Regression-Based Mixed Frequency Granger Causality Tests

Eric Ghysels*         Jonathan B. Hill†         Kaiji Motegi‡

First Draft: October 1, 2013
This Draft: August 14, 2014

Abstract

This paper proposes a new mixed frequency Granger causality test that achieves high power even when we have a small sample size and a large ratio of sampling frequencies. We postulate multiple parsimonious regression models where each model regresses a low frequency variable onto only one individual lag or lead of a high frequency variable. We then formulate what we call a max test statistic by picking the largest squared estimator among all parsimonious regression models. We show via Monte Carlo simulations that the max test is more powerful than existing mixed frequency Granger causality tests in small sample. In empirical application, the max test yields a plausible result that weekly U.S. interest rate spread used to cause U.S. real growth until about the year 2000 but such causality has vanished more recently.

Keywords: Granger causality test, Local asymptotic power, Max test, Mixed data sampling (MIDAS), Sims test, Temporal aggregation.

*Department of Economics and Department of Finance, Kenan-Flagler Business School, University of North Carolina at Chapel Hill. E-mail: eghysels@unc.edu
†Department of Economics, University of North Carolina at Chapel Hill. E-mail: jbhill@email.unc.edu
‡Faculty of Political Science and Economics, Waseda University. E-mail: motegi@aoni.waseda.jp
1 Introduction

Time series are often sampled at different frequencies, and it is well known that temporal aggregation may hide or generate Granger’s (1969) causality. Existing Granger causality tests typically ignore this issue and they merely aggregate data to the common lowest frequency, which may result in spurious non-causality or spurious causality. See Zellner and Montmarquette (1971) and Amemiya and Wu (1972) for early contributions. This subject has been extensively researched ever since, e.g. Granger (1980), Granger (1988), Lütkepohl (1993), Granger (1995), Renault, Sekkat, and Szafarz (1998), Marcellino (1999), Breitung and Swanson (2002), McCrorie and Chambers (2006), Silvestrini and Veredas (2008), among others.

One of the most popular Granger causality tests is a Wald test based on multi-step ahead vector autoregression (VAR) models. This approach can handle causal chains among more than two variables. See Lütkepohl (1993), Dufour and Renault (1998), Dufour, Pelletier, and Renault (2006), and Hill (2007). Since standard VAR models are designed for single-frequency data, these tests often suffer from the adverse effect of temporal aggregation. To alleviate this problem, Ghysels, Hill, and Motegi (2013) develop a set of Granger causality tests that explicitly take advantage of data sampled at mixed frequencies. They extend Dufour, Pelletier, and Renault’s (2006) VAR-based causality test using Ghysels’ (2012) mixed frequency vector autoregressive (MF-VAR) models. MF-VAR models avoid temporal aggregation by stacking all observations of high frequency variables.\(^1\)

Ghysels, Hill, and Motegi’s (2013) tests have low power when the ratio of sampling frequencies \(m\) is large and sample size is small. An essential reason for the low power is that the dimension of MF-VAR models soars as \(m\) increases, resulting in parameter proliferation. Imposing parametric restrictions might be one simple solution, but there is always a chance of misspecification. It is thus desired to establish a new test that imposes no parametric constraints and still achieves high power for large \(m\) and small sample size. Such a contribution would be especially relevant for applied macroeconomics, where we tend to have a small sample size and Granger causality has been of great interest since the empirical study of Sims (1972) and Sims (1980).

The present paper proposes a regression-based mixed frequency Granger causality test that is based on Sims’ (1972) two-sided regression. We postulate multiple parsimonious regression models where the \(j\)-th model regresses a low frequency variable \(x_L\) onto the only \(j\)-th lag or lead of a high frequency variable \(x_H\). Our test statistic is the maximum among squared estimators scaled and weighted properly. In this sense we call it the max test for short.

While the max test statistic follows a non-standard asymptotic distribution under the null hypothesis of Granger non-causality, a simulated \(p\)-value is readily available through an arbitrary number of draws from the null distribution. The max test is thus straightforward to implement in practice.

\(^1\)MIDAS, standing for Mi(xed) Da(ta) S(ampling), regression models have been put forward in recent work by Ghysels, Santa-Clara, and Valkanov (2004), Ghysels, Santa-Clara, and Valkanov (2006), and Andreou, Ghysels, and Kourtellos (2010). See Andreou, Ghysels, and Kourtellos (2011) and Armesto, Engemann, and Owyang (2010) for surveys. VAR models for mixed frequency data were independently introduced by Anderson, Deistler, Felsenstein, Funovits, Zadrozny, Eichler, Chen, and Zamani (2012), Ghysels (2012), and McCracken, Owyang, and Sekhposyan (2013). An early example of related ideas appears in Friedman (1962). Foroni, Ghysels, and Marcellino (2013) provide a survey of mixed frequency VAR models and related literature.
Through local asymptotic power analysis and Monte Carlo simulations, we compare the max test based on mixed frequency data (henceforth "MF max test"), a Wald test based on mixed frequency data ("MF Wald test"), the max test based on low frequency data ("LF max test"), and a Wald test based on low frequency data ("LF Wald test"). It will turn out that MF tests are more robust against complex (but realistic) causal patterns than LF tests in both local asymptotics and finite sample. The MF max test and the MF Wald test are roughly as powerful as each other in local asymptotics, but the former is clearly more powerful than the latter in finite sample.

For Granger causality from $x_H$ to $x_L$, we prove the consistency of MF max test. We also show by counter-examples that LF tests are not consistent. For Granger causality from $x_L$ to $x_H$, proving the consistency of MF max test remains as an open question.

As an empirical application, we conduct a rolling window analysis on weekly interest rate spread and real GDP growth in the U.S. The MF max test yields an intuitive result that the weekly spread used to cause real growth until about the year 2000 but such causality has vanished more recently.

This paper is organized as follows. Section 2 derives the max test and derives its asymptotic properties. In Section 3 we conduct local power analysis. In Section 4 we run Monte Carlo simulations. Section 5 presents the empirical application on interest rate spread and GDP. Section 6 concludes the paper. All tables and figures are collected after Section 6. Proofs for all theorems as well as some theoretical details are provided in Technical Appendices.

2 Methodology

This paper focuses on a bivariate case where we have a high frequency variable $x_H$ and a low frequency variable $x_L$. A trivariate case should await future research since it involves an extra complexity of causal chains (see Dufour and Renault (1998), Dufour, Pelletier, and Renault (2006) and Hill (2007)).

To discuss Granger causality between $x_H$ and $x_L$, we need to formulate a data generating process (DGP) governing these variables. For each low frequency time period $\tau_L \in \mathbb{Z}$, we have $m$ high frequency time periods. $m$ is often called the ratio of sampling frequencies. We sequentially observe $\{x_H(\tau_L, 1), \ldots, x_H(\tau_L, m), x_L(\tau_L)\}$ in a period $\tau_L$. A simple example would be a month vs. quarter case, where $m = 3$ since each quarter has three months. $x_H(\tau_L, 1)$ is the first monthly observation of $x_H$ in quarter $\tau_L$, $x_H(\tau_L, 2)$ is the second, and $x_H(\tau_L, 3)$ is the third. We then observe $x_L(\tau_L)$, the quarterly observation of $x_L$. The assumption that $x_L(\tau_L)$ is observed after $x_H(\tau_L, m)$ is just by convention. See Figure 1 for a visual explanation of these notations.

**Example 2.1** (Mixed Frequency Data in Economic Applications). 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 unemployment, oil prices, inflation, interest rates, etc. aggregated into quarters (see Hill (2007) for references). Consider monthly CPI inflation $[x_H(\tau_L, 1), x_H(\tau_L, 2), x_H(\tau_L, 3)]'$ in quarter $\tau_L$. According to the Bureau of Economic Analysis, GDP is announced in advance roughly one month after the quarter, with subsequent updates over the following two months (e.g. the 2014 first quarter advanced estimate is due on April 30, 2014). By comparison, the monthly CPI is
announced roughly three weeks after the month. Since the CPI inflation is announced before the GDP, 
\{x_H(\tau_L, 1), x_H(\tau_L, 2), x_H(\tau_L, 3), x_L(\tau_L)\} is an appropriate order.

The ratio of sampling frequencies, \(m\), depends on \(\tau_L\) in some applications like week vs. month, 
where \(m\) is four or five. This paper postpones such a case to the future work since time-dependent \(m\) 
complicates our statistical theory substantially.

We collect all observations in period \(\tau_L\) to define a \(K \times 1\) mixed frequency vector \(X(\tau_L) = [x_H(\tau_L, 1), \ldots, x_H(\tau_L, m), x_L(\tau_L)]^t\). We have that \(K = m + 1\) since we are considering a bivariate case with time-independent \(m\). Define the filtration \(\mathcal{F}_{\tau_L} = \sigma(X(\tau) : \tau \leq \tau_L)\). Following Ghysels (2012) and Ghysels, Hill, and Motegi (2013), we assume that \(E[X(\tau_L)|\mathcal{F}_{\tau_L-1}]\) has a version that is almost surely linear in 
\{\(X(\tau_L-1), \ldots, X(\tau_L - p)\}\} for some finite \(p \geq 1\).

**Assumption 2.1.** The mixed frequency vector \(X(\tau_L)\) is governed by MF-VAR\((p)\) for some finite \(p \geq 1\):

\[
\begin{bmatrix}
    x_H(\tau_L, 1) \\
    \vdots \\
    x_H(\tau_L, m) \\
    x_L(\tau_L) \\
\end{bmatrix} \overset{\text{def}}{=} \sum_{k=1}^{p} \begin{bmatrix}
    d_{11,k} & \cdots & d_{1m,k} & c_{(k-1)m+1} \\
    \vdots & \ddots & \vdots & \vdots \\
    d_{m1,k} & \cdots & d_{mm,k} & c_{km} \\
    b_{km} & \cdots & b_{(k-1)m+1} & a_k \\
\end{bmatrix} \begin{bmatrix}
    x_H(\tau_L - k, 1) \\
    \vdots \\
    x_H(\tau_L - k, m) \\
    x_L(\tau_L - k) \\
\end{bmatrix} + \begin{bmatrix}
    \epsilon_H(\tau_L, 1) \\
    \vdots \\
    \epsilon_H(\tau_L, m) \\
    \epsilon_L(\tau_L) \\
\end{bmatrix}
\]

or compactly \(X(\tau_L) = \sum_{k=1}^{p} A_k X(\tau_L - k) + \epsilon(\tau_L)\). \{\epsilon(\tau_L)\} is a strictly stationary martingale difference sequence (mds) with respect to increasing \(\mathcal{F}_{\tau_L} \subset \mathcal{F}_{\tau_L+1}\), where \(\Omega \equiv E[\epsilon(\tau_L)\epsilon(\tau_L)^t]\) is positive definite.

**Remark 2.1.** Note that Assumption 2.1 assumes the mds error, which is weaker than i.i.d. No theorems in this paper require i.i.d. errors. A leading economic example of mds error that is not i.i.d. is generalized autoregressive conditional heteroskedasticity (GARCH) studied by Bollerslev (1986).

A constant term is omitted from (2.1) for algebraic simplicity, but can be easily added if desired. Coefficients \(d\)'s govern the autoregressive property of \(x_H\), while coefficients \(a\)'s govern the autoregressive property of \(x_L\).

Coefficients \(b\)'s and \(c\)'s are more relevant in view of Granger causality, so we explain how they are labeled in (2.1). \(b_1\) is the impact of the most recent past observation of \(x_H\) (i.e. \(x_H(\tau_L - 1, m)\)) on \(x_L(\tau_L)\), \(b_2\) is the impact of the second most recent past observation of \(x_H\) (i.e. \(x_H(\tau_L - 1, m - 1)\)) on \(x_L(\tau_L)\), and so on through \(b_{pm}\). In general, \(b_k\) represents the impact of \(x_H\) on \(x_L\) when there are \(k\) high frequency periods apart from each other.

Similarly, \(c_1\) is the impact of \(x_L(\tau_L - 1)\) on the nearest observation of \(x_H\) (i.e. \(x_H(\tau_L, 1)\)), \(c_2\) is the impact of \(x_L(\tau_L - 1)\) on the second nearest observation of \(x_H\) (i.e. \(x_H(\tau_L, 2)\)), \(c_{m+1}\) is the impact of \(x_L(\tau_L - 2)\) on the \((m + 1)\)-st nearest observation of \(x_H\) (i.e. \(x_H(\tau_L, 1)\)), and so on. Finally, \(c_{pm}\) is the impact of \(x_L(\tau_L - p)\) on \(x_H(\tau_L, m)\). In general, \(c_k\) represents the impact of \(x_L\) on \(x_H\) when there are \(k\) high frequency periods apart from each other.

To proceed further, we impose the stability condition of the MF-VAR system as well as the \(\alpha\)-mixing property of the mixed frequency vector \(X(\tau_L)\) and mds error \(\epsilon(\tau_L)\).
Assumption 2.2. All roots of the polynomial det$(I_K - \sum_{k=1}^{p} A_k x^k) = 0$ lie outside the unit circle, where det$(\cdot)$ means the determinant.

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

Remark 2.2. While Assumptions 2.1 and 2.2 ensure the covariance stationarity of $\{x_H(\tau_L, j)\}_{\tau_L}$ for each $j \in \{1, \ldots, m\}$, they do not ensure the covariance stationarity of the entire high frequency series $\{\{x_H(\tau_L, j)\}_{j=1}^{m}\}_{\tau_L}$. A simple counter-example is $X(\tau_L) = \epsilon(\tau_L)$ and a diagonal error covariance matrix $\Omega = [\sigma_j^2]_{j=1}^{K}$ with $\sigma_i^2 \neq \sigma_j^2$. In this case the entire high frequency series $\{\{x_H(\tau_L, j)\}_{j=1}^{m}\}_{\tau_L}$ is heteroskedastic since $E[x_H(\tau_L, j)^2] = \sigma_j^2$. No theoretical results in this paper require the covariance stationarity of the entire high frequency series, so we do not assume it.

We are now ready to discuss testing strategies for Granger causality between $x_H$ and $x_L$. Section 2.1 is concerned with high-to-low causality (i.e. causality from $x_H$ to $x_L$), while Section 2.2 is concerned with low-to-high causality (i.e. causality from $x_L$ to $x_H$).

2.1 High-to-Low Granger Causality

To focus on high-to-low Granger causality, we pick the last row of the entire system (2.1):

$$x_L(\tau_L) = \sum_{k=1}^{p} a_k x_L(\tau_L - k) + \sum_{j=1}^{pm} b_j x_H(\tau_L - 1, m+1-j) + \epsilon_L(\tau_L),$$

$$\epsilon_L(\tau_L) \overset{mds.}{\sim} (0, \sigma_L^2), \quad \sigma_L^2 > 0.$$  \hfill (2.2)

In (2.2), index $j \in \{1, \ldots, pm\}$ is in terms of high frequency and the second argument of $x_H, m+1-j,$ will go below 1 when $j > m$. Allowing any integer value (e.g. smaller than 1 or larger than $m$) in the the second argument of $x_H$ does not cause any confusion. That makes analytical work much easier as well. In this paper $x_H(\tau_L, 0)$ is understood as $x_H(\tau_L - 1, m)$; $x_H(\tau_L, -1)$ is understood as $x_H(\tau_L - 1, m-1)$; $x_H(\tau_L, m+1)$ is understood as $x_H(\tau_L + 1, 1)$. More generally, we can interchangeably write $x_H(\tau_L - i, j) = x_H(\tau_L, j - im)$ for $j = 1, \ldots, m$ and $i \geq 0$. Complete details of these notational conventions are given in Appendix A, and we exploit these notations throughout the paper.

To rewrite (2.2) in matrix form, define $X_L(\tau_L - 1) = [x_L(\tau_L - 1), \ldots, x_L(\tau_L - p)]'$, $X_H(\tau_L - 1) = [x_H(\tau_L - 1, m+1), \ldots, x_H(\tau_L - 1, m+1-pm)]'$, and $b = [b_1, \ldots, b_{pm}]'$. Then, (2.2) can be rewritten as

$$x_L(\tau_L) = X_L(\tau_L - 1)' a + X_H(\tau_L - 1)' b + \epsilon_L(\tau_L).$$  \hfill (2.3)

It is evident from (2.1)-(2.3) that high-to-low Granger causality has a strong connection with coefficient $b$. Based on the classic theory of Dufour and Renault (1998) and the mixed frequency extension made by Ghysels, Hill, and Motegi (2013), we know that $x_H$ does not Granger cause $x_L$ given the mixed frequency information set $\mathcal{F}_{\tau_L} = \sigma(X(\tau) : \tau \leq \tau_L)$ if and only if $b = 0_{pm \times 1}$. In other words, DGP (2.2) reduces to a pure AR($p$) process under non-causality. Our main concern is how to construct a desirable test statistic with respect to $b = 0_{pm \times 1}$. It is desired that the test statistic is consistent (i.e. we can
detect any form of causality with power approaching 1), it achieves high power in local asymptotics and finite sample, and it does not produce size distortions in small sample. Section 2.1.1 discusses the mixed frequency approach which works on high frequency observations of $x_H$, while Section 2.1.2 discusses the conventional low frequency approach which works on an aggregated $x_H$. It will turn out that only the former allows us to construct a consistent test.

### 2.1.1 Mixed Frequency Approach

Before presenting our own test, let us review the existing mixed frequency Granger causality test proposed by Ghysels, Hill, and Motegi (2013). Essentially, their methodology is running a Wald test based on a naïve regression model that regresses $x_L$ onto its own low frequency lags and high frequency lags of $x_H$:

$$x_L(\tau_L) = \sum_{k=1}^{q} \alpha_k x_L(\tau_L - k) + \sum_{j=1}^{h} \beta_j x_H(\tau_L - 1, m + 1 - j) + u_L(\tau_L)$$

(2.4)

for $\tau_L = 1, \ldots, T_L$. Ghysels, Hill, and Motegi (2013) fit OLS to (2.4) and then test $H_0: \beta_1 = \cdots = \beta_h = 0$ via a Wald test. Model (2.4) contains DGP (2.2) as a special case when $q \geq p$ and $h \geq pm$. Hence the Wald test is trivially consistent if $q \geq p$ and $h \geq pm$. A potential problem here is that $pm$, the true lag order of $x_H$, may be quite large in some applications even when $p$ is fairly small. Consider a week vs. quarter case for instance, then the MF-V AR lag order $p$ is in terms of quarter and $m = 13$ approximately. We thus have $pm = 39$ when $p = 3$, and $pm = 52$ when $p = 4$, etc. Therefore, including sufficiently many high frequency lags $h \geq pm$ likely results in size distortions when sample size $T_L$ is small and $m$ is large. Size distortions may be deleted by bootstrap, but then finite sample power may get quite low. If in turn we take a small number of lags $h < pm$, then there may be less size distortions. But the test no longer has power approaching 1 when there exists Granger causality involving lags beyond $h$.

A main contribution of this paper is to resolve this trade-off by combining multiple parsimonious regression models:

**Mixed Frequency Parsimonious Regression Models for High-to-Low Granger Causality**

$$x_L(\tau_L) = \sum_{k=1}^{q} \alpha_{k,j} x_L(\tau_L - k) + \beta_j x_H(\tau_L - 1, m + 1 - j) + u_{L,j}(\tau_L), \quad j = 1, \ldots, h.$$  

(2.5)

In a matrix form, model $j$ is rewritten as

$$x_L(\tau_L) = \left[ X_L^{(q)}(\tau_L - 1)' x_H(\tau_L - 1, m + 1 - j) \right] \equiv \theta_j + u_{L,j}(\tau_L), \quad j = 1, \ldots, h.$$  

(2.6)
where \( \mathbf{X}_L^{(q)}(\tau_L - 1) = [x_L(\tau_L - 1), \ldots, x_L(\tau_L - q)]' \). Note that model \( j \) contains \( q \) low frequency autoregressive lags of \( x_L \) as well as the only \( j \)-th high frequency lag of \( x_H \). The number of parameters in model \( j \) is thus \( q + 1 \), which tends to be much smaller than the number of parameters in the naïve regression model (2.4), \( q + h \). This feature alleviates size distortions for large \( m \) and small \( T_L \). For each parsimonious regression model to be correctly specified under the null hypothesis of high-to-low non-causality, we need to assume that the autoregressive part of (2.5) has enough lags: \( q \geq p \). We impose the same assumption on the naïve regression model (2.4) in order to focus on the causality component, not autoregressive component.

**Assumption 2.4.** The number of autoregressive lags included in the the naïve regression model (2.4) and each parsimonious regression model (2.5), \( q \), is larger than or equal to the true autoregressive lag order \( p \) in (2.2).

We describe how to combine all \( h \) parsimonious models to get a test statistic. Consider fitting OLS for each of parsimonious regression models (2.5). Since we are assuming that \( q \geq p \), each model is correctly specified under the null hypothesis of high-to-low non-causality. OLS estimators \( \hat{\beta}_1, \ldots, \hat{\beta}_h \) should therefore converge to zeros in probability under non-causality. This implies that even the maximum among \( \{\hat{\beta}_1, \ldots, \hat{\beta}_h\} \) should converge to zero under non-causality. Using this property, we propose a test statistic:

**Mixed Frequency Max Test Statistic for High-to-Low Granger Causality**

\[
\mathcal{T} = \max_{1 \leq j \leq h} \left( \sqrt{\mathcal{T}_L w_{T_L,j} \hat{\beta}_j} \right)^2 . \tag{2.7}
\]

We call \( \mathcal{T} \) the *mixed frequency max test statistic* since it takes the maximum of the square of properly scaled individual OLS estimators, using mixed frequency data. \( \{w_{T_L,j} : j = 1, \ldots, h\} \) is a sequence of \( \sigma(\mathbf{X}(\tau_L - k) : k \geq 1) \)-measurable \( L_2 \)-bounded non-negative scalars with non-random mean-squared-error limits \( \{w_j\} \). As a standardization, we assume that \( \sum_{j=1}^h w_{T_L,j} = 1 \) without loss of generality. Let \( \mathbf{W}_{T_L,h} \) be an \( h \times h \) diagonal matrix whose diagonal elements are \( w_{T_L,1}, \ldots, w_{T_L,h} \). Similarly, let \( \mathbf{W}_h \) be an \( h \times h \) diagonal matrix whose diagonal elements are \( w_1, \ldots, w_h \). When we do not own any prior information about the weighting structure, a trivial choice of \( w_{T_L,j} \) is 1/h, non-random equal weights. We can consider any other weighting structure by choosing desired \( \{w_{T_L,1}, \ldots, w_{T_L,h}\} \) (cfr. Andrews and Ploberger (1994)).

We derive the asymptotic distribution of \( \mathcal{T} \) under \( H_0 : \mathbf{b} = \mathbf{0}_{pm \times 1} \). First, recall from (2.6) that all parameters in model \( j \) are expressed as \( \mathbf{\theta}_j \). We stack all parameters across all \( h \) models \( \mathbf{\theta} = [\mathbf{\theta}_1', \ldots, \mathbf{\theta}_h'] \) and then construct a selection matrix \( \mathbf{R} \) such that \( \mathbf{\beta} = [\hat{\beta}_1, \ldots, \hat{\beta}_h]' = \mathbf{R}\mathbf{\theta} \). Specifically, \( \mathbf{R} \) is an \( h \times (q + 1)h \) matrix whose \((j, (q + 1)j)\) element is 1 for \( j = 1, \ldots, h \) and all others are zeros. Under Assumptions 2.1-2.4, it is not hard to prove the asymptotic normality of \( \mathbf{\theta} \) and hence \( \mathbf{\beta} \). We then rely on the Cramér-Wold theorem to combine \( \hat{\beta}_1, \ldots, \hat{\beta}_h \).

**Theorem 2.1.** Let Assumptions 2.1, 2.2, 2.3, and 2.4 hold. Under \( H_0 : \mathbf{b} = \mathbf{0}_{pm \times 1} \), we have that

\[ T \overset{d}{\to} \max_{1 \leq j \leq h} \mathcal{N}_j^2 \text{ as } T_L \to \infty. \]

\( \mathcal{N} = [\mathcal{N}_1', \ldots, \mathcal{N}_h']' \) is a vector-valued random variable drawn from
We now investigate the asymptotic property of the mixed frequency max test statistic. If an estimator $\tilde{W}$ with $H$ matrix $H$ without imposing under a general alternative hypothesis $T$ where $N$ the continuous mapping theorem and Slutsky's theorem, consistent estimators can be obtained directly from (2.9). Hence, the availability of a consistent estimator $\hat{\sigma}^2_L$ can be calculated by fitting an AR($q$) model for $x_L$ and computing the sample variance of residuals (recall Assumption 2.4 ensuring $q \geq p$). If we did not impose $H_0$ then we could not get consistent $\hat{\sigma}^2_L$ due to the misspecification of each parsimonious regression model, but all we need for statistical inference is a consistent estimator for $V$ under $H_0$. Therefore, the mixed frequency max test can be implemented through the asymptotic $p$-value approximation $\hat{p}$.

**Consistency** We now investigate the asymptotic property of the mixed frequency max test statistic $T$ under a general alternative hypothesis $H_1 : b \neq 0_{pm \times 1}$. When there exists Granger causality, each parsimonious regression model is in general misspecified and hence the resulting estimator will be biased due to omitted regressors. We consider how it is biased by characterizing the pseudo-true value of $\beta = [\beta_1, \ldots, \beta_h]'$, denoted as $\beta^* = [\beta^*_1, \ldots, \beta^*_h]'$, in terms of underlying parameters $a, b, \sigma_L^2$ as well.
as some population moments of \(x_H\) and \(x_L\). Stack all parameters \(\theta = [\theta'_1, \ldots, \theta'_h]'\) and let \(\theta^*\) be the pseudo-true value of \(\theta\).

**Theorem 2.2.** Let Assumptions 2.1, 2.2, and 2.4 hold. Then, the pseudo-true value of \(\beta\) associated with OLS is given by \(\beta^* = R\theta^*, \theta^* = [\theta'_1, \ldots, \theta'_h]'\), and

\[
\theta^*_j \equiv \begin{bmatrix}
\alpha^*_1_{,j} \\
\vdots \\
\alpha^*_p_{,j} \\
\alpha^*_{p+1, j} \\
\vdots \\
\alpha^*_q_{,j} \\
\beta^*_j
\end{bmatrix} = \begin{bmatrix}
a_1 \\
\vdots \\
a_p \\
0 \\
\vdots \\
0 \\
0
\end{bmatrix} + \left[ E \left[ x_j(\tau_L - 1)x_j(\tau_L - 1)' \right] \right]^{-1} E \left[ x_j(\tau_L - 1)X_H(\tau_L - 1)' \right] b, \tag{2.10}
\]

where \(x_j(\tau_L - 1)\) is a vector of all regressors in each parsimonious regression model (cfr. (2.6)) while \(X_H(\tau_L - 1)\) is a vector of \(pm\) high frequency lags of \(x_H\) (cfr. (2.3)).

**Proof 2.2.** See Appendix C.

The population covariance terms \(\Gamma_{j,j}\) and \(C_j\) can be characterized by the underlying parameters \(a, b,\) and \(\sigma^2_L\) although algebra is very complicated. We will elaborate it in local asymptotic power analysis (cfr. Section 3).

Theorem 2.2 provides useful insights on the relationship between the underlying coefficient \(b\) and the pseudo-true value for the parameter \(\beta\). First, it is clear that \(\beta^* = 0_{h \times 1}\) whenever there is non-causality (i.e. \(b = 0_{pm \times 1}\)), regardless of the relative magnitude of \(h\) and \(pm\). We have already exploited this fact when we derive the asymptotic distribution of \(T\) under non-causality (cfr. Theorem 2.1). Second, it can be shown that \(b = 0_{pm \times 1}\) whenever \(\beta^* = 0_{h \times 1}\), assuming \(h \geq pm\).

**Theorem 2.3.** Let Assumptions 2.1, 2.2, and 2.4 hold. When \(h \geq pm\), we have that \(\beta^* = 0_{h \times 1} \Rightarrow b = 0_{pm \times 1}\).

**Proof 2.3.** See Appendix D.

Assuming \(w_j > 0\) for all \(j = 1, \ldots, h\), (2.7) indicates that \(T \xrightarrow{p} \infty\) if and only if \(\beta^* \neq 0_{h \times 1}\). Theorem 2.3 essentially states that \(\beta^* \neq 0_{h \times 1}\) under a general alternative hypothesis \(H_1 : b \neq 0_{pm \times 1}\), given \(h \geq pm\). The consistency of the mixed frequency max test is an immediate implication of these properties.

**Theorem 2.4.** Let Assumptions 2.1, 2.2, and 2.4 hold. Assume \(w_j > 0\) for all \(j = 1, \ldots, h\). Given \(h \geq pm, T \xrightarrow{p} \infty\) under a general alternative hypothesis \(H_1 : b \neq 0_{pm \times 1}\).

