYSELS, AND M. MARCELLINO (2013): "Mixed Frequency Approaches for Vector Autoregressions," in *VAR Models in Macroeconomics, Financial Econometrics, and Forecasting - Advances in Econometrics*, ed. by T. Fomby, and L. Killian, vol. 31.

FORONI, C., M. MARCELLINO, AND C. SCHUMACHER (2013): "U-MIDAS: MIDAS Regressions with Unrestricted Lag Polynomials," *Journal of the Royal Statistical Society*, Series A (fortcoming).

FRIEDMAN, M. (1962): "The Interpolation of Time Series by Related Series," *Journal of the American Statistical Association*, 57, 729–757.

GHYSELS, E. (2012): "Macroeconomics and the Reality of Mixed Frequency Data," Working paper, UNC Chapel Hill.

GHYSELS, E., J. B. HILL, AND K. MOTEGI (2013): "Regression-Based Mixed Frequency Granger Causality Tests," Work in progress, UNC Chapel Hill.

GHYSELS, E., P. SANTA-CLARA, AND R. VALKANOV (2004): "The MIDAS Touch: Mixed Data Sampling Regression Models," Working Paper, UCLA and UNC.

——— (2005): "There is a Risk-Return Trade-off After All," *Journal of Financial Economics*, 76, 509–548.

——— (2006): "Predicting volatility: Getting the Most out of Return Data Sampled at Different Frequencies," *Journal of Econometrics*, 131, 59–95.

GRANGER, C. (1969): "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica*, 3, 424–438.

——— (1980): "Testing for Causality: A Personal Viewpoint," *Journal of Economic Dynamics and Control*, 2, 329–352.

——— (1988): "Some Recent Developments in a Concept of Causality," *Journal of Econometrics*, 39, 199–211.

——— (1995): "Causality in the Long Run," *Econometric Theory*, 11, 530–536.

HILL, J. (2007): "Efficient Tests of Long-run Causation in Trivariate VAR Processes with a Rolling Window Study of the Money-Income Relationship," *Journal of Applied Econometrics*, 22, 747–765.

IBRAGIMOV, I. (1975): "A Note on the Central Limit Theorem for Dependent Random Variables," *Theory of Probability and its Applications*, 20, 135–141.

KUZIN, V., M. MARCELLINO, AND C. SCHUMACHER (2011): "MIDAS versus Mixed-Frequency VAR: Nowcasting GDP in the Euro Area," *International Journal of Forecasting*, 27, 529–542.

LEWIS, R., AND G. REINSEL (1985): "Prediction of Multivariate Time Series by Autoregressive Model Fitting," *Journal of Multivariate Analysis*, 16, 393–411.

LÜTKEPOHL, H. (1984): "Linear Transformation of Vector ARMA Processes," *Journal of Econometrics*, 26, 283–293.

——— (1987): *Forecasting Aggregated Vector ARMA Processes*. Berlin: Springer-Verlag.

——— (1993): "Testing for Causation Between Two Variables in Higher Dimensional VAR Models," in *Studies in Applied Econometrics*, ed. by H. Schneeweiss, and K. Zimmerman, p. 75. Springer-Verlag, Heidelberg.

LÜTKEPOHL, H., AND D. POSKITT (1996): "Testing for Causation Using Infinite Order Vector Autoregressive Processes," *Econometric Theory*, 12, 61–87.

MARCELLINO, M. (1999): "Some Consequences of Temporal Aggregation in Empirical Analysis," *Journal of Business and Economic Statistics*, 17, 129–136.

MCCRACKEN, M., M. OWYANG, AND T. SEKHPOSYAN (2013): "Real-Time Forecasting with a Large Bayesian Block Model," Discussion Paper, Federal Reserve Bank of St. Louis and Bank of Canada.

MCCRORIE, J., AND M. CHAMBERS (2006): "Granger Causality and the Sampling of Economic Processes," *Journal of Econometrics*, 132, 311–336.

NEWEY, W., AND K. WEST (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

——— (1994): "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631–653.

PETERSEN, K. (1983): *Ergodic Theory*. Cambridge Univ. Press.

RENAULT, E., K. SEKKAT, AND A. SZAFARZ (1998): "Testing for Spurious Causality in Exchange Rates," *Journal of Empirical Finance*, 5, 47–66.

SAIKKONEN, P., AND H. LÜTKEPOHL (1996): "Infinite-Order Cointegrated Vector Autoregressive Processes: Estimation and Inference," *Econometric Theory*, 12, 814–844.

SALAMALIKI, P., AND I. VENETIS (2013): "Energy Consumption and Real GDP in G-7: Multi-Horizon Causality Testing in the Presence of Capital Stock," *Energy Economics*, forthcoming.

SILVESTRINI, A., AND D. VEREDAS (2008): "Temporal Aggregation of Univariate and Multivariate Time Series Models: A Survey," *Journal of Economic Surveys*, 22, 458–497.

SIMS, C. A. (1972): "Money, Income, and Causality," *American Economic Review*, 62, 540–552.

YAMAMOTO, T., AND E. KUROZUMI (2006): "Tests for Long-Run Granger Non-Causality in Cointegrated Systems," *Journal of Time Series Analysis*, 27, 703–723.

ZELLNER, A., AND C. MONTMARQUETTE (1971): "A Study of Some Aspects of Temoral Aggregation Problems in Econometric Analyses," *Review of Economics and Statistics*, 53, 335–342.

Technical Appendices

A Asymptotic Properties of MF-VAR Parameter Estimators

In this section we derive the asymptotic distribution of the MF-VAR parameter estimators leading to the proofs of Theorems 2.1 and 2.2. We additionally present a simple consistent *almost surely* positive semi-definite estimator of the least squares asymptotic variance that we use in the simulation study.

A.1 Least Squares Estimator and Asymptotic Variance

In this subsection we present the compact model that leads to the least squares estimator $\hat{\boldsymbol{B}}(h)$ of the parameter set $\boldsymbol{B}(h)$ appearing in equation (2.6). We then characterize the matrix components that enter into the least squares asymptotic covariance $\boldsymbol{\Sigma}_p(h) = (\boldsymbol{I}_K \otimes \boldsymbol{\Gamma}_{p,0}^{-1})\boldsymbol{D}_p(h)(\boldsymbol{I}_K \otimes \boldsymbol{\Gamma}_{p,0}^{-1})'$ appearing in the proof of Theorem 2.1 below. We save notation by writing $\boldsymbol{\Sigma}_p$ instead of $\boldsymbol{\Sigma}_p(h)$ throughout the appendix. We then explicitly derive the covariance matrices $\boldsymbol{\Gamma}_{p,0}$ and $\boldsymbol{D}_p(h)$. Finally, we present a simple consistent HAC estimator of $\boldsymbol{\Sigma}_p$ that satisfies the requirements of Theorem 2.2. The proofs of Theorems 2.1 and 2.2 are presented in Appendix A.2 where we explicitly verify the form of $\boldsymbol{\Sigma}_p$.

