GRANGER CAUSALITY IN MIXED FREQUENCY TIME SERIES

Kaiji Motegi

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Economics.

Chapel Hill
2014

Approved by:

Eric Ghysels

Michael Aguilar

Saraswata Chaudhuri

Neville Francis

Jonathan B. Hill

ABSTRACT

Kaiji Motegi: Granger Causality in Mixed Frequency Time Series
(Under the direction of Eric Ghysels)

It is a classic topic in time series econometrics to test Granger causality among multiple variables. While many Granger causality tests have been invented in the literature, they are often vulnerable to temporal aggregation which potentially generates or hides causality. Based on the growing literature of Mixed Data Sampling (MIDAS) analysis, this dissertation proposes a set of mixed frequency Granger causality tests which are robust against temporal aggregation. The mixed frequency causality tests take an explicit treatment of data sampled at different frequencies, and hence enable more accurate statistical inference than the conventional approach that aggregates all time series into the common lowest frequency.

Depending on the magnitude of the ratio of sampling frequencies, this dissertation proposes two types of mixed frequency causality tests. The first one handles a small ratio of sampling frequencies like month vs. quarter. Exploiting Ghysels' mixed frequency vector autoregressive (MF-VAR) models, we extend Dufour, Pelletier, and Renault's VAR-based causality test to the mixed frequency context. We prove that the mixed frequency approach better recovers the underlying causal patterns than the existing low frequency approach. Moreover, we demonstrate via local asymptotic power analysis and simulations that the mixed frequency test has higher power than the low frequency test in both large sample and small sample. In an empirical application on U.S. macroeconomy, we show that the mixed frequency approach and the low frequency approach produce very different causal implications, with the former yielding more intuitive results.

The second part of this dissertation deals with a relatively large ratio of sampling frequencies like month vs. year. Inspired by Sims' regression-based causality tests, we develop a new test that achieves higher power than the conventional test in both large sample and small sample. In this framework, a larger ratio of sampling frequencies is likely to improve power since our

methodology circumvents parameter proliferation. We apply our test to weekly interest rate spread and quarterly GDP in the U.S. The empirical result shows that the interest rate spread used to be a valid predictor of GDP but its predictability has declined more recently.

ACKNOWLEDGEMENTS

I am very grateful for the guidance I have received from the five dissertation committee members. First and foremost, I would like to thank my dissertation advisor, Dr. Eric Ghysels, for his great direction and patience throughout my Ph.D. research. I am largely indebted to Dr. Jonathan B. Hill, another coauthor of mine, for his expert help and suggestions over the last few years. Also, I learned a great deal of econometric theory and methods from Dr. Saraswata Chaudhuri as a student and teaching assistant of his advanced econometrics course. I thank Dr. Michael Aguilar for his detailed and inspiring suggestions, which have strengthened the empirical aspect of this research. I am thankful to Dr. Neville Francis for his professional comments and encouragement as the placement director. Besides these five professors, I also thank all participants of the UNC econometrics seminar, organized by Dr. Chaudhuri, for their helpful comments. Financial supports from the Institute of International Education (Fulbright Scholarship) and Yoshida Scholarship Foundation are greatly appreciated. Last but not least, I am grateful for Dr. Akira Sadahiro, Dr. Kenichiro Tamaki, and Dr. Hajime Yamashita at Waseda University, where I began my exciting academic journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

It is a classic topic in time series econometrics to test Granger's (1969) causality among multiple variables. Many kinds of Granger causality tests have been invented in the past fifty years, and the most prominent ones include Dufour, Pelletier, and Renault's (2006) test based on vector autoregression (VAR) models and Sims' (1972) regression-based test. A well-known problem of these existing tests is that they are often vulnerable to temporal aggregation which potentially generates or hides causality, as noted in Granger (1980) and Granger (1988) among many others. Such a misleading causality is called *spurious causality* (cfr. Dufour and Renault (1998)). Since economic time series are often sampled at different frequencies (e.g. daily financial variables, monthly business-cycle indicators, quarterly gross domestic product), we need new causality tests that can control spurious causality.

To this end, we propose mixed frequency causality tests based on the growing literature of Mixed Data Sampling (MIDAS) analysis. Originated with Ghysels, Santa-Clara, and Valkanov (2004), Ghysels, Santa-Clara, and Valkanov (2006), etc., the MIDAS approach works on data sampled at different frequencies instead of working on data aggregated to the common lowest frequency. By expanding the notion of Granger causality into the MIDAS framework, this dissertation establishes mixed frequency Granger causality tests which give us improved statistical accuracy, namely higher power, than the conventional low frequency approach does.

Depending on the magnitude of the ratio of sampling frequencies m, this dissertation proposes two types of mixed frequency causality tests. The first one handles a small m like month vs. quarter ($m = 3$). Exploiting Ghysels' (2012) mixed frequency vector autoregressive (MF-VAR) models, we extend Dufour, Pelletier, and Renault's (2006) VAR-based causality test to

the mixed frequency framework. We prove that the mixed frequency approach better recovers the underlying causal patterns than the existing low frequency approach. Moreover, we demonstrate via local asymptotic power analysis and Monte Carlo simulations that the mixed frequency test has higher power than the low frequency test in both large sample and small sample. In an empirical application involving U.S. macroeconomic indicators, we show that the mixed frequency approach and the low frequency approach produce very different causal implications, with the former yielding more intuitive results.

The second part of this dissertation deals with a relatively large ratio of sampling frequencies like month vs. year ($m = 12$). Inspired by Sims' (1972) regression-based causality tests and Andrews and Ploberger's (1994) optimal tests involving a nuisance parameter, we develop a new test that achieves higher power than the conventional test in both large sample and small sample. We combine multiple parsimonious regression models where the i-th model regresses a low frequency variable x_L onto the i-th high frequency lag or lead of a high frequency variable x_H for $i \in \{1, \ldots, h\}$. Let $\hat{\beta}_i$ be an estimator for the loading of the i-th high frequency lag or lead, then our test statistic basically takes the maximum among $\{\hat{\beta}_1^2, \ldots, \hat{\beta}_h^2\}$. In this framework, a larger m is likely to improve power since our methodology circumvents parameter proliferation. We apply our test to weekly interest rate spread and quarterly GDP in the U.S. The empirical result shows that the interest rate spread used to be a valid predictor of GDP but its predictability has declined more recently.

CHAPTER 2

VAR-BASED TEST

2.1 Introduction

It is well known that temporal aggregation may have spurious effects on testing for Granger causality, as noted by Clive Granger himself in a number of papers, see e.g. Granger (1980), Granger (1988), Granger (1995). In this paper we deal with what might be an obvious, yet largely overlooked remedy. Time series processes are often sampled at different frequencies and then typically aggregated to the common lowest frequency to test for Granger causality. The analysis of the present paper pertains to comparing testing for Granger causality with all series aggregated to the common lowest frequency, and testing for Granger causality taking advantage of all the series sampled at whatever frequency they are available. We rely on mixed frequency vector autoregressive models to implement a new class of Granger causality tests.[1]

We show that mixed frequency Granger causality tests better recover causality patterns in an underlying high frequency process compared to the traditional low frequency approach. We also formally prove that mixed frequency causality tests have higher asymptotic power against local alternatives and show via simulation that this also holds in finite samples involving realistic data generating processes. The simulations indicate that the mixed frequency VAR approach works well for small differences in sampling frequencies - like quarterly/monthly mixtures.

We apply the mixed frequency causality test to monthly inflation, monthly crude oil price

[1]MIDAS, meaning Mi(xed) Da(ta) S(ampling), regression models have been put forward in recent work by Ghysels, Santa-Clara, and Valkanov (2004), Ghysels, Santa-Clara, and Valkanov (2006) and Andreou, Ghysels, and Kourtellos (2010). See Andreou, Ghysels, and Kourtellos (2011) and Armesto, Engemann, and Owyang (2010) for surveys. VAR models for mixed frequency data were independently introduced by Anderson, Deistler, Felsenstein, Funovits, Zadrozny, Eichler, Chen, and Zamani (2012), Ghysels (2012) and McCracken, Owyang, and Sekhposyan (2013). An early example of related ideas appears in Friedman (1962). Foroni, Ghysels, and Marcellino (2013) provide a survey of mixed frequency VAR models and related literature.

fluctuations, the real GDP growth in the U.S. We also apply the conventional low frequency causality test to the aggregated quarterly price series and real GDP for comparison. These two approaches yield very different causal implications. In particular, significant causality from oil prices to inflation is detected by the mixed frequency approach but not by the low frequency approach. The result suggests that the quarterly frequency is too coarse to capture such causality.

The paper is organized as follows. In Section 2.2 we first briefly review the Granger causality and MIDAS literatures and then frame mixed frequency VAR models. In Section 2.3 we develop the mixed frequency causality tests. Section 2.4 discusses how we can recover underlying causality using a mixed frequency approach compared to a traditional low frequency approach. Section 2.5 shows that the mixed frequency causality tests have higher local asymptotic power than the low frequency ones do. Section 2.6 reports Monte Carlo simulation results and documents the finite sample power improvements achieved by the mixed frequency causality test. In Section 2.7 we apply the mixed frequency and low frequency causality tests to U.S. macroeconomic data. Finally, Section 2.8 provides some concluding remarks. All tables and Figures are provided after Section 2.8. Proofs for all theorems as well as some theoretical details are provided in Technical Appendices A.

2.2 Mixed Frequency Data Model Specifications

In this section we frame a mixed frequency vector autoregressive (henceforth MF-VAR) model and derive some asymptotic properties. We first provide a short review of the related literature. We then formally present the MF-VAR model. Finally, we establish large sample results for parameter estimators and corresponding Wald statistics.

We will use the following notational conventions throughout. Let $\boldsymbol{A} \in \mathcal{R}^{n\times l}$. The l_2-norm is $|\boldsymbol{A}| := (\sum_{i=1}^{n}\sum_{j=1}^{l} a_{ij}^2)^{1/2} = (\text{tr}[\boldsymbol{A}'\boldsymbol{A}])^{1/2}$; the L_r-norm is $\|\boldsymbol{A}\|_r := (\sum_{i=1}^{n}\sum_{j=1}^{l} E|a_{ij}|^r)^{1/r}$; the determinant is $\det(\boldsymbol{A})$; and the transpose is $\boldsymbol{A}'$. $\boldsymbol{0}_{n\times l}$ is an $n \times l$ matrix of zeros. $\boldsymbol{I}_K$ is the K-dimensional identity matrix. $\text{Var}[\boldsymbol{A}]$ is the variance-covariance matrix of a stochastic matrix $\boldsymbol{A}$. $\boldsymbol{B} \circ \boldsymbol{C}$ denotes element-by-element multiplication for conformable vectors $\boldsymbol{B}, \boldsymbol{C}$.

2.2.1 Brief Literature Review

The notion of causality introduced by Granger (1969) is defined in terms of incremental predictive ability, beyond the past observations of a time series process X, by past observations of another time series process Y. Although so-called *Granger causality* has been extended to fairly general settings including nonlinear and random volatility models, it is typically discussed in a linear regression framework, in particular since Sims (1972).

Early contributions by Zellner and Montmarquette (1971) and Amemiya and Wu (1972) pointed out the potentially adverse effects of temporal aggregation on testing for Granger causality. The subject has been extensively researched ever since, e.g. Granger (1980), Granger (1988), Lütkepohl (1993), Granger (1995), Renault, Sekkat, and Szafarz (1998), Marcellino (1999), Breitung and Swanson (2002), McCrorie and Chambers (2006), Silvestrini and Veredas (2008), among others. It is worth noting that whenever Granger causality and temporal aggregation are discussed, it is typically done in a setting where *all* series are subject to temporal aggregation. In such a setting it is well-known that even the simplest models, like a bivariate VAR(1) with stock (or skipped) sampling, may suffer from spuriously hidden or generated causality, and recovering the original causal pattern is very hard or even impossible in general.

As in the single frequency VAR literature, exploring mixed frequency Granger causality among more than two variables invariably relates to the notion of multi-horizon causality studied by Lütkepohl (1993), Dufour and Renault (1998) and Hill (2007). Of direct interest to us is Dufour and Renault (1998) who generalized the original definition of single-horizon or short run causality to multiple-horizon or long run causality to handle causality chains: in the presence of an auxiliary variable Z, Y may be useful for a multiple-step ahead prediction of X even if it is useless for the one-step ahead prediction. Dufour and Renault (1998) formalize the relationship between VAR coefficients and multiple-horizon causality and Dufour, Pelletier, and Renault (2006) formulate accordingly single step Wald tests of multiple-horizon non-causality. Their framework will be used extensively in our analysis. See Hill (2007) for a sequential method of testing for multiple-horizon non-causality.

In addition to the causality literature, the present paper also draws upon and contributes to the MIDAS literature originated by Ghysels, Santa-Clara, and Valkanov (2004) and Ghysels,

Santa-Clara, and Valkanov (2005). A number of papers have linked MIDAS regressions to (latent) high frequency VAR models, such as Foroni, Marcellino, and Schumacher (2013) and Kuzin, Marcellino, and Schumacher (2011), whereas Ghysels (2012) discusses the link between mixed frequency VAR models and MIDAS regressions. None of these papers study in any detail the issue of Granger causality.

2.2.2 Mixed Frequency VAR Models

We want to characterize three settings which we will refer to as HF, MF and LF - respectively high, mixed and low frequency. We begin by considering a partially latent underlying HF process. Using the notation of Ghysels (2012), the HF process contains $\{\{\boldsymbol{x}_H(\tau_L,k)\}_{k=1}^m\}_{\tau_L}$ and $\{\{\boldsymbol{x}_L(\tau_L,k)\}_{k=1}^m\}_{\tau_L}$, where $\tau_L \in \{0,\ldots,T_L\}$ is the LF time index (e.g. quarter), $k \in \{1, \ldots,m\}$ denotes the HF (e.g. month), and m is the number of HF time periods between LF time indices. In the month versus quarter case, for example, m equals three since one quarter has three months. Observations $\boldsymbol{x}_H(\tau_L,k) \in \mathcal{R}^{K_H\times 1}$, $K_H \geq 1$, are called HF variables, whereas $\boldsymbol{x}_L(\tau_L,k) \in \mathcal{R}^{K_L\times 1}$, $K_L \geq 1$, are latent LF variables because they are not observed at high frequencies - as only some temporal aggregates are available.

Note that two simplifying assumptions have implicitly been made. First, there are assumed to be only two sampling frequencies. Second, it is assumed that m is fixed and does not depend on τ_L. Both assumptions can be relaxed at the cost of much more complex notation and algebra which we avoid for expositional purpose - again see Ghysels (2012).

In reality the analyst's choice is limited to MF and LF cases. Only low frequency variables have been aggregated from a latent HF process in a MF setting, whereas both low *and* high frequency variables are aggregated from the latent HF process to form a LF process. Following Lütkepohl (1987) we consider only linear aggregation schemes involving weights $\boldsymbol{w} = [w_1,\ldots, w_m]'$ such that:

$$\boldsymbol{x}_H(\tau_L) = \sum_{k=1}^{m} w_k \boldsymbol{x}_H(\tau_L,k) \quad \text{and} \quad \boldsymbol{x}_L(\tau_L) = \sum_{k=1}^{m} w_k \boldsymbol{x}_L(\tau_L,k). \tag{2.2.1}$$

Two cases are of special interest given their broad use: (1) *stock* or *skipped* sampling, where

$w_k = I(k = m)$; and (2) *flow* sampling, where $w_k = 1$ for $k = 1, \ldots, m$.[2] In summary, we observe:

- all high and low frequency variables $\{\{\boldsymbol{x}_H(\tau_L, j)\}_{j=1}^m\}_{\tau_L}$ and $\{\{\boldsymbol{x}_L(\tau_L, j)\}_{j=1}^m\}_{\tau_L}$ in a HF process;
- all high frequency variables $\{\{\boldsymbol{x}_H(\tau_L, j)\}_{j=1}^m\}_{\tau_L}$ but only aggregated low frequency variables $\{\boldsymbol{x}_L(\tau_L)\}_{\tau_L}$ in a MF process;
- only aggregated high and low frequency variables $\{\boldsymbol{x}_H(\tau_L)\}_{\tau_L}$ and $\{\boldsymbol{x}_L(\tau_L)\}_{\tau_L}$ in a LF process.

A key idea of MF-VAR models is to stack everything observable given a MF process according to their order over time. This results in the following $K = K_L + mK_H$ dimensional vector:

$$\boldsymbol{X}(\tau_L) = [\boldsymbol{x}_H(\tau_L, 1)', \ldots, \boldsymbol{x}_H(\tau_L, m)', \boldsymbol{x}_L(\tau_L)']'. \tag{2.2.2}$$

Note that $\boldsymbol{x}_L(\tau_L)$ is the last block in the stacked vector - a conventional assumption implying that it is observed after $\boldsymbol{x}_H(\tau_L, m)$. Any other order is conceptually the same, except that it implies a different timing of information about the respective processes. We will work with the specification appearing in (2.2.2) as it is most convenient.

Example 1 : *Quarterly Real GDP* : A leading example of how a mixed frequency model is useful in macroeconomics concerns quarterly real GDP growth $x_L(\tau_L)$, where existing studies of causal patterns use monthly unemployment, oil prices, inflation, interest rates, etc. aggregated into quarters (see Hill (2007) for references). Consider the monthly oil price changes and CPI inflation $[\boldsymbol{x}_H(\tau_L, 1)', \ldots, \boldsymbol{x}_H(\tau_L, 3)']'$, which will be actually analyzed in Section 2.7. Note that τ_L represents a quarter and $\boldsymbol{x}_H$ is a 2×1 vector. According to the Bureau of Economic Analysis, GDP is announced in advance roughly one month after the quarter, with subsequent updates over the following two months (e.g. the 2014 first quarter advanced estimate is due April 30, 2014). By comparison, oil prices are available on a daily basis and hence their monthly data

[2]One can equivalently let $w_k = 1/m$ for $k = 1, \ldots, m$ in flow sampling if the average is preferred to a summation.

can be calculated immediately after the month. Also, the monthly CPI is announced roughly three weeks after the month. Since the two monthly series are announced before the GDP, the ordering is exactly as shown in (2.2.2).

In order to proceed, we will make a number of standard regulatory assumptions. Let $\mathcal{F}_{\tau_L} \equiv \sigma(\boldsymbol{X}(t) : t \leq \tau_L)$. In particular we assume $E[\boldsymbol{X}(\tau_L)|\mathcal{F}_{\tau_L-1}]$ has a version that is *almost surely* linear in $\{\boldsymbol{X}(\tau_L - 1), ..., \boldsymbol{X}(\tau_L - p)\}$ for some finite $p \geq 1$.

Assumption 2.2.1. The process $\boldsymbol{X}(\tau_L)$ is governed by a VAR(p) for some $p \geq 1$:

$$\boldsymbol{X}(\tau_L) = \sum_{k=1}^{p} \boldsymbol{A}_k \boldsymbol{X}(\tau_L - k) + \boldsymbol{\epsilon}(\tau_L). \tag{2.2.3}$$

The coefficients $\boldsymbol{A}_k$ are $K \times K$ matrices for $k = 1, \ldots, p$. The $K \times 1$ error vector $\boldsymbol{\epsilon}(\tau_L) = [\epsilon_1(\tau_L), \ \ldots, \ \epsilon_K(\tau_L)]'$ is a strictly stationary martingale difference with respect to increasing $\mathcal{F}_{\tau_L} \subset \mathcal{F}_{\tau_L+1}$, where $\boldsymbol{\Omega} \equiv E[\boldsymbol{\epsilon}(\tau_L)\boldsymbol{\epsilon}(\tau_L)']$ is positive definite.

Remark 1. Martingale difference errors $E[\boldsymbol{\epsilon}(\tau_L)|\mathcal{F}_{\tau_L-1}] = \mathbf{0}$ allow for conditional heteroskedasticity of unknown form. Nevertheless, in order to test for non-causality using (2.2.3) we estimate a parameter subset from a (p, h)-autoregression defined in (2.2.5), below. Asymptotics for M-estimators of the resulting parameter involve finite sums of martingale differences which are not in general martingale differences, and anyway the martingale difference property alone does not suffice for Gaussian asymptotics of M-estimators (cf. McLeish (1974), Hall and Heyde (1980)). We therefore impose a mixing condition in Assumption 2.2.3 below.

