Regression-Based Mixed Frequency Granger Causality Tests

Eric Ghysels*  Jonathan B. Hill†  Kaiji Motegi‡

First Draft: October 1, 2013
This Draft: May 12, 2014

Abstract

This paper proposes a new Granger causality test that achieves higher power than existing tests in both local asymptotics and small sample. We postulate multiple parsimonious regression models where the $j$-th model regresses a low frequency variable onto only the $j$-th lag or lead of a high frequency variable. We then formulate what we call max test statistic by picking the largest squared estimator among all parsimonious regression models. The max test based on mixed frequency data is consistent under any type of Granger causality. In local power analysis and Monte Carlo simulations, we show that the max test based on mixed frequency data is more powerful than the standard Wald test based on mixed frequency data, the max test based on low frequency data, and the Wald test based on low frequency data. We apply these four tests to weekly interest rate spread and quarterly GDP growth in the U.S. The max test based on mixed frequency data yield a plausible result that the spread used to cause GDP until about 1980 but the causality has vanished since then.

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 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 since 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). This test often suffers from the adverse effect of temporal aggregation since standard VAR models require to work on a single frequency. 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.

Ghysels, Hill, and Motegi’s (2013) tests have higher power than the conventional low frequency causality tests in large sample, but they suffer from size distortions in small sample with a large ratio of sampling frequencies, $m$. The essential reason for the size distortions is that the dimension of MF-VAR models soars as $m$ increases. It is desired to invent a mixed frequency Granger causality test that performs well even when $m$ is large or sample size is small. Such a contribution would be especially relevant for multivariate macroeconomic time series analysis, where we tend to have a small sample size and Granger causality has been of great interest since the applied work of Sims (1972, 1980) among others.

Based on this motivation, the present paper proposes a Granger causality test that is based on Sims’ (1972) two-sided regression, not on MF-VAR. 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 an estimator for the parameter of the $j$-th lag or lead of $x_H$, then our test statistic is the maximum among $\{\hat{\beta}_{j1}^2, \ldots, \hat{\beta}_{jh}^2\}$ 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.

---

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.
from the null distribution. The max test is thus very easy to implement in practice.

The max test based on mixed frequency data (henceforth "MF max test") is consistent under any type of Granger causality like decaying or lagged causality. The standard Wald test based on mixed frequency data ("MF Wald test"), which is essentially what Ghysels, Hill, and Motegi (2013) proposed, satisfies consistency as well. The max test based on low frequency data ("LF max test") and the Wald test based on low frequency data ("LF Wald test") do not satisfy consistency, which suggests the relevance of mixed frequency approach. We will present an example of Granger causality that cannot be captured by the LF max test or LF Wald test no matter how many leads or lags are included.

In local power analysis, we show that the MF max test and MF Wald test are more robust against some tricky (but realistic) causal patterns than their low frequency counterparts are. There is no clear raking between the MF max test and MF Wald test in terms of local power.

We show via Monte Carlo simulations that the MF max test is clearly more powerful than the MF Wald test in finite sample. The MF max test is thus preferred to any other test when the ratio of sampling frequencies $m$ is large and sample size is small.

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 1980 but the causality has vanished since then.

This paper is structured as follows. Section 2 derives the MF max test and proves its consistency formally. We also prove that the LF max test and LF Wald test do not satisfy consistency. In Section 3 we conduct local power analysis to compare the local asymptotic power of all four tests. In Section 4 we run Monte Carlo simulations to compare the finite sample size and power of these tests. 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 causality chains (see Dufour and Renault (1998) and Dufour, Pelletier, and Renault (2006)).

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. 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.

Example 2.1. 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
The mixed frequency vector \( z \) property of \( x \) complicates our statistical theory substantially. CPI inflation is announced before the GDP, \( \tau_L \), 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, \( \{x_H(\tau_L, 1), x_H(\tau_L, 2), x_H(\tau_L, 3), x_L(\tau_L)\} \) can be thought of as an appropriate ordering.

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:

\[
X(\tau_L) = [x_H(\tau_L, 1), \ldots, x_H(\tau_L, m), x_L(\tau_L)]',
\]

where \( K = m + 1 \) since we are considering a bivariate case with time-independent \( m \). Define the filtration \( \mathcal{F}_t = \sigma(X(t) : t \leq \tau_L) \). 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 \( p \geq 1 \):

\[
\begin{bmatrix}
x_H(\tau_L, 1) \\
\vdots \\
x_L(\tau_L)
\end{bmatrix}
= \sum_{k=1}^{p} \begin{bmatrix}
d_{11,k} & \ldots & d_{1m,k} \\
\vdots & \ddots & \vdots \\
b_{km} & \ldots & b_{(k-1)m+1}
\end{bmatrix}
\begin{bmatrix}
x_H(\tau_L - k, 1) \\
\vdots \\
x_L(\tau_L - k)
\end{bmatrix}
+ \begin{bmatrix}
\epsilon_H(\tau_L, 1) \\
\vdots \\
\epsilon_L(\tau_L)
\end{bmatrix}
\]

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 \( \Omega \equiv E[\epsilon(\tau_L)\epsilon(\tau_L)'] \) 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) in 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., \( x_H(\tau_L - m) \)) on \( x_L(\tau_L) \), \( b_2 \) is the impact of the second most recent past observation of \( x_H \) (i.e., \( x_H(\tau_L - m - 1) \)) 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. $x_H(\tau_L, 2)$), 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 $\det(I_K - \sum_{k=1}^p A_k z^k) = 0$ 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 $\{x_H(\tau_L, j)\}_{j=1}^m$. A simple counter-example is that $X(\tau_L) = \epsilon(\tau_L)$ and a diagonal error covariance matrix $\Omega = [\sigma_{ij}^2]_{ij=1}^K$ with $\sigma_{11}^2 \neq \sigma_{22}^2$. In this case the entire high frequency series $\{x_H(\tau_L, j)\}_{j=1}^m$ 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 investigate Granger causality between $x_H$ and $x_L$. Section 2.1 discusses high-to-low causality (i.e. causality from $x_H$ to $x_L$), while Section 2.2 discusses 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):

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

$$\epsilon_L(\tau_L) \overset{m.d.}{\sim} (0, \sigma^2_L), \quad \sigma^2_L > 0. \tag{2.2}$$

Notice in (2.2) that 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 and is very convenient for analytical work. Specifically, $x_H(\tau_L, 0)$ 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 $X_L(\tau_L - 1) = [x_L(\tau_L - 1), \ldots, x_L(\tau_L - p)]', X_H(\tau_L - 1) = [x_H(\tau_L - 1, m + 1 - 1), \ldots, x_H(\tau_L - 1, m + 1 - pm)]'$, $a = [a_1, \ldots, a_p]'$, and $b = [b_1, \ldots, b_{pm}]'$. Then, (2.2) can be rewritten as

$$x_L(\tau_L) = X_L(\tau_L - 1)'a + X_H(\tau_L - 1)'b + \epsilon_L(\tau_L). \tag{2.3}$$
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 \( F_{\tau_L} = \sigma(X(t) : t \leq \tau_L) \) 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. The question 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

The existing MIDAS literature takes a naïve approach regressing \( x_L \) onto its own low frequency lags and high frequency lags of \( x_H \):

**Mixed Frequency Naïve Regression Model**

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

(2.4)

Consider fitting OLS to (2.4) and then testing \( H_0 : \beta_1 = \cdots = \beta_h = 0 \) by a Wald test. The existing mixed frequency causality test developed by Ghysels, Hill, and Motegi (2013) follows this procedure essentially. Comparing (2.2) and (2.4), this test trivially has power approaching 1 if \( q \geq p \) and \( h \geq pm \) since the model embraces the DGP as a special case. 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 month vs. year case for instance, then the MF-VAR lag order \( p \) is in terms of year and \( m = 12 \). We thus have \( pm = 36 \) when \( p = 3 \), and \( pm = 48 \) when \( p = 4 \), etc. Therefore, including sufficiently many high frequency lags \( h \geq pm \) results in size distortions when \( T_L \) is small and \( m \) is large. The size distortions may be deleted after bootstrapping, but then finite sample power may be 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 purpose of this paper is to resolve this trade-off by combining multiple parsimonious regression models:

**Mixed Frequency Parsimonious Regression Models**

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

(2.5)
In a matrix form, model $j$ is rewritten as

$$x_L(\tau_L) = \begin{bmatrix} X_L^{(q)}(\tau_L - 1)' & x_H(\tau_L - 1, m + 1 - j) \end{bmatrix} \begin{bmatrix} \alpha_{1,j} \\ \vdots \\ \alpha_{q,j} \\ \beta_j \end{bmatrix} + u_{L,j}(\tau_L),$$  \hspace{1cm} (2.6)

where $X_L^{(q)}(\tau_L - 1) = [x_L(\tau_L - 1), \ldots, x_L(\tau_L - q)]'$. 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 is much smaller than the number of parameters in the naïve regression model (2.4), $q + pm$. This feature alleviates size distortions when $m$ is large and $T_L$ is small. 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) since we are mainly interested in the causality part, not autoregressive part.

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

Two important issues related with the parsimonious regression approach (2.5) are (i) how to combine $h$ parsimonious models to test high-to-low causality and (ii) how to choose $h$ relative to the true, potentially large lag order $pm$. We first describe how to combine all $h$ parsimonious models to get a test statistic. Consider fitting OLS for each model (2.5). Since we are assuming that $q \geq p$, each parsimonious model is correctly specified under the null hypothesis of high-to-low non-causality. When there is causality in general, the model is in general misspecified and hence the resulting estimator will be biased due to omitted regressors. We thus need to characterize the pseudo-true value of $\beta = [\beta_1, \ldots, \beta_h]'$, denoted by $\beta^* = [\beta_1^*, \ldots, \beta_h^*]'$, in terms of underlying parameters $a, b, \sigma^2_L$ as well as 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 $\theta$. We construct a selection matrix $R$ such that $\beta = R \theta^*$. 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:

$$R_{h \times (q + 1)h} = \begin{bmatrix} 0 & \ldots & 0 & 1 & \ldots & \ldots & \ldots & 0 & \ldots & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \ldots & 0 & 0 & \ldots & \ldots & \ldots & 0 & \ldots & 0 & 1 \end{bmatrix}.$$  \hspace{1cm} (2.7)

**Theorem 2.1.** Let Assumptions 2.1, 2.2, and 2.4 hold. Then, the pseudo-true value of $\beta$ associated with
OLS is given by

\[
\beta^* = R\theta^*, \quad \theta^* = \left[\theta_1^*, \ldots, \theta_h^* \right]'
\]

\[
\theta_j^* = \begin{bmatrix}
\alpha_{1,j}^* \\
\vdots \\
\alpha_{p,j}^* \\
\vdots \\
\alpha_{q,j}^* \\
\beta_j^*
\end{bmatrix}
= \begin{bmatrix}
\alpha_1 \\
\vdots \\
\alpha_p \\
\vdots \\
0 \\
0
\end{bmatrix} + 
\begin{bmatrix}
E \left[ x_j(\tau_L - 1) x_j(\tau_L - 1) \right]^{-1} \\
\vdots \\
E \left[ x_j(\tau_L - 1) X_H(\tau_L - 1) \right] \\
\vdots \\
E \left[ x_j(\tau_L - 1) X_H(\tau_L - 1) \right] \\
\vdots \\
E \left[ x_j(\tau_L - 1) X_H(\tau_L - 1) \right]
\end{bmatrix} b.
\]

(2.8)

**Proof 2.1.** See Appendix B.

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

Theorem 2.1 provides a useful insight 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 \(b = 0_{pm \times 1}\)). The converse is also true assuming \(h \geq pm\); \(b = 0_{pm \times 1}\) whenever \(\beta^* = 0_{h \times 1}\).

**Theorem 2.2.** Let Assumptions 2.1, 2.2, and 2.4 hold. We have that \(b = 0_{pm \times 1} \Rightarrow \beta^* = 0_{h \times 1}\) regardless of the relative magnitude of \(p\) and \(h\). When \(h \geq pm\), the converse is also true: \(\beta^* = 0_{h \times 1} \Rightarrow b = 0_{pm \times 1}\).

**Proof 2.2.** See Appendix C.

We are interested in the null hypothesis of Granger non-causality, \(H_0: b = 0_{pm \times 1}\), and the first part of Theorem 2.2 implies that \(\beta^* = 0_{h \times 1}\) under \(H_0\). Moreover, the second part of Theorem 2.2 essentially states that \(\beta^* \neq 0_{h \times 1}\) under a general alternative hypothesis \(H_1: b \neq 0_{pm \times 1}\), given \(h \geq pm\). Exploiting these properties, we construct a test statistic for \(H_0\). For each parsimonious regression model (2.5) we get \(\hat{\beta}_j\), the OLS estimator for \(\beta_j\). Define \(\hat{\beta} = [\hat{\beta}_1, \ldots, \hat{\beta}_h]'\). The basic idea of our test, partly inspired by Andrews and Ploberger (1994), is to look at the maximum value among \(\hat{\beta}_1^2, \ldots, \hat{\beta}_h^2\) with a certain weighting scheme.

Let \(\{w_{TL,j} : j = 1, \ldots, h\}\) be a sequence of \(\sigma(X(\tau_L - k) : k \geq 1)\)-measurable \(L_2\)-bounded non-negative scalars with non-random mean-squared-error limits \(\{w_j\}\). As a standardization, we assume that \(\sum_{j=1}^{h} w_{TL,j} = 1\) without loss of generality. We write

\[
W_{TL} = \begin{bmatrix}
w_{TL,1} & \ldots & 0 \\
\vdots & \ddots & \vdots \\
0 & \ldots & w_{TL,h}
\end{bmatrix}
\quad \text{and} \quad
W = \begin{bmatrix}
w_1 & \ldots & 0 \\
\vdots & \ddots & \vdots \\
0 & \ldots & w_h
\end{bmatrix}.
\]

(2.9)
A trivial choice of \( w_{T_L,j} \) is \( 1/h \), non-random equal weights, but we can consider any other weighting structure as well.

We propose a test statistic:

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

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

We call this the *max test statistic* since it takes the maximum of the square of properly scaled individual OLS estimators. We first derive the asymptotic distribution of the mixed frequency max test statistic \( T \) under \( H_0 \).

**Theorem 2.3.** Let Assumptions 2.1, 2.2, 2.3, and 2.4 hold. Under \( H_0 : b = 0_{pm \times 1} \), we have that \( T \overset{d}{\rightarrow} \max_{1 \leq j \leq h} \mathcal{N}^2 \) as \( T_L \rightarrow \infty \). \( \mathcal{N} = [\mathcal{N}_1, \ldots, \mathcal{N}_h] \) is a vector-valued random variable drawn from \( N(0_{h \times 1}, V) \), where

\[
V_{h \times h} = \sigma_L^2 W R \Sigma W', \quad \Sigma_{(q+1)h \times (q+1)h} = \begin{bmatrix} \Sigma_{1,1} & \cdots & \Sigma_{1,h} \\ \vdots & \ddots & \vdots \\ \Sigma_{h,1} & \cdots & \Sigma_{h,h} \end{bmatrix},
\]

\[
\Sigma_{j,i} = \Gamma^{-1}_{j,j} \Gamma_{j,i} \Gamma_{i,i}^{-1}, \quad \Gamma_{j,i} = E [x_j(\tau_L - 1)x_i(\tau_L - 1)', \quad \sigma_L^2 \text{ is the error variance appearing in (2.2), } R \text{ is the selection matrix defined in (2.7), and } W \text{ is the diagonal weighting matrix defined in (2.9).}
\]

**Proof 2.3.** See Appendix D.

Although the max test statistic \( T \) has a non-standard limit distribution under \( H_0 \), a simulated \( p \)-value is easily available via simulation from the null distribution.

**Simulation from Null Distribution** If an estimator \( \hat{V} \) that is consistent for \( V \) under \( H_0 \) is available, then we can simply draw \( R \) samples \( \mathcal{N}^{(1)}, \ldots, \mathcal{N}^{(R)} \) independently from \( N(0_{h \times 1}, \hat{V}) \) and calculate artificial test statistics \( T_r = \max_{1 \leq j \leq h} (\mathcal{N}_j^{(r)})^2 \). Then we can get an asymptotic \( p \)-value approximation

\[
\hat{p} = (1/R) \sum_{r=1}^R I (T_r > T).
\]

It turns out that we can compute a consistent estimator for \( V \) under \( H_0 \) although we cannot without imposing \( H_0 \). Recall the definition of \( V \) in (2.11) to see this point. First, the selection matrix \( R \) is simply given in (2.7). Second, \( W_{T_L} \overset{p}{\rightarrow} W \) by assumption, so \( W \) in (2.11) can be replaced with \( W_{T_L} \). Third, \( \hat{\Gamma}_{j,i} = 1/T_L \sum_{\tau_L=1}^{T_L} x_j(\tau_L - 1)x_i(\tau_L - 1)' \overset{p}{\rightarrow} \Gamma_{j,i} \) under Assumptions 2.1-2.4. Using the continuous mapping theorem and Slutsky’s theorem, consistent estimators \( \hat{\Sigma}_{j,i} \overset{p}{\rightarrow} \Sigma_{j,i} \) and \( \hat{\Sigma} \overset{p}{\rightarrow} \Sigma \) can be obtained directly from the definitions in (2.11). Hence, the availability of consistent estimator \( \hat{V} \) depends entirely
on the availability of consistent estimator $\hat{\sigma}^2_h$. Since the DGP (2.2) reduces to a pure AR($p$) process under $H_0$, $\hat{\sigma}^2_h$ 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 do not impose $H_0$ then we cannot get consistent $\hat{\sigma}^2_h$ 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 is implementable through asymptotic p-value approximation in (2.12).

We now consider the max test statistic $T$ under a general alternative hypothesis $H_1 : b \neq 0_{pm \times 1}$. In view of Theorem 2.2, it is straightforward to see that the max test is consistent as long as $h \geq pm$.

**Theorem 2.4.** Let Assumptions 2.1, 2.2, and 2.4 hold. 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.

Although a formal proof is provided in Appendix E, Theorem 2.4 is intuitively clear from Theorem 2.2. Equation (2.10) indicates that $T \xrightarrow{p} \infty$ if and only if $\beta^* \neq 0_{h \times 1}$. Theorem 2.2 states that, as long as $h \geq pm$, nonzero $b$ implies nonzero $\beta^*$. The mixed frequency max test is therefore consistent.

If one happens to choose $h$ that is smaller than $pm$, there is a certain form of causality such that the power does not approach 1. We provide such an example below.

**Example 2.2.** Consider a very simple DGP with $m = 2$ and $p = 1$:

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

(2.13)

$$
\epsilon(\tau_L) \overset{m.d.}{\sim} \begin{bmatrix}
  1 & \rho & 0 \\
  \rho & 1 & 0 \\
  0 & 0 & 1
\end{bmatrix}, \quad \rho \neq 0, \ |\rho| < 1.
$$

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 $\Gamma_{1,1}$ defined in (2.8). The DGP (2.13) implies that $E[x_H(\tau_L - 1, 1)^2] = E[x_H(\tau_L - 1, 2)^2] = 1$, $E[x_L(\tau_L - 1)^2] = 1/\rho^2$, and

$$
E[x_L(\tau_L - 1)x_H(\tau_L - 2)] = E \left[ \left( -\frac{1}{\rho}x_H(\tau_L - 2, 1) + x_H(\tau_L - 2, 2) + \epsilon_L(\tau_L - 1) \right) x_H(\tau_L - 1, 2) \right] = 0.
$$

Using these results and the definition in (2.6) that $x_1(\tau_L - 1) = [x_L(\tau_L - 1), x_H(\tau_L - 1, 2)]'$, we have that

$$
\Gamma_{1,1} = E \left[ x_1(\tau_L - 1) x_1(\tau_L - 1) \right]
= \begin{bmatrix}
  E[x_L(\tau_L - 1)^2] & E[x_L(\tau_L - 1)x_H(\tau_L - 1, 2)] \\
  E[x_H(\tau_L - 1, 2)x_L(\tau_L - 1)] & E[x_H(\tau_L - 1, 2)^2]
\end{bmatrix}
= \begin{bmatrix}
1/\rho^2 & 0 \\
0 & 1
\end{bmatrix}.
$$

9
Similarly, $\Gamma_{2,2} = \Gamma_{1,1}$ and hence
\[
\Gamma_{1,1}^{-1} = \Gamma_{2,2}^{-1} = \begin{bmatrix} \rho^2 & 0 \\ 0 & 1 \end{bmatrix}.
\]
Recalling $pm = 1 \times 2 = 2$, we have from (2.3) that $X_H(\tau_L - 1) = [x_H(\tau_L - 1, 2), x_H(\tau_L - 1, 1)]'$. We also have that $x_2(\tau_L - 1) = [x_L(\tau_L - 1), x_H(\tau_L - 1, 1)]'$ as defined in (2.6). Thus,
\[
C_1 \equiv E [x_1(\tau_L - 1)X_H(\tau_L - 1)'] = \begin{bmatrix} 0 \\ 0 \\ 1 \\ \rho \end{bmatrix} \quad \text{and} \quad C_2 \equiv E [x_2(\tau_L - 1)X_H(\tau_L - 1)'] = \begin{bmatrix} 0 \\ 0 \\ \rho \\ 1 \end{bmatrix}.
\]
In view of (2.8), we get that
\[
\left[ \begin{array}{c} \alpha_{1,1}^1 \\ \beta_1^1 \end{array} \right] = \left[ \begin{array}{ccc} \rho^2 & 0 & 0 \\ 0 & 1 & \rho \\ 1 & -1/\rho & 0 \\ 0 & 0 & 1 \end{array} \right] \left[ \begin{array}{c} 0 \\ 0 \\ \rho \\ 1 \end{array} \right] = \left[ \begin{array}{c} 0 \\ 0 \\ \rho \\ 1 \end{array} \right]
\quad \text{and} \quad
\left[ \begin{array}{c} \alpha_{1,2}^1 \\ \beta_2^1 \end{array} \right] = \left[ \begin{array}{ccc} \rho^2 & 0 & 0 \\ 0 & 1 & \rho \\ 1 & -1/\rho & 0 \\ 0 & 0 & 1 \end{array} \right] \left[ \begin{array}{c} 0 \\ 0 \\ \rho \\ 1 \end{array} \right] = \left[ \begin{array}{c} 0 \\ 0 \\ \rho \\ 1 \end{array} \right].
\]
Note that $\beta_1^1 = 0$ and $\beta_2^1 = \rho - 1/\rho \neq 0$ since $|\rho| < 1$. Therefore, if we choose $h = 1$, the max test statistic $T$ in (2.10) 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 $\hat{b}_2$, i.e. $w_2 > 0$).

