# Testing a Large Set of Zero Restrictions in Regression Models, with an Application to Mixed Frequency Granger Causality*

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

This Draft: March 10, 2017

**Abstract**

This paper proposes a test for a large set of zero restrictions in regression models based on a seemingly overlooked, but simple, dimension reduction technique. The procedure involves multiple parsimonious regression models where key regressors are split across simple regressions: each parsimonious model has one key regressor, and other regressors that are not associated with the null hypothesis. The test is based on the maximum key squared parameter among all parsimonious regressions. Parsimony ensures sharper estimates and therefore improves power in small samples. We present the general theory of the max test and focus on mixed frequency Granger causality as a prominent application since parameter proliferation is a major challenge in mixed frequency settings.

**Keywords:** Granger causality test, max test, Mixed Data Sampling (MIDAS), Sims test, temporal aggregation.

**JEL Classification:** C12, C22, C51.

*We thank Manfred Deistler, Atsushi Inoue, Helmut Lütkepohl, Massimiliano Marcellino, Yasumasa Matsuda, Daisuke Nagakura, Eric Renault, Peter Robinson, Peter R. Hansen, seminar participants at NYU Stern and Université Libre de Bruxelles, conference participants at the Recent Developments in Time Series and Related Fields, 2015 Japanese Joint Statistical Meeting, the 11th World Congress of the Econometric Society, the 11th International Symposium on Econometric Theory and Applications, the 25th $EC^2$ Conference - Advances in Forecasting, and the 2014 NBER-NSF Time Series Conference for helpful comments. The third author is grateful for financial supports of KAKENHI (Grant Number: 16K17104).

[†]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

[§]Graduate School of Economics, Kobe University. E-mail: motegi@econ.kobe-u.ac.jp

# 1 Introduction

We propose a new test designed for a large set of zero restrictions in regression models. The test is based on a seemingly overlooked, but simple, dimension reduction technique for regression models. Suppose that the underlying data generating process for some observed scalar $y_t$ is

$$y_t = z_t'a + x_t'b + \epsilon_t,$$

$z_t$ is assumed to have a small dimension while $x_t$ may have a large but finite dimension $h$. We want to test the null hypothesis $H_0 : b = 0$ against $H_1 : b \neq 0$.

A classic approach exploits what we call a *naïve regression model* $y_t = z_t'\alpha + x_t'\beta + u_t$ and computes a Wald statistic in order to test $H_0$. This approach may produce an imprecise result when the dimension of $b$ is large relative to sample size $n$. The asymptotic $\chi^2$-test may suffer from size distortions due to parameter proliferation. A bootstrap method can be employed to improve empirical size, but this generally results in the *size corrected* bootstrap test having low power due to large size distortions (cfr. Davidson and MacKinnon (2006)). A shrinkage estimator can be used, including Lasso, Adaptive Lasso, or Ridge Regression, but these are valid only under a sparsity assumption, and therefore we cannot test $H_0 : b = 0$ against a general alternative hypothesis $H_1 : b \neq 0$.

We propose splitting each of the key regressors $x_t = [x_{1t}, \ldots, x_{ht}]'$ across separate regression models. This results in what we call *parsimonious regression models*:

$$y_t = z_t'\alpha_i + \beta_i x_{it} + u_{it} \text{ for } i = 1, \ldots, h.$$

The $i^{th}$ parsimonious regression model has the $i^{th}$ element of $x_t$ only, so that parameter proliferation is not an issue. We then consider a max test statistic:

$$\hat{\mathcal{T}}_n = \max\{(\sqrt{n}\hat{\beta}_{n1})^2, \ldots, (\sqrt{n}\hat{\beta}_{nh})^2\}.$$

The asymptotic distribution of $\hat{\mathcal{T}}_n$ is non-standard under $H_0 : b = 0$, but an approximate p-value is readily available by drawing directly from an asymptotically valid approximation of the asymptotic distribution. Under $H_1 : b \neq 0$, at least one of $\{\hat{\beta}_{n1}, \ldots, \hat{\beta}_{nh}\}$ has a nonzero probability limit under fairly weak conditions. This key result ensures the consistency of the max test. The testing procedure method obviously cannot identify an entire vector $b$ when the null is false, but we can identify that the null is false asymptotically with probability approaching one for any direction $b \neq 0$ under the alternative.

The maximum of a sequence of 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). The use of a max statistic across econometric models is used in White's (2000) predictive model selection criterion. Evidently ours is the first application of a maximum statistic to a regression model as a core test statistic for zero restrictions.

After presenting the general theory of the max test, we focus on mixed frequency Granger causality as a prominent application which involves many zero restrictions. Time series are often sampled at different frequencies, and it is well known that temporal aggregation adversely affects Granger's (1969) notion of 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, see in particular Lütkepohl (1993), Dufour and Renault (1998), Dufour, Pelletier, and Renault (2006), and Hill (2007). Since standard VAR models are designed for single-frequency data, these tests often suffer from the adverse effect of temporal aggregation. 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.[2] Although Ghysels, Hill, and Motegi's (2016) tests avoid the undesirable effects of temporal aggregation, their applicability is limited because parameter proliferation in MF-VAR models makes the tests imprecise. 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 the correct size but low size-corrected power, a common occurrence when bootstrapping a size distorted asymptotic test (cfr. Davidson and MacKinnon (2006)).[3]

In order to circumvent the adverse impact of parameter proliferation on empirical power, we run parsimonious regressions and max tests for mixed frequency Granger causality. We consider a bivariate case with a high frequency variable $x_H$ and a low frequency variable $x_L$. Max tests can be applied to both causality from $x_H$ to $x_L$ (high-to-low causality) and causality from $x_L$ to $x_H$ (low-to-high causality). We compare the finite sample performance of the max test and Wald test in Monte Carlo simulations. We use mixed frequency [MF] regression models and aggregated low frequency [LF] models. We show that MF tests are better capable of detecting complex causal patterns than LF tests. 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.

As an empirical application, we analyze Granger causality between a weekly interest rate

[1]Existing Granger causality tests typically ignore this issue. They aggregate 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]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.

[3]Götz, Hecq, and Smeekes (2016), using an approach similar to Ghysels, Hill, and Motegi (2016), propose reduced rank regression and Bayesian estimation in order to run a large-dimensional MF-VAR. However, unlike Ghysels, Hill, and Motegi (2016), they do not consider a true high frequency process that governs the data, hence their focus is not on the causality mis-specification that arises from aggregation.

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.

The remainder of the paper is organized as follows. In Section 2, we present general theory of parsimonious regressions and max tests. We focus on mixed frequency Granger causality tests as a specific application in Section 3. In Section 4, we perform Monte Carlo simulations, an empirical application follows in Section 5, and Section 6 concludes the paper. Technical Appendices follow with omitted proofs and some technical details. See the supplemental material Ghysels, Hill, and Motegi (2017) for extra simulation results.

# 2 Methodology

Consider a data generating process in which a univariate time series $\{y_t\}$ depends linearly on two groups of regressors $z_t = [z_{1t}, \dots, z_{pt}]'$ and $x_t = [x_{1t}, \dots, x_{ht}]'$. Define the $\sigma$-field $\mathcal{F}_t = \sigma(Y_\tau : \tau \leq t)$ with all variables $Y_t = [y_t, X_t']'$ and all regressors $X_t = [z_t', x_t']'$.

**Assumption 2.1.** The true DGP is

$$y_t = \sum_{k=1}^{p} a_k z_{kt} + \sum_{i=1}^{h} b_i x_{it} + \epsilon_t. \tag{2.1}$$

The error $\{\epsilon_t\}$ is a stationary martingale difference sequence (mds) with respect to the increasing $\sigma$-field filtration $\mathcal{F}_t \subset \mathcal{F}_{t+1}$, and $\sigma^2 \equiv E[\epsilon_t^2] > 0$.

**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 $\sigma$-field $\mathcal{F}_t$. The mds assumption can be relaxed at the expense of more technical proofs, and a more complicated asymptotic variance structure and subsequent estimator. Since this is all well known, we do not consider such generalizations here.

Define an $n \times (p+h)$ matrix of regressors $X = [X_1, \dots, X_n]'$. The following rules out perfect multicollinearity in the regressors, a standard in the literature.

**Assumption 2.2.** $X$ is of full column rank $p+h$ *almost surely.*

We also impose a weak dependence property in order to ensure standard asymptotics. In the following, we assume that $Y_t = [y_t, X_t']'$ and $\epsilon_t$ are stationary $\alpha$-mixing.[4]

**Assumption 2.3.** $Y_t$ and $\epsilon_t$ are strictly stationary $\alpha$-mixing with mixing coefficients $\alpha_j$ that satisfy $\sum_{j=0}^{\infty} \alpha_{2^j} < \infty$.

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

**Remark 2.2.** The condition $\sum_{j=0}^{\infty} \alpha_{2^j} < \infty$ is quite general, allowing for geometric or hyperbolic memory decay in $\epsilon_t$, hence conditional volatility with a broad range of dynamics. We also assume the regressors $X_t$ and infinite order lag function $y_t$ of $\epsilon_t$ are mixing as a simplifying assumption, since underlying sufficient conditions for $y_t$ are rather technical if $\{\epsilon_t\}$ is a non-finite dependent process (see Chapter 2.3.2 in Doukhan (1994)).

Using standard vector notations, e.g., $a = [a_1, \ldots, a_p]'$, model (2.1) is rewritten as

$$y_t = z_t' a + x_t' b + \epsilon_t. \tag{2.2}$$

where $\{z_t\}$ serves as auxiliary regressors whose coefficients are not our main target. The number of those regressors, $p$, is assumed to be relatively small. We want to test for the zero restrictions with respect to main regressors $\{x_t\}$:

$$H_0 : b = 0. \tag{2.3}$$

The number of zero restrictions, $h$, is assumed to be large but finite, particularly in practice large relative to the sample size.

A classical approach of testing for $H_0 : b = 0$ is the Wald test based on what we call a *naïve regression model*:

$$y_t = \sum_{k=1}^{p} \alpha_k z_{kt} + \sum_{i=1}^{h} \beta_i x_{it} + u_t. \tag{2.4}$$

We assume that the model is correctly specified in order to focus ideas.

Based on (2.4), it is straightforward to compute a Wald statistic with respect to $H_0 : b = 0$. The statistic has an asymptotic $\chi^2$ distribution with $h$ degrees of freedom under Assumptions 2.1-2.3. A potential problem with this classic approach is that the asymptotic approximation may be poor when there are many zero restrictions relative to sample size $n$. A parametric or wild bootstrap can be used to control for the size of the test, but this typically leads to comparatively low size-corrected power. It is therefore of interest to propose a new test that achieves a sharper size and higher power when the Wald approach faces parameter proliferation.

In order to resolve the problem of high dimensional parameter restrictions, we propose *parsimonious regression models*:

$$y_t = \sum_{k=1}^{p} \alpha_{ki} z_{kt} + \beta_i x_{it} + u_{it}, \quad i = 1, \ldots, h. \tag{2.5}$$

There are therefore $h$ models and the key regressor $x_{it}$ along with the $z_{kt}'s$ appear in the $i^{th}$ model. In general the parameters $\alpha_{ki}$ may differ across model $i$, and unless the null is true, they are generally not equal to $a_{ki}$ in the true DGP 2.1. We fit least squares for each model to get $\hat{\beta}_{n1}, \ldots, \hat{\beta}_{nh}$. Then we formulate a *max-text statistic*

$$\hat{\mathcal{T}}_n = \max \left\{ (\sqrt{n}\hat{\beta}_{n1})^2, \ldots, (\sqrt{n}\hat{\beta}_{nh})^2 \right\}. \tag{2.6}$$

Equations (2.5) and (2.6) form a core part of our approach. The number of regressors in each parsimonious regression model is $p+1$, which is much smaller than $p+h$ in the naïve regression model (2.4). As a result, the precision of $\hat{\beta}_{ni}$ improves toward its probability limit for each $i$.

Under $H_0 : b = 0$, each parsimonious regression model is correctly specified and therefore $\hat{\beta}_{ni} \overset{p}{\to} \beta_i^* = 0$ straightforwardly. The asymptotic distribution of $\hat{\mathcal{T}}_n$ under $H_0$ is non-standard, but we can compute a simulation-based p-value by drawing from an asymptotically valid approximation to the asymptotic distribution directly (see Theorems 2.1-2.3). Under $H_1 : b \neq 0$, each parsimonious regression model is in general misspecified due to an omitted regressor, and therefore $\hat{\beta}_{ni} \overset{p}{\to} \beta_i^* \neq b_i$. A key result which is proven in Theorems 2.4-2.5 is that at least one of $\{\beta_1^*, \ldots, \beta_h^*\}$ must be nonzero under $H_1 : b \neq 0$. Hence the max test achieves consistency $\hat{\mathcal{T}}_n \overset{p}{\to} \infty$, although $\hat{\beta}_{ni}$ itself may not Fisher consistent for $b_i$.