Remark 2. Unless $\{\boldsymbol{\epsilon}(\tau_L)\}$ is an i.i.d. process, the VAR coefficients $\boldsymbol{A}_k$ do not necessarily carry all the usual information about higher order causation, including volatility spillover (cfr. King, Sentana, and Wadhwani (1994), Caporale, Pittis, and Spagnolo (2006)). This is irrelevant for our purposes, however, because in the tradition of Dufour and Renault (1998) our analysis is primarily about deducing nonlinear restrictions on $\{\boldsymbol{A}_1, ..., \boldsymbol{A}_p\}$ that relate information about first order predictive ability in $\boldsymbol{X}(\tau_L)$, and about recovering information on (non-)causation in HF-VAR by using MF- or LF-VAR models. Nevertheless, without independence the close relationship between Granger's (1969) and Sims' (1972) notions of causality in terms of linear predictive improvement breaks down, as shown in Florens and Mouchart (1982).

Remark 3. We assume the lag order p is either known, or the true order resulting in a martingale difference error $\boldsymbol{\epsilon}(\tau_L)$ is at least as large as p. In practice standard methods for selecting the lag apply in a mixed frequency environment, including tests of white noise. Indeed, in general regression model specification tests easily extend to mixed frequency data. A large lag order and/or a large number of variables, moreover, may lead to empirical size distortions in our asymptotic chi-squared test statistic. This is particularly relevant in a mixed frequency VAR since m and therefore K may be large. This topic is well known with bootstrap-based solutions (e.g. Dufour, Pelletier, and Renault (2006) when regression errors are i.i.d, and Gonçalves and Killian (2004) when errors may be heteroskedastic of unknown form). See Section 2.6, below where simulation evidence clearly shows a bootstrap approach for approximating our test statistic's critical values work well.

In addition, the following standard assumptions ensure stationarity and α-mixing of the observed time series and the MF-VAR errors.[3] Define $\mathcal{G}_s^t \equiv \sigma(\{\boldsymbol{X}(i), \boldsymbol{\epsilon}(i)\} : s \leq i \leq t)$ and mixing coefficients $\alpha_h \equiv \sup_{\mathcal{A} \subset \mathcal{G}_{-\infty}^t, \mathcal{B} \subset \mathcal{G}_{t+h}^{\infty}} |P(\mathcal{A} \cap \mathcal{B}) - P(\mathcal{A})P(\mathcal{B})|$ (cfr. Rosenblatt (1956) and Ibragimov (1975)).

Assumption 2.2.2. All roots of the polynomial $\det(\boldsymbol{I}_K - \sum_{k=1}^{p} \boldsymbol{A}_k z^k) = 0$ lie outside the unit circle.

Assumption 2.2.3. $\boldsymbol{X}(\tau_L)$ and $\boldsymbol{\epsilon}(\tau_L)$ are α-mixing: $\sum_{h=0}^{\infty} \alpha_{2^h} < \infty$.

Remark 4. Recall that α-mixing implies mixing in the ergodic sense, and therefore ergodicity (see Petersen (1983)). Hence by Assumptions 2.2.1-2.2.3 $\{\boldsymbol{X}(\tau_L), \boldsymbol{\epsilon}(\tau_L)\}$ are stationary and ergodic.

Remark 5. Asymptotics for our estimator only requires $\sum_{h=0}^{\infty} \alpha_{2^h} < \infty$ because under Assumptions 2.2.1 and 2.2.2 $\boldsymbol{X}(\tau_L)$ has a positive bounded spectral density (cfr. Ibragimov (1975)). This allows for geometric memory decay $\alpha_h = O(\rho^h)$ for some $\rho \in (0, 1)$, as well as much slower decay and therefore persistence since $\alpha_h = O(h^{-\iota})$ for tiny $\iota > 0$ suffices for Gaussian asymptotics (cfr. Ibragimov (1975)). Our Wald statistic requires a greater constraint on the

[3] Although a large body of literature exists on Granger causality in non-stationary or cointegrated systems (e.g. Yamamoto and Kurozumi (2006)), the generalization is beyond the scope of this paper.

mixing coefficients α_h due to a kernel variance estimator: see Theorem 2.2.2 and Appendix A.1.1, below.

Remark 6. A mixing property for the scalar components of the error $\boldsymbol{\epsilon}(\tau_L)$ covers a great variety of conditional volatility processes including GARCH and many asymmetric GARCH processes (e.g. Boussama (1998), Carrasco and Chen (2002), Meitz and Saikkonen (2008)). Conditions for geometric ergodicity and therefore α-mixing for the BEKK class of multivariate strong GARCH(p, q) processes are known and carry over to a latent HF multivariate GARCH(p, q) process (see Boussama, Fuchs, and Stelzer (2011)). Any finite lag measurable transform of α-mixing $\boldsymbol{\epsilon}(\tau_L)$ is α-mixing, hence the mixing property for the HF process carries over to the MF process. If $\boldsymbol{\epsilon}(\tau_L)$ is i.i.d. and has a continuous bounded joint distribution then from stationarity Assumption 2.2.2 it follows $\boldsymbol{X}(\tau_L)$ is geometrically α-mixing (see § 2.3.1 in Doukhan (1994)). Otherwise in general an α-mixing property for $\boldsymbol{\epsilon}(\tau_L)$ implies $\boldsymbol{X}(\tau_L)$ is also α-mixing when the joint distribution of $\boldsymbol{\epsilon}(\tau_L)$ conditional on its history is absolutely continuous and bounded *almost surely* (see § 2.3.2 in Doukhan (1994)).

Note also that we do not include a constant term in (2.2.3) solely to reduce notation, thus $\boldsymbol{X}(\tau_L)$ should be thought of as a de-meaned process. Finally, it is straightforward to allow an infinite order VAR structure, and estimate a truncated finite order VAR model as in Lewis and Reinsel (1985), Lütkepohl and Poskitt (1996), and Saikkonen and Lütkepohl (1996).

2.2.3 Estimators and Their Large Sample Properties

If the VAR(p) model appearing in (2.2.3) were standard, then the off-diagonal elements of any matrix $\boldsymbol{A}_k$ would tell us something about causal relationships for some specific horizon. The fact that MF-VAR models involve stacked replicas of high frequency data sampled across different (high frequency) periods implies that potentially multi-horizon causal patterns reside inside any of the matrices $\boldsymbol{A}_k$. It is therefore natural to start with a multi-horizon setting. We do so, at first, focusing on multiple low frequency prediction horizons which we will denote by $h \in \mathcal{N}$.[4]

[4]Another reason for studying multiple horizons is the potential of causality chains when $K_H > 1$ or $K_L > 1$. Note, however, that despite the MF-VAR being by design multi-dimensional there are no causality chains when $K_H = K_L = 1$ since the $m \times 1$ vector of the high frequency observations refers to a single variable.

It is convenient to iterate (2.2.3) over the desired test horizon in order to deduce simple testable parameter restrictions for non-causality. Recall that under Assumption 2.2.2 a unique stationary and ergodic solution to (2.2.3) exists:

$$\boldsymbol{X}(\tau_L) = \sum_{k=0}^{\infty} \boldsymbol{\Psi}_k \boldsymbol{\epsilon}(\tau_L - k), \tag{2.2.4}$$

where $\boldsymbol{\Psi}_k$ satisfies $\boldsymbol{\Psi}_0 = \boldsymbol{I}_K$, $\boldsymbol{\Psi}_k = \sum_{s=1}^{p} \boldsymbol{A}_s \boldsymbol{\Psi}_{k-s}$ for $k \geq 1$ and $\boldsymbol{\Psi}_k = \boldsymbol{0}_{K\times K}$ for $k < 0$, and $|\boldsymbol{\Psi}_k| = O(\rho^k)$ for some $\rho \in (0,1)$. We then have what Dufour, Pelletier, and Renault (2006) labeled as a (p,h)-autoregression:

$$\boldsymbol{X}(\tau_L + h) = \sum_{k=1}^{p} \boldsymbol{A}_k^{(h)} \boldsymbol{X}(\tau_L + 1 - k) + \sum_{k=0}^{h-1} \boldsymbol{\Psi}_k \boldsymbol{\epsilon}(\tau_L + h - k), \tag{2.2.5}$$

where

$$\boldsymbol{A}_k^{(1)} = \boldsymbol{A}_k \text{ and } \boldsymbol{A}_k^{(i)} = \boldsymbol{A}_{k+i-1} + \sum_{l=1}^{i-1} \boldsymbol{A}_{i-l} \boldsymbol{A}_k^{(l)} \text{ for } i \geq 2.$$

By convention $\boldsymbol{A}_k = \boldsymbol{0}_{K\times K}$ whenever $k > p$. The MF-VAR causality test exploits Wald statistics based on the OLS estimator of the (p,h)-autoregression parameter set

$$\boldsymbol{B}(h) = \left[\boldsymbol{A}_1^{(h)}, \ldots, \boldsymbol{A}_p^{(h)}\right]' \in \mathcal{R}^{pK\times K}. \tag{2.2.6}$$

If all variables were aggregated into a common low frequency and expanded into a (p,h)-autoregression, then h-step ahead non-causality has a simple parametric expression in terms of $\boldsymbol{B}(h)$; cfr. Dufour, Pelletier, and Renault (2006). Recall, however, that the MF-VAR has a special structure because of the stacked HF vector. This implies that the Wald-type test for non-causality that we derive is slightly more complicated than those considered by Dufour, Pelletier, and Renault (2006) since in MF-VAR models the restrictions will often deal with linear parametric restrictions across multiple equations. Nevertheless, in a generic sense we show in Section 2.3 that non-causality between any set of variables in a MF-VAR model can be expressed as linear constraints with respect to $\boldsymbol{B}(h)$. Hence, the null hypothesis of interest is a linear restriction:

$$H_0(h) : \boldsymbol{R}\text{vec}\left[\boldsymbol{B}(h)\right] = \boldsymbol{r}, \tag{2.2.7}$$

where $\boldsymbol{R}$ is a $q \times pK^2$ selection matrix of full row rank q, and $\boldsymbol{r} \in \mathcal{R}^q$. We leave complete details of the construction of $\boldsymbol{R}$ for Section 2.3.

The OLS estimator $\hat{\boldsymbol{B}}(h)$ of $\boldsymbol{B}(h)$ is

$$\hat{\boldsymbol{B}}(h) \equiv \underset{\boldsymbol{B}(h)}{arg\ min} \left\{ \text{vec}\left[\boldsymbol{U}_h(h)\right]' \text{vec}\left[\boldsymbol{U}_h(h)\right] \right\} = \left[\overline{\boldsymbol{W}}_p(h)'\overline{\boldsymbol{W}}_p(h)\right]^{-1} \overline{\boldsymbol{W}}_p(h)'\boldsymbol{W}_h(h),$$

where $\boldsymbol{U}_h(h)$ is a matrix of stacked sums of $\{\boldsymbol{\Psi}_k\}$ and $\{\boldsymbol{\epsilon}(\tau_L)\}$ while $\overline{\boldsymbol{W}}_p(h)$ and $\boldsymbol{W}_h(h)$ are matrices of stacked $\{\boldsymbol{X}(\tau_L)\}$. See Appendix A.1.1 for derivation of $\{\boldsymbol{U}_h(h), \overline{\boldsymbol{W}}_p(h), \boldsymbol{W}_h(h)\}$.

Assumptions 2.2.1-2.2.3 suffice for $\hat{\boldsymbol{B}}(h)$ to be consistent for $\boldsymbol{B}(h)$ and asymptotically normal. Limits are with respect to $T_L \to \infty$ hence $T_L^* \to \infty$, where $T_L^* = T_L - h + 1$ is the effective sample size for the (p, h)-autoregression.

Theorem 2.2.1. Under Assumptions 2.2.1-2.2.3 $\hat{\boldsymbol{B}}(h) \xrightarrow{p} \boldsymbol{B}(h)$ and

$$\sqrt{T_L^*}\text{vec}\left[\hat{\boldsymbol{B}}(h) - \boldsymbol{B}(h)\right] \xrightarrow{d} N\left(\boldsymbol{0}_{pK^2 \times 1}, \boldsymbol{\Sigma}_p(h)\right), \tag{2.2.8}$$

where $\boldsymbol{\Sigma}_p(h)$ is positive definite.

Remark 7. See Appendix A.1.2 for a proof, and see Appendices A.1.1-A.1.2 for a complete characterization of the asymptotic covariance matrix $\boldsymbol{\Sigma}_p(h)$.

If we have a consistent estimator $\hat{\boldsymbol{\Sigma}}_p(h)$ for $\boldsymbol{\Sigma}_p(h)$ which is *almost surely* positive semi-definite for $T_L^* \geq 1$, we can define the Wald statistic

$$W[H_0(h)] \equiv T_L^* \left(\boldsymbol{R}\text{vec}\left[\hat{\boldsymbol{B}}(h)\right] - \boldsymbol{r}\right)' \times \left(\boldsymbol{R}\hat{\boldsymbol{\Sigma}}_p(h)\boldsymbol{R}'\right)^{-1} \times \left(\boldsymbol{R}\text{vec}\left[\hat{\boldsymbol{B}}(h)\right] - \boldsymbol{r}\right). \tag{2.2.9}$$

Implicitly, of course, $\boldsymbol{R}\hat{\boldsymbol{\Sigma}}_p(h)\boldsymbol{R}'$ must be non-singular for any $\boldsymbol{R} \in \mathcal{R}^{q \times pK^2}$ with full row rank. In view of positive definiteness of $\boldsymbol{\Sigma}_p(h)$ by Theorem 2.2.1, and the supposition $\hat{\boldsymbol{\Sigma}}_p(h) = \boldsymbol{\Sigma}_p(h)$ + $o_p(1)$, it follows $(\boldsymbol{R}\hat{\boldsymbol{\Sigma}}_p(h)\boldsymbol{R}')^{-1}$ is well defined asymptotically with probability approaching one.

We therefore obtain the following result, which we prove in Appendix A.1.2.

Theorem 2.2.2. Let $\hat{\boldsymbol{\Sigma}}_p(h)$ be a consistent estimator for $\boldsymbol{\Sigma}_p(h)$ that is *almost surely* positive semi-definite for $T_L^* \geq 1$. Then under Assumptions 2.2.1-2.2.3, $W[H_0(h)] \xrightarrow{d} \chi_q^2$ under $H_0(h)$.

Remark 8. A consistent, *almost surely* positive semi-definite estimator $\hat{\boldsymbol{\Sigma}}_p(h)$ is easily constructed by using Newey and West's (1987) HAC estimator, given the stronger moment and mixing conditions $||\boldsymbol{\epsilon}(\tau_L)||^{4+\delta} < \infty$ and $\alpha_h = O(h^{(4+\delta)\backslash\delta})$ for some $\delta > 0$. See Appendix A.1.1 for complete details.

In the remainder of the paper we will provide various tests for Granger causality which are special cases of the generic framework derived so far.

2.3 Testing Causality with Mixed Frequency Data

In this section we define non-causality when data are sampled at mixed frequencies and describe Wald-type tests associated with it. We first cover some preliminary notions of multiple-horizon causality and extend it to the mixed sampling frequency case. We discuss in detail testing non-causality from one variable to another, and whether they are high or low frequency variables. We also cover non-causality from all high frequency variables to all low frequency variables and vice versa, cases for which we give explicit formulae for the selection matrix $\boldsymbol{R}$ used in the null hypothesis (2.2.7) and test statistic (2.2.9).

2.3.1 Preliminaries

We start with adopting the notion of non-causality to a mixed sampling frequency data filtration setting. Using the notation of Dufour and Renault (1998) we define the relevant information sets for the purpose of characterizing non-causality. In particular, let L^2 be a Hilbert space of covariance stationary real-valued random variables defined on a common probability space, and the covariance as inner product. Moreover, let $\mathcal{I}(\tau_L)$ be a closed increasing subspace of L^2 such that $\mathcal{I}(\tau_L) \subset \mathcal{I}(\tau_L')$ whenever $\tau_L < \tau_L'$, where $\tau_L, \tau_L' \in \mathcal{Z}$.

Furthermore, define the indices $i \in \{1, \ldots, K_H\}$ and $j \in \{1, \ldots, K_L\}$, and write $\tilde{\boldsymbol{x}}_{H,i}(\tau_L) = [x_{H,i}(\tau_L, 1), \ldots, x_{H,i}(\tau_L, m)]'$. Note that $\tilde{\boldsymbol{x}}_{H,i}(\tau_L)$ is a vector *stacking* all m observations of the i-th high frequency variable available at period τ_L. We are putting a tilde in order to distinguish it from the *aggregated* high frequency variable $x_{H,i}(\tau_L) = \sum_{k=1}^{m} w_k x_{H,i}(\tau_L, k)$ defined in (2.2.1). Similarly, let $x_{L,j}(\tau_L)$ be a scalar low frequency observation of the j-th low frequency variable in period τ_L.

Denote by $\boldsymbol{x}(-\infty, \tau_L]$ the Hilbert space spanned by $\{\boldsymbol{x}(\tau)|\ \tau \leq \tau_L\}$. The information set $\mathcal{I}$ is said to be *conformable* with $\boldsymbol{x}$ if $\boldsymbol{x}(-\infty, \tau_L] \subset \mathcal{I}(\tau_L)$ for all τ_L. We call the information set derived from $\mathcal{I}(\tau_L) = \boldsymbol{X}(-\infty, \tau_L]$, where $\boldsymbol{X}(\tau_L)$ is given in (2.2.2), as the *MF reference information set in period* τ_L, whereas $\mathcal{I} = \{\mathcal{I}(\tau_L)|\ \tau_L \in \mathcal{Z}\}$ is the *MF reference information set*. Therefore, the only information available up to period τ_L is the high frequency observations of all high frequency variables and the low frequency observations of all low frequency variables. In addition, let $\mathcal{I}_{(H,i)}$ denote the MF reference information set except for the i-th high frequency variable $x_{H,i}$, and let $\mathcal{I}_{(L,j)}$ denote the information set except for $x_{L,j}$. Similarly, $\mathcal{I}_{(H)}$ is the MF reference information set except for all high frequency variables $x_{H,1}, \ldots, x_{H,K_H}$. $\mathcal{I}_{(L)}$ is the MF reference information set except for all low frequency variables $x_{L,1}, \ldots, x_{L,K_L}$. Note that since the stacked high frequency observations $\tilde{\boldsymbol{x}}_{H,i}(\tau_L)$ and the low frequency observation $x_{L,j}(\tau_L)$ belong to $\boldsymbol{X}(\tau_L)$ for all $i \in \{1, \ldots, K_H\}$ and $j \in \{1, \ldots, K_L\}$, it is clear that the MF reference information set $\mathcal{I} = \{\mathcal{I}(\tau_L)|\ \tau_L \in \mathcal{Z}\}$ is conformable with $\tilde{\boldsymbol{x}}_{H,i}(\tau_L)$ and $x_{L,j}(\tau_L)$.

Finally, let E and F be two subspaces of L^2, and let $E + F$ denote the Hilbert subspace generated by the elements of E and F. Let $P[\boldsymbol{x}(\tau_L + h)|\ \mathcal{I}(\tau_L)]$ be the best linear forecast of $\boldsymbol{x}(\tau_L + h)$ based on $\mathcal{I}(\tau_L)$ in the sense of a covariance orthogonal projection.

For any generic information set and pair of processes (high or low frequency) the notion of non-causality is defined as follows.

