Simple Granger Causality Tests for Mixed Frequency Data

Eric Ghysels* Jonathan B. Hill† Kaiji Motegi‡

First Draft: October 1, 2013
This Draft: June 9, 2015

Abstract

This paper presents simple Granger causality tests applicable to any mixed frequency sampling data setting, and feature remarkable power properties even with relatively small low frequency data samples and a considerable wedge between sampling frequencies (for example, quarterly and daily or weekly data). Our tests are based on a seemingly overlooked, but simple, dimension reduction technique for regression models. If the number of parameters of interest is large then in small or even large samples any of the trilogy test statistics may not be well approximated by their asymptotic distribution. A bootstrap method can be employed to improve empirical test size, but this generally results in a loss of power. A shrinkage estimator can be employed, including Lasso, Adaptive Lasso, or Ridge Regression, but these are valid only under a sparsity assumption which does not apply to Granger causality tests. The procedure, which is of general interest when testing potentially large sets of parameter restrictions, involves multiple parsimonious regression models where each model regresses a low frequency variable onto only one individual lag or lead of a high frequency series, where that lag or lead slope parameter is necessarily zero under the null hypothesis of non-causality. Our test is then based on a max test statistic that selects the largest squared estimator among all parsimonious regression models. Parsimony ensures sharper estimates and therefore improved power in small samples. Inference requires a simple simulation-bootstrap step since the test statistic has a non-standard limit distribution. We show via Monte Carlo simulations that the max test is more powerful than existing mixed frequency Granger causality tests in small samples. An empirical application examines Granger causality over rolling windows of U.S. macroeconomic data from 1962-2013 using a mixture of high and low frequency data.

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 adversely affects Granger's (1969) causality.[1] One of the most popular Granger causality tests is a Wald statistic based on multi-step ahead vector autoregression (VAR) models. Its appeal is that the approach can handle *causal chains* among more than two variables.[2] Since standard VAR models are designed for single-frequency data, these tests often suffer from the adverse effect of temporal aggregation. In order 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 accomplish this by extending Dufour, Pelletier, and Renault's (2006) VAR-based causality test using Ghysels' (2014) mixed frequency vector autoregressive (MF-VAR) models.[3] Although these tests avoid the undesirable effects of temporal aggregation, their applicability is limited because parameter proliferation in MF-VAR models adversely affects the power of the tests, even after bootstrapping a test statistic p-value. Indeed, if we let m be the ratio of high and low frequencies (e.g. $m = 3$ in mixed monthly and quarterly data), then for bivariate mixed frequency settings the MF-VAR is of dimension $m + 1$. Parameter proliferation occurs when m is large, and becomes precipitously worse as the VAR order increases. In these cases, Ghysels, Hill, and Motegi's (2013) Wald test of non-causality exhibits size distortions, while a bootstrapped Wald test results in correct size but low size-corrected power, a common occurrence when bootstrapping a size distorted asymptotic test (cfr. Davidson and MacKinnon (2006)).

The present paper proposes a remarkably simple regression-based mixed frequency Granger causality test that, in the case of testing non-causality from a low to a high frequency variable, exploits Sims' (1972) two-sided regression model. The tests we propose have several advantages: (i) they are regression-based and simple to implement, and (ii) they apply to any m, large or small, and even time-varying m, for example the number of days in a month.[4] We postulate multiple parsimonious regression models where the j^{th} model regresses a low frequency variable x_L onto lags of x_L and only the j^{th} lag or lead of a high frequency variable x_H. Our test statistic is the *maximum* among squared estimators scaled and weighted properly. Although 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 therefore straightforward to implement in practice.

Our tests are based on a seemingly overlooked, but simple, dimension reduction technique

[1] Existing Granger causality tests typically ignore this issue since they are based on aggregating data to the common lowest frequency, leading possibly to spurious (non)causality. See Zellner and Montmarquette (1971) and Amemiya and Wu (1972) for early contributions. This subject has been subsequently extensively researched: see, for example, Granger (1980), Granger (1988), Lütkepohl (1993), Granger (1995), Renault, Sekkat, and Szafarz (1998), Marcellino (1999), Breitung and Swanson (2002), and McCrorie and Chambers (2006), among others.

[2] See Lütkepohl (1993), Dufour and Renault (1998), Dufour, Pelletier, and Renault (2006), and Hill (2007).

[3] An early example of ideas related to mixed frequency VAR models appeared in Friedman (1962). Foroni, Ghysels, and Marcellino (2013) provide a survey of mixed frequency VAR models.

[4] We assume m is constant in order to conserve notation, but all of our theoretical results extend to time varying m in a straightforward way.

for regression models. The merits can be easily understood if we focus on a low frequency data generating process $y(\tau) = a'x(\tau) + b'z(\tau) + \epsilon(\tau)$, where $y(\tau)$ is a scalar, $\epsilon(\tau)$ is an idiosyncratic term, $a \in \mathbb{R}^k$ and $b \in \mathbb{R}^h$, and $(x(\tau), z(\tau))$ are regressors. Consider testing the hypothesis that $b = 0$, and suppose that standard asymptotics apply for estimating $[a', b']'$. If the number of parameters of interest h is large then in small or even large samples any of the trilogy test statistics may not be well approximated by their asymptotic χ^2 distribution. A bootstrap method can be employed to improve empirical test size, but this generally results in a loss of power since (i) bootstrap samples only approximate the true data generating process; and (ii) at best *size corrected* bootstrap test power nearly matches *size corrected* asymptotic test power, where the latter may be quite low due to large size distortions (cfr. Davidson and MacKinnon (2006)). A shrinkage estimator can be employed, including Lasso, Adaptive Lasso, or Ridge Regression, but these are valid only under the sparsity assumption $b = 0$, hence we cannot test $b = 0$ against $b \neq 0$. In the mixed frequency literature, MIDAS polynomials are proposed as an ad hoc dimension reduction, but these models may generally be mis-specified and therefore result in low or no power in some directions from the null of Granger non-causality.

Our contribution is to build parsimonious regression models $y(\tau) = a_j'x(\tau) + \beta_j z(\tau, j) + u(\tau, j)$ for $j = 1, \ldots, h$, where $z(\tau, j)$ is the j^{th} component of $z(\tau)$, and $\tau = 1, \ldots, T$ with sample size T. Write $\beta = [\beta_j]_{j=1}^h$. Provided the covariance matrix for the regressors $[x(\tau)', z(\tau)']'$ is non-singular, it can be shown $b = 0$ *if and only if* $\beta = 0$.[5] If $\hat{\beta}_j$ estimates β_j then our test in mixed frequencies is based on a slightly generalized form of a max test statistic $\max_{j=1,\ldots,h} \{(\sqrt{T}\hat{\beta}_j)^2\}$. This is a boon for a small sample asymptotic test because for even large h the dimension reduction leads to sharp estimates of β_j, and therefore accurate empirical size, as long as the remaining regressor set $x(\tau)$ does not have a large dimension. This is precisely the case for a test of non-causality from a high frequency (e.g. week) to a low frequency (e.g. quarter) variable. Further, in the above framework $b = 0$ *if and only if* $\beta = 0$ implies the max test is consistent against any deviation from the null. This method obviously cannot identify b when the null is false, but we can identify that $b = 0$ is false asymptotically with probability one for any direction $b \neq 0$. The testing approach we propose can be applied in many other settings, whenever hypothesis testing involves a zero restriction on a large parameter set.

In mixed frequencies, consistency carries over to a test of non-causality from a high to a low frequency variable. There are, however, unresolved difficulties in identifying causality from a low to high frequency variable, although the max test trivially identifies non-causality. As the above example reveals, our method has broad applications for inference in time series regressions with either same or mixed frequency data, and in cross sections where a penalized regression under a sparsity assumption is increasingly common. Our focus here is mixed frequency time series models where parameter proliferation is typical, and no broadly accepted solution exists for sharp small sample inference based on an asymptotic test.

[5] We prove the claim for a test of no causality from a high to a low frequency variable in Theorem 2.4, but the proof trivially carries over to the present low frequency data generating process when $\epsilon(\tau)$ is a stationary martingale difference with respect to the sigma field $\sigma(x(t) : t \leq \tau)$.

In our theoretical analysis, we compare the max test based on mixed frequency [MF] data, with a Wald test based on mixed frequency data, a max test based on low frequency [LF] data, and a Wald test based on low frequency data. We prove the consistency of MF max test for Granger causality from high frequency data x_H to low frequency x_L. We also show by counter-examples that LF tests need *not* be consistent. In the case of Granger causality from x_L to x_H, proving the consistency of MF max test remains an open question. Moreover, relative to LF tests, we show that MF tests are more robust against complex (but realistic) causal patterns both in terms of local asymptotics and in finite samples. In addition, we also show that the MF max and the MF Wald tests are roughly equally powerful in terms of local asymptotics, but the former is clearly more powerful in finite samples. Local power is similar for max and Wald tests precisely because power is asymptotic, and these statistics may have different dispersions.

The remainder of the paper is organized as follows. Sections 2 and 3 present the max test statistic and derive its asymptotic properties for the two cases of testing for non-causality from high-to-low and low-to-high frequencies. In Section 4 we conduct local power analysis. In Section 5 we run Monte Carlo simulations, Section 6 presents an empirical application, and Section 7 concludes the paper. Proofs for all theorems are provided in Technical Appendices, and tables and figures are collected at the end.

2 High-to-Low Frequency Data Granger Causality

This paper focuses on a bivariate case where we have a high frequency variable x_H and a low frequency variable x_L, so we need to formulate a data generating process (DGP) governing these variables.[6] We denote by m the number of high frequency time periods for each low frequency time period $\tau_L \in \mathbb{Z}$, often called *the ratio of sampling frequencies*. We assume throughout that m is fixed (e.g. $m = 3$ months per quarter) in order to focus ideas and reduce notation, but all of our main results carry over to time-varying sequences $m(\tau_L)$ in a straightforward way.

Example 2.1 (Mixed Frequency Data - Quarterly and Monthly)**.** A simple example is when the high frequency is monthly observations combined with quarterly low frequency data, hence $m = 3$. We let $x_H(\tau_L, 1)$ be the first monthly observation of x_H in quarter τ_L, $x_H(\tau_L, 2)$ is the second, and $x_H(\tau_L, 3)$ is the third. A leading example in macroeconomics is quarterly real GDP growth $x_L(\tau_L)$, where existing analyses of causal patterns use unemployment, oil prices, inflation, interest rates, etc., aggregated into quarters (see Hill (2007) for references). Consider monthly CPI inflation in quarter τ_L, denoted $[x_H(\tau_L, 1), x_H(\tau_L, 2), x_H(\tau_L, 3)]'$ and the resulting stacked system for quarter τ_L therefore is $\{x_H(\tau_L, 1), x_H(\tau_L, 2), x_H(\tau_L, 3), x_L(\tau_L)\}$. The assumption that $x_L(\tau_L)$ is observed *after* $x_H(\tau_L, m)$ is merely a convention.

In the bivariate case with one high and one low frequency variable, we have a $K \times 1$ *mixed frequency vector* $\boldsymbol{X}(\tau_L) = [x_H(\tau_L, 1), \ldots, x_H(\tau_L, m), x_L(\tau_L)]'$, where $K = m + 1$. Define the

[6] The trivariate case involves causality chains in mixed frequency which are far more complicated, and detract us from the main theme of dimension reduction covered in this paper. See Dufour and Renault (1998), Dufour, Pelletier, and Renault (2006) and Hill (2007) for further discussion.

σ-field $\mathcal{F}_{\tau_L} \equiv \sigma(\boldsymbol{X}(\tau) : \tau \leq \tau_L)$. We assume as in Ghysels, Hill, and Motegi (2013) and Ghysels (2014) that $E[\boldsymbol{X}(\tau_L)|\mathcal{F}_{\tau_L-1}]$ has a version that is *almost surely* linear in $\{\boldsymbol{X}(\tau_L-1), \ldots, \boldsymbol{X}(\tau_L-p)\}$ for some finite $p \geq 1$.

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

$$\underbrace{\begin{bmatrix} x_H(\tau_L,1) \\ \vdots \\ x_H(\tau_L,m) \\ x_L(\tau_L) \end{bmatrix}}_{=\boldsymbol{X}(\tau_L)} = \sum_{k=1}^{p} \underbrace{\begin{bmatrix} d_{11,k} & \ldots & d_{1m,k} & c_{(k-1)m+1} \\ \vdots & \ddots & \vdots & \vdots \\ d_{m1,k} & \ldots & d_{mm,k} & c_{km} \\ b_{km} & \ldots & b_{(k-1)m+1} & a_k \end{bmatrix}}_{\equiv \boldsymbol{A}_k} \underbrace{\begin{bmatrix} x_H(\tau_L-k,1) \\ \vdots \\ x_H(\tau_L-k,m) \\ x_L(\tau_L-k) \end{bmatrix}}_{=\boldsymbol{X}(\tau_L-k)} + \underbrace{\begin{bmatrix} \epsilon_H(\tau_L,1) \\ \vdots \\ \epsilon_H(\tau_L,m) \\ \epsilon_L(\tau_L) \end{bmatrix}}_{\equiv \boldsymbol{\epsilon}(\tau_L)} \tag{2.1}$$

or compactly

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

The error $\{\boldsymbol{\epsilon}(\tau_L)\}$ is a strictly stationary martingale difference sequence (mds) with respect to increasing $\mathcal{F}_{\tau_L} \subset \mathcal{F}_{\tau_L+1}$, with positive definite covariance matrix $\boldsymbol{\Omega} \equiv E[\boldsymbol{\epsilon}(\tau_L)\boldsymbol{\epsilon}(\tau_L)']$.

Two remarks regarding the above assumption are worth making.

Remark 2.1. The mds assumption allows for conditional heteroscedasticity of unknown form, including GARCH-type processes. We can also easily allow for stochastic volatility or other random volatility errors by expanding the definition of the σ-fields $\mathcal{F}_{\tau_L}$.

Remark 2.2. A constant term is omitted from (2.1) for simplicity, but can be easily added if desired. Therefore, $\boldsymbol{X}(\tau_L)$ is mean centered. The coefficients d (a) govern the autoregressive property of x_H (x_L).

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

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

Since $\{\boldsymbol{\epsilon}(\tau_L)\}$ is not i.i.d. we must impose a weak dependence property in order to ensure standard asymptotics. In the following we assume $\boldsymbol{\epsilon}(\tau_L)$ and $\boldsymbol{X}(\tau_L)$ are stationary α-mixing.[7]

[7]See Doukhan (1994) for compendium details on mixing sequences.

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

Assumption 2.3. $\boldsymbol{X}(\tau_L)$ and $\boldsymbol{\epsilon}(\tau_L)$ are α-mixing with mixing coefficients α_h that satisfy $\sum_{h=0}^{\infty} \alpha_{2^h} < \infty$.

Remark 2.3. Note that $\boldsymbol{\Omega} \equiv E[\boldsymbol{\epsilon}(\tau_L)\boldsymbol{\epsilon}(\tau_L)']$ allows for the high frequency innovations $\epsilon_H(\tau_L, j)$ to have a different variance for each j. Therefore, while Assumptions 2.1 and 2.2 imply $\{x_H(\tau_L, j)\}_{\tau_L}$ is covariance stationary *for each fixed* $j \in \{1, \ldots, m\}$, they do not imply covariance stationarity for *the entire high frequency array* $\{\{x_H(\tau_L, j)\}_{j=1}^{m}\}_{\tau_L}$.

Remark 2.4. The condition $\sum_{h=0}^{\infty} \alpha_{2^h} < \infty$ is quite general, allowing for geometric or hyperbolic memory decay in the innovations $\boldsymbol{\epsilon}(\tau_L)$, hence conditional volatility with a broad range of dynamics. We impose the infinite order lag function $\boldsymbol{X}(\tau_L)$ of $\boldsymbol{\epsilon}(\tau_L)$ to also be mixing as a simplifying assumption since underlying sufficient conditions are rather technical if $\{\boldsymbol{\epsilon}(\tau_L)\}$ is a non-finite dependent process (see Chapter 2.3.2 in Doukhan (1994)).

Since there are fundamentally different challenges when testing for non-causality from high-to-low or low-to-high frequency, we restrict attention to the former in this section, and treat the latter in Section 3.

We first pick the last row of the entire system (2.1):

$$\begin{aligned} x_L(\tau_L) &= \sum_{k=1}^{p} a_k x_L(\tau_L - k) + \sum_{j=1}^{pm} b_j x_H(\tau_L - 1, m + 1 - j) + \epsilon_L(\tau_L), \\ \epsilon_L(\tau_L) &\overset{mds}{\sim} (0, \sigma_L^2), \quad \sigma_L^2 > 0. \end{aligned} \tag{2.2}$$

The index $j \in \{1, \ldots, pm\}$ is in high frequency terms, and the second argument $m + 1 - j$ of $x_H(\tau_L - 1, m + 1 - j)$ can be less than 1 since $j > m$ occurs when $p > 1$. Allowing any integer value in the second argument of $x_H(\tau_L - 1, m + 1 - j)$, including those smaller than 1 or larger than m, does not cause any confusion, and simplifies analytical arguments below. We can therefore interchangeably write $x_H(\tau_L - i, j) = x_H(\tau_L, j - im)$ for $j = 1, \ldots, m$ and $i \geq 0$, for example: $x_H(\tau_L, 0) = x_H(\tau_L - 1, m)$, $x_H(\tau_L, -1) = x_H(\tau_L - 1, m - 1)$, and $x_H(\tau_L, m + 1) = x_H(\tau_L + 1, 1)$.[8]

Now define $\boldsymbol{X}_L(\tau_L - 1) = [x_L(\tau_L - 1), \ldots, x_L(\tau_L - p)]'$, $\boldsymbol{X}_H(\tau_L - 1) = [x_H(\tau_L - 1, m + 1 - 1), \ldots, x_H(\tau_L - 1, m + 1 - pm)]'$, $\boldsymbol{a} = [a_1, \ldots, a_p]'$, and $\boldsymbol{b} = [b_1, \ldots, b_{pm}]'$. Then, (2.2) becomes:

$$x_L(\tau_L) = \boldsymbol{X}_L(\tau_L - 1)'\boldsymbol{a} + \boldsymbol{X}_H(\tau_L - 1)'\boldsymbol{b} + \epsilon_L(\tau_L). \tag{2.3}$$

Based on the classic theory of Dufour and Renault (1998) and the mixed frequency extension made by Ghysels, Hill, and Motegi (2013), we know that x_H does not Granger cause x_L given the mixed frequency information set $\mathcal{F}_{\tau_L} = \sigma(\boldsymbol{X}(\tau) : \tau \leq \tau_L)$ *if and only if* $\boldsymbol{b} = \boldsymbol{0}_{pm \times 1}$. In

[8] Complete details on mixed frequency notation conventions are given in Appendix A.

order to test the non-causality hypothesis $H_0 : \boldsymbol{b} = \mathbf{0}_{pm\times 1}$, we want a test statistic that obtains asymptotic power of one against any deviation from non-causality (i.e. it is consistent), achieves high power in local asymptotics and finite samples, and does not produce size distortions in small samples when pm is large. We divide the topic into mixed and low frequency approaches.

2.1 Max Test: High-to-Low Granger Causality

Before presenting the new test, it is helpful to review the existing mixed frequency Granger causality test proposed by Ghysels, Hill, and Motegi (2013). They work with a regression model that regresses x_L onto q low frequency lags and h high frequency lags of x_H:

$$x_L(\tau_L) = \sum_{k=1}^{q} \alpha_k x_L(\tau_L - k) + \sum_{j=1}^{h} \beta_j x_H(\tau_L - 1, m + 1 - j) + u_L(\tau_L) \tag{2.4}$$

for $\tau_L = 1, \ldots, T_L$. Ghysels, Hill, and Motegi (2013) estimate the parameters in (2.4) by least square sand then test $H_0 : \beta_1 = \cdots = \beta_h = 0$ via a Wald test. Model (2.4) contains DGP (2.2) as a special case when $q \geq p$ and $h \geq pm$, hence the Wald test is trivially consistent if $q \geq p$ and $h \geq pm$.

A potential problem here is that pm, the true lag order of x_H, may be quite large in some applications, even when the AR order p is fairly small. Consider a weekly versus quarterly data case for instance, hence the MF-VAR lag order p is in terms of quarters and $m = 13$ approximately. Then $pm = 39$ when $p = 3$, and $pm = 52$ when $p = 4$, etc. Including sufficiently many high frequency lags $h \geq pm$ generally results in size distortions for an asymptotic Wald test when the sample size T_L is small and pm is large. Davidson and MacKinnon (2006), however, show that the power of bootstrap tests is close to the asymptotic power for *size-corrected* tests. An asymptotic Wald test of mixed frequency non-causality can exhibit substantial size distortions, implying size-corrected power well below one in finite samples. Of course, we may use a small number of lags $h < pm$ to ensure the Wald statistic is well characterized by its χ^2 limit distribution, but this results in an inconsistent test when there exists Granger causality involving lags beyond h. Conversely, we may use a MIDAS type parametric dimension reduction, but a mis-specified MIDAS polynomial again results in an inconsistent test (cfr. Ghysels, Santa-Clara, and Valkanov (2004), Ghysels, Santa-Clara, and Valkanov (2006), Ghysels, Hill, and Motegi (2013)). Finally, we can simply bypass a mixed frequency approach in order to reduce dimensionality, but a Wald test of non-causality in a low frequency model is not consistent (see Section 2.2). This is in a nutshell the problem of parameter proliferation in mixed frequency models.

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

$$x_L(\tau_L) = \sum_{k=1}^{q} \alpha_{k,j} x_L(\tau_L - k) + \beta_j x_H(\tau_L - 1, m + 1 - j) + u_{L,j}(\tau_L), \quad j = 1, \ldots, h. \tag{2.5}$$

We abuse notation since the key parameter β_j in (2.5) in generally not equivalent to β_j in (2.4) when there is causality. We are, however, concerned with a test of non-causality, in which case there is little loss of generality, and a gain of notation simplicity. Moreover, as we show below, the hypothesis $\boldsymbol{b} = \mathbf{0}_{pm\times 1}$, corresponding to high-to-low non-causality in (2.3), holds *if and only if* $\beta_j = 0$ for each $j = 1, ..., h$ in (2.5), provided $q \geq p$ and $h \geq pm$. Hence, the parsimonious models allow us to identify null and alternative hypotheses.

Model j is compactly rewritten as

$$\begin{aligned} x_L(\tau_L) &= \left[\boldsymbol{X}_L^{(q)}(\tau_L - 1)' \quad x_H(\tau_L - 1, m + 1 - j)\right] \begin{bmatrix} \alpha_{1,j} \\ \vdots \\ \alpha_{q,j} \\ \beta_j \end{bmatrix} + u_{L,j}(\tau_L) \qquad (2.6) \\ &= \boldsymbol{x}_j(\tau_L - 1)'\boldsymbol{\theta}_j + u_{L,j}(\tau_L), \end{aligned}$$

say, where

$$\boldsymbol{X}_L^{(q)}(\tau_L - 1) \equiv [x_L(\tau_L - 1), \ldots, x_L(\tau_L - q)]'.$$

The j^{th} model contains q low frequency autoregressive lags of x_L as well as *only* the j^{th} high frequency lag of x_H. Therefore, the number of parameters, $q + 1$, is typically much smaller than the number of parameters restrictions equal to $q + h$ in the naïve regression model (2.4). This advantage helps to alleviate size distortions for large m and small T_L.

In order 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 regression model (2.4) in order to focus on the causality component, and not the autoregressive component.

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

The parsimonious regression models obviously reveal non-causality from high-to-low frequency since $\beta_j = 0$ for each j in (2.4) implies $\beta_j = 0$ in each j^{th} equation in (2.5). The subtler challenge is showing that (2.5) reveals any departure from non-causality in (2.4), hence (2.5) can be used as a valid and consistent test of high-to-low non-causality. We first describe how to combine all h parsimonious models to get a test statistic for testing non-causality, and then show how the resulting test identifies the (non)-causality.

Since we are assuming that $q \geq p$, each model (2.5) is correctly specified under the null hypothesis of high-to-low non-causality. Hence, if there is non-causality from high-to-low frequency, the least squares estimators $\hat{\beta}_j \xrightarrow{p} 0$, hence $\max_{1\leq j\leq h} \{\hat{\beta}_j^2\} \xrightarrow{p} 0$. Using this property, we propose a max test statistic:

$$\hat{\mathcal{T}} \equiv \max_{1\leq j\leq h} \left(\sqrt{T_L} w_{T_L,j} \hat{\beta}_j\right)^2, \qquad (2.7)$$

where $\{w_{T_L,j} : j = 1, \ldots, h\}$ is a sequence of $\sigma(\boldsymbol{X}(\tau_L - k) : k \geq 1)$-measurable L_2-bounded non-negative scalar weights with non-random probability limits $\{w_j\}$. As a standardization, we assume $\sum_{j=1}^{h} w_{T_L,j} = 1$ without loss of generality. When we do not have any prior information about the weighting structure, a straightforward choice of $w_{T_L,j}$ is the non-random flat weight $1/h$. We can consider any other weighting structure by choosing desired $\{w_{T_L,1}, \ldots, w_{T_L,h}\}$, and other measurable mappings from $\mathbb{R}^h$ to $[0, \infty)$ like the average $\sum_{j=1}^{h}(\sqrt{T_L} w_{T_L,j} \hat{\beta}_j)^2$ (cfr. Andrews and Ploberger (1994)). Finally, we will also need to make an assumption about how many models $j = 1, \ldots, h$ we run to perform the test, namely:

Assumption 2.5. The number of models h used to compute $\hat{\mathcal{T}}$ in (2.7) satisfies, $h \geq pm$.

2.1.1 Asymptotics under Non-Causality from High-to-Low Frequency

We stack all parameters across the h models (2.6) and write:

$$\boldsymbol{\theta} \equiv [\boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_h']',$$

and construct a selection matrix $\boldsymbol{R}$ such that

$$\boldsymbol{\beta} \equiv [\beta_1, \ldots, \beta_h]' = \boldsymbol{R}\boldsymbol{\theta}.$$

Therefore, $\boldsymbol{R}$ is an $h \times (q+1)h$ matrix with $\boldsymbol{R}_{j,(q+1)j} = 1$ for $j = 1, \ldots, h$, and all other elements are zero. Let $\boldsymbol{W}_{T_L,h}$ be an $h \times h$ diagonal matrix whose diagonal elements are $w_{T_L,1}$, …, $w_{T_L,h}$. Similarly, let $\boldsymbol{W}_h$ be an $h \times h$ diagonal matrix whose diagonal elements are w_1, …, w_h.

Under Assumptions 2.1–2.4, it is not hard to prove the asymptotic normality of $\hat{\boldsymbol{\theta}}$ and hence $\hat{\boldsymbol{\beta}}$. A simple weak convergence argument then suffices for the max test statistic.

Theorem 2.1. Let Assumptions 2.1-2.4 hold. Under $H_0 : \boldsymbol{b} = \boldsymbol{0}_{pm\times 1}$, we have that $\hat{\mathcal{T}} \xrightarrow{d} \max_{1\leq j\leq h} \mathcal{N}_j^2$ as $T_L \to \infty$, where $\boldsymbol{\mathcal{N}} \equiv [\mathcal{N}_1, \ldots, \mathcal{N}_h]'$ is distributed $N(\boldsymbol{0}_{h\times 1}, \boldsymbol{V})$ with positive definite covariance matrix:

$$\boldsymbol{V} \equiv \sigma_L^2 \boldsymbol{W}_h \boldsymbol{R}\boldsymbol{\mathcal{S}}\boldsymbol{R}'\boldsymbol{W}_h \in \mathbb{R}^{h\times h}, \tag{2.8}$$

where $\sigma_L^2 \equiv E[\epsilon_L^2(\tau_L)]$, and

$$\boldsymbol{\mathcal{S}} \equiv \begin{bmatrix} \boldsymbol{\Sigma}_{1,1} & \ldots & \boldsymbol{\Sigma}_{1,h} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{h,1} & \ldots & \boldsymbol{\Sigma}_{h,h} \end{bmatrix} \in \mathbb{R}^{(q+1)h\times(q+1)h} \text{ and } \boldsymbol{\Sigma}_{j,i} \equiv \boldsymbol{\Gamma}_{j,j}^{-1}\boldsymbol{\Gamma}_{j,i}\boldsymbol{\Gamma}_{i,i}^{-1} \in \mathbb{R}^{(q+1)\times(q+1)},$$

$$\boldsymbol{\Gamma}_{j,i} \equiv E\left[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{x}_i(\tau_L - 1)'\right] \in \mathbb{R}^{(q+1)\times(q+1)} \text{ and } \boldsymbol{R} \equiv \begin{bmatrix} \boldsymbol{0}_{1\times q} & 1 & \cdots & \cdots & \boldsymbol{0}_{1\times q} & 0 \\ \vdots & \vdots & \ldots & \ldots & \vdots & \vdots \\ \boldsymbol{0}_{1\times q} & 0 & \cdots & \cdots & \boldsymbol{0}_{1\times q} & 1 \end{bmatrix} \in \mathbb{R}^{h\times(q+1)h}. \tag{2.9}$$

Proof. See Appendix B.

