Simple Granger Causality Tests for Mixed Frequency Data*

Eric Ghysels[†] Jonathan B. Hill[‡] Kaiji Motegi[§]

July 30, 2016

Abstract

The paper presents simple Granger causality tests applicable to any mixed frequency sampling data setting, which feature remarkable power properties even with a relatively small sample size. 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. We show via Monte Carlo simulations that the max test is particularly powerful for causality with a large time lag.

Keywords: Granger causality test, Local asymptotic power, Max test, MIDAS regression, Sims test, Temporal aggregation.

---

*We would like to thank Manfred Deistler, Atsushi Inoue, Helmut Lütkepohl, Massimiliano Marcellino, Yasumasa Matsuda, Daisuke Nagakura, Eric Renault, Peter Robinson, seminar participants at NYU Stern and Université Libre de Bruxelles, conference participants at the Recent Developments in Time Series and Related Fields (Tohoku University), 2015 Japanese Joint Statistical Meeting (Okayama University), the 11th World Congress of the Econometric Society (Montréal), the 11th International Symposium on Econometric Theory and Applications (Hitotsubashi University), the 25th EC^2 Conference - Advances in Forecasting (Universitat Pompeu Fabra), and the 2014 NBER-NSF Time Series Conference (Federal Reserve Bank of St. Louis) for helpful comments.

[†]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 test 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 (2016) 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' (2016) 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 lag order increases. In these cases, Ghysels, Hill, and Motegi's (2016) Wald test 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}

[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 and Lin (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.

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 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

[5] We prove the claim for a test of non-causality from a high to a low frequency variable in Theorem 2.5, 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 $\sigma(x(t) : t \leq \tau)$.

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 approaching one for any direction $b \neq 0$. Our test approach can be applied in many other settings, whenever hypothesis testing involves many zero restrictions on a parameter set.

As the above example reveals, our method has broad applications for inference in time series regressions with either the 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.

In our theoretical analysis, we compare the max test based on mixed frequency [MF] data, with a max-test based on low frequency [LF] data, and Wald tests variously based on mixed or 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 have roughly equal power in most cases, but the former better captures causality with a large time lag.

The remainder of the paper is organized as follows. Sections 2 and 3 present the max test statistic and derive its asymptotic properties, respectively for testing non-causality from high-to-low and low-to-high frequencies. We analyze local power in Section 4. In Section 5 we run Monte Carlo simulations, Section 6 presents an empirical application, and Section 7 concludes the paper.

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.[6] We first formulate a data generating process (DGP) governing these variables. Let m

[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. See Dufour and Renault (1998), Dufour, Pelletier, and Renault (2006) and Hill (2007) for further discussion.

denote the *ratio of sampling frequencies*, i.e. the number of high frequency time periods for each low frequency time period $\tau_L \in \mathbb{Z}$. 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 σ-field $\mathcal{F}_{\tau_L} \equiv \sigma(\boldsymbol{X}(\tau) : \tau \leq \tau_L)$. We assume as in Ghysels (2016) and Ghysels, Hill, and Motegi (2016) 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 finite $p \geq 1$:

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

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 a positive definite covariance matrix $\boldsymbol{\Omega} \equiv E[\boldsymbol{\epsilon}(\tau_L)\boldsymbol{\epsilon}(\tau_L)']$.

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 and a govern the autoregressive property of x_H and $x_L, respectively.$

The coefficients b and c in (2.1) are relevant for Granger causality, so we explain how they are labeled. 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 they 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 they 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]

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* $\{\{x_H(\tau_L, j)\}_{j=1}^{m}\}_{\tau_L}$.

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

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 here, 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 $j \in \{1, \dots, pm\}$ 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 $j = 1, \dots, m$ 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 $\boldsymbol{X}_L(\tau_L - 1) = [x_L(\tau_L - 1), \dots, x_L(\tau_L - p)]'$, $\boldsymbol{X}_H(\tau_L - 1) = [x_H(\tau_L - 1, m + 1 - 1), \dots, x_H(\tau_L - 1, m + 1 - pm)]'$, $\boldsymbol{a} = [a_1, \dots, a_p]'$, and $\boldsymbol{b} = [b_1, \dots, b_{pm}]'$. 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 (2016), 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 order to test the non-causality hypothesis $H_0 : \boldsymbol{b} = \boldsymbol{0}_{pm \times 1}$, 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

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

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 (2016). 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 $\tau_L = 1, \ldots, T_L$. They estimate the parameters in (2.4) by least squares and then test $H_0 : \beta_1 = \cdots = \beta_h = 0$ via a Wald test. Model (2.4) contains DGP (2.2) as a special case when $q \geq p$ and $h \geq pm$, hence the Wald test is trivially consistent if $q \geq p$ and $h \geq pm$.

A potential problem here is that pm, the true lag order of x_H, may be quite large in some applications, even when 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 (2016)). 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). In a nutshell, this is 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 can be compactly written as

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

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 statistic identifies the hypotheses.

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:

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

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. Assume the weights $w_{T_L} > 0$ *a.s.* for all T_L and $j = 1, \ldots, h$ and limits $w_j > 0$ for all $j = 1, \ldots, h$ so that we have a non-trivial result under the alternative. 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 $\sum_{j=1}^{h} (\sqrt{T_L} w_{T_L,j} \hat{\beta}_j)^2$.

The maximum of a sequence of a statistics with (or without) pointwise Gaussian limits has a long history (e.g. Gnedenko (1943)), including the maximum correlation over an increasing sequence of integer displacements. See, e.g., Berman (1964) and Hannan (1974). More recently, test statistic functionals when there is a nuisance parameter under the alternative include the supremum and average, e.g. Chernoff and Zacks (1964) and Davies (1977). See Andrews and Ploberger (1994).

Finally, we need to restrict h, the number of models estimated.

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 High-to-Low Non-Causality

Stack all parameters across the h models (2.6), 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 $w_{T_L,1}, \ldots, w_{T_L,h}$. Similarly, let $\boldsymbol{W}_h$ be an $h \times h$ diagonal matrix whose diagonal elements are $w_1, \ldots, w_h$.

Under Assumptions 2.1–2.4, it is not hard to prove the asymptotic normality of a suitably normalized $\hat{\boldsymbol{\theta}}$ and hence $\hat{\boldsymbol{\beta}}$. A simple weak convergence argument then suffices for the max test statistic. All proofs are presented in Appendix B.

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 covariance matrix:

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

where

$$\boldsymbol{\mathcal{S}} \equiv [\boldsymbol{\Sigma}_{j,i}] \in \mathbb{R}^{(q+1)h\times(q+1)h} \text{ and } \boldsymbol{\Sigma}_{j,i} \equiv \boldsymbol{\Gamma}_{j,j}^{-1} \tilde{\boldsymbol{\Gamma}}_{j,i} \boldsymbol{\Gamma}_{i,i}^{-1} \in \mathbb{R}^{(q+1)\times(q+1)} \tag{2.9}$$

$$\boldsymbol{\Gamma}_{j,i} \equiv E\left[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{x}_i(\tau_L - 1)'\right] \text{ and } \tilde{\boldsymbol{\Gamma}}_{j,i} \equiv E\left[\epsilon_L^2(\tau_L)\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{x}_i(\tau_L - 1)'\right]$$

$$\boldsymbol{R} \equiv \begin{bmatrix} \boldsymbol{0}_{1\times q} & 1 & \cdots & \cdots & \boldsymbol{0}_{1\times q} & 0 \\ \vdots & \vdots & \cdots & \cdots & \vdots & \vdots \\ \boldsymbol{0}_{1\times q} & 0 & \cdots & \cdots & \boldsymbol{0}_{1\times q} & 1 \end{bmatrix} \in \mathbb{R}^{h\times(q+1)h}.$$

Remark 2.5. Under Assumption 2.1 the errors $\epsilon_L(\tau_L)$ is an adapted martingale difference. Suppose also that $\epsilon_L^2(\tau_L) - \sigma_L^2$ is an adapted martingale difference with $\sigma_L^2 \equiv E[\epsilon_L^2(\tau_L)]$ such that there are no conditional volatility effects. Then the covariance matrix $\boldsymbol{V}$ reduces to

$$\boldsymbol{V} \equiv \sigma_L^2 \boldsymbol{W}_h \boldsymbol{R}\boldsymbol{\mathcal{S}}\boldsymbol{R}' \boldsymbol{W}_h \in \mathbb{R}^{h\times h} \text{ where } \boldsymbol{\mathcal{S}} \equiv [\boldsymbol{\Sigma}_{j,i}] \text{ with } \boldsymbol{\Sigma}_{j,i} \equiv \boldsymbol{\Gamma}_{j,j}^{-1}\boldsymbol{\Gamma}_{j,i}\boldsymbol{\Gamma}_{i,i}^{-1}. \tag{2.10}$$

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 $N(\boldsymbol{0}_{h\times 1}, \hat{\boldsymbol{V}}_{T_L})$. 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 $\hat{p} = \bar{F}^0(\hat{\mathcal{T}}) + o_p(1)$.

A consistent estimator $\hat{\boldsymbol{V}}_{T_L}$ of $\boldsymbol{V}$ in (2.8) is computed as follows. Define $\hat{\boldsymbol{\Gamma}}_{j,i} \equiv 1/T_L \sum_{\tau_L=1}^{T_L} \boldsymbol{x}_j(\tau_L - 1)\boldsymbol{x}_i(\tau_L-1)'$ and $\hat{\hat{\boldsymbol{\Gamma}}}_{j,i} \equiv 1/T_L \sum_{\tau_L=1}^{T_L} \hat{u}_{L,i}^2(\tau_L)\boldsymbol{x}_j(\tau_L-1)\boldsymbol{x}_i(\tau_L-1)'$ with residuals $\hat{u}_{L,i}(\tau_L) \equiv x_L(\tau_L) - \boldsymbol{x}_i(\tau_L-1)'\hat{\boldsymbol{\theta}}_i$ from model (2.5). Now define $\hat{\boldsymbol{\Sigma}}_{j,i} \equiv \hat{\boldsymbol{\Gamma}}_{j,j}^{-1}\hat{\hat{\boldsymbol{\Gamma}}}_{j,i}\hat{\boldsymbol{\Gamma}}_{i,i}^{-1}$, $\hat{\boldsymbol{\mathcal{S}}} \equiv [\hat{\boldsymbol{\Sigma}}_{i,j}]$ and

$$\hat{\boldsymbol{V}}_{T_L} \equiv \boldsymbol{W}_{T_L,h}\boldsymbol{R}\hat{\boldsymbol{\mathcal{S}}}\boldsymbol{R}'\boldsymbol{W}_{T_L,h}. \tag{2.11}$$

Theorem 2.3. Under Assumptions 2.1 - 2.4 $\hat{\boldsymbol{V}}_{T_L} \xrightarrow{p} \widetilde{\boldsymbol{V}}$ where $\widetilde{\boldsymbol{V}}$ is some matrix that satisfies $||\widetilde{\boldsymbol{V}}|| < \infty$. Under the null hypothesis $\widetilde{\boldsymbol{V}} = \boldsymbol{V}$, where $\boldsymbol{V}$ is the asymptotic covariance matrix defined in (2.8).

2.1.3 Identification of Null and Alternative Hypotheses

We now show that as long as $h \geq pm$, that is 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* $\boldsymbol{\beta} = \mathbf{0}_{h \times 1}$.

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 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 2.4. 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

$$\boldsymbol{\theta}_j^* \equiv \begin{bmatrix} \alpha_{1,j}^* \\ \vdots \\ \alpha_{p,j}^* \\ \alpha_{p+1,j}^* \\ \vdots \\ \alpha_{q,j}^* \\ \beta_j^* \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_p \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} + \underbrace{\left[E\left[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{x}_j(\tau_L - 1)'\right]\right]^{-1}}_{=\boldsymbol{\Gamma}_{j,j}^{-1}\colon\, (q+1)\times(q+1)} \times \underbrace{E\left[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{X}_H(\tau_L - 1)'\right]}_{\equiv \boldsymbol{C}_j\colon\, (q+1)\times pm} \boldsymbol{b}, \quad (2.12)$$

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 $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}^* = \boldsymbol{R}\boldsymbol{\theta}^*$.

Remark 2.6. 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 2.4 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 $\boldsymbol{\beta}^* = \mathbf{0}_{h\times 1}$ whenever there is non-causality (i.e $\boldsymbol{b} = \mathbf{0}_{pm\times 1}$), regardless of the relative magnitude of h and pm. Second, as the next result proves, $\boldsymbol{b} = \mathbf{0}_{pm\times 1}$ whenever $\boldsymbol{\beta}^* = \mathbf{0}_{h\times 1}$, 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 2.5. Let Assumptions 2.1, 2.2, 2.4 and 2.5 hold. Then $\boldsymbol{\beta}^* = \mathbf{0}_{h\times 1}$ implies $\boldsymbol{b} = \mathbf{0}_{pm\times 1}$, hence $\boldsymbol{\beta}^* = \mathbf{0}_{h\times 1}$ *if and only if* $\boldsymbol{b} = \mathbf{0}_{pm\times 1}$. Therefore $\hat{\boldsymbol{\beta}} \xrightarrow{p} \mathbf{0}_{pm\times 1}$ *if and only if* $\boldsymbol{b} = \mathbf{0}_{pm\times 1}$.

Theorems 2.1 and 2.5 together imply the max test statistic has its intended limit properties under either hypothesis. First, observe that the max test statistic construction (2.7) with non-trivial weight limits $w_j > 0$ 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 2.4 and 2.5 $\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}^* \neq \mathbf{0}_{h\times 1}$ 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 2.6. 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.

An immediate consequence of the limit distribution Theorem 2.1, identification Theorem 2.5 and consistency Theorem 2.6, is the limiting null distribution arises *if and only if* H_0 is true.

Corollary 2.7. 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 $p = 1$:

$$\begin{bmatrix} x_H(\tau_L, 1) \\ x_H(\tau_L, 2) \\ x_L(\tau_L) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1/\rho & 1 & 0 \end{bmatrix} \begin{bmatrix} x_H(\tau_L - 1, 1) \\ x_H(\tau_L - 1, 2) \\ x_L(\tau_L - 1) \end{bmatrix} + \begin{bmatrix} \epsilon_H(\tau_L, 1) \\ \epsilon_H(\tau_L, 2) \\ \epsilon_L(\tau_L) \end{bmatrix} \tag{2.13}$$

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

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 C for a proof.

Now 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:

$$
\begin{aligned}
x_L(\tau_L) &= \sum_{k=1}^{q} \alpha_k x_L(\tau_L - k) + \sum_{j=1}^{h} \beta_j x_H(\tau_L - j) + u_L(\tau_L) \\
&= [\boldsymbol{X}_L^{(q)}(\tau_L - 1)', \underbrace{x_H(\tau_L - 1), \ldots, x_H(\tau_L - h)}_{\equiv \underline{\boldsymbol{X}}_H(\tau_L - 1)'}] \underbrace{\begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_q \\ \beta_1 \\ \vdots \\ \beta_h \end{bmatrix}}_{\equiv \boldsymbol{\theta}^{(LF)}} + u_L(\tau_L) \\
&= \underbrace{[\boldsymbol{X}_L^{(q)}(\tau_L - 1)', \underline{\boldsymbol{X}}_H(\tau_L - 1)']}_{\equiv \underline{\boldsymbol{x}}(\tau_L - 1)'} \boldsymbol{\theta}^{(LF)} + u_L(\tau_L).
\end{aligned}
\tag{2.14}
$$

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 the low frequency framework.[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:

$$
\boldsymbol{\theta}^{(LF)*} \equiv \begin{bmatrix} \alpha_1^* \\ \vdots \\ \alpha_p^* \\ \alpha_{p+1}^* \\ \vdots \\ \alpha_q^* \\ \boldsymbol{\beta}^* \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_p \\ 0 \\ \vdots \\ 0 \\ \mathbf{0}_{h \times 1} \end{bmatrix} + \underbrace{\left[E\left[\underline{\boldsymbol{x}}(\tau_L - 1) \underline{\boldsymbol{x}}(\tau_L - 1)' \right] \right]^{-1}}_{\equiv \underline{\boldsymbol{\Gamma}}^{-1}:\ (q+h) \times (q+h)} \underbrace{E\left[\underline{\boldsymbol{x}}(\tau_L - 1) \boldsymbol{X}_H(\tau_L - 1)' \right]}_{\equiv \underline{\boldsymbol{C}}:\ (q+h) \times pm} \boldsymbol{b}, \tag{2.15}
$$

where $\boldsymbol{\beta}^* = [\beta_1^*, \ldots, \beta_h^*]'$. The derivation of (2.15) is omitted since it is similar to the proof of Theorem 2.4.

