Testing a large set of zero restrictions in regression models, with an application to mixed frequency Granger causality —Short version—*

Eric Ghysels† Jonathan B. Hill‡ Kaiji Motegi§

June 28, 2019

Abstract

This paper proposes a new 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 regression model has one key regressor and other regressors not associated with the null hypothesis. The test is based on the maximum of the squared parameters of the key regressors. Parsimony ensures sharper estimates and therefore improves power in small sample. We present the general theory of our test and focus on mixed frequency Granger causality as a prominent application involving many zero restrictions.

Keywords: dimension reduction, Granger causality test, max test, Mixed Data Sampling (MIDAS), parsimonious regression models.

JEL Classification: C12, C22, C51.

1 Introduction

We propose a new test designed for a large set of zero restrictions in regression models. Suppose that the underlying data generating process is $y_t = \boldsymbol{z}_t'\boldsymbol{a} + \boldsymbol{x}_t'\boldsymbol{b} + \epsilon_t$, where $\boldsymbol{z}_t$ is assumed to have a small dimension p while $\boldsymbol{x}_t$ may have a large but finite dimension h. We want to test the null hypothesis $H_0 : \boldsymbol{b} = \boldsymbol{0}$ against a general alternative hypothesis $H_1 : \boldsymbol{b} \neq \boldsymbol{0}$.

A classical approach for testing H_0 relies on what we call a *full regression model* $y_t = \boldsymbol{z}_t'\boldsymbol{\alpha} + \boldsymbol{x}_t'\boldsymbol{\beta} + u_t$ and a Wald test. This approach may result in imprecise inference when $\boldsymbol{b}$ has a large

*This short version is prepared for the memorial lecture of the 33rd JSS Ogawa Award at the 2019 Japanese Joint Statistical Meeting, which is held at Shiga University in Japan in September 2019. The full version which contains all technical and numerical details is forthcoming in the Journal of Econometrics, and available at `http://www2.kobe-u.ac.jp/~motegi/research.html`

†Department of Economics and Department of Finance, Kenan-Flagler Business School, University of North Carolina at Chapel Hill, and CEPR. 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`

dimension relative to sample size n. The asymptotic χ^2-test may suffer from size distortions in small sample due to parameter proliferation. A bootstrap method can be employed to improve empirical size, but this generally results in the *size-corrected* test having low power.

To circumvent the issue of parameter proliferation, we propose to split each of the key regressors $\boldsymbol{x}_t = [x_{1t}, \ldots, x_{ht}]'$ across separate regression models. This approach leads to what we call *parsimonious regression models*: $y_t = \boldsymbol{z}_t'\boldsymbol{\alpha}_i + \beta_i x_{it} + u_{it}$ for $i = 1, \ldots, h$. The i^{th} parsimonious regression model has $\boldsymbol{z}_t$ and the i^{th} element of $\boldsymbol{x}_t$ only, so that parameter proliferation is not an issue. We then compute 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 $H_0 : \boldsymbol{b} = \boldsymbol{0}$, but it is easy to approximate a p-value by drawing from the asymptotic distribution. We will prove that, under $H_1 : \boldsymbol{b} \neq \boldsymbol{0}$, at least one of $\{\hat{\beta}_{n1}, \ldots, \hat{\beta}_{nh}\}$ has a nonzero probability limit under fairly weak conditions. This result ensures that, although the proposed approach does not deliver a consistent estimator of $\boldsymbol{b}$, we can reject the null hypothesis with probability approaching one for any direction $\boldsymbol{b} \neq \boldsymbol{0}$ under the alternative hypothesis. The max test is therefore a consistent test.

After presenting the general theory of the max test, we focus on Mixed Data Sampling (MIDAS) Granger causality as a prominent application involving many zero restrictions. Standard VAR models are designed for single-frequency data, and causality tests based on those models may produce misleading results of spurious (non-)causality. To alleviate the adverse impact of temporal aggregation, Ghysels, Hill, and Motegi (2016) develop Granger causality tests that explicitly take advantage of mixed frequency data. They extend Dufour, Pelletier, and Renault's (2006) VAR-based causality test, exploiting Ghysels' (2016) mixed frequency vector autoregressive (MF-VAR) models.