2.1.2 Simulated P-Value

The mixed frequency max test statistic $\hat{\mathcal{T}}$ has a non-standard limit distribution under H_0 that can be easily simulated in order to compute an approximate p-value. Let $\hat{\boldsymbol{V}}_{T_L}$ be a consistent estimator of $\boldsymbol{V}$ (see below), and draw R samples $\boldsymbol{\mathcal{N}}^{(1)}, \ldots, \boldsymbol{\mathcal{N}}^{(R)}$ independently from $N(\mathbf{0}_{h\times 1}, \hat{\boldsymbol{V}}_{T_L})$. Now compute artificial test statistics $\hat{\mathcal{T}}^{(r)} \equiv \max_{1\leq j\leq h}(\boldsymbol{\mathcal{N}}_j^{(r)})^2$. An asymptotic p-value approximation for $\hat{\mathcal{T}}$ is

$$\hat{p} = \frac{1}{R}\sum_{r=1}^{R} I\left(\hat{\mathcal{T}}^{(r)} > \hat{\mathcal{T}}\right).$$

Since $\boldsymbol{\mathcal{N}}^{(r)}$ are i.i.d., and R can be made arbitrarily large, by the Glivenko-Cantelli Theorem $\hat{p}$ can be made arbitrarily close to $P(\hat{\mathcal{T}}^{(1)} > \hat{\mathcal{T}})$.

Define the max test limit distribution under H_0: $F^0(c) \equiv P(\max_{1\leq j\leq h}(\boldsymbol{\mathcal{N}}_j^{(r)})^2 \leq c)$. The asymptotic p-value is therefore $\bar{F}^0(\hat{\mathcal{T}}) \equiv 1 - F^0(\hat{\mathcal{T}}) = P(\max_{1\leq j\leq h}(\boldsymbol{\mathcal{N}}_j^{(r)})^2 \geq \hat{\mathcal{T}})$. By an argument identical to Theorem 2 in Hansen (1996), we have the following link between the p-value approximation $P(\hat{\mathcal{T}}^{(1)} > \hat{\mathcal{T}})$ and the asymptotic p-value for $\hat{\mathcal{T}}$.

Theorem 2.2. Under Assumptions 2.1 - 2.4 $P(\hat{\mathcal{T}}^{(1)} > \hat{\mathcal{T}}) = \bar{F}^0(\hat{\mathcal{T}}) + o_p(1)$, hence $\hat{p} = \bar{F}^0(\hat{\mathcal{T}}) + o_p(1)$.

Proof. See Appendix C.

A consistent estimator of $\boldsymbol{V}$ in (2.8) is easily obtained. Simply note that the weights $\boldsymbol{W}_{T_L,h} \xrightarrow{p} \boldsymbol{W}_h$ by assumption, and $\hat{\boldsymbol{\Gamma}}_{j,i} \equiv 1/T_L \sum_{\tau_L=1}^{T_L} \boldsymbol{x}_j(\tau_L - 1)\boldsymbol{x}_i(\tau_L - 1)' \xrightarrow{p} \boldsymbol{\Gamma}_{j,i}$ under Assumptions 2.1-2.4. Consistent estimators $\hat{\boldsymbol{\Sigma}}_{j,i} \xrightarrow{p} \boldsymbol{\Sigma}_{j,i}$ and $\hat{\boldsymbol{\mathcal{S}}} \xrightarrow{p} \boldsymbol{\mathcal{S}}$ can then be obtained directly from (2.9). Next, a consistent estimator $\hat{\sigma}_L^2$ of $\sigma_L^2 \equiv E[\epsilon_L^2(\tau_L)]$ can be obtained by computing residuals $\hat{\epsilon}_L(\tau_L)$ from model (2.4). Note, though, that we only require consistency for the true σ_L^2 under H_0 since power only requires an estimator with a constant finite probability limit. As a bonus, estimating σ_L^2 under H_0 can be done simply by fitting an AR(q) model for x_L and computing the sample variance of the residuals.

2.1.3 Identification of Null and Alternative Hypotheses

We now show that as long as the number of high frequency lags h used across the parsimonious regression models (2.5) is at least as large as the dimension pm of the parameters $\boldsymbol{b}$ from the true DGP (2.3), then the parsimonious regression parameters $\boldsymbol{\beta}$ identify null and alternative hypotheses in the sense $\boldsymbol{b} = \mathbf{0}_{pm\times 1}$ *if and only if* $\boldsymbol{\beta} = \mathbf{0}_{h\times 1}$. Under the condition $h \geq pm$ the max test statistic $\hat{\mathcal{T}}$ then has its intended limit properties under either hypothesis.

If there is Granger causality then the estimator $\hat{\beta}_j$ of β_j in (2.5) is in general not Fisher consistent for the true b_j in DGP (2.2) due to omitted regressors. The next result shows that the least squares first order equations for (2.5) identify some so-called pseudo-true values $\boldsymbol{\beta}^* = [\beta_1^*, \ldots, \beta_h^*]'$, which are identically the probability limits of $\hat{\beta}_j$. Since this in turn is a function of underlying parameters $\boldsymbol{a}$, $\boldsymbol{b}$, and σ_L^2, as well as population moments of x_H and x_L, the resulting relationship can be exploited to identify null and alternative hypotheses.

Stack again all parameters $\boldsymbol{\theta}_j$ from (2.6) and write $\boldsymbol{\theta} = [\boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_h']'$, and let $\hat{\boldsymbol{\theta}}$ be the least squares estimator.

Theorem 2.3. Let 2.1-2.4 hold. Then $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}^* \equiv [\boldsymbol{\theta}_1^{*\prime}, \ldots, \boldsymbol{\theta}_h^{*\prime}]'$, the unique pseudo-true value of $\boldsymbol{\theta}$ that satisfies

$$\boldsymbol{\theta}_j^* \equiv \begin{bmatrix} \alpha_{1,j}^* \\ \vdots \\ \alpha_{p,j}^* \\ \alpha_{p+1,j}^* \\ \vdots \\ \alpha_{q,j}^* \\ \beta_j^* \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_p \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} + \underbrace{\left[E\left[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{x}_j(\tau_L - 1)'\right]\right]^{-1}}_{=\boldsymbol{\Gamma}_{j,j}^{-1}:\ (q+1)\times(q+1)} \times \underbrace{E\left[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{X}_H(\tau_L - 1)'\right]}_{\equiv \boldsymbol{C}_j:\ (q+1)\times pm} \boldsymbol{b}, \quad (2.10)$$

where $\boldsymbol{x}_j(\tau_L - 1)$ is a vector of all regressors in each parsimonious regression model (cfr. (2.6)) while $\boldsymbol{X}_H(\tau_L - 1)$ is a vector of pm high frequency lags of x_H (cfr. (2.3)). Therefore $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}^* = \boldsymbol{R}\boldsymbol{\theta}^*$.

Proof. See Appendix D.

Remark 2.5. Although tedious, the population covariance terms $\boldsymbol{\Gamma}_{j,j}$ and $\boldsymbol{C}_j$ can be characterized by the underlying parameters $\boldsymbol{a}$, $\boldsymbol{b}$, and σ_L^2. For an example, see the local asymptotic power analysis in Section 4.

Theorem 2.3 provides useful insights on the relationship between the underlying coefficient $\boldsymbol{b}$ and the pseudo-true value $\boldsymbol{\beta}^*$ of $\boldsymbol{\beta}$ in general. First, as noted in the discussion leading to Theorem 2.1, trivially $\boldsymbol{\beta}^* = \boldsymbol{0}_{h\times 1}$ whenever there is non-causality (i.e $\boldsymbol{b} = \boldsymbol{0}_{pm\times 1}$), regardless of the relative magnitude of h and pm. Second, as the next result proves, $\boldsymbol{b} = \boldsymbol{0}_{pm\times 1}$ whenever $\boldsymbol{\beta}^* = \boldsymbol{0}_{h\times 1}$, provided $h \geq pm$. This follows ultimately from Assumption 2.1 which ensures covariance matrices of $\boldsymbol{X}_L^{(q)}(\tau_L)$ and $\boldsymbol{X}_H(\tau_L)$ are non-singular, which in turns allows us to exactly identify the null, and therefore identify when a deviation from the null takes place. Of course, we cannot generally identify the true β_j in model (2.4) under the alternative, but we can identify that the alternative must be true: that some β_j in (2.4) is non-zero.

Theorem 2.4. Let Assumptions 2.1, 2.2, 2.4 and 2.5 hold. Then $\boldsymbol{\beta}^* = \boldsymbol{0}_{h\times 1}$ implies $\boldsymbol{b} = \boldsymbol{0}_{pm\times 1}$, hence $\boldsymbol{\beta}^* = \boldsymbol{0}_{h\times 1}$ *if and only if* $\boldsymbol{b} = \boldsymbol{0}_{pm\times 1}$.

Proof. See Appendix E.

Theorems 2.1 and 2.4 together imply the max test statistic has its intended limit properties under either hypothesis. Assume the weight limits $w_j > 0$ for all $j = 1, \ldots, h$ so that we have a non-trivial result under the alternative. In view of Theorem 2.4, we also assume h is sufficiently large to allow the parsimonious regression models to identify the hypotheses.

We first tackle the limit when H_0 is false in order to definitively show that the max test statistic null distribution limit holds *if and only if* H_0 is true. The max test statistic construction (2.7) with non-trivial weights $w_j > 0$ for all $j = 1, \ldots, h$ indicates that $\hat{\mathcal{T}} \xrightarrow{p} \infty$ *if and only if* $\boldsymbol{\beta}^* \neq \mathbf{0}_{h\times 1}$, and by Theorems 2.3 and 2.4 $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}^* \neq \mathbf{0}_{h\times 1}$ under a general alternative hypothesis $H_1 : \boldsymbol{b} \neq \mathbf{0}_{pm\times 1}$, given $h \geq pm$. This proves consistency of the mixed frequency max test.

Theorem 2.5. Let Assumptions 2.1-2.5 hold, and assume $w_j > 0$ for all $j = 1, \ldots, h$. Then $\hat{\mathcal{T}} \xrightarrow{p} \infty$ *if and only if* $H_1 : \boldsymbol{b} \neq \mathbf{0}_{pm\times 1}$ is true.

As an immediate consequence of the limit distribution Theorem 2.1, identification Theorem 2.4 and consistency Theorem 2.5, is that the limiting null distribution arises *if and only if* H_0 is true.

Corollary 2.6. Let Assumptions 2.1-2.5 hold, and assume $w_j > 0$ for all $j = 1, \ldots, h$. Then $\hat{\mathcal{T}} \xrightarrow{d} \max_{1\leq j\leq h} \mathcal{N}_j^2$ as $T_L \to \infty$ *if and only if* $H_0 : \boldsymbol{b} = \mathbf{0}_{pm\times 1}$ is true.

If we choose $h < pm$ then it is possible for asymptotic power to be less than unity, as the following example reveals.

Example 2.2 (Inconsistency due to Small h)**.** Consider a simple DGP with $m = 2$ and $p = 1$:

$$\begin{bmatrix} x_H(\tau_L, 1) \\ x_H(\tau_L, 2) \\ x_L(\tau_L) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1/\rho & 1 & 0 \end{bmatrix} \begin{bmatrix} x_H(\tau_L - 1, 1) \\ x_H(\tau_L - 1, 2) \\ x_L(\tau_L - 1) \end{bmatrix} + \begin{bmatrix} \epsilon_H(\tau_L, 1) \\ \epsilon_H(\tau_L, 2) \\ \epsilon_L(\tau_L) \end{bmatrix} \tag{2.11}$$

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

If $(q, h) = (1, 1)$ then asymptotic power of the max test is zero (above the nominal size), and if $(q, h) = (1, 2)$ the asymptotic power is 1. See Appendix E for a proof.

Assume $\rho > 0$ for simplicity. The simple explanation behind the lack of power is that the positive impact of $x_H(\tau_L - 1, 2)$ on $x_L(\tau_L)$, the negative impact of $x_H(\tau_L - 1, 1)$ on $x_L(\tau_L)$, and the positive autocorrelation of x_H all offset each other to make the pseudo-true $\beta_1^* = 0$.

2.2 Low Frequency Approach

The mixed frequency Wald test based on (2.4) and the mixed frequency max test based on model (2.5) are consistent as long as $h \geq pm$. If we instead work with an aggregated x_H, then under DGP (2.1) neither test would be consistent no matter how many low frequency lags of x_H we included.

This is verified by using the following linear aggregation scheme for the high frequency variable:

$$x_H(\tau_L) = \sum_{j=1}^{m} \delta_j x_H(\tau_L, j) \text{ where } \delta_j \geq 0 \text{ for all } j = 1, \ldots, m \text{ and } \sum_{j=1}^{m} \delta_j = 1,$$

where δ_j is a user chosen quantity which determines the aggregation scheme. The 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.

We impose Assumption 2.4 such that $q \geq p$ in order to focus on testing for causality.

2.2.1 Low Frequency Naïve Regression : Wald Test

The low frequency naïve regression model is:

$$\begin{aligned} x_L(\tau_L) &= \sum_{k=1}^{q} \alpha_k x_L(\tau_L - k) + \sum_{j=1}^{h} \beta_j x_H(\tau_L - j) + u_L(\tau_L) \\ &= [\boldsymbol{X}_L^{(q)}(\tau_L - 1)', \underbrace{x_H(\tau_L - 1), \ldots, x_H(\tau_L - h)}_{\equiv \underline{\boldsymbol{X}}_H(\tau_L - 1)'}] \underbrace{\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_q \\ \beta_1 \\ \vdots \\ \beta_h \end{bmatrix}}_{\equiv \boldsymbol{\theta}^{(LF)}} + u_L(\tau_L) \\ &= \underbrace{[\boldsymbol{X}_L^{(q)}(\tau_L - 1)', \underline{\boldsymbol{X}}_H(\tau_L - 1)']}_{\equiv \underline{\boldsymbol{x}}(\tau_L - 1)'} \boldsymbol{\theta}^{(LF)} + u_L(\tau_L). \end{aligned} \tag{2.12}$$

Note that $\underline{\boldsymbol{X}}_H(\tau_L - 1)$ is an $h \times 1$ vector stacking *aggregated* x_H, and $\underline{\boldsymbol{x}}(\tau_L - 1)$ is a $(q + h) \times 1$ vector of all regressors. The superscript "LF" in $\boldsymbol{\theta}^{(LF)}$ emphasizes that we are working on a low frequency model here.[9]

Since (2.1) governs the data generating process, the pseudo-true value for $\boldsymbol{\theta}^{(LF)}$, denoted $\boldsymbol{\theta}^{(LF)*}$, can be derived easily:

$$\boldsymbol{\theta}^{(LF)*} \equiv \begin{bmatrix} \alpha_1^* \\ \vdots \\ \alpha_p^* \\ \alpha_{p+1}^* \\ \vdots \\ \alpha_q^* \\ \boldsymbol{\beta}^* \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_p \\ 0 \\ \vdots \\ 0 \\ \mathbf{0}_{h \times 1} \end{bmatrix} + \underbrace{\left[E\left[\underline{\boldsymbol{x}}(\tau_L - 1)\underline{\boldsymbol{x}}(\tau_L - 1)'\right]\right]^{-1}}_{\equiv \underline{\boldsymbol{\Gamma}}^{-1}:\ (q+h)\times(q+h)} \underbrace{E\left[\underline{\boldsymbol{x}}(\tau_L - 1)\boldsymbol{X}_H(\tau_L - 1)'\right]}_{\equiv \underline{\boldsymbol{C}}:\ (q+h)\times pm} \boldsymbol{b}, \tag{2.13}$$

where $\boldsymbol{\beta}^* = [\beta_1^*, \ldots, \beta_h^*]'$. The derivation of (2.13) is omitted since it is similar to the proof of Theorem 2.3.

[9] Technically "LF" should also be put on α's, β's, and $u_L(\tau_L)$ since they are generally different from the parameters and error term in the mixed frequency naïve regression model (2.4). We refrain from doing so for the sake of notational brevity.

A low frequency Wald statistic W_{LF} is simply a classic Wald statistic with respect to the restriction $\boldsymbol{\beta} \equiv [\beta_1, \ldots, \beta_h]' = \mathbf{0}_{h\times 1}$. Consistency requires $W_{LF} \xrightarrow{p} \infty$ whenever there is high-to-low non-causality $\boldsymbol{b} \neq \mathbf{0}_{pm\times 1}$ in (2.3). We present a counter-example where high-to-low Granger causality exists such that $\boldsymbol{b} \neq \mathbf{0}_{pm\times 1}$, yet in the LF model (2.13) $\boldsymbol{\beta}^* = \mathbf{0}_{h\times 1}$.

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

$$\begin{bmatrix} x_H(\tau_L,1) \\ x_H(\tau_L,2) \\ x_L(\tau_L) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ b_2 & b_1 & 0 \end{bmatrix} \begin{bmatrix} x_H(\tau_L-1,1) \\ x_H(\tau_L-1,2) \\ x_L(\tau_L-1) \end{bmatrix} + \begin{bmatrix} \epsilon_H(\tau_L,1) \\ \epsilon_H(\tau_L,2) \\ \epsilon_L(\tau_L) \end{bmatrix}, \ \boldsymbol{\epsilon}(\tau_L) \overset{mds}{\sim} (\mathbf{0}_{3\times 1}, \boldsymbol{I}_{3\times 1}), \ \boldsymbol{b} \neq \mathbf{0}_{2\times 1}. \quad (2.14)$$

The linear aggregation scheme when $m = 2$ is $x_H(\tau_L) = \delta_1 x_H(\tau_L, 1) + (1-\delta_1)x_H(\tau_L, 2)$.

Test consistency requires that $\boldsymbol{\beta}^* \neq \mathbf{0}_{h\times 1}$ for any deviation $\boldsymbol{b} \neq \mathbf{0}_{2\times 1}$ from the null hypothesis. Below we show that, for any given aggregation scheme δ_1, there exists $\boldsymbol{b} \neq \mathbf{0}_{2\times 1}$ such that $\boldsymbol{\beta}^* = \mathbf{0}_{h\times 1}$ for any lag length h. Given (2.14) it follows that:[10]

$$\underline{\boldsymbol{C}} \equiv E\left[\underline{\boldsymbol{x}}(\tau_L - 1)\boldsymbol{X}_H(\tau_L - 1)'\right] = \begin{bmatrix} 0 & 1-\delta_1 & 0 & \cdots & 0 \\ 0 & \delta_1 & 0 & \cdots & 0 \end{bmatrix}'$$

Once we choose δ_1 we can find a deviation from the null $\boldsymbol{b} \neq \mathbf{0}_{2\times 1}$ such that $\underline{\boldsymbol{C}}\boldsymbol{b} = \mathbf{0}_{(h+1)\times 1}$. Simply let $b_1 = 0$ and $b_2 \neq 0$ if $\delta_1 = 0$; let $b_1 \neq 0$ and $b_2 = -b_1(1-\delta_1)/\delta_1$ if $\delta_1 \in (0, 1)$; or let $b_1 \neq 0$ and $b_2 = 0$ if $\delta_1 = 1$. In each case $\underline{\boldsymbol{C}}\boldsymbol{b} = \mathbf{0}_{(h+1)\times 1}$ hence $\boldsymbol{\beta}^* = \mathbf{0}_{h\times 1}$ in view of (2.13).

An intuition behind the above choice of $\boldsymbol{b}$ is that 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. Hence, high-to-low causal effects are offset by each other after temporal aggregation.

2.2.2 Low Frequency Parsimonious Regression : Max Test

Now consider regressing x_L onto its own low frequency lags and only one low frequency lag of aggregated x_H:

$$\begin{aligned} x_L(\tau_L) &= \sum_{k=1}^{q} \alpha_{k,j} x_L(\tau_L - k) + \beta_j x_H(\tau_L - j) + u_{L,j}(\tau_L) \\ &= \underbrace{[\boldsymbol{X}_L^{(q)}(\tau_L - 1)', x_H(\tau_L - j)]}_{\equiv \underline{\boldsymbol{x}}_j(\tau_L-1)'} \underbrace{\begin{bmatrix} \alpha_{1,j} \\ \vdots \\ \alpha_{q,j} \\ \beta_j \end{bmatrix}}_{\equiv \boldsymbol{\theta}_j^{(LF)}} + u_{L,j}(\tau_L), \quad j = 1, \ldots, h. \end{aligned} \quad (2.15)$$

[10] Equation (2.14) immediately implies that $x_L(\tau_L - 1) = b_1\epsilon_H(\tau_L - 2, 2) + b_2\epsilon_H(\tau_L - 2, 1) + \epsilon_L(\tau_L - 1)$ and therefore $E[x_L(\tau_L - 1)x_H(\tau_L - 1, 2)] = b_1 E[\epsilon_H(\tau_L - 2, 2)\epsilon_H(\tau_L - 1, 2)] + b_2 E[\epsilon_H(\tau_L - 2, 1)\epsilon_H(\tau_L - 1, 2)] + \epsilon_L(\tau_L - 1)\epsilon_H(\tau_L - 1, 2) = 0$. Similarly, $E[x_L(\tau_L - 1)x_H(\tau_L - 1, 1)] = 0$. In addition, assuming a general linear aggregation scheme, $E[x_H(\tau_L - j)x_H(\tau_L - 1, 2)] = E[(\delta_1 x_H(\tau_L - j, 1) + (1 - \delta_1)x_H(\tau_L - j, 2))x_H(\tau_L - 1, 2)] = (1 - \delta_1)I(j = 1)$. Similarly, $E[x_H(\tau_L - j)x_H(\tau_L - 1, 1)] = \delta_1 I(j = 1)$. Therefore, the second row of $\underline{\boldsymbol{C}}$ is $[1 - \delta_1, \delta_1]$ and all other rows are zeros.

Under the mixed frequency DGP (2.1) the pseudo-true value $\boldsymbol{\theta}_j^{(LF)*}$ for $\boldsymbol{\theta}_j^{(LF)}$ can be easily derived by replacing $\boldsymbol{x}_j(\tau_L-1)'$ with $\underline{\boldsymbol{x}}_j(\tau_L-1)'$ in (2.10):

$$\boldsymbol{\theta}_j^{(LF)*} \equiv \begin{bmatrix} \alpha_{1,j}^* \\ \vdots \\ \alpha_{p,j}^* \\ \alpha_{p+1,j}^* \\ \vdots \\ \alpha_{q,j}^* \\ \beta_j^* \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_p \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} + \underbrace{\left[E\left[\underline{\boldsymbol{x}}_j(\tau_L-1)\underline{\boldsymbol{x}}_j(\tau_L-1)'\right]\right]^{-1}}_{\equiv \underline{\boldsymbol{\Gamma}}_{j,j}^{-1}:\ (q+1)\times(q+1)} \underbrace{E\left[\underline{\boldsymbol{x}}_j(\tau_L-1)\boldsymbol{X}_H(\tau_L-1)'\right]}_{\equiv \underline{\boldsymbol{C}}_j:\ (q+1)\times pm} \boldsymbol{b}. \tag{2.16}$$

The low frequency max test statistic is constructed in the same way as (2.7):

$$\hat{\mathcal{T}}^{(LF)} \equiv \max_{1\le j\le h} (\sqrt{T_L} w_{T_L,j} \hat{\beta}_j)^2.$$

The limit distribution of $\hat{\mathcal{T}}^{(LF)}$ under $H_0 : \boldsymbol{b} = \boldsymbol{0}_{pm\times 1}$ has the same structure as the distribution limit in Theorem 2.1, except that $\underline{\boldsymbol{x}}_j(\tau_L-1)$ replaces $\boldsymbol{x}_j(\tau_L-1)$, hence the Gaussian limit distribution covariance $\boldsymbol{V} \equiv \sigma_L^2 \boldsymbol{W}_h \boldsymbol{R}\boldsymbol{\mathcal{S}}\boldsymbol{R}'\boldsymbol{W}_h \in \mathbb{R}^{h\times h}$ in (2.8) is now defined with a different $\boldsymbol{\mathcal{S}}$ based on $\underline{\boldsymbol{x}}_j(\tau_L-1)$. In the spirit of Example 2.3, we can easily show that $\hat{\mathcal{T}}^{(LF)}$ is inconsistent: asymptotic power is not one in all deviations from the null hypothesis for all linear aggregation schemes.

3 Low-to-High Frequency Data Granger Causality

We now consider testing for Granger causality from low frequency x_L to high frequency x_H, both in mixed and low frequency settings. The null hypothesis based on model (2.1) is $H_0 : \boldsymbol{c} = \boldsymbol{0}_{pm\times 1}$. Under the MF-VAR($p$) data generating process (2.1) we derive Wald and max test statistics, and discuss an additional dimension reduction step based on MIDAS polynomials.

3.1 Mixed Frequency Approach

A natural extension of Sims' (1972) two-sided regression model to the mixed frequency framework allows for a simple Wald test. The model regresses x_L onto q low frequency lags of x_L, h high frequency lags of x_H, and $r \ge 1$ high frequency leads of x_H:

$$x_L(\tau_L) = \sum_{k=1}^{q} \alpha_k x_L(\tau_L - k) + \sum_{j=1}^{h} \beta_j x_H(\tau_L - 1, m+1-j) + \sum_{j=1}^{r} \gamma_j x_H(\tau_L + 1, j) + u_L(\tau_L). \tag{3.1}$$

Low-to-high non-causality $\boldsymbol{c} = \boldsymbol{0}_{pm\times 1}$ in model (2.1) therefore implies $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_r]' = \boldsymbol{0}_{r\times 1}$. Under $H_0 : \boldsymbol{c} = \boldsymbol{0}_{pm\times 1}$ a Wald statistic derived from a least squares estimator of $\boldsymbol{\gamma}$ has a χ_r^2 limit distribution under Assumptions 2.1-2.4, as long as $q \ge p$ and $h \ge pm$ which ensures (3.1) contains the true DGP.

3.1.1 Parsimonious Regressions: Max Test

Parsimonious regression models inspired by (3.1) are, for $j = 1, \ldots, r$:

$$x_L(\tau_L) = \sum_{k=1}^{q} \alpha_{k,j} x_L(\tau_L - k) + \sum_{k=1}^{h} \beta_{k,j} x_H(\tau_L - 1, m + 1 - k) + \gamma_j x_H(\tau_L + 1, j) + u_{L,j}(\tau_L) \quad (3.2)$$

hence only the j^{th} high frequency lead of x_H is included. As before, we abuse notation since under causation γ_j in (3.1) and (3.2) are generally not equivalent.

Let $n \equiv q + h + 1$ denote the number of regressors in each model, and define $n \times 1$ vectors $\boldsymbol{y}_j(\tau_L - 1) = [x_L(\tau_L - 1), \ldots, x_L(\tau_L - q), x_H(\tau_L - 1, m + 1 - 1), \ldots, x_H(\tau_L - 1, m + 1 - h), x_H(\tau_L + 1, j)]'$ and $\boldsymbol{\phi}_j = [\alpha_{1,j}, \ldots, \alpha_{q,j}, \beta_{1,j}, \ldots, \beta_{h,j}, \gamma_j]'$. Therefore $\boldsymbol{y}_j(\tau_L - 1)$ is a vector of all regressors, and $\boldsymbol{\phi}_j$ is a vector of all parameters in model j, hence we can write:

$$x_L(\tau_L) = \boldsymbol{y}_j(\tau_L - 1)' \boldsymbol{\phi}_j + u_{L,j}(\tau_L).$$

Now stack the least squares estimator $\hat{\gamma}_j$ for γ_j into $\hat{\boldsymbol{\gamma}} = [\hat{\gamma}_1, \ldots, \hat{\gamma}_r]'$. Low-to-high non-causality $H_0 : \boldsymbol{c} = \boldsymbol{0}_{pm \times 1}$ implies $\boldsymbol{\gamma} = \boldsymbol{0}_{r \times 1}$ for any $r \geq 1$, which justifies a mixed frequency max test statistic for low-to-high causality:

$$\hat{\mathcal{U}} \equiv \max_{1 \leq j \leq r} \left(\sqrt{T_L} w_{T_L,j} \hat{\gamma}_j \right)^2. \quad (3.3)$$

The asymptotic null distribution of $\hat{\mathcal{U}}$ can be derived in the same way as in Theorem 2.1 under Assumptions 2.1-2.5, hence the proof is omitted.