A.1.1 Least Squares Estimator

We require a more compact notation in order to derive the least squares estimator $\hat{\boldsymbol{B}}(h)$. Define

$$\begin{aligned}
\boldsymbol{W}_h(k) &= [\boldsymbol{X}(h), \boldsymbol{X}(1+h), \ldots, \boldsymbol{X}(T_L - k + h)]' \in \mathbb{R}^{(T_L-k+1)\times K} \\
\boldsymbol{W}(\tau_L, p) &= [\boldsymbol{X}(\tau_L)', \boldsymbol{X}(\tau_L - 1)', \ldots, \boldsymbol{X}(\tau_L - p + 1)']' \in \mathbb{R}^{pK\times 1} \\
\overline{\boldsymbol{W}}_p(h) &= [\boldsymbol{W}(0,p), \boldsymbol{W}(1,p), \ldots, \boldsymbol{W}(T_L - h, p)]' \in \mathbb{R}^{(T_L-h+1)\times pK},
\end{aligned} \tag{A.1}$$

and define the error

$$\boldsymbol{u}^{(h)}(\tau_L) = \sum_{k=0}^{h-1} \boldsymbol{\Psi}_k \boldsymbol{\epsilon}(\tau_L - k) \tag{A.2}$$

stacked as follows:

$$\boldsymbol{U}_l(k) = \left[\boldsymbol{u}^{(h)}(l), \boldsymbol{u}^{(h)}(1+l), \ldots, \boldsymbol{u}^{(h)}(T_L - k + l)\right]' \in \mathbb{R}^{(T_L-k+1)\times K}. \tag{A.3}$$

Then the (p, h)-autoregression appearing in (2.5) has the equivalent representation

$$\boldsymbol{W}_h(h) = \overline{\boldsymbol{W}}_p(h)\boldsymbol{B}(h) + \boldsymbol{U}_h(h). \tag{A.4}$$

The estimator $\hat{\boldsymbol{B}}(h) = [\overline{\boldsymbol{W}}_p(h)'\overline{\boldsymbol{W}}_p(h)]^{-1}\overline{\boldsymbol{W}}_p(h)'\boldsymbol{W}_h(h)$ then follows.

A.1.2 Asymptotic Variance Components: Covariance Matrices

We now derive the components $\boldsymbol{\Gamma}_{p,0}$ and $\boldsymbol{D}_p(h)$ of the asymptotic variance $\boldsymbol{\Sigma}_p$. First, let $\boldsymbol{\Gamma}_{p,0}$ denote the variance matrices for $\boldsymbol{W}(\tau_L, p)$ in (A.1):

$$\boldsymbol{\Gamma}_{p,0} \equiv E\left[\boldsymbol{W}(\tau_L, p)\boldsymbol{W}(\tau_L, p)'\right].$$

By Assumptions 2.1 through 2.3 it is easily verified that $\mathbf{\Gamma}_{p,0}$ is positive definite. Second, by a standard first order expansion we require the long-run variance of a vectorized $\boldsymbol{W}(\tau_L,p)\boldsymbol{u}(\tau_L+h)'$, denoted

$$\boldsymbol{Y}(\tau_L+h,p) \equiv \text{vec}\left[\boldsymbol{W}(\tau_L,p)\boldsymbol{u}^{(h)}(\tau_L+h)'\right] = \left(\boldsymbol{I}_K \otimes \boldsymbol{W}(\tau_L,p)\right)\boldsymbol{u}^{(h)}(\tau_L+h) \in \mathbb{R}^{pK^2\times 1}. \tag{A.5}$$

By construction $\{\boldsymbol{Y}(\tau_L+h,p)\}_{\tau_L}$ is a zero mean covariance stationary process with auto-covariances

$$\mathbf{\Delta}_{p,s}(h) \equiv E\left[\boldsymbol{Y}(\tau_L+h+s,p)\boldsymbol{Y}(\tau_L+h,p)'\right] \text{ where } \mathbf{\Delta}_{p,s}(h) = 0 \; \forall s \geq h.$$

Of particular note, although in general $\boldsymbol{Y}(\tau_L+h,p)$ is not finite dependent, it has zero auto-covariances at lag $s \geq h$. Analytical characterizations of $\mathbf{\Gamma}_{p,0}$ and $\mathbf{\Delta}_{p,s}(h)$ are presented below. The partial sum variance of $\boldsymbol{Y}(\tau_L+h,p)$ is therefore:

$$\begin{aligned}
\boldsymbol{D}_{p,T_L^*}(h) &\equiv \text{Var}\left[\frac{1}{\sqrt{T_L^*}}\sum_{\tau_L=0}^{T_L^*-1}\boldsymbol{Y}(\tau_L+h,p)\right] \\
&= \mathbf{\Delta}_{p,0}(h) + \sum_{s=1}^{h-1}\left[1-\frac{s}{T_L^*}\right]\times\left[\mathbf{\Delta}_{p,s}(h)+\mathbf{\Delta}_{p,s}(h)'\right] \\
&= \mathbf{\Delta}_{p,0}(1) \text{ if } h=1,
\end{aligned} \tag{A.6}$$

where $T_L^* = T_L - h + 1$. We define $\boldsymbol{D}_p(h)$ as the long-run variance of $\boldsymbol{Y}(\tau_L+h,p)$:

$$\begin{aligned}
\boldsymbol{D}_p(h) &\equiv \lim_{T_L^*\to\infty}\boldsymbol{D}_{p,T_L^*}(h) = \mathbf{\Delta}_{p,0}(h) + \sum_{s=1}^{h-1}\left[\mathbf{\Delta}_{p,s}(h)+\mathbf{\Delta}_{p,s}(h)'\right] \\
&= \mathbf{\Delta}_{p,0}(1) \text{ if } h=1.
\end{aligned} \tag{A.7}$$