---

[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 (2.15) $\boldsymbol{\beta}^* = \mathbf{0}_{h\times 1}$.

Example 2.3 (Inconsistency of Low Frequency Wald Test)**.** Consider an even simpler DGP than (2.13) with $m = 2$ and $p = 1$:

$$\begin{bmatrix} x_H(\tau_L,1) \\ x_H(\tau_L,2) \\ x_L(\tau_L) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ b_2 & b_1 & 0 \end{bmatrix} \begin{bmatrix} x_H(\tau_L-1,1) \\ x_H(\tau_L-1,2) \\ x_L(\tau_L-1) \end{bmatrix} + \begin{bmatrix} \epsilon_H(\tau_L,1) \\ \epsilon_H(\tau_L,2) \\ \epsilon_L(\tau_L) \end{bmatrix}, \ \boldsymbol{\epsilon}(\tau_L) \overset{mds}{\sim} (\mathbf{0}_{3\times 1}, \boldsymbol{I}_{3\times 1}), \ \boldsymbol{b} \neq \mathbf{0}_{2\times 1}. \quad (2.16)$$

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 (2.16) 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 (2.15).

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.

[10] Equation (2.16) 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.

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:

$$\begin{aligned} x_L(\tau_L) &= \sum_{k=1}^{q} \alpha_{k,j} x_L(\tau_L - k) + \beta_j x_H(\tau_L - j) + u_{L,j}(\tau_L) \\ &= \underbrace{[\boldsymbol{X}_L^{(q)}(\tau_L-1)', x_H(\tau_L - j)]}_{\equiv \underline{\boldsymbol{x}}_j(\tau_L-1)'} \underbrace{\begin{bmatrix} \alpha_{1,j} \\ \vdots \\ \alpha_{q,j} \\ \beta_j \end{bmatrix}}_{\equiv \boldsymbol{\theta}_j^{(LF)}} + u_{L,j}(\tau_L), \quad j = 1, \ldots, h. \end{aligned} \tag{2.17}$$

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 (2.12):

$$\boldsymbol{\theta}_j^{(LF)*} \equiv \begin{bmatrix} \alpha_{1,j}^* \\ \vdots \\ \alpha_{p,j}^* \\ \alpha_{p+1,j}^* \\ \vdots \\ \alpha_{q,j}^* \\ \beta_j^* \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_p \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix} + \underbrace{\left[E\left[\underline{\boldsymbol{x}}_j(\tau_L-1)\underline{\boldsymbol{x}}_j(\tau_L-1)'\right]\right]^{-1}}_{\equiv \underline{\boldsymbol{\Gamma}}_{j,j}^{-1}:\ (q+1)\times(q+1)} \underbrace{E\left[\underline{\boldsymbol{x}}_j(\tau_L-1)\boldsymbol{X}_H(\tau_L-1)'\right]}_{\equiv \underline{\boldsymbol{C}}_j:\ (q+1)\times pm} \boldsymbol{b}. \tag{2.18}$$

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

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

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 $\boldsymbol{x}_j(\tau_L - 1)$. The Gaussian limit distribution covariance $\boldsymbol{V} \equiv \boldsymbol{W}_h \boldsymbol{R}\boldsymbol{\mathcal{S}}\boldsymbol{R}'\boldsymbol{W}_h \in \mathbb{R}^{h\times h}$ in (2.8) is therefore now defined with a different $\boldsymbol{\mathcal{S}}$ based on $\underline{\boldsymbol{x}}_j(\tau_L - 1)$, and residuals are based on (2.17). Using the same logic as Example 2.3, $\hat{\mathcal{T}}^{(LF)}$ 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 the MF-VAR(p) data generating process (2.1) is $H_0 : \boldsymbol{c} = \boldsymbol{0}_{pm \times 1}$.

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 $r \geq 1$ 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^2_r limit distribution under Assumptions 2.1-2.4, as long as $q \geq p$ and $h \geq pm$ which ensures (3.1) contains the true DGP.

3.1.1 Parsimonious Regressions: Max Test

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

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

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:

$$x_L(\tau_L) = \boldsymbol{y}_j(\tau_L - 1)' \boldsymbol{\phi}_j + u_{L,j}(\tau_L). \tag{3.3}$$

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

: $\boldsymbol{c} = \mathbf{0}_{pm\times 1}$ implies $\boldsymbol{\gamma} = \mathbf{0}_{r\times 1}$ for any $r \geq 1$, which justifies a mixed frequency max test statistic for low-to-high causality:

$$\hat{\mathcal{U}} \equiv \max_{1\leq j\leq r} \left(\sqrt{T_L} w_{T_L,j} \hat{\gamma}_j\right)^2 . \tag{3.4}$$

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 $H_0 : \boldsymbol{c} = \mathbf{0}_{pm\times 1}$ 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 $N(\mathbf{0}_{r\times 1}, \tilde{\boldsymbol{V}})$ with positive definite covariance matrix: $\tilde{\boldsymbol{V}} \equiv \boldsymbol{W}_r \tilde{\boldsymbol{R}} \tilde{\boldsymbol{\mathcal{S}}} \tilde{\boldsymbol{R}}' \boldsymbol{W}_r \in \mathbb{R}^{r\times r}$, where $\tilde{\boldsymbol{\mathcal{S}}}$ 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 computing residuals from (3.3); and selection matrix $\tilde{\boldsymbol{R}}$ is an 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 2.7 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 $\boldsymbol{\gamma} \neq \mathbf{0}_{r\times 1}$ in (3.2) follows under low-to-high causality $\boldsymbol{c} \neq \mathbf{0}_{pm\times 1}$, 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]

[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 $T_L = 120$

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 parsimonious models now become for $j = 1, \ldots, r$:

$$x_L(\tau_L) = \sum_{k=1}^{q} \alpha_k x_L(\tau_L - k) + \sum_{k=1}^{h} \omega_k(\boldsymbol{\pi}) x_H(\tau_L - 1, m + 1 - k) + \gamma_j x_H(\tau_L + 1, j) + u_{L,j}(\tau_L), \qquad (3.5)$$

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 (2016)). In our simulation study we use the Almon polynomial $\omega_k(\boldsymbol{\pi}) = \sum_{l=1}^{s} \pi_l k^l$, hence model (3.5) 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).

3.2 Low Frequency Approach

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

$$x_L(\tau_L) = \sum_{k=1}^{q} \alpha_{k,j} x_L(\tau_L - k) + \sum_{k=1}^{h_{LF}} \beta_{k,j} x_H(\tau_L - k) + \gamma_j x_H(\tau_L + j) + u_{L,j}(\tau_L), \quad j = 1, \ldots, r_{LF}. \quad (3.6)$$

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.

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 (3.6) by least squares, and use a low frequency max test statistic as in (3.4):

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

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 (3.6) 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. The aggregated high frequency variable is:

$$\begin{aligned}
\sum_{k=1}^{h_{LF}} \beta_k x_H(\tau_L - k) &= \sum_{k=1}^{h_{LF}} \beta_k \sum_{l=1}^{m} \delta_l x_H(\tau_L - k, l) \\
&= \beta_1 \delta_m x_H(\tau_L - 1, m + 1 - 1) + \cdots + \beta_1 \delta_1 x_H(\tau_L - 1, m + 1 - m) \\
&\quad + \cdots + \beta_{h_{LF}} \delta_m x_H(\tau_L - h_{LF}, m + 1 - 1) + \cdots + \beta_{h_{LF}} \delta_1 x_H(\tau_L - h_{LF}, m + 1 - m) \\
&= \sum_{l=1}^{h_{LF} m} \beta_{\lceil l/m \rceil} \delta_{\lceil l/m \rceil m + 1 - l} x_H(\tau_L - 1, m + 1 - l),
\end{aligned} \tag{3.7}$$

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 (3.7) 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 (3.6) 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 Power Analysis of High-to-Low Causality Tests

The results of Section 2 characterize the asymptotic global power properties high-to-low frequency non-causality tests. 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

$$H_1^L : \boldsymbol{b} = (1/\sqrt{T_L}) \times \boldsymbol{\nu},$$

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 of Mixed Frequency Tests

4.1.1 Mixed Frequency Max Test

Our first result gives the asymptotic distribution of the max statistic $\hat{\mathcal{T}}$ in (2.7) 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 by (2.8) and (2.9), and

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

In order to compute local power, we need the mean vector $\boldsymbol{\mu}$ and variance matrix $\boldsymbol{V}$, and therefore explicit characterizations of $\boldsymbol{\Gamma}_{j,i}$, $\tilde{\boldsymbol{\Gamma}}_{j,i}$ and $\boldsymbol{C}_j$ in terms of underlying parameters $(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_p)$ and $\boldsymbol{\Omega}$ in (2.1). Recall $\tilde{\boldsymbol{\Gamma}}_{j,i}$ is used to compute $\boldsymbol{V}$: see (2.8) and (2.9). Define

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

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. Write $\boldsymbol{\Upsilon}_k \equiv E[\boldsymbol{X}(\tau_L)\boldsymbol{X}(\tau_L - k)']$ with (s,t) element $\boldsymbol{\Upsilon}_k(s,t)$, and $\tilde{\boldsymbol{\Upsilon}}_k \equiv E[\epsilon_L^2(\tau_L)\boldsymbol{X}(\tau_L)\boldsymbol{X}(\tau_L - k)']$ with (s,t) element $\tilde{\boldsymbol{\Upsilon}}_k(s,t)$.

(a) For $j, i \in \{1, \ldots, h\}$:

$$\underbrace{\mathbf{\Gamma}_{j,i}}_{(q+1)\times(q+1)} = \begin{bmatrix} \Upsilon_{1-1}(K,K) & \ldots & \Upsilon_{1-q}(K,K) & \Upsilon_{-f(i)}(g(i),K) \\ \vdots & \ddots & \vdots & \vdots \\ \Upsilon_{q-1}(K,K) & \ldots & \Upsilon_{q-q}(K,K) & \Upsilon_{(q-1)-f(i)}(g(i),K) \\ \Upsilon_{f(j)}(K,g(j)) & \ldots & \Upsilon_{f(j)-(q-1)}(K,g(j)) & \Upsilon_{f(j)-f(i)}(g(i),g(j)) \end{bmatrix} \tag{4.3}$$

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

$\tilde{\mathbf{\Gamma}}_{j,i}$ relates to $\tilde{\mathbf{\Upsilon}}_k$ in the same manner that $\mathbf{\Gamma}_{j,i}$ relates to $\mathbf{\Upsilon}_k$.

(b) Let $\mathbf{\Upsilon}$ be a $pK \times pK$ matrix whose (i,j) block is $\mathbf{\Upsilon}_{j-i}$ for $i, j \in \{1, \ldots, p\}$. Let $\bar{\mathbf{\Omega}}$ be a $pK \times pK$ matrix whose $(1,1)$ block is $\mathbf{\Omega}$ and all other blocks are $\mathbf{0}_{K\times K}$. Then $\mathbf{\Upsilon}_k$ satisfies:

$$\begin{aligned} &\text{vec}[\mathbf{\Upsilon}_k] = \left(\boldsymbol{I}_{(pK)^2} - \boldsymbol{A} \otimes \boldsymbol{A}\right)^{-1} \text{vec}[\bar{\mathbf{\Omega}}] \text{ for } k \in \{1-p, \ldots, p-1\} \\ &\mathbf{\Upsilon}_k = \sum_{l=1}^{p} \boldsymbol{A}_l \mathbf{\Upsilon}_{k-l} \text{ for } k \geq p \;\text{ and }\; \mathbf{\Upsilon}_k = \mathbf{\Upsilon}'_{-k} \text{ for } k \leq -p \end{aligned} \tag{4.4}$$

Remark 4.1. The matrices in (4.3) are always well-defined. Simply note that $g(j)$ takes a natural number between 1 and $m+1$ for any $j \in \mathbb{N}$ because $\mathbf{\Upsilon}_k$ is $(m+1)\times(m+1)$, in particular $g(j)$ takes m, $m-1$, $\ldots$, 1 as j runs from $(k-1)m+1$ to km for any $k \in \mathbb{N}$.[12] Further, $f(j)$ 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.

Remark 4.2. An explicit iterative formula for $\tilde{\mathbf{\Upsilon}}_k$ similar to (4.4) does not exist, unless second order properties of $\boldsymbol{\epsilon}(\tau_L)$ like a GARCH structure are given. If $\boldsymbol{\epsilon}(\tau_L)\boldsymbol{\epsilon}(\tau_L)' - \mathbf{\Omega}$ is an adapted martingale difference then, in view of Remark 2.5 and equation (2.10), we only need to characterize $\mathbf{\Upsilon}_k$. We maintain this case for the remainder of this section in order to obtain local power results without a

[12] See Table T.1 in Ghysels, Hill, and Motegi (2015).

second order model for the errors. Therefore, assume:

$$E\left[\boldsymbol{\epsilon}(\tau_L)\boldsymbol{\epsilon}(\tau_L)'\middle|\,\mathcal{F}_{\tau_L-1}\right] = \boldsymbol{\Omega} \text{ \textit{a.s.} for each } \tau_L. \tag{4.5}$$

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

Step 1 Calculate $\boldsymbol{\mu}$ and $\boldsymbol{V}$ from underlying parameters $(\boldsymbol{A}_1, \ldots, \boldsymbol{A}_p)$, 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 $N(\boldsymbol{0}_{h\times 1}, \boldsymbol{V})$, 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:

$$\begin{aligned} x_L(\tau_L) &= \sum_{k=1}^{q} x_L(\tau_L - k)\alpha_k + \sum_{j=1}^{h} x_H(\tau_L - 1, m+1-j)\beta_j + u_L(\tau_L) \\ &= \begin{bmatrix} \boldsymbol{X}_L^{(q)}(\tau_L-1)' & \boldsymbol{X}_H^{(h)}(\tau_L-1)' \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\beta} \end{bmatrix} + u_L(\tau_L) = \boldsymbol{x}(\tau_L-1)'\boldsymbol{\Theta} + u_L(\tau_L). \end{aligned} \tag{4.6}$$

$\boldsymbol{X}_H^{(h)}(\tau_L - 1)$ 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 Local Power of Low Frequency Tests

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 Max Test

We compute the low frequency max test statistic $\hat{\mathcal{T}}^{(LF)}$ based on h low frequency parsimonious regression models (2.17): $x_L(\tau_L) = \underline{\boldsymbol{x}}_j(\tau_L - 1)'\boldsymbol{\theta}_j^{(LF)} + u_{L,j}(\tau_L)$, 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 Wald Test