Theorem 3.1. Let Assumptions 2.1-2.5 hold. Under $H_0 : \boldsymbol{c} = \boldsymbol{0}_{pm \times 1}$ we have that $\hat{\mathcal{U}} \xrightarrow{d} \max_{1 \leq j \leq r} \tilde{\mathcal{N}}_j^2$ as $T_L \to \infty$, where $\tilde{\boldsymbol{\mathcal{N}}} \equiv [\tilde{\mathcal{N}}_1, \ldots, \tilde{\mathcal{N}}_r]'$ is distributed $N(\boldsymbol{0}_{r \times 1}, \tilde{\boldsymbol{V}})$ with positive definite covariance matrix: $\tilde{\boldsymbol{V}} \equiv \sigma_L^2 \boldsymbol{W}_r \tilde{\boldsymbol{R}} \tilde{\boldsymbol{\mathcal{S}}} \tilde{\boldsymbol{R}}' \boldsymbol{W}_r \in \mathbb{R}^{r \times r}$, where $\sigma_L^2 \equiv E[\epsilon_L^2(\tau_L)]$; $\tilde{\boldsymbol{\mathcal{S}}}$ is defined the same way as $\boldsymbol{\mathcal{S}}$ in (2.9) by replacing the regressors $\boldsymbol{x}_j(\tau_L - 1)$ with $\boldsymbol{y}_j(\tau_L - 1)$; and selection matrix $\tilde{\boldsymbol{R}}$ is a r-by-$(q + h + 1)r$ matrix that picks $[\gamma_1, .., \gamma_r]'$ out of $[\boldsymbol{\phi}_1', ..., \boldsymbol{\phi}_r']'$.

Remark 3.1. We require Assumption 2.5, such that the number of high frequency lags h in models (3.1) and (3.2) is at least as large as the true lag length pm, in order to ensure that the true DGP is contained in the two-sided models (3.2) under the null hypothesis of no causation from low-to-high frequency: under no causation $\hat{\boldsymbol{\gamma}} \xrightarrow{p} 0$ only if the model is otherwise correctly specified vis-à-vis DGP (2.1). Conversely, the high-to-low frequency max test limit distribution in Theorem 2.1 applies without $h \geq pm$ precisely because for *any* $h \geq 1$ the coefficients $\boldsymbol{\beta}$ are identically 0 under the null of no causality from high-to-low frequency. We imposed $h \geq pm$ solely to deduce by Corollary 2.6 that the limit distribution applies *if and only if* no causation from high-to-low frequency is true.

Remark 3.2. In general, it cannot be shown that $\boldsymbol{\gamma} \neq \boldsymbol{0}_{r \times 1}$ in (3.2) follows under low-to-high causality $\boldsymbol{c} \neq \boldsymbol{0}_{pm \times 1}$, even if $r \geq pm$. Therefore, consistency of the low-to-high Wald and max

tests is an open question, and evidently not yet resolved by our methods.

3.1.2 MIDAS Polynomials in the Max Test

In a low-to-high frequency causality test the max statistic only operates on the *lead* parameters γ_j, while our simulation study reveals a large *lag* h can prompt size distortions. In general a comparatively large low frequency sample size is needed for the max test empirical size to be very close to the nominal level.[11]

One option is to use a bootstrap procedure for p-value computation, but we find that a wild bootstrap similar to Gonçalves and Killian's (2004) does not alleviate size distortions.

Another approach is to exploit a MIDAS polynomial for the high-to-low causality part in order to reduce the impact of large h, and keep the low-to-high causality part unrestricted (cfr. Ghysels, Santa-Clara, and Valkanov (2006), Ghysels, Sinko, and Valkanov (2007), among others). The model now becomes:

$$x_L(\tau_L) = \sum_{k=1}^{q} \alpha_k x_L(\tau_L - k) + \sum_{k=1}^{h} \omega_k(\boldsymbol{\pi}) x_H(\tau_L - 1, 12 + 1 - k) + \gamma_j x_H(\tau_L + 1, j) + u_L(\tau_L),\ j = 1, \ldots, r, \quad (3.4)$$

where $\omega_k(\boldsymbol{\pi})$ represents a MIDAS polynomial with a parameter vector $\boldsymbol{\pi} \in \mathbb{R}^s$ of small dimension $s < h$.

There are a variety of possible polynomials in the literature (see e.g. Technical Appendix A of Ghysels (2014)). In our simulation study we use the Almon polynomial $\omega_k(\boldsymbol{\pi}) = \sum_{l=1}^{s} \pi_l k^l$, hence model (3.4) is linear in $\boldsymbol{\pi}$, allowing for least squares estimation. Another important characteristic of the Almon polynomial is that it allows negative and positive values in general (e.g. $w_k(\boldsymbol{\pi}) \geq 0$ for $k < 3$ and $w_k(\boldsymbol{\pi}) < 0$ for $k \geq 4$, etc.). Many other MIDAS polynomials, like the beta probability density or exponential Almon, assume a single sign for all lags.

MIDAS regressions, of course, may be misspecified. Therefore, the least squares estimator of $\boldsymbol{\gamma}$ may not be consistent for 0 under the null, but rather may be consistent for some non-zero pseudo-true value identified by the resulting first order moment conditions. Nevertheless, we show that a model with mis-specified MIDAS polynomials leads to a dramatic improvement in empirical size, even though the max test statistic for that model does not have its intended null limit distribution. We also show that size distortions vanish with a large enough sample size (cfr. Footnote 11).

[11] In our simulation study where $m = 12$, we find $T_L \in \{40, 80\}$ is not large enough but $T_L \geq 120$ is large enough for sharp max test empirical size. If the low frequency is years, such that there are $m = 12$ high frequency months, then $T_L = 120$ years is obviously too large for practical applications in macroeconomics and finance, outside of deep historical studies. If the low frequency is quarters such that the high frequency is approximately $m = 12$ weeks, then $T_L = 120$ quarters, or 30 years, is reasonable.

3.2 Low Frequency Approach

Consider a low frequency counterpart to the parsimonious regression models (3.2), with aggregated high frequency variable $x_H(\tau_L) = \sum_{j=1}^{m} \delta_j x_H(\tau_L, j)$:

$$x_L(\tau_L) = \sum_{k=1}^{q} \alpha_{k,j} x_L(\tau_L - k) + \sum_{k=1}^{h_{LF}} \beta_{k,j} x_H(\tau_L - k) + \gamma_j x_H(\tau_L + j) + u_{L,j}(\tau_L), \quad j = 1, \ldots, r_{LF}. \quad (3.5)$$

The subscript "LF" on h and r emphasizes that these are the number of *low frequency* lags and leads of aggregated x_H. We estimate the parsimonious model (3.5) by least squares, and use a low frequency max test statistic as in (3.3):

$$\hat{\mathcal{U}}^{(LF)} \equiv \max_{1 \le j \le r_{LF}} (\sqrt{T_L} w_{T_L,j} \hat{\gamma}_j)^2.$$

Deriving the limit distribution of $\hat{\mathcal{U}}^{(LF)}$ under $H_0 : x_L \nrightarrow x_H$ requires an assumption in addition to Assumptions 2.1-2.5. Under the null hypothesis of low-to-high non-causality, a correctly specified two-sided MF regression reduces to (2.3). In general, each low frequency parsimonious regression model (3.5) does *not* contain (2.3) as a special case. The true high-to-low causal pattern based on the non-aggregated x_H, i.e. $\sum_{l=1}^{pm} b_l x_H(\tau_L - 1, m + 1 - l)$, may not be fully captured by the low frequency lags of aggregated x_H, i.e. $\sum_{k=1}^{h_{LF}} \beta_{k,j} x_H(\tau_L - k)$, no matter what the lag length h_{LF} is. See Examples 3.1 and 3.2 below.

In order to find a condition that ensures each low frequency parsimonious regression model contains (2.3) as a special case, we elaborate the relationship between the (non)aggregated causal terms $\sum_{l=1}^{pm} b_l x_H(\tau_L - 1, m + 1 - l)$ and $\sum_{k=1}^{h_{LF}} \beta_{k,j} x_H(\tau_L - k)$. In the following we write β_k instead of $\beta_{k,j}$ since it is irrelevant which j^{th} lead term of x_H is included in the model. Observe that the aggregated high frequency variable is:

$$\begin{aligned}
&\sum_{k=1}^{h_{LF}} \beta_k x_H(\tau_L - k) = \sum_{k=1}^{h_{LF}} \beta_k \sum_{l=1}^{m} \delta_l x_H(\tau_L - k, l) \\
&= \beta_1 \delta_m x_H(\tau_L - 1, m + 1 - 1) + \cdots + \beta_1 \delta_1 x_H(\tau_L - 1, m + 1 - m) \\
&\quad + \cdots + \beta_{h_{LF}} \delta_m x_H(\tau_L - h_{LF}, m + 1 - 1) + \cdots + \beta_{h_{LF}} \delta_1 x_H(\tau_L - h_{LF}, m + 1 - m) \\
&= \sum_{l=1}^{h_{LF} m} \beta_{\lceil l/m \rceil} \delta_{\lceil l/m \rceil m + 1 - l} x_H(\tau_L - 1, m + 1 - l),
\end{aligned} \quad (3.6)$$

where $\lceil z \rceil$ is the smallest integer not smaller than z. The last equality exploits the notational convention that the second argument of x_H can go below 1 (see Appendix A). Now compare the last term in (3.6) with the true high-to-low causal pattern $\sum_{l=1}^{pm} b_l x_H(\tau_L - 1, m + 1 - l)$. The following assumption gives a sufficient property for the true $\boldsymbol{b}$ in model (2.3) in order for the parsimonious regressions with aggregated x_H to contain the true DGP under the null.

Assumption 3.1. Fix the linear aggregation scheme $\boldsymbol{\delta} = [\delta_1, \ldots, \delta_m]'$, and fix the true causal pattern from x_H to x_L: $\boldsymbol{b} = [b_1, \ldots, b_{pm}]'$. There exists $\boldsymbol{\beta}^* = [\beta_1^*, \ldots, \beta_p^*]'$ such that $b_l =$

$\beta^*_{\lceil l/m \rceil}\delta_{\lceil l/m \rceil m+1-l}$ for all $l \in \{1, \ldots, pm\}$ provided a sufficiently large lag length $h_{LF} \geq p$ is chosen.

Remark 3.3. Assumption 3.1 is in some sense a low frequency version of Assumption 2.5, but with a deeper restriction that the DGP *can* be aggregated and still retain identification of underlying causal patterns. It ensures that there exists a pseudo-true $\boldsymbol{\beta}^*$ such that $\sum_{k=1}^{h_{LF}} \beta^*_k$ $x_H(\tau_L - k) = \sum_{l=1}^{pm} b_l \; x_H(\tau_L - 1, m + 1 - l)$, in which case each parsimonious regression model (3.5) is correctly specified under $H_0 : x_L \nrightarrow x_H$. Therefore, for a given aggregation scheme $\boldsymbol{\delta}$ it assumes the DGP itself, and therefore $\boldsymbol{b} = [b_1, \ldots, b_{pm}]'$, allows for identification of the DGP under low-to-high non-causality using an aggregated high frequency variable x_H.

If there is high-to-low non-causality, that is $\boldsymbol{b} = \mathbf{0}_{pm\times 1}$ in (2.3), then Assumption 3.1 is trivially satisfied by choosing any $h_{LF} \in \mathbb{N}$ and letting $\beta^*_k = 0$ for all $k \in \{1, \ldots, h_{LF}\}$. Under high-to-low causality $\boldsymbol{b} \neq \mathbf{0}_{pm\times 1}$, however, Assumption 3.1 is a relatively stringent restriction on the DGP. The following examples show that some DGP's *cannot* satisfy Assumption 3.1, in particular that a low frequency test may not be able to reveal whether there is low-to-high frequency causation.

Example 3.1 (Causality with Stock Sampling)**.** Assume the sampling frequency ratio is $m = 3$, the AR order is $p = 2$, and in model (2.3) consider lagged causality $b_l = b \times I(l = 4)$ for $l \in \{1, \ldots, 6\}$ with $b \neq 0$. This causal pattern can be captured by the low frequency parsimonious regression models *if and only if* the aggregation scheme is stock sampling.

A proof of this claim is as follows. Since stock sampling is represented as $\delta_l = I(l = 3)$ for $l \in \{1, 2, 3\}$, the summation term included in each parsimonious regression models, $\sum_{k=1}^{h_{LF}} \beta_k x_H(\tau_L - k)$, can be rewritten as $\sum_{k=1}^{h_{LF}} \beta_k x_H(\tau_L - k, 3)$. Therefore, we can simply choose $h_{LF} = 2$, $\beta^*_1 = 0$, and $\beta^*_2 = b$ to replicate the true causal pattern.

Conversely, assume $\delta_l \geq 0$ for all $l \in \{1, 2, 3\}$, $\delta_3 < 1$, and $\sum_{l=1}^{3} \delta_l = 1$, which allows for any aggregation *except* stock sampling. Since $\delta_3 < 1$, either δ_1 and/or δ_2 should have a positive value, so assume $\delta_1 > 0$ without loss of generality. Assumption 3.1 therefore requires $b_4 = \beta^*_2\delta_3$ and $b_6 = \beta^*_2\delta_1$, but the true causal pattern implies $b_4 = b \neq 0$ and $b_6 = 0$. Since $\delta_1 > 0$, there does not exist any β^*_2 that satisfies all four equalities.

Example 3.2 (Flow Sampling)**.** Under flow sampling we have $\delta_l = 1/m$ for all $l \in \{1, \ldots, m\}$, in which case Assumption 3.1 requires $b_l = \beta^*_{\lceil l/m \rceil}/m$, hence $b_1 = \cdots = b_m$, $b_{m+1} = \cdots = b_{2m}$, and so on. In other words, Assumption 3.1 holds only when all m high frequency lags of x_H in each low frequency period have an identical coefficient. This is quite severe since empirical evidence (here and elsewhere) points to more nuanced patterns of inter-period dynamics, including parameter values with different signs, lagged causality (some parameters values are zeros), or decaying causality (parameter values decline).

The limit distribution of $\hat{\mathcal{U}}^{(LF)}$ is straightforward to derive under $H_0 : x_L \nrightarrow x_H$. Define an

$n \times 1$ vector of all regressors in model j:

$$\underline{\boldsymbol{y}}_j(\tau_L - 1) \equiv [x_L(\tau_L - 1), \ldots, x_L(\tau_L - q), x_H(\tau_L - 1), \ldots, x_H(\tau_L - h), x_H(\tau_L + j)]'.$$

Under Assumptions 2.1 - 2.4 and 3.1, the asymptotic distribution of $\hat{\mathcal{U}}^{(LF)}$ under $H_0 : x_L \nrightarrow x_H$ is identical to that in Theorem 3.1, except we replace regressors $\boldsymbol{y}_j(\tau_L - 1)$ with $\underline{\boldsymbol{y}}_j(\tau_L - 1)$.

4 Local Asymptotic Power Analysis for Tests of High-to-Low Non-Causality

The results of Section 2 characterize the asymptotic global power properties of MF and LF Wald and max tests of high-to-low frequency non-causality. The MF max and Wald tests are consistent as long as the selected number of high frequency lags h is larger than or equal to the true lag order pm. Conversely, the LF tests are sensitive to the chosen aggregation scheme: for some DGP's and aggregation schemes power is trivial, hence these tests are not generally consistent. In this section we study the local power properties of each test for the high-to-low case. We do not treat the low-to-high frequency case since identification of causation within Sims' (1972) two-sided regression model is unresolved.

As Example 2.3 suggests, the LF tests have asymptotic power of one in some cases depending on the aggregation scheme and DGP, even though these tests do not have asymptotic power of one against *all* deviations from non-causality. An advantage of LF tests of course is that they require fewer parameters than MF tests, hence in some cases local power for LF tests may be actually *higher* than for MF tests.

We impose Assumptions 2.1-2.4, and consider the usual mixed frequency DGP (2.1). The high-to-low non-causality null hypothesis is $H_0 : \boldsymbol{b} = \boldsymbol{0}_{pm\times 1}$, hence the local alternative hypothesis under regular asymptotics is

$$H_1^L : \boldsymbol{b} = \boldsymbol{\nu}/\sqrt{T_L},$$

where $\boldsymbol{\nu} = [\nu_1, \ldots, \nu_{pm}]'$ is the drift parameter. Under H_1^L, model (2.3) becomes

$$x_L(\tau_L) = \boldsymbol{X}_L(\tau_L - 1)'\boldsymbol{a} + \boldsymbol{X}_H(\tau_L - 1)'\left(\frac{1}{\sqrt{T_L}}\boldsymbol{\nu}\right) + \epsilon_L(\tau_L) \tag{4.1}$$

where as before $\boldsymbol{X}_L(\tau_L - 1) = [x_L(\tau_L - 1), \ldots, x_L(\tau_L - p)]'$, $\boldsymbol{X}_H(\tau_L - 1) = [x_H(\tau_L - 1, m + 1 - 1), \ldots, x_H(\tau_L - 1, m + 1 - pm)]'$, and $\boldsymbol{a} = [a_1, \ldots, a_p]'$.

4.1 Local Power for Tests with Mixed Frequency Data

4.1.1 Mixed Frequency Max Test

Our first result gives the asymptotic distribution of the max statistic under H_1^L. Define covariance matrices $\boldsymbol{\Gamma}_{j,j} \equiv E[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{x}_j(\tau_L - 1)']$ and $\boldsymbol{C}_j \equiv E[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{X}_H(\tau_L - 1)']$), and recall the

weighting scheme $\boldsymbol{W}_h$ in (2.7) and selection matrix $\boldsymbol{R}$ in (2.9).

Theorem 4.1. Under Assumptions 2.1-2.4 and H_1^L we have $\hat{\mathcal{T}} \xrightarrow{d} \max_{1\leq i\leq h} \mathcal{M}_i^2$ as $T_L \to \infty$, where $\boldsymbol{\mathcal{M}} = [\mathcal{M}_1, \ldots, \mathcal{M}_h]'$ is distributed $N(\boldsymbol{\mu}, \boldsymbol{V})$, $\boldsymbol{V}$ is defined in (2.8), and

$$\boldsymbol{\mu} = \boldsymbol{W}_h \boldsymbol{R} \begin{bmatrix} \boldsymbol{\Gamma}_{1,1}^{-1} \boldsymbol{C}_1 \\ \vdots \\ \boldsymbol{\Gamma}_{h,h}^{-1} \boldsymbol{C}_h \end{bmatrix} \boldsymbol{\nu} \in \mathbb{R}^{h\times 1}. \tag{4.2}$$

Proof. See Appendix F.

In order to compute local power we need the Gaussian law mean $\boldsymbol{\mu}$ and therefore explicit characterizations of $\boldsymbol{\Gamma}_{j,i}$ and $\boldsymbol{C}_j$ in terms of underlying parameters $(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_p)$ and $\boldsymbol{\Omega}$ in (2.1).

Lemma 4.2. Let Assumptions 2.1 and 2.2 hold, and define $f(j) \equiv \lceil (j-m)/m \rceil$ and $g(j) \equiv mf(j) + m + 1 - j$, where $\lceil x \rceil$ is the smallest integer not smaller than x. Let $\Upsilon_k(s,t)$ be the (s,t) element of $\boldsymbol{\Upsilon}_k \equiv E[\boldsymbol{X}(\tau_L)\boldsymbol{X}(\tau_L - k)']$. Then, for $j, i \in \{1, \ldots, h\}$:

$$\underbrace{\boldsymbol{\Gamma}_{j,i}}_{(q+1)\times(q+1)} = \begin{bmatrix} \Upsilon_{1-1}(K,K) & \ldots & \Upsilon_{1-q}(K,K) & \Upsilon_{-f(i)}(g(i),K) \\ \vdots & \ddots & \vdots & \vdots \\ \Upsilon_{q-1}(K,K) & \ldots & \Upsilon_{q-q}(K,K) & \Upsilon_{(q-1)-f(i)}(g(i),K) \\ \Upsilon_{f(j)}(K,g(j)) & \ldots & \Upsilon_{f(j)-(q-1)}(K,g(j)) & \Upsilon_{f(j)-f(i)}(g(i),g(j)) \end{bmatrix},$$

$$\underbrace{\boldsymbol{C}_j}_{(q+1)\times pm} = \begin{bmatrix} \Upsilon_{f(1)}(K,g(1)) & \ldots & \Upsilon_{f(pm)}(K,g(pm)) \\ \vdots & \ddots & \vdots \\ \Upsilon_{f(1)-(q-1)}(K,g(1)) & \ldots & \Upsilon_{f(pm)-(q-1)}(K,g(pm)) \\ \Upsilon_{f(j)-f(1)}(g(1),g(j)) & \ldots & \Upsilon_{f(j)-f(pm)}(g(pm),g(j)) \end{bmatrix}. \tag{4.3}$$

Proof. See Appendix G.

Remark 4.1. Considering that $\boldsymbol{\Upsilon}_k$ is $(m+1)\times(m+1)$, it must be the case that $g(j)$ takes a natural number between 1 and $m+1$ for any $j \in \mathbb{N}$. We verify this in Table T.1 of the supplemental material Ghysels, Hill, and Motegi (2015). By definition, $f(j)$ is a step function taking 0 for $j = 1, \ldots, m$, 1 for $j = m+1, \ldots, 2m$, 2 for $j = 2m+1, \ldots, 3m$, etc. $g(j)$ takes m, $m-1$, …, 1 as j runs from $(k-1)m+1$ to km for any $k \in \mathbb{N}$. Hence, the matrices in (4.3) are always well-defined.

Remark 4.2. Let $\boldsymbol{\Upsilon}$ be a $pK \times pK$ matrix whose (i,j) block is $\boldsymbol{\Upsilon}_{j-i}$ for $i, j \in \{1, \ldots, p\}$. Let

$\bar{\boldsymbol{\Omega}}$ be a $pK \times pK$ matrix whose $(1,1)$ block is $\boldsymbol{\Omega}$ and all other blocks are $\mathbf{0}_{K\times K}$. Define

$$\boldsymbol{A} \equiv \begin{bmatrix} \boldsymbol{A}_1 & \dots & \boldsymbol{A}_{p-1} & \boldsymbol{A}_p \\ \boldsymbol{I}_K & \dots & \mathbf{0}_{K\times K} & \mathbf{0}_{K\times K} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_{K\times K} & \dots & \boldsymbol{I}_K & \mathbf{0}_{K\times K} \end{bmatrix} \in \mathbb{R}^{pK\times pK}.$$

Using the discrete Lyapunov equation, $\boldsymbol{\Upsilon}_k$ for $k \in \{1-p, \dots, p-1\}$ can be computed by the following well-known formula:

$$\text{vec}[\boldsymbol{\Upsilon}] = \left(\boldsymbol{I}_{(pK)^2} - \boldsymbol{A} \otimes \boldsymbol{A}\right)^{-1} \text{vec}[\bar{\boldsymbol{\Omega}}], \tag{4.4}$$

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

Local asymptotic power can be easily computed numerically for a given DGP. See Section 4.3 for numerical experiments.

Step 1 Calculate $\boldsymbol{\mu}$ and $\boldsymbol{V}$ from underlying parameters $(\boldsymbol{A}_1, \dots, \boldsymbol{A}_p)$, and $\boldsymbol{\Omega}$ in model (2.1) by using Theorems 2.1 and 4.1, and Lemma 4.2.

Step 2 Draw random vectors $\{[\boldsymbol{\mathcal{N}}_i^{(r)}]_{i=1}^h\}_{r=1}^{R_1}$ independently from $N(\mathbf{0}_{h\times 1}, \boldsymbol{V})$, and calculate test statistics $\mathcal{T}_r = \max_{1\leq i\leq h}(\boldsymbol{\mathcal{N}}_i^{(r)})^2$, where R_1 is a large integer. The $100(1-\alpha)\%$ empirical quantile $q_{R_1}^{\alpha}$ of $\{\mathcal{T}_r\}_{r=1}^{R_1}$ is an asymptotically valid approximation of the α-level asymptotic critical value of $\hat{\mathcal{T}}$.

Step 3 Draw random vectors $\{[\boldsymbol{\mathcal{M}}_i^{(r)}]_{i=1}^h\}_{r=1}^{R_2}$ independently from $N(\boldsymbol{\mu}, \boldsymbol{V})$, and calculate test statistics $\tilde{\mathcal{T}}_r = \max_{1\leq i\leq h}(\boldsymbol{\mathcal{M}}_i^{(r)})^2$, where R_2 is a large integer. Empirical local asymptotic power is computed as $\hat{\mathcal{P}} \equiv (1/R_2)\sum_{r=1}^{R_2} I(\tilde{\mathcal{T}}_r > q_{R_1}^{\alpha})$.

By construction $q_{R_1}^{\alpha}$ estimates the $100(1-\alpha)\%$ quantile q^{α} of the max test statistic limit law $\max_{1\leq i\leq h}\mathcal{N}_i^2$. By independence of the sample draws, $q_{R_1}^{\alpha} \xrightarrow{p} q^{\alpha}$ as $R_1 \to \infty$. Indeed, since we can choose $\{R_1, R_2\}$ to be arbitrarily large, and the samples are independently drawn, empirical power $\hat{\mathcal{P}}$ can be made arbitrarily close to local asymptotic power $\lim_{T_L\to\infty} P(\hat{\mathcal{T}} > q^{\alpha}|H_1^L)$ by the Glivenko-Cantelli theorem. See the proof of Theorem 2.2 for related details.

4.1.2 Mixed Frequency Wald Test

Rewrite model (2.4) in matrix form:

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

$$= \begin{bmatrix} \boldsymbol{X}_L^{(q)}(\tau_L - 1)' & \boldsymbol{X}_H^{(h)}(\tau_L - 1)' \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} + u_L(\tau_L) = \boldsymbol{x}(\tau_L - 1)'\boldsymbol{\Theta} + u_L(\tau_L). \tag{4.5}$$

$\boldsymbol{X}_H^{(h)}(\tau_L - 1)$ is a vector stacking h high frequency lags of x_H, while $\boldsymbol{X}_H(\tau_L - 1)$ is a vector stacking pm high frequency lags of x_H.

Let $\hat{W}$ denote the least squares based Wald statistic for testing $H_0 : \boldsymbol{b} = \boldsymbol{0}_{pm\times 1}$, and Assumptions 2.1-2.4 hold. Then $\hat{W}$ is asymptotically $\chi^2_h(\kappa)$ distributed under H_1^L, where $\chi^2_h(\kappa)$ denotes the noncentral χ^2 distribution with degrees of freedom h and noncentrality κ that is a function of drift $\boldsymbol{\nu}$, and the covariance matrices $\boldsymbol{\Gamma} \equiv E[\boldsymbol{x}(\tau_L-1)\boldsymbol{x}(\tau_L-1)']$ and $\boldsymbol{C} \equiv E[\boldsymbol{x}(\tau_L-1)\boldsymbol{X}_H(\tau_L-1)']$. In particular $\kappa = 0$ *if and only if* $\boldsymbol{\nu} = 0$ such that the null is true. The covariances $\boldsymbol{\Gamma}$ and $\boldsymbol{C}$ can be derived in terms of underlying parameters, analogous to Lemma 4.2. See Section B.1 in the supplemental material Ghysels, Hill, and Motegi (2015) for complete derivations of the Wald statistic, covariance matrices, and noncentrality parameter.

Since $\hat{W}$ has asymptotic χ^2 distributions under H_0 and H_1^L, local power of the mixed frequency Wald test is $\mathcal{P} \equiv 1 - F_1[F_0^{-1}(1-\alpha)]$, where $\alpha \in (0,1)$ is the nominal size, F_0 is the asymptotic null χ^2_h distribution, and F_1 is the asymptotic local alternative $\chi^2_h(\kappa)$ distribution. Noncentrality κ is computed from $\boldsymbol{\nu}$, $\boldsymbol{\Gamma}$ and $\boldsymbol{C}$, and therefore from $(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_p)$ and $\boldsymbol{\Omega}$, cfr. Theorem B.1 and Lemma B.2 in the supplemental material.