One might argue that the example DGP (2.13) 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 $x_H(\tau_L - 1, 2)$ 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 naïve regression model (2.4) and the max test based on the mixed frequency parsimonious regression models (2.5) are consistent (as long as $h \geq pm$). The key aspect of these two tests is that we 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 get consistency (i.e. there would be a certain form of Granger causality which could not be detected with probability approaching 1), no matter how many low frequency lags of $x_H$ we included.

To verify this point, we formulate a max test based on a low frequency version of parsimonious regression models and a Wald test based on low frequency version of naïve regression model. 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 fixed by the researcher.

Using this notation, we start with low frequency parsimonious regression models and then move on to naïve regression model.

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

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

\[ \equiv \alpha_{L,j}^* \]

\[ \equiv \beta_j^L \]

\[ \equiv \theta_j^L \]

Note that each lag of \( x_H \) is taken in terms of low frequency period. We still impose Assumption 2.4 that \( q \geq p \) so that we can focus on the magnitude of \( h \). Since we are assuming the same DGP (2.1) as in the mixed frequency case, the pseudo-true value for \( \theta_j^L \), denoted as \( (\theta_j^L)_* \), can be easily derived by replacing \( x_j(\tau_L - 1)' \) with \( x_j(\tau_L - 1)' \) in (2.8):

\[ (\theta_j^L)_* = \left[ (\alpha_{1,j}^*)^*, \ldots, (\alpha_{p,j}^*)^*, (\alpha_{p+1,j}^*)^*, \ldots, (\alpha_{q,j}^*)^*, (\beta_j^L)^* \right] \]

\[ \equiv \left[ \begin{array}{c} \alpha_1^* \\ \vdots \\ \alpha_p^* \\ 0 \\ \vdots \\ 0 \end{array} \right] + \left[ E \left[ x_j(\tau_L - 1) x_j(\tau_L - 1)' \right] \right]^{-1} E \left[ x_j(\tau_L - 1) x_H(\tau_L - 1)' \right] \]

\[ \equiv \Sigma_{ij}^{-1}, \quad (q+1) \times (q+1) \]

\[ \equiv \Sigma_j, \quad (q+1) \times pm \]

Refer to Appendix B for more detailed derivation. We formulate a low frequency version of max test in an exactly analogous fashion to (2.10).

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

\[ T_{LF} \equiv \max_{1 \leq j \leq h} (\sqrt{T_{LW} T_{j_{LF}}} \beta_j^L)^2. \]

(2.16)

Deriving the limit distribution of \( T_{LF} \) under \( H_0 \) is omitted here since it will be covered in local power analysis (cfr. Theorem 3.5). Here we rather focus on the limit property of \( T_{LF} \) under \( H_1 \). A crucial point here is that having zero pseudo-true values (i.e. \( (\beta_j^L)^* \equiv [(\beta_1^L)^*, \ldots, (\beta_h^L)^*]' = 0_{h \times 1} \)) does not imply non-causality (i.e. \( b = 0_{pm \times 1} \)), no matter what a linear aggregation scheme and the choice of \( h \) are. In other words, we can create an example where there is causality (i.e. \( b \neq 0_{pm \times 1} \)) but the low frequency max test has no power (i.e. \( (\beta_j^L)^* = 0_{h \times 1} \)) regardless of the aggregation scheme and the choice of \( h \in \mathbb{N} \). Below we present such an example.
Example 2.3. Consider an even simpler DGP than (2.13) with \( m = 2 \) and \( p = 1 \):

\[
\begin{bmatrix}
  x_H(\tau_L, 1) \\
  x_H(\tau_L, 2) \\
  x_L(\tau_L)
\end{bmatrix}
= X(\tau_L)
= A_1
= \begin{bmatrix}
  0 & 0 & 0 \\
  0 & 0 & 0 \\
  b_2 & b_1 & 0
\end{bmatrix}
\begin{bmatrix}
  x_H(\tau_L - 1, 1) \\
  x_H(\tau_L - 1, 2) \\
  x_L(\tau_L - 1)
\end{bmatrix}
+ \begin{bmatrix}
  \epsilon_H(\tau_L, 1) \\
  \epsilon_H(\tau_L, 2) \\
  \epsilon_L(\tau_L)
\end{bmatrix}, \quad \epsilon(\tau_L) \sim \mathcal{N}(0, \mathbf{I}_3), \quad b \neq 0_{2 \times 1}. \quad (2.17)
\]

We want to show that \((\beta_{1j}^{LF})^* = 0\) for any \( j \). To this end, we first do some computation to calculate \( C_j \) in (2.15). First, \( E[x_L(\tau_L - 1)x_H(\tau_L - 1, 2)] = E[x_L(\tau_L - 1)x_H(\tau_L - 1, 1)] = 0 \). Second, assuming a general linear aggregation scheme, \( E[x_H(\tau_L - j)x_H(\tau_L - 1, 2)] = E[(\delta_1 x_H(\tau_L - j, 1) + (1 - \delta_1) x_H(\tau_L - j + 2)) x_H(\tau_L - 1, 2)] = (1 - \delta_1) I(j = 1) \). Similarly, \( E[x_H(\tau_L - j)x_H(\tau_L - 1, 1)] = \delta_1 I(j = 1) \). Recall from (2.15) that \( \mathbf{x}_j(\tau_L - 1) = [x_L(\tau_L - 1), x_H(\tau_L - j)]' \). Recall from (2.3) that \( X_H(\tau_L - 1) = [x_H(\tau_L - 1, 2), x_H(\tau_L - 1, 1)]' \). Thus, we have that

\[
C_j = E[\mathbf{x}_j(\tau_L - 1) X_H(\tau_L - 1)] = \begin{bmatrix}
  E[x_L(\tau_L - 1) x_H(\tau_L - 1, 2)] & E[x_L(\tau_L - 1) x_H(\tau_L - 1, 1)] \\
  E[x_H(\tau_L - j) x_H(\tau_L - 1, 2)] & E[x_H(\tau_L - j) x_H(\tau_L - 1, 1)]
\end{bmatrix}
= \begin{bmatrix}
  0 & 0 \\
  1 - \delta_1 & \delta_1
\end{bmatrix} I(j = 1). \quad (2.18)
\]

Evidently, \( C_j = 0_{2 \times 2} \) for any \( j \geq 2 \), which implies that \( (\beta_{1j}^{LF})^* = 0 \) for any \( j \geq 2 \) in view of (2.15). To find the condition which yields \( C_1 = 0_{2 \times 2} \) and thus \( (\beta_{1j}^{LF})^* = 0 \), we consider three cases depending on \( \delta_1 \).

First, assume \( \delta_1 = 0 \). Then having \( b_1 = 0 \) and any nonzero \( b_2 \) implies that \( C_1 b = 0_{2 \times 1} \). This implies that \( (\beta_{1j}^{LF})^* = 0 \), so the low frequency max test statistic in (2.16) converges to the asymptotic distribution under the null hypothesis of non-causality, resulting in no power. (See Theorem 3.5 for the exact form of the null distribution.)

Second, assume \( \delta_1 \in (0, 1) \). Then having any nonzero \( b_1 \) and \( b_2 = -b_1 (1 - \delta_1) / \delta_1 \) implies that \( C_1 b = 0_{2 \times 1} \).

Third, assume \( \delta_1 = 1 \). Then having any nonzero \( b_1 \) and \( b_2 = 0 \) implies that \( C_1 b = 0_{2 \times 1} \).

Thus, the low frequency max test is inconsistent whatever the aggregation scheme and the choice of \( h \) are. 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 Naïve Regression** We next discuss a low frequency naïve regression model which regresses \( x_L \) onto its own low frequency lags and \( h \) low frequency lags of \( x_H \).
Low Frequency Naïve Regression Model

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

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

Note that \( h \) low frequency lags of \( x_H \) are taken in the model. We still impose Assumption 2.4 that \( q \geq p \)
so that we can focus on the magnitude of \( h \). 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 easily derived:

\[ (\theta^{LF})^* = \begin{bmatrix} (\alpha_1^{LF})^* \\ \vdots \\ (\alpha_p^{LF})^* \\ (\alpha_{p+1}^{LF})^* \\ \vdots \\ (\alpha_q^{LF})^* \\ (\beta^{LF})^* \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_p \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} + \left[ E \left[ x(\tau_L - 1) x(\tau_L - 1)' \right] \right]^{-1} E \left[ x(\tau_L - 1) X_H(\tau_L - 1)' \right] b, \]

\[ \equiv \Gamma^{-1}: (q + h) \times (q + h) \equiv \Gamma: (q + h) \times pm \]

where \( (\beta^{LF})^* = [(\beta_1^{LF})^*, \ldots, (\beta_h^{LF})^*]' \). The derivation of (2.20) is omitted since it is very similar to the proof of Theorem 2.1; see Appendix B.

We explain how to get a Wald statistic with respect to a hypothesis \( \beta^{LF} \equiv [\beta_1^{LF}, \ldots, \beta_h^{LF}]' = 0_{h \times 1} \).
First, fit OLS for model (2.19) and get \( \hat{\beta}^{LF} \). Second, calculate \( \tilde{\Gamma} \equiv (1/T_L) \sum_{\tau_L=1}^{T_L} x(\tau_L - 1) x(\tau_L - 1)' \overset{P}{\rightarrow} \Gamma \). Third, get \( \hat{\sigma}_L^2 \) by fitting pure AR(q) to \( x_L \) and calculating the variance of the residual. It can be shown that \( \hat{\sigma}_L^2 \overset{P}{\rightarrow} \sigma_L^2 \) under non-causality since (2.20) implies that \( b = 0_{pm \times 1} \Rightarrow (\beta^{LF})^* = 0_{h \times 1} \).
Finally, compute the Wald statistic

\[ W_{LF} = T_L \left( \hat{\beta}^{LF} \right)' \left( \hat{\sigma}_L^2 \tilde{R} \tilde{R}' \right)^{-1} \hat{\beta}^{LF}, \]

where \( \tilde{R} = [0_{h \times q}, I_h] \). It is evident that \( W_{LF} \overset{d}{\rightarrow} \chi^2_h \). Consistency requires \( W_{LF} \overset{P}{\rightarrow} \infty \) whenever \( b \neq 0_{pm \times 1} \). In view of (2.21), consistency requires that \( b \neq 0_{pm \times 1} \Rightarrow (\beta^{LF})^* \neq 0_{h \times 1} \). We present a counter-example where there is Granger causality (i.e. \( b \neq 0_{pm \times 1} \)) but the low frequency Wald test has no power (i.e. \( (\beta^{LF})^* = 0_{h \times 1} \)) regardless of linear aggregation scheme and the choice of \( h \in \mathbb{N} \).
Example 2.4. Let us simply continue Example 2.3. Using (2.18), we have that

\[
C \equiv E[x(\tau_L - 1)X_H(\tau_L - 1)']
\]

\[
\begin{bmatrix}
E[x(\tau_L - 1)x_H(\tau_L - 1)] & E[x(\tau_L - 1)x_H(\tau_L - 1, 2)] \\
E[x_H(\tau_L - 1)x_H(\tau_L - 1, 2)] & E[x_H(\tau_L - 1)x_H(\tau_L - 1, 1)] \\
E[x_H(\tau_L - 2)x_H(\tau_L - 1, 2)] & E[x_H(\tau_L - 2)x_H(\tau_L - 1, 1)] \\
\vdots & \vdots \\
E[x_H(\tau_L - h)x_H(\tau_L - 1, 2)] & E[x_H(\tau_L - h)x_H(\tau_L - 1, 1)]
\end{bmatrix}
= \begin{bmatrix}
0 & 0 \\
0 & 1 - \delta_1 & \delta_1 \\
0 & 0 & \ddots & \ddots \\
0 & 0 & \cdots & \cdots & 0 & 0
\end{bmatrix}.
\]

We can make \( Cb = 0 \) by changing \( b \) according to \( \delta_1 \) exactly like Example 2.3. 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 case, we have that \( Cb = 0 \) and \( (\beta^{LF})^* = 0 \) in view of (2.20).

The low frequency Wald test statistic \( W_{LF} \) therefore converges to the asymptotic distribution under the null hypothesis of non-causality, \( \chi^2_{h} \), which results in no power.

In summary, we introduced four different high-to-low Granger causality tests associated with the DGP (2.1). 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 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 \( pm \). In contrast, the 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.

2.2 Low-to-High Granger Causality

Consider the same MF-VAR(\( p \)) data generating process (2.1) as before. A possible way of testing 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.

Naïve Regression Model

\[
x_L(\tau_L) = \sum_{k=1}^{p} \alpha_k x_L(\tau_L - k) + \sum_{j=1}^{pm} \beta_j x_H(\tau_L - 1, m + 1 - j) + \sum_{j=1}^{h} \gamma_j x_H(\tau_L + 1, j) + u_L(\tau_L), \tag{2.22}
\]

Instruments: \{all \( p + pm + h \) regressors, \( x_H(\tau_L, 1), \ldots, x_H(\tau_L, m) \}\).

Here we are assuming that the true MF-VAR lag order \( p \) is known.\(^2\) Besides all explanatory variables, we include \( m \) contemporaneous high frequency observations of \( x_H \) in the set of instruments in order to

\(^2\)This assumption should be relaxed in the future.
handle simultaneity between $x_L$ and $x_H$. We test the significance of $\gamma_1, \ldots, \gamma_h$, the parameters on the high frequency leads of $x_H$. Under the null hypothesis of low-to-high non-causality, all those parameters should be equal to zero.

A potential problem of this approach is that there may be parameter proliferation as in the naïve regression model for high-to-low causality. We thus propose more parsimonious models:

\[
\text{Parsimonious Regression Model } j \in \{1, \ldots, h\}
\]

\[
x_L(\tau_L) = \sum_{k=1}^{p} \alpha_{k,j} x_L(\tau_L - k) + \sum_{k=1}^{pm} \beta_{k,j} x_H(\tau_L - 1, m + 1 - k) + \gamma_j x_H(\tau_L + 1, j) + u_{L,j}(\tau_L),
\]

(2.23)

Instruments: \{ all $p + pm + 1$ regressors in model $j, x_H(\tau_L, 1), \ldots, x_H(\tau_L, m) \}.

We are combining $h$ parsimonious regression models, and the $j$-th model contains the $j$-th high frequency lead of $x_H$. As in the naïve regression model (2.22), we include $m$ contemporaneous high frequency observations of $x_H$ as instruments.

A key insight is that the pseudo-true values of $\gamma_1, \ldots, \gamma_h$ are all zeros under the null hypothesis that $x_L$ does not Granger cause $x_H$. Using this property, our test strategy is to get the generalized instrumental variable estimator (GIVE) for $\gamma_j$ and formulate the max test statistic:

\[
T = \max_{1 \leq j \leq h} \left( \sqrt{T_L w_{T_L,j} \hat{\gamma}_j} \right)^2, \tag{2.24}
\]

where $w_{T_L} = [w_{T_L,1}, \ldots, w_{T_L,h}]'$ is a weighting scheme such that $w_{T_L} \overset{L^2}{\to} w$. We will derive the asymptotic null distribution of $T$ under the null hypothesis of non-causality $H_0: x_L \not\rightarrow x_H$.

**Theorem 2.5.** Under $H_0 : x_L \not\rightarrow x_H$, it follows that $T \equiv \max_{1 \leq j \leq h} \left( \sqrt{T_L w_{T_L,j} \hat{\gamma}_j} \right)^2 \overset{d}{\to} \max_{1 \leq j \leq h} \mathcal{N}_j^2$, where $\mathcal{N} = [\mathcal{N}_1, \ldots, \mathcal{N}_h]' \sim N(0_{h \times 1}, U)$.

**Proof 2.5.** See Appendix F. The covariance matrix $U$ is derived there.

A consistent estimator for the covariance matrix $U$ can be constructed from sample moments in an analogous fashion with high-to-low causality, so the testing procedure is not described in detail here.\(^3\)

### 3 Local Asymptotic Power Analysis

The goal of this section is to compare the local asymptotic power of the four different high-to-low Granger causality tests listed in Table 1: MF max test, MF Wald test, LF max test, and LF Wald test.\(^4\) 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.

\(^3\)Consistency of the low-to-high max test should be discussed in the future.

\(^4\)The low-to-high case remains as a future task.
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.

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. \( H^0 : b = 0_{pm \times 1} \)), but here we consider a local alternative hypothesis \( H^l_1 : b = (1/p^T_L) \nu \). In the hypothesis testing literature \( \nu = [\nu_1, \ldots, \nu_{pm}]' \) is often called the Pitman drift. Under \( H^l_1 \), the DGP is written as

\[
X(\tau_L) = \sum_{k=1}^{p} \begin{bmatrix}
  d_{11,k} & \cdots & d_{1m,k} & e_{(k-1)m+1} \\
  \vdots & \ddots & \vdots & \vdots \\
  d_{m1,k} & \cdots & d_{mm,k} & e_{km} \\
\nu_{km}/\sqrt{T_L} & \cdots & \nu_{(k-1)m+1}/\sqrt{T_L} & a_k
\end{bmatrix}
X(\tau_L - k) + \epsilon(\tau_L).
\]

Picking the last row, we have that

\[
x_L(\tau_L) = \sum_{k=1}^{p} a_k x_L(\tau_L - k) + \sum_{j=1}^{pm} \nu_j \epsilon_H(\tau_L - 1, m + 1 - j) + \epsilon_L(\tau_L)
\]

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

Based on this DGP, we derive the limit distribution of each test statistic under \( H^l_1 \).

### 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 max test statistic \( T \) defined in (2.10). The asymptotic distribution under \( H^0_0 : b = 0_{pm \times 1} \) is already derived in Theorem 2.3. Here we derive the asymptotic distribution under \( H^l_1 : b = (1/\sqrt{T_L}) \nu \).

**Theorem 3.1.** Let Assumptions 2.1, 2.2, 2.3, and 2.4 hold. Then, we have that \( T \xrightarrow{d} \max_{1 \leq i \leq h} M_i^2 \) as \( T_L \to \infty \) under \( H^l_1 : b = (1/\sqrt{T_L}) \nu \). \( \mathcal{M} = [\mathcal{M}_1, \ldots, \mathcal{M}_h]' \) is a vector-valued random variable drawn from \( N(\mu, V) \). \( V \) is defined in (2.11) and

\[
\mu_{h \times 1} = WR \begin{bmatrix}
\Gamma_{11}^{-1} C_1 \\
\vdots \\
\Gamma_{h,h}^{-1} C_h
\end{bmatrix} \nu,
\]

16
where $\Gamma_{j,j} = E[x_j(\tau_L - 1)x_j(\tau_L - 1)']$ and $C_j = E[x_j(\tau_L - 1)X_H(\tau_L - 1)']$ as defined in (2.8).

**Proof 3.1.** See Appendix G.

Comparing Theorems 2.3 and 3.1, the asymptotic distribution under $H_0$ and the asymptotic distribution under $H_1^+$ share the same covariance matrix $V$ but only the latter has a nonzero location parameter $\mu$. We see that $V$ and $\mu$ depend on two population covariance terms: $\Gamma_{j,i}$ and $C_j$ for $j, i \in \{1, \ldots, h\}$. Recall from (2.11) that $\Gamma_{j,i} = E[x_j(\tau_L - 1)x_i(\tau_L - 1)']$ 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 (2.8) that $C_j = E[x_j(\tau_L - 1)X_H(\tau_L - 1)']$ is the $(q + 1) \times pm$ covariance matrix between all regressors in parsimonious regression model $j$ and $pm$ high frequency lags of $x_H$.

An analytical computation of local asymptotic power therefore requires an explicit characterization of $\Gamma_{j,i}$ 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:

\[
\begin{align*}
\Upsilon_k = E[X(\tau_L)X(\tau_L - k)'],
\Upsilon_p = \begin{bmatrix}
\Upsilon_{11} & \cdots & \Upsilon_{1p - 1} \\
\vdots & \ddots & \vdots \\
\Upsilon_{1p} & \cdots & \Upsilon_{p-p}
\end{bmatrix},
\end{align*}
\]

\[
A = \begin{bmatrix}
A_1 & \cdots & A_{p - 1} & A_p \\
I_K & \cdots & 0_{K \times K} & 0_{K \times K} \\
\vdots & \ddots & \vdots & \vdots \\
0_{K \times K} & \cdots & I_K & 0_{K \times K}
\end{bmatrix},
\]

\[
\Omega = \begin{bmatrix}
\Omega & 0_{K \times K} & \cdots & 0_{K \times K} \\
0_{K \times K} & \Omega & \cdots & 0_{K \times K} \\
\vdots & \ddots & \ddots & \vdots \\
0_{K \times K} & 0_{K \times K} & \cdots & \Omega
\end{bmatrix}.
\]

Based on these notations and the discrete Lyapunov equation, $\Upsilon_k$ for $k \in \{1 - p, \ldots, p - 1\}$ can be computed by the following well-known formula:

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

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

Now that we have characterized $\{\Upsilon_k\}_{k \in \mathbb{Z}}$ in terms of $A_1, \ldots, A_p, \Omega$, the next step is to characterize $\Gamma_{j,i}$ and $C_j$ in terms of $\{\Upsilon_k\}_{k \in \mathbb{Z}}$. 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 in Theorem 3.2 and put detailed derivation in Appendix H.

**Theorem 3.2.** Let Assumptions 2.1 and 2.2 hold. Define $f(j) = \lceil (j - m)/m \rceil$ and $g(j) = mf(j) + m + 1 - j$, where $\lceil x \rceil$ is the smallest integer not smaller than $x$. Let $\Upsilon_k(s, t)$ be the $(s, t)$ element of $\Upsilon_k$. 

17
Then,

\[
\Gamma_{j,i}^{(q+1)\times(q+1)} = \begin{bmatrix}
\Upsilon_{1-1}(K,K) & \ldots & \Upsilon_{1-q}(K,K) & \Upsilon_{-f(i)}(g(i),K) \\
\vdots & \ddots & \vdots & \vdots \\
\Upsilon_{q-1}(K,K) & \ldots & \Upsilon_{q-q}(K,K) & \Upsilon_{(q-1)-f(i)}(g(i),K) \\
\Upsilon_{f(j)}(K,g(j)) & \ldots & \Upsilon_{f((q-1))}(g(j)) & \Upsilon_{f-j-f(i)}(g(i),g(j))
\end{bmatrix},
\]

(3.4)

and

\[
C_{j}^{(q+1)\times pm} = \begin{bmatrix}
\Upsilon_{f(j)}(K,g(1)) & \ldots & \Upsilon_{f(pm)}(K,g(pm)) \\
\vdots & \ddots & \vdots \\
\Upsilon_{f(j)-(q-1)}(g(1),g(j)) & \ldots & \Upsilon_{f(j)-(pm)}(g(pm),g(j))
\end{bmatrix}
\]

for \(j, i \in \{1, \ldots, h\}\).

**Proof 3.2.** See Appendix H.

Theorem 3.2 holds for an arbitrary choice of \(h \in \mathbb{N}\). In fact, we see by construction that \(g(j) \in \{1, \ldots, m\}\) for any \(j \in \{1, \ldots, h\}\). Hence, (3.4) is well-defined for any choice of \(h\).