The low frequency naïve regression model is (2.14): $x_L(\tau_L) = \underline{\boldsymbol{x}}(\tau_L - 1)'\boldsymbol{\theta}^{(LF)} + u_L(\tau_L)$, 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. yearly LF increment with monthly HF increment, or quarterly LF increment with weekly HF increment approximately):

$$\underbrace{\begin{bmatrix} 1 & 0 & \ldots & \ldots & \ldots & 0 \\ -d & 1 & \ddots & \ddots & \ddots & 0 \\ 0 & -d & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & -d & 1 & 0 \\ 0 & 0 & \ldots & 0 & 0 & 1 \end{bmatrix}}_{\equiv \boldsymbol{N}} \underbrace{\begin{bmatrix} x_H(\tau_L, 1) \\ \vdots \\ x_H(\tau_L, 12) \\ x_L(\tau_L) \end{bmatrix}}_{= \boldsymbol{X}(\tau_L)} = \underbrace{\begin{bmatrix} 0 & 0 & \ldots & d & c_1 \\ 0 & 0 & \ldots & 0 & c_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & c_{12} \\ b_{12} & b_{11} & \ldots & b_1 & a \end{bmatrix}}_{\equiv \boldsymbol{M}} \underbrace{\begin{bmatrix} x_H(\tau_L - 1, 1) \\ \vdots \\ x_H(\tau_L - 1, 12) \\ x_L(\tau_L - 1) \end{bmatrix}}_{= \boldsymbol{X}(\tau_L - 1)} + \underbrace{\begin{bmatrix} \eta_H(\tau_L, 1) \\ \vdots \\ \eta_H(\tau_L, 12) \\ \eta_L(\tau_L) \end{bmatrix}}_{\equiv \boldsymbol{\eta}(\tau_L)}, \tag{4.7}$$

where $\boldsymbol{\eta}(\tau_L) \sim (\boldsymbol{0}_{13\times 1}, \boldsymbol{I}_{13})$, and $\boldsymbol{\eta}(\tau_L)$ and $\boldsymbol{\eta}(\tau_L)\boldsymbol{\eta}(\tau_L)' - \boldsymbol{I}_{13}$ are martingale differences 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, $\boldsymbol{c} = [c_1, \ldots, c_{12}]'$ represents Granger causality from x_L to x_H, and our interest lies in $\boldsymbol{b} = [b_1, \ldots, b_{12}]'$ since it expresses Granger causality from x_H to x_L. Since

$$\boldsymbol{N}^{-1} = \begin{bmatrix} 1 & 0 & \ldots & \ldots & \ldots & 0 \\ d & 1 & \ddots & \ddots & \ddots & 0 \\ d^2 & d & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ d^{11} & d^{10} & \ldots & d & 1 & 0 \\ 0 & 0 & \ldots & 0 & 0 & 1 \end{bmatrix} \text{ thus } \boldsymbol{A} \equiv \boldsymbol{N}^{-1}\boldsymbol{M} = \begin{bmatrix} 0 & 0 & \ldots & d & \sum_{i=1}^{1} d^{1-i}c_i \\ 0 & 0 & \ldots & d^2 & \sum_{i=1}^{2} d^{2-i}c_i \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & d^{12} & \sum_{i=1}^{12} d^{12-i}c_i \\ b_{12} & b_{11} & \ldots & b_1 & a \end{bmatrix}, \tag{4.8}$$

the reduced form of (4.7) 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 four types of drift $\boldsymbol{\nu} = [\nu_1, \ldots, \nu_{12}]'$. First, *decaying causality* with hyperbolic decay and alternating signs: $\nu_j = (-1)^{j-1} \times 2.5/j$ for $j = 1, \ldots, 12$. Second, *lagged causality* with $\nu_j = 2 \times I(j = 12)$, 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. Such a relationship may exist in macroeconomic processes due to lagged information transmission, seasonality, feedback effects, and ambiguous theoretical relations in terms of signs. Fourth, *uniform causality* with $\nu_j = 0.3$ for all j.

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: $c_j = (-1)^{j-1} \times 0.4/j$ for $j = 1, \ldots, 12$.

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 by drawing $100,000$ 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 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 provide now a discussion for each causal pattern.

Decaying Causality, Low Persistence in x_H: $d = 0.2$ The MF max and Wald tests have moderately high power between .350 and .570. The LF tests with flow sampling have 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. Under stock sampling, however, local power is much larger, ranging from .473 to .648. The reason for the improved performance is that 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 higher 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 .075, .177, and .770 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}. The power of the LF max test, for instance, is .453, .466, and .413 when h_{LF} is 1, 2, and 3, respectively. A reason for this good performance is that the causal effect is unambiguously positive. The lagged causality is preserved under the flow aggregation since we have one large positive coefficient $\nu_{12} = 2$ and no negative coefficients.

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 $h_{LF} = 2$ (.675 for max test and .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. Focusing on $d = 0.2$, the MF max test has power .390, .318, and .678 when h_{MF} is 4, 8, and 12, and MF Wald test power is .365, .291, and .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 .070) due to their vulnerability to alternating signs and lagged causality as seen above.

Uniform Causality Focusing on $d = 0.8$, flow sampling tests have highest power (ranging from .956 to .988), mixed frequency tests are second (ranging from .753 to .956), and stock sampling tests have lowest power (ranging from .633 to .736). Since the uniform causality involves a uniquely positive sign, flow sampling tests achieve the highest power.

MF Max Test versus MF Wald Test It is not the case that the MF max test is always preferred to the MF Wald test. Max test power surpasses Wald-test power in 15 cases out of 24, but the max test yields lower power in the other 9 cases. An important result, however, is that the max test is clearly more powerful than the Wald test under lagged causality. When $d = 0.8$ and $h_{MF} = 12$, the max test power is .770 while the Wald test power is .559. This result suggests that the max test is better capable of detecting a lagged impact from x_H to x_L due to its robustness against large parameter dimensions. This feature is confirmed by our Monte Carlo simulations below.

5 Monte Carlo Simulations

We conduct simulation experiments in order to compare the MF and LF max and Wald tests. High-to-low causality is tackled in Section 5.1 and then low-to-high causality in Section 5.2.

5.1 High-to-Low Granger Causality

5.1.1 MF-VAR(1)

The benchmark process is MF-VAR(1). We then consider MF-VAR(2) as a robustness check.

Data Generating Processes We begin with the MF-VAR(1) process (4.7) with $m = 12$. The structural error $\boldsymbol{\eta}(\tau_L)$ is either i.i.d. or a GARCH process. Let $\boldsymbol{\xi}(\tau_L) \overset{i.i.d.}{\sim} N(\mathbf{0}_{13\times1}, \boldsymbol{I}_{13})$. In the i.i.d. case $\boldsymbol{\eta}(\tau_L) = \boldsymbol{\xi}(\tau_L)$. In the GARCH case $\boldsymbol{\eta}(\tau_L) = \boldsymbol{H}(\tau_L)^{1/2}\boldsymbol{\xi}(\tau_L)$ where the conditional covariance matrix $\boldsymbol{H}(\tau_L)$ is a BEKK process (cfr. Engle and Kroner (1995)):

$$\boldsymbol{H}(\tau_L) \equiv \boldsymbol{H}(\tau_L)^{1/2}\boldsymbol{H}(\tau_L)^{1/2'} = \boldsymbol{C}_s\boldsymbol{C}_s' + \boldsymbol{A}_s\boldsymbol{\eta}(\tau_L-1)\boldsymbol{\eta}(\tau_L-1)'\boldsymbol{A}_s' + \boldsymbol{B}_s\boldsymbol{H}(\tau_L-1)\boldsymbol{B}_s'.$$

The reduced-form error $\boldsymbol{\epsilon}(\tau_L) = \boldsymbol{N}^{-1}\boldsymbol{\eta}(\tau_L)$ is conditionally $N(\mathbf{0}_{13\times1}, \boldsymbol{\Omega}(\tau_L))$ distributed, where

$$\boldsymbol{\Omega}(\tau_L) = \boldsymbol{C}\boldsymbol{C}' + \boldsymbol{A}\boldsymbol{\epsilon}(\tau_L-1)\boldsymbol{\epsilon}(\tau_L-1)'\boldsymbol{A}' + \boldsymbol{B}\boldsymbol{\Omega}(\tau_L-1)\boldsymbol{B}'$$

with $\boldsymbol{C} = \boldsymbol{N}^{-1}\boldsymbol{C}_s$, $\boldsymbol{A} = \boldsymbol{N}^{-1}\boldsymbol{A}_s\boldsymbol{N}$, and $\boldsymbol{B} = \boldsymbol{N}^{-1}\boldsymbol{B}_s\boldsymbol{N}$. For simplicity we impose a diagonal structure $\boldsymbol{C}_s = \sqrt{0.1} \times \boldsymbol{N}$, $\boldsymbol{A}_s = \sqrt{0.2} \times \boldsymbol{I}_{13}$, and $\boldsymbol{B}_s = \sqrt{0.4} \times \boldsymbol{I}_{13}$, hence the reduced-form parameters boil down to $\boldsymbol{C} = \sqrt{0.1} \times \boldsymbol{I}_{13}$, $\boldsymbol{A} = \sqrt{0.2} \times \boldsymbol{I}_{13}$, and $\boldsymbol{B} = \sqrt{0.4} \times \boldsymbol{I}_{13}$ so that

$$\boldsymbol{\Omega}(\tau_L) = 0.1 \times \boldsymbol{I}_{13} + 0.2 \times \boldsymbol{\epsilon}(\tau_L - 1)\boldsymbol{\epsilon}(\tau_L - 1)' + 0.4 \times \boldsymbol{\Omega}(\tau_L - 1).$$

We consider non-causality $\boldsymbol{b} = \boldsymbol{0}_{12\times 1}$ and four causal patterns as in the local power analysis. The first causal pattern is *decaying causality* with alternating signs: $b_j = (-1)^{j-1} \times 0.3/j$ for $j = 1, \ldots, 12$. The second is *lagged causality*: $b_j = 0.3 \times I(j = 12)$ for all j. The third is *sporadic causality*: $(b_3, b_7, b_{10}) = (0.2, 0.05, -0.3)$ and all other $b_j = 0$. The fourth is *uniform causality*: $b_j = 0.02$ for all j.

We again assume a weak autoregressive property for x_L (i.e. $a = 0.2$). The choice of a does not appear to significantly influence rejection frequencies.[13] 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 $T_L \in \{80, 160\}$. Since $m = 12$, our experimental design can approximately be thought as week versus quarter, matching our empirical applications in Section 6. Hence $T_L = 80$ or 160 implies that the low frequency sample size is 20 or 40 years.

Model Estimation All regression models include two low frequency lags of x_L (i.e. $q = 2$). Setting $q = 1$ would be sufficient when 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 $h_{MF} \in \{4, 8, 12, 24\}$. The number of low frequency lags of aggregated x_H used in the LF tests is $h_{LF} \in \{1, 2, 3, 4\}$. 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 5,000 for p-value computation.

Given the large ratio $m = 12$, the MF (and possibly even LF) Wald test may suffer from size distortions if we use the asymptotic chi-square distribution. We therefore use Gonçalves and Kilian's (2004) recursive design parametric wild bootstrap which allows for conditionally heteroscedastic errors

[13] Simulation results with $a = 0.8$ are not reported to conserve space, but available upon request.

of unknown form. Their bootstrap p-value is computed with 1,000 bootstrap samples.[14]

The key covariance matrix $\boldsymbol{V}$ for the max test is computed with the heteroscedasticity-robust estimator (2.11). In Tables T.6-T.9 of the supplemental material Ghysels, Hill, and Motegi (2015), we also use an estimate of the non-robust (2.10). Bootstrapped Wald tests exhibit size distortions when (2.11) is used, hence we only use the non-robust covariance. Evidently the distortions arise from the added sampling error in this more complex robust estimator. We keep the number of bootstrap samples at just 1,000 to control for the rather large computation time.[15]

The number of Monte Carlo samples is 5,000 for max tests, and 1,000 for bootstrapped Wald tests due to the substantial computation time.

Results Table 2 compiles simulation results for the case of GARCH errors. (Results for the i.i.d. error are collected in Table T.2 of Ghysels, Hill, and Motegi (2015) in order to save space. The two errors yield similar results in general.) Nominal size is fixed at 0.05. Empirical size in both tests is fairly sharp, ranging from .031 to .063. The max tests have sharp size evidently due to its relatively more parsimonious specification, while the Wald test has sharp size due to bootstrapping the p-value (with the simpler non-robust covariance matrix). Regarding power, as in the local power study MF tests are better than LF tests in terms of detecting complicated causal patterns like sporadic causality.

Consider the relative power performance of the MF max and Wald tests. In most cases across causal patterns $\boldsymbol{b}$, lag length h_{MF}, persistence d, and sample size T_L, max and Wald tests have similar power. An advantage of the max test is highlighted under lagged causality with $d = 0.2$. When $T_L = 160$, the max test power is .763 for $h_{MF} = 12$ and .685 for $h_{MF} = 24$. The Wald test power is, by comparison, .610 for $h_{MF} = 12$ and .434 for $h_{MF} = 24$. By switching from 12 lags to 24 lags, the max test loses power by only $.763 - .685 = .078$ while the Wald test loses power by $.610 - .434 = .176$. This result suggests that the max test is more robust against large parameter dimensions than the Wald test. The max test therefore better captures a lagged impact from x_H to x_L.

The max test focuses on the largest squared parameter estimate, and therefore discards the other identical parameter values in the case of uniform causality. It seems *prima facia* that that should disad-

[14] Consider bootstrapping in the MF case with model (4.6), the LF case being similar. Suppose that $\hat{\boldsymbol{\Theta}}$ is 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 $\hat{u}_L(\tau_L)$, and the Wald statistic is $\hat{W}$. Simulate $N = 1,000$ bootstrap samples from $x_L(\tau_L) = \boldsymbol{x}(\tau_L - 1)'\tilde{\boldsymbol{\Theta}}_0 + \hat{u}_L(\tau_L)v(\tau_L)$, where $\tilde{\boldsymbol{\Theta}}_0$ is $\hat{\boldsymbol{\Theta}}$ with the null hypothesis of non-causality imposed, and $v(\tau_L) \overset{i.i.d.}{\sim} N(0,1)$. For each sample $i = 1, \ldots, N$, compute the Wald statistic $\tilde{W}_i$. The bootstrapped p-value is $p_N = (1/N) \times \sum_{i=1}^{N} I(\tilde{W}_i \geq \hat{W})$.

[15] Results for the bootstrapped Wald test with robust covariance matrix (2.11) are available upon request.

vantage the max test relative to the Wald test, because the latter reduces to a squared linear combination of *all* positive parameter estimates. When the error is conditionally heteroskedastic, however, both MF and LF tests yield essentially trivial power, while under i.i.d. errors the two tests are comparable with strong power, and neither dominates across cases. The Wald test does not dominate precisely because it uses all parameter estimates from model (2.2), hence greater dispersion exists in the parameter estimates and therefore the Wald statistic.

5.1.2 MF-VAR(2)

As a further analysis, we use Monte Carlo samples 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 (4.7), the extra coefficient matrix $\boldsymbol{M}_2$ is parameterized as

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

Non-causality is now expressed as $\boldsymbol{b} = \mathbf{0}_{24\times 1}$; *decaying causality* is $b_j = (-1)^{j-1} \times 0.3/j$ for $j = 1, \ldots, 24$; *lagged causality* is $b_j = 0.3 \times I(j = 24)$ for all j; *sporadic causality* is $(b_5, b_{12}, b_{17}, b_{19}) = (-0.2, 0.1, 0.2, -0.35)$ and all other $b_j = 0$; and *uniform causality* is $b_j = 0.02$ for all j.

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 $j = 1, \ldots, 12$; $q = 2$; $\boldsymbol{W}_h = (1/h) \times \boldsymbol{I}_h$; $T_L \in \{80, 160\}$; and nominal size is 0.05. The number of high frequency lags of x_H used in the MF tests is $h_{MF} \in \{16, 20, 24\}$, while the number of low frequency lags of aggregated x_H used in LF tests is $h_{LF} \in \{1, 2, 3\}$.

Rejection frequencies with GARCH errors are compiled in Table 3. (The i.i.d. error case is covered in Table T.4 of Ghysels, Hill, and Motegi (2015) in order to save space.) There are virtually no size distortions for both MF max and bootstrapped MF Wald tests. In most cases the MF max test and the MF Wald test exhibit similar empirical power. The former surpasses the latter under lagged causality with $d = 0.2$ and $h_{MF} = 24$. When $T_L = 160$, the max test power is .682 whereas the Wald test power is .530. These results are consistent with the MF-VAR(1) scenario.