**Proof 2.4.** See Appendix E.

If one happens to choose \(h\) that is smaller than \(pm\), there is a certain form of causality such that power does not approach 1. We provide such an example below.
Example 2.2 (Inconsistency due to Small \( h \)). Consider a very simple DGP with \( m = 2 \) and \( p = 1 \):

\[
\begin{bmatrix}
  x_H(\tau_L, 1) \\
  x_H(\tau_L, 2) \\
  x_L(\tau_L)
\end{bmatrix}
= \begin{bmatrix}
  0 & 0 & 0 \\
  0 & 0 & 0 \\
  -1/\rho & 1 & 0
\end{bmatrix}
\begin{bmatrix}
  x_H(\tau_L - 1, 1) \\
  x_H(\tau_L - 1, 2) \\
  x_L(\tau_L - 1)
\end{bmatrix}
+ \begin{bmatrix}
  \epsilon_H(\tau_L, 1) \\
  \epsilon_H(\tau_L, 2) \\
  \epsilon_L(\tau_L)
\end{bmatrix},
\]

where \( \epsilon(\tau_L) \overset{mds.}{\sim} (0_{3\times 1}, \Omega) \),  \( \Omega = \begin{bmatrix}
  1 & \rho & 0 \\
  \rho & 1 & 0 \\
  0 & 0 & 1
\end{bmatrix}, \rho \neq 0, |\rho| < 1. \)

Given this DGP, we will show that choosing \((q, h) = (1, 1)\) provides no power and choosing \((q, h) = (1, 2)\) provides power approaching 1. We first compute \( \Gamma_{1,1} = E[x_1(\tau_L - 1)x_1(\tau_L - 1)'] \), where \( x_1(\tau_L - 1) = [x_L(\tau_L - 1), x_H(\tau_L - 1, 2)]' \) as defined in (2.6) and (2.10). Given the DGP (2.11), it follows that\(^2\)

\[
\Gamma_{1,1} = \begin{bmatrix}
  1/\rho^2 & 0 \\
  0 & 1
\end{bmatrix}.
\]

Next we consider \( \Gamma_{2,2} = E[x_2(\tau_L - 1)x_2(\tau_L - 1)'] \), where \( x_2(\tau_L - 1) = [x_L(\tau_L - 1), x_H(\tau_L - 1, 1)]' \).

It is easy to show that \( \Gamma_{2,2} = \Gamma_{1,1} \) and hence

\[
\Gamma_{1,1}^{-1} = \Gamma_{2,2}^{-1} = \begin{bmatrix}
  \rho^2 & 0 \\
  0 & 1
\end{bmatrix}.
\]

We next compute \( C_1 \equiv E[x_1(\tau_L - 1)X_H(\tau_L - 1)'] \), where \( X_H(\tau_L - 1) = [x_H(\tau_L - 1, 2), x_H(\tau_L - 1, 1)]' \) as defined in (2.3). It is evident that

\[
C_1 \equiv E\left[ x_1(\tau_L - 1)X_H(\tau_L - 1)'ight]
= \begin{bmatrix}
  E[x_L(\tau_L - 1)x_H(\tau_L - 1, 2)] & E[x_L(\tau_L - 1)x_H(\tau_L - 1, 1)] \\
  E[x_H(\tau_L - 1, 2)x_H(\tau_L - 1)] & E[x_H(\tau_L - 1, 2)x_H(\tau_L - 1, 1)]
\end{bmatrix}
= \begin{bmatrix}
  0 & 0 \\
  1 & \rho
\end{bmatrix}
\]

and similarly

\[
C_2 \equiv E\left[ x_2(\tau_L - 1)X_H(\tau_L - 1)'ight]
= \begin{bmatrix}
  E[x_L(\tau_L - 1)x_H(\tau_L - 1, 2)] & E[x_L(\tau_L - 1)x_H(\tau_L - 1, 1)] \\
  E[x_H(\tau_L - 1, 1)x_H(\tau_L - 1, 2)] & E[x_H(\tau_L - 1, 1)x_H(\tau_L - 1, 1)]
\end{bmatrix}
= \begin{bmatrix}
  0 & \rho \\
  \rho & 1
\end{bmatrix}.
\]

\(^2\)DGP (2.11) implies that \( E[x_H(\tau_L - 1, 1)'] = E[x_H(\tau_L - 1, 2)'] = 1, E[x_L(\tau_L - 1)^2] = 1/\rho^2, \) and \( E[x_L(\tau_L - 1) x_H(\tau_L - 1, 2)] = E\left[ (-\frac{1}{\rho} x_H(\tau_L - 2, 1) + x_H(\tau_L - 2, 2) + \epsilon_L(\tau_L - 1)) x_H(\tau_L - 1, 2)\right] = 0. \) Using these results, we have that

\[
\Gamma_{1,1} \equiv E\left[ x_1(\tau_L - 1)x_1(\tau_L - 1)'ight]
= \begin{bmatrix}
  E[x_L(\tau_L - 1)^2] & E[x_L(\tau_L - 1)x_H(\tau_L - 1, 2)] \\
  E[x_H(\tau_L - 1, 2)x_L(\tau_L - 1)] & E[x_H(\tau_L - 1, 2)^2]
\end{bmatrix}
= \begin{bmatrix}
  1/\rho^2 & 0 \\
  0 & 1
\end{bmatrix}.
\]
In view of (2.10), we get that
\[
\begin{bmatrix}
\alpha_{1,1}^* \\
\beta_1^*
\end{bmatrix} = \begin{bmatrix}
\rho^2 & 0 \\
0 & 1
\end{bmatrix} \begin{bmatrix}
0 & 0 \\
1 & -1/\rho
\end{bmatrix} = \begin{bmatrix}
0 \\
0
\end{bmatrix}
\quad \text{and} \quad
\begin{bmatrix}
\alpha_{1,2}^* \\
\beta_2^*
\end{bmatrix} = \begin{bmatrix}
\rho^2 & 0 \\
0 & 1
\end{bmatrix} \begin{bmatrix}
0 & 0 \\
\rho & -1/\rho
\end{bmatrix} = \begin{bmatrix}
0 \\
\rho - 1/\rho
\end{bmatrix}.
\]

Note that \( \beta_1^* = 0 \) and \( \beta_2^* = \rho - 1/\rho \neq 0 \) since \( |\rho| < 1 \). Therefore, if we choose \( h = 1 \), the mixed frequency max test statistic \( T \) converges to the asymptotic distribution under the null hypothesis of non-causality, resulting in no power. If we choose \( h = 2 \), we have that \( T \xrightarrow{P} \infty \) and there is power approaching 1 (assuming a positive weight is assigned on \( \hat{\beta}_2 \), i.e. \( w_2 > 0 \)).

An intuition behind no power, assuming \( \rho > 0 \) for a purely expositional purpose, is that the positive impact of \( x_H(\tau_L - 1, 2) \) on \( x_L(\tau_L) \), the negative impact of \( x_H(\tau_L - 1, 1) \) on \( x_L(\tau_L) \), and the positive autocorrelation of \( x_H \) offset each other to make \( \beta_1^* = 0 \).

One might argue that the example DGP (2.11) is unrealistic since \( |b_2| > |b_1| \) (i.e. \( x_H(\tau_L - 1, 1) \) has a larger impact on \( x_L(\tau_L) \) than a more recent observation \( x_H(\tau_L - 1, 2) \) does). But some applications may have such a tricky Granger causality due to lagged information transmission or seasonal effects. It is thus advised to take a sufficiently large \( h \) when the possibility of lagged causality cannot be ruled out.

### 2.1.2 Low Frequency Approach

We have shown in Section 2.1.1 that the Wald test based on the mixed frequency na"ive regression model (2.4) and the max test based on the mixed frequency parsimonious regression models (2.5) are both consistent (as long as \( h \geq pm \)). These two tests are sharing a key feature that they are working on high frequency observations of \( x_H \). If we worked on an aggregated \( x_H \), then neither Wald test nor max test would be consistent no matter how many low frequency lags of \( x_H \) we included.

To verify this point, we formulate a Wald statistic based on low frequency version of na"ive regression model and a max test statistic based on a low frequency version of parsimonious regression models. We introduce linear aggregation scheme \( x_H(\tau_L) = \sum_{j=1}^{m} \delta_j x_H(\tau_L, j) \) with \( \delta_j \geq 0 \) for all \( j = 1, \ldots, m \) and \( \sum_{j=1}^{m} \delta_j = 1 \). The linear aggregation scheme is sufficiently general for most economic applications since it includes flow sampling (i.e. \( \delta_j = 1/m \) for \( j = 1, \ldots, m \)) and stock sampling (i.e. \( \delta_j = I(j = m) \) for \( j = 1, \ldots, m \)) as special cases. Note that \( \delta_j \) is not a parameter to estimate; it is a fixed quantity that determines an aggregation scheme.

We start with low frequency na"ive regression models and then move on to parsimonious regression models.

**Low Frequency Na"ive Regression** We first discuss a low frequency na"ive regression model which regresses \( x_L \) onto its own low frequency lags and \( h \) low frequency lags of aggregated \( x_H \).
Low Frequency Na"ıve Regression Model for High-to-Low Granger Causality

\[ x_L(\tau_L) = \sum_{k=1}^{q} \alpha_k x_L(\tau_L - k) + \sum_{j=1}^{h} \beta_j x_H(\tau_L - j) + u_L(\tau_L) \]

\[ = [X_L^{(q)}(\tau_L - 1)', x_H(\tau_L - 1), \ldots, x_H(\tau_L - h)] \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_{q} \\ \beta_1 \\ \vdots \\ \beta_h \end{bmatrix} + u_L(\tau_L) \]

\[ = [X_L^{(q)}(\tau_L - 1)', x_H(\tau_L - 1)]\theta^{LF} + u_L(\tau_L). \]

(2.12)

Note that \( X_H(\tau_L - 1) \) is an \( h \times 1 \) vector stacking aggregated \( x_H \). \( \bar{x}(\tau_L - 1) \) is a \( (q + h) \times 1 \) vector of all regressors. Superscript "LF" is put on the parameter vector \( \theta^{LF} \) to emphasize that we are working on a low frequency model here. Strictly speaking, "LF" should also be put on \( \alpha \)'s, \( \beta \)'s, and residual \( u_L(\tau_L) \) since they are generally different from what is appearing in the mixed frequency na"ıve regression model (2.4). We refrain from doing that for the sake of notational brevity.

We still impose Assumption 2.4 that \( q \geq p \) so that we can focus on the causality part. Since we are assuming the same DGP (2.1) as in the mixed frequency case, the pseudo-true value for \( \theta^{LF} \), denoted as \( (\theta^{LF})^* \), can be derived easily:

\[
(\theta^{LF})^* \equiv \begin{bmatrix} \alpha_1^* \\ \vdots \\ \alpha_p^* \\ \alpha_{p+1}^* \\ \vdots \\ \alpha_q^* \\ \beta^* \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_p \\ 0 \\ \vdots \\ \alpha_q \\ 0 \end{bmatrix} + \left[ E[x(\tau_L - 1)x(\tau_L - 1)'] \right]^{-1} E[x(\tau_L - 1)X_H(\tau_L - 1)] b,
\]

(2.13)

where \( \beta^* = [\beta_1^*, \ldots, \beta_h^*]' \). The derivation of (2.13) is omitted since it is very similar to the proof of Theorem 2.2.

Low frequency Wald statistic \( W_{LF} \) is simply a classic Wald statistic with respect to a hypothesis \( \beta \equiv [\beta_1, \ldots, \beta_h]' = 0_{h \times 1} \). Consistency requires that \( W_{LF} \overset{p}{\rightarrow} \infty \) whenever \( b \neq 0_{pn \times 1} \). We present a counter-example of Granger causality that makes \( \beta^* = 0_{h \times 1} \). We can create such an example for any linear aggregation scheme and \( h \in \mathbb{N} \).

**Example 2.3** (Inconsistency of Low Frequency Wald Test). Consider an even simpler DGP than (2.11)
with $m = 2$ and $p = 1$:

\[
\begin{bmatrix}
  x_H(\tau_L, 1) \\
  x_H(\tau_L, 2) \\
  x_L(\tau_L)
\end{bmatrix} =
\begin{bmatrix}
  0 & 0 & 0 \\
  0 & 0 & 0 \\
  b_2 & b_1 & 0
\end{bmatrix}
\begin{bmatrix}
  x_H(\tau_L - 1, 1) \\
  x_H(\tau_L - 1, 2) \\
  x_L(\tau_L - 1)
\end{bmatrix} +
\begin{bmatrix}
  \epsilon_H(\tau_L, 1) \\
  \epsilon_H(\tau_L, 2) \\
  \epsilon_L(\tau_L)
\end{bmatrix} +
\begin{bmatrix}
  \epsilon(\tau_L, 1)
\end{bmatrix} =
\begin{bmatrix}
  \mathbf{X}(\tau_L) \\
  \mathbf{X}(\tau_L - 1)
\end{bmatrix} +
\begin{bmatrix}
  \mathbf{e}(\tau_L)
\end{bmatrix}
\]

Recall the linear aggregation scheme $x_H(\tau_L) = \delta_1 x_H(\tau_L, 1) + (1 - \delta_1) x_H(\tau_L, 2)$. Fixing $\delta_1$ and $h$, we want to show that $\beta^* = 0_{(h \times 1)}$ regardless of linear aggregation scheme and the choice of $h$. To this end, we need to elaborate a key population moment in (2.13):

\[
\mathbf{C} \equiv E[\mathbf{e}(\tau_L - 1) \mathbf{X}(\tau_L - 1)'] =
\begin{bmatrix}
  E[x_L(\tau_L - 1)x_H(\tau_L - 1, 1)] & E[x_L(\tau_L - 1)x_H(\tau_L - 1, 1)] \\
  E[x_H(\tau_L - 1)x_H(\tau_L - 1, 2)] & E[x_H(\tau_L - 1)x_H(\tau_L - 1, 2)] \\
  \vdots & \vdots \\
  E[x_H(\tau_L - h)x_H(\tau_L - 1, 2)] & E[x_H(\tau_L - h)x_H(\tau_L - 1, 1)]
\end{bmatrix}.
\]

Given (2.14), it follows that the second row of $\mathbf{C}$ is $[1 - \delta_1, \delta_1]$ and all other rows are zeros.$^3$

Given, $\delta_1$, we can find a $\mathbf{b} \neq 0_{2 \times 1}$ such that $\mathbf{C} \mathbf{b} = 0_{(h+1 \times 1)}$. First, let $b_1 = 0$ and $b_2 \neq 0$ if $\delta_1 = 0$. Second, let $b_1 \neq 0$ and $b_2 = -b_1(1 - \delta_1)/\delta_1$ if $\delta_1 \in (0, 1)$. Third, let $b_1 \neq 0$ and $b_2 = 0$ if $\delta_1 = 1$. For any of these three cases, we have that $\mathbf{C} \mathbf{b} = 0_{(h+1 \times 1)}$ and thus $\beta^* = 0_{h \times 1}$ in view of (2.13). The low frequency Wald statistic $W_{LF}$ therefore converges to the asymptotic distribution under the null hypothesis of non-causality, $\chi^2_h$, which produces no power.

An intuition behind this result is quite simple. When the impact of $x_H(\tau_L - 1, 1)$ on $x_L(\tau_L)$ and the impact of $x_H(\tau_L - 1, 2)$ on $x_L(\tau_L)$ are inversely proportional to the aggregation scheme, the causal effects are offset by each other after temporal aggregation.

**Low Frequency Parsimonious Regression** We now consider regressing $x_L$ onto its own low frequency lags and only one low frequency lag of aggregated $x_H$:

**Low Frequency Parsimonious Regression Models for High-to-Low Granger Causality**

\[
x_L(\tau_L) = \sum_{k=1}^{q} \alpha_{k,j}x_L(\tau_L - k) + \beta_j x_H(\tau_L - j) + u_{L,j}(\tau_L)
\]

\[
= \begin{bmatrix}
  X^{[g]}_L(\tau_L - 1)^t, x_H(\tau_L - j)
\end{bmatrix}
\begin{bmatrix}
  \alpha_{1,j} \\
  \vdots \\
  \alpha_{q,j} \\
  \beta_j
\end{bmatrix} + u_{L,j}(\tau_L), \quad j = 1, \ldots, h. \tag{2.15}
\]

$^3$Equation (2.14) immediately implies that $x_L(\tau_L - 1) = b_1 \epsilon_H(\tau_L - 2, 2) + b_2 \epsilon_H(\tau_L - 2, 1) + \epsilon_L(\tau_L - 1)$ and therefore $E[x_L(\tau_L - 1)x_H(\tau_L - 1, 2)] = b_1 E[\epsilon_H(\tau_L - 2, 2)\epsilon_H(\tau_L - 1, 2)] + b_2 E[\epsilon_H(\tau_L - 2, 1)\epsilon_H(\tau_L - 1, 2)] + \epsilon_L(\tau_L - 1)\epsilon_H(\tau_L - 1, 2) = 0$. Similarly, $E[x_L(\tau_L - j)x_H(\tau_L - 1, 1)] = 0$. In addition, assuming a general linear aggregation scheme, $E[x_H(\tau_L - j)x_H(\tau_L - j, 1)] = E[(\delta_1 x_H(\tau_L - j, 1) + (1 - \delta_1) x_H(\tau_L - j, 2))x_H(\tau_L - 1, 2)] = (1 - \delta_1)I(j = 1)$. Similarly, $E[x_H(\tau_L - j)x_H(\tau_L - 1, 1)] = \delta_1 I(j = 1)$. Thus, the second row of $\mathbf{C}$ is $[1 - \delta_1, \delta_1]$ and all other rows are zeros.
We still impose Assumption 2.4 that \( q \geq p \). Since we are assuming the same DGP (2.1) as in the mixed frequency case, the pseudo-true value for \( \theta_{LF}^j \), denoted as \((\theta_{LF}^j)^*\), can be easily derived by replacing \( x_j(\tau_L - 1)' \) with \( x_j(\tau_L - 1)' \) in (2.10):

\[
(\theta_{LF}^j)^* = \begin{bmatrix}
\alpha_{1,j}^* \\
\vdots \\
\alpha_{p,j}^* \\
\alpha_{p+1,j}^* \\
\vdots \\
\alpha_{q,j}^* \\
\beta_j^*
\end{bmatrix} = \begin{bmatrix}
a_1 \\
\vdots \\
a_p \\
0 \\
\vdots \\
0
\end{bmatrix} + E \left[ x_j(\tau_L - 1)x_j(\tau_L - 1)' \right]^{-1} E \left[ x_j(\tau_L - 1)X_H(\tau_L - 1)' \right] b. \tag{2.16}
\]

Low frequency max test \( T_{LF} \) is constructed in the same way as (2.7): \( T_{LF} \equiv \max_{1 \leq j \leq h} (\sqrt{T_L} w_{T_L,j} \hat{\beta}_j)^2 \). The limit distribution of \( T_{LF} \) under \( H_0 : b = 0_{pm \times 1} \) is same as Theorem 2.1, the only difference being that \( x_j(\tau_L - 1) \) there should be replaced with \( x_j(\tau_L - 1) \).

Continuing Example 2.3, we can easily show that \( T_{LF} \) is inconsistent. Algebraic details are omitted to save space but available upon request.

Summarizing Section 2.1, we have introduced four different high-to-low Granger causality tests. They are categorized in Table 1 depending on the sampling frequency of \( x_H \) and model specification. "Mixed Frequency" works on high frequency observations of \( x_H \), while "Low Frequency" works on an aggregated \( x_H \). "Parsimonious" specification prepares \( h \) separate models with the \( j \)-th model containing only the \( j \)-th lag of \( x_H \) (either high frequency or low frequency lag). "Naïve" specification prepares only one model which contains all \( h \) lags of \( x_H \) (either high frequency or low frequency). The MF Wald test tends to involve many parameters, while the MF max test as well as LF tests involve fewer parameters. MF tests are consistent if the selected number of high frequency lags \( h \) is larger than or equal to the true lag order \( pm \). In contrast, LF tests are inconsistent no matter how many lags are included in the model and no matter which linear aggregation scheme is used.

### 2.2 Low-to-High Granger Causality

We now consider testing for Granger causality from \( x_L \) to \( x_H \). Section 2.2.1 discusses mixed frequency tests which work on high frequency observations of \( x_H \), while Section 2.2.2 discusses low frequency tests which work on aggregated \( x_H \).

#### 2.2.1 Mixed Frequency Approach

Consider the same MF-VAR(\( p \)) data generating process (2.1) as before. One possible way of testing for low-to-high causality (i.e. causality from \( x_L \) to \( x_H \)) is a Wald test based on the naïve regression model below, which is a natural extension of Sims’ (1972) two-sided regression model to the mixed frequency framework.
Mixed Frequency Naive Regression Model for Low-to-High Granger Causality

\[ x_L(\tau_L) = \sum_{k=1}^{q} \alpha_k x_L(\tau_L - k) + \sum_{j=1}^{h} \beta_j x_H(\tau_L - 1, m + 1 - j) + \sum_{j=1}^{r} \gamma_j x_H(\tau_L + 1, j) + u_L(\tau_L). \quad (2.17) \]

Model (2.17) regresses \( x_L \) onto \( q \) low frequency lags of \( x_L \), \( h \) high frequency lags of \( x_H \), and \( r \) high frequency leads of \( x_H \). Suppose that we run OLS for model (2.17) and then implement a Wald test with respect to \( \gamma = [\gamma_1, \ldots, \gamma_r]' = 0_{r \times 1} \). Under the null hypothesis of low-to-high non-causality, all \( \gamma \)'s should be equal to zero and the Wald statistic should converge to \( \chi^2_r \) (given sufficiently large \( q \) and \( h \)).

A potential problem of the mixed frequency Wald test based on model (2.17) is that there may be a potential problem of the mixed frequency Wald test based on model (2.17) is that there may be parameter proliferation. The number of parameters, \( q + h + r \) may get quite large in some applications. We thus propose more parsimonious models:

Mixed Frequency Parsimonious Regression Models for Low-to-High Granger Causality

\[ x_L(\tau_L) = \sum_{k=1}^{q} \alpha_{k,j} x_L(\tau_L - k) + \sum_{k=1}^{h} \beta_{k,j} x_H(\tau_L - 1, m + 1 - k) + \gamma_j x_H(\tau_L + 1, j) + u_{L,j}(\tau_L), \quad j = 1, \ldots, r. \quad (2.18) \]

Model \( j \) regresses \( x_L \) onto \( q \) low frequency lags of \( x_L \), \( h \) high frequency lags of \( x_H \), and only the \( j \)-th high frequency lead of \( x_H \). Let \( \hat{\gamma}_j \) be the OLS estimator for \( \gamma_j \) from model \( j \). Stack \( \hat{\gamma} = [\hat{\gamma}_1, \ldots, \hat{\gamma}_r]' \). Under the null hypothesis of low-to-high non-causality, we have that \( \hat{\gamma} \overset{p}{\rightarrow} 0_{r \times 1} \). Using this property, we formulate a mixed frequency max test statistic for low-to-high causality:

Mixed Frequency Max Test Statistic for Low-to-High Granger Causality

\[ U = \max_{1 \leq j \leq r} \left( \sqrt{T_L} w_{L,j} \hat{\gamma}_j \right)^2, \quad (2.19) \]

where \( w_{TL} = [w_{TL,1}, \ldots, w_{TL,r}]' \) is a weighting scheme such that \( w_{TL} \overset{L^2}{\rightarrow} w \), as in high-to-low causality. We will derive the asymptotic distribution of \( U \) under the null hypothesis of low-to-high non-causality. To this end, we keep imposing Assumption 2.4 that \( q \geq p \) and an extra assumption that \( h \geq pm \). The latter is formally stated as Assumption 2.5.

Assumption 2.5. The number of high frequency lags of \( x_H \) included in models (2.17) and (2.18), \( h \), is larger than or equal to the true lag length \( pm \).

We rewrite the parsimonious regression models (2.18) in matrix form. Let \( n = q + h + 1 \), the number of regressors in each model. Define \( n \times 1 \) vectors \( y_j(\tau_L - 1) = [x_L(\tau_L - 1), \ldots, x_L(\tau_L - q), x_H(\tau_L - 1, m + 1 - 1), \ldots, x_H(\tau_L - 1, m + 1 - h), x_H(\tau_L + 1, j)]' \) and \( \phi_j = [\alpha_{1,j}, \ldots, \alpha_{q,j}, \beta_{1,j}, \ldots, \beta_{h,j}, \gamma_j]' \). \( y_j(\tau_L - 1) \) is a vector of all regressors while \( \phi_j \) is a vector of all parameters in model \( j \). Using these notations, model \( j \) can be rewritten as \( x_L(\tau_L) = y_j(\tau_L - 1)' \phi_j + u_{L,j}(\tau_L) \).

The asymptotic distribution of \( U \) under non-causality can be derived in the same way as in Theorem 2.1. Simply replace the regressors \( x_j(\tau_L - 1) \) with \( y_j(\tau_L - 1) \). Also, remember that here we need Assumption 2.5 additionally. The final result and derivation are omitted, but available upon request.\(^4\)

\(^4\)Consistency of the low-to-high max test is an open question at this position.
While MF parsimonious regression models (2.18) certainly have a fewer number of parameters than the MF naive regression model (2.17), the parsimonious regression models may still have many parameters. Since $h$ and $r$ sometimes take a large number relative to sample size, there would be size distortions if we worked on a completely unrestricted model. We are interested in low-to-high causality here, so we impose a MIDAS polynomial for the high-to-low causality part and keep the low-to-high causality part unrestricted.

$$x_L(\tau_L) = \sum_{k=1}^{q} \alpha_k x_L(\tau_L - k) + \sum_{k=1}^{h} \omega_k(\pi)x_H(\tau_L - 1, 12 + 1 - k) + \gamma_j x_H(\tau_L + 1, j) + u_L(\tau_L), \quad j = 1, \ldots, r.$$  \hspace{1cm} (2.20)

$\omega_k(\pi)$ represents a MIDAS polynomial with a small-dimensional parameter vector $\pi \in \mathbb{R}^s$. Using a MIDAS polynomial with small $s$ is a common technique to save the number of parameters.

There are a variety of choice for the MIDAS polynomial $\omega_k(\pi)$ in the literature (see Technical Appendix A of Ghysels (2012)). This paper uses the Almon polynomial with dimension $s$, namely $\omega_k(\pi) = \sum_{l=1}^{s} \pi_l k^l$. A convenient characteristic of this polynomial is that the model is linear in $\pi$, which allows us to run OLS (not nonlinear least squares):

$$\sum_{k=1}^{h_{MF}} \omega_k(\pi)x_H(\tau_L - 1, 12 + 1 - k) = \sum_{k=1}^{h_{MF}} \left( \sum_{l=1}^{s} \pi_l k^l \right) x_H(\tau_L - 1, 12 + 1 - k)$$

$$= [x_H(\tau_L - 1, 12 + 1 - 1), \ldots, x_H(\tau_L - 1, 12 + 1 - h_{MF})] \begin{bmatrix} (1 - 1)^{1-1} \cdots (1 - 1)^{s-1} \\ \vdots \\ (h_{MF} - 1)^{1-1} \cdots (h_{MF} - 1)^{s-1} \end{bmatrix} [\pi_1 \cdots \pi_s]. $$

Another important characteristic of the Almon polynomial is that it allows both signs in general (e.g. $w_k(\pi) \geq 0$ for $k < 3$ and $w_k(\pi) < 0$ for $k \geq 4$, etc.). Many other MIDAS polynomials, like the beta probability density or exponential Almon, assume a single sign for all lags (e.g. $w_k(\pi) \geq 0$ for all $k$).

### 2.2.2 Low Frequency Approach

Using an aggregated high frequency variable $x_H(\tau_L) = \sum_{j=1}^{m} \delta_j x_H(\tau_L, j)$, let us consider a low frequency counterpart to the parsimonious regression models (2.18):

**Low Frequency Parsimonious Regression Models for Low-to-High Granger Causality**

$$x_L(\tau_L) = \sum_{k=1}^{q} \alpha_{k,j} x_L(\tau_L - k) + \sum_{k=1}^{h_{LF}} \beta_{k,j} x_H(\tau_L - k) + \gamma_{j} x_H(\tau_L + j) + u_L,j(\tau_L), \quad j = 1, \ldots, r_{LF}.$$  \hspace{1cm} (2.21)

Subscript "LF" is put on $h$ and $r$ in order to emphasize that they signify the number of low frequency lags and leads of aggregated $x_H$, respectively. The subscript will be dropped for brevity when there is not a potential confusion. We run OLS for each parsimonious model (2.21) and then formulate a low frequency max test statistic as in (2.19): $U_{LF} = \max_{1 \leq j \leq r_{LF}} \left( \sqrt{T_L w_{T_L,j \tau_L} \hat{\gamma}_j} \right)^2$.

