



Testing for Granger causality with mixed frequency data[☆]

Eric Ghysels ^{a,*}, Jonathan B. Hill ^b, Kaiji Motegi ^c

^a Department of Economics and Department of Finance, Kenan-Flagler Business School, University of North Carolina at Chapel Hill, United States

^b Department of Economics, UNC Chapel Hill, United States

^c Faculty of Political Science and Economics, Waseda University, Japan



ARTICLE INFO

Article history:

Received 13 March 2014

Received in revised form

1 July 2015

Accepted 9 July 2015

Available online 29 December 2015

Keywords:

Granger causality test

Local asymptotic power

Mixed data sampling (MIDAS)

Temporal aggregation

Vector autoregression (VAR)

ABSTRACT

We develop Granger causality tests that apply directly to 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 new causality tests have higher local asymptotic power as well as more power in finite samples compared to conventional tests. In an empirical application involving U.S. macroeconomic indicators, we show that the mixed frequency approach and the low frequency approach produce very different causal implications, with the former yielding more intuitively appealing result.

© 2016 Elsevier B.V. All rights reserved.

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, 1988) and Granger and Lin (1995).¹ 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.

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 then 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 a common low 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, henceforth MF-VAR, models to implement a new class of Granger causality tests.

We show that mixed frequency Granger causality tests better recover causal patterns in an underlying high frequency process compared to the traditional low frequency, henceforth LF, approach. We also formally prove that mixed frequency, henceforth MF, 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 MF-VAR approach works well for small differences in sampling frequencies—like quarterly/monthly mixtures.

We apply the MF causality tests to monthly U.S. inflation, monthly crude oil price fluctuations and quarterly real GDP growth. We also apply the conventional causality test to the aggregated quarterly price series and real GDP for comparison. These two approaches yield very different conclusions regarding causal patterns. In particular, significant causality from oil prices

[☆] 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. We thank participants at 2014 Japanese Joint Statistical Meeting (University of Tokyo, September) for helpful comments. We also thank two referees, an associate editor, and co-editor Oliver Linton for helpful comments.

* Corresponding author.

E-mail addresses: eghysels@unc.edu (E. Ghysels), jbhill@email.unc.edu (J.B. Hill), motegi@aoni.waseda.jp (K. Motegi).

¹ 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 notably Lütkepohl (1993), Renault et al. (1998), Marcellino (1999), Breitung and Swanson (2002), McCrorie and Chambers (2006), and Silvestrini and Veredas (2008), among others.

to inflation is detected by the MF approach but not when applying conventional Granger causality tests based on LF data. The result suggests that the quarterly frequency is too coarse to capture such causality.

The nature of MF implies that we are potentially dealing with multi-horizon Granger causality since more than one period high frequency (HF) observations are collected within a single LF time span. Moreover, as in the standard (single frequency) VAR literature, exploring MF Granger causality among more than two series also invariably relates to the notion of multi-horizon causality, see in particular 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, say 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 et al. (2006) formulate Wald tests for multiple-horizon non-causality. Their framework will be used extensively in our analysis.

In addition to the causality literature, the present paper also draws upon and contributes to the MIDAS literature originated by Ghysels et al. (2004, 2005). A number of papers have linked MIDAS regressions to (latent) high frequency VAR models, such as Kuzin et al. (2011) and Foroni et al. (2015), whereas Ghysels (forthcoming) discusses the link between MF-VAR models and MIDAS regressions. None of these papers study in any detail the issue of Granger causality, which is the topic of the present paper.

The paper is organized as follows. In Section 2 we frame MF-VAR models and present core assumptions. In Section 3 we derive the asymptotic properties of the least squares estimator, and a Wald statistic for testing arbitrary linear restrictions on an h -step ahead autoregression. We then develop the mixed frequency causality tests by extending ideas on h -step ahead non-causality tests in Dufour and Renault (1998) and Dufour et al. (2006) to mixed frequencies. Section 4 discusses how we can recover underlying causality using a mixed frequency approach compared to a traditional LF approach. Section 5 shows that the mixed frequency causality tests have higher local asymptotic power than the LF ones do. Section 6 reports Monte Carlo simulation results and documents the finite sample power improvements achieved by the mixed frequency causality test. In Section 7 we apply the mixed frequency and LF causality tests to U.S. macroeconomic data. Finally, Section 8 provides some concluding remarks.

We will use the following notational conventions throughout. Let $\mathbf{A} \in \mathbb{R}^{n \times l}$. The l_2 -norm is $|\mathbf{A}| := (\sum_{i=1}^n \sum_{j=1}^l a_{ij}^2)^{1/2} = (\text{tr}[\mathbf{A}'\mathbf{A}])^{1/2}$; the L_r -norm is $\|\mathbf{A}\|_r := (\sum_{i=1}^n \sum_{j=1}^l E|a_{ij}|^r)^{1/r}$; the determinant is $\det(\mathbf{A})$; and the transpose is \mathbf{A}' . $\mathbf{0}_{n \times 1}$ is an $n \times 1$ matrix of zeros. \mathbf{I}_K is the K -dimensional identity matrix. $\text{Var}[\mathbf{A}]$ is the variance-covariance matrix of a stochastic matrix \mathbf{A} . $\mathbf{B} \circ \mathbf{C}$ denotes element-by-element multiplication for conformable vectors \mathbf{B}, \mathbf{C} .

2. Mixed frequency data model specifications

In this section we present the MF-VAR model and three main assumptions. We want to characterize three settings, respectively high, mixed and low frequency or HF, MF and LF. We begin by considering a partially latent underlying HF process. Using the notation of Ghysels (forthcoming), the HF process contains $\{\mathbf{x}_H(\tau_L, k)\}_{k=1}^m$ and $\{\mathbf{x}_L(\tau_L, k)\}_{k=1}^m$, where $\tau_L \in \{0, \dots, T_L\}$ is the LF time index (e.g. quarterly), $k \in \{1, \dots, m\}$ denotes the HF (e.g. monthly), 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 $\mathbf{x}_H(\tau_L, k) \in \mathbb{R}^{K_H \times 1}$, $K_H \geq 1$, are called HF variables, whereas $\mathbf{x}_L(\tau_L, k) \in \mathbb{R}^{K_L \times 1}$, $K_L \geq 1$, are latent LF variables because they are not observed at high frequencies—as only some temporal aggregates, denoted $\mathbf{x}_H(\tau_L)$, 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 (forthcoming). In reality the econometrician's choice is limited to MF and LF cases. Only LF variables have been aggregated from a latent HF process in a MF setting, whereas both low and high frequency variables are aggregated from the latent HF process to form a LF process. Following Lütkepohl (1987) we consider only linear aggregation schemes involving weights $\mathbf{w} = [w_1, \dots, w_m]'$ such that:

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

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 $k = 1, \dots, m$.² In summary, we observe:

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

A key idea of MF-VAR models is to stack everything observable given a MF process in what we call the *mixed frequency vector*:

$$\mathbf{X}(\tau_L) = [\mathbf{x}_H(\tau_L, 1)', \dots, \mathbf{x}_H(\tau_L, m)', \mathbf{x}_L(\tau_L)']. \quad (2.2)$$

The dimension of the mixed frequency vector is $K = K_L + mK_H$. Note that $\mathbf{x}_L(\tau_L)$ is the last block in the mixed frequency vector—a conventional assumption implying that it is observed after $\mathbf{x}_H(\tau_L, m)$. 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.

Example 1 (*Quarterly Real GDP*). A leading example of how a mixed frequency model is useful in macroeconomics concerns quarterly real GDP growth $x_L(\tau_L)$, where existing studies of causal patterns use monthly unemployment, oil prices, inflation, interest rates, etc. aggregated into quarters (see e.g. Hill (2007) for references). Consider the monthly oil price changes and CPI inflation stacked into a 6×1 vector (since we have two series for three months) $[\mathbf{x}_H(\tau_L, 1)', \dots, \mathbf{x}_H(\tau_L, 3)', \mathbf{x}_L(\tau_L)']'$, which concatenated with quarterly GDP yields the vector $\mathbf{X}(\tau_L)$ appearing in (2.2), which will be further analyzed in Section 7.

We will make a number of standard regulatory assumptions. Let $\mathcal{F}_{\tau_L} \equiv \sigma(\mathbf{X}(t) : t \leq \tau_L)$. In particular we assume $E[\mathbf{X}(\tau_L)|\mathcal{F}_{\tau_L-1}]$ has a version that is *almost surely* linear in $\{\mathbf{X}(\tau_L-1), \dots, \mathbf{X}(\tau_L-p)\}$ for some finite $p \geq 1$.

Assumption 2.1. The process $\mathbf{X}(\tau_L)$ is governed by a VAR(p) for some $p \geq 1$:

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

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

The coefficients \mathbf{A}_k are $K \times K$ matrices for $k = 1, \dots, p$. The $K \times 1$ error vector $\boldsymbol{\epsilon}(\tau_L) = [\epsilon_1(\tau_L), \dots, \epsilon_K(\tau_L)]'$ is a strictly stationary martingale difference with respect to increasing $\mathcal{F}_{\tau_L} \subset \mathcal{F}_{\tau_L+1}$, where $\Omega \equiv E[\boldsymbol{\epsilon}(\tau_L)\boldsymbol{\epsilon}(\tau_L)']$ is positive definite.

Remark 1. We do not include a constant term in (2.3) solely to reduce notation, therefore $\mathbf{X}(\tau_L)$ 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).

In addition, the following standard assumptions ensure stationarity and α -mixing of the observed time series and the MF-VAR errors.³

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

Remark 2. The condition for stationarity here is the same as for a standard LF-VAR. This follows since the MF-VAR(p) in (2.3) has a VAR(1) representation by stacking variables, and a VAR(1) is a first order Markov process. Stationarity allows us to focus on the pure idea of identifying latent causal patterns under mixed frequency. We rule out non-stationary VAR's and cointegration since these ideas under mixed frequency are themselves substantial.⁴

Moreover, let $\mathcal{G}_s^t \equiv \sigma(\{\mathbf{X}(i), \boldsymbol{\epsilon}(i)\} : s \leq i \leq t)$ and define the mixing coefficients $\alpha_h \equiv \sup_{\mathcal{A} \subset \mathcal{G}_{-\infty}^t, \mathcal{B} \subset \mathcal{G}_{t+h}^\infty} |P(\mathcal{A} \cap \mathcal{B}) - P(\mathcal{A})P(\mathcal{B})|$ (cfr. Rosenblatt (1956) and Ibragimov (1975)).

Assumption 2.3. $\mathbf{X}(\tau_L)$ and $\boldsymbol{\epsilon}(\tau_L)$ are α -mixing: $\sum_{h=0}^{\infty} \alpha_{2^h} < \infty$.

Remark 3. Recall that α -mixing implies mixing in the ergodic sense, and therefore ergodicity (see Petersen (1983)). Hence by Assumptions 2.1–2.3 $\{\mathbf{X}(\tau_L), \boldsymbol{\epsilon}(\tau_L)\}$ are stationary and ergodic.

Remark 4. We must impose a weak dependence property on $\boldsymbol{\epsilon}(\tau_L)$ since it is not i.i.d., and the martingale difference sequence property (MDS) alone does not suffice for a central limit theory (cfr. McLeish (1974)), hence we impose a mixing condition. An infinite order lag function $\mathbf{X}(\tau_L)$ of mixing $\boldsymbol{\epsilon}(\tau_L)$ need not be mixing, hence we also assume $\mathbf{X}(\tau_L)$ is mixing (see Chapter 2.3.1 Doukhan (1994)). Asymptotics for our estimator only requires $\sum_{h=0}^{\infty} \alpha_{2^h} < \infty$ because under Assumptions 2.1 and 2.2 $\mathbf{X}(\tau_L)$ has a positive bounded spectral density (cfr. Ibragimov (1975)). Asymptotics for a Wald statistic naturally requires a greater moment bound and a more restrictive mixing condition. See Theorem 3.2.

Remark 5. Assumptions 2.1–2.3 are identical in form to the set of assumptions we may want to impose for a LF-VAR.

Example 2 (Bivariate Structural MF-VAR(1) with $m = 3$). For illustration we present a concrete example of structural MF-VAR model and then derive a reduced form. Consider a bivariate case

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 See, e.g., Seong et al. (2013), Miller (2014) and Ghysels and Miller (forthcoming). See, especially, Götz et al. (2013) and Ghysels and Miller (forthcoming) for tests of cointegration in mixed frequency data. See also Marcellino (1999) for a broad treatment of (non)stationarity under temporal aggregation.

with $m = 3$ (e.g. monthly inflation vs. quarterly GDP). Suppose that the structural form is given by

$$\begin{aligned} & \underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ -d & 1 & 0 & 0 \\ 0 & -d & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}}_{\equiv N} \begin{bmatrix} x_H(\tau_L, 1) \\ x_H(\tau_L, 2) \\ x_H(\tau_L, 3) \\ x_L(\tau_L) \end{bmatrix} \\ &= \underbrace{\begin{bmatrix} 0 & 0 & d & c_1 \\ 0 & 0 & 0 & c_2 \\ 0 & 0 & 0 & c_3 \\ b_3 & b_2 & b_1 & a \end{bmatrix}}_{\equiv M} \begin{bmatrix} x_H(\tau_L - 1, 1) \\ x_H(\tau_L - 1, 2) \\ x_H(\tau_L - 1, 3) \\ x_L(\tau_L - 1) \end{bmatrix} + \underbrace{\begin{bmatrix} \xi_H(\tau_L, 1) \\ \xi_H(\tau_L, 2) \\ \xi_H(\tau_L, 3) \\ \xi_L(\tau_L) \end{bmatrix}}_{\equiv \xi(\tau_L)} \end{aligned}$$

or $\mathbf{N}\mathbf{x}(\tau_L) = \mathbf{M}\mathbf{x}(\tau_L - 1) + \boldsymbol{\xi}(\tau_L)$. It is assumed that x_H follows a high frequency AR(1) with coefficient d . The impact of lagged x_L on x_H is governed by c_1 , c_2 , and c_3 . x_L follows a low frequency AR(1) with coefficient a . The impact of lagged x_H on x_L is governed by b_1 , b_2 , and b_3 . Premultiply both sides of the structural form by

$$\mathbf{N}^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ d & 1 & 0 & 0 \\ d^2 & d & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

to get the reduced form $\mathbf{X}(\tau_L) = \mathbf{A}_1 \mathbf{X}(\tau_L - 1) + \boldsymbol{\epsilon}(\tau_L)$, where

$$\mathbf{A}_1 = \mathbf{N}^{-1} \mathbf{M} = \begin{bmatrix} 0 & 0 & d & \sum_{i=1}^1 d^{1-i} c_i \\ 0 & 0 & d^2 & \sum_{i=1}^2 d^{2-i} c_i \\ 0 & 0 & d^3 & \sum_{i=1}^3 d^{3-i} c_i \\ b_3 & b_2 & b_1 & a \end{bmatrix} \quad \text{and}$$

$$\boldsymbol{\epsilon}(\tau_L) = \mathbf{N}^{-1} \boldsymbol{\xi}(\tau_L).$$

Since this paper focuses on Granger causality, the off-diagonal blocks of \mathbf{A}_1 are of particular interest. Below we demonstrate that x_H does not Granger cause x_L in the mixed frequency sense if and only if $b_1 = b_2 = b_3 = 0$. Similarly, x_L does not Granger cause x_H in the mixed frequency sense if and only if $\sum_{i=1}^j d^{j-i} c_i$ for $j = 1, 2, 3$. The precise meaning of “the mixed frequency sense” will be clarified in Section 3.1.

As seen in the upper-right block of \mathbf{A}_1 , non-causality from x_L to x_H involves d , the AR(1) coefficient of x_H . In this specific example a , the AR(1) coefficient of x_L , does not show up in the off-diagonal blocks of \mathbf{A}_1 . The persistence of x_L can be relevant in more general settings (e.g. longer autoregressive lags, trivariate models, etc.). Generally, not only the structural interdependence between x_H and x_L (i.e. b 's and c 's) but also the persistence of each variable plays a role in Granger causality.

Before proceeding, it is worth explaining why we adopt the MF-VAR approach discussed so far and do not use other mixed frequency model specifications. The framework we use is different from the typical MF approaches which involve latent processes. Several papers, Kuzin et al. (2011), Eraker et al. (forthcoming), Foroni et al. (2015), among others, have proposed HF-VAR models and apply a state space framework to link latent processes with observables. State space models are, using the terminology of Cox (1981), parameter-driven models. The MF-VAR models we use are, using again the same terminology, observation-driven models as they are formulated exclusively in terms of observable data. This means that our approach directly relates to standard VAR model settings and therefore allows us to exploit many testing

tools available to us. The same argument also applies to MF modeling approaches which rely on a factor structure, see e.g., [Mariano and Murasawa \(2010\)](#), [Giannone et al. \(2008\)](#) and [Baíbura and Rünstler \(2011\)](#). Such a framework would not be suitable to discuss Granger causality, not only because of the presence of latent variables but also because of the assumed factor structure.

3. Testing causality with mixed frequency data

In this section we cover 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 LF variables. We cover non-causality from all high frequency variables to all LF variables and vice versa, cases for which we give explicit formulae for the selection matrices used to compute relevant Wald statistics. We also establish large sample results for parameter estimators, and corresponding Wald statistics. The latter are framed for tests of linear restrictions in a h -step ahead autoregression used for testing multi-step ahead non-causality (cf. [Dufour et al. \(2006\)](#)).

3.1. Non-causality in mixed frequencies

The definition of Granger (non-)causality is associated with information sets. Before discussing mixed frequency cases, we briefly review a generic notion of Granger non-causality at different horizons (see [Dufour and Renault \(1998\)](#) for complete details). Extension to mixed frequency cases will be fairly straightforward.

For convenience we use the low frequency time index τ_L . Consider a vector of random variables $\mathbf{W}(\tau_L) = [\mathbf{x}(\tau_L)', \mathbf{y}(\tau_L)', \mathbf{z}(\tau_L)']'$. Let $\mathbf{x}(-\infty, \tau_L]$ denote the Hilbert space spanned by $\{\mathbf{x}(\tau) | \tau \leq \tau_L\}$. $\mathbf{y}(-\infty, \tau_L]$ and $\mathbf{z}(-\infty, \tau_L]$ are defined in the same way. Similarly, let $\mathcal{J}(\tau_L) = \mathbf{W}(-\infty, \tau_L]$ be the Hilbert space spanned by the entire vector $\{\mathbf{W}(\tau) | \tau \leq \tau_L\}$. We can write $\mathcal{J}(\tau_L) = \mathbf{x}(-\infty, \tau_L] + \mathbf{y}(-\infty, \tau_L] + \mathbf{z}(-\infty, \tau_L]$, where $\mathbf{x}(-\infty, \tau_L] + \mathbf{y}(-\infty, \tau_L]$ denotes the Hilbert subspace generated by the elements of $\mathbf{x}(-\infty, \tau_L]$ and $\mathbf{y}(-\infty, \tau_L]$. Let $P[\mathbf{x}(\tau_L + h) | \mathcal{J}(\tau_L)]$ be the best linear forecast of $\mathbf{x}(\tau_L + h)$ based on $\mathcal{J}(\tau_L)$ in the sense of a covariance orthogonal projection.

Definition 3.1 (*Non-causality at Different Horizons*). (i) \mathbf{y} does not cause \mathbf{x} at horizon h given \mathcal{J} (denoted by $\mathbf{y} \not\rightarrow_h \mathbf{x} | \mathcal{J}$):

$$P[\mathbf{x}(\tau_L + h) | \mathbf{x}(-\infty, \tau_L] + \mathbf{z}(-\infty, \tau_L)] = P[\mathbf{x}(\tau_L + h) | \mathcal{J}(\tau_L)] \\ \forall \tau_L \in \mathbb{Z}.$$

Moreover, (ii) \mathbf{y} does not cause \mathbf{x} up to horizon h given \mathcal{J} (denoted by $\mathbf{y} \not\rightarrow_{(h)} \mathbf{x} | \mathcal{J}$) if $\mathbf{y} \not\rightarrow_k \mathbf{x} | \mathcal{J}$ for all $k \in \{1, \dots, h\}$.

Definition 3.1 corresponds to [Dufour and Renault's \(1998\)](#) Definition 2.2 with covariance stationary processes. **Definition 3.1** states that non-causality from \mathbf{y} to \mathbf{x} at horizon h means that the h -step ahead prediction of \mathbf{x} is unchanged whether the past and present values of \mathbf{y} are available or not. While **Definition 3.1** covers causality from \mathbf{y} to \mathbf{x} , other cases (e.g. causality from \mathbf{z} to \mathbf{y}) can be defined analogously.

We now discuss mixed frequency cases. For exposition we assume that there two high frequency variables $x_{H,1}$ and $x_{H,2}$ and one LF variables x_L . The ratio of sampling frequencies is $m = 3$. In this case the mixed frequency vector $\mathbf{X}(\tau_L)$ defined in (2.2) becomes a 7×1 vector as in [Example 1](#):

$$\mathbf{X}(\tau_L) = [x_{H,1}(\tau_L, 1), x_{H,2}(\tau_L, 1), x_{H,1}(\tau_L, 2), x_{H,2}(\tau_L, 2), \\ x_{H,1}(\tau_L, 3), x_{H,2}(\tau_L, 3), x_L(\tau_L)]'.$$

We partition the mixed frequency vector by three subvectors: $\tilde{\mathbf{x}}_{H,1}(\tau_L) = [x_{H,1}(\tau_L, 1), x_{H,1}(\tau_L, 2), x_{H,1}(\tau_L, 3)]'$, $\tilde{\mathbf{x}}_{H,2}(\tau_L) =$

$[x_{H,2}(\tau_L, 1), x_{H,2}(\tau_L, 2), x_{H,2}(\tau_L, 3)]'$, and $x_L(\tau_L)$. Hence, the partition of $\mathbf{X}(\tau_L)$ is made variable by variable, and the m observations of a high frequency variable are stacked.