We can potentially generalize (2.6) by adding a weight to each term:

$$\hat{\mathcal{T}}_n = \max\left\{(\sqrt{n} w_{n1} \hat{\beta}_{n1})^2, \ldots, (\sqrt{n} w_{nh} \hat{\beta}_{nh})^2\right\}, \tag{2.7}$$

where $\{w_{n1}, \ldots, w_{nh}\}$ is a sequence of possibly stochastic $L_2$-bounded positive scalar weights with non-random positive probability limits. We restrict ourselves to the equal weights $w_{ni} = 1$ in order to focus on the key implications from (2.5) and (2.6). Choosing non-equal weights is a possible future task.[5]

## 2.1 Asymptotics under the Null Hypothesis

We derive the asymptotic distribution of $\hat{\mathcal{T}}_n$ in (2.6) under $H_0 : b = 0$. Rewrite each parsimonious regression model (2.5) as

$$y_t = X_{it}' \theta_i + u_{it}, \quad i = 1, \ldots, h, \tag{2.8}$$

where

$$X_{it} = [z_{1t}, \ldots, z_{pt}, x_{it}]' \text{ and } \theta_i = [\alpha_i', \beta_i]' = [\alpha_{1i}, \ldots, \alpha_{pi}, \beta_i]'.$$

Stack all parameters across the $h$ models as $\theta = [\theta_1', \ldots, \theta_h']'$. Define a selection matrix $R$ that selects $\beta = [\beta_1, \ldots, \beta_h]'$ from $\theta$. $R$ is an $h \times (p+1)h$ full row rank matrix such that $\beta = R\theta$, hence

$$R = \begin{bmatrix} 0_{1\times p} & 1 & 0_{1\times p} & 0 & \ldots & 0_{1\times p} & 0 \\ 0_{1\times p} & 0 & 0_{1\times p} & 1 & \ldots & 0_{1\times p} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0_{1\times p} & 0 & 0_{1\times p} & 0 & \ldots & 0_{1\times p} & 1 \end{bmatrix}. \tag{2.9}$$

**Theorem 2.1.** Under the null hypothesis $H_0 : b = 0$, we have that $\hat{\mathcal{T}}_n \overset{d}{\to} \max\{\mathcal{N}_1^2, \ldots, \mathcal{N}_h^2\}$ as

[5] Logical weights include $w_{ni} = 1$, in which case we operate on the maximum (in absolute value) key parsimonious regression parameter estimator. Another obvious choice is the inverted standard error $\hat{\mathcal{V}}_{ni}^{-1/2}$ where $\hat{\mathcal{V}}_{n,i}$ is a consistent estimator the the asymptotic variance of $\sqrt{n}\hat{\beta}_{ni}$ under the null. This allows for control of different sampling and asymptotic dispersions of the estimators across parsimonious models, which may improve empirical and local power. For the sake of compactness, we do not consider such generality here.

$n \to \infty$, where $\mathcal{N} = [\mathcal{N}_1, \ldots, \mathcal{N}_h]'$ is distributed $N(0, V)$ with covariance matrix:

$$V = RSR' \in \mathbb{R}^{h \times h}, \tag{2.10}$$

where

$$\begin{aligned} &S = \begin{bmatrix} \Sigma_{11} & \ldots & \Sigma_{1h} \\ \vdots & \ddots & \vdots \\ \Sigma_{h1} & \ldots & \Sigma_{hh} \end{bmatrix} \in \mathbb{R}^{(p+1)h \times (p+1)h}, \\ &\Gamma_{ij} = E[X_{it}X_{jt}'], \quad \Lambda_{ij} = E[\epsilon_t^2 X_{it}X_{jt}'], \quad \Sigma_{ij} = \Gamma_{ii}^{-1}\Lambda_{ij}\Gamma_{jj}^{-1}; \quad i, j \in \{1, \ldots, h\}. \end{aligned} \tag{2.11}$$

All proofs appear in the Appendix section A.

**Remark 2.3.** Under Assumption 2.3, the error $\epsilon_t$ is an adapted martingale difference. Suppose also that $\epsilon_t^2 - \sigma^2$ is an adapted martingale difference with $\sigma^2 = E[\epsilon_t^2]$ (i.e. conditional homoskedasticity). Then, we have the simplification $\Lambda_{ij} = \sigma^2\Gamma_{ij}$.

**Remark 2.4.** Each $X_{it}$ contains the same subset of regressors $z_t$. It is therefore easy to prove that $S$ is positive semi-definite and singular.[6]

**Remark 2.5.** We do not have a general proof that the key asymptotic covariance matrix $V$ is nonsingular, except for a simple case with $(p, h) = (1, 2)$. This is irrelevant for performing the max test since we do not invert $V$. Further, as we show below, we can easily bootstrap an asymptotically valid p-value by drawing from an asymptotically valid approximation to the distribution $N(0, V)$.

## 2.2 Simulated P-Value

While the max test statistic $\hat{\mathcal{T}}_n$ has a non-standard limit distribution under $H_0$, an approximate p-value is readily available by drawing from an asymptotically valid approximation to the limit distribution directly. Let $\hat{V}_n$ be a consistent estimator for $V$ (see Theorem 2.3, below), and draw $M$ samples of vectors $\{\mathcal{N}^{(1)}, \ldots, \mathcal{N}^{(M)}\}$ independently from $N(0, \hat{V}_n)$. Now compute artificial test statistics $\hat{\mathcal{T}}_n^{(j)} = \max\{(\mathcal{N}_1^{(j)})^2, \ldots, (\mathcal{N}_h^{(j)})^2\}$ for $j = 1, \ldots, M$. An asymptotic p-value approximation for $\hat{\mathcal{T}}_n$ is

$$\hat{p}_{n,M} = \frac{1}{M}\sum_{j=1}^{M} I\left(\hat{\mathcal{T}}_n^{(j)} > \hat{\mathcal{T}}_n\right). \tag{2.12}$$

Since $\mathcal{N}^{(j)}$ are i.i.d. over $j$, and $M$ can be made arbitrarily large, by the Glivenko-Cantelli Theorem $\hat{p}_{n,M}$ can be made arbitrarily close to $P(\hat{\mathcal{T}}_n^{(1)} > \hat{\mathcal{T}}_n)$. The proposed max test is to

[6] Observe that $\Sigma_{ij} = E[\epsilon_t^2 \tilde{X}_{it}\tilde{X}_{jt}']$, where $\tilde{X}_{it} = E\left[X_{it}X_{it}'\right]^{-1} X_{it}$. Now define the set of all (non-unique) regressors across all parsimonious regression models: $\tilde{X}_t = [\tilde{X}_{1t}', ..., \tilde{X}_{ht}']' \in \mathbb{R}^{(p+1)h}$. Let $\lambda = [\lambda_1', ..., \lambda_h']'$ where $\lambda_i = [\lambda_{ij}]_{j=1}^{p+1}$ is $(p+1) \times 1$ and $\lambda'\lambda = 1$. Then $\lambda' S \lambda = \sum_{i=1}^{h}\sum_{j=1}^{h} E[\epsilon_t^2 \lambda_i' \tilde{X}_{it}\tilde{X}_{jt}'\lambda_j] = E[\epsilon_t^2(\lambda'\tilde{X}_t)^2] \geq 0$. Because each $\tilde{X}_{it}$ contains $z_t$ it is possible to find a non-zero $\lambda$ such that $\lambda'\tilde{X}_t = 0$, e.g. $\lambda_{1p+1} = 0$, $\lambda_2 = -\lambda_1$, and all other $\lambda_i = 0$. Therefore $S$ is singular and positive semi-definite.

reject $H_0$ at level $\alpha$ when $\hat{p}_{n,M_n} < \alpha$, where $\{M_n\}_{n\geq 1}$ is a sequence of positive integers that satisfies $M_n \to \infty$.

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

**Theorem 2.2.** Let $\{M_n\}_{n\geq 1}$ be a sequence of positive integers, $M_n \to \infty$. Under Assumptions 2.1-2.3 $P(\hat{\mathcal{T}}_n^{(1)} > \hat{\mathcal{T}}_n) = \bar{F}^0(\hat{\mathcal{T}}_n) + o_p(1)$, hence $\hat{p}_{n,n} = \bar{F}^0(\hat{\mathcal{T}}_n) + o_p(1)$. Therefore under $H_0$, $P(\hat{p}_{n,M_n} < \alpha) \to \alpha$ for any $\alpha \in (0,1)$.

A consistent estimator $\hat{V}_n$ for $V$ in (2.10) is computed as follows. Run least squares for each parsimonious regression model to get $\hat{\theta}_{ni} = [\hat{\alpha}_{ni}', \hat{\beta}_{ni}]'$ and residuals $\hat{u}_{it} = y_t - X_{it}'\hat{\theta}_{ni}$. Define $\hat{\Gamma}_{ij} = (1/n)\sum_{t=1}^n X_{it}X_{jt}'$, $\hat{\Lambda}_{ij} = (1/n)\sum_{t=1}^n \hat{u}_{it}^2 X_{it}X_{jt}'$, $\hat{\Sigma}_{ij} = \hat{\Gamma}_{ii}^{-1}\hat{\Lambda}_{ij}\hat{\Gamma}_{jj}^{-1}$, $\hat{S} = [\hat{\Sigma}_{ij}]_{i,j}$, and

$$\hat{V}_n = R\hat{S}R'. \tag{2.13}$$

**Theorem 2.3.** Under Assumptions 2.1-2.3, $\hat{V}_n \xrightarrow{p} \bar{V}$ where $\bar{V}$ is some matrix that satisfies $||\bar{V}|| < \infty$. Moreover, $\bar{V} = V$ under $H_0$.

### 2.3 Identification of the Null and Alternative Hypotheses

Under the alternative $H_1 : b \neq 0$, $\hat{\beta}_{ni}$ is in general not Fisher consistent for the true $b_i$ due to omitted regressors. Let $\beta_i^* = \text{plim}_{n\to\infty}\hat{\beta}_{ni}$ be the so-called pseudo-true value of $\beta_i$. The same notation applies to $\alpha_i^*$ and $\theta_i^* = [\alpha_i^{*\prime}, \beta_i^*]'$. We can characterize $\theta_i^*$ as follows.

**Theorem 2.4.** Let Assumptions 2.1-2.3 hold. Let $\Gamma_{ii} = E[X_{it}X_{it}'] \in \mathbb{R}^{(p+1)\times(p+1)}$ and $C_i = E[X_{it}x_t'] \in \mathbb{R}^{(p+1)\times h}$. Then, $\hat{\theta}_n \xrightarrow{p} \theta^* = [\theta_1^{*\prime}, \ldots, \theta_h^{*\prime}]'$ where:

$$\theta_i^* = \begin{bmatrix} \alpha_{1i}^* \\ \vdots \\ \alpha_{pi}^* \\ \beta_i^* \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_p \\ 0 \end{bmatrix} + \Gamma_{ii}^{-1}C_i b, \quad i = 1, \ldots, h. \tag{2.14}$$

Therefore, $\hat{\beta}_n \xrightarrow{p} \beta^* = R\theta^*$ by construction.

Theorem 2.4 provides useful insights on the relationship between the underlying coefficient $b$ and the pseudo-true value $\beta^*$. First, it is clear from (2.14) that $\beta^* = 0$ whenever $b = 0$. This is an intuitive result since each parsimonious regression model is correctly specified under $H_0 : b = 0$. Second, as the next result proves, $b = 0$ whenever $\beta^* = 0$. This is a useful result that allows us to identify the null and alternative hypotheses exactly. Of course, our approach cannot identify *all* of $\{b_1, \ldots, b_h\}$ under $H_1$. We can, however, identify that *at least one* of $\{b_1, \ldots, b_h\}$ must be non-zero, which is sufficient for rejecting $H_0 : b = 0$.

**Theorem 2.5.** Let Assumptions 2.1-2.3 hold. Then $\beta^* = 0$ implies $b = 0$, hence $\beta^* = 0$ *if and only if* $b = 0$. Therefore $\hat{\beta}_n \xrightarrow{p} 0$ *if and only if* $b = 0$.

A proof of Theorem 2.5 exploits the non-singularity of $E[z_t z_t']$ and $E[x_t x_t']$, which is ensured by Assumption 2.2. We next present a simple example which illustrates Theorem 2.5.

**Example 2.1** (Identification)**.** Suppose that the true DGP is

$$y_t = a_1 + b_1 x_{1t} + b_2 x_{2t} + \epsilon_t$$

and we run two parsimonious regression models

$$y_t = \alpha_{11} + \beta_1 x_{1t} + u_{1t} \quad \text{and} \quad y_t = \alpha_{12} + \beta_2 x_{2t} + u_{2t}.$$

Notice that we have only one common regressor $z_{1t} = 1$ and two main regressors $\{x_{1t}, x_{2t}\}$. Assume for simplicity that $E[x_{it}] = 0$ and $E[x_{it}^2] = 1$ for $i = 1, 2$. Finally, since Assumption 2.2 rules out perfect multicollinearity, we have that $\rho_{12} = E[x_{1t}x_{2t}] \in (-1, 1)$.