Definition 2.3.1. (Non-causality at Different Horizons). Suppose that $\mathcal{I}$ is conformable with $\boldsymbol{x}$. (i) $\boldsymbol{y}$ *does not cause* $\boldsymbol{x}$ *at horizon* h *given* $\mathcal{I}$ (denoted by $\boldsymbol{y} \nrightarrow_h \boldsymbol{x}|\ \mathcal{I}$) if:

$$P[\boldsymbol{x}(\tau_L + h)|\mathcal{I}(\tau_L)] = P[\boldsymbol{x}(\tau_L + h)|\mathcal{I}(\tau_L) + \boldsymbol{y}(-\infty, \tau_L]] \qquad \forall \tau_L \in \mathcal{Z}.$$

Moreover, (ii) $\boldsymbol{y}$ *does not cause* $\boldsymbol{x}$ *up to horizon* h *given* $\mathcal{I}$ (denoted by $\boldsymbol{y} \nrightarrow_{(h)} \boldsymbol{x}|\ \mathcal{I}$) if $\boldsymbol{y} \nrightarrow_k \boldsymbol{x}|\ \mathcal{I}$ for all $k \in \{1, \ldots, h\}$.

Definition 2.3.1 applies to a mixed sampling frequency setting when suitable information set and processes are used.[5] Consider, for example, non-causality from the j-th low frequency variable $x_{L,j}$ to the i-th high frequency variable $x_{H,i}$. We say $x_{L,j}$ *does not cause* $x_{H,i}$ *at horizon*

[5]Definition 2.3.1 corresponds to Definition 2.2 in Dufour and Renault (1998) for covariance stationary processes.

h given $\mathcal{I}$ (denoted by $x_{L,j} \nrightarrow_h x_{H,i} | \, \mathcal{I}$) if $P[\tilde{\boldsymbol{x}}_{H,i}(\tau_L+h) | \, \mathcal{I}_{(L,j)}(\tau_L)] = P\,[\tilde{\boldsymbol{x}}_{H,i}(\tau_L+h) | \, \mathcal{I}(\tau_L)\,]$ for all $\tau_L \in \mathcal{Z}$. A key here is that we treat the m-dimensional stacked vector of $x_{H,i}$ as one block. This treatment allows us to apply Definition 2.3.1 to mixed frequency frameworks without any theoretical complications.

When we consider non-causality between a pair of high frequency series, namely $x_{H,i_1} \nrightarrow_h x_{H,i_2} | \, \mathcal{I}$ for $i_1, i_2 \in \{1, \ldots, K_H\}$, it should be noted that we focus exclusively on low frequency horizons h, or equivalently horizons $h \times m$. Any other horizon, not a multiple of m, is not considered here. They can be handled with the existing same frequency setting of Dufour and Renault (1998).

We often treat all K_H high frequency variables as a group and all K_L low frequency variables as the other group, so provide the explicit definition of non-causality in such a case. We say all low frequency variables do not cause all high frequency variables at horizon h given $\mathcal{I}$ (denoted by $\boldsymbol{x}_L \nrightarrow_h \boldsymbol{x}_H | \, \mathcal{I}$) if $P[\tilde{\boldsymbol{x}}_H(\tau_L + h) | \, \mathcal{I}_{(L)}(\tau_L)] = P\,[\tilde{\boldsymbol{x}}_H(\tau_L + h) | \, \mathcal{I}(\tau_L)\,]$ for all $\tau_L \in \mathcal{Z}$, where $\tilde{\boldsymbol{x}}_H(\tau_L) = [\tilde{\boldsymbol{x}}_{H,1}(\tau_L)', \ldots, \tilde{\boldsymbol{x}}_{H,K_H}(\tau_L)']'$.

In summary, there are six basic cases to consider in a mixed frequency setting.

Case 1 (low to low) Non-causality from the j_1-th *low* frequency variable, x_{L,j_1}, to the j_2-th *low* frequency variable, x_{L,j_2}, at horizon h. The null hypothesis is written as $H_0^1(h) : x_{L,j_1} \nrightarrow_h x_{L,j_2} | \mathcal{I}$.

Case 2 (high to low) $H_0^2(h) : x_{H,i_1} \nrightarrow_h x_{L,j_1} | \, \mathcal{I}$.

Case 3 (low to high) $H_0^3(h) : x_{L,j_1} \nrightarrow_h x_{H,i_1} | \, \mathcal{I}$.

Case 4 (high to high) $H_0^4(h) : x_{H,i_1} \nrightarrow_h x_{H,i_2} | \, \mathcal{I}$.

Case I (all high to all low) Non-causality from all high frequency variables $x_{H,1}, \ldots, x_{H,K_H}$ to all low frequency variables $x_{L,1}, \ldots, x_{L,K_L}$ at horizon h. The null hypothesis is written as $H_0^I(h) : \boldsymbol{x}_H \nrightarrow_h \boldsymbol{x}_L | \, \mathcal{I}$.

Case II (all low to all high) $H_0^{II}(h) : \boldsymbol{x}_L \nrightarrow_h \boldsymbol{x}_H | \, \mathcal{I}$.

Cases 1 through 4 handle individual variables, while Cases I and II handle entire groups of variables. In the sequel we often consider Cases I and II for simplicity since - viewed as

a bivariate system - causality chains can be excluded in both cases. In the bivariate system non-causality at one horizon is synonymous to non-causality at all horizons (see Dufour and Renault (1998: Proposition 2.3), cfr. Florens and Mouchart (1982: p. 590)). In order to avoid tedious matrix notation, we do not treat in detail cases involving non-causation from a subset of all variables to another subset. Our results straightforwardly apply, however, in such cases as well.

2.3.2 Causality Tests in Mixed Frequency VAR Models

Our next task is to construct the selection matrices $\boldsymbol{R}$ for the various null hypotheses (2.2.7) associated with the six generic cases. This requires deciphering parameter restrictions for non-causation based on the (p, h)-autoregression appearing in equation (2.2.5).

Characterizing restrictions on $\boldsymbol{A}_k^{(h)}$ for each case above requires some additional matrix notation. Let $\boldsymbol{N} \in \mathcal{R}^{n\times n}$, and let $a, b, c, d, \iota, \iota' \in \{1, \ldots, n\}$ with $a \leq b$, $c \leq d$, and $(b-a)/\iota$ and $(d-c)/\iota'$ being nonnegative integers. Then we define $\boldsymbol{N}(a : \iota : b, c : \iota' : d)$ as the $(\frac{b-a}{\iota}+1)\times(\frac{d-c}{\iota'}+1)$ matrix which consists of the a-th, $(a+\iota)$-th, $(a+2\iota)$-th, ..., b-th rows and c-th, $(c+\iota')$-th, $(c+2\iota')$-th, ..., d-th columns of $\boldsymbol{N}$. Put differently, a signifies the first element to pick, b is the last, and ι is the increment with respect to rows. c, d, and ι' play analogous roles with respect to columns. It is clear that:

$$\boldsymbol{N}(a : \iota : b, c : \iota' : d)' = \boldsymbol{N}'(c : \iota' : d, a : \iota : b). \tag{2.3.1}$$

A short-hand notation is used when $a = b : \boldsymbol{N}(a : \iota : b, c : \iota' : d) = \boldsymbol{N}(a, c : \iota' : d)$. When $\iota = 1$, we write: $\boldsymbol{N}(a : \iota : b, c : \iota' : d) = \boldsymbol{N}(a : b, c : \iota' : d)$. Analogous notations are used when $c = d$ or $\iota' = 1$, respectively.

By Theorem 3.1 in Dufour and Renault (1998) and from model (2.2.5), it follows that $H_0^i(h)$ are equivalent to:

$$\boldsymbol{A}_k^{(h)}(a : \iota : b, c : \iota' : d) = \boldsymbol{0} \text{ for each } k \in \{1, \ldots, p\}, \tag{2.3.2}$$

where a, ι, b, c, ι', d, and the size of the null vector differ across cases $i = 1, \ldots, 4$ and I and

II.[6] In Table 2.1 we detail the specifics for a, ι, b, c, ι', d in these quantities for each of the six cases.

Each case in Table 2.1 can be interpreted as follows. In Case 1, the (mK_H+j_2, mK_H+j_1)-th element of $\boldsymbol{A}_k^{(h)}$ (i.e., the impact of the j_1-th low frequency variable on the j_2-th low frequency variable) is zero if and only if $H_0^1(h)$ is true. Likewise, in Case 2, the $(mK_H + j_1, i_1)$-th, $(mK_H + j_1, i_1 + K_H)$-th, ..., $(mK_H + j_1, i_1 + (m-1)K_H)$-th elements of $\boldsymbol{A}_k^{(h)}$ are all zeros under $H_0^2(h)$. Note that we are testing whether or not all mp coefficients of the i_1-th high frequency variable on the j_1-th low frequency variable are zeros, i.e., the i_1-th high frequency variable has no impact as a whole on the j_1-th low frequency variable at a given horizon h.

When $H_0^3(h)$ holds, all mp coefficients of the j_1-th low frequency variable on the i_1-th high frequency variable are zeros at horizon h. Note that the parameter constraints run across the i_1-th, $(i_1 + K_H)$-th, ..., $(i_1 + (m-1)K_H)$-th rows of $\boldsymbol{A}_k^{(h)}$, not columns. This means that we are testing *simultaneous* linear restrictions *across multiple equations*, unlike Dufour, Pelletier, and Renault (2006) who focus mainly on *simultaneous* linear restrictions *within one equation*, and unlike Hill (2007) who focuses on *sequential* linear restrictions *across multiple equations*.

In Case 4, the i_1-th high frequency variable has no impact on the i_2-th high frequency variable if and only if $H_0^4(h)$ is true. In this case m^2 elements out of $\boldsymbol{A}_k^{(h)}$ are restricted to be zeros for each k, so the total number of zero restrictions is pm^2. Under $H_0^I(h)$, the $K_L \times mK_H$ lower-left block of $\boldsymbol{A}_k^{(h)}$ is a null matrix. Finally, in Case II, the $mK_H \times K_L$ upper-right block of $\boldsymbol{A}_k^{(h)}$ is a null matrix if and only if $H_0^{II}(h)$ is true.

We can now combine the (p,h)-autoregression parameter set $\boldsymbol{B}(h)$ in (2.2.6) with the matrix construction (2.3.1), its implication for testable restrictions (2.3.2), and Table 2.1, to obtain generic formulae for $\boldsymbol{R}$ and $\boldsymbol{r}$ so that all six cases can be treated as special cases of (2.2.7).

The above can be summarized as follows:

Theorem 2.3.1. All hypotheses $H_0^i(h)$ for $i \in \{1,2,3,4,I,II\}$ are special cases of $H_0(h)$ with

$$\boldsymbol{R} = \left[\boldsymbol{\Lambda}(\delta_1)', \boldsymbol{\Lambda}(\delta_2)', \ldots, \boldsymbol{\Lambda}(\delta_{g(a,\iota,b)p})'\right]' \tag{2.3.3}$$

[6]Recall that $x_{L,j}(\tau_L)$ and $\tilde{\boldsymbol{x}}_{H,i}(\tau_L) = [x_{H,i}(\tau_L,1), \ldots, x_{H,i}(\tau_L,m)]'$ belong to $\boldsymbol{X}$ in (2.2.2) for all $j \in \{1, \ldots, K_L\}$ and $i \in \{1, \ldots, K_H\}$. This is why non-causality under mixed frequencies is well-defined and Theorem 3.1 in Dufour and Renault (1998) can be applied directly.

and

$$\boldsymbol{r} = \mathbf{0}_{g(a,\iota,b)g(c,\iota',d)p\times 1}, \tag{2.3.4}$$

where $g(a,\iota,b) = (b-a)/\iota + 1$, $\delta_1 = pK(a-1) + c$,

$$\delta_l = \delta_{l-1} + K + pK(\iota-1)I(l-1 = zp \text{ for some } z \in \mathcal{N}) \tag{2.3.5}$$

for $l = 2, \ldots, g(a,\iota,b)p$, and $\boldsymbol{\Lambda}(\delta)$ is a $g(c,\iota',d) \times pK^2$ matrix whose $(j,\ \delta+(j-1)\iota')$-th element is 1 for $j \in \{1, \ldots, g(c,\iota',d)\}$ and all other elements are zeros.

Several key points will help us understand (2.3.3) through (2.3.5). First, $g(a,\iota,b)$ and $g(c,\iota',d)$ represent how many rows and columns of $\boldsymbol{A}_k^{(h)}$ have zero restrictions for each $k \in \{1,\ldots,p\}$, respectively. The total number of zero restrictions is therefore $q = g(a,\iota,b)g(c,\iota',d)p$ as in (2.3.4). Second, $\boldsymbol{\Lambda}(\delta)$ has only one nonzero element in each row that is identically 1, signifying which element of $\text{vec}[\boldsymbol{B}(h)]$ is supposed to be zero. The location of 1 is determined by $\delta_1, \ldots, \delta_{g(a,\iota,b)p}$, which are recursively updated according to (2.3.5). As seen in (2.3.5), the increment of δ_l is basically K, but an extra increment of $pK(\iota-1)$ is added when $l-1$ is a multiple of p in order to skip some columns of $\boldsymbol{B}(h)$.

Theorem 2.3.1 provides unified testing for non-causality as summarized below.

Step 1 For a given VAR lag order p and test horizon h, estimate a (p,h)-autoregression.[7]

Step 2 Calculate a, ι, b, c, ι', d according to Table 2.1 for a given case of non-causality relation. Put those quantities into (2.3.3) and (2.3.4) to get $\boldsymbol{R}$ and $\boldsymbol{r}$.

Step 3 Use $\boldsymbol{R}$ and $\boldsymbol{r}$ in order to calculate the Wald test statistic $W[H_0(h)]$ in (2.2.9).

Example 2 : *Selection Matrices* $\boldsymbol{R}$ *and* $\boldsymbol{r}$: Since Table 2.1 and Theorem 2.3.1 are rather abstract, we present a concrete example of how $\boldsymbol{R}$ and $\boldsymbol{r}$ are constructed in our trivariate simulation and empirical application. In Section 2.6.2 and Section 2.7, we fit a MF-VAR(1)

[7] A potential drawback of our approach as well as Dufour, Pelletier, and Renault (2006) is that the prediction horizon h is fixed at each test and thus the entire set of results for multiple h's may yield a contradiction. See footnote 2 in Hill (2007). Hill (2007) avoids this problem by a sequential multiple-horizon non-causation test, in which a series of individual non-causation tests are performed to deduce causal chains and causation horizon. The present paper takes the Dufour, Pelletier, and Renault (2006) approach because of its simplicity. See Hill (2007) and Salamaliki and Venetis (2013) for a comparison of the two methods.

model with prediction horizons $h \in \{1, 2, 3\}$ to two high frequency variables X and Y and one low frequency variable Z with $m = 3$. In this case the mixed frequency vector appearing in (2.2.2) can be written as:

$$\boldsymbol{W}(\tau_L) = [X(\tau_L, 1), Y(\tau_L, 1), X(\tau_L, 2), Y(\tau_L, 2), X(\tau_L, 3), Y(\tau_L, 3), Z(\tau_L)]'.$$

Note that $K_H = 2$, $K_L = 1$, and hence $K = 7$ in this example. Although the construction of $\boldsymbol{R}$ and $\boldsymbol{r}$ do not depend on the value of h, consider $h = 1$ for simplicity, and write the parameter matrix:

$$\boldsymbol{A}_1 = \begin{bmatrix} a_{11} & \dots & a_{17} \\ \vdots & \ddots & \vdots \\ a_{71} & \dots & a_{77} \end{bmatrix} \quad \text{or} \quad \boldsymbol{A}_1' = \begin{bmatrix} a_{11} & \dots & a_{71} \\ \vdots & \ddots & \vdots \\ a_{17} & \dots & a_{77} \end{bmatrix}.$$

Since $p = h = 1$, $\boldsymbol{B}(h)$ appearing in (2.2.6) is simply $\boldsymbol{A}_1'$.

Consider the null hypothesis that Z does not cause X at horizon 1. This null hypothesis is equivalently $a_{17} = a_{37} = a_{57} = 0$ since a_{17}, a_{37}, and a_{57} represent the impact of $Z(\tau_L - 1)$ on $X(\tau_L, 1)$, $X(\tau_L, 2)$, and $X(\tau_L, 3)$, respectively. Note that a_{17}, a_{37}, and a_{57} are respectively the 7th, 21st, and 35th element of vec$[\boldsymbol{B}(h)]$ appearing in (2.2.7). Hence, the proper choice of $\boldsymbol{R}$ and $\boldsymbol{r}$ is:

$$\boldsymbol{R} = \begin{bmatrix} \mathbf{0}_{1\times 6} & 1 & \mathbf{0}_{1\times 13} & 0 & \mathbf{0}_{1\times 13} & 0 & \mathbf{0}_{1\times 14} \\ \mathbf{0}_{1\times 6} & 0 & \mathbf{0}_{1\times 13} & 1 & \mathbf{0}_{1\times 13} & 0 & \mathbf{0}_{1\times 14} \\ \mathbf{0}_{1\times 6} & 0 & \mathbf{0}_{1\times 13} & 0 & \mathbf{0}_{1\times 13} & 1 & \mathbf{0}_{1\times 14} \end{bmatrix} \quad \text{and} \quad \boldsymbol{r} = \mathbf{0}_{3\times 1}. \tag{2.3.6}$$

We now confirm that the same $\boldsymbol{R}$ and $\boldsymbol{r}$ can be obtained via Table 2.1 and Theorem 2.3.1. Non-causality from Z to X falls in Case 3 with $i_1 = j_1 = 1$ (i.e. non-causality from the first low frequency variable to the first high frequency variable). Using Table 2.1, we have that $(a, \iota, b, c, \iota', d) = (1, 2, 5, 7, 1, 7)$ and therefore $g(a, \iota, b) = 3$, $g(c, \iota', d) = 1$, and $\{\delta_1, \delta_2, \delta_3\} = \{7, 21, 35\}$ by application of Theorem 2.3.1. This implies that $\boldsymbol{r} = \mathbf{0}_{3\times 1}$ and $\boldsymbol{R} = [\boldsymbol{\Lambda}(7)', \boldsymbol{\Lambda}(21)', \boldsymbol{\Lambda}(35)']'$, where $\boldsymbol{\Lambda}(\delta)$ is a 1×49 vector whose δ-th element is 1 and all other elements are zeros for $\delta \in \{7, 21, 35\}$. We can therefore confirm that Table 2.1 and Theorem 2.3.1 provide correct $\boldsymbol{R}$ and $\boldsymbol{r}$ shown in (2.3.6).

2.4 Recovery of High Frequency Causality

The existing literature on Granger causality and temporal aggregation has three key ingredients. Starting with (1) a data generating process (DGP) for HF data, and (2) specifying a (linear) aggregation scheme, one is interested in (3) the relationship between causal patterns - or lack thereof - among the HF series and the inference obtained from LF data when *all* HF series are aggregated. So far, we refrained from (1) specifying a DGP for HF series and (2) specifying an aggregation scheme. We will proceed along the same path as the existing literature in this section with a different purpose, namely to show that the MF approach recovers more underlying causal patterns than the standard LF approach does. While conducting Granger causality tests with MF series does not resolve all HF causal patterns, using MF instead of using exclusively LF series promotes sharper inference.

We first start with a fairly straightforward extension of Lütkepohl (1984), establishing the link between HF-VAR and MF data representations. We then analyze the link between HF, MF and LF causality.

2.4.1 Temporal Aggregation of VAR Processes

Lütkepohl (1984) provides a comprehensive analysis of temporal aggregation and VAR processes. We extend his analysis to a MF setting. While the extension is straightforward, it provides us with a framework that will be helpful for the analysis in the rest of the paper.

Let $K^* = K_H + K_L$, and define $\overline{\boldsymbol{X}}(\tau_L, k) = [\boldsymbol{x}_H(\tau_L, k)',\ \boldsymbol{x}_L(\tau_L, k)']' \in \mathcal{R}^{K^*}$ for $k = 1, \ldots, m$. Note that part of the $\overline{\boldsymbol{X}}$ vector process is obviously latent, namely the high frequency observations of the LF process, represented by the $\boldsymbol{x}_L(\tau_L, k)$ elements of the vector process.