A challenge not addressed by Ghysels, Hill, and Motegi (2016) is parameter proliferation in MF-VAR models which has a negative impact on the finite sample performance of their tests. When the dimension of MF-VAR model is large, an asymptotic Wald test results in size distortions while a bootstrapped Wald test features low size-corrected power. The max test proposed in the present paper is a useful solution to such a large-dimension problem.

We show via Monte Carlo simulations that the proposed max test dominates the existing Wald test. As an empirical application, we analyze Granger causality from a weekly interest rate spread to quarterly real GDP growth in the United States. The proposed max test yields more intuitive empirical results than the existing Wald test.

The remainder of the paper is organized as follows. In Section 2, we present general theory of the parsimonious regression models and max tests. In Section 3, we focus on mixed frequency Granger causality tests as a specific application. In Section 4, we perform Monte Carlo simulations. The empirical application is presented in Section 5.

2 General theory

Consider a DGP in which $\{y_t\}$ depends linearly on $\boldsymbol{z}_t = [z_{1t}, \ldots, z_{pt}]'$ and $\boldsymbol{x}_t = [x_{1t}, \ldots, x_{ht}]'$. Define the σ-field $\mathcal{F}_t = \sigma(\boldsymbol{Y}_\tau : \tau \leq t+1)$ with all variables $\boldsymbol{Y}_t = [y_{t-1}, \boldsymbol{X}_t']'$, where $\boldsymbol{X}_t = [\boldsymbol{z}_t', \boldsymbol{x}_t']'$.

Thus, $\mathcal{F}_{t-1}$ contains information on $\{y_{t-1}, y_{t-2}, ...\}$ and the regressors $\{(\boldsymbol{z}_t, \boldsymbol{x}_t), (\boldsymbol{z}_{t-1}, \boldsymbol{x}_{t-1}), \dots\}$.

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 σ-field filtration $\mathcal{F}_t \subset \mathcal{F}_{t+1}$, and $\sigma^2 \equiv E[\epsilon_t^2] > 0$.

Let n be a sample size, and define an $n \times (p+h)$ matrix of regressors $\boldsymbol{X} = [\boldsymbol{X}_1, \dots, \boldsymbol{X}_n]'$.

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

Assumption 2.3. $\{\boldsymbol{X}_t, \epsilon_t\}$ are strictly stationary and ergodic. $\boldsymbol{X}_t$ is square integrable.

Using standard vector notations, e.g., $\boldsymbol{a} = [a_1, \dots, a_p]'$, DGP (2.1) is rewritten as $y_t = \boldsymbol{z}_t'\boldsymbol{a} + \boldsymbol{x}_t'\boldsymbol{b} + \epsilon_t$, where $\{\boldsymbol{z}_t\}$ are auxiliary regressors whose coefficients are not our main target. The number of the auxiliary regressors, p, is assumed to be relatively small. We want to test for the zero restrictions with respect to main regressors $\boldsymbol{x}_t$, i.e., $H_0 : \boldsymbol{b} = \boldsymbol{0}$. The number of zero restrictions, h, is assumed to be finite but potentially large relative to sample size n.

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

$$y_t = \sum_{k=1}^{p} \alpha_k z_{kt} + \sum_{i=1}^{h} \beta_i x_{it} + u_t. \tag{2.2}$$

Given model (2.2), it is straightforward to compute a Wald statistic with respect to $H_0 : \boldsymbol{b} = \boldsymbol{0}$. The statistic has an asymptotic χ^2 distribution with h degrees of freedom under Assumptions 2.1-2.3. A potential problem of this approach is that the asymptotic approximation may be poor when there are many zero restrictions relative to sample size. A parametric or wild bootstrap can be used to control for the size of the test, but this typically leads to low size-corrected power. It is therefore of use to propose a new test that achieves a sharper size and higher power.