In this setting, it follows that $X_{it} = [1, x_{it}]'$, $x_t = [x_{1t}, x_{2t}]'$, and

$$\Gamma_{ii} = E[X_{it}X_{it}'] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad C_1 = E[X_{1t}x_t'] = \begin{bmatrix} 0 & 0 \\ 1 & \rho_{12} \end{bmatrix}, \quad C_2 = E[X_{2t}x_t'] = \begin{bmatrix} 0 & 0 \\ \rho_{12} & 1 \end{bmatrix}.$$

Substitute these quantities into (2.14) to get

$$\beta_1^* = b_1 + \rho_{12} b_2 \quad \text{and} \quad \beta_2^* = b_2 + \rho_{12} b_1 \text{ where } \rho_{12} = E[x_{1t}x_{2t}]. \tag{2.15}$$

Now we verify Theorem 2.5 by showing that $\beta^* = 0 \Longrightarrow b = 0$. Assume $\beta_1^* = \beta_2^* = 0$. If $\rho_{12} = 0$, then it is trivial from (2.15) that $b_1 = b_2 = 0$. However, $\rho_{12} \neq 0$, (2.15) implies that

$$\frac{(1 + \rho_{12})(1 - \rho_{12})}{\rho_{12}} \times b_1 = 0.$$

Since $\rho_{12} \in (-1, 1)$, it must be the case that $b_1 = 0$ and therefore $b_2 = 0$.

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 (2.6) indicates that $\hat{\mathcal{T}}_n \xrightarrow{p} \infty$ *if and only if* $\beta^* \neq 0$, and by Theorems 2.4 and 2.5 $\hat{\beta}_n \xrightarrow{p} \beta^* \neq 0$ under a general alternative hypothesis $H_1 : b \neq 0$. In conjunction with p-value approximation consistency Theorem 2.2, this proves consistency of the max test.

**Theorem 2.6.** Let Assumptions 2.1-2.3 hold, then $\hat{\mathcal{T}}_n \xrightarrow{p} \infty$ and therefore $P(\hat{p}_{n,M_n} < \alpha) \to 1$ for any $\alpha \in (0, 1)$ *if and only if* $H_1 : b \neq 0$ is true.

An immediate consequence of the limit distribution Theorem 2.1, p-value approximation consistency Theorem 2.2, 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 $\{M_n\}_{n\geq 1}$ be a sequence of positive constants, $M_n \to \infty$. Let Assumptions 2.1-2.3 hold. Then $\hat{\mathcal{T}}_n \xrightarrow{d} \max\{\mathcal{N}_1^2, \ldots, \mathcal{N}_h^2\}$ as $n \to \infty$ and therefore $P(\hat{p}_{n,M_n} < \alpha) \to \alpha$ for any $\alpha \in (0,1)$ *if and only if* $H_0 : b = 0$ is true.

To conclude this section, we briefly discuss a misspecification problem. Consider the case where a true DGP has $h$ regressors $x_{1t}, \ldots, x_{ht}$ while we run only $\tilde{h}$ parsimonious regression models with $\tilde{h} < h$. In such a case, the max test is generally inconsistent, just as is the Wald test. Observe that (2.14) still holds for $i = 1, \ldots, \tilde{h}$ when there is underspecification. (See the proof of Theorem 2.4 for derivation.) When $\tilde{h} < h$, having $\beta_1^* = \cdots = \beta_{\tilde{h}}^* = 0$ does not imply that $b = 0$. Below is a simple counter-example.

**Example 2.2** (Non-identification due to Underspecification)**.** Continue Example 2.1 except for that $\tilde{h} = 1$ now – we run only one regression model $y_t = \alpha_{11} + \beta_1 x_{1t} + u_{1t}$. In this simple set-up, the parsimonious regression approach is identical to the naïve regression approach. By (2.15), $\beta_1^* = b_1 + \rho_{12} b_2$. For any given $\rho_{12} \in (-1, 1)$, having $\beta_1^* = 0$ does not imply $b_1 = b_2 = 0$ because $(b_1, b_2)$ can take any value as long as $b_1 = -\rho_{12} b_2$.

# 3 Mixed Frequency Granger Causality

In this section, we focus on mixed frequency Granger causality tests in the setting of Ghysels, Hill, and Motegi (2016). Testing for Granger causality with mixed frequency data is a prominent example of testing for many zero restrictions. We restrict ourselves to a bivariate case where we have a high frequency variable $x_H$ and a low frequency variable $x_L$.[7]

Following the notation of Ghysels, Hill, and Motegi (2016), we first formulate a DGP for $x_H$ and $x_L$. Let $m$ denote the *ratio of sampling frequencies*, i.e. the number of high frequency time periods in 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. All of our main results would carry over to time-varying sequences $m(\tau_L)$ in a straightforward way, e.g. when $x_H$ is daily and $x_L$ is monthly.

**Example 3.1** (Mixed Frequency Data - Quarterly and Monthly)**.** A simple example of mixed frequency data is when we analyze a monthly variable $x_H$ and a quarterly variable $x_L$, hence $m = 3$. Suppose that $x_H(\tau_L, 1)$ is the first monthly observation in quarter $\tau_L$, $x_H(\tau_L, 2)$ 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 $\tau_L$, denoted $[x_H(\tau_L, 1),\ x_H(\tau_L, 2),\ x_H(\tau_L, 3)]'$. The resulting stacked system 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.

[7] 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.

In the bivariate case, 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 $\sigma$-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 $p \geq 1$.[8]

**Assumption 3.1.** The mixed frequency vector $\boldsymbol{X}(\tau_L)$ is governed by a MF-VAR($p$) for finite $p \geq 1$:

$$\underbrace{\begin{bmatrix} x_H(\tau_L, 1) \\ \vdots \\ x_H(\tau_L, m) \\ x_L(\tau_L) \end{bmatrix}}_{\equiv \boldsymbol{X}(\tau_L)} = \sum_{k=1}^{p} \underbrace{\begin{bmatrix} d_{11,k} & \ldots & d_{1m,k} & c_{(k-1)m+1} \\ \vdots & \ddots & \vdots & \vdots \\ d_{m1,k} & \ldots & d_{mm,k} & c_{km} \\ b_{km} & \ldots & b_{(k-1)m+1} & a_k \end{bmatrix}}_{\equiv \boldsymbol{A}_k} \underbrace{\begin{bmatrix} x_H(\tau_L - k, 1) \\ \vdots \\ x_H(\tau_L - k, m) \\ x_L(\tau_L - k) \end{bmatrix}}_{\equiv \boldsymbol{X}(\tau_L - k)} + \underbrace{\begin{bmatrix} \epsilon_H(\tau_L, 1) \\ \vdots \\ \epsilon_H(\tau_L, m) \\ \epsilon_L(\tau_L) \end{bmatrix}}_{\equiv \boldsymbol{\epsilon}(\tau_L)} \tag{3.1}$$

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 3.1.** A constant term is omitted from (3.1) 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 (3.1) 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, $b_j$ represents the impact of $x_H$ on $x_L$ with $j$ high frequency lags.

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 $x_H$ with $j$ high frequency lags.

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 $\alpha$-mixing.

**Assumption 3.2.** 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.

[8]Complete details on the mixed frequency notations are presented in Appendix B.

**Assumption 3.3.** $\boldsymbol{X}(\tau_L)$ and $\boldsymbol{\epsilon}(\tau_L)$ are $\alpha$-mixing with mixing coefficients $\alpha_h$ that satisfy $\sum_{h=0}^{\infty} \alpha_{2^h} < \infty$.

**Remark 3.2.** Note that $\boldsymbol{\Omega} \equiv E[\boldsymbol{\epsilon}(\tau_L)\boldsymbol{\epsilon}(\tau_L)']$ allows for the high frequency innovations $\epsilon_H(\tau_L, i)$ to have a different variance for each $i$. Therefore, while Assumptions 3.1 and 3.2 imply $\{x_H(\tau_L, i)\}_{\tau_L}$ is covariance stationary *for each fixed* $i \in \{1, \ldots, m\}$, they do not imply covariance stationarity for *the entire high frequency array* $\{\{x_H(\tau_L, i)\}_{i=1}^m\}_{\tau_L}$.

Granger causality from $x_H$ to $x_L$ is what we call *high-to-low causality.* Granger causality from $x_L$ to $x_H$ is what we call *low-to-high causality.* Since there are fundamentally different challenges when testing for high-to-low causality and low-to-high causality, we treat the former in Section 3.1, and treat the latter in Section 3.2.

## 3.1 High-to-Low Frequency Data Granger Causality

Pick the last row of the entire system (3.1):

$$\begin{aligned} x_L(\tau_L) &= \sum_{k=1}^{p} a_k x_L(\tau_L - k) + \sum_{i=1}^{pm} b_i x_H(\tau_L - 1, m + 1 - i) + \epsilon_L(\tau_L), \\ \epsilon_L(\tau_L) &\overset{mds}{\sim} (0, \sigma_L^2), \quad \sigma_L^2 > 0. \end{aligned} \tag{3.2}$$

The index $i \in \{1, \ldots, pm\}$ is in high frequency terms, and the second argument $m + 1 - i$ of $x_H$ can be less than 1 since $i > m$ occurs when $p > 1$. Allowing any integer value in the second argument of $x_H$, including those smaller than 1 or larger than $m$, does not cause any confusion, and simplifies analytical arguments below. It is understood, for example, 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 $x_H(\tau_L, m + 1) = x_H(\tau_L + 1, 1)$. More generally, we interchangeably write $x_H(\tau_L - \tau, i) = x_H(\tau_L, i - m\tau)$ for any $\tau_L, \tau, i \in \mathbb{Z}$.

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*

$$H_0 : b_1 = \cdots = b_{pm} = 0.$$

A well-known issue here is that the number of zero restrictions, $pm$, may be quite large in some applications, depending on the ratio of sampling frequencies $m$. Consider a weekly versus quarterly data case for instance.[9] The MF-VAR lag length $p$ is in terms of quarters and $m = 12$ approximately. Then $pm = 36$ when $p = 3$, and $pm = 48$ when $p = 4$, etc. In order to deal with the many zero restrictions arising from a large $m$, it is of use to frame the problem in terms of parsimonious regression models and perform a max test.

There is a clear correspondence between the general linear DGP (2.1) and the mixed frequency DGP (3.2). The regressand $y_t$ is $x_L(\tau_L)$; common regressors $\{z_{1t}, \ldots, z_{pt}\}$ are identi-

[9] In Section 5, we analyze Granger causality between quarterly GDP and weekly interest rate spread in the U.S.

cally the low frequency $\{x_L(\tau_L - 1), \ldots, x_L(\tau_L - p)\}$; and the main regressors that are split into parsimonious regression models $\{x_{1t}, \ldots, x_{ht}\}$ are the high frequency $\{x_H(\tau_L - 1, m + 1 - 1), \ldots, x_H(\tau_L - 1, m + 1 - pm)\}$. The parsimonious regression models are therefore:

$$x_L(\tau_L) = \sum_{k=1}^{p} \alpha_{k,i} x_L(\tau_L - k) + \beta_i x_H(\tau_L - 1, m + 1 - i) + u_{L,i}(\tau_L), \quad i = 1, \ldots, pm. \tag{3.3}$$

Here we are using $h = pm$ models, which matches the true high frequency lag length. If $h < pm$, then there is underspecification and the max test loses consistency (cfr. Example 2.2). If $h > pm$, then consistency is ensured although finite sample performance may be poorer due to redundant regressors.

Assumptions 3.1-3.3 imply Assumptions 2.1-2.3. Thus, under Assumptions 3.1-3.3, Theorems 2.1-2.6 and Corollary 2.7 carry over to a high-to-low causality max test.

### 3.2 Low-to-High Frequency Data Granger Causality

We now consider testing for Granger causality from the low frequency variable $x_L$ to the high frequency variable $x_H$. Recall from (3.1) that the DGP is MF-VAR($p$). As defined in Ghysels, Hill, and Motegi (2016), the null hypothesis of low-to-high non-causality is written as $H_0 : c_1 = \cdots = c_{pm} = 0$ or compactly $\boldsymbol{c} = [c_1, \ldots, c_{pm}]' = \mathbf{0}_{pm\times 1}$.