Treating the m observations of a high frequency variable as a stacked bundle, we are now ready to apply [Definition 3.1](#) to the mixed frequency case. Take $\mathcal{J}(\tau_L) = \mathbf{X}(-\infty, \tau_L]$, then it can be understood that $\mathcal{J}(\tau_L) = \tilde{\mathbf{x}}_{H,1}(-\infty, \tau_L] + \tilde{\mathbf{x}}_{H,2}(-\infty, \tau_L] + x_L(-\infty, \tau_L]$. $\mathcal{J}(\tau_L)$ is called the *mixed frequency reference information set in period τ_L* . We also call $\mathcal{J} = \{\mathcal{J}(\tau_L) | \tau_L \in \mathbb{Z}\}$ the *mixed frequency reference information set*. $x_{H,1}$ does not cause x_L at horizon h given \mathcal{J} (denoted by $x_{H,1} \not\rightarrow_h x_L | \mathcal{J}$) if:

$$P[x_L(\tau_L + h) | \tilde{\mathbf{x}}_{H,2}(-\infty, \tau_L] + x_L(-\infty, \tau_L)] \\ = P[x_L(\tau_L + h) | \mathcal{J}(\tau_L)] \quad \forall \tau_L \in \mathbb{Z}.$$

In words, non-causality from $x_{H,1}$ to x_L at horizon h means that the h -step ahead prediction of x_L is unchanged whether the past and present high frequency values of $x_{H,1}$ are available or not. The reverse direction is defined analogously: x_L does not cause $x_{H,1}$ at horizon h given \mathcal{J} (denoted by $x_L \not\rightarrow_h x_{H,1} | \mathcal{J}$) if:

$$P[\tilde{\mathbf{x}}_{H,1}(\tau_L + h) | \tilde{\mathbf{x}}_{H,1}(-\infty, \tau_L] + \tilde{\mathbf{x}}_{H,2}(-\infty, \tau_L)] \\ = P[\tilde{\mathbf{x}}_{H,1}(\tau_L + h) | \mathcal{J}(\tau_L)] \quad \forall \tau_L \in \mathbb{Z}.$$

In general we have K_H high frequency variables and K_L LF variables. For non-causality from a single variables to another single variable, we have four cases:

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 is written as $H_0^1(h) : x_{L,j_1} \not\rightarrow_h x_{L,j_2} | \mathcal{J}$.

Case 2 (high to low) Non-causality from the i_1 th high frequency variable, x_{H,i_1} , to the j_1 th low frequency variable, x_{L,j_1} , at horizon h . The null hypothesis is written as $H_0^2(h) : x_{H,i_1} \not\rightarrow_h x_{L,j_1} | \mathcal{J}$.

Case 3 (low to high) $H_0^3(h) : x_{L,j_1} \not\rightarrow_h x_{H,i_1} | \mathcal{J}$.

Case 4 (high to high) $H_0^4(h) : x_{H,i_1} \not\rightarrow_h x_{H,i_2} | \mathcal{J}$.

We often treat all K_H high frequency variables as a group and all K_L LF variables as the other group. Non-causality from one group to the other is summarized as follows.

Case I (all high to all low) Non-causality from all high frequency variables $x_{H,1}, \dots, x_{H,K_H}$ to all LF variables $x_{L,1}, \dots, x_{L,K_L}$ at horizon h . The null hypothesis is written as $H_0^I(h) : \mathbf{x}_H \not\rightarrow_h \mathbf{x}_L | \mathcal{J}$. The precise meaning is that $P[\mathbf{x}_L(\tau_L + h) | \mathbf{x}_H(-\infty, \tau_L)] = P[\mathbf{x}_L(\tau_L + h) | \mathcal{J}(\tau_L)]$ for all $\tau_L \in \mathbb{Z}$, where $\mathbf{x}_L(\tau_L) = [x_{L,1}(\tau_L), \dots, x_{L,K_L}(\tau_L)]'$.

Case II (all low to all high) $H_0^{II}(h) : \mathbf{x}_L \not\rightarrow_h \mathbf{x}_H | \mathcal{J}$. The precise meaning is that $P[\tilde{\mathbf{x}}_H(\tau_L + h) | \tilde{\mathbf{x}}_H(-\infty, \tau_L)] = P[\tilde{\mathbf{x}}_H(\tau_L + h) | \mathcal{J}(\tau_L)]$ for all $\tau_L \in \mathbb{Z}$, where $\tilde{\mathbf{x}}_H(\tau_L) = [\tilde{\mathbf{x}}_{H,1}(\tau_L)', \dots, \tilde{\mathbf{x}}_{H,K_H}(\tau_L)']'$.

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. In the bivariate system non-causality at one horizon is synonymous to non-causality at all horizons (see [Dufour and Renault \(1998, Proposition 2.3\)](#) and [Florens and Mouchart \(1982, p. 590\)](#)). In order 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. OLS estimator, Wald statistic and asymptotic properties

We now detail estimation and inference for MF-VAR model (2.3). If the VAR(p) model in (2.3) were standard, then the off-diagonal elements of any matrix \mathbf{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 \mathbf{A}_k . It is therefore natural to start with a multi-horizon setting. We do so, at first, focusing on multiple LF prediction horizons which we will denote by $h \in \mathbb{N}$.⁵

It is convenient to iterate (2.3) over the desired test horizon h in order to deduce simple testable parameter restrictions for non-causality at horizon h . The result is the (p, h) -autoregression (cfr. Dufour et al. (2006)):

$$\mathbf{X}(\tau_L + h) = \sum_{k=1}^p \mathbf{A}_k^{(h)} \mathbf{X}(\tau_L + 1 - k) + \mathbf{u}^{(h)}(\tau_L), \quad (3.1)$$

where

$$\begin{aligned} \mathbf{A}_k^{(1)} &= \mathbf{A}_k \quad \text{and} \quad \mathbf{A}_k^{(i)} = \mathbf{A}_{k+i-1} + \sum_{l=1}^{i-1} \mathbf{A}_{l-i} \mathbf{A}_k^{(l)} \quad \text{for } i \geq 2, \\ \mathbf{u}^{(h)}(\tau_L) &= \sum_{k=0}^{h-1} \Psi_k \boldsymbol{\epsilon}(\tau_L - k). \end{aligned} \quad (3.2)$$

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

$$\mathbf{B}(h) = [\mathbf{A}_1^{(h)}, \dots, \mathbf{A}_p^{(h)}]' \in \mathbb{R}^{pK \times K}. \quad (3.3)$$

If all variables were aggregated into a common LF and expanded into a (p, h) -autoregression, then h -step ahead non-causality has a simple parametric expression in terms of $\mathbf{B}(h)$; cfr. Dufour et al. (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 et al. (2006) since in MF-VAR models the restrictions will often deal with linear parametric restrictions across multiple equations. Nevertheless, in a generic sense we show in Section 3 that non-causality at horizon h between any set of variables in a MF-VAR model can be expressed as linear constraints with respect to $\mathbf{B}(h)$. Hence, the null hypothesis of interest is a linear restriction:

$$H_0(h) : \mathbf{R} \operatorname{vec}[\mathbf{B}(h)] = \mathbf{r}, \quad (3.4)$$

where \mathbf{R} is a $q \times pK^2$ selection matrix of full row rank q , and $\mathbf{r} \in \mathbb{R}^q$. We leave complete details of the construction of \mathbf{R} for Section 3.3.

In view of frequency mixing, we require a more compact notation for the least squares estimator $\hat{\mathbf{B}}(h)$. Define

$$\mathbf{W}_h(h) = [\mathbf{X}(h), \mathbf{X}(1+h), \dots, \mathbf{X}(T_L - k + h)]' \in \mathbb{R}^{(T_L - k + 1) \times K}$$

$$\mathbf{W}(\tau_L, p) = [\mathbf{X}(\tau_L)', \mathbf{X}(\tau_L - 1)', \dots, \mathbf{X}(\tau_L - p + 1)']' \in \mathbb{R}^{pK \times 1}$$

$$\overline{\mathbf{W}}_p(h) = [\mathbf{W}(0, p), \mathbf{W}(1, p), \dots, \mathbf{W}(T_L - h, p)]' \in \mathbb{R}^{(T_L - h + 1) \times pK}.$$

Now stack the error $\mathbf{u}^{(h)}(\tau_L)$ from (3.1) and (3.2) as follows:

$$\mathbf{U}_l(k) = [\mathbf{u}^{(h)}(l), \mathbf{u}^{(h)}(1+l), \dots, \mathbf{u}^{(h)}(T_L - k + l)]' \in \mathbb{R}^{(T_L - k + 1) \times K}.$$

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.

Then (p, h) -autoregression (3.1) has the equivalent representation

$$\mathbf{W}_h(h) = \overline{\mathbf{W}}_p(h) \mathbf{B}(h) + \mathbf{U}_h(h), \quad (3.5)$$

hence the OLS estimator $\hat{\mathbf{B}}(h)$ of $\mathbf{B}(h)$ is simply

$$\hat{\mathbf{B}}(h) \equiv [\overline{\mathbf{W}}_p(h)' \overline{\mathbf{W}}_p(h)]^{-1} \overline{\mathbf{W}}_p(h)' \mathbf{W}_h(h).$$

The asymptotic variance of $\hat{\mathbf{B}}(h)$ follows by noting the easily verified relation:

$$\begin{aligned} &\sqrt{T_L^*} \operatorname{vec} [\hat{\mathbf{B}}(h) - \mathbf{B}(h)] \\ &= [\mathbf{I}_K \otimes \hat{\Gamma}_{p,0}^{-1}] \frac{1}{\sqrt{T_L^*}} \sum_{\tau_L=0}^{T_L^*-1} \mathbf{Y}(\tau_L + h, p) \end{aligned} \quad (3.6)$$

where

$$\mathbf{Y}(\tau_L + h, p) \equiv \operatorname{vec} [\mathbf{W}(\tau_L, p) \mathbf{u}^{(h)}(\tau_L + h)'], \quad (3.7)$$

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

and $T_L^* \equiv T_L - h + 1$ is the effective sample size of the (p, h) -autoregression. Now define the covariance matrices for $\mathbf{W}(\tau_L, p)$ and $\mathbf{Y}(\tau_L + h, p)$, and the long run variance for $\mathbf{Y}(\tau_L + h, p)$:

$$\Gamma_{p,0} \equiv E [\mathbf{W}(\tau_L, p) \mathbf{W}(\tau_L, p)'] \quad \text{and}$$

$$\Delta_{p,s}(h) \equiv E [\mathbf{Y}(\tau_L + h + s, p) \mathbf{Y}(\tau_L + h, p)'].$$

$$\mathbf{D}_{p,T_L^*}(h) \equiv \operatorname{Var} \left[\frac{1}{\sqrt{T_L^*}} \sum_{\tau_L=0}^{T_L^*-1} \mathbf{Y}(\tau_L + h, p) \right] \quad \text{and}$$

$$\mathbf{D}_p(h) \equiv \lim_{T_L^* \rightarrow \infty} \mathbf{D}_{p,T_L^*}(h).$$

Assumptions 2.1–2.3 suffice for $\hat{\mathbf{B}}(h)$ to be consistent for $\mathbf{B}(h)$ and asymptotically normal. Limits are with respect to $T_L \rightarrow \infty$ hence $T_L^* \rightarrow \infty$.

Theorem 3.1. Under Assumptions 2.1–2.3 $\hat{\mathbf{B}}(h) \xrightarrow{p} \mathbf{B}(h)$ and

$$\sqrt{T_L^*} \operatorname{vec} [\hat{\mathbf{B}}(h) - \mathbf{B}(h)] \xrightarrow{d} N (\mathbf{0}_{pK^2 \times 1}, \Sigma_p(h)),$$

where

$$\Sigma_p(h) \equiv (\mathbf{I}_K \otimes \Gamma_{p,0}^{-1}) \mathbf{D}_p(h) (\mathbf{I}_K \otimes \Gamma_{p,0}^{-1})'$$

is positive definite. Moreover:

$$\mathbf{D}_p(h) = \Delta_{p,0}(h) + \sum_{s=1}^{h-1} [\Delta_{p,s}(h) + \Delta_{p,s}(h)']. \quad (3.8)$$

See Appendix A for a proof.

If we have a consistent estimator $\hat{\Sigma}_p(h)$ for $\Sigma_p(h) \equiv (\mathbf{I}_K \otimes \Gamma_{p,0}^{-1}) \mathbf{D}_p(h) (\mathbf{I}_K \otimes \Gamma_{p,0}^{-1})'$ which is almost surely positive semi-definite for $T_L^* \geq 1$, we can define the Wald statistic

$$\begin{aligned} W_{T_L^*}[H_0(h)] &\equiv T_L^* \left(\mathbf{R} \operatorname{vec} [\hat{\mathbf{B}}(h)] - \mathbf{r} \right)' \times \left(\mathbf{R} \hat{\Sigma}_p(h) \mathbf{R}' \right)^{-1} \\ &\quad \times \left(\mathbf{R} \operatorname{vec} [\hat{\mathbf{B}}(h)] - \mathbf{r} \right). \end{aligned} \quad (3.9)$$

Positive definiteness of $\hat{\Sigma}_p(h)$ and consistency ensure $\mathbf{R} \hat{\Sigma}_p(h) \mathbf{R}'$ is, with probability approaching one, non-singular for any $\mathbf{R} \in \mathbb{R}^{q \times pK^2}$ with full row rank.

A natural estimator of $\Gamma_{p,0}$ is the sample conjugate:

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

Under Assumptions 2.1–2.2 $\hat{\Gamma}_{p,0}$ is almost surely positive definite.

Now consider the long-run variance $\mathbf{D}_p(h)$. Denote the least squares residual $\hat{\mathbf{U}}_h(h) \equiv \mathbf{W}_h(h) - \overline{\mathbf{W}}_p(h)\hat{\mathbf{B}}(h)$ for model (3.5) and the resulting residual $\hat{\mathbf{u}}^{(h)}(\tau_L) \equiv \mathbf{X}(\tau_L) - \sum_{k=1}^p \hat{\mathbf{A}}_k^{(h)} \mathbf{X}(\tau_L-h+1-k)$ computed from (3.1), compute the sample version of $\mathbf{Y}(\tau_L+h, p)$ in (3.7),

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

and compute the sample covariance:

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

If $h = 1$ then from (3.8) the estimator of $\mathbf{D}_p(h)$ need only be $\hat{\mathbf{D}}_{p,T_L^*}(1) = \hat{\Delta}_{p,0}(1)$. Otherwise, a naïve estimator of $\mathbf{D}_p(h)$ simply substitutes $\hat{\Delta}_{p,s}(h)$ for $\Delta_{p,s}(h)$ in the right-hand side of (3.8), but such an estimator may not be positive semi-definite when $h > 1$.

We therefore exploit Newey and West's (1987) Bartlett kernel-based HAC estimator which ensures almost sure positive semi-definiteness for any $T_L^* \geq 1$ ⁶:

$$\hat{\mathbf{D}}_{p,T_L^*}(h) = \hat{\Delta}_{p,0}(h) + \sum_{s=1}^{n_{T_L^*}-1} \left(1 - \frac{s}{n_{T_L^*}}\right) (\hat{\Delta}_{p,s}(h) + \hat{\Delta}_{p,s}(h)') \quad (3.10)$$

with bandwidth $n_{T_L^*}: h \leq n_{T_L^*} \leq T_L^*$, $n_{T_L^*} \rightarrow \infty$ and $n_{T_L^*} = o(T_L^*)$. Notice $\hat{\Delta}_{p,0}(h) + \sum_{s=1}^{h-1} (1-s/n_{T_L^*}) (\hat{\Delta}_{p,s}(h) + \hat{\Delta}_{p,s}(h)')$ is also a valid estimator since $\mathbf{D}_p(h)$ depends on only $h-1$ lags of $\Delta_{p,s}(h)$, but this estimator too need not be positive semi-definite in small samples.

Our proposed estimator of Σ_p is therefore

$$\hat{\Sigma}_p = (\mathbf{I}_K \otimes \hat{\Gamma}_{p,0}^{-1}) \times \hat{\mathbf{D}}_{p,T_L^*}(h) \times (\mathbf{I}_K \otimes \hat{\Gamma}_{p,0}^{-1}). \quad (3.11)$$

Almost sure positive definiteness of $\hat{\Gamma}_{p,0}$ and positive semi-definiteness of $\hat{\mathbf{D}}_{p,T_L^*}(h)$ imply $\hat{\Sigma}_p$ is almost surely positive semi-definite. Consistency requires stronger moment and mixing conditions.

Assumption 3.1. For some $\delta > 0$ let $E|\epsilon_i(\tau_L)|^{4+\delta} < \infty$ for each $i = 1, \dots, K$.

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

Theorem 3.2. Under Assumptions 2.1–2.3 and 3.1, $W_{T_L^*}[H_0(h)] \xrightarrow{d} \chi_q^2$ under $H_0(h)$.

3.3. Selection matrices for mixed frequency causality tests

Our next task is to construct the selection matrices \mathbf{R} for the various null hypotheses (3.4) associated with the six generic cases. This requires deciphering parameter restrictions for non-causation based on the (p, h) -autoregression appearing in (3.1).

⁶ There is a large choice of valid kernels, including Parzen and Tukey–Hanning. See Andrews (1991) and de Jong and Davidson (2000).

Characterizing restrictions on $\mathbf{A}_k^{(h)}$ for each case above requires some additional matrix notation. Let $\mathbf{N} \in \mathbb{R}^{n \times n}$, and let $a, b, c, d, \iota, \iota' \in \{1, \dots, n\}$ with $a \leq b, c \leq d$, and $(b-a)/\iota$ and $(d-c)/\iota'$ being nonnegative integers. Then we define $\mathbf{N}(a : \iota : b, c : \iota' : d)$ 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, \dots , b th rows and c th, $(c+\iota')$ th, $(c+2\iota')$ th, \dots , d th columns of \mathbf{N} . 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:

$$\mathbf{N}(a : \iota : b, c : \iota' : d)' = \mathbf{N}'(c : \iota' : d, a : \iota : b). \quad (3.12)$$

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

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

$$\mathbf{A}_k^{(h)}(a : \iota : b, c : \iota' : d) = \mathbf{0} \quad \text{for each } k \in \{1, \dots, p\}, \quad (3.13)$$

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

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 $\mathbf{A}_k^{(h)}$ (i.e., the impact of the j_1 th LF variable on the j_2 th LF 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, \dots , $(mK_H + j_1, i_1 + (m-1)K_H)$ th elements of $\mathbf{A}_k^{(h)}$ 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 LF variable are zeros, i.e., the i_1 th high frequency variable has no impact as a whole on the j_1 th LF variable at a given horizon h .

When $H_0^3(h)$ holds, all mp coefficients of the j_1 th LF 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, \dots , $(i_1 + (m-1)K_H)$ th rows of $\mathbf{A}_k^{(h)}$, not columns. This means that we are testing simultaneous linear restrictions across multiple equations, unlike Dufour et al. (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 $\mathbf{A}_k^{(h)}$ are restricted to be zeros for each k , so the total number of zero restrictions is pm^2 . Under $H_0^4(h)$, the $K_L \times mK_H$ lower-left block of $\mathbf{A}_k^{(h)}$ is a null matrix. Finally, in Case II, the $mK_H \times K_L$ upper-right block of $\mathbf{A}_k^{(h)}$ is a null matrix if and only if $H_0^{II}(h)$ is true.

We can now combine the (p, h) -autoregression parameter set $\mathbf{B}(h)$ in (3.3) with the matrix construction (3.12), its implication for testable restrictions (3.13), and Table 1, to obtain generic formulae for \mathbf{R} and \mathbf{r} so that all six cases can be treated as special cases of (3.4).

The above can be summarized as follows:

Theorem 3.3. All hypotheses $H_0^i(h)$ for $i \in \{1, 2, 3, 4, I, II\}$ are special cases of $H_0(h)$ with

$$\mathbf{R} = [\Lambda(\delta_1)', \Lambda(\delta_2)', \dots, \Lambda(\delta_{g(a,\iota,b)p})']' \quad (3.14)$$

⁷ Recall that $x_{L,j}(\tau_L)$ and $\tilde{x}_{H,i}(\tau_L) = [x_{H,i}(\tau_L, 1), \dots, x_{H,i}(\tau_L, m)]'$ belong to \mathbf{X} in (2.2) for all $j \in \{1, \dots, K_L\}$ and $i \in \{1, \dots, K_H\}$. This is why non-causality under mixed frequencies is well-defined and Theorem 3.1 in Dufour and Renault (1998) can be applied directly.

Table 1

Linear parametric restrictions of non-causality.