5.2 Low-to-High Granger Causality

We now focus on low-to-high causality $\boldsymbol{c} = [c_1, \ldots, c_{12}]'$ in the structural MF-VAR(1) in (4.7) with $m = 12$.[16]

5.2.1 Design

We consider non-causality and the usual four causal 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 (4.8), 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* $c_j = (-1)^{j-1} \times 0.3/j$ for $j = 1, \ldots, 12$, assuming $d = 0.2$. For *lagged causality* $c_j = 0.25 \times I(j = 12)$ for all j, the upper-right block of $\boldsymbol{A}_1$ is identically $\boldsymbol{c}$ regardless of d. In the case of *sporadic causality* $(c_3, c_7, c_{10}) = (0.3, 0.15, -0.3)$ a similar pattern arises, assuming $d = 0.2$. Uniform causality $c_j = 0.07$ for all j is also preserved (though not perfectly). 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 $j = 1, \ldots, 12$. The sample size is again $T_L \in \{80, 160\}$. We implement the MF max test based on MF parsimonious regression model (3.2), and the bootstrapped Wald test based on the MF naïve regression model (3.1). For both tests lags and leads are taken from $h_{MF}, r_{MF} \in \{4, 8, 12, 24\}$.

The max test exhibits size distortions when $h_{MF} = 24$, especially in the smaller sample $T_L = 80$. This is a natural consequence of parameter proliferation. As a second max test we therefore use the MF parsimonious regression models with a MIDAS polynomial on the high frequency lags, as in (3.5), as an ad hoc attempt to tackle 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 a Wald test based on the MF naïve regression model (3.1) with lags of x_H replaced with the Almon polynomial.

[16] MF-VAR(2) cases are covered in Tables T.12, T.13, T.16, and T.17 of the supplemental material Ghysels, Hill, and Motegi (2015). See Tables T.12 and T.13 for bootstrapped Wald tests and max test with the robust covariance matrix. See Tables T.16 and T.17 for max test with the non-robust covariance matrix.

We also perform the LF max test based on LF parsimonious regression models, for $j = 1, \ldots, r_{LF}$:

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

The lead lengths r_{LF} and h_{LF} are taken from $\{1, 2, 3, 4\}$. We do not exploit a MIDAS polynomial here since the lag length h_{LF} is sufficiently small to avoid size distortions. 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:

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

with $h_{LF}, r_{LF} \in \{1, 2, 3, 4\}$. As before, a MIDAS polynomial is not required.

We use flat weights for the max test and 5,000 draws from the asymptotic distribution to compute p-values. The robust covariance matrix is used for max tests. (See Tables T.14 and T.15 of Ghysels, Hill, and Motegi (2015) for results with the non-robust covariance matrix.) For Wald tests, we generate 1,000 bootstrap samples. The number of Monte Carlo samples is 5,000 for max tests and 1,000 for the Wald tests, and nominal size is 5%.

5.2.2 Results

Table 4 presents rejection frequencies for the GARCH error case. The i.i.d. case is covered in Table T.10 of Ghysels, Hill, and Motegi (2015): the two errors yield similar results in general. First, the MF max test without MIDAS polynomials exhibits size distortions when $T_L = 80$. Empirical size is .167 for $h_{MF} = r_{MF} = 24$, while it is .055 for $h_{MF} = r_{MF} = 4$ (Panel A.1.1). This is a logical result since fewer parameters align with sharper empirical size. When $T_L = 160$ empirical size is much sharper, although it is still .082 when $h_{MF} = r_{MF} = 24$ (Panel A.2.1). Second, the max test with MIDAS polynomials exhibits nearly perfect empirical size, despite the inherent mis-specification of the estimated model. This follows because, within our design, the quasi-true parameters $\beta_{k,j}$ associated with the lagged x_H in (3.2) are well approximated by the Almon coefficients $\omega_k(\boldsymbol{\pi})$ in (3.5). The bootstrapped Wald test is correctly sized with or without MIDAS polynomials (Panels A.1.2 and A.2.2).

MF versus LF Tests We now compare the empirical power of MF tests versus LF tests. Under decaying causality (Panel B), we see a clear advantage of MF tests compared to LF tests. When $T_L = 160$, the MF tests with the MIDAS polynomial have empirical power of at least .590 (and much higher in many cases), whereas the LF test power is at most .159. In order to understand why the LF tests suffer from such low power, consider stock sampling first. As seen in (5.1) and (5.2), lead terms used in those tests are $x_H(\tau_L + 1, 12), x_H(\tau_L + 2, 12), \ldots, x_H(\tau_L + r_{LF}, 12)$, all of which have small coefficients under decaying causality. The stock sampling test, in other words, is missing the most important lead term $x_H(\tau_L + 1, 1)$ and thus suffer from a poor signal relative to noise. Under flow sampling, averaging $x_H(\tau_L + 1, 1)$ through $x_H(\tau_L + 1, 12)$ results in an offset of positive and negative impacts, hence again there is a poor causation signal. This has been well documented in the literature: temporal aggregation can obfuscate true underlying causality.

Next, consider lagged causality (Panel C). The MF tests have little power when $r_{MF} < 12$ because the only relevant term is $x_H(\tau_L + 1, 12)$ by construction. When $r_{MF} = 12$ then power improves sharply to about .2 for $T_L = 80$ and .5 for $T_L = 160$. LF tests with stock sampling, by contrast, obtain much higher power than the MF tests for any $h_{LF}, r_{LF} \in \{1, 2, 3, 4\}$ (Panels C.1.4 and C.2.4). This occurs because the LF models with stock sampling contain the relevant lead term $x_H(\tau_L + 1, 12)$, and require fewer estimated parameters.

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 $c_{10} = -0.3$. When $T_L = 160$ and $r_{MF} = 12$, MF tests have power above .9 (Panel D.2.2). LF tests, by contrast, have negligible power in all cases. The low frequency leads and lags of x_H are too coarse to capture the complicated causal pattern with unevenly-spaced lags, alternating signs, and non-decaying structure.

Under uniform causality (Panel E), power is greatest when flow sampling is used. This result is reasonable since the uniform causal pattern is preserved under flow sampling.

MF Max versus MF Wald Tests The max test has higher power than the Wald test under lagged causality. When $r_{MF} = 24$ and $T_L = 160$, the max test power is .438 on average while the Wald test power is .332 on average (Panel C.2.2). We can therefore conclude that the max test is better capable of detecting lagged impact from x_L to x_H due to its robustness against large parameter dimensions. The two tests are similar for the remaining causal patterns.

6 Empirical Application

As an empirical illustration, we study Granger causality between a weekly term spread (long and short term interest rate spread) and quarterly real GDP growth in the U.S. We analyze both high-to-low causality (spread to GDP) and low-to-high causality (GDP to spread), although we are particularly interested in the former. A decline in the interest rate spread has historically been regarded as a strong predictor of a recession, but recent events place doubt on its use for such prediction.[17] 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. four-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.[18]

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.

[17] See Stock and Watson (2003) for a survey of the historical relationship between term spread and business cycle.

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

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 allows us to include more leads and lags in our models. Furthermore, our simulation experiments in Section 5 reveal our tests work well for $T_L = 80$.

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, for $j = 1, \ldots, 24$:

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

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

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

This has $q = 2$ quarters of lagged x_L) and $h_{LF} = 3$ quarters of lagged x_H^*. 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:

$$x_L(\tau_L) = \alpha_0 + \sum_{k=1}^{2} \alpha_k x_L(\tau_L - k) + \sum_{j=1}^{3} \beta_j x_H^*(\tau_L - j) + u_L(\tau_L). \tag{6.3}$$

Wald statistic p-values are computed based on the non-robust covariance matrix and Gonçalves and Kilian's (2004) bootstrap, with $N = 1,000$ replications. Max statistic p-values are computed based on the robust covariance matrix with 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 (6.3) 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 $\{M_1, M_2\} = \{999, 249\}$ for the first and second stage. We perform Q tests with 4, 8, or 12 lags for each window and model.

When the Q-test bootstrap block size is $b = 4$, 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. The MF model again produces fewer rejections than the LF model under larger block sizes $b \in \{10, 20\}$. (See Table T.18 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 1979:I-1998:IV, the LF max test detects significant causality prior to 1975:III-1995: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 only *after* subsample 1990:IV-2010:III, which is somewhat counter-intuitive. This result may stem from parameter proliferation. As seen from (6.1)-(6.3), the MF naïve regression model has many more parameters than any other model. 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 10,000.

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 the 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 .037, hence we reject non-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 .048 for the max test and .085 for the Wald 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:

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

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

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

in each case $j = 1, \ldots, 24$. We include $q = 2$ quarters of lagged x_L, $h_{MF} = 24$ weeks of lagged x_H^*, and $r_{MF} = 24$ weeks of led x_H^*.

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 interest 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: $x_L(\tau_L) = \alpha_0 + \sum_{k=1}^{2} \alpha_k x_L(\tau_L - k) + \sum_{k=1}^{3} \beta_k x_H^*(\tau_L - k) + \sum_{j=1}^{3} \gamma_j x_H^*(\tau_L + j) + u_L(\tau_L)$.

Wald test p-values are bootstrapped with $N = 1,000$ bootstrap samples, and the max test p-values are computed using 100,000 draws from limit distributions under non-causality. The non-robust covariance matrix is used for the bootstrapped Wald test, while the robust covariance matrix is used for the max test. 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.[19]

[19] When the block size is $b = 4$, the MF model without a MIDAS polynomial rejects the null hypothesis of uncorrelated

Figure 4 plots p-values for the causality tests over the 129 subsamples. While MF tests 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 middle 1983:IV-2005:II (cfr. Panel (e)). The LF Wald test shows significant causality in approximately the last 20% 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.[20]

The MF max and Wald tests without a MIDAS polynomial have p-values .041 and .265, respectively, and with a MIDAS polynomial the p-values are .160 and .686. The LF max and Wald test have p-values .135 and .215. Thus, only the MF max test with MIDAS points to causality. Overall, we do not observe strong evidence for low-to-high causality in general. This result is consistent with the rolling window analysis above.

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 $j \in \{1, \dots, h\}$. The resulting max test statistic then operates on the largest j^{th} lag or lead estimated parameter. This method extends to any regression setting where many parameters have the value zero under the null hypothesis, and should therefore be of general interest

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.18 of Ghysels, Hill, and Motegi (2015) for complete results.

[20] 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 and therefore weaker evidence of residual correlatedness in general.

when many possibly irrelevant regressors are available.

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 easily generalize the test for an increasing number of parameters, and would therefore apply to, for example, nonparametric regression 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 (see e.g. Xiao and Wu (2014) and Hill and Motegi (2016)). 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.

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 equal power in most cases, but the former is more powerful under causality with a large time lag.

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 1990s, 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.

ANDERSEN, N. T., AND V. DOBRIC (1987): "The Central Limit Theorem for Stochastic Processes," *Annals of Probability*, 15, 164–177.

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

BERMAN, S. M. (1964): "Limit Theorems for the Maximum Term in Stationary Sequences," *Annals of Mathematical Statistics*, 35, 502–516.

BILLINGSLEY, P. (1961): "The Lindeberg-Lvy Theorem for Martingales," *Proceedings of the American Mathematical Society*, 12, 788–792.

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

CHERNOFF, H., AND S. ZACKS (1964): "Estimating the Current Mean of a Normal Distribution which Is Subject to Changes in Time," *Annals of Mathematical Statistics*, 35, 999–1028.

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

DAVIES, R. (1977): "Hypothesis Testing when a Nuisance Parameter is Present Only under the Alternative," *Biometrika*, 64, 247–254.

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

DRAPER, N. R., H. SMITH, AND E. POWNELL (1966): *Applied Regression Analysis*. Wiley New York.

DUDLEY, R. M. (1978): "Central Limit Theorems for Empirical Measures," *Annals of Probability*, 6, 899–929.

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

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

ENGLE, R. F., AND K. F. KRONER (1995): "Multivariate Simultaneous Generalized ARCH," *Econometric Theory*, 11, 122–150.

FORONI, C., E. GHYSELS, AND M. MARCELLINO (2013): "Mixed-Frequency Vector Autoregressive Models," in *VAR Models in Macroeconomics – New Developments and Applications: Essays in Honor of Christopher A. Sims*, ed. by T. B. Fomby, L. Kilian, and A. Murphy, vol. 32, pp. 247–272. Emerald Group Publishing Limited.

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. (2016): "Macroeconomics and the Reality of Mixed Frequency Data," *Journal of Econometrics*, 193, 294–314.

GHYSELS, E., J. B. HILL, AND K. MOTEGI (2015): "Supplemental Material for *Simple Granger Causality Tests for Mixed Frequency Data*," Dept. of Economics, University of North Carolina.

——— (2016): "Testing for Granger Causality with Mixed Frequency Data," *Journal of Econometrics*, 192, 207–230.

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.

GNEDENKO, B. V. (1943): "Sur la Distribution Limite du Terms Maximum d'une Série Aléatoire," *Annals of Mathematics*, 44, 423–453.

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

GRANGER, C. W. J. (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.

GRANGER, C. W. J., AND J. L. LIN (1995): "Causality in the Long Run," *Econometric Theory*, 11, 530–536.

HANNAN, E. J. (1974): "The Uniform Convergence of Autocovariances," *Annals of Statistics*, 2, 803–806.

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

HILL, J. B. (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.

HILL, J. B., AND K. MOTEGI (2016): "A Max-Correlation White Noise Test for Weakly Dependent Time Series," Working Paper, Department of Economics, University of North Carolina at Chapel Hill.

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.

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. Zimmermann, pp. 75–91. Physica-Verlag Heidelberg.

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

MCCRORIE, J. R., AND M. J. 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.

STOCK, J. H., AND M. W. WATSON (2003): “Forecasting Output and Inflation: The Role of Asset Prices,” *Journal of Economic Literature*, 41, 788–829.

XIAO, H., AND W. B. WU (2014): “Portmanteau Test and Simultaneous Inference for Serial Covariances,” *Statistica Sinica*, 24, 577–600.

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 Proofs of Main Results

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

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 $\{\sqrt{T_L} w_{T_L,j} \hat{\beta}_j\}_{j=1}^h$ and stochastic equicontinuity (e.g. Dudley (1978), Andersen and Dobric (1987)). By discreteness of j note $\forall(\epsilon, \eta) > 0 \, \exists \delta \in (0,1)$ such that $\sup_{0 \leq j \leq h : |j - \tilde{j}| \leq \delta} |\sqrt{T_L} w_{T_L,j} \hat{\beta}_j - \sqrt{T_L} w_{T_L,\tilde{j}} \hat{\beta}_{\tilde{j}}| = 0 \; a.s.$ Therefore $\lim_{n \to \infty} P(\sup_{0 \leq j \leq h : |j - \tilde{j}| \leq \delta} |\sqrt{T_L} w_{T_L,j} \hat{\beta}_j - \sqrt{T_L} w_{T_L,\tilde{j}} \hat{\beta}_{\tilde{j}}| > \eta) \leq \varepsilon$ for some $\delta > 0$, hence $\{\sqrt{T_L} w_{T_L,j} \hat{\beta}_j\}_{j=1}^h$ is stochastically equicontinuous.

It remains to show the finite dimensional distributions for any h are $\sqrt{T_L} \boldsymbol{W}_{T_L,h} \hat{\boldsymbol{\beta}} \xrightarrow{d} N(\mathbf{0}_{h \times 1}, \boldsymbol{V})$ where $\boldsymbol{V} \equiv \boldsymbol{W}_h \boldsymbol{R} \boldsymbol{\mathcal{S}} \boldsymbol{R}' \boldsymbol{W}_h$, and $\boldsymbol{\mathcal{S}}$ is defined in (2.9). 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 $N(\mathbf{0}_{h \times 1}, \boldsymbol{V})$, then follows instantly from the continuous mapping theorem since $\max\{\cdot\}$ is a continuous function.