We propose *parsimonious regression models*:

$$y_t = \sum_{k=1}^{p} \alpha_{ki} z_{kt} + \beta_i x_{it} + u_{it}, \quad i = 1, \dots, h. \tag{2.3}$$

A crucial difference between the full regression model (2.2) and the parsimonious regression models (2.3) is that, in the latter, there are h equations and the key regressor x_{it} along with the auxiliary regressors $\{z_{1t}, \dots, z_{pt}\}$ appear in the i^{th} equation. The parameters $\{\alpha_{1i}, \dots, \alpha_{pi}\}$ may differ across the equations $i = 1, \dots, h$, and unless the null hypothesis is true, they are generally not equal to the true values $\{a_1, \dots, a_p\}$. We run least squares for each parsimonious regression model to get $\{\hat{\beta}_{n1}, \dots, \hat{\beta}_{nh}\}$. Then we formulate a *max test statistic*

$$\hat{\mathcal{T}}_n = \max\left\{(\sqrt{n}\hat{\beta}_{n1})^2, \dots, (\sqrt{n}\hat{\beta}_{nh})^2\right\}. \tag{2.4}$$

Equations (2.3) and (2.4) 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 full regression model. As a result, the precision of $\hat{\beta}_{ni}$ improves toward its probability limit $\beta_i^* = \text{plim}_{n\to\infty}\hat{\beta}_{ni}$ for each i.

2.1 Asymptotic theory under the null hypothesis

We derive the asymptotic distribution of $\hat{\mathcal{T}}_n$ in (2.4) under $H_0 : \boldsymbol{b} = \boldsymbol{0}$. Rewrite each parsimonious regression model (2.3) as $y_t = \boldsymbol{X}_{it}'\boldsymbol{\theta}_i + u_{it}$ for $i = 1, \ldots, h$, where $\boldsymbol{X}_{it} = [z_{1t}, \ldots, z_{pt}, x_{it}]'$ and $\boldsymbol{\theta}_i = [\boldsymbol{\alpha}_i', \beta_i]' = [\alpha_{1i}, \ldots, \alpha_{pi}, \beta_i]'$. Stack all parameters across the h models as $\boldsymbol{\theta} = [\boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_h']'$. Define a selection matrix $\boldsymbol{R}$ that selects $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_h]'$ from $\boldsymbol{\theta}$. $\boldsymbol{R}$ is an $h \times (p+1)h$ full row rank matrix such that $\boldsymbol{\beta} = \boldsymbol{R}\boldsymbol{\theta}$.

Theorem 2.1. Under $H_0 : \boldsymbol{b} = \boldsymbol{0}$, we have that $\hat{\mathcal{T}}_n \xrightarrow{d} \max\{\mathcal{N}_1^2, \ldots, \mathcal{N}_h^2\}$ as $n \to \infty$, where $\boldsymbol{\mathcal{N}} = [\mathcal{N}_1, \ldots, \mathcal{N}_h]'$ is distributed as $N(\boldsymbol{0}, \boldsymbol{V})$ with covariance matrix $\boldsymbol{V} \equiv \boldsymbol{R}\boldsymbol{S}\boldsymbol{R}' \in \mathbb{R}^{h\times h}$, where $\boldsymbol{S} = [\boldsymbol{\Sigma}_{ij}]$, $\boldsymbol{\Sigma}_{ij} = \boldsymbol{\Gamma}_{ii}^{-1}\boldsymbol{\Lambda}_{ij}\boldsymbol{\Gamma}_{jj}^{-1}$, $\boldsymbol{\Gamma}_{ij} = E[\boldsymbol{X}_{it}\boldsymbol{X}_{jt}']$, and $\boldsymbol{\Lambda}_{ij} = E[\epsilon_t^2\boldsymbol{X}_{it}\boldsymbol{X}_{jt}']$ for $i, j \in \{1, \ldots, h\}$.

2.2 Simulated p-value

Let $\hat{\boldsymbol{V}}_n$ be a consistent estimator for $\boldsymbol{V}$, and draw M samples of vectors $\{\boldsymbol{\mathcal{N}}^{(1)}, \ldots, \boldsymbol{\mathcal{N}}^{(M)}\}$ independently from $N(\boldsymbol{0}, \hat{\boldsymbol{V}}_n)$. 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 $\hat{p}_{n,M} = (1/M)\sum_{j=1}^{M} I(\hat{\mathcal{T}}_n^{(j)} > \hat{\mathcal{T}}_n)$, where $I(A)$ is the indicator function that equals 1 if A occurs and 0 otherwise. Since $\boldsymbol{\mathcal{N}}^{(j)}$ are i.i.d. over j, and M can be made arbitrarily large, $\hat{p}_{n,M}$ can be made arbitrarily close to $P(\hat{\mathcal{T}}_n^{(1)} > \hat{\mathcal{T}}_n)$. The proposed max test is to reject H_0 at level α when $\hat{p}_{n,M_n} < \alpha$ and to accept H_0 when $\hat{p}_{n,M_n} \geq \alpha$, 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 $\bar{F}^0(\hat{\mathcal{T}}_n) \equiv 1 - F^0(\hat{\mathcal{T}}_n) = P(\max_{1\leq i\leq h}(\mathcal{N}_i^{(1)})^2 \geq \hat{\mathcal{T}}_n)$. 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 $\hat{p}_{n,M_n} = \bar{F}^0(\hat{\mathcal{T}}_n) + o_p(1)$. Therefore under H_0, $P(\hat{p}_{n,M_n} < \alpha) \to \alpha$ for any $\alpha \in (0, 1)$.