In order to account for the difficulty of explaining the stacked high frequency dependent variables $[x_H(\tau_L, 1), \ldots, x_H(\tau_L, m)]'$ with the lagged low frequency regressors $x_L(\tau_L - k)$ in (3.1), and to utilize parsimonious regression models, we exploit Sims' (1972) two-sided regression model. This can be extended to a mixed frequency case in a straightforward way:

$$x_L(\tau_L) = \sum_{k=1}^{p} \alpha_k x_L(\tau_L - k) + \sum_{j=1}^{pm} \beta_j x_H(\tau_L - 1, m + 1 - j) + \sum_{i=1}^{r} \gamma_i x_H(\tau_L + 1, i) + u_L(\tau_L). \tag{3.4}$$

Model (3.4) regresses $x_L$ onto $p$ low frequency lags of $x_L$, $pm$ high frequency lags of $x_H$, and $r \geq 1$ high frequency *leads* of $x_H$. Low-to-high non-causality $H_0 : \boldsymbol{c} = \mathbf{0}_{pm\times 1}$ from (3.1) implies $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_r]' = \mathbf{0}_{r\times 1}$ from (3.4).

Model (3.4) can be thought of as a naïve regression model in the sense that all leads of $x_H$ are included in one model. We therefore propose parsimonious regression models

$$x_L(\tau_L) = \sum_{k=1}^{p} \alpha_{k,i} x_L(\tau_L - k) + \sum_{j=1}^{pm} \beta_{ji} x_H(\tau_L - 1, m + 1 - j) + \gamma_i x_H(\tau_L + 1, i) + u_{L,i}(\tau_L), \quad i = 1, \ldots, r. \tag{3.5}$$

Note that the $i^{th}$ parsimonious regression model has only the $i^{th}$ high frequency lead of $x_H$. We take the lagged low and high frequency $\{x_L(\tau_L - 1), \ldots, x_L(\tau_L - p), x_H(\tau_L - 1, m + 1 - 1), \ldots, x_H(\tau_L - 1, m + 1 - pm)\}$ as the common regressors, and take the lead high frequency $\{x_H(\tau_L + 1, 1), \ldots, x_H(\tau_L + 1, r)\}$ as the main regressors that are split into each parsimonious regression model.

Under non-causality, Assumptions 2.1-2.3 hold and hence Theorems 2.1-2.3 carry over. Under $H_1 : \boldsymbol{c} \neq \mathbf{0}_{r\times 1}$, there does not exist a clear characterization of a pseudo-true value $\gamma_i^*$. Proving consistency is therefore an open question.

### 3.2.1 MIDAS Polynomials in the Max Test

In a low-to-high frequency causality test, the max statistic only operates on the *lead* parameters $\gamma_i$, while our simulation study reveals a large *lag* length $pm$ 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.[10]

One option is to use a bootstrap procedure for p-value computation, but we find that a wild bootstrap similar to Gonçalves and Kilian'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 $pm$, and keep the low-to-high causality part unrestricted (cfr. Ghysels, Santa-Clara, and Valkanov (2006), Ghysels, Sinko, and Valkanov (2007), among others). Modified parsimonious models are

$$x_L(\tau_L) = \sum_{k=1}^{p} \alpha_{k,i} x_L(\tau_L - k) + \sum_{j=1}^{pm} \omega_j(\boldsymbol{\pi}_i) x_H(\tau_L - 1, m + 1 - j) + \gamma_i x_H(\tau_L + 1, i) + u_{L,i}(\tau_L), \quad i = 1, \ldots, r, \tag{3.6}$$

where $\omega_j(\boldsymbol{\pi}_i)$ represents a MIDAS polynomial with a parameter vector $\boldsymbol{\pi}_i \in \mathbb{R}^s$ of small dimension $s \ll pm$.

Various software packages including the MIDAS Matlab Toolbox (Ghysels (2013)), the R Package *midasr* (Ghysels, Kvedaras, and Zemlys (2016)), EViews and Gretl cover a variety of polynomial specifications. In our simulation study we use the Almon polynomial $\omega_j(\boldsymbol{\pi}) = \sum_{l=1}^{s} \pi_l j^l$, hence model (3.6) 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. $w_j(\boldsymbol{\pi}) \geq 0$ for $j < 3$ and $w_j(\boldsymbol{\pi}) < 0$ for $j \geq 4$, 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 10).

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

# 4 Monte Carlo Simulations

In this section, we perform Monte Carlo simulations in order to compare max tests and Wald tests in finite sample. We consider a mixed frequency environment because it naturally lends to parameter proliferation. We begin with a MF-VAR data generating process and then fit Granger causality tests based on the max approach and the Wald approach. We then aggregate the simulated mixed frequency data into low frequency and again fit causality tests based on both approaches, allowing for a direct comparison between mixed frequency (MF) and the traditional low frequency (LF) methods. We discuss high-to-low causality in Section 4.1 and low-to-high causality in Section 4.2.

## 4.1 High-to-Low Granger Causality

We first take a MF-VAR(1) as a benchmark DGP, and then work with a MF-VAR(2) for a robustness check.

### 4.1.1 MF-VAR(1)

**Data Generating Process** We work with the following structural MF-VAR(1) process with $m = 12$:

$$\underbrace{\begin{bmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ -d & 1 & \ddots & \ddots & \ddots & 0 \\ 0 & -d & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & -d & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 \end{bmatrix}}_{=\boldsymbol{N}} \underbrace{\begin{bmatrix} x_H(\tau_L,1) \\ \vdots \\ x_H(\tau_L,12) \\ x_L(\tau_L) \end{bmatrix}}_{=\boldsymbol{X}(\tau_L)} = \underbrace{\begin{bmatrix} 0 & 0 & \dots & d & c_1 \\ 0 & 0 & \dots & 0 & c_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & c_{12} \\ b_{12} & b_{11} & \dots & b_1 & a \end{bmatrix}}_{=\boldsymbol{M}} \underbrace{\begin{bmatrix} x_H(\tau_L-1,1) \\ \vdots \\ x_H(\tau_L-1,12) \\ x_L(\tau_L-1) \end{bmatrix}}_{=\boldsymbol{X}(\tau_L-1)} + \underbrace{\begin{bmatrix} \eta_H(\tau_L,1) \\ \vdots \\ \eta_H(\tau_L,12) \\ \eta_L(\tau_L) \end{bmatrix}}_{=\boldsymbol{\eta}(\tau_L)}, \quad (4.1)$$

or compactly $\boldsymbol{N}\boldsymbol{X}(\tau_L) = \boldsymbol{M}\boldsymbol{X}(\tau_L-1)+\boldsymbol{\eta}(\tau_L)$. Setting $m = 12$ occurs in practice for, e.g., yearly low frequency increment with monthly high frequency increment, or quarterly low frequency increment with weekly high frequency increment approximately.

Coefficient $a$ governs the autoregressive property of $x_L$, $d$ governs the autoregressive property of $x_H$, $\boldsymbol{c} = [c_1, \dots, c_{12}]'$ represents Granger causality from $x_L$ to $x_H$, and our interest lies in $\boldsymbol{b} = [b_1, \dots, b_{12}]'$ since it expresses Granger causality from $x_H$ to $x_L$. Since

$$
\boldsymbol{N}^{-1} = \begin{bmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ d & 1 & \ddots & \ddots & \ddots & 0 \\ d^2 & d & \ddots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ d^{11} & d^{10} & \dots & d & 1 & 0 \\ 0 & 0 & \dots & 0 & 0 & 1 \end{bmatrix} \text{ therefore } \boldsymbol{A} \equiv \boldsymbol{N}^{-1}\boldsymbol{M} = \begin{bmatrix} 0 & 0 & \dots & d & \sum_{i=1}^{1} d^{1-i} c_i \\ 0 & 0 & \dots & d^2 & \sum_{i=1}^{2} d^{2-i} c_i \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & d^{12} & \sum_{i=1}^{12} d^{12-i} c_i \\ b_{12} & b_{11} & \dots & b_1 & a \end{bmatrix}, \tag{4.2}
$$

the reduced form of (4.1) 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 non-causality $\boldsymbol{b} = \boldsymbol{0}_{12\times 1}$ and four causal patterns. The first causal pattern is *decaying causality* with alternating signs: $b_j = (-1)^{j-1} \times 0.3/j$ for $j = 1, \dots, 12$. 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 $b_j = 0$. Such a relationship may exist in macroeconomic processes due to lagged information transmission, seasonality, feedback effects, and ambiguous theoretical relations in terms of signs. The fourth is *uniform causality*: $b_j = 0.02$ for all $j$.

We 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.[11] 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 $j = 1, \dots, 12$.

The structural error $\boldsymbol{\eta}(\tau_L)$ is either i.i.d. or a GARCH process. Let $\boldsymbol{\xi}(\tau_L) \overset{i.i.d.}{\sim} N(\boldsymbol{0}_{13\times 1}, \boldsymbol{I}_{13})$. 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 $\boldsymbol{H}(\tau_L)$ follows 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 $N(\boldsymbol{0}_{13\times 1}, \boldsymbol{\Omega}(\tau_L))$ 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

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

Sample size in terms of low frequency is $n \in \{80, 160\}$. Since $m = 12$, our experimental design can approximately be thought as week versus quarter, matching our empirical applications in Section 5. Hence $n = 80$ or $160$ implies that the low frequency sample size is 20 or 40 years.

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

**Model Estimation** The mixed frequency naïve regression model is

$$x_L(\tau_L) = \sum_{k=1}^{q} \alpha_k x_L(\tau_L - k) + \sum_{i=1}^{h_{MF}} \beta_i x_H(\tau_L - 1, m + 1 - i) + u_L(\tau_L). \tag{4.3}$$

Given DGP (4.1), the true lag length is $q = p = 1$ and $h_{MF} = pm = 12$. In this experiment, we consider $q = 2$ and $h_{MF} \in \{4, 8, 12, 24\}$ for comparison. The mixed frequency parsimonious regression models are

$$x_L(\tau_L) = \sum_{k=1}^{q} \alpha_{ki} x_L(\tau_L - k) + \beta_i x_H(\tau_L - 1, m + 1 - i) + u_{L,i}(\tau_L), \quad i = 1, \ldots, h_{MF}.$$

In order to perform tests at a common low frequency, we also perform flow aggregation $x_H(\tau_L) = (1/m) \sum_{j=1}^{m} x_H(\tau_L, j)$ and stock aggregation $x_H(\tau_L) = x_H(\tau_L, m)$. A low frequency naïve regression model is

$$x_L(\tau_L) = \sum_{k=1}^{q} \alpha_k x_L(\tau_L - k) + \sum_{i=1}^{h_{LF}} \beta_i x_H(\tau_L - i) + u_L(\tau_L).$$

When a MF-VAR is aggregated into a LF-VAR, the resulting lag length is infinite in general (cfr. Ghysels, Hill, and Motegi (2016)). We consider $q = 2$ and $h_{LF} \in \{1, 2, 3, 4\}$ for comparison. The relevant low frequency parsimonious regression models are

$$x_L(\tau_L) = \sum_{k=1}^{q} \alpha_{ki} x_L(\tau_L - k) + \beta_i x_H(\tau_L - i) + u_{L,i}(\tau_L), \quad i = 1, \ldots, h_{LF}.$$

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 heteroskedastic errors of unknown form. Their bootstrap p-value is computed with 1,000 bootstrap samples.[12]

In order to compute a max test approximate p-value, we draw 5,000 samples from an approximation to the limit distribution under $H_0$. That requires estimation of the key covariance matrix $V$ in (2.10). We compute a heteroscedasticity-robust estimator $\hat{V}_n$ using (2.13). In Tables T.5-T.8 of the supplemental material Ghysels, Hill, and Motegi (2017), we also use a simplified or non-robust estimator based on Remark 2.3. Bootstrapped Wald tests exhibit size distortions when (2.13) is used, hence we only use the non-robust covariance. Evidently the distortions arise

[12] Consider bootstrapping in the MF naïve regression model (4.3), the LF case being similar. Rewrite the model in matrix form as $x_L(\tau_L) = \boldsymbol{x}(\tau_L - 1)'\boldsymbol{\theta} + u_L(\tau_L)$, using standard vector notations. Let $\hat{\boldsymbol{\theta}}_n$ be an unrestricted least squares estimator for $\boldsymbol{\theta}$. Let $\hat{u}_L(\tau_L) = x_L(\tau_L) - \boldsymbol{x}(\tau_L - 1)'\hat{\boldsymbol{\theta}}_n$. Let $\hat{W}_n$ be a Wald test statistic with respect to $H_0 : \boldsymbol{\beta} = \mathbf{0}_{h_{MF} \times 1}$. Simulate $M = 1,000$ bootstrap samples from $x_L(\tau_L) = \boldsymbol{x}(\tau_L - 1)'\tilde{\boldsymbol{\theta}}_{n0} + \hat{u}_L(\tau_L)v(\tau_L)$, where $\tilde{\boldsymbol{\theta}}_{n0}$ is $\hat{\boldsymbol{\theta}}_n$ with the null hypothesis of non-causality imposed, and $v(\tau_L) \overset{i.i.d.}{\sim} N(0, 1)$. Compute bootstrapped Wald statistics $\tilde{W}_{n1}, \ldots, \tilde{W}_{nM}$ for each sample. The bootstrapped p-value is $p_M = (1/M) \times \sum_{j=1}^{M} I(\tilde{W}_{nj} \geq \hat{W}_n)$.