Observe that $\boldsymbol{D}_p(h)$ is well defined for any $h \geq 1$ in view of the stationary geometric strong mixing property of $\boldsymbol{W}(\tau_L,p)$, the i.i.d. property for $\boldsymbol{\epsilon}(\tau_L)$, and the $L_{2+\delta}$-boundedness. In particular $\boldsymbol{D}_p(h)$, and $\boldsymbol{D}_{p,T_L^*}(h)$ for T_L^* sufficiently large, are positive definite. Simply note that since $\boldsymbol{X}(\tau_L)$ is a stationary VAR with an i.i.d. error $\boldsymbol{\epsilon}(\tau_L)$ that has a positive definite variance, it follows $\boldsymbol{X}(\tau_L)$ and therefore $\boldsymbol{Y}(\tau_L+h,p)$ have a continuous bounded spectral density that is positive definite at frequency zero. Therefore $\boldsymbol{a}'\boldsymbol{D}_p(h)\boldsymbol{a} > 0$ for all conformable $\boldsymbol{a} \neq 0$ (see Theorem 2.2 in Ibragimov (1975)). Hence $\boldsymbol{D}_p(h)$ is positive definite, and since $\boldsymbol{D}_{p,T_L^*}(h) = \boldsymbol{D}_p(h) + o(1)$ it follows for $N \in \mathbb{N}$ sufficiently large and all $T_L^* \geq N$ that $\boldsymbol{D}_{p,T_L^*}(h)$ is positive definite.

We now explicitly characterize the covariance matrices $\mathbf{\Gamma}_{p,0} \equiv E[\boldsymbol{W}(\tau_L,p)\boldsymbol{W}(\tau_L,p)']$ and $\mathbf{\Delta}_{p,s}(h) \equiv E[\boldsymbol{Y}(\tau_L + h + s,p)\boldsymbol{Y}(\tau_L+h,p)']$. Denote the auto-covariances of $\boldsymbol{X}(\tau_L)$ as

$$\mathbf{\Upsilon}_s = [\upsilon_{ij,s}]_{i,j=1}^K \equiv E\left[\boldsymbol{X}(\tau_L+s)\boldsymbol{X}(\tau_L)'\right] = \begin{cases} \sum_{k=0}^{\infty}\mathbf{\Psi}_{s+k}\mathbf{\Omega}\mathbf{\Psi}_k' & \text{if } s \geq 0 \\ \mathbf{\Upsilon}_{-s}' & \text{if } s < 0, \end{cases} \tag{A.8}$$

where $\mathbf{\Psi}_k$ is defined by the moving average representation (2.4). In view of $|\mathbf{\Psi}_k| = O(\rho^h)$ for $\rho \in (0,1)$ it follows

$\sum_{s=-\infty}^{\infty} |\upsilon_{ij,s}| < \infty$ for any i, j. The process $\{\boldsymbol{W}(\tau_L, p)\}_{\tau_L}$ defined by (A.1) therefore has auto-covariances

$$\boldsymbol{\Gamma}_{p,s} \equiv E\left[\boldsymbol{W}(\tau_L + s, p)\boldsymbol{W}(\tau_L, p)'\right] = \begin{bmatrix} \boldsymbol{\Upsilon}_s & \boldsymbol{\Upsilon}_{s+1} & \cdots & \boldsymbol{\Upsilon}_{s+p-1} \\ \boldsymbol{\Upsilon}_{s-1} & \boldsymbol{\Upsilon}_s & \cdots & \boldsymbol{\Upsilon}_{s+p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Upsilon}_{s-p+1} & \boldsymbol{\Upsilon}_{s-p+2} & \cdots & \boldsymbol{\Upsilon}_s \end{bmatrix}. \tag{A.9}$$

Further, $\boldsymbol{u}^{(h)}(\tau_L)$ has auto-covariances

$$\boldsymbol{Q}_s(h) \equiv E\left[\boldsymbol{u}^{(h)}(\tau_L + s)\boldsymbol{u}^{(h)}(\tau_L)'\right] = \begin{cases} \sum_{k=0}^{h-s-1} \boldsymbol{\Psi}_{s+k}\boldsymbol{\Omega}\boldsymbol{\Psi}_k' & \text{if } 0 \leq s < h \\ \boldsymbol{Q}_{-s}(h)' & \text{if } -h < s < 0 \\ \boldsymbol{0}_{K\times K} & \text{if } |s| \geq h. \end{cases} \tag{A.10}$$

Using (A.10) and $\boldsymbol{Y}(\tau_L + h, p) \equiv (\boldsymbol{I}_K \otimes \boldsymbol{W}(\tau_L, p))\boldsymbol{u}^{(h)}(\tau_L + h)$, the auto-covariances of $\boldsymbol{Y}(\tau_L + h, p)$ are

$$\boldsymbol{\Delta}_{p,s}(h) \equiv E[\boldsymbol{Y}(\tau_L + h + s, p)\boldsymbol{Y}(\tau_L + h, p)'] = \begin{cases} \boldsymbol{Q}_0(h) \otimes \boldsymbol{\Gamma}_{p,0} & \text{if } s = 0 \\ \boldsymbol{\Delta}_{p,-s}(h)' & \text{if } -h < s < 0 \\ \boldsymbol{0}_{pK^2\times pK^2} & \text{if } |s| \geq h. \end{cases} \tag{A.11}$$

Notice $\boldsymbol{Y}(\tau_L + h, p)$ is serially uncorrelated at lag $|s| \geq h$, although in general we cannot say $\boldsymbol{Y}(\tau_L + h, p)$ is $h - 1$ dependent. Evidently a convenient expression for $\boldsymbol{\Delta}_{p,s}(h)$ does not exist when $s \in \{1, \ldots, h-1\}$.