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 2.3, 3.1, and 3.2 to calculate \(\mu\) and \(V\).

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

**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 \(d^*\).

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

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

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

\[
= [x_L(\tau_L - 1), \ldots, x_L(\tau_L - q)] \begin{bmatrix} \alpha \end{bmatrix}^{\alpha} + [x_H(\tau_L - 1, m + 1 - 1), \ldots, x_H(\tau_L - 1, m + 1 - h)] \begin{bmatrix} \beta \end{bmatrix}^{\beta} + u_L(\tau_L)
\]

\[
= \begin{bmatrix} X_L^{(q)}(\tau_L - 1)^\prime & X_H^{(b)}(\tau_L - 1)^\prime \end{bmatrix} \begin{bmatrix} \alpha \\beta \end{bmatrix}^{\alpha\beta} + u_L(\tau_L).
\]

18
The true DGP for $x_L$ is the same as before and can be found in (2.2). We keep imposing Assumption 2.4 so that the autoregressive lag order $q$ is larger than or equal to the true lag order $p$. 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 $pm$. Clearly, the Wald test with respect to $H_0 : b = 0_{pm \times 1}$ is consistent if $h \geq pm$. If $h < pm$, then the Wald test may not have power approaching 1 for some type of Granger causality.

We want to formulate a Wald statistic $W$ with respect to $H_0$ and find its asymptotic distribution under $H_0$ as well as $H_1: b = 1/\sqrt{T_L}\nu$. First, the OLS estimator associated with model (3.5) is given by

$$\hat{\theta} = \left[\sum_{\tau_L=1}^{T_L} x(\tau_L-1)x(\tau_L-1)' \right]^{-1} \left[\sum_{\tau_L=1}^{T_L} x(\tau_L-1)x_L(\tau_L)\right].$$

Second, construct an $h \times (q+h)$ selection matrix $R$ such that $\beta = R\theta$. Specifically, we let $R = [0_{h \times q} I_h]$. Third, define two population covariance terms:

$$\Gamma \equiv \begin{bmatrix} \Gamma_{UL} & \Gamma_{UR} \\ \Gamma'_{UR} & \Gamma_{LR} \end{bmatrix}$$

$$C \equiv \begin{bmatrix} C_{11} & \cdots & C_{1m} \\ \vdots & \ddots & \vdots \\ C_{m1} & \cdots & C_{mm} \end{bmatrix}$$

where $x(\tau_L-1)$ and $X_H(\tau_L-1)$ are defined in (3.5) and (2.3), respectively.

Using these quantities, we can establish the asymptotic distribution of $W$ under $H_1$.

**Theorem 3.3.** Let Assumptions 2.1, 2.2, 2.3, and 2.4 hold. Let $W$ be the Wald statistic with respect to $H_0$ and find its asymptotic distribution under $H_0$ as well as $H_1: b = 1/\sqrt{T_L}\nu$. First, the OLS estimator associated with model (3.5) is given by

$$\hat{\theta} = \left[\sum_{\tau_L=1}^{T_L} x(\tau_L-1)x(\tau_L-1)' \right]^{-1} \left[\sum_{\tau_L=1}^{T_L} x(\tau_L-1)x_L(\tau_L)\right].$$

Second, construct an $h \times (q+h)$ selection matrix $R$ such that $\beta = R\theta$. Specifically, we let $R = [0_{h \times q} I_h]$. Third, define two population covariance terms:

$$\Gamma \equiv \begin{bmatrix} \Gamma_{UL} & \Gamma_{UR} \\ \Gamma'_{UR} & \Gamma_{LR} \end{bmatrix}$$

$$C \equiv \begin{bmatrix} C_{11} & \cdots & C_{1m} \\ \vdots & \ddots & \vdots \\ C_{m1} & \cdots & C_{mm} \end{bmatrix}$$

where $x(\tau_L-1)$ and $X_H(\tau_L-1)$ are defined in (3.5) and (2.3), respectively.

Using these quantities, we can establish the asymptotic distribution of $W$ under $H_1$.

**Proof 3.3.** See Appendix I.

As seen in (3.7), the noncentrality $\kappa$ depends on $\Gamma$ and $C$ defined in (3.6). We thus need to characterize these two covariances in terms of $\Upsilon_k$, the autocovariance of the mixed frequency vector $X(\tau_L)$, as we did for $\Gamma_{j,i}$ and $C_{j}$ in Theorem 3.2.

**Theorem 3.4.** Let Assumptions 2.1 and 2.2 hold. Let $f(j) = [(j-m)/m]$ and $g(j) = m f(j) + m + 1 - j$ as in Theorem 3.2. Let $\Upsilon_k(s,t)$ be the $(s,t)$ element of $\Upsilon_k$. Then,

$$\Gamma \equiv \begin{bmatrix} \Gamma_{UL} & \Gamma_{UR} \\ \Gamma'_{UR} & \Gamma_{LR} \end{bmatrix}$$
with
\[ \Gamma_{UL} = \begin{bmatrix} \Upsilon_{1-1}(K, K) & \cdots & \Upsilon_{1-q}(K, K) \\ \vdots & \ddots & \vdots \\ \Upsilon_{q-1}(K, K) & \cdots & \Upsilon_{q-q}(K, K) \end{bmatrix}, \quad \Gamma_{UR} = \begin{bmatrix} \Upsilon_{-f(1)}(g(1), K) & \cdots & \Upsilon_{-f(h)}(g(h), K) \\ \vdots & \ddots & \vdots \\ \Upsilon_{(q-1)-f(1)}(g(1), K) & \cdots & \Upsilon_{(q-1)-f(h)}(g(h), K) \end{bmatrix}, \]
\[ \Gamma_{LR} = \begin{bmatrix} \Upsilon_{f(1)-f(1)}(g(1), g(1)) & \cdots & \Upsilon_{f(1)-f(h)}(g(h), g(1)) \\ \vdots & \ddots & \vdots \\ \Upsilon_{f(h)-f(1)}(g(1), g(h)) & \cdots & \Upsilon_{f(h)-f(h)}(g(h), g(h)) \end{bmatrix} \]
and
\[ C = \begin{bmatrix} \Upsilon_{f(1)}(K, g(1)) & \cdots & \Upsilon_{f(pm)}(K, g(pm)) \\ \vdots & \ddots & \vdots \\ \Upsilon_{f(1)-(q-1)}(K, g(1)) & \cdots & \Upsilon_{f(pm)-(q-1)}(K, g(pm)) \\ \Upsilon_{f(1)-f(1)}(g(1), g(1)) & \cdots & \Upsilon_{f(1)-f(pm)}(g(pm), g(1)) \\ \vdots & \ddots & \vdots \\ \Upsilon_{f(h)-f(1)}(g(1), g(h)) & \cdots & \Upsilon_{f(h)-f(pm)}(g(pm), g(h)) \end{bmatrix}. \]

**Proof 3.4.** See Appendix J.

We are now ready to compute the local asymptotic power associated with the mixed frequency Wald test. Unlike the max test, the asymptotic distributions under \( H_0 \) and \( H_1 \) are well-known and thus we can do exact computation. The procedure is as follows.

**Step 1** Starting with underlying parameters \( A_1, \ldots, A_p, \Omega \), use Theorems 3.3 and 3.4 to calculate non-centrality \( \kappa \).

**Step 2** Calculate local asymptotic power \( P \) according to \( P = 1 - F_1[F_0^{-1}(1 - \alpha)] \), 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. \( \chi^2_h \)), while \( F_1 \) is the c.d.f. of \( W \) under \( H_1 \) (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.

**Low Frequency Parsimonious Regression** We combine \( h \) low frequency parsimonious regression models (2.14) in order to formulate the max test statistic \( T_{LF} \) defined in (2.16). Here we derive the asymptotic distribution of \( T_{LF} \) under \( H_1^1 : b = (1/\sqrt{T_L})\nu \). The asymptotic distribution under \( H_0 : b = 0_{pm \times 1} \) can be derived simply by setting \( \nu = 0_{pm \times 1} \).

**Theorem 3.5.** Let Assumptions 2.1, 2.2, 2.3, and 2.4 hold. Then, we have that \( T_{LF} \xrightarrow{d} \max_{1 \leq i \leq h} \mathcal{M}_i^2 \) as \( T_L \to \infty \) under \( H_1^1 : b = (1/\sqrt{T_L})\nu \). \( \mathcal{M} = [\mathcal{M}_1, \ldots, \mathcal{M}_h]' \) is a vector-valued random variable.
drawn from $N(\mu, \Sigma)$, where

$$\begin{bmatrix} \mu \\ \Sigma_{h \times 1} \end{bmatrix} = WR \begin{bmatrix} \Gamma^{-1} \Sigma_{i,i} \\ \vdots \\ \Gamma^{-1} \Sigma_{h,h} \end{bmatrix} \nu, \quad \begin{bmatrix} \Sigma_{h \times h} \end{bmatrix} = \sigma^2 W R \Sigma R W,$$

$$\Sigma = \begin{bmatrix} \Sigma_{1,1} & \cdots & \Sigma_{1,h} \\ \vdots & \ddots & \vdots \\ \Sigma_{h,1} & \cdots & \Sigma_{h,h} \end{bmatrix}, \quad \Sigma_{j,i} = \Sigma_{j,i}^{-1}, \quad \Sigma_{j,i} = E[(\mathbf{z}_j - \mathbf{1})^\prime (\mathbf{z}_j - \mathbf{1})],$$

where $\Sigma_{j,i} = E[(\mathbf{z}_j - \mathbf{1})^\prime \mathbf{X}_H (\mathbf{z}_j - \mathbf{1})]$ as defined in (2.15).

**Proof 3.5.** See Appendix K.

We see in Theorem 3.5 that $\mu$ and $\Sigma$ depend on two population covariance terms: $\Sigma_{j,i}$ and $\Sigma_{j}$ for $j, i \in \{1, \ldots, h\}$. Recall from (3.8) that $\Sigma_{j,i} = E[\mathbf{z}_j (\tau_L - 1)^\prime \mathbf{z}_i (\tau_L - 1)]$ is the covariance matrix between all regressors in low frequency parsimonious regression model $j$ and all regressors in model $i$. $\Sigma_{j} = E[\mathbf{z}_j (\tau_L - 1)^\prime \mathbf{X}_H (\tau_L - 1)^\prime]$ is the covariance between all regressors in low frequency parsimonious regression model $j$ and $pm$ high frequency lags of $x_H$.

An analytical computation of local asymptotic power therefore requires an explicit characterization of $\Sigma_{j,i}$ and $\Sigma_{j}$ in terms of $\{\mathbf{Y}_k\}$ elaborated in (3.3) and around.

**Theorem 3.6.** Let Assumptions 2.1 and 2.2 hold. As in Theorem 3.2, suppose that $f(j) = [(j - m)/m]$, $g(j) = mf(j) + m + 1 - j$, and $\mathbf{Y}_k(s, t)$ is the $(s, t)$ element of $\mathbf{Y}_k$. Let $\{\delta_1, \ldots, \delta_m\}$ represent the linear aggregation scheme: $x_H(\tau_L) = \sum_{l=1}^{m} \delta_l x_H(\tau_L, j)$ with $\delta_l \geq 0$ and $\sum_{l=1}^{m} \delta_l = 1$. Then,

$$\Sigma_{j,i} = \begin{bmatrix} \mathbf{Y}_{1-1}(K, K) & \cdots & \mathbf{Y}_{1-q}(K, K) & \sum_{l=1}^{m} \delta_l \mathbf{Y}_{l-1}(K, l) \\ \vdots & \ddots & \vdots \\ \mathbf{Y}_{q-1}(K, K) & \cdots & \mathbf{Y}_{q-q}(K, K) & \sum_{l=1}^{m} \delta_l \mathbf{Y}_{l-q}(K, l) \\ \sum_{l=1}^{m} \delta_l \mathbf{Y}_{j-1}(K, l) & \cdots & \sum_{l=1}^{m} \delta_l \mathbf{Y}_{j-q}(K, l) & \sum_{l=1}^{m} \delta_l \mathbf{Y}_{j-1}(l, l) \end{bmatrix},$$

for $j, i \in \{1, \ldots, h\}$.

**Proof 3.6.** See Appendix L.

The 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 the local asymptotic power associated with the low frequency naïve regression model (2.19). The true DGP is the same as before and can be found in (2.2). We keep imposing Assumption 2.4 so that the autoregressive lag order $q$ is larger than or equal to the
true lag order $p$. 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 $pm$.

We want to formulate a Wald statistic $W_{LF}$ with respect to $H_0 : b = 0_{pm \times 1}$ and find its asymptotic distribution under the null of non-causality (i.e. $H_0 : b = 0_{pm \times 1}$) as well as local alternative (i.e. $H_1 : b = (1/\sqrt{T_L})\nu$). First, the OLS estimator associated with model (2.19) is given by $\hat{\theta} = [\sum_{\tau_L=1}^{T_L} \mathbf{x}(\tau_L - 1)]^{-1} \sum_{\tau_L=1}^{T_L} \mathbf{x}(\tau_L - 1)x_L(\tau_L)$. Second, construct an $h \times (q + h)$ selection matrix $R$ such that $\beta = R\theta$. Specifically, we let $R = [0_{h \times q} \mathbf{I}_h]$. Third, define two population covariance terms:

$$
\Gamma = E[\mathbf{x}(\tau_L - 1)\mathbf{x}(\tau_L - 1)^\prime] \quad \text{and} \quad \Sigma = E[\mathbf{x}(\tau_L - 1)\mathbf{X}(\tau_L - 1)^\prime].
$$

(3.10)

Using these quantities, we can establish the asymptotic distribution of $W_{LF}$ under $H_0$ and $H_1$ below.

**Theorem 3.7.** Let Assumptions 2.1, 2.2, 2.3, and 2.4 hold. Let $W_{LF}$ be the Wald statistic with respect to $H_0 : b = 0_{pm \times 1}$. Then, the asymptotic distribution of $W_{LF}$ is $\chi^2_\nu$ under $H_0$ and $\chi^2_\nu(\kappa)$ under $H_1$. $\chi^2_\nu(\kappa)$ is the noncentral chi-squared distribution with degrees of freedom $\nu$ and noncentrality:

$$
\kappa = \frac{1}{\sigma_L^2} \mathbf{C} \mathbf{G}^{-1} \mathbf{R}^\prime (\mathbf{R}\mathbf{G}^{-1}\mathbf{R}^\prime)^{-1} \mathbf{R}\mathbf{G}^{-1}\mathbf{C} \mathbf{G}.
$$

(3.11)

**Proof 3.7.** See Appendix M.

As seen in (3.11), the noncentrality $\kappa$ depends on $\mathbf{G}$ and $\mathbf{C}$ defined in (3.10). We thus need to characterize these two covariances in terms of $\mathbf{Y}_k$, the autocovariance of the mixed frequency vector $\mathbf{X}(\tau_L)$, as we did for $\mathbf{G}$ and $\mathbf{C}$ in Theorem 3.4.

**Theorem 3.8.** Let Assumptions 2.1 and 2.2 hold. As in Theorem 3.2, suppose that $f(j) = [(j - m)/m]$, $g(j) = mf(j) + m + 1 - j$, and $\mathbf{Y}_k(s, t)$ is the $(s, t)$ element of $\mathbf{Y}_k$. Let $\{\delta_1, \ldots, \delta_m\}$ represent the linear aggregation scheme: $x_H(\tau_L) = \sum_{j=1}^{m} \delta_j x_H(\tau_L, j)$ with $\delta_j \geq 0$ and $\sum_{j=1}^{m} \delta_j = 1$. Then,

$$
\mathbf{G} = \begin{bmatrix} \mathbf{G}_{UL} & \mathbf{G}_{UR} \\ \mathbf{G}_{UR}^\prime & \mathbf{G}_{LR} \end{bmatrix}
$$

with

$$
\mathbf{G}_{UL} = \sum_{l=1}^{m} \delta_l \begin{bmatrix} \mathbf{Y}_{1-1}(K, l) & \ldots & \mathbf{Y}_{h-1}(K, l) \\ \vdots & \ddots & \vdots \\ \mathbf{Y}_{1-q}(K, l) & \ldots & \mathbf{Y}_{h-q}(K, l) \end{bmatrix}, \quad \mathbf{G}_{UR} = \sum_{l=1}^{m} \sum_{k=1}^{m} \delta_k \begin{bmatrix} \mathbf{Y}_{1-1}(k, l) & \ldots & \mathbf{Y}_{h-1}(k, l) \\ \vdots & \ddots & \vdots \\ \mathbf{Y}_{1-q}(k, l) & \ldots & \mathbf{Y}_{h-q}(k, l) \end{bmatrix}
$$

and

$$
\mathbf{C} = \begin{bmatrix} \mathbf{Y}_{f(1)}(K, g(1)) & \ldots & \mathbf{Y}_{f(pm)}(K, g(pm)) \\ \vdots & \ddots & \vdots \\ \sum_{l=1}^{m} \delta_l \mathbf{Y}_{f(1)-(q-1)}(K, g(1)) & \ldots & \sum_{l=1}^{m} \delta_l \mathbf{Y}_{f(pm)-(q-1)}(K, g(pm)) \\ \sum_{l=1}^{m} \delta_l \mathbf{Y}_{f(1)-(q-1)}(g(1), l) & \ldots & \sum_{l=1}^{m} \delta_l \mathbf{Y}_{f(pm)-(q-1)}(g(pm), l) \\ \vdots & \ddots & \vdots \\ \sum_{l=1}^{m} \delta_l \mathbf{Y}_{h-1-f(1)}(g(1), l) & \ldots & \sum_{l=1}^{m} \delta_l \mathbf{Y}_{h-1-f(pm)}(g(pm), l) \end{bmatrix}.
$$
where $\Gamma_{UL}$ is given in Theorem 3.4.

**Proof 3.8.** See Appendix N.

The procedure for calculating the local power associated with the low frequency Wald test is identical to the mixed frequency Wald test.

### 3.3 Numerical Examples

Through numerical examples we compare the four different tests listed in Table 1 and discussed above: mixed frequency max test, mixed frequency Wald test, low frequency max test, and low frequency Wald test. We initially consider a MF-VAR(1) data generating process to highlight an advantage of the mixed frequency tests relative to the low frequency tests. We then consider a MF-VAR(2) data generating process for robustness check.

**MF-VAR(1)** We assume that the true DGP follows a structural MF-VAR(1) process with $m = 12$:

$$
\begin{bmatrix}
1 & 0 & \ldots & \ldots & 0 \\
-\nu & 1 & \ldots & \ldots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & -\nu & \ldots & \ldots & 0 \\
0 & 0 & \ldots & 0 & 1
\end{bmatrix}
\begin{bmatrix}
x_{H}(\tau_{L}, 1) \\
x_{H}(\tau_{L}, 12) \\
x_{L}(\tau_{L}) \\
\vdots \\
b_{12}
\end{bmatrix}
= \begin{bmatrix}
0 & 0 & \ldots & d & c_{1} \\
0 & 0 & \ldots & 0 & c_{2} \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & \ldots & 0 & c_{12} \\
b_{12} & b_{11} & \ldots & b_{1}
\end{bmatrix}
\begin{bmatrix}
x_{H}(\tau_{L} - 1) \\
x_{H}(\tau_{L} - 1, 12) \\
x_{L}(\tau_{L} - 1) \\
\vdots \\
\eta(\tau_{L})
\end{bmatrix}
+ \begin{bmatrix}
\eta(\tau_{L}, 1) \\
\vdots \\
\eta(\tau_{L}, 12) \\
\eta(\tau_{L})
\end{bmatrix},
$$

(3.12)

where $\eta(\tau_{L}) \overset{m.d.}{\sim} (0_{13 \times 1}, I_{13})$. Coefficient $a$ governs the autoregressive property of $x_{L}$, while $d$ governs the autoregressive property of $x_{H}$. Besides these AR(1) structures, the DGP (3.12) has coefficients $c = [c_{1}, \ldots, c_{12}]$, which represent Granger causality from $x_{L}$ to $x_{H}$. Our main focus lies on coefficients $b = [b_{1}, \ldots, b_{12}]$, which represent Granger causality from $x_{H}$ to $x_{L}$.

To derive the reduced form of (3.12), note that

$$
N^{-1} = \begin{bmatrix}
1 & 0 & \ldots & \ldots & 0 \\
d & 1 & \ldots & \ldots & 0 \\
d^{2} & d & \ldots & \ldots & 0 \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
d^{11} & d^{10} & \ldots & d & 1 & 0 \\
0 & 0 & \ldots & 0 & 0 & 1
\end{bmatrix}
$$

and thus $A_{1} \equiv N^{-1} M_{1} = \begin{bmatrix}
0 & 0 & \ldots & d & \sum_{i=1}^{1} d^{i-1} c_{i} \\
0 & 0 & \ldots & d^{2} & \sum_{i=1}^{2} d^{2-i} c_{i} \\
\vdots & \vdots & \ddots & \ddots & \vdots \\
0 & 0 & \ldots & d^{12} & \sum_{i=1}^{12} d^{12-i} c_{i} \\
b_{12} & b_{11} & \ldots & b_{1} & a
\end{bmatrix}$.

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 (3.12) 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 $H_{1}^{L} : b = (1/\sqrt{T_{L}}) \nu$. 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, $\nu = [2.5, -1.25, 0.83, \ldots, -0.21]'$. 

23
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 and so on.

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 the 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 and we will consider MF-VAR(2) data generating process later. For the max tests, the weighting scheme is simply an equal scheme: \( W = (1/h) \times I_h \). 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 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 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. \( \delta_k = I(k = 12) \) for \( k = 1, \ldots, 12 \)). Nominal size \( \alpha \) is fixed at 0.05.

Table 2 compares the local asymptotic power of the four tests listed in Table 1: 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 hence 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 exactly included in the low frequency models with stock sampling coincidentally.
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).

We next turn on to Panel B, Lagged Causality. Focusing on the high persistence case $d = 0.8$ in Panel B.2, the 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 is reflecting 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 the 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 pattern 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. 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. Such a spillover effect exists 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 such a tricky (but realistic) causality, the mixed frequency tests with sufficiently many lags show a great advantage in terms of local asymptotic power.

While we have highlighted the advantage of the mixed frequency tests relative to the low frequency tests, it is not clear from Table 2 how the MF max test is preferred to the MF Wald test. There are $3 \times 3 \times 2 = 18$ ways of comparing them depending on causal patterns $\nu$, lag length $h_{MF}$, and 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 0.209 for Lagged Causality with $(d, h_{MF}) = (0.8, 12)$, it takes the smallest value of -0.182 for Sporadic Causality with $(d, h_{MF}) = (0.8, 12)$. We thus cannot firmly assert that the MF max test achieves higher 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).