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.[13]

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 1 compiles simulation results for the case of GARCH errors. Results for the i.i.d. error are collected in Table T.1 of Ghysels, Hill, and Motegi (2017) 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, MF tests are better than LF tests in terms of detecting complicated causal patterns like sporadic causality. See Panel D.1.2 for example, where there is sporadic causality with $d = 0.2$, $n = 160$, and $h_{MF} = 24$. Empirical power is .721 for the MF max test and .644 for the MF Wald test. Empirical power of LF tests is less than .100 regardless of test types, aggregation schemes, and $h_{LF}$.

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 $n$, max and Wald tests have similar power. An advantage of the max test is highlighted under lagged causality with $d = 0.2$. When $n = 160$, the max test power is .763 for $h_{MF} = 12$ and .685 for $h_{MF} = 24$ (see Panel C.1.2). 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 facie* that such a feature should disadvantage 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. In the latter i.i.d. environment when there is also high persistence in $x_H$, the max-test generally performs better in MF and LF (flow) cases. In some cases the difference is quite stark: see Table T.1 in Ghysels, Hill, and Motegi (2017). Thus, when causality is more easily detected, the max test offers an advantage garnered directly from its parsimonious use of model parameters. The Wald test does not dominate in any case precisely because the statistic uses all parameter estimates from model (4.3), hence greater dispersion exists in the parameter estimates and therefore the Wald statistic.

[13] Results for the bootstrapped Wald test with robust covariance matrix (2.13) are available upon request.

### 4.1.2 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 (4.1), the extra coefficient $\boldsymbol{M}_2$ is parameterized as

$$\boldsymbol{M}_2 = \begin{bmatrix} \mathbf{0}_{12\times1} & \dots & \mathbf{0}_{12\times1} & \mathbf{0}_{12\times1} \\ b_{24} & \dots & b_{13} & 0 \end{bmatrix}.$$

*Non-causality* is now expressed as $\boldsymbol{b} = \mathbf{0}_{24\times1}$; *decaying causality* is $b_j = (-1)^{j-1} \times 0.3/j$ for $j = 1, \dots, 24$; *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 in Section 4.1.1: $a = 0.2$; $d \in \{0.2, 0.8\}$; $c_j = (-1)^{j-1} \times 0.4/j$ for $j = 1, \dots, 12$; $q = 2$; $n \in \{80, 160\}$; 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 2. The i.i.d. error case is reported in Table T.3 of Ghysels, Hill, and Motegi (2017). 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 $n = 160$, the max test power is .682 whereas the Wald test power is .530. These results are consistent with the MF-VAR(1) scenario.

## 4.2 Low-to-High Granger Causality

We now focus on low-to-high causality $\boldsymbol{c} = [c_1, \dots, c_{12}]'$ in the structural MF-VAR(1) in (4.1) with $m = 12$.[14]

### 4.2.1 Design

We consider non-causality and the four causal patterns as above. In each case we need to be careful about how $\boldsymbol{c}$ is transferred to the upper-right block $[\sum_{i=1}^{1} d^{1-i} c_i, \ \dots, \ \sum_{i=1}^{12} d^{12-i} c_i]'$ of $\boldsymbol{A}_1$, the low-to-high causality pattern in the reduced form (4.2). For *non-causality* $\boldsymbol{c} = \mathbf{0}_{12\times1}$, the upper-right block of $\boldsymbol{A}_1$ is a null vector regardless of $d$, the AR(1) coefficient of $x_H$. A similar pattern arises for *decaying causality* $c_j = (-1)^{j-1} \times 0.3/j$ for $j = 1, \dots, 12$, 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.

[14] MF-VAR(2) cases are covered in Tables T.11, T.12, T.15, and T.16 of the supplemental material Ghysels, Hill, and Motegi (2017). See Tables T.11 and T.12 for bootstrapped Wald tests and max test with the robust covariance matrix. See Tables T.15 and T.16 for max test based on a simplified or non-robust covariance matrix estimator.

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 $n \in \{80, 160\}$. We implement the MF max test based on MF parsimonious regression model (3.5), and the bootstrapped Wald test based on the MF naïve regression model (3.4). For both tests, lags and leads of $x_H$ are taken from $h_{MF}, r_{MF} \in \{4, 8, 12, 24\}$ for comparison.

The max test exhibits size distortions when $h_{MF} = 24$, especially in the smaller sample $n = 80$. 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 (3.6), as an ad hoc attempt to tackle parameter proliferation. We use the Almon polynomial of dimension $s = 3$ (cfr. Section 3.2.1). In order to make a direct comparison with the Wald test, we also perform a Wald test based on the MF naïve regression model (3.4) with lags of $x_H$ replaced with the Almon polynomial.

We also perform the max test based on LF parsimonious regression models

$$x_L(\tau_L) = \alpha_{1i} x_L(\tau_L - 1) + \sum_{j=1}^{h_{LF}} \beta_{ji} x_H(\tau_L - j) + \gamma_i x_H(\tau_L + i) + u_{L,i}(\tau_L), \quad i = 1, \ldots, r_{LF}. \tag{4.4}$$

$r_{LF}$ and $h_{LF}$ are both 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 bootstrapped Wald test is based on a LF naïve regression model:

$$x_L(\tau_L) = \alpha_1 x_L(\tau_L - 1) + \sum_{j=1}^{h_{LF}} \beta_j x_H(\tau_L - j) + \sum_{i=1}^{r_{LF}} \gamma_i x_H(\tau_L + i) + u_L(\tau_L) \tag{4.5}$$

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

We use 5,000 draws from an asymptotically valid approximation of the asymptotic distribution to compute max test p-values. The robust covariance matrix estimator is used for max tests. (See Tables T.13 and T.14 of Ghysels, Hill, and Motegi (2017) for results with the simplified or non-robust covariance matrix estimator.) 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 Wald tests, and nominal size is 5%.

### 4.2.2 Results

Table 3 presents rejection frequencies for the GARCH error case. The i.i.d. case is covered in Table T.9 of Ghysels, Hill, and Motegi (2017): the two errors yield similar results in general. First, the MF max test without MIDAS polynomials exhibits size distortions when $n = 80$. 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 plausible result since, although mis-specified, the MIDAS approach promotes parsimony which aids in achieving sharper empirical size. When $n = 160$ 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 $\beta_{ji}$ associated with the lagged $x_H$ in (3.5) are well approximated by the Almon coefficients $\omega_j(\boldsymbol{\pi}_i)$ in (3.6). 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 $n = 160$, 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 (4.4) and (4.5), 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 therefore 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 $n = 80$ and .5 for $n = 160$. 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 $n = 160$ 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 $n = 160$, 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 a 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.

# 5 Empirical Application

As an empirical illustration, we analyze Granger causality between a weekly term spread (long and short term interest rate spread) and quarterly real GDP growth in the U.S. We test for 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.[15] 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 interest rates used for the term spread are respectively the federal funds (FF) rate and 10-year Treasury constant maturity rate. We aggregate each daily series into weekly series by picking the last observation in each week (recall that interest rates are stock variables). The sample period is January 5, 1962 to December 31, 2013, covering 2,736 weeks or 208 quarters.[16]

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

Table 4 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 $\tau_L$ is not constant, which we denote as $m(\tau_L)$: 13 quarters have 12 weeks each, 150 quarters have 13 weeks each, and 45 quarters have 14 weeks each. While the max test can be applied with varying $m(\tau_L)$, we simplify the analysis by taking a sample average at the end of each $\tau_L$, resulting in the following modified spread $\{x_H^*(\tau_L, j)\}_{j=1}^{12}$:

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

This modification gives us a dataset with $n = 208$, $m = 12$, and therefore $T = mn = 2,496$ high

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

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

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 4 reveal our tests work well for $n = 80$.

### 5.1 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 a MF-VAR(2) specification since the resulting residuals from the naïve model (5.2), below, appear to be serially uncorrelated (all models also include a constant term). The MF max test operates on parsimonious regression models

$$x_L(\tau_L) = \alpha_{0i} + \sum_{k=1}^{2} \alpha_{ki} x_L(\tau_L - k) + \beta_i x_H^*(\tau_L - 1, 12 + 1 - i) + u_{L,i}(\tau_L), \quad i = 1, \ldots, 24, \quad (5.1)$$

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:

$$x_L(\tau_L) = \alpha_0 + \sum_{k=1}^{2} \alpha_k x_L(\tau_L - k) + \sum_{i=1}^{24} \beta_i x_H^*(\tau_L - 1, 12 + 1 - i) + u_L(\tau_L). \quad (5.2)$$

The LF max test is based on parsimonious models:

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

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:

$$x_L(\tau_L) = \alpha_0 + \sum_{k=1}^{2} \alpha_k x_L(\tau_L - k) + \sum_{i=1}^{3} \beta_i x_H^*(\tau_L - i) + u_L(\tau_L). \quad (5.3)$$

Wald statistic p-values are computed based on the non-robust covariance matrix and Gonçalves and Kilian's (2004) bootstrap, with $M = 1,000$ replications. Max statistic p-values are computed based on the robust covariance matrix with 100,000 draws from an approximation to the limit distribution under non-causality.

We perform the Ljung-Box $Q$ test of serial uncorrelatedness of the least squares residuals from the MF model (5.2) and LF model (5.3) 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 only {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.17 of Ghysels, Hill, and Motegi (2017) 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 4.1. These three tests all agree that there is non-causality in recent periods, possibly reflecting some structural change in the middle of the entire sample.

The MF Wald test, in contrast, suggests that there is significant causality only *after* subsample 1990:IV-2010:III, which is somewhat counter-intuitive. This result may stem from parameter proliferation. As seen from (5.1)-(5.3), 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.

### 5.2 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

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

or the restricted models with Almon polynomial $\omega_j(\boldsymbol{\pi}_i)$ of order $s = 3$:

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

in each case $i = 1, \ldots, 24$. 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:

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

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

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

The LF max test is based on the unrestricted parsimonious regression models $x_L(\tau_L) = \alpha_{0i} + \sum_{k=1}^{2} \alpha_{ki} x_L(\tau_L - k) + \sum_{j=1}^{3} \beta_{ji} x_H^*(\tau_L - j) + \gamma_i x_H^*(\tau_L + i) + u_{L,i}(\tau_L), \quad i = 1, 2, 3$. 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: $x_L(\tau_L) = \alpha_0 + \sum_{k=1}^{2} \alpha_k x_L(\tau_L - k) + \sum_{j=1}^{3} \beta_j x_H^*(\tau_L - j) + \sum_{i=1}^{3} \gamma_i x_H^*(\tau_L + i) + u_L(\tau_L)$.

Wald test p-values are bootstrapped with $M = 1,000$ bootstrap samples, and the max test p-values are computed using 100,000 draws from limit distribution 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.[17]

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

Figure 4 plots p-values for the causality tests over the 129 subsamples. While MF tests without a MIDAS polynomial find significant causality in some subsamples (cfr. Panels (a) and (b)), MF tests with a 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.[18]

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.

# 6 Conclusions

We propose a new test designed for many zero restrictions in regression models. A classical Wald test approach may have a poor finite sample performance when the number of zero restrictions is relatively large. We tackle the dimensionality problem head on by splitting key regressors across many parsimonious regression models. The $i^{th}$ parsimonious regression model contains the $i^{th}$ individual regressor only, so that parameter proliferation is less an issue. We then take the maximum of the squared estimators across all parsimonious regression models.

The asymptotic distribution of our max test statistic is non-standard under the null hypothesis, but an approximate p-value is readily available by drawing from an asymptotically valid approximation to the asymptotic distribution directly. Under the alternative hypothesis, at least one of the key estimators has a nonzero probability limit under fairly weak conditions. The max test is therefore consistent.

After presenting the general theory of the max test, the paper focuses on mixed frequency

uncorrelated residuals at the 5% level in {4, 8, 5} windows out of 129 for lags {4, 8, 12}. When a MIDAS polynomial is used, the null hypothesis is rejected in {25, 14, 16} windows. In the LF model the null hypothesis is rejected in {31, 17, 26} windows. If we raise the block size to 10 or 20, rejections occur in only a few windows across all models. See Table T.17 of Ghysels, Hill, and Motegi (2017) for complete results.

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

Granger causality as a prominent application. Through 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 finite sample. The MF max and Wald tests have roughly equal power in many cases, but the former is more powerful under causality with a large time lag.

As an empirical application, we investigate Granger causality 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 the 1990s, after which causality vanishes, while Wald and LF tests yield mixed results.

Finally, the max test has wide 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 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 (2017)). 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.