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 $H_0 : \boldsymbol{b} = \mathbf{0}_{pm \times 1}$. Under H_0 the parsimonious model parameters $\boldsymbol{\beta} \equiv [\beta_1, \ldots, \beta_h]' = \mathbf{0}_{h \times 1}$, 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 and Assumption 2.4, the i^{th} parsimonious model (2.5) becomes $x_L(\tau_L) = \boldsymbol{x}_j(\tau_L - 1)'\boldsymbol{\theta}_{0,j} + \epsilon_L(\tau_L)$. The Assumption 2.3 α-mixing property implies ergodicity. Define

$$\boldsymbol{\Gamma}_{j,i} \equiv E\left[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{x}_i(\tau_L - 1)'\right] \text{ and } \tilde{\boldsymbol{\Gamma}}_{j,i} \equiv E\left[\epsilon_L^2(\tau_L)\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{x}_i(\tau_L - 1)\right].$$

Stationarity, square integrability, and the ergodic theorem therefore yield:

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

a positive definite matrix since $\boldsymbol{x}_j(\tau_L - 1)$ has a strictly increasing σ-field under Assumption 2.1. By the construction of $\hat{\boldsymbol{\theta}}_j$ for the parsimonious model (2.6), and (B.1):

$$\begin{aligned} \sqrt{T_L}\left(\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_{0,j}\right) &= \sqrt{T_L}\left[\sum_{\tau_L=1}^{T_L} \boldsymbol{x}_j(\tau_L - 1)\boldsymbol{x}_j(\tau_L - 1)'\right]^{-1} \sum_{\tau_L=1}^{T_L} \boldsymbol{x}_j(\tau_L - 1)\epsilon_L(\tau_L) \qquad \text{(B.2)} \\ &= \boldsymbol{\Gamma}_{j,j}^{-1} \frac{1}{\sqrt{T_L}} \sum_{\tau_L=1}^{T_L} \boldsymbol{x}_j(\tau_L - 1)\epsilon_L(\tau_L) + o_p(1). \end{aligned}$$

Therefore, for any $\boldsymbol{\lambda} = [\boldsymbol{\lambda}_1', \ldots, \boldsymbol{\lambda}_h']'$, $\boldsymbol{\lambda}_j \in \mathbb{R}^{q+1}$, $\boldsymbol{\lambda}_j'\boldsymbol{\lambda}_j = 1$:

$$\boldsymbol{\lambda}' \times \sqrt{T_L}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = \frac{1}{\sqrt{T_L}} \sum_{\tau_L=1}^{T_L} X(\tau_L - 1, \boldsymbol{\lambda})\epsilon_L(\tau_L) + o_p(1), \tag{B.3}$$

where $X(\tau_L - 1, \boldsymbol{\lambda}) \equiv \sum_{j=1}^h \boldsymbol{\lambda}_j' \boldsymbol{\Gamma}_{j,j}^{-1} \boldsymbol{x}_j(\tau_L - 1)$.

Observe that:

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

where $\boldsymbol{\mathcal{S}}\in\mathbb{R}^{(q+1)h\times(q+1)h}$ has (i,j) block $\boldsymbol{\mathcal{S}}_{i,j}=\boldsymbol{\Sigma}_{i,j}\equiv\boldsymbol{\Gamma}_{j,j}^{-1}\tilde{\boldsymbol{\Gamma}}_{j,i}\boldsymbol{\Gamma}_{i,i}^{-1}$, cfr. (2.9). Under Assumptions 2.1-2.3, $\{\sum_{j=1}^{h}\boldsymbol{\lambda}_j'\boldsymbol{x}_j(\tau_L-1)\epsilon_L(\tau_L)\}$ is a stationary, ergodic, square integrable martingale difference. Therefore Billingsley's (1961) central limit theorem applies: $1/\sqrt{T_L}\sum_{\tau_L=1}^{T_L}X(\tau_L-1,\boldsymbol{\lambda})\epsilon_L(\tau_L)\overset{d}{\to}N\left(0,\boldsymbol{\lambda}'\boldsymbol{\mathcal{S}}\boldsymbol{\lambda}\right)$, hence $\sqrt{T_L}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\overset{d}{\to}N(\mathbf{0}_{(q+1)h\times1},\boldsymbol{\mathcal{S}})$ by the Cramér-Wold theorem. Now use $w_{T_L,j}\overset{p}{\to}w_j$, $\boldsymbol{R}\hat{\boldsymbol{\theta}}=\hat{\boldsymbol{\beta}}$, and $\boldsymbol{R}\boldsymbol{\theta}_0=\mathbf{0}_{h\times1}$ under H_0 to deduce as required:

$$\sqrt{T_L}\boldsymbol{W}_{T_L,h}\hat{\boldsymbol{\beta}}=\sqrt{T_L}\boldsymbol{W}_{T_L,h}\boldsymbol{R}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\overset{d}{\to}N(\mathbf{0}_{h\times1},\boldsymbol{W}_h\boldsymbol{R}\boldsymbol{\mathcal{S}}\boldsymbol{R}'\boldsymbol{W}_h).\ \mathcal{QED}\tag{B.4}$$

Proof of Theorem 2.2 $\hat{\mathcal{T}}\equiv\max_{1\leq t\leq h}(\sqrt{T_L}w_{T_L,j}\hat{\beta}_j)^2$ operates on a discrete valued stochastic function $g_{T_L}(j)\equiv w_{T_L,j}\hat{\beta}_j$. As shown in the proof of Theorem 2.1, this implies weak convergence for $\{g_{T_L}(j)\ :\ j\in\{1,...,h\}\}$ is identical to convergence in the finite dimensional distributions of $\{g_{T_L}(j)\ :\ j\in\{1,...,h\}\}$. Hansen's (1996) proof of his Theorem 2 therefore carries over to prove the present claim. $\mathcal{QED}$

Proof of Theorem 2.3 The weights $\boldsymbol{W}_{T_L,h}\overset{p}{\to}\boldsymbol{W}_h$ by assumption. Under Assumptions 2.1-2.3 $\{x_L(\tau_L),\boldsymbol{x}_j(\tau_L-1),\epsilon_L(\tau_L)\}$ are square integrable, stationary α-mixing, and therefore ergodic. Therefore $\hat{\boldsymbol{\Gamma}}_{j,i}\overset{p}{\to}\boldsymbol{\Gamma}_{j,i}$ and $\hat{\boldsymbol{\theta}}_j\overset{p}{\to}(E[\boldsymbol{x}_j(\tau_L-1)\boldsymbol{x}_j(\tau_L-1)'])^{-1}E[\boldsymbol{x}_j(\tau_L-1)x_L(\tau_L)]\equiv\boldsymbol{\theta}_j^*$. Under H_0 the parsimonious model is identically $x_L(\tau_L)=\boldsymbol{x}_j(\tau_L-1)'\boldsymbol{\theta}_{0,j}+\epsilon_L(\tau_L)$ where $\boldsymbol{\theta}_{0,j}\equiv[\alpha_{1,j},\ldots,\alpha_{q,j},0]'$, thus $\hat{\boldsymbol{\theta}}_j\overset{p}{\to}\boldsymbol{\theta}_{0,j}$, hence $\boldsymbol{\theta}_j^*=\boldsymbol{\theta}_{0,j}$. Combined with stationarity, ergodicity and square integrability: $\hat{\tilde{\boldsymbol{\Gamma}}}_{j,i}\overset{p}{\to}\tilde{\boldsymbol{\Gamma}}_{j,i}^*\equiv E[(x_L(\tau_L)-\boldsymbol{x}_j(\tau_L-1)'\boldsymbol{\theta}_j^*)^2\boldsymbol{x}_j(\tau_L-1)\boldsymbol{x}_i(\tau_L-1)']$ satisfies $||\tilde{\boldsymbol{\Gamma}}_{j,i}^*||<\infty$, and identically $\tilde{\boldsymbol{\Gamma}}_{j,i}^*=\tilde{\boldsymbol{\Gamma}}_{j,i}$ under H_0. Therefore $\hat{\boldsymbol{V}}\overset{p}{\to}\boldsymbol{V}$ under H_0, and under the alternative $\hat{\boldsymbol{V}}\overset{p}{\to}\boldsymbol{W}_{T_L,h}\boldsymbol{R}\boldsymbol{\mathcal{S}}^*\boldsymbol{R}'\boldsymbol{W}_{T_L,h}$, where $\boldsymbol{\mathcal{S}}^*\equiv[\boldsymbol{\Gamma}_{j,j}^{-1}\tilde{\boldsymbol{\Gamma}}_{j,i}^*\boldsymbol{\Gamma}_{i,i}^{-1}]$ hence $||\boldsymbol{W}_{T_L,h}\boldsymbol{R}\boldsymbol{\mathcal{S}}^*R'\boldsymbol{W}_{T_L,h}||<\infty$. $\mathcal{QED}$

Proof of Theorem 2.4 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, the least squares estimator satisfies $\hat{\boldsymbol{\theta}}_j\overset{p}{\to}\boldsymbol{\theta}_j^*$, 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:

$$\begin{aligned}\boldsymbol{\theta}_j^*&=\left[E\left[\boldsymbol{x}_j(\tau_L-1)\boldsymbol{x}_j(\tau_L-1)'\right]\right]^{-1}E\left[\boldsymbol{x}_j(\tau_L-1)\left\{\boldsymbol{X}_L(\tau_L-1)'\boldsymbol{a}+\boldsymbol{X}_H(\tau_L-1)'\boldsymbol{b}+\epsilon_L(\tau_L)\right\}\right]\\&=\left[E\left[\boldsymbol{x}_j(\tau_L-1)\boldsymbol{x}_j(\tau_L-1)'\right]\right]^{-1}\left\{E\left[\boldsymbol{x}_j(\tau_L-1)\boldsymbol{X}_L(\tau_L-1)'\right]\boldsymbol{a}+E\left[\boldsymbol{x}_j(\tau_L-1)\boldsymbol{X}_H(\tau_L-1)'\right]\boldsymbol{b}\right\},\end{aligned}\tag{B.5}$$

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:

$$\boldsymbol{X}_L(\tau_L-1)=[\boldsymbol{I}_p,\ \mathbf{0}_{p\times(q-p+1)}]\boldsymbol{x}_j(\tau_L-1)\tag{B.6}$$

hence

$$E\left[\boldsymbol{x}_j(\tau_L-1)\boldsymbol{X}_L(\tau_L-1)'\right]=E\left[\boldsymbol{x}_j(\tau_L-1)\boldsymbol{x}_j(\tau_L-1)'\right]\begin{bmatrix}\boldsymbol{I}_p\\ \boldsymbol{0}_{(q-p+1)\times p}\end{bmatrix}. \tag{B.7}$$

Substituting (B.7) into (B.5), we obtain the desired result (2.12). $\mathcal{QED}$

Proof of Theorem 2.5 Pick the last row of (2.12). 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 (2.12) is $n_j^{-1}\boldsymbol{d}_j'$, where

$$\begin{aligned}\boldsymbol{d}_j\equiv\ &E\left[\boldsymbol{X}_H(\tau_L-1)x_H(\tau_L-1,m+1-j)\right]\\&-E\left[\boldsymbol{X}_H(\tau_L-1)\boldsymbol{X}_L^{(q)}(\tau_L-1)'\right]\left\{E\left[\boldsymbol{X}_L^{(q)}(\tau_L-1)\boldsymbol{X}_L^{(q)}(\tau_L-1)'\right]\right\}^{-1}\\&\times E\left[\boldsymbol{X}_L^{(q)}(\tau_L-1)x_H(\tau_L-1,m+1-j)\right].\end{aligned} \tag{B.8}$$

If $\boldsymbol{\beta}^*=\boldsymbol{0}_{h\times 1}$ then $n_j^{-1}\boldsymbol{d}_j'\boldsymbol{b}=0$ in view of (2.12). 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 $\boldsymbol{D}\boldsymbol{b}=\boldsymbol{0}_{h\times 1}$ and thus $\boldsymbol{b}'\boldsymbol{D}'\boldsymbol{D}\boldsymbol{b}=0$, where $\boldsymbol{D}\equiv[\boldsymbol{d}_1,\ldots,\boldsymbol{d}_h]'$.

The claim $\boldsymbol{b}=\boldsymbol{0}_{pm\times 1}$ 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 (B.8) 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 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.2) 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 $\mathbf{\Gamma}_{j,j} = E[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{x}_j(\tau_L - 1)']$ and $\boldsymbol{C}_j = E[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{X}_H(\tau_L - 1)']$. The remainder of the proof proceeds along the lines of the proof of Theorem 2.1. $\mathcal{QED}$

Proof of Lemma 4.2

Claim (a). We exploit covariance stationarity without explicitly stating it. Recall $\mathbf{\Upsilon}_k = \mathbf{\Upsilon}'_{-k}$ for $k < 0$, and $\mathbf{\Upsilon}_k \equiv E[\boldsymbol{X}(\tau_L)\boldsymbol{X}(\tau_L - k)']$ for $k \geq 0$, hence

$$\mathbf{\Upsilon}_k = \begin{bmatrix} E[x_H(\tau_L,1)x_H(\tau_L-k,1)] & \dots & E[x_H(\tau_L,1)x_H(\tau_L-k,m)] & E[x_H(\tau_L,1)x_L(\tau_L-k)] \\ \vdots & \ddots & \vdots & \vdots \\ E[x_H(\tau_L,m)x_H(\tau_L-k,1)] & \dots & E[x_H(\tau_L,m)x_H(\tau_L-k,m)] & E[x_H(\tau_L,m)x_L(\tau_L-k)] \\ E[x_L(\tau_L)x_H(\tau_L-k,1)] & \dots & E[x_L(\tau_L)x_H(\tau_L-k,m)] & E[x_L(\tau_L)x_L(\tau_L-k)]. \end{bmatrix} \tag{B.9}$$

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

$$\begin{aligned} \mathbf{\Gamma}_{j,i} &\equiv E[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{x}_i(\tau_L - 1)'] = E[\boldsymbol{x}_j(\tau_L)\boldsymbol{x}_i(\tau_L)'] \\ &= E\begin{bmatrix} x_L(\tau_L) \\ \vdots \\ x_L(\tau_L-(q-1)) \\ x_H(\tau_L, m+1-j) \end{bmatrix} \begin{bmatrix} x_L(\tau_L) & \dots & x_L(\tau_L-(q-1)) & x_H(\tau_L, m+1-i) \end{bmatrix}. \end{aligned} \tag{B.10}$$

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 $\mathbf{\Gamma}_{j,i}$ is identical to which element of $\mathbf{\Upsilon}_k$. In order to ensure that the second argument of x_H lies in $\{1,\ldots,m\}$, we use the high frequency simplification (A.1):

$$\begin{aligned} x_H(\tau_L, m+1-j) &= x_H\left(\tau_L - \left\lceil \frac{1-(m+1-j)}{m} \right\rceil, m\left\lceil \frac{1-(m+1-j)}{m} \right\rceil + (m+1-j)\right) \\ &= x_H\left(\tau_L - \left\lceil \frac{j-m}{m} \right\rceil, m\left\lceil \frac{j-m}{m} \right\rceil + m+1-j\right) = x_H(\tau_L - f(j),\, g(j)), \end{aligned} \tag{B.11}$$

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 $g(j) \in \{1,\ldots,m\}$ for any j as desired. Substituting (B.11) into (B.10), $\mathbf{\Gamma}_{j,i}$ can be rewritten as follows.

$$\begin{bmatrix} E[x_L(\tau_L)x_L(\tau_L)] & \dots & E[x_L(\tau_L)x_L(\tau_L-(q-1))] & E[x_H(\tau_L-f(i),g(i))x_L(\tau_L)] \\ \vdots & \ddots & \vdots & \vdots \\ E[x_L(\tau_L-(q-1))x_L(\tau_L)] & \dots & E[x_L(\tau_L-(q-1))x_L(\tau_L-(q-1))] & E[x_H(\tau_L-f(i),g(i))x_L(\tau_L-(q-1))] \\ E[x_H(\tau_L-f(j),g(j))x_L(\tau_L)] & \dots & E[x_H(\tau_L-f(j),g(j))x_L(\tau_L-(q-1))] & E[x_H(\tau_L-f(j),g(j))x_H(\tau_L-f(i),g(i))] \end{bmatrix}. \tag{B.12}$$

Now consider how $\mathbf{\Gamma}_{j,i}$ relates to $\mathbf{\Upsilon}_k$, 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 (B.9). 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 (B.9). Combine the two cases to deduce $E[x_H(\tau_L - f(j), g(j)) \times x_L(\tau_L - (q-1))] = \Upsilon_{f(j)-(q-1)}(K, g(j))$ for any j. Now apply the same argument to each element of $\mathbf{\Gamma}_{j,i}$ appearing in (B.12) 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 (B.11) 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 $\boldsymbol{\Upsilon}_k$. 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 (B.12), 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.