Cases	a	ι	b	c	ι'	d	$\mathbf{0}$
$H_0^1(h)$	$mK_H + j_2$	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	$mK_H \times K_L$

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

and

$$\mathbf{r} = \mathbf{0}_{g(a, \iota, b)g(c, \iota', d)p \times 1}, \quad (3.15)$$

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

$$\delta_l = \delta_{l-1} + K + pK(\iota - 1)I(l - 1 = zp \text{ for some } z \in \mathbb{N}) \quad (3.16)$$

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

Several key points will help us understand (3.14)–(3.16). First, $g(a, \iota, b)$ and $g(c, \iota', d)$ represent how many rows and columns of $\mathbf{A}_k^{(h)}$ have zero restrictions for each $k \in \{1, \dots, p\}$, respectively. The total number of zero restrictions is therefore $q = g(a, \iota, b)g(c, \iota', d)p$ as in (3.15). Second, $\Lambda(\delta)$ has only one nonzero element in each row that is identically 1, signifying which element of $\text{vec}[\mathbf{B}(h)]$ is supposed to be zero. The location of 1 is determined by $\delta_1, \dots, \delta_{g(a, \iota, b)p}$, which are recursively updated according to (3.16). As seen in (3.16), 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 $\mathbf{B}(h)$.

Theorem 3.3 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
- Step 2** Calculate $a, \iota, b, c, \iota', d$ according to Table 1 for a given case of non-causality relation. Put those quantities into (3.14) and (3.15) to get \mathbf{R} and \mathbf{r} .
- Step 3** Use \mathbf{R} and \mathbf{r} in order to calculate the Wald test statistic $W_{T_L^*}[H_0(h)]$ in (3.9).

Example 3 (Selection Matrices \mathbf{R} and \mathbf{r}). Since Table 1 and Theorem 3.3 are rather abstract, we present a concrete example of how \mathbf{R} and \mathbf{r} are constructed in our trivariate simulation and empirical application. In Sections 6.2 and 7, 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 LF variable Z with $m = 3$. In this case the mixed frequency vector appearing in (2.2) can be written as:

$$\begin{aligned} \mathbf{W}(\tau_L) &= [X(\tau_L, 1), Y(\tau_L, 1), X(\tau_L, 2), Y(\tau_L, 2), X(\tau_L, 3), \\ &\quad Y(\tau_L, 3), Z(\tau_L)]'. \end{aligned}$$

⁸ A potential drawback of our approach as well as Dufour et al. (2006) is that the prediction horizon h is fixed at each test and therefore 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-causality test, in which a series of individual non-causality tests are performed to deduce causal chains and causation horizon. The present paper takes the Dufour et al. (2006) approach because of its simplicity. See Hill (2007) and Salamaliki and Venetis (2013) for a comparison of the two methods.

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

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

Since $p = h = 1$, $\mathbf{B}(h)$ appearing in (3.3) is simply \mathbf{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 $\text{vec}[\mathbf{B}(h)]$ appearing in (3.4). Hence, the proper choice of \mathbf{R} and \mathbf{r} is:

$$\mathbf{R} = \begin{bmatrix} \mathbf{0}_{1 \times 6} & 1 & \mathbf{0}_{1 \times 13} & 0 & \mathbf{0}_{1 \times 13} & 0 & \mathbf{0}_{1 \times 14} \\ \mathbf{0}_{1 \times 6} & 0 & \mathbf{0}_{1 \times 13} & 1 & \mathbf{0}_{1 \times 13} & 0 & \mathbf{0}_{1 \times 14} \\ \mathbf{0}_{1 \times 6} & 0 & \mathbf{0}_{1 \times 13} & 0 & \mathbf{0}_{1 \times 13} & 1 & \mathbf{0}_{1 \times 14} \end{bmatrix} \text{ and} \quad (3.17)$$

$$\mathbf{r} = \mathbf{0}_{3 \times 1}.$$

We now confirm that the same \mathbf{R} and \mathbf{r} can be obtained via Table 1 and Theorem 3.3. Non-causality from Z to X falls in Case 3 with $i_1 = j_1 = 1$ (i.e. non-causality from the first LF 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 $\{\delta_1, \delta_2, \delta_3\} = \{7, 21, 35\}$ by application of Theorem 3.3. This implies that $\mathbf{r} = \mathbf{0}_{3 \times 1}$ and $\mathbf{R} = [\Lambda(7)', \Lambda(21)', \Lambda(35)']'$, where $\Lambda(\delta)$ 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 3.3 provide correct \mathbf{R} and \mathbf{r} shown in (3.17).

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 series does not resolve all HF causal patterns, using MF instead of using exclusively LF series promotes sharper inference. A fundamental lesson from this section is that the convention in the macroeconomics literature of using a common LF can erode, or eradicate, the possibility of detecting true HF causal patterns.

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. Here our analysis extends Marcellino (1999).

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.

Denote by $\{\mathbf{x}_L(\tau_L, k)\}$ the latent high frequency version of the observed LF variable $\mathbf{x}_L(\tau_L)$. The latent high frequency variable is then

$$\bar{\mathbf{X}}(\tau_L, k) = [\mathbf{x}_H(\tau_L, k)', \mathbf{x}_L(\tau_L, k)']' \in \mathbb{R}^{K^*}$$

for $k = 1, \dots, m$, where $K^* = K_H + K_L$.

We require the *high frequency* lag operator \mathcal{L}_H , which satisfies:

$$\mathcal{L}_H^l \bar{\mathbf{X}}(\tau_L, k) = \bar{\mathbf{X}}(\tau_L - l, l')$$

with

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

Note that $\lfloor x \rfloor$ is the largest integer not larger than x . For example, $\mathcal{L}_H^2 \bar{\mathbf{X}}(\tau_L, 2) = \bar{\mathbf{X}}(\tau_L, 1)$ and $\mathcal{L}_H^1 \bar{\mathbf{X}}(\tau_L, 1) = \bar{\mathbf{X}}(\tau_L - 1, m)$. Similarly, the *low frequency* lag operator \mathcal{L}_L satisfies $\mathcal{L}_L^l \bar{\mathbf{X}}(\tau_L, 1) = \mathcal{L}_H^m \bar{\mathbf{X}}(\tau_L, 1) = \bar{\mathbf{X}}(\tau_L - 1, 1)$.

Now assume that the latent high frequency process $\{\bar{\mathbf{X}}(\tau_L, k)\}_{\tau_L}$ follows a VAR(p) with $p \in \mathbb{N} \cup \{\infty\}$:

$$\bar{\mathbf{X}}(\tau_L, k) = \sum_{l=1}^p \Phi_l \mathcal{L}_H^l \bar{\mathbf{X}}(\tau_L, k) + \bar{\eta}(\tau_L, k). \quad (4.1)$$

The coefficient matrix Φ_l is partitioned in the following manner:

$$\Phi_l = \begin{bmatrix} \Phi_{HH,l} & \Phi_{HL,l} \\ \Phi_{LH,l} & \Phi_{LL,l} \end{bmatrix}, \quad (4.2)$$

where $\Phi_{yz,l} \in \mathbb{R}^{K_y \times K_z}$ with $y, z \in \{H, L\}$. The error $\bar{\eta}(\tau_L, k)$ satisfies a HF martingale difference property similar to the LF based Assumption 2.1 in Section 2.⁹ It is therefore helpful to define a HF sigma field using a single-index version of $\bar{\mathbf{X}}(\tau_L, k)$. Simply write (4.1) as

$$\mathbf{Y}_t = \sum_{l=1}^p \Phi_l \mathbf{Y}_{t-l} + \xi_t, \quad (4.3)$$

where $\mathbf{Y}_t \in \mathbb{R}^{K^*}$ with $K^* = K_H + K_L$ is a single-index version of $\bar{\mathbf{X}}(\tau_L, k)$. One way of mapping (τ_L, k) to t is to let $t = m(\tau_L - 1) + k$ so that \mathbf{Y}_1 corresponds to $\bar{\mathbf{X}}(1, 1)$. The same mapping is used between ξ_t and $\bar{\eta}(\tau_L, k)$. Then $\xi_t = \bar{\eta}(\tau_L, k)$ is a stationary martingale difference with respect to the σ -field $\sigma(\mathbf{Y}_s : s \leq t)$, with variance $\mathbf{V} \equiv E[\bar{\eta}(\tau_L, k)\bar{\eta}(\tau_L, k)']$.

As stated in (2.1), a general linear aggregation scheme is considered: $\mathbf{x}_H(\tau_L) = \sum_{k=1}^m w_k \mathbf{x}_H(\tau_L, k)$ and $\mathbf{x}_L(\tau_L) = \sum_{k=1}^m w_k \mathbf{x}_L(\tau_L, k)$. By an application of Theorem 1 in Lütkepohl (1984), the

mixed frequency vector $\mathbf{X}(\tau_L)$ defined in (2.2) and the LF vector defined as

$$\underline{\mathbf{X}}(\tau_L) = [\mathbf{x}_H(\tau_L)', \mathbf{x}_L(\tau_L)']' \in \mathbb{R}^{K^*} \quad (4.4)$$

follow VARMA processes. We formally prove this in Theorem 4.1.

In order to bound the VARMA orders, we require a VAR representation of the latent HF process $\bar{\mathbf{X}}(\tau_L, k)$. Define $mK^* \times 1$ vectors:

$$\bar{\mathbf{X}}(\tau_L) = [\mathbf{X}(\tau_L, 1)', \dots, \mathbf{X}(\tau_L, m)']' \quad \text{and}$$

$$\bar{\eta}(\tau_L) = [\eta(\tau_L, 1)', \dots, \eta(\tau_L, m)']'.$$

The HF-VAR(p) process in (4.1) satisfies:

$$\mathbf{N} \bar{\mathbf{X}}(\tau_L) = \sum_{k=1}^s \mathbf{M}_k \bar{\mathbf{X}}(\tau_L - k) + \bar{\eta}(\tau_L), \quad (4.5)$$

where

$$\mathbf{N} = \begin{bmatrix} \mathbf{I}_{K^*} & \mathbf{0}_{K^* \times K^*} & \dots & \mathbf{0}_{K^* \times K^*} \\ -\Phi_1 & \mathbf{I}_{K^*} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0}_{K^* \times K^*} \\ -\Phi_{m-1} & \dots & -\Phi_1 & \mathbf{I}_{K^*} \end{bmatrix} \quad \text{and}$$

$$\mathbf{M}_k = \begin{bmatrix} \Phi_{km} & \Phi_{km-1} & \dots & \Phi_{(k-1)m+1} \\ \Phi_{km+1} & \Phi_{km} & \dots & \Phi_{(k-1)m+2} \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_{(k+1)m-1} & \Phi_{(k+1)m-2} & \dots & \Phi_{km} \end{bmatrix}$$

for $k = 1, \dots, s$. It is understood that $\Phi_k = \mathbf{0}_{K^* \times K^*}$ whenever $k > p$. We have:

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

where $\mathbf{N}_1 = \mathbf{I}_{K^*}$ and $\mathbf{N}_k = \sum_{l=1}^{k-1} \Phi_{k-l} \mathbf{N}_l$ for $k = 2, \dots, m$. Now rewrite (4.5) as:

$$\bar{\mathbf{A}}(\mathcal{L}_L) \bar{\mathbf{X}}(\tau_L) = \bar{\epsilon}(\tau_L), \quad (4.6)$$

where \mathcal{L}_L is the LF lag operator,

$$\bar{\mathbf{A}}(\mathcal{L}_L) \equiv \mathbf{I}_{mK^*} - \sum_{k=1}^s \bar{\mathbf{A}}_k \mathcal{L}_L^k \quad \text{and} \quad \bar{\mathbf{A}}_k = \mathbf{N}^{-1} \mathbf{M}_k, \quad (4.7)$$

and $\bar{\epsilon}(\tau_L) = \mathbf{N}^{-1} \bar{\eta}(\tau_L)$ is second order stationary and serially uncorrelated by the stationary martingale difference property of $\bar{\eta}(\tau_L)$.

The matrix $\bar{\mathbf{A}}(\mathcal{L}_L)$ is key for bounding the ARMA orders for low and mixed frequency processes. We now have the following result which we prove in Appendix B.

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 LF process is a VARMA(p_L, q_L). Moreover,

$$p_M \leq \deg[\det(\bar{\mathbf{A}}(\mathcal{L}_L))] \equiv g \quad \text{and} \quad p_L \leq g,$$

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

$$q_M \leq \max \{ \deg[\bar{\mathbf{A}}_{kl}(\mathcal{L}_L)] - g + p_M \mid k, l = 1, \dots, mK^* \},$$

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

$$q_L \leq \max \{ \deg[\bar{\mathbf{A}}_{kl}(\mathcal{L}_L)] - g + p_L \mid k, l = 1, \dots, mK^* \}.$$

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

⁹ Lütkepohl (1984) only requires the VAR error to be vector white noise. We impose the martingale difference assumption here for continuity with the paper in general.

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 $\{\mathbf{X}(\tau_L, k)\}$ is governed by a VAR(p) then the MF and LF processes $\{\underline{\mathbf{X}}(\tau_L)\}$ and $\{\mathbf{X}(\tau_L)\}$ have VARMA representations, and therefore VAR(∞) representations under the assumption of invertibility.¹⁰ Therefore, 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 4 (Stock Sampling With $p = 1$). Suppose that an underlying HF process follows a VAR(1) $\bar{\mathbf{X}}(\tau_L, k) = \Phi_1 \mathcal{L}_H^1 \bar{\mathbf{X}}(\tau_L, k) + \bar{\eta}(\tau_L, k)$ where $\bar{\eta}(\tau_L, k)$ is a stationary martingale difference with respect to the HF sigma field $\sigma(\mathbf{Y}_s : s \leq t)$, \mathbf{Y}_t is a single-index version of $\bar{\mathbf{X}}(\tau_L, k)$, and $\mathbf{V} \equiv E[\bar{\eta}(\tau_L, k)\bar{\eta}(\tau_L, k)']$.

It is easy to show that the corresponding MF process also follows a VAR(1) if we consider stock sampling:

$$\mathbf{X}(\tau_L) = \mathbf{A}_1 \mathbf{X}(\tau_L - 1) + \epsilon(\tau_L). \quad (4.8)$$

The parameter \mathbf{A}_1 is

$$\mathbf{A}_1 = \begin{bmatrix} \mathbf{0}_{K_H \times (m-1)K_H} & \Phi_{HH,1}^{[1]} & \Phi_{HL,1}^{[1]} \\ \vdots & \vdots & \vdots \\ \mathbf{0}_{K_H \times (m-1)K_H} & \Phi_{HH,1}^{[m]} & \Phi_{HL,1}^{[m]} \\ \mathbf{0}_{K_L \times (m-1)K_H} & \Phi_{LH,1}^{[m]} & \Phi_{LL,1}^{[m]} \end{bmatrix}, \quad (4.9)$$

where

$$\Phi_l^k = \begin{bmatrix} \Phi_{HH,l}^{[k]} & \Phi_{HL,l}^{[k]} \\ \Phi_{LH,l}^{[k]} & \Phi_{LL,l}^{[k]} \end{bmatrix}$$

with the matrix Φ_l is defined in Eq. (4.2). Moreover, by construction

$$\epsilon(\tau_L) = \begin{bmatrix} \sum_{k=1}^1 \left[\Phi_{HH,1}^{[1-k]} \Phi_{HL,1}^{[1-k]} \right] \eta(\tau_L, k) \\ \vdots \\ \sum_{k=1}^m \left[\Phi_{HH,1}^{[m-k]} \Phi_{HL,1}^{[m-k]} \right] \eta(\tau_L, k) \\ \sum_{k=1}^m \left[\Phi_{LH,1}^{[m-k]} \Phi_{LL,1}^{[m-k]} \right] \eta(\tau_L, k) \end{bmatrix},$$

and therefore $\epsilon(\tau_L)$ is a stationary martingale difference with respect to the MF sigma field $\sigma(\mathbf{X}(t) : t \leq \tau_L)$, where $\Omega \equiv E[\epsilon(\tau_L)\epsilon(\tau_L)']$ can be explicitly characterized as a function of Φ_1 and \mathbf{V} . The covariance matrix Ω has a block representation

$$\Omega = \begin{bmatrix} \Omega_{1,1} & \dots & \Omega_{1,m} & \Omega_{1,m+1} \\ \vdots & \ddots & \vdots & \vdots \\ \Omega'_{1,m} & \dots & \Omega_{m,m} & \Omega_{m,m+1} \\ \Omega'_{1,m+1} & \dots & \Omega'_{m,m+1} & \Omega_{m+1,m+1} \end{bmatrix} \in \mathbb{R}^{K^* \times K^*}, \quad (4.10)$$

¹⁰ VARMA representations discussed in Theorem 4.1 should not be mixed up with the (p, h) -autoregression (3.1)–(3.2). The latter is a redundant VARMA designed purely for multihorizon prediction, and it boils down to the original VAR(p) after canceling out the common roots in AR and MA components. The VARMA discussed in Theorem 4.1 is an outcome of temporal aggregation, and it is generally non-redundant although its analytical expression is unavailable (cfr. Lütkepohl (1984)).

with components

$$\Omega_{i,j} = \sum_{k=1}^i \begin{bmatrix} \Phi_{HH,1}^{[i-k]} & \Phi_{HL,1}^{[i-k]} \end{bmatrix} \mathbf{V} \begin{bmatrix} \Phi_{HH,1}^{[j-k]'} \\ \Phi_{HL,1}^{[j-k]'} \end{bmatrix} \quad (4.11)$$

for $i, j \in \{1, \dots, m\}$ and $i \leq j$,

$$\Omega_{i,i+1} = \sum_{k=1}^i \begin{bmatrix} \Phi_{HH,1}^{[i-k]} & \Phi_{HL,1}^{[i-k]} \end{bmatrix} \mathbf{V} \begin{bmatrix} \Phi_{LH,1}^{[m-k]'} \\ \Phi_{LL,1}^{[m-k]'} \end{bmatrix} \quad (4.11)$$

for $i \in \{1, \dots, m\}$

and

$$\Omega_{m+1,m+1} = \sum_{k=1}^m \begin{bmatrix} \Phi_{LH,1}^{[m-k]} & \Phi_{LL,1}^{[m-k]} \end{bmatrix} \mathbf{V} \begin{bmatrix} \Phi_{LH,1}^{[m-k]'} \\ \Phi_{LL,1}^{[m-k]'} \end{bmatrix}. \quad (4.12)$$

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

$$\underline{\mathbf{X}}(\tau_L) = \mathbf{A}_1 \underline{\mathbf{X}}(\tau_L - 1) + \underline{\epsilon}(\tau_L), \quad (4.13)$$

where

$$\mathbf{A}_1 = \Phi_1^m, \quad (4.14)$$

and $\underline{\epsilon}(\tau_L)$ is a stationary martingale difference with respect to the LF sigma field $\sigma(\underline{\mathbf{X}}(t) : t \leq \tau_L)$, with $\underline{\Omega} \equiv E[\underline{\epsilon}(\tau_L)\underline{\epsilon}(\tau_L)']$. Simply note $\underline{\epsilon}(\tau_L) = \sum_{k=1}^m \Phi_1^{m-k} \eta(\tau_L, k)$ to deduce the covariance matrix structure:

$$\underline{\Omega} = \sum_{k=1}^m \Phi_1^{m-k} \mathbf{V} (\Phi_1^{m-k})' \in \mathbb{R}^{K^* \times K^*}. \quad (4.15)$$

4.2. Causality and temporal aggregation

Anderson et al. (2015) find sufficient conditions for a full identification of HF-VAR coefficients based on MF data. If their conditions are satisfied, then recovery of HF causality is trivially feasible by looking at off-diagonal elements of the identified coefficients. Their conditions are stringent in reality, and when they are not satisfied the full identification of HF-VAR coefficients is generally impossible.

Contrary to Anderson et al. (2015), the research interest of this paper lies on a *partial* identification of HF-VAR coefficients. Since we are interested in Granger causality, identifying relevant off-diagonal elements suffices for us. Below we show that there are some cases where we can identify HF (non-)causality based on MF data if not the full HF-VAR coefficients.

Since the notion of Granger causality is associated with information sets, we need to define reference information sets for HF- and LF-VAR processes. Information sets are traditionally indexed with a single time unit, hence for added clarity we work with the single time index latent higher frequency variable $\mathbf{Y}_t \in \mathbb{R}^{K^*}$, $K^* = K_H + K_L$, defined in (4.3).

Recall from Section 3.1 that $\mathcal{I}(\tau_L)$ and $\underline{\mathcal{I}} = \{\mathcal{I}(\tau_L) \mid \tau_L \in \mathbb{Z}\}$ are the MF reference information set in period τ_L and the MF reference information set. Also recall the LF vector $\underline{\mathbf{X}}(\tau_L)$ from (4.4). Then the HF reference information set at time t and the HF reference information set are $\mathcal{I}(t) = \mathbf{Y}(-\infty, t]$ and $\underline{\mathcal{I}} = \{\mathcal{I}(t) \mid t \in \mathbb{Z}\}$. The prediction horizon for non-causality given $\underline{\mathcal{I}}$ is in terms of the high frequency, denoted by $\bar{h} \in \mathbb{Z}$. Similarly, the LF reference information set at time τ_L and the LF reference information set are $\underline{\mathcal{I}}(\tau_L) = \underline{\mathbf{X}}(-\infty, \tau_L]$ and $\underline{\underline{\mathcal{I}}} = \{\underline{\mathcal{I}}(\tau_L) \mid \tau_L \in \mathbb{Z}\}$.

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 causal 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 $\mathbf{x}_H \not\rightarrow \mathbf{x}_L | \mathcal{I}$, then $\mathbf{x}_H \not\rightarrow \mathbf{x}_L | \underline{\mathcal{I}}$. (ii) If $\mathbf{x}_L \not\rightarrow \mathbf{x}_H | \bar{\mathcal{I}}$, then $\mathbf{x}_L \not\rightarrow \mathbf{x}_H | \underline{\mathcal{I}}$.