It is important to note that deriving the limit distribution of $U_{LF}$ under $H_0 : x_L \not\rightarrow x_H$ requires an extra assumption other than Assumptions 2.1-2.5. Under $H_0$, the true DGP boils down to (2.2). In general, each low frequency parsimonious regression model does not contain (2.2) as a special case. The true high-to-low causal pattern, $\sum_{l=1}^{m} b_l x_H(\tau_L - 1, m + 1 - l)$, may not be fully captured by the low
frequency lags of aggregated $x_H$, $\sum_{k=1}^{h_{LF}} \beta_{k,j}x_H(\tau_L - k)$, no matter what the lag length $h_{LF}$ is.

To find a condition ensuring that each low frequency parsimonious regression model contains (2.2) as a special case, we elaborate the relationship between the two summation terms $\sum_{l=1}^{pm} b_l x_H(\tau_L - 1, m + 1 - l)$ and $\sum_{k=1}^{h_{LF}} \beta_{k,j}x_H(\tau_L - k)$. We can drop subscript $j \in \{1, \ldots, r_{LF}\}$ here since it is irrelevant which lead term of $x_H$ is included in the model. We thus write $\beta_k$ instead of $\beta_{k,j}$. Observe:

$$\sum_{k=1}^{h_{LF}} \beta_k x_H(\tau_L - k) = \sum_{k=1}^{h_{LF}} \beta_k \sum_{l=1}^{m} \delta_l x_H(\tau_L - k, l)$$

$$= \beta_1 \delta_1 x_H(\tau_L - 1, m + 1 - 1) + \cdots + \beta_1 \delta_1 x_H(\tau_L - 1, m + 1 - m)$$

$$+ \cdots + \beta_{h_{LF}} \delta_{h_{LF}} x_H(\tau_L - h_{LF}, m + 1 - 1) + \cdots + \beta_{h_{LF}} \delta_{h_{LF}} x_H(\tau_L - h_{LF}, m + 1 - m)$$

(2.22)

$$= \sum_{l=1}^{h_{LF}} \beta_{[l/m]} \delta_{[l/m]} m + 1 - l x_H(\tau_L - 1, m + 1 - l),$$

where $[x]$ is the smallest integer not smaller than $x$. The last equality of (2.22) exploits the notational convention that the second argument of $x_H$ can go below 1. See Appendix A for more details. Comparing the last term in (2.22) and the true causal pattern $\sum_{l=1}^{pm} b_l x_H(\tau_L - 1, m + 1 - l)$, we impose the following extra assumption.

**Assumption 2.6.** Fix a linear aggregation scheme $\delta = [\delta_1, \ldots, \delta_m]'$. Also fix $b = [b_1, \ldots, b_{pm}]'$, a true causal pattern from $x_H$ to $x_L$. We assume that there exists a real vector $\beta^* = [\beta^*_1, \ldots, \beta^*_p]'$ such that $b_l = \beta^*_{[l/m]} \delta_{[l/m]} m + 1 - l$ for all $l \in \{1, \ldots, pm\}$ and that a sufficiently large lag length $h_{LF} \geq p$ is chosen.

Assumption 2.6 ensures that the parameter vector $\beta$ has pseudo-true values such that $\sum_{k=1}^{h_{LF}} \beta_k^* x_H(\tau_L - k) = \sum_{l=1}^{pm} b_l x_H(\tau_L - 1, m + 1 - l)$ and therefore each parsimonious regression model (2.21) is correctly specified under $H_0 : x_L \rightarrow x_H$.

If there is non-causality from $x_H$ to $x_L$ (i.e. $b = 0_{pm \times 1}$), then Assumption 2.6 is trivially satisfied by choosing any $h_{LF} \in \mathbb{N}$ and letting $\beta_k^* = 0$ for all $k \in \{1, \ldots, h_{LF}\}$. If there is causality (i.e. $b \neq 0_{pm \times 1}$), then Assumption 2.6 is a relatively stringent assumption as described in Examples 2.4 and 2.5.

**Example 2.4** (Lagged Causality and Stock Sampling). Assume $m = 3$ and $p = 2$ and consider lagged causality $b_l = b \times I(l = 4)$ for $l \in \{1, \ldots, 6\}$ with $b \neq 0$. This means that only $x_H(\tau_L - 2, 3)$ has a nonzero coefficient $b$ and all other terms $x_H(\tau_L - 1, 3), x_H(\tau_L - 1, 2), x_H(\tau_L - 1, 1), x_H(\tau_L - 2, 2), x_H(\tau_L - 2, 1)$ have no impact on $x_L(\tau_L)$. This can be thought of as an extreme example of delayed information transmission or seasonality. We will show that such a causal pattern can be captured by the low frequency parsimonious regression models if and only if a linear aggregation scheme is stock sampling.

We first show that the true causal pattern can be captured under stock sampling. Since stock sampling is represented as $\delta_l = I(l = 3)$ for $l \in \{1, 2, 3\}$, the summation term included in each parsimonious regression models, $\sum_{k=1}^{h_{LF}} \beta_k x_H(\tau_L - k)$, can be rewritten as $\sum_{k=1}^{h_{LF}} (\beta_k x_H(\tau_L - k, 3)$. Thus, we can simply choose $h_{LF} = 2, \beta_1^* = 0$, and $\beta_2^* = b$ to replicate the true causal pattern.

16
Next we show that the true causal pattern cannot be captured under any other linear aggregation scheme than stock sampling. Assume $\delta_l \geq 0$ for all $l \in \{1, 2, 3\}$, $\delta_3 < 1$, and $\sum_{l=1}^{3} \delta_l = 1$. This aggregation scheme excluded stock sampling but includes any other. Since $\delta_3 < 1$, at least one of $\delta_1$ and $\delta_2$ should have a positive value. Assume $\delta_1 > 0$ without loss of generality. Assumption 2.6, on one hand, requires that $b_4 = \beta_2^* \delta_3$ and $b_6 = \beta_2^* \delta_1$. The true causal pattern, on the other hand, implies that $b_4 = b_6 = 0$ and $b_6 = 0$. Since $\delta_1 > 0$, there does not exist any $\beta_2^*$ that satisfies these four equalities simultaneously.

**Example 2.5** (Flow Sampling). When flow sampling is considered, each low frequency parsimonious regression model is most likely misspecified under realistic causal patterns. Since $\delta_l = 1/m$ for all $l \in \{1, \ldots, m\}$, Assumption 2.6 requires that $b_l = \beta_{[l/m]}^* m$. This implies that $b_1 = \cdots = b_{m}$, $b_{m+1} = \cdots = b_{2m}$, and so on. In words, Assumption 2.6 holds only when all $m$ high frequency lags of $x_H$ in each low frequency period have an identical coefficient. This is an unrealistically strong requirement since we often have both positive and negative signs, lagged causality, or decaying causality in practice.

It is straightforward to derive the asymptotic distribution of low frequency max test statistic $U_{LF}$ under $H_0 : x_L \rightarrow x_H$. We write $h$ and $r$ instead of $h_{LF}$ and $r_{LF}$ for brevity. Define an $n \times 1$ vector $y_j(\tau_L - 1) = [x_L(\tau_L - 1), \ldots, x_L(\tau_L - q), x_H(\tau_L - 1), \ldots, x_H(\tau_L - h), x_H(\tau_L + j)]'$. $y_j(\tau_L - 1)$ is a vector of all regressors in model $j$. Given Assumptions 2.1, 2.2, 2.3, 2.4, and 2.6, we can derive the asymptotic distribution of $U_{LF}$ under $H_0 : x_L \rightarrow x_H$. Just repeat the derivation used in the low-to-high max test, replacing the regressors $y_j(\tau_L - 1)$ with $y_j(\tau_L - 1)$. Note that here we need Assumption 2.6, not 2.5. Assumption 2.6 is a strong restriction excluding many kinds of high-to-low causality, as shown in Examples 2.4 and 2.5.

### 3 Local Asymptotic Power Analysis

For each of high-to-low causality and low-to-high causality, Section 2 has introduced four different tests: mixed frequency max test, mixed frequency Wald test, low frequency max test, and low frequency Wald test. The goal of Section 3 is to compare the local asymptotic power of these four tests, focusing on high-to-low causality.\(^5\) In terms of consistency, Table 1 shows that the MF max test and the MF Wald test are equally good. Both of them achieve power approaching 1 for any form of causality, given the same condition that the selected number of high frequency lags $h$ is larger than or equal to the true lag order $pm$. It is of interest to check if there is any difference between them in terms of local power.

While Table 1 says that the LF max test and LF Wald test are inconsistent, it does not mean that they are always useless. Their power still approaches 1 under *some* form of Granger causality, though not all. An advantage of the low frequency approach is that we often have fewer parameters than in the mixed frequency approach. Hence, there is a chance that the low frequency approach has in fact higher local power than the mixed frequency approach does, depending on causal patterns. It is thus worth comparing all four tests carefully.

\(^5\)The low-to-high case remains as a future task.
We keep imposing Assumptions 2.1, 2.2, 2.3, and 2.4. Consider the same DGP (2.1) again. Our null hypothesis is the same as before (i.e. $H_0 : \mathbf{b} = \mathbf{0}_{pm \times 1}$), but here we consider a local alternative hypothesis $H_1' : \mathbf{b} = (1/\sqrt{T_L})\mathbf{\nu}$. In the hypothesis testing literature $\mathbf{\nu} = [\nu_1, \ldots, \nu_{pm}]'$ is often called the Pitman drift. Under $H_1'$, the last row of the MF-VAR data generating process is written as

$$
x_L(\tau_L) = \sum_{k=1}^{p} a_k x_L(\tau_L - k) + \sum_{j=1}^{pm} \frac{\nu_j}{\sqrt{T_L}} x_H(\tau_L - 1, m + 1 - j) + \epsilon_L(\tau_L)
$$

$$= \mathbf{X}_L(\tau_L - 1)' \mathbf{a} + \mathbf{X}_H(\tau_L - 1)' \left( \frac{1}{\sqrt{T_L}} \mathbf{\nu} \right) + \epsilon_L(\tau_L),$$

where $\mathbf{X}_L(\tau_L - 1) = [x_L(\tau_L - 1), \ldots, x_L(\tau_L - p)]'$, $\mathbf{X}_H(\tau_L - 1) = [x_H(\tau_L - 1, m + 1 - 1), \ldots, x_H(\tau_L - 1, m + 1 - pm)]'$, and $\mathbf{a} = [a_1, \ldots, a_p]'$. Based on this DGP, we derive the limit distribution of each of the four test statistics under $H_1'$.

### 3.1 Mixed Frequency Approach

In this section we consider the mixed frequency parsimonious regression and then the mixed frequency naïve regression.

**Mixed Frequency Parsimonious Regression** We combine $h$ parsimonious regression models (2.5) in order to formulate the mixed frequency max test statistic $T$ defined in (2.7). The asymptotic distribution under $H_0 : \mathbf{b} = \mathbf{0}_{pm \times 1}$ is already derived in Theorem 2.1. Here we derive the asymptotic distribution under $H_1' : \mathbf{b} = (1/\sqrt{T_L})\mathbf{\nu}$.

**Theorem 3.1.** Let Assumptions 2.1, 2.2, 2.3, and 2.4 hold. Then, we have that $T \overset{d}{\rightarrow} \max_{1 \leq i \leq h} \mathcal{M}_i^2$ as $T_L \rightarrow \infty$ under $H_1' : \mathbf{b} = (1/\sqrt{T_L})\mathbf{\nu}$. \(\mathcal{M} = [\mathcal{M}_1, \ldots, \mathcal{M}_h]'\) is a vector-valued random variable drawn from $N(\mathbf{\mu}, \mathbf{V})$. $\mathbf{V}$ is defined in (2.8) and

$$\mathbf{\mu}_{h \times 1} = \mathbf{W}_h \mathbf{R} \begin{bmatrix} \Gamma_{1,1}^{-1} \mathbf{C}_1 \\ \vdots \\ \Gamma_{h,h}^{-1} \mathbf{C}_h \end{bmatrix} \mathbf{\nu},$$

where $\Gamma_{j,i} = E[\mathbf{x}_j(\tau_L - 1)\mathbf{x}_j(\tau_L - 1)']$ and $\mathbf{C}_j = E[\mathbf{x}_j(\tau_L - 1)\mathbf{X}_H(\tau_L - 1)']$ as defined in (2.10). See (2.7) and (2.9) for the definitions of weighting scheme $\mathbf{W}_h$ and selection matrix $\mathbf{R}$, respectively.

**Proof 3.1.** See Appendix F.

Comparing Theorems 2.1 and 3.1, the asymptotic distribution under $H_0$ and the asymptotic distribution under $H_1'$ share the same covariance matrix $\mathbf{V}$ but only the latter has a nonzero location parameter $\mathbf{\mu}$. We see that $\mathbf{V}$ and $\mathbf{\mu}$ depend on two population covariance terms: $\Gamma_{j,i}$ and $\mathbf{C}_j$ for $j, i \in \{1, \ldots, h\}$. Recall from (2.9) that $\Gamma_{j,i} = E[\mathbf{x}_j(\tau_L - 1)\mathbf{x}_i(\tau_L - 1)']$ is the $(q + 1) \times (q + 1)$ covariance matrix between all regressors in parsimonious regression model $j$ and all regressors in model $i$. Recall from (2.10) that
Let Assumptions 2.1 and 2.2 hold. Define the stationarity of \( \Upsilon \) be recursively computed by the Yule-Walker equation:

\[
\text{Theorem 3.2.}
\]

In general, an analytical computation of local asymptotic power requires an explicit characterization of \( \Gamma_j, \) and \( C_j \) in terms of underlying parameters \( A_1, \ldots, A_p, \Omega \) appearing in (2.1). To this end, we first construct the autocovariance matrix of the mixed frequency vector \( X(\tau_L) \). Recall that the dimension of \( X(\tau_L) \) is \( K = m + 1 \) and it follows covariance stationary MF-VAR(\( p \)) as stated in Assumptions 2.1 and 2.2. Hence, the autocovariance matrix of \( X(\tau_L) \) can be calculated from the classic argument involving the multivariate Yule-Walker equation. Define some matrices:

\[
\begin{align*}
\mathbf{\Upsilon}_k & = E[ X(\tau_L) X(\tau_L - k)^\top], \\
\mathbf{\Upsilon} & = \begin{bmatrix}
\mathbf{\Upsilon}_{1-1} & \cdots & \mathbf{\Upsilon}_{p-1} \\
\vdots & \ddots & \vdots \\
\mathbf{\Upsilon}_{1-p} & \cdots & \mathbf{\Upsilon}_{p-p}
\end{bmatrix}, \\
\mathbf{A} & = \begin{bmatrix}
\mathbf{A}_1 & \cdots & \mathbf{A}_{p-1} & \mathbf{A}_p \\
\mathbf{I}_K & \cdots & \mathbf{0}_{K \times K} & \mathbf{0}_{K \times K} \\
\vdots & \ddots & \vdots & \vdots \\
\mathbf{0}_{K \times K} & \cdots & \mathbf{I}_K & \mathbf{0}_{K \times K}
\end{bmatrix}, \\
\mathbf{\Omega} & = \begin{bmatrix}
\mathbf{\Omega} & \mathbf{0}_{K \times K} & \cdots & \mathbf{0}_{K \times K} \\
\mathbf{0}_{K \times K} & \mathbf{0}_{K \times K} & \cdots & \mathbf{0}_{K \times K} \\
\vdots & \ddots & \vdots & \vdots \\
\mathbf{0}_{K \times K} & \mathbf{0}_{K \times K} & \cdots & \mathbf{0}_{K \times K}
\end{bmatrix}.
\end{align*}
\]

Using the discrete Lyapunov equation, \( \mathbf{\Upsilon}_k \) for \( k \in \{1, \ldots, p-1\} \) can be computed by the following well-known formula:

\[
\text{vec}[\mathbf{\Upsilon}] = (I_{(pK)^2} - \mathbf{A} \otimes \mathbf{A})^{-1} \text{vec}[\mathbf{\Omega}],
\]  

where \( \text{vec}[\cdot] \) is a column-wise vectorization operator and \( \otimes \) is the Kronecker product. \( \mathbf{\Upsilon}_k \) for \( k \geq p \) can be recursively computed by the Yule-Walker equation: \( \mathbf{\Upsilon}_k = \sum_{l=1}^{p} \mathbf{A}_l \mathbf{\Upsilon}_{k-l} \). Finally, the covariance stationarity of \( X(\tau_L) \) ensures that \( \mathbf{\Upsilon}_k \) for \( k \leq -p \) can be computed by \( \mathbf{\Upsilon}_k = \mathbf{\Upsilon}'_k \).

Now that we have characterized \( \{ \mathbf{\Upsilon}_k \}_{k \in \mathbb{Z}} \) in terms of \( \mathbf{A}_1, \ldots, \mathbf{A}_p, \mathbf{\Omega} \), the next step is to characterize \( \Gamma_{j,i} \) and \( C_j \) in terms of \( \{ \mathbf{\Upsilon}_k \}_{k \in \mathbb{Z}} \). This requires heavy matrix algebra since indices \( j, i \in \{1, \ldots, h\} \) may go beyond \( m \) or even \( pm \). We thus present only the final outcome as Theorem 3.2 and put detailed derivation in Appendix G.

**Theorem 3.2.** Let Assumptions 2.1 and 2.2 hold. Define \( f(j) = [(j - m)/m] \) and \( g(j) = mf(j) + m + 1 - j \), where \( \lceil x \rceil \) is the smallest integer not smaller than \( x \). Let \( \mathbf{T}_k(s,t) \) be the \((s,t)\) element of \( \mathbf{\Upsilon}_k \equiv E[ X(\tau_L) X(\tau_L - k)^\top]. \) Then,

\[
\begin{align*}
\mathbf{\Gamma}_{j,i} & = \begin{bmatrix}
\mathbf{\Upsilon}_{1-1}(K,K) & \cdots & \mathbf{\Upsilon}_{1-q}(K,K) & \mathbf{\Upsilon}_{f(i)}(g(i),K) \\
\vdots & \ddots & \vdots & \vdots \\
\mathbf{\Upsilon}_{q-1}(K,K) & \cdots & \mathbf{\Upsilon}_{q-q}(K,K) & \mathbf{\Upsilon}_{(f(i)-1)}(g(i),K) \\
\mathbf{\Upsilon}_{f(j)}(K,g(j)) & \cdots & \mathbf{\Upsilon}_{f(j)-(q-1)}(K,g(j)) & \mathbf{\Upsilon}_{f(j)-f(i)}(g(i),g(j))
\end{bmatrix},
\end{align*}
\]

\[
\begin{align*}
\mathbf{C}_j & = \begin{bmatrix}
\mathbf{\Upsilon}_{f(1)}(K,g(1)) & \cdots & \mathbf{\Upsilon}_{f(pm)}(K,g(pm)) \\
\vdots & \ddots & \vdots \\
\mathbf{\Upsilon}_{f(1)-(q-1)}(K,g(1)) & \cdots & \mathbf{\Upsilon}_{f(pm)-(q-1)}(K,g(pm)) \\
\mathbf{\Upsilon}_{f(j)-f(1)}(g(1),g(j)) & \cdots & \mathbf{\Upsilon}_{f(j)-f(pm)}(g(pm),g(j))
\end{bmatrix}
\end{align*}
\]

for \( j, i \in \{1, \ldots, h\} \).
Proof 3.2. See Appendix G.

While a formal proof is collected in Appendix G, we briefly confirm that (3.4) is well-defined here. Note that $\Upsilon_k$ is $(m+1) \times (m+1)$. In order for (3.4) to make sense, it should be the case that $g(j)$ takes a natural number between 1 and $m+1$ for any $j \in \mathbb{N}$. We can verify that this is in fact the case. See Table 2 for illustration. By definition, $f(j)$ is a step function taking 0 for $j = 1, \ldots, m$, 1 for $j = m+1, \ldots, 2m$, 2 for $j = 2m+1, \ldots, 3m$, etc. $g(j)$ takes $m$, $m-1$, $\ldots$, 1 as $j$ runs from $(k-1)m+1$ to $km$ for any $k \in \mathbb{N}$. Hence, (3.4) is always well-defined.

We are now ready to compute the local asymptotic power associated with the mixed frequency max test numerically. The procedure is as follows.

Step 1 Starting with underlying parameters $A_1, \ldots, A_p, \Omega$, use Theorems 2.1, 3.1, and 3.2 to calculate $\mu$ and $V$.

Step 2 Draw $R_1$ samples $\mathcal{N}^{(1)}, \ldots, \mathcal{N}^{(R_1)}$ independently from the limit distribution under $H_0$, $N(0_{h \times 1}, V)$, and calculate a set of test statistics $T_r = \max_{1 \leq i \leq h}(\mathcal{N}_i^{(r)})^2$.

Step 3 Sort the test statistics $T_{(1)} \leq \cdots \leq T_{(R_1)}$ and take the $100(1-\alpha)$% quantile, which is a numerical approximation of the critical value associated with a nominal size $\alpha$. Call that quantile $d^\star$.

Step 4 Draw $R_2$ samples $\mathcal{M}^{(1)}, \ldots, \mathcal{M}^{(R_2)}$ independently from the limit distribution under $H_1^1$, $N(\mu, V)$, and calculate another set of test statistics $\tilde{T}_r = \max_{1 \leq i \leq h}(\mathcal{M}_i^{(r)})^2$. Local asymptotic power $\mathcal{P}$ is given by $\mathcal{P} = (1/R_2) \sum_{r=1}^{R_2} I(\tilde{T}_r > d^\star)$.

Mixed Frequency Naïve Regression To derive local asymptotic power associated with the mixed frequency Wald test, we rewrite the mixed frequency naive regression model (2.4) in matrix form:

$$x_L(\tau_L) = \sum_{k=1}^{q} \alpha_k x_L(\tau_L - k) + \sum_{j=1}^{h} \beta_j x_H(\tau_L - 1, m+1-j) + u_L(\tau_L)$$

$$= [x_L(\tau_L - 1), \ldots, x_L(\tau_L - q)] \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_q \end{bmatrix} + x_H(\tau_L - 1, m+1-1), \ldots, x_H(\tau_L - 1, m+1-h)] \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_h \end{bmatrix} + u_L(\tau_L)$$

$$= \mathbf{X}_L^{(q)}(\tau_L - 1)' \mathbf{\alpha} + u_L(\tau_L).$$

$$= \mathbf{X}_L^{(q)}(\tau_L - 1)' \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + u_L(\tau_L).$$

$$= \mathbf{X}_H^{(h)}(\tau_L - 1)' \begin{bmatrix} \alpha \\ \beta \end{bmatrix} + u_L(\tau_L).$$

Note that $\mathbf{X}_H^{(h)}(\tau_L - 1)$ is a vector stacking $h$ high frequency lags of $x_H$, while $\mathbf{X}_H(\tau_L - 1)$ is a vector stacking $pm$ high frequency lags of $x_H$. The true DGP for $x_L$ is the same as before and can be found in (2.2). We keep imposing Assumption 2.4 that $q \geq p$.

We formulate a Wald statistic $W$ with respect to $H_0 : b = 0_{pm \times 1}$. As is well known, its asymptotic distribution under $H_0$ is $\chi^2_{k}$. It is also well-known that the asymptotic distribution of $W$ under $H_1^1 : b =$
Starting with underlying parameters depend on population moments. We combine calculate local asymptotic power. We now derive local asymptotic power associated with the low frequency max test. The simulation procedure for computing the local asymptotic power associated with the low frequency Wald test exactly. The procedure is summarized as follows.

Step 1 Starting with underlying parameters $A_1, \ldots, A_p, \Omega$, calculate noncentrality $\kappa$.

Step 2 Calculate local asymptotic power $\mathcal{P}$ according to $\mathcal{P} = 1 - F_1[F_0^{-1}(1 - \alpha)]$, where $\alpha$ is a nominal size. $F_0$ is the cumulative distribution function (c.d.f.) of the Wald statistic $W$ under $H_0$ (i.e. $\chi^2_h(\kappa)$), while $F_1$ is the c.d.f. of $W$ under $H_1^\text{L}$ (i.e. $\chi^2_h(\kappa)$).

3.2 Low Frequency Approach

In this section we consider the low frequency parsimonious regression and then the low frequency naïve regression. We keep imposing Assumption 2.4 that $q \geq p$. We do not impose any assumption about the magnitude of $h$, however. The number of high frequency lags $h$ can be smaller than, equal to, or larger than the truth $pm$.

Low Frequency Parsimonious Regression We combine $h$ low frequency parsimonious regression models (2.15): $x_L(\tau_L) = \bar{x}_j(\tau_L - 1) \theta_{LF} + u_{LF}(\tau_L)$, where $\bar{x}_j(\tau_L - 1) = [x_L(\tau_L - 1), \ldots, x_L(\tau_L - q), x_H(\tau_L - j)]$ with $x_H(\tau_L - j) = \sum_{l=1}^m \delta_l x_H(\tau_L - j, l)$. We then formulate the low frequency max test statistic $T_{LF}$. The asymptotic distribution of $T_{LF}$ under $H_1^\text{L} : b = (1/\sqrt{T_L})\nu$ is the same as in Theorem 3.1 with the only difference being that $x_j(\tau_L - 1)$ there should be replaced with $\bar{x}_j(\tau_L - 1)$. As a result, computing the local power of the low frequency max test requires an analytical characterization of $\sum_{j,i} E[\bar{x}_j(\tau_L - 1)\bar{x}_i(\tau_L - 1)]$ and $\mathcal{C}_j = E[\bar{x}_j(\tau_L - 1)X_H(\tau_L - 1)]$. Algebra is even messier than in Theorem 3.2 due to the presence of aggregation scheme $\{\delta_1, \ldots, \delta_m\}$. We do not report final results and derivations in order to save space, but they are available upon request.

The simulation procedure for computing the local asymptotic power associated with the low frequency max test is omitted since it is identical to the mixed frequency case.

Low Frequency Naïve Regression We now derive local asymptotic power associated with the low frequency naïve regression model (2.12): $x_L(\tau_L) = \bar{x}(\tau_L - 1) \theta_{LF} + u_L(\tau_L)$, where $\bar{x}(\tau_L - 1) = [x_L(\tau_L - 1), \ldots, x_L(\tau_L - q), x_H(\tau_L - 1), \ldots, x_H(\tau_L - h)]$. The true DGP is the same as before and can be found in (2.2).

We formulate a Wald statistic $W_{LF}$ with respect to $H_0 : b = 0_{pm \times 1}$. Its asymptotic distribution is $\chi^2_h$ under $H_0$ and $\chi^2_h(\kappa)$ under $H_1^\text{L}$. Noncentrality $\kappa$ depend on population moments $\Gamma = E[\bar{x}(\tau_L -
Decaying Causality

Lagged Causality

where the true DGP follows a structural MF-VAR(1) process with $\nu$

\[ (\nu_1, \ldots, \nu_1) \]

By numerical examples we compare the four different tests discussed above: mixed frequency max test, mixed frequency Wald test, low frequency max test, and low frequency Wald test. We assume that the true DGP follows a structural MF-VAR(1) process with $m = 12$:

$$
\begin{bmatrix}
1 & 0 & \ldots & \ldots & 0 \\
-d & 1 & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
0 & \ldots & -d & 1 & 0 \\
0 & 0 & \ldots & 0 & 1
\end{bmatrix}
\begin{bmatrix}
\eta(\tau_L, 1) \\
\eta(\tau_L, 12) \\
\vdots \\
\eta(\tau_L)
\end{bmatrix}
\begin{bmatrix}
x_H(\tau_L, 1) \\
x_H(\tau_L, 12) \\
\vdots \\
x_H(\tau_L - 1)
\end{bmatrix}
= X(\tau_L)
\begin{bmatrix}
0 & 0 & \ldots & \ldots & 0 \\
0 & 0 & \ldots & \ldots & 0 \\
0 & 0 & \ldots & \ldots & 0 \\
0 & 0 & \ldots & \ldots & 0
\end{bmatrix}
\begin{bmatrix}
c_1 \\
c_2 \\
\vdots \\
\alpha
\end{bmatrix}
+ \begin{bmatrix}
0 \\
\vdots \\
0 \\
\eta(\tau_L)
\end{bmatrix}

\text{(3.6)}
$$

and thus $A_1 \equiv N^{-1} M_1 =

$$
\begin{bmatrix}
1 & 0 & \ldots & \ldots & 0 \\
\vdots \\
0 & 0 & \ldots & 0 & 1
\end{bmatrix}
\begin{bmatrix}
0 & 0 & \ldots & \ldots & 0 \\
0 & 0 & \ldots & \ldots & 0 \\
0 & 0 & \ldots & \ldots & 0 \\
0 & 0 & \ldots & \ldots & 0
\end{bmatrix}
\begin{bmatrix}
\sum_{i=1}^{12} d^{1-i} c_i \\
\sum_{i=1}^{2} d^{2-i} c_i \\
\vdots \\
\sum_{i=1}^{12} d^{12-i} c_i
\end{bmatrix}
+ \begin{bmatrix}
0 \\
\vdots \\
0 \\
\alpha
\end{bmatrix}

\text{(3.7)}
$$

Define $\epsilon(\tau_L) = N^{-1} \eta(\tau_L)$, then $\Omega \equiv E[\epsilon(\tau_L) \epsilon(\tau_L)'] = N^{-1} N^{-1}$. The reduced form of (3.6) is then simply

\[ X(\tau_L) = A_1 X(\tau_L - 1) + \epsilon(\tau_L). \]

As in the previous sections, we consider the local alternative hypothesis $H_1' : b = (1/\sqrt{T}) \nu$. We prepare three types of $\nu = [\nu_1, \ldots, \nu_{12}]'$.