Claim (b). The first equality follows from the discrete Lyapunov equation; the second from the Yule-Walker equations; and the third from covariance stationarity. $\mathcal{QED}$

C Proof of Example 2.2

We require inverses of $\mathbf{\Gamma}_{1,1} \equiv E[\boldsymbol{x}_1(\tau_L - 1)\boldsymbol{x}_1(\tau_L - 1)']$, where $\boldsymbol{x}_1(\tau_L - 1) = [x_L(\tau_L - 1), x_H(\tau_L - 1, 2)]'$ as defined in (2.6) and (2.12), and $\mathbf{\Gamma}_{2,2} = E[\boldsymbol{x}_2(\tau_L - 1)\boldsymbol{x}_2(\tau_L - 1)']$, where $\boldsymbol{x}_2(\tau_L - 1) = [x_L(\tau_L - 1), x_H(\tau_L - 1, 1)]'$. By construction $\mathbf{\Gamma}_{2,2} = \mathbf{\Gamma}_{1,1}$. Under (2.13) 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

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

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

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

Now use (2.12) to deduce:

$$\begin{bmatrix} \alpha^*_{1,1} \\ \beta^*_1 \end{bmatrix} = \begin{bmatrix} \rho^2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 1 & \rho \end{bmatrix} \begin{bmatrix} 1 \\ -1/\rho \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad , \quad \begin{bmatrix} \alpha^*_{1,2} \\ \beta^*_2 \end{bmatrix} = \begin{bmatrix} \rho^2 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ \rho & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -1/\rho \end{bmatrix} = \begin{bmatrix} 0 \\ \rho - 1/\rho \end{bmatrix}.$$

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 $\hat{\mathcal{T}} \stackrel{p}{\to} \infty$.

Tables and Figures

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

A. Mixed Frequency								
	Decaying Causality		Lagged Causality		Sporadic Causality		Uniform Causality	
	$d = 0.2$	$d = 0.8$	$d = 0.2$	$d = 0.8$	$d = 0.2$	$d = 0.8$	$d = 0.2$	$d = 0.8$
h_{MF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald	Max, Wald	Max, Wald	Max, Wald	Max, Wald
4	.489, .570	.769, .736	.048, .050	.075, .066	.390, .365	.722, .614	.083, .080	.870, .753
8	.403, .472	.705, .621	.051, .050	.177, .125	.318, .291	.667, .679	.090, .092	.949, .829
12	.350, .407	.657, .542	.247, .206	.770, .559	.678, .761	.687, .872	.095, .100	.956, .810

B. Low Frequency with Flow Sampling								
	Decaying Causality		Lagged Causality		Sporadic Causality		Uniform Causality	
	$d = 0.2$	$d = 0.8$	$d = 0.2$	$d = 0.8$	$d = 0.2$	$d = 0.8$	$d = 0.2$	$d = 0.8$
h_{LF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald	Max, Wald	Max, Wald	Max, Wald	Max, Wald
1	.075, .076	.302, .302	.093, .094	.453, .454	.070, .070	.170, .168	.247, .248	.987, .988
2	.070, .067	.238, .244	.078, .080	.466, .459	.063, .063	.133, .132	.191, .190	.977, .972
3	.061, .063	.206, .208	.073, .074	.413, .402	.060, .060	.119, .115	.167, .163	.971, .956

C. Low Frequency with Stock Sampling								
	Decaying Causality		Lagged Causality		Sporadic Causality		Uniform Causality	
	$d = 0.2$	$d = 0.8$	$d = 0.2$	$d = 0.8$	$d = 0.2$	$d = 0.8$	$d = 0.2$	$d = 0.8$
h_{LF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald	Max, Wald	Max, Wald	Max, Wald	Max, Wald
1	.648, .643	.860, .862	.049, .050	.060, .059	.052, .051	.435, .435	.066, .067	.644, .644
2	.551, .538	.796, .786	.064, .063	.675, .664	.050, .051	.357, .356	.061, .062	.691, .736
3	.493, .473	.755, .728	.059, .060	.624, .598	.050, .051	.307, .305	.059, .059	.633, .674

The DGP is MF-VAR(1) with a ratio of sampling frequencies $m = 12$. From left to right, there is Pitman 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 $j = 1, \ldots, 12$; sporadic causality $(\nu_3, \nu_9, \nu_{11}) = (2.1, -2.8, 1.9)$ with all other $\nu_j = 0$; and uniform causality $\nu_j = 0.3$ for $j = 1, \ldots, 12$. The high frequency variable x_H has low or high persistence: $d \in \{0.2, 0.8\}$. The low frequency variable x_L has low persistence: $a = 0.2$. There exists low-to-high decaying causality with alternating signs: $c_j = (-1)^{j-1} \times 0.8/j$ for $j = 1, \ldots, 12$. The max test uses 100000 draws from the limit distributions under H_0 and H_1^L. The weights are $\boldsymbol{W}_h = (1/h) \times \boldsymbol{I}_h$, and nominal size is $\alpha = .05$.

Table 2: Rejection Frequencies of High-to-Low Causality Tests Based on MF-VAR(1) - GARCH Error and Robust Covariance Matrix for Max Tests

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

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

A.1.1. $T_L = 80$						A.1.2. $T_L = 160$				
	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
4	.055, .046	1	.057, .054	.059, .057		4	.052, .053	1	.052, .063	.053, .060
8	.046, .045	2	.056, .050	.051, .039		8	.052, .045	2	.053, .048	.053, .045
12	.048, .041	3	.050, .051	.051, .041		12	.045, .046	3	.042, .053	.049, .051
24	.043, .046	4	.054, .038	.054, .047		24	.041, .031	4	.049, .045	.051, .046

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

A.2.1. $T_L = 80$						A.2.2. $T_L = 160$				
	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
4	.056, .036	1	.057, .043	.054, .050		4	.051, .032	1	.054, .052	.048, .041
8	.054, .040	2	.053, .038	.054, .040		8	.049, .050	2	.049, .040	.052, .046
12	.047, .034	3	.050, .039	.049, .039		12	.052, .061	3	.054, .048	.046, .056
24	.043, .041	4	.058, .051	.056, .036		24	.040, .034	4	.049, .040	.050, .047

B. Decaying Causality: $b_j = (-1)^{j-1}0.3/j$ for $j = 1, \ldots, 12$

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

B.1.1. $T_L = 80$						B.1.2. $T_L = 160$				
	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
4	.523, .624	1	.089, .083	.681, .649		4	.878, .920	1	.118, .109	.932, .919
8	.410, .481	2	.078, .073	.574, .540		8	.809, .868	2	.103, .094	.880, .859
12	.361, .434	3	.068, .054	.518, .481		12	.770, .826	3	.090, .078	.852, .846
24	.270, .250	4	.072, .065	.493, .445		24	.692, .639	4	.078, .081	.832, .790

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

B.2.1. $T_L = 80$						B.2.2. $T_L = 160$				
	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
4	.533, .631	1	.073, .058	.676, .679		4	.881, .906	1	.088, .079	.930, .899
8	.443, .487	2	.068, .055	.583, .553		8	.833, .879	2	.070, .070	.885, .866
12	.413, .434	3	.059, .050	.515, .529		12	.816, .826	3	.070, .067	.861, .835
24	.311, .251	4	.065, .050	.482, .426		24	.733, .675	4	.071, .054	.837, .790

There is weak 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 weight for max tests is $\boldsymbol{W}_h = (1/h) \times \boldsymbol{I}_h$, and p-value is computed using 5000 draws from the null limit distribution. The Wald test p-value is computed using the parametric bootstrap based on Gonçalves and Kilian (2004), with 1000 bootstrap replications. Nominal size is $\alpha = 0.05$. The number of Monte Carlo samples is 5000 for max tests and 1000 for Wald tests.

Table 2: Continued

C. Lagged Causality: $b_j = 0.3 \times I(j = 12)$ for $j = 1, \ldots, 12$										
C.1. $d = 0.2$ (low persistence in x_H)										
C.1.1. $T_L = 80$						C.1.2. $T_L = 160$				
	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
4	.066, .054	1	.135, .107	.066, .047		4	.071, .058	1	.198, .184	.066, .068
8	.059, .063	2	.101, .108	.073, .087		8	.063, .058	2	.143, .159	.098, .073
12	.327, .275	3	.089, .093	.065, .053		12	.763, .610	3	.135, .126	.079, .074
24	.235, .161	4	.085, .078	.068, .071		24	.685, .434	4	.130, .114	.071, .062
C.2. $d = 0.8$ (high persistence in x_H)										
C.2.1. $T_L = 80$						C.2.2. $T_L = 160$				
	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
4	.073, .060	1	.477, .443	.066, .043		4	.079, .069	1	.778, .749	.063, .062
8	.094, .098	2	.387, .353	.407, .371		8	.115, .140	2	.699, .677	.713, .699
12	.346, .430	3	.351, .284	.347, .337		12	.854, .862	3	.660, .607	.667, .614
24	.276, .237	4	.323, .242	.315, .258		24	.788, .703	4	.623, .546	.624, .575
D. Sporadic Causality: $(b_3, b_7, b_{10}) = (0.2, 0.05, -0.3)$ and $b_j = 0$ for other j's										
D.1. $d = 0.2$ (low persistence in x_H)										
D.1.1. $T_L = 80$						D.1.2. $T_L = 160$				
	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
4	.218, .231	1	.069, .058	.076, .068		4	.451, .451	1	.075, .061	.064, .059
8	.170, .174	2	.062, .055	.053, .056		8	.374, .346	2	.063, .071	.065, .054
12	.373, .386	3	.057, .053	.059, .049		12	.809, .812	3	.060, .042	.059, .045
24	.276, .255	4	.065, .057	.056, .044		24	.721, .644	4	.057, .060	.062, .055
D.2. $d = 0.8$ (high persistence in x_H)										
D.2.1. $T_L = 80$						D.2.2. $T_L = 160$				
	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
4	.195, .183	1	.138, .110	.071, .068		4	.395, .378	1	.215, .206	.074, .067
8	.160, .152	2	.105, .099	.139, .118		8	.331, .325	2	.174, .155	.226, .234
12	.363, .427	3	.099, .094	.115, .103		12	.807, .844	3	.156, .148	.202, .186
24	.268, .266	4	.096, .079	.102, .094		24	.723, .674	4	.132, .136	.167, .165
E. Uniform Causality: $b_j = 0.02$ for $j = 1, \ldots, 12$										
E.1. $d = 0.2$ (low persistence in x_H)										
E.1.1. $T_L = 80$						E.1.2. $T_L = 160$				
	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
4	.061, .049	1	.103, .107	.065, .056		4	.071, .049	1	.148, .128	.062, .054
8	.053, .048	2	.075, .063	.053, .042		8	.061, .055	2	.107, .101	.057, .052
12	.056, .043	3	.071, .079	.056, .054		12	.061, .060	3	.093, .089	.056, .044
24	.048, .042	4	.062, .060	.056, .042		24	.052, .063	4	.085, .087	.057, .047
E.2. $d = 0.8$ (high persistence in x_H)										
E.2.1. $T_L = 80$						E.2.2. $T_L = 160$				
	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
4	.066, .058	1	.154, .132	.065, .051		4	.067, .069	1	.247, .241	.064, .067
8	.064, .046	2	.106, .072	.089, .056		8	.077, .071	2	.186, .180	.126, .114
12	.069, .050	3	.103, .094	.075, .051		12	.089, .081	3	.165, .146	.101, .113
24	.064, .054	4	.094, .070	.073, .072		24	.083, .061	4	.156, .138	.104, .098

Table 3: Rejection Frequencies of High-to-Low Causality Tests Based on MF-VAR(2) - GARCH Error and Robust Covariance Matrix for Max Tests

A. Non-Causality: $\boldsymbol{b} = \mathbf{0}_{24\times1}$										
A.1. $d = 0.2$ (low persistence in x_H)										
A.1.1. $T_L = 80$						A.1.2. $T_L = 160$				
	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	.046, .043	1	.060, .054	.055, .044		16	.043, .052	1	.054, .049	.049, .037
20	.045, .040	2	.060, .055	.060, .044		20	.053, .040	2	.049, .053	.059, .055
24	.045, .048	3	.051, .038	.046, .037		24	.047, .040	3	.047, .055	.052, .031
A.2. $d = 0.8$ (high persistence in x_H)										
A.2.1. $T_L = 80$						A.2.2. $T_L = 160$				
	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	.046, .037	1	.054, .046	.054, .050		16	.052, .037	1	.054, .048	.055, .042
20	.047, .048	2	.054, .042	.056, .046		20	.043, .046	2	.048, .048	.056, .056
24	.041, .043	3	.054, .042	.049, .042		24	.046, .042	3	.050, .048	.053, .046
B. Decaying Causality: $b_j = (-1)^{j-1} 0.3/j$ for $j = 1, \ldots, 24$										
B.1. $d = 0.2$ (low persistence in x_H)										
B.1.1. $T_L = 80$						B.1.2. $T_L = 160$				
	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	.312, .337	1	.094, .087	.680, .651		16	.736, .785	1	.117, .100	.930, .931
20	.273, .314	2	.082, .063	.564, .553		20	.707, .716	2	.097, .100	.890, .883
24	.259, .239	3	.074, .057	.514, .507		24	.697, .654	3	.093, .069	.855, .810
B.2. $d = 0.8$ (high persistence in x_H)										
B.2.1. $T_L = 80$						B.2.2. $T_L = 160$				
	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	.354, .324	1	.080, .069	.689, .671		16	.771, .780	1	.099, .091	.928, .923
20	.312, .280	2	.064, .076	.571, .563		20	.741, .724	2	.087, .092	.897, .879
24	.315, .262	3	.073, .045	.523, .469		24	.731, .679	3	.081, .088	.860, .829

There is weak 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 weight for max tests is $\boldsymbol{W}_h = (1/h) \times \boldsymbol{I}_h$, and p-value is computed using 5000 draws from the null limit distribution. The Wald test p-value is computed using the parametric bootstrap based on Gonçalves and Kilian (2004), with 1000 bootstrap replications. Nominal size is $\alpha = 0.05$. The number of Monte Carlo samples is 5000 for max tests and 1000 for Wald tests.