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.¹¹

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

$$\Phi_1 = \begin{bmatrix} \phi_{HH,1} & 0 \\ \phi_{LH} & 0 \end{bmatrix}, \quad \text{and} \quad \Phi_2 = \begin{bmatrix} \phi_{HH,2} & 0 \\ 0 & 0 \end{bmatrix}.$$

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

$$\mathbf{A}_1 = \begin{bmatrix} \phi_{HH,2} & \phi_{HH,1} & 0 \\ \phi_{HH,1}\phi_{HH,2} & \phi_{HH,1}^2 + \phi_{HH,2} & 0 \\ \phi_{LH}\phi_{HH,2} & \phi_{LH}\phi_{HH,1} & 0 \end{bmatrix}, \quad (4.16)$$

while the LF coefficient is

$$\mathbf{A}_2 = \begin{bmatrix} \phi_{HH,1}^2 + \phi_{HH,2} & \phi_{HH,1}\phi_{HH,2}/\phi_{LH} \\ \phi_{LH}\phi_{HH,1} & \phi_{HH,2} \end{bmatrix}. \quad (4.17)$$

Eqs. (4.16) and (4.17) indicate that x_L does not cause x_H given \mathcal{I} , but x_L does cause x_H given $\underline{\mathcal{I}}$. therefore, we confirm that non-causality from all LF 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. \mathbf{x}_L for (i) and \mathbf{x}_H for (ii)) is not lost by temporal aggregation. Since both x_L and x_H are aggregated in the LF approach, neither high-to-low non-causality nor low-to-high non-causality in the HF data generating process can be recovered. Since only x_L is aggregated in the MF approach, low-to-high non-causality in the HF data generating process can be recovered (if not high-to-low non-causality). Hence, the MF approach is more desired than the LF approach especially when low-to-high (non-)causality is of interest.¹²

We conclude this subsection by again focusing on stock sampling with $p = 1$ as this particular case yields much sharper results.

¹¹ Theoretical results in the presence of auxiliary variables are seemingly intractable since potential causal chains may complicate causal patterns substantially.

¹² In practice, $x_L(\tau_L)$ is often announced with a large time lag. It may be announced even after $x_H(\tau_L + 1, 1), x_H(\tau_L + 1, 2), \dots$. One simple example would quarterly GDP versus monthly stock prices. In such a scenario low-to-high causality is less of interest than high-to-low causality since $x_L(\tau_L)$ and $x_H(\tau_L + 1, 1), x_H(\tau_L + 1, 2), \dots$ are effectively contemporaneous observations. In this sense the advantage of the MF approach over the LF approach should be interpreted with caution.

Example 5 (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 (4.8) and (4.13). Eq. (4.9) highlights what kind of causality information gets lost by switching from a HF- to MF-VAR. Similarly, (4.14) reveals the information loss when moving from a MF- to LF-VAR. This brings us to the following proposition.

Proposition 4.1. 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 $\not\rightarrow$ low) and Case 2 (high $\not\rightarrow$ low),
 - Non-causation up to HF horizon m given the HF information set $\bar{\mathcal{I}}$ implies non-causation at horizon 1 given the MF information set \mathcal{I} .
 - Non-causation at horizon 1 given \mathcal{I} is necessary and sufficient for non-causation at horizon 1 given the LF information set $\underline{\mathcal{I}}$.
- ii. In Case 3 (low $\not\rightarrow$ high) and Case 4 (high $\not\rightarrow$ high),
 - Non-causation up to HF horizon m given \mathcal{I} is necessary and sufficient for non-causation at horizon 1 given \mathcal{I} .
 - Non-causation at horizon 1 given \mathcal{I} implies non-causation at horizon 1 given $\underline{\mathcal{I}}$.
- iii. In Case I (all high $\not\rightarrow$ all low),
 - Non-causation up to HF horizon 1 given $\bar{\mathcal{I}}$ implies non-causation at horizon 1 given \mathcal{I} .
 - Non-causation at horizon 1 given \mathcal{I} is necessary and sufficient for non-causation at horizon 1 given $\underline{\mathcal{I}}$.
- iv. In Case II (all low $\not\rightarrow$ all high),
 - Non-causation up to HF horizon 1 given $\bar{\mathcal{I}}$ is necessary and sufficient for non-causation at horizon 1 given \mathcal{I} .
 - Non-causation at horizon 1 given \mathcal{I} implies non-causation at horizon 1 given $\underline{\mathcal{I}}$.

Proof. See Appendix D.

Note the difference between the two phrases repeatedly used in Proposition 4.1: “ A is necessary and sufficient for B ” and “ A implies B ”. The former means that A holds if and only if B holds. The latter means that B holds if A holds (but the converse is not true in general).

In Case 3 (low $\not\rightarrow$ high), Case 4 (high $\not\rightarrow$ high), and Case II (all low $\not\rightarrow$ all high), there is a “necessary and sufficient” relationship between non-causation given the HF information set and non-causation given the MF information set. We therefore do not lose any causal implications by switching from the HF information set to the MF information set. In contrast, non-causation given the MF information set implies (but is not implied by) non-causation given the LF information set. We do lose some causal implications by switching from the MF information set to the LF information set, suggesting an advantage of the MF approach.

In Case 1 (low $\not\rightarrow$ low), Case 2 (high $\not\rightarrow$ low), and Case I (all high $\not\rightarrow$ all low), we do not lose any causal implications by switching from the MF information set to the LF information set. We however lose some causal implications by switching from the HF information set to the MF information set. Hence, there are not a clear advantage of the MF approach over the LF approach in these cases (as far as stock sampling with $p = 1$ is concerned).

Summarizing these results, the MF causality test should not perform worse than the LF causality test in any case (apart from a parameter proliferation issue), 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.¹³

¹³ The authors thank a referee for pointing out that there remains a challenging yet important question after Section 4. The sampling frequency of our underlying

5. Local asymptotic power analysis

The goal of this section is to show that the MF causality tests have higher local asymptotic power than the LF causality tests. Comparing local power requires a simple DGP whose HF, MF, and LF representations are analytically tractable. We therefore have to restrict ourselves to 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.¹⁴

We first compute the local asymptotic power functions of MF and LF tests in Case I (i.e. causality from the high frequency variable to the LF variable). We then repeat the same steps for Case II (i.e. causality from the LF variable to the high frequency variable). Finally, we plot the local asymptotic power functions in a numerical exercise. Since we work with a HF process, define the HF sample size $T \equiv T_L \times m$.

Case I: high-to-low causality. In order to characterize local asymptotic power, assume that the high frequency DGP is given by:

$$\mathbf{X}(\tau_L, k) = \Phi(v/\sqrt{T}) \mathcal{L}_H \mathbf{X}(\tau_L, k) + \eta(\tau_L, k), \quad (5.1)$$

where

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

with $\rho_H, \rho_L \in (-1, 1)$, where $v \in \mathbb{R}$ is the usual Pitman drift parameter. Assume for computational simplicity that $\eta(\tau_L, k) \stackrel{i.i.d.}{\sim} (\mathbf{0}_{2 \times 1}, \mathbf{I}_2)$, hence $\mathbf{X}(\tau_L, k)$ has a strictly stationary solution and model (5.1) fully describes the causal structure of $\mathbf{X}(\tau_L, k)$. In the true DGP, the LF variable does not cause the high frequency variable, while for $v \neq 0$ the high frequency variable causes the LF variable with a marginal impact of v/\sqrt{T} which vanishes as $T \rightarrow \infty$. First note we have $p = h = 1$. We will therefore simplify notation, namely denote the least squares asymptotic covariance matrix $\Sigma_p(h)$ as Σ_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:

$$\mathbf{X}(\tau_L) = \mathbf{A}(v/\sqrt{T}) \mathbf{X}(\tau_L - 1) + \epsilon(\tau_L), \quad (5.2)$$

where

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

VAR process is set to be the observed sampling frequency of x_H (e.g. monthly VAR for inflation and GDP). One could technically begin with a VAR process of even higher sampling frequency (e.g. daily VAR for inflation and GDP). In that case, our starting point itself may be a wrong or highly restricted DGP because of temporal aggregation. It is an open question to recover (non-)causality in a latent VAR process where even x_H is not observed. A possible direction is to rely on state-space methods elaborated by Zadrozny (1988) and Harvey (1989) (see Foroni et al. (2013) for surveys).

¹⁴ In Section 3 we described the procedure of MF causality tests in general settings (i.e. an arbitrary number of variables and an arbitrary VAR lag order). It is admittedly a strong restriction that local power comparison is possible for bivariate VAR(1) only. The general framework in Section 3 is still worth keeping though, since it enables applied researchers to implement the MF causality test in broad applications.

and $\epsilon(\tau_L) \stackrel{i.i.d.}{\sim} (\mathbf{0}_{K \times 1}, \underline{\Omega})$. See (4.10)–(4.12) in Section 4.1 for a characterization of $\underline{\Omega}$. The MF-VAR(1) being estimated is:

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

with coefficient matrix $\mathbf{A} = \mathbf{A}(v/\sqrt{T})$. Table 1 and Theorem 3.3 provide us the Case I selection matrix \mathbf{R} to formulate the null hypothesis of high-to-low non-causality:

$$H_0^l : \mathbf{R} \text{ vec} [\mathbf{A}'] = \mathbf{0}_{m \times 1} \quad \text{where } \mathbf{R} \in \mathbb{R}^{m \times K^2}$$

therefore, the corresponding local alternatives $H_A^{l,L}$ are written as

$$H_A^{l,L} : \mathbf{R} \text{ vec} [\mathbf{A}'] = (v/\sqrt{T}) \mathbf{a},$$

where by (5.3) it follows \mathbf{a} is the $m \times 1$ vector $[0, \dots, 0, \sum_{k=1}^m \rho_H^{k-1} \rho_L^{m-k}]'$. Now let $\hat{\mathbf{A}}$ be the least squares estimator of \mathbf{A} . Theorem 3.2 implies that $W_{T_L^*}[H_0^l] \xrightarrow{d} \chi_m^2$ as $T \rightarrow \infty$ under H_0^l . Similarly, by classic arguments it is easy to verify under $H_A^{l,L}$ that $W_{T_L^*}[H_A^{l,L}] \xrightarrow{d} \chi_m^2(\kappa_{MF})$, where $\chi_m^2(\kappa_{MF})$ is the non-central chi-squared distribution with m degrees of freedom and non-centrality parameter κ_{MF} :

$$\kappa_{MF} = v^2 \mathbf{a}' [\mathbf{R} \Sigma_1 \mathbf{R}']^{-1} \mathbf{a}, \quad (5.4)$$

where Σ_1 is the asymptotic variance of $\hat{\mathbf{A}}$, in particular

$$\Sigma_1 = \underline{\Omega} \otimes \mathbf{\Upsilon}_0^{-1} \quad \text{with}$$

$$\mathbf{\Upsilon}_0 = \sum_{i=0}^{\infty} \mathbf{A}^i \underline{\Omega} \mathbf{A}^{i'} \quad \text{where } \mathbf{A} \equiv \lim_{T \rightarrow \infty} \mathbf{A}(v/\sqrt{T}). \quad (5.5)$$

The above equation can be obtained from non-local least squares asymptotics with $\mathbf{A} \equiv \lim_{T \rightarrow \infty} \mathbf{A}(v/\sqrt{T})$. See the technical appendix Ghysels et al. (2015) for a characterization Σ_1 in terms of underlying parameters. Using the discrete Lyapunov equation, $\mathbf{\Upsilon}_0$ can be characterized by:

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

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

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

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:

$$\underline{\mathbf{X}}(\tau_L) = \underline{\mathbf{A}}(v/\sqrt{T}) \underline{\mathbf{X}}(\tau_L - 1) + \underline{\epsilon}(\tau_L),$$

where

$$\underline{\mathbf{A}}(v/\sqrt{T}) = \begin{bmatrix} \rho_H^m & 0 \\ \sum_{k=1}^m \rho_H^{k-1} \rho_L^{m-k} (v/\sqrt{T}) & \rho_L^m \end{bmatrix} \quad (5.7)$$

and $\underline{\epsilon}(\tau_L) \stackrel{i.i.d.}{\sim} (\mathbf{0}_{2 \times 1}, \underline{\Omega})$. Note that $\underline{\Omega}$ is characterized in (4.15).

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

$$H_0^l : \mathbf{R} \text{ vec} [\underline{\mathbf{A}}'] = 0,$$

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

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

Let $\hat{\mathbf{A}}$ be the least squares estimator of $\underline{\mathbf{A}}$. We have that $W_{T_L^*}[H_0^I] \xrightarrow{d} \chi_1^2$ as $T \rightarrow \infty$ under H_0^I , while $W_{T_L^*}[H_A^{I,L}] \xrightarrow{d} \chi_1^2(\kappa_{LF})$ under $H_A^{I,L}$ with κ_{LF} given by:

$$\kappa_{LF} = \frac{\left(\nu \sum_{k=1}^m \rho_H^{k-1} \rho_L^{m-k} \right)^2}{\mathbf{R} \underline{\Sigma}_1 \mathbf{R}'},$$

where $\underline{\Sigma}_1$ is the asymptotic variance of $\hat{\mathbf{A}} \equiv \lim_{T \rightarrow \infty} \{\underline{\mathbf{A}}(\nu / \sqrt{T})\}$, in particular as in (5.5) it can be shown $\underline{\Sigma}_1 = \underline{\Omega} \otimes \underline{\Upsilon}_0^{-1}$ with $\underline{\Upsilon}_0 = \sum_{i=0}^{\infty} \underline{\mathbf{A}}^i \underline{\Omega} \underline{\mathbf{A}}^i'$. 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_1^2 and F_1 is the c.d.f. of $\chi_1^2(\kappa_{LF})$.

Case II: low-to-high causality. Assume that the true DGP is given by (5.1) with

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

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

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

$$\mathbf{A}(\nu / \sqrt{T}) = \begin{bmatrix} \mathbf{0}_{1 \times (m-1)} & \rho_H & \sum_{k=1}^1 \rho_H^{k-1} \rho_L^{1-k} (\nu / \sqrt{T}) \\ \vdots & \vdots & \vdots \\ \mathbf{0}_{1 \times (m-1)} & \rho_H^m & \sum_{k=1}^m \rho_H^{k-1} \rho_L^{m-k} (\nu / \sqrt{T}) \\ \mathbf{0}_{1 \times (m-1)} & 0 & \rho_L^m \end{bmatrix}. \quad (5.8)$$

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, the selection matrix \mathbf{R} is specified according to Case II in Section 3.3. Second, \mathbf{a} in the low-to-high case is defined as $\mathbf{a} = [\sum_{k=1}^1 \rho_H^{k-1} \rho_L^{1-k}, \dots, \sum_{k=1}^m \rho_H^{k-1} \rho_L^{m-k}]'$, while it was $\mathbf{a} = [0, \dots, 0, \sum_{k=1}^m \rho_H^{k-1} \rho_L^{m-k}]'$ in the high-to-low case.

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

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

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 $\mathbf{R} = [0, 1, 0, 0]$ here. A key difference between $\mathbf{A}(\nu / \sqrt{T})$ in (5.8) and $\underline{\mathbf{A}}(\nu / \sqrt{T})$ in (5.9) is that the $m - 1$ terms $\{\sum_{k=1}^1 \rho_H^{k-1} \rho_L^{1-k}, \dots, \sum_{k=1}^{m-1} \rho_H^{k-1} \rho_L^{(m-1)-k}\}$ have disappeared due to temporal aggregation. This information loss results in lower local power as seen below.

Numerical exercises. In order to study the local asymptotic power analysis more directly, we rely on some numerical calculations. In Fig. 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 $m \in \{3, \dots, 12\}$. 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), (5.7), and Proposition 4.1) and the former has more parameters: recall that \mathbf{A} is $(m + 1) \times (m + 1)$ while $\underline{\mathbf{A}}$ is 2×2 . 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 stock-sampling (or skip-sampling) a high frequency variable is less severe when it is more persistent. Persistent x_H means that $x_H(\tau_L, 1), \dots, x_H(\tau_L, m)$ take similar values. A single observation $x_H(\tau_L, m)$ therefore represents the previous $m - 1$ high frequency observations relatively well. Conversely, transitory x_H means that $x_H(\tau_L, m)$ may be far different from the previous $m - 1$ high frequency observations.

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 (5.8), (5.9), and Proposition 4.1, 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 stock-sampling 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 $(\rho_H, \rho_L) = (0.75, 0.75)$. In order to interpret this fact, let $\rho_H = \rho_L = \rho$ and consider a key quantity in the upper-right block of \mathbf{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 \mathbf{A} has $f(1), \dots, f(m)$ while that of $\underline{\mathbf{A}}$ has $f(m)$ only, therefore it is $\{f(1), \dots, f(m-1)\}$ 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. Fig. 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 $\rho = 0.75$. Therefore, the power ratio is decreasing in m for $\rho = 0.25$ and increasing in m for $\rho = 0.75$.

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 LF variable. Moreover, the power ratio increases in m for persistent series, while it decreases in m for transitory series.

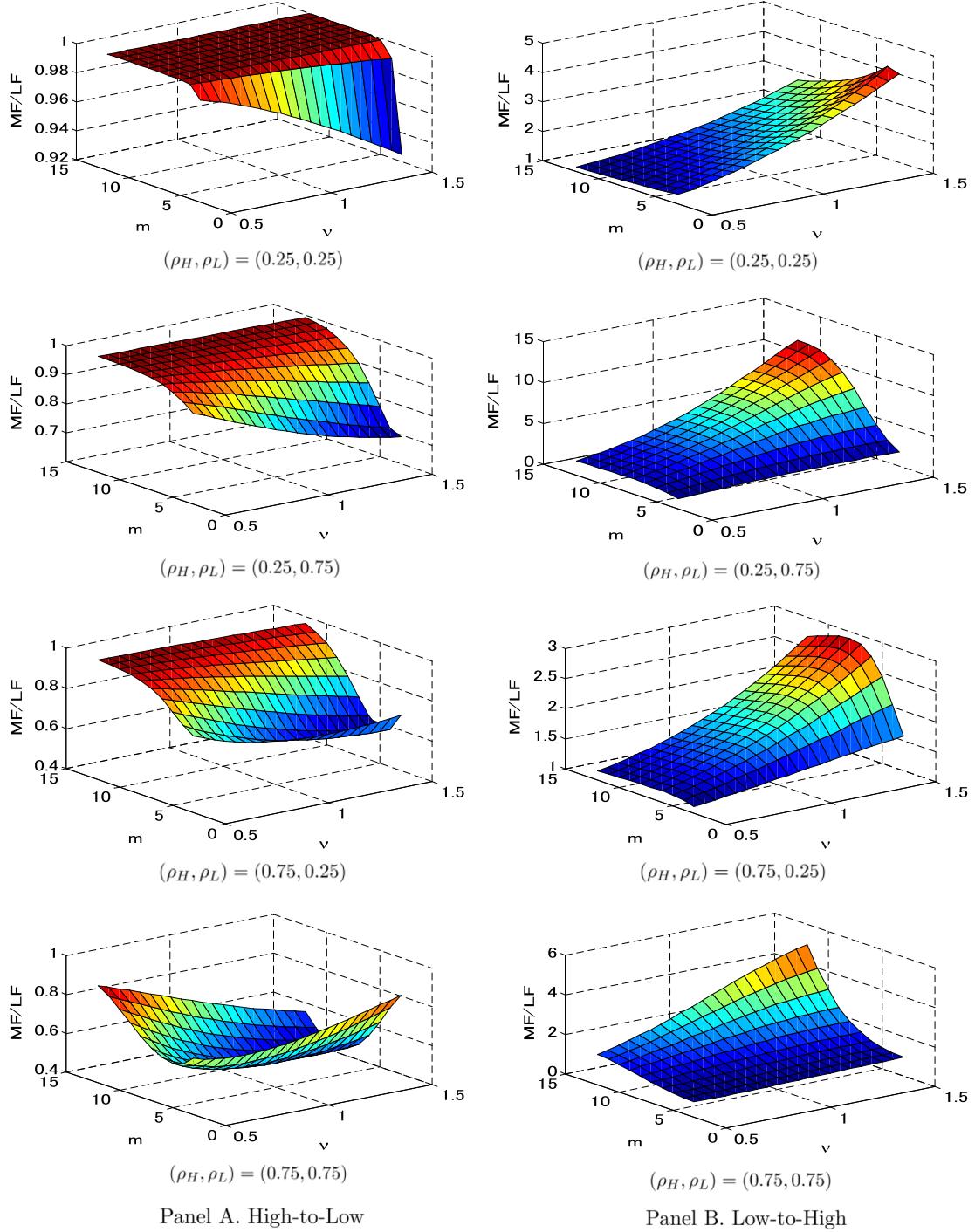


Fig. 1. Local asymptotic power of mixed and low frequency causality tests. The z-axis of each figure has the power ratio (i.e. the ratio of the local asymptotic power of the MF causality test to that of the LF causality test). Note that the scale of each z-axis is different. The x-axis has $v \in [0.5, 1.5]$, while the y-axis has $m \in \{3, \dots, 12\}$.

6. Power improvements in finite samples

This section conducts Monte Carlo simulations for bivariate cases and trivariate cases to evaluate the finite sample performance of the MF 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 spurious causality or non-causality, so our main exercise there is to evaluate the performance of the MF causality test itself

by checking empirical size and power. In particular, we will show that the MF causality test can capture causality chains under a realistic simulation design. All tests in this section are performed at the 5% level.