4.2 Low Frequency Approach

As usual, we impose Assumption 2.4 that $q \geq p$, but we do not assume anything about the magnitude of the number of included high frequency lags h relative to the true lag order pm.

4.2.1 Low Frequency Parsimonious Regression

We compute the low frequency max test statistic $\hat{\mathcal{T}}^{(LF)}$ based on h low frequency parsimonious regression models (2.15): $x_L(\tau_L) = \underline{\boldsymbol{x}}_j(\tau_L - 1)'\boldsymbol{\theta}_j^{(LF)} + u_{L,j}(\tau_L)$, where $\underline{\boldsymbol{x}}_j(\tau_L - 1) \equiv [x_L(\tau_L - 1), \ldots, x_L(\tau_L - q), x_H(\tau_L - j)]'$ with aggregation $x_H(\tau_L - j) \equiv \sum_{l=1}^m \delta_l x_H(\tau_L - j, l)$. The asymptotic distribution of $\hat{\mathcal{T}}^{(LF)}$ under $H_1^L : \boldsymbol{b} = (1/\sqrt{T_L})\boldsymbol{\nu}$ is the same as in Theorem 4.1, except $\boldsymbol{x}_j(\tau_L - 1)$ there is replaced with $\underline{\boldsymbol{x}}_j(\tau_L - 1)$. Computing local power for the low frequency max test therefore requires an analytical characterization of $\underline{\boldsymbol{\Gamma}}_{j,i} = E[\underline{\boldsymbol{x}}_j(\tau_L - 1)\underline{\boldsymbol{x}}_i(\tau_L - 1)']$ and $\underline{\boldsymbol{C}}_j = E[\underline{\boldsymbol{x}}_j(\tau_L - 1)\boldsymbol{X}_H(\tau_L - 1)']$. See Section B.2 in the supplemental material Ghysels, Hill, and Motegi (2015).

4.2.2 Low Frequency Naïve Regression

The low frequency naïve regression model is (2.12): $x_L(\tau_L) = \underline{\boldsymbol{x}}(\tau_L - 1)'\boldsymbol{\theta}^{(LF)} + u_L(\tau_L)$, where $\underline{\boldsymbol{x}}(\tau_L - 1) = [x_L(\tau_L - 1), \ldots, x_L(\tau_L - q), x_H(\tau_L - 1), \ldots, x_H(\tau_L - h)]'$.

Let $\hat{W}^{(LF)}$ denote the Wald statistic for testing $H_0 : \boldsymbol{b} = \boldsymbol{0}_{pm\times 1}$. The asymptotic distribution of $\hat{W}^{(LF)}$ is χ^2_h under H_0, and $\chi^2_h(\underline{\kappa})$ under H_1^L, where noncentrality $\underline{\kappa}$ depends on the covariances $\underline{\boldsymbol{\Gamma}} = E[\underline{\boldsymbol{x}}(\tau_L - 1)\underline{\boldsymbol{x}}(\tau_L - 1)']$ and $\underline{\boldsymbol{C}} = E[\underline{\boldsymbol{x}}(\tau_L - 1)\boldsymbol{X}_H(\tau_L - 1)']$. Complete analytical details are presented in Section B.3 of the supplemental material Ghysels, Hill, and Motegi (2015).

4.3 Numerical Examples

We now compare the local power of MF and LF max and Wald test.

4.3.1 Design

We work with a structural MF-VAR(1) process with $m = 12$ (e.g. the LF increment is one year and the HF increment is one month):

$$\underbrace{\begin{bmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ -d & 1 & \ddots & \ddots & \ddots & 0 \\ 0 & -d & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & -d & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 \end{bmatrix}}_{\equiv \boldsymbol{N}} \underbrace{\begin{bmatrix} x_H(\tau_L, 1) \\ \vdots \\ x_H(\tau_L, 12) \\ x_L(\tau_L) \end{bmatrix}}_{=\boldsymbol{X}(\tau_L)} = \underbrace{\begin{bmatrix} 0 & 0 & \dots & d & c_1 \\ 0 & 0 & \dots & 0 & c_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & c_{12} \\ b_{12} & b_{11} & \dots & b_1 & a \end{bmatrix}}_{\equiv \boldsymbol{M}} \underbrace{\begin{bmatrix} x_H(\tau_L - 1, 1) \\ \vdots \\ x_H(\tau_L - 1, 12) \\ x_L(\tau_L - 1) \end{bmatrix}}_{=\boldsymbol{X}(\tau_L - 1)} + \underbrace{\begin{bmatrix} \eta_H(\tau_L, 1) \\ \vdots \\ \eta_H(\tau_L, 12) \\ \eta_L(\tau_L) \end{bmatrix}}_{\equiv \boldsymbol{\eta}(\tau_L)}, \quad (4.6)$$

where $\boldsymbol{\eta}(\tau_L) \sim (\boldsymbol{0}_{13\times 1}, \boldsymbol{I}_{13})$ is an mds with respect to increasing $\mathcal{F}_{\tau_L} \equiv \sigma(\boldsymbol{X}(\tau) : \tau \leq \tau_L)$. Coefficient a governs the autoregressive property of x_L, d governs the autoregressive property of x_H, $\boldsymbol{c} = [c_1, \dots, c_{12}]'$ represents Granger causality from x_L to x_H, and our interest lies in $\boldsymbol{b} = [b_1, \dots, b_{12}]'$ since it expresses Granger causality from x_H to x_L. Since

$$\boldsymbol{N}^{-1} = \begin{bmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ d & 1 & \ddots & \ddots & \ddots & 0 \\ d^2 & d & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ d^{11} & d^{10} & \dots & d & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 \end{bmatrix} \text{ thus } \boldsymbol{A} \equiv \boldsymbol{N}^{-1}\boldsymbol{M} = \begin{bmatrix} 0 & 0 & \dots & d & \sum_{i=1}^{1} d^{1-i} c_i \\ 0 & 0 & \dots & d^2 & \sum_{i=1}^{2} d^{2-i} c_i \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & d^{12} & \sum_{i=1}^{12} d^{12-i} c_i \\ b_{12} & b_{11} & \dots & b_1 & a \end{bmatrix}, \quad (4.7)$$

the reduced form of (4.6) is $\boldsymbol{X}(\tau_L) = \boldsymbol{A}\boldsymbol{X}(\tau_L - 1) + \boldsymbol{\epsilon}(\tau_L)$, where $\boldsymbol{\epsilon}(\tau_L) = \boldsymbol{N}^{-1}\boldsymbol{\eta}(\tau_L)$ and $\boldsymbol{\Omega} \equiv E[\boldsymbol{\epsilon}(\tau_L)\boldsymbol{\epsilon}(\tau_L)'] = \boldsymbol{N}^{-1}\boldsymbol{N}^{-1'}$.

We consider three types of drift $\boldsymbol{\nu} = [\nu_1, \dots, \nu_{12}]'$. First, $\nu_j = (-1)^{j-1} \times 2.5/j$ for $j = 1, \dots, 12$, hence there is *decaying causality* from x_H to x_L with hyperbolic decay and alternating signs. Second, *lagged causality* with $\nu_j = 2 \times I(j = 12)$ for $j = 1, \dots, 12$, hence only $\nu_{12} \neq 0$. Third, *sporadic causality* with $(\nu_3, \nu_9, \nu_{11}) = (2.1, -2.8, 1.9)$ and all other ν_j's are zeros.[12]

Other parameters in the DGP are as follows. Since local power is not significantly affected by the choice of a, we simply set $a = 0.2$ such that the autoregressive property for x_L is fairly weak. There are two values for the persistence of x_H: $d \in \{0.2, 0.8\}$, and decaying causality with alternating signs for low-to-high causality: $c_j = (-1)^{j-1} \times 0.8/j$ for $j = 1, \dots, 12$.

In the regression models used as the premise for the four tests, we include two low frequency lags of x_L (i.e. $q = 2$), although $q = 1$ would suffice since the true DGP is MF-VAR(1). Max text local power is then computed using draws of 100,000 random variables from the limit distributions under H_0 and H_1^L, and a flat weight: $\boldsymbol{W}_h = (1/h) \times \boldsymbol{I}_h$. We use a flat weight as

[12] Such relationships may exist in macroeconomic processes due to lagged information transmission, seasonality, feedback effects, and ambiguous theoretical relations in terms of signs.

a convention in the absence of information about the relative magnitudes of the parsimonious regression slopes β_j.

The number of high frequency lags of x_H used in the mixed frequency tests is $h_{MF} \in \{4, 8, 12\}$, and the number of low frequency lags of aggregated x_H used in the low frequency tests is $h_{LF} \in \{1, 2, 3\}$. In the low frequency tests, for aggregating x_H we use flow sampling (i.e. $\delta_k = 1/12$ for $k = 1, \ldots, 12$) and stock sampling (i.e. $\delta_k = I(k = 12)$ for $k = 1, \ldots, 12$). Finally, the nominal size α is 0.05.

4.3.2 Results

Table 1 contains all local power results. We distinguish different cases which we characterize as follows:

Decaying Causality, Low Persistence in x_H: $d = 0.2$ The MF max and Wald tests have moderately high power between 0.346 and 0.570 (e.g. max and Wald tests with $h_{MF} = 4$ have respective power 0.487 and 0.570).

The LF tests with flow sampling have very little power above nominal size, regardless of the number of lags $h_{LF} \in \{1, 2, 3\}$, since alternating signs in ν_j and flow aggregation combine together to offset causality. The lowest value is 0.063 and the largest value is 0.076. Under stock sampling, however, local power is much larger, ranging from 0.495 to 0.643. For example, the LF Wald test with stock sampling and $h_{LF} = 1$ has power 0.643, while the MF Wald test power is .570. The reason for the improved performance is the largest coefficient $\nu_1 = 2.5$ is assigned to $x_H(\tau_L - 1, 12)$, which is precisely the regressor included in the low frequency models with stock sampling.

Decaying Causality, High Persistence in x_H: $d = 0.8$ Local power rises in general when there is high persistence in the high frequency variable, but otherwise the above results carry over qualitatively.

Lagged Causality Consider the high persistence case $d = 0.8$ (low persistence leads to similar results with lower power). Mixed frequency tests have power that is increasing in h_{MF}, for example max-test power is 0.075, 0.181, and 0.769 when h_{MF} is 4, 8, and 12. This reflects the causal pattern that only the coefficient ν_{12} on $x_H(\tau_L - 1, 1)$ is 2 and all other ν's are zeros. It is thus important to include sufficiently many lags when we apply MF tests.

The LF tests with flow sampling have reasonably high power regardless of h_{LF}, ultimately because the low frequency models work on an aggregated x_H and hence taking only a few lags tends to be enough. The power of the LF max test, for instance, is 0.455, 0.468, and 0.415 when h_{LF} is 1, 2, and 3, respectively. Another important reason for this good performance is that the causal effect is unambiguously positive; we have one large positive coefficient $\nu_{12} = 2$ and no negative coefficients, while flow aggregation preserves such causality.

The LF tests with stock sampling, by contrast, have nearly no power at $h_{LF} = 1$. This is expected since $x_H(\tau_L - 1, 12)$ has a zero coefficient by construction. They have high power, however, when $h_{LF} = 2$ (0.676 for max test and 0.664 for Wald test) because the extra regressor $x_H(\tau_L - 2, 12)$ has a strong correlation with the adjacent term $x_H(\tau_L - 1, 1)$, which has a nonzero coefficient $\nu_{12} = 2$. Such spillover monotonically adds to local power as the persistence of x_H (i.e. d) is larger.

Sporadic Causality The case of sporadic causality highlights the advantage of the mixed frequency approach. Consider low persistence in the high frequency variable: $d = 0.2$. The MF max test has power 0.391, 0.323, and 0.677 when h_{MF} is 4, 8, and 12, and MF Wald test power is 0.365, 0.291, and 0.761. Their power declines when switching from $h_{MF} = 4$ to $h_{MF} = 8$ since ν_5, ν_6, ν_7, and ν_8 are all zeros, and thus a penalty arises due to the extra number of parameters. The LF tests, whether flow sampling or stock sampling, have nearly no power (at most 0.072) due to their vulnerability to alternating signs and lagged causality as seen above.

Max versus Wald Test It is not clear from Table 1 whether the MF max test with a flat weight is preferred to the MF Wald test. Across our experiments, the max-test power surpasses the Wald-test power in 12 cases out of 18. The difference between max and Wald tests takes the largest value $0.769 - 0.560 = 0.209$ for lagged causality with $(d, h_{MF}) = (0.8, 12)$, and smallest value $0.690 - 0.872 = -0.182$ for sporadic causality with $(d, h_{MF}) = (0.8, 12)$. In the absence of other experiment designs or weighting schemes, we cannot in general conclude that the max text dominates the Wald test in terms of local power.[13]

This is not surprising since both test statistics are simple functions of least squares estimators, and local power is asymptotic. Indeed, although the max statistic operates on the largest estimator across parsimonious regression models, that need not imply higher local power since the statistic may asymptotically have greater dispersion. As we show below, however, the flat weighted max text statistic has the advantage of small sample size accuracy and in general higher power.

5 Monte Carlo Simulations

Our next task is to conduct simulation experiments in order to compare the MF and LF max and Wald tests.

5.1 High-to-Low Granger Causality

We first investigate tests of high-to-low non-causality $H_0 : \boldsymbol{b} = \mathbf{0}_{pm\times 1}$ against general causality $H_1 : \boldsymbol{b} \neq \mathbf{0}_{pm\times 1}$.

[13] In experiments not reported here, we tried a MF-VAR(2) data generating process as a robustness check, and obtained qualitative the same results as in MF-VAR(1).

5.1.1 MF-VAR(1)

We initially work with the MF-VAR(1) process (4.6) with $m = 12$, which serves as a benchmark. We consider a MF-VAR(2), below, as a robustness check.

Data Generating Processes The error term $\boldsymbol{\eta}(\tau_L)$ is mutually and serially independent standard normal random distributed.

There are four (non)causality cases, similar to those in the local power experiments: *non-causality*; *decaying causality* with alternating signs: $b_j = (-1)^{j-1} \times 0.3/j$ for $j = 1, \ldots, 12$; *lagged causality*: $b_j = 0.3 \times I(j = 12)$ for $j = 1, \ldots, 12$; and *sporadic causality*: $(b_3, b_7, b_{10}) = (0.2, 0.05, -0.3)$ and all other $b_j = 0$.

As in the local power study, we assume a weak autoregressive property for x_L (i.e. $a = 0.2$). See Table T.2 in the supplemental material for high-to-low causality tests when $a = 0.8$: the choice of a does not appear to significantly influence rejection frequencies. There are two values for the persistence of x_H: $d \in \{0.2, 0.8\}$, and decaying low-to-high causality with alternating signs: $c_j = (-1)^{j-1} \times 0.4/j$ 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 thought as month versus year, or approximately as week versus quarter. In the latter case, $T_L = 40$ or 80 imply that the low frequency sample size is 10 or 20 years, which are respectively small and medium spans of time. In the former case $T_L = 40$ implies that the low frequency sample size is a relatively large span of 40 years.

Model Estimation We estimate regression models that in all cases include two low frequency lags of x_L (i.e. $q = 2$). Setting $q = 1$ would be sufficient since the true DGP is MF-VAR(1), but the true lag order is typically unknown. The number of high frequency lags of x_H used in the MF tests is $h_{MF} \in \{4, 8, 12\}$ as well as $h_{MF} = 24$ when $T_L = 80$, and the number of low frequency lags of aggregated x_H used in LF tests is $h_{LF} \in \{1, 2, 3\}$ as well as $h_{LF} = 4$ when $T_L = 80$. Low frequency tests use flow and stock sampling for the high frequency variable. The max test weighting scheme is flat $\boldsymbol{W}_h = (1/h) \times \boldsymbol{I}_h$, and the number of draws from the limit distributions under H_0 is 1,000 for p-value computation.

Given the large ratio $m = 12$, the MF Wald test (and possibly even the LF Wald test) may suffer from size distortions if we use the asymptotic chi-square distribution. We therefore use Gonçalves and Killian's (2004) parametric bootstrap in order to better approximate the small sample Wald statistic distribution. As a bonus, their bootstrap allows for conditionally heteroscedastic errors of unknown form, while our innovations are i.i.d. We use their bootstrap p-value with 499 bootstrap samples in order to match the empirical study below, where the innovations are unlikely to be i.i.d. [14]

[14] Consider bootstrapping in the MF case with model (4.5), the LF case being similar. Let $\hat{\boldsymbol{\Theta}}$ be the unrestricted least squares estimator for $\boldsymbol{\Theta}$ in $x_L(\tau_L) = \boldsymbol{x}(\tau_L - 1)'\boldsymbol{\Theta} + u_L(\tau_L)$, the residual is $\tilde{u}_L(\tau_L)$, and the Wald statistic is $\hat{W}$. Compute the least squares $\tilde{\boldsymbol{\Theta}}_0$ for $\boldsymbol{\Theta}_0 = [\boldsymbol{\alpha}', \boldsymbol{0}_{1\times h}]'$ in the null model $x_L(\tau_L) = \boldsymbol{x}(\tau_L - 1)'\boldsymbol{\Theta}_0 + u_L(\tau_L)$. Then simulate N samples from $x_L(\tau_L) = \boldsymbol{x}(\tau_L - 1)'\tilde{\boldsymbol{\Theta}}_0 + \tilde{u}_L(\tau_L)v(\tau_L)$, where $v(\tau_L) \overset{i.i.d.}{\sim} N(0, 1)$, and compute the Wald

The number of Monte Carlo samples drawn is 5,000 for max tests and 1,000 for bootstrapped Wald tests (due to the added computation time), and size α is fixed at 0.05.

Results Table 2 compiles the simulation results: Panel A contains empirical size, while Panels B-D contain empirical power. Empirical size in both tests is fairly sharp, ranging across cases between 0.032 and 0.068. The max tests has sharp size evidently due to its relatively more parsimonious specification, while the Wald test has sharp size due to bootstrapping the p-value.

Panels B-D provide the same implications for MF versus LF tests as in the local power study, cfr. Table 1. In particular, MF tests are better capable of detecting complicated causal patterns like sporadic causality.

Consider the relative power performance of the MF max and Wald tests. In a strong majority of cases across causal patterns $\boldsymbol{b}$, lag length h_{MF}, persistence d, and sample size T_L, the max test has higher power than the Wald test. In a few exceptions the differences are negligible, where the greatest spread being $0.482 - 0.527 = -0.045$ when there is decaying causality, $h_{MF} = 4$ and $T_L = 80$ (Panel B.1.2). In the cases where the max test performs better, the difference in power is often substantial. Under lagged causality with $(d, T_L, h_{MF}) = (0.8, 40, 12)$, for example, max test power is 0.576 but Wald test has power is only 0.255 (Panel C.2.1). Similarly, max and Wald tests have powers 0.907 and 0.498 under lagged causality with $(d, T_L, h_{MF}) = (0.8, 80, 24)$ (Panel C.2.2). The apparent reason is the bootstrapped Wald test achieves a *size corrected* power that approximates the *size corrected* power of the asymptotic test (cfr. Davidson and MacKinnon (2006)). In simulations not reported here we find large to massive size distortions for the Wald test, hence the bootstrapped version has sharp size and comparatively low power. Therefore, in view of relatively sharp size for both tests, the max test dominates in general.

5.1.2 MF-VAR(2)

As a robustness check, we use observations drawn from a structural MF-VAR(2) $\boldsymbol{N}\boldsymbol{X}(\tau_L) = \sum_{i=1}^{2} \boldsymbol{M}_i \boldsymbol{X}(\tau_L - i) + \boldsymbol{\eta}(\tau_L)$ with $m = 12$. Relative to the MF-VAR(1) in (4.6), the extra coefficient matrix $\boldsymbol{M}_2$ is parameterized as

$$\boldsymbol{M}_2 = \begin{bmatrix} \mathbf{0}_{12\times 1} & \dots & \mathbf{0}_{12\times 1} & \mathbf{0}_{12\times 1} \\ b_{24} & \dots & b_{13} & 0 \end{bmatrix}.$$

The four causal patterns are *non-causality*: $\boldsymbol{b} = \mathbf{0}_{24\times 1}$; *decaying causality*: $b_j = (-1)^{j-1} \times 0.3/j$ for $j = 1, \dots, 24$; *lagged causality*: $b_j = 0.3 \times I(j = 24)$ for $j = 1, \dots, 24$; and *sporadic causality*: $(b_5, b_{12}, b_{17}, b_{19}) = (-0.2, 0.1, 0.2, -0.35)$ and all other $b_j = 0$.

Other quantities are similar to those used above: $a = 0.2$; $d \in \{0.2, 0.8\}$; $c_j = (-1)^{j-1} \times 0.4/j$ for $j = 1, \dots, 12$; $q = 2$; $\boldsymbol{W}_h = (1/h) \times \boldsymbol{I}_h$; $T_L \in \{40, 80\}$; and size is $\alpha = 0.05$. Rejection frequencies are similar when $a = 0.8$: see Table T.3 in the supplemental material Ghysels, Hill, and Motegi (2015). The number of high frequency lags of x_H used in the MF tests is

statistic $\tilde{W}_i$. The bootstrapped p-value is $p_N = [1 + \sum_{i=1}^{N} I(\tilde{W}_i \geq W)]/(N+1)$.

$h_{MF} \in \{16, 20, 24\}$, while the number of low frequency lags of aggregated x_H used in LF tests is $h_{LF} \in \{1, 2, 3\}$.

Rejection frequencies are compiled in Table 3. There are no serious size distortions as in the MF-VAR(1) case, and the relative performance of MF and LF tests is also the same. In a vast majority of cases the MF max test has higher empirical power than the MF Wald test, and the power advantage is often quite 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 (Panel C.2.2). There are two exceptions where the max test is less powerful than the Wald test, but the power differences are negligible: $0.284 - 0.290 = -0.006$ (Panel B.1.2: decaying causality) and $0.621 - 0.655 = -0.034$ (Panel D.2.2: sporadic causality). Therefore, again the max test dominates, and it has an even greater advantage precisely in a model with greater parameter proliferation. In general, therefore, the max test is best for MF high-to-low causality tests.

5.2 Low-to-High Granger Causality

Our focus now is low-to-high causality $\boldsymbol{c} = [c_1, \ldots, c_{12}]'$ in the structural MF-VAR(1) in (4.6) with $m = 12$.

5.2.1 Design

We use the usual four causality patterns. In each case we need to be careful about how $\boldsymbol{c}$ is transferred to the upper-right block $[\sum_{i=1}^{1} d^{1-i}c_i, \ldots, \sum_{i=1}^{12} d^{12-i}c_i]'$ of $\boldsymbol{A}_1$, the low-to-high causality pattern in the reduced form (4.7), where d is the AR(1) coefficient of x_H. For *non-causality* $\boldsymbol{c} = \boldsymbol{0}_{12\times 1}$, the upper-right block of $\boldsymbol{A}_1$ is a null vector regardless of d. A similar pattern arises for *decaying causality* $c_j = (-1)^{j-1} \times 0.45/j$ for $j = 1, \ldots, 12$, assuming $d = 0.2$. For *lagged causality* $c_j = 0.4 \times I(j = 12)$ for $j = 1, \ldots, 12$, the upper-right block of $\boldsymbol{A}_1$ is identically $\boldsymbol{c}$ regardless of d. In the case of *sporadic causality* $(c_3, c_7, c_{10}) = (0.4, 0.25, -0.5)$ a similar pattern arises, assuming $d = 0.2$. Consult Figure 1 for a graphical representation.

We impose weak autoregressive properties $a = d = 0.2$ for x_L and x_H, and decaying high-to-low causality with alternating signs: $b_j = (-1)^{j-1} \times 0.2/j$ for $j = 1, \ldots, 12$. Sample size is $T_L \in \{40, 80\}$.

As a benchmark, the Wald test is based on the MF naïve regression model (3.1) with a bootstrapped p-value and 499 bootstrapped samples, and the MF max test is based on MF parsimonious regression model (3.2).

In small samples $T_L = 40$ the max test exhibits relatively large size distortions due to the number of included high frequency lags. The asymptotic distribution is therefore a poor approximation of the small sample distribution. In larger samples $T_L \geq 120$, however, size distortions decrease precipitously: see the discussion below. In view of our relatively small sample sizes, as a second max test we use the MF parsimonious regression models with a MIDAS polynomial on the high frequency lags, as in (3.4), as an ad hoc attempt to tackle remaining

parameter proliferation. We use the Almon polynomial of dimension $s = 3$ (cfr. Section 3.1). In order to make a direct comparison with the Wald test, we also perform the Wald test on MF naïve regression model (3.1) with lags of x_H replaced with the Almon polynomial.

Besides the MF max test and MF Wald test, we perform the LF max test based on LF parsimonious regression models:

$$x_L(\tau_L) = \alpha_{1,j} x_L(\tau_L - 1) + \sum_{k=1}^{h_{LF}} \beta_{k,j} x_H(\tau_L - k) + \gamma_j x_H(\tau_L + j) + u_{L,j}(\tau_L), \quad j = 1, \ldots, r_{LF}. \tag{5.1}$$

We do not exploit a MIDAS polynomial since h_{LF} takes small values in our design. We consider both stock and flow sampling for aggregating x_H.

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

$$x_L(\tau_L) = \alpha_1 x_L(\tau_L - 1) + \sum_{k=1}^{h_{LF}} \beta_k x_H(\tau_L - k) + \sum_{j=1}^{r_{LF}} \gamma_j x_H(\tau_L + j) + u_L(\tau_L), \tag{5.2}$$

and 499 bootstrapped samples.

The estimated MF models use leads and lags of x_H taken from $h_{MF}, r_{MF} \in \{4, 8, 12\}$, and the estimated LF models use leads and lags of aggregated x_H taken from $h_{LF}, r_{LF} \in \{1, 2, 3\}$. We compute flat weighted max statistics, and use 1,000 draws from the asymptotic distribution under low-to-high non-causality for p-value computation. The number of Monte Carlo samples drawn is 5,000 for max tests and 1,000 for Wald tests, and size is fixed at 5%.

5.2.2 Results

Table 4 presents the results. First, the MF max test exhibits large (small) size distortions when T_L is 40 (80). At worst, empirical size is 0.225 at the 5% level (Panel A.2.1: $T_L = 80$ and $h_{MF} = r_{MF} = 24$), and at best size is 0.058 (Panel A.2.1: $T_L = 80$, $h_{MF} = 4$ and $r_{MF} = 8$), logically since fewer parameters align with sharper empirical size. If $T_L = 80$, then sizes are between 0.058 and 0.097 for $h_{MF} \in \{4, 8, 12\}$, generally revealing greater parameter proliferation is aligned with a greater size distortion. If we increase the sample size to $T_L = 120$, then empirical size is much sharper: see Table T.4 in the supplemental material Ghysels, Hill, and Motegi (2015).

Second, the max test with MIDAS polynomials exhibits relatively sharp empirical size, despite the inherent mis-specification of the estimated model. The Wald test works well with or without the MIDAS polynomials: without the mis-specified polynomials, empirical size is sharp due to the bootstrapped p-value.

In the case of decaying causality (Panel B), we see a clear advantage of MF tests compared to LF tests when $h_{MF} \in \{4, 8, 12\}$. For example, when $T_L = 80$ the MF max test without MIDAS has size adjusted empirical power of at least 0.728-0.097 = 0.631 (see Panel A.2.1 for size and B.2.1 for power), while the LF test has unadjusted empirical power at most at 0.137 across both stock and flow sampling (see Panel B.2.4: max test with flow sampling). In order

to understand why, consider stock sampling first. As seen in (5.1) and (5.2), lead terms used in those tests are $x_H(\tau_L+1,12), x_H(\tau_L+2,12), \ldots, x_H(\tau_L+r_{LF},12)$ and all these terms have small coefficients due to the decaying structure of true causality. The LF test statistics, in other words, are missing the most important lead term $x_H(\tau_L+1,1)$ and thus suffer from a poor signal relative to noise. In the case of flow sampling, averaging $x_H(\tau_L+1,1)$ through $x_H(\tau_L+1,12)$ makes positive impacts and negative impacts offset each other, hence again there is a very poor causation signal. This has been well documented in the literature: temporal aggregation can obfuscate or eliminate true underlying causality.