In order to proceed, let $\mathcal{L}_H$ denote the *high frequency* lag operator, in particular

$$\mathcal{L}_H^l \overline{\boldsymbol{X}}(\tau_L, k) = \overline{\boldsymbol{X}}(\tau_L - \iota, \iota')$$

with

$$\iota = \begin{cases} 0 & \text{if} \quad 0 \leq l < k \\ 1 + \lfloor \frac{l-k}{m} \rfloor & \text{if} \quad l \geq k \end{cases} \quad \text{and} \quad \iota' = \begin{cases} k - l & \text{if} \quad 0 \leq l < k \\ \iota m + k - l & \text{if} \quad l \geq k. \end{cases}$$

Note that $\lfloor x \rfloor$ is the largest integer not larger than x. For example, $\mathcal{L}_H \overline{\boldsymbol{X}}(\tau_L, 2) = \overline{\boldsymbol{X}}(\tau_L, 1)$ and $\mathcal{L}_H \overline{\boldsymbol{X}}(\tau_L, 1) = \overline{\boldsymbol{X}}(\tau_L - 1, m)$. Letting $\mathcal{L}_L$ be the *low frequency* lag operator, we have that $\mathcal{L}_L \overline{\boldsymbol{X}}(\tau_L, 1) = \mathcal{L}_H^m \overline{\boldsymbol{X}}(\tau_L, 1) = \overline{\boldsymbol{X}}(\tau_L - 1, 1)$.

Assume that $\{\{\overline{\boldsymbol{X}}(\tau_L, k)\}_k\}_{\tau_L}$ follows a VAR(p) process with $p \in \mathcal{N} \cup \{\infty\}$:

$$\overline{\boldsymbol{X}}(\tau_L, k) = \sum_{l=1}^{p} \boldsymbol{\Phi}_l \mathcal{L}_H^l \overline{\boldsymbol{X}}(\tau_L, k) + \overline{\boldsymbol{\eta}}(\tau_L, k). \tag{2.4.1}$$

The coefficient matrix $\boldsymbol{\Phi}_l$ is partitioned in the following manner:

$$\boldsymbol{\Phi}_l = \begin{bmatrix} \boldsymbol{\Phi}_{HH,l} & \boldsymbol{\Phi}_{HL,l} \\ \boldsymbol{\Phi}_{LH,l} & \boldsymbol{\Phi}_{LL,l} \end{bmatrix},$$

where $\boldsymbol{\Phi}_{yz,l} \in \mathcal{R}^{K_y \times K_z}$ with $y, z \in \{H, L\}$. The error $\overline{\boldsymbol{\eta}}(\tau_L, k)$ satisfies a HF martingale difference property similar to the LF based Assumption 2.2.1 in Section 2.2.2.[8] It is therefore helpful to define a HF sigma field using a single-index version of $\overline{\boldsymbol{X}}(\tau_L, k)$. Simply write (2.4.1) as $\boldsymbol{Y}_t = \sum_{l=1}^{p} \boldsymbol{\Phi}_l \boldsymbol{Y}_{t-l} + \boldsymbol{\xi}_t$, where $\{\boldsymbol{Y}_t, \boldsymbol{\xi}_t\} \in \mathcal{R}^{K^*}$ are single-index versions of $\{\overline{\boldsymbol{X}}(\tau_L, k), \overline{\boldsymbol{\eta}}(\tau_L, k)\}$, e.g. $t = m(\tau_L - 1) + k$, so that $\boldsymbol{Y}_1$ corresponds to $\overline{\boldsymbol{X}}(1, 1)$. See also Section 2.4.2 below. Then $\boldsymbol{\xi}_t = \overline{\boldsymbol{\eta}}(\tau_L, k)$ is a stationary martingale difference with respect to $\sigma(\boldsymbol{Y}_s : s \leq t)$ with variance $\boldsymbol{V} \equiv E[\overline{\boldsymbol{\eta}}(\tau_L, k)\overline{\boldsymbol{\eta}}(\tau_L, k)']$.

As stated in (2.2.1), a general linear aggregation scheme is considered: $\boldsymbol{x}_H(\tau_L) = \sum_{k=1}^{m} w_k \boldsymbol{x}_H(\tau_L, k)$ and $\boldsymbol{x}_L(\tau_L) = \sum_{k=1}^{m} w_k \boldsymbol{x}_L(\tau_L, k)$. By an application of Theorem 1 in Lütkepohl (1984), the mixed frequency vector $\boldsymbol{X}(\tau_L)$ defined in (2.2.2) and the low frequency vector defined as

$$\underline{\boldsymbol{X}}(\tau_L) = [\boldsymbol{x}_H(\tau_L)', \boldsymbol{x}_L(\tau_L)']' \in \mathcal{R}^{K^*} \tag{2.4.2}$$

follow VARMA processes. More specifically, we have the following.

Theorem 2.4.1. Suppose that an underlying high frequency process follows a VAR(p). Then the corresponding MF process is a VARMA(p_M, q_M), and the corresponding low frequency

[8] Lütkepohl (1984) only requires the VAR error to be vector white noise. We impose the martingale difference assumption here for continuity with the paper in general.

process is a VARMA(p_L, q_L). Moreover,

$$p_M \leq \deg\left[\det(\overline{\mathcal{A}}(\mathcal{L}_L))\right] \equiv g \text{ and } p_L \leq g,$$

where g is the degree of polynomial of $\det(\overline{\mathcal{A}}(\mathcal{L}_L))$. Furthermore,

$$q_M \leq \max\left\{\deg\left[\overline{\mathcal{A}}_{kl}(\mathcal{L}_L)\right] - g + p_M |\ k,l = 1,\ldots,mK^*\right\},$$

where $\overline{\mathcal{A}}_{kl}(\mathcal{L}_L)$ is the (k,l)-th cofactor of $\overline{\mathcal{A}}(\mathcal{L}_L)$. Similarly,

$$q_L \leq \max\left\{\deg\left[\overline{\mathcal{A}}_{kl}(\mathcal{L}_L)\right] - g + p_L |\ k,l = 1,\ldots,mK^*\right\}.$$

Finally, if the high frequency VAR process is stationary then so are the mixed and low frequency VARMA processes.

Remark 9. See Appendix A.2 for a proof, and for completeness the construction of $\overline{\mathcal{A}}(\mathcal{L}_L)$.

In general it is impossible to characterize p_M, q_M, p_L, or q_L exactly (cfr. Lütkepohl (1984)). Nevertheless, if the HF process $\{\boldsymbol{X}(\tau_L,k)\}$ is governed by a VAR(p) then the MF and LF processes $\{\boldsymbol{X}(\tau_L)\}$ and $\{\underline{\boldsymbol{X}}(\tau_L)\}$ have VARMA representations, and therefore VAR(∞) representations under the assumption of invertibility. Thus, one can still estimate those invertible VARMA processes by using a finite order approximation as in Lewis and Reinsel (1985), Lütkepohl and Poskitt (1996), and Saikkonen and Lütkepohl (1996). Moreover, the VARMA order can be characterized under certain simple cases such as stock sampling with $p = 1$.

Example 3 : *stock sampling with $p = 1$*: Suppose that an underlying HF process follows a VAR(1) $\overline{\boldsymbol{X}}(\tau_L,k) = \boldsymbol{\Phi}_1\mathcal{L}_H^1\overline{\boldsymbol{X}}(\tau_L,k) + \overline{\boldsymbol{\eta}}(\tau_L,k)$ where $\overline{\boldsymbol{\eta}}(\tau_L,k)$ is a stationary martingale difference with respect to the HF sigma field $\sigma(\boldsymbol{Y}_s : s \leq t)$, $\boldsymbol{Y}_t$ is a single-index version of $\overline{\boldsymbol{X}}(\tau_L,k)$, and $\boldsymbol{V} \equiv E[\overline{\boldsymbol{\eta}}(\tau_L,k)\overline{\boldsymbol{\eta}}(\tau_L,k)']$.

It is easy to show that the corresponding MF process also follows a VAR(1) if we consider stock sampling:

$$\boldsymbol{X}(\tau_L) = \boldsymbol{A}_1\boldsymbol{X}(\tau_L - 1) + \boldsymbol{\epsilon}(\tau_L). \tag{2.4.3}$$

The parameter $\boldsymbol{A}_1$ is

$$\boldsymbol{A}_1 = \begin{bmatrix} \mathbf{0}_{K_H \times (m-1)K_H} & \boldsymbol{\Phi}_{HH,1}^{[1]} & \boldsymbol{\Phi}_{HL,1}^{[1]} \\ \vdots & \vdots & \vdots \\ \mathbf{0}_{K_H \times (m-1)K_H} & \boldsymbol{\Phi}_{HH,1}^{[m]} & \boldsymbol{\Phi}_{HL,1}^{[m]} \\ \mathbf{0}_{K_L \times (m-1)K_H} & \boldsymbol{\Phi}_{LH,1}^{[m]} & \boldsymbol{\Phi}_{LL,1}^{[m]} \end{bmatrix}, \tag{2.4.4}$$

where

$$\boldsymbol{\Phi}_l^k \equiv \begin{bmatrix} \boldsymbol{\Phi}_{HH,l}^{[k]} & \boldsymbol{\Phi}_{HL,l}^{[k]} \\ \boldsymbol{\Phi}_{LH,l}^{[k]} & \boldsymbol{\Phi}_{LL,l}^{[k]} \end{bmatrix}.$$

By construction

$$\boldsymbol{\epsilon}(\tau_L) = \begin{bmatrix} \sum_{k=1}^{1} \begin{bmatrix} \boldsymbol{\Phi}_{HH,1}^{[1-k]} & \boldsymbol{\Phi}_{HL,1}^{[1-k]} \end{bmatrix} \boldsymbol{\eta}(\tau_L, k) \\ \vdots \\ \sum_{k=1}^{m} \begin{bmatrix} \boldsymbol{\Phi}_{HH,1}^{[m-k]} & \boldsymbol{\Phi}_{HL,1}^{m-k]} \end{bmatrix} \boldsymbol{\eta}(\tau_L, k) \\ \sum_{k=1}^{m} \begin{bmatrix} \boldsymbol{\Phi}_{LH,1}^{[m-k]} & \boldsymbol{\Phi}_{LL,1}^{[m-k]} \end{bmatrix} \boldsymbol{\eta}(\tau_L, k) \end{bmatrix},$$

hence $\boldsymbol{\epsilon}(\tau_L)$ is a stationary martingale difference with respect to the MF sigma field $\sigma(\boldsymbol{X}(t) : t \leq \tau_L)$, where $\boldsymbol{\Omega} \equiv E[\boldsymbol{\epsilon}(\tau_L)\boldsymbol{\epsilon}(\tau_L)']$ can be explicitly characterized as a function of $\boldsymbol{\Phi}_1$ and $\boldsymbol{V}$. The covariance matrix $\boldsymbol{\Omega}$ has a block representation

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Omega}_{1,1} & \dots & \boldsymbol{\Omega}_{1,m} & \boldsymbol{\Omega}_{1,m+1} \\ \vdots & \ddots & \vdots & \vdots \\ \boldsymbol{\Omega}_{1,m}' & \dots & \boldsymbol{\Omega}_{m,m} & \boldsymbol{\Omega}_{m,m+1} \\ \boldsymbol{\Omega}_{1,m+1}' & \dots & \boldsymbol{\Omega}_{m,m+1}' & \boldsymbol{\Omega}_{m+1,m+1} \end{bmatrix} \in \mathcal{R}^{K \times K}, \tag{2.4.5}$$

with components

$$\boldsymbol{\Omega}_{i,j} = \sum_{k=1}^{i} \begin{bmatrix} \boldsymbol{\Phi}_{HH,1}^{[i-k]} & \boldsymbol{\Phi}_{HL,1}^{[i-k]} \end{bmatrix} \boldsymbol{V} \begin{bmatrix} \boldsymbol{\Phi}_{HH,1}^{[j-k]'} \\ \boldsymbol{\Phi}_{HL,1}^{[j-k]'} \end{bmatrix} \text{ for } i,j \in \{1,\dots,m\} \text{ and } i \leq j, \tag{2.4.6}$$

$$\boldsymbol{\Omega}_{i,m+1} = \sum_{k=1}^{i} \begin{bmatrix} \boldsymbol{\Phi}_{HH,1}^{[i-k]} & \boldsymbol{\Phi}_{HL,1}^{[i-k]} \end{bmatrix} \boldsymbol{V} \begin{bmatrix} \boldsymbol{\Phi}_{LH,1}^{[m-k]'} \\ \boldsymbol{\Phi}_{LL,1}^{[m-k]'} \end{bmatrix} \text{ for } i \in \{1,\dots,m\}$$

and

$$\boldsymbol{\Omega}_{m+1,m+1} = \sum_{k=1}^{m} \begin{bmatrix} \boldsymbol{\Phi}_{LH,1}^{[m-k]} & \boldsymbol{\Phi}_{LL,1}^{[m-k]} \end{bmatrix} \boldsymbol{V} \begin{bmatrix} \boldsymbol{\Phi}_{LH,1}^{[m-k]'} \\ \boldsymbol{\Phi}_{LL,1}^{[m-k]'} \end{bmatrix}. \tag{2.4.7}$$

Similarly, the LF process follows a VAR(1):

$$\underline{\boldsymbol{X}}(\tau_L) = \underline{\boldsymbol{A}}_1 \underline{\boldsymbol{X}}(\tau_L - 1) + \underline{\boldsymbol{\epsilon}}(\tau_L), \tag{2.4.8}$$

where

$$\underline{\boldsymbol{A}}_1 = \boldsymbol{\Phi}_1^m, \tag{2.4.9}$$

and $\underline{\boldsymbol{\epsilon}}(\tau_L)$ is a stationary martingale difference with respect to the LF sigma field $\sigma(\underline{\boldsymbol{X}}(t) : t \leq \tau_L)$, with $\underline{\boldsymbol{\Omega}} \equiv E[\underline{\boldsymbol{\epsilon}}(\tau_L)\underline{\boldsymbol{\epsilon}}(\tau_L)']$. Simply note $\underline{\boldsymbol{\epsilon}}(\tau_L) = \sum_{k=1}^{m} \boldsymbol{\Phi}_1^{m-k} \boldsymbol{\eta}(\tau_L, k)$ to deduce the covariance matrix structure:

$$\underline{\boldsymbol{\Omega}} = \sum_{k=1}^{m} \boldsymbol{\Phi}_1^{m-k} \boldsymbol{V} (\boldsymbol{\Phi}_1^{m-k})' \in \mathcal{R}^{K^* \times K^*}. \tag{2.4.10}$$

2.4.2 Causality and Temporal Aggregation

Felsenstein et al. (2013) explore conditions for identifying a HF process based on MF data. When their conditions are satisfied, recovery of HF causality is trivially feasible by looking at off-diagonal elements of the identified HF-VAR coefficients. The conditions for identification are stringent, however, and one may therefore wonder what happens if they are not satisfied. In this subsection we fill some of the gap by focusing on testing for causality since this does not require full identification of the entire HF process.

Since Granger causality is based on information sets, we need to define reference information sets for HF- and LF-VAR processes. Toward this end, we rewrite a HF-VAR(p) process in (2.4.1) with a single time index t: $\boldsymbol{Y}_t = \sum_{l=1}^{p} \boldsymbol{\Phi}_l \boldsymbol{Y}_{t-l} + \boldsymbol{\xi}_t$, where $\boldsymbol{Y}_t \in \mathcal{R}^{K^*}$ is simply a single-index version of $\boldsymbol{X}(\tau_L, k)$. One way of mapping (τ_L, k) to t is to let $t = m(\tau_L - 1) + k$ so that $\boldsymbol{Y}_1$ corresponds to $\boldsymbol{X}(1, 1)$. The same mapping is used between $\boldsymbol{\xi}_t$ and $\boldsymbol{\eta}(\tau_L, k)$. Recall from Section 2.3.1 that $\mathcal{I}(\tau_L)$ is the MF *reference information set in period* τ_L, while $\mathcal{I} = \{\mathcal{I}(\tau_L) | \ \tau_L \in \mathcal{Z}\}$ is the MF *reference information set*. We now introduce HF and LF versions of the information set.

The *HF reference information set at time t* is defined as $\overline{\mathcal{I}}(t) = \boldsymbol{Y}(-\infty, t]$. The *HF reference information set* is defined as $\overline{\mathcal{I}} = \{\overline{\mathcal{I}}(t) \,|\, t \in \mathcal{Z}\}$. The prediction horizon for non-causality given $\overline{\mathcal{I}}$ is in terms of the high frequency, denoted by $\overline{h} \in \mathcal{Z}$. For example, non-causality from all high frequency variables to all low frequency variables at high frequency horizon $\overline{h}$ given $\overline{\mathcal{I}}$ is written as $\boldsymbol{x}_H \nrightarrow_{\overline{h}} \boldsymbol{x}_L \,|\, \overline{\mathcal{I}}$. Similarly, the *LF reference information set at time* τ_L is defined as $\underline{\mathcal{I}}(\tau_L) = \underline{\boldsymbol{X}}(-\infty, \tau_L]$, where $\underline{\boldsymbol{X}}(\tau_L)$ is given in (2.4.2). The *LF reference information set* is defined as $\underline{\mathcal{I}} = \{\underline{\mathcal{I}}(\tau_L) \,|\, \tau_L \in \mathcal{Z}\}$. Whether (non-)causality is preserved under temporal aggregation depends mainly on three conditions: an aggregation scheme, VAR lag order p, and the presence of an auxiliary variable and therefore the possibility of causality chains. The existing literature has found that temporal aggregation may hide or generate causality even in very simple cases. We show that the MF approach recovers underlying causality patterns better than the traditional LF approach.

Theorem 2.4.2. Consider the linear aggregation scheme appearing in (2.2.1) and assume a HF-VAR(p) with $p \in \mathcal{N} \cup \{\infty\}$. Then, the following two properties hold when applied respectively to all low and all high frequency processes: (i) If $\boldsymbol{x}_H \nrightarrow \boldsymbol{x}_L \,|\, \mathcal{I}$, then $\boldsymbol{x}_H \nrightarrow \boldsymbol{x}_L \,|\, \underline{\mathcal{I}}$. (ii) If $\boldsymbol{x}_L \nrightarrow \boldsymbol{x}_H \,|\, \overline{\mathcal{I}}$, then $\boldsymbol{x}_L \nrightarrow \boldsymbol{x}_H \,|\, \mathcal{I}$.

Proof: See Appendix A.3.

Note that the prediction horizon in Theorem 2.4.2 is arbitrary since there are no auxiliary variables involved. This follows since we only examine the relationship between all low and all high frequency processes respectively.[9]

Theorem 2.4.2 part (i) states that non-causality from all high frequency variables to all low frequency variables is preserved between MF and LF processes, while part (ii) states that non-causality from all low frequency variables to all high frequency variables is preserved between HF and MF processes. One might incorrectly guess from Theorem 2.4.2 part (ii) that $\boldsymbol{x}_L \nrightarrow \boldsymbol{x}_H \,|\, \mathcal{I}$ $\Rightarrow$ $\boldsymbol{x}_L \nrightarrow \boldsymbol{x}_H \,|\, \underline{\mathcal{I}}$. This statement does not hold in general. A simple counter-example is a HF-VAR(2) process with stock sampling, $m = 2$, $K_H = K_L = 1$,

[9]Theoretical results in the presence of auxiliary variables are seemingly intractable since potential causal chains complicate causality patterns substantially.

$$\mathbf{\Phi}_1 = \begin{bmatrix} \phi_{HH,1} & 0 \\ \phi_{LH} & 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{\Phi}_2 = \begin{bmatrix} \phi_{HH,2} & 0 \\ 0 & 0 \end{bmatrix}.$$

Assume that $\phi_{HH,1}$, $\phi_{HH,2}$, and ϕ_{LH} are all nonzero. Note that, given $\overline{\mathcal{I}}$, x_L does not cause x_H while x_H does cause x_L. In this particular case, we can derive the corresponding MF-VAR(1) and LF-VAR(1) processes. The MF coefficient is

$$\boldsymbol{A}_1 = \begin{bmatrix} \phi_{HH,2} & \phi_{HH,1} & 0 \\ \phi_{HH,1}\phi_{HH,2} & \phi_{HH,1}^2 + \phi_{HH,2} & 0 \\ \phi_{LH}\phi_{HH,2} & \phi_{LH}\phi_{HH,1} & 0 \end{bmatrix}, \tag{2.4.11}$$

while the LF coefficient is

$$\underline{\boldsymbol{A}}_1 = \begin{bmatrix} \phi_{HH,1}^2 + \phi_{HH,2} & \phi_{HH,1}\phi_{HH,2}/\phi_{LH} \\ \phi_{LH}\phi_{HH,1} & \phi_{HH,2} \end{bmatrix}. \tag{2.4.12}$$

Equations (2.4.11) and (2.4.12) indicate that x_L does not cause x_H given $\mathcal{I}$, but x_L does cause x_H given $\underline{\mathcal{I}}$. Thus, we confirm that non-causality from all low frequency variables to all high frequency variables is *not necessarily* preserved between MF and LF processes.