6.1. Bivariate case

This subsection considers a bivariate HF-VAR(1) process with stock sampling as in Section 5 so that the corresponding MF- and LF-VAR processes are known. One drawback of this experimental design is that we cannot easily study flow sampling

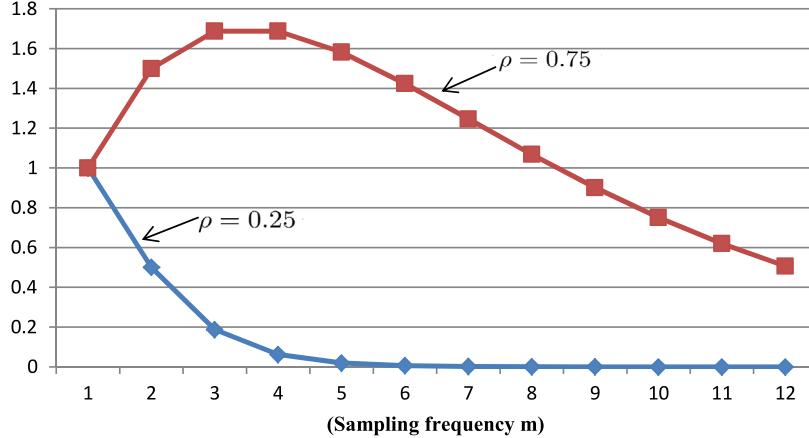


Fig. 2. Plot of the function $m\rho^{m-1}$ —driver of local asymptotic power ratios. The horizontal axis has $m \in \{1, \dots, 12\}$, while the vertical axis has $m\rho^{m-1}$ for $\rho \in \{0.25, 0.75\}$.

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 4.1.¹⁵

6.1.1. Simulation design

We draw J independent samples from a HF-VAR(1) process $\{\bar{\mathbf{X}}(\tau_L, k)\}$ according to (4.1) with Φ_1 partitioned in two possible ways:

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

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

therefore we have in (a) unidirectional causality from the high frequency variable to the LF variable and in (b) unidirectional causality from the LF 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 parameterization.

The innovations are either mutually and serially independent standard normal $\eta(\tau_L, k) \stackrel{i.i.d.}{\sim} N(\mathbf{0}_{2 \times 1}, \mathbf{I}_2)$, or follow multivariate GARCH since many macroeconomic and financial time series exhibit volatility clustering. The latter is best represented using the single-index representation of (4.1): $\mathbf{Y}_t = \Phi_1 \mathbf{Y}_{t-1} + \xi_t$. We assume $\xi_t = [\xi_{1t}, \xi_{2t}]'$ follows diagonal BEKK(1, 1), one of the most well-known form of multivariate GARCH (cfr. Engle and Kroner (1995)). Let $\mathbf{z}_t \stackrel{i.i.d.}{\sim} N(\mathbf{0}_{2 \times 1}, \mathbf{I}_2)$, then diagonal BEKK(1, 1) is written as $\xi_t = \mathbf{H}_t^{1/2} \mathbf{z}_t$ with

$$\mathbf{H}_t \equiv \begin{bmatrix} H_{11t} & H_{12t} \\ H_{21t} & H_{22t} \end{bmatrix} = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix} + \begin{bmatrix} 0.05 & 0 \\ 0 & 0.05 \end{bmatrix} \\ \times \begin{bmatrix} \xi_{1t-1}^2 & \xi_{1t-1}\xi_{2t-1} \\ \xi_{2t-1}\xi_{1t-1} & \xi_{2t-1}^2 \end{bmatrix} \begin{bmatrix} 0.05 & 0 \\ 0 & 0.05 \end{bmatrix} \\ + \begin{bmatrix} 0.9 & 0 \\ 0 & 0.9 \end{bmatrix} \begin{bmatrix} H_{11t-1} & H_{12t-1} \\ H_{21t-1} & H_{22t-1} \end{bmatrix} \begin{bmatrix} 0.9 & 0 \\ 0 & 0.9 \end{bmatrix}. \quad (6.1)$$

$\mathbf{H}_t^{1/2}$ is the lower-triangular Cholesky factor such that $\mathbf{H}_t \equiv \mathbf{H}_t^{1/2}(\mathbf{H}_t^{1/2})'$. The chosen parameter values are similar to those found in many macroeconomic and financial time series. While

(6.1) has a diagonal structure, the conditional covariance is still time-varying: $H_{12t} = 0.05^2 \xi_{1t-1} \xi_{2t-1} + 0.9^2 H_{12t-1}$. In view of i.i.d. normality for \mathbf{z}_t , the HF error process $\{\xi_t\}$ is stationary geometrically α -mixing (cfr. Boussama (1998)), hence MF and LF errors are also geometrically α -mixing.

The LF 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. When m takes a much larger value (e.g. $m = 12$ in month vs. year), our methodology loses practical applicability due to parameter proliferations. Handling a large m remains as a future research question.¹⁶

We aggregate the HF data into MF data $\{\mathbf{X}(\tau_L)\}_1^{T_L}$ and LF data $\{\bar{\mathbf{X}}(\tau_L)\}_1^{T_L}$ using stock sampling; see (2.2) and (4.4). We then fit MF-VAR(1) and LF-VAR(1), which are correctly specified. Finally, we compute Wald statistics for two separate null hypotheses of high-to-low non-causality $H_{H \rightarrow L}: x_H \not\rightarrow x_L$ and low-to-high non-causality $H_{L \rightarrow H}: x_L \not\rightarrow x_H$, each for horizon $h = 1$.¹⁷ The Wald statistic shown in (3.9) is computed by OLS with two covariance matrix estimators. The first one exploits the OLS variance estimator (3.11) based on the Bartlett kernel HAC estimator with bandwidth $n_{T_L^*} \equiv \max\{1, \lambda(T_L^*)^{1/3}\}$ 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.¹⁸ 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, \mathbf{A} and Ω in that expression should be replaced with $\underline{\mathbf{A}}$ and $\underline{\Omega}$, respectively (see (4.9), (4.10), (4.14) and (4.15)).

¹⁶ A recent work by Götz and Hecq (2014) attempts to resolve the parameter proliferation issue by using reduced-rank regressions and Bayesian estimation.

¹⁷ Note from (4.8) and (4.13) that $H_{H \rightarrow L}$ corresponds to $\mathbf{A}_1(m+1, 1 : m) = \mathbf{0}_{1 \times m}$ in the MF-VAR and to $\underline{\mathbf{A}}_1(2, 1) = 0$ in the LF-VAR models, while $H_{L \rightarrow H}$ corresponds to $\mathbf{A}_1(1 : m, m+1) = \mathbf{0}_{m \times 1}$ in the MF-VAR and to $\underline{\mathbf{A}}_1(1, 2) = 0$ in the LF-VAR models.

¹⁸ 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 (3.10) and the discussion leading to it. 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, hence we do not report them here.

¹⁵ In simulations not reported here we explored Lütkepohl and Poskitt's (1996) finite-order approximation for $\text{VAR}(\infty)$. The resulting test exhibited large empirical size distortions and was therefore not considered in this paper.

We circumvent size distortions for small samples $T_L \in \{50, 100\}$ by employing parametric bootstraps in Dufour et al. (2006) and Gonçalves and Kilian (2004).¹⁹ Dufour, Pelletier and Renault's (2006) procedure assumes i.i.d. errors with a known distribution while Gonçalves and Kilian's (2004) wild bootstrap does not require knowledge of the true error distribution and is robust to conditional heteroskedasticity of unknown form. Although $p = h = 1$ in this specific experiment, we present the bootstrap procedures with general p and h for completeness.

We use Dufour, Pelletier and Renault's (2006) [DPR] parametric bootstrap for the model with i.i.d. errors. The model with GARCH errors leads to greater size distortions, hence in that case we use Gonçalves and Kilian's (2004) [GK] wild bootstrap detailed below. The bootstrap procedure for the MF-VAR case is described below, the LF-VAR case being similar.

- Step 1** Fit an unrestricted MF-VAR(p) model for prediction horizon one to get $\hat{\mathbf{B}}(1)$ (which is simply $\hat{\mathbf{A}}_1, \dots, \hat{\mathbf{A}}_p$) and residuals $\{\hat{\epsilon}(1), \dots, \hat{\epsilon}(T_L^*)\}$ (cfr. (2.3) and (3.3)). Also fit an unrestricted MF-VAR(p) model for prediction horizon h to get $\hat{\mathbf{B}}(h)$ (cfr. (3.1)).
- Step 2** Using (3.9), compute the Wald test statistic based on the actual data, $W_{T_L^*}[H_0(h)]$.
- Step 3** Simulate N samples of artificial errors $\{\tilde{\epsilon}_i(1), \dots, \tilde{\epsilon}_i(T_L^*) | i = 1, \dots, N\}$. In the DPR bootstrap $\tilde{\epsilon}_i(\tau_L) \stackrel{i.i.d.}{\sim} N(\mathbf{0}_{K \times 1}, \hat{\Omega})$ with $\hat{\Omega} = (1/T_L^*) \sum_{\tau_L=1}^{T_L^*} \hat{\epsilon}(\tau_L) \hat{\epsilon}(\tau_L)'$. In the GK bootstrap $\tilde{\epsilon}_i(\tau_L) = \hat{\epsilon}(\tau_L) \circ \xi_i(\tau_L)$ with $\xi_i(\tau_L) \stackrel{i.i.d.}{\sim} N(\mathbf{0}_{K \times 1}, \mathbf{I}_K)$, where \circ means element-by-element multiplication.
- Step 4** For each replication $i \in \{1, \dots, N\}$, simulate the (p, h) -autoregression (3.1) using $\mathbf{B}(h) = \hat{\mathbf{B}}(h)$ with the null hypothesis $H_0(h)$ imposed. The way $H_0(h)$ is imposed on $\mathbf{B}(h)$ can be found in (3.13) and Table 1. The compound error for the i th replication is given by $\mathbf{u}_i^{(h)}(\tau_L) = \sum_{k=0}^{h-1} \hat{\Psi}_k \tilde{\epsilon}_i(\tau_L - k)$. $\hat{\Psi}_k$ is recursively computed from (3.2), using $\hat{\mathbf{A}}_1, \dots, \hat{\mathbf{A}}_p$. We denote by $W_i[H_0(h)]$ the Wald test statistic based on the i th simulated sample.
- Step 5** Finally, compute the resulting p-value $\hat{p}_N(W_{T_L^*}[H_0(h)])$, defined as

$$\begin{aligned} \hat{p}_N & \left(W_{T_L^*}[H_0(h)] \right) \\ &= \frac{1}{N+1} \left(1 + \sum_{i=1}^N I \left(W_i[H_0(h)] \geq W_{T_L^*}[H_0(h)] \right) \right). \end{aligned}$$

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

For small sample sizes $T_L \in \{50, 100\}$, we draw $J = 1000$ samples with $N = 499$ bootstrap replications. 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 Proposition 4.1 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 Tables 2–4 we report rejection frequencies. These three tables are different in terms of the error structure and bootstrap method: i.i.d. error with the DPR bootstrap in Table 2, i.i.d. error with the GK bootstrap in Table 3, and GARCH error with the GK bootstrap in Table 4. Also, the benchmark case with analytical covariance matrices is omitted in Tables 3 and 4 since the HAC case and the benchmark case produce very similar results as shown in Table 2. Finally, the large sample case $T_L = 500$ with i.i.d. errors and without bootstrap is omitted in Table 3 simply because that is covered in Table 2.

Note 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.

Consider the model with i.i.d. error and use of the DPR bootstrap: Table 2. Empirical size varies within [0.039, 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 frequencies for MF and LF causality tests for $H_{H \rightarrow L}$ 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 0.128 while power for the LF high-to-low causality test is 0.189.

Second, case (b) shows that the MF low-to-high causality test always has higher power than the LF counterpart. For example, the MF-based power is 0.415 and the LF-based power is 0.163 when $(m, T_L) = (3, 100)$ with HAC estimator. The difference between MF-based power and LF-based power 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.

The remaining simulation results are not too surprising. When Gonçalves and Kilian's (2004) bootstrap is used for i.i.d. errors, the rejection frequencies are similar to when i.i.d. normality is merely assumed. Rejection frequencies in the GARCH case are similar with those in the i.i.d. cases. Concerning low-to-high causality, the MF test again achieves higher power than the LF test in all cases.

6.2. Trivariate case

We now focus on a trivariate MF-VAR model with multiple prediction horizons in order to see if the MF causality test can capture causality chains properly. While there is no clear theory on how causality is linked between MF- and LF-VAR processes in the presence of causality chains, we also consider LF-VAR models with flow sampling and stock sampling for comparison.

6.2.1. Simulation design

Suppose that there are two high frequency variables X and Y and one LF variable Z with the ratio of sampling frequencies $m = 3$. Hence we have that $K_H = 2$, $K_L = 1$, and $K = mK_H + K_L = 7$. The LF sample size is $T_L = 100$. This setting matches with the empirical application in Section 7, where we analyze monthly inflation, monthly oil price changes, and quarterly real GDP growth covering 300 months (100 quarters, 25 years).

¹⁹ Chauvet et al. (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.

Table 2
Rejection frequencies (bivariate VAR with i.i.d. error and DPR bootstrap).

	Case (a)		Case (b)	
	$m = 2$	$m = 3$	$m = 2$	$m = 3$
Sample size $T_L = 50$ (DPR bootstrap)				
Size	MF: 0.063(0.059)	MF: 0.053(0.045)	MF: 0.056(0.055)	MF: 0.039(0.051)
	LF: 0.057(0.059)	LF: 0.063(0.054)	LF: 0.053(0.051)	LF: 0.044(0.046)
Power	MF: 0.128(0.155)	MF: 0.060(0.068)	MF: 0.241(0.266)	MF: 0.187(0.237)
	LF: 0.189(0.198)	LF: 0.072(0.088)	LF: 0.175(0.205)	LF: 0.097(0.110)
Sample size $T_L = 100$ (DPR bootstrap)				
Size	MF: 0.051(0.062)	MF: 0.045(0.040)	MF: 0.050(0.051)	MF: 0.060(0.069)
	LF: 0.047(0.056)	LF: 0.042(0.051)	LF: 0.049(0.050)	LF: 0.053(0.056)
Power	MF: 0.221(0.262)	MF: 0.098(0.120)	MF: 0.456(0.506)	MF: 0.415(0.454)
	LF: 0.311(0.338)	LF: 0.133(0.150)	LF: 0.323(0.340)	LF: 0.163(0.168)
Sample size $T_L = 500$				
Size	MF: 0.059(0.051)	MF: 0.064(0.051)	MF: 0.060(0.052)	MF: 0.066(0.052)
	LF: 0.056(0.052)	LF: 0.055(0.051)	LF: 0.056(0.050)	LF: 0.056(0.053)
Power	MF: 0.900(0.898)	MF: 0.414(0.390)	MF: 0.998(0.998)	MF: 0.997(0.997)
	LF: 0.943(0.944)	LF: 0.557(0.551)	LF: 0.943(0.944)	LF: 0.556(0.550)

Rejection frequencies at the 5% level for mixed and low frequency causality tests at horizon $h = 1$. The error term in the true DGP is i.i.d. Stock sampling is used when we aggregate MF data into LF data. 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. Dufour, Pelletier, and Renault's (2006) [DPR] 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 LF.

Table 3

Rejection frequencies (bivariate VAR with i.i.d. error and GK bootstrap).

	Case (a)		Case (b)	
	$m = 2$	$m = 3$	$m = 2$	$m = 3$
Sample size $T_L = 50$ (GK bootstrap)				
Size	MF: 0.071	MF: 0.037	MF: 0.061	MF: 0.049
	LF: 0.055	LF: 0.063	LF: 0.054	LF: 0.045
Power	MF: 0.135	MF: 0.076	MF: 0.216	MF: 0.161
	LF: 0.187	LF: 0.073	LF: 0.173	LF: 0.102
Sample size $T_L = 100$ (GK bootstrap)				
Size	MF: 0.054	MF: 0.038	MF: 0.050	MF: 0.066
	LF: 0.047	LF: 0.043	LF: 0.046	LF: 0.056
Power	MF: 0.238	MF: 0.117	MF: 0.435	MF: 0.386
	LF: 0.293	LF: 0.134	LF: 0.318	LF: 0.158

Rejection frequencies at the 5% level for mixed and low frequency causality tests at horizon $h = 1$. The error term in the DGP is i.i.d. Stock sampling is used when we aggregate MF data into LF data. 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. The HAC covariance estimator with Newey and West's (1994) automatic bandwidth selection is used. Gonçalves and Kilian's (2004) [GK] wild bootstrap is employed for $T_L \in \{50, 100\}$ to avoid size distortions. See Table 2 for the result with $T_L = 500$ and without bootstrap. m is the sampling frequency and T_L is the sample size in terms of LF.

Define a MF vector: $\mathbf{W}(\tau_L) = [X(\tau_L, 1), Y(\tau_L, 1), X(\tau_L, 2), Y(\tau_L, 2), X(\tau_L, 3), Y(\tau_L, 3), Z(\tau_L)]'$. Our true DGP is MF-VAR(1)²⁰:

$$\mathbf{W}(\tau_L) = \mathbf{A} \times \mathbf{W}(\tau_L - 1) + \boldsymbol{\epsilon}(\tau_L). \quad (6.2)$$

²⁰ We do not start with a HF-VAR process, unlike in Section 6.1. In the bivariate case there are clear theoretical results on how HF, MF, and LF processes are connected with each other. Section 6.1 therefore started with HF-VAR in order to investigate the impact of temporal aggregation on Granger causality. In trivariate cases, connection among HF, MF, and LF processes is generally unknown due to causality chains. Hence in this section we put more emphasis on the performance of the MF causality test itself than the impact of temporal aggregation.

Table 4
Rejection frequencies (bivariate VAR with GARCH error and GK bootstrap).

	Case (a)		Case (b)	
	$m = 2$	$m = 3$	$m = 2$	$m = 3$
Sample size $T_L = 50$ (GK bootstrap)				
Size	MF: 0.056	MF: 0.037	MF: 0.047	MF: 0.064
	LF: 0.064	LF: 0.064	LF: 0.050	LF: 0.052
Power	MF: 0.146	MF: 0.093	MF: 0.230	MF: 0.182
	LF: 0.183	LF: 0.083	LF: 0.195	LF: 0.094
Sample size $T_L = 100$ (GK bootstrap)				
Size	MF: 0.061	MF: 0.036	MF: 0.056	MF: 0.064
	LF: 0.062	LF: 0.047	LF: 0.047	LF: 0.049
Power	MF: 0.258	MF: 0.100	MF: 0.446	MF: 0.349
	LF: 0.318	LF: 0.143	LF: 0.311	LF: 0.133
Sample size $T_L = 500$				
Size	MF: 0.059	MF: 0.066	MF: 0.059	MF: 0.063
	LF: 0.054	LF: 0.056	LF: 0.055	LF: 0.055
Power	MF: 0.899	MF: 0.414	MF: 0.998	MF: 0.996
	LF: 0.941	LF: 0.557	LF: 0.941	LF: 0.556

Rejection frequencies at the 5% level for mixed and low frequency causality tests at horizon $h = 1$. The error term in the true DGP follows a multivariate GARCH process. Stock sampling is used when we aggregate MF data into LF data. 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. The HAC covariance estimator with Newey and West's (1994) automatic bandwidth selection is used. Gonçalves and Kilian's (2004) [GK] wild 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.

We assume the errors $\{\boldsymbol{\epsilon}(\tau_L)\}$ follow diagonal BEKK(1, 1) to match conditional volatility dynamics in macroeconomic and financial

²¹ We assume that $\boldsymbol{\epsilon}(\tau_L) = \mathbf{H}(\tau_L)^{1/2} \mathbf{z}(\tau_L)$ where $\mathbf{z}(\tau_L) \stackrel{i.i.d.}{\sim} \mathbf{N}(\mathbf{0}_{7 \times 1}, \mathbf{I}_7)$. As in (6.1), $\mathbf{H}(\tau_L) = 0.1\mathbf{I}_7 + 0.05^2 \boldsymbol{\epsilon}(\tau_L - 1)\boldsymbol{\epsilon}(\tau_L - 1)'$

²¹ In extra experiments not reported here, we also tried mutually and serially independent standard normal errors. Simulation results in the i.i.d. case and in the

+ 0.9² $\mathbf{H}(\tau_L - 1)$. We use Gonçalves and Kilian's (2004) bootstrap in order to control size.

The coefficient matrix \mathbf{A} in the DGP (6.2) is specified in two ways.

Case 1 (Unambiguously positive impact of X on Y.) In the first case we set \mathbf{A} as follows.

$$\mathbf{A} = \begin{bmatrix} 0.2 & 0 & -0.3 & 0 & 0.6 & 0 & 0 \\ \underline{0.3} & 0.3 & \underline{0.3} & -0.4 & \underline{0.4} & 0.5 & 0 \\ 0 & 0 & -0.2 & 0 & 0.4 & 0 & 0 \\ \underline{0} & 0 & \underline{0.2} & 0.2 & \underline{0.2} & 0.4 & 0 \\ 0 & 0 & 0 & 0 & 0.3 & 0 & 0 \\ \underline{0} & 0 & \underline{0} & 0 & \underline{0.3} & 0.3 & 0 \\ 0 & \underline{0.3} & 0 & \underline{0.3} & 0 & \underline{0.4} & 0.6 \end{bmatrix}, \quad (6.3)$$

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 bold elements represent the impact of Y on Z . All other non-zero elements are autoregressive coefficients, so not directly relevant for causal patterns. Eq. (6.3) therefore implies that there are only two channels of causality at $h = 1$: $X \rightarrow_1 Y | \mathcal{I}$ and $Y \rightarrow_1 Z | \mathcal{I}$. X has an unambiguously positive impact on Y , and Y has an unambiguously positive impact on Z .