We now prove $\boldsymbol{\Delta}_{p,s}(h) = \boldsymbol{0}_{pK^2\times pK^2}$ for $|s| \geq h$. Assume without loss of generality that $s \geq h$. Equation (A.5) and the definition of $\boldsymbol{\Delta}_{p,s}(h)$ imply that

$$\boldsymbol{\Delta}_{p,s}(h) = E\left[\left(\boldsymbol{I}_K \otimes \boldsymbol{W}(\tau_L + s, p)\right)\boldsymbol{u}^{(h)}(\tau_L + s + h)\boldsymbol{u}^{(h)}(\tau_L + h)'\left(\boldsymbol{I}_K \otimes \boldsymbol{W}(\tau_L, p)'\right)\right]. \tag{A.12}$$

Let $I(\tau_L + s) = \sigma\{\boldsymbol{\epsilon}(\tau)|\tau \leq \tau_L + s\}$. Note that $\boldsymbol{W}(\tau_L, p)$, $\boldsymbol{W}(\tau_L + s, p)$, and $\boldsymbol{u}^{(h)}(\tau_L + h)$ are all known at period $\tau_L + s$, while $\boldsymbol{u}^{(h)}(\tau_L + s + h)$ depends only on $\{\boldsymbol{\epsilon}(\tau_L + s + 1), \ldots, \boldsymbol{\epsilon}(\tau_L + s + h)\}$ and therefore $E[\boldsymbol{u}^{(h)}(\tau_L + s + h)|I(\tau_L + s)] = E[\boldsymbol{u}^{(h)}(\tau_L + s + h)] = \boldsymbol{0}_{K\times 1}$ by the i.i.d. assumption. We can thus get the desired result by applying the law of iterated expectations to (A.12). Similarly, $\boldsymbol{\Delta}_{p,0}(h) = \boldsymbol{Q}_0(h) \otimes \boldsymbol{\Gamma}_{p,0}$ can be shown by applying the law of iterated expectations given $I(\tau_L)$ to (A.12).

Example ($h = 1$): It is useful to derive the least squares asymptotic variance $\boldsymbol{\Sigma}_p = (\boldsymbol{I}_K \otimes \boldsymbol{\Gamma}_{p,0}^{-1})\boldsymbol{D}_p(h)(\boldsymbol{I}_K \otimes \boldsymbol{\Gamma}_{p,0}^{-1})'$ for the case $h = 1$. Use (A.8) and (A.9) to deduce $\boldsymbol{\Gamma}_{p,0} = \boldsymbol{\Upsilon}_0 = \sum_{k=0}^{\infty} \boldsymbol{\Psi}_k\boldsymbol{\Omega}\boldsymbol{\Psi}_k'$. Next, use A.7 and A.11 to deduce $\boldsymbol{D}_p(1) = \boldsymbol{\Delta}_{p,0}(1) = \boldsymbol{Q}_0(1) \otimes \boldsymbol{\Gamma}_{p,0}$, hence by (A.9) and (A.10) it follows $\boldsymbol{D}_p(1) = \boldsymbol{\Omega} \otimes \boldsymbol{\Gamma}_{p,0} = \boldsymbol{\Omega} \otimes \sum_{k=0}^{\infty} \boldsymbol{\Psi}_k\boldsymbol{\Omega}\boldsymbol{\Psi}_k'$. Kronecker product propertiesF therefore imply $\boldsymbol{\Sigma}_p$ is identically $\boldsymbol{\Omega} \otimes \boldsymbol{\Gamma}_{p,0}^{-1} = \boldsymbol{\Omega} \otimes \boldsymbol{\Upsilon}_0^{-1} = \boldsymbol{\Omega} \otimes (\sum_{k=0}^{\infty} \boldsymbol{\Psi}_k\boldsymbol{\Omega}\boldsymbol{\Psi}_k')^{-1}$.

A.1.3 Consistent and Almost Surely Positive Semi-Definite HAC Estimator

We need only estimate the components of $\boldsymbol{\Sigma}_p = (\boldsymbol{I}_K \otimes \boldsymbol{\Gamma}_{p,0}^{-1})\boldsymbol{D}_p(h)(\boldsymbol{I}_K \otimes \boldsymbol{\Gamma}_{p,0}^{-1})'$. A natural estimator of $\boldsymbol{\Gamma}_{p,0}$ is the sample conjugate:

$$\hat{\boldsymbol{\Gamma}}_{p,0} = \frac{1}{T_L^*} \sum_{\tau_L=0}^{T_L^*-1} \boldsymbol{W}(\tau_L, p)\boldsymbol{W}(\tau_L, p)'.$$

Under Assumptions 2.1-2.3 $\hat{\boldsymbol{\Gamma}}_{p,0}$ is *almost surely* positive definite.

Turning to the long-run variance $\boldsymbol{D}_p(h)$, denote the least squares residual $\hat{\boldsymbol{U}}_h(h) \equiv \boldsymbol{W}_h(h) - \overline{\boldsymbol{W}}_p(h)\hat{\boldsymbol{B}}(h)$ for model (A.4) and the resulting residual $\hat{\boldsymbol{u}}^{(h)}(\tau_L) \equiv \boldsymbol{X}(\tau_L) - \sum_{k=1}^{p} \hat{\boldsymbol{A}}_k^{(h)} \boldsymbol{X}(\tau_L - h + 1 - k)$ computed from (A.3). Now compute the sample version of $\boldsymbol{Y}(\tau_L + h, p)$ defined in (A.5),

$$\hat{\boldsymbol{Y}}(\tau_L + h, p) = \text{vec}\left[\boldsymbol{W}(\tau_L, p)\hat{\boldsymbol{u}}^{(h)}(\tau_L + h)'\right],$$

and compute

$$\hat{\boldsymbol{\Delta}}_{p,s}(h) = \frac{1}{T_L^*} \sum_{\tau_L=s}^{T_L^*-1} \hat{\boldsymbol{Y}}(\tau_L + h, p)\hat{\boldsymbol{Y}}(\tau_L + h - s, p)'.$$

If $h = 1$ then from (A.6) the estimator of $\boldsymbol{D}_p(h)$ need only be $\hat{\boldsymbol{D}}_{p,T_L^*}(1) = \hat{\boldsymbol{\Delta}}_{p,0}(1)$. Otherwise, a naïve estimator of $\boldsymbol{D}_p(h)$ simply substitutes $\hat{\boldsymbol{\Delta}}_{p,s}(h)$ for $\boldsymbol{\Delta}_{p,s}(h)$ in the right-hand side of (A.6), but it is well-known that such an estimator may not be positive semi-definite unless $h = 1$.