Summarizing Theorem 2.4.2 and the counter-example above, a crucial condition for non-causality preservation is that the information for the "right-hand side" variables (i.e. $\boldsymbol{x}_L$ for (i) and $\boldsymbol{x}_H$ for (ii)) is not lost by temporal aggregation. In this sense, the MF approach yields more implications on hidden causality patterns than the LF approach, which switches directly from a HF process by aggregating all variables.

We conclude this subsection by again focusing on stock sampling with p =1 as this particular case yields much sharper results.

Example 4: *stock sampling with p = 1:* When $p = 1$ and stock sampling is of interest, the exact functional form for the MF and LF processes is known and appear in (2.4.3) and (2.4.8). Equation (2.4.4) highlights what kind of causality information gets lost by switching from a HF- to MF-VAR. Similarly, (2.4.9) reveals the information loss when moving from a MF- to LF-VAR. This brings us to the following theorem.

Theorem 2.4.3. Consider stock sampling with $p = 1$. Then, the corresponding MF-VAR and LF-VAR processes are also of order 1. Furthermore, non-causation among the HF-, MF-, and LF-VAR processes is related as follows.

i. *In Case 1 (low $\nrightarrow$ low) and Case 2 (high $\nrightarrow$ low), non-causation up to HF horizon m given the HF information set $\overline{\mathcal{I}}$ implies non-causation at horizon* 1 *given the MF information set $\mathcal{I}$, which is necessary and sufficient for non-causation at horizon* 1 *given the LF information set $\underline{\mathcal{I}}$.*

ii. *In Case 3 (low $\nrightarrow$ high) and Case 4 (high $\nrightarrow$ high), non-causation up to HF horizon m given $\overline{\mathcal{I}}$ is necessary and sufficient for non-causation at horizon* 1 *given $\mathcal{I}$, which implies non-causation at horizon* 1 *given $\underline{\mathcal{I}}$.*

iii. *In Case I (all high $\nrightarrow$ all low), non-causation at HF horizon* 1 *given $\overline{\mathcal{I}}$ implies non-causation at horizon* 1 *given $\mathcal{I}$, which is necessary and sufficient for non-causation at horizon* 1 *given $\underline{\mathcal{I}}$.*

iv. *In Case II (all low $\nrightarrow$ all high), non-causation at HF horizon* 1 *given $\overline{\mathcal{I}}$ is necessary and sufficient for non-causation at horizon* 1 *given $\mathcal{I}$, which implies non-causation at horizon* 1 *given $\underline{\mathcal{I}}$.*

Proof: See Appendix A.4.

Although Theorem 2.4.3 is much sharper than Theorem 2.4.2 due to much stronger assumptions, they share an interesting feature that causality tends to be contaminated more when temporal aggregation discards information for "right-hand side" variables. For example, item 2 shows that no relevant information for testing low-to-high or high-to-high causality is lost when moving from $\overline{\mathcal{I}}$ to $\mathcal{I}$ (i.e., when aggregating low frequency variables), while some information is lost when moving from $\mathcal{I}$ to $\underline{\mathcal{I}}$ (i.e., when aggregating high frequency variables).

Theorem 2.4.3 suggests that the MF causality test should never perform worse than the low frequency causality test, and the former should be more powerful than the latter especially when Cases 3, 4, and II are of interest. Sections 2.5 and 2.6 verify this point by a local asymptotic power analysis and a Monte Carlo simulation, respectively.

2.5 Local Asymptotic Power Analysis

The goal of this section is to show that the MF causality tests have higher local asymptotic power compared to the LF causality test. We need to constrain our attention to analytically tractable DGPs, which is why we consider a bivariate HF-VAR(1) process with stock sampling. As shown in the previous section, for the bivariate HF-VAR(1) one can derive analytically the corresponding MF- and LF-VAR(1) processes. Recall that Case I considers unidirectional causality from the high frequency variable to the low frequency variable, while Case II considers unidirectional causality from the low frequency variable to the high frequency variable.

We first compute the local asymptotic power functions for both cases, and then plot them in a numerical exercise. Since we work with a HF process, define the HF sample size $T \equiv T_L \times m$.

Case I: High-to-Low Causality In order to characterize local asymptotic power, assume that the high frequency DGP is given by:

$$\boldsymbol{X}(\tau_L, k) = \boldsymbol{\Phi}(\nu/\sqrt{T})\mathcal{L}_H \boldsymbol{X}(\tau_L, k) + \boldsymbol{\eta}(\tau_L, k), \tag{2.5.1}$$

where

$$\boldsymbol{\Phi}(\nu/\sqrt{T}) = \begin{bmatrix} \rho_H & 0 \\ \nu/\sqrt{T} & \rho_L \end{bmatrix}$$

with $\rho_H, \rho_L \in (-1, 1)$, where $\nu \in \mathcal{R}$ is the usual Pitman drift parameter. Assume for computational simplicity that $\boldsymbol{\eta}(\tau_L, k) \overset{i.i.d.}{\sim} (\boldsymbol{0}_{2\times 1}, \boldsymbol{I}_2)$, hence $\boldsymbol{X}(\tau_L, k)$ has a strictly stationary solution and model (2.5.1) fully describes the causal structure of $\boldsymbol{X}(\tau_L, k)$. In the true DGP, the low frequency variable does not cause the high frequency variable, while for $\nu \neq 0$ the high frequency variable causes the low frequency variable with a marginal impact of $\nu/\sqrt{T}$ which vanishes as $T \to \infty$. First note we have $p = h = 1$. We will therefore simplify notation, namely denote the least squares asymptotic covariance matrix $\boldsymbol{\Sigma}_p(h)$ as $\boldsymbol{\Sigma}_1$.

Assuming stock sampling and general $m \in \mathcal{N}$, the corresponding MF-VAR(1) process of dimension $K = m + 1$ (since $K_H = K_L = 1$) is as follows:

$$\boldsymbol{X}(\tau_L) = \boldsymbol{A}(\nu/\sqrt{T})\boldsymbol{X}(\tau_L - 1) + \boldsymbol{\epsilon}(\tau_L), \tag{2.5.2}$$

where

$$\boldsymbol{A}(\nu/\sqrt{T}) = \begin{bmatrix} \mathbf{0}_{1\times(m-1)} & \rho_H & 0 \\ \vdots & \vdots & \vdots \\ \mathbf{0}_{1\times(m-1)} & \rho_H^m & 0 \\ \mathbf{0}_{1\times(m-1)} & \sum_{k=1}^{m} \rho_H^{k-1}\rho_L^{m-k}(\nu/\sqrt{T}) & \rho_L^m \end{bmatrix} \tag{2.5.3}$$

and $\boldsymbol{\epsilon}(\tau_L) \overset{i.i.d.}{\sim} (\mathbf{0}_{K\times 1}, \boldsymbol{\Omega})$ See (2.4.5)-(2.4.7) in Section 2.4.1 for a characterization of $\boldsymbol{\Omega}$. The MF-VAR(1) being estimated is:

$$\boldsymbol{X}(\tau_L) = \boldsymbol{A} \times \boldsymbol{X}(\tau_L - 1) + \boldsymbol{\epsilon}(\tau_L)$$

with coefficient matrix $\boldsymbol{A} = \boldsymbol{A}(\nu/\sqrt{T})$. Table 2.1 and Theorem 2.3.1 provide us the Case I selection matrix $\boldsymbol{R}$ to formulate the null hypothesis of high-to-low non-causality:

$$H_0^I : \boldsymbol{R}\text{vec}\left[\boldsymbol{A}'\right] = \mathbf{0}_{m\times 1} \quad \text{where} \quad \boldsymbol{R} \in \mathcal{R}^{m\times K^2}.$$

Thus, the corresponding local alternatives $H_A^{I,L}$ are written as

$$H_A^{I,L} : \boldsymbol{R}\text{vec}\left[\boldsymbol{A}'\right] = (\nu/\sqrt{T})\boldsymbol{a},$$

where by (2.5.3) it follows $\boldsymbol{a}$ is the $m \times 1$ vector $[0, ..., 0, \sum_{k=1}^{m} \rho_H^{k-1}\rho_L^{m-k}]'$. Now let $\hat{\boldsymbol{A}}$ be the least squares estimator of $\boldsymbol{A}$. Theorem 2.2.2 implies that $W[H_0^I] \xrightarrow{d} \chi_m^2$ as $T \to \infty$ under H_0^I. Similarly, by classic arguments it is easy to verify under $H_A^{I,L}$ that $W[H_A^{I,L}] \xrightarrow{d} \chi_m^2(\kappa_{MF})$, where $\chi_m^2(\kappa_{MF})$ is the non-central chi-squared distribution with m degrees of freedom and non-centrality parameter κ_{MF} :

$$\kappa_{MF} = \nu^2 \boldsymbol{a}' \left[\boldsymbol{R}\boldsymbol{\Sigma}_1\boldsymbol{R}'\right]^{-1} \boldsymbol{a}, \tag{2.5.4}$$

where $\boldsymbol{\Sigma}_1$ is the asymptotic variance of $\hat{\boldsymbol{A}}$, in particular

$$\boldsymbol{\Sigma}_1 = \boldsymbol{\Omega} \otimes \boldsymbol{\Upsilon}_0^{-1} \quad \text{with} \quad \boldsymbol{\Upsilon}_0 = \sum_{i=0}^{\infty} \boldsymbol{A}^i \boldsymbol{\Omega} \boldsymbol{A}^{i'} \quad \text{where} \quad \boldsymbol{A} \equiv \lim_{T\to\infty} \boldsymbol{A}(\nu/\sqrt{T}). \tag{2.5.5}$$

Equation (2.5.5) can be obtained from non-local least squares asymptotics with $\boldsymbol{A} \equiv \lim_{T\to\infty} \boldsymbol{A}$ $(\nu/\sqrt{T})$. See Appendix A.1.1 for details on deriving $\boldsymbol{\Sigma}_1$ in (2.5.5). Using the discrete Lyapunov equation, $\boldsymbol{\Upsilon}_0$ can be characterized by:

$$\text{vec}[\boldsymbol{\Upsilon}_0] = (\boldsymbol{I}_{K^2} - \boldsymbol{A} \otimes \boldsymbol{A})^{-1}\text{vec}[\boldsymbol{\Omega}].$$

Let $F_0 : \mathcal{R} \to [0,1]$ be the cumulative distribution function (c.d.f.) of the null distribution, χ^2_m. Similarly, let $F_1 : \mathcal{R} \to [0,1]$ be the c.d.f. of the alternative distribution, $\chi^2_m(\kappa_{MF})$. The local asymptotic power of the MF high-to-low causality test, $\mathcal{P}$, is given by:

$$\mathcal{P} = 1 - F_1\left[F_0^{-1}(1-\alpha)\right], \tag{2.5.6}$$

where $\alpha \in [0,1]$ is a nominal size.

We now derive the local asymptotic power of the LF high-to-low causality test. First, the LF-VAR(1) process corresponding to (2.5.1) is given by:

$$\underline{\boldsymbol{X}}(\tau_L) = \underline{\boldsymbol{A}}(\nu/\sqrt{T})\underline{\boldsymbol{X}}(\tau_L - 1) + \underline{\boldsymbol{\epsilon}}(\tau_L), \tag{2.5.7}$$

where

$$\underline{\boldsymbol{A}}(\nu/\sqrt{T}) = \begin{bmatrix} \rho_H^m & 0 \\ \sum_{k=1}^m \rho_H^{k-1}\rho_L^{m-k}(\nu/\sqrt{T}) & \rho_L^m \end{bmatrix} \tag{2.5.8}$$

and $\underline{\boldsymbol{\epsilon}}(\tau_L) \overset{i.i.d.}{\sim} (\boldsymbol{0}_{2\times 1}, \underline{\boldsymbol{\Omega}})$. Note that $\underline{\boldsymbol{\Omega}}$ is characterized in (2.4.10).

Suppose that we fit a LF-VAR(1) model with coefficient matrix $\underline{\boldsymbol{A}} \in \mathcal{R}^{2\times 2}$, that is $\underline{\boldsymbol{X}}(\tau_L) = \underline{\boldsymbol{A}} \times \underline{\boldsymbol{X}}(\tau_L - 1) + \underline{\boldsymbol{\epsilon}}(\tau_L)$. The null hypothesis of high-to-low non-causality is that the lower-left element of $\underline{\boldsymbol{A}}$ is zero:

$$H_0^I : \boldsymbol{R}\text{vec}\left[\underline{\boldsymbol{A}}'\right] = 0,$$

where $\boldsymbol{R} = [0,0,1,0]$. The corresponding local alternative hypothesis is:

$$H_A^{I,L} : \boldsymbol{R}\text{vec}\left[\underline{\boldsymbol{A}}'\right] = \sum_{k=1}^m \rho_H^{k-1}\rho_L^{m-k}(\nu/\sqrt{T}).$$

Let $\hat{\underline{\boldsymbol{A}}}$ be the least squares estimator of $\underline{\boldsymbol{A}}$. We have that $W[H_0^I] \xrightarrow{d} \chi_1^2$ as $T \to \infty$ under H_0^I, while $W[H_A^{I,L}] \xrightarrow{d} \chi_1^2(\kappa_{LF})$ under $H_A^{I,L}$ with κ_{LF} given by:

$$\kappa_{LF} = \frac{\left(\nu \sum_{k=1}^{m} \rho_H^{k-1} \rho_L^{m-k}\right)^2}{\boldsymbol{R}\underline{\boldsymbol{\Sigma}}_{\mathbf{1}}\boldsymbol{R}'},$$

where $\underline{\boldsymbol{\Sigma}}_{\mathbf{1}}$ is the asymptotic variance of $\hat{\underline{\boldsymbol{A}}} \equiv \lim_{T\to\infty}\{\underline{\boldsymbol{A}}(\nu/\sqrt{T})\}$, in particular as in (2.5.5) it can be shown $\underline{\boldsymbol{\Sigma}}_1 = \underline{\boldsymbol{\Omega}} \otimes \underline{\boldsymbol{\Upsilon}}_0^{-1}$ with $\underline{\boldsymbol{\Upsilon}}_0 = \sum_{i=0}^{\infty} \underline{\boldsymbol{A}}^i \underline{\boldsymbol{\Omega}} \underline{\boldsymbol{A}}^{i'}$. The local asymptotic power of the LF high-to-low causality test is given by (2.5.6), where F_0 is the c.d.f. of χ_1^2 and F_1 is the c.d.f. of $\chi_1^2(\kappa_{LF})$.

Case II: Low-to-High Causality Assume that the true DGP is given by (2.5.1) with

$$\boldsymbol{\Phi}(\nu/\sqrt{T}) = \begin{bmatrix} \rho_H & \nu/\sqrt{T} \\ 0 & \rho_L \end{bmatrix}$$

with ρ_H, $\rho_L \in (-1, 1)$. Assume again that $\boldsymbol{\eta}(\tau_L, k) \overset{i.i.d.}{\sim} (\mathbf{0}_{2\times 1}, \boldsymbol{I}_2)$. In the true DGP, the high frequency variable does not cause the low frequency variable, while the low frequency variable causes the high frequency variable, a relationship which vanishes as $T \to \infty$.

Assuming stock sampling and general $m \in \mathcal{N}$, the corresponding MF-VAR(1) process is given by (2.5.2) with

$$\boldsymbol{A}(\nu/\sqrt{T}) = \begin{bmatrix} \mathbf{0}_{1\times(m-1)} & \rho_H & \sum_{k=1}^{1} \rho_H^{k-1} \rho_L^{1-k}(\nu/\sqrt{T}) \\ \vdots & \vdots & \vdots \\ \mathbf{0}_{1\times(m-1)} & \rho_H^m & \sum_{k=1}^{m} \rho_H^{k-1} \rho_L^{m-k}(\nu/\sqrt{T}) \\ \mathbf{0}_{1\times(m-1)} & 0 & \rho_L^m \end{bmatrix}. \tag{2.5.9}$$

Our model is again a MF-VAR(1) model, so the local asymptotic power of the MF low-to-high causality test can be computed exactly as in Case I with only two changes. First, $\boldsymbol{a}$ in (2.5.4) has different elements here: $\boldsymbol{a} = [\sum_{k=1}^{1} \rho_H^{k-1}\rho_L^{1-k}, \ldots, \sum_{k=1}^{m} \rho_H^{k-1}\rho_L^{m-k}]'$. Second, the selection matrix $\boldsymbol{R}$ is specified according to Case II in Section 2.3.2. These differences will produce an interesting asymmetry between the MF high-to-low causality test and the MF low-to-high

causality test.

We now consider the LF low-to-high causality test. The LF-VAR(1) process is given by:

$$\underline{\boldsymbol{A}}(\nu/\sqrt{T}) = \begin{bmatrix} \rho_H^m & \sum_{k=1}^m \rho_H^{k-1}\rho_L^{m-k}(\nu/\sqrt{T}) \\ 0 & \rho_L^m \end{bmatrix}. \tag{2.5.10}$$

The local asymptotic power of the LF low-to-high causality test can again be computed exactly as in Case I with the only difference being that $\boldsymbol{R} = [0, 1, 0, 0]$ here, so there is no asymmetry between the LF high-to-low causality test and the LF low-to-high causality test.

Numerical Exercises In order to study the local asymptotic power analysis more directly, we rely on some numerical calculations. In Figure 2.1 we plot the ratio of the local asymptotic power of the MF causality test to that of the LF causality test, which we call the *power ratio* hereafter. We assume a nominal size $\alpha = 0.05$. Panel A focuses on high-to-low causality, while Panel B focuses on low-to-high causality. Each panel has four figures depending on $\rho_H, \rho_L \in \{0.25, 0.75\}$. The x-axis of each figure has $\nu \in [0.5, 1.5]$, while the y-axis has $m \in \{3, \ldots, 12\}$. The case that $m = 3$ can be thought of as the month versus quarter case, while the case that $m = 12$ can be thought of as the month versus year case. Note that the scale of each z-axis is different.

In Panel A, the power ratio varies within $[0.5, 1]$, hence the MF causality test is as powerful as, or is in fact *less* powerful than, the LF causality test. This is reasonable since a MF process contains the same information about high-to-low causality test as the corresponding LF process does (cfr. (2.5.3), (2.5.8), and Theorem 2.4.3) and the former has more parameters: recall that $\boldsymbol{A}$ is $(m+1)\times(m+1)$ while $\underline{\boldsymbol{A}}$ is 2×2. The power ratio tends to be low in the bottom figures of Panel A, where $\rho_H = 0.75$. This result is also understandable since the information loss caused by aggregating a high frequency variable is less severe when it is more persistent.

Panel B highlights the advantage of the MF approach over the LF approach. Note that the power ratio always exceeds one and the largest value of the z-axis is 5, 15, 3, or 6 when (ρ_H, ρ_L) $= (0.25, 0.25)$, $(0.25, 0.75)$, $(0.75, 0.25)$, or $(0.75, 0.75)$, respectively. This result is consistent with (2.5.9), (2.5.10), and Theorem 2.4.3, where we show that a MF process contains more information about low-to-high causality test than the corresponding LF process does. Given

the same ρ_L, the power ratio tends to be low when the high frequency variable is more persistent. The reason for this result is again that aggregating a high frequency variable produces less severe information loss when it is more persistent.