Next, consider lagged causality (Panel C). Consider $T_L = 80$ and $h_{MF} \leq 12$ so that size distortions are less an issue. The MF tests have no or little power net of size, when $r_{MF} = 4$ or $r_{MF} = 8$. This is understandable since the only relevant term is $x_H(\tau_L+1,12)$ by construction. When $r_{MF} = 12$ then power improves sharply, but parametric proliferation evidently augments noise in the least squares estimator and therefore diminishes power in either max or Wald test statistics. If $T_L = 80$, for example, then the MF max test has empirical power within $[0.563, 0.581]$ without MIDAS and $[0.582, 0.592]$ with MIDAS, while the MF Wald test has empirical power within $[0.415, 0.456]$ without MIDAS and $[0.497, 0.516]$ with MIDAS (see Panels C.2.1 and C.2.2). LF tests with stock sampling, by contrast, obtain much higher power, within $[0.683, 0.919]$ (see Panel C.2.3). This occurs precisely because the LF models contain the relevant lead term $x_H(\tau_L+1,12)$, and requires fewer estimated parameters.

Lastly, under sporadic causality (Panel D) MF tests exhibit very high power, especially when the number of lead terms is $r_{MF} = 12$ since this takes into account $c_{10} = -0.5$, the largest coefficient in absolute value. When $T_L = 80$, $h_{MF} \in \{4, 8, 12\}$ and $r_{MF} = 12$, the MF max test has size corrected power above 0.81 (0.87) without (with) a MIDAS polynomial (see Panels D.2.1 and D.2.2). LF tests, by contrast, have no or negligible power in all cases. The low frequency leads and lags of x_H are too coarse to capture the complicated causality pattern with unevenly-spaced lags, alternating signs, and non-decaying structure.

Finally, the max test has higher power than the Wald test in a strong majority of cases. Hence, based on our simulation design, the MF max and Wald tests are roughly equally powerful under decaying causality, but the max test is more powerful under lagged and sporadic causality. Therefore, the max test again dominates overall as a test of causality when there is parameter proliferation.[15]

6 Empirical Application

As an empirical illustration, we study Granger causality between a weekly term spread (short and long term interest rate spread) and quarterly real GDP growth in the U.S. We analyze both high-to-low causality (i.e. from spread to GDP) and low-to-high causality (i.e. from GDP to spread), although we are particularly interested in the former. A decline in the interest rate

[15] In Table T.5 of the supplemental material Ghysels, Hill, and Motegi (2015), we try MF-VAR(2) as a data generating process. The results are generally similar to the MF-VAR(1) case.

spread has historically been regarded as a strong predictor of a recession, but recent events place doubt on its use for such prediction. Recall that in 2005 the interest rate spread fell substantially due to a relatively constant long-term rate and an increasing short-term rate (also known as "Greenspan's Conundrum"), yet a recession did not follow immediately. The subprime mortgage crisis started nearly 2 years later, in December 2007, and therefore may not be directly related to the 2005 plummet in the interest rate spread.

We use seasonally-adjusted quarterly real GDP growth as a business cycle measure. In order to remove potential seasonal effects remaining after seasonal adjustment, we use annual growth (i.e. 4 quarter log-difference $\ln(y_t) - \ln(y_{t-4})$). The short and long term interests rates used for the term spread are respectively the federal funds (FF) rate and 10-year Treasury constant maturity rate. We aggregate each daily series into weekly series by picking the last observation in each week (recall that interest rates are stock variables). The sample period is January 5, 1962 to December 31, 2013, covering 2,736 weeks or 208 quarters.[16]

Figure 2 shows the weekly 10-year rate, weekly FF rate, their spread (10Y - FF), and quarterly GDP growth from January 5, 1962 through December 31, 2013. The shaded areas represent recession periods defined by the National Bureau of Economic Research (NBER). In the first half of the sample period, a sharp decline of the spread seems to be immediately followed by a recession. In the second half of the sample period there appears to be a weaker association, and a larger time lag between a spread drop and a recession.

Table 5 contains sample statistics. The 10-year rate is about 1% point higher than the FF rate on average, while average GDP growth is 3.15%. The spread has a relatively large kurtosis of 5.61, whereas GDP growth has a smaller kurtosis of 3.54.

The number of weeks contained in each quarter τ_L is not constant, which we denote as $m(\tau_L)$: 13 quarters have 12 weeks each, 150 quarters have 13 weeks each, and 45 quarters have 14 weeks each. While the max test can be applied with varying $m(\tau_L)$, we simplify the analysis by forcing a constant $m = 12$ by taking a sample average at the end of each τ_L, resulting in the following modified spread $\{x^*_H(\tau_L, j)\}_{j=1}^{12}$:

$$x^*_H(\tau_L, j) = \begin{cases} x_H(\tau_L, j) & \text{for} \quad j = 1, \ldots, 11, \\ \frac{1}{m(\tau_L)-11} \sum_{k=12}^{m(\tau_L)} x_H(\tau_L, k) & \text{for} \quad j = 12. \end{cases}$$

This modification gives us a dataset with $T_L = 208$, $m = 12$, and thus $T = mT_L = 2,496$ high frequency observations.

In view of our 52-year sample period, we implement a rolling window analysis with a window width of 80 quarters (i.e. 20 years). The first subsample covers the first quarter of 1962 through the fourth quarter of 1981 (written as 1962:I-1981:IV), the second one is 1962:II-1982:I, and the last one is 1994:I-2013:IV, equaling 129 subsamples. The trade-off between small and large window widths is that the latter is more likely to contain a structural break, but a 20 year window

[16] All data are downloaded from the Saint Louis Federal Reserve Bank data archive.

allows us to include more leads and lags in our models. Furthermore, our simulation experiments in Section 5.1.2 reveal our tests work well for similar parsimonious and naïve regression models with $T_L = 80$.

6.1 Granger Causality from Interest Rate Spread to GDP Growth

We first consider causality from the high frequency interest rate spread (x_H^*) to low frequency GDP growth (x_L). We use an MF-VAR(2) specification since the resulting residuals from the naïve model (6.2), below, appear to be serially uncorrelated (all models also include a constant term). The MF max test operates on parsimonious regression models:

$$x_L(\tau_L) = \alpha_{0,j} + \sum_{k=1}^{2} \alpha_{k,j} x_L(\tau_L - k) + \beta_j x_H^*(\tau_L - 1, 12 + 1 - j) + u_{L,j}(\tau_L). \quad j = 1, \ldots, 24, \tag{6.1}$$

which includes two quarters of lagged GDP growth (x_L) hence $q = 2$, and 24 weeks or about 2 quarter) of lagged interest rate spread (x_H^*) hence $h_{MF} = 24$.

The MF Wald test operates on:

$$x_L(\tau_L) = \alpha_0 + \sum_{k=1}^{2} \alpha_k x_L(\tau_L - k) + \sum_{j=1}^{24} \beta_j x_H^*(\tau_L - 1, 12 + 1 - j) + u_L(\tau_L). \tag{6.2}$$

The LF max test is based on parsimonious models:

$$x_L(\tau_L) = \alpha_{0,j} + \sum_{k=1}^{2} \alpha_{k,j} x_L(\tau_L - k) + \beta_j x_H^*(\tau_L - j) + u_{L,j}(\tau_L), \quad j = 1, 2, 3, \tag{6.3}$$

which has two quarters of lagged x_L (i.e. $q = 2$) and three quarters of lagged x_H^* (i.e. $h_{LF} = 3$). Since the interest rate spread is a stock variable, we let the aggregated high frequency variable be $x_H^*(\tau_L) = x_H^*(\tau_L, 12)$. Finally, the LF Wald test is performed on:

$$x_L(\tau_L) = \alpha_0 + \sum_{k=1}^{2} \alpha_k x_L(\tau_L - k) + \sum_{j=1}^{3} \beta_j x_H^*(\tau_L - j) + u_L(\tau_L). \tag{6.4}$$

Wald statistic p-values are computed based on Gonçalves and Killian's (2004) bootstrap, with $N = 999$ replications. Max statistic p-values are computed based on 100,000 draws from the limit distributions under non-causality.

We perform the Ljung-Box Q tests of serial uncorrelatedness of the least squares residuals from the MF model (6.2) and LF model (6.4) in order to check whether these models are well specified. Since the true innovations are not likely to be independent, we use Horowitz, Lobato, Nankervis and Savin's (2006) double blocks-of-blocks bootstrap with block size $b \in \{4, 10, 20\}$. The number of bootstrap samples is $M_1 = 999$ for the first stage and $M_2 = 249$ for the second stage. In each of the 129 windows and for each model we implement the Q tests with 4, 8, or 12 lags.

When $b = 4$, the null hypothesis of residual uncorrelatedness in the MF case is rejected at the 5% level in {13, 5, 1} windows out of 129 for tests with {4, 8, 12} lags, suggesting the MF model is well specified. In the LF case, the null hypothesis is rejected at the 5% level in {51, 23, 33} windows with {4, 8, 12} lags, hence the LF model may not be well specified. For larger block size $b \in \{10, 20\}$, the MF model again produces fewer rejections than the LF model. (See Table T.6 of Ghysels, Hill, and Motegi (2015) for complete results.) Overall, the MF model seems to yield a better fit than the LF model in terms of residual uncorrelatedness.

Figure 3 plots p-values for tests of non-causality over the 129 subsamples. Unless otherwise stated, the significance level is 5%. All tests except for the MF Wald test find significant causality in early periods. The MF max test detects significant causality prior to 1981:IV-2001:III, the LF max test detects significant causality prior to 1980:III-2000:II, and the LF Wald test detects significant causality prior to 1974:III-1994:II. The MF max test has the longest period of significant causality, arguably due to its high power, as shown in Section 5.1. 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 at the 5% level only *after* subsample 1990:III-2010:II, which is somewhat counter-intuitive. This result may stem from parameter proliferation. As seen from (6.1)-(6.4), the MF naïve regression model has many more parameters than any other model. The MF Wald test does, however, give weak evidence of causality through the 1973:II-1993:I, where we reject non-causality at levels below 10% for a few early windows. In view of the intuitive test results, the MF max test seems to be preferred to the MF Wald test when the ratio of sampling frequencies m is large.

We also implement the four tests for the full sample covering 52 years from January 1962 through December 2013. We try models with more lags than in the rolling window analysis, taking advantage of the greater sample size: $(q, h_{MF}, h_{LF}) = (4, 48, 6)$. This specification means that (i) each model has 4 quarters of low frequency lags of x_L, (ii) each mixed frequency model has 48 weeks of high frequency lags of x^*_H, and (iii) each low frequency model has 6 quarters of low frequency lags of x^*_H. The number of bootstrap replications for the Wald tests is 9,999.

We first implement the bootstrapped Ljung-Box Q test with 4, 8, or 12 lags on the least squares residuals from MF and LF models. When block size is $b = 4$, p-values from the MF model are {.107, .180, .084} for lags {4, 8, 12}. The null hypothesis of residual uncorrelatedness is not rejected at the 5% level for any lag (although it is rejected at the 10% level for lag 12). The MF model is therefore well specified in general. P-values from the LF model are {.021, .066, .024} for lags {4, 8, 12}, suggesting that the LF model is not well specified. Similar results appear when we change the block size to 10 or 20. As in the rolling window analysis, the MF model yields a better fit than the LF model in terms of residual uncorrelatedness.

The p-value for the MF max test is .000, hence we strongly reject causality. Conversely, we fail to reject non-causality at any conventional level by the MF Wald test (p-value .465), possibly due to lower power relative to the max test in view of parameter proliferation. The LF

p-values are .001 for the max test and .086 for the Wald test, demonstrating stronger evidence of causality by the max test, and the strongest by the MF max test. Overall, there is strong evidence for causality from interest rate spread to GDP based on the max test, and only weak or partial evidence based on Wald tests.

6.2 Granger Causality from GDP Growth to Interest Rate Spread

We now consider causality from GDP growth to the interest rate spread, hence low-to-high causality. The MF max test is either based on the unrestricted parsimonious regression models:

$$x_L(\tau_L) = \alpha_{0,j} + \sum_{k=1}^{2} \alpha_{k,j} x_L(\tau_L - k) + \sum_{k=1}^{24} \beta_k x_H^*(\tau_L - 1, 12 + 1 - k) + \gamma_j x_H^*(\tau_L + 1, j) + u_{L,j}(\tau_L),\ j = 1, \ldots, 24,$$

or the restricted models with Almon polynomial $\omega_k(\boldsymbol{\pi})$ of order 3:

$$x_L(\tau_L) = \alpha_{0,j} + \sum_{k=1}^{2} \alpha_{k,j} x_L(\tau_L - k) + \sum_{k=1}^{24} \omega_k(\boldsymbol{\pi}) x_H^*(\tau_L - 1, 12 + 1 - k) + \gamma_j x_H^*(\tau_L + 1, j) + u_{L,j}(\tau_L),\ j = 1, \ldots, 24.$$

We include two quarters of lagged x_L (i.e. $q = 2$), 24 weeks of lagged x_H^* (i.e. $h_{MF} = 24$), and 24 weeks of led x_H^* (i.e. $r_{MF} = 24$).

The Wald test is based on either an unrestricted naïve regression model:

$$x_L(\tau_L) = \alpha_0 + \sum_{k=1}^{2} \alpha_k x_L(\tau_L - k) + \sum_{k=1}^{24} \beta_k x_H^*(\tau_L - 1, 12 + 1 - k) + \sum_{j=1}^{24} \gamma_j x_H^*(\tau_L + 1, j) + u_L(\tau_L),$$

or a restricted model with Almon polynomial $\omega_k(\boldsymbol{\pi})$:

$$x_L(\tau_L) = \alpha_0 + \sum_{k=1}^{2} \alpha_k x_L(\tau_L - k) + \sum_{k=1}^{24} \omega_k(\boldsymbol{\pi}) x_H^*(\tau_L - 1, 12 + 1 - k) + \sum_{j=1}^{24} \gamma_j x_H^*(\tau_L + 1, j) + u_L(\tau_L).$$

The LF max test is based on the unrestricted parsimonious regression models $x_L(\tau_L) = \alpha_{0,j} + \sum_{k=1}^{2} \alpha_{k,j} x_L(\tau_L - k) + \sum_{k=1}^{3} \beta_{k,j} x_H^*(\tau_L - k) + \gamma_j x_H^*(\tau_L + j) + u_{L,j}(\tau_L),\ \ j = 1, 2, 3$. Since the rate spread is a stock variable, we let $x_H^*(\tau_L) = x_H^*(\tau_L, 12)$. We include two quarters of lagged x_L (i.e. $q = 2$), three quarters of lagged x_H^* (i.e. $h_{LF} = 3$), and three quarters of lead x_H^* (i.e. $r_{LF} = 3$). Finally, the LF Wald test uses the naïve regression model: $x_L(\tau_L) = \alpha_0 + \sum_{k=1}^{2} \alpha_k x_L(\tau_L - k) + \sum_{k=1}^{3} \beta_k x_H^*(\tau_L - k) + \sum_{j=1}^{3} \gamma_j x_H^*(\tau_L + j) + u_L(\tau_L)$.

As usual, Wald test p-values are bootstrapped with $N = 999$ bootstrap samples, and the max test p-values are computed using 100,000 draws from limit distributions under non-causality. Bootstrapped Ljung-Box Q tests with lags 4, 8, or 12 suggest that the MF models produce uncorrelated residuals in more windows than the LF model.[17]

Figure 4 plots p-values for the causality tests over the 129 subsamples. While MF tests

[17] When the block size is $b = 4$, the MF model without a MIDAS polynomial rejects the null hypothesis of uncorrelated residuals at the 5% level in {4, 8, 5} windows out of 129 for lags {4, 8, 12}. When a MIDAS polynomial is used, the null hypothesis is rejected in {25, 14, 16} windows. In the LF model the null hypothesis is rejected in {31, 17, 26} windows. If we raise the block size to 10 or 20, rejections occur in only a few windows across all models. See Table T.6 of Ghysels, Hill, and Motegi (2015) for complete results.

without MIDAS polynomial find significant causality in some subsamples (cfr. Panels (a) and (b)), MF tests with MIDAS polynomial find non-causality in all subsamples (cfr. Panels (c) and (d)). The LF max test shows significant causality in only a few subsamples around 1984:I-2003:IV (cfr. Panel (e)). The LF Wald test shows significant causality starting in subsample 1989:IV-2009:III, amounting to roughly the last 15% of the subsamples (cfr. Panel (f)).

Finally, we conduct the four tests on the full sample based on one specification $(q, h_{MF}, r_{MF}, h_{LF}, r_{LF}) = (2, 24, 24, 3, 3)$, hence: (i) each model has 2 quarters of low frequency lags of x_L, (ii) each MF model has 24 weeks of high frequency leads and lags of x_H^* each, and (iii) each LF model has 3 quarters of low frequency leads and lags of x_H^* each. Considering that we already have 52 total leads and lags, we do not treat another specification with more lags. The number of bootstrap replications for the Wald test is 9,999. Bootstrapped Ljung-Box Q tests again suggest that residuals from the MF models have a weaker degree of autocorrelation than residuals from the LF model.[18]

The MF max and Wald tests without a MIDAS polynomial have p-values .036 and .261, respectively, and with a MIDAS polynomial the p-values are .090 and .689. The LF max and Wald test have p-values .098 and .222. Therefore, the max test leads to the strongest evidence for causality, and overall the clearest evidence comes from the max test in a model without a MIDAS polynomial.

7 Conclusions

This paper proposes a new mixed frequency Granger causality test that achieves high power even when the ratio of sampling frequencies m is large. This is accomplished by exploiting 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 $j \in \{1, \ldots, h\}$. Our resulting max test statistic then operates on the largest j^{th} lag or lead estimated parameter. This method of inference extends to any regression setting where many parameters have the value zero under the null hypothesis, and should therefore be of general interest when many possibly irrelevant regressors are available.

Although the max test statistic follows a non-standard asymptotic distribution under the null hypothesis of non-causality, a p-value can be easily computed by drawing a large number observations from the null limit distribution. We prove the MF max test obtains an asymptotic power of one for a test of non-causality from *high to low* frequency, but consistency in the case of *low to high* frequency remains as an open question.

Our max test has wider applicability. One can easy generalize the test for an increasing number of parameters, and would therefore apply to, for example, nonparametric regression

[18] When the block size is $b = 4$, the p-values for the MF model without a MIDAS polynomial are {.076, .280, .054} for lags {4, 8, 12}. When a MIDAS polynomial is used the p-values are {.035, .179, .019}. In the LF model the p-values are {.043, .041, .001}. If we raise the block size to 10 or 20, we observe larger p-values (i.e. weaker rejections of the null hypothesis of residual uncorrelatedness) in general.

models using Fourier flexible forms (Gallant and Souza (1991)), Chebyshev, Laguerre or Hermite polynomials (see e.g. Draper, Smith, and Pownell (1966)), and splines (Rice and Rosenblatt (1983), Friedman (1991)) - where our test has use for determining whether a finite or infinite number of terms are redundant. Similarly, a max test of white noise is another application since bootstrapped Q-tests have comparatively lower power. These are only a few examples involving a large - possibly infinite - set of parametric zero restrictions. We leave this as an area of future research.

In terms of the specific application in this paper - we can summarize the findings as follows. Through local asymptotic power analysis and Monte Carlo simulations, we compare the max and Wald tests based on mixed or low frequency data. We show that MF tests are better able to detect complex causal patterns than LF tests in both local asymptotics and finite samples. The MF max and Wald tests have roughly the same local asymptotic power, but the max test is generally more powerful, and in many cases substantially more powerful, than the Wald test in finite samples since (i) only bootstrapping the Wald test can correct for size distortions due to parameter proliferation, and (ii) size-corrected bootstrapped power for the Wald test is only as good as size-corrected asymptotic test power, which is generally low due to the asymptotic test size distortions.

We study causality patterns between a weekly interest rate spread and real GDP growth in the U.S., over rolling sample windows. The MF max test yields an intuitive result that the interest rate spread causes GDP growth until about the year 2000, after which causality vanishes, while Wald and LF tests yield mixed results.

References

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

Andrews, D. W. K., and W. Ploberger (1994): ""Optimal Tests when a Nuisance Parameter is Present Only under the Alternative," *Econometrica*, 62, 1383–1414.

Billingsley, P. (1999): *Convergence of Probability Measures*. Wiley, New York.

Bradley, R. (1993): "Equivalent Mixing Conditions for Random Fields," *Annals of Probability*, 21, 1921–1926.

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

Davidson, R., and J. G. MacKinnon (2006): "The Power of Bootstrap and Asymptotic Tests," *Journal of Econometrics*, 133, 421–441.

Doukhan, P. (1994): *Mixing: Properties and Examples*. Springer-Verlag.

Draper, N. R., H. Smith, and E. Pownell (1966): *Applied Regression Analysis*. Wiley New York.

Dufour, J., D. Pelletier, and E. Renault (2006): "Short Run and Long Run Causality in Time Series: Inference," *Journal of Econometrics*, 132, 337–362.

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

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

Friedman, J. H. (1991): "Multivariate Adaptive Regression Splines," *Annals of Statistics*, 19, 1–67.

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

Gallant, A. R., and G. Souza (1991): "On the Asymptotic Normality of Fourier Flexible Form Estimates," *Journal of Econometrics*, 50, 329–353.

Ghysels, E. (2014): "Macroeconomics and the Reality of Mixed Frequency Data," *Journal of Econometrics*, (forthcoming).

Ghysels, E., J. B. Hill, and K. Motegi (2013): "Testing for Granger Causality with Mixed Frequency Data," Working paper, Dept. of Economics, University of North Carolina.

——— (2015): "Supplemental Material for *Simple Granger Causality Tests for Mixed Frequency Data*," Dept. of Economics, University of North Carolina.

Ghysels, E., P. Santa-Clara, and R. Valkanov (2004): "The MIDAS Touch: Mixed Data Sampling Regression Models," Working Paper, Dept. of Economics, UNC and UCLA.

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

Ghysels, E., A. Sinko, and R. Valkanov (2007): "MIDAS Regressions: Further Results and New Directions," *Econometric Reviews*, 26, 53–90.

Gonçalves, S., and L. Killian (2004): "Bootstrapping Autoregressions with Conditional Heteroskedasticity of Unknown Form," *Journal of Econometrics*, 123, 89–120.

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

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

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

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

Hansen, B. (1996): "Inference when a Nuisance Parameter is not Identified under the Null Hypothesis," *Econometrica*, 64, 413–430.

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

Horowitz, J., I. N. Lobato, J. C. Nankervis, and N. E. Savin (2006): "Bootstrapping the Box-Pierce Q Test: A Robust Test of Uncorrelatedness," *Journal of Econometrics*, 133, 841–862.

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

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

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

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

Renault, E., K. Sekkat, and A. Szafarz (1998): "Testing for Spurious Causality in Exchange Rates," *Journal of Empirical Finance*, 5, 47–66.

Rice, J., and M. Rosenblatt (1983): "Smoothing Splines: Regression, Derivatives and Deconvolution," *Annals of Statistics*, 11, 141–156.

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

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

Technical Appendices

A Double Time Indices

Consider low and high frequency variables x_L and x_H. $x_L(\tau_L)$ obviously only requires a single time index $\tau_L \in \mathbb{Z}$. $x_H(\tau_L, j)$, however, has two time indices: the low frequency increments $\tau_L \in \mathbb{Z}$ and the higher frequency increment $j \in \{1, \ldots, m\}$, e.g. $m = 3$ for low frequency quarter τ_L and high frequency month $j \in \{1, \ldots, m\}$.

It is often useful to use a notation convention that allows the second argument of x_H to be any integer. It is, for example, 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, the following notation can be used as a *high frequency simplification*:

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

$\lceil x \rceil$ is the smallest integer not smaller than x, while $\lfloor x \rfloor$ is the largest integer not larger than x. 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. Indeed, $m\left\lceil \frac{1-j}{m} \right\rceil + j \in \{1, \ldots, m\}$ when $j \leq 0$, and $j - m\left\lfloor \frac{j-1}{m} \right\rfloor \in \{1, \ldots, m\}$ when $j \geq m + 1$.

Since the high frequency simplification allows both arguments of x_H to be any integer, we can verify the following *low frequency simplification*:

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

Therefore, any lag or lead i in the first argument of x_H can be deleted by modifying the second argument appropriately. The second argument may therefore become an integer that is non-positive or larger than m, but such a case is covered by the high frequency simplification.

B Proof of Theorem 2.1

Parsimonious regression model j is written as $x_L(\tau_L) = \boldsymbol{x}_j(\tau_L - 1)'\boldsymbol{\theta}_j + u_{L,j}(\tau_L)$, where $\boldsymbol{x}_j(\tau_L - 1) \equiv [x_L(\tau_L - 1), \ldots, x_L(\tau_L - q), x_H(\tau_L - 1, m + 1 - j)]'$ and $\boldsymbol{\theta}_j \equiv [\alpha_{1,j}, \ldots, \alpha_{q,j}, \beta_j]'$. Write $\boldsymbol{\theta} \equiv [\boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_h']'$.

Recall $\boldsymbol{W}_{T_L,h}$ is an $h \times h$ diagonal matrix whose diagonal elements are the stochastic max test weights$w_{T_L,1}$, …, $w_{T_L,h}$. Similarly, let $\boldsymbol{W}_h$ be an $h \times h$ diagonal matrix whose diagonal elements are the max test weight non-random probability limits w_1, …, w_h.

In order to characterize the distribution limit of $\hat{\mathcal{T}} \equiv \max_{1 \leq j \leq h}(\sqrt{T_L} w_{T_L,j} \hat{\beta}_j)^2$, we must show convergence of the finite dimensional distributions of $[w_{T_L,j}\hat{\beta}_j]_{j=1}^h$ and tightness (e.g. Theorem 7.1 in Billingsley (1999)). Since j is discrete valued, tightness is trivial (e.g. Theorem 1.3 in Billingsley (1999)). Hence, we only need the to show the finite dimensional distributions for any h are $\sqrt{T_L}\boldsymbol{W}_{T_L,h}\hat{\boldsymbol{\beta}} \xrightarrow{d} N(\boldsymbol{0}_{h\times 1}, \boldsymbol{V})$ where $\boldsymbol{V} \equiv \sigma_L^2 \boldsymbol{W}_h \boldsymbol{R}\boldsymbol{\mathcal{S}}\boldsymbol{R}'\boldsymbol{W}_h$. The claim $\max_{1 \leq j \leq h}(\sqrt{T_L} w_{T_L,j}\hat{\beta}_j)^2 \xrightarrow{d} \max_{1 \leq j \leq h} \mathcal{N}_j^2$, where $\boldsymbol{\mathcal{N}} = [\mathcal{N}_1, \ldots, \mathcal{N}_h]'$ is a vector-valued random variable drawn from $N(\boldsymbol{0}_{h\times 1}, \boldsymbol{V})$, then follows instantly from the continuous mapping theorem.

B.1 Convergence in Finite Dimensional Distributions

It is easier to work with the identity $\boldsymbol{\beta} = \boldsymbol{R}\boldsymbol{\theta}$, where the selection matrix $\boldsymbol{R} \in \mathbb{R}^{h \times (q+1)h}$ satisfies $\boldsymbol{R}_{j,(q+1)j} = 1$ for $j = 1, \ldots, h$ and all other elements are zero. Let $\hat{\boldsymbol{\theta}}_j$ be the least squares estimators and define $\hat{\boldsymbol{\theta}} \equiv [\hat{\boldsymbol{\theta}}_1', \ldots, \hat{\boldsymbol{\theta}}_h']'$. We begin by deriving the finite dimensional distributions of $\{\hat{\boldsymbol{\theta}}_j\}_{j=1}^h$ under $H_0 : \boldsymbol{b} = \boldsymbol{0}_{pm\times 1}$. Under H_0 the parsimonious model parameters $\boldsymbol{\beta} \equiv [\beta_1, \ldots, \beta_h]' = 0$, so define $\boldsymbol{\theta}_{0,j} \equiv [\alpha_{1,j}, \ldots, \alpha_{q,j}, 0]'$ and $\boldsymbol{\theta}_0 \equiv [\boldsymbol{\theta}_{0,1}', \ldots, \boldsymbol{\theta}_{0,h}']'$.