We therefore exploit Newey and West (1987)'s Bartlett kernel-based HAC estimator which ensures *almost sure* positive semi-definiteness for any $T_L^* \geq 1$ (see Newey and West (1987) and Andrews (1991)):[14]

$$\hat{\boldsymbol{D}}_{p,T_L^*}(h) = \hat{\boldsymbol{\Delta}}_{p,0}(h) + \sum_{s=1}^{n_{T_L^*}-1} \left(1 - \frac{s}{n_{T_L^*}}\right) (\hat{\boldsymbol{\Delta}}_{p,s}(h) + \hat{\boldsymbol{\Delta}}_{p,s}(h)') \tag{A.13}$$

with bandwidth $n_{T_L^*}$: $h \leq n_{T_L^*} \leq T_L^*$, $n_{T_L^*} \to \infty$ and $n_{T_L^*} = o(T_L^*)$. Intuitively since $\boldsymbol{Y}(\tau_L, p)$ is serially uncorrelated for all lags above $h - 1$, and $\hat{\boldsymbol{\Delta}}_{p,s}(h) = 1/T_L^* \sum_{\tau_L=s}^{T_L^*-1} \boldsymbol{Y}(\tau_L + h, p)\boldsymbol{Y}(\tau_L + h - s, p)' + o_p(1)$ is easily verified, we only need $h - 1$ lags, that is $\hat{\boldsymbol{\Delta}}_{p,0}(h) + \sum_{s=1}^{h-1}(1 - s/n_{T_L^*})(\hat{\boldsymbol{\Delta}}_{p,s}(h) + \hat{\boldsymbol{\Delta}}_{p,s}(h)')$ is a valid estimator in place of (A.13). But this estimator also need not be positive semi-definite in small samples.

Our proposed estimator of $\boldsymbol{\Sigma}_p$ is therefore

$$\hat{\boldsymbol{\Sigma}}_p = \left(\boldsymbol{I}_K \otimes \hat{\boldsymbol{\Gamma}}_{p,0}^{-1}\right) \times \hat{\boldsymbol{D}}_{p,T_L^*}(h) \times \left(\boldsymbol{I}_K \otimes \hat{\boldsymbol{\Gamma}}_{p,0}^{-1}\right). \tag{A.14}$$

In view of *almost sure* positive definiteness of $\hat{\boldsymbol{\Gamma}}_{p,0}$ and positive semi-definiteness of $\hat{\boldsymbol{D}}_{p,T_L^*}(h)$ it follows $\hat{\boldsymbol{\Sigma}}_p$ is *almost surely* positive semi-definite. Consistency can be shown given a stronger moment condition for the error term.

Lemma A.1. Let Assumptions 2.1 through 2.3 hold and assume $||\boldsymbol{\epsilon}(\tau_L)||^{4+\delta} < \infty$ for some $\delta > 0$. Then $\hat{\boldsymbol{\Sigma}}_p$ is *almost surely* positive semi-definite for any $T_L^* \geq 1$, and $\hat{\boldsymbol{\Sigma}}_p \xrightarrow{p} \boldsymbol{\Sigma}_p$ where $\boldsymbol{\Sigma}_p$ is positive definite.

Proof. *Almost sure* positive semi-definiteness of $\hat{\boldsymbol{\Sigma}}_p$ follows from *almost sure* positive definiteness of $\hat{\boldsymbol{\Gamma}}_{p,0}$ under Assumptions 2.1 - 2.3, and *almost sure* positive semi-definiteness of $\hat{\boldsymbol{D}}_{p,T_L^*}(h)$ by Theorem 1 in Newey and West (1987). In view of $\boldsymbol{\Sigma}_p = (\boldsymbol{I}_K \otimes \boldsymbol{\Gamma}_{p,0}^{-1}) \times \boldsymbol{D}_p(h) \times (\boldsymbol{I}_K \otimes \boldsymbol{\Gamma}_{p,0}^{-1})$ and the fact that by stationarity, ergodicity and $L_{4+\delta}$-boundedness a consistent estimator for $\boldsymbol{\Gamma}_{p,0}$ is immediately available from its sample counterpart $\hat{\boldsymbol{\Gamma}}_{p,0}$, it is sufficient to show the consistency of $\hat{\boldsymbol{D}}_{p,T_L^*}(h)$. This can be done by verifying Assumptions 1-4 in de Jong and Davidson (2000) due to their Theorem 2.2.

First, the Bartlett kernel satisfies their Assumption 1. Second, their Assumptions 2 and 3 hold since $n_{T_L^*} \to \infty$ as $T_L^* \to \infty$, $n_{T_L^*} = o(T_L^*)$, and by independence and $L_{4+\delta}$-boundedness of $\boldsymbol{\epsilon}(\tau_L)$, and the fact that $\boldsymbol{W}(\tau_L, p)$ is stationary geometric strong mixing, it follows by measurability that $\{1/\sqrt{T_L^*}\boldsymbol{Y}(\tau_L + h, p)\}_{\tau_L}$ is an $L_{4+\delta}$-bounded

[14] There is a large choice of valid kernels, including Parzen and Tukey-Hanning. See de Jong and Davidson (2000).

geometrically strong mixing process.[15]

Finally, in order to verify their Assumption 4, define the regression error function $\boldsymbol{u}^{(h)}(\tau_L, \tilde{\boldsymbol{B}}) \equiv \boldsymbol{X}(\tau_L) - \sum_{k=1}^{p} \tilde{\boldsymbol{A}}_k \boldsymbol{X}(\tau_L - h + 1 - k)$ for any conformable $\tilde{\boldsymbol{A}}_k$ where $\tilde{\boldsymbol{B}} \equiv [\tilde{\boldsymbol{A}}_1, \ldots, \tilde{\boldsymbol{A}}_p]'$, and $\boldsymbol{Y}(\tau_L + h, p, \tilde{\boldsymbol{B}}) = (\boldsymbol{I}_K \otimes \boldsymbol{W}(\tau_L, p))\boldsymbol{u}^{(h)}(\tau_L + h, \tilde{\boldsymbol{B}})$. Now define $\boldsymbol{Z}(\tau_L + h, p, \tilde{\boldsymbol{B}}) \equiv \boldsymbol{Y}(\tau_L + h, p, \tilde{\boldsymbol{B}})/\sqrt{T_L^*}$ and note $\hat{\boldsymbol{Y}}(\tau_L + h, p) = \boldsymbol{Y}(\tau_L + h, p, \hat{\boldsymbol{B}}(h))$. In order to match (de Jong and Davidson 2000)'s standardization, we work with $\boldsymbol{Z}(\tau_L + h, p, \tilde{\boldsymbol{B}})$. Assumption 4 consists of three parts, (a)-(c), with a scale factor κ_n that is simply $\boldsymbol{I}_{pK^2}$ in our case. Part (a) applies since $\hat{\boldsymbol{B}}(h)$ is $\sqrt{T_L^*}$-convergent. Next, (b) applies since under our assumptions and by model linearity it follows $1/\sqrt{T_L^*}\sum_{\tau_L=0}^{T_L^*-1} E[(\partial/\partial\tilde{\boldsymbol{B}})\boldsymbol{Z}(\tau_L + h, p, \tilde{\boldsymbol{B}})]$ is trivially continuous at $\boldsymbol{B}(h)$ uniformly in T_L^*. Finally, (c) involves a uniform LLN for $(\partial/\partial\tilde{\boldsymbol{B}})\boldsymbol{Z}(\tau_L + h, p, \tilde{\boldsymbol{B}})$. The latter is not a function of $\tilde{\boldsymbol{B}}$ in view of linearity, hence a uniform LLN reduces to a pointwise LLN which hold instantly under Assumptions 2.1-2.3. $\mathcal{QED}$.