1. Decaying Causality: $\nu_j = (-1)^{j-1} \times 2.5/j$ for $j = 1, \ldots, 12$. In this case the impact of $x_H$ on $x_L$ decays gradually with signs alternating. Specifically, $\nu = [2.5, -1.25, 0.83, \ldots, -0.21]'$.

2. Lagged Causality: $\nu_j = 2 \times I(j = 12)$ for $j = 1, \ldots, 12$. Only $\nu_{12}$ is $2$ and all others are zeros. This case is an extreme example of seasonality or lagged response of $x_L$ to $x_H$.

3. Sporadic Causality: $(\nu_3, \nu_9, \nu_{11}) = (2.1, -2.8, 1.9)$ and all other $\nu$’s are zeros. This is a complicated (but realistic) case where we have both positive and negative signs, lags are unevenly-lagged.
spaced, and the coefficients are not monotonically decreasing in absolute values. Such an economic interrelationship should not be uncommon in practice due to lagged information transmission, seasonality, feedback effects, ambiguous theoretical relations in terms of signs, etc.

Other parameters are specified as follows. First, we assume fairly weak autoregressive property for $x_L$ (i.e. $a = 0.2$). The choice of $a$ does not affect local power much, however. Numerical evidence for different $a$’s is available upon request. Second, we consider two values for the persistence of $x_H$: $d \in \{0.2, 0.8\}$. Third, we assume decaying causality with alternating signs for low-to-high causality: $c_j = (-1)^{j-1} \times 0.8/j$ for $j = 1, \ldots, 12$.

We now explain our models. For all four tests, we include two low frequency lags of $x_L$ (i.e. $q = 2$). Having $q = 1$ would be sufficient since the true DGP is MF-VAR(1) here, but the true lag order is typically unknown. For max tests, the weighting scheme is simply an equal scheme: $W_h = (1/h) \times I_h$. The number of draws from the limit distributions under $H_0$ and $H_1$ is 100,000 each. The number of high frequency lags of $x_H$ taken by mixed frequency tests is $h_{MF} \in \{4, 8, 12\}$, while the number of low frequency lags of $x_H$ taken by low frequency tests is $h_{LF} \in \{1, 2, 3\}$. For the low frequency tests, we consider both flow sampling (i.e. $\delta_k = 1/12$ for $k = 1, \ldots, 12$) and stock sampling (i.e. $\delta_k = 1(k = 12)$ for $k = 1, \ldots, 12$). Nominal size $\alpha$ is fixed at 0.05.

Table 3 compares the local asymptotic power of MF max test, MF Wald test, LF max test, and LF Wald test. Panel A considers Decaying Causality, Panel B considers Lagged Causality, and Panel C considers Sporadic Causality. For each panel we consider low persistence of $x_H$ (i.e. $d = 0.2$) and high persistence (i.e. $d = 0.8$).

We start with Panel A, Decaying Causality. Focusing on the low persistence case $d = 0.2$ in Panel A.1, the mixed frequency cases have moderately high power between 0.346 and 0.570. For example, the mixed frequency max test with $h_{MF} = 4$ has power 0.487, while the mixed frequency Wald test with $h_{MF} = 4$ has power 0.570.

The low frequency tests with flow sampling have absolutely no power regardless of the number of lags $h_{LF} \in \{1, 2, 3\}$. The lowest value is 0.063 and the largest value is 0.076. An essential reason for this poor performance is that we have alternating signs in $\nu$ and flow aggregation offsets all those effects. The low frequency tests with stock sampling, in contrast, have very high power between 0.495 and 0.643, which is often higher than the mixed frequency cases. For example, the low frequency Wald test with stock sampling and $h_{LF} = 1$ has power 0.643. The reason for this high performance is that the largest coefficient $\nu_1 = 2.5$ is assigned to $x_H(\tau_L - 1, 12)$, which is a regressor exactly included in the low frequency models with stock sampling by coincidence.

If the persistence parameter $d$ is raised from 0.2 to 0.8, local power rises in general, but all qualitative implications above still hold (cfr. Panel A.2). While mixed frequency tests have high power, low frequency tests with stock sampling often have even higher power. Low frequency tests with flow sampling have low power.

We turn on to Panel B, Lagged Causality. Focusing on the high persistence case $d = 0.8$ in Panel B.2, mixed frequency cases have power that is increasing in $h_{MF}$. For example, local power of the mixed
frequency max test is 0.075, 0.181, and 0.769 when $h_{MF}$ is 4, 8, and 12, respectively. This reflects the causal pattern that only $\nu_{12}$, coefficient of $x_H(\tau_L - 1, 1)$, is 2 and all other $\nu$'s are zeros. It is thus important to include sufficiently many lags when we apply mixed frequency tests.

The low frequency tests with flow sampling have reasonably high power regardless of $h_{LF}$. The power of the low frequency max test, for instance, is 0.455, 0.468, and 0.415 when $h_{LF}$ is 1, 2, and 3, respectively. This is because the low frequency models work on an aggregated $x_H$ and hence taking only a few lags tends to be enough. Another important reason for this good performance is that our causal effect is unambiguously positive; we have only one positive coefficient $\nu_{12} = 2$ and no negative coefficients at all. Flow aggregation preserves the original causality in such a case. The low frequency tests with stock sampling, in contrast, have absolutely no power at $h_{LF} = 1$. This is expected since the only regressor $x_H(\tau_L - 1, 12)$ has a zero coefficient by construction. They have high power when $h_{LF} = 2$ (0.676 for max test and 0.664 for Wald test). This is because their extra regressor $x_H(\tau_L - 2, 12)$ has a strong correlation with the adjacent term $x_H(\tau_L - 1, 1)$, which has a nonzero coefficient $\nu_{12} = 2$. Such a spillover effect matters only when coefficient $d$, the persistence of $x_H$, is relatively large.

When the persistence parameter $d$ decreases from 0.8 to 0.2, local power falls sharply in general, but the implications above still hold (cfr. Panel B.1). The mixed frequency max test and Wald test with $h_{MF} = 12$ have power 0.247 and 0.206 respectively. The low frequency tests with flow sampling have power around 0.1. The low frequency tests with stock sampling have absolutely no power as expected.

We now discuss Sporadic Causality in Panel C, which highlights the full advantage of the mixed frequency tests over the low frequency tests. Fixing $d = 0.2$, the mixed frequency max test has power 0.391, 0.323, and 0.677 when $h_{MF}$ is 4, 8, and 12, respectively. Similarly, the mixed frequency Wald test has power 0.365, 0.291, and 0.761. Their power declines when switching from $h_{MF} = 4$ to $h_{MF} = 8$ since $\nu_5$, $\nu_6$, $\nu_7$, and $\nu_8$ are all zeros and thus the extra number of parameters gets penalized. The low frequency tests, whether flow sampling or stock sampling, have absolutely no power. The smallest value is 0.051 and the largest value is 0.072. This is an expected consequence since we learned from Panel A that flow sampling is vulnerable to alternating signs and from Panel B that stock sampling is vulnerable to lagged causality. In the presence of a tricky (but realistic) causality like Sporadic Causality, mixed frequency tests with sufficiently many lags show a great advantage in terms of local asymptotic power.

While we have highlighted the advantage of mixed frequency tests relative to low frequency tests, it is not clear from Table 3 how the MF max test is preferred to the MF Wald test. There are $3 \times 3 \times 2 = 18$ ways of comparing them since we have three causal patterns, three values for lag length $h_{MF}$, and two values for persistence $d$. In 12 out of 18 cases the MF max test has higher power than the MF Wald test, but not in the other 6 cases. While the difference between the max test and the Wald test takes the largest value of $0.769 - 0.560 = 0.209$ for Lagged Causality with $(d, h_{MF}) = (0.8, 12)$, it takes the smallest value of $0.690 - 0.872 = -0.182$ for Sporadic Causality with $(d, h_{MF}) = (0.8, 12)$. We thus cannot firmly assert that the MF max test achieves higher local power than the MF Wald test. This is not surprising since the superiority of the former is supposed to appear in small sample (cfr. Section 4).

We tried a MF-VAR(2) data generating process for robustness check, and got the same implications as in MF-VAR(1). All results are omitted to save space but available upon request.
4 Monte Carlo Simulations

In this section we run Monte Carlo simulations to compare the finite sample power of the mixed frequency max test, mixed frequency Wald test, low frequency max test, and low frequency Wald test. Section 4.1 is concerned with high-to-low causality, while Section 4.2 is concerned with low-to-high causality.

4.1 High-to-Low Granger Causality

We first assume that the true DGP follows a structural MF-VAR(1) process with $m = 12$. We then consider MF-VAR(2) as a robustness check.

**MF-VAR(1)** Unlike the local power analysis, we actually have to simulate samples from the MF-VAR(1) data generating process (3.6). We draw mutually and serially independent standard normal random numbers for $\eta(\tau_L)$, the error term in the structural form.

The null hypothesis is again non-causality $H_0: \mathbf{b} = 0_{12 \times 1}$, while the alternative hypothesis is general causality $H_1: \mathbf{b} \neq 0_{12 \times 1}$. We consider four causal patterns below.

1. **Non-causality**: $\mathbf{b} = 0_{12 \times 1}$. In this case we can check the empirical size of each test.

2. **Decaying Causality**: $b_j = (-1)^{j-1} \times 0.3/j$ for $j = 1, \ldots, 12$. In this case the impact of $x_H$ on $x_L$ decays gradually with signs alternating. Specifically, $\mathbf{b} = [0.3, -0.15, 0.1, \ldots, -0.025]'$.

3. **Lagged Causality**: $b_j = 0.3 \times I(j = 12)$ for $j = 1, \ldots, 12$. Only $b_{12}$ is nonzero at 0.3 and all others are zeros. This case is an extreme example of seasonality or lagged response of $x_L$ to $x_H$.

4. **Sporadic Causality**: $(b_3, b_7, b_{10}) = (0.2, 0.05, -0.3)$ and all other $b$’s are zeros. This is a realistic case where we have both positive and negative signs as well as unevenly-spaced lags. Such a complicated economic interrelationship should not be uncommon in practice.

Other quantities are chosen as follows. First, we assume fairly weak autoregressive property for $x_L$ (i.e. $a = 0.2$). The choice of $a$ does not affect rejection frequencies much, however. Second, we consider two values for the persistence of $x_H$: $d \in \{0.2, 0.8\}$. Third, we assume decaying causality with alternating signs for low-to-high causality: $c_j = (-1)^{j-1} \times 0.4/j$ for $j = 1, \ldots, 12$.

Sample size in terms of low frequency is $T_L \in \{40, 80\}$. Since $m = 12$, our experimental design can be approximately thought of as week vs. quarter. In this case having $T_L = 40$ means that the low frequency sample size is 10 years, a fairly small sample. Having $T_L = 80$ means that the low frequency sample size is 20 years, a medium sample size. Alternatively, we could think of month vs. year in which case $m$ is exactly 12. In that scenario, having $T_L = 40$ means that the low frequency sample size is 40 years, a large sample but not totally unrealistic. We focus on the week vs. quarter scenario hereafter.

Given the large ratio of sampling frequencies $m = 12$, the mixed frequency Wald test (and potentially low frequency Wald test too) would suffer from serious size distortions without bootstrap. We thus use a parametric bootstrap of Gonçalves and Killian (2004), which is designed to be robust against
conditionally heteroskedastic errors, for both MF and LF Wald tests. Our actual DGP has i.i.d. errors, but in practice we do not have such a prior knowledge and often use Gonçalves and Killian’s (2004) bootstrap to take care of potential GARCH effects. The procedure of their bootstrap is as follows.

**Step 1** Let us take the mixed frequency naïve regression model (3.5) as an example. The low frequency case is completely analogous. Run unrestricted OLS for $x_L(\tau_L) = x(\tau_L - 1)\Theta + u_L(\tau_L)$ to get $\hat{\Theta}$ and then compute the Wald statistic $W$.

**Step 2** Run OLS for a restricted model $x_L(\tau_L) = x(\tau_L - 1)\Theta_0 + u_L(\tau_L)$ where $\Theta_0 = [\alpha', 0_{1\times h}]'$ to get $\hat{\Theta}_0$ and $\hat{u}_L(\tau_L)$. Note that this is a pure AR($q$) model since the null hypothesis of non-causality is imposed.

**Step 3** Simulate $N$ samples from $x_L(\tau_L) = x(\tau_L - 1)\hat{\Theta}_0 + \hat{u}_L(\tau_L)v(\tau_L)$, where $v(\tau_L)$ i.i.d. $\sim N(0, 1)$.

**Step 4** For each sample compute Wald statistics $\tilde{W}_1, \ldots, \tilde{W}_N$ and calculate the bootstrapped $p$-value: $p_N = [1 + \sum_{i=1}^N I(\tilde{W}_i \geq W)]/(N + 1)$. The null hypothesis of non-causality is rejected at level $\alpha$ if $p_N \leq \alpha$.

We now explain the details of our models and simulation design. For all four tests, we include two low frequency lags of $x_L$ (i.e., $q = 2$). Having $q = 1$ would be sufficient since the true DGP is MF-VAR(1) here, but the true lag order is typically unknown. For max tests, the weighting scheme is simply an equal scheme: $W_h = (1/h) \times I_h$. The number of draws from the limit distributions under $H_0$ is 5,000. The number of high frequency lags of $x_H$ taken by mixed frequency tests is $h_{MF} \in \{4, 8, 12\}$, while the number of low frequency lags of aggregated $x_H$ taken by low frequency tests is $h_{LF} \in \{1, 2, 3\}$. For low frequency tests, we consider both flow sampling and stock sampling. Nominal size $\alpha$ is fixed at 0.05. The number of Monte Carlo iterations is 5,000 for max tests and 1,000 for bootstrapped Wald tests. The number of bootstrap replications is $N = 499$.

Table 4 compares the rejection frequencies of MF max test, MF Wald test, LF max test, and LF Wald test. Panel A considers Non-causality, Panel B considers Decaying Causality, Panel C considers Lagged Causality, and Panel D considers Sporadic Causality. For each panel we consider low persistence of $x_H$ (i.e., $d = 0.2$) and high persistence (i.e., $d = 0.8$) as well as small sample size (i.e., $T_L = 40$) and medium sample size ($T_L = 80$).

As seen in Panel A, no tests have size distortions. The smallest rejection frequency out of all 72 slots in Panel A is 0.035, while the largest one is 0.068. All values are thus close to the nominal size 0.05. Max tests have correct size due to their parsimonious specification, while Wald tests have correct size due to Gonçalves and Killian’s (2004) parametric bootstrap. We can thus compare empirical power of each test meaningfully.

In terms of mixed frequency tests versus low frequency tests, Panels B, C, and D of Table 4 provide the same implications as Table 3. In the local power analysis, we show that mixed frequency tests are more robust against complicated causal patterns like Sporadic Causality than low frequency tests are (cfr. Table 3). The same goes for finite sample power as well (cfr. Table 4). We refrain from repeating detailed discussion to save space.
We rather focus on the relative performance of the MF max test and MF Wald test. There are $3 \times 3 \times 2 \times 2 = 36$ ways of comparing them depending on causal patterns $b$ (except for Non-causality), lag length $h_{MF}$, persistence $d$, and sample size $T_L$. In 30 out of the 36 cases the MF max test has higher power than the MF Wald test. In the 6 exceptions the difference between the MF max test and the MF Wald test is negligibly small, at most $0.482 - 0.527 = -0.045$ (cfr. $h_{MF} = 4$ in Panel B.1.2). In the 30 cases where the max test performs better, the difference in power is often substantial. For example, the max test has power 0.576 and the Wald test has power 0.255 under Lagged Causality with $(d, T_L, h_{MF}) = (0.8, 40, 12)$. Thus, we can conclude that the max test achieves higher power than the Wald test in small or medium sample.

**MF-VAR(2)** For robustness check we also consider a structural MF-VAR(2) data generating process $NX(t_L) = \sum_{i=1}^{2} M_i X(t_L - i) + \eta(t_L)$ with $m = 12$. The extra coefficient matrix $M_2$ is parameterized as

$$M_2 = \begin{bmatrix} 0_{12 \times 1} & \cdots & 0_{12 \times 1} & 0_{12 \times 1} \\ b_{24} & \cdots & b_{13} & 0 \end{bmatrix}.$$ 

We prepare four causal patterns:

1. **Non-causality:** $b = 0_{24 \times 1}$.

2. **Decaying Causality:** $b_j = (-1)^{j-1} \times 0.3/j$ for $j = 1, \ldots, 24$.

3. **Lagged Causality:** $b_j = 0.3 \times I(j = 24)$ for $j = 1, \ldots, 24$.

4. **Sporadic Causality:** $(b_5, b_{12}, b_{17}, b_{19}) = (-0.2, 0.1, 0.2, -0.35)$ and all other $b$’s are zeros.

Other quantities are mostly as before; $a = 0.2$; $d \in \{0.2, 0.8\}$; $c_j = (-1)^{j-1} \times 0.4/j$ for $j = 1, \ldots, 12$; $T_L \in \{40, 80\}$; $q = 2$; $W_h = (1/h) \times I_h$; $\alpha = 0.05$. The number of high frequency lags of $x_H$ taken by mixed frequency tests is $h_{MF} \in \{16, 20, 24\}$, while the number of low frequency lags of aggregated $x_H$ taken by low frequency tests is $h_{LF} \in \{1, 2, 3\}$. For low frequency tests, we consider both flow sampling and stock sampling.

Table 5 compares the rejection frequencies of each test. No tests have size distortions as seen in Panel A. We refrain from comparing MF tests and LF tests since their relationship is exactly analogous to the MF-VAR(1) case above. We rather focus on the difference between the MF max test and MF Wald test. There are $3 \times 3 \times 2 \times 2 = 36$ ways of comparing them again. As many as 34 out of all 36 slots indicate that the max test is more powerful. The difference in power in those 34 slots is often very large. For example, fixing $(d, T_L, h_{MF}) = (0.8, 80, 24)$, empirical power under Lagged Causality is 0.896 for the max test and 0.480 for the Wald test. Moreover, the 2 exceptional cases have negligible power difference of $0.284 - 0.290 = -0.006$ and $0.621 - 0.655 = -0.034$ (cfr. $h_{MF} = 16$ in Panel B.1.2 and $h_{MF} = 20$ in Panel D.2.2). Hence, the superiority of the MF max test to the MF Wald test is even more emphasized in MF-VAR(2) than in MF-VAR(1). This is reasonable since parameter proliferation is more an issue in MF-VAR(2).
4.2 Low-to-High Granger Causality

In this section we focus on the finite sample power of low-to-high causality tests. We assume that the true DGP follows a structural MF-V AR(1) process with \( m = 12 \) as in (3.6). Our main focus here lies on \( c = [c_1, \ldots, c_{12}]' \), which represents Granger causality from \( x_L \) to \( x_H \) in the structural form. We need to be careful about how \( c \) is transferred to the upper-right block of \( A_1 \), which represents low-to-high causality in the reduced form. As shown in (3.7), the upper-right block of \( A_1 \) is equal to \( \left[ \sum_{i=1}^{12} d^{1-i} c_i, \ldots, \sum_{i=1}^{12} d^{12-i} c_i \right]' \), where \( d \) is the AR(1) coefficient of \( x_H \).

We prepare four causal patterns:

1. **Non-causality**: \( c = 0_{12 \times 1} \). In this case we can check empirical size. The upper-right block of \( A_1 \) is a null vector regardless of \( d \).

2. **Decaying Causality**: \( c_j = (-1)^{j-1} \times 0.45/j \) for \( j = 1, \ldots, 12 \). In this case the impact of \( x_L \) on \( x_H \) decays gradually with signs alternating. The upper-right block of \( A_1 \) has a similar pattern, assuming \( d = 0.2 \) (cfr. Figure 2).

3. **Lagged Causality**: \( c_j = 0.4 \times I(j = 12) \) for \( j = 1, \ldots, 12 \). This case is an extreme example of seasonality or lagged response of \( x_H \) to \( x_L \). The upper-right block of \( A_1 \) is identically \( c \) regardless of \( d \) (cfr. Figure 2).

4. **Sporadic Causality**: \( (c_3, c_7, c_{10}) = (0.4, 0.25, -0.5) \) and all other \( c \)'s are zeros. This is a tricky (but realistic) case where we have both signs, lags are unevenly spaced, and the largest coefficient in absolute value does not show up first. The upper-right block of \( A_1 \) has a similar pattern, assuming \( d = 0.2 \) (cfr. Figure 2).

Other quantities are specified as follows. First, we assume fairly weak autoregressive property for \( x_L \) and \( x_H \) (i.e. \( a = d = 0.2 \)). Second, we assume decaying causality with alternating signs for high-to-low causality: \( b_j = (-1)^{j-1} \times 0.2/j \) for \( j = 1, \ldots, 12 \).

We now explain how to implement each of mixed frequency max test, mixed frequency Wald test, low frequency max test, and low frequency Wald test. First, the MF max test is based on MF parsimonious regression models (2.20) with a MIDAS polynomial:

\[
x_L(\tau_L) = \alpha_1 x_L(\tau_L - 1) + \sum_{k=1}^{k_{MF}} \omega_k(\pi)x_H(\tau_L - 1, 12 + 1 - k) + \gamma_j x_H(\tau_L + 1, j) + u_L(\tau_L), \quad j = 1, \ldots, \tau_{MF}.
\]

For \( \omega_k(\pi) \) we use the Almon polynomial of dimension \( s = 3 \) (cfr. Section 2.2.1).

Second, the MF Wald test is based on a MF naïve regression model (2.17):

\[
x_L(\tau_L) = \alpha_1 x_L(\tau_L - 1) + \sum_{k=1}^{k_{MF}} \beta_k x_H(\tau_L - 1, 12 + 1 - k) + \sum_{j=1}^{r_{MF}} \gamma_j x_H(\tau_L + 1, j) + u_L(\tau_L).
\]

This model often has too many parameters relative to sample size since we are including all leads and lags of \( x_H \) at the same time. Using a MIDAS polynomial for the high-to-low causality part only, on one hand, is not enough to delete size distortions since there are still many parameters in the lead part. Using
a MIDAS polynomial for the low-to-high part too, on the other hand, is not desired since the low-to-high Wald test may lose power due to misspecification. Hence we keep the model itself unrestricted and use Gonçalves and Killian’s (2004) bootstrap with \( N = 499 \) replications, as we did in the MF high-to-low Wald test.

Third, the LF max test is based on a LF parsimonious regression model (2.21):

\[
x_L(\tau_L) = \alpha_{1,j} x_L(\tau_L - 1) + \sum_{k=1}^{h_{LF}} \beta_{k,j} x_H(\tau_L - k) + \sum_{j=1}^{r_{LF}} \gamma_{j} x_H(\tau_L + j) + u_{LF,j}(\tau_L), \quad j = 1, \ldots, r_{LF}.
\]  

(4.1)

Here we do not need a MIDAS polynomial or bootstrap since \( h_{LF} \) typically takes a small value. We consider both stock sampling \( x_H(\tau_L) = x_H(\tau_L, 12) \) and flow sampling \( x_H(\tau_L) = (1/12) \sum_{j=1}^{12} x_H(\tau_L, j) \).

Fourth, the LF Wald test is based on a LF naïve regression model:

\[
x_L(\tau_L) = \alpha_{1} x_L(\tau_L - 1) + \sum_{k=1}^{h_{LF}} \beta_{k} x_H(\tau_L - k) + \sum_{j=1}^{r_{LF}} \gamma_{j} x_H(\tau_L + j) + u_{LF}^L(\tau_L).
\]  

(4.2)

We use Gonçalves and Killian’s (2004) bootstrap with \( N = 499 \) replications since \( h_{LF} + r_{LF} \) may take a large value.

Other details of the simulation design are in order. The number of Monte Carlo simulations is 5,000 for max tests and 1,000 for Wald tests. For the latter we use Gonçalves and Killian’s (2004) bootstrap with \( N = 499 \) replications as stated above. For mixed frequency models, the number of leads and lags of \( x_H \) is taken from \( h_{MF}, r_{MF} \in \{4, 8, 12\} \). For low frequency models, the number of leads and lags of aggregated \( x_H \) is taken from \( h_{LF}, r_{LF} \in \{1, 2, 3\} \). We compute max statistics based on the equal weighting scheme and 1,000 draws from the asymptotic distribution under low-to-high non-causality. Since we fix \( m = 12 \), our simulation study can be thought of as a week vs. quarter case, approximately. Sample size is either \( T_L = 40 \) quarters (small sample) or \( T_L = 80 \) quarters (medium sample). The nominal size is fixed at 5%.

Table 6 shows rejection frequencies for low-to-high Granger causality. Panel A assumes Non-causality, Panel B assumes Decaying Causality, Panel C assumes Lagged Causality, and Panel D assumes Sporadic Causality. For each panel we have small sample size (40 quarters) and medium sample size (80 quarters). For each sample size we have mixed frequency tests, low frequency tests with stock sampling, and low frequency tests with flow sampling. Both max tests and Wald tests are implemented for each of these three cases.

First of all, we confirm from Panel A that no tests have size distortions. Empirical size varies between 0.032 and 0.073, a fairly accurate interval containing the nominal size 5% (cfr. Panels A.1.1 and A.2.2). The mixed frequency max test has correct size due to the Almon polynomial. The low frequency max test has correct size due to its parsimonious specification. Wald test has correct size due to Gonçalves and Killian’s (2004) bootstrap.

\(^6\)If we assume alternatively that our simulation study is on month vs. year, then \( T_L = 40 \) years is a very large sample but not totally unrealistic.
We now focus on Panel B (Decaying Causality) to show the superiority of mixed frequency tests to low frequency tests. See Panel B.2 (medium sample) for example. Mixed frequency tests have empirical power at least 0.739, while low frequency tests have empirical power at most 0.130 (whether stock sampling or flow sampling is considered). We first explain why the low frequency tests with stock sampling have low power. As seen in (4.1) and (4.2), lead terms used in those tests are $x_H(\tau_L + 1, 12), x_H(\tau_L + 2, 12), \ldots, x_H(\tau_L + r_{LF}, 12)$ and all these terms have small coefficients due to the decaying structure of true causality. These tests, in other words, are missing the most important lead term $x_H(\tau_L + 1, 1)$ and thus suffering from low power. Second, low frequency tests with flow sampling have low power since adding $x_H(\tau_L + 1, 1)$ through $x_H(\tau_L + 1, 12)$ eliminates the true, decaying causal pattern. These are typical examples where temporal aggregation can hide underlying causality.

Next we discuss Panel C (Lagged Causality). Mixed frequency tests have no power unless $r_{MF} = 12$, which is understandable since the only relevant term is $x_H(\tau_L + 1, 12)$ by construction. When $r_{MF} = 12$, they get some power but not so high due to too many parameters. In medium sample, for example, the mixed frequency max test has empirical power within $[0.582, 0.592]$ while the mixed frequency Wald test has empirical power within $[0.415, 0.456]$ (cfr. Panel C.2.1). Low frequency tests with stock sampling, in contrast, get very high power within $[0.766, 0.919]$ (cfr. Panel C.2.2). This is because their models contain the relevant lead term $x_H(\tau_L + 1, 12)$ exactly. Hence, if the a causal pattern happens to be Lagged Causality, low frequency tests with stock sampling perform best due to its parsimonious and coincidentally correct specification.