Note from (6.3) 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 : $X \rightarrow_h Y | \mathcal{I}$, $Y \rightarrow_h Z | \mathcal{I}$, and $X \rightarrow_h Z | \mathcal{I}$ (cfr. Dufour and Renault (1998)). This point is verified by observing \mathbf{A}^2 and \mathbf{A}^3 :

$$\mathbf{A}^2 = \begin{bmatrix} 0.04 & 0 & 0 & 0 & 0.18 & 0 & 0 \\ \underline{0.15} & 0.09 & \underline{-0.14} & -0.04 & \underline{0.61} & 0.14 & 0 \\ 0 & 0 & 0.04 & 0 & 0.04 & 0 & 0 \\ \underline{0} & 0 & \underline{-0.08} & 0.04 & \underline{0.22} & 0.04 & 0 \\ 0 & 0 & 0 & 0 & 0.09 & 0 & 0 \\ \underline{0} & 0 & \underline{0} & 0 & \underline{0.18} & 0.09 & 0 \\ 0.09 & \underline{0.27} & \underline{0.15} & \underline{0} & \underline{0.30} & \underline{0.63} & 0.36 \end{bmatrix} \quad (6.4)$$

and

$$\mathbf{A}^3 = \begin{bmatrix} 0.01 & 0 & -0.01 & 0 & 0.08 & 0 & 0 \\ \underline{0.06} & 0.03 & \underline{0.00} & -0.03 & \underline{0.29} & 0.07 & 0 \\ 0 & 0 & -0.01 & 0 & 0.03 & 0 & 0 \\ \underline{0} & 0 & \underline{0.02} & -0.01 & \underline{0.05} & 0.03 & 0 \\ 0 & 0 & 0 & 0 & 0.03 & 0 & 0 \\ \underline{0} & 0 & \underline{0} & 0 & \underline{0.08} & 0.03 & 0 \\ 0.10 & \underline{0.19} & \underline{0.02} & \underline{-0.00} & \underline{0.50} & \underline{0.47} & 0.22 \end{bmatrix}. \quad (6.5)$$

In summary, X causes Y at $h = 1, 2, 3$; X does not cause Z at $h = 1$ but does cause Z at $h = 2, 3$; Y causes Z at $h = 1, 2, 3$. There does not exist any other causality.

Case 2 (Positive and negative impacts of X on Y.) In the second case we specify \mathbf{A} as follows.

$$\mathbf{A} = \begin{bmatrix} 0.2 & 0 & -0.3 & 0 & 0.6 & 0 & 0 \\ \underline{0.3} & 0.3 & \underline{0.3} & -0.4 & \underline{-0.4} & 0.5 & 0 \\ 0 & 0 & -0.2 & 0 & 0.4 & 0 & 0 \\ \underline{0} & 0 & \underline{0.2} & 0.2 & \underline{-0.2} & 0.4 & 0 \\ 0 & 0 & 0 & 0 & 0.3 & 0 & 0 \\ \underline{0} & 0 & \underline{0} & 0 & \underline{-0.3} & 0.3 & 0 \\ 0 & \underline{0.3} & 0 & \underline{0.3} & 0 & \underline{0.4} & 0.6 \end{bmatrix}. \quad (6.6)$$

The only difference between Case 1 and Case 2 is that the minus sign is put on the (2, 5), (4, 5), and (6, 5) elements. In Case 1 X

BEKK case are nearly identical, so we refrain from reporting the i.i.d. case in order to save space.

has an unambiguously positive impact on Y , while in Case 2 X has both positive and negative impacts on Y . Such a difference does not affect rejection frequencies of the MF causality test, but does affect rejection frequencies of the LF causality test. We will elaborate these aspects in Section 6.2.2.

\mathbf{A}^2 and \mathbf{A}^3 in Case 2 are as follows.

$$\mathbf{A}^2 = \begin{bmatrix} 0.04 & 0 & 0 & 0 & 0.18 & 0 & 0 \\ \underline{0.15} & 0.09 & \underline{-0.14} & -0.04 & \underline{-0.01} & 0.14 & 0 \\ 0 & 0 & 0.04 & 0 & 0.04 & 0 & 0 \\ \underline{0} & 0 & \underline{-0.08} & 0.04 & \underline{-0.06} & 0.04 & 0 \\ 0 & 0 & 0 & 0 & 0.09 & 0 & 0 \\ \underline{0} & 0 & \underline{0} & 0 & \underline{-0.18} & 0.09 & 0 \\ 0.09 & \underline{0.27} & \underline{0.15} & \underline{0} & \underline{-0.30} & \underline{0.63} & 0.36 \end{bmatrix}$$

and

$$\mathbf{A}^3 = \begin{bmatrix} 0.01 & 0 & -0.01 & 0 & 0.08 & 0 & 0 \\ \underline{0.06} & 0.03 & \underline{0.00} & -0.03 & \underline{-0.04} & 0.07 & 0 \\ 0 & 0 & -0.01 & 0 & 0.03 & 0 & 0 \\ \underline{0} & 0 & \underline{0.02} & -0.01 & \underline{-0.07} & 0.03 & 0 \\ 0 & 0 & 0 & 0 & 0.03 & 0 & 0 \\ \underline{0} & 0 & \underline{0} & 0 & \underline{-0.08} & 0.03 & 0 \\ 0.10 & \underline{0.19} & \underline{0.02} & \underline{-0.00} & \underline{-0.27} & \underline{0.47} & 0.22 \end{bmatrix}.$$

X causes Y at $h = 1, 2, 3$; X does not cause Z at $h = 1$ but does cause Z at $h = 2, 3$; Y causes Z at $h = 1, 2, 3$. There does not exist any other causality. All these aspects are shared by Cases 1 and 2 (note however that the degree of causality from X to Y at $h = 2, 3$ is weaker in Case 2 than in Case 1, since the positive and negative impacts at $h = 1$ offset each other when the model is iterated).

For each of Cases 1 and 2, we fit a (p, h) -autoregression with $p \in \{1, 2\}$ and $h \in \{1, 2, 3\}$ to implement the MF 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 $h = 2, 3$. Since the true DGP is MF-VAR(1), setting $p = 1$ should produce sharper results than setting $p = 2$. By comparing $p = 1$ and $p = 2$, we can see the sensitivity of the MF causality test to parameter proliferation. We draw $J = 1000$ samples and $N = 499$ bootstrap samples to avoid size distortions. Gonçalves and Kilian's (2004) [GK] bootstrap is used to account for conditional heteroskedasticity. The HAC covariance estimator with Newey and West's (1994) automatic bandwidth selection is used as in the bivariate simulation.

We consider both flow sampling and stock sampling in order to aggregate the MF data $\{\mathbf{W}(\tau_L)\}$ into LF. We then fit a trivariate LF (p, h) -autoregression with $p, h \in \{1, 2, 3\}$ and then repeat the individual Granger causality tests. Causal chains in the MF DGP make the corresponding LF DGP intractable. First, the true lag order of LF-VAR is unknown (and it is infinity in general). We therefore try $p = 1, 2, 3$ to see how rejection frequencies evolve as the model gets larger. Second, there is no theoretical conjecture on LF causal patterns. There is not a guarantee that the MF causal patterns are preserved. Hence the rejection frequencies we report below should be interpreted with caution.

As in the MF case, we draw $J = 1000$ samples and $N = 499$ GK bootstrap samples to avoid size distortions.

6.2.2. Simulation results

We report and discuss simulation results. We start with the MF cases and then move on to the LF cases.

Mixed frequency cases. Table 5 reports the rejection frequencies for the MF case. Empirical size lies in $[0.037, 0.060]$ when $p = 1$ and in $[0.040, 0.073]$ when $p = 2$ (recall that the rejection frequency for $X \rightarrow_1 Z$ should be interpreted as size, not power). We have fairly accurate size for both $p = 1$ and $p = 2$ due to the GK

Table 5

Rejection frequencies for trivariate MF-VAR.

Lag length \ Prediction horizon	$h = 1$	$h = 2$	$h = 3$
Case 1. Unambiguously positive impact of X on Y			
$p = 1$	$\begin{bmatrix} - & 0.053 & 0.052 \\ 0.993 & - & 0.057 \\ 0.058 & 1.000 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.050 & 0.041 \\ 0.763 & - & 0.057 \\ 0.520 & 0.988 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.043 & 0.060 \\ 0.118 & - & 0.054 \\ 0.649 & 0.716 & - \end{bmatrix}$
$p = 2$	$\begin{bmatrix} - & 0.048 & 0.054 \\ 0.973 & - & 0.048 \\ 0.073 & 0.993 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.060 & 0.060 \\ 0.647 & - & 0.051 \\ 0.463 & 0.942 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.059 & 0.057 \\ 0.114 & - & 0.055 \\ 0.481 & 0.463 & - \end{bmatrix}$
Case 2. Positive and negative impacts of X on Y			
$p = 1$	$\begin{bmatrix} - & 0.048 & 0.037 \\ 0.984 & - & 0.051 \\ 0.055 & 1.000 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.042 & 0.043 \\ 0.185 & - & 0.059 \\ 0.521 & 0.988 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.046 & 0.051 \\ 0.059 & - & 0.041 \\ 0.218 & 0.664 & - \end{bmatrix}$
$p = 2$	$\begin{bmatrix} - & 0.064 & 0.055 \\ 0.951 & - & 0.040 \\ 0.071 & 0.996 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.055 & 0.052 \\ 0.173 & - & 0.059 \\ 0.401 & 0.940 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.047 & 0.046 \\ 0.065 & - & 0.048 \\ 0.162 & 0.476 & - \end{bmatrix}$

Rejection frequencies at the 5% level based on MF (p, h)-autoregression with $p \in \{1, 2\}$ and $h \in \{1, 2, 3\}$, where we have two high frequency variables X and Y and one LF variable Z with $m = 3$. Each test deals with the null hypothesis of non-causality from an individual variable to another at horizon h . The upper-right triangular matrices have empirical size for $Y \not\rightarrow_h X$, $Z \not\rightarrow_h X$, and $Z \not\rightarrow_h Y$. Also, the rejection frequency for $X \not\rightarrow_1 Z$ is regarded as empirical size since X indeed does not cause Z at horizon 1. All other slots represent empirical power. In Case 1 X has an unambiguously positive impact on Y (cfr. Eq. (6.3)), while in Case 2 X has both positive and negative impacts on Y (cfr. Eq. (6.6)). The HAC covariance estimator with Newey and West's (1994) automatic bandwidth selection is used. The error term in the true DGP follows a multivariate GARCH process, and we use Gonçalves and Kilian's (2004) bootstrapped p-value. We draw $J = 1000$ samples and $N = 499$ bootstrap replications.

$$\text{Null Hypothesis: } \begin{bmatrix} - & Y \not\rightarrow_h X & Z \not\rightarrow_h X \\ X \not\rightarrow_h Y & - & Z \not\rightarrow_h Y \\ X \not\rightarrow_h Z & Y \not\rightarrow_h Z & - \end{bmatrix}.$$

bootstrap. Since size is well controlled in all cases, we can compare power meaningfully.

We first focus on Case 1, where X has an unambiguously positive impact on Y . Fix $p = 1$ first. Empirical power for the test of $X \not\rightarrow_h Y$ is 0.993, 0.763, and 0.118 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 (6.3)–(6.5).

Power for $Y \not\rightarrow_h Z$ vanishes more slowly as h increases: 1.000, 0.988, and 0.716 for horizons 1, 2, and 3, respectively. In fact the bold elements of \mathbf{A}^2 and \mathbf{A}^3 have 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 (6.3).

Rejection frequencies for $X \not\rightarrow_h Z$ are 0.058, 0.520, and 0.649 for horizons 1, 2, and 3, respectively. The rejection frequency 0.058 should be understood as size (not power) since there is *non-causality* from X to Z at horizon 1. The MF causality test properly captures the non-causality at horizon 1. For horizons 2 and 3, we have relatively high power for the indirect impact of X on Z via Y (see the underlined elements in (6.3)–(6.5)). Therefore, the MF causality test performs well even in the presence of a causality chain.

Increasing the lag length from $p = 1$ to $p = 2$ results in lower empirical power as expected. See $Y \not\rightarrow_h Z$ for example. Power for $Y \not\rightarrow_1 Z$ is 1.000 for $p = 1$ and 0.993 for $p = 2$ (0.007% decrease). Power for $Y \not\rightarrow_2 Z$ is 0.988 for $p = 1$ and 0.942 for $p = 2$ (4.656% decrease). Power for $Y \not\rightarrow_3 Z$ is 0.716 for $p = 1$ and 0.463 for $p = 2$ (35.34% decrease). Taking another example, power for $X \not\rightarrow_2 Z$ is 0.520 for $p = 1$ and 0.463 for $p = 2$ (10.96% decrease). Power for $X \not\rightarrow_3 Z$ is 0.649 for $p = 1$ and 0.481 for $p = 2$ (25.89% decrease). These results suggest that adding a redundant lag can have a large adverse impact on power, especially when prediction horizon h is large.

We now discuss Case 2, where X has both positive and negative impacts on Y . Fixing $p = 1$, power for $X \not\rightarrow_1 Y$ is 0.984, while it is 0.993 in Case 1. Hence, the presence of positive and negative impacts does not have a substantial impact on power at horizon $h = 1$. The power for $h = 2$ is 0.763 in Case 1 and 0.185 in Case

2. The power for $h = 3$ is 0.118 in Case 1 and 0.059 in Case 2. The lower power in Case 2 is a consequence of the lower degree of causality from X to Y at $h = 2, 3$ (compare \mathbf{A}^2 and \mathbf{A}^3 in both cases). In Case 2, the degree of causality at $h = 2, 3$ declines since the positive impact and negative impact at $h = 1$ offset each other when the model is iterated.

In general, the presence of positive and negative impacts does not lower power for $h = 1$ though does lower power for $h \geq 2$. Similar patterns can be observed for all pairs $X \not\rightarrow_h Y$, $Y \not\rightarrow_h Z$, and $X \not\rightarrow_h Z$.

Low frequency cases. Here we review the results for LF-VAR. See Table 6 for flow sampling and Table 7 for stock sampling. We first discuss Case 1, where X has an unambiguously positive impact on Y in the MF-VAR data generating process.

The LF test with flow sampling recovers the underlying causal patterns with high power. First, the rejection frequencies on $X \not\rightarrow_h Y$ are high at $h = 1$ but decay quickly. Second, the rejection frequencies on $Y \not\rightarrow_h Z$ are high and decay much more slowly. Third, the rejection frequency on $X \not\rightarrow_h Z$ is close to the nominal size 5% for $h = 1$ and soars for $h = 2, 3$. All these results resemble the MF case. Finally, empirical power decreases as the LF-VAR lag length increases from $p = 1$ to $p = 3$, which suggests that including one lag is enough to capture all causal patterns.

Turning on to stock sampling, there is an interesting difference from flow sampling. The rejection frequency on $X \not\rightarrow_1 Z$ is 0.060, 0.311, and 0.353 when p is 1, 2, and 3, respectively. Recall that, in the flow sampling, the rejection frequency on $X \not\rightarrow_1 Z$ is 0.052, 0.062, and 0.057 when p is 1, 2, and 3, respectively. We can therefore see that different aggregation schemes may produce different results. In this experiment, the stock-sampling test with more than one lag claims erroneously that X causes Z at horizon $h = 1$ (i.e. spurious causality).

Summarizing Case 1, the stock sampling case suffers from spurious causality while the flow sampling case does not. The reason for the superiority of flow-sampling test is that the underlying causal patterns have the unambiguous sign (all positive).

We next consider Case 2, where X has positive and negative impacts on Y . The flow-sampling test loses power for $X \not\rightarrow_h Y$ at

Table 6

Rejection frequencies for trivariate LF-VAR (flow sampling).

Lag length \ Prediction horizon	$h = 1$	$h = 2$	$h = 3$
Case 1. Unambiguously positive impact of X on Y			
$p = 1$	$\begin{bmatrix} - & 0.047 & 0.041 \\ 0.995 & - & 0.048 \\ 0.052 & 1.000 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.054 & 0.062 \\ 0.474 & - & 0.047 \\ 0.696 & 0.995 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.059 & 0.057 \\ 0.160 & - & 0.045 \\ 0.579 & 0.754 & - \end{bmatrix}$
$p = 2$	$\begin{bmatrix} - & 0.068 & 0.050 \\ 0.984 & - & 0.054 \\ 0.062 & 1.000 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.060 & 0.064 \\ 0.400 & - & 0.044 \\ 0.607 & 0.912 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.056 & 0.064 \\ 0.140 & - & 0.068 \\ 0.534 & 0.504 & - \end{bmatrix}$
$p = 3$	$\begin{bmatrix} - & 0.054 & 0.061 \\ 0.982 & - & 0.054 \\ 0.057 & 0.999 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.063 & 0.058 \\ 0.318 & - & 0.066 \\ 0.541 & 0.870 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.056 & 0.043 \\ 0.124 & - & 0.052 \\ 0.427 & 0.432 & - \end{bmatrix}$
Case 2. Positive and negative impacts of X on Y			
$p = 1$	$\begin{bmatrix} - & 0.059 & 0.041 \\ 0.053 & - & 0.063 \\ 0.040 & 1.000 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.050 & 0.055 \\ 0.091 & - & 0.065 \\ 0.051 & 0.974 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.054 & 0.051 \\ 0.044 & - & 0.056 \\ 0.059 & 0.634 & - \end{bmatrix}$
$p = 2$	$\begin{bmatrix} - & 0.058 & 0.042 \\ 0.101 & - & 0.053 \\ 0.060 & 1.000 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.044 & 0.050 \\ 0.102 & - & 0.053 \\ 0.075 & 0.925 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.040 & 0.040 \\ 0.061 & - & 0.050 \\ 0.072 & 0.571 & - \end{bmatrix}$
$p = 3$	$\begin{bmatrix} - & 0.067 & 0.046 \\ 0.088 & - & 0.055 \\ 0.053 & 0.999 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.064 & 0.059 \\ 0.082 & - & 0.054 \\ 0.074 & 0.880 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.042 & 0.055 \\ 0.048 & - & 0.062 \\ 0.063 & 0.464 & - \end{bmatrix}$

Rejection frequencies at the 5% level based on LF (p, h)-autoregression with $p, h \in \{1, 2, 3\}$, where we have two high frequency variables X and Y and one LF variable Z with $m = 3$. The high frequency variables X and Y are aggregated into LF using flow sampling. Each test deals with the null hypothesis of non-causality from an individual variable to another at horizon h . In Case 1 X has an unambiguously positive impact on Y (cfr. Eq. (6.3)), while in Case 2 X has both positive and negative impacts on Y (cfr. Eq. (6.6)). The HAC covariance estimator with Newey and West's (1994) automatic bandwidth selection is used. The error term in the true MF DGP follows a multivariate GARCH process, and we use Gonçalves and Kilian's (2004) bootstrapped p-value. We draw $J = 1000$ samples and $N = 499$ bootstrap replications.

$$\text{Null hypothesis: } \begin{bmatrix} - & Y \xrightarrow{h} X & Z \xrightarrow{h} X \\ X \xrightarrow{h} Y & - & Z \xrightarrow{h} Y \\ X \xrightarrow{h} Z & Y \xrightarrow{h} Z & - \end{bmatrix}.$$

Table 7

Rejection frequencies for trivariate LF-VAR (Stock Sampling).