Another interesting finding from Panel B is that the power ratio is decreasing in m for (ρ_H, ρ_L) = $(0.25, 0.25)$ and increasing in m for $(\rho_H, \rho_L) = (0.75, 0.75)$. In order to interpret this fact, let $\rho_H = \rho_L = \rho$ and consider a key quantity in the upper-right block of $\boldsymbol{A}$, $\sum_{k=1}^{m} \rho_H^{k-1} \rho_L^{m-k} = m\rho^{m-1} \equiv f(m)$. Given m, the upper-right block of $\boldsymbol{A}$ has $f(1), \ldots, f(m)$ while that of $\underline{\boldsymbol{A}}$ has $f(m)$ only, therefore it is $\{f(1), \ldots, f(m-1)\}$ that determines the power ratio. Hence, whether the power ratio increases or decreases by switching from m to $m+1$ depends on the magnitude of $f(m)$. If $f(m)$ is close to zero, then the power ratio decreases due to more parameters in a MF-VAR model and negligible informational gain from $f(m)$. If $f(m)$ is away from zero, then the power ratio increases since such a large coefficient helps us reject the incorrect null hypothesis of low-to-high non-causality. Figure 2.2 plots $f(m)$ for $\rho \in \{0.25, 0.75\}$. It shows that $f(m)$ converges to zero quickly as m grows when $\rho = 0.25$, while it does much more slowly when $\rho = 0.75$. Thus, the power ratio is decreasing in m for $\rho = 0.25$ and increasing in m for ρ = 0.75.

In summary, the local asymptotic power of the MF low-to-high causality test is higher than that of the LF counterpart. The ratio of the former to the latter increases as a high frequency variable gets less persistent, given the persistence of a low frequency variable. Moreover, the power ratio increases in m for persistent series, while it decreases in m for transitory series.

2.6 Power Improvements in Finite Samples

This section conducts Monte Carlo simulations for bivariate cases and trivariate cases to evaluate the finite sample performance of the mixed frequency causality test. In bivariate cases with stock sampling, we know how causality is transferred among HF-, MF-, and LF-VAR processes and hence we can compare the finite sample power of MF and LF causality tests. In trivariate cases we have little theoretical results on how causality is transferred because of potential spurious causality or non-causality, so our main exercise there is to evaluate the performance of the MF causality test itself by checking empirical size and power. In particular, we will show

that the mixed frequency causality test can capture causality chains under a realistic simulation design. All tests in this section are performed at the 5% level.

2.6.1 Bivariate Case

This subsection considers a bivariate HF-VAR(1) process with stock sampling as in Section 2.5 so that the corresponding MF- and LF-VAR processes are known. One drawback of this experimental design is that we cannot easily study flow sampling since the corresponding MF and LF processes only have VARMA representations of unknown order, and therefore may not have a finite order VAR representation, by Theorem 2.4.1.[10]

Simulation Design

We draw J independent samples from a HF-VAR(1) process $\{\overline{\boldsymbol{X}}(\tau_L, k)\}$ according to (2.4.1) with $\boldsymbol{\Phi}_1$ partitioned in two possible ways:

$$(a) \quad \begin{bmatrix} \phi_{HH,1} & \phi_{HL,1} \\ \phi_{LH,1} & \phi_{LL,1} \end{bmatrix} = \begin{bmatrix} 0.4 & 0.0 \\ 0.2 & 0.4 \end{bmatrix} \quad \text{and} \quad (b) \quad \begin{bmatrix} \phi_{HH,1} & \phi_{HL,1} \\ \phi_{LH,1} & \phi_{LL,1} \end{bmatrix} = \begin{bmatrix} 0.4 & 0.2 \\ 0.0 & 0.4 \end{bmatrix}.$$

Thus we have in (a) unidirectional causality from the high frequency variable to the low frequency variable and in (b) unidirectional causality from the low frequency variable to the high frequency variable. Since we assume stock sampling here, these causal patterns carry over to the corresponding MF- and LF-VAR processes under this parameterization. The innovations are either mutually and serially independent standard normal $\boldsymbol{\eta}(\tau_L, k) \overset{i.i.d.}{\sim} N(\mathbf{0}_{2\times 1}, \boldsymbol{I}_2)$, or follow a GARCH(1,1) process since many macroeconomic and financial time series exhibit volatility clustering. The latter is best represented using the single-index representation of (2.4.1): $\boldsymbol{Y}_t = \boldsymbol{\Phi}_1 \boldsymbol{Y}_{t-1} + \boldsymbol{\xi}_t$. The components $\xi_{i,t}$ of $\boldsymbol{\xi}_t$ are mutually independent GARCH(1,1) with the

[10] In simulations not reported here we explored Lütkepohl and Poskitt's (1996) finite-order approximation for VAR(∞). The resulting test exhibited large empirical size distortions and was therefore not considered in this paper.

same feedback structure:

$$\begin{aligned} &\xi_{i,t} = \sigma_{i,t} z_{i,t}, \quad \boldsymbol{z}_t \overset{i.i.d.}{\sim} N(\mathbf{0}_{2\times 1}, \boldsymbol{I}_2), \\ &\sigma^2_{i,t} = 0.1 + 0.05\xi^2_{i,t-1} + 0.9\sigma^2_{i,t-1}. \end{aligned} \tag{2.6.1}$$

The chosen parameter values are similar to those found in many macroeconomic and financial time series. In view of i.i.d. normality for the GARCH innovations the HF error process $\{\boldsymbol{\xi}_t\}$ is stationary geometrically α-mixing (cfr. Boussama (1998)), hence MF and LF errors are also geometrically α-mixing.

The low frequency sample size is $T_L \in \{50, 100, 500\}$. The sampling frequency is taken from $m \in \{2, 3\}$, so the high frequency sample size is $T = mT_L \in \{100, 150, 200, 300, 1000, 1500\}$. The case that $(m, T_L) = (3, 100)$ can be thought of as a month versus quarter case covering 25 years. When m takes a much larger value (e.g. $m = 12$ in month vs. year), our methodology loses practical applicability due to parameter proliferations. Handling a large m remains as a future research question.

We aggregate the HF data into MF data $\{\boldsymbol{X}(\tau_L)\}_1^{T_L}$ and LF data $\{\underline{\boldsymbol{X}}(\tau_L)\}_1^{T_L}$ using stock sampling; see (2.2.2) and (2.4.2). We then fit MF-VAR(1) and LF-VAR(1), which are correctly specified. Finally, we compute Wald statistics for two separate null hypotheses of high-to-low non-causality $H_{H\nrightarrow L}$: $x_H \nrightarrow x_L$ and low-to-high non-causality $H_{L\nrightarrow H}$: $x_L \nrightarrow x_H$, each for horizon $h = 1$.[11] The Wald statistic shown in (2.2.9) is computed by OLS with two covariance matrix estimators. The first one is based on the Bartlett kernel HAC estimator discussed in Appendix A.1.1. We use a bandwidth of the form $n_{T_L^*} \equiv \max\{1, \lambda(T_L^*)^{1/3}\}$ since this optimizes the estimator's rate of convergence (Newey and West (1994)), while λ is determined by Newey and West's (1994) automatic bandwidth selection. This so-called *HAC case* corresponds to a situation where the researcher merely uses one robust covariance estimation technique irrespective of theory results.[12] The second covariance matrix is the true analytical matrix, and is

[11] Note from (2.4.3) and (2.4.8) that $H_{H\nrightarrow L}$ corresponds to $\boldsymbol{A}_1(m+1, 1:m) = \mathbf{0}_{1\times m}$ in the MF-VAR and to $\underline{\boldsymbol{A}}_1(2,1) = 0$ in the LF-VAR models, while $H_{L\nrightarrow H}$ corresponds to $\boldsymbol{A}_1(1:m, m+1) = \mathbf{0}_{m\times 1}$ in the MF-VAR and to $\underline{\boldsymbol{A}}_1(1,2) = 0$ in the LF-VAR models.

[12] In the special case when $h = 1$, a consistent and *almost surely* positive definite least squares asymptotic variance estimator is easily computed without a long-run variance HAC estimator (see Appendix A.1.1). Based on this insight, we also tried a sufficiently small λ instead of Newey and West's (1994) automatic selection. The

therefore called the *benchmark case*. This case corresponds to a complete-information situation where the researcher knows the true parameters. The benchmark covariance matrix for the MF-VAR model can be computed according to (2.5.5). In the LF-VAR model, $\boldsymbol{A}$ and $\boldsymbol{\Omega}$ in that expression should be replaced with $\underline{\boldsymbol{A}}$ and $\underline{\boldsymbol{\Omega}}$, respectively (see (2.4.4), (2.4.5), (2.4.9), and (2.4.10)).

We circumvent size distortions for small samples $T_L \in \{50, 100\}$ by employing parametric bootstraps in Dufour, Pelletier, and Renault (2006) and Gonçalves and Killian (2004).[13] Dufour, Pelletier and Renault's (2006) procedure assumes i.i.d. errors with a known distribution while Gonçalves and Killian's (2004) wild bootstrap does not require knowledge of the true error distribution and is robust to conditional heteroskedasticity of unknown form. Although $p = h = 1$ in this specific experiment, we present the bootstrap procedures with general p and h for completeness. We present the concrete procedures with respect to $H_0^2(h) : \boldsymbol{x}_{H,i_1} \nrightarrow_h x_{L,j_1} | \mathcal{I}_{(H,i_1)}$, non-causality from the i_1-th high frequency variable to the j_1-th low frequency variable, but all other cases can be treated analogously.

We use Dufour, Pelletier and Renault's (2006) [DPR] parametric bootstrap for the model with i.i.d. errors. The model with GARCH errors leads to greater size distortions, hence in that case we use Gonçalves and Killian's (2004) [GK] wild bootstrap detailed below. The DPR bootstrap procedure in the MF-VAR case follows, the LF-VAR case being similar.

Step 1 We fit an unrestricted MF-VAR(p) model for prediction horizon one to get $\hat{\boldsymbol{B}}(1)$ and $\hat{\boldsymbol{\Omega}}$ (cfr. (2.2.3) and (2.2.6)). We also fit an unrestricted MF-VAR(p) model for prediction horizon h to get $\hat{\boldsymbol{B}}(h)$ (cfr. (2.2.5)).

Step 2 Using (2.2.9), we compute the Wald test statistic based on the actual data, $W[H_0^2(h)]$.

Step 3 We simulate N samples from (2.2.5) using $\boldsymbol{B}(h) = \hat{\boldsymbol{B}}(h)$ and $\boldsymbol{\Omega} = \hat{\boldsymbol{\Omega}}$ and the correct assumption that $\boldsymbol{\epsilon}(\tau_L)$ is jointly standard normal, where we impose parametric constraints corresponding to $H_0^2(h)$, found in (2.3.2) and Table 2.1. Estimates of the impulse response

---

results were similar to those of the HAC case, hence we do not reported them here.

[13] Chauvet, Götz, and Hecq (2013) explore an alternative approach of parameter reductions based on reduced rank conditions, the imposition of an ARX(1) structure on the high frequency variables, and the transformation of MF-VAR into LF-VAR models.

coefficients $\boldsymbol{\Psi}_k$ can be obtained using $\hat{\boldsymbol{B}}(1)$ and (2.2.4). We denote by $W_i[H_0^2(h)]$ the Wald test statistic based on the i-th simulated sample, where $i \in \{1, \ldots, N\}$.

Step 4 Finally, we compute the resulting p-value $\hat{p}_N(W[H_0^2(h)])$, defined as

$$\hat{p}_N(W[H_0^2(h)]) \equiv \frac{1}{N+1}\left(1 + \sum_{i=1}^{N} I(W_i[H_0^2(h)] \geq W[H_0^2(h)])\right).$$

The null hypothesis $H_0^2(h)$ is rejected at level α if $\hat{p}_N(W[H_0^2(h)]) \leq \alpha$.

We use the GK bootstrap for all models, hence for i.i.d. or GARCH errors. In this case bootstrap errors are drawn as $\hat{\boldsymbol{\epsilon}}(\tau_L) \circ \boldsymbol{\xi}(\tau_L)$ with $\boldsymbol{\xi}(\tau_L) \overset{i.i.d.}{\sim} N(\mathbf{0}_{K\times 1}, \boldsymbol{I}_K)$. All other steps remain the same as the DPR procedure above.

For small sample sizes $T_L \in \{50, 100\}$, we draw $J = 1,000$ samples with $N = 499$ bootstrap replications. For the larger sample size $T_L = 500$, we draw $J = 100,000$ samples without bootstrap since size distortions do not occur.

We expect the following two results based on Theorem 2.4.3 and Section 2.5. First, the MF high-to-low causality test should have the same or lower power than the LF high-to-low causality test does since they contain the same amount of causal information and the former entails more parameters. Second, the MF low-to-high causality test should have higher power than the LF low-to-high causality test does since the former contains more causal information than the latter.

Simulation Results

In Tables 2.2-2.4 we report rejection frequencies. These three tables are different in terms of the error structure and bootstrap method: i.i.d. error with the DPR bootstrap in Table 2.2, i.i.d. error with the GK bootstrap in Table 2.3, and GARCH error with the GK bootstrap in Table 2.4. Also, the benchmark case with analytical covariance matrices is omitted in Tables 2.3 and 2.4 since the HAC case and the benchmark case produce very similar results as shown in Table 2.2. Finally, the large sample case $T_L = 500$ with i.i.d. errors and *without* bootstrap is omitted in Table 2.3 simply because that is covered in Table 2.2.

Note that, in case (a), size is computed with respect to low-to-high causality while power is

computed with respect to high-to-low causality. In case (b), size is computed with respect to high-to-low causality, while power is computed with respect to low-to-high causality. Values in parentheses are the benchmark rejection frequencies based on the analytical covariance matrix, and values not in parentheses concern the HAC case.

Consider the model with i.i.d. error and use of the DPR bootstrap: Table 2.2. Empirical size varies within $[0.039, 0.069]$, so there are no serious size distortions in any case. Focusing on power, the results are consistent with the two conjectures above. First, the gap between rejection frequencies for MF and LF causality tests for $H_{H \nrightarrow L}$ is not large (see case (a) in Table 2.2). For example, when $(m, T_L) = (2, 50)$ and the HAC covariance matrix is used, power for the MF high-to-low causality test is 0.128 while power for the LF high-to-low causality test is 0.189. Second, the MF low-to-high causality test has clearly higher power than the LF counterpart (see case (b)). This difference is most prominent for the largest m and T_L, where the rejection frequencies in the HAC case are 0.997 and 0.556 for the MF- and LF-VAR models, respectively. All these implications hold for both the HAC case and the benchmark case.

The remaining simulation results are not too surprising. When Gonçalves and Killian's (2004) bootstrap is used for i.i.d. errors, the rejection frequencies are similar to when i.i.d. normality is merely assumed. In the GARCH case, empirical power tends to be slightly lower than the i.i.d. case, logically following from the added noise to the VAR signal.

2.6.2 Trivariate Case

We now focus on a trivariate MF-VAR model with multiple prediction horizons in order to see if the mixed frequency causality test can capture causality chains properly. While there is no clear theory on how causality is linked between MF- and LF-VAR processes in the presence of causality chains, we also consider LF-VAR models with flow sampling and stock sampling for comparison. We also allow for non-i.i.d. errors to better match conditional volatility dynamics in macroeconomic and financial data.

Simulation Design

Suppose that there are two high frequency variables X and Y and one low frequency variable Z with sampling frequency $m = 3$ so that $K_H = 2$, $K_L = 1$, and $K = 7$. The low frequency

sample size is $T_L = 100$. This setting matches with the empirical application in Section 2.7, where we analyze monthly inflation, monthly oil price changes, and quarterly real GDP growth covering 300 months (100 quarters, 25 years).

Define a mixed frequency vector:

$$\boldsymbol{W}(\tau_L) = [X(\tau_L, 1), Y(\tau_L, 1), X(\tau_L, 2), Y(\tau_L, 2), X(\tau_L, 3), Y(\tau_L, 3), Z(\tau_L)]'.$$

Our true DGP is MF-VAR(1):

$$\boldsymbol{W}(\tau_L) = \boldsymbol{A} \times \boldsymbol{W}(\tau_L - 1) + \boldsymbol{\epsilon}(\tau_L). \tag{2.6.2}$$

As in the bivariate model, we assume the errors $\{\boldsymbol{\epsilon}(\tau_L)\}$ are either mutually and serially independent standard normal, or are mutually independent GARCH. Taking the error $\epsilon_Z(\tau_L)$ for the low frequency variable Z as an example, the GARCH parameterization is identical to (2.6.1): $\epsilon_Z(\tau_L) = \sigma_Z(\tau_L)\eta(\tau_L)$ where $\eta(\tau_L) \overset{i.i.d.}{\sim} N(0,1)$ and $\sigma_Z^2(\tau_L) = 0.1 + 0.05\epsilon_Z^2(\tau_L - 1) + 0.9\sigma_Z^2(\tau_L - 1)$. The same GARCH structure is applied for high frequency errors of X and Y. The MF errors $\boldsymbol{\epsilon}(\tau_L)$ are therefore stationary and geometrically α-mixing (cfr. Nelson (1990), Boussama (1998), Carrasco and Chen (2002)). In the case of i.i.d. errors we use either Dufour, Pelletier and Renault's (2006) or Gonçalves and Killian's (2004) bootstrap, and in the case of GARCH errors we use Gonçalves and Killian's (2004) bootstrap since otherwise size distortions exist.

The coefficient matrix $\boldsymbol{A}$ in the DGP (2.6.2) is set as follows.

$$\boldsymbol{A} = \begin{bmatrix} 0.2 & 0 & -0.3 & 0 & 0.6 & 0 & 0 \\ \boxed{0.3} & 0.3 & \boxed{0.3} & -0.4 & \boxed{0.4} & 0.5 & 0 \\ 0 & 0 & -0.2 & 0 & 0.4 & 0 & 0 \\ \boxed{0} & 0 & \boxed{0.2} & 0.2 & \boxed{0.2} & 0.4 & 0 \\ 0 & 0 & 0 & 0 & 0.3 & 0 & 0 \\ \boxed{0} & 0 & \boxed{0} & 0 & \boxed{0.3} & 0.3 & 0 \\ \underline{0} & \boxed{0.3} & \underline{0} & \boxed{0.3} & \underline{0} & \boxed{0.4} & 0.6 \end{bmatrix}, \tag{2.6.3}$$

where the nine elements in rectangles represent the impact of X on Y, the three underlined elements represent the impact of X on Z, and the three boxed elements represent the impact of Y on Z. All other non-zero elements are autoregressive coefficients, so not directly relevant for causal patterns. Equation (2.6.3) thus implies that there are only two channels of causality at $h = 1$: $X \to_1 Y \,|\, \mathcal{I}$ and $Y \to_1 Z \,|\, \mathcal{I}$. In particular, note that X does not cause Z at $h = 1$. For $h \geq 2$, we have three channels of causality because of a causal chain from X to Z via Y: $X \to_h Y \,|\, \mathcal{I}$, $Y \to_h Z \,|\, \mathcal{I}$, and $X \to_h Z \,|\, \mathcal{I}$ (cfr. Dufour and Renault (1998)). This point is verified by observing $\boldsymbol{A}^2$ and $\boldsymbol{A}^3$:

$$\boldsymbol{A}^2 = \begin{bmatrix} 0.04 & 0 & 0 & 0 & 0.18 & 0 & 0 \\ \boxed{0.15} & 0.09 & \boxed{\text{-}0.14} & -0.04 & \boxed{0.61} & 0.14 & 0 \\ 0 & 0 & 0.04 & 0 & 0.04 & 0 & 0 \\ \boxed{0} & 0 & \boxed{\text{-}0.08} & 0.04 & \boxed{0.22} & 0.04 & 0 \\ 0 & 0 & 0 & 0 & 0.09 & 0 & 0 \\ \boxed{0} & 0 & \boxed{0} & 0 & \boxed{0.18} & 0.09 & 0 \\ \underline{0.09} & \boxed{0.27} & \underline{0.15} & \boxed{0} & \underline{0.30} & \boxed{0.63} & 0.36 \end{bmatrix} \tag{2.6.4}$$

and

$$\boldsymbol{A}^3 = \begin{bmatrix} 0.01 & 0 & -0.01 & 0 & 0.08 & 0 & 0 \\ \boxed{0.06} & 0.30 & \boxed{0.00} & -0.03 & \boxed{0.29} & 0.07 & 0 \\ 0 & 0 & -0.01 & 0 & 0.03 & 0 & 0 \\ \boxed{0} & 0 & \boxed{0.02} & -0.01 & \boxed{0.05} & 0.03 & 0 \\ 0 & 0 & 0 & 0 & 0.03 & 0 & 0 \\ \boxed{0} & 0 & \boxed{0} & 0 & \boxed{0.08} & 0.03 & 0 \\ \underline{0.10} & \boxed{0.19} & \underline{0.02} & \boxed{\text{-}0.00} & \underline{0.50} & \boxed{0.47} & 0.22 \end{bmatrix}. \tag{2.6.5}$$

We fit a (p, h)-autoregression with $p = 1$ and $h \in \{1, 2, 3\}$ to implement the mixed frequency causality test from an individual variable to another. We are particularly interested in whether we can find *non-causality* from X to Z at $h = 1$ and *causality* from X to Z at $h = 2, 3$. We draw $J = 1,000$ samples and $N = 499$ bootstrap samples to avoid size distortions. The HAC covariance estimator with Newey and West's (1994) automatic bandwidth selection is used as

in the bivariate simulation.