Table 3: Continued

C. Lagged Causality: $b_j = 0.3 \times I(j = 24)$ for $j = 1, \ldots, 24$										
C.1. $d = 0.2$ (low persistence in x_H)										
C.1.1. $T_L = 80$						C.1.2. $T_L = 160$				
	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	.051, .061	1	.056, .046	.055, .044		16	.051, .053	1	.056, .053	.049, .048
20	.048, .050	2	.115, .107	.064, .043		20	.056, .054	2	.179, .165	.056, .054
24	.205, .158	3	.099, .077	.061, .062		24	.682, .530	3	.164, .143	.088, .077
C.2. $d = 0.8$ (high persistence in x_H)										
C.2.1. $T_L = 80$						C.2.2. $T_L = 160$				
	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	.058, .046	1	.061, .060	.055, .040		16	.060, .040	1	.065, .056	.055, .051
20	.067, .051	2	.433, .402	.056, .041		20	.096, .112	2	.760, .752	.057, .065
24	.206, .281	3	.376, .362	.352, .311		24	.770, .740	3	.729, .709	.673, .640
D. Sporadic Causality: $(b_5, b_{12}, b_{17}, b_{19}) = (-0.2, 0.1, 0.2, -0.35)$ and $b_j = 0$ for other j's										
D.1. $d = 0.2$ (low persistence in x_H)										
D.1.1. $T_L = 80$						D.1.2. $T_L = 160$				
	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	.113, .117	1	.067, .063	.062, .055		16	.290, .276	1	.077, .057	.062, .053
20	.414, .440	2	.078, .064	.058, .057		20	.883, .886	2	.085, .070	.066, .063
24	.391, .373	3	.071, .055	.065, .058		24	.858, .881	3	.077, .079	.060, .068
D.2. $d = 0.8$ (high persistence in x_H)										
D.2.1. $T_L = 80$						D.2.2. $T_L = 160$				
	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	.111, .124	1	.068, .055	.061, .051		16	.256, .284	1	.069, .061	.057, .057
20	.429, .454	2	.088, .063	.078, .050		20	.877, .888	2	.105, .104	.087, .091
24	.415, .376	3	.078, .062	.072, .082		24	.883, .875	3	.093, .091	.097, .084
E. Uniform Causality: $b_j = 0.02$ for $j = 1, \ldots, 24$										
E.1. $d = 0.2$ (low persistence in x_H)										
E.1.1. $T_L = 80$						E.1.2. $T_L = 160$				
	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	.051, .061	1	.091, .078	.054, .055		16	.062, .064	1	.152, .140	.063, .060
20	.050, .053	2	.107, .103	.062, .054		20	.060, .052	2	.166, .158	.060, .047
24	.049, .042	3	.091, .070	.054, .041		24	.055, .072	3	.140, .139	.058, .056
E.2. $d = 0.8$ (high persistence in x_H)										
E.2.1. $T_L = 80$						E.2.2. $T_L = 160$				
	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	.079, .077	1	.173, .175	.060, .051		16	.116, .130	1	.312, .292	.063, .061
20	.072, .070	2	.198, .187	.111, .107		20	.118, .092	2	.380, .391	.187, .176
24	.077, .065	3	.165, .148	.116, .104		24	.112, .121	3	.319, .311	.198, .215

Table 4: Rejection Frequencies of Low-to-High Causality Tests Based on MF-VAR(1) - GARCH Error and Robust Covariance Matrix for Max Tests

A. Non-Causality: $\boldsymbol{c} = \mathbf{0}_{12\times 1}$									
A.1 $T_L = 80$					A.2 $T_L = 160$				
A.1.1 Mixed Frequency Tests without MIDAS					A.2.1 Mixed Frequency Tests without MIDAS				
	$r_{MF} = 4$	$r_{MF} = 8$	$r_{MF} = 12$	$r_{MF} = 24$		$r_{MF} = 4$	$r_{MF} = 8$	$r_{MF} = 12$	$r_{MF} = 24$
h_{MF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald	h_{MF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald
4	.055, .054	.051, .039	.048, .049	.046, .046	4	.055, .058	.051, .050	.050, .048	.045, .038
8	.068, .047	.064, .045	.068, .047	.061, .050	8	.053, .044	.055, .049	.050, .033	.058, .036
12	.080, .047	.079, .041	.080, .047	.076, .046	12	.060, .054	.061, .044	.065, .047	.057, .039
24	.139, .044	.144, .062	.157, .046	.167, .052	24	.079, .052	.083, .037	.077, .046	.082, .046
A.1.2 Mixed Frequency Tests with MIDAS					A.2.2 Mixed Frequency Tests with MIDAS				
	$r_{MF} = 4$	$r_{MF} = 8$	$r_{MF} = 12$	$r_{MF} = 24$		$r_{MF} = 4$	$r_{MF} = 8$	$r_{MF} = 12$	$r_{MF} = 24$
h_{MF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald	h_{MF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald
4	.051, .054	.051, .045	.046, .044	.042, .043	4	.054, .063	.050, .047	.047, .052	.041, .042
8	.056, .047	.049, .044	.051, .051	.042, .047	8	.048, .046	.046, .055	.046, .034	.050, .043
12	.056, .061	.051, .044	.049, .045	.043, .044	12	.049, .056	.048, .053	.049, .043	.044, .043
24	.055, .052	.049, .060	.046, .039	.043, .035	24	.051, .058	.049, .048	.046, .046	.046, .047
A.1.3 Low Frequency Tests (Flow Sampling)					A.2.3 Low Frequency Tests (Flow Sampling)				
	$r_{LF} = 1$	$r_{LF} = 2$	$r_{LF} = 3$	$r_{LF} = 4$		$r_{LF} = 1$	$r_{LF} = 2$	$r_{LF} = 3$	$r_{LF} = 4$
h_{LF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald	h_{LF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald
1	.056, .044	.053, .050	.052, .050	.043, .049	1	.053, .046	.051, .043	.054, .049	.042, .044
2	.063, .042	.053, .047	.047, .039	.050, .055	2	.054, .051	.054, .049	.049, .047	.051, .045
3	.060, .058	.056, .051	.052, .045	.055, .040	3	.053, .050	.054, .048	.049, .041	.052, .044
4	.060, .064	.057, .041	.053, .051	.050, .043	4	.053, .053	.058, .059	.055, .035	.049, .052
A.1.4 Low Frequency Tests (Stock Sampling)					A.2.4 Low Frequency Tests (Stock Sampling)				
	$r_{LF} = 1$	$r_{LF} = 2$	$r_{LF} = 3$	$r_{LF} = 4$		$r_{LF} = 1$	$r_{LF} = 2$	$r_{LF} = 3$	$r_{LF} = 4$
h_{LF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald	h_{LF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald
1	.054, .057	.050, .041	.044, .049	.044, .050	1	.047, .029	.050, .048	.049, .056	.051, .045
2	.057, .039	.052, .042	.051, .046	.047, .051	2	.052, .054	.050, .057	.049, .053	.048, .052
3	.062, .034	.055, .048	.051, .046	.048, .047	3	.055, .057	.055, .036	.052, .061	.049, .047
4	.057, .044	.054, .043	.060, .049	.056, .046	4	.050, .042	.049, .052	.053, .032	.054, .039

The AR(1) coefficient for x_L is $a = 0.2$, while the AR(1) coefficient for x_H is $d = 0.2$. There is high-to-low decaying causality with alternating signs: $b_j = (-1)^{j-1} \times 0.2/j$ for $j = 1, \ldots, 12$. The models estimated have two low frequency lags of x_L (i.e. $q = 2$). The max test statistic is computed using a flat weight $\boldsymbol{W}_r = (1/r) \times \boldsymbol{I}_r$, and p-value is computed using 5000 draws from the null limit distribution. The Wald test p-value is computed using the parametric bootstrap based on Gonçalves and Kilian (2004), with 1000 bootstrap replications. We implement mixed frequency tests with and without an Almon MIDAS polynomial of dimension $s = 3$. Nominal size is $\alpha = 0.05$. The number of Monte Carlo samples is 5000 for max tests and 1000 for Wald tests.

Table 4: Continued

B. Decaying Causality: $c_j = (-1)^{j-1} \times 0.3/j$ for $j = 1, \ldots, 12$									
B.1 $T_L = 80$					B.2 $T_L = 160$				
B.1.1 Mixed Frequency Tests without MIDAS					B.2.1 Mixed Frequency Tests without MIDAS				
	$r_{MF} = 4$	$r_{MF} = 8$	$r_{MF} = 12$	$r_{MF} = 24$		$r_{MF} = 4$	$r_{MF} = 8$	$r_{MF} = 12$	$r_{MF} = 24$
h_{MF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald	h_{MF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald
4	.459, .522	.353, .401	.290, .371	.208, .231	4	.823, .893	.739, .825	.695, .735	.578, .588
8	.471, .481	.357, .417	.295, .335	.220, .208	8	.818, .846	.744, .803	.687, .733	.585, .570
12	.464, .445	.370, .397	.320, .324	.234, .189	12	.824, .851	.740, .786	.677, .724	.596, .555
24	.508, .354	.430, .280	.393, .228	.333, .156	24	.813, .831	.751, .753	.702, .688	.614, .508
B.1.2 Mixed Frequency Tests with MIDAS					B.2.2 Mixed Frequency Tests with MIDAS				
	$r_{MF} = 4$	$r_{MF} = 8$	$r_{MF} = 12$	$r_{MF} = 24$		$r_{MF} = 4$	$r_{MF} = 8$	$r_{MF} = 12$	$r_{MF} = 24$
h_{MF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald	h_{MF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald
4	.470, .548	.360, .431	.293, .381	.206, .250	4	.837, .901	.757, .835	.708, .747	.590, .623
8	.477, .523	.358, .464	.293, .392	.208, .232	8	.835, .881	.766, .844	.705, .778	.602, .626
12	.472, .534	.361, .465	.306, .396	.202, .252	12	.840, .894	.759, .847	.705, .804	.611, .634
24	.479, .540	.361, .470	.309, .380	.201, .237	24	.838, .909	.766, .853	.723, .799	.620, .623
B.1.3 Low Frequency Tests (Flow Sampling)					B.2.3 Low Frequency Tests (Flow Sampling)				
	$r_{LF} = 1$	$r_{LF} = 2$	$r_{LF} = 3$	$r_{LF} = 4$		$r_{LF} = 1$	$r_{LF} = 2$	$r_{LF} = 3$	$r_{LF} = 4$
h_{LF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald	h_{LF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald
1	.096, .103	.083, .073	.065, .070	.066, .072	1	.159, .137	.125, .125	.093, .096	.090, .076
2	.102, .081	.082, .070	.070, .058	.061, .063	2	.155, .142	.119, .123	.108, .102	.093, .094
3	.107, .089	.084, .085	.072, .071	.069, .077	3	.145, .142	.116, .123	.107, .103	.094, .107
4	.107, .085	.086, .082	.073, .068	.069, .069	4	.150, .146	.119, .122	.102, .108	.086, .092
B.1.4 Low Frequency Tests (Stock Sampling)					B.2.4 Low Frequency Tests (Stock Sampling)				
	$r_{LF} = 1$	$r_{LF} = 2$	$r_{LF} = 3$	$r_{LF} = 4$		$r_{LF} = 1$	$r_{LF} = 2$	$r_{LF} = 3$	$r_{LF} = 4$
h_{LF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald	h_{LF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald
1	.056, .054	.055, .052	.049, .049	.047, .040	1	.055, .065	.056, .059	.054, .049	.052, .058
2	.057, .038	.052, .035	.046, .058	.050, .043	2	.062, .054	.057, .051	.054, .050	.055, .047
3	.057, .053	.055, .045	.049, .048	.052, .057	3	.062, .053	.057, .062	.051, .057	.058, .049
4	.063, .059	.058, .049	.057, .046	.051, .044	4	.056, .055	.061, .060	.053, .062	.057, .036

Table 4: Continued

C. Lagged Causality: $c_j = 0.25 \times I(j = 12)$ for $j = 1, \ldots, 12$									
C.1 $T_L = 80$					C.2 $T_L = 160$				
C.1.1 Mixed Frequency Tests without MIDAS					C.2.1 Mixed Frequency Tests without MIDAS				
	$r_{MF} = 4$	$r_{MF} = 8$	$r_{MF} = 12$	$r_{MF} = 24$		$r_{MF} = 4$	$r_{MF} = 8$	$r_{MF} = 12$	$r_{MF} = 24$
h_{MF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald	h_{MF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald
4	.054, .045	.054, .039	.211, .205	.139, .143	4	.054, .044	.053, .051	.513, .476	.429, .343
8	.068, .048	.065, .043	.208, .182	.167, .101	8	.056, .050	.054, .046	.522, .469	.425, .289
12	.081, .047	.080, .049	.240, .169	.181, .106	12	.066, .055	.058, .044	.519, .409	.428, .268
24	.132, .053	.144, .040	.309, .122	.271, .094	24	.083, .050	.084, .045	.528, .380	.450, .272
C.1.2 Mixed Frequency Tests with MIDAS					C.2.2 Mixed Frequency Tests with MIDAS				
	$r_{MF} = 4$	$r_{MF} = 8$	$r_{MF} = 12$	$r_{MF} = 24$		$r_{MF} = 4$	$r_{MF} = 8$	$r_{MF} = 12$	$r_{MF} = 24$
h_{MF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald	h_{MF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald
4	.055, .036	.052, .047	.210, .213	.140, .155	4	.051, .041	.051, .048	.521, .478	.442, .343
8	.055, .042	.046, .045	.199, .196	.149, .121	8	.048, .048	.048, .042	.537, .476	.439, .309
12	.055, .047	.050, .050	.210, .205	.146, .126	12	.053, .060	.044, .047	.538, .472	.428, .318
24	.053, .055	.050, .035	.215, .213	.145, .129	24	.052, .046	.053, .044	.536, .487	.442, .356
C.1.3 Low Frequency Tests (Flow Sampling)					C.2.3 Low Frequency Tests (Flow Sampling)				
	$r_{LF} = 1$	$r_{LF} = 2$	$r_{LF} = 3$	$r_{LF} = 4$		$r_{LF} = 1$	$r_{LF} = 2$	$r_{LF} = 3$	$r_{LF} = 4$
h_{LF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald	h_{LF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald
1	.093, .082	.088, .080	.072, .063	.063, .078	1	.156, .157	.126, .114	.111, .103	.104, .113
2	.103, .099	.086, .076	.075, .071	.068, .063	2	.150, .162	.128, .132	.110, .132	.096, .107
3	.108, .088	.087, .067	.077, .064	.075, .049	3	.153, .164	.124, .129	.110, .107	.099, .085
4	.101, .091	.088, .084	.078, .087	.077, .063	4	.151, .132	.133, .111	.110, .125	.100, .103
C.1.4 Low Frequency Tests (Stock Sampling)					C.2.4 Low Frequency Tests (Stock Sampling)				
	$r_{LF} = 1$	$r_{LF} = 2$	$r_{LF} = 3$	$r_{LF} = 4$		$r_{LF} = 1$	$r_{LF} = 2$	$r_{LF} = 3$	$r_{LF} = 4$
h_{LF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald	h_{LF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald
1	.552, .553	.455, .439	.402, .398	.363, .340	1	.847, .870	.787, .785	.746, .752	.720, .705
2	.557, .525	.457, .450	.397, .374	.355, .304	2	.848, .863	.795, .775	.742, .735	.707, .650
3	.554, .521	.453, .415	.402, .378	.364, .342	3	.845, .836	.784, .763	.745, .714	.696, .651
4	.538, .501	.456, .415	.394, .344	.356, .283	4	.844, .848	.787, .778	.743, .740	.698, .662

Table 4: Continued

D. Sporadic Causality: $(c_3, c_7, c_{10}) = (0.3, 0.15, -0.3)$ and $c_j = 0$ for other j's									
D.1 $T_L = 80$					D.2 $T_L = 160$				
D.1.1 Mixed Frequency Tests without MIDAS					D.2.1 Mixed Frequency Tests without MIDAS				
	$r_{MF} = 4$	$r_{MF} = 8$	$r_{MF} = 12$	$r_{MF} = 24$		$r_{MF} = 4$	$r_{MF} = 8$	$r_{MF} = 12$	$r_{MF} = 24$
h_{MF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald	h_{MF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald
4	.471, .445	.386, .454	.488, .675	.378, .445	4	.834, .832	.810, .835	.913, .968	.860, .900
8	.464, .444	.385, .436	.495, .649	.382, .400	8	.831, .831	.791, .820	.908, .958	.853, .901
12	.464, .422	.410, .372	.510, .587	.391, .365	12	.834, .808	.794, .815	.908, .972	.855, .887
24	.510, .314	.471, .299	.574, .465	.492, .246	24	.825, .762	.787, .790	.908, .928	.854, .832
D.1.2 Mixed Frequency Tests with MIDAS					D.2.2 Mixed Frequency Tests with MIDAS				
	$r_{MF} = 4$	$r_{MF} = 8$	$r_{MF} = 12$	$r_{MF} = 24$		$r_{MF} = 4$	$r_{MF} = 8$	$r_{MF} = 12$	$r_{MF} = 24$
h_{MF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald	h_{MF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald
4	.473, .467	.402, .490	.502, .698	.384, .467	4	.842, .843	.821, .839	.920, .974	.870, .911
8	.476, .493	.388, .489	.509, .700	.380, .473	8	.852, .857	.806, .840	.920, .972	.864, .929
12	.463, .487	.403, .468	.508, .697	.374, .476	12	.853, .850	.816, .858	.921, .987	.869, .920
24	.503, .464	.403, .496	.506, .698	.385, .445	24	.852, .864	.821, .852	.933, .971	.868, .927
D.1.3 Low Frequency Tests (Flow Sampling)					D.2.3 Low Frequency Tests (Flow Sampling)				
	$r_{LF} = 1$	$r_{LF} = 2$	$r_{LF} = 3$	$r_{LF} = 4$		$r_{LF} = 1$	$r_{LF} = 2$	$r_{LF} = 3$	$r_{LF} = 4$
h_{LF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald	h_{LF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald
1	.084, .067	.061, .056	.065, .065	.055, .043	1	.111, .088	.091, .092	.079, .071	.075, .075
2	.080, .070	.069, .061	.059, .055	.059, .066	2	.112, .097	.087, .068	.083, .076	.080, .090
3	.079, .075	.071, .067	.062, .055	.060, .055	3	.110, .102	.093, .093	.083, .071	.073, .064
4	.091, .062	.068, .060	.060, .056	.066, .047	4	.103, .088	.089, .098	.081, .071	.083, .054
D.1.4 Low Frequency Tests (Stock Sampling)					D.2.4 Low Frequency Tests (Stock Sampling)				
	$r_{LF} = 1$	$r_{LF} = 2$	$r_{LF} = 3$	$r_{LF} = 4$		$r_{LF} = 1$	$r_{LF} = 2$	$r_{LF} = 3$	$r_{LF} = 4$
h_{LF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald	h_{LF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald
1	.058, .058	.046, .039	.050, .043	.050, .041	1	.058, .052	.046, .051	.049, .048	.048, .047
2	.061, .050	.054, .040	.051, .052	.049, .033	2	.061, .047	.054, .039	.053, .042	.052, .051
3	.055, .040	.060, .046	.045, .049	.052, .044	3	.052, .044	.049, .046	.051, .045	.053, .053
4	.057, .063	.053, .036	.055, .041	.054, .048	4	.062, .050	.055, .059	.052, .054	.055, .057