Lag length \ Prediction horizon	$h = 1$	$h = 2$	$h = 3$
Case 1. Unambiguously positive impact of X on Y			
$p = 1$	$\begin{bmatrix} - & 0.047 & 0.042 \\ 0.780 & - & 0.050 \\ 0.060 & 0.990 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.065 & 0.057 \\ 0.284 & - & 0.046 \\ 0.602 & 0.995 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.040 & 0.058 \\ 0.087 & - & 0.040 \\ 0.769 & 0.828 & - \end{bmatrix}$
$p = 2$	$\begin{bmatrix} - & 0.053 & 0.067 \\ 0.689 & - & 0.053 \\ 0.311 & 0.981 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.053 & 0.059 \\ 0.238 & - & 0.043 \\ 0.615 & 0.958 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.051 & 0.060 \\ 0.068 & - & 0.063 \\ 0.725 & 0.609 & - \end{bmatrix}$
$p = 3$	$\begin{bmatrix} - & 0.055 & 0.049 \\ 0.601 & - & 0.061 \\ 0.353 & 0.956 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.070 & 0.038 \\ 0.176 & - & 0.055 \\ 0.563 & 0.928 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.052 & 0.046 \\ 0.078 & - & 0.051 \\ 0.611 & 0.537 & - \end{bmatrix}$
Case 2. Positive and negative impacts of X on Y			
$p = 1$	$\begin{bmatrix} - & 0.063 & 0.049 \\ 0.800 & - & 0.056 \\ 0.045 & 0.986 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.043 & 0.044 \\ 0.287 & - & 0.051 \\ 0.443 & 0.984 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.053 & 0.049 \\ 0.088 & - & 0.066 \\ 0.264 & 0.753 & - \end{bmatrix}$
$p = 2$	$\begin{bmatrix} - & 0.050 & 0.059 \\ 0.706 & - & 0.052 \\ 0.175 & 0.961 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.053 & 0.058 \\ 0.254 & - & 0.053 \\ 0.400 & 0.940 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.036 & 0.047 \\ 0.077 & - & 0.038 \\ 0.232 & 0.590 & - \end{bmatrix}$
$p = 3$	$\begin{bmatrix} - & 0.052 & 0.050 \\ 0.601 & - & 0.068 \\ 0.193 & 0.940 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.055 & 0.066 \\ 0.199 & - & 0.044 \\ 0.323 & 0.932 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.053 & 0.060 \\ 0.078 & - & 0.055 \\ 0.212 & 0.513 & - \end{bmatrix}$

Rejection frequencies at the 5% level based on LF (p, h)-autoregression with $p, h \in \{1, 2, 3\}$, where we have two high frequency variables X and Y and one LF variable Z with $m = 3$. The high frequency variables X and Y are aggregated into LF using stock sampling. Each test deals with the null hypothesis of non-causality from an individual variable to another at horizon h . In Case 1 X has an unambiguously positive impact on Y (cfr. Eq. (6.3)), while in Case 2 X has both positive and negative impacts on Y (cfr. Eq. (6.6)). The HAC covariance estimator with Newey and West's (1994) automatic bandwidth selection is used. The error term in the true MF DGP follows a multivariate GARCH process, and we use Gonçalves and Kilian's (2004) bootstrapped p-value. We draw $J = 1000$ samples and $N = 499$ bootstrap replications.

$$\text{Null hypothesis: } \begin{bmatrix} - & Y \xrightarrow{h} X & Z \xrightarrow{h} X \\ X \xrightarrow{h} Y & - & Z \xrightarrow{h} Y \\ X \xrightarrow{h} Z & Y \xrightarrow{h} Z & - \end{bmatrix}.$$

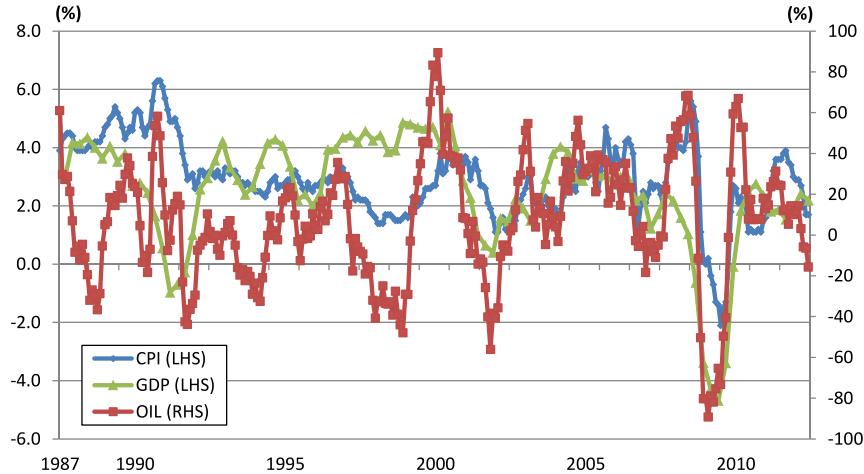


Fig. 3. Time series plot of CPI, OIL, and GDP. Year-to-year growth rates of U.S. monthly CPI, monthly spot West Texas Intermediate oil price, and quarterly real GDP. The sample period is July 1987 through June 2012, totaling 300 months (100 quarters, 25 years). The left-axis is for CPI and GDP and the right axis is for OIL.

any $h = 1, 2, 3$. As presented in Table 6, power for $X \rightarrow_1 Y$ is 0.053 for $p = 1$, 0.101 for $p = 2$, and 0.088 for $p = 3$; power for $X \rightarrow_2 Y$ is 0.091 for $p = 1$, 0.102 for $p = 2$, and 0.082 for $p = 3$; power for $X \rightarrow_3 Y$ is 0.044 for $p = 1$, 0.061 for $p = 2$, and 0.048 for $p = 3$. Taking a sample average of X and Y offsets the positive effect and negative effect of X on Y . The stock-sampling test, by comparison, has high power for $X \rightarrow_h Y$ (cfr. Table 7). For example, power for $X \rightarrow_1 Y$ is 0.800 for $p = 1$, 0.706 for $p = 2$, and 0.601 for $p = 3$.

Concerning $X \rightarrow_h Z$, the flow-sampling test suffers from spurious *non-causality* (or no power) at $h = 2, 3$, while the stock-sampling test suffers from spurious *causality* at $h = 1$. Since the latter phenomenon is already observed and discussed in Case 1, we focus on the spurious non-causality under flow sampling. As shown in Table 6, power for $X \rightarrow_2 Z$ is 0.051 for $p = 1$, 0.075 for $p = 2$, and 0.074 for $p = 3$; power for $X \rightarrow_3 Z$ is 0.059 for $p = 1$, 0.072 for $p = 2$, and 0.063 for $p = 3$. Hence, the flow-sampling test does not capture the underlying causality from X to Z at $h = 2, 3$.

Summarizing Case 2, the causality test under flow sampling loses power when positive effects and negative effects coexist in the underlying DGP. The stock sampling case suffers from spurious causality as in Case 1. Overall, LF causality tests often produce misleading results depending on aggregation schemes and underlying causal patterns. Case 1 has a relatively simple causal pattern so that the flow-sampling test performs as well as the MF test (stock-sampling test suffers from spurious causality). Case 2 has a more challenging causal pattern of mixed signs so that flow sampling generates spurious non-causality and stock sampling generates spurious causality. In practice we do not know what kind of causal pattern exists. It is therefore advised to take the MF approach in order to avoid spurious (non-)causality.

7. Empirical application

In this section we apply the MF causality test to U.S. macroeconomic data. We consider $100 \times$ annual log-differences of the U.S. monthly consumer price index for all items (CPI), monthly West Texas Intermediate spot oil price (OIL), and quarterly real GDP from July 1987 through June 2012 as an illustrative example. We use year-to-year growth rates to control for likely seasonality in each series. CPI, OIL and GDP data are made publicly available by the U.S. Department of Labor, Energy Information Administration, and Bureau of Economic Analysis, respectively.

The causal relationship between oil and the macroeconomy has been a major applied research area as surveyed in Hamilton (2008). See Payne (2010) for an extensive survey on the use of causality

Table 8
Sample statistics.

	Mean	Median	Std. dev.	Skewness	Kurtosis
CPI	2.913	2.900	1.316	-0.392	4.495
OIL	6.979	7.777	30.60	-0.312	3.485
GDP	2.513	2.783	1.882	-1.670	6.625

Sample statistics for $100 \times$ annual log-differences of monthly U.S. CPI, monthly spot West Texas Intermediate oil price, and quarterly real GDP. The sample period is July 1987 through June 2012.

tests to determine the relationship between energy consumption and economic growth. We introduce the MF concept into these literatures by analyzing CPI, OIL, and GDP. In particular, we expect significant causality from OIL to CPI since (i) oil products form a component of the all-item CPI and (ii) crude oil is one of the most important raw materials for a wide range of sectors (e.g. electricity, manufacturing, transportation, etc.).

Fig. 3 plots the three series, while Table 8 presents sample statistics. There is fairly strong positive correlation between CPI and OIL, and the latter is much more volatile than the former. The sample standard deviation is 1.316% for CPI and 30.60% for OIL. The sample correlation coefficient between these two is 0.512 with the 95% confidence interval based on the Fisher transformation being [0.423, 0.591]. Since CPI, OIL, and GDP have a positive sample mean of 2.913%, 6.979%, and 2.513%, we de-mean each series and fit VAR without a constant term. The sample kurtosis is 4.495 for CPI, 3.485 for OIL, and 6.625 for GDP. These figures suggest that the three series follow non-normal distributions, but note that the asymptotic theory of the MF causality test is free of the normality assumption (cfr. Section 2).

Using mean-centered $100 \times$ annual log-differenced data, we fit an unrestricted MF-VAR model with LF prediction horizon $h \in \{1, \dots, 5\}$ to monthly CPI, monthly OIL, and quarterly GDP. We therefore have $K_H = 2$, $K_L = 1$, $m = 3$, $K = 7$, $T_L = 100$, and $T = 300$. This setting matches the one used in trivariate simulation study in Section 6.2.

We choose the MF-VAR lag order to be one. Since the dimension of MF vector is $K = 7$, there are as many as 49 parameters even with the lag order one. As discussed in Section 6.2, including redundant lags has a large adverse impact on power (especially for larger prediction horizon h). Ghysels (forthcoming) proposes a variety of parsimonious specifications based on the MIDAS literature, but they involve nonlinear parameter constraints that may not be describe the true data generating process. The trade-off between unrestricted and restricted MIDAS regressions is discussed in Foroni et al. (2015). A general consensus is that the

unrestricted approach achieves higher prediction accuracy when m is small, such as monthly and quarterly ($m = 3$).

All six causal patterns ($\text{CPI} \rightarrow \text{OIL}$, $\text{CPI} \rightarrow \text{GDP}$, $\text{OIL} \rightarrow \text{GDP}$ and their converses) are tested. We use Newey and West's (1987) kernel-based HAC covariance estimator with Newey and West's (1994) automatic lag selection. In order to avoid potential size distortions and to allow for conditional heteroskedasticity of unknown form, we use Gonçalves and Kilian's (2004) bootstrap with $N = 999$ replications. See Section 6 for the details.

For the purpose of comparison, we also fit an unrestricted LF-VAR(4) model with LF prediction horizon $h \in \{1, \dots, 5\}$ to quarterly CPI, quarterly OIL, and quarterly GDP. Since parameter proliferation is less of an issue in LF-VAR, we let the lag order be 4 in order to take potential seasonality into account.

As discussed in the trivariate simulation study of Section 6.2, temporal aggregation can easily produce spurious (non-)causality. It is not surprising at all if empirical results based on MF-VAR and on LF-VAR are different from each other. It is of particular interest to see if these models detect significant causality from OIL to CPI, what we expect to exist from a theoretical point of view.

Table 9 presents bootstrapped p-values for MF and LF tests at each quarterly horizon $h \in \{1, \dots, 5\}$ (recall h is the *low frequency* prediction horizon). We denote whether rejection occurs at the 5% or 10% level. Note that the MF and LF approaches result in very different conclusions at standard levels of significance. At the 5% level, for example, the MF case reveals three significant causal patterns: CPI causes GDP at horizon 3, OIL causes CPI at horizons 1 and 4, and GDP causes CPI at horizon 1. The LF case, however, has two different significant causal patterns: CPI causes OIL at horizon 1 and OIL causes GDP at horizons 2 and 4.

Note that significant causality from OIL to CPI is found by the MF approach but not by the LF approach, whether the 5% level or 10% level is used. In this sense the former is producing a more intuitive result than the latter. Our result suggests that the quarterly frequency is too coarse to capture the OIL-to-CPI causality while the MF data contain enough information for us to capture it successfully.

None of the LF causal patterns appears in the MF results. For example, in the LF case CPI causes OIL at horizon 1 at the 5% level. The p-value is 0.035, roughly 1/10th the magnitude of the MF p-value. Similarly, OIL causes GDP in the LF case with p-values less than 1/10th the MF p-values.²² Based on our trivariate simulation study, the large difference between the MF results and LF results likely stems from spurious (non-)causality due to temporal aggregation. Choice of sampling frequencies therefore has a considerable impact on empirical applications.

8. 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.

²² In view of Theorem 3.2 of Hill (2007), our empirical results have some conflicts with causation theory. Focusing on the MF case in **Table 9**, the significant causation from CPI to GDP at horizon 3 implies a causal chain via OIL at least from a theoretical point of view. We do not observe significant causation from CPI to OIL or causation from OIL to GDP, however. The LF approach is facing a similar problem; OIL causes GDP at horizons 2 and 4 but we do not observe significant causation from OIL to CPI or causation from CPI to GDP. As noted in footnote 8, this sort of discrepancy stems from treating each prediction horizon h separately. Hill (2007) proposes a sequential multi-horizon test as a solution which in principle can be applied to the present MF context for small horizon causation tests, e.g. $h \leq 3$.

Table 9
Granger causality tests for CPI, OIL, and GDP.

h	Panel A. Mixed Frequency				
	1	2	3	4	5
CPI \rightarrow OIL	0.391	0.128	0.559	0.636	0.165
CPI \rightarrow GDP	0.195	0.098°	0.049°	0.100	0.180
OIL \rightarrow GDP	0.680	0.548	0.236	0.300	0.196
OIL \rightarrow CPI	0.002°	0.182	0.439	0.029°	0.605
GDP \rightarrow CPI	0.015°	0.570	0.583	0.125	0.500
GDP \rightarrow OIL	0.724	0.833	0.895	0.855	0.946

h	Panel B. Low Frequency				
	1	2	3	4	5
CPI \rightarrow OIL	0.035°	0.095°	0.095°	0.116	0.492
CPI \rightarrow GDP	0.380	0.215	0.272	0.238	0.683
OIL \rightarrow GDP	0.145	0.044°	0.088°	0.027°	0.066°
OIL \rightarrow CPI	0.206	0.320	0.986	0.710	0.521
GDP \rightarrow CPI	0.680	0.497	0.323	0.596	0.645
GDP \rightarrow OIL	0.095°	0.164	0.516	0.376	0.541

The MF approach uses monthly CPI, monthly OIL, and quarterly GDP. The LF approach uses all quarterly series. A $\boxed{\text{box}}$ indicates rejection at the 5% level of the null hypothesis of non-causality at the quarterly horizon $h \in \{1, \dots, 5\}$. A circle \circ indicates rejection at the 10% level. All data are mean centered annual log-differences. The sample period covers July 1987 through June 2012, which has 300 months (100 quarters, 25 years). We use Newey and West's (1987) kernel-based HAC covariance estimator with Newey and West's (1994) automatic lag selection, and Gonçalves and Kilian's (2004) bootstrapped p-value with $N = 999$ replications.

We rely on MF-VAR models to implement the new class of Granger causality tests.

We show that MF causality tests better recover causal patterns in an underlying high frequency process compared to the traditional LF approach. Moreover, we show formally that MF 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 MF-VAR approach works well for small differences in sampling frequencies like month versus quarter.

We apply the MF causality test to monthly consumer price index series, monthly crude oil prices, and quarterly real GDP in the U.S. We also apply the conventional LF causality test to the aggregated quarterly price series and the real GDP for comparison. These two approaches produce very different results at any standard level of significance. In particular, significant causality from oil prices to CPI is detected by the MF approach but not by the LF approach. This result suggests that the quarterly frequency is too coarse to capture such causality while the MF data contain enough information for us to capture that successfully.

Appendix A. Proofs of Theorems 3.1 and 3.2

Recall from (3.1) and (3.2) $\mathbf{u}^{(h)}(\tau_L) \equiv \sum_{k=0}^{h-1} \Psi_k \epsilon(\tau_L - k)$, and define:

$$\mathbf{Y}(\tau_L + h, p) \equiv \text{vec} [\mathbf{W}(\tau_L, p) \mathbf{u}^{(h)}(\tau_L + h)']$$

$$\mathbf{D}_{p,T_L^*}(h) \equiv \text{Var} \left[\frac{1}{\sqrt{T_L^*}} \sum_{\tau_L=0}^{T_L^*-1} \mathbf{Y}(\tau_L + h, p) \right] \quad \text{and}$$

$$\mathbf{D}_p(h) \equiv \lim_{T_L^* \rightarrow \infty} \mathbf{D}_{p,T_L^*}(h)$$

$$\Delta_{p,s}(h) \equiv E [\mathbf{Y}(\tau_L + h + s, p) \mathbf{Y}(\tau_L + h, p)'] .$$

The proof of **Theorem 3.1** exploits the following central limit theorem.

Lemma A.1. Under **Assumptions 2.1–2.3** $1/\sqrt{T_L^*} \sum_{\tau_L=0}^{T_L^*-1} \mathbf{Y}(\tau_L + h, p) \xrightarrow{d} N(\mathbf{0}_{pK^2 \times 1}, \mathbf{D}_p(h))$ where $\mathbf{D}_p(h)$ is positive definite.

Proof. By the Cramér–Wold theorem it is necessary and sufficient to show $1/\sqrt{T_L^*} \sum_{\tau_L=0}^{T_L^*-1} \mathbf{a}' \mathbf{Y}(\tau_L + h, p) \xrightarrow{d} N(0, \mathbf{a}' \mathbf{D}_p(h) \mathbf{a})$ for any conformable \mathbf{a} , $\mathbf{a}' \mathbf{a} = 1$. By construction, measurability and Assumptions 2.1–2.3 it follows $\{\mathbf{a}' \mathbf{Y}(\tau_L + h, p)\}_{\tau_L}$ is a zero mean, $L_{2+\delta}$ -bounded α -mixing process with coefficients that satisfy $\sum_{h=0}^{\infty} \alpha_h < \infty$.

Further, by the mds property of $\epsilon(\tau_L)$ it follows $\mathbf{u}^{(h)}(\tau_L) \equiv \sum_{k=0}^{h-1} \Psi_k \epsilon(\tau_L - k)$, and therefore $\mathbf{Y}(\tau_L + h, p) \equiv \text{vec}[\mathbf{W}(\tau_L, p) \mathbf{u}^{(h)}(\tau_L + h)']$, is uncorrelated at lag h , hence $\Delta_{p,s}(h) = \mathbf{0}_{pK^2 \times pK^2} \forall s \geq h$. See the technical appendix Ghysels et al. (2015) for a complete derivation of $\Delta_{p,s}(h)$. Therefore:

$$\begin{aligned} \mathbf{D}_{p,T_L^*}(h) &= \Delta_{p,0}(h) + \sum_{s=1}^{h-1} \left[1 - \frac{s}{T_L^*} \right] \times [\Delta_{p,s}(h) + \Delta_{p,s}(h)'] \\ &\rightarrow \Delta_{p,0}(h) + \sum_{s=1}^{h-1} \times [\Delta_{p,s}(h) + \Delta_{p,s}(h)'] = \mathbf{D}_p(h). \end{aligned}$$

Observe that $\mathbf{D}_{p,T_L^*}(h)$ for T_L^* sufficiently large is positive definite, hence $\mathbf{D}_p(h)$ is positive definite. Simply note that by stationarity and spectral density positiveness for $\mathbf{X}(\tau_L)$, it follows $\mathbf{a}' \mathbf{Y}(\tau_L + h, p) \mathbf{a}$ is for any conformable $\mathbf{a} \neq \mathbf{0}$, $\mathbf{a}' \mathbf{a} = 1$, stationary and has a continuous, bounded everywhere positive spectral density $f_a(\lambda)$. Therefore $\mathbf{a}' \mathbf{D}_{p,T_L^*}(h) \mathbf{a} = 2\pi f_a(0) + o(1) > 0$ for T_L^* sufficiently large (see Eq. (1.7) in Ibragimov (1962)). Therefore $1/\sqrt{T_L^*} \sum_{\tau_L=0}^{T_L^*-1} \mathbf{a}' \mathbf{Y}(\tau_L + h, p) / (\mathbf{a}' \mathbf{D}_{p,T_L^*}(h) \mathbf{a}) \xrightarrow{d} N(0, 1)$ by Theorem 2.2 in Ibragimov (1975). Since $\mathbf{a}' \mathbf{D}_{p,T_L^*}(h) \mathbf{a} \rightarrow \mathbf{a}' \mathbf{D}_p(h) \mathbf{a} > 0$ the claim now follows from Cramér's Theorem. \square

The proof of Theorem 3.2 requires consistency of the kernel variance estimator $\hat{\Sigma}_p = (\mathbf{I}_K \otimes \hat{\Gamma}_{p,0}^{-1}) \times \hat{\mathbf{D}}_{p,T_L^*}(h) \times (\mathbf{I}_K \otimes \hat{\Gamma}_{p,0}^{-1})$ in (3.11).

Lemma A.2. Under Assumptions 2.1–2.2 and 3.1 $\hat{\Sigma}_p$ is almost surely positive semi-definite for any $T_L^* \geq 1$, and $\hat{\Sigma}_p \xrightarrow{p} \Sigma_p$ where Σ_p is positive definite.

Proof. Almost sure positive semi-definiteness of $\hat{\Sigma}_p$ follows from almost sure positive definiteness of $\hat{\Gamma}_{p,0}$ under Assumptions 2.1–2.2, and almost sure positive semi-definiteness of $\hat{\mathbf{D}}_{p,T_L^*}(h)$ by Theorem 1 in Newey and West (1987). Further $\hat{\Gamma}_{p,0} \xrightarrow{p} \Gamma_{p,0}$ by the ergodic theorem given stationarity, ergodicity, and square integrability under Assumptions 2.1–2.3. Since $\Sigma_p = (\mathbf{I}_K \otimes \Gamma_{p,0}^{-1}) \times \mathbf{D}_p(h) \times (\mathbf{I}_K \otimes \Gamma_{p,0}^{-1})$ it therefore suffices to show $\hat{\mathbf{D}}_{p,T_L^*}(h) \xrightarrow{p} \mathbf{D}_p(h)$. The latter follows from Theorem 2.2 in de Jong and Davidson (2000) [JD] if we verify their Assumptions 1–4.