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

Theorem 2.3. Under Assumptions 2.1-2.3, $\hat{\boldsymbol{V}}_n \xrightarrow{p} \bar{\boldsymbol{V}}$ where $\bar{\boldsymbol{V}}$ is some matrix that satisfies $||\bar{\boldsymbol{V}}|| < \infty$. Specifically, $\bar{\boldsymbol{V}} = \boldsymbol{V} \equiv \boldsymbol{R}\boldsymbol{S}\boldsymbol{R}'$ under H_0, and under H_1 it follows $\boldsymbol{R}\boldsymbol{S}^*\boldsymbol{R}'$ where $\boldsymbol{S}^* = [\boldsymbol{\Gamma}_{ii}^{-1}\boldsymbol{\Lambda}_{ij}^*\boldsymbol{\Gamma}_{jj}^{-1}]_{i,j}$.

2.3 Limits under the alternative hypothesis

Under $H_1 : \boldsymbol{b} \neq \boldsymbol{0}$, $\hat{\beta}_{ni}$ in general does not converge in probability to the true value b_i due to omitted regressors. Let $\beta_i^* = \text{plim}_{n\to\infty}\hat{\beta}_{ni}$ be the pseudo-true value of β_i. The same notation applies to $\boldsymbol{\alpha}_i^*$ and $\boldsymbol{\theta}_i^* = [\boldsymbol{\alpha}_i^{*\prime}, \beta_i^*]'$. Let $\hat{\boldsymbol{\beta}}_n = [\hat{\beta}_{n1}, \ldots, \hat{\beta}_{nh}]'$ and $\boldsymbol{\beta}^* = [\beta_1^*, \ldots, \beta_h^*]'$.

Theorem 2.4. Let Assumptions 2.1-2.3 hold. Let $\boldsymbol{\Gamma}_{ii} = E[\boldsymbol{X}_{it}\boldsymbol{X}_{it}'] \in \mathbb{R}^{(p+1)\times(p+1)}$ and $\boldsymbol{C}_i = E[\boldsymbol{X}_{it}\boldsymbol{x}_t'] \in \mathbb{R}^{(p+1)\times h}$. Then, $\hat{\boldsymbol{\theta}}_n \xrightarrow{p} \boldsymbol{\theta}^* = [\boldsymbol{\theta}_1^{*\prime}, \ldots, \boldsymbol{\theta}_h^{*\prime}]'$, where

$$\boldsymbol{\theta}_i^* = \begin{bmatrix} \alpha_{1i}^* \\ \vdots \\ \alpha_{pi}^* \\ \beta_i^* \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_p \\ 0 \end{bmatrix} + \boldsymbol{\Gamma}_{ii}^{-1}\boldsymbol{C}_i\boldsymbol{b}, \quad i = 1, \ldots, h. \tag{2.5}$$

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

Theorem 2.4 provides useful insights on the relationship between the underlying coefficient $\boldsymbol{b}$ and the pseudo-true value $\boldsymbol{\beta}^*$. First, it is clear from (2.5) that $\boldsymbol{\beta}^* = \boldsymbol{0}$ whenever $\boldsymbol{b} = \boldsymbol{0}$. This is an intuitive result since each parsimonious regression model is correctly specified under $H_0 : \boldsymbol{b} = \boldsymbol{0}$. Second, as the next theorem proves, $\boldsymbol{b} = \boldsymbol{0}$ whenever $\boldsymbol{\beta}^* = \boldsymbol{0}$. This is a key 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 $H_0 : \boldsymbol{b} = \boldsymbol{0}$.