Now we discuss Panel D (Sporadic Causality), which is a more realistic example than Decaying Causality or Lagged Causality. Mixed frequency tests are showing very high power, especially when $r_{MF} = 12$. Having twelve lead terms improves power since it takes into account $c_{10} = -0.5$, the largest coefficient in absolute value. In medium sample with $r_{MF} = 12$, mixed frequency tests have empirical power more than 0.900 (cfr. Panel D.2.1). Low frequency tests, in contrast, do not have any power no matter what $h_{LF}, r_{LF}$ and aggregation schemes are. The low frequency leads and lags of $x_H$ are too coarse to capture the complicated causality with unevenly-spaced leads, alternating signs, and non-decaying structure.

Finally, we compare the finite sample power of the mixed frequency max test and mixed frequency Wald test. There are three causal patterns (not counting Non-causality), two values for sample size, three values for $h_{MF}$, and three values for $r_{MF}$. This means that there are $3 \times 2 \times 3 \times 3 = 54$ ways to compare the MF max test and MF Wald test. Of those 54 slots, the max test has higher power than the Wald test in 41 slots (7 slots in Decaying Causality, all 18 slots in Lagged Causality, and 16 slots in Sporadic Causality). The maximum difference between the power of max test and the power of Wald test is $0.592 - 0.415 = 0.177$, which occurs in Panel C.2.1 (Lagged Causality, medium sample, $(h_{MF}, r_{MF}) = (12, 12)$). The smallest difference between the power of max test and the power of Wald test is $0.739 - 0.834 = -0.095$, which occurs in Panel B.2.1 (Decaying Causality, medium sample, $(h_{MF}, r_{MF}) = (4, 12)$). These results suggest that the MF max test and the MF Wald test are roughly as powerful as each other under Decaying Causality but the former is much more powerful than the latter under Lagged Causality and Sporadic Causality.

Summarizing Table 6, mixed frequency tests are more robust against a complex causal pattern like
Sporadic Causality than low frequency tests are. In particular, the mixed frequency max test is even more robust than the mixed frequency Wald test.

5 Empirical Application

As an empirical illustration, this section studies Granger causality between weekly interest rate spread and quarterly real GDP growth in the U.S. While we analyze both high-to-low causality (i.e. causality from spread to GDP) and low-to-high causality (i.e. causality from GDP to spread), we are particularly interested in the former. Declined interest rate spread used to be regarded as a strong predictor of recession, but more recent evidence questions its predictability. One well-known episode is "Greenspan’s Conundrum" around 2005, when interest rate spread declined substantially due to constant long-term rates and increased short-term rates. While this sharp decline of interest rate spread itself was referred as a conundrum by Alan Greenspan, Chairman of the Federal Reserve of the U.S. from 1987 to 2006, another interesting phenomenon is that the U.S. macroeconomy did not run into recession at that position. Although they did get a serious recession due to the subprime mortgage crisis starting December 2007, the time lag between the declined spread and the recession seems much larger than it used to be. Based on this motivation, we investigate how Granger causality from interest rate spread to GDP growth evolved over time.

As a business cycle measure, seasonally-adjusted quarterly real GDP growth is used. The data can be found at Federal Reserve Economic Data (FRED) maintained by the Federal Reserve Bank of St. Louis. To remove potential seasonal effects remaining after seasonal adjustment, we use percentage growth rates from previous year.

For short-term and long-term interest rates, we first download daily series of federal funds (FF) rate and 10-year Treasury constant maturity rate at FRED. While we could directly work on the daily interest rates and the quarterly GDP, the ratio of sampling frequencies \( m \) would seem too large to ensure reasonable size and power. We thus aggregate the daily series into weekly by picking the last observation in each week (recall that interest rates are stock variables). Finally, we calculate interest rate spread as the difference between the weekly 10-year rate and the weekly FF rate.

Figure 3 shows the weekly 10-year rate, weekly FF rate, their spread (10Y - FF), and the quarterly GDP growth from January 5, 1962 through December 31, 2013. The blue, solid line represents the 10-year rate. The red, dashed line represents the FF rate. The gray, solid line represents the spread. The yellow, solid line represents GDP. The full sample period covers 2,736 weeks or 208 quarters. The shaded areas represent recession periods defined by the National Bureau of Economic Research (NBER). In the first half of the entire sample period, a sharp decline of the spread seems to be immediately followed by recession. In the second half, we find a weaker evidence or at least there seems a larger time lag between declined spread and recession.

Table 7 shows sample statistics of the weekly 10-year rate, weekly FF rate, their spread, and the quarterly real GDP growth. The 10-year rate is about 1% point higher than the FF rate on average. The average GDP growth is 3.151%, indicating a fairly steady growth of the U.S. economy during the sample period. The spread has a relatively large kurtosis of 5.611, while the GDP growth has a smaller kurtosis.
When we apply mixed frequency tests, a slightly inconvenient aspect of our data is that the number of weeks contained in each quarter is not constant. Specifically, (i) 13 quarters have 12 weeks each, (ii) 150 quarters have 13 weeks each, and (iii) 45 quarters have 14 weeks each. Since our asymptotic theory requires \( m \) to be constant, we assume \( m = 13 \) by making the following modification. We (i) duplicate the twelfth observation once when a quarter contains 12 weeks, (ii) do nothing when it contains 13 weeks, and (iii) cut the last observation when it contains 14 weeks. This gives us a manageable dataset with \( T_L = 208 \), \( m = 13 \), and thus \( T = mT_L = 2,704 \).

Since our entire sample size is as large as 52 years, we implement rolling window analysis with the window width 80 quarters (i.e. 20 years). The first subsample is 1962:I-1981:IV, the second one is 1962:II-1982:I, and so on. There are 129 subsamples in total, the last one being 1994:I-2013:IV. The trade-off between a small window width and a large one is that the large window is more likely to contain a structural break but it allows us to include more leads and lags in our models.

**Granger Causality from Spread to GDP** We first consider high-to-low causality (i.e. causality from spread to GDP). Recall that the window width is 80 quarters. In view of our MF-VAR(2) simulation results with \( T_L = 80 \), mixed frequency parsimonious regression models are specified as follows.

\[
x_L(\tau_L) = \alpha_{0,j} + \sum_{k=1}^{2} \alpha_{k,j} x_L(\tau_L - k) + \beta_j x_H(\tau_L - 1, 13 + 1 - j) + u_{L,j}(\tau_L). \quad j = 1, \ldots, 26. \tag{5.1}
\]

\( x_L \) signifies GDP growth rate, while \( x_H \) signifies interest rate spread. We are including two quarters of lagged \( x_L \) (i.e. \( q = 2 \)) and 26 weeks of lagged \( x_H \) (i.e. \( h_{MF} = 26 \)). Based on this model we compute the MF max test statistic.

Similarly, the mixed frequency naive regression model is specified as follows.

\[
x_L(\tau_L) = \alpha_0 + \sum_{k=1}^{2} \alpha_k x_L(\tau_L - k) + \sum_{j=1}^{26} \beta_j x_H(\tau_L - 1, 13 + 1 - j) + u_L(\tau_L). \tag{5.2}
\]

Based on this model we compute the MF Wald statistic. Note that \( \beta_j \) in (5.1) and \( \beta_j \) in (5.2) are not restricted to be the same. These are estimated via OLS separately.

The low frequency parsimonious regression models are specified as follows.

\[
x_L(\tau_L) = \alpha_{0,j} + \sum_{k=1}^{2} \alpha_{k,j} x_L(\tau_L - k) + \beta_j x_H(\tau_L - j) + u_{L,j}(\tau_L), \quad j = 1, 2, 3. \tag{5.3}
\]

Since interest rate spread is a stock variable, we let \( x_H(\tau_L) = x_H(\tau_L, 13) \). We are including two quarters of lagged \( x_L \) (i.e. \( q = 2 \)) and three quarters of lagged \( x_H \) (i.e. \( h_{LF} = 3 \)). Based on this model we compute the LF max test statistic.
Finally, the low frequency naïve regression model is specified as follows.

\[ x_L(\tau_L) = \alpha_0 + \sum_{k=1}^{2} \alpha_k x_L(\tau_L - k) + \sum_{j=1}^{3} \beta_j x_H(\tau_L - j) + u_L(\tau_L). \]  

(5.4)

Based on this model we compute the LF Wald statistic.

To control size, the parametric bootstrap of Gonçalves and Killian (2004) with replications \( N = 999 \) is used for Wald tests. For max tests, the number of draws from limit distributions under non-causality is 100,000.

Figure 4 plots \( p \)-values with respect to the null hypothesis of high-to-low non-causality. The horizontal axis takes 129 different subsamples. See Panel (a) for the MF max test, Panel (b) for the MF Wald test, Panel (c) for the LF max test, and Panel (d) for the LF Wald test. If a \( p \)-value plotted is inside the shaded area, then it means that we have a \( p \)-value smaller than nominal size 5% and thus there exists significant causality from interest rate spread to GDP in that subsample.

As seen in Panels (a), (c), and (d), all tests except for the MF Wald test find significant causality in early periods. At the 5% level, the MF max test always detects significant causality until subsample 1981:III-2001:II, the LF max test always detects significant causality until subsample 1979:III-1999:II, and the LF Wald test always detects significant causality until subsample 1973:III-1993:II. The MF max test has the longest period of significant causality likely due to its high power, as shown in Section 4.1. These three tests all agree that there is non-causality in recent periods, possibly reflecting some structural change in the middle of the entire sample.

The MF Wald test, in contrast, suggests that there is significant causality only after subsample 1989:IV-2009:III, which is somewhat counter-intuitive. This result may stem from parameter proliferation. As seen from (5.1)-(5.4), the MF naïve regression model has much more parameters than any other model. In this sense the MF max test seems to be preferred to the MF Wald test when the ratio of sampling frequencies \( m \) is large.

As a supplemental analysis, we implement the four tests for the full sample covering 52 years from January 1962 through December 2013. We try two specifications for completeness. Specification 1 is the same as (5.1)-(5.4) and assumes \((q, h_{MF}, h_{LF}) = (2, 26, 3)\). Specification 2 assumes larger models taking advantage of the very long sample period: \((q, h_{MF}, h_{LF}) = (4, 52, 6)\). This specification means that (i) each model has 4 quarters of low frequency lags of \( x_L \), (ii) each mixed frequency model has 52 weeks of high frequency lags of \( x_H \), and (iii) each low frequency model has 6 quarters of low frequency lags of \( x_H \). For both Specifications 1 and 2, the number of bootstrap replications for Wald tests is 9,999.

See Table 8 for \( p \)-values with respect to the null hypothesis of non-causality. Two asterisks are put for a \( p \)-value less than 1%, indicating a strong rejection of non-causality. One asterisk is put for a \( p \)-value less than 5% but greater than 1%, indicating a rejection of non-causality. The MF max test and the LF max test yield strong rejections under both Specifications 1 and 2. The MF Wald test does not reject the null hypothesis in either Specification 1 or Specification 2, possibly suggesting low power due to parameter proliferation. The LF Wald test rejects non-causality at 5% (but not 1%) under Specification 1, but it does not reject the null hypothesis under Specification 2. Overall, significant causality from
We now consider low-to-high causality (i.e. causality from GDP to spread). The mixed frequency parsimonious regression models are specified as \( x_L(\tau_L) = \alpha_{0,j} + \sum_{k=1}^{2} \alpha_k x_L(\tau_L - k) + \sum_{j=1}^{26} \omega_j x_H(\tau_L - 1, 13 + 1 - k) + \gamma_j x_H(\tau_L + 1, j) + u_L(\tau_L), \) \( j = 1, \ldots, 26. \) We are including two quarters of lagged \( x_L \) (i.e. \( q = 2 \)), 26 weeks of lagged \( x_H \) (i.e. \( h_{MF} = 26 \)), and 26 weeks of led \( x_H \) (i.e. \( r_{MF} = 26 \)). We use the Almon polynomial of order 3 for \( \omega_k(\pi) \). Based on this model we compute the MF max test statistic.

Similarly, the mixed frequency naïve regression model is specified as \( x_L(\tau_L) = \alpha_0 + \sum_{k=1}^{2} \alpha_k x_L(\tau_L - k) + \sum_{j=1}^{26} \beta_j x_H(\tau_L - k) + \gamma_j x_H(\tau_L + j) + u_L(\tau_L). \) Based on this model we compute the MF Wald statistic. Note that \( \gamma_j \) in the MF parsimonious regression models and \( \gamma_j \) in the MF naïve regression model are not restricted to be the same. These are estimated via OLS separately.

Low frequency parsimonious regression models are specified as \( x_L(\tau_L) = \alpha_{0,j} + \sum_{k=1}^{2} \alpha_k x_L(\tau_L - k) + \sum_{j=1}^{26} \beta_j x_H(\tau_L - k) + \gamma_j x_H(\tau_L + j) + u_L(\tau_L), \) \( j = 1, 2, 3. \) Since interest rate spread is a stock variable, we let \( x_H(\tau_L) = x_L(\tau_L, 13) \). We are including two quarters of lagged \( x_L \) (i.e. \( q = 2 \)), three quarters of lagged \( x_H \) (i.e. \( h_{LF} = 3 \)), and three quarters of led \( x_H \) (i.e. \( r_{LF} = 3 \)). Based on this model we compute the LF max test statistic.

Finally, the low frequency naïve regression model is specified as \( x_L(\tau_L) = \alpha_0 + \sum_{k=1}^{2} \alpha_k x_L(\tau_L - k) + \sum_{j=1}^{3} \beta_j x_H(\tau_L - k) + \sum_{j=1}^{3} \gamma_j x_H(\tau_L + j) + u_L(\tau_L). \) Based on this model we compute the LF Wald statistic.

To control size, the parametric bootstrap of Gonçalves and Killian (2004) with replications \( N = 999 \) is used for Wald tests. For max tests the number of draws from limit distributions under non-causality is 100,000.

Figure 5 plots \( p \)-values with respect to the null hypothesis of low-to-high non-causality. The horizontal axis takes 129 different subsamples. See Panel (a) for the MF max test, Panel (b) for the MF Wald test, Panel (c) for the LF max test, and Panel (d) for the LF Wald test. If a \( p \)-value plotted is inside the shaded area, then it means that we have a \( p \)-value smaller than nominal size 5% and thus there exists significant causality from GDP to spread in that subsample.

In most subsamples we find non-causality from GDP to interest rate spread. Based on the MF max test, \( p \)-values never go below 10%. Also, \( p \)-values based on the MF Wald test and the LF max test go below 10% very rarely. The LF Wald test shows significant causality in subsample 1983:II-2003:I and after (about one-third of the total number of subsamples), but we find non-causality before that. Non-causality from GDP to interest rate spread we found is not a surprising result since we do not have any strong theoretical or empirical conjecture supporting causality.

As a supplemental analysis, we conduct the four tests on the full sample covering 52 years from January 1962 through December 2013. We assume \((q, h_{MF}, r_{MF}, h_{LF}, r_{LF}) = (2, 26, 26, 3, 3)\). This specification means that (i) each model has 2 quarters of low frequency lags of \( x_L \), (ii) each mixed frequency model has 26 weeks of high frequency leads and lags of \( x_H \) each, and (iii) each low frequency model has 3 quarters of low frequency leads and lags of \( x_H \) each. The number of bootstrap replications for each naïve regression model is 9,999.

Granger Causality from GDP to Spread We now consider low-to-high causality (i.e. causality from GDP to spread). The mixed frequency parsimonious regression models are specified as \( x_L(\tau_L) = \alpha_{0,j} + \sum_{k=1}^{2} \alpha_k x_L(\tau_L - k) + \sum_{j=1}^{26} \omega_j x_H(\tau_L - 1, 13 + 1 - k) + \gamma_j x_H(\tau_L + 1, j) + u_L(\tau_L), \) \( j = 1, \ldots, 26. \) We are including two quarters of lagged \( x_L \) (i.e. \( q = 2 \)), 26 weeks of lagged \( x_H \) (i.e. \( h_{MF} = 26 \)), and 26 weeks of led \( x_H \) (i.e. \( r_{MF} = 26 \)). We use the Almon polynomial of order 3 for \( \omega_k(\pi) \). Based on this model we compute the MF max test statistic.

Similarly, the mixed frequency naïve regression model is specified as \( x_L(\tau_L) = \alpha_0 + \sum_{k=1}^{2} \alpha_k x_L(\tau_L - k) + \sum_{j=1}^{26} \beta_j x_H(\tau_L - k) + \gamma_j x_H(\tau_L + j) + u_L(\tau_L). \) Based on this model we compute the MF Wald statistic. Note that \( \gamma_j \) in the MF parsimonious regression models and \( \gamma_j \) in the MF naïve regression model are not restricted to be the same. These are estimated via OLS separately.

Low frequency parsimonious regression models are specified as \( x_L(\tau_L) = \alpha_{0,j} + \sum_{k=1}^{2} \alpha_k x_L(\tau_L - k) + \sum_{j=1}^{26} \beta_j x_H(\tau_L - k) + \gamma_j x_H(\tau_L + j) + u_L(\tau_L), \) \( j = 1, 2, 3. \) Since interest rate spread is a stock variable, we let \( x_H(\tau_L) = x_L(\tau_L, 13) \). We are including two quarters of lagged \( x_L \) (i.e. \( q = 2 \)), three quarters of lagged \( x_H \) (i.e. \( h_{LF} = 3 \)), and three quarters of led \( x_H \) (i.e. \( r_{LF} = 3 \)). Based on this model we compute the LF max test statistic.

Finally, the low frequency naïve regression model is specified as \( x_L(\tau_L) = \alpha_0 + \sum_{k=1}^{2} \alpha_k x_L(\tau_L - k) + \sum_{j=1}^{3} \beta_j x_H(\tau_L - k) + \sum_{j=1}^{3} \gamma_j x_H(\tau_L + j) + u_L(\tau_L). \) Based on this model we compute the LF Wald statistic.

To control size, the parametric bootstrap of Gonçalves and Killian (2004) with replications \( N = 999 \) is used for Wald tests. For max tests the number of draws from limit distributions under non-causality is 100,000.

Figure 5 plots \( p \)-values with respect to the null hypothesis of low-to-high non-causality. The horizontal axis takes 129 different subsamples. See Panel (a) for the MF max test, Panel (b) for the MF Wald test, Panel (c) for the LF max test, and Panel (d) for the LF Wald test. If a \( p \)-value plotted is inside the shaded area, then it means that we have a \( p \)-value smaller than nominal size 5% and thus there exists significant causality from GDP to spread in that subsample.

In most subsamples we find non-causality from GDP to interest rate spread. Based on the MF max test, \( p \)-values never go below 10%. Also, \( p \)-values based on the MF Wald test and the LF max test go below 10% very rarely. The LF Wald test shows significant causality in subsample 1983:II-2003:I and after (about one-third of the total number of subsamples), but we find non-causality before that. Non-causality from GDP to interest rate spread we found is not a surprising result since we do not have any strong theoretical or empirical conjecture supporting causality.

As a supplemental analysis, we conduct the four tests on the full sample covering 52 years from January 1962 through December 2013. We assume \((q, h_{MF}, r_{MF}, h_{LF}, r_{LF}) = (2, 26, 26, 3, 3)\). This specification means that (i) each model has 2 quarters of low frequency lags of \( x_L \), (ii) each mixed frequency model has 26 weeks of high frequency leads and lags of \( x_H \) each, and (iii) each low frequency model has 3 quarters of low frequency leads and lags of \( x_H \) each. The number of bootstrap replications for each naïve regression model is 9,999.
The MF max test, MF Wald test, LF max test, and LF Wald test have p-values 0.090, 0.089, 0.145, and 0.299, respectively. MF tests find significant causality from GDP to spread at 10% but not at 5%. LF tests do not find significant causality even at 10%. These results again suggest that there is no strong evidence for causality from GDP to spread.

6 Conclusions

This paper proposes a new mixed frequency Granger causality test that achieves high power even when the ratio of sampling frequencies $m$ is large. We postulate multiple parsimonious regression models where the $j$-th model regresses a low frequency variable $x_L$ onto the $j$-th lag or lead of a high frequency variable $x_H$ for $j \in \{1, \ldots, h\}$. Let $\hat{\beta}_j$ be a least squares estimator for the parameter of the $j$-th lag or lead of $x_H$, then our test statistic is the maximum among $\{\hat{\beta}_1^2, \ldots, \hat{\beta}_h^2\}$ scaled and weighted properly. In this sense we call it the max test for short.

While the max test statistic follows a non-standard asymptotic distribution under the null hypothesis of Granger non-causality, a simulated $p$-value is readily available through an arbitrary number of draws from the null distribution. The max test is thus very easy to implement in practice.

Through local asymptotic power analysis and Monte Carlo simulations, we compare the max test based on mixed frequency data (MF max test), a Wald test based on mixed frequency data (MF Wald test), the max test based on low frequency data (LF max test), and a Wald test based on low frequency data (LF Wald test). We have shown that MF tests are more robust against complex (but realistic) causal patterns than LF tests in both local asymptotics and finite sample. The MF max test and the MF Wald test are roughly as powerful as each other in local asymptotics, but the former is clearly more powerful than the latter in finite sample.

For Granger causality from $x_H$ to $x_L$, we have proved the consistency of MF max test. We have also proved by counter-examples that LF tests are not consistent. For Granger causality from $x_L$ to $x_H$, proving the consistency of MF max test remains as an open question.

As an empirical application, we conduct a rolling window analysis on weekly interest rate spread and real GDP growth in the U.S. The MF max test yields an intuitive result that the interest rate spread used to cause GDP until about the year 2000 but such causality has vanished more recently.

References


199–211.


Springer-Verlag, Heidelberg.


Bayesian Block Model,” Discussion Paper, Federal Reserve Bank of St. Louis and Bank of Canada.

MCCRORIE, J., AND M. CHAMBERS (2006): “Granger Causality and the Sampling of Economic Pro-


Tables and Figures

Table 1: Four Different Tests for High-to-Low Granger Causality

This table lists four different high-to-low Granger causality tests: mixed frequency max test, mixed frequency Wald test, low frequency max test, and low frequency Wald test. They can be distinguished by the sampling frequency of $x_H$ and model specification. "Mixed Frequency" works on high frequency observations of $x_H$, while "Low Frequency" works on an aggregated $x_H$. "Parsimonious" specification, on which max statistics are constructed, prepares $h$ separate models with the $i$-th model containing only the $i$-th lag of $x_H$ (either high frequency or low frequency lag). "Naïve" specification, on which Wald statistics are constructed, prepares only one model which contains all $h$ lags of $x_H$ (either high frequency or low frequency). MF tests are consistent (i.e. power approaches 1 under any form of Granger causality) if the selected number of high frequency lags $h$ is larger than or equal to the true lag order $pm$, where $p$ is the lag order of MF-VAR data generating process and $m$ is the ratio of sampling frequencies. In contrast, LF tests are inconsistent (i.e. there exists some form of Granger causality where power does not approach 1) no matter how many lags are included in the model and no matter which linear aggregation scheme is used. The MF naïve specification often entails many parameters since the lags of $x_H$ are taken in terms of high frequency. In contrast, the LF naïve specification often entails few parameters since the lags of $x_H$ are taken in terms of low frequency. The number of parameters in each parsimonious regression model is often small since only one lag of $x_H$ is included.

<table>
<thead>
<tr>
<th>Model \ Frequency</th>
<th>Mixed Frequency</th>
<th>Low Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Parsimonious</strong></td>
<td>MF Max Test</td>
<td>LF Max Test</td>
</tr>
<tr>
<td>1. Consistent if $h \geq pm$</td>
<td>1. Inconsistent regardless of $h \in \mathbb{N}$</td>
<td></td>
</tr>
<tr>
<td>2. Few parameters</td>
<td>2. Few parameters</td>
<td></td>
</tr>
<tr>
<td><strong>Naïve</strong></td>
<td>MF Wald Test</td>
<td>LF Wald Test</td>
</tr>
<tr>
<td>1. Consistent if $h \geq pm$</td>
<td>1. Inconsistent regardless of $h \in \mathbb{N}$</td>
<td></td>
</tr>
<tr>
<td>2. Many parameters</td>
<td>2. Few parameters</td>
<td></td>
</tr>
</tbody>
</table>
Table 2: $f(j)$ and $g(j)$ – Useful Indices in Local Asymptotic Power Analysis

For an arbitrary natural number $j \in \mathbb{N}$ and the ratio of sampling frequencies $m \in \mathbb{N}$, indices $f(j) = \lfloor (j - m)/m \rfloor$ and $g(j) = m f(j) + m + 1 - j$ play an important role in local asymptotic power analysis, where $\lfloor x \rfloor$ is the smallest integer not smaller than $x$. This table illustrates how $f(j)$ and $g(j)$ are constructed. $f(j)$ is a step function taking 0 for $j = 1, \ldots, m$, 1 for $j = m + 1, \ldots, 2m$, 2 for $j = 2m + 1, \ldots, 3m$, etc. $g(j)$ takes $m, m - 1, \ldots, 1$ as $j$ runs from $(k - 1)m + 1$ to $km$ for any $k \in \mathbb{N}$.

<table>
<thead>
<tr>
<th>$j$</th>
<th>1</th>
<th>\ldots</th>
<th>$m$</th>
<th>$m + 1$</th>
<th>\ldots</th>
<th>$2m$</th>
<th>$2m + 1$</th>
<th>\ldots</th>
<th>$3m$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f(j)$</td>
<td>0</td>
<td>\ldots</td>
<td>0</td>
<td>1</td>
<td>\ldots</td>
<td>1</td>
<td>2</td>
<td>\ldots</td>
<td>2</td>
</tr>
<tr>
<td>$g(j)$</td>
<td>$m$</td>
<td>\ldots</td>
<td>1</td>
<td>$m$</td>
<td>\ldots</td>
<td>1</td>
<td>$m$</td>
<td>\ldots</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 3: Local Asymptotic Power of High-to-Low Causality Tests

This table compares the local asymptotic power of four different high-to-low causality tests: mixed frequency max test, mixed frequency Wald test, low frequency max test, and low frequency Wald test. The DGP is MF-VAR(1) with \( m = 12 \). We have three panels depending on the Pitman drift \( \nu = [\nu_1, \ldots, \nu_{12}]' \), where \( \nu_j / \sqrt{T_L} \) signifies the impact of \( x_H(\tau_L - 1, m + 1 - j) \) on \( x_L(\tau_L) \). Panel A deals with Decaying Causality: \( \nu_j = (-1)^{j-1} \times 2.5/j \) for \( j = 1, \ldots, 12 \). Panel B deals with Lagged Causality: \( \nu_j = 2 \times I(j = 12) \) for \( j = 1, \ldots, 12 \). Only \( \nu_{12} \) is 2 and all others are zeros. Panel C deals with Sporadic Causality: \( (\nu_3, \nu_9, \nu_{11}) = (2.1, -2.8, 1.9) \), and all other \( \nu \)'s are zeros. For each panel we consider low persistence of \( x_H \) (i.e. \( d = 0.2 \)) and high persistence (i.e. \( d = 0.8 \)). Other parameters are specified as follows.

First, we assume fairly weak autoregressive property for \( x_L \) (i.e. \( a = 0.2 \)). Second, we assume decaying causality with alternating signs for low-to-high causality: \( c_j = (-1)^{j-1} \times 0.8/j \) for \( j = 1, \ldots, 12 \). Nominal size is \( \alpha = 0.05 \). For all four models, we include two low frequency lags of \( x_L \) (i.e. \( q = 2 \)). For the max tests, the number of draws from the limit distributions under \( H_0 \) and \( H_1 \) is 100,000 each. The weighting scheme is simply an equal scheme: \( W_h = (1/h) \times I_h \). The number of high frequency lags of \( x_H \) taken by the mixed frequency approaches is \( h_{MF} \in \{4, 8, 12\} \), while the number of low frequency lags of \( x_H \) taken by the low frequency approaches is \( h_{LF} \in \{1, 2, 3\} \).

For low frequency tests we consider both flow sampling and stock sampling.