**MF-VAR(2)** For robustness check we also consider a structural MF-VAR(2) data generating process with \( m = 12 \):  
\[
NX(\tau_L) = \sum_{k=1}^{2} M_k X(\tau_L - k) + \eta(\tau_L).
\]

All matrices except for \( M_2 \) are already defined in (3.12). \( M_2 \) is simply defined as follows.

\[
M_2 = \begin{bmatrix}
0 & \ldots & 0 & 0 \\
: & \ddots & \ddots & \vdots \\
0 & \ldots & 0 & 0 \\
b_{24} & \ldots & b_{13} & 0
\end{bmatrix}.
\]  

(3.13)

The reduced form is trivially \( X(\tau_L) = \sum_{k=1}^{2} A_k X(\tau_L - k) + \epsilon(\tau_L) \) with \( A_2 \equiv N^{-1} M_2 = M_2 \).

We prepare three causal patterns depending on \( \nu = [\nu_1, \ldots, \nu_{24}]' \).

1. **Decaying Causality**: \( \nu_j = (-1)^{j-1} \times 2.5/j \) for \( j = 1, \ldots, 24 \). The decaying structure continues up to lag 24.

2. **Lagged Causality**: \( \nu_j = 2 \times I(j = 24) \) for \( j = 1, \ldots, 24 \). Only 24th lag, not 12th, is nonzero at 2.

3. **Sporadic Causality**: \( (\nu_5, \nu_{12}, \nu_{17}, \nu_{19}) = (-1.2, 0.6, 2.8, -1.5) \) and all other \( \nu \)'s are zeros. There are no clear patterns in terms of signs, magnitude, and lags.

Other parameters are mostly as before; \( a = 0.2 \); \( d \in \{0.2, 0.8\} \); \( c_j = (-1)^{j-1} \times 0.8/j \) for \( j = 1, \ldots, 12 \); \( q = 2 \); \( W = (1/h) \times I_h \); \( \alpha = 0.05 \); the number of draws from the limit distributions under \( H_0 \) and \( H_1^f \) is 100,000 each.

The number of high frequency lags of \( x_H \) taken by the mixed frequency approaches is \( h_{MF} \in \{16, 20, 24\} \), while the number of low frequency lags of \( x_H \) taken by the low frequency approaches is \( h_{LF} \in \{1, 2, 3\} \). For the low frequency approaches, we consider both flow sampling and stock sampling.

Table 3 compares the local asymptotic power of the four tests listed in Table 1: MF max test, MF Wald test, LF max test, and LF Wald test. We first provide a quick review on the relative performance of the mixed frequency tests to the low frequency tests. The low frequency tests with flow sampling perform well under Lagged Causality (cfr. Panel B) but perform very poorly under Decaying Causality (cfr. Panel A). The low frequency tests with stock sampling perform much better than the mixed frequency tests under Decaying Causality but perform poorly under Lagged Causality. Under Sporadic Causality (cfr. Panel C), all low frequency tests have low power. We thus confirm our previous conclusion; low frequency tests are sometimes even more powerful than mixed frequency tests, but they easily lose power depending on causal patterns.

We now compare the MF max test and MF Wald test. There are \( 3 \times 3 \times 2 = 18 \) ways of comparing them depending on causal patterns \( \nu \), lag length \( h_{MF} \), and persistence \( d \). There are again 12 slots
where the max test is more powerful than the Wald test and 6 slots where the max test is less powerful. While the difference between the max test and the Wald test takes the largest value of 0.218 for Lagged Causality with \((d, h_{MF}) = (0.8, 24)\), it takes the smallest value of -0.087 for Lagged Causality with \((d, h_{MF}) = (0.2, 24)\). One might be tempted to argue that the smallest value has changed substantially from -0.182 to -0.087 by switching from MF-VAR(1) to MF-VAR(2), but this argument would not be convincing since Sporadic Causality between MF-VAR(1) and MF-VAR(2) is hard to compare due to its own nature of sporadicness. Our conclusion is therefore that, as long as local power framework is concerned, there is no clear ranking between the MF max test and the MF Wald test. It will turn out in the following Monte Carlo simulations that the max test achieves clearly higher power than the Wald test in small and medium sample size.

4 Monte Carlo Simulations

In this section we run Monte Carlo simulations to examine the finite sample performance 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.\(^5\)

4.1 High-to-Low Causality

**MF-VAR(1)** As in (3.12), we first assume that the true DGP follows a structural MF-VAR(1) process with \(m = 12\). Unlike the local power analysis, we actually have to simulate samples from the DGP (3.12). We simply 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 \(H_0 : b = 0_{12 \times 1}\), while the alternative hypothesis is general causality \(H_1 : b \neq 0_{12 \times 1}\). We prepare four types of \(b = [b_1, \ldots, b_{12}]'\) depending on causal patterns.

1. **Non-causality** \(b = 0_{12 \times 1}\). 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, \(b = [0.3, -0.15, 0.1, \ldots, -0.025]'\).
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-lagged causality. Such a complicated economic interrelationship should not be uncommon in practice.

\(^5\)The low-to-high part is currently concerned with the max tests only. The Wald tests should be added in the near future.
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 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 short 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 long sample but not totally unrealistic. We hereafter focus on the week vs. quarter scenario.

Given this small or medium sample size relative to the 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 Kilian (2004), which is designed to be robust against conditional heteroskedastic errors, for both MF and LF Wald tests. In our specific experiment the error terms is i.i.d., but in practice we do not have such a prior knowledge and often use the GK bootstrap to take care of potential GARCH effects. The procedure for the GK 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 \( x_L(\tau_L) = x(\tau_L - 1)^\prime \theta + u_L(\tau_L) \) to get \( \hat{\theta} \) and then compute the Wald statistic \( W \).

**Step 2** Run OLS for a restricted model \( x_L(\tau_L) = x(\tau_L - 1)^\prime \theta_0 + u_L(\tau_L) \) where \( \theta_0 = [\alpha', 0_{1 \times h}]' \) to get \( \tilde{\theta}_0 \) and \( \tilde{u}_L(\tau_L) \).

**Step 3** Simulate \( N \) samples from \( x_L(\tau_L) = x(\tau_L - 1)^\prime \tilde{\theta}_0 + \tilde{u}_L(\tau_L)v(\tau_L) \), where \( v(\tau_L) \sim i.i.d. N(0, 1) \). Note that this process is essentially a pure AR(\( q \)) process since the null hypothesis of non-causality is imposed.

**Step 4** For each sample compute Wald statistics \( \tilde{W}_1, \ldots, \tilde{W}_N \) and calculate the bootstrapped \( p \)-value:

\[
p_N = \frac{1}{N+1} \left( 1 + \sum_{i=1}^{N} I(\tilde{W}_i \geq W) \right).
\]

The null hypothesis of non-causality is rejected at level \( \alpha \) if \( p_N \leq \alpha \).

We now explain the details of 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 and we will consider MF-VAR(2) data generating process later. For the max tests, the weighting scheme is simply an equal scheme: \( W = (1/h) \times I_h \). 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 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 the low frequency approaches, 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 the max tests and 1,000 for the bootstrapped Wald tests. The number of bootstrap replications is \(N = 499\).

Table 4 compares the rejection frequencies of the four tests listed in Table 1: 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. The max tests have correct size due to their parsimonious specification, while the Wald tests have correct size due to Gonçalves and Killian’s (2004) parametric bootstrap. We can thus compare empirical power of each test meaningfully.

We start comparing empirical power with Panel B, Decaying Causality. Focusing on the low persistence case \(d = 0.2\) with medium sample \(T_L = 80\) in Panel B.1.2, the mixed frequency tests have moderately high power between 0.332 and 0.527. For example, the mixed frequency max test with \(h_{MF} = 4\) has power 0.482, while the mixed frequency Wald test with \(h_{MF} = 4\) has power 0.527.

The low frequency tests with flow sampling have absolutely no power regardless of the number of lags \(h_{LF} \in \{1, 2, 3\}\). The smallest value is 0.052 and the largest value is 0.088. An essential reason for this poor performance is that we have alternating signs in \(b\) and hence flow aggregation offsets all those effects. The low frequency tests with stock sampling, in contrast, have very high power between 0.429 and 0.657, 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 0.597. The reason for this high performance is that the largest coefficient \(b_1 = 0.3\) is assigned to \(x_H(\tau_L - 1, 12)\), which is exactly included in the low frequency models with stock sampling coincidentally.

If the persistence parameter \(d\) is raised from 0.2 to 0.8, empirical power rises in general (cfr. Panel B.2.2). If the sample size gets smaller from 80 to 40, empirical power gets lower as expected (cfr. Panels B.1.1 and B.2.1). In either case, the relative performance of each test is all unchanged.

We next move on to Panel C, Lagged Causality. Focusing on the high persistence case \(d = 0.8\) in Panel C.2, the mixed frequency cases have power that is increasing in \(h_{MF}\). For example, empirical power of the mixed frequency max test with \(T_L = 80\) is 0.084, 0.227, and 0.931 when \(h_{MF} = 4, 8,\) and 12, respectively. This is reflecting the causal pattern that only \(b_{12}\), coefficient of \(x_H(\tau_L - 1, 1)\), is 0.3 and all other \(b\)'s are zeros. It is thus important to include sufficiently many lags when we apply the 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 0.592, 0.606, and 0.539 when \(h_{LF} = 1, 2,\) 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
pattern is unambiguously positive; we have only one positive coefficient $b_{12} = 0.3$ 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. They have high power when $h_{LF} = 2$ (0.523 for max test and 0.418 for Wald test) since 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. Such a spillover effect exists only when coefficient $d$, the persistence of $x_H$, is relatively large.

When the persistence parameter $d$ decreases from 0.8 to 0.2, empirical power falls sharply in general, but the implications above still hold (cfr. Panels C.1.1 and C.1.2). The mixed frequency max test and Wald test with $(h_{MF}, T_L) = (12, 80)$ have power 0.405 and 0.264, 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 D, which highlights the full advantage of the mixed frequency tests over the low frequency tests. Fixing $(d, T_L) = (0.2, 80)$, the mixed frequency max test has power 0.248, 0.184, and 0.442 when $h_{MF}$ is 4, 8, and 12, respectively. Similarly, the mixed frequency Wald test has power 0.207, 0.167, and 0.416. Their power declines when switching from $h_{MF} = 4$ to $h_{MF} = 8$ since $b_5$, $b_6$, and $b_8$ are all zeros and $b_7 = 0.05$ is a relatively small value. The low frequency tests, whether flow sampling or stock sampling, have absolutely no power. The smallest value is 0.050 and the largest value is 0.062. This is an expected consequence since we learned from Panel B that flow sampling is vulnerable to alternating signs and from Panel C that stock sampling is vulnerable to lagged causality. In the presence of such a tricky (but realistic) causality, the mixed frequency tests with sufficiently many lags show a great advantage in terms of empirical power.

Now that we have highlighted the advantage of the mixed frequency tests relative to the low frequency tests, we now compare 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 in power is negligibly small, at most 0.045. 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 have found 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 with $m = 12$. The additional coefficient matrix $M_2$ is parameterized as in (3.13), so our DGP is identical to that in the local power section.

We prepare four causal patterns depending on $b = [b_1, \ldots, b_{24}]'$.

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

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

30
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 parameters are mostly as before: \( a = 0.2; \, d \in \{0.2, 0.8\}; \, c_j = (-1)^{j-1} \times 0.4/j \) for \( j = 1, \ldots, 12; \, T_L \in \{40, 80\}; \, q = 2; \, W = (1/h) \times I_h; \, \alpha = 0.05 \).

The number of high frequency lags of \( x_H \) taken by the mixed frequency tests is \( h_{MF} \in \{16, 20, 24\} \), while the number of low frequency lags of \( x_H \) taken by the low frequency tests is \( h_{LF} \in \{1, 2, 3\} \). For the 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 the MF tests and LF tests since their relationship is exactly analogous to the MF-V AR(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 0.006 and 0.034. Hence, the superiority of the max test to the Wald test is even more emphasized in MF-V AR(2) than in MF-V AR(1). This is reasonable since parameter proliferation is a more severe issue in MF-V AR(2).

4.2 Low-to-High Causality

We now consider Granger causality from \( x_L \) to \( x_H \). Assume that the true DGP is again (3.12), i.e. a bivariate structural MF-V AR(1) with the ratio of sampling frequencies \( m = 12 \). We assume that \( a = d = 0.2 \) and \( b_j = 0.2/j \) for \( j = 1, \ldots, 12 \). As in Section 4.1, we draw mutually and serially independent standard normal random numbers for the error term \( \eta(\tau_L) \). A key parameter vector \( c = [c_1, \ldots, c_{12}]' \) represents low-to-high causality, and we consider the following three cases:

1. Non-causality: \( c = 0_{12 \times 1} \). In this case we can check the empirical size of our tests.

2. Decaying Causality: \( c_j = 0.3/j \) for \( j = 1, \ldots, 12 \).

3. Lagged Causality: \( c_j = 0.4 \times I(j = 12) \) for \( j = 1, \ldots, 12 \). Only \( c_{12} \) is 0.4 and all other coefficients are zeros.

Having \( m = 12 \) can be thought of as a week vs. quarter case approximately, so we take \( T_L \in \{40, 80, 120\} \). These values can be thought of as 40 quarters (i.e. 10 years) through 120 quarters (i.e. 30 years).

\[6\] Sporadic Causality should be considered as well in the future. Also, alternating signs for \( b \) and \( c \) should be considered. Trying a more persistent case \( d = 0.8 \) is another future task.
For the mixed frequency max test, we combine \( h_{MF} \in \{4, 8, 12\} \) parsimonious regression models, and the \( j \)-th model is specified as

\[
x_{L}(\tau_{L}) = \alpha_{1,j} x_{L}(\tau_{L} - 1) + \sum_{k=1}^{r_{MF}} \beta_{k,j} x_{H}(\tau_{L} - 1, 12 + 1 - k) + \gamma_{j} x_{H}(\tau_{L} + 1, j) + u_{L,j}(\tau_{L}),
\]

(4.1)

Instruments: \( \{ \text{all } r_{MF} + 2 \text{ regressors in model } j, x_{H}(\tau_{L}, 1), \ldots, x_{H}(\tau_{L}, 12) \} \).

The number of high frequency lags of \( x_{H} \) is taken from \( r_{MF} \in \{4, 8, 12\} \).

For the purpose of comparison, we also formulate a low frequency counterpart to the model (4.1). We aggregate \( x_{H} \) using the linear aggregation scheme:

\[
x_{L}(\tau_{L}) = \sum_{j=1}^{12} \delta_{j} x_{H}(\tau_{L}, j).
\]

In particular, we focus on stock sampling \( \delta_{j} = I(j = 12) \) and flow sampling \( \delta_{j} = 1/12 \) in this simulation study. Using the aggregated \( x_{H} \), we combine \( h_{LF} \in \{1, 2, 3\} \) parsimonious regression models, and the \( j \)-th model is specified as

\[
x_{L}(\tau_{L}) = \alpha_{1,j}^{LF} x_{L}(\tau_{L} - 1) + \sum_{k=1}^{r_{LF}} \beta_{k,j}^{LF} x_{H}(\tau_{L} - k) + \gamma_{j}^{LF} x_{H}(\tau_{L} + j) + u_{L,j}^{LF}(\tau_{L}),
\]

(4.2)

Instruments: \( \{ \text{all } r_{LF} + 2 \text{ regressors in model } j, x_{H}(\tau_{L}) \} \).

The number of low frequency lags of \( x_{H} \) is taken from \( r_{LF} \in \{1, 2, 3\} \). We can formulate the max test corresponding to the low frequency model (4.2) in a completely analogous fashion with the mixed frequency case.\(^7\)

All max test statistics are computed based on the equal weighting scheme \( w_{j} = 1/h \). The number of Monte Carlo replications is 5,000, while the number of draws from the asymptotic null distribution is 1,000. The nominal size is 5%.

See Table 6 for the rejection frequencies. Panel A has Non-causality, Panel B has Decaying Causality, and Panel C has Lagged Causality. For each panel we have the mixed frequency case, low frequency case with stock sampling, and low frequency case with flow sampling. Different \( h \)’s are put vertically while different \( r \)’s are put horizontally. We first focus on Panel A to check empirical size. Since the low frequency approach involves few parameters, the empirical size is always very close to the nominal size 5% (cfr. Panels A.2 and A.3). The mixed frequency approach involves more parameters, so there is a size distortion issue when \( T_{L} \) is as small as 40 (cfr. Panel A.1). The worst empirical size of 0.176 occurs when \((h_{MF}, r_{MF}, T_{L}) = (12, 12, 40)\). The empirical size converges to the nominal size 5% quickly as \( T_{L} \) grows to 80, however.

We now focus on Panel B: Decaying Causality. For the mixed frequency case, the empirical power is at most 0.375 when \( T_{L} = 40 \), 0.629 when \( T_{L} = 80 \), and 0.827 when \( T_{L} = 120 \) (cfr. Panel B.1). Fixing \( r_{MF} \), a larger \( h_{MF} \) always produces lower power. This is reasonable since having more leads of \( x_{H} \) is not so informative given the decaying structure of low-to-high causality while the increased number of parameters certainly contributes to lower power. Similarly, having larger \( r_{MF} \) given a fixed level of \( h_{MF} \) does not always improve power due to the decaying pattern of high-to-low causality.

\(^7\)MF and LF Wald tests should also be considered in the future.
Panel B.2 indicates that the low frequency test with stock sampling has absolutely no power. In contrast, Panel B.3 indicates that the low frequency test with flow sampling is in fact more powerful than the mixed frequency test. For example, the low frequency test with flow sampling and \((h_{LF}, r_{LF}, T_L) = (1, 1, 40)\) yields the rejection frequency of 0.370, while the mixed frequency test at \((h_{MF}, r_{MF}, T_L) = (4, 4, 40)\) yields 0.305. This result suggests that flow sampling preserves the original pattern of Decaying Causality well.

We now focus on Panel C: Lagged Causality. For the mixed frequency case, having \(h_{MF} = 4, 8\) produces no power but having \(h_{MF} = 12\) produces high power, as desired (cfr. Panel C.1). Fixing \((h_{MF}, r_{MF}) = (12, 4)\), the empirical power is 0.209 when \(T_L = 40\), 0.582 when \(T_L = 80\), and 0.844 when \(T_L = 120\). These results are understandable since the twelfth lead of \(x_H\) is crucial for capturing the lagged low-to-high causality. As in Panel B, having larger \(r_{MF}\) does not always improve power due to the decaying pattern of high-to-low causality.

Panel C.2 indicates that the low frequency test with stock sampling is much more powerful than the mixed frequency test. For example, the low frequency test with stock sampling at \((h_{LF}, r_{LF}, T_L) = (1, 1, 40)\) yields the rejection frequency of 0.633, while the mixed frequency test at \((h_{MF}, r_{MF}, T_L) = (12, 4, 40)\) yields only 0.209. This is not surprising since the low frequency test with stock sampling works on \(x_H(\tau_L, 12)\), exactly relevant observations for Lagged Causality.

We summarize our comparison of the mixed frequency approach and the low frequency approach. The former always provides reasonable power by picking appropriate \(h_{MF}\), regardless of causal patterns. The low frequency approach with flow sampling performs better than the mixed frequency approach under Decaying Causality, but it does not work at all under Lagged Causality. The low frequency approach with stock sampling performs much better than the mixed frequency approach under Lagged Causality, but it does not work at all under Decaying Causality. In reality we do not know what kind of causality exists, so taking the mixed frequency approach is encouraged in order to avoid spurious non-causality.

### 5 Empirical Application

As an illustrative example of high-to-low max tests, this section investigates Granger causality from weekly interest rate spread to quarterly real GDP growth in the U.S. As in Sections 3 and 4, we compare all four tests listed in Table 1: mixed frequency max test, mixed frequency Wald test, low frequency max test, and low frequency Wald test. Interest rate spread used to be regarded as a strong predictor of business cycle, 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 in interest rate spread itself was referred to 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 past fifty years.

As a business cycle measure, seasonally-adjusted quarterly real GDP growth is used. The data can be found at Federal Reserve Economic Data (FRED). To remove potential seasonal effects remaining after seasonal adjustment, we use percentage growth rate 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 1 shows the weekly 10-year rate, weekly FF rate, their spread, and the quarterly GDP growth from January 5, 1962 through December 31, 2013. This 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, however, we find a weaker evidence or at least there is 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 over the past 52 years. The spread has a relatively large kurtosis of 5.611, while the GDP growth has a smaller kurtosis of 3.543.

When we apply the 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-1982:1, and so on. In this setting there are 129 subsamples in total.

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 model. Based on our simulation results for \( T_L = 80 \), the 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 naïve regression model is specified as follows.

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

(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 two models are estimated separately.

The low frequency parsimonious regression models are specified as follows.

$$x_L(\tau_L) = \alpha_{0,j}^{LF} + \sum_{k=1}^{2} \alpha_{k,j}^{LF} x_L(\tau_L - k) + \beta_{j}^{LF} x_H(\tau_L - j) + u_{L,j}^{LF}(\tau_L), \quad j = 1, \ldots, 3.$$ 

(5.3)

Since interest rate spread is a stock variable, we have that $x_H(\tau_L) = x_H(\tau_L, 13)$. We are including two quarters of lagged $x_L$ (i.e. $q = 2$) and three quarters 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.

$$x_L(\tau_L) = \alpha_{0}^{LF} + \sum_{k=1}^{2} \alpha_{k}^{LF} x_L(\tau_L - k) + \sum_{j=1}^{3} \beta_{j}^{LF} x_H(\tau_L - j) + u_{L}^{LF}(\tau_L).$$

(5.4)

Based on this model we compute the LF Wald statistic. Note again that $\beta_{j}^{LF}$ in (5.3) and $\beta_{j}^{LF}$ in (5.4) are not restricted to be the same.

As in Section 4.1, the parametric bootstrap of Gonçalves and Killian (2004) with replications $N = 999$ is used for the Wald tests in order to control size. For the max tests, the number of draws from limit distributions under the null of non-causality is 100,000.

Figure 2 plots the asymptotic $p$-values with respect to the null hypothesis of high-to-low non-causality. Panel (a) deals with the MF max test, Panel (b) deals with the MF Wald test, Panel (c) deals with the LF max test, and Panel (d) deals with the LF Wald test. If a $p$-value plotted is below a nominal size (say 0.05 or 0.1), then it means that 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 possibly 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 prolif-
eration. As seen in (5.1)-(5.4) or Table 1, the MF naive 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.

6 Conclusions

This paper proposes a new 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 \( j \)-th lag or lead of a high frequency variable \( x_H \) for \( j \in \{1, \ldots, h\} \). Let \( \hat{\beta}_j \) be an 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.

The max test based on mixed frequency data is consistent under any type of Granger causality like decaying or lagged causality. The standard Wald test based on mixed frequency data, which is essentially what Ghysels, Hill, and Motegi (2013) proposed, satisfies consistency as well. The max test based on low frequency data and the Wald test based on low frequency data do not satisfy consistency, which suggests the relevance of mixed frequency approach. We present an example of Granger causality that cannot be captured by the LF max test or LF Wald test no matter how many leads or lags are included and no matter what kind of linear aggregation scheme is used.