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

Theorems 2.1-2.5 together imply the max test statistic has its intended limit properties under either hypothesis. First, (2.4) indicates that $\hat{\mathcal{T}}_n \xrightarrow{p} \infty$ at rate n *if and only if* $\boldsymbol{\beta}^* \neq \boldsymbol{0}$, and by Theorems 2.4 and 2.5 $\hat{\boldsymbol{\beta}}_n \xrightarrow{p} \boldsymbol{\beta}^* \neq \boldsymbol{0}$ under $H_1 : \boldsymbol{b} \neq \boldsymbol{0}$. In conjunction with Theorem 2.2, the consistency of the max test follows.

Theorem 2.6. Let Assumptions 2.1-2.3 hold, then $\hat{\mathcal{T}}_n \xrightarrow{p} \infty$ at rate n and therefore $P(\hat{p}_{n,M_n} < \alpha) \to 1$ for any $\alpha \in (0, 1)$ *if and only if* $H_1 : \boldsymbol{b} \neq \boldsymbol{0}$ is true. In particular, $H_1 : \boldsymbol{b} \neq \boldsymbol{0}$ is true *if and only if* $\hat{\mathcal{T}}_n/n \xrightarrow{p} \max\left\{(\boldsymbol{\beta}_1^*)^2, \ldots, (\boldsymbol{\beta}_h^*)^2\right\} > 0$.

3 Mixed frequency Granger causality

In this section, we discuss testing for Granger causality between bivariate mixed frequency time series as a prominent example involving a large set of zero restrictions. We restrict ourselves to the bivariate case where we have a high frequency variable x_H and a low frequency variable 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 $\tau \in \mathbb{Z}$. A $K \times 1$ *mixed frequency vector* is defined as $\boldsymbol{X}(\tau) = [x_H(\tau, 1), \ldots, x_H(\tau, m), x_L(\tau)]'$, where $K = m + 1$ and $x_H(\tau, j)$ is the realization of x_H

at the j-th high frequency time period within a low frequency time period τ. Define the σ-field $\mathcal{F}_\tau \equiv \sigma(\boldsymbol{X}(\tau') : \tau' \leq \tau)$. We assume that $E[\boldsymbol{X}(\tau)|\mathcal{F}_{\tau-1}]$ has a version that is *almost surely* linear in $\{\boldsymbol{X}(\tau-1), \ldots, \boldsymbol{X}(\tau-p)\}$ for some finite $p \geq 1$.

Assumption 3.1. $\boldsymbol{X}(\tau)$ follows MF-VAR(p) for finite $p \geq 1$:

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

or compactly $\boldsymbol{X}(\tau) = \sum_{k=1}^{p} \boldsymbol{A}_k \boldsymbol{X}(\tau-k) + \boldsymbol{\epsilon}(\tau)$. The error $\{\boldsymbol{\epsilon}(\tau)\}$ is a strictly stationary martingale difference sequence (mds) with respect to increasing $\mathcal{F}_\tau \subset \mathcal{F}_{\tau+1}$, with a positive definite covariance matrix $E[\boldsymbol{\epsilon}(\tau)\boldsymbol{\epsilon}(\tau)']$.

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.

Assumption 3.3. $\boldsymbol{X}(\tau)$ and $\boldsymbol{\epsilon}(\tau)$ are ergodic.

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

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

x_H does not Granger cause x_L given the mixed frequency information set $\mathcal{F}_\tau = \sigma(\boldsymbol{X}(\tau') : \tau' \leq \tau)$ *if and only if* $H_0 : b_1 = \cdots = b_{pm} = 0$. The number of zero restrictions, pm, may be very large in some applications. Running parsimonious regression models and the max test is our proposed solution to the parameter proliferation arising from a large pm.

There is a clear correspondence between the general linear DGP (2.1) and the mixed frequency DGP (3.2). The regressand y_t is $x_L(\tau)$; common regressors $\{z_{1t}, \ldots, z_{pt}\}$ are identically the low frequency lags $\{x_L(\tau-1), \ldots, x_L(\tau-p)\}$; the main regressors split across parsimonious regression models $\{x_{1t}, \ldots, x_{ht}\}$ are the high frequency lags $\{x_H(\tau-1, m+1-1), \ldots, x_H(\tau-1, m+1-pm)\}$. The parsimonious regression models are therefore written as $x_L(\tau) = \sum_{k=1}^{p} \alpha_{k,i} x_L(\tau-k) + \beta_i x_H(\tau-1, m+1-i) + u_{L,i}(\tau)$ with $i = 1, \ldots, pm$. Assumptions 3.1-3.3 imply Assumptions 2.1-2.3. Thus, under Assumptions 3.1-3.3, Theorems 2.1-2.6 carry over to the max test for high-to-low causality.