### Panel A. Decaying Causality: \( \nu_j = (-1)^{j-1} \times 2.5/j \)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( h_{MF} = 4 )</td>
<td>0.487</td>
<td>0.570</td>
<td>0.074</td>
<td>0.076</td>
<td>0.638</td>
<td>0.643</td>
</tr>
<tr>
<td>( h_{MF} = 8 )</td>
<td>0.396</td>
<td>0.472</td>
<td>0.068</td>
<td>0.067</td>
<td>0.554</td>
<td>0.538</td>
</tr>
<tr>
<td>( h_{MF} = 12 )</td>
<td>0.346</td>
<td>0.407</td>
<td>0.063</td>
<td>0.063</td>
<td>0.495</td>
<td>0.473</td>
</tr>
</tbody>
</table>

### Panel A.2. Decaying Causality: \( \nu_j = (-1)^{j-1} \times 2.5/j \)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( h_{MF} = 4 )</td>
<td>0.773</td>
<td>0.736</td>
<td>0.301</td>
<td>0.302</td>
<td>0.863</td>
<td>0.862</td>
</tr>
<tr>
<td>( h_{MF} = 8 )</td>
<td>0.699</td>
<td>0.621</td>
<td>0.237</td>
<td>0.244</td>
<td>0.801</td>
<td>0.786</td>
</tr>
<tr>
<td>( h_{MF} = 12 )</td>
<td>0.657</td>
<td>0.542</td>
<td>0.206</td>
<td>0.208</td>
<td>0.754</td>
<td>0.728</td>
</tr>
</tbody>
</table>

40
Table 3: Local Asymptotic Power of High-to-Low Causality Tests (Continued)

Panel B. Lagged Causality: $\nu_j = 2 \times I(j = 12)$

<table>
<thead>
<tr>
<th></th>
<th>$d = 0.2$ (low persistence in $x_H$)</th>
<th>$d = 0.8$ (high persistence in $x_H$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MF</td>
<td>LF (flow)</td>
</tr>
<tr>
<td>$h_{MF} = 4$</td>
<td>0.051 0.050</td>
<td>0.092 0.094</td>
</tr>
<tr>
<td>$h_{MF} = 8$</td>
<td>0.052 0.050</td>
<td>0.080 0.080</td>
</tr>
<tr>
<td>$h_{MF} = 12$</td>
<td>0.247 0.206</td>
<td>0.073 0.074</td>
</tr>
<tr>
<td></td>
<td>MF</td>
<td>LF (flow)</td>
</tr>
<tr>
<td>$h_{MF} = 4$</td>
<td>0.075 0.066</td>
<td>0.455 0.454</td>
</tr>
<tr>
<td>$h_{MF} = 8$</td>
<td>0.181 0.125</td>
<td>0.468 0.459</td>
</tr>
<tr>
<td>$h_{MF} = 12$</td>
<td>0.769 0.560</td>
<td>0.415 0.402</td>
</tr>
</tbody>
</table>

Panel C. Sporadic Causality: $(\nu_8, \nu_9, \nu_{11}) = (2.1, -2.8, 1.9)$

<table>
<thead>
<tr>
<th></th>
<th>$d = 0.2$ (low persistence in $x_H$)</th>
<th>$d = 0.8$ (high persistence in $x_H$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MF</td>
<td>LF (flow)</td>
</tr>
<tr>
<td>$h_{MF} = 4$</td>
<td>0.391 0.365</td>
<td>0.072 0.070</td>
</tr>
<tr>
<td>$h_{MF} = 8$</td>
<td>0.323 0.291</td>
<td>0.061 0.063</td>
</tr>
<tr>
<td>$h_{MF} = 12$</td>
<td>0.677 0.761</td>
<td>0.060 0.060</td>
</tr>
<tr>
<td></td>
<td>MF</td>
<td>LF (flow)</td>
</tr>
<tr>
<td>$h_{MF} = 4$</td>
<td>0.716 0.614</td>
<td>0.171 0.169</td>
</tr>
<tr>
<td>$h_{MF} = 8$</td>
<td>0.667 0.679</td>
<td>0.132 0.132</td>
</tr>
<tr>
<td>$h_{MF} = 12$</td>
<td>0.690 0.872</td>
<td>0.118 0.115</td>
</tr>
</tbody>
</table>
The table compares the rejection frequencies of four different high-to-low causality tests: mixed frequency max test, mixed frequency Wald test, low frequency max test, and low frequency Wald test. The DGP is MF-VAR(1) with \( m = 12 \). We have four panels depending on the key parameter \( b = [b_1, \ldots, b_{12}]' \), where \( b_j \) signifies the impact of \( x_H(\tau_L - 1, m + 1 - j) \) on \( x_L(\tau_L) \). Panel A deals with Non-causality: \( b = \mathbf{0}_{12 \times 1} \). Panel B deals with Decaying Causality: \( b_j = (-1)^{j-1} \times 0.3/j \) for \( j = 1, \ldots, 12 \). Panel C deals with Lagged Causality: \( b_j = 0.3 \times I(j = 12) \) for \( j = 1, \ldots, 12 \). Panel D deals with Sporadic Causality: \( (b_3, \nu_7, \nu_{10}) = (0.2, 0.05, -0.3) \), and all other \( b \)'s are zeros. For each panel we consider low persistence of \( x_H \) (i.e. \( d = 0.2 \)) and high persistence (i.e. \( d = 0.8 \)). We also try small sample size \( T_L = 40 \) and medium sample size \( T_L = 80 \) for each case. Other parameters are specified as follows. First, we assume fairly weak autoregressive property for \( x_L \) (i.e. \( \alpha = 0.2 \)). Second, we assume decaying causality with alternating signs for low-to-high causality: \( c_j = (-1)^{j-1} \times 0.4/j \) for \( j = 1, \ldots, 12 \). Nominal size is \( \alpha = 0.05 \). For all four models, we include two low frequency lags of \( x_L \) (i.e. \( q = 2 \)). For max tests, the number of draws from the limit distributions under \( H_0 \) is 5,000. The weighting scheme is simply an equal scheme: \( W_h = (1/h) \times I_h \). The number of high frequency lags of \( x_H \) taken by the mixed frequency tests is \( h_{MF} \in \{4, 8, 12\} \), while the number of low frequency lags of \( x_H \) taken by the low frequency tests is \( h_{LF} \in \{1, 2, 3\} \). For low frequency tests, we consider both flow sampling and stock sampling. For Wald tests we use the parametric bootstrap based on Gonçalves and Kilian (2004) in order to control empirical size. The number of bootstrap replications is \( N = 499 \). The number of Monte Carlo iterations is 5,000 for max tests and 1,000 for Wald tests.

<table>
<thead>
<tr>
<th>Panel A. Non-causality: ( b = \mathbf{0}_{12 \times 1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.1. ( d = 0.2 ) (low persistence in ( x_H ))</td>
</tr>
<tr>
<td>A.1.1. ( T_L = 40 ) (small sample size)</td>
</tr>
<tr>
<td>( h_{MF} = 4 )</td>
</tr>
<tr>
<td>( h_{LF} = 1 )</td>
</tr>
<tr>
<td>0.061</td>
</tr>
<tr>
<td>( h_{MF} = 8 )</td>
</tr>
<tr>
<td>( h_{MF} = 12 )</td>
</tr>
<tr>
<td>A.1.2. ( T_L = 80 ) (medium sample size)</td>
</tr>
<tr>
<td>( h_{MF} = 4 )</td>
</tr>
<tr>
<td>( h_{LF} = 1 )</td>
</tr>
<tr>
<td>0.055</td>
</tr>
<tr>
<td>( h_{MF} = 8 )</td>
</tr>
<tr>
<td>( h_{MF} = 12 )</td>
</tr>
<tr>
<td>A.2. ( d = 0.8 ) (high persistence in ( x_H ))</td>
</tr>
<tr>
<td>A.2.1. ( T_L = 40 ) (small sample)</td>
</tr>
<tr>
<td>( h_{MF} = 4 )</td>
</tr>
<tr>
<td>( h_{LF} = 1 )</td>
</tr>
<tr>
<td>0.058</td>
</tr>
<tr>
<td>( h_{MF} = 8 )</td>
</tr>
<tr>
<td>( h_{MF} = 12 )</td>
</tr>
<tr>
<td>A.2.2. ( T_L = 80 ) (medium sample)</td>
</tr>
<tr>
<td>( h_{MF} = 4 )</td>
</tr>
<tr>
<td>( h_{LF} = 1 )</td>
</tr>
<tr>
<td>0.060</td>
</tr>
<tr>
<td>( h_{MF} = 8 )</td>
</tr>
<tr>
<td>( h_{MF} = 12 )</td>
</tr>
</tbody>
</table>
Table 4: Rejection Frequencies of High-to-Low Causality Tests Based on MF-VAR(1) (Continued)

Panel B. Decaying Causality: $b_j = (-1)^{j-1}0.3/j$

B.1. $d = 0.2$ (low persistence in $x_H$)

<table>
<thead>
<tr>
<th>$T_L = 40$ (small sample size)</th>
<th>$T_L = 80$ (medium sample size)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MF</strong></td>
<td><strong>LF (flow)</strong></td>
</tr>
<tr>
<td>Max Wald</td>
<td>Max Wald</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>$h_{MF} = 4$</td>
<td>0.228 0.241</td>
</tr>
<tr>
<td>$h_{MF} = 8$</td>
<td>0.163 0.157</td>
</tr>
<tr>
<td>$h_{MF} = 12$</td>
<td>0.128 0.136</td>
</tr>
</tbody>
</table>

B.1.1. $T_L = 40$ (small sample size)

<table>
<thead>
<tr>
<th>$T_L = 80$ (medium sample size)</th>
<th>$T_L = 80$ (medium sample size)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MF</strong></td>
<td><strong>LF (flow)</strong></td>
</tr>
<tr>
<td>Max Wald</td>
<td>Max Wald</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>---------------------------------</td>
</tr>
<tr>
<td>$h_{MF} = 4$</td>
<td>0.482 0.527</td>
</tr>
<tr>
<td>$h_{MF} = 8$</td>
<td>0.374 0.412</td>
</tr>
<tr>
<td>$h_{MF} = 12$</td>
<td>0.332 0.335</td>
</tr>
</tbody>
</table>

B.2. $d = 0.8$ (high persistence in $x_H$)

<table>
<thead>
<tr>
<th><strong>B.2.1. $T_L = 40$ (small sample size)</strong></th>
<th><strong>B.2.2. $T_L = 80$ (medium sample size)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MF</strong></td>
<td><strong>LF (flow)</strong></td>
</tr>
<tr>
<td>Max Wald</td>
<td>Max Wald</td>
</tr>
<tr>
<td>------------------------------------------</td>
<td>------------------------------------------</td>
</tr>
<tr>
<td>$h_{MF} = 4$</td>
<td>0.444 0.305</td>
</tr>
<tr>
<td>$h_{MF} = 8$</td>
<td>0.343 0.222</td>
</tr>
<tr>
<td>$h_{MF} = 12$</td>
<td>0.272 0.163</td>
</tr>
</tbody>
</table>

B.2.1. $T_L = 40$ (small sample size)

<table>
<thead>
<tr>
<th><strong>B.2.2. $T_L = 80$ (medium sample size)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MF</strong></td>
</tr>
<tr>
<td>Max Wald</td>
</tr>
<tr>
<td>------------------------------------------</td>
</tr>
<tr>
<td>$h_{MF} = 4$</td>
</tr>
<tr>
<td>$h_{MF} = 8$</td>
</tr>
<tr>
<td>$h_{MF} = 12$</td>
</tr>
</tbody>
</table>
Table 4: Rejection Frequencies of High-to-Low Causality Tests Based on MF-VAR(1) (Continued)

<table>
<thead>
<tr>
<th>Panel C. Lagged Causality: $b_j = 0.3 \times I(j = 12)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.1. $d = 0.2$ (low persistence in $x_H$)</td>
</tr>
<tr>
<td>C.1.1. $T_L = 40$ (small sample size)</td>
</tr>
<tr>
<td>MF</td>
</tr>
<tr>
<td>Max</td>
</tr>
<tr>
<td>----</td>
</tr>
<tr>
<td>$h_{MF} = 4$</td>
</tr>
<tr>
<td>$h_{MF} = 8$</td>
</tr>
<tr>
<td>$h_{MF} = 12$</td>
</tr>
<tr>
<td>C.1.2. $T_L = 80$ (medium sample size)</td>
</tr>
<tr>
<td>MF</td>
</tr>
<tr>
<td>Max</td>
</tr>
<tr>
<td>----</td>
</tr>
<tr>
<td>$h_{MF} = 4$</td>
</tr>
<tr>
<td>$h_{MF} = 8$</td>
</tr>
<tr>
<td>$h_{MF} = 12$</td>
</tr>
<tr>
<td>C.2. $d = 0.8$ (high persistence in $x_H$)</td>
</tr>
<tr>
<td>C.2.1. $T_L = 40$ (small sample size)</td>
</tr>
<tr>
<td>MF</td>
</tr>
<tr>
<td>Max</td>
</tr>
<tr>
<td>----</td>
</tr>
<tr>
<td>$h_{MF} = 4$</td>
</tr>
<tr>
<td>$h_{MF} = 8$</td>
</tr>
<tr>
<td>$h_{MF} = 12$</td>
</tr>
<tr>
<td>C.2.2. $T_L = 80$ (medium sample size)</td>
</tr>
<tr>
<td>MF</td>
</tr>
<tr>
<td>Max</td>
</tr>
<tr>
<td>----</td>
</tr>
<tr>
<td>$h_{MF} = 4$</td>
</tr>
<tr>
<td>$h_{MF} = 8$</td>
</tr>
<tr>
<td>$h_{MF} = 12$</td>
</tr>
</tbody>
</table>
Table 4: Rejection Frequencies of High-to-Low Causality Tests Based on MF-VAR(1) (Continued)

Panel D. Sporadic Causality: \((b_3, b_7, b_{10}) = (0.2, 0.05, -0.3)\)

<table>
<thead>
<tr>
<th></th>
<th>(MF = 4)</th>
<th>(LF = 1)</th>
<th>(LF = 2)</th>
<th>(LF = 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Max</td>
<td>Wald</td>
<td>Max</td>
<td>Wald</td>
</tr>
<tr>
<td>(d = 0.2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(T_L = 40)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(h_{MF} = 4)</td>
<td>0.119 0.075</td>
<td>(h_{LF} = 1)</td>
<td>0.060 0.043</td>
</tr>
<tr>
<td></td>
<td>(h_{MF} = 8)</td>
<td>0.101 0.079</td>
<td>(h_{LF} = 2)</td>
<td>0.061 0.043</td>
</tr>
<tr>
<td></td>
<td>(h_{MF} = 12)</td>
<td>0.173 0.136</td>
<td>(h_{LF} = 3)</td>
<td>0.067 0.048</td>
</tr>
<tr>
<td>(T_L = 80)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(h_{MF} = 4)</td>
<td>0.248 0.207</td>
<td>(h_{LF} = 1)</td>
<td>0.075 0.057</td>
</tr>
<tr>
<td></td>
<td>(h_{MF} = 8)</td>
<td>0.184 0.167</td>
<td>(h_{LF} = 2)</td>
<td>0.052 0.049</td>
</tr>
<tr>
<td></td>
<td>(h_{MF} = 12)</td>
<td>0.442 0.416</td>
<td>(h_{LF} = 3)</td>
<td>0.050 0.055</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>(MF = 8)</th>
<th>(LF = 1)</th>
<th>(LF = 2)</th>
<th>(LF = 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Max</td>
<td>Wald</td>
<td>Max</td>
<td>Wald</td>
</tr>
<tr>
<td>(d = 0.8)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(T_L = 40)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(h_{MF} = 4)</td>
<td>0.243 0.163</td>
<td>(h_{LF} = 1)</td>
<td>0.075 0.047</td>
</tr>
<tr>
<td></td>
<td>(h_{MF} = 8)</td>
<td>0.197 0.180</td>
<td>(h_{LF} = 2)</td>
<td>0.093 0.057</td>
</tr>
<tr>
<td></td>
<td>(h_{MF} = 12)</td>
<td>0.402 0.260</td>
<td>(h_{LF} = 3)</td>
<td>0.082 0.063</td>
</tr>
<tr>
<td>(T_L = 80)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(h_{MF} = 4)</td>
<td>0.459 0.305</td>
<td>(h_{LF} = 1)</td>
<td>0.076 0.052</td>
</tr>
<tr>
<td></td>
<td>(h_{MF} = 8)</td>
<td>0.404 0.404</td>
<td>(h_{LF} = 2)</td>
<td>0.128 0.101</td>
</tr>
<tr>
<td></td>
<td>(h_{MF} = 12)</td>
<td>0.803 0.740</td>
<td>(h_{LF} = 3)</td>
<td>0.107 0.067</td>
</tr>
</tbody>
</table>
Table 5: Rejection Frequencies of High-to-Low Causality Tests Based on MF-VAR(2)

This table compares the rejection frequencies of four different high-to-low causality tests: mixed frequency max test, mixed frequency Wald test, low frequency max test, and low frequency Wald test. The DGP is MF-VAR(2) with $m = 12$. We have four panels depending on the key parameter $b = [b_1, \ldots, b_{24}]$, where $b_j$ signifies the impact of $x_H(\tau_L - 1, m + 1 - j)$ on $x_L(\tau_L)$. Panel A deals with Non-causality: $b = 0_{24 \times 1}$. Panel B deals with Decaying Causality: $b_j = (-1)^{j-1} \times 0.3/j$ for $j = 1, \ldots, 24$. Panel C deals with Lagged Causality: $b_j = 0.3 \times I(j = 24)$ for $j = 1, \ldots, 12$. Panel D deals with Sporadic Causality: $(b_5, b_{12}, b_{17}, b_{19}) = (-0.2, 0.1, 0.2, -0.35)$, and all other $b$'s are zeros. For each panel we consider low persistence of $x_H$ (i.e. $d = 0.2$) and high persistence (i.e. $d = 0.8$). We also try small sample size $T_L = 40$ and medium sample size $T_L = 80$ for each case. Other parameters are specified as follows. First, we assume fairly weak autoregressive property for $x_L$ (i.e. $\alpha = 0.2$). Second, for low-to-high causality, we assume decaying causality with alternating signs up to low frequency lag $1$: $c_j = (-1)^{j-1} \times 0.4/j$ for $j = 1, \ldots, 12$. Nominal size is $\alpha = 0.05$. For all four models, we include two low frequency lags of $x_L$ (i.e. $q = 2$). For max tests, the number of draws from the limit distributions under $H_0$ is 5,000. The weighting scheme is simply an equal scheme: $W_h = (1/h) \times I_h$. The number of high frequency lags of $x_H$ taken by mixed frequency tests is $h_{MF} \in \{16, 20, 24\}$, while the number of low frequency lags of $x_H$ taken by low frequency tests is $h_{LF} \in \{1, 2, 3\}$. For low frequency tests, we consider both flow sampling and stock sampling. For Wald tests we use the parametric bootstrap based on Gonçalves and Killian (2004) in order to control empirical size. The number of bootstrap replications is $N = 499$. The number of Monte Carlo iterations is 5,000 for max tests and 1,000 for Wald tests.

Panel A. Non-causality: $b = 0_{24 \times 1}$

<table>
<thead>
<tr>
<th>$A.1. d = 0.2$ (low persistence in $x_H$)</th>
<th>$A.2. d = 0.8$ (high persistence in $x_H$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A.1. $T_L = 40$ (small sample size)</strong></td>
<td><strong>A.2. $T_L = 40$ (small sample size)</strong></td>
</tr>
<tr>
<td>MF</td>
<td>LF (flow)</td>
</tr>
<tr>
<td>Max</td>
<td>Wald</td>
</tr>
<tr>
<td>$h_{MF} = 16$</td>
<td>0.062</td>
</tr>
<tr>
<td>$h_{MF} = 20$</td>
<td>0.061</td>
</tr>
<tr>
<td>$h_{MF} = 24$</td>
<td>0.055</td>
</tr>
<tr>
<td><strong>A.1. $T_L = 80$ (medium sample size)</strong></td>
<td><strong>A.2. $T_L = 80$ (medium sample size)</strong></td>
</tr>
<tr>
<td>MF</td>
<td>LF (flow)</td>
</tr>
<tr>
<td>Max</td>
<td>Wald</td>
</tr>
<tr>
<td>$h_{MF} = 16$</td>
<td>0.064</td>
</tr>
<tr>
<td>$h_{MF} = 20$</td>
<td>0.056</td>
</tr>
<tr>
<td>$h_{MF} = 24$</td>
<td>0.050</td>
</tr>
</tbody>
</table>

46
Table 5: Rejection Frequencies of High-to-Low Causality Tests Based on MF-VAR(2) (Continued)

Panel B. Decaying Causality: $b_j = (-1)^{j-1}0.3/j$

<table>
<thead>
<tr>
<th></th>
<th>MF (flow)</th>
<th>LF (stock)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Max</td>
<td>Wald</td>
</tr>
<tr>
<td><strong>B.1. $d = 0.2$ (low persistence in $x_H$)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>B.1.1. $T_L = 40$ (small sample size)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$h_{MF} = 16$</td>
<td>0.118</td>
<td>0.095</td>
</tr>
<tr>
<td>$h_{MF} = 20$</td>
<td>0.109</td>
<td>0.092</td>
</tr>
<tr>
<td>$h_{MF} = 24$</td>
<td>0.094</td>
<td>0.076</td>
</tr>
<tr>
<td><strong>B.1.2. $T_L = 80$ (medium sample size)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$h_{MF} = 16$</td>
<td>0.284</td>
<td>0.290</td>
</tr>
<tr>
<td>$h_{MF} = 20$</td>
<td>0.258</td>
<td>0.250</td>
</tr>
<tr>
<td>$h_{MF} = 24$</td>
<td>0.246</td>
<td>0.221</td>
</tr>
<tr>
<td><strong>B.2. $d = 0.8$ (high persistence in $x_H$)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>B.2.1. $T_L = 40$ (small sample size)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$h_{MF} = 16$</td>
<td>0.233</td>
<td>0.120</td>
</tr>
<tr>
<td>$h_{MF} = 20$</td>
<td>0.200</td>
<td>0.091</td>
</tr>
<tr>
<td>$h_{MF} = 24$</td>
<td>0.190</td>
<td>0.077</td>
</tr>
<tr>
<td><strong>B.2.2. $T_L = 80$ (medium sample size)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$h_{MF} = 16$</td>
<td>0.610</td>
<td>0.398</td>
</tr>
<tr>
<td>$h_{MF} = 20$</td>
<td>0.563</td>
<td>0.356</td>
</tr>
<tr>
<td>$h_{MF} = 24$</td>
<td>0.545</td>
<td>0.271</td>
</tr>
</tbody>
</table>
Table 5: Rejection Frequencies of High-to-Low Causality Tests Based on MF-VAR(2) (Continued)

Panel C. Lagged Causality: $b_j = 0.3 \times I(j = 24)$

<table>
<thead>
<tr>
<th></th>
<th>MF (flow)</th>
<th>LF (stock)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Max Wald</td>
<td>Max Wald</td>
</tr>
<tr>
<td>$h_{MF} = 16$</td>
<td>0.060 0.044</td>
<td>0.061 0.036</td>
</tr>
<tr>
<td>$h_{MF} = 20$</td>
<td>0.060 0.046</td>
<td>0.082 0.051</td>
</tr>
<tr>
<td>$h_{MF} = 24$</td>
<td>0.091 0.055</td>
<td>0.083 0.055</td>
</tr>
</tbody>
</table>

C.1. $d = 0.2$ (low persistence in $x_H$)

<table>
<thead>
<tr>
<th></th>
<th>MF (flow)</th>
<th>LF (stock)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Max Wald</td>
<td>Max Wald</td>
</tr>
<tr>
<td>$h_{MF} = 16$</td>
<td>0.069 0.053</td>
<td>0.067 0.045</td>
</tr>
<tr>
<td>$h_{MF} = 20$</td>
<td>0.162 0.096</td>
<td>0.515 0.464</td>
</tr>
<tr>
<td>$h_{MF} = 24$</td>
<td>0.896 0.480</td>
<td>0.588 0.505</td>
</tr>
</tbody>
</table>

C.1. $T_L = 40$ (small sample size)

<table>
<thead>
<tr>
<th></th>
<th>MF (flow)</th>
<th>LF (stock)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Max Wald</td>
<td>Max Wald</td>
</tr>
<tr>
<td>$h_{MF} = 16$</td>
<td>0.060 0.044</td>
<td>0.061 0.036</td>
</tr>
<tr>
<td>$h_{MF} = 20$</td>
<td>0.060 0.046</td>
<td>0.082 0.051</td>
</tr>
<tr>
<td>$h_{MF} = 24$</td>
<td>0.091 0.055</td>
<td>0.083 0.055</td>
</tr>
</tbody>
</table>

C.1. $T_L = 80$ (medium sample size)

<table>
<thead>
<tr>
<th></th>
<th>MF (flow)</th>
<th>LF (stock)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Max Wald</td>
<td>Max Wald</td>
</tr>
<tr>
<td>$h_{MF} = 16$</td>
<td>0.069 0.053</td>
<td>0.067 0.045</td>
</tr>
<tr>
<td>$h_{MF} = 20$</td>
<td>0.162 0.096</td>
<td>0.515 0.464</td>
</tr>
<tr>
<td>$h_{MF} = 24$</td>
<td>0.896 0.480</td>
<td>0.588 0.505</td>
</tr>
</tbody>
</table>

C.2. $d = 0.8$ (high persistence in $x_H$)

<table>
<thead>
<tr>
<th></th>
<th>MF (flow)</th>
<th>LF (stock)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Max Wald</td>
<td>Max Wald</td>
</tr>
<tr>
<td>$h_{MF} = 16$</td>
<td>0.063 0.042</td>
<td>0.064 0.040</td>
</tr>
<tr>
<td>$h_{MF} = 20$</td>
<td>0.095 0.073</td>
<td>0.260 0.196</td>
</tr>
<tr>
<td>$h_{MF} = 24$</td>
<td>0.438 0.108</td>
<td>0.285 0.189</td>
</tr>
</tbody>
</table>

C.2. $T_L = 40$ (small sample size)

<table>
<thead>
<tr>
<th></th>
<th>MF (flow)</th>
<th>LF (stock)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Max Wald</td>
<td>Max Wald</td>
</tr>
<tr>
<td>$h_{MF} = 16$</td>
<td>0.063 0.042</td>
<td>0.064 0.040</td>
</tr>
<tr>
<td>$h_{MF} = 20$</td>
<td>0.095 0.073</td>
<td>0.260 0.196</td>
</tr>
<tr>
<td>$h_{MF} = 24$</td>
<td>0.438 0.108</td>
<td>0.285 0.189</td>
</tr>
</tbody>
</table>

C.2. $T_L = 80$ (medium sample size)

<table>
<thead>
<tr>
<th></th>
<th>MF (flow)</th>
<th>LF (stock)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Max Wald</td>
<td>Max Wald</td>
</tr>
<tr>
<td>$h_{MF} = 16$</td>
<td>0.069 0.053</td>
<td>0.067 0.045</td>
</tr>
<tr>
<td>$h_{MF} = 20$</td>
<td>0.162 0.096</td>
<td>0.515 0.464</td>
</tr>
<tr>
<td>$h_{MF} = 24$</td>
<td>0.896 0.480</td>
<td>0.588 0.505</td>
</tr>
</tbody>
</table>
Table 5: Rejection Frequencies of High-to-Low Causality Tests Based on MF-VAR(2) (Continued)

Panel D. Sporadic Causality: \((b_5, b_{12}, b_{17}, b_{19}) = (−0.2, 0.1, 0.2, −0.35)\)

<table>
<thead>
<tr>
<th></th>
<th>MF Max Wald</th>
<th>LF (flow) Max Wald</th>
<th>LF (stock) Max Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>D.1.  (d = 0.2) (low persistence in (x_H))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D.1.1. (T_L = 40) (small sample size)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(h_{MF} = 16)</td>
<td>0.090 0.064</td>
<td>(h_{LF} = 1) 0.072 0.039</td>
<td>0.066 0.043</td>
</tr>
<tr>
<td>(h_{MF} = 20)</td>
<td>0.186 0.146</td>
<td>(h_{LF} = 2) 0.067 0.065</td>
<td>0.057 0.056</td>
</tr>
<tr>
<td>(h_{MF} = 24)</td>
<td>0.153 0.089</td>
<td>(h_{LF} = 3) 0.066 0.059</td>
<td>0.062 0.043</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D.1.2. (T_L = 80) (medium sample size)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(h_{MF} = 16)</td>
<td>0.141 0.118</td>
<td>(h_{LF} = 1) 0.066 0.063</td>
<td>0.056 0.048</td>
</tr>
<tr>
<td>(h_{MF} = 20)</td>
<td>0.502 0.474</td>
<td>(h_{LF} = 2) 0.073 0.053</td>
<td>0.055 0.040</td>
</tr>
<tr>
<td>(h_{MF} = 24)</td>
<td>0.461 0.419</td>
<td>(h_{LF} = 3) 0.068 0.071</td>
<td>0.057 0.061</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D.2.  (d = 0.8) (high persistence in (x_H))</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D.2.1. (T_L = 40) (small sample size)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(h_{MF} = 16)</td>
<td>0.170 0.090</td>
<td>(h_{LF} = 1) 0.158 0.106</td>
<td>0.117 0.067</td>
</tr>
<tr>
<td>(h_{MF} = 20)</td>
<td>0.254 0.187</td>
<td>(h_{LF} = 2) 0.175 0.137</td>
<td>0.101 0.079</td>
</tr>
<tr>
<td>(h_{MF} = 24)</td>
<td>0.256 0.136</td>
<td>(h_{LF} = 3) 0.148 0.097</td>
<td>0.100 0.075</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>D.2.2. (T_L = 80) (medium sample size)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(h_{MF} = 16)</td>
<td>0.390 0.234</td>
<td>(h_{LF} = 1) 0.272 0.243</td>
<td>0.151 0.118</td>
</tr>
<tr>
<td>(h_{MF} = 20)</td>
<td>0.621 0.655</td>
<td>(h_{LF} = 2) 0.311 0.252</td>
<td>0.134 0.132</td>
</tr>
<tr>
<td>(h_{MF} = 24)</td>
<td>0.623 0.606</td>
<td>(h_{LF} = 3) 0.269 0.226</td>
<td>0.163 0.152</td>
</tr>
</tbody>
</table>
Table 6: Rejection Frequencies for Low-to-High Causality Tests