Under the null we may write $x_L(\tau_L) = \boldsymbol{x}_j(\tau_L - 1)'\boldsymbol{\theta}_{0,j} + \epsilon_L(\tau_L)$. Therefore, by the construction of $\hat{\boldsymbol{\theta}}_j$

for the parsimonious model (2.6), and law of large numbers (B.4):

$$\begin{aligned}\sqrt{T_L}\left(\hat{\boldsymbol{\theta}}_j-\boldsymbol{\theta}_{0,j}\right) &= \sqrt{T_L}\left[\sum_{\tau_L=1}^{T_L}\boldsymbol{x}_j(\tau_L-1)\boldsymbol{x}_j(\tau_L-1)'\right]^{-1}\sum_{\tau_L=1}^{T_L}\boldsymbol{x}_j(\tau_L-1)\epsilon_L(\tau_L) \\ &= \left(E\left[\boldsymbol{x}_j(\tau_L-1)\boldsymbol{x}_j(\tau_L-1)'\right]\right)^{-1}\frac{1}{\sqrt{T_L}}\sum_{\tau_L=1}^{T_L}\boldsymbol{x}_j(\tau_L-1)\epsilon_L(\tau_L)+o_p(1) \\ &= \boldsymbol{\Gamma}_{j,j}^{-1}\frac{1}{\sqrt{T_L}}\sum_{\tau_L=1}^{T_L}\boldsymbol{x}_j(\tau_L-1)\epsilon_L(\tau_L)+o_p(1).\end{aligned} \tag{B.1}$$

Therefore, for any $\boldsymbol{\lambda}=[\boldsymbol{\lambda}_1',\ldots,\boldsymbol{\lambda}_h']'$, $\boldsymbol{\lambda}'\boldsymbol{\lambda}=1$:

$$\boldsymbol{\lambda}'\times\sqrt{T_L}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)=\frac{1}{\sqrt{T_L}}\sum_{\tau_L=1}^{T_L}\sum_{j=1}^{h}\boldsymbol{\lambda}_j'\boldsymbol{\Gamma}_{j,j}^{-1}\boldsymbol{x}_j(\tau_L-1)\epsilon_L(\tau_L)+o_p(1)=\frac{1}{\sqrt{T_L}}\sum_{\tau_L=1}^{T_L}X(\tau_L-1,\boldsymbol{\lambda})\epsilon_L(\tau_L)+o_p(1), \tag{B.2}$$

say. Now define $\boldsymbol{\Gamma}_{j,i}=E[\boldsymbol{x}_j(\tau_L-1)\boldsymbol{x}_i(\tau_L-1)']$, and note that:

$$E\left[X(\tau_L-1,\boldsymbol{\lambda})^2\right]=\sum_{j=1}^{h}\sum_{i=1}^{h}\boldsymbol{\lambda}_j'\boldsymbol{\Gamma}_{j,j}^{-1}\boldsymbol{\Gamma}_{j,i}\boldsymbol{\Gamma}_{i,i}^{-1}\boldsymbol{\lambda}_i=\sum_{j=1}^{h}\sum_{i=1}^{h}\boldsymbol{\lambda}_j'\boldsymbol{\Sigma}_{j,i}\boldsymbol{\lambda}_i=\boldsymbol{\lambda}'\boldsymbol{\mathcal{S}}\boldsymbol{\lambda},$$

where $\boldsymbol{\mathcal{S}}\in\mathbb{R}^{(q+1)h\times(q+1)h}$ has (i,j) block $\boldsymbol{\mathcal{S}}_{i,j}=\boldsymbol{\Sigma}_{i,j}$, and is $\boldsymbol{\Sigma}_{i,j}$ defined by (2.9). Under Assumptions 2.1-2.3 it is easily verified that $\boldsymbol{\mathcal{S}}$ is positive definite. Now apply central limit theorem (B.5) to (B.2) in order to obtain that $\boldsymbol{\lambda}'\times\sqrt{T_L}\,(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\xrightarrow{d}N(0,\boldsymbol{\lambda}'(\sigma_L^2\boldsymbol{\mathcal{S}})\,\boldsymbol{\lambda})$. Hence, by the Cramér-Wold theorem $\sqrt{T_L}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\xrightarrow{d}N(\mathbf{0}_{(q+1)h\times1},\sigma_L^2\boldsymbol{\mathcal{S}})$. Now use $w_{T_L,j}\xrightarrow{p}w_j$ and the constructions $\hat{\boldsymbol{\beta}}=\boldsymbol{R}\hat{\boldsymbol{\theta}}$ and $0=\boldsymbol{\beta}=\boldsymbol{R}\boldsymbol{\theta}_0$ under H_0 to deduce

$$\sqrt{T_L}\boldsymbol{W}_{T_L,h}\hat{\boldsymbol{\beta}}=\sqrt{T_L}\boldsymbol{W}_{T_L,h}\boldsymbol{R}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)=\sqrt{T_L}\boldsymbol{W}_h\boldsymbol{R}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)+o_p(1)\xrightarrow{d}N(\mathbf{0}_{h\times1},\boldsymbol{V}) \tag{B.3}$$

where $\boldsymbol{V}\equiv\sigma_L^2\boldsymbol{W}_h\boldsymbol{R}\boldsymbol{\mathcal{S}}\boldsymbol{R}'\boldsymbol{W}_h$. Finally, $\boldsymbol{V}$ is positive definite by the positive definiteness of $\boldsymbol{\mathcal{S}}$, the assumption $\boldsymbol{W}_h\neq0$, and by the construction of $\boldsymbol{R}$.

B.2 LLN and CLT

The Assumption 2.3 α-mixing property implies mixing in the ergodic sense, hence ergodicity. Therefore, by stationarity, square integrability, and the ergodic theorem:

$$\frac{1}{T_L}\sum_{\tau_L=1}^{T_L}\boldsymbol{x}_j(\tau_L-1)\boldsymbol{x}_j(\tau_L-1)'\xrightarrow{p}E\left[\boldsymbol{x}_j(\tau_L-1)\boldsymbol{x}_j(\tau_L-1)'\right]. \tag{B.4}$$

Next, we want a central limit theorem for $1/\sqrt{T_L}\sum_{\tau_L=1}^{T_L}\sum_{j=1}^{h}\boldsymbol{r}_j'\boldsymbol{x}_j(\tau_L-1)\epsilon_L(\tau_L)$ for $\{\boldsymbol{r}_j\}_{j=1}^h$, $\boldsymbol{r}_j'\boldsymbol{r}_j=1$ under H_0. Under $H_0:\boldsymbol{b}=\mathbf{0}_{pm\times1}$ it follows $x_L(\tau_L)=\boldsymbol{X}_L(\tau_L-1)'\boldsymbol{a}+\epsilon_L(\tau_L)$ with mds $\epsilon_L(\tau_L)$, and by Assumption 2.3 and measurability, $\sum_{j=1}^{h}\boldsymbol{r}_j'\boldsymbol{x}_j(\tau_L-1)\epsilon_L(\tau_L)$ is α-mixing with coefficients $\sum_{h=0}^{\infty}\alpha_{2^h}<\infty$. Further, by the mds property $\sum_{j=1}^{h}\boldsymbol{r}_j'\boldsymbol{x}_j(\tau_L-1)\epsilon_L(\tau_L)$ is square integrable white noise, measurable with respect to strictly increasing σ-fields $\mathcal{F}_{\tau_L}$, and therefore has variance

$$E\left[\left(\sum_{j=1}^{h}\boldsymbol{r}_j'\boldsymbol{x}_j(\tau_L-1)\right)^2\epsilon_L(\tau_L)\right]=\sum_{i,j=1}^{h}\boldsymbol{r}_j'E\left[\boldsymbol{x}_j(\tau_L-1)\boldsymbol{x}_i(\tau_L-1)'\right]\boldsymbol{r}_i\times\sigma_L^2=\sum_{i,j=1}^{h}\boldsymbol{r}_j'\Gamma_{j,i}\boldsymbol{r}_i\times\sigma_L^2>0.$$

Hence $\sum_{j=1}^{h}\boldsymbol{r}_j'\boldsymbol{x}_j(\tau_L-1)\epsilon_L(\tau_L)$ has a spectral density $f(\lambda)$ that satisfies $f(0)>0$. In view of Theorem 1.a of Bradley (1993) it therefore follows that Theorem 2.2 of Ibragimov (1975) applies, hence

$$\frac{1}{\sqrt{T_L}}\sum_{\tau_L=1}^{T_L}\sum_{j=1}^{h}\boldsymbol{r}_j'\boldsymbol{x}_j(\tau_L-1)\epsilon_L(\tau_L)\xrightarrow{d}N\left(0,\sum_{i,j=1}^{h}\boldsymbol{r}_j'\Gamma_{j,i}\boldsymbol{r}_i\times E\left[\epsilon_L^2(\tau_L)\right]\right). \tag{B.5}$$

C Proof of Theorem 2.2

The max test statistic $\hat{\mathcal{T}} \equiv \max_{1\leq t\leq h}(\sqrt{T_L}w_{T_L,j}\hat{\beta}_j)^2$ operates on a discrete valued stochastic function $g_{T_L}(j) \equiv w_{T_L,j}\hat{\beta}_j$. Therefore, weak convergence for $\{g_{T_L}(j) \ : j \in \{1,...,h\}\}$ is identical to convergence in the finite dimensional distributions of $\{g_{T_L}(j) \ : j \in \{1,...,h\}\}$, cfr. Theorem 1.3 in Billingsley (1999). Hansen's (1996) proof of his Theorem 2 therefore carries over to prove the present claim.

D Proof of Theorem 2.3

Recall the parsimonious regression model j is $x_L(\tau_L) = \boldsymbol{x}_j(\tau_L-1)'\boldsymbol{\theta}_j + u_{L,j}(\tau_L)$. In view of stationarity, square integrability, and ergodicity (see Appendix B), the least squares estimator satisfies $\hat{\boldsymbol{\theta}}_j \xrightarrow{p} \boldsymbol{\theta}_j^*$, where:

$$\boldsymbol{\theta}_j^* = \left[E\left[\boldsymbol{x}_j(\tau_L-1)\boldsymbol{x}_j(\tau_L-1)'\right]\right]^{-1} E\left[\boldsymbol{x}_j(\tau_L-1)x_L(\tau_L)\right].$$

Now, recall the DGP $x_L(\tau_L) = \boldsymbol{X}_L(\tau_L-1)'\boldsymbol{a} + \boldsymbol{X}_H(\tau_L-1)'\boldsymbol{b} + \epsilon_L(\tau_L)$. Therefore:

$$\begin{aligned}\boldsymbol{\theta}_j^* &= \left[E\left[\boldsymbol{x}_j(\tau_L-1)\boldsymbol{x}_j(\tau_L-1)'\right]\right]^{-1} E\left[\boldsymbol{x}_j(\tau_L-1)\left\{\boldsymbol{X}_L(\tau_L-1)'\boldsymbol{a} + \boldsymbol{X}_H(\tau_L-1)'\boldsymbol{b} + \epsilon_L(\tau_L)\right\}\right] \\ &= \left[E\left[\boldsymbol{x}_j(\tau_L-1)\boldsymbol{x}_j(\tau_L-1)'\right]\right]^{-1}\left\{E\left[\boldsymbol{x}_j(\tau_L-1)\boldsymbol{X}_L(\tau_L-1)'\right]\boldsymbol{a} + E\left[\boldsymbol{x}_j(\tau_L-1)\boldsymbol{X}_H(\tau_L-1)'\right]\boldsymbol{b}\right\},\end{aligned} \tag{D.1}$$

where the second equality holds from the mds assumption of ϵ_L. By Assumption 2.4, the number of autoregressive lags q in the parsimonious models (2.6) is at least as large as the true lag order p in the true data generating process (2.3). Therefore, the low frequency regressor set from (2.3) satisfies:

$$\boldsymbol{X}_L(\tau_L-1) = [\boldsymbol{I}_p,\ \boldsymbol{0}_{p\times(q-p+1)}]\boldsymbol{x}_j(\tau_L-1) \tag{D.2}$$

hence

$$E\left[\boldsymbol{x}_j(\tau_L-1)\boldsymbol{X}_L(\tau_L-1)'\right] = E\left[\boldsymbol{x}_j(\tau_L-1)\boldsymbol{x}_j(\tau_L-1)'\right]\begin{bmatrix}\boldsymbol{I}_p \\ \boldsymbol{0}_{(q-p+1)\times p}\end{bmatrix}. \tag{D.3}$$

Substituting (D.3) into (D.1), we obtain the desired result (2.10).

E Proof of Theorem 2.4 and Example 2.2

Proof of Theorem 2.4 Pick the last row of (2.10). The lower left block of $\{E\left[\boldsymbol{x}_j(\tau_L-1)\boldsymbol{x}_j(\tau_L-1)'\right]\}^{-1}$ is

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

while the lower right block is simply n_j^{-1}, where

$$\begin{aligned}n_j \equiv\ & E\left[x_H(\tau_L-1,m+1-j)^2\right] \\ & - E\left[x_H(\tau_L-1,m+1-j)\boldsymbol{X}_L^{(q)}(\tau_L-1)'\right]\left\{E\left[\boldsymbol{X}_L^{(q)}(\tau_L-1)\boldsymbol{X}_L^{(q)}(\tau_L-1)'\right]\right\}^{-1} \\ & \times E\left[\boldsymbol{X}_L^{(q)}(\tau_L-1)x_H(\tau_L-1,m+1-j)\right].\end{aligned}$$

Hence, the last row of $[E[\boldsymbol{x}_j(\tau_L-1)\boldsymbol{x}_j(\tau_L-1)']]^{-1}\times E[\boldsymbol{x}_j(\tau_L-1)\boldsymbol{X}_H(\tau_L-1)']$ appearing in (2.10) is $n_j^{-1}\boldsymbol{d}_j'$, where

$$\begin{aligned}\boldsymbol{d}_j \equiv\ & E\left[\boldsymbol{X}_H(\tau_L-1)x_H(\tau_L-1,m+1-j)\right] \\ & - E\left[\boldsymbol{X}_H(\tau_L-1)\boldsymbol{X}_L^{(q)}(\tau_L-1)'\right]\left\{E\left[\boldsymbol{X}_L^{(q)}(\tau_L-1)\boldsymbol{X}_L^{(q)}(\tau_L-1)'\right]\right\}^{-1} \\ & \times E\left[\boldsymbol{X}_L^{(q)}(\tau_L-1)x_H(\tau_L-1,m+1-j)\right].\end{aligned} \tag{E.1}$$

If $\boldsymbol{\beta}^* = \mathbf{0}_{h\times 1}$ then $n_j^{-1}\boldsymbol{d}_j'\boldsymbol{b} = 0$ in view of (2.10). Since n_j is a nonzero finite scalar for any $j = 1, \ldots, h$ by the nonsingularity of $E\left[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{x}_j(\tau_L - 1)'\right]$, it follows $\boldsymbol{d}_j'\boldsymbol{b} = 0$. Stacking these h equations, we have that $\boldsymbol{D}\boldsymbol{b} = \mathbf{0}_{h\times 1}$ and thus $\boldsymbol{b}'\boldsymbol{D}'\boldsymbol{D}\boldsymbol{b} = 0$, where $\boldsymbol{D} \equiv [\boldsymbol{d}_1, \ldots, \boldsymbol{d}_h]'$.

The claim $\boldsymbol{b} = \mathbf{0}_{pm\times 1}$ follows provided we show that $\boldsymbol{D}'\boldsymbol{D}$ is positive definite. It is sufficient to show that $\boldsymbol{D}$ is of full column rank pm. Since we are assuming that $h \geq pm$, we only have to show that $\boldsymbol{D}_{pm} \equiv [\boldsymbol{d}_1, \ldots, \boldsymbol{d}_{pm}]'$, the first pm rows of $\boldsymbol{D}$, is of full column rank pm or equivalently non-singular. Equation (E.1) implies that

$$\begin{aligned}\boldsymbol{D}_{pm} =& E\left[\boldsymbol{X}_H(\tau_L - 1)\boldsymbol{X}_H(\tau_L - 1)'\right] \\ &- E\left[\boldsymbol{X}_H(\tau_L - 1)\boldsymbol{X}_L^{(q)}(\tau_L - 1)'\right]\left[E\left[\boldsymbol{X}_L^{(q)}(\tau_L - 1)\boldsymbol{X}_L^{(q)}(\tau_L - 1)'\right]\right]^{-1} E\left[\boldsymbol{X}_L^{(q)}(\tau_L - 1)\boldsymbol{X}_H(\tau_L - 1)'\right].\end{aligned}$$

Now define

$$\boldsymbol{\Delta} \equiv E\left[\begin{bmatrix}\boldsymbol{X}_L^{(q)}(\tau_L - 1) \\ \boldsymbol{X}_H(\tau_L - 1)\end{bmatrix}\begin{bmatrix}\boldsymbol{X}_L^{(q)}(\tau_L - 1)' & \boldsymbol{X}_H(\tau_L - 1)'\end{bmatrix}\right].$$

$\boldsymbol{\Delta}$ is trivially non-singular by Assumption 2.1. But $\boldsymbol{D}_{pm}$ is the Schur complement of $\boldsymbol{\Delta}$ with respect to $E[\boldsymbol{X}_L^{(q)}(\tau_L - 1)\boldsymbol{X}_L^{(q)}(\tau_L - 1)']$. Therefore, by the classic argument of partitioned matrix inversion, $\boldsymbol{D}_{pm}$ is non-singular as desired. $\mathcal{QED}$

Proof of Example 2.2 We require inverses of $\boldsymbol{\Gamma}_{1,1} \equiv E[\boldsymbol{x}_1(\tau_L - 1)\boldsymbol{x}_1(\tau_L - 1)']$, where $\boldsymbol{x}_1(\tau_L - 1) = [x_L(\tau_L - 1), x_H(\tau_L - 1, 2)]'$ as defined in (2.6) and (2.10, and $\boldsymbol{\Gamma}_{2,2} = E[\boldsymbol{x}_2(\tau_L - 1)\boldsymbol{x}_2(\tau_L - 1)']$, where $\boldsymbol{x}_2(\tau_L - 1) = [x_L(\tau_L - 1), x_H(\tau_L - 1, 1)]'$. By construction $\boldsymbol{\Gamma}_{2,2} = \boldsymbol{\Gamma}_{1,1}$. Under (2.11) we have $E[x_H(\tau_L - 1, 1)^2] = E[x_H(\tau_L - 1, 2)^2] = 1$, $E[x_L(\tau_L - 1)^2] = 1/\rho^2$, and $E[\ x_L(\tau_L - 1)\ x_H(\tau_L - 1, 2)\ ] = E\ [\ (-\frac{1}{\rho}\ x_H(\tau_L - 2, 1) + x_H(\tau_L - 2, 2) + \epsilon_L(\tau_L - 1))\ x_H(\tau_L - 1, 2)] = 0$, hence

$$\boldsymbol{\Gamma}_{1,1} = \begin{bmatrix} 1/\rho^2 & 0 \\ 0 & 1 \end{bmatrix} \text{ and } \boldsymbol{\Gamma}_{1,1}^{-1} = \begin{bmatrix} \rho^2 & 0 \\ 0 & 1 \end{bmatrix}$$

Next, recall $\boldsymbol{X}_H(\tau_L - 1) = [x_H(\tau_L - 1, 2), x_H(\tau_L - 1, 1)]'$ in (2.3). Then

$$\begin{aligned}\boldsymbol{C}_1 &\equiv E\left[\boldsymbol{x}_1(\tau_L - 1)\boldsymbol{X}_H(\tau_L - 1)'\right] \\ &= \begin{bmatrix} E[x_L(\tau_L - 1)x_H(\tau_L - 1, 2)] & E[x_L(\tau_L - 1)x_H(\tau_L - 1, 1)] \\ E[x_H(\tau_L - 1, 2)x_H(\tau_L - 1, 2)] & E[x_H(\tau_L - 1, 2)x_H(\tau_L - 1, 1)] \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & \rho \end{bmatrix}\end{aligned}$$

and

$$\begin{aligned}\boldsymbol{C}_2 &\equiv E\left[\boldsymbol{x}_2(\tau_L - 1)\boldsymbol{X}_H(\tau_L - 1)'\right] \\ &= \begin{bmatrix} E[x_L(\tau_L - 1)x_H(\tau_L - 1, 2)] & E[x_L(\tau_L - 1)x_H(\tau_L - 1, 1)] \\ E[x_H(\tau_L - 1, 1)x_H(\tau_L - 1, 2)] & E[x_H(\tau_L - 1, 1)x_H(\tau_L - 1, 1)] \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ \rho & 1 \end{bmatrix}.\end{aligned}$$

Now use (2.10) to deduce:

$$\begin{bmatrix}\alpha_{1,1}^* \\ \beta_1^*\end{bmatrix} = \begin{bmatrix}\rho^2 & 0 \\ 0 & 1\end{bmatrix}\begin{bmatrix}0 & 0 \\ 1 & \rho\end{bmatrix}\begin{bmatrix}1 \\ -1/\rho\end{bmatrix} = \begin{bmatrix}0 \\ 0\end{bmatrix} \quad \text{and} \quad \begin{bmatrix}\alpha_{1,2}^* \\ \beta_2^*\end{bmatrix} = \begin{bmatrix}\rho^2 & 0 \\ 0 & 1\end{bmatrix}\begin{bmatrix}0 & 0 \\ \rho & 1\end{bmatrix}\begin{bmatrix}1 \\ -1/\rho\end{bmatrix} = \begin{bmatrix}0 \\ \rho - 1/\rho\end{bmatrix}.$$

Hence $\beta_1^* = 0$ and $\beta_2^* = \rho - 1/\rho \neq 0$ since $|\rho| < 1$. Therefore, if $h = 1$ then the MF max test statistic $\hat{\mathcal{T}}$ converges to the Theorem 2.1 asymptotic null distribution, resulting in no power (above the nominal size). However, $h = 2$ and assign positive weight $w_2 > 0$ to $\hat{\beta}_2$, then $\hat{\mathcal{T}} \xrightarrow{p} \infty$. $\mathcal{QED}$

F Proof of Theorem 4.1

This argument mirrors the proof of Theorem 2.1. The DGP under $H_1^L : \boldsymbol{b} = (1/\sqrt{T_L})\boldsymbol{\nu}$ is described in (4.1). Hence (B.1) becomes

$$\sqrt{T_L}(\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_0) = \boldsymbol{\Gamma}_{j,j}^{-1}\boldsymbol{C}_j\boldsymbol{\nu} + \boldsymbol{\Gamma}_{j,j}^{-1}(1/\sqrt{T_L}) \sum_{\tau_L=1}^{T_L} \boldsymbol{x}_j(\tau_L - 1)\epsilon_L(\tau_L) + o_p(1),$$

where $\boldsymbol{\Gamma}_{j,j} = E[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{x}_j(\tau_L - 1)']$ and $\boldsymbol{C}_j = E[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{X}_H(\tau_L - 1)']$. Repeating (B.2), we get $\boldsymbol{\lambda}' \times \sqrt{T_L}(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_0) \xrightarrow{d} N(\boldsymbol{\lambda}'\boldsymbol{u},\, \boldsymbol{\lambda}'(\sigma_L^2\boldsymbol{\Sigma})\boldsymbol{\lambda})$, for any $\boldsymbol{\lambda}'\boldsymbol{\lambda} = 1$, where $\boldsymbol{u} \equiv [(\boldsymbol{\Gamma}_{1,1}^{-1}\boldsymbol{C}_1)', \cdots, (\boldsymbol{\Gamma}_{h,h}^{-1}\boldsymbol{C}_h)']'\boldsymbol{\nu}$, hence by the Cramér-Wold theorem $\sqrt{T_L}(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_0) \xrightarrow{d} N(\boldsymbol{u},\, \sigma_L^2\boldsymbol{\Sigma})$. Now repeat (B.3) to get $\sqrt{T_L}\boldsymbol{W}_{T_L,h}\hat{\boldsymbol{\beta}} \xrightarrow{d} N(\boldsymbol{\mu}, \boldsymbol{V})$, where $\boldsymbol{\mu} = \boldsymbol{W}_h\boldsymbol{R}\boldsymbol{u}$ and $\boldsymbol{V} = \sigma_L^2\boldsymbol{W}_h\boldsymbol{R}\boldsymbol{\Sigma}\boldsymbol{R}'\boldsymbol{W}_h$. The remainder of the proof proceeds along the lines of the proof of Theorem 2.1.

G Proof of Lemma 4.2

We exploit covariance stationarity throughout without explicitly stating so. Recall:

$$\begin{aligned}
\boldsymbol{\Upsilon}_k &\equiv E[\boldsymbol{X}(\tau_L)\boldsymbol{X}(\tau_L - k)'] \\
&= \begin{bmatrix}
E[x_H(\tau_L, 1)x_H(\tau_L - k, 1)] & \dots & E[x_H(\tau_L, 1)x_H(\tau_L - k, m)] & E[x_H(\tau_L, 1)x_L(\tau_L - k)] \\
\vdots & \ddots & \vdots & \vdots \\
E[x_H(\tau_L, m)x_H(\tau_L - k, 1)] & \dots & E[x_H(\tau_L, m)x_H(\tau_L - k, m)] & E[x_H(\tau_L, m)x_L(\tau_L - k)] \\
E[x_L(\tau_L)x_H(\tau_L - k, 1)] & \dots & E[x_L(\tau_L)x_H(\tau_L - k, m)] & E[x_L(\tau_L)x_L(\tau_L - k)]
\end{bmatrix}
\end{aligned} \tag{G.1}$$

for $k \geq 0$, and $\boldsymbol{\Upsilon}_k = \boldsymbol{\Upsilon}'_{-k}$ for $k < 0$.

We have for $\boldsymbol{\Gamma}_{j,i}$:

$$\begin{aligned}
\boldsymbol{\Gamma}_{j,i} &\equiv E[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{x}_i(\tau_L - 1)'] = E[\boldsymbol{x}_j(\tau_L)\boldsymbol{x}_i(\tau_L)'] \\
&= E\begin{bmatrix} x_L(\tau_L) \\ \vdots \\ x_L(\tau_L - (q-1)) \\ x_H(\tau_L, m + 1 - j) \end{bmatrix} \begin{bmatrix} x_L(\tau_L) & \dots & x_L(\tau_L - (q-1)) & x_H(\tau_L, m + 1 - i) \end{bmatrix}.
\end{aligned} \tag{G.2}$$

Since j and i may be larger than m, the second argument of x_H may be smaller than 1, hence it is not immediately clear which element of $\boldsymbol{\Gamma}_{j,i}$ is identical to which element of $\boldsymbol{\Upsilon}_k$. In order to ensure that the second argument of x_H lies in $\{1, \dots, m\}$, we use the high frequency simplification (A.1):

$$\begin{aligned}
x_H(\tau_L, m + 1 - j) &= x_H\left(\tau_L - \left\lceil \frac{1 - (m + 1 - j)}{m} \right\rceil, m \left\lceil \frac{1 - (m + 1 - j)}{m} \right\rceil + (m + 1 - j)\right) \\
&= x_H\left(\tau_L - \left\lceil \frac{j - m}{m} \right\rceil, m \left\lceil \frac{j - m}{m} \right\rceil + m + 1 - j\right) = x_H(\tau_L - f(j),\, g(j)),
\end{aligned} \tag{G.3}$$

where the last equality follows from the definitions $f(j) \equiv \lceil (j - m)/m \rceil$ and $g(j) \equiv mf(j) + m + 1 - j$. Note that $f(j) \geq 0$ and $g(j) \in \{1, \dots, m\}$ for any j as desired. Substituting (G.3) into (G.2), $\boldsymbol{\Gamma}_{j,i}$ can be rewritten as follows.

$$\begin{bmatrix}
E[x_L(\tau_L)x_L(\tau_L)] & \dots & E[x_L(\tau_L)x_L(\tau_L - (q-1))] & E[x_H(\tau_L - f(i), g(i))x_L(\tau_L)] \\
\vdots & \ddots & \vdots & \vdots \\
E[x_L(\tau_L - (q-1))x_L(\tau_L)] & \dots & E[x_L(\tau_L - (q-1))x_L(\tau_L - (q-1))] & E[x_H(\tau_L - f(i), g(i))x_L(\tau_L - (q-1))] \\
E[x_H(\tau_L - f(j), g(j))x_L(\tau_L)] & \dots & E[x_H(\tau_L - f(j), g(j))x_L(\tau_L - (q-1))] & E[x_H(\tau_L - f(j), g(j))x_H(\tau_L - f(i), g(i))]
\end{bmatrix}. \tag{G.4}$$

Now consider how $\boldsymbol{\Gamma}_{j,i}$ relates to $\boldsymbol{\Upsilon}_k$, and take as an example $E[x_H(\tau_L - f(j), g(j))\ x_L(\tau_L - (q-1))\ ]$,

the $(q+1,q)$ element of $\mathbf{\Gamma}_{j,i}$. There are two cases: $f(j) \geq q-1$ or $f(j) < q-1$. For the first case:

$$\begin{aligned} E[x_H(\tau_L - f(j), g(j))x_L(\tau_L - (q-1))] &= E[x_H(\tau_L - (q-1) - (f(j) - (q-1)), g(j))x_L(\tau_L - (q-1))] \\ &= E[x_H(\tau_L - (f(j) - (q-1)), g(j))x_L(\tau_L)] = \Upsilon_{f(j)-(q-1)}(K, g(j)), \end{aligned}$$

where $\Upsilon_{f(j)-(q-1)}(K, g(j))$ denotes the $(K, g(j))$ element of $\mathbf{\Upsilon}_{f(j)-(q-1)}$, and the third equality follows from (G.1). For the second case:

$$\begin{aligned} E[x_H(\tau_L - f(j), g(j))x_L(\tau_L - (q-1))] &= E[x_H(\tau_L - f(j), g(j))x_L(\tau_L - f(j) - (q-1-f(j)))] \\ &= E[x_H(\tau_L, g(j))x_L(\tau_L - (q-1-f(j)))] \\ &= \Upsilon_{(q-1)-f(j)}(g(j), K) = \Upsilon_{f(j)-(q-1)}(K, g(j)). \end{aligned}$$

where the third equality follows from (G.1). Combine the two cases to deduce $E[x_H(\tau_L - f(j), g(j))x_L(\tau_L - (q-1))] = \Upsilon_{f(j)-(q-1)}(K, g(j))$ for any j. Now apply the same argument to each element of $\mathbf{\Gamma}_{j,i}$ appearing in (G.4) to deduce the claimed representations of $\mathbf{\Gamma}_{j,i}$ in (4.3).