4 Monte Carlo simulations

In this section, we perform Monte Carlo simulations and compare the finite sample performance of the max tests and Wald tests. In the present version of the paper, we focus on benchmark scenarios that include cross-section and single-frequency time series data. See the full version of the paper for more simulation results related with MIDAS scenarios.

4.1 Simulation design

Consider a DGP that $y_t = \boldsymbol{x}_t'\boldsymbol{b} + \epsilon_t$ with $\epsilon_t \overset{i.i.d.}{\sim} N(0,1)$. For simplicity, we do not include any common regressor z_t (i.e. $p = 0$) and only include key regressors $\boldsymbol{x}_t$. The dimension of $\boldsymbol{x}_t$ is $h \in \{10, 100, 200\}$, and sample size is $n \in \{10h, 50h\}$. Assume that $\boldsymbol{x}_t \overset{i.i.d.}{\sim} N(\boldsymbol{0}, \boldsymbol{\Gamma})$, where the diagonal elements of $\boldsymbol{\Gamma}$ are all equal to 1 and the off-diagonal elements are all equal to $\rho = 0.3$. We consider two patterns for the true coefficients $\boldsymbol{b} = [b_1, \ldots, b_h]'$:

DGP-1 $b_i = 0$ for all $i \in \{1, \ldots, h\}$.

DGP-2 $b_i = I(i < \frac{1}{10}h) \times \frac{10}{20h}$ for all $i \in \{1, \ldots, h\}$.

DGP-1 is prepared for investigating the empirical size of the max and Wald tests. Nonzero coefficients are put on the first 10% of the entire regressors under DGP-2, and the magnitude of causality sums up to $\sum_{i=1}^{h} b_i = 1/20$.

We run parsimonious regression models $y_t = \beta_i x_{it} + u_{it}$ for $i = 1, \ldots, h$, and compute the max test statistic $\hat{\mathcal{T}}_n = \max\{(\sqrt{n}\hat{\beta}_{n1})^2, \ldots, (\sqrt{n}\hat{\beta}_{nh})^2\}$. To compute a p-value, we draw $M = 5000$ samples from the asymptotic distribution under $H_0 : \boldsymbol{b} = \boldsymbol{0}$.

For comparison, we also run a full regression model $y_t = \boldsymbol{x}_t'\boldsymbol{\beta} + u_t$ and perform the asymptotic and bootstrapped Wald tests. The asymptotic Wald test is a well-known χ^2-test with degrees of freedom h. For the bootstrapped test, generate $B = 1000$ bootstrap samples based on Gonçalves and Kilian's (2004) recursive-design wild bootstrap with standard normal innovations.

For each test, we draw $J = 1000$ Monte Carlo samples and compute rejection frequencies with respect to nominal size $\alpha = 0.05$.

4.2 Simulation results

See Table 1 for rejection frequencies. Focusing on DGP-1, the asymptotic Wald test suffers from severe size distortions. See, for example, the case with $(h, n) = (200, 2000)$, where the empirical size of the asymptotic Wald test is 0.320. The max test and the bootstrapped Wald test are correctly sized for all cases.

Focusing on DGP-2, we observe that the max test dominates the bootstrapped Wald test in terms of empirical power. The empirical power of the max test is sometimes more than three times as high as the empirical power of the Wald test. For $(h, n) = (100, 1000)$, the empirical power is 0.833 for the max test and 0.258 for the Wald test.

An intuitive explanation for the remarkably high power of the max test is as follows. Performing the Wald test requires the computation of $\hat{\boldsymbol{\Gamma}} = (1/n)\sum_{t=1}^{n} \boldsymbol{x}_t\boldsymbol{x}_t'$. The resulting $\hat{\boldsymbol{\Gamma}}$ is more and more an imprecise estimator for $\boldsymbol{\Gamma}$ as the dimension h gets larger. The bootstrapped Wald test tends to lose finite sample power in such a case. The max test, instead, bypasses the computation of $\hat{\boldsymbol{\Gamma}}$ since it only requires the computation of $\hat{\Gamma}_{ii} = (1/n)\sum_{t=1}^{n} x_{it}^2$ for $i = 1, \ldots, h$. Since $\hat{\Gamma}_{ii}$ has a much smaller dimension than $\hat{\boldsymbol{\Gamma}}$, the accuracy of estimation improves considerably. Hence the max test keeps high power when the dimension h is relatively large. Large dimensionality is indeed an issue in many economic applications, and the max test resolves it.