Aggregating the mixed frequency data $\{\boldsymbol{W}(\tau_L)\}$ into low frequency, we also fit a trivariate low frequency (p,h)-autoregression with $p,h \in \{1,2,3\}$ and then repeat the individual Granger causality tests. Given the presence of causal chains in the mixed frequency DGP, there is no theoretical conjecture on the causal pattern on the low frequency basis. Our exercise is thus simply observing how rejection frequencies change after temporal aggregation. As in the MF study, we draw $J = 1,000$ samples and $N = 499$ GK bootstrap samples to avoid size distortions.

Simulation Results

Table 2.5 reports the rejection frequencies on the mixed frequency basis. Empirical size always lies in $[0.037, 0.071]$, a fairly accurate result due to the DPR or GK bootstrap. Empirical size is in general more accurate when the errors are i.i.d., as expected; the boundary values 0.037 and 0.071 indeed realized in the GARCH case as seen in Panel C.

In the remaining discussion we will focus on empirical power in the case of GARCH errors with the GK bootstrapped p-values: Panel C. The other two panels have very similar implications and hence we will not mention them. Empirical power for the test of $X \nrightarrow_h Y$ is 0.994, 0.754, and 0.128 for horizons 1, 2, and 3, respectively. Diminishing power is reasonable given the diminishing impact of X on Y; see the elements in rectangles in (2.6.3), (2.6.4), and (2.6.5).

Power for the test of $Y \nrightarrow_h Z$ vanishes more slowly as h increases: 0.999, 0.989, and 0.724 for horizons 1, 2, and 3, respectively. In fact the boxed elements of $\boldsymbol{A}^2$ and $\boldsymbol{A}^3$ contain relatively large loadings 0.63 and 0.47, respectively. The intuitive reason for this slower decay is that Y has a more persistent impact on Z than X does on Y; see the upper triangular structure of the rectangles in (2.6.3).

Finally, the rejection frequency for $X \nrightarrow_h Z$ is 0.050, 0.594, and 0.648 for horizons 1, 2, and 3, respectively. At horizon 1 we get the desired result of *non-causality* from X to Z, while we have relatively high power for $h = 2, 3$ capturing the indirect impact of X on Z via Y (see the underlined elements in (2.6.3)-(2.6.5)). Thus, our mixed frequency causality test performs very well even in the presence of a causality chain.

We now review the results for LF-VAR. See Table 2.6 for flow sampling and Table 2.7 for stock sampling. In Panel A, the underlying mixed frequency error is i.i.d. and we use the GK

bootstrap. In Panel B, the underlying mixed frequency error is GARCH and we use the GK bootstrap. The DPR bootstrap is not considered since there is no theoretical guarantee that the LF-VAR process has i.i.d. errors even if the mixed frequency error is i.i.d. Focusing on flow sampling, we find that the rejection frequencies have very similar patterns with the mixed frequency experiments. First, the rejection frequencies on $X \nrightarrow_h Y$ are high at $h = 1$ but decay quickly. Second, the rejection frequencies on $Y \nrightarrow_h Z$ are high and decay much more slowly. Third, the rejection frequency on $X \nrightarrow_h Z$ is close to the nominal size 0.050 for $h = 1$ but soars to 0.442 - 0.690 for $h = 2, 3$. These results suggest that all causal patterns in the MF-VAR are preserved under flow sampling. Finally, empirical power tends to decrease as the LF-VAR lag length p increases from 1 to 3, which suggests that including one lag is enough to capture all causality patterns.

Turning on to stock sampling, there is an interesting difference from flow sampling. Focusing on Panel B, the rejection frequency on $X \nrightarrow_h Z$ for $h = 1$ is 0.073, 0.310, and 0.329 when p is 1, 2, and 3, respectively. This result suggests that X does cause Z at horizon $h = 1$ (i.e. spurious causality) while that can only be captured by including at least two lags (i.e. delayed causality). Since the rejection frequency on $X \nrightarrow_h Z$ for $h = 1$ was always close to the nominal size in the flow sampling case, we can conclude that different aggregation schemes may produce different causal patterns.

2.7 Empirical Application

In this section we apply the mixed frequency causality test to U.S. macroeconomic data. We consider 100× annual log-differences of the U.S. monthly consumer price index for all items (CPI), monthly West Texas Intermediate spot oil price (OIL), and quarterly real GDP from July 1987 through June 2012 as an illustrative example. We use year-to-year growth rates to control for likely seasonality in each series. CPI, OIL and GDP data are made publicly available by the U.S. Department of Labor, Energy Information Administration, and Bureau of Economic Analysis, respectively.

The causal relationship between oil and the macroeconomy has been a major applied research area as surveyed in Hamilton (2008). See Payne (2010) for an extensive survey on the

use of causality tests to determine the relationship between energy consumption and economic growth. We introduce the mixed frequency concept into these literatures by analyzing CPI, OIL, and GDP.

Figure 2.3 plots the three series, while Table 2.8 presents sample statistics. There is fairly strong positive correlation between CPI and OIL although the latter is much more volatile than the former. The sample standard deviation is 1.316% for CPI and 30.60% for OIL. The sample correlation coefficient between these two is 0.512 with the 95% confidence interval based on the Fisher transformation being [0.423, 0.591]. Since CPI, OIL, and GDP have a positive sample mean of 2.913%, 6.979%, and 2.513%, we de-mean each series and fit VAR without a constant term. The sample kurtosis is 4.495 for CPI, 3.485 for OIL, and 6.625 for GDP. These figures suggest that the three series follow non-normal distributions, but note that the asymptotic theory of the mixed frequency causality test is free of the normality assumption (cfr. Section 2.2).

Using mean-centered 100× annual log-differenced data, we fit an unrestricted MF-VAR(1) model with low frequency prediction horizon $h \in \{1, \ldots, 5\}$ to monthly CPI, monthly OIL, and quarterly GDP. We therefore have $K_H = 2$, $K_L = 1$, $m = 3$, $K = 7$, $T_L = 100$, and $T = 300$. This setting matches the one used in trivariate simulation study in Section 2.6.2. Since the dimension of the MF-VAR is $K = 7$, there are as many as 49 parameters even with the lag order one. Ghysels (2012) proposes a variety of parsimonious specifications based on the MIDAS literature, but they involve nonlinear parameter constraints. The trade-off between unrestricted and restricted MIDAS regressions is discussed in Foroni, Marcellino, and Schumacher (2013). A general consensus is that the unrestricted approach achieves higher prediction accuracy when m is small, such as monthly and quarterly ($m = 3$).

All six causal patterns (CPI$\nrightarrow$OIL, CPI$\nrightarrow$GDP, OIL$\nrightarrow$GDP and their converses) are tested. We use Newey and West's (1987) kernel-based HAC covariance estimator with Newey and West's (1994) automatic lag selection. In order to avoid potential size distortions and to allow for conditional heteroskedasticity of unknown form, we use Gonçalves and Killian's (2004) bootstrap with $N = 999$ replications. See Section 2.6 for the details.

For the purpose of comparison, we also fit an unrestricted LF-VAR(4) model with low frequency prediction horizon $h \in \{1, \ldots, 5\}$ to quarterly CPI, quarterly OIL, and quarterly

GDP. Since parameter proliferation is less of an issue in LF-VAR, we let the lag order be 4 in order to take potential seasonality into account.

Table 2.9 presents bootstrapped p-values for MF and LF tests at each quarterly horizon $h \in \{1, \ldots, 5\}$ (recall h is the *low* frequency prediction horizon). We denote whether rejection occurs at the 5% or 10% level. Note that the MF and LF approaches result in very different conclusions at standard levels of significance. At the 5% level, for example, the MF case reveals three significant causal patterns: CPI causes GDP at horizon 3, OIL causes CPI at horizons 1 and 4, and GDP causes CPI at horizon 1. The LF case, however, has two different significant causal patterns: CPI causes OIL at horizon 1 and OIL causes GDP at horizons 2 and 4.

Note that significant causality from OIL to CPI is found by the MF approach but not by the LF approach, whether the 5% level or 10% level is used. Intuitively, such a causality should exist since (i) oil products are a component of the all-item CPI and (ii) crude oil is a key natural resource for most sectors. Our result suggests that the quarterly frequency is too coarse to capture the OIL-to-CPI causality while the mixed frequency data contain enough information for us to capture it successfully. Conversely, none of the LF causal patterns appear in the MF data. For example, in LF data at the 5% level CPI causes OIL at horizon 1. The p-value is .035, roughly $1/10^{th}$ the magnitude of the MF p-value. Similarly, OIL causes GDP in low frequency data with p-values less than $1/10^{th}$ the MF p-values.[14]

2.8 Concluding Remarks

Time series processes are often sampled at different frequencies and are typically aggregated to the common lowest frequency to test for Granger causality. This paper compares testing for Granger causality with all series aggregated to the common lowest frequency, and testing for Granger causality taking advantage of all the series sampled at whatever frequency they are

[14] In view of Theorem 3.2 of Hill (2007), our empirical results have some conflicts with causation theory. Focusing on the mixed frequency case in Table 2.9, the significant causation from CPI to GDP at horizon 3 implies a causal chain via OIL at least from a theoretical point of view. We do not observe significant causation from CPI to OIL or causation from OIL to GDP, however. The LF approach is facing a similar problem; OIL causes GDP at horizons 2 and 4 but we do not observe significant causation from OIL to CPI or causation from CPI to GDP. As noted in footnote 7, this sort of discrepancy stems from treating each prediction horizon h separately. Hill (2007) proposes a sequential multi-horizon test as a solution which in principle can be applied to the present MF context for small horizon causation tests, e.g. $h \leq 3$.

available. We rely on mixed frequency vector autoregressive models to implement the new class of Granger causality tests.

We show that mixed frequency causality tests better recover causality patterns in an underlying high frequency process compared to the traditional low frequency approach. Moreover, we show formally that mixed frequency causality tests have higher asymptotic power against local alternatives and show via simulation that this also holds in finite samples involving realistic data generating processes. The simulations indicate that the mixed frequency VAR approach works well for small differences in sampling frequencies like month versus quarter.

We apply the mixed frequency causality test to a monthly consumer price index, monthly crude oil prices, and the real GDP in the U.S. We also apply the conventional low frequency causality test to the aggregated quarterly price series and the real GDP for comparison. These two approaches produce very different results at any standard level of significance. In particular, significant causality from oil prices to CPI is detected by the mixed frequency approach but not by the low frequency approach. This result suggests that the quarterly frequency is too coarse to capture such causality while the mixed frequency data contain enough information for us to capture that successfully.

Table 2.1: Linear Parametric Restrictions of Non-causality

The null hypotheses of non-causality cases $H_0^i(h)$ for $i = 1, \ldots, 4$ and I and II. can be written as $\boldsymbol{A}_k^{(h)}(a : \iota : b, c : \iota' : d) = \mathbf{0}$ for all $k \in \{1, \ldots, p\}$, where a, ι, b, c, ι', d, and the size of the null vector appear as entries to the table.

Cases	a	ι	b	c	ι'	d	$\mathbf{0}$
$H_0^1(h)$	$mK_H + j_2$	1	$mK_H + j_2$	$mK_H + j_1$	1	$mK_H + j_1$	1×1
$H_0^2(h)$	$mK_H + j_1$	1	$mK_H + j_1$	i_1	K_H	$i_1 + (m-1)K_H$	$1 \times m$
$H_0^3(h)$	i_1	K_H	$i_1 + (m-1)K_H$	$mK_H + j_1$	1	$mK_H + j_1$	$m \times 1$
$H_0^4(h)$	i_2	K_H	$i_2 + (m-1)K_H$	i_1	K_H	$i_1 + (m-1)K_H$	$m \times m$
$H_0^I(h)$	$mK_H + 1$	1	K	1	1	mK_H	$K_L \times mK_H$
$H_0^{II}(h)$	1	1	mK_H	$mK_H + 1$	1	K	$mK_H \times K_L$

Table 2.2: Rejection Frequencies (Bivariate VAR with i.i.d. Error and DPR Bootstrap)

Rejection frequencies at the 5% level for mixed and low frequency causality tests at horizon $h = 1$. The error term in the true DGP is i.i.d. Stock sampling is used when we aggregate mixed frequency data into low frequency data. The two cases are (a) $\phi_{HL,1} = 0$ and $\phi_{LH,1} = 0.2$ (unidirectional high-to-low causality) and (b) $\phi_{HL,1} = 0.2$ and $\phi_{LH,1} = 0$ (unidirectional low-to-high causality). In case (a), size is computed with respect to low-to-high causality, while power is computed with respect to high-to-low causality. In case (b), size is computed with respect to high-to-low causality, while power is computed with respect to low-to-high causality. Entries in parentheses are based on the benchmark analytical covariance matrix, and entries not in parentheses are based on the HAC estimator. Dufour, Pelletier, and Renault's (2006) [DPR] parametric bootstrap is employed for $T_L \in \{50, 100\}$ to avoid size distortions. m is the sampling frequency and T_L is the sample size in terms of low frequency.

Sample Size $T_L = 50$ (DPR bootstrap)

	Case (a)			Case (b)	
	m=2	m=3		m=2	m=3
Size	MF: 0.063(0.059) LF: 0.057(0.059)	MF: 0.053(0.045) LF: 0.063(0.054)		MF: 0.056(0.055) LF: 0.053(0.051)	MF: 0.039(0.051) LF: 0.044(0.046)
Power	MF: 0.128(0.155) LF: 0.189(0.198)	MF: 0.060(0.068) LF: 0.072(0.088)		MF: 0.241(0.266) LF: 0.175(0.205)	MF: 0.187(0.237) LF: 0.097(0.110)

Sample Size $T_L = 100$ (DPR bootstrap)

	Case (a)			Case (b)	
	m=2	m=3		m=2	m=3
Size	MF: 0.051(0.062) LF: 0.047(0.056)	MF: 0.045(0.040) LF: 0.042(0.051)		MF: 0.050(0.051) LF: 0.049(0.050)	MF: 0.060(0.069) LF: 0.053(0.056)
Power	MF: 0.221(0.262) LF: 0.311(0.338)	MF: 0.098(0.120) LF: 0.133(0.150)		MF: 0.456(0.506) LF: 0.323(0.340)	MF: 0.415(0.454) LF: 0.163(0.168)

Sample Size $T_L = 500$

	Case (a)			Case (b)	
	m=2	m=3		m=2	m=3
Size	MF: 0.059(0.051) LF: 0.056(0.052)	MF: 0.064(0.051) LF: 0.055(0.051)		MF: 0.060(0.052) LF: 0.056(0.050)	MF: 0.066(0.052) LF: 0.056(0.053)
Power	MF: 0.900(0.898) LF: 0.943(0.944)	MF: 0.414(0.390) LF: 0.557(0.551)		MF: 0.998(0.998) LF: 0.943(0.944)	MF: 0.997(0.997) LF: 0.556(0.550)

Table 2.3: Rejection Frequencies (Bivariate VAR with i.i.d. Error and GK Bootstrap)

Rejection frequencies at the 5% level for mixed and low frequency causality tests at horizon $h = 1$. The error term in the DGP is i.i.d. Stock sampling is used when we aggregate mixed frequency data into low frequency data. The two cases are (a) $\phi_{HL,1} = 0$ and $\phi_{LH,1} = 0.2$ (unidirectional high-to-low causality) and (b) $\phi_{HL,1} = 0.2$ and $\phi_{LH,1} = 0$ (unidirectional low-to-high causality). In case (a), size is computed with respect to low-to-high causality, while power is computed with respect to high-to-low causality. In case (b), size is computed with respect to high-to-low causality, while power is computed with respect to low-to-high causality. The HAC covariance estimator with Newey and West's (1994) automatic bandwidth selection is used. Gonçalves and Killian's (2004) [GK] wild bootstrap is employed for $T_L \in \{50, 100\}$ to avoid size distortions. See Table 2.2 for the result with $T_L = 500$ and without bootstrap. m is the sampling frequency and T_L is the sample size in terms of low frequency.

Sample Size $T_L = 50$ (GK bootstrap)

	Case (a)			Case (b)	
	m=2	m=3		m=2	m=3
Size	MF: 0.071 LF: 0.055	MF: 0.037 LF: 0.063		MF: 0.061 LF: 0.054	MF: 0.049 LF: 0.045
Power	MF: 0.135 LF: 0.187	MF: 0.076 LF: 0.073		MF: 0.216 LF: 0.173	MF: 0.161 LF: 0.102

Sample Size $T_L = 100$ (GK bootstrap)

	Case (a)			Case (b)	
	m=2	m=3		m=2	m=3
Size	MF: 0.054 LF: 0.047	MF: 0.038 LF: 0.043		MF: 0.050 LF: 0.046	MF: 0.066 LF: 0.056
Power	MF: 0.238 LF: 0.293	MF: 0.117 LF: 0.134		MF: 0.435 LF: 0.318	MF: 0.386 LF: 0.158

Table 2.4: Rejection Frequencies (Bivariate VAR with GARCH Error and GK Bootstrap)

Rejection frequencies at the 5% level for mixed and low frequency causality tests at horizon $h = 1$. The error term in the true DGP follows a GARCH process. Stock sampling is used when we aggregate mixed frequency data into low frequency data. The two cases are (a) $\phi_{HL,1} = 0$ and $\phi_{LH,1} = 0.2$ (unidirectional high-to-low causality) and (b) $\phi_{HL,1} = 0.2$ and $\phi_{LH,1} = 0$ (unidirectional low-to-high causality). In case (a), size is computed with respect to low-to-high causality, while power is computed with respect to high-to-low causality. In case (b), size is computed with respect to high-to-low causality, while power is computed with respect to low-to-high causality. The HAC covariance estimator with Newey and West's (1994) automatic bandwidth selection is used. Gonçalves and Killian's (2004) [GK] wild bootstrap is employed for $T_L \in \{50, 100\}$ to avoid size distortions. m is the sampling frequency and T_L is the sample size in terms of low frequency.

Sample Size $T_L = 50$ (GK bootstrap)

	Case (a)			Case (b)	
	m=2	m=3		m=2	m=3
Size	MF: 0.053 LF: 0.047	MF: 0.035 LF: 0.039		MF: 0.066 LF: 0.056	MF: 0.055 LF: 0.043
Power	MF: 0.136 LF: 0.143	MF: 0.079 LF: 0.083		MF: 0.228 LF: 0.188	MF: 0.142 LF: 0.090

Sample Size $T_L = 100$ (GK bootstrap)

	Case (a)			Case (b)	
	m=2	m=3		m=2	m=3
Size	MF: 0.050 LF: 0.044	MF: 0.047 LF: 0.052		MF: 0.056 LF: 0.051	MF: 0.057 LF: 0.048
Power	MF: 0.227 LF: 0.314	MF: 0.092 LF: 0.132		MF: 0.416 LF: 0.306	MF: 0.353 LF: 0.146

Sample Size $T_L = 500$

	Case (a)			Case (b)	
	m=2	m=3		m=2	m=3
Size	MF: 0.060 LF: 0.055	MF: 0.068 LF: 0.057		MF: 0.061 LF: 0.057	MF: 0.065 LF: 0.056
Power	MF: 0.894 LF: 0.937	MF: 0.418 LF: 0.554		MF: 0.997 LF: 0.937	MF: 0.996 LF: 0.553

Table 2.5: Rejection Frequencies for Trivariate MF-VAR

Rejection frequencies at the 5% level based on (p, h)-autoregression with $p = 1$ and $h \in \{1, 2, 3\}$, where we have two high frequency variables X and Y and one low frequency variable Z with $m = 3$. Each test deals with the null hypothesis of non-causality from an individual variable to another at horizon h. The upper-right triangular matrices have empirical size for $Y \nrightarrow_h X$, $Z \nrightarrow_h X$, and $Z \nrightarrow_h Y$. Also, the rejection frequency for $X \nrightarrow_1 Z$ is regarded as empirical size since X indeed does not cause Z at horizon 1. All other slots represent empirical power. We draw $J = 1,000$ samples and $N = 499$ bootstrap replications. The HAC covariance estimator with Newey and West's (1994) automatic bandwidth selection is used. The error term in the true DGP follows either an i.i.d. or a GARCH process, and we use either Dufour, Pelletier and Renault's (2006) [DPR] or Gonçalves and Killian's (2004) [GK] bootstrapped p-value.