Next, consider $\boldsymbol{C}_j$ for $j \in \{1, \ldots, h\}$. We have:

$$\begin{aligned} \boldsymbol{C}_j &\equiv E[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{X}_H(\tau_L - 1)'] = E[\boldsymbol{x}_j(\tau_L)\boldsymbol{X}_H(\tau_L)'] \\ &= E\begin{bmatrix} x_L(\tau_L) \\ \vdots \\ x_L(\tau_L - (q-1)) \\ x_H(\tau_L, m+1-j) \end{bmatrix} \begin{bmatrix} x_H(\tau_L, m+1-1) & \ldots & x_H(\tau_L, m+1-pm) \end{bmatrix}. \end{aligned}$$

Since the second argument of x_H may be non-positive, we apply the high frequency simplification (G.3) to get:

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

We now map each element of $\boldsymbol{C}_j$ to an appropriate element of $\mathbf{\Upsilon}_k$. Consider $E[x_L(\tau_L - (q-1))x_H(\tau_L - f(pm), g(pm))]$, the (q, pm) element of $\boldsymbol{C}_j$, as an example. In view of (G.4), this quantity is equal to the $(q+1, q)$ element of $\mathbf{\Gamma}_{pm,i}$ with an arbitrary i. We already know from (4.3) that the $(q+1, q)$ element of $\mathbf{\Gamma}_{pm,i}$ is equal to $\Upsilon_{f(pm)-(q-1)}(K, g(pm))$. Applying the same argument to each element of $\boldsymbol{C}_j$ to complete the proof.

Tables and Figures

Table 1: Local Asymptotic Power of High-to-Low Causality Tests

A. Mixed Frequency												
	Decaying Causality				Lagged Causality				Sporadic Causality			
	$d = 0.2$		$d = 0.8$		$d = 0.2$		$d = 0.8$		$d = 0.2$		$d = 0.8$	
h_{MF}	Max	Wald	Max	Wald	Max	Wald	Max	Wald	Max	Wald	Max	Wald
4	.487	.570	.773	.736	.051	.050	.075	.066	.391	.365	.716	.614
8	.396	.472	.699	.621	.052	.050	.181	.125	.323	.291	.667	.679
12	.346	.407	.657	.542	.247	.206	.769	.560	.677	.761	.690	.872

B. Low Frequency with Flow Sampling												
	Decaying Causality				Lagged Causality				Sporadic Causality			
	$d = 0.2$		$d = 0.8$		$d = 0.2$		$d = 0.8$		$d = 0.2$		$d = 0.8$	
h_{LF}	Max	Wald	Max	Wald	Max	Wald	Max	Wald	Max	Wald	Max	Wald
1	.074	.076	.301	.302	.092	.094	.455	.454	.072	.070	.171	.169
2	.068	.067	.237	.244	.080	.080	.468	.459	.061	.063	.132	.132
3	.063	.063	.206	.208	.073	.074	.415	.402	.060	.060	.118	.115

C. Low Frequency with Stock Sampling												
	Decaying Causality				Lagged Causality				Sporadic Causality			
	$d = 0.2$		$d = 0.8$		$d = 0.2$		$d = 0.8$		$d = 0.2$		$d = 0.8$	
h_{LF}	Max	Wald	Max	Wald	Max	Wald	Max	Wald	Max	Wald	Max	Wald
1	.638	.643	.863	.862	.050	.050	.061	.059	.051	.051	.437	.435
2	.554	.538	.801	.786	.062	.063	.676	.664	.052	.051	.355	.356
3	.495	.473	.754	.728	.059	.060	.626	.598	.051	.051	.306	.305

The DGP is MF-VAR(1) with a ratio of sampling frequencies $m = 12$. Panels A, B, and C concern tests in mixed frequency, low frequency with flow sampling, and low frequency with stock sampling. In each case, from left to right, there is drift $\boldsymbol{\nu}$ representing decaying causality $\nu_j = (-1)^{j-1} \times 2.5/j$ for $j = 1, \ldots, 12$, lagged causality $\nu_j = 2 \times I(j = 12)$ for $j = 1, \ldots, 12$, and sporadic causality $(\nu_3, \nu_9, \nu_{11}) = (2.1, -2.8, 1.9)$ with all other $\nu_j = 0$. The high frequency variable x_H has low or high persistence: $d \in \{0.2, 0.8\}$, and the low frequency variable x_L has weak persistence: $a = 0.2$. Low-to-high causality is decaying with alternating signs: $c_j = (-1)^{j-1} \times 0.8/j$ for $j = 1, \ldots, 12$. In the models used as the premise for our tests, the number of high frequency lags of x_H for the mixed frequency tests are $h_{MF} \in \{4, 8, 12\}$, and the number of low frequency lags of aggregated x_H for the low frequency tests are $h_{LF} \in \{1, 2, 3\}$. The max text uses $100,000$ draws from the limit distributions under H_0 and H_1^L, and the weights are $\boldsymbol{W}_h = (1/h) \times \boldsymbol{I}_h$. Nominal size is $\alpha = .05$.

Table 2: Rejection Frequencies of High-to-Low Causality Tests Based on MF-VAR(1)

A. Non-Causality: $\boldsymbol{b} = \mathbf{0}_{12\times 1}$

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

A.1.1 $T_L = 40$

	MF			LF (flow)		LF (stock)	
h_{MF}	Max	Wald	h_{LF}	Max	Wald	Max	Wald
4	.061	.035	1	.064	.044	.063	.045
8	.056	.049	2	.061	.037	.055	.042
12	.063	.038	3	.058	.049	.061	.046

A.1.2 $T_L = 80$

	MF			LF (flow)		LF (stock)	
h_{MF}	Max	Wald	h_{LF}	Max	Wald	Max	Wald
4	.055	.054	1	.058	.048	.056	.045
8	.057	.037	2	.055	.040	.058	.039
12	.052	.048	3	.058	.065	.058	.042
24	.056	.038	4	.050	.054	.059	.039

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

A.2.1 $T_L = 40$

	MF			LF (flow)		LF (stock)	
h_{MF}	Max	Wald	h_{LF}	Max	Wald	Max	Wald
4	.058	.050	1	.065	.041	.063	.048
8	.066	.041	2	.063	.050	.057	.053
12	.058	.032	3	.068	.043	.060	.038

A.2.2 $T_L = 80$

	MF			LF (flow)		LF (stock)	
h_{MF}	Max	Wald	h_{LF}	Max	Wald	Max	Wald
4	.060	.035	1	.058	.038	.057	.045
8	.058	.037	2	.056	.054	.053	.055
12	.055	.043	3	.054	.039	.058	.038
24	.052	.058	4	.059	.041	.059	.048

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

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

B.1.1 $T_L = 40$

	MF			LF (flow)		LF (stock)	
h_{MF}	Max	Wald	h_{LF}	Max	Wald	Max	Wald
4	.228	.241	1	.072	.047	.383	.321
8	.163	.157	2	.068	.050	.276	.239
12	.128	.136	3	.061	.040	.228	.178

B.1.2 $T_L = 80$

	MF			LF (flow)		LF (stock)	
h_{MF}	Max	Wald	h_{LF}	Max	Wald	Max	Wald
4	.482	.527	1	.088	.062	.657	.597
8	.374	.412	2	.071	.058	.554	.469
12	.332	.335	3	.069	.052	.503	.429
24	.257	.229	4	.069	.053	.451	.342

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

B.2.1 $T_L = 40$

	MF			LF (flow)		LF (stock)	
h_{MF}	Max	Wald	h_{LF}	Max	Wald	Max	Wald
4	.444	.305	1	.185	.138	.585	.485
8	.343	.222	2	.145	.096	.451	.374
12	.272	.163	3	.122	.087	.376	.313

B.2.2 $T_L = 80$

	MF			LF (flow)		LF (stock)	
h_{MF}	Max	Wald	h_{LF}	Max	Wald	Max	Wald
4	.794	.700	1	.309	.265	.883	.834
8	.697	.564	2	.235	.205	.798	.764
12	.642	.480	3	.198	.159	.748	.697
24	.534	.272	4	.189	.147	.719	.619

The DGP is MF-VAR(1) with a ratio of sampling frequencies $m = 12$, where b_j signifies the impact of $x_H(\tau_L - 1, m + 1 - j)$ on $x_L(\tau_L)$. Panels A and B concern high-to-low non-causality and decaying causality. We consider low and higher persistence in x_H : $d \in \{0.2, 0.8\}$. Other parameters are specified as follows. There is weak autoregressive persistence in x_L ($a = 0.2$) and low-to-high decaying causality with alternating signs: $c_j = (-1)^{j-1} \times 0.4/j$ for $j = 1, \ldots, 12$. The models estimated have two low frequency lags of x_L (i.e. $q = 2$); the number of high frequency lags of x_H used in the mixed frequency tests is $h_{MF} \in \{4, 8, 12\}$ and $h_{MF} = 24$ when $T_L = 80$; and the number of low frequency lags of x_H used in the low frequency tests is $h_{LF} \in \{1, 2, 3\}$ and $h_{LF} = 4$ when $T_L = 80$. Low frequency tests use either flow or stock sampling. The max test statistic is computed using a flat weight $\boldsymbol{W}_h = (1/h) \times \boldsymbol{I}_h$, and the p-value is computed using 5,000 draws from the null limit distribution. The Wald test p-value is computed using the parametric bootstrap based on Gonçalves and Killian (2004), with 499 bootstrap replications. Sample sizes are small and medium $T_L \in \{40, 80\}$. Nominal size is $\alpha = 0.05$. The number of Monte Carlo samples drawn is 5,000 for max tests and 1,000 for Wald tests.

Table 2: Rejection Frequencies of High-to-Low Causality Tests Based on MF-VAR(1) : Continued

C. Lagged Causality: $b_j = 0.3 \times I(j = 12)$

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

C.1.1 $T_L = 40$

	MF			LF (flow)		LF (stock)	
h_{MF}	Max	Wald	h_{LF}	Max	Wald	Max	Wald
4	.064	.047	1	.101	.082	.059	.047
8	.062	.043	2	.081	.071	.068	.052
12	.146	.115	3	.073	.059	.071	.037

C.1.2 $T_L = 80$

	MF			LF (flow)		LF (stock)	
h_{MF}	Max	Wald	h_{LF}	Max	Wald	Max	Wald
4	.053	.040	1	.125	.102	.061	.056
8	.052	.043	2	.104	.088	.082	.056
12	.405	.264	3	.088	.074	.068	.049
24	.325	.155	4	.083	.059	.061	.060

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

C.2.1 $T_L = 40$

	MF			LF (flow)		LF (stock)	
h_{MF}	Max	Wald	h_{LF}	Max	Wald	Max	Wald
4	.081	.041	1	.337	.258	.063	.039
8	.132	.067	2	.320	.282	.523	.418
12	.576	.255	3	.274	.200	.446	.332

C.2.2 $T_L = 80$

	MF			LF (flow)		LF (stock)	
h_{MF}	Max	Wald	h_{LF}	Max	Wald	Max	Wald
4	.084	.060	1	.592	.518	.073	.052
8	.227	.141	2	.606	.562	.845	.779
12	.931	.731	3	.539	.527	.804	.749
24	.907	.498	4	.493	.438	.769	.662

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

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

D.1.1 $T_L = 40$

	MF			LF (flow)		LF (stock)	
h_{MF}	Max	Wald	h_{LF}	Max	Wald	Max	Wald
4	.119	.075	1	.060	.043	.061	.055
8	.101	.079	2	.061	.043	.056	.042
12	.173	.136	3	.067	.048	.056	.051

D.1.2 $T_L = 80$

	MF			LF (flow)		LF (stock)	
h_{MF}	Max	Wald	h_{LF}	Max	Wald	Max	Wald
4	.248	.207	1	.058	.057	.062	.048
8	.184	.167	2	.052	.049	.051	.049
12	.442	.416	3	.050	.055	.054	.051
24	.352	.250	4	.059	.039	.058	.061

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

D.2.1 $T_L = 40$

	MF			LF (flow)		LF (stock)	
h_{MF}	Max	Wald	h_{LF}	Max	Wald	Max	Wald
4	.243	.163	1	.075	.047	.161	.135
8	.197	.180	2	.093	.057	.221	.164
12	.402	.260	3	.082	.063	.175	.173

D.2.2 $T_L = 80$

	MF			LF (flow)		LF (stock)	
h_{MF}	Max	Wald	h_{LF}	Max	Wald	Max	Wald
4	.459	.305	1	.076	.052	.273	.228
8	.404	.404	2	.128	.101	.391	.402
12	.803	.740	3	.107	.067	.332	.325
24	.677	.524	4	.101	.071	.292	.335

The DGP is MF-VAR(1) with a ratio of sampling frequencies $m = 12$, where b_j signifies the impact of $x_H(\tau_L - 1, m + 1 - j)$ on $x_L(\tau_L)$. Panels C and D concern high-to-low lagged and sporadic causality. We consider low and higher persistence in x_H : $d \in \{0.2, 0.8\}$. Other parameters are specified as follows. There is weak autoregressive persistence in x_L ($a = 0.2$) and low-to-high decaying causality with alternating signs: $c_j = (-1)^{j-1} \times 0.4/j$ for $j = 1, \ldots, 12$. The models estimated have two low frequency lags of x_L (i.e. $q = 2$); the number of high frequency lags of x_H used in the mixed frequency tests is $h_{MF} \in \{4, 8, 12\}$ and $h_{MF} = 24$ when $T_L = 80$; and the number of low frequency lags of x_H used in the low frequency tests is $h_{LF} \in \{1, 2, 3\}$ and $h_{LF} = 4$ when $T_L = 80$. Low frequency tests use either flow or stock sampling. The max test statistic is computed using a flat weight $\boldsymbol{W}_h = (1/h) \times \boldsymbol{I}_h$, and the p-value is computed using 5,000 draws from the null limit distribution. The Wald test p-value is computed using the parametric bootstrap based on Gonçalves and Killian (2004), with 499 bootstrap replications. Sample sizes are small and medium $T_L \in \{40, 80\}$. Nominal size is $\alpha = 0.05$. The number of Monte Carlo samples drawn is 5,000 for max tests and 1,000 for Wald tests.

Table 3: Rejection Frequencies of High-to-Low Causality Tests Based on MF-VAR(2)

A. Non-Causality: $\boldsymbol{b} = \mathbf{0}_{24\times1}$								B. Decaying Causality: $b_j = (-1)^{j-1}0.3/j$							
A.1 $d = 0.2$ (low persistence in x_H)								B.1 $d = 0.2$ (low persistence in x_H)							
A.1.1 $T_L = 40$								B.1.1 $T_L = 40$							
	MF				LF (flow)		LF (stock)			MF				LF (flow)		LF (stock)	
h_{MF}	Max	Wald		h_{LF}	Max	Wald	Max	Wald	h_{MF}	Max	Wald		h_{LF}	Max	Wald	Max	Wald
16	.062	.047		1	.061	.054	.060	.056	16	.118	.095		1	.079	.045	.386	.302
20	.061	.047		2	.056	.033	.065	.042	20	.109	.092		2	.069	.055	.286	.250
24	.055	.035		3	.060	.046	.061	.045	24	.094	.076		3	.063	.036	.223	.174
A.1.2 $T_L = 80$								B.1.2 $T_L = 80$							
	MF				LF (flow)		LF (stock)			MF				LF (flow)		LF (stock)	
h_{MF}	Max	Wald		h_{LF}	Max	Wald	Max	Wald	h_{MF}	Max	Wald		h_{LF}	Max	Wald	Max	Wald
16	.059	.041		1	.053	.045	.053	.065	16	.284	.290		1	.080	.067	.664	.599
20	.055	.043		2	.049	.045	.056	.048	20	.258	.250		2	.067	.057	.566	.507
24	.058	.043		3	.043	.043	.053	.048	24	.246	.221		3	.067	.053	.475	.454
A.2 $d = 0.8$ (high persistence in x_H)								B.2 $d = 0.8$ (high persistence in x_H)							
A.2.1 $T_L = 40$								B.2.1 $T_L = 40$							
	MF				LF (flow)		LF (stock)			MF				LF (flow)		LF (stock)	
h_{MF}	Max	Wald		h_{LF}	Max	Wald	Max	Wald	h_{MF}	Max	Wald		h_{LF}	Max	Wald	Max	Wald
16	.064	.038		1	.069	.044	.068	.051	16	.233	.120		1	.212	.215	.575	.468
20	.056	.031		2	.055	.045	.053	.051	20	.200	.091		2	.140	.110	.446	.390
24	.050	.035		3	.061	.039	.067	.039	24	.190	.077		3	.125	.094	.376	.292
A.2.2 $T_L = 80$								B.2.2 $T_L = 80$							
	MF				LF (flow)		LF (stock)			MF				LF (flow)		LF (stock)	
h_{MF}	Max	Wald		h_{LF}	Max	Wald	Max	Wald	h_{MF}	Max	Wald		h_{LF}	Max	Wald	Max	Wald
16	.054	.049		1	.056	.035	.061	.044	16	.610	.398		1	.335	.273	.876	.836
20	.053	.036		2	.056	.050	.057	.046	20	.563	.356		2	.247	.218	.805	.748
24	.052	.047		3	.054	.050	.054	.051	24	.545	.271		3	.217	.185	.743	.682

The DGP is MF-VAR(2) with a ratio of sampling frequencies $m = 12$, where b_j signifies the impact of $x_H(\tau_L - 1, m + 1 - j)$ on $x_L(\tau_L)$. Panels A and B concern high-to-low non-causality and decaying causality. We consider low and higher persistence in x_H : $d \in \{0.2, 0.8\}$. Other parameters are specified as follows. There is weak autoregressive persistence in x_L ($a = 0.2$) and low-to-high decaying causality with alternating signs: $c_j = (-1)^{j-1} \times 0.4/j$ for $j = 1, \ldots, 12$. The models estimated have two low frequency lags of x_L (i.e. $q = 2$); the number of high frequency lags of x_H used in the mixed frequency tests is $h_{MF} \in \{16, 20, 24\}$; and the number of low frequency lags of x_H used in the low frequency tests is $h_{LF} \in \{1, 2, 3\}$. Low frequency tests use either flow or stock sampling. The max test statistic is computed using a flat weight $\boldsymbol{W}_h = (1/h) \times \boldsymbol{I}_h$, and the p-value is computed using 5,000 draws from the null limit distribution. The Wald test p-value is computed using the parametric bootstrap based on Gonçalves and Killian (2004), with 499 bootstrap replications. Sample sizes are small and medium $T_L \in \{40, 80\}$. Nominal size is $\alpha = 0.05$. The number of Monte Carlo samples drawn is 5,000 for max tests and 1,000 for Wald tests.

Table 3: Rejection Frequencies of High-to-Low Causality Tests Based on MF-VAR(2) : Continued

C. Lagged Causality: $b_j = 0.3 \times I(j = 24)$							
C.1 $d = 0.2$ (low persistence in x_H)							
C.1.1 $T_L = 40$							
	MF			LF (flow)		LF (stock)	
h_{MF}	Max	Wald	h_{LF}	Max	Wald	Max	Wald
16	.060	.044	1	.061	.036	.064	.042
20	.060	.046	2	.082	.051	.059	.037
24	.091	.055	3	.083	.055	.066	.041
C.1.2 $T_L = 80$							
	MF			LF (flow)		LF (stock)	
h_{MF}	Max	Wald	h_{LF}	Max	Wald	Max	Wald
16	.057	.050	1	.055	.041	.057	.046
20	.051	.046	2	.104	.092	.054	.054
24	.273	.195	3	.099	.073	.073	.069
C.2 $d = 0.8$ (high persistence in x_H)							
C.2.1 $T_L = 40$							
	MF			LF (flow)		LF (stock)	
h_{MF}	Max	Wald	h_{LF}	Max	Wald	Max	Wald
16	.063	.042	1	.064	.040	.065	.041
20	.095	.073	2	.260	.196	.064	.054
24	.438	.108	3	.285	.189	.462	.325
C.2.2 $T_L = 80$							
	MF			LF (flow)		LF (stock)	
h_{MF}	Max	Wald	h_{LF}	Max	Wald	Max	Wald
16	.069	.053	1	.057	.045	.058	.049
20	.162	.096	2	.515	.464	.054	.060
24	.896	.480	3	.588	.505	.808	.725

D. Sporadic: $(b_5, b_{12}, b_{17}, b_{19}) = (-0.2, 0.1, 0.2, -0.35)$							
D.1 $d = 0.2$ (low persistence in x_H)							
D.1.1 $T_L = 40$							
	MF			LF (flow)		LF (stock)	
h_{MF}	Max	Wald	h_{LF}	Max	Wald	Max	Wald
16	.090	.064	1	.072	.039	.066	.043
20	.186	.146	2	.067	.065	.057	.056
24	.153	.089	3	.066	.059	.062	.043
D.1.2 $T_L = 80$							
	MF			LF (flow)		LF (stock)	
h_{MF}	Max	Wald	h_{LF}	Max	Wald	Max	Wald
16	.141	.118	1	.066	.063	.056	.048
20	.502	.474	2	.073	.053	.055	.040
24	.461	.419	3	.068	.071	.057	.061
D.2 $d = 0.8$ (high persistence in x_H)							
D.2.1 $T_L = 40$							
	MF			LF (flow)		LF (stock)	
h_{MF}	Max	Wald	h_{LF}	Max	Wald	Max	Wald
16	.170	.090	1	.158	.106	.117	.067
20	.254	.187	2	.175	.137	.101	.079
24	.256	.136	3	.148	.097	.100	.075
D.2.2 $T_L = 80$							
	MF			LF (flow)		LF (stock)	
h_{MF}	Max	Wald	h_{LF}	Max	Wald	Max	Wald
16	.390	.234	1	.272	.243	.151	.118
20	.621	.655	2	.311	.252	.134	.132
24	.623	.606	3	.269	.226	.163	.152

The DGP is MF-VAR(2) with a ratio of sampling frequencies $m = 12$, where b_j signifies the impact of $x_H(\tau_L - 1, m + 1 - j)$ on $x_L(\tau_L)$. Panels C and D concern high-to-low lagged and sporadic causality. We consider low and higher persistence in x_H : $d \in \{0.2, 0.8\}$. Other parameters are specified as follows. There is weak autoregressive persistence in x_L ($a = 0.2$) and low-to-high decaying causality with alternating signs: $c_j = (-1)^{j-1} \times 0.4/j$ for $j = 1, \ldots, 12$. The models estimated have two low frequency lags of x_L (i.e. $q = 2$); the number of high frequency lags of x_H used in the mixed frequency tests is $h_{MF} \in \{16, 20, 24\}$; and the number of low frequency lags of x_H used in the low frequency tests is $h_{LF} \in \{1, 2, 3\}$. Low frequency tests use either flow or stock sampling. The max test statistic is computed using a flat weight $\boldsymbol{W}_h = (1/h) \times \boldsymbol{I}_h$, and the p-value is computed using 5,000 draws from the null limit distribution. The Wald test p-value is computed using the parametric bootstrap based on Gonçalves and Killian (2004), with 499 bootstrap replications. Sample sizes are small and medium $T_L \in \{40, 80\}$. Nominal size is $\alpha = 0.05$. The number of Monte Carlo samples drawn is 5,000 for max tests and 1,000 for Wald tests.

Table 4: Rejection Frequencies for Low-to-High Causality Tests Based on MF-VAR(1)

A. Non-Causality: $\boldsymbol{c} = \mathbf{0}_{12\times 1}$

A.1 $T_L = 40$

A.1.1 Mixed Frequency Tests

	$r_{MF}=4$		$r_{MF}=8$		$r_{MF}=12$	
h_{MF}	Max	Wald	Max	Wald	Max	Wald
4	.076	.044	.080	.049	.076	.044
8	.103	.044	.117	.033	.113	.044
12	.156	.037	.176	.041	.193	.040
-	-	-	-	-	-	-

A.1.2 Mixed Frequency Tests with MIDAS

	$r_{MF}=4$		$r_{MF}=8$		$r_{MF}=12$	
h_{MF}	Max	Wald	Max	Wald	Max	Wald
4	.071	.040	.069	.041	.071	.050
8	.073	.038	.070	.050	.066	.051
12	.066	.030	.068	.041	.068	.042
-	-	-	-	-	-	-

A.1.3 Low Frequency Tests (Stock Sampling)

	$r_{LF}=1$		$r_{LF}=2$		$r_{LF}=3$	
h_{LF}	Max	Wald	Max	Wald	Max	Wald
1	.057	.045	.053	.042	.057	.053
2	.064	.046	.062	.048	.056	.044
3	.062	.041	.068	.046	.068	.047
-	-	-	-	-	-	-

A.1.4 Low Frequency Tests (Flow Sampling)

	$r_{LF}=1$		$r_{LF}=2$		$r_{LF}=3$	
h_{LF}	Max	Wald	Max	Wald	Max	Wald
1	.057	.048	.054	.059	.050	.050
2	.063	.050	.053	.041	.064	.052
3	.062	.041	.065	.038	.064	.053
-	-	-	-	-	-	-

A.2 $T_L = 80$

A.2.1 Mixed Frequency Tests

	$r_{MF}=4$		$r_{MF}=8$		$r_{MF}=12$		$r_{MF}=24$	
h_{MF}	Max	Wald	Max	Wald	Max	Wald	Max	Wald
4	.060	.060	.058	.056	.062	.060	.062	.048
8	.074	.048	.079	.050	.077	.043	.081	.056
12	.082	.044	.095	.037	.097	.048	.093	.047
24	.161	.069	.176	.052	.200	.046	.225	.050

A.2.2 Mixed Frequency Tests with MIDAS

	$r_{MF}=4$		$r_{MF}=8$		$r_{MF}=12$		$r_{MF}=24$	
h_{MF}	Max	Wald	Max	Wald	Max	Wald	Max	Wald
4	.055	.049	.057	.060	.054	.051	.058	.045
8	.056	.050	.051	.048	.055	.055	.061	.044
12	.060	.048	.058	.040	.052	.049	.056	.040
24	.055	.047	.062	.035	.061	.052	.052	.056

A.2.3 Low Frequency Tests (Stock Sampling)

	$r_{LF}=1$		$r_{LF}=2$		$r_{LF}=3$		$r_{LF}=4$	
h_{LF}	Max	Wald	Max	Wald	Max	Wald	Max	Wald
1	.058	.043	.051	.046	.050	.050	.049	.052
2	.059	.048	.051	.058	.054	.046	.048	.035
3	.061	.044	.055	.032	.057	.047	.057	.047
4	.065	.043	.055	.045	.060	.062	.056	.058

A.2.4 Low Frequency Tests (Flow Sampling)

	$r_{LF}=1$		$r_{LF}=2$		$r_{LF}=3$		$r_{LF}=4$	
h_{LF}	Max	Wald	Max	Wald	Max	Wald	Max	Wald
1	.054	.045	.046	.049	.050	.038	.053	.050
2	.052	.044	.050	.043	.052	.054	.055	.055
3	.057	.050	.051	.054	.056	.050	.056	.051
4	.063	.047	.062	.044	.052	.039	.061	.053

The DGP is MF-VAR(1) with a ratio of sampling frequencies $m = 12$. In Panel A there exists non-causality from x_L to x_H. The AR(1) coefficient of x_H is 0.2, and the AR(1) coefficient of x_L is also 0.2. Mixed frequency statistics use $r_{MF} \in \{4, 8, 12\}$ high frequency leads of x_H, and $h_{MF} \in \{4, 8, 12\}$ high frequency lags of x_H, and when $T_L = 80$ we also use $r_{MF} = 24$ and $h_{MF} = 24$. Low frequency tests use $r_{LF} \in \{1, 2, 3\}$ low frequency leads of aggregated x_H, and $h_{LF} \in \{1, 2, 3\}$ low frequency lags of aggregated x_H, and when $T_L = 80$ we also use $r_{LF} = 4$ and $h_{LF} = 4$. The max test statistic is computed using a flat weight $\boldsymbol{W}_r = (1/r) \times \boldsymbol{I}_r$, and the p-value is computed using 1,000 draws from the null distribution. Wald tests are performed using Gonçalves and Killian's (2004) bootstrap with $N = 499$ replications. Mixed frequency max and Wald tests are computed from models with and without an Almon (MIDAS) polynomial with dimension $s = 3$ for the lag terms of x_H. Sample sizes are $T_L \in \{40, 80\}$. Nominal size is $\alpha = 0.05$. The number of Monte Carlo samples drawn is 5,000 for max tests and 1,000 for Wald tests.