In local power analysis, we show that the MF max test and MF Wald test are more robust against some tricky (but realistic) causal patterns than their low frequency counterparts are. There is no clear ranking between the MF max test and MF Wald test in terms of local power.

We show via Monte Carlo simulations that the MF max test is clearly more powerful than the MF Wald test in finite sample. The MF max test is thus preferred to any other test when the ratio of sampling frequencies \( m \) is large and sample size is small.

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 1980 but the causality has vanished since then.

References


Tables and Figures

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

This table lists the four different high-to-low Granger causality tests. 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 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 prepares only one model which contains all $h$ lags of $x_H$ (either high frequency or low frequency). The 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, the 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 there.

<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></td>
<td>1. Consistent if $h \geq pm$</td>
<td>1. Inconsistent regardless of $h \in \mathbb{N}$</td>
</tr>
<tr>
<td></td>
<td>2. Few parameters</td>
<td>2. Few parameters</td>
</tr>
<tr>
<td><strong>Naive</strong></td>
<td>MF Wald Test</td>
<td>LF Wald Test</td>
</tr>
<tr>
<td></td>
<td>1. Consistent if $h \geq pm$</td>
<td>1. Inconsistent regardless of $h \in \mathbb{N}$</td>
</tr>
<tr>
<td></td>
<td>2. Many parameters</td>
<td>2. Few parameters</td>
</tr>
</tbody>
</table>

39
Table 2: Local Asymptotic Power of High-to-Low Causality Tests Based on MF-VAR(1)

This table compares the local asymptotic power of the four different high-to-low causality tests listed in Table 1: 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 key parameter \( \nu = [\nu_1, \ldots, \nu_{12}]' \), where \( \nu_j \) 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 \( j = 1, \ldots, 12 \). In this case the impact of \( x_H \) on \( x_L \) decays gradually with signs alternating. 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. This case is an extreme example of seasonality or lagged response of \( x_L \) to \( x_H \). 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. This is a realistic case where we have both positive and negative signs as well as unevenly-lagged causality. 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: \( W = (1/h) \times I_h \). 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 the low frequency approaches, 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>( h_{MF} )</th>
<th>( h_{LF} )</th>
<th>( \alpha )</th>
<th>( \alpha )</th>
<th>( \alpha )</th>
<th>( \alpha )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Max</td>
<td>Wald</td>
<td>LF (flow)</td>
<td>LF (flow)</td>
<td>LF (stock)</td>
</tr>
<tr>
<td>4</td>
<td>0.487</td>
<td>0.570</td>
<td>0.074</td>
<td>0.076</td>
<td>0.638</td>
</tr>
<tr>
<td>8</td>
<td>0.396</td>
<td>0.472</td>
<td>0.068</td>
<td>0.067</td>
<td>0.554</td>
</tr>
<tr>
<td>12</td>
<td>0.346</td>
<td>0.407</td>
<td>0.063</td>
<td>0.063</td>
<td>0.495</td>
</tr>
</tbody>
</table>

A.1. \( d = 0.2 \) (low persistence in \( x_H \))

<table>
<thead>
<tr>
<th>( h_{MF} )</th>
<th>( h_{LF} )</th>
<th>( \alpha )</th>
<th>( \alpha )</th>
<th>( \alpha )</th>
<th>( \alpha )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Max</td>
<td>Wald</td>
<td>LF (flow)</td>
<td>LF (flow)</td>
<td>LF (stock)</td>
</tr>
<tr>
<td>4</td>
<td>0.773</td>
<td>0.736</td>
<td>0.301</td>
<td>0.302</td>
<td>0.863</td>
</tr>
<tr>
<td>8</td>
<td>0.699</td>
<td>0.621</td>
<td>0.237</td>
<td>0.244</td>
<td>0.801</td>
</tr>
<tr>
<td>12</td>
<td>0.657</td>
<td>0.542</td>
<td>0.206</td>
<td>0.208</td>
<td>0.754</td>
</tr>
</tbody>
</table>

A.2. \( d = 0.8 \) (high persistence in \( x_H \))
Table 2: Local Asymptotic Power of High-to-Low Causality Tests Based on MF-VAR(1) (Continued)

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

<table>
<thead>
<tr>
<th></th>
<th>MF</th>
<th></th>
<th>LF (flow)</th>
<th></th>
<th>LF (stock)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Max</td>
<td>Wald</td>
<td>Max</td>
<td>Wald</td>
<td>Max</td>
<td>Wald</td>
</tr>
<tr>
<td>$h_{MF} = 4$</td>
<td>0.051</td>
<td>0.050</td>
<td>$h_{LF} = 1$</td>
<td>0.092</td>
<td>0.094</td>
<td>0.050</td>
</tr>
<tr>
<td>$h_{MF} = 8$</td>
<td>0.052</td>
<td>0.050</td>
<td>$h_{LF} = 2$</td>
<td>0.080</td>
<td>0.080</td>
<td>0.062</td>
</tr>
<tr>
<td>$h_{MF} = 12$</td>
<td>0.247</td>
<td>0.206</td>
<td>$h_{LF} = 3$</td>
<td>0.073</td>
<td>0.074</td>
<td>0.059</td>
</tr>
</tbody>
</table>

Panel B. Lagged Causality: $d = 0.2$ (low persistence in $x_H$)

<table>
<thead>
<tr>
<th></th>
<th>MF</th>
<th></th>
<th>LF (flow)</th>
<th></th>
<th>LF (stock)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Max</td>
<td>Wald</td>
<td>Max</td>
<td>Wald</td>
<td>Max</td>
<td>Wald</td>
</tr>
<tr>
<td>$h_{MF} = 4$</td>
<td>0.075</td>
<td>0.066</td>
<td>$h_{LF} = 1$</td>
<td>0.455</td>
<td>0.454</td>
<td>0.061</td>
</tr>
<tr>
<td>$h_{MF} = 8$</td>
<td>0.181</td>
<td>0.125</td>
<td>$h_{LF} = 2$</td>
<td>0.468</td>
<td>0.459</td>
<td>0.676</td>
</tr>
<tr>
<td>$h_{MF} = 12$</td>
<td>0.769</td>
<td>0.560</td>
<td>$h_{LF} = 3$</td>
<td>0.415</td>
<td>0.402</td>
<td>0.626</td>
</tr>
</tbody>
</table>

Panel B. Lagged Causality: $d = 0.8$ (high persistence in $x_H$)

<table>
<thead>
<tr>
<th></th>
<th>MF</th>
<th></th>
<th>LF (flow)</th>
<th></th>
<th>LF (stock)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Max</td>
<td>Wald</td>
<td>Max</td>
<td>Wald</td>
<td>Max</td>
<td>Wald</td>
</tr>
<tr>
<td>$h_{MF} = 4$</td>
<td>0.391</td>
<td>0.365</td>
<td>$h_{LF} = 1$</td>
<td>0.072</td>
<td>0.070</td>
<td>0.051</td>
</tr>
<tr>
<td>$h_{MF} = 8$</td>
<td>0.323</td>
<td>0.291</td>
<td>$h_{LF} = 2$</td>
<td>0.061</td>
<td>0.063</td>
<td>0.052</td>
</tr>
<tr>
<td>$h_{MF} = 12$</td>
<td>0.677</td>
<td>0.761</td>
<td>$h_{LF} = 3$</td>
<td>0.060</td>
<td>0.060</td>
<td>0.051</td>
</tr>
</tbody>
</table>

Panel C. Sporadic Causality: $(\nu_5, \nu_9, \nu_{11}) = (2.1, -2.8, 1.9)$

Panel C. Sporadic Causality: $d = 0.2$ (low persistence in $x_H$)

<table>
<thead>
<tr>
<th></th>
<th>MF</th>
<th></th>
<th>LF (flow)</th>
<th></th>
<th>LF (stock)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Max</td>
<td>Wald</td>
<td>Max</td>
<td>Wald</td>
<td>Max</td>
<td>Wald</td>
</tr>
<tr>
<td>$h_{MF} = 4$</td>
<td>0.716</td>
<td>0.614</td>
<td>$h_{LF} = 1$</td>
<td>0.171</td>
<td>0.169</td>
<td>0.437</td>
</tr>
<tr>
<td>$h_{MF} = 8$</td>
<td>0.667</td>
<td>0.679</td>
<td>$h_{LF} = 2$</td>
<td>0.132</td>
<td>0.132</td>
<td>0.355</td>
</tr>
<tr>
<td>$h_{MF} = 12$</td>
<td>0.690</td>
<td>0.872</td>
<td>$h_{LF} = 3$</td>
<td>0.118</td>
<td>0.115</td>
<td>0.306</td>
</tr>
</tbody>
</table>
Table 3: Local Asymptotic Power of High-to-Low Causality Tests Based on MF-VAR(2)

This table compares the local asymptotic power of the four different high-to-low causality tests listed in Table 1: 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 three panels depending on the key parameter \( \nu = [\nu_1, \ldots, \nu_{24}]^\top \), where \( \nu_j \) signifies the impact of \( x_H(t_L - 1, m + 1 - j) \) on \( x_L(t_L) \). Panel A deals with Decaying Causality: \( \nu_j = (-1)^{j-1} \times 2.5/j \) for \( j = 1, \ldots, 24 \). In this case the impact of \( x_H \) on \( x_L \) decays gradually with signs alternating. Panel B deals with Lagged Causality: \( \nu_j = 2 \times I(j = 24) \) for \( j = 1, \ldots, 24 \). Only \( \nu_{24} \) is 2 and all others are zeros. This case is an extreme example of seasonality or lagged response of \( x_L \) to \( x_H \). Panel C deals with Sporadic Causality: \( (\nu_5, \nu_{12}, \nu_{17}, \nu_{19}) = (-1.2, 0.6, 2.8, -1.5) \), and all other \( \nu \)'s are zeros. This is a realistic case where we have both positive and negative signs as well as unevenly-lagged causality. 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, for low-to-high causality, we assume decaying causality with alternating signs up to low frequency lag 1: \( c_j = (-1)^{j-1} \times 0.8/j \) 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: \( W = (1/h) \times I_h \). The number of high frequency lags of \( x_H \) taken by the mixed frequency approaches is \( h_{MF} \in \{16, 20, 24\} \), while the number of low frequency lags of \( x_H \) taken by the low frequency approaches is \( h_{LF} \in \{1, 2, 3\} \). For the low frequency approaches, 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>( h_{MF} )</th>
<th>MF Max</th>
<th>MF Wald</th>
<th>LF (flow) Max</th>
<th>LF (flow) Wald</th>
<th>LF (stock) Max</th>
<th>LF (stock) Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>0.322</td>
<td>0.360</td>
<td>0.079</td>
<td>0.077</td>
<td>0.646</td>
<td>0.643</td>
</tr>
<tr>
<td>20</td>
<td>0.298</td>
<td>0.326</td>
<td>0.068</td>
<td>0.068</td>
<td>0.546</td>
<td>0.539</td>
</tr>
<tr>
<td>24</td>
<td>0.288</td>
<td>0.300</td>
<td>0.065</td>
<td>0.064</td>
<td>0.491</td>
<td>0.475</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( h_{MF} )</th>
<th>MF Max</th>
<th>MF Wald</th>
<th>LF (flow) Max</th>
<th>LF (flow) Wald</th>
<th>LF (stock) Max</th>
<th>LF (stock) Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>0.634</td>
<td>0.485</td>
<td>0.331</td>
<td>0.330</td>
<td>0.864</td>
<td>0.865</td>
</tr>
<tr>
<td>20</td>
<td>0.606</td>
<td>0.440</td>
<td>0.259</td>
<td>0.266</td>
<td>0.802</td>
<td>0.788</td>
</tr>
<tr>
<td>24</td>
<td>0.584</td>
<td>0.405</td>
<td>0.224</td>
<td>0.227</td>
<td>0.764</td>
<td>0.731</td>
</tr>
</tbody>
</table>

42
Table 3: Local Asymptotic Power of High-to-Low Causality Tests Based on MF-VAR(2) (Continued)

Panel B. Lagged Causality: $\nu_j = 2 \times I(j = 24)$

<table>
<thead>
<tr>
<th></th>
<th>MF</th>
<th>LF (flow)</th>
<th>LF (stock)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Max Wald</td>
<td>Max Wald</td>
<td>Max Wald</td>
</tr>
<tr>
<td>$h_{MF} = 16$</td>
<td>0.052 0.051</td>
<td>0.050 0.050</td>
<td>0.049 0.050</td>
</tr>
<tr>
<td>$h_{MF} = 20$</td>
<td>0.051 0.054</td>
<td>0.106 0.109</td>
<td>0.052 0.051</td>
</tr>
<tr>
<td>$h_{MF} = 24$</td>
<td>0.151 0.238</td>
<td>0.095 0.098</td>
<td>0.059 0.060</td>
</tr>
</tbody>
</table>

Panel C. Sporadic Causality: $(\nu_5, \nu_{12}, \nu_{17}, \nu_{19}) = (-1.2, 0.6, 2.8, -1.5)$

<table>
<thead>
<tr>
<th></th>
<th>MF</th>
<th>LF (flow)</th>
<th>LF (stock)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Max Wald</td>
<td>Max Wald</td>
<td>Max Wald</td>
</tr>
<tr>
<td>$h_{MF} = 16$</td>
<td>0.106 0.109</td>
<td>0.057 0.058</td>
<td>0.049 0.050</td>
</tr>
<tr>
<td>$h_{MF} = 20$</td>
<td>0.493 0.491</td>
<td>0.072 0.075</td>
<td>0.051 0.051</td>
</tr>
<tr>
<td>$h_{MF} = 24$</td>
<td>0.470 0.451</td>
<td>0.068 0.069</td>
<td>0.053 0.051</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>MF</th>
<th>LF (flow)</th>
<th>LF (stock)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_{MF} = 16$</td>
<td>0.613 0.507</td>
<td>0.072 0.073</td>
<td>0.102 0.099</td>
</tr>
<tr>
<td>$h_{MF} = 20$</td>
<td>0.753 0.730</td>
<td>0.390 0.447</td>
<td>0.342 0.374</td>
</tr>
<tr>
<td>$h_{MF} = 24$</td>
<td>0.734 0.688</td>
<td>0.337 0.403</td>
<td>0.296 0.321</td>
</tr>
</tbody>
</table>
Table 4: Rejection Frequencies of High-to-Low Causality Tests Based on MF-VAR(1)

This table compares the rejection frequencies of the four different high-to-low causality tests listed in Table 1: 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: \( b = 0_{12 \times 1} \). 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: \((b_3, \tau_7, \nu_{10}) = (0.2, 0.05, -0.3)\), and all other \( b_j \)'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. \( a = 0.2 \)). 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 the max tests, the number of draws from the limit distributions under \( H_0 \) and \( H_1 \) is 5,000 each. The weighting scheme is simply an equal scheme: \( W = (1/h) \times I_h \). 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 the low frequency approaches, we consider both flow sampling and stock sampling. For the 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 the max tests and 1,000 for the Wald tests.

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

#### A.1. \( d = 0.2 \) (low persistence in \( x_H \))

**A.1.1. \( T_L = 40 \) (small sample size)**

<table>
<thead>
<tr>
<th></th>
<th>MF (Max)</th>
<th>MF (Wald)</th>
<th>LF (flow) (Max)</th>
<th>LF (flow) (Wald)</th>
<th>LF (stock) (Max)</th>
<th>LF (stock) (Wald)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( h_{MF} = 4 )</td>
<td>0.061</td>
<td>0.035</td>
<td>0.064</td>
<td>0.044</td>
<td>0.063</td>
<td>0.045</td>
</tr>
<tr>
<td>( h_{MF} = 8 )</td>
<td>0.056</td>
<td>0.049</td>
<td>0.061</td>
<td>0.037</td>
<td>0.055</td>
<td>0.042</td>
</tr>
<tr>
<td>( h_{MF} = 12 )</td>
<td>0.063</td>
<td>0.038</td>
<td>0.058</td>
<td>0.049</td>
<td>0.061</td>
<td>0.046</td>
</tr>
</tbody>
</table>

#### A.1.2. \( T_L = 80 \) (medium sample size)

<table>
<thead>
<tr>
<th></th>
<th>MF (Max)</th>
<th>MF (Wald)</th>
<th>LF (flow) (Max)</th>
<th>LF (flow) (Wald)</th>
<th>LF (stock) (Max)</th>
<th>LF (stock) (Wald)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( h_{MF} = 4 )</td>
<td>0.055</td>
<td>0.054</td>
<td>0.058</td>
<td>0.048</td>
<td>0.056</td>
<td>0.045</td>
</tr>
<tr>
<td>( h_{MF} = 8 )</td>
<td>0.057</td>
<td>0.037</td>
<td>0.055</td>
<td>0.040</td>
<td>0.058</td>
<td>0.039</td>
</tr>
<tr>
<td>( h_{MF} = 12 )</td>
<td>0.052</td>
<td>0.048</td>
<td>0.058</td>
<td>0.065</td>
<td>0.058</td>
<td>0.042</td>
</tr>
</tbody>
</table>

#### A.2. \( d = 0.8 \) (high persistence in \( x_H \))

**A.2.1. \( T_L = 40 \) (small sample)**

<table>
<thead>
<tr>
<th></th>
<th>MF (Max)</th>
<th>MF (Wald)</th>
<th>LF (flow) (Max)</th>
<th>LF (flow) (Wald)</th>
<th>LF (stock) (Max)</th>
<th>LF (stock) (Wald)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( h_{MF} = 4 )</td>
<td>0.058</td>
<td>0.050</td>
<td>0.065</td>
<td>0.041</td>
<td>0.063</td>
<td>0.048</td>
</tr>
<tr>
<td>( h_{MF} = 8 )</td>
<td>0.066</td>
<td>0.041</td>
<td>0.063</td>
<td>0.050</td>
<td>0.057</td>
<td>0.053</td>
</tr>
<tr>
<td>( h_{MF} = 12 )</td>
<td>0.058</td>
<td>0.032</td>
<td>0.068</td>
<td>0.043</td>
<td>0.060</td>
<td>0.038</td>
</tr>
</tbody>
</table>

**A.2.2. \( T_L = 80 \) (medium sample)**

<table>
<thead>
<tr>
<th></th>
<th>MF (Max)</th>
<th>MF (Wald)</th>
<th>LF (flow) (Max)</th>
<th>LF (flow) (Wald)</th>
<th>LF (stock) (Max)</th>
<th>LF (stock) (Wald)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( h_{MF} = 4 )</td>
<td>0.060</td>
<td>0.035</td>
<td>0.058</td>
<td>0.038</td>
<td>0.057</td>
<td>0.045</td>
</tr>
<tr>
<td>( h_{MF} = 8 )</td>
<td>0.058</td>
<td>0.037</td>
<td>0.056</td>
<td>0.054</td>
<td>0.053</td>
<td>0.055</td>
</tr>
<tr>
<td>( h_{MF} = 12 )</td>
<td>0.055</td>
<td>0.043</td>
<td>0.054</td>
<td>0.039</td>
<td>0.058</td>
<td>0.038</td>
</tr>
</tbody>
</table>
Table 4: Rejection Frequencies of High-to-Low Causality Tests Based on MF-VAR(1) (Continued)