Table 4: Continued

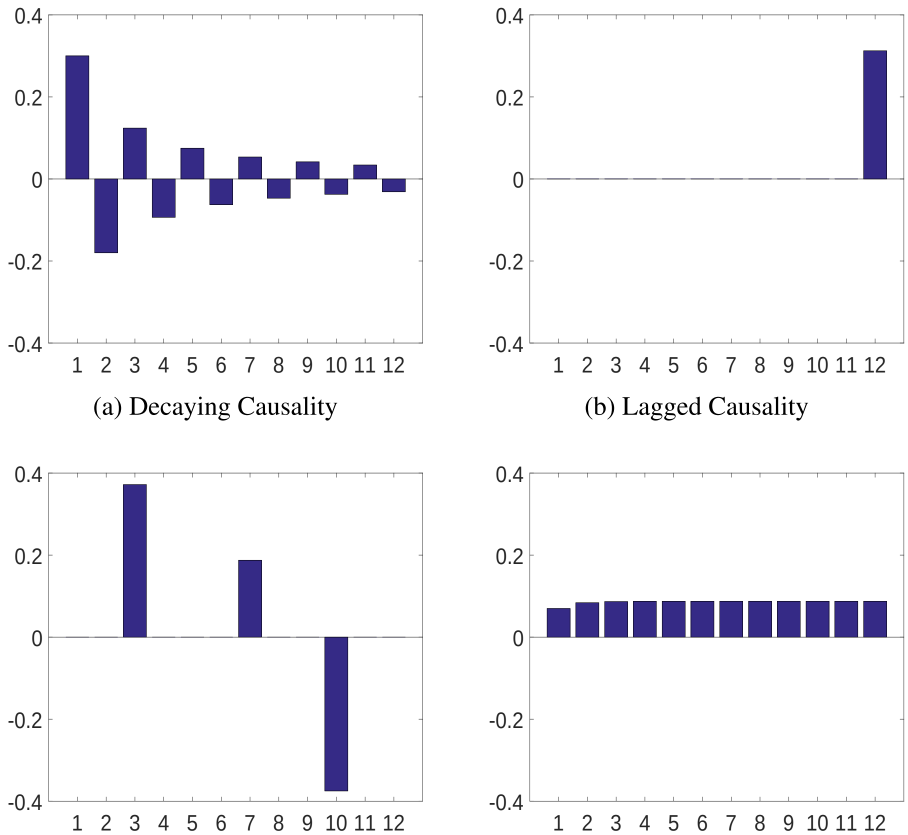
E. Uniform Causality: $c_j = 0.07$ for $j = 1, \ldots, 12$									
E.1 $T_L = 80$					E.2 $T_L = 160$				
E.1.1 Mixed Frequency Tests without MIDAS					E.2.1 Mixed Frequency Tests without MIDAS				
	$r_{MF} = 4$	$r_{MF} = 8$	$r_{MF} = 12$	$r_{MF} = 24$		$r_{MF} = 4$	$r_{MF} = 8$	$r_{MF} = 12$	$r_{MF} = 24$
h_{MF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald	h_{MF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald
4	.131, .171	.150, .210	.164, .269	.103, .165	4	.247, .321	.290, .497	.319, .622	.220, .460
8	.140, .162	.179, .191	.179, .260	.126, .166	8	.245, .315	.302, .481	.319, .597	.243, .425
12	.166, .155	.180, .185	.203, .229	.158, .147	12	.254, .278	.313, .464	.340, .590	.252, .435
24	.221, .106	.282, .144	.302, .200	.261, .122	24	.286, .252	.346, .397	.383, .571	.296, .397
E.1.2 Mixed Frequency Tests with MIDAS					E.2.2 Mixed Frequency Tests with MIDAS				
	$r_{MF} = 4$	$r_{MF} = 8$	$r_{MF} = 12$	$r_{MF} = 24$		$r_{MF} = 4$	$r_{MF} = 8$	$r_{MF} = 12$	$r_{MF} = 24$
h_{MF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald	h_{MF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald
4	.133, .168	.151, .219	.159, .287	.106, .169	4	.250, .332	.296, .527	.322, .635	.216, .463
8	.130, .169	.160, .230	.157, .304	.108, .185	8	.242, .337	.297, .528	.316, .658	.232, .471
12	.140, .172	.144, .226	.157, .294	.109, .183	12	.252, .309	.293, .487	.321, .683	.235, .505
24	.136, .152	.156, .221	.156, .307	.100, .197	24	.256, .305	.301, .506	.319, .677	.230, .512
E.1.3 Low Frequency Tests (Flow Sampling)					E.2.3 Low Frequency Tests (Flow Sampling)				
	$r_{LF} = 1$	$r_{LF} = 2$	$r_{LF} = 3$	$r_{LF} = 4$		$r_{LF} = 1$	$r_{LF} = 2$	$r_{LF} = 3$	$r_{LF} = 4$
h_{LF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald	h_{LF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald
1	.703, .680	.610, .592	.552, .520	.512, .475	1	.951, .944	.911, .907	.891, .875	.867, .850
2	.705, .701	.617, .590	.563, .530	.510, .473	2	.946, .952	.914, .906	.881, .870	.867, .826
3	.681, .701	.604, .580	.556, .527	.505, .472	3	.947, .934	.919, .909	.892, .867	.862, .822
4	.684, .667	.596, .554	.529, .498	.491, .439	4	.942, .935	.911, .896	.888, .866	.854, .820
E.1.4 Low Frequency Tests (Stock Sampling)					E.2.4 Low Frequency Tests (Stock Sampling)				
	$r_{LF} = 1$	$r_{LF} = 2$	$r_{LF} = 3$	$r_{LF} = 4$		$r_{LF} = 1$	$r_{LF} = 2$	$r_{LF} = 3$	$r_{LF} = 4$
h_{LF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald	h_{LF}	Max, Wald	Max, Wald	Max, Wald	Max, Wald
1	.120, .118	.099, .095	.076, .080	.071, .076	1	.195, .180	.145, .156	.133, .115	.109, .116
2	.119, .112	.096, .092	.084, .067	.071, .077	2	.198, .180	.146, .143	.133, .130	.109, .117
3	.115, .128	.095, .089	.083, .071	.078, .071	3	.197, .187	.151, .158	.128, .127	.108, .126
4	.120, .111	.105, .089	.079, .071	.080, .077	4	.186, .172	.151, .149	.129, .118	.117, .103

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



(a) Decaying Causality

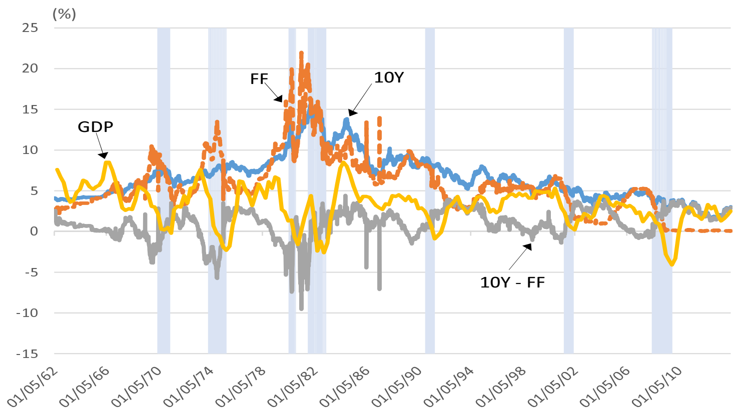
(b) Lagged Causality

(c) Sporadic Causality

(d) Uniform Causality

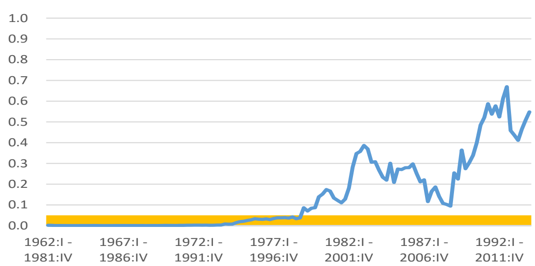
In the low-to-high causality simulation experiment, we start with a structural MF-VAR(1) data generating process, and transform it to a reduced-form MF-VAR(1). This figure shows how each causal pattern in the structural form is transformed in the reduced form. The AR(1) parameter of x_H is fixed at $d = 0.2$. The horizontal axis has the 1st through 12th lags, and the vertical axis has the reduced form coefficient for each lag. In the structural form, the decaying causality is $c_j = (-1)^{j-1} \times 0.3/j$ for $j = 1, \dots, 12$; the lagged causality is $c_j = 0.25 \times I(j = 12)$; the sporadic causality is $(c_3, c_7, c_{10}) = (0.3, 0.15, -0.3)$ and $c_j = 0$ for all other j's; the uniform causality is $c_j = 0.07$ for all j's. As indicated in Panels (a)-(d), the causal patterns in the reduced form resemble the structural causal patterns.

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

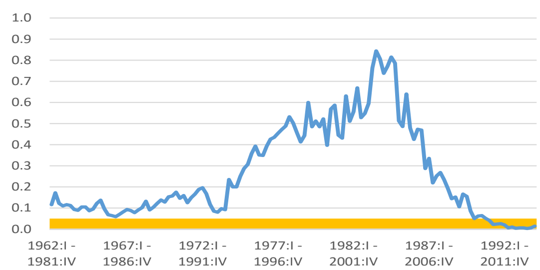


Note: This figure plots weekly 10-year Treasury constant maturity rate (blue, solid line), weekly effective federal funds rate (red, dashed line), their spread 10Y−FF (gray, solid line), and the quarterly real GDP growth from previous year (yellow, solid line). The sample period covers January 5, 1962 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).

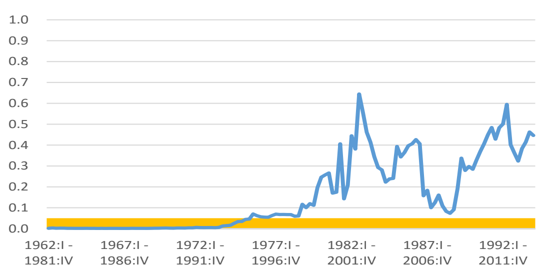
Figure 3: P-values for Tests of Non-Causality from Interest Rate Spread to GDP



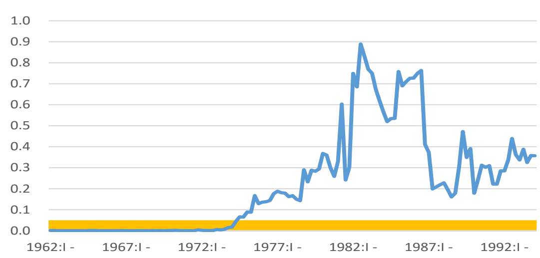
(a) MF Max Test



(b) MF Wald Test



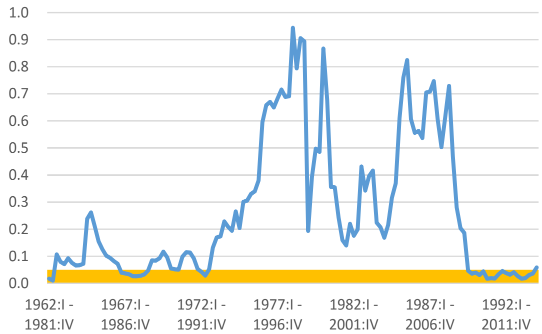
(c) LF Max Test



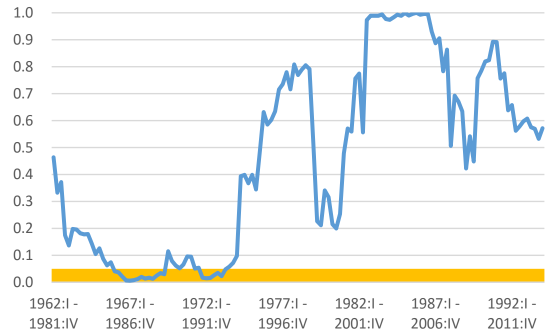
(d) LF Wald Test

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.

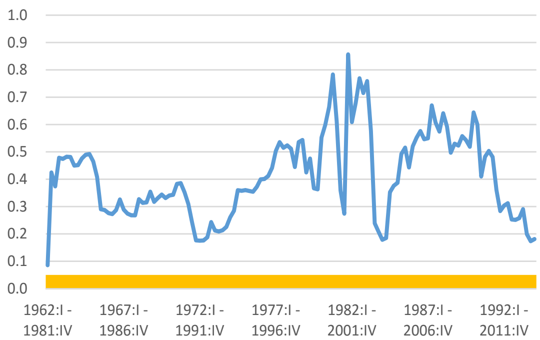
Figure 4: Rolling Window P-values for Tests of Non-Causality from GDP to Interest Rate Spread



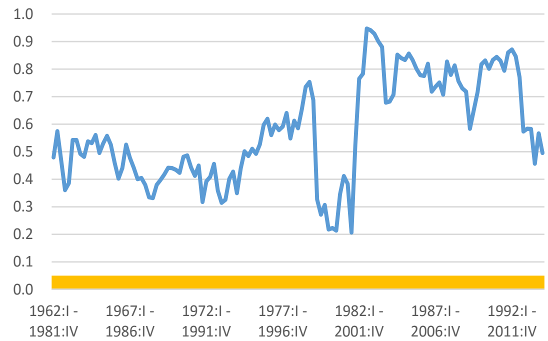
(a) MF Max Test (w/o MIDAS)



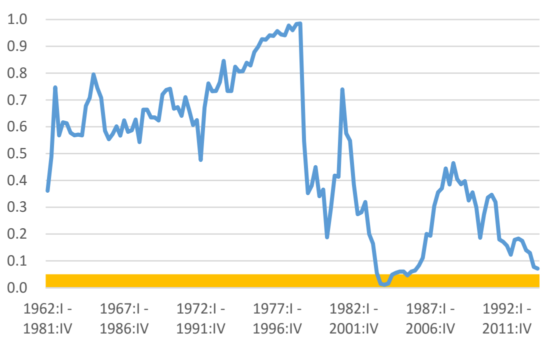
(b) MF Wald Test (w/o MIDAS)



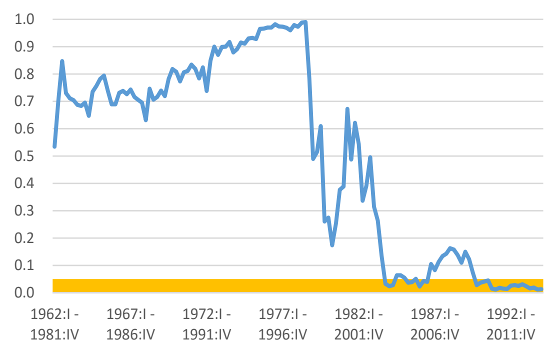
(c) MF Max Test (w/ MIDAS)



(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.