First, the Bartlett kernel satisfies JD's Assumption 1. Second, JD's Assumptions 2 and 3 hold since $n_{T_L^*} \rightarrow \infty$ as $T_L^* \rightarrow \infty$, $n_{T_L^*} = o(T_L^*)$, and by Assumptions 2.1–2.3 and 3.1 and given measurability, $\{1/\sqrt{T_L^*} \mathbf{Y}(\tau_L + h, p) : 1 \leq \tau_L \leq T_L^*\}_{T_L^* \geq 1}$ forms an $L_{2+\delta}$ -bounded α -mixing triangular array with coefficients $\alpha_h = O(h^{(4+\delta)/\delta})$.²³

Finally, in order to verify JD's Assumption 4, define the regression error function $\mathbf{u}^{(h)}(\tau_L, \tilde{\mathbf{B}}) \equiv \mathbf{X}(\tau_L) - \sum_{k=1}^p \tilde{\mathbf{A}}_k \mathbf{X}(\tau_L - h + 1 - k)$ for any conformable $\tilde{\mathbf{A}}_k$ where $\tilde{\mathbf{B}} \equiv [\tilde{\mathbf{A}}_1, \dots, \tilde{\mathbf{A}}_p]'$, and $\mathbf{Y}(\tau_L + h, p, \tilde{\mathbf{B}}) = (\mathbf{I}_K \otimes \mathbf{W}(\tau_L, p)) \mathbf{u}^{(h)}(\tau_L + h, \tilde{\mathbf{B}})$. Now define

$\mathbf{Z}(\tau_L + h, p, \tilde{\mathbf{B}}) \equiv \mathbf{Y}(\tau_L + h, p, \tilde{\mathbf{B}}) / \sqrt{T_L^*}$ and note $\hat{\mathbf{Y}}(\tau_L + h, p) = \mathbf{Y}(\tau_L + h, p, \hat{\mathbf{B}}(h))$. In order to match JD's standardization, we work with $\mathbf{Z}(\tau_L + h, p, \tilde{\mathbf{B}})$. Assumption 4 consists of three parts, (a)–(c), with a scale factor κ_n that is simply \mathbf{I}_{pK^2} in our case. Part (a) applies since $\hat{\mathbf{B}}(h)$ is $\sqrt{T_L^*}$ -convergent by Theorem 3.1. Next, (b) applies since under our assumptions and by model linearity $1/\sqrt{T_L^*} \sum_{\tau_L=0}^{T_L^*-1} E[(\partial/\partial \tilde{\mathbf{B}}) \mathbf{Z}(\tau_L + h, p, \tilde{\mathbf{B}})]$ is trivially continuous at $\mathbf{B}(h)$ uniformly in T_L^* . Finally, (c) involves a uniform LLN for $\mathbf{D}\mathbf{Z}(\tau_L + h, p, \tilde{\mathbf{B}}) \equiv (\partial/\partial \tilde{\mathbf{B}}) \mathbf{Z}(\tau_L + h, p, \tilde{\mathbf{B}})$. The latter is not a function of $\tilde{\mathbf{B}}$ in view of linearity (i.e. $\mathbf{D}\mathbf{Z}(\tau_L + h, p, \tilde{\mathbf{B}}) = \mathbf{D}\mathbf{Z}(\tau_L + h, p)$), hence a uniform LLN reduces to a pointwise LLN which holds by the ergodic theorem given stationarity, ergodicity, and integrability of $\mathbf{D}\mathbf{Z}(\tau_L + h, p)$ which follows from stationarity and square integrability of $\epsilon(\tau_L)$. \square

We now prove Theorems 3.1 and 3.2. Recall the least square expansion (3.6). By stationarity, ergodicity and square integrability $\hat{\Gamma}_{p,0} = 1/T_L^* \sum_{\tau_L=0}^{T_L^*-1} \mathbf{W}(\tau_L, p) \mathbf{W}(\tau_L, p)' \xrightarrow{p} E[\mathbf{W}(\tau_L, p) \mathbf{W}(\tau_L, p)'] = \Gamma_{p,0}$. Define $\mathbf{D}_p(h) \equiv \mathbf{D}_{p,T_L^*}(h) = \text{Var}[1/\sqrt{T_L^*} \sum_{\tau_L=0}^{T_L^*-1} \mathbf{Y}(\tau_L + h, p)]$. Now use $\Sigma_p \equiv (\mathbf{I}_K \otimes \Gamma_{p,0}^{-1}) \times \mathbf{D}_p(h) \times (\mathbf{I}_K \otimes \Gamma_{p,0}^{-1})'$, combined with Lemma A.1, and Slutsky's and Cramér's Theorems to deduce $\sqrt{T_L^*} \text{vec}[\hat{\mathbf{B}}(h) - \mathbf{B}(h)] \xrightarrow{d} N(\mathbf{0}_{pK^2 \times 1}, \Sigma_p)$. Finally, Σ_p is positive definite given the positive definiteness of $\Gamma_{p,0}$ and $\mathbf{D}_p(h)$ as discussed in the proof of Lemma A.1. This proves Theorem 3.1.

The proof of Theorem 3.2 follows instantly from Theorem 3.1, the assumption $\hat{\Sigma}_p \xrightarrow{p} \Sigma_p$, and the mapping theorem.

Appendix B. Proof of Theorem 4.1

In view of Theorem 1 in Lütkepohl (1984) it suffices to show that $\mathbf{X}(\tau_L)$ in (2.2) and $\underline{\mathbf{X}}(\tau_L)$ in (4.4) are linear transformations of a VAR process. Lütkepohl (1984) defines a VAR process as having a vector white noise error term, hence any subsequent VAR process need only have a second order stationary and serially uncorrelated error.

First, recall that $\{\bar{\mathbf{X}}(\tau_L)\}$ follows a VAR(s) process by (4.6). The proof is therefore complete if we show that $\mathbf{X}(\tau_L)$ and $\underline{\mathbf{X}}(\tau_L)$ are linear functions of $\{\bar{\mathbf{X}}(\tau_L)\}$. Recall the generic aggregation schemes (2.1) with selection vector \mathbf{w} . Define $\mathbf{H} = [\mathbf{I}_{K_H}, \mathbf{0}_{K_H \times K_L}]$, $\mathbf{L} = [\mathbf{0}_{K_L \times K_H}, \mathbf{I}_{K_L}]$, $\mathbf{F}_{M \rightarrow L} = [\mathbf{w}' \otimes \mathbf{I}_{K_H} \quad \mathbf{0}_{K_H \times K_L}]$, $\mathbf{F}_{H \rightarrow M} = [\mathbf{w} \otimes \mathbf{H}', \mathbf{w} \otimes \mathbf{L}']$, and

$$\mathbf{F}_{M \rightarrow L} = \begin{bmatrix} \mathbf{w}' \otimes \mathbf{I}_{K_H} & \mathbf{0}_{K_H \times K_L} \\ \mathbf{0}_{K_L \times m K_H} & \mathbf{I}_{K_L} \end{bmatrix}.$$

Then: $\mathbf{X}(\tau_L) = \mathbf{F}_{H \rightarrow M} \bar{\mathbf{X}}(\tau_L)$ and $\underline{\mathbf{X}}(\tau_L) = \mathbf{F}_{H \rightarrow L} \bar{\mathbf{X}}(\tau_L)$, where $\mathbf{F}_{H \rightarrow L} = \mathbf{F}_{M \rightarrow L} \mathbf{F}_{H \rightarrow M} = [\mathbf{w} \otimes \mathbf{H}', \mathbf{w} \otimes \mathbf{L}']'$. Moreover, in view of the transformation being a finite order, if $\bar{\mathbf{X}}(\tau_L)$ is stationary then so are $\mathbf{X}(\tau_L)$ and $\underline{\mathbf{X}}(\tau_L)$.

Appendix 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 $\bar{\mathcal{I}}(t)$ and the mapping between single time index t and double time indices (τ_L, k) is that $t = m(\tau_L - 1) + k$. Also recall our notation that $\tilde{\mathbf{x}}_H(\tau_L) = [\tilde{\mathbf{x}}_{H,1}(\tau_L)', \dots, \tilde{\mathbf{x}}_{H,K_H}(\tau_L)']'$ and $\tilde{\mathbf{x}}_{H,i}(\tau_L) = [x_{H,i}(\tau_L, 1), \dots, x_{H,i}(\tau_L, m)]'$. We have that:

$$\begin{aligned} P[\tilde{\mathbf{x}}_H(\tau_L + 1) | \mathcal{I}(\tau_L)] &= P[P[\tilde{\mathbf{x}}_H(\tau_L + 1) | \bar{\mathcal{I}}(m\tau_L)] | \mathcal{I}(\tau_L)] \\ &= P[P[\tilde{\mathbf{x}}_H(\tau_L + 1) | \bar{\mathcal{I}}_{(L)}(m\tau_L)] | \mathcal{I}(\tau_L)] \\ &= P[P[\tilde{\mathbf{x}}_H(\tau_L + 1) | \mathcal{I}_{(L)}(\tau_L)] | \mathcal{I}(\tau_L)] \\ &= P[\tilde{\mathbf{x}}_H(\tau_L + 1) | \mathcal{I}_{(L)}(\tau_L)]. \end{aligned}$$

²³ 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.

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 $\mathbf{x}_L \not\rightarrow \mathbf{x}_H \mid \bar{\mathcal{I}}$; and the third holds because $\bar{\mathcal{I}}_{(L)}(m\tau_L) = \mathcal{I}_{(L)}(\tau_L)$. Hence $\mathbf{x}_L \not\rightarrow \mathbf{x}_H \mid \underline{\mathcal{I}}$ as claimed.

Appendix D. Proof of Proposition 4.1

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

Case 1 (low $\not\rightarrow$ low). Suppose that x_{L,j_1} does not cause x_{L,j_2} up to high frequency horizon m given $\bar{\mathcal{I}}$ (i.e., $x_{L,j_1} \not\rightarrow_{(m)} x_{L,j_2} \mid \bar{\mathcal{I}}$). Then, $\Phi_{LL,1}^{[k]}(j_2, j_1) = 0$ for any $k \in \{1, \dots, m\}$ and hence x_{L,j_1} does not cause x_{L,j_2} at horizon 1 given $\underline{\mathcal{I}}$ (i.e., $x_{L,j_1} \not\rightarrow x_{L,j_2} \mid \underline{\mathcal{I}}$) in view of (4.9). 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

$$\Phi_1 = \begin{bmatrix} \phi_{HH} & 0.3 & \phi_{HL} \\ \phi_{LH} & 0.2 & \phi_{LL} \\ -0.1 & 0.1 & 0.1 \end{bmatrix},$$

where $\phi_{HH}, \phi_{HL}, \phi_{LH}$, and ϕ_{LL} are arbitrary coefficients. It is evident that $\Phi_{LL,1}(2, 1) = 0.1$ and $\Phi_{LL,1}^{[2]}(2, 1) = 0$. The former denies that $x_{L,j_1} \not\rightarrow_{(m)} x_{L,j_2} \mid \bar{\mathcal{I}}$, while the latter implies that $x_{L,j_1} \not\rightarrow x_{L,j_2} \mid \underline{\mathcal{I}}$.

Suppose now that $x_{L,j_1} \not\rightarrow x_{L,j_2} \mid \underline{\mathcal{I}}$. Then, $\Phi_{LL,1}^{[m]}(j_2, j_1) = 0$ and hence $x_{L,j_1} \not\rightarrow x_{L,j_2} \mid \underline{\mathcal{I}}$ in view of (4.14). The converse is also true.

Case 2 (high $\not\rightarrow$ low). Suppose that $x_{H,i_1} \not\rightarrow_{(m)} x_{L,j_1} \mid \bar{\mathcal{I}}$. Then, $\Phi_{LH,1}^{[k]}(j_1, i_1) = 0$ for any $k \in \{1, \dots, m\}$ and hence $x_{H,i_1} \not\rightarrow x_{L,j_1} \mid \underline{\mathcal{I}}$. The converse does not necessarily hold.

Suppose now that $x_{H,i_1} \not\rightarrow x_{L,j_1} \mid \underline{\mathcal{I}}$. Then, $\Phi_{LH,1}^{[m]}(j_1, i_1) = 0$ and hence $x_{H,i_1} \not\rightarrow x_{L,j_1} \mid \underline{\mathcal{I}}$. The converse is also true.

Appendix E. Supplementary data

Supplementary material related to this article can be found online at <http://dx.doi.org/10.1016/j.jeconom.2015.07.007>.

References

- Amemiya, T., Wu, R.Y., 1972. The effect of aggregation on prediction in the autoregressive model. *J. Amer. Statist. Assoc.* 67, 628–632.
- Anderson, B.D.O., Deistler, M., Felsenstein, E., Funovits, B., Koelbl, L., Zamani, M., 2015. Multivariate AR systems and mixed frequency data: G-identifiability and estimation. *Econometric Theory* (open access).
- Andrews, D.W.K., 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.
- Bańbura, M., Rünstler, G., 2011. A look into the factor model black box: Publication lags and the role of hard and soft data in forecasting GDP. *Int. J. Forecast.* 27, 333–346.
- Boussama, F., 1998. Ergodicity, Mixing and Estimation in GARCH Models (Ph.D. thesis). University of Paris.
- Breitung, J., Swanson, N.R., 2002. Temporal aggregation and spurious instantaneous causality in multiple time series models. *J. Time Series Anal.* 23, 651–665.
- Chauvet, M., Götz, T., Hecq, A., 2013. Realized Volatility and Business Cycle Fluctuations: A Mixed-Frequency VAR Approach, Working Paper, University of California Riverside and Maastricht University.
- Cox, D., 1981. Statistical analysis of time series: Some recent developments [with discussion and reply]. *Scand. J. Stat.* 8, 93–115.
- Davidson, J., 1994. Stochastic Limit Theory. Oxford University Press, New York.
- de Jong, R.M., Davidson, J., 2000. Consistency of kernel estimators of heteroscedastic and autocorrelated covariance matrices. *Econometrica* 68, 407–423.
- Doukhan, P., 1994. Mixing: Properties and Examples. Springer-Verlag.
- Dufour, J.M., Pelletier, D., Renault, E., 2006. Short run and long run causality in time series: Inference. *J. Econometrics* 132, 337–362.
- Dufour, J.M., Renault, E., 1998. Short run and long run causality in time series: Theory. *Econometrica* 66, 1099–1125.
- Engle, R.F., Kroner, K.F., 1995. Multivariate simultaneous generalized ARCH. *Econometric Theory* 11, 122–150.
- Eraker, B., Chiu, C.W.J., Foerster, A.T., Kim, T.B., Seoane, H.D., 2014. Bayesian mixed frequency VARs. *J. Financial Econometrics* (forthcoming).
- Florens, J.P., Mouchart, M., 1982. A note on noncausality. *Econometrica* 50, 583–591.
- Foroni, C., Ghysels, E., Marcellino, M., 2013. Mixed-frequency vector autoregressive models. In: Fomby, T.B., Kilian, L., Murphy, A. (Eds.), *VAR Models in Macroeconomics—New Developments and Applications: Essays in Honor of Christopher A. Sims*, vol. 32. Emerald Group Publishing Limited, pp. 247–272.
- Foroni, C., Marcellino, M., Schumacher, C., 2015. Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. *J. Roy. Statist. Soc. Ser. A* 178, 57–82.
- Ghysels, E., 2014. Macroeconomics and the reality of mixed frequency data. *J. Econometrics* (forthcoming).
- Ghysels, E., Hill, J.B., Motegi, K., 2015. Technical Appendix for ‘Testing for Granger Causality with Mixed Frequency Data’. Dept. of Economics, University of North Carolina.
- Ghysels, E., Miller, J.I., 2015. Testing for cointegration with temporally aggregated and mixed-frequency time series. *J. Time Series Anal.* (forthcoming).
- Ghysels, E., Santa-Clara, P., Valkanov, R., 2004. The MIDAS Touch: Mixed Data Sampling Regression Models, Working Paper, Dept. of Economics, UNC and UCLA.
- Ghysels, E., Santa-Clara, P., Valkanov, R., 2005. There is a risk-return trade-off after all. *J. Financial Economics* 76, 509–548.
- Giannone, D., Reichlin, L., Small, D., 2008. Nowcasting: The real-time informational content of macroeconomic data. *J. Monetary Economics* 55, 665–676.
- Gonçalves, S., Kilian, L., 2004. Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. *J. Econometrics* 123, 89–120.
- Götz, T., Hecq, A., 2014. Testing for Granger Causality in Large Mixed-Frequency VARs, Working Paper at Maastricht University.
- Götz, T.B., Hecq, A., Urbain, J.-P., 2013. Testing for common cycles in non-stationary VARs with varied frequency data. In: Fomby, T.B., Kilian, L., Murphy, A. (Eds.), *VAR Models in Macroeconomics—New Developments and Applications: Essays in Honor of Christopher A. Sims*, vol. 32. Emerald Group Publishing Limited, pp. 361–393.
- Granger, C.W.J., 1980. Testing for causality: A Personal Viewpoint. *J. Econom. Dynam. Control* 2, 329–352.
- Granger, C.W.J., 1988. Some recent developments in a concept of causality. *J. Econometrics* 39, 199–211.
- Granger, C.W.J., Lin, J.L., 1995. Causality in the long run. *Econometric Theory* 11, 530–536.
- Hamilton, J.D., 2008. In: Durlauf, S.N., Blume, L.E. (Eds.), *Oil and the Macroeconomy*, second ed. In: *The New Palgrave Dictionary of Economics*, vol. 6. Palgrave Macmillan, Basingstoke, Hampshire; New York, pp. 172–175.
- Harvey, A.C., 1989. Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press, Cambridge.
- Hill, J.B., 2007. Efficient tests of long-run causation in trivariate VAR processes with a rolling window study of the money-income relationship. *J. Appl. Econometrics* 22, 747–765.
- Ibragimov, I.A., 1962. Some limit theorems for stationary processes. *Theory Probab. Appl.* 7, 349–382.
- Ibragimov, I.A., 1975. A note on the central limit theorem for dependent random variables. *Theory Probab. Appl.* 20, 135–141.
- Kuzin, V., Marcellino, M., Schumacher, C., 2011. MIDAS vs. Mixed-frequency VAR: Nowcasting GDP in the Euro Area. *Int. J. Forecast.* 27, 529–542.
- Lewis, R., Reinsel, G.C., 1985. Prediction of multivariate time series by autoregressive model fitting. *J. Multivariate Anal.* 16, 393–411.
- Lütkepohl, H., 1984. Linear transformations of vector ARMA processes. *J. Econometrics* 26, 283–293.
- Lütkepohl, H., 1987. Forecasting Aggregated Vector ARMA Processes. Springer-Verlag, Berlin.
- Lütkepohl, H., 1993. Testing for causation between two variables in higher-dimensional VAR models. In: Schneeweis, H., Zimmermann, K. (Eds.), *Studies in Applied Econometrics*. Physica-Verlag, Heidelberg, pp. 75–91.
- Lütkepohl, H., Poskitt, D.S., 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. *J. Bus. Econom. Statist.* 17, 129–136.
- Mariano, R.S., Murasawa, Y., 2010. A coincident index, common factors, and monthly real GDP. *Oxford Bull. Econ. Stat.* 72, 27–46.
- McCrorie, J.R., Chambers, M.J., 2006. Granger causality and the sampling of economic processes. *J. Econometrics* 132, 311–336.
- McLeish, D.L., 1974. Dependent central limit theorems and invariance principles. *Ann. Probab.* 2, 620–628.
- Miller, J.I., 2014. Mixed-frequency cointegrating regressions with parsimonious distributed lag structures. *J. Financial Econometrics* 12, 584–614.
- Newey, W.K., West, K.D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.
- Newey, W.K., West, K.D., 1994. Automatic lag selection in covariance matrix estimation. *Rev. Econom. Stud.* 61, 631–653.
- Payne, J.E., 2010. Survey of the international evidence on the causal relationship between energy consumption and growth. *J. Economic Studies* 37, 53–95.
- Petersen, K., 1983. Ergodic Theory. Cambridge University Press.
- Renault, E., Sekkat, K., Szafarz, A., 1998. Testing for spurious causality in exchange rates. *J. Empirical Finance* 5, 47–66.
- Rosenblatt, M., 1956. A central limit theorem and a strong mixing condition. *Proc. Natl. Acad. Sci. USA* 42, 43–47.
- SaiKKonen, P., Lütkepohl, H., 1996. Infinite-order cointegrated vector autoregressive processes: Estimation and inference. *Econometric Theory* 12, 814–844.
- Salamaliki, P.K., Venetis, I.A., 2013. Energy consumption and real GDP in G-7: Multi-horizon causality testing in the presence of capital stock. *Energy Econ.* 39, 108–121.

- Seong, B., Ahn, S.K., Zadrozny, P.A., 2013. Estimation of vector error correction models with mixed frequency data. *J. Time Series Anal.* 34, 194–205.
- Silvestrini, A., Veredas, D., 2008. Temporal aggregation of univariate and multivariate time series models: A survey. *J. Economic Surveys* 22, 458–497.
- Yamamoto, T., Kurozumi, E., 2006. Tests for long-run granger non-causality in cointegrated systems. *J. Time Series Anal.* 27, 703–723.
- Zadrozny, P., 1988. Gaussian likelihood of continuous-time ARMAX models when data are stocks and flows at different frequencies. *Econometric Theory* 4, 108–124.
- Zellner, A., Montmarquette, C., 1971. A study of some aspects of temporal aggregation problems in econometric analyses. *Rev. Econ. Stat.* 53, 335–342.