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

<table>
<thead>
<tr>
<th></th>
<th>MF (flow)</th>
<th>LF (stock)</th>
</tr>
</thead>
<tbody>
<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>$h_{MF} = 4$</td>
<td>0.228</td>
<td>0.241</td>
</tr>
<tr>
<td>$h_{LF} = 1$</td>
<td>0.072</td>
<td>0.047</td>
</tr>
<tr>
<td>$h_{MF} = 8$</td>
<td>0.163</td>
<td>0.157</td>
</tr>
<tr>
<td>$h_{LF} = 2$</td>
<td>0.068</td>
<td>0.050</td>
</tr>
<tr>
<td>$h_{MF} = 12$</td>
<td>0.128</td>
<td>0.136</td>
</tr>
<tr>
<td>$h_{LF} = 3$</td>
<td>0.061</td>
<td>0.040</td>
</tr>
<tr>
<td><strong>B.1.2. $T_L = 80$ (medium sample size)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$h_{MF} = 4$</td>
<td>0.482</td>
<td>0.527</td>
</tr>
<tr>
<td>$h_{LF} = 1$</td>
<td>0.088</td>
<td>0.062</td>
</tr>
<tr>
<td>$h_{MF} = 8$</td>
<td>0.374</td>
<td>0.412</td>
</tr>
<tr>
<td>$h_{LF} = 2$</td>
<td>0.071</td>
<td>0.058</td>
</tr>
<tr>
<td>$h_{MF} = 12$</td>
<td>0.332</td>
<td>0.335</td>
</tr>
<tr>
<td>$h_{LF} = 3$</td>
<td>0.069</td>
<td>0.052</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>$h_{MF} = 4$</td>
<td>0.444</td>
<td>0.305</td>
</tr>
<tr>
<td>$h_{LF} = 1$</td>
<td>0.185</td>
<td>0.138</td>
</tr>
<tr>
<td>$h_{MF} = 8$</td>
<td>0.343</td>
<td>0.222</td>
</tr>
<tr>
<td>$h_{LF} = 2$</td>
<td>0.145</td>
<td>0.096</td>
</tr>
<tr>
<td>$h_{MF} = 12$</td>
<td>0.272</td>
<td>0.163</td>
</tr>
<tr>
<td>$h_{LF} = 3$</td>
<td>0.122</td>
<td>0.087</td>
</tr>
<tr>
<td><strong>B.2.2. $T_L = 80$ (medium sample size)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$h_{MF} = 4$</td>
<td>0.794</td>
<td>0.700</td>
</tr>
<tr>
<td>$h_{LF} = 1$</td>
<td>0.309</td>
<td>0.265</td>
</tr>
<tr>
<td>$h_{MF} = 8$</td>
<td>0.697</td>
<td>0.564</td>
</tr>
<tr>
<td>$h_{LF} = 2$</td>
<td>0.235</td>
<td>0.205</td>
</tr>
<tr>
<td>$h_{MF} = 12$</td>
<td>0.642</td>
<td>0.480</td>
</tr>
<tr>
<td>$h_{LF} = 3$</td>
<td>0.198</td>
<td>0.159</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><strong>C.1. $d = 0.2$ (low persistence in $x_H$)</strong></td>
</tr>
<tr>
<td><strong>C.1.1. $T_L = 40$ (small sample size)</strong></td>
</tr>
<tr>
<td>MF</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><strong>C.1.2. $T_L = 80$ (medium sample size)</strong></td>
</tr>
<tr>
<td>MF</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><strong>C.2. $d = 0.8$ (high persistence in $x_H$)</strong></td>
</tr>
<tr>
<td><strong>C.2.1. $T_L = 40$ (small sample size)</strong></td>
</tr>
<tr>
<td>MF</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><strong>C.2.2. $T_L = 80$ (medium sample size)</strong></td>
</tr>
<tr>
<td>MF</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><strong>MF</strong></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Max</td>
<td>Wald</td>
<td>Max</td>
<td>Wald</td>
<td>Max</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D.1. (d = 0.2) (low persistence in (x_H))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D.1.1. (T_L = 40) (small sample size)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(h_{MF} = 4)</td>
<td>0.119</td>
<td>0.075</td>
<td>(h_{LF} = 1)</td>
<td>0.060</td>
<td>0.043</td>
</tr>
<tr>
<td>(h_{MF} = 8)</td>
<td>0.101</td>
<td>0.079</td>
<td>(h_{LF} = 2)</td>
<td>0.061</td>
<td>0.043</td>
</tr>
<tr>
<td>(h_{MF} = 12)</td>
<td>0.173</td>
<td>0.136</td>
<td>(h_{LF} = 3)</td>
<td>0.067</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D.1.2. (T_L = 80) (medium sample size)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(h_{MF} = 4)</td>
<td>0.248</td>
<td>0.207</td>
<td>(h_{LF} = 1)</td>
<td>0.058</td>
<td>0.057</td>
</tr>
<tr>
<td>(h_{MF} = 8)</td>
<td>0.184</td>
<td>0.167</td>
<td>(h_{LF} = 2)</td>
<td>0.052</td>
<td>0.049</td>
</tr>
<tr>
<td>(h_{MF} = 12)</td>
<td>0.442</td>
<td>0.416</td>
<td>(h_{LF} = 3)</td>
<td>0.050</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D.2. (d = 0.8) (high persistence in (x_H))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D.2.1. (T_L = 40) (small sample size)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(h_{MF} = 4)</td>
<td>0.243</td>
<td>0.163</td>
<td>(h_{LF} = 1)</td>
<td>0.075</td>
<td>0.047</td>
</tr>
<tr>
<td>(h_{MF} = 8)</td>
<td>0.197</td>
<td>0.180</td>
<td>(h_{LF} = 2)</td>
<td>0.093</td>
<td>0.057</td>
</tr>
<tr>
<td>(h_{MF} = 12)</td>
<td>0.402</td>
<td>0.260</td>
<td>(h_{LF} = 3)</td>
<td>0.082</td>
<td>0.063</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D.2.2. (T_L = 80) (medium sample size)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(h_{MF} = 4)</td>
<td>0.459</td>
<td>0.305</td>
<td>(h_{LF} = 1)</td>
<td>0.076</td>
<td>0.052</td>
</tr>
<tr>
<td>(h_{MF} = 8)</td>
<td>0.404</td>
<td>0.404</td>
<td>(h_{LF} = 2)</td>
<td>0.128</td>
<td>0.101</td>
</tr>
<tr>
<td>(h_{MF} = 12)</td>
<td>0.803</td>
<td>0.740</td>
<td>(h_{LF} = 3)</td>
<td>0.107</td>
<td>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 the four different high-to-low causality tests listed in Table 1: 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 \( b = [b_1, \ldots, b_2]^\prime \), 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: \( (b_0, b_12, b_{17}, b_{19}) = (-0.2, 0.1, 0.2, -0.35) \), and all other \( b_j \)'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 the max tests, the number of draws from the limit distributions under \( H_0 \) and \( H_1^L \) is 5,000 each. The weighting scheme is simply an equal scheme: \( W = (1/h) \times I_n \). The number of high frequency lags of \( x_H \) taken by the mixed frequency approaches is \( h_{MF} \in \{16, 20, 24\} \), while the number of low frequency lags of \( x_H \) taken by the low frequency approaches is \( h_{LF} \in \{1, 2, 3\} \). For the low frequency approaches, we consider both flow sampling and stock sampling. For the 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 the max tests and 1,000 for the Wald tests.

<table>
<thead>
<tr>
<th>Panel A. Non-causality: ( b = 0_{24 \times 1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A.1. ( d = 0.2 ) (low persistence in ( x_H ))</strong></td>
</tr>
<tr>
<td>( 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} = 16 )</td>
</tr>
<tr>
<td>( h_{MF} = 20 )</td>
</tr>
<tr>
<td>( h_{MF} = 24 )</td>
</tr>
<tr>
<td><strong>A.1.2. ( T_L = 80 ) (medium sample size)</strong></td>
</tr>
<tr>
<td>MF</td>
</tr>
<tr>
<td>Max</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>( h_{MF} = 16 )</td>
</tr>
<tr>
<td>( h_{MF} = 20 )</td>
</tr>
<tr>
<td>( h_{MF} = 24 )</td>
</tr>
<tr>
<td><strong>A.2. ( d = 0.8 ) (high persistence in ( x_H ))</strong></td>
</tr>
<tr>
<td><strong>A.2.1. ( T_L = 40 ) (small sample)</strong></td>
</tr>
<tr>
<td>MF</td>
</tr>
<tr>
<td>Max</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>( h_{MF} = 16 )</td>
</tr>
<tr>
<td>( h_{MF} = 20 )</td>
</tr>
<tr>
<td>( h_{MF} = 24 )</td>
</tr>
<tr>
<td><strong>A.2.2. ( T_L = 80 ) (medium sample)</strong></td>
</tr>
<tr>
<td>MF</td>
</tr>
<tr>
<td>Max</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>( h_{MF} = 16 )</td>
</tr>
<tr>
<td>( h_{MF} = 20 )</td>
</tr>
<tr>
<td>( h_{MF} = 24 )</td>
</tr>
</tbody>
</table>

48
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>$\text{MF}$</th>
<th>$\text{LF (flow)}$</th>
<th>$\text{LF (stock)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{Max}$</td>
<td>$\text{Wald}$</td>
<td>$\text{Max}$</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>
</tbody>
</table>

B.1. $d = 0.2$ (low persistence in $x_H$)
B.1.1. $T_L = 40$ (small sample size)

<table>
<thead>
<tr>
<th>$\text{MF}$</th>
<th>$\text{LF (flow)}$</th>
<th>$\text{LF (stock)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{Max}$</td>
<td>$\text{Wald}$</td>
<td>$\text{Max}$</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>
</tbody>
</table>

B.1.2. $T_L = 80$ (medium sample size)

<table>
<thead>
<tr>
<th>$\text{MF}$</th>
<th>$\text{LF (flow)}$</th>
<th>$\text{LF (stock)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{Max}$</td>
<td>$\text{Wald}$</td>
<td>$\text{Max}$</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>

B.2. $d = 0.8$ (high persistence in $x_H$)
B.2.1. $T_L = 40$ (small sample size)

<table>
<thead>
<tr>
<th>$\text{MF}$</th>
<th>$\text{LF (flow)}$</th>
<th>$\text{LF (stock)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{Max}$</td>
<td>$\text{Wald}$</td>
<td>$\text{Max}$</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>
</tbody>
</table>

B.2.2. $T_L = 80$ (medium sample size)
Table 5: Rejection Frequencies of High-to-Low Causality Tests Based on MF-VAR(2) (Continued)

<table>
<thead>
<tr>
<th>Panel C. Lagged Causality: $b_j = 0.3 \times I(j = 24)$</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>C.1. $d = 0.2$ (low persistence in $x_H$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.1.1. $T_L = 40$ (small sample size)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MF</th>
<th>LF (flow) Max</th>
<th>Wald</th>
<th>LF (stock) Max</th>
<th>Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_{MF} = 16$</td>
<td>0.060</td>
<td>0.044</td>
<td>$h_{LF} = 1$</td>
<td>0.061</td>
</tr>
<tr>
<td>$h_{MF} = 20$</td>
<td>0.060</td>
<td>0.046</td>
<td>$h_{LF} = 2$</td>
<td>0.082</td>
</tr>
<tr>
<td>$h_{MF} = 24$</td>
<td>0.091</td>
<td>0.055</td>
<td>$h_{LF} = 3$</td>
<td>0.083</td>
</tr>
</tbody>
</table>

| C.1.2. $T_L = 80$ (medium sample size) |

<table>
<thead>
<tr>
<th>MF</th>
<th>LF (flow) Max</th>
<th>Wald</th>
<th>LF (stock) Max</th>
<th>Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_{MF} = 16$</td>
<td>0.057</td>
<td>0.050</td>
<td>$h_{LF} = 1$</td>
<td>0.055</td>
</tr>
<tr>
<td>$h_{MF} = 20$</td>
<td>0.051</td>
<td>0.046</td>
<td>$h_{LF} = 2$</td>
<td>0.104</td>
</tr>
<tr>
<td>$h_{MF} = 24$</td>
<td>0.273</td>
<td>0.195</td>
<td>$h_{LF} = 3$</td>
<td>0.099</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>C.2. $d = 0.8$ (high persistence in $x_H$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.2.1. $T_L = 40$ (small sample size)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MF</th>
<th>LF (flow) Max</th>
<th>Wald</th>
<th>LF (stock) Max</th>
<th>Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_{MF} = 16$</td>
<td>0.063</td>
<td>0.042</td>
<td>$h_{LF} = 1$</td>
<td>0.064</td>
</tr>
<tr>
<td>$h_{MF} = 20$</td>
<td>0.095</td>
<td>0.073</td>
<td>$h_{LF} = 2$</td>
<td>0.260</td>
</tr>
<tr>
<td>$h_{MF} = 24$</td>
<td>0.438</td>
<td>0.108</td>
<td>$h_{LF} = 3$</td>
<td>0.285</td>
</tr>
</tbody>
</table>

| C.2.2. $T_L = 80$ (medium sample size) |

<table>
<thead>
<tr>
<th>MF</th>
<th>LF (flow) Max</th>
<th>Wald</th>
<th>LF (stock) Max</th>
<th>Wald</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_{MF} = 16$</td>
<td>0.069</td>
<td>0.053</td>
<td>$h_{LF} = 1$</td>
<td>0.057</td>
</tr>
<tr>
<td>$h_{MF} = 20$</td>
<td>0.162</td>
<td>0.096</td>
<td>$h_{LF} = 2$</td>
<td>0.515</td>
</tr>
<tr>
<td>$h_{MF} = 24$</td>
<td>0.896</td>
<td>0.480</td>
<td>$h_{LF} = 3$</td>
<td>0.588</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)</th>
<th>(LF) (flow)</th>
<th>(LF) (stock)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Max</td>
<td>Wald</td>
<td>Max</td>
</tr>
<tr>
<td>(D.1. d = 0.2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(T_L = 40)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(h_{MF} = 16)</td>
<td>0.090</td>
<td>0.064</td>
<td>(h_{LF} = 1)</td>
</tr>
<tr>
<td>(h_{MF} = 20)</td>
<td>0.186</td>
<td>0.146</td>
<td>(h_{LF} = 2)</td>
</tr>
<tr>
<td>(h_{MF} = 24)</td>
<td>0.153</td>
<td>0.089</td>
<td>(h_{LF} = 3)</td>
</tr>
<tr>
<td>(D.1.1. d = 0.2) (T_L = 80) (medium sample size)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(h_{MF} = 16)</td>
<td>0.141</td>
<td>0.118</td>
<td>(h_{LF} = 1)</td>
</tr>
<tr>
<td>(h_{MF} = 20)</td>
<td>0.502</td>
<td>0.474</td>
<td>(h_{LF} = 2)</td>
</tr>
<tr>
<td>(h_{MF} = 24)</td>
<td>0.461</td>
<td>0.419</td>
<td>(h_{LF} = 3)</td>
</tr>
<tr>
<td>(D.2. d = 0.8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(T_L = 40)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(h_{MF} = 16)</td>
<td>0.170</td>
<td>0.090</td>
<td>(h_{LF} = 1)</td>
</tr>
<tr>
<td>(h_{MF} = 20)</td>
<td>0.254</td>
<td>0.187</td>
<td>(h_{LF} = 2)</td>
</tr>
<tr>
<td>(h_{MF} = 24)</td>
<td>0.256</td>
<td>0.136</td>
<td>(h_{LF} = 3)</td>
</tr>
<tr>
<td>(D.2.2. T_L = 80) (medium sample size)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(h_{MF} = 16)</td>
<td>0.390</td>
<td>0.234</td>
<td>(h_{LF} = 1)</td>
</tr>
<tr>
<td>(h_{MF} = 20)</td>
<td>0.621</td>
<td>0.655</td>
<td>(h_{LF} = 2)</td>
</tr>
<tr>
<td>(h_{MF} = 24)</td>
<td>0.623</td>
<td>0.606</td>
<td>(h_{LF} = 3)</td>
</tr>
</tbody>
</table>
Table 6: Rejection Frequencies of Max Test for Low-to-High Granger Causality

This table shows the rejection frequencies of the max tests for low-to-high Granger causality. Panel A assumes Non-causality, Panel B assumes Decaying Causality, and Panel C assumes Lagged Causality. For each panel we have the mixed frequency case, low frequency case with stock sampling, and low frequency case with flow sampling. For the mixed frequency case we combine $h_{MF} \in \{4, 8, 12\}$ parsimonious regression models, and the number of high frequency lags of $x_H$ is taken from $r_{MF} \in \{4, 8, 12\}$. For the low frequency case we combine $h_{LF} \in \{1, 2, 3\}$ parsimonious regression models, and the number of low frequency lags of $x_H$ is taken from $r_{LF} \in \{1, 2, 3\}$. Different $h$'s are put vertically while different $r$’s are put horizontally. All max tests use the equal weighting scheme, and the test statistic is computed based on 1,000 draws from the asymptotic null distribution. We fix $m = 12$, which can be thought of as a week vs. quarter case approximately. The sample size $T_L$ is 40, 80, or 120 quarters. There is decaying Granger causality from $x_H$ to $x_L$ in the sense that $x_L(\tau_L)$ depends on $\sum_{j=1}^{12} (0.2/j)x_H(\tau_L - 1, m + 1 - j)$. 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. The nominal size is 5%.

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

#### Panel A.1. Mixed Frequency

<table>
<thead>
<tr>
<th>$h_{MF}$ \ $r_{MF}$</th>
<th>$T_L = 40$</th>
<th>$T_L = 80$</th>
<th>$T_L = 120$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.071</td>
<td>0.100</td>
<td>0.154</td>
</tr>
<tr>
<td>8</td>
<td>0.070</td>
<td>0.106</td>
<td>0.163</td>
</tr>
<tr>
<td>12</td>
<td>0.071</td>
<td>0.105</td>
<td>0.176</td>
</tr>
</tbody>
</table>

#### Panel A.2. Low Frequency (Stock Sampling)

<table>
<thead>
<tr>
<th>$h_{LF}$ \ $r_{LF}$</th>
<th>$T_L = 40$</th>
<th>$T_L = 80$</th>
<th>$T_L = 120$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.056</td>
<td>0.063</td>
<td>0.067</td>
</tr>
<tr>
<td>2</td>
<td>0.054</td>
<td>0.056</td>
<td>0.061</td>
</tr>
<tr>
<td>3</td>
<td>0.048</td>
<td>0.054</td>
<td>0.057</td>
</tr>
</tbody>
</table>

#### Panel A.3. Low Frequency (Flow Sampling)

<table>
<thead>
<tr>
<th>$h_{LF}$ \ $r_{LF}$</th>
<th>$T_L = 40$</th>
<th>$T_L = 80$</th>
<th>$T_L = 120$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.055</td>
<td>0.056</td>
<td>0.060</td>
</tr>
<tr>
<td>2</td>
<td>0.051</td>
<td>0.051</td>
<td>0.060</td>
</tr>
<tr>
<td>3</td>
<td>0.045</td>
<td>0.051</td>
<td>0.058</td>
</tr>
</tbody>
</table>
Table 6: Rejection Frequencies of Max Test for Low-to-High Granger Causality (Continued)

<table>
<thead>
<tr>
<th>Panel B. Decaying Causality ($c_j = 0.3/j, j = 1, \ldots, 12$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel B.1. Mixed Frequency</strong></td>
</tr>
<tr>
<td>$T_L = 40$</td>
</tr>
<tr>
<td>$h_{MF \backslash T_{MF}}$</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Panel B.2. Low Frequency (Stock Sampling)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_L = 40$</td>
</tr>
<tr>
<td>$h_{LF \backslash T_{LF}}$</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Panel B.3. Low Frequency (Flow Sampling)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_L = 40$</td>
</tr>
<tr>
<td>$h_{LF \backslash T_{LF}}$</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Panel C. Lagged Causality ($c_j = 0.4 \times I(j = 12), j = 1, \ldots, 12$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel C.1. Mixed Frequency</strong></td>
</tr>
<tr>
<td>$T_L = 40$</td>
</tr>
<tr>
<td>$h_{MF \backslash T_{MF}}$</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Panel C.2. Low Frequency (Stock Sampling)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_L = 40$</td>
</tr>
<tr>
<td>$h_{LF \backslash T_{LF}}$</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Panel C.3. Low Frequency (Flow Sampling)</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>$T_L = 40$</td>
</tr>
<tr>
<td>$h_{LF \backslash T_{LF}}$</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</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>
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 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 1: Time Series Plot of U.S. Interest Rates and Real GDP Growth
Note: This figure plots asymptotic \( p \)-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. The mixed frequency tests deal with weekly interest rate spread and quarterly GDP growth, while the 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 below a nominal size (say 0.05 or 0.1), then it means that there exists significant causality from interest rate spread to GDP in that subsample.

Figure 2: Asymptotic \( p \)-values for Causality from Interest Rate Spread to GDP Growth
Technical Appendices

A Double Time Indices

Throughout this paper we 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, on the other hand, 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**

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

(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 \( m \left\lfloor \frac{j}{m} \right\rfloor + j \in \{1, \ldots, m\} \) when \( j \leq 0 \), and \( j - m \left\lfloor \frac{j}{m} \right\rfloor \in \{1, \ldots, m\} \) 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**

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

(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 non-positive or larger than \( m \), but such a case is covered by (A.1).

B Proof of Theorem 2.1

Recall that the parsimonious regression model \( j \) given in (2.5) is written as

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

or in matrix form (2.6)

\[
x_L(\tau_L) = x_j(\tau_L - 1)\theta_j + u_{L,j}(\tau_L),
\]

where \( x_j(\tau_L - 1) = [x_L(\tau_L - 1), \ldots, x_L(\tau_L - q), x_H(\tau_L - 1, m + 1 - j)]' = [X_L^q(\tau_L - 1)', x_H(\tau_L - 1, m + 1 - j)]' \).

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:

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

(B.1)
Recall that the DGP in matrix form is given in (2.3):

\[
x_L(\tau_L) = \left[ x_L(\tau_L - 1), \ldots, x_L(\tau_L - p) \right] + \varepsilon_L(\tau_L).
\]

Substituting this into (B.1), we get

\[
\theta^*_j = \left[ E \left[ x_j(\tau_L - 1) x_j(\tau_L - 1)' \right] \right]^{-1} E \left[ x_j(\tau_L - 1) \left\{ X_L(\tau_L - 1)' a + X_H(\tau_L - 1)' b + \varepsilon_L(\tau_L) \right\} \right]
= \left[ E \left[ x_j(\tau_L - 1) x_j(\tau_L - 1)' \right] \right]^{-1} \left\{ E \left[ x_j(\tau_L - 1) X_L(\tau_L - 1)' a + E \left[ x_j(\tau_L - 1) X_H(\tau_L - 1)' b \right] \right] \right),
\]

where the second equality holds from the mds. assumption of \( \varepsilon_L \). 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

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

and hence

\[
E \left[ x_j(\tau_L - 1) X_L(\tau_L - 1)' \right] = E \left[ x_j(\tau_L - 1) x_j(\tau_L - 1)' \right] \left[ I_p 0_{(q-p+1) \times p} \right] = [I_p] 0_{(q-p+1) \times p}.
\]

Substituting (B.5) into (B.3), we obtain

\[
\theta^*_j = \begin{bmatrix} \alpha^*_{1,j} \\ \vdots \\ \alpha^*_{p,j} \\ \alpha^*_{p+1,j} \\ \vdots \\ \alpha^*_{q,j} \\ \beta^*_{j} \end{bmatrix} = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_p \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \left[ E \left[ x_j(\tau_L - 1) x_j(\tau_L - 1)' \right] \right]^{-1} \left[ E \left[ x_j(\tau_L - 1) X_H(\tau_L - 1)' \right] b \right].
\]

Finally, it is easy to express the pseudo-true value of \( \beta = [\beta_1, \ldots, \beta_h]' \), written as \( \beta^* \), by constructing an appropriate selection matrix \( R \) such that \( \beta^* = R \theta^* \), where \( \theta^* = [\theta^*_1, \ldots, \theta^*_h]' \). 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:

\[
R \overset{h \times (q+1)h}{=} = \begin{bmatrix} 0 & \ldots & 0 & 1 & \ldots & \ldots & 0 & 0 & \ldots & 0 \\ \vdots & \ddots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & 0 & \ldots & \ldots & 0 & 0 & \ldots & 1 \end{bmatrix}.
\]

C Proof of Theorem 2.2

We first show that \( b = 0_{pm \times 1} \Rightarrow \beta^* = 0_{h \times 1} \). Under Assumptions 2.1, 2.2, and 2.4, we can derive (B.6). If \( b = 0_{pm \times 1} \), then (B.6) implies that \( \beta^*_j = 0 \) for any \( j = 1, \ldots, h \). We thus have that \( \beta^* = 0_{h \times 1} \).

We now show that \( \beta^* = 0_{h \times 1} \Rightarrow b = 0_{pm \times 1} \), assuming that \( h \geq pm \). We pick the last row of (B.6). The
where As seen in (2.6), the parsimonious regression model \( j \) is trivially non-singular by Assumption 2.1 with positive definite error covariance matrix \( \Omega \). Now define

\[
\beta = \left[ \begin{array}{c}
\theta_j \\
\end{array} \right] = \alpha_j + \beta_j,
\]

having \( \beta^* = 0_{h \times 1} \) implies that \( n_j^{-1}d_j' \beta = 0 \) in view of (B.6). Since \( n_j \) is a nonzero finite scalar for any \( j = 1, \ldots, h \) by the non-singularity of \( E[x_j(\tau_L - 1)x_j(\tau_L - 1)'] \), it has to be the case that \( d_j' \beta = 0 \). Stacking these \( h \) equations, we have that

\[
\begin{bmatrix}
d_1' \\
\vdots \\
d_h'
\end{bmatrix} \times \beta = 0_{h \times 1} \quad \text{and hence} \quad b'D'Db = 0.
\]

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 \( D_{pm} = [d_1, \ldots, d_{pm}]' \), the first \( pm \) rows of \( D \), is of full column rank \( pm \) or equivalently non-singular. Equation (C.1) implies that

\[
D_{pm} = E[x_H(\tau_L - 1)x_H(\tau_L - 1)'] - E[x_H(\tau_L - 1)x_L(q)(\tau_L - 1)'] E[x_L(q)(\tau_L - 1)x_L(q)(\tau_L - 1)']^{-1} E[x_L(q)(\tau_L - 1)x_H(\tau_L - 1)'].
\]

Now define

\[
\Delta = E\left[ \begin{bmatrix}
x_L(q)(\tau_L - 1) \\
x_H(\tau_L - 1)
\end{bmatrix} \begin{bmatrix}
x_L(q)(\tau_L - 1)' \\
x_H(\tau_L - 1)'
\end{bmatrix} \right],
\]

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 \( E[x_L(q)(\tau_L - 1)x_L(q)(\tau_L - 1)'] \). Thus, by the classic argument of partitioned matrix inversion, \( D_{pm} \) is non-singular as desired.