This table shows rejection frequencies for low-to-high Granger causality. We fix $m = 12$, which can be thought of as a week vs. quarter case approximately. Panel A assumes Non-causality, Panel B assumes Decaying Causality, Panel C assumes Lagged Causality, and Panel D assumes Sporadic Causality. For each panel we have small sample size (40 quarters) and medium sample size (80 quarters). For each sample size we have mixed frequency tests, low frequency tests with stock sampling, and low frequency tests with flow sampling. Both max tests and Wald tests are implemented for each of these three cases. For mixed frequency tests, the number of high frequency leads of $x_H$ is taken from $r_{MF} \in \{4, 8, 12\}$ while the number of high frequency lags of $x_H$ is taken from $h_{MF} \in \{4, 8, 12\}$. For low frequency tests, the number of low frequency leads of aggregated $x_H$ is taken from $r_{LF} \in \{1, 2, 3\}$ while the number of low frequency lags of aggregated $x_H$ is taken from $h_{LF} \in \{1, 2, 3\}$. Max tests use the equal weighting scheme, and the test statistic is computed based on 1,000 draws from the asymptotic null distribution. In particular, the mixed frequency max test uses the Almon polynomial with dimension $s = 3$ for the lag terms of $x_H$. Wald tests use Gonçalves and Kilian’s (2004) bootstrap with $N = 499$ replications. There is decaying Granger causality from $x_H$ to $x_L$ in the sense that $x_L(\tau_L)$ depends on $\sum_{j=1}^{12} (-1)^{j-1} (0.2/j) x_H(\tau_L - 1, m + 1 - j)$. The high frequency AR(1) coefficient of $x_H$ is 0.2, and the low frequency AR(1) coefficient of $x_L$ is also 0.2. The number of Monte Carlo replications is 5,000 for max tests and 1,000 for Wald tests. The nominal size is 5%.
Table 6: Rejection Frequencies for Low-to-High Causality Tests (Continued)

Panel A. Non-causality: \( c = 0_{12 \times 1} \)

Panel A.1. \( T_L = 40 \) (Small Sample)

Panel A.1.1. Mixed Frequency Tests

<table>
<thead>
<tr>
<th>( r_{MF} )</th>
<th>( h_{MF} )</th>
<th>( h_{MF} )</th>
<th>( h_{MF} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wald max</td>
<td>Wald max</td>
<td>Wald max</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.071</td>
<td>0.069</td>
<td>0.071</td>
</tr>
<tr>
<td>8</td>
<td>0.073</td>
<td>0.070</td>
<td>0.066</td>
</tr>
<tr>
<td>12</td>
<td>0.066</td>
<td>0.068</td>
<td>0.068</td>
</tr>
</tbody>
</table>

Panel A.1.2. Low Frequency Tests (Stock Sampling)

<table>
<thead>
<tr>
<th>( r_{LF} )</th>
<th>( h_{LF} )</th>
<th>( h_{LF} )</th>
<th>( h_{LF} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wald max</td>
<td>Wald max</td>
<td>Wald max</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.057</td>
<td>0.053</td>
<td>0.057</td>
</tr>
<tr>
<td>2</td>
<td>0.064</td>
<td>0.062</td>
<td>0.056</td>
</tr>
<tr>
<td>3</td>
<td>0.062</td>
<td>0.068</td>
<td>0.068</td>
</tr>
</tbody>
</table>

Panel A.1.3. Low Frequency Tests (Flow Sampling)

<table>
<thead>
<tr>
<th>( r_{LF} )</th>
<th>( h_{LF} )</th>
<th>( h_{LF} )</th>
<th>( h_{LF} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wald max</td>
<td>Wald max</td>
<td>Wald max</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.058</td>
<td>0.054</td>
<td>0.050</td>
</tr>
<tr>
<td>2</td>
<td>0.059</td>
<td>0.051</td>
<td>0.054</td>
</tr>
<tr>
<td>3</td>
<td>0.061</td>
<td>0.055</td>
<td>0.057</td>
</tr>
</tbody>
</table>

Panel A.2. \( T_L = 80 \) (Medium Sample)

Panel A.2.1. Mixed Frequency Tests

<table>
<thead>
<tr>
<th>( r_{MF} )</th>
<th>( h_{MF} )</th>
<th>( h_{MF} )</th>
<th>( h_{MF} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wald max</td>
<td>Wald max</td>
<td>Wald max</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.055</td>
<td>0.057</td>
<td>0.054</td>
</tr>
<tr>
<td>8</td>
<td>0.056</td>
<td>0.051</td>
<td>0.055</td>
</tr>
<tr>
<td>12</td>
<td>0.060</td>
<td>0.058</td>
<td>0.052</td>
</tr>
</tbody>
</table>

Panel A.2.2. Low Frequency Tests (Stock Sampling)

<table>
<thead>
<tr>
<th>( r_{LF} )</th>
<th>( h_{LF} )</th>
<th>( h_{LF} )</th>
<th>( h_{LF} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wald max</td>
<td>Wald max</td>
<td>Wald max</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.058</td>
<td>0.051</td>
<td>0.050</td>
</tr>
<tr>
<td>2</td>
<td>0.059</td>
<td>0.051</td>
<td>0.054</td>
</tr>
<tr>
<td>3</td>
<td>0.061</td>
<td>0.055</td>
<td>0.057</td>
</tr>
</tbody>
</table>

Panel A.2.3. Low Frequency Tests (Flow Sampling)

<table>
<thead>
<tr>
<th>( r_{LF} )</th>
<th>( h_{LF} )</th>
<th>( h_{LF} )</th>
<th>( h_{LF} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wald max</td>
<td>Wald max</td>
<td>Wald max</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.054</td>
<td>0.046</td>
<td>0.050</td>
</tr>
<tr>
<td>2</td>
<td>0.052</td>
<td>0.050</td>
<td>0.052</td>
</tr>
<tr>
<td>3</td>
<td>0.057</td>
<td>0.051</td>
<td>0.056</td>
</tr>
</tbody>
</table>
Table 6: Rejection Frequencies for Low-to-High Causality Tests (Continued)

Panel B. Decaying Causality: $c_j = (-1)^{j-1} \times 0.45/j$, $j = 1, \ldots, 12$

Panel B.1. $T_L = 40$ (Small Sample)

<table>
<thead>
<tr>
<th></th>
<th>$r_{MF} = 4$ max Wald</th>
<th>$r_{MF} = 8$ max Wald</th>
<th>$r_{MF} = 12$ max Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_{MF} = 4$</td>
<td>0.481</td>
<td>0.515</td>
<td>0.344</td>
</tr>
<tr>
<td>$h_{MF} = 8$</td>
<td>0.482</td>
<td>0.453</td>
<td>0.339</td>
</tr>
<tr>
<td>$h_{MF} = 12$</td>
<td>0.487</td>
<td>0.371</td>
<td>0.338</td>
</tr>
</tbody>
</table>

Panel B.1.1. Mixed Frequency Tests

<table>
<thead>
<tr>
<th></th>
<th>$r_{LF} = 1$ max Wald</th>
<th>$r_{LF} = 2$ max Wald</th>
<th>$r_{LF} = 3$ max Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_{LF} = 1$</td>
<td>0.061</td>
<td>0.047</td>
<td>0.058</td>
</tr>
<tr>
<td>$h_{LF} = 2$</td>
<td>0.064</td>
<td>0.050</td>
<td>0.054</td>
</tr>
<tr>
<td>$h_{LF} = 3$</td>
<td>0.070</td>
<td>0.058</td>
<td>0.068</td>
</tr>
</tbody>
</table>

Panel B.1.2. Low Frequency Tests (Stock Sampling)

<table>
<thead>
<tr>
<th></th>
<th>$r_{LF} = 1$ max Wald</th>
<th>$r_{LF} = 2$ max Wald</th>
<th>$r_{LF} = 3$ max Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_{LF} = 1$</td>
<td>0.090</td>
<td>0.072</td>
<td>0.083</td>
</tr>
<tr>
<td>$h_{LF} = 2$</td>
<td>0.093</td>
<td>0.071</td>
<td>0.083</td>
</tr>
<tr>
<td>$h_{LF} = 3$</td>
<td>0.097</td>
<td>0.058</td>
<td>0.088</td>
</tr>
</tbody>
</table>

Panel B.1.3. Low Frequency Tests (Flow Sampling)

<table>
<thead>
<tr>
<th></th>
<th>$r_{LF} = 1$ max Wald</th>
<th>$r_{LF} = 2$ max Wald</th>
<th>$r_{LF} = 3$ max Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_{LF} = 1$</td>
<td>0.130</td>
<td>0.120</td>
<td>0.105</td>
</tr>
<tr>
<td>$h_{LF} = 2$</td>
<td>0.129</td>
<td>0.106</td>
<td>0.103</td>
</tr>
<tr>
<td>$h_{LF} = 3$</td>
<td>0.119</td>
<td>0.113</td>
<td>0.106</td>
</tr>
</tbody>
</table>

Panel B.2. $T_L = 80$ (Medium Sample)

Panel B.2.1. Mixed Frequency Tests

<table>
<thead>
<tr>
<th></th>
<th>$r_{MF} = 4$ max Wald</th>
<th>$r_{MF} = 8$ max Wald</th>
<th>$r_{MF} = 12$ max Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_{MF} = 4$</td>
<td>0.886</td>
<td>0.945</td>
<td>0.804</td>
</tr>
<tr>
<td>$h_{MF} = 8$</td>
<td>0.894</td>
<td>0.911</td>
<td>0.818</td>
</tr>
<tr>
<td>$h_{MF} = 12$</td>
<td>0.895</td>
<td>0.889</td>
<td>0.806</td>
</tr>
</tbody>
</table>

Panel B.2.2. Low Frequency Tests (Stock Sampling)

<table>
<thead>
<tr>
<th></th>
<th>$r_{LF} = 1$ max Wald</th>
<th>$r_{LF} = 2$ max Wald</th>
<th>$r_{LF} = 3$ max Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_{LF} = 1$</td>
<td>0.058</td>
<td>0.046</td>
<td>0.047</td>
</tr>
<tr>
<td>$h_{LF} = 2$</td>
<td>0.069</td>
<td>0.060</td>
<td>0.059</td>
</tr>
<tr>
<td>$h_{LF} = 3$</td>
<td>0.065</td>
<td>0.044</td>
<td>0.058</td>
</tr>
</tbody>
</table>

Panel B.2.3. Low Frequency Tests (Flow Sampling)

<table>
<thead>
<tr>
<th></th>
<th>$r_{LF} = 1$ max Wald</th>
<th>$r_{LF} = 2$ max Wald</th>
<th>$r_{LF} = 3$ max Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_{LF} = 1$</td>
<td>0.130</td>
<td>0.120</td>
<td>0.105</td>
</tr>
<tr>
<td>$h_{LF} = 2$</td>
<td>0.129</td>
<td>0.106</td>
<td>0.103</td>
</tr>
<tr>
<td>$h_{LF} = 3$</td>
<td>0.119</td>
<td>0.113</td>
<td>0.106</td>
</tr>
</tbody>
</table>

52
Table 6: Rejection Frequencies for Low-to-High Causality Tests (Continued)

Panel C. Lagged Causality: $c_j = 0.4 \times I(j = 12), j = 1, \ldots, 12$

Panel C.1. $T_L = 40$ (Small Sample)

Panel C.1.1. Mixed Frequency Tests

<table>
<thead>
<tr>
<th>$r_{MF}$</th>
<th>max Wald</th>
<th>max Wald</th>
<th>max Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_{MF} = 4$</td>
<td>0.076</td>
<td>0.054</td>
<td>0.064</td>
</tr>
<tr>
<td>$h_{MF} = 8$</td>
<td>0.067</td>
<td>0.040</td>
<td>0.068</td>
</tr>
<tr>
<td>$h_{MF} = 12$</td>
<td>0.068</td>
<td>0.042</td>
<td>0.069</td>
</tr>
</tbody>
</table>

Panel C.1.2. Low Frequency Tests (Stock Sampling)

<table>
<thead>
<tr>
<th>$r_{LF}$</th>
<th>max Wald</th>
<th>max Wald</th>
<th>max Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_{LF} = 1$</td>
<td>0.640</td>
<td>0.560</td>
<td>0.510</td>
</tr>
<tr>
<td>$h_{LF} = 2$</td>
<td>0.616</td>
<td>0.525</td>
<td>0.518</td>
</tr>
<tr>
<td>$h_{LF} = 3$</td>
<td>0.616</td>
<td>0.485</td>
<td>0.501</td>
</tr>
</tbody>
</table>

Panel C.1.3. Low Frequency Tests (Flow Sampling)

<table>
<thead>
<tr>
<th>$r_{LF}$</th>
<th>max Wald</th>
<th>max Wald</th>
<th>max Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_{LF} = 1$</td>
<td>0.110</td>
<td>0.072</td>
<td>0.081</td>
</tr>
<tr>
<td>$h_{LF} = 2$</td>
<td>0.106</td>
<td>0.070</td>
<td>0.086</td>
</tr>
<tr>
<td>$h_{LF} = 3$</td>
<td>0.107</td>
<td>0.083</td>
<td>0.089</td>
</tr>
</tbody>
</table>

Panel C.2. $T_L = 80$ (Medium Sample)

Panel C.2.1. Mixed Frequency Tests

<table>
<thead>
<tr>
<th>$r_{MF}$</th>
<th>max Wald</th>
<th>max Wald</th>
<th>max Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_{MF} = 4$</td>
<td>0.059</td>
<td>0.047</td>
<td>0.060</td>
</tr>
<tr>
<td>$h_{MF} = 8$</td>
<td>0.058</td>
<td>0.046</td>
<td>0.064</td>
</tr>
<tr>
<td>$h_{MF} = 12$</td>
<td>0.058</td>
<td>0.041</td>
<td>0.059</td>
</tr>
</tbody>
</table>

Panel C.2.2. Low Frequency Tests (Stock Sampling)

<table>
<thead>
<tr>
<th>$r_{LF}$</th>
<th>max Wald</th>
<th>max Wald</th>
<th>max Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_{LF} = 1$</td>
<td>0.915</td>
<td>0.892</td>
<td>0.876</td>
</tr>
<tr>
<td>$h_{LF} = 2$</td>
<td>0.919</td>
<td>0.912</td>
<td>0.863</td>
</tr>
<tr>
<td>$h_{LF} = 3$</td>
<td>0.908</td>
<td>0.883</td>
<td>0.855</td>
</tr>
</tbody>
</table>

Panel C.2.3. Low Frequency Tests (Flow Sampling)

<table>
<thead>
<tr>
<th>$r_{LF}$</th>
<th>max Wald</th>
<th>max Wald</th>
<th>max Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_{LF} = 1$</td>
<td>0.135</td>
<td>0.115</td>
<td>0.121</td>
</tr>
<tr>
<td>$h_{LF} = 2$</td>
<td>0.143</td>
<td>0.128</td>
<td>0.108</td>
</tr>
<tr>
<td>$h_{LF} = 3$</td>
<td>0.132</td>
<td>0.118</td>
<td>0.121</td>
</tr>
</tbody>
</table>
Table 6: Rejection Frequencies for Low-to-High Causality Tests (Continued)

Panel D. Sporadic Causality: \((c_3, c_7, c_{10}) = (0.4, 0.25, -0.5)\)

Panel D.1. \(T_L = 40\) (Small Sample)

### Panel D.1.1. Mixed Frequency Tests

<table>
<thead>
<tr>
<th>(r_{MF} = 4)</th>
<th>(r_{MF} = 8)</th>
<th>(r_{MF} = 12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>max Wald</td>
<td>max Wald</td>
<td>max Wald</td>
</tr>
<tr>
<td>(h_{MF} = 4)</td>
<td>0.373</td>
<td>0.304</td>
</tr>
<tr>
<td>(h_{MF} = 8)</td>
<td>0.364</td>
<td>0.261</td>
</tr>
<tr>
<td>(h_{MF} = 12)</td>
<td>0.363</td>
<td>0.250</td>
</tr>
</tbody>
</table>

### Panel D.1.2. Low Frequency Tests (Stock Sampling)

<table>
<thead>
<tr>
<th>(r_{LF} = 1)</th>
<th>(r_{LF} = 2)</th>
<th>(r_{LF} = 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>max Wald</td>
<td>max Wald</td>
<td>max Wald</td>
</tr>
<tr>
<td>(h_{LF} = 1)</td>
<td>0.067</td>
<td>0.047</td>
</tr>
<tr>
<td>(h_{LF} = 2)</td>
<td>0.061</td>
<td>0.044</td>
</tr>
<tr>
<td>(h_{LF} = 3)</td>
<td>0.067</td>
<td>0.055</td>
</tr>
</tbody>
</table>

### Panel D.1.3. Low Frequency Tests (Flow Sampling)

<table>
<thead>
<tr>
<th>(r_{LF} = 1)</th>
<th>(r_{LF} = 2)</th>
<th>(r_{LF} = 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>max Wald</td>
<td>max Wald</td>
<td>max Wald</td>
</tr>
<tr>
<td>(h_{LF} = 1)</td>
<td>0.066</td>
<td>0.049</td>
</tr>
<tr>
<td>(h_{LF} = 2)</td>
<td>0.070</td>
<td>0.053</td>
</tr>
<tr>
<td>(h_{LF} = 3)</td>
<td>0.077</td>
<td>0.031</td>
</tr>
</tbody>
</table>

Panel D.2. \(T_L = 80\) (Medium Sample)

### Panel D.2.1. Mixed Frequency Tests

<table>
<thead>
<tr>
<th>(r_{MF} = 4)</th>
<th>(r_{MF} = 8)</th>
<th>(r_{MF} = 12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>max Wald</td>
<td>max Wald</td>
<td>max Wald</td>
</tr>
<tr>
<td>(h_{MF} = 4)</td>
<td>0.762</td>
<td>0.730</td>
</tr>
<tr>
<td>(h_{MF} = 8)</td>
<td>0.762</td>
<td>0.691</td>
</tr>
<tr>
<td>(h_{MF} = 12)</td>
<td>0.772</td>
<td>0.637</td>
</tr>
</tbody>
</table>

### Panel D.2.2. Low Frequency Tests (Stock Sampling)

<table>
<thead>
<tr>
<th>(r_{LF} = 1)</th>
<th>(r_{LF} = 2)</th>
<th>(r_{LF} = 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>max Wald</td>
<td>max Wald</td>
<td>max Wald</td>
</tr>
<tr>
<td>(h_{LF} = 1)</td>
<td>0.059</td>
<td>0.051</td>
</tr>
<tr>
<td>(h_{LF} = 2)</td>
<td>0.063</td>
<td>0.046</td>
</tr>
<tr>
<td>(h_{LF} = 3)</td>
<td>0.058</td>
<td>0.043</td>
</tr>
</tbody>
</table>

### Panel D.2.3. Low Frequency Tests (Flow Sampling)

<table>
<thead>
<tr>
<th>(r_{LF} = 1)</th>
<th>(r_{LF} = 2)</th>
<th>(r_{LF} = 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>max Wald</td>
<td>max Wald</td>
<td>max Wald</td>
</tr>
<tr>
<td>(h_{LF} = 1)</td>
<td>0.072</td>
<td>0.078</td>
</tr>
<tr>
<td>(h_{LF} = 2)</td>
<td>0.074</td>
<td>0.065</td>
</tr>
<tr>
<td>(h_{LF} = 3)</td>
<td>0.082</td>
<td>0.069</td>
</tr>
</tbody>
</table>
Table 7: Sample Statistics of U.S. Interest Rates and Real GDP Growth

Sample statistics of weekly 10-year Treasury constant maturity rate, weekly effective federal fund rate, their spread (10Y - FF), and the quarterly real GDP growth from previous year. All these series are in terms of percentage. The sample period covers January 5, 1962 through December 31, 2013, which has 2,736 weeks or 208 quarters.

<table>
<thead>
<tr>
<th></th>
<th>mean</th>
<th>median</th>
<th>std. dev.</th>
<th>skewness</th>
<th>kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-Year</td>
<td>6.555</td>
<td>6.210</td>
<td>2.734</td>
<td>0.781</td>
<td>3.488</td>
</tr>
<tr>
<td>FF</td>
<td>5.563</td>
<td>5.250</td>
<td>3.643</td>
<td>0.928</td>
<td>4.615</td>
</tr>
<tr>
<td>Spread</td>
<td>0.991</td>
<td>1.160</td>
<td>1.800</td>
<td>-1.198</td>
<td>5.611</td>
</tr>
<tr>
<td>GDP</td>
<td>3.151</td>
<td>3.250</td>
<td>2.349</td>
<td>-0.461</td>
<td>3.543</td>
</tr>
</tbody>
</table>
Table 8: Granger Causality from Interest Rate Spread to Real GDP Growth (Full Sample Analysis)

We implement the mixed frequency max test, mixed frequency Wald test, low frequency max test, and low frequency Wald test on the full sample covering 52 years from January 1962 through December 2013. The null hypothesis is Granger non-causality from interest rate spread to real GDP growth. Mixed frequency tests work on weekly spread and quarterly GDP, while low frequency tests work on quarterly spread and quarterly GDP. Specification 1 assumes \((q, h_{MF}, h_{LF}) = (2, 26, 3)\), which means that (i) each model has 2 quarters of low frequency lags of \(x_L\), (ii) each mixed frequency model has 26 weeks of high frequency lags of \(x_H\), and (iii) each low frequency model has 3 quarters of low frequency lags of \(x_H\). Specification 2 assumes larger models: \((q, h_{MF}, h_{LF}) = (4, 52, 6)\). For both Specifications 1 and 2, the number of bootstrap replications for Wald tests is 9,999. The table below reports the \(p\)-values of each test with respect to the null hypothesis of non-causality. Two asterisks are put for a \(p\)-value less than 1%, indicating a strong rejection of non-causality. One asterisk is put for a \(p\)-value less than 5% but greater than 1%, indicating a rejection of non-causality.

<table>
<thead>
<tr>
<th>Specification</th>
<th>MF Max</th>
<th>MF Wald</th>
<th>LF Max</th>
<th>LF Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specification 1</td>
<td>0.000**</td>
<td>0.322</td>
<td>0.000**</td>
<td>0.017*</td>
</tr>
<tr>
<td>Specification 2</td>
<td>0.001**</td>
<td>0.367</td>
<td>0.002**</td>
<td>0.140</td>
</tr>
</tbody>
</table>
Note: This figure explains a standard notation used in the Mixed Data Sampling (MIDAS) literature. Assume there are only one high frequency variable \( x_H \) and only one low frequency variable \( x_L \). In low frequency period \( \tau_L \), we sequentially observe \( x_H(\tau_L, 1), x_H(\tau_L, 2), \ldots, x_H(\tau_L, m), x_L(\tau_L) \).

Figure 1: Visual Explanation of Mixed Frequency Time Series
Note: In our Monte Carlo simulation for low-to-high causality, we start with a structural MF-VAR(1) data generating process. It can be transformed to a reduced-form MF-VAR(1), and this figure shows how each causal pattern in the structural form is transformed in the reduced form. We assume that $d$, the AR(1) parameter of $x_H$, is fixed at 0.2. The horizontal axis has the first lag through the twelfth lag in the reduced form, and the vertical axis has a coefficient for each of them. The blue, solid line with circles is a reduced-form causal pattern implied by Decaying Causality: $c_j = (-1)^{j-1} \times 0.45/j$. The red, dashed line with squares is a reduced-form causal pattern implied by Lagged Causality: $c_j = 0.4 \times I(j = 12)$. The gray, dotted line with triangles is a reduced-form causal pattern implied by Sporadic Causality: $(c_3, c_7, c_{10}) = (0.4, 0.25, -0.5)$ and all other $c$’s are zeros. It is evident from the figure that each causal pattern in the structural form is basically preserved in the reduced form.

Figure 2: Low-to-High Causal Patterns in Reduced Form
Note: This figure plots weekly 10-year Treasury constant maturity rate, weekly effective federal funds rate, their spread (10Y - FF), and the quarterly real GDP growth from previous year. The blue, solid line represents the 10-year rate. The red, dashed line represents the FF rate. The gray, solid line represents the spread. The yellow, solid line represents GDP. The sample period covers January 5, 1962 through December 31, 2013, which has 2,736 weeks or 208 quarters. The shaded areas represent recession periods defined by the National Bureau of Economic Research (NBER).

Figure 3: Time Series Plot of U.S. Interest Rates and Real GDP Growth
Note: This figure plots p-values with respect to the null hypothesis of high-to-low non-causality (i.e. non-causality from interest rate spread to GDP). Panel (a) considers the mixed frequency max test, Panel (b) considers the mixed frequency Wald test, Panel (c) considers the low frequency max test, and Panel (d) considers the low frequency Wald test. Mixed frequency tests deal with weekly interest rate spread and quarterly GDP growth, while low frequency tests deal with quarterly interest rate spread and quarterly GDP growth. The entire sample period covers January 5, 1962 through December 31, 2013, which has 2,736 weeks or 208 quarters. In this figure, rolling window analysis with 80-quarter window size is conducted. If a p-value plotted is inside the shaded area, then it means that we have a p-value smaller than nominal size 5% and thus there exists significant causality from interest rate spread to GDP in that subsample.

Figure 4: p-values for Causality from Interest Rate Spread to GDP Growth
Note: This figure plots $p$-values with respect to the null hypothesis of low-to-high non-causality (i.e., non-causality from GDP to interest rate spread). Panel (a) considers the mixed frequency max test, Panel (b) considers the mixed frequency Wald test, Panel (c) considers the low frequency max test, and Panel (d) considers the low frequency Wald test. Mixed frequency tests deal with weekly interest rate spread and quarterly GDP growth, while low frequency tests deal with quarterly interest rate spread and quarterly GDP growth. The entire sample period covers January 5, 1962 through December 31, 2013, which has 2,736 weeks or 208 quarters. In this figure, rolling window analysis with 80-quarter window size is conducted. If a $p$-value plotted is inside the shaded area, then it means that we have a $p$-value smaller than nominal size 5% and thus there exists significant causality from GDP to interest rate spread in that subsample.

Figure 5: $p$-values for Causality from GDP to Interest Rate Spread
Technical Appendices

A Double Time Indices

Consider a low frequency variable \( x_L \) and a high frequency variable \( x_H \). The low frequency variable has a single time index \( x_L(\tau_L) \) for \( \tau_L \in \mathbb{Z} \), as in the usual time series literature. The high frequency variable, in contrast, has two time indices \( x_H(\tau_L, j) \) for \( \tau_L \in \mathbb{Z} \) and \( j \in \{1, \ldots, m\} \).

When we derive time series properties of \( x_H \), it is useful to introduce a notational convention that allows the second argument of \( x_H \) to be any integer. For example, it is understood that \( x_H(\tau_L, 0) = x_H(\tau_L - 1, m) \), \( x_H(\tau_L, -1) = x_H(\tau_L - 1, m - 1) \), and \( x_H(\tau_L, m + 1) = x_H(\tau_L + 1, 1) \). In general, we can introduce the following notation without any confusion:

**High Frequency Simplification**

\[
x_H(\tau_L, j) = \begin{cases} x_H(\tau_L - \left\lfloor \frac{j}{m} \right\rfloor, m \left\lfloor \frac{j - 1}{m} \right\rfloor + j) & \text{if } j \leq 0, \\ x_H(\tau_L + \left\lfloor \frac{j}{m} \right\rfloor, j - m \left\lfloor \frac{j - 1}{m} \right\rfloor) & \text{if } j \geq m + 1. \end{cases} \tag{A.1}
\]

\([x]\) is the smallest integer not smaller than \( x \), while \([x]\) is the largest integer not larger than \( x \). We call (A.1) the high frequency simplification in the sense that any integer put in the second argument of \( x_H \) can be transformed to a natural number between 1 and \( m \) by modifying the first argument appropriately. In fact, we can verify that \( m \left\lfloor \frac{j - 1}{m} \right\rfloor + j \in \{1, \ldots, m\} \) when \( j \leq 0 \), and \( j - m \left\lfloor \frac{j - 1}{m} \right\rfloor \in \{1, \ldots, m\} \) when \( j \geq m + 1 \).