Panel A: i.i.d. Error and DPR Bootstrap

Null Hypothesis	$h = 1$	$h = 2$	$h = 3$
$\begin{bmatrix} - & Y \nrightarrow_h X & Z \nrightarrow_h X \\ X \nrightarrow_h Y & - & Z \nrightarrow_h Y \\ X \nrightarrow_h Z & Y \nrightarrow_h Z & - \end{bmatrix}$	$\begin{bmatrix} - & 0.050 & 0.052 \\ 0.997 & - & 0.050 \\ 0.051 & 0.999 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.057 & 0.041 \\ 0.822 & - & 0.054 \\ 0.567 & 0.996 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.049 & 0.062 \\ 0.154 & - & 0.041 \\ 0.648 & 0.711 & - \end{bmatrix}$

Panel B: i.i.d. Error and GK Bootstrap

Null Hypothesis	$h = 1$	$h = 2$	$h = 3$
$\begin{bmatrix} - & Y \nrightarrow_h X & Z \nrightarrow_h X \\ X \nrightarrow_h Y & - & Z \nrightarrow_h Y \\ X \nrightarrow_h Z & Y \nrightarrow_h Z & - \end{bmatrix}$	$\begin{bmatrix} - & 0.054 & 0.062 \\ 0.998 & - & 0.055 \\ 0.064 & 0.999 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.055 & 0.046 \\ 0.794 & - & 0.053 \\ 0.568 & 0.997 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.045 & 0.065 \\ 0.134 & - & 0.047 \\ 0.633 & 0.703 & - \end{bmatrix}$

Panel C: GARCH Error and GK Bootstrap

Null Hypothesis	$h = 1$	$h = 2$	$h = 3$
$\begin{bmatrix} - & Y \nrightarrow_h X & Z \nrightarrow_h X \\ X \nrightarrow_h Y & - & Z \nrightarrow_h Y \\ X \nrightarrow_h Z & Y \nrightarrow_h Z & - \end{bmatrix}$	$\begin{bmatrix} - & 0.061 & 0.053 \\ 0.994 & - & 0.044 \\ 0.050 & 0.999 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.054 & 0.071 \\ 0.754 & - & 0.050 \\ 0.594 & 0.989 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.047 & 0.064 \\ 0.128 & - & 0.037 \\ 0.648 & 0.724 & - \end{bmatrix}$

Table 2.6: Rejection Frequencies for Trivariate LF-VAR (Flow Sampling)

Rejection frequencies at the 5% level based on (p, h)-autoregression with $p, h \in \{1, 2, 3\}$, where we have two high frequency variables X and Y and one low frequency variable Z with $m = 3$. The high frequency variables X and Y are aggregated into low frequency using flow sampling. Each test deals with the null hypothesis of non-causality from an individual variable to another at horizon h. We draw $J = 1,000$ samples and $N = 499$ bootstrap replications. The HAC covariance estimator with Newey and West's (1994) automatic bandwidth selection is used. The error term in the true mixed frequency DGP follows either an i.i.d. or a GARCH process, and we use Gonçalves and Killian's (2004) [GK] bootstrapped p-value.

$$\text{Null hypothesis: } \begin{bmatrix} - & Y \nrightarrow_h X & Z \nrightarrow_h X \\ X \nrightarrow_h Y & - & Z \nrightarrow_h Y \\ X \nrightarrow_h Z & Y \nrightarrow_h Z & - \end{bmatrix}$$

Panel A: i.i.d. Error and GK Bootstrap

Lag length \ Prediction horizon	$h = 1$	$h = 2$	$h = 3$
$p = 1$	$\begin{bmatrix} - & 0.052 & 0.065 \\ 0.999 & - & 0.051 \\ 0.063 & 0.999 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.037 & 0.058 \\ 0.516 & - & 0.069 \\ 0.690 & 0.989 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.044 & 0.049 \\ 0.151 & - & 0.063 \\ 0.570 & 0.730 & - \end{bmatrix}$
$p = 2$	$\begin{bmatrix} - & 0.057 & 0.055 \\ 0.994 & - & 0.056 \\ 0.045 & 0.998 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.052 & 0.051 \\ 0.374 & - & 0.040 \\ 0.607 & 0.925 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.046 & 0.052 \\ 0.138 & - & 0.048 \\ 0.534 & 0.521 & - \end{bmatrix}$
$p = 3$	$\begin{bmatrix} - & 0.047 & 0.057 \\ 0.989 & - & 0.046 \\ 0.049 & 0.998 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.054 & 0.058 \\ 0.307 & - & 0.049 \\ 0.569 & 0.894 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.052 & 0.056 \\ 0.115 & - & 0.056 \\ 0.442 & 0.411 & - \end{bmatrix}$

Panel B: GARCH Error and GK Bootstrap

Lag length \ Prediction horizon	$h = 1$	$h = 2$	$h = 3$
$p = 1$	$\begin{bmatrix} - & 0.048 & 0.046 \\ 0.999 & - & 0.060 \\ 0.053 & 0.999 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.058 & 0.064 \\ 0.466 & - & 0.041 \\ 0.713 & 0.992 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.043 & 0.051 \\ 0.163 & - & 0.048 \\ 0.579 & 0.773 & - \end{bmatrix}$
$p = 2$	$\begin{bmatrix} - & 0.045 & 0.062 \\ 0.996 & - & 0.063 \\ 0.052 & 0.998 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.046 & 0.060 \\ 0.374 & - & 0.045 \\ 0.591 & 0.936 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.049 & 0.049 \\ 0.115 & - & 0.059 \\ 0.501 & 0.532 & - \end{bmatrix}$
$p = 3$	$\begin{bmatrix} - & 0.049 & 0.056 \\ 0.988 & - & 0.050 \\ 0.052 & 0.999 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.048 & 0.065 \\ 0.315 & - & 0.053 \\ 0.541 & 0.876 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.063 & 0.071 \\ 0.101 & - & 0.050 \\ 0.426 & 0.416 & - \end{bmatrix}$

Table 2.7: Rejection Frequencies for Trivariate LF-VAR (Stock Sampling)

Rejection frequencies at the 5% level based on (p, h)-autoregression with $p, h \in \{1, 2, 3\}$, where we have two high frequency variables X and Y and one low frequency variable Z with $m = 3$. The high frequency variables X and Y are aggregated into low frequency using stock sampling. Each test deals with the null hypothesis of non-causality from an individual variable to another at horizon h. We draw $J = 1,000$ samples and $N = 499$ bootstrap replications. The HAC covariance estimator with Newey and West's (1994) automatic bandwidth selection is used. The error term in the true mixed frequency DGP follows either an i.i.d. or a GARCH process, and we use Gonçalves and Killian's (2004) [GK] bootstrapped p-value.

$$\text{Null hypothesis: } \begin{bmatrix} - & Y \nrightarrow_h X & Z \nrightarrow_h X \\ X \nrightarrow_h Y & - & Z \nrightarrow_h Y \\ X \nrightarrow_h Z & Y \nrightarrow_h Z & - \end{bmatrix}$$

Panel A: i.i.d. Error and GK Bootstrap

Lag length \ Prediction horizon	$h = 1$	$h = 2$	$h = 3$
$p = 1$	$\begin{bmatrix} - & 0.043 & 0.046 \\ 0.822 & - & 0.057 \\ 0.058 & 0.984 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.041 & 0.061 \\ 0.310 & - & 0.052 \\ 0.604 & 0.998 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.050 & 0.052 \\ 0.079 & - & 0.055 \\ 0.788 & 0.828 & - \end{bmatrix}$
$p = 2$	$\begin{bmatrix} - & 0.043 & 0.059 \\ 0.697 & - & 0.039 \\ 0.295 & 0.979 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.053 & 0.041 \\ 0.254 & - & 0.048 \\ 0.629 & 0.962 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.052 & 0.056 \\ 0.069 & - & 0.043 \\ 0.710 & 0.638 & - \end{bmatrix}$
$p = 3$	$\begin{bmatrix} - & 0.045 & 0.051 \\ 0.618 & - & 0.061 \\ 0.338 & 0.947 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.055 & 0.061 \\ 0.173 & - & 0.043 \\ 0.556 & 0.918 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.056 & 0.053 \\ 0.073 & - & 0.057 \\ 0.647 & 0.518 & - \end{bmatrix}$

Panel B: GARCH Error and GK Bootstrap

Lag length \ Prediction horizon	$h = 1$	$h = 2$	$h = 3$
$p = 1$	$\begin{bmatrix} - & 0.051 & 0.046 \\ 0.823 & - & 0.053 \\ 0.073 & 0.998 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.047 & 0.050 \\ 0.283 & - & 0.056 \\ 0.620 & 0.998 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.045 & 0.054 \\ 0.087 & - & 0.039 \\ 0.784 & 0.820 & - \end{bmatrix}$
$p = 2$	$\begin{bmatrix} - & 0.053 & 0.053 \\ 0.705 & - & 0.050 \\ 0.310 & 0.975 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.045 & 0.053 \\ 0.230 & - & 0.063 \\ 0.585 & 0.959 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.042 & 0.066 \\ 0.080 & - & 0.068 \\ 0.699 & 0.630 & - \end{bmatrix}$
$p = 3$	$\begin{bmatrix} - & 0.063 & 0.057 \\ 0.618 & - & 0.062 \\ 0.329 & 0.966 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.061 & 0.056 \\ 0.198 & - & 0.069 \\ 0.555 & 0.936 & - \end{bmatrix}$	$\begin{bmatrix} - & 0.046 & 0.053 \\ 0.081 & - & 0.058 \\ 0.617 & 0.521 & - \end{bmatrix}$

Table 2.8: Sample Statistics

Sample statistics for 100× annual log-differences of monthly U.S. CPI, monthly spot West Texas Intermediate oil price, and quarterly real GDP. The sample period is July 1987 through June 2012.

	mean	median	std. dev.	skewness	kurtosis
CPI	2.913	2.900	1.316	-0.392	4.495
OIL	6.979	7.777	30.60	-0.312	3.485
GDP	2.513	2.783	1.882	-1.670	6.625

Table 2.9: Granger Causality Tests for CPI, OIL, and GDP

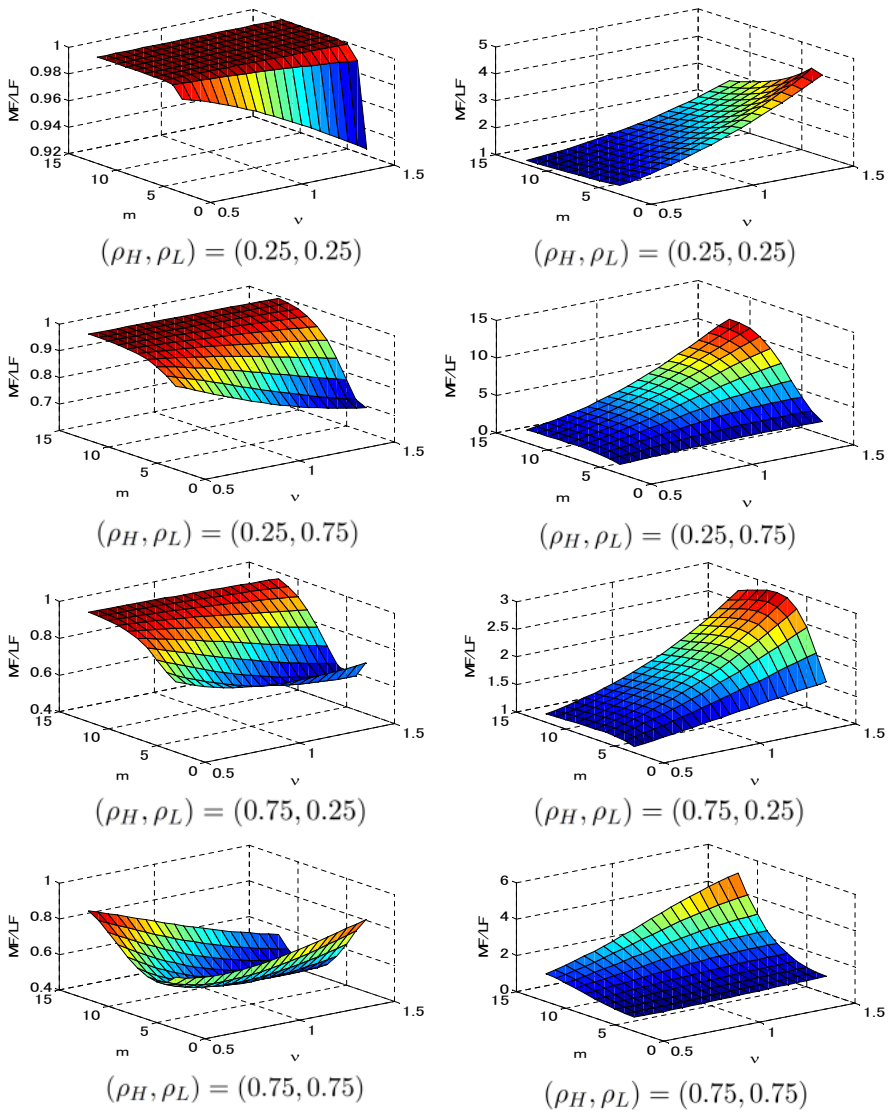
The mixed frequency approach uses monthly CPI, monthly OIL, and quarterly GDP. The low frequency approach uses all quarterly series. A $\boxed{\text{box}}$ indicates rejection at the 5% level of the null hypothesis of non-causality at the quarterly horizon $h \in \{1, \ldots, 5\}$. A circle $\circ$ indicates rejection at the 10% level. All data are mean centered annual log-differences. The sample period covers July 1987 through June 2012, which has 300 months (100 quarters, 25 years). We use Newey and West's (1987) kernel-based HAC covariance estimator with Newey and West's (1994) automatic lag selection, and Gonçalves and Killian's (2004) bootstrapped p-value with $N = 999$ replications.

Panel A. Mixed Frequency

h	1	2	3	4	5
CPI$\nrightarrow$OIL	0.391	0.128	0.559	0.636	0.165
CPI$\nrightarrow$GDP	0.195	0.098°	$\boxed{0.049}^\circ$	0.100	0.180
OIL$\nrightarrow$GDP	0.680	0.548	0.236	0.300	0.196
OIL$\nrightarrow$CPI	$\boxed{0.002}^\circ$	0.182	0.439	$\boxed{0.029}^\circ$	0.605
GDP$\nrightarrow$CPI	$\boxed{0.015}^\circ$	0.570	0.583	0.125	0.500
GDP$\nrightarrow$OIL	0.724	0.833	0.895	0.855	0.946

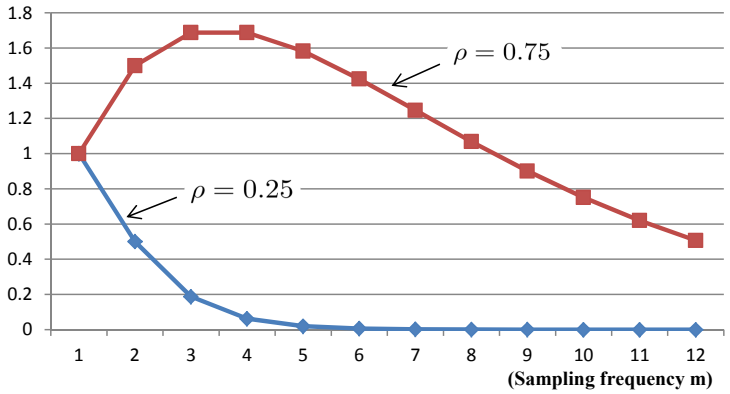
Panel B. Low Frequency

h	1	2	3	4	5
CPI$\nrightarrow$OIL	$\boxed{0.035}^\circ$	0.095°	0.095°	0.116	0.492
CPI$\nrightarrow$GDP	0.380	0.215	0.272	0.238	0.683
OIL$\nrightarrow$GDP	0.145	$\boxed{0.044}^\circ$	0.088°	$\boxed{0.027}^\circ$	0.066°
OIL$\nrightarrow$CPI	0.206	0.320	0.986	0.710	0.521
GDP$\nrightarrow$CPI	0.680	0.497	0.323	0.596	0.645
GDP$\nrightarrow$OIL	0.095°	0.164	0.516	0.376	0.541



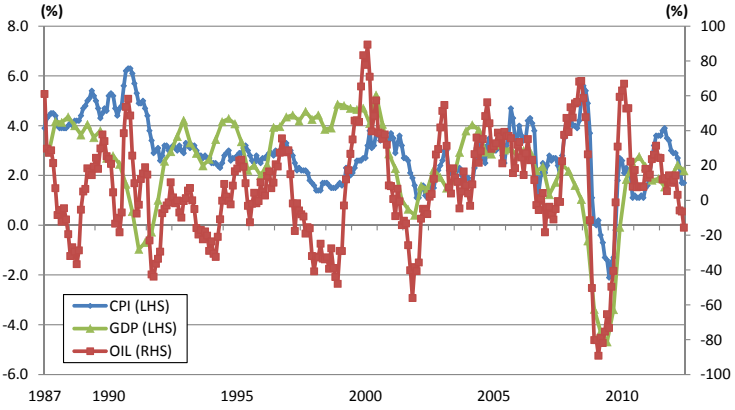
Note: The z-axis of each figure has the power ratio (i.e. the ratio of the local asymptotic power of the MF causality test to that of the low frequency causality test). Note that the scale of each z-axis is different. The x-axis has $\nu \in [0.5, 1.5]$, while the y-axis has $m \in \{3, \ldots, 12\}$.

Figure 2.1: Local Asymptotic Power of Mixed and Low Frequency Causality Tests



Note: The horizontal axis has $m \in \{1, \ldots, 12\}$, while the vertical axis has $m\rho^{m-1}$ for $\rho \in \{0.25, 0.75\}$.

Figure 2.2: Plot of the Function $m\rho^{m-1}$ - Driver of Local Asymptotic Power Ratios



Note: Year-to-year growth rates of U.S. monthly CPI, monthly spot West Texas Intermediate oil price, and quarterly real GDP. The sample period is July 1987 through June 2012, totalling 300 months (100 quarters, 25 years). The left-axis is for CPI and GDP and the right axis is for OIL.

Figure 2.3: Time Series Plot of CPI, OIL, and GDP

CHAPTER 3

REGRESSION-BASED TEST

3.1 Introduction

Time series are often sampled at different frequencies, and it is well known that temporal aggregation may hide or generate Granger causality. Existing Granger causality tests typically ignore this issue and they merely aggregate date to the common lowest frequency, which may result in spurious non-causality or spurious causality. See Zellner and Montmarquette (1971) and Amemiya and Wu (1972) for early contributions. This subject has been extensively researched ever since, e.g. Granger (1980), Granger (1988), Lütkepohl (1993), Granger (1995), Renault, Sekkat, and Szafarz (1998), Marcellino (1999), Breitung and Swanson (2002), McCrorie and Chambers (2006), Silvestrini and Veredas (2008), among others.

One of the most popular Granger causality tests is a Wald test based on multi-step ahead vector autoregression (VAR) models since this approach