**D Proof of Theorem 2.3**

As seen in (2.6), the parsimonious regression model \( j \) in matrix form is given by

\[
x_L(\tau_L) = x_j(\tau_L - 1)' \theta_j + u_L, j = 1, \ldots, h,
\]

where \( \theta_j = [\alpha_{1,j}, \ldots, \alpha_{q,j}, \beta_j]' \). We collect all parameters across the \( h \) models as \( \theta = [\theta_1', \ldots, \theta_h'] \).
Deriving the asymptotic distribution of the max test statistic \( T = \max_{1 \leq j \leq h} \left( \sqrt{T_L} w_{T_L,j} \beta_j \right)^2 \) under \( H_0 : b = 0_{m \times 1} \) can be done by deriving the asymptotic distribution of \( \sqrt{T_L} \beta \) under \( H_0 \), where \( \beta = [\beta_1, \ldots, \beta_h]' \). Working on \( \sqrt{T_L} \beta \) directly is rather cumbersome, so we work on \( R \times \sqrt{T_L} (\hat{\beta} - \theta_0) \), where the selection matrix \( R \) is such that \( \beta = R \theta \) as shown in (B.7). 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 \( \sqrt{T_L} \beta = R \times \sqrt{T_L} (\hat{\theta} - \theta_0) \). The most convenient choice satisfying this condition is \( \theta_0 = \nu_h \otimes \theta_0 \) with \( \theta_0 = [a_1, \ldots, a_p, 0_{1 \times (q-p+1)}]' \), where \( \nu_h \) 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 \( \sqrt{T_L} (\hat{\beta} - \theta_0) \) under \( H_0 \). By the construction of \( \theta_0 \), the DGP appearing in (2.3), \( x_L (\tau_L) = X_L (\tau_L - 1)' a + X_H (\tau_L - 1)' b + \epsilon_L (\tau_L) \), 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
\[
\sqrt{T_L} (\hat{\theta}_j - \theta_0) = \sqrt{T_L} \left[ \sum_{\tau_L = 1}^{T_L} x_j (\tau_L - 1) \tau_L (\tau_L - 1)' \right]^{-1} \sum_{\tau_L = 1}^{T_L} x_j (\tau_L - 1) x_L (\tau_L) - \sqrt{T_L} \theta_0
\]
\[
= \sqrt{T_L} \left[ \sum_{\tau_L = 1}^{T_L} x_j (\tau_L - 1) \tau_L (\tau_L - 1)' \right]^{-1} \sum_{\tau_L = 1}^{T_L} x_j (\tau_L - 1) x_j (\tau_L - 1)' \theta_0 + \epsilon_L (\tau_L) - \sqrt{T_L} \theta_0
\]
\[
= \sqrt{T_L} \left[ \sum_{\tau_L = 1}^{T_L} x_j (\tau_L - 1) \tau_L (\tau_L - 1)' \right]^{-1} \sum_{\tau_L = 1}^{T_L} x_j (\tau_L - 1) \epsilon_L (\tau_L)
\]
\[
= \left[ E [x_j (\tau_L - 1) x_j (\tau_L - 1)'] \right]^{-1} \frac{1}{\sqrt{T_L}} \sum_{\tau_L = 1}^{T_L} x_j (\tau_L - 1) \epsilon_L (\tau_L) + o_p(1),
\]
\[
= \Gamma^{-1} \frac{1}{\sqrt{T_L}} \sum_{\tau_L = 1}^{T_L} x_j (\tau_L - 1) \epsilon_L (\tau_L) + o_p(1),
\]

where the last equality follows just by definition in (B.6). Using (D.2), we now deduce the asymptotic distribution of \( \sqrt{T_L} (\hat{\theta} - \theta_0) \). 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 \( \lambda' \times \sqrt{T_L} (\hat{\theta} - \theta_0) \). We have that
\[
\lambda' \times \sqrt{T_L} (\hat{\theta} - \theta_0) = \sum_{j=1}^{h} \lambda_j' \times \sqrt{T_L} (\hat{\beta}_j - \theta_0)
\]
\[
= \sum_{j=1}^{h} \lambda_j' \Gamma^{-1} \frac{1}{\sqrt{T_L}} \sum_{\tau_L = 1}^{T_L} x_j (\tau_L - 1) \epsilon_L (\tau_L) + o_p(1)
\]
\[
= \frac{1}{\sqrt{T_L}} \sum_{\tau_L = 1}^{T_L} \sum_{j=1}^{h} \lambda_j' \Gamma^{-1} x_j (\tau_L - 1) \epsilon_L (\tau_L) + o_p(1),
\]
\[
\equiv \lambda' \epsilon (\tau_L - 1, \lambda)
\]

where the second equality follows from (D.2).

Define \( \Gamma_{j,i} = E[x_j (\tau_L - 1) x_i (\tau_L - 1)'] \), then we have that
\[
E \left[ X (\tau_L - 1, \lambda)^2 \right] = \sum_{j=1}^{h} \sum_{i=1}^{h} \lambda_j' \Gamma^{-1}_{j,j} \Gamma^{-1}_{i,i} \lambda_i = \lambda' \Sigma \lambda,
\]
\[
E \left[ \epsilon (\tau_L - 1, \lambda)^2 \right] = \sum_{j=1}^{h} \sum_{i=1}^{h} \lambda_j' \Gamma^{-1}_{j,j} \Gamma^{-1}_{i,i} \lambda_i = \lambda' \Sigma \lambda,
\]
\[
(4.4)
\]

60
where

\[ \Sigma = \begin{bmatrix} \Sigma_{1,1} & \cdots & \Sigma_{1,h} \\ \vdots & \ddots & \vdots \\ \Sigma_{h,1} & \cdots & \Sigma_{h,h} \end{bmatrix}. \]

Using (D.4), we apply a central limit theorem to (D.3) in order to obtain that \( \lambda' \times \sqrt{T_L} (\hat{\theta} - \theta_0) \overset{d}{\rightarrow} N(0, \lambda' \Sigma \lambda) \). By the Cramér-Wold theorem, we get that \( \sqrt{T_L} (\hat{\theta} - \theta_0) \overset{d}{\rightarrow} N(0_{(q+1)h \times 1}, \sigma_L^2 \Sigma) \). Hence,

\[
\sqrt{T_L} \widetilde{W}_{TL} \hat{\beta} = \sqrt{T_L} \widetilde{W}_{TL} R(\hat{\theta} - \theta_0) \overset{d}{\rightarrow} N(0_{h \times 1}, \sigma_L^2 WR \Sigma R'W). \tag{D.5}
\]

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

### E  Proof of Theorem 2.4

Recall that the max test statistic is given by \( T = \max_{1 \leq j \leq h} (\sqrt{T_L} w_{TL,j} \hat{\beta}_j)^2 \). Since \( \beta_j^* \) is defined as the pseudo-true value of \( \beta_j \), we have that \( \hat{\beta}_j \overset{p}{\rightarrow} \beta_j^* \) by construction. Hence, \( T \overset{p}{\rightarrow} \infty \Leftrightarrow \beta^* \neq 0_{h \times 1} \). Given \( h \geq pm \), Theorem 2.2 ensures that \( b \neq 0_{pm \times 1} \Rightarrow \beta^* \neq 0_{h \times 1} \). Therefore, the test statistic \( T \) diverges in probability under a general alternative hypothesis \( H_1 : b \neq 0_{pm \times 1} \).

### F  Proof of Theorem 2.5

Recall the parsimonious regression models (2.23):

\[
x_L(\tau_L) = \sum_{k=1}^{p} \alpha_{k,j} x_L(\tau_L - k) + \sum_{k=1}^{pm} \beta_{k,j} x_H(\tau_L - 1, m + 1 - k) + \gamma_j x_H(\tau_L + 1, j) + u_{L,j}(\tau_L),
\]

Instruments: \{all \( p + pm + 1 \) regressors in model \( j, x_H(\tau_L, 1), \ldots, x_H(\tau_L, m) \)\}.

To rewrite them in a matrix form, define

\[
\begin{pmatrix} x_L(\tau_L - 1) \\
\vdots \\
x_L(\tau_L - p) \\
x_H(\tau_L - 1, m + 1 - 1) \\
\vdots \\
x_H(\tau_L - 1, m + 1 - pm) \\
x_H(\tau_L + 1, j) \end{pmatrix}_{n \times 1}, \quad \begin{pmatrix} \alpha_{1,j} \\
\vdots \\
\alpha_{p,j} \\
\beta_{1,j} \\
\vdots \\
\beta_{pm,j} \\
\gamma_j \end{pmatrix}_{(n+1) \times 1}, \text{ and } \begin{pmatrix} \bar{x}_j(\tau_L) \\
x_H(\tau_L, 1) \\
\vdots \\
x_H(\tau_L, m) \end{pmatrix}_{(n+1)m \times 1},
\]

where \( n = p + pm + 1 \). \( \bar{x}_j(\tau_L) \) is a vector of all explanatory variables while \( \theta_j \) is a vector of all parameters in model \( j \). \( z_j(\tau_L) \) is a vector of instruments consisting of all \( n \) explanatory variables and \( m \) contemporaneous high frequency observations of \( x_H \).

Using these notations, model (2.23) can be rewritten as

\[
x_L(\tau_L) = \bar{x}_j(\tau_L)' \theta_j + u_{L,j}(\tau_L) \text{ with instruments } z_j(\tau_L), \quad j = 1, \ldots, h.
\]
To derive the GIVE for $\theta_j$, define sample moments

$$
\hat{S}_j^{(n+m)\times n} = \frac{1}{T_L} \sum_{\tau_L=1}^{T_L} z_j(\tau_L) x_j(\tau_L)',
$$

$$
\hat{\Sigma}_j^{(n+m)\times (n+m)} = \frac{1}{T_L} \sum_{\tau_L=1}^{T_L} z_j(\tau_L) z_j(\tau_L)'.
$$

Using these matrices, the GIVE for $\theta_j$ is given by

$$
\hat{\theta}_j = \left( \hat{S}_j' \hat{\Sigma}_j^{-1} \hat{S}_j \right)^{-1} \hat{S}_j' \hat{\Sigma}_j^{-1} \hat{\theta}_j.
$$

To derive the limit distribution of $\hat{\theta}_j$ under $H_0$, consider a hypothesized value:

$$
\theta_{0,j} = [\alpha_{1,j}, \ldots, \alpha_{p,j}, \beta_{1,j}, \ldots, \beta_{pm,j}, 0]^\prime; 
$$

where the asterisk signifies the pseudo-true value. We do not know the pseudo-true values of $\alpha$’s and $\beta$’s in practice, but that does not matter since we are only interested in the zero hypothesis with respect to $\gamma_j$. Eq. (F.1) is the most convenient choice of a hypothesized value in terms of mathematical derivation.

Under $H_0: x_L \Rightarrow x_H$, we have that

$$
\hat{s}_j = \frac{1}{T_L} \sum_{\tau_L=1}^{T_L} z_j(\tau_L) [x_j(\tau_L)\theta_{0,j} + \epsilon_L(\tau_L)] = \hat{S}_j \theta_{0,j} + \frac{1}{T_L} \sum_{\tau_L=1}^{T_L} z_j(\tau_L)\epsilon_L(\tau_L)
$$

and thus

$$
\sqrt{T_L}(\hat{\theta}_j - \theta_{0,j}) = \left( \hat{S}_j' \hat{\Sigma}_j^{-1} \hat{S}_j \right)^{-1} \hat{S}_j' \hat{\Sigma}_j^{-1} \hat{s}_j = \frac{1}{\sqrt{T_L}} \sum_{\tau_L=1}^{T_L} z_j(\tau_L)\epsilon_L(\tau_L), \quad j = 1, \ldots, h. 
$$

We have that

$$
\hat{S}_j \xrightarrow{p} E[z_j(\tau_L)x_j(\tau_L)'] \equiv S_j \quad \text{and} \quad \hat{\Sigma}_j \xrightarrow{p} E[z_j(\tau_L)z_j(\tau_L)'] \equiv \Sigma_j. 
$$

Using (F.3), we apply the Cramér-Wold theorem to (F.2) in order to combine all $h$ parsimonious regression models. To this end, define $\lambda = [x_1', \ldots, x_h'] \in \mathbb{R}^{nh}$ as well as

$$
\hat{\theta}_n = \begin{bmatrix} \hat{\theta}_1 \\ \vdots \\ \hat{\theta}_h \end{bmatrix} \quad \text{and} \quad \theta_{0} = \begin{bmatrix} \theta_{0,1} \\ \vdots \\ \theta_{0,h} \end{bmatrix}. 
$$
Then we have that

$$
X' \sqrt{T_L} (\theta - \theta_0) = \sum_{j=1}^{h} X_j' \sqrt{T_L} (\theta_j - \theta_{0,j}) \\
= \sum_{j=1}^{h} X_j' \left[ \left( S_j' \Sigma_j^{-1} S_j \right)^{-1} S_j' \Sigma_j^{-1} \mathbf{1} \mathbf{1}' \mathbf{1} \right] \times \frac{1}{\sqrt{T_L}} \sum_{r=1}^{T_L} z_j (\tau_L) \epsilon_L (\tau_L) \\
= \frac{1}{\sqrt{T_L}} \sum_{r=1}^{T_L} \left( \sum_{j=1}^{h} X_j' \left( S_j' \Sigma_j^{-1} S_j \right)^{-1} S_j' \Sigma_j^{-1} z_j (\tau_L) \right) \epsilon_L (\tau_L) + o_p (1).
$$

Define

$$
\Sigma_{j,i} = E [z_j (\tau_L) z_i (\tau_L)'],
$$

then we have that

$$
E \left[ Z(\tau_L)^2 \right] = \sum_{j=1}^{h} \sum_{i=1}^{h} X_j' \left( S_j' \Sigma_j^{-1} S_j \right)^{-1} S_j' \Sigma_j^{-1} \Sigma_{j,i} \Sigma_{j,i}^{-1} S_i \left( S_i' \Sigma_i^{-1} S_i \right)^{-1} \lambda_i \equiv \Psi_{j,i}: n \times n
$$

where

$$
\Psi = \begin{bmatrix} \Psi_{1,1} & \cdots & \Psi_{1,h} \\ \vdots & \ddots & \vdots \\ \Psi_{h,1} & \cdots & \Psi_{h,h} \end{bmatrix}_{nh \times nh}.
$$

Applying a central limit theorem to (F.4) using (F.5), we get that

$$
X' \sqrt{T_L} (\bar{\theta} - \theta_0) \overset{d}{\rightarrow} N (0, \lambda' \Psi \lambda).
$$

(F.5)

Then by the Cramér-Wold theorem, we obtain that

$$
\sqrt{T_L} (\theta - \theta_0) \overset{d}{\rightarrow} N (0_{nh \times 1}, \sigma^2_L \Psi).
$$

(F.6)

Define

$$
\hat{\gamma} \overset{h \times 1}{= \begin{bmatrix} \hat{\gamma}_1 \\ \vdots \\ \hat{\gamma}_h \end{bmatrix}}, \quad R \overset{h \times nh}{= \begin{bmatrix} O_{1 \times (n-1)} & \cdots & O_{1 \times (n-1)} \\ \vdots & \ddots & \vdots \\ O_{1 \times (n-1)} & \cdots & O_{1 \times (n-1)} \end{bmatrix}}_{h \times nh}, \quad W \overset{h \times h}{= \begin{bmatrix} w_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_h \end{bmatrix}}.
$$

(F.7)

\( R \) is a selection matrix choosing \( \gamma \)'s out of the entire parameter vector \( \theta \), while \( W \) is a diagonal matrix having the \( L^2 \) limit of the weighting scheme \( wT_L \). Equations (F.6) and (F.7) imply that

$$
\sqrt{T_L} W \hat{\gamma} \overset{d}{\rightarrow} N (0_{h \times 1}, \sigma^2_L \Psi R R' W) \quad \Rightarrow U
$$

under \( H_0 : x_L \overset{d}{\rightarrow} x_H \).
G  Proof of Theorem 3.1

This proof is identical to the proof for Theorem 2.3 except for that we impose \( H_1^' : b = (1/\sqrt{TL}) \nu \) instead of \( H_0 : b = 0_{pm \times 1} \) when we derive (D.2). Recall (3.1), the DGP under \( H_1^' : \)

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

where the third equality follows from (B.4). Based on this equation, (D.2) should be modified as follows.

\[
\sqrt{TL}(\hat{\theta}_j - \theta_0) = \sqrt{TL} \left[ \sum_{\tau_L = 1}^{T_L} \mathbf{x}_j(\tau_L - 1) \mathbf{x}_j(\tau_L - 1)' \right]^{-1} \sum_{\tau_L = 1}^{T_L} \mathbf{x}_j(\tau_L - 1)x_L(\tau_L) - \sqrt{TL} \theta_0
\]

\[
= \sqrt{TL} \left[ \sum_{\tau_L = 1}^{T_L} \mathbf{x}_j(\tau_L - 1) \mathbf{x}_j(\tau_L - 1)' \right]^{-1} \sum_{\tau_L = 1}^{T_L} \mathbf{x}_j(\tau_L - 1)x_L(\tau_L) - \sqrt{TL} \theta_0
\]

\[
= \sqrt{TL} \left[ \sum_{\tau_L = 1}^{T_L} \mathbf{x}_j(\tau_L - 1) \mathbf{x}_j(\tau_L - 1)' \right]^{-1} \sum_{\tau_L = 1}^{T_L} \mathbf{x}_j(\tau_L - 1)x_L(\tau_L) - \sqrt{TL} \theta_0
\]

\[
= \left[ E[\mathbf{x}_j(\tau_L - 1) \mathbf{x}_j(\tau_L - 1)'] \right]^{-1} E[\mathbf{x}_j(\tau_L - 1)X_H(\tau_L - 1)'] \nu
\]

\[
+ \left[ E[\mathbf{x}_j(\tau_L - 1) \mathbf{x}_j(\tau_L - 1)'] \right]^{-1} \frac{1}{\sqrt{TL}} \sum_{\tau_L = 1}^{T_L} \mathbf{x}_j(\tau_L - 1) \epsilon_L(\tau_L) + o_p(1),
\]

where the last equality follows simply from the definitions in (B.6).

Repeating (D.3), we get

\[
\lambda' \times \sqrt{TL}(\hat{\theta}_j - \theta_0) = \sum_{j=1}^{h} \lambda_j' \Gamma_{j, j}^{-1} C_j \nu + \frac{1}{\sqrt{TL}} \sum_{\tau_L = 1}^{T_L} X(\tau_L - 1, \lambda) \epsilon_L(\tau_L) + o_p(1)
\]

\[
\overset{d}{\rightarrow} N(\lambda' \mathbf{u}, \lambda' (\tau_L^2 \Sigma) \lambda),
\]

where

\[
\mathbf{u}_{(q+1)h \times 1} \equiv \begin{bmatrix} \Gamma_{1, 1}^{-1} C_1 \\ \vdots \\ \Gamma_{h, h}^{-1} C_h \end{bmatrix} \nu.
\]

64
By the Cramér-Wold theorem, we have that $\sqrt{T} \left( \hat{\theta} - \theta_0 \right) \xrightarrow{d} N(\mu, \sigma^2 \Sigma)$. Now repeat (D.5) to get
\[
\sqrt{T} W_{TL} \hat{\beta} = \sqrt{T} W_{TL} R(\hat{\theta} - \theta_0) \xrightarrow{d} N(\mu^T R_{u} \cdot w, \sigma^2 W R \Sigma R^{T} W).
\]

Recall that the max test statistic is given by $T = \max_{1 \leq s \leq k} \left( \sqrt{T} w_{TL, s} \hat{\beta} \right)^2$. Hence we have that $T \xrightarrow{d} \max_{1 \leq s \leq k} \mathcal{M}_s^2$, where $\mathcal{M} = \{ \mathcal{M}_1, \ldots, \mathcal{M}_k \}'$ is a vector-valued random variable drawn from $N(\mu, \Sigma)$.

### H Proof of Theorem 3.2

Recall that $\mathbf{Y}_k$, autocovariance matrix of the mixed frequency vector $\mathbf{X}(\tau_L)$ of order $k$, is constructed as follows:

\[
\mathbf{Y}_k \equiv E[\mathbf{X}(\tau_L) \mathbf{X}(\tau_L - k)']
\]

\[
= \begin{bmatrix}
E[x_H(\tau_L, 1)x_H(\tau_L - k, 1)] & \cdots & E[x_H(\tau_L, 1)x_H(\tau_L - k, m)] \\
\vdots & \ddots & \vdots \\
E[x_L(\tau_L, k)x_H(\tau_L - k, 1)] & \cdots & E[x_L(\tau_L, k)x_H(\tau_L - k, m)]
\end{bmatrix}
\]  \hspace{1cm} (H.1)

for $k \geq 0$, and $\mathbf{Y}_k = \mathbf{Y}_{-k}'$ for $k < 0$. Recall that $\mathbf{Y}_k$ 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 $\mathbf{Y}_k$ as given here, and characterize $\mathbf{G}_{j,i}$ and $C_j$ in terms of $\mathbf{Y}_k$ for $j, i \in \{1, \ldots, h\}$.