A.2 Proof of Theorems 2.1 and 2.2

Recall $\boldsymbol{D}_{p,T_L^*}(h) \equiv \text{Var}[1/\sqrt{T_L^*}\sum_{\tau_L=0}^{T_L^*-1} \boldsymbol{Y}(\tau_L + h, p)]$ in (A.6) and $\boldsymbol{D}_p(h) \equiv \lim_{T_L^*\to\infty} \boldsymbol{D}_{p,T_L^*}(h)$. To prove Theorem 2.1 we require the following central limit theorem.

Lemma A.2. $1/\sqrt{T_L^*}\sum_{\tau_L=0}^{T_L^*-1} \boldsymbol{Y}(\tau_L + h, p) \xrightarrow{d} N(\boldsymbol{0}_{pK^2\times 1}, \boldsymbol{D}_p\,(h))$ where $\boldsymbol{D}_p(h)$ is positive definite.

Proof. By the Cramér-Wold theorem it is necessary and sufficient to show $\frac{1}{\sqrt{T_L^*}}\sum_{\tau_L=0}^{T_L^*-1} \boldsymbol{\alpha}'\boldsymbol{Y}(\tau_L + h, p) \xrightarrow{d} N(0, \boldsymbol{\alpha}'\boldsymbol{D}_p(h)\boldsymbol{\alpha})$ for any $pK^2 \times 1$ non-zero vector $\boldsymbol{\alpha}$. Observe that $\{\boldsymbol{\alpha}'\boldsymbol{Y}(\tau_L + h, p)\}_{\tau_L}$ is a zero mean geometrically strongly mixing process. This follows from $\boldsymbol{Y}(\tau_L + h, p) \equiv \text{vec}[\boldsymbol{W}(\tau_L, p)\boldsymbol{u}^{(h)}(\tau_L + h)']$, $\boldsymbol{W}(\tau_L, p)$ is a finite dimensional vector of geometrically strong mixing $\boldsymbol{X}(\tau_L)$, and $\boldsymbol{u}^{(h)}(\tau_L) = \sum_{k=0}^{h-1} \boldsymbol{\Psi}_k\boldsymbol{\epsilon}(\tau_L - k)$ is a finite lag of i.i.d. $\boldsymbol{\epsilon}(\tau_L)$. Moreover $||\boldsymbol{\alpha}'\boldsymbol{Y}(\tau_L + h, p)||_{2+\delta} < \infty$ by the moving average representation (2.4), the independence of $\boldsymbol{\epsilon}(\tau_L)$, and $||\boldsymbol{\epsilon}(\tau_L)||_{2+\delta} < \infty$ under Assumption 2.3.

Further, since $\boldsymbol{Y}(\tau_L + h, p)$ is auto-correlated only up to lag $h-1$ it follows $\lim_{T_L^*\to\infty} \boldsymbol{D}_{p,T_L^*}(h) = \boldsymbol{\Delta}_{p,0}(h) + \sum_{s=1}^{h-1}[\boldsymbol{\Delta}_{p,s}(h) + \boldsymbol{\Delta}_{p,s}(h)']$. Both $\lim_{T_L^*\to\infty} \boldsymbol{D}_{p,T_L^*}(h)$, and $\boldsymbol{D}_{p,T_L^*}(h)$ for sufficiently large T_L^*, are positive definite by the discussion in Appendix A.1. Therefore $1/\sqrt{T_L^*}\sum_{\tau_L=0}^{T_L^*-1} \boldsymbol{\alpha}'\boldsymbol{Y}(\tau_L + h, p)/(\boldsymbol{\alpha}'\boldsymbol{D}_{p,T_L^*}(h)\boldsymbol{\alpha}) \xrightarrow{d} N(0,1)$ by Theorem 2.2 in Ibragimov (1975). In view of $\boldsymbol{\alpha}'\boldsymbol{D}_{p,T_L^*}(h)\boldsymbol{\alpha} \to \boldsymbol{\alpha}'\boldsymbol{D}_p(h)\boldsymbol{\alpha}$ the claim now follows from Cramér's Theorem. $\mathcal{QED}$.

We now prove Theorems 2.1 and 2.2. By the construction of $\hat{\boldsymbol{B}}(h)$, $\hat{\boldsymbol{\Gamma}}_{p,0}$ and $\boldsymbol{Y}(\tau_L + h, p)$ it follows

$$\begin{aligned}
\sqrt{T_L^*}\text{vec}\left[\hat{\boldsymbol{B}}(h) - \boldsymbol{B}(h)\right] &= \sqrt{T_L^*}\text{vec}\left[\left(\overline{\boldsymbol{W}}_p(h)'\overline{\boldsymbol{W}}_p(h)\right)^{-1}\overline{\boldsymbol{W}}_p(h)'\boldsymbol{U}_h(h)\right] \\
&= \left[\boldsymbol{I}_K \otimes \left(\frac{1}{T_L^*}\overline{\boldsymbol{W}}_p(h)'\overline{\boldsymbol{W}}_p(h)\right)^{-1}\right] \times \text{vec}\left[\frac{1}{\sqrt{T_L^*}}\overline{\boldsymbol{W}}_p(h)'\boldsymbol{U}_h(h)\right] \\
&= \left[\boldsymbol{I}_K \otimes \hat{\boldsymbol{\Gamma}}_{p,0}^{-1}\right] \times \frac{1}{\sqrt{T_L^*}}\sum_{\tau_L=0}^{T_L^*-1} \boldsymbol{Y}(\tau_L + h, p).
\end{aligned}$$