## 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.

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-Lévy 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.

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

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

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

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. (2013): "Matlab Toolbox for Mixed Sampling Frequency Data Analysis using MIDAS Regression Models," Available at `http://www.mathworks.com/matlabcentral/fileexchange/45150-midas-matlab-toolbox`.

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

Ghysels, E., J. B. Hill, and K. Motegi (2016): "Testing for Granger Causality with Mixed Frequency Data," *Journal of Econometrics*, 192, 207–230.

——— (2017): "Supplemental Material for '*Testing a Large Set of Zero Restrictions in Regression Models, with an Application to Mixed Frequency Granger Causality*'," Department of Economics, University of North Carolina at Chapel Hill.

Ghysels, E., V. Kvedaras, and V. Zemlys (2016): "Mixed Frequency Data Sampling Regression Models: The R Package midasr," *Journal of Statistical Software*, 72, 1–35.

Ghysels, E., P. Santa-Clara, and R. Valkanov (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.

Götz, T. B., A. Hecq, and S. Smeekes (2016): "Testing for Granger Causality in Large Mixed-Frequency VARs," *Journal of Econometrics*, 193, 418–432.

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 (2017): "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.

White, H. (2000): "A Reality Check for Data Snooping," *Econometrica*, 68, 1097–1126.

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 Proofs of Main Results

We present proofs of Theorems 2.1-2.5. Proofs of Theorem 2.6 and Corollary 2.7 are omitted since they are self-explanatory.

**Proof of Theorem 2.1** Recall from (2.8) that the $i^{th}$ parsimonious regression model is written as $y_t = X_{it}'\theta_i + u_{it}$, where $X_{it} = [z_{1t}, \ldots, z_{pt}, x_{it}]'$ and $\theta_i = [\alpha_i', \beta_i]' = [\alpha_{1i}, \ldots, \alpha_{pi}, \beta_i]'$. Let $\hat{\theta}_{ni} = [\hat{\alpha}_{ni}', \hat{\beta}_{ni}]' = [\hat{\alpha}_{n1i}, \ldots, \hat{\alpha}_{npi}, \hat{\beta}_{ni}]'$ be the least squares estimator for $\theta_i$.

In order to characterize the distribution limit of $\hat{\mathcal{T}}_n = \max_{1\le i\le h}(\sqrt{n}\hat{\beta}_{ni})^2$, we must show convergence of the finite dimensional distributions of $\{\sqrt{n}\hat{\beta}_{ni}\}_{i=1}^h$ and stochastic equicontinuity (e.g. Dudley (1978), Andersen and Dobric (1987)). By discreteness of $i$, note $\forall(\epsilon,\eta) > 0\ \exists\delta \in (0,1)$ such that $\sup_{1\le i\le h:\,|i-\tilde{i}|\le\delta} |\sqrt{n}\hat{\beta}_{ni} - \sqrt{n}\hat{\beta}_{n\tilde{i}}| = 0\ a.s.$ Therefore $\lim_{n\to\infty} P(\sup_{1\le i\le h:\,|i-\tilde{i}|\le\delta} |\sqrt{n}\hat{\beta}_{ni} - \sqrt{n}\hat{\beta}_{n\tilde{i}}| > \eta) \le \varepsilon$ for some $\delta > 0$, hence $\{\sqrt{n}\hat{\beta}_{ni}\}_{i=1}^h$ is stochastically equicontinuous.

Let $\beta = [\beta_1, \ldots, \beta_h]'$ and $\hat{\beta}_n = [\hat{\beta}_{n1}, \ldots, \hat{\beta}_{nh}]'$. Stack all parameters across the $h$ models as $\theta = [\theta_1', \ldots, \theta_h']'$ and $\hat{\theta}_n = [\hat{\theta}_{n1}', \ldots, \hat{\theta}_{nh}']'$. Define an $h \times (p+1)h$ full row rank selection matrix $R$ such that $\hat{\beta}_n = R\hat{\theta}_n$. See (2.9) for the exact construction of $R$.

In order to prove Theorem 2.1, it suffices to show that $\sqrt{n}\hat{\beta}_n \xrightarrow{d} N(0,V)$ under $H_0 : b = 0$, where $V$ is defined in (2.10). The main statement that $\hat{\mathcal{T}}_n = \max_{1\le i\le h}(\sqrt{n}\hat{\beta}_{ni})^2 \xrightarrow{d} \max_{1\le i\le h}\mathcal{N}_i^2$, where $\mathcal{N} = [\mathcal{N}_1, \ldots, \mathcal{N}_h]' \sim N(0,V)$, then follows instantly from the continuous mapping theorem since $\max\{\cdot\}$ is a continuous function.

Stack the underlying coefficients as $\theta_{0i} = [a_1, \ldots, a_p, b_i]'$ and $\theta_0 = [\theta_{01}', \ldots, \theta_{0h}']'$. It is sufficient to show that $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0,S)$ under $H_0 : b = 0$, where $S$ is defined in (2.11). This result suffices since $\sqrt{n}\hat{\beta}_n = R \times \sqrt{n}(\hat{\theta}_n - \theta_0)$ under $H_0$ and $V = RSR'$ by construction.

Under $H_0$, we have that $\theta_{0i} = [a_1, \ldots, a_p, 0]'$ for $i = 1, \ldots, h$. Hence the $i^{th}$ parsimonious regression model includes the true DGP as $y_t = X_{it}'\theta_{0i} + \epsilon_t$. Define $\Gamma_{ij} = E[X_{it}X_{jt}']$, $\Lambda_{ij} = E[\epsilon_t^2 X_{it}X_{jt}']$, $\Sigma_{ij} = \Gamma_{ii}^{-1}\Lambda_{ij}\Gamma_{jj}^{-1}$, and $S = [\Sigma_{ij}]$ for $i,j \in \{1,\ldots,h\}$, as in (2.11).

The $\alpha$-mixing property of Assumption 2.3 implies ergodicity. Stationarity, square integrability, and the ergodic theorem yield

$$\hat{\Gamma}_{ii} = \frac{1}{n}\sum_{t=1}^{n} X_{it}X_{it}' \xrightarrow{p} \Gamma_{ii}, \tag{A.1}$$

which is a positive definite matrix under Assumption 2.2. By (A.1),

$$\sqrt{n}(\hat{\theta}_{ni} - \theta_{0i}) = \sqrt{n}\left(\sum_{t=1}^{n} X_{it}X_{it}'\right)^{-1}\sum_{t=1}^{n} X_{it}\epsilon_t = \Gamma_{ii}^{-1} \times \frac{1}{\sqrt{n}}\sum_{t=1}^{n} X_{it}\epsilon_t + o_p(1). \tag{A.2}$$

Pick any $\lambda = [\lambda_1', \ldots, \lambda_h']'$ with $\lambda_i \in \mathbb{R}^{q+1}$ and $\lambda_i'\lambda_i = 1$. Define $X_t(\lambda) = \sum_{i=1}^{h} \lambda_i'\Gamma_{ii}^{-1}X_{it}$. We have that

$$\begin{aligned}\lambda' \times \sqrt{n}(\hat{\theta}_n - \theta_0) &= \sum_{i=1}^{h} \lambda_i' \sqrt{n}(\hat{\theta}_{ni} - \theta_{0i}) = \sum_{i=1}^{h} \lambda_i' \left( \Gamma_{ii}^{-1} \times \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_{it}\epsilon_t \right) + o_p(1) \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( \sum_{i=1}^{h} \lambda_i' \Gamma_{ii}^{-1} X_{it} \right) \epsilon_t + o_p(1) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} X_t(\lambda)\epsilon_t + o_p(1).\end{aligned} \tag{A.3}$$

Observe that

$$E[X_t(\lambda)^2 \epsilon_t^2] = \sum_{i=1}^{h} \sum_{j=1}^{h} \lambda_i' \Gamma_{ii}^{-1} \Lambda_{ij} \Gamma_{jj}^{-1} \lambda_j = \lambda' S \lambda < \infty.$$

Under Assumptions 2.1-2.3, $\{\sum_{i=1}^{h} \lambda_i' X_{it}\epsilon_t\}$ is a stationary, ergodic, square integrable martingale difference. Therefore Billingsley's (1961) central limit theorem applies to yield $(1/\sqrt{n}) \sum_{t=1}^{n} X_t(\lambda)\epsilon_t \overset{d}{\to} N(0, \lambda' S \lambda)$. By the Cramér-Wold theorem, $\sqrt{n}(\hat{\theta}_n - \theta_0) \overset{d}{\to} N(0, S)$. Hence,

$$\sqrt{n}\hat{\beta}_n = R \times \sqrt{n}(\hat{\theta}_n - \theta_0) \overset{d}{\to} N(0, V) \tag{A.4}$$

and therefore

$$\hat{\mathcal{T}}_n = \max\left\{ (\sqrt{n}\hat{\beta}_{n1})^2, \ldots, (\sqrt{n}\hat{\beta}_{nh})^2 \right\} \overset{d}{\to} \max\left\{ \mathcal{N}_1^2, \ldots, \mathcal{N}_h^2 \right\}. \qquad \mathcal{QED}$$

**Proof of Theorem 2.2** $\hat{\mathcal{T}}_n = \max_{1 \le i \le h}(\sqrt{n}\hat{\beta}_{ni})^2$ operates on a discrete-valued stochastic function $g_n(i) \equiv \hat{\beta}_{ni}$. As shown in the proof of Theorem 2.1, this implies that weak convergence for $\{g_n(1), \ldots, g_n(h)\}$ is identical to convergence in the finite dimensional distributions of $\{g_n(1), \ldots, g_n(h)\}$. Hansen's (1996) proof of his Theorem 2 therefore carries over to prove the present claim. $\mathcal{QED}$

**Proof of Theorem 2.3** Under Assumptions 2.1-2.3, $\{y_t, X_{it}, \epsilon_t\}$ are square integrable, stationary $\alpha$-mixing, and therefore ergodic. Therefore $\hat{\Gamma}_{ij} \overset{p}{\to} \Gamma_{ij}$ and $\hat{\theta}_{ni} \overset{p}{\to} (E[X_{it}X_{it}'])^{-1}E[X_{it}y_t] \equiv \theta_i^*$. Under $H_0$, the parsimonious model is identically $y_t = X_{it}'\theta_{0i} + \epsilon_t$ with $\theta_{0i} = [a_1, \ldots, a_p, 0]'$ for $i = 1, \ldots, h$. Hence $\hat{\theta}_{ni} \overset{p}{\to} \theta_{0i}$ and $\theta_i^* = \theta_{0i}$.

Combined with stationarity, ergodicity and square integrability, $\hat{\Lambda}_{ij} \overset{p}{\to} \Lambda_{ij}^* \equiv E[(y_t - X_{it}'\theta_i^*)^2 X_{it}X_{jt}']$ with $\|\Lambda_{ij}^*\| < \infty$. Also, $\Lambda_{ij}^* = \Lambda_{ij}$ under $H_0$. Therefore $\hat{V}_n \overset{p}{\to} V$ under $H_0$. Under $H_1 : b \neq 0$, $\hat{V}_n \overset{p}{\to} RS^*R'$ with $S^* = [\Gamma_{ii}^{-1}\Lambda_{ij}^*\Gamma_{jj}^{-1}]_{i,j}$ and $\|RS^*R'\| < \infty$. $\mathcal{QED}$

**Proof of Theorem 2.4** Recall the $i^{th}$ parsimonious regression model is $y_t = X_{it}'\theta_i + u_{it}$. In view of stationarity, square integrability, and ergodicity, the least squares estimator satisfies $\hat{\theta}_{ni} \overset{p}{\to} \theta_i^*$, where $\theta_i^* = [E[X_{it}X_{it}']]^{-1}E[X_{it}y_t]$. Recall from (2.2) that the DGP is $y_t = z_t'a + x_t'b + \epsilon_t$. Therefore

$$\begin{aligned}\theta_i^* &= \left[E\left[X_{it}X_{it}'\right]\right]^{-1} E\left[X_{it}(z_t'a + x_t'b + \epsilon_t)\right] = \left[E\left[X_{it}X_{it}'\right]\right]^{-1} (E[X_{it}z_t']a + E[X_{it}x_t']b + E[X_{it}\epsilon_t]) \\ &= \left[E\left[X_{it}X_{it}'\right]\right]^{-1} (E[X_{it}z_t']a + E[X_{it}x_t']b) = \Gamma_{ii}^{-1}(E[X_{it}z_t']a + C_i b),\end{aligned} \tag{A.5}$$

where the third equality holds from the mds assumption of $\epsilon_t$. Note that $z_t = [I_p, 0_{p\times 1}]X_{it}$ for

$i \in \{1, \ldots, h\}$ by construction. Hence

$$E\left[X_{it} z_t'\right] = E\left[X_{it} X_{it}'\right] \times \begin{bmatrix} I_p \\ 0_{1\times p} \end{bmatrix} = \Gamma_{ii} \times \begin{bmatrix} I_p \\ 0_{1\times p} \end{bmatrix}. \tag{A.6}$$

Substitute (A.6) into (A.5) to get the desired result (2.14). $\mathcal{QED}$

**Proof of Theorem 2.5** Pick the last row of (2.14). The lower left block of $\Gamma_{ii}^{-1}$ is

$$-n_i^{-1} E\left[x_{it} z_t'\right] \left[E\left[z_t z_t'\right]\right]^{-1}$$

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

$$n_i = E\left[x_{it}^2\right] - E\left[x_{it} z_t'\right] \left\{E\left[z_t z_t'\right]\right\}^{-1} E\left[z_t x_{it}\right].$$

Hence, the last row of $\Gamma_{ii}^{-1} C_i$ appearing in (2.14) is $n_i^{-1} d_i'$, where

$$d_i = E\left[x_t x_{it}\right] - E\left[x_t z_t'\right] \left\{E\left[z_t z_t'\right]\right\}^{-1} E\left[z_t x_{it}\right]. \tag{A.7}$$

If $\beta^* = 0$, then $n_i^{-1} d_i' b = 0$ for any $i \in \{1, \ldots, h\}$ in view of (2.14). $n_i$ is a nonzero finite scalar for any $i \in \{1, \ldots, h\}$ by the nonsingularity of $E\left[X_{it} X_{it}'\right]$. Hence we have that $d_i' b = 0$ for all $i \in \{1, \ldots, h\}$. Stack these equations to get that $Db = 0$, where $D = [d_1, \ldots, d_h]' \in \mathbb{R}^{h\times h}$.