Table 1: Rejection frequencies after Monte Carlo experiments

DGP-1: $y_t = \sum_{i=1}^{h} b_i x_{it} + \epsilon_t$ with $b_i = 0$

	$h = 10$		$h = 100$		$h = 200$	
	$n = 100$	$n = 500$	$n = 1000$	$n = 5000$	$n = 2000$	$n = 10000$
Max	.031	.051	.046	.054	.047	.033
Wald (asy)	.101	.052	.210	.085	.320	.083
Wald (boot)	.066	.050	.043	.058	.062	.061

DGP-2: $y_t = \sum_{i=1}^{h} b_i x_{it} + \epsilon_t$ with $b_i = I(i < \frac{1}{10}h) \times \frac{10}{5h}$

	$h = 10$		$h = 100$		$h = 200$	
	$n = 100$	$n = 500$	$n = 1000$	$n = 5000$	$n = 2000$	$n = 10000$
Max	.232	.949	.833	1.000	.992	1.000
Wald (asy)	.320	.879	.537	0.994	.748	1.000
Wald (boot)	.235	.877	.258	0.991	.368	1.000

5 Empirical application

We test for Granger causality from a weekly term spread (i.e. difference between long-term and short-term interest rates) to quarterly real GDP growth in the United States. 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. $\log x_L(\tau) - \log x_L(\tau - 4)$). The short and long term interest rates used for the term spread are respectively the effective 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. The sample period covers January 5, 1962 to December 31, 2013, which contains 2,736 weeks or 208 quarters.

The number of weeks contained in each quarter τ, denoted by $m(\tau)$, is not constant. While the max test can be applied with time-varying $m(\tau)$, we simplify the analysis by taking a sample average at the end of each quarter, resulting in the following modified term spread $\{x_H^*(\tau, i)\}_{i=1}^{12}$:

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

This modification gives us $n = 208$ quarters, $m = 12$, and $T = m \times n = 2,496$ weeks.

5.1 Models and tests

We perform a rolling window analysis to observe a dynamic evolution of Granger causality from term spread to GDP. We set a window size to be 80 quarters (i.e. 20 years) so that there are 129 windows in total. The first subsample covers the first quarter of 1962 through the fourth quarter of 1981 (written as 1962:I-1981:IV), the second one covers 1962:II-1982:I, and the last one covers 1994:I-2013:IV.

A full regression model with mixed frequency (MF) data is specified as

$$x_L(\tau) = \alpha_0 + \sum_{k=1}^{2} \alpha_k x_L(\tau - k) + \sum_{i=1}^{24} \beta_i x_H^*(\tau - 1, 12 + 1 - i) + u_L(\tau). \tag{5.1}$$

We are regressing the quarterly GDP growth $x_L(\tau)$ onto a constant, $p = 2$ quarters of lagged GDP growth, and $h_{MF} = 24$ weeks of lagged interest rate spread. We perform the Wald test based on model (5.1). P-values are computed after generating $B = 1000$ bootstrap samples based on Gonçalves and Kilian's (2004) bootstrap.

Naturally, parsimonious regression models with MF data are specified as

$$x_L(\tau) = \alpha_{0i} + \sum_{k=1}^{2} \alpha_{ki} x_L(\tau - k) + \beta_i x_H^*(\tau - 1, 12 + 1 - i) + u_{L,i}(\tau), \quad i = 1, \ldots, 24. \tag{5.2}$$

We perform the max test based on model (5.2). P-values are computed with $M = 100000$ draws from an approximation to the limit distribution under non-causality.

For comparison, we also perform low frequency (LF) analysis by aggregating the weekly term spread to the quarterly level. Aggregate the term spread as $x_H^*(\tau) = x_H^*(\tau, m)$. A full regression model with LF data is specified as

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

This model has $q = 2$ quarters of lagged x_L and $h_{LF} = 3$ quarters of lagged x_H^*. We perform the Wald test based on model (5.3) again with Gonçalves and Kilian's (2004) bootstrap.