Since the high frequency simplification allows both arguments of \( x_H \) to be any integer, we can verify the following relationship.

**Low Frequency Simplification**

\[
x_H(\tau_L - i, j) = x_H(\tau_L, j - im), \quad \forall i, j, \tau_L \in \mathbb{Z}. \tag{A.2}
\]

We call (A.2) the low frequency simplification in the sense that any lag or lead \( i \) put in the first argument of \( x_H \) can be deleted by modifying the second argument appropriately. As a result the second argument may become an integer that is non-positive or larger than \( m \), but such a case is covered by (A.1).

B Proof of Theorem 2.1

Recall that the DGP is written as \( x_L(\tau_L) = \sum_{k=1}^{p} \alpha_k x_L(\tau_L - k) + \sum_{j=1}^{pm} b_j x_H(\tau_L - 1, m + 1 - j) + \epsilon_L(\tau_L) \). In matrix form it is rewritten as \( x_L(\tau_L) = X_L(\tau_L - 1)^\prime \alpha + X_H(\tau_L - 1)^\prime b + \epsilon_L(\tau_L) \), where \( X_L(\tau_L - 1) = [x_L(\tau_L - 1), \ldots, x_L(\tau_L - p)]^\prime \), \( X_H(\tau_L - 1) = [x_H(\tau_L - 1, m + 1 - 1), \ldots, x_H(\tau_L - 1, m + 1 - pm)]^\prime \), \( \alpha = [a_1, \ldots, a_p]^\prime \), and \( b = [b_1, \ldots, b_{pm}]^\prime \).

Parsimonious regression model \( j \) is written as \( x_L(\tau_L) = \sum_{k=1}^{q} \alpha_{k,j} x_L(\tau_L - k) + \beta_j x_H(\tau_L - 1, m + 1 - j) + u_{L,j}(\tau_L) \) for \( j = 1, \ldots, h \). In matrix form it is rewritten as \( x_L(\tau_L) = x_L(\tau_L - 1)^\prime \theta_j + u_{L,j}(\tau_L) \), where \( x_L(\tau_L - 1) = [x_L(\tau_L - 1), \ldots, x_L(\tau_L - q), x_H(\tau_L - 1, m + 1 - j)]^\prime \) and \( \theta_j = [\alpha_{1,j}, \ldots, a_q,j, \beta_j]^\prime \). We collect all parameters across the \( h \) models as \( \theta = [\theta'_1, \ldots, \theta'_h]^\prime \).

Deriving the asymptotic distribution of the max test statistic \( T = \max_{1 \leq j \leq h} \langle \sqrt{T_L w_{T_L,j} \hat{\beta}_j} \rangle^2 \) under \( H_0 : b = 0_{pm \times 1} \) can be done by deriving the asymptotic distribution of \( \sqrt{T_L} \hat{\beta} \) under \( H_0 \), where \( \hat{\beta} = [\hat{\beta}_1, \ldots, \hat{\beta}_h]^\prime \). Working on \( \sqrt{T_L} \hat{\beta} \) directly is rather cumbersome, so we work on \( R \times \sqrt{T_L} (\theta - \theta_0) \), where the selection matrix \( R \) is such
that \( \hat{\beta} = R \hat{\theta} \). Specifically, \( R \) is an \( h \times (q + 1)h \) matrix whose \( (j, (q + 1)j) \) element is 1 for \( j = 1, \ldots, h \) and all others are zeros. Note that \( \theta_0 \), a hypothesized value for the pseudo-true value of \( \theta \), can be arbitrarily chosen as long as \( R \theta_0 = 0_{h \times 1} \). This condition guarantees that \( \sqrt{T L} \hat{\beta} = R \sqrt{T L} (\hat{\theta} - \theta_0) \). The most convenient choice satisfying this condition is \( \theta_0 = \lambda_n \otimes \theta_0 \) with \( \theta_0 = [a_1, \ldots, a_p, 0_{1 \times (q-p+1)}]' \), where \( \lambda_n \) is an \( h \times 1 \) vector of ones. \( \theta_0 \) is a hypothesized value for \( \theta_j \), all parameters in parsimonious regression model \( j \). Although it contains unknown quantities \( a_1, \ldots, a_p \), it does not violate our theory since the last element of \( \theta_0 \) is 0 and hence \( R \theta_0 = 0_{h \times 1} \).

We first derive the asymptotic distribution of \( \sqrt{T L} (\hat{\theta}_j - \theta_0) \) under \( H_0 \). By the construction of \( \theta_0 \), the DGP can be rewritten as \( x_L(\tau_L) = x_j(\tau_L - 1)' \theta_0 + \epsilon_L(\tau_L) \) under \( H_0 \). Using this, we have that

\[
\sqrt{T_L}(\hat{\theta}_j - \theta_0) = \sqrt{T_L} \left[ \sum_{\tau_L=1}^{T_L} \left( \sum_{\tau_L=1}^{T_L} \epsilon_L(\tau_L) \right)^{-1} \sum_{\tau_L=1}^{T_L} \epsilon_L(\tau_L) \right]^{-1} \sum_{\tau_L=1}^{T_L} \epsilon_L(\tau_L) + o_p(1) \tag{B.1}
\]

Using (B.1), we now deduce the asymptotic distribution of \( \sqrt{T L} (\hat{\theta}_j - \theta_0) \). To rely on the Cramér-Wold theorem, we define a \((q + 1)h \times 1\) nonzero vector \( \lambda = [\lambda_1', \ldots, \lambda_h']' \) and consider \( \lambda' \times \sqrt{T_L} (\hat{\theta}_j - \theta_0) \). We have that

\[
\lambda' \times \sqrt{T_L}(\hat{\theta}_j - \theta_0) = \sum_{j=1}^{h} \lambda_j' \times \sqrt{T_L}(\hat{\theta}_j - \theta_0) \\
= \sum_{j=1}^{h} \lambda_j' \left\{ \Gamma_{j,j}^{-1} \frac{1}{\sqrt{T_L}} \sum_{\tau_L=1}^{T_L} x_j(\tau_L - 1) \epsilon_L(\tau_L) \right\} + o_p(1) \\
= \frac{1}{\sqrt{T_L}} \sum_{j=1}^{h} \left\{ \sum_{j=1}^{h} \lambda_j' \Gamma_{j,j}^{-1} x_j(\tau_L - 1) \right\} \epsilon_L(\tau_L) + o_p(1) \tag{B.2}
\]

where the second equality follows from (B.1).

Define \( \Gamma_{j,i} = E[x_j(\tau_L - 1)x_i(\tau_L - 1)'] \), then we have that

\[
E \left[ \epsilon(\tau_L - 1, \lambda)^2 \right] = \sum_{j=1}^{h} \sum_{i=1}^{h} \lambda_j' \Gamma_{j,j}^{-1} \Gamma_{j,i} \lambda_i = \lambda' \Sigma \lambda, \tag{B.3}
\]

where \( \Sigma \) is a \((q + 1)h \times (q + 1)h\) matrix whose \((i, j)\) block is \( \Sigma_{i,j} \). Using (B.3), we apply a central limit theorem to (B.2) in order to obtain that \( \lambda' \times \sqrt{T_L}(\hat{\theta}_j - \theta_0) \overset{d}{\rightarrow} N(0, \lambda' \Sigma \lambda) \). By the Cramér-Wold theorem, we get that \( \sqrt{T_L}(\hat{\theta}_j - \theta_0) \overset{d}{\rightarrow} N(0_{h \times 1}, \sigma_j^2 \Sigma) \). Hence,

\[
\sqrt{T_L} W_{T_L,h} \hat{\beta} = \sqrt{T_L} W_{T_L,h} R(\hat{\theta} - \theta_0) \overset{d}{\rightarrow} N(0_{h \times 1}, \sigma_j^2 W_h R \Sigma R' W_h). \tag{B.4}
\]

63
Recall that the max test statistic is given by $T = \max_{1 \leq j \leq h}(\sqrt{T_L} w_{T_L,j} \hat{\beta}_j)^2$. Hence we have that $T \xrightarrow{d} \max_{1 \leq j \leq h} \mathcal{N}_j^2$, where $\mathcal{N} = [\mathcal{N}_1, \ldots, \mathcal{N}_h]'$ is a vector-valued random variable drawn from $N(0_{h \times 1}, V)$.

**C Proof of Theorem 2.2**

Recall that parsimonious regression model $j$ is written as $x_L(\tau_L) = \sum_{k=1}^{q} \alpha_{k,j} x_L(\tau_L - k) + \beta_j x_H(\tau_L - 1, m + 1 - j) + u_{L,j}(\tau_L)$ for $j = 1, \ldots, h$. In matrix form it is rewritten as $x_L(\tau_L) = x_j(\tau_L - 1)' \theta_j + u_{L,j}(\tau_L)$. The moment condition with respect to OLS is that $E[x_j(\tau_L - 1)u_{L,j}(\tau_L)] = 0_{(q+1) \times 1}$, so the pseudo-true value of $\theta_j$, denoted by $\theta_j^*$, is as follows:

$$\theta_j^* = [E[x_j(\tau_L - 1)x_j(\tau_L - 1)']}^{-1} E[x_j(\tau_L - 1)x_L(\tau_L)].$$

(C.1)

Recall that the DGP in matrix form is given by $x_L(\tau_L) = X_L(\tau_L - 1)' \alpha + X_H(\tau_L - 1)' b + \epsilon_L(\tau_L)$. Substituting this into (C.1), we get

$$\theta_j^* = [E[x_j(\tau_L - 1)x_j(\tau_L - 1)']]^{-1} E[x_j(\tau_L - 1)\{X_L(\tau_L - 1)' \alpha + X_H(\tau_L - 1)' b + \epsilon_L(\tau_L)\}]$$

$$= [E[x_j(\tau_L - 1)x_j(\tau_L - 1)']]^{-1} \{E[x_j(\tau_L - 1)X_L(\tau_L - 1)' \alpha] + E[x_j(\tau_L - 1)X_H(\tau_L - 1)' b]\}.$$  

(C.2)

where the second equality holds from the mds. assumption of $\epsilon_L$. Assumption 2.4 ensures that $q \geq p$ (i.e. the number of autoregressive lags in our model is at least as large as the true lag order $p$), we have that

$$X_L(\tau_L - 1) = [I_p, 0_{p \times (q-p+1)}] x_j(\tau_L - 1)$$

(C.3)

and hence

$$E[x_j(\tau_L - 1)X_L(\tau_L - 1)'] = E[x_j(\tau_L - 1)x_j(\tau_L - 1)'] \begin{bmatrix} I_p \\ 0_{(q-p+1) \times p} \end{bmatrix}.$$ (C.4)

Substituting (C.4) into (C.2), we obtain

$$\theta_j^* \equiv \begin{bmatrix} \alpha_{1,j}^* \\ \vdots \\ \alpha_{p,j}^* \\ \alpha_{p+1,j}^* \\ \vdots \\ \alpha_{q,j}^* \\ \beta_j^* \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_p \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} + [E[x_j(\tau_L - 1)x_j(\tau_L - 1)']]^{-1} E[x_j(\tau_L - 1)X_H(\tau_L - 1)' b].$$

(C.5)

Finally, it is easy to express the pseudo-true value of $\beta = [\beta_1, \ldots, \beta_h]'$, written as $\beta^*$, by observing $\beta^* = R\theta^*$. As in Appendix B, $R$ is an $h \times (q+1)h$ matrix whose $(j, (q+1)j)$ element is 1 for $j = 1, \ldots, h$ and all others are zeros.

**D Proof of Theorem 2.3**

We want to show that $\beta^* = 0_{h \times 1} \Rightarrow b = 0_{pm \times 1}$, assuming that $h \geq pm$. We pick the last row of (C.5). The lower left block of $[E[x_j(\tau_L - 1)x_j(\tau_L - 1)']]^{-1}$ is

$$-n_j^{-1} E \left[ x_H(\tau_L - 1, m + 1 - j)X_L^{(q)}(\tau_L - 1)' \right] \left[ E \left[ X_L^{(q)}(\tau_L - 1)X_L^{(q)}(\tau_L - 1)' \right] \right]^{-1}$$

64
while the lower right block is simply $n_j^{-1}$, where
\[
n_j \equiv E \left[ x_H (\tau_L - 1, m + 1 + j) \right] - E \left[ x_H (\tau_L - 1, m + 1 - j) \right] X_L^T \left( \tau_L - 1 \right)^{-1} E \left[ X_L^{(q)} (\tau_L - 1) x_H (\tau_L - 1, m + 1 - j) \right].
\]
Hence, the last row of $[E[x_j(\tau_L - 1)x_j(\tau_L - 1)^T]]^{-1} \times [E[x_j(\tau_L - 1)X_H(\tau_L - 1)^T]$ appearing in (C.5) is $n_j^{-1}d_j^T$,
where
\[
d_j \equiv E \left[ X_H (\tau_L - 1) x_H (\tau_L - 1, m + 1 - j) \right] - E \left[ X_H (\tau_L - 1) X_L^{(q)} (\tau_L - 1)^T \right] \left[ E \left[ X_L^{(q)} (\tau_L - 1) X_L^{(q)} (\tau_L - 1)^T \right] \right]^{-1} E \left[ X_L^{(q)} (\tau_L - 1) x_H (\tau_L - 1, m + 1 - j) \right].
\]
Having $\beta^* = 0_{h \times 1}$ implies that $n_j^{-1}d_j^T b = 0$ in view of (C.5). Since $n_j$ is a nonzero finite scalar for any $j = 1, \ldots, h$ by the non-singularity of $E[x_j(\tau_L - 1)x_j(\tau_L - 1)^T]$, it has to be the case that $d_j^T b = 0$. Stacking these $h$ equations, we have that $Db = 0_{h \times 1}$ and thus $b'D'Db = 0$, where $D \equiv [d_1, \ldots, d_h]^T$.

To conclude that $b = 0_{pm \times 1}$, it is sufficient to show that $D'D$ is positive definite. Hence it is sufficient to show that $D$ is of full column rank $pm$. Since we are assuming that $h \geq pm$, we only have to show that $D_{pm} \equiv [d_1, \ldots, d_{pm}]^T$, the first $pm$ rows of $D$, is of full column rank $pm$ or equivalently non-singular. Equation (D.1) implies that
\[
D_{pm} = E \left[ X_H (\tau_L - 1) X_H (\tau_L - 1)^T \right] - E \left[ X_H (\tau_L - 1) X_L^{(q)} (\tau_L - 1)^T \right] \left[ E \left[ X_L^{(q)} (\tau_L - 1) X_L^{(q)} (\tau_L - 1)^T \right] \right]^{-1} E \left[ X_L^{(q)} (\tau_L - 1) X_H (\tau_L - 1)^T \right].
\]
Now define
\[
\Delta \equiv E \left[ \begin{bmatrix} X_L^{(q)} (\tau_L - 1) \\ X_H (\tau_L - 1) \end{bmatrix} \begin{bmatrix} X_L^{(q)} (\tau_L - 1) & X_H (\tau_L - 1)^T \end{bmatrix} \right],
\]
which is trivially non-singular by Assumption 2.1 with positive definite error covariance matrix $\Omega$. Evidently, $D_{pm}$ is the Schur complement of $\Delta$ with respect to $E[X_L^{(q)}(\tau_L - 1)X_L^{(q)}(\tau_L - 1)^T]$. Thus, by the classic argument of partitioned matrix inversion, $D_{pm}$ is non-singular as desired.

## E  Proof of Theorem 2.4

Recall that the max test statistic is given by $T = \max_{1 \leq j \leq h} \left( \sqrt{T_L w_{T,L,j}} \hat{\beta}_j \right)^2$. Since $\beta^*_j$ is defined as the pseudo-true value of $\beta_j$, we have that $\hat{\beta}_j \overset{D}{=} \beta^*_j$ by construction. Hence, assuming $w_{j} \equiv \text{plim} w_{T,L,j} > 0$ for all $j = 1, \ldots, h$, we have that $T \overset{p}{\to} \infty \Leftrightarrow \beta^* \neq 0_{h \times 1}$. Given $h \geq pm$, Theorem 2.3 ensures that $b \neq 0_{pm \times 1} \Rightarrow \beta^* \neq 0_{h \times 1}$. Therefore, $T \overset{p}{\to} \infty$ under a general alternative hypothesis $H_1 : b \neq 0_{pm \times 1}$.
F Proof of Theorem 3.1

This proof is identical to the proof for Theorem 2.1 except for that we impose \( H_1^t : b = (1/\sqrt{T_L})\nu \) instead of \( H_0 : b = 0_{p_m \times 1} \). The DGP under \( H_1^t \) is:

\[
x_L(\tau_L) = \sum_{k=1}^{p} a_k x_L(\tau_L - k) + \sum_{j=1}^{p_m} \frac{\nu_j}{\sqrt{T_L}} x_H(\tau_L - 1, m + 1 - j) + \epsilon_L(\tau_L)
\]

\[
= X_L(\tau_L - 1)' a + X_H(\tau_L - 1)' \left( \frac{1}{\sqrt{T_L}} \nu \right) + \epsilon_L(\tau_L)
\]

\[
= x_j(\tau_L - 1)' \left[ I_q \right] a + X_H(\tau_L - 1)' \left( \frac{1}{\sqrt{T_L}} \nu \right) + \epsilon_L(\tau_L)
\]

where the third equality follows from (C.3). Based on this equation, (B.1) should be modified as \( \sqrt{T_L}(\hat{\theta}_j - \theta_0) = \Gamma_{j,j}^{-1} C_j \nu + \Gamma_{j,j}^{-1} (1/\sqrt{T_L}) \sum_{\tau_L=1}^{T_L} (x_j(\tau_L - 1) - a L(\tau_L) + o_p(1), \quad \text{where} \quad \Gamma_{j,j} = E[x_j(\tau_L - 1)x_j(\tau_L - 1)'] \) and \( C_j = E[x_j(\tau_L - 1)X_H(\tau_L - 1)'] \).

Repeating (B.2), we get \( X' \times \sqrt{T_L}(\hat{\theta} - \theta_0) \overset{d}{\rightarrow} N(\lambda', \lambda' \Sigma \lambda) \), where

\[
\begin{bmatrix}
C_1 \\
\vdots \\
C_h
\end{bmatrix} = \begin{bmatrix}
\Gamma_{1,1}^{-1} C_1 \\
\vdots \\
\Gamma_{h,h}^{-1} C_h
\end{bmatrix} \nu.
\]

By the Cramér-Wold theorem, we have that \( \sqrt{T_L}(\hat{\theta} - \theta_0) \overset{d}{\rightarrow} N(u, \sigma^2 \Sigma) \). Now repeat (B.4) to get \( \sqrt{T_L}W_{L,L,h}\hat{\beta} \overset{d}{\rightarrow} N(\mu, V) \), where \( \mu = W_h R u \) and \( V = \sigma^2 L W_h R \Sigma R' W_h \).

Recall that the max test statistic is given by \( T = \max_{1 \leq j \leq k} (\sqrt{T_L}w_T, j)^2 \). Hence we have that \( T \overset{d}{\rightarrow} \max_{1 \leq j \leq h} M_j^2 \), where \( M = [M_1, \ldots, M_h]' \) is a vector-valued random variable drawn from \( N(\mu, V) \).

G Proof of Theorem 3.2

Recall that \( \Upsilon_k \), autocovariance matrix of the mixed frequency vector \( X(\tau_L) \) of order \( k \), is constructed as follows:

\[
\Upsilon_k \equiv E[X(\tau_L)X'(\tau_L - k)]
\]

\[
= \begin{bmatrix}
E[x_H(\tau_L, 1)x_H(\tau_L - k, 1)] & \ldots & E[x_H(\tau_L, 1)x_H(\tau_L - k, m)] & E[x_H(\tau_L, 1)x_L(\tau_L - k)] \\
& \ddots & \vdots & \vdots \\
E[x_H(\tau_L, m)x_H(\tau_L - k, 1)] & \ldots & E[x_H(\tau_L, m)x_H(\tau_L - k, m)] & E[x_H(\tau_L, m)x_L(\tau_L - k)] \\
E[x_L(\tau_L)x_H(\tau_L - k, 1)] & \ldots & E[x_L(\tau_L)x_H(\tau_L - k, m)] & E[x_L(\tau_L)x_L(\tau_L - k)]
\end{bmatrix}
\]

(G.1)

for \( k \geq 0 \), and \( \Upsilon_k = \Upsilon_k' \) for \( k < 0 \). Recall that \( \Upsilon_k \) is already characterized in terms of underlying parameters \( A_1, \ldots, A_p, \Omega \) for any \( k \in \mathbb{Z} \) through the discrete Lyapunov equation and multivariate Yule-Walker equation; see (3.3) and around. We thus take \( \Upsilon_k \) as given here, and characterize \( \Gamma_{j,i} \) and \( C_j \) in terms of \( \Upsilon_k \) for \( j, i \in \{1, \ldots, h\} \).
We begin with $\Gamma_{j,i}$. By the covariance stationarity of $X(\tau_L)$, we have that

$$\Gamma_{j,i} \equiv E[x_j(\tau_L - 1)x_i(\tau_L - 1)'] = E[x_j(\tau_L)x_i(\tau_L)']$$

$$= E\left[\begin{array}{c} x_L(\tau_L) \\ \vdots \\ x_L(\tau_L - (q - 1)) \\ x_H(\tau_L, m + 1 - j) \end{array}\right] \left[\begin{array}{c} x_L(\tau_L) \\ \vdots \\ x_L(\tau_L - (q - 1)) \\ x_H(\tau_L, m + 1 - i) \end{array}\right].$$

(G.2)

The problem here is that indices $j$ and $i$ may be larger than $m$ and thus the second argument of $x_H$ may be smaller than 1. In such a case it is not immediately clear which element of $\Gamma_{j,i}$ is identical to which element of $\Upsilon_k$. To ensure that the second argument of $x_H$ lies in \{1, \ldots, m\}, we use the high frequency simplification (A.1):

$$x_H(\tau_L, m + 1 - j) = x_H(\tau_L - \left[\frac{1 - (m + 1 - j)}{m}\right], m \left[\frac{1 - (m + 1 - j)}{m}\right] + (m + 1 - j))$$

$$= x_H(\tau_L - \frac{j - m}{m}, m \left[\frac{j - m}{m}\right] + m + 1 - j)$$

(G.3)

where the last equality follows simply from the definitions $f(j) = [(j - m)/m]$ and $g(j) = mf(j) + m + 1 - j$. Note that $f(j) \geq 0$ and $g(j) \in \{1, \ldots, m\}$ for any $j$ as desired. Substituting (G.3) into (G.2), $\Gamma_{j,i}$ can be rewritten as follows.

$$\begin{bmatrix}
E[x_L(\tau_L)x_L(\tau_L)] & \cdots & E[x_L(\tau_L)x_L(\tau_L - (q - 1))] & E[x_L(\tau_L - f(i), g(i))x_L(\tau_L)] \\
E[x_L(\tau_L - (q - 1))x_L(\tau_L)] & \cdots & E[x_L(\tau_L - (q - 1))x_L(\tau_L - (q - 1))] & E[x_L(\tau_L - f(i), g(i))x_L(\tau_L - (q - 1))] \\
E[x_H(\tau_L - f(i), g(i))x_L(\tau_L)] & \cdots & E[x_H(\tau_L - f(i), g(i))x_L(\tau_L - (q - 1))] & E[x_H(\tau_L - f(i), g(i))x_L(\tau_L - f(i), g(i))] \\
E[x_H(\tau_L - f(i), g(i))x_H(\tau_L)] & \cdots & E[x_H(\tau_L - f(i), g(i))x_H(\tau_L - (q - 1))] & E[x_H(\tau_L - f(i), g(i))x_H(\tau_L - f(i), g(i))] \\
\end{bmatrix}$$

(G.4)

We now consider which element of $\Gamma_{j,i}$ is identical to which element of $\Upsilon_k$. Take $E[x_H(\tau_L - f(j), g(j))x_L(\tau_L - (q - 1))]$, the $(q + 1, q)$ element of $\Gamma_{j,i}$, as an example. Depending on the magnitude of $j$, we have two cases to consider: $f(j) \geq q - 1$ or $f(j) < q - 1$. For the first case, we have that

$$E[x_H(\tau_L - f(j), g(j))x_L(\tau_L - (q - 1))] = E[x_H(\tau_L - (q - 1) - (f(j) - (q - 1)), g(j))x_L(\tau_L - (q - 1))]$$

$$= E[x_H(\tau_L - f(j) - (q - 1)), g(j))x_L(\tau_L)]$$

$$= \Upsilon_{f(j)-(q-1)}(K, g(j)),$$

where $\Upsilon_{f(j)-(q-1)}(K, g(j))$ means the $(K, g(j))$ element of $\Upsilon_{f(j)-(q-1)}$. The second equality follows from covariance stationarity, while the third equality follows from (G.1). For the second case that $f(j) < q - 1$, we get

$$E[x_H(\tau_L - f(j), g(j))x_L(\tau_L - (q - 1))] = E[x_H(\tau_L - f(j), g(j))x_L(\tau_L - f(j) - (q - 1 - f(j)))$$

$$= E[x_H(\tau_L, g(j))x_L(\tau_L - (q - 1 - f(j)))$$

$$= \Upsilon_{f(j)-(q-1)}(g(j), K)$$

$$= \Upsilon_{f(j)-(q-1)}(K, g(j)).$$

The second and fourth equalities follow from covariance stationarity, while the third equality follows from (G.1). Combining the two cases, we get that $E[x_H(\tau_L - f(j), g(j))x_L(\tau_L - (q - 1))] = \Upsilon_{f(j)-(q-1)}(K, g(j))$ for any
j. Applying the same argument to each element of $\Gamma_{j,i}$ appearing in (G.4), we obtain

$$\Gamma_{j,i} = \begin{bmatrix}
\Upsilon_{1-1}(K, K) & \ldots & \Upsilon_{1-q}(K, K) & \Upsilon_{f(i)}(g(i), K) \\
\vdots & \ddots & \vdots & \vdots \\
\Upsilon_{q-1}(K, K) & \ldots & \Upsilon_{q-q}(K, K) & \Upsilon_{(q-1)f(i)}(g(i), K) \\
\Upsilon_{f(j)}(K, g(j)) & \ldots & \Upsilon_{f(j)-(q-1)}(K, g(j)) & \Upsilon_{f(j)-(f(i))}(g(i), g(j))
\end{bmatrix}.$$  \hspace{1cm} (G.5)

We now discuss $C_j$ for $j \in \{1, \ldots, h\}$. By covariance stationarity, we have that

$$C_j \equiv E[x_j(\tau_L - 1)X_H(\tau_L - 1)] = E[x_j(\tau_L)X_H(\tau_L)'].$$

$$= E\begin{bmatrix}
x_L(\tau_L) \\
\vdots \\
x_L(\tau_L - (q - 1)) \\
x_L(\tau_L, m + 1 - j)
\end{bmatrix} \begin{bmatrix}
x_H(\tau_L, m + 1 - 1) & \ldots & x_H(\tau_L, m + 1 - pm)
\end{bmatrix}.

We need the high frequency simplification (G.3) since the second argument of $x_H$ may go nonpositive. Then we get

$$C_j = \begin{bmatrix}
E[x_L(\tau_L)x_H(\tau_L - f(1), g(1))] & \ldots & E[x_L(\tau_L)x_H(\tau_L - f(pm), g(pm))] \\
\vdots & \ddots & \vdots \\
E[x_L(\tau_L - (q - 1))x_H(\tau_L - f(1), g(1))] & \ldots & E[x_L(\tau_L - (q - 1))x_H(\tau_L - f(pm), g(pm))] \\
E[x_H(\tau_L - f(j), g(j))x_H(\tau_L - f(1), g(1))] & \ldots & E[x_H(\tau_L - f(j), g(j))x_H(\tau_L - f(pm), g(pm))]
\end{bmatrix}.

We now map each element of $C_j$ to an appropriate element of $Y_k$. Consider $E[x_L(\tau_L - (q - 1))x_H(\tau_L - f(pm), g(pm))]$, the $(q, pm)$ element of $C_j$, as an example. In view of (G.4), this quantity is equal to $\Upsilon_{f(pm)-(q-1)}(K, g(pm))$. Applying the same argument to each element of $C_j$, we obtain

$$C_j = \begin{bmatrix}
\Upsilon_{f(1)}(K, g(1)) & \ldots & \Upsilon_{f(pm)}(K, g(pm)) \\
\vdots & \ddots & \vdots \\
\Upsilon_{f(1)-(q-1)}(K, g(1)) & \ldots & \Upsilon_{f(pm)-(q-1)}(K, g(pm)) \\
\Upsilon_{f(j)-(f(1))}(g(1), g(j)) & \ldots & \Upsilon_{f(j)-(f(pm))}(g(pm), g(j))
\end{bmatrix}.$$