To prove the main statement that $b = 0$, it is sufficient to show that $D$ is non-singular. Equation (A.7) implies that

$$D = E[x_t x_t'] - E[x_t z_t'] \left[E[z_t z_t']\right]^{-1} E[z_t x_t].$$

Now define

$$\Delta = E\left[\begin{bmatrix} z_t \\ x_t \end{bmatrix} \begin{bmatrix} z_t' & x_t' \end{bmatrix}\right].$$

$\Delta$ is trivially non-singular by Assumption 2.2. Note that $D$ is the Schur complement of $\Delta$ with respect to $E[z_t z_t']$. Therefore, by the classic argument of partitioned matrix inversion, $D$ is non-singular as desired. $\mathcal{QED}$

# B Double Time Indices

In this section, we introduce useful notations for mixed frequency data. Consider a high frequency variable $x_H$ and a low frequency variable $x_L$. Let $\tau_L \in \mathbb{Z}$ be a low frequency time period. Suppose that each low frequency time period has $m$ high frequency time periods. In period $\tau_L$, we observe $\{x_H(\tau_L, 1), x_H(\tau_L, 2), \ldots, x_H(\tau_L, m), x_L(\tau_L)\}$ sequentially.

It is often useful to use a notational convention that allows the second argument of $x_H$ to be an arbitrary integer. It is understood, for example, 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, what we call a *high frequency simplification* operates as follows.

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

$\lceil x \rceil$ is the smallest integer not smaller than $x$, while $\lfloor x \rfloor$ is the largest integer not larger than $x$. By applying the high frequency simplification, any integer put in the second argument of $x_H$ can be transformed to a natural number between 1 and $m$ with the first argument being modified appropriately. One can indeed verify that $m \left\lceil \frac{1-i}{m} \right\rceil + i \in \{1, \dots, m\}$ when $i \leq 0$, and $i - m \left\lfloor \frac{i-1}{m} \right\rfloor \in \{1, \dots, m\}$ when $i \geq m + 1$.

Since the high frequency simplification allows both arguments of $x_H$ to be any integer, the following *low frequency simplification* is well defined.

$$x_H(\tau_L - \tau, i) = x_H(\tau_L, i - m\tau), \quad \forall \tau_L, \tau, i \in \mathbb{Z}. \tag{B.2}$$

Equation (B.2) states that any lag or lead $\tau$ appearing in the first argument can be deleted by modifying the second argument appropriately. The second argument may go below 1 or above $m$, but such a case is covered by the high frequency simplification (B.1). Equations (B.1) and (B.2) are of practical use when one writes DGPs or regression models for mixed frequency data.

Table 1: 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. $n = 80$ | | | | | A.1.2. $n = 160$ | | | | |
| | 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. $n = 80$ | | | | | A.2.2. $n = 160$ | | | | |
| | 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. $n = 80$ | | | | | B.1.2. $n = 160$ | | | | |
| | 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. $n = 80$ | | | | | B.2.2. $n = 160$ | | | | |
| | 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 max test 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 5%. The number of Monte Carlo samples is 5000 for max tests and 1000 for Wald tests.

Table 1: 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. $n = 80$ | | | | | C.1.2. $n = 160$ | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 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. $n = 80$ | | | | | C.2.2. $n = 160$ | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 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. $n = 80$ | | | | | D.1.2. $n = 160$ | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 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. $n = 80$ | | | | | D.2.2. $n = 160$ | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 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. $n = 80$ | | | | | E.1.2. $n = 160$ | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 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. $n = 80$ | | | | | E.2.2. $n = 160$ | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 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 2: 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**: $\boldsymbol{b} = \mathbf{0}_{24\times 1}$ | | | | | | | | | | |
| A.1. $d = 0.2$ (low persistence in $x_H$) | | | | | | | | | | |
| A.1.1. $n = 80$ | | | | | | A.1.2. $n = 160$ | | | | |
| | 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. $n = 80$ | | | | | | A.2.2. $n = 160$ | | | | |
| | 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. $n = 80$ | | | | | | B.1.2. $n = 160$ | | | | |
| | 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. $n = 80$ | | | | | | B.2.2. $n = 160$ | | | | |
| | 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 max test 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**: $b_j = 0.3 \times I(j = 24)$ for $j = 1, \ldots, 24$ | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| C.1. $d = 0.2$ (low persistence in $x_H$) | | | | | | | | | | |
| C.1.1. $n = 80$ | | | | | | C.1.2. $n = 160$ | | | | |
| | 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. $n = 80$ | | | | | | C.2.2. $n = 160$ | | | | |
| | 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. $n = 80$ | | | | | | D.1.2. $n = 160$ | | | | |
| | 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. $n = 80$ | | | | | | D.2.2. $n = 160$ | | | | |
| | 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. $n = 80$ | | | | | | E.1.2. $n = 160$ | | | | |
| | 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. $n = 80$ | | | | | | E.2.2. $n = 160$ | | | | |
| | 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 3: 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 $n = 80$ | | | | | A.2 $n = 160$ | | | | | |
| 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 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 3: Continued

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

B.1 $n = 80$

| | | | | |
|---|---|---|---|---|
| **B.1.1 Mixed Frequency Tests without MIDAS** | | | | |
| | $r_{MF} = 4$ | $r_{MF} = 8$ | $r_{MF} = 12$ | $r_{MF} = 24$ |
| $h_{MF}$ | Max, Wald | Max, Wald | Max, Wald | Max, Wald |
| 4 | .459, .522 | .353, .401 | .290, .371 | .208, .231 |
| 8 | .471, .481 | .357, .417 | .295, .335 | .220, .208 |
| 12 | .464, .445 | .370, .397 | .320, .324 | .234, .189 |
| 24 | .508, .354 | .430, .280 | .393, .228 | .333, .156 |
| **B.1.2 Mixed Frequency Tests with MIDAS** | | | | |
| | $r_{MF} = 4$ | $r_{MF} = 8$ | $r_{MF} = 12$ | $r_{MF} = 24$ |
| $h_{MF}$ | Max, Wald | Max, Wald | Max, Wald | Max, Wald |
| 4 | .470, .548 | .360, .431 | .293, .381 | .206, .250 |
| 8 | .477, .523 | .358, .464 | .293, .392 | .208, .232 |
| 12 | .472, .534 | .361, .465 | .306, .396 | .202, .252 |
| 24 | .479, .540 | .361, .470 | .309, .380 | .201, .237 |
| **B.1.3 Low Frequency Tests (Flow Sampling)** | | | | |
| | $r_{LF} = 1$ | $r_{LF} = 2$ | $r_{LF} = 3$ | $r_{LF} = 4$ |
| $h_{LF}$ | Max, Wald | Max, Wald | Max, Wald | Max, Wald |
| 1 | .096, .103 | .083, .073 | .065, .070 | .066, .072 |
| 2 | .102, .081 | .082, .070 | .070, .058 | .061, .063 |
| 3 | .107, .089 | .084, .085 | .072, .071 | .069, .077 |
| 4 | .107, .085 | .086, .082 | .073, .068 | .069, .069 |
| **B.1.4 Low Frequency Tests (Stock Sampling)** | | | | |
| | $r_{LF} = 1$ | $r_{LF} = 2$ | $r_{LF} = 3$ | $r_{LF} = 4$ |
| $h_{LF}$ | Max, Wald | Max, Wald | Max, Wald | Max, Wald |
| 1 | .056, .054 | .055, .052 | .049, .049 | .047, .040 |
| 2 | .057, .038 | .052, .035 | .046, .058 | .050, .043 |
| 3 | .057, .053 | .055, .045 | .049, .048 | .052, .057 |
| 4 | .063, .059 | .058, .049 | .057, .046 | .051, .044 |

B.2 $n = 160$

| | | | | |
|---|---|---|---|---|
| **B.2.1 Mixed Frequency Tests without MIDAS** | | | | |
| | $r_{MF} = 4$ | $r_{MF} = 8$ | $r_{MF} = 12$ | $r_{MF} = 24$ |
| $h_{MF}$ | Max, Wald | Max, Wald | Max, Wald | Max, Wald |
| 4 | .823, .893 | .739, .825 | .695, .735 | .578, .588 |
| 8 | .818, .846 | .744, .803 | .687, .733 | .585, .570 |
| 12 | .824, .851 | .740, .786 | .677, .724 | .596, .555 |
| 24 | .813, .831 | .751, .753 | .702, .688 | .614, .508 |
| **B.2.2 Mixed Frequency Tests with MIDAS** | | | | |
| | $r_{MF} = 4$ | $r_{MF} = 8$ | $r_{MF} = 12$ | $r_{MF} = 24$ |
| $h_{MF}$ | Max, Wald | Max, Wald | Max, Wald | Max, Wald |
| 4 | .837, .901 | .757, .835 | .708, .747 | .590, .623 |
| 8 | .835, .881 | .766, .844 | .705, .778 | .602, .626 |
| 12 | .840, .894 | .759, .847 | .705, .804 | .611, .634 |
| 24 | .838, .909 | .766, .853 | .723, .799 | .620, .623 |
| **B.2.3 Low Frequency Tests (Flow Sampling)** | | | | |
| | $r_{LF} = 1$ | $r_{LF} = 2$ | $r_{LF} = 3$ | $r_{LF} = 4$ |
| $h_{LF}$ | Max, Wald | Max, Wald | Max, Wald | Max, Wald |
| 1 | .159, .137 | .125, .125 | .093, .096 | .090, .076 |
| 2 | .155, .142 | .119, .123 | .108, .102 | .093, .094 |
| 3 | .145, .142 | .116, .123 | .107, .103 | .094, .107 |
| 4 | .150, .146 | .119, .122 | .102, .108 | .086, .092 |
| **B.2.4 Low Frequency Tests (Stock Sampling)** | | | | |
| | $r_{LF} = 1$ | $r_{LF} = 2$ | $r_{LF} = 3$ | $r_{LF} = 4$ |
| $h_{LF}$ | Max, Wald | Max, Wald | Max, Wald | Max, Wald |
| 1 | .055, .065 | .056, .059 | .054, .049 | .052, .058 |
| 2 | .062, .054 | .057, .051 | .054, .050 | .055, .047 |
| 3 | .062, .053 | .057, .062 | .051, .057 | .058, .049 |
| 4 | .056, .055 | .061, .060 | .053, .062 | .057, .036 |

Table 3: Continued

| **C. Lagged Causality**: $c_j = 0.25 \times I(j = 12)$ for $j = 1, \ldots, 12$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| C.1 $n = 80$ | | | | | C.2 $n = 160$ | | | | |
| 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 3: 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 $n = 80$ | | | | | D.2 $n = 160$ | | | | |
| 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 3: Continued