Notice $\hat{\boldsymbol{\Gamma}}_{p,0} = 1/T_L^*\sum_{\tau_L=0}^{T_L^*-1} \boldsymbol{W}(\tau_L, p)\boldsymbol{W}(\tau_L, p)' \xrightarrow{p} E[\boldsymbol{W}(\tau_L, p)\boldsymbol{W}(\tau_L, p)'] = \boldsymbol{\Gamma}_{p,0}$ in view of stationarity, ergodicity and L_2-boundedness of $\boldsymbol{W}(\tau_L, p)$. Further, $\boldsymbol{D}_{p,T_L^*}(h) = \text{Var}[1/\sqrt{T_L^*}\sum_{\tau_L=0}^{T_L^*-1} \boldsymbol{Y}(\tau_L + h, p)] \to \boldsymbol{D}_p(h)$. Now use

$$\boldsymbol{\Sigma}_p \equiv (\boldsymbol{I}_K \otimes \boldsymbol{\Gamma}_{p,0}^{-1}) \times \boldsymbol{D}_p(h) \times (\boldsymbol{I}_K \otimes \boldsymbol{\Gamma}_{p,0}^{-1})',$$

[15] See Chapter 17 in Davidson (1994) for verification that geometric strong mixing satisfies the Near Epoch Dependence property in de Jong and Davidson's (2000) Assumption 2.

combined with Lemma A.2, and Slutsky's and Cramér's Theorems, to deduce $\sqrt{T_L^*}\text{vec}[\hat{\boldsymbol{B}}(h)-\boldsymbol{B}(h)] \xrightarrow{d} N(\boldsymbol{0}_{pK^2\times 1}, \boldsymbol{\Sigma}_p)$. Finally, $\boldsymbol{\Sigma}_p$ is positive definite in view of the positive definiteness of $\boldsymbol{\Gamma}_{p,0}$ and $\boldsymbol{D}_p(h)$ as discussed in Appendix A.1. This proves Theorem 2.1.

The proof of Theorem 2.2 follows instantly from Theorem 2.1, the assumption $\hat{\boldsymbol{\Sigma}}_p \xrightarrow{p} \boldsymbol{\Sigma}_p$, and the mapping theorem.

B Proof of Theorem 4.1

In view of Theorem 1 in Lütkepohl (1984) it suffices to show that $\boldsymbol{X}(\tau_L)$ and $\underline{\boldsymbol{X}}(\tau_L)$ are linear transformations of a VAR process. Define $mK^* \times 1$ vectors:

$$\overline{\boldsymbol{X}}(\tau_L) = [\boldsymbol{X}(\tau_L,1)', \ldots, \boldsymbol{X}(\tau_L,m)']' \quad \text{and} \quad \overline{\boldsymbol{\eta}}(\tau_L) = [\boldsymbol{\eta}(\tau_L,1)', \ldots, \boldsymbol{\eta}(\tau_L,m)']'.$$

We first show that $\{\overline{\boldsymbol{X}}(\tau_L)\}$ follows a VAR(s) process with $s = \lceil p/m \rceil$, the smallest integer not smaller than p/m. We then prove the claim.

The HF-VAR(p) process in (4.1) implies that:

$$\boldsymbol{N}\overline{\boldsymbol{X}}(\tau_L) = \sum_{k=1}^{s} \boldsymbol{M}_k \overline{\boldsymbol{X}}(\tau_L - k) + \overline{\boldsymbol{\eta}}(\tau_L), \tag{B.1}$$

where

$$\boldsymbol{N} = \begin{bmatrix} \boldsymbol{I}_{K^*} & \boldsymbol{0}_{K^*\times K^*} & \ldots & \boldsymbol{0}_{K^*\times K^*} \\ -\boldsymbol{\Phi}_1 & \boldsymbol{I}_{K^*} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \boldsymbol{0}_{K^*\times K^*} \\ -\boldsymbol{\Phi}_{m-1} & \ldots & -\boldsymbol{\Phi}_1 & \boldsymbol{I}_{K^*} \end{bmatrix} \text{ and } \boldsymbol{M}_k = \begin{bmatrix} \boldsymbol{\Phi}_{km} & \boldsymbol{\Phi}_{km-1} & \ldots & \boldsymbol{\Phi}_{(k-1)m+1} \\ \boldsymbol{\Phi}_{km+1} & \boldsymbol{\Phi}_{km} & \ldots & \boldsymbol{\Phi}_{(k-1)m+2} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Phi}_{(k+1)m-1} & \boldsymbol{\Phi}_{(k+1)m-2} & \ldots & \boldsymbol{\Phi}_{km} \end{bmatrix}$$

for $k = 1, \ldots, s$. It is understood that $\boldsymbol{\Phi}_k = \boldsymbol{0}_{K^*\times K^*}$ whenever $k > p$. We have that:

$$\boldsymbol{N}^{-1} = \begin{bmatrix} \boldsymbol{N}_1 & \boldsymbol{0}_{K^*\times K^*} & \ldots & \boldsymbol{0}_{K^*\times K^*} \\ \boldsymbol{N}_2 & \boldsymbol{N}_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \boldsymbol{0}_{K^*\times K^*} \\ \boldsymbol{N}_m & \ldots & \boldsymbol{N}_2 & \boldsymbol{N}_1 \end{bmatrix},$$

where $\boldsymbol{N}_1 = \boldsymbol{I}_{K^*}$ and $\boldsymbol{N}_k = \sum_{l=1}^{k-1} \boldsymbol{\Phi}_{k-l}\boldsymbol{N}_l$ for $k = 2, \ldots, m$. Using this property, (B.1) can be rewritten as follows:

$$\overline{\mathbb{A}}(\mathcal{L}_L)\overline{\boldsymbol{X}}(\tau_L) = \overline{\boldsymbol{\epsilon}}(\tau_L),$$

where $\mathcal{L}_L$ is the low frequency lag operator, $\overline{\mathbb{A}}(\mathcal{L}_L) = \boldsymbol{I}_{mK^*} - \sum_{k=1}^{s} \overline{\boldsymbol{A}}_k \mathcal{L}_L^k$, $\overline{\boldsymbol{A}}_k = \boldsymbol{N}^{-1}\boldsymbol{M}_k$, and $\overline{\boldsymbol{\epsilon}}(\tau_L) = \boldsymbol{N}^{-1}\overline{\boldsymbol{\eta}}(\tau_L)$. Hence, $\{\overline{\boldsymbol{X}}(\tau_L)\}$ follows a VAR(s) process.