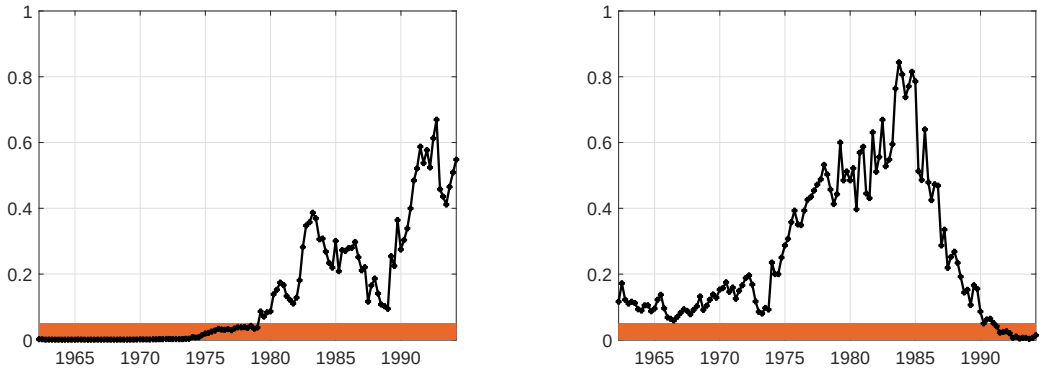
Similarly, parsimonious regression models with LF data are specified as $x_L(\tau) = \alpha_{0i} + \sum_{k=1}^{2} \alpha_{ki} x_L(\tau - i) + \beta_i x_H^*(\tau - i) + u_{L,i}(\tau)$ for $i = 1, 2, 3$. Perform the max test based on this model with $M = 100000$ draws from the limit distribution under non-causality.

5.2 Empirical results

Figure 1 plots rolling window p-values for each causality test over the 129 windows. All tests except for the MF Wald test find significant causality from term spread to GDP in early periods. Significant causality is detected up until 1979:I-1998:IV by the MF max test; 1975:III-1995:II by the LF max test; 1974:III-1994:II by the LF Wald test. The MF max test has the longest period of significant causality. 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.

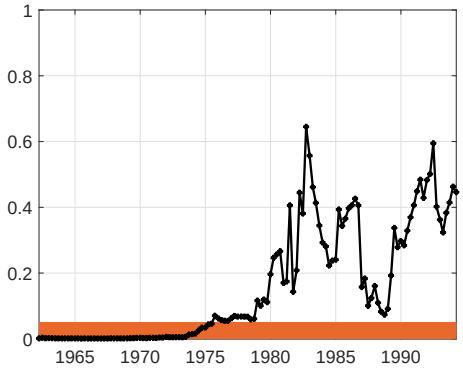
The MF Wald test, in contrast, suggests that there is significant causality only *after* subsample 1990:IV-2010:III, which is somewhat counter-intuitive. This result may stem from parameter proliferation. The full regression model with mixed frequency data has many more parameters than any other model. The results of the MF max test seem to be more intuitive and preferable than the results of the MF Wald test.

Figure 1: P-values for tests of non-causality from interest rate spread to GDP

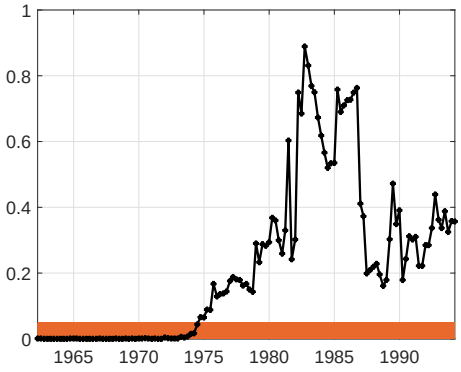


(a) Mixed frequency max test

(b) Mixed frequency Wald test



(c) Low frequency max test



(d) Low frequency Wald test

Rolling window p-values of each causality test are plotted. Each point of the x-axis represents the beginning date of a given window. The shaded area is $[0, 0.05]$, hence any p-value in that range indicates rejection of non-causality from the interest rate spread to GDP growth at the 5% level in a given window.

References

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.

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

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

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