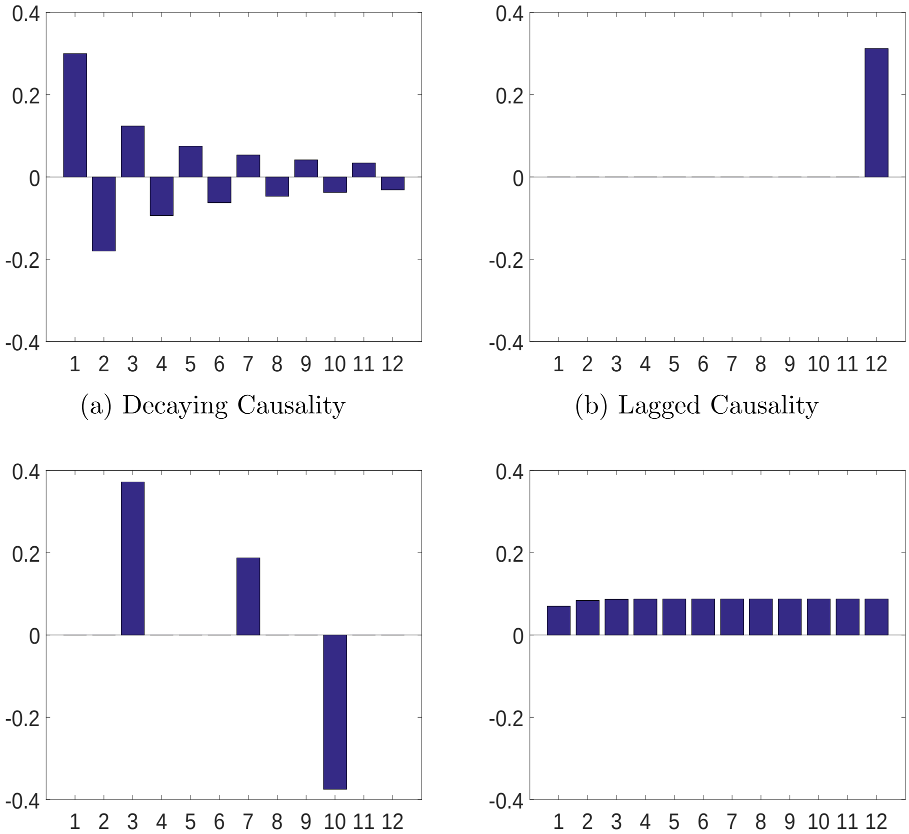
| **E. Uniform Causality**: $c_j = 0.07$ for $j = 1, \ldots, 12$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| E.1 $n = 80$ | | | | | E.2 $n = 160$ | | | | |
| E.1.1 Mixed Frequency Tests without MIDAS | | | | | E.2.1 Mixed Frequency Tests without MIDAS | | | | |
| | $r_{MF} = 4$ | $r_{MF} = 8$ | $r_{MF} = 12$ | $r_{MF} = 24$ | | $r_{MF} = 4$ | $r_{MF} = 8$ | $r_{MF} = 12$ | $r_{MF} = 24$ |
| $h_{MF}$ | 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 4: 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

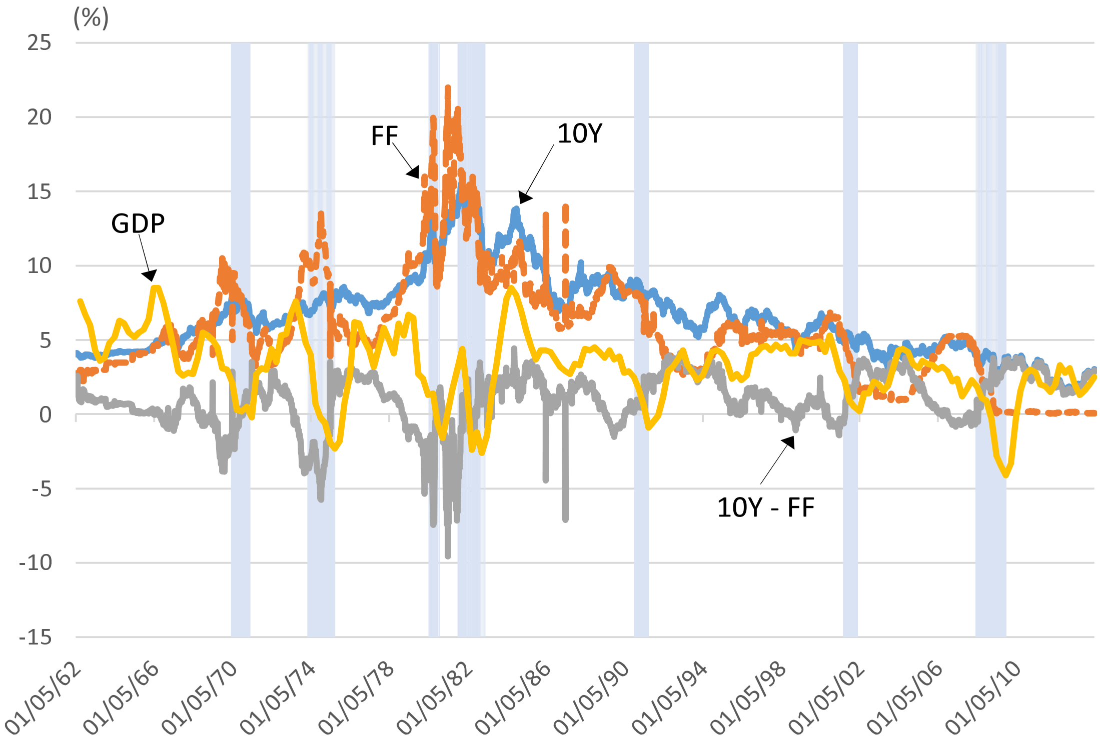



(a) Decaying Causality (b) Lagged Causality

(c) Sporadic Causality (d) Uniform Causality

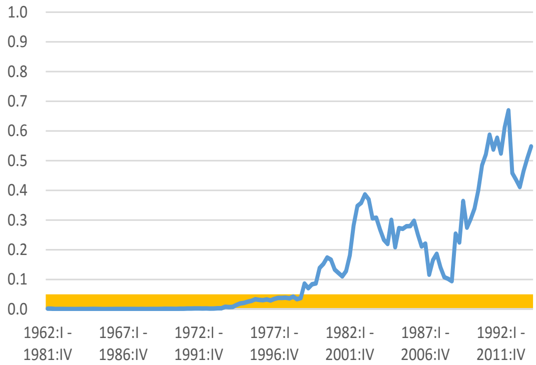
In the low-to-high causality simulation experiment, we start with a structural MF-VAR(1) data generating process, and transform it to a reduced-form MF-VAR(1). This figure shows how each causal pattern in the structural form is transformed in the reduced form. The AR(1) parameter of $x_H$ is fixed at $d = 0.2$. 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 $j = 1, \ldots, 12$; 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

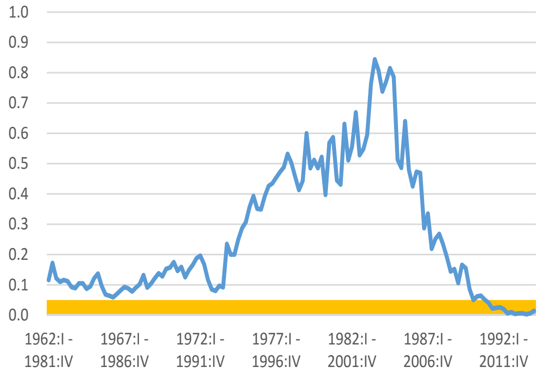



This figure plots weekly 10-year Treasury constant maturity rate (blue, solid line), weekly effective federal funds rate (red, dashed line), their spread 10Y−FF (gray, solid line), and the quarterly real GDP growth from previous year (yellow, solid line). The sample period covers January 5, 1962 through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

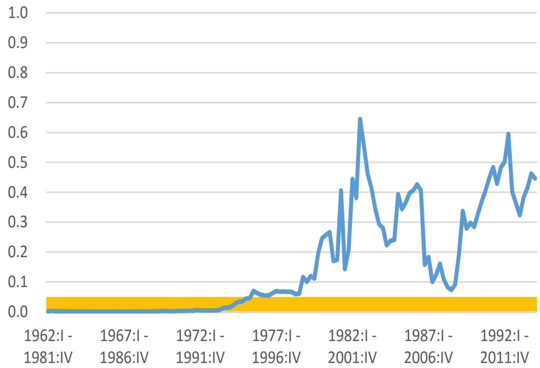
Figure 3: P-values for Tests of Non-Causality from Interest Rate Spread to GDP



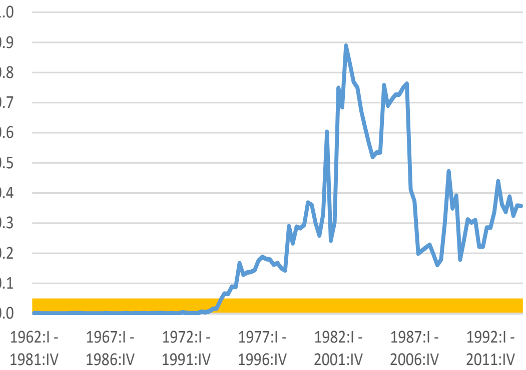

(a) MF Max Test




(b) MF Wald Test



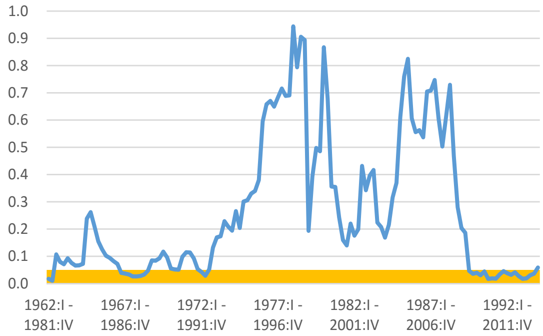

(c) LF Max Test



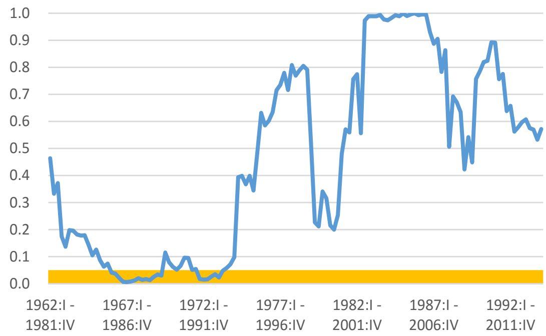

(d) LF Wald Test

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.

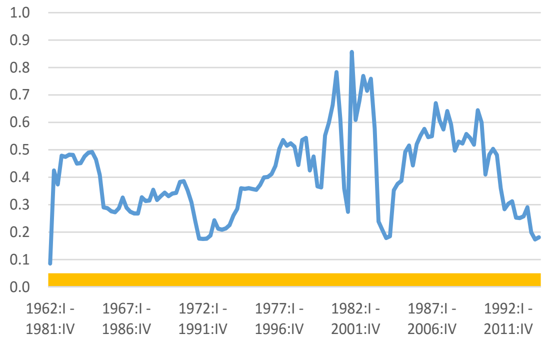
Figure 4: Rolling Window P-values for Tests of Non-Causality from GDP to Interest Rate Spread



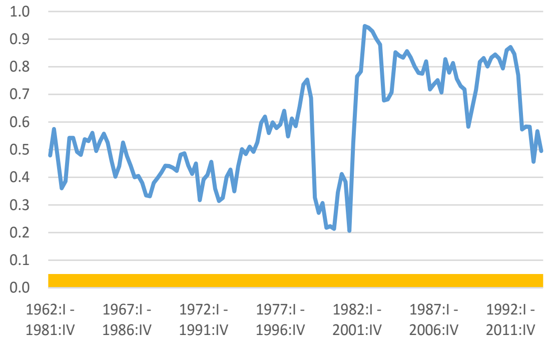

(a) MF Max Test (w/o MIDAS)



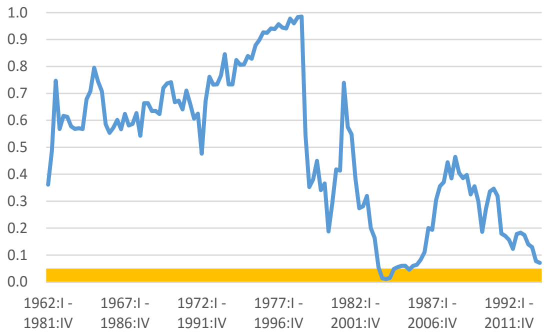

(b) MF Wald Test (w/o MIDAS)



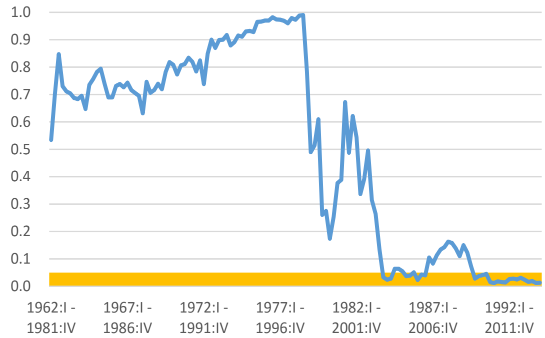

(c) MF Max Test (w/ MIDAS)




(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.