Table 4: Rejection Frequencies for Low-to-High Causality Tests Based on MF-VAR(1) : Continued

B. Decaying Causality: $c_j = (-1)^{j-1}0.45/j$

B.1 $T_L = 40$

B.1.1 Mixed Frequency Tests

	$r_{MF}=4$		$r_{MF}=8$		$r_{MF}=12$	
h_{MF}	Max	Wald	Max	Wald	Max	Wald
4	.490	.515	.337	.367	.279	.304
8	.495	.453	.370	.336	.315	.225
12	.510	.371	.431	.258	.385	.188
-	-	-	-	-	-	-

B.1.2 Mixed Frequency Tests with MIDAS

	$r_{MF}=4$		$r_{MF}=8$		$r_{MF}=12$	
h_{MF}	Max	Wald	Max	Wald	Max	Wald
4	.481	.539	.344	.426	.261	.308
8	.482	.553	.339	.401	.271	.314
12	.487	.566	.338	.426	.269	.338
-	-	-	-	-	-	-

B.1.3 Low Frequency Tests (Stock Sampling)

	$r_{LF}=1$		$r_{LF}=2$		$r_{LF}=3$	
h_{LF}	Max	Wald	Max	Wald	Max	Wald
1	.061	.047	.058	.054	.050	.058
2	.064	.050	.054	.054	.067	.043
3	.070	.058	.068	.043	.066	.044
-	-	-	-	-	-	-

B.1.4 Low Frequency Tests (Flow Sampling)

	$r_{LF}=1$		$r_{LF}=2$		$r_{LF}=3$	
h_{LF}	Max	Wald	Max	Wald	Max	Wald
1	.090	.072	.083	.060	.064	.052
2	.093	.071	.083	.071	.080	.055
3	.097	.058	.088	.062	.082	.057
-	-	-	-	-	-	-

B.2 $T_L = 80$

B.2.1 Mixed Frequency Tests

	$r_{MF}=4$		$r_{MF}=8$		$r_{MF}=12$		$r_{MF}=24$	
h_{MF}	Max	Wald	Max	Wald	Max	Wald	Max	Wald
4	.875	.945	.789	.879	.723	.834	.616	.581
8	.867	.911	.791	.866	.736	.789	.621	.525
12	.878	.889	.792	.807	.728	.748	.624	.451
24	.854	.796	.797	.652	.756	.592	.664	.280

B.2.2 Mixed Frequency Tests with MIDAS

	$r_{MF}=4$		$r_{MF}=8$		$r_{MF}=12$		$r_{MF}=24$	
h_{MF}	Max	Wald	Max	Wald	Max	Wald	Max	Wald
4	.886	.945	.804	.896	.739	.837	.629	.619
8	.894	.935	.818	.885	.747	.815	.629	.629
12	.895	.945	.806	.891	.746	.836	.620	.612
24	.894	.940	.803	.861	.741	.822	.620	.629

B.2.3 Low Frequency Tests (Stock Sampling)

	$r_{LF}=1$		$r_{LF}=2$		$r_{LF}=3$		$r_{LF}=4$	
h_{LF}	Max	Wald	Max	Wald	Max	Wald	Max	Wald
1	.058	.046	.047	.047	.059	.050	.056	.051
2	.069	.060	.059	.053	.055	.044	.057	.053
3	.065	.044	.058	.032	.060	.058	.054	.043
4	.063	.076	.068	.047	.070	.056	.061	.046

B.2.4 Low Frequency Tests (Flow Sampling)

	$r_{LF}=1$		$r_{LF}=2$		$r_{LF}=3$		$r_{LF}=4$	
h_{LF}	Max	Wald	Max	Wald	Max	Wald	Max	Wald
1	.130	.120	.105	.090	.084	.081	.081	.080
2	.129	.106	.103	.085	.087	.095	.081	.071
3	.119	.113	.106	.080	.089	.092	.080	.077
4	.137	.106	.106	.080	.085	.071	.088	.068

The DGP is MF-VAR(1) with a ratio of sampling frequencies $m = 12$. In Panel B there exists decaying causality from x_L to x_H. The AR(1) coefficient of x_H is 0.2, and the AR(1) coefficient of x_L is also 0.2. Mixed frequency statistics use $r_{MF} \in \{4, 8, 12\}$ high frequency leads of x_H, and $h_{MF} \in \{4, 8, 12\}$ high frequency lags of x_H, and when $T_L = 80$ we also use $r_{MF} = 24$ and $h_{MF} = 24$. Low frequency tests use $r_{LF} \in \{1, 2, 3\}$ low frequency leads of aggregated x_H, and $h_{LF} \in \{1, 2, 3\}$ low frequency lags of aggregated x_H, and when $T_L = 80$ we also use $r_{LF} = 4$ and $h_{LF} = 4$. The max test statistic is computed using a flat weight $\boldsymbol{W}_r = (1/r) \times \boldsymbol{I}_r$, and the p-value is computed using 1,000 draws from the null distribution. Wald tests are performed using Gonçalves and Killian's (2004) bootstrap with $N = 499$ replications. Mixed frequency max and Wald tests are computed from models with and without an Almon (MIDAS) polynomial with dimension $s = 3$ for the lag terms of x_H. Sample sizes are $T_L \in \{40, 80\}$. Nominal size is $\alpha = 0.05$. The number of Monte Carlo samples drawn is 5,000 for max tests and 1,000 for Wald tests.

Table 4: Rejection Frequencies for Low-to-High Causality Tests Based on MF-VAR(1) : Continued

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

C.1 $T_L = 40$

C.1.1 Mixed Frequency Tests

	$r_{MF}=4$		$r_{MF}=8$		$r_{MF}=12$	
h_{MF}	Max	Wald	Max	Wald	Max	Wald
4	.079	.054	.080	.050	.201	.163
8	.114	.040	.123	.046	.244	.145
12	.161	.042	.182	.048	.327	.097
-	-	-	-	-	-	-

C.1.2 Mixed Frequency Tests with MIDAS

	$r_{MF}=4$		$r_{MF}=8$		$r_{MF}=12$	
h_{MF}	Max	Wald	Max	Wald	Max	Wald
4	.076	.051	.064	.055	.197	.162
8	.067	.048	.068	.044	.189	.157
12	.068	.040	.069	.048	.196	.161
-	-	-	-	-	-	-

C.1.3 Low Frequency Tests (Stock Sampling)

	$r_{LF}=1$		$r_{LF}=2$		$r_{LF}=3$	
h_{LF}	Max	Wald	Max	Wald	Max	Wald
1	.640	.560	.510	.464	.462	.404
2	.616	.525	.518	.415	.443	.356
3	.616	.485	.501	.368	.441	.310
-	-	-	-	-	-	-

C.1.4 Low Frequency Tests (Flow Sampling)

	$r_{LF}=1$		$r_{LF}=2$		$r_{LF}=3$	
h_{LF}	Max	Wald	Max	Wald	Max	Wald
1	.110	.072	.081	.063	.080	.061
2	.106	.070	.086	.053	.081	.064
3	.107	.083	.089	.072	.087	.067
-	-	-	-	-	-	-

C.2 $T_L = 80$

C.2.1 Mixed Frequency Tests

	$r_{MF}=4$		$r_{MF}=8$		$r_{MF}=12$		$r_{MF}=24$	
h_{MF}	Max	Wald	Max	Wald	Max	Wald	Max	Wald
4	.061	.047	.063	.048	.581	.456	.463	.291
8	.073	.046	.073	.045	.574	.437	.470	.279
12	.088	.041	.088	.053	.563	.415	.489	.237
24	.138	.041	.180	.049	.615	.295	.536	.164

C.2.2 Mixed Frequency Tests with MIDAS

	$r_{MF}=4$		$r_{MF}=8$		$r_{MF}=12$		$r_{MF}=24$	
h_{MF}	Max	Wald	Max	Wald	Max	Wald	Max	Wald
4	.059	.056	.060	.035	.585	.516	.476	.302
8	.058	.061	.064	.059	.582	.497	.467	.313
12	.058	.060	.059	.045	.592	.504	.472	.338
24	.059	.054	.060	.042	.585	.495	.467	.325

C.2.3 Low Frequency Tests (Stock Sampling)

	$r_{LF}=1$		$r_{LF}=2$		$r_{LF}=3$		$r_{LF}=4$	
h_{LF}	Max	Wald	Max	Wald	Max	Wald	Max	Wald
1	.915	.892	.876	.852	.835	.781	.789	.750
2	.919	.912	.863	.817	.812	.785	.791	.715
3	.908	.883	.855	.809	.817	.766	.783	.721
4	.907	.851	.842	.799	.815	.793	.769	.683

C.2.4 Low Frequency Tests (Flow Sampling)

	$r_{LF}=1$		$r_{LF}=2$		$r_{LF}=3$		$r_{LF}=4$	
h_{LF}	Max	Wald	Max	Wald	Max	Wald	Max	Wald
1	.135	.115	.121	.107	.098	.074	.091	.092
2	.143	.128	.108	.100	.103	.084	.101	.090
3	.132	.118	.121	.087	.106	.102	.094	.086
4	.135	.108	.125	.114	.110	.086	.099	.096

The DGP is MF-VAR(1) with a ratio of sampling frequencies $m = 12$. In Panel C there exists lagged causality from x_L to x_H. The AR(1) coefficient of x_H is 0.2, and the AR(1) coefficient of x_L is also 0.2. Mixed frequency statistics use $r_{MF} \in \{4, 8, 12\}$ high frequency leads of x_H, and $h_{MF} \in \{4, 8, 12\}$ high frequency lags of x_H, and when $T_L = 80$ we also use $r_{MF} = 24$ and $h_{MF} = 24$. Low frequency tests use $r_{LF} \in \{1, 2, 3\}$ low frequency leads of aggregated x_H, and $h_{LF} \in \{1, 2, 3\}$ low frequency lags of aggregated x_H, and when $T_L = 80$ we also use $r_{LF} = 4$ and $h_{LF} = 4$. The max test statistic is computed using a flat weight $\boldsymbol{W}_r = (1/r) \times \boldsymbol{I}_r$, and the p-value is computed using 1,000 draws from the null distribution. Wald tests are performed using Gonçalves and Killian's (2004) bootstrap with $N = 499$ replications. Mixed frequency max and Wald tests are computed from models with and without an Almon (MIDAS) polynomial with dimension $s = 3$ for the lag terms of x_H. Sample sizes are $T_L \in \{40, 80\}$. Nominal size is $\alpha = 0.05$. The number of Monte Carlo samples drawn is 5,000 for max tests and 1,000 for Wald tests.

Table 4: Rejection Frequencies for Low-to-High Causality Tests Based on MF-VAR(1) : Continued

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

D.1 $T_L = 40$

D.1.1 Mixed Frequency Tests

	$r_{MF}=4$		$r_{MF}=8$		$r_{MF}=12$	
h_{MF}	Max	Wald	Max	Wald	Max	Wald
4	.360	.304	.295	.281	.425	.440
8	.383	.261	.352	.258	.462	.342
12	.412	.250	.411	.189	.519	.268
-	-	-	-	-	-	-

D.1.2 Mixed Frequency Tests with MIDAS

	$r_{MF}=4$		$r_{MF}=8$		$r_{MF}=12$	
h_{MF}	Max	Wald	Max	Wald	Max	Wald
4	.373	.344	.302	.319	.425	.474
8	.364	.348	.310	.311	.432	.473
12	.363	.344	.308	.281	.423	.469
-	-	-	-	-	-	-

D.1.3 Low Frequency Tests (Stock Sampling)

	$r_{LF}=1$		$r_{LF}=2$		$r_{LF}=3$	
h_{LF}	Max	Wald	Max	Wald	Max	Wald
1	.067	.047	.059	.050	.049	.041
2	.061	.044	.065	.035	.064	.052
3	.067	.055	.066	.049	.067	.048
-	-	-	-	-	-	-

D.1.4 Low Frequency Tests (Flow Sampling)

	$r_{LF}=1$		$r_{LF}=2$		$r_{LF}=3$	
h_{LF}	Max	Wald	Max	Wald	Max	Wald
1	.066	.049	.064	.054	.058	.038
2	.070	.053	.067	.056	.067	.057
3	.077	.031	.071	.042	.066	.054
-	-	-	-	-	-	-

D.2 $T_L = 80$

D.2.1 Mixed Frequency Tests

	$r_{MF}=4$		$r_{MF}=8$		$r_{MF}=12$		$r_{MF}=24$	
h_{MF}	Max	Wald	Max	Wald	Max	Wald	Max	Wald
4	.745	.730	.731	.735	.919	.948	.845	.820
8	.749	.691	.724	.699	.919	.928	.849	.755
12	.739	.637	.723	.671	.915	.909	.856	.720
24	.724	.539	.745	.520	.914	.800	.854	.531

D.2.2 Mixed Frequency Tests with MIDAS

	$r_{MF}=4$		$r_{MF}=8$		$r_{MF}=12$		$r_{MF}=24$	
h_{MF}	Max	Wald	Max	Wald	Max	Wald	Max	Wald
4	.762	.745	.736	.772	.931	.962	.854	.834
8	.762	.759	.748	.773	.935	.963	.857	.828
12	.772	.747	.742	.761	.935	.969	.857	.851
24	.762	.730	.736	.769	.927	.965	.851	.872

D.2.3 Low Frequency Tests (Stock Sampling)

	$r_{LF}=1$		$r_{LF}=2$		$r_{LF}=3$		$r_{LF}=4$	
h_{LF}	Max	Wald	Max	Wald	Max	Wald	Max	Wald
1	.059	.051	.054	.042	.055	.050	.048	.031
2	.063	.046	.055	.044	.058	.045	.057	.042
3	.058	.043	.058	.055	.058	.053	.052	.049
4	.067	.052	.064	.063	.059	.049	.066	.047

D.2.4 Low Frequency Tests (Flow Sampling)

	$r_{LF}=1$		$r_{LF}=2$		$r_{LF}=3$		$r_{LF}=4$	
h_{LF}	Max	Wald	Max	Wald	Max	Wald	Max	Wald
1	.072	.078	.066	.069	.064	.062	.059	.051
2	.074	.065	.064	.064	.070	.078	.063	.048
3	.082	.069	.065	.068	.057	.047	.066	.068
4	0.72	.079	.069	.073	.071	.054	.074	.070

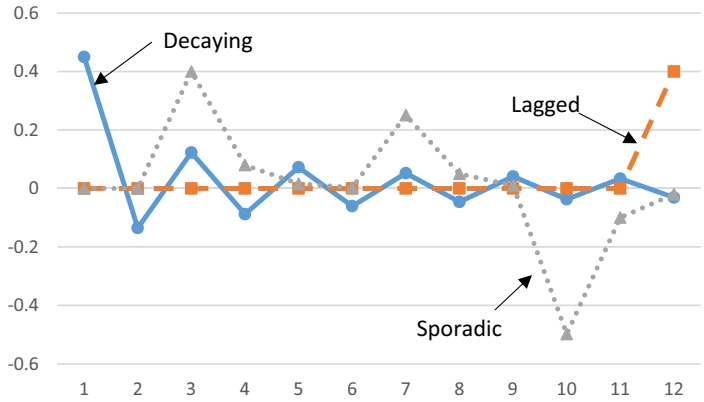
The DGP is MF-VAR(1) with a ratio of sampling frequencies $m = 12$. In Panel D there exists sporadic causality from x_L to x_H. The AR(1) coefficient of x_H is 0.2, and the AR(1) coefficient of x_L is also 0.2. Mixed frequency statistics use $r_{MF} \in \{4, 8, 12\}$ high frequency leads of x_H, and $h_{MF} \in \{4, 8, 12\}$ high frequency lags of x_H, and when $T_L = 80$ we also use $r_{MF} = 24$ and $h_{MF} = 24$. Low frequency tests use $r_{LF} \in \{1, 2, 3\}$ low frequency leads of aggregated x_H, and $h_{LF} \in \{1, 2, 3\}$ low frequency lags of aggregated x_H, and when $T_L = 80$ we also use $r_{LF} = 4$ and $h_{LF} = 4$. The max test statistic is computed using a flat weight $\boldsymbol{W}_r = (1/r) \times \boldsymbol{I}_r$, and the p-value is computed using 1,000 draws from the null distribution. Wald tests are performed using Gonçalves and Killian's (2004) bootstrap with $N = 499$ replications. Mixed frequency max and Wald tests are computed from models with and without an Almon (MIDAS) polynomial with dimension $s = 3$ for the lag terms of x_H. Sample sizes are $T_L \in \{40, 80\}$. Nominal size is $\alpha = 0.05$. The number of Monte Carlo samples drawn is 5,000 for max tests and 1,000 for Wald tests.

Table 5: Sample Statistics of U.S. Interest Rates and Real GDP Growth

	mean	median	std. dev.	skewness	kurtosis
weekly 10 Year Treasury constant maturity rate	6.555	6.210	2.734	0.781	3.488
weekly Federal Funds rate	5.563	5.250	3.643	0.928	4.615
spread (10-Year T-bill minus Fed. Funds)	0.991	1.160	1.800	-1.198	5.611
percentage growth rate of quarterly GDP	3.151	3.250	2.349	-0.461	3.543

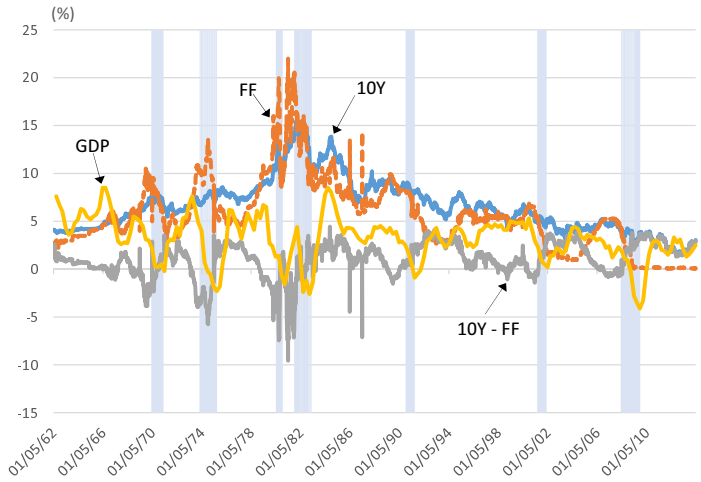
The sample period is January 5, 1962 through December 31, 2013, covering 2,736 weeks or 208 quarters.

Figure 1: Low-to-High Causal Patterns in Reduced Form



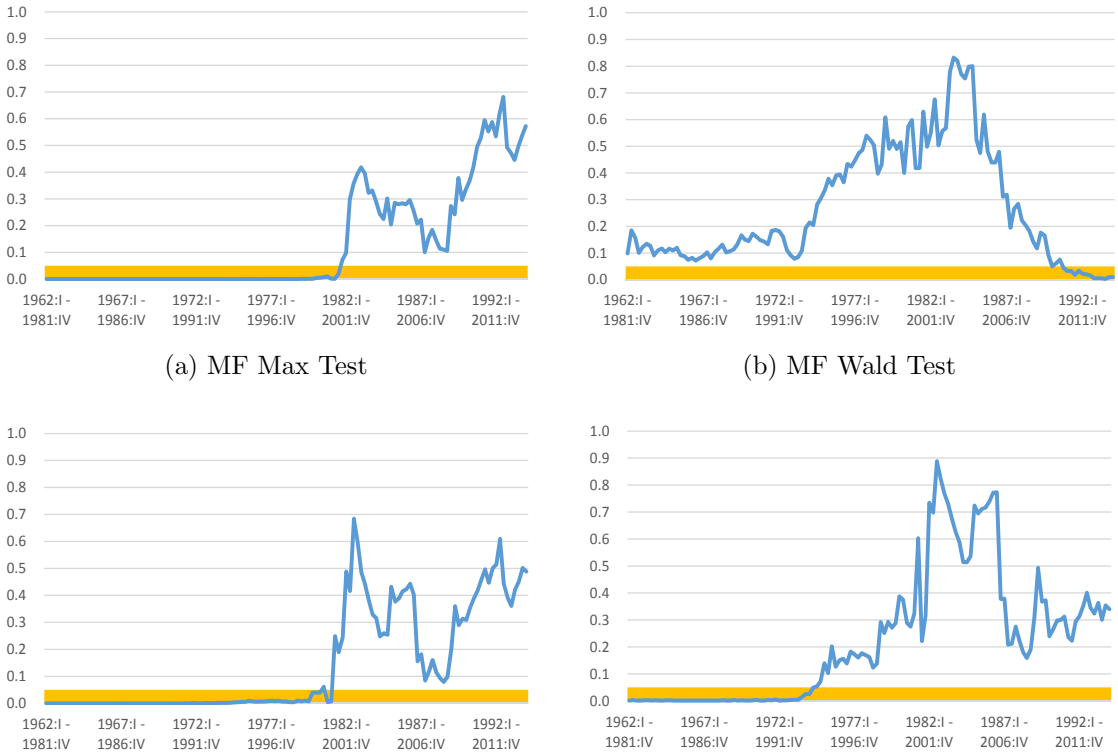
Note: In the low-to-high causality simulation experiment, we start with a structural MF-VAR(1) data generating process, and transform it to a reduced-form MF-VAR(1). This figure shows how each causal pattern in the structural form is transformed in the reduced form. The AR(1) parameter of x_H is fixed at $d = 0.2$. The horizontal axis has the first lag through the twelfth lag in the reduced form, and the vertical axis has a coefficient of each lag. The blue, solid line with circles is a reduced-form causal pattern implied by *decaying causality*: $c_j = (-1)^{j-1} \times 0.45/j$. The red, dashed line with squares is a reduced-form causal pattern implied by *lagged causality*: $c_j = 0.4 \times I(j = 12)$. The gray, dotted line with triangles is a reduced-form causal pattern implied by *sporadic causality*: $(c_3, c_7, c_{10}) = (0.4, 0.25, -0.5)$ and all other c's are zeros.

Figure 2: Time Series Plot of U.S. Interest Rates and Real GDP Growth



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

Figure 3: P-values for Tests of Non-Causality from Interest Rate Spread to GDP



(a) MF Max Test

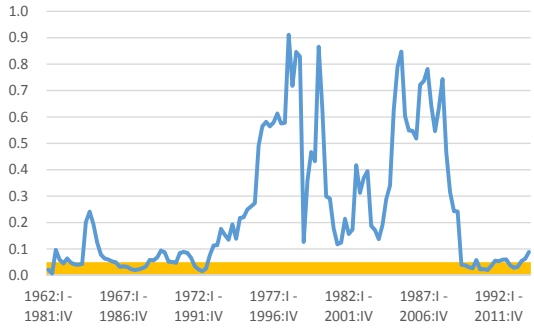
(b) MF Wald Test

(c) LF Max Test

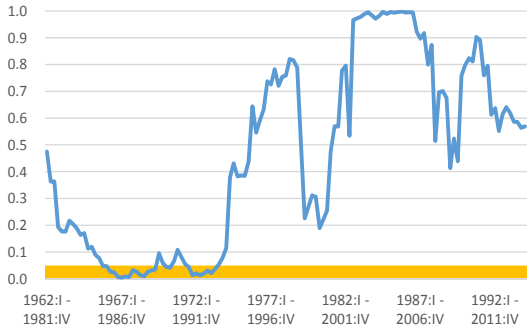
(d) LF Wald Test

Panel (a) contains rolling window p-values for the MF max test, Panel (b) represents the MF Wald test, Panel (c) the LF max test, and Panel (d) the LF Wald test. MF tests concern weekly interest rate spread and quarterly GDP growth, while LF tests concern quarterly interest rate spread and GDP growth. The sample period is January 5, 1962 through December 31, 2013, covering 2,736 weeks or 208 quarters. The window size is 80-quarters. The shaded area is $[0, 0.05]$, hence any p-value in that range suggests rejection of non-causality from the interest rate spread to GDP growth at the 5% level for that window.

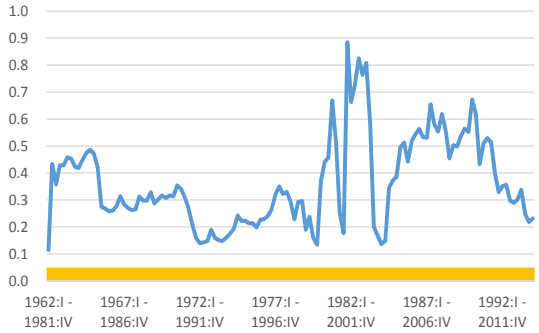
Figure 4: Rolling Window P-values for Tests of Non-Causality from GDP to Interest Rate Spread



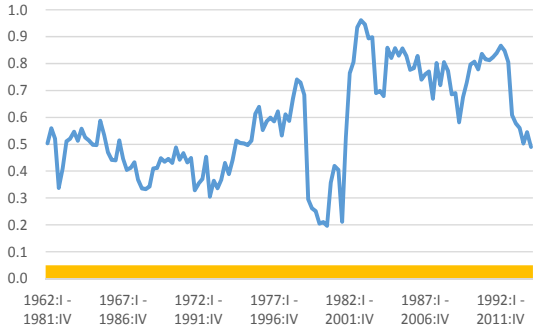
(a) MF Max Test (w/o MIDAS)



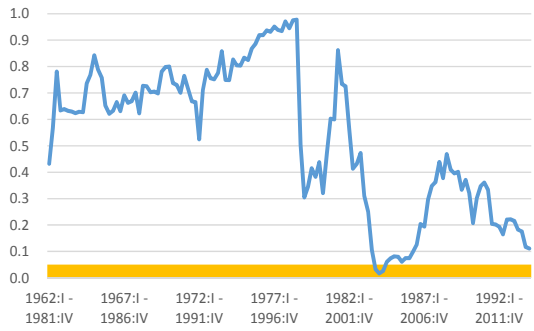
(b) MF Wald Test (w/o MIDAS)



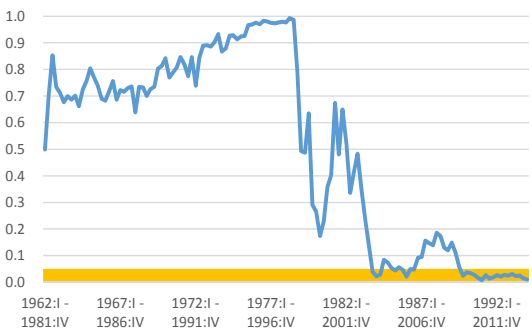
(c) MF Max Test (w/ MIDAS)



(d) MF Wald Test (w/ MIDAS)



(e) LF Max Test



(f) LF Wald Test

Panel (a) contains rolling window p-values for the MF max test without MIDAS polynomial, Panel (b) represents the MF Wald test without MIDAS polynomial, Panel (c) the MF max test with MIDAS polynomial, Panel (d) the MF Wald test with MIDAS polynomial, Panel (e) the LF max test, and Panel (f) the LF Wald test. MF tests concern weekly interest rate spread and quarterly GDP growth, while LF tests concern quarterly interest rate spread and GDP growth. The sample period is January 5, 1962 through December 31, 2013, covering 2,736 weeks or 208 quarters. The window size is 80-quarters. The shaded area is $[0, 0.05]$, hence any p-value in that range suggests rejection of non-causality from GDP growth to the interest rate spread at the 5% level for that window.