We begin with $\mathbf{G}_{j,i}$. By the covariance stationarity of $\mathbf{X}(\tau_L)$, we have that

\[
\mathbf{G}_{j,i} \equiv E[x_j(\tau_L - 1)x_i(\tau_L - 1)'] = E[x_j(\tau_L)x_i(\tau_L)']
\]

\[
= E \begin{bmatrix}
x_L(\tau_L) \\
x_L(\tau_L - (q - 1)) \\
x_H(\tau_L, m + 1 - j)
\end{bmatrix} \begin{bmatrix}
x_L(\tau_L) & \cdots & x_L(\tau_L - (q - 1)) & x_H(\tau_L, m + 1 - i)
\end{bmatrix}.
\]  \hspace{1cm} (H.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 $\mathbf{G}_{j,i}$ is identical to which element of $\mathbf{Y}_k$. To ensure that the second argument of $x_H$ lies in $\{1, \ldots, m\}$, we use the high frequency simplification (A.1):

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

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

\[
= x_H(\tau_L - f(j), g(j)), \hspace{1cm} (H.3)
\]

where the last equality follows simply from the definitions $f(j) = \lfloor (j - m)/m \rfloor$ 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 (H.3) into (H.2), $\mathbf{G}_{j,i}$ can be rewritten as follows.

\[
\begin{bmatrix}
E[x_L(\tau_L)x_L(\tau_L)] & \cdots & E[x_L(\tau_L)x_L(\tau_L - (q - 1))] \\
\vdots & \ddots & \vdots \\
E[x_L(\tau_L - (q - 1))x_L(\tau_L)] & \cdots & E[x_L(\tau_L - (q - 1))x_L(\tau_L - (q - 1))]
\end{bmatrix}
\]

\[
\begin{bmatrix}
E[x_H(\tau_L - f(j), g(j))x_L(\tau_L)] & \cdots & E[x_H(\tau_L - f(j), g(j))x_L(\tau_L - (q - 1))] \\
\vdots & \ddots & \vdots \\
E[x_H(\tau_L - f(j), g(j))x_L(\tau_L)] & \cdots & E[x_H(\tau_L - f(j), g(j))x_L(\tau_L - (q - 1))]
\end{bmatrix}.
\]  \hspace{1cm} (H.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:

- If \( f(j) \geq q - 1 \) or \( f(j) < q - 1 \). For the first case, we have that

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

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 (H.1). For the second case that \( f(j) < q - 1 \), we get

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

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

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

The second and fourth equalities follow from covariance stationarity, while the third equality follows from (H.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 (H.4), we obtain

\[
\Gamma_{j,i} = \begin{bmatrix}
\Upsilon_{1-1}(K, K) & \ldots & \Upsilon_{1-q}(K, K) & \Upsilon_{f(i)}(g(i), K) \\
\vdots & \ddots & \vdots & \vdots \\
\Upsilon_{q-1}(K, K) & \ldots & \Upsilon_{q-q}(K, K) & \Upsilon_{(q-1)-f(i)}(g(i), K) \\
\Upsilon_{f(j)}(K, g(j)) & \ldots & \Upsilon_{f(j)-(q-1)}(K, g(j)) & \Upsilon_{f(j)-(q-1)}(K, g(j))
\end{bmatrix}.
\]

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

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

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

(6.6)

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

\[
C_j = \begin{bmatrix}
E[x_L(\tau_L)X_H(\tau_L - f(1), g(1))] & \ldots & E[x_L(\tau_L)X_H(\tau_L - f(pm), g(pm))]
\end{bmatrix}
\]

\[
= \begin{bmatrix}
E[x_L(\tau_L - (q - 1))X_H(\tau_L - f(1), g(1))] & \ldots & E[x_L(\tau_L - (q - 1))X_H(\tau_L - f(pm), g(pm))]
\end{bmatrix}
\]

(6.7)

We now map each element of \( C_j \) to an appropriate element of \( \Upsilon_k \). 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 (H.4), this quantity is equal to the \((q + 1, q)\) element of \( \Gamma_{pm,i} \) with an arbitrary \( i \). We already know from (H.5) that the \((q + 1, q)\) element of \( \Gamma_{pm,i} \) is equal to
The definition of $X_L(T_L - 1)$ in (I.1) implies that $X_L(T_L - 1) = [x_L(T_L - 1), \ldots, x_L(T_L - p)]' = [I_p, 0_{p \times (q - p + h)}] x(T_L - 1)$. Hence, the DGP (B.2) can be rewritten as follows.

$$x_L(T_L) = X_L(T_L - 1)'a + X_H(T_L - 1)'b + \epsilon_L(T_L)$$

$$= x(T_L - 1)' \begin{bmatrix} I_p \\ 0_{(q-p+h) \times p} \end{bmatrix} a + X_H(T_L - 1)'b + \epsilon_L(T_L)$$

$$= x(T_L - 1)' \theta_0 + X_H(T_L - 1)'b + \epsilon_L(T_L),$$

where $\theta_0 = [a', 0_{(q-p+h)}]'$.

We formulate a Wald statistic $W$ with respect to $H_0 : b = 0_{pm \times 1}$ and find its asymptotic distribution under $H^1_0 : b = (1/\sqrt{T_L})\nu$. First, the OLS estimator associated with model (I.1) is given by $\hat{\theta} = (\sum_{T_L=1}^{T_L} x(T_L - 1)x(T_L - 1)' )^{-1} \sum_{T_L=1}^{T_L} x(T_L - 1)x(T_L - 1)'$. Second, define two population covariance terms:

$$\sum_{(q+h) \times (q+h)} = E[x(T_L - 1)x(T_L - 1)'] \quad \text{and} \quad \sum_{(q+h) \times pm} = E[x(T_L - 1)X_H(T_L - 1)'].$$

where $x(T_L - 1)$ and $X_H(T_L - 1)$ are defined in (I.1) and (B.2), respectively.
Using (I.2) as well as (I.3) and imposing $H_1^1$, we have that

$$\sqrt{T_L}(\hat{\theta} - \theta_0)$$

$$= \sqrt{T_L} \left[ \sum_{\tau_L=1}^{T_L} x(\tau_L - 1)x(\tau_L - 1)' \right]^{-1} \frac{T_L}{\sqrt{T_L}} \sum_{\tau_L=1}^{T_L} x(\tau_L - 1)\epsilon_L(\tau_L) - \sqrt{T_L}\theta_0$$

$$= \sqrt{T_L} \left[ \sum_{\tau_L=1}^{T_L} x(\tau_L - 1)x(\tau_L - 1)' \right]^{-1} \frac{T_L}{\sqrt{T_L}} \sum_{\tau_L=1}^{T_L} x(\tau_L - 1) \left[ x(\tau_L - 1)'\theta_0 + X_H(\tau_L - 1) \times \frac{1}{\sqrt{T_L}} \nu + \epsilon_L(\tau_L) \right] - \sqrt{T_L}\theta_0$$

$$= \left[ \frac{1}{T_L} \sum_{\tau_L=1}^{T_L} x(\tau_L - 1)x(\tau_L - 1)' \right]^{-1} \left[ \frac{1}{T_L} \sum_{\tau_L=1}^{T_L} x(\tau_L - 1)X_H(\tau_L - 1)' \right] \nu$$

$$+ \left[ \frac{1}{T_L} \sum_{\tau_L=1}^{T_L} x(\tau_L - 1)x(\tau_L - 1)' \right]^{-1} \times \frac{1}{\sqrt{T_L}} \sum_{\tau_L=1}^{T_L} x(\tau_L - 1)\epsilon_L(\tau_L)$$

$$= \Gamma^{-1}C\nu + \Gamma^{-1} \times \frac{1}{\sqrt{T_L}} \sum_{\tau_L=1}^{T_L} x(\tau_L - 1)\epsilon_L(\tau_L) + o_p(1)$$

$$\xrightarrow{d} N(\Gamma^{-1}C\nu, \sigma^2_L \Gamma^{-1}).$$

Construct an $h \times (q + h)$ selection matrix $\bar{R}$ such that $\beta = \bar{R}\hat{\theta}$. Specifically, we let $\bar{R} = [0_h \times I_h]$. Then we have that $\sqrt{T_L}\beta = \bar{R} \times \sqrt{T_L}(\hat{\theta} - \theta_0) \xrightarrow{d} N(\bar{R}\Gamma^{-1}C\nu, \sigma^2_L \bar{R}\Gamma^{-1}\bar{R})$.

As usual, the Wald statistic $W$ is defined as $W = T_L \beta'(\sigma^2_L \bar{R}\Gamma^{-1}\bar{R})^{-1} \beta$, where $\bar{\Gamma}$ is a consistent estimator for $\Gamma$. We have that $W \xrightarrow{d} \chi^2_N(k), \nu$.

The asymptotic distribution of $W$ under $H_0 : b = 0_{pm \times 1}$ can be derived simply by letting $\nu = 0_{pm \times 1}$ above. Specifically, $W \xrightarrow{d} \chi^2_N(k)$ under $H_0$.

**J Proof of Theorem 3.4**

We consider $\Gamma$ first. Under Assumptions 2.1 and 2.2, we have that

$$\Gamma = \begin{bmatrix} G_{UL} & G_{UR} \\ G_{UR} & G_{LR} \end{bmatrix} = E[x(\tau_L - 1)x(\tau_L - 1)'] = E[x(\tau_L)x(\tau_L)']$$

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

Comparing (J.1) with (H.2) and (H.5), we get the following results immediately. First, $G_{UL}$ is identical to the upper-left block of $\Gamma_{J,:}$:

$$G_{UL} = \begin{bmatrix} \Upsilon_{1-1}(K, K) & \ldots & \Upsilon_{1-q}(K, K) \\ \vdots & \ddots & \vdots \\ \Upsilon_{q-1}(K, K) & \ldots & \Upsilon_{q-q}(K, K) \end{bmatrix}.$$
Second, $\Gamma_{UR}$ is aligned upper-right blocks of $\Gamma_{j,1}, \ldots, \Gamma_{j,h}$:

$$
\Gamma_{UR} = \begin{bmatrix}
\Upsilon_{f(1)}(g(1), K) & \ldots & \Upsilon_{f(h)}(g(h), K) \\
\vdots & \ddots & \vdots \\
\Upsilon_{(q-1)-f(1)}(g(1), K) & \ldots & \Upsilon_{(q-1)-f(h)}(g(h), K)
\end{bmatrix}.
$$

Third, $\Gamma_{LR}$ is aligned lower-right blocks of $\Gamma_{j,i}$ with $j, i \in \{1, \ldots, h\}$:

$$
\Gamma_{LR} = \begin{bmatrix}
\Upsilon_{f(1)-f(1)}(g(1), g(1)) & \ldots & \Upsilon_{f(1)-f(h)}(g(h), g(1)) \\
\vdots & \ddots & \vdots \\
\Upsilon_{f(h)-f(1)}(g(1), g(h)) & \ldots & \Upsilon_{f(h)-f(h)}(g(h), g(h))
\end{bmatrix}.
$$

We now consider $C$. Under Assumptions 2.1 and 2.2, we have that

$$
C = \begin{bmatrix} C_U \\ C_L \end{bmatrix} = E[\mathbf{x}_L(\tau_L - 1) \mathbf{X}_L(\tau_L - 1)^\prime] = E[\mathbf{x}_L(\tau_L) \mathbf{X}_L(\tau_L)^\prime]
$$

(3)

Comparing (3) with (H.6) and (H.8), we get the following results immediately. First, $C_U$ is identical to the first block of $C_j$:

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

Second, $C_L$ is aligned second blocks of $C_1, \ldots, C_h$:

$$
C_L = \begin{bmatrix}
\Upsilon_{f(1)-f(1)}(g(1), g(1)) & \ldots & \Upsilon_{f(1)-f(pm)}(g(pm), g(1)) \\
\vdots & \ddots & \vdots \\
\Upsilon_{f(h)-f(1)}(g(1), g(h)) & \ldots & \Upsilon_{f(h)-f(pm)}(g(pm), g(h))
\end{bmatrix}.
$$

# Proof of Theorem 3.5

This proof is almost identical to the proof of Theorem 3.1 shown in Appendix G. The only difference is that we worked on the mixed frequency parsimonious regression model in Theorem 3.1, while we work on the low frequency parsimonious regression model here in Theorem 3.5. The regressors in the former are $x_j(\tau_L - 1) = [x_L(\tau_L - 1), \ldots, x_L(\tau_L - q), x_h(\tau_L - 1, m + 1 - j)]^\prime$, while the regressors in the latter are $\mathbf{x}_L(\tau_L - 1) = [x_L(\tau_L - 1), \ldots, x_L(\tau_L - q), x_h(\tau_L - j)]^\prime$. By replacing $x_j(\tau_L - 1)$ with $\mathbf{x}_L(\tau_L - 1)$, we can reuse Appendix G to derive the asymptotic distribution of the low frequency max test statistic $T_{LF}$ under the local alternative hypothesis $H_1^L : \mathbf{b} = (1/\sqrt{T_L})\nu$. Specifically, $T_{LF} \overset{d}{\rightarrow} \max_{1 \leq i \leq h} M_i^2$, where $M = [M_1, \ldots, M_h]^\prime$ is a vector-
valued random variable drawn from \( N(\mu, V) \). \( \mu \) and \( V \) are defined as follows.

\[
\mu \xrightarrow{h \times 1} \mathbf{W} \mathbf{R} \begin{bmatrix} \Sigma_{1,1} & \cdots & \Sigma_{1,h} \\ \vdots & \ddots & \vdots \\ \Sigma_{h,1} & \cdots & \Sigma_{h,h} \end{bmatrix} \nu, \quad V \xrightarrow{h \times h} \sigma^2 \mathbf{W} \mathbf{R}^\top \mathbf{W}, \quad C_j \xrightarrow{(q+1) \times pm} E[\mathbf{u}_j(\tau_L - 1)X_H(\tau_L - 1)].
\]

The upper-left block of \( \Gamma \) follows from (H.4) and (H.5). Hence, the lower-left block of \( \Gamma \) is given by

\[
\Sigma_{j,i} = \begin{bmatrix} \Sigma_{1,1} & \cdots & \Sigma_{1,h} \\ \vdots & \ddots & \vdots \\ \Sigma_{h,1} & \cdots & \Sigma_{h,h} \end{bmatrix}, \quad \Sigma_{j,i} = \Gamma_{j,i} \Gamma_{i,j}, \quad \Gamma_{i,j} = E[\mathbf{u}_i(\tau_L - 1)\mathbf{u}_j(\tau_L - 1)].
\]

### L Proof of Theorem 3.6

We consider \( \Sigma_{j,i} \) first.

\[
\Sigma_{j,i} = E[\mathbf{u}_i(\tau_L - 1)\mathbf{u}_j(\tau_L - 1)]
\]

\[
= E\begin{bmatrix} x_L(\tau_L - 1) \\ \vdots \\ x_L(\tau_L - q) \\ x_H(\tau_L - j) \end{bmatrix} \begin{bmatrix} x_L(\tau_L - 1) & \ldots & x_L(\tau_L - q) & x_H(\tau_L - j) \end{bmatrix}
\]

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

As seen in (H.2) and (H.5), the upper-left block of \( \Sigma_{j,i} \) can be characterized as

\[
\begin{bmatrix} \Upsilon_{1-1}(K, K) & \ldots & \Upsilon_{1-q}(K, K) \\ \vdots & \ddots & \vdots \\ \Upsilon_{q-1}(K, K) & \ldots & \Upsilon_{q-q}(K, K) \end{bmatrix}
\]

We next consider the lower-left block of \( \Sigma_{j,i} \). We have that

\[
E[x_H(\tau_L - j)x_L(\tau_L - k)] = E[\sum_{l=1}^{m} \delta_l x_H(\tau_L - j, l)x_L(\tau_L - k)] = \sum_{l=1}^{m} \delta_l E[x_H(\tau_L - j, l)x_L(\tau_L - k)] = \sum_{l=1}^{m} \delta_l \Upsilon_{j-1}(k, l), \text{ where the last equality follows from (H.4) and (H.5).} \]

Hence, the lower-left block of \( \Sigma_{j,i} \) is characterized as \( \sum_{l=1}^{m} \delta_l \Upsilon_{j-1}(K, l), \ldots, \sum_{l=1}^{m} \delta_l \Upsilon_{j-1}(K, l) \). The upper-right block of \( \Sigma_{j,i} \) is exactly analogous.

Finally, we consider the lower-right block of \( \Sigma_{j,i} \). We have that

\[
E[x_H(\tau_L - j)x_H(\tau_L - i)] = E[ \sum_{l=1}^{m} \delta_l x_H(\tau_L - j, l)x_H(\tau_L - i, k)] = \sum_{l=1}^{m} \sum_{k=1}^{m} \delta_l \delta_k E[x_H(\tau_L - j, l)x_H(\tau_L - i, k)] = \sum_{l=1}^{m} \sum_{k=1}^{m} \delta_l \delta_k \Upsilon_{j-1}(k, l), \text{ where the last equality follows from (H.4) and (H.5).} \]

Therefore, we conclude that

\[
\Sigma_{j,i} = \begin{bmatrix} \Upsilon_{1-1}(K, K) & \ldots & \Upsilon_{1-q}(K, K) & \sum_{l=1}^{m} \delta_l \Upsilon_{1-1}(K, l) \\ \vdots & \ddots & \vdots & \vdots \\ \Upsilon_{q-1}(K, K) & \ldots & \Upsilon_{q-q}(K, K) & \sum_{l=1}^{m} \delta_l \Upsilon_{q-1}(K, l) \\ \sum_{l=1}^{m} \delta_l \Upsilon_{j-1}(K, l) & \ldots & \sum_{l=1}^{m} \delta_l \Upsilon_{j-1}(K, l) & \sum_{l=1}^{m} \sum_{k=1}^{m} \delta_l \delta_k \Upsilon_{j-1}(k, l) \end{bmatrix}.
\]
We now consider $C_j$.

\[
C_j \equiv E[\xi_j(\tau_L - 1)X_H(\tau_L - 1)'] = E[\xi_j(\tau_L)X_H(\tau_L)']
\]

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

(L.3)

Comparing (H.6) and (L.3), the first $q$ rows of $C_j$ are identical to the first $q$ rows of $C$ and thus we can use (H.8) directly.

For the last row of $C_j$, we have that

\[
E[x_H(\tau_L - (j - 1))x_H(\tau_L, m + 1 - k)] = E \left[ \sum_{l=1}^{m} \delta_l x_H(\tau_L - (j - 1), l) x_H(\tau_L, m + 1 - k) \right]
\]

\[
= \sum_{l=1}^{m} \delta_l E[x_H(\tau_L - (j - 1), l) x_H(\tau_L, m + 1 - k)]
\]

\[
= \sum_{l=1}^{m} \delta_l E[x_H(\tau_L - (j - 1), l) x_H(\tau_L - f(k), g(k))]
\]

\[
= \sum_{l=1}^{m} \delta_l \Upsilon_{(j-1)-f(k)}(g(k), l).
\]

The third equality follows from the high frequency simplification (H.3), while the last equality follows from (H.7) and (H.8).

Thus, we conclude that

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

(L.4)

### M Proof of Theorem 3.7

This proof is almost identical to the proof of Theorem 3.3 shown in Appendix I. The only difference is that we worked on the mixed frequency naïve regression model in Theorem 3.3, while we work on the low frequency naïve regression model here in Theorem 3.7. The regressors in the former are $x(\tau_L - 1) = [x_L(\tau_L - 1), \ldots, x_L(\tau_L - q), x_H(\tau_L - 1, m + 1 - h)]'$, while the regressors in the latter are $\bar{x}(\tau_L - 1) = [x_L(\tau_L - 1), \ldots, x_L(\tau_L - q), x_H(\tau_L - 1), \ldots, x_H(\tau_L - h)]'$. By replacing $x(\tau_L - 1)$ with $\bar{x}(\tau_L - 1)$, we can reuse Appendix I to derive the asymptotic distribution of the low frequency Wald statistic $W_{LF}$. Under $H_0 : b = 0_{pm \times 1}$, we have
that \( W_{LF} \overset{d}{\to} \chi^2_h \). Under \( H_1: b = (1/\sqrt{T_L}) \nu \), we have that \( W_{LF} \overset{d}{\to} \chi^2_h(\nu) \) with

\[
\nu = \frac{1}{\sigma_L^2} \nu' \Gamma^{(1)}(R \Gamma^{(1)} R')^{-1} \Gamma^{(1)} C \nu,
\]

where \( \bar{R} = [0_{h \times q}, I_h], \Gamma = E[\varphi(\tau - 1)\varphi(\tau - 1)'], \) and \( C = E[\varphi(\tau - 1)X_H(\tau - 1)'] \).

### N Proof of Theorem 3.8

We start with \( \Gamma \).

\[
\Gamma \equiv \begin{bmatrix} \Gamma_{UL} & \Gamma_{UR} \\ \Gamma_{LR} & \Gamma_{LR} \end{bmatrix} = E[\varphi(\tau - 1)\varphi(\tau - 1)']
\]

\[
= E \begin{bmatrix} X_L^{(q)}(\tau - 1) \\ x_H(\tau - 1) \\ \vdots \\ x_H(\tau - h) \end{bmatrix} [X_L^{(q)}(\tau - 1)', x_H(\tau - 1), \ldots, x_H(\tau - h)]
\]

(N.1)

Comparing (N.1) and (J.1), it follows that \( \Gamma_{UL} = \Gamma_{UL} \). We can therefore simply use (J.2).

In view of (L.1), \( \Gamma_{UR} \) is aligned upper-right blocks of \( \Gamma_{j,1}, \ldots, \Gamma_{j,h} \). Hence, we have from (L.2) that

\[
\Gamma_{UR} = \sum_{i=1}^{m} \delta_i \begin{bmatrix} \Upsilon_{1-1}(K, l) & \ldots & \Upsilon_{h-1}(K, l) \\ \vdots & \ddots & \vdots \\ \Upsilon_{1-q}(K, l) & \ldots & \Upsilon_{h-q}(K, l) \end{bmatrix}.
\]

Similarly, \( \Gamma_{LR} \) is aligned lower-right blocks of \( \Gamma_{j,i} \) for \( j, i \in \{1, \ldots, h\} \). Hence, we have from (L.2) that

\[
\Gamma_{LR} = \sum_{i=1}^{m} \sum_{k=1}^{m} \delta_i \delta_k \begin{bmatrix} \Upsilon_{1-1}(k, l) & \ldots & \Upsilon_{1-h}(k, l) \\ \vdots & \ddots & \vdots \\ \Upsilon_{h-1}(k, l) & \ldots & \Upsilon_{h-h}(k, l) \end{bmatrix}.
\]

We now discuss \( C \).

\[
C \equiv E[\varphi(\tau - 1)X_H(\tau - 1)'] = \begin{bmatrix} E[X_L^{(q)}(\tau - 1)X_H(\tau - 1)'] \\ E[x_H(\tau - 1)X_H(\tau - 1)'] \\ \vdots \\ E[x_H(\tau - h)X_H(\tau - 1)'] \end{bmatrix}.
\]

In view of (L.3), the first \( q \) rows of \( C \) are identical to the first \( q \) rows of \( C_j \). We can thus use the first \( q \) rows of (L.4) directly.
Moreover, the last $h$ rows of $C$ are aligned last rows of $C_1, \ldots, C_h$. Hence,

$$C = \begin{bmatrix}
Y_{f(1)}(K, g(1)) & \cdots & Y_{f(pm)}(K, g(pm)) \\
\vdots & \ddots & \vdots \\
\sum_{l=1}^{m} \delta_l Y_{(1-1)-f(1)}(g(1), l) & \cdots & \sum_{l=1}^{m} \delta_l Y_{(1-1)-f(pm)}(g(pm), l) \\
\sum_{l=1}^{m} \delta_l Y_{(h-1)-f(1)}(g(1), l) & \cdots & \sum_{l=1}^{m} \delta_l Y_{(h-1)-f(pm)}(g(pm), l)
\end{bmatrix}$$