Now consider $\boldsymbol{X}(\tau_L)$ and $\underline{\boldsymbol{X}}(\tau_L)$. Recall the generic aggregation schemes (2.1) detailed in Section 2 with selection vector $\boldsymbol{w}$. Define $\boldsymbol{H} = [\boldsymbol{I}_{K_H}, \boldsymbol{0}_{K_H\times K_L}]$, $\boldsymbol{L} = [\boldsymbol{0}_{K_L\times K_H}, \boldsymbol{I}_{K_L}]$, $\boldsymbol{F}_{H\to M} = [\boldsymbol{I}_m \otimes \boldsymbol{H}', \boldsymbol{w} \otimes \boldsymbol{L}']'$, and

$$\boldsymbol{F}_{M\to L} = \begin{bmatrix} \boldsymbol{w}' \otimes \boldsymbol{I}_{K_H} & \boldsymbol{0}_{K_H\times K_L} \\ \boldsymbol{0}_{K_L\times mK_H} & \boldsymbol{I}_{K_L} \end{bmatrix}.$$

Observe that $\boldsymbol{X}(\tau_L)$ and $\underline{\boldsymbol{X}}(\tau_L)$ are finite order linear transformations of $\overline{\boldsymbol{X}}(\tau_L)$: $\boldsymbol{X}(\tau_L) = \boldsymbol{F}_{H\to M}\overline{\boldsymbol{X}}(\tau_L)$ and $\underline{\boldsymbol{X}}(\tau_L) = \boldsymbol{F}_{H\to L}\overline{\boldsymbol{X}}(\tau_L)$, where $\boldsymbol{F}_{H\to L} = \boldsymbol{F}_{M\to L}\boldsymbol{F}_{H\to M} = [\boldsymbol{w}\otimes\boldsymbol{H}', \boldsymbol{w}\otimes\boldsymbol{L}']'$. Moreover, in view of the transformation being a finite order, if $\overline{\boldsymbol{X}}(\tau_L)$ is stationary then so are $\boldsymbol{X}(\tau_L)$ and $\underline{\boldsymbol{X}}(\tau_L)$.

C Proof of Theorem 4.2

We prove only part (ii) since part (i) is similar or even simpler. Recall that the high frequency reference information set at time t is expressed as $\overline{\mathcal{I}}(t)$ and the mapping between single time index t and double time indices (τ_L, k) is that $t = m(\tau_L - 1) + k$. We have that:

$$\begin{aligned} P[\boldsymbol{x}_H(\tau_L+1)\,|\,\mathcal{I}(\tau_L)] &= P\left[P[\boldsymbol{x}_H(\tau_L+1)\,|\,\overline{\mathcal{I}}(m\tau_L)]\,|\,\mathcal{I}(\tau_L)\right] \\ &= P\left[P[\boldsymbol{x}_H(\tau_L+1)\,|\,\overline{\mathcal{I}}_{(L)}(m\tau_L)]\,|\,\mathcal{I}(\tau_L)\right] \\ &= P\left[P[\boldsymbol{x}_H(\tau_L+1)\,|\,\mathcal{I}_{(L)}(\tau_L)]\,|\,\mathcal{I}(\tau_L)\right] \\ &= P[\boldsymbol{x}_H(\tau_L+1)\,|\,\mathcal{I}_{(L)}(\tau_L)]. \end{aligned}$$

The first equality follows from the law of iterated projections for orthogonal projections on a Hilbert space; the second from the linear aggregation scheme and the assumption that $\boldsymbol{x}_L \nrightarrow \boldsymbol{x}_H\,|\,\overline{\mathcal{I}}$; and the third holds because $\overline{\mathcal{I}}_{(L)}(m\tau_L) = \mathcal{I}_{(L)}(\tau_L)$. Hence $\boldsymbol{x}_L \nrightarrow \boldsymbol{x}_H\,|\,\mathcal{I}$ as claimed.

D Proof of Theorem 4.3

We prove claim (i) only since parts (ii)-(iv) are analogous. The following two cases complete part 1:

Case 1 (low $\nrightarrow$ low). Suppose that x_{L,j_1} does not cause x_{L,j_2} up to HF horizon m given $\overline{\mathcal{I}}$ (i.e., $x_{L,j_1} \nrightarrow_{(m)} x_{L,j_2}|\,\overline{\mathcal{I}}$). Then, $\boldsymbol{\Phi}^{[k]}_{LL,1}(j_2,j_1) = 0$ for any $k \in \{1,\ldots,m\}$ and hence x_{L,j_1} does not cause x_{L,j_2} at horizon 1 given $\mathcal{I}$ (i.e., $x_{L,j_1} \nrightarrow x_{L,j_2}|\,\mathcal{I}$) in view of (4.4). The converse does not necessarily hold; a simple counter-example is that $K_H = 1$, $K_L = 2$, $m = 2$, $(j_1, j_2) = (1, 2)$, and

$$\boldsymbol{\Phi}_1 = \begin{bmatrix} \phi_{HH} & 0.3 & \phi_{HL} \\ \phi_{LH} & 0.2 & \phi_{LL} \\ -0.1 & 0.1 & 0.1 \end{bmatrix},$$

where ϕ_{HH}, ϕ_{HL}, ϕ_{LH}, and ϕ_{LL} are arbitrary coefficients. It is evident that $\boldsymbol{\Phi}_{LL,1}(2,1) = 0.1$ and $\boldsymbol{\Phi}^{[2]}_{LL,1}(2,1) = 0$. The former denies that $x_{L,j_1} \nrightarrow_{(m)} x_{L,j_2}|\,\overline{\mathcal{I}}$, while the latter implies that $x_{L,j_1} \nrightarrow x_{L,j_2}|\,\mathcal{I}$.

Suppose now that $x_{L,j_1} \nrightarrow x_{L,j_2}|\,\mathcal{I}$. Then, $\boldsymbol{\Phi}^{[m]}_{LL,1}(j_2,j_1) = 0$ and hence $x_{L,j_1} \nrightarrow x_{L,j_2}|\,\underline{\mathcal{I}}$ in view of (4.9). The converse is also true.

Case 2 (high $\nrightarrow$ low). Suppose that $x_{H,i_1} \nrightarrow_{(m)} x_{L,j_1}|\,\overline{\mathcal{I}}$. Then, $\boldsymbol{\Phi}^{[k]}_{LH,1}(j_1,i_1) = 0$ for any $k \in \{1,\ldots,m\}$ and hence $\boldsymbol{x}_{H,i_1} \nrightarrow x_{L,j_1}|\,\mathcal{I}$. The converse does not necessarily hold.

Suppose now that $\boldsymbol{x}_{H,i_1} \nrightarrow x_{L,j_1}|\,\mathcal{I}$. Then, $\boldsymbol{\Phi}^{[m]}_{LH,1}(j_1,i_1) = 0$ and hence $x_{H,i_1} \nrightarrow x_{L,j_1}|\,\underline{\mathcal{I}}$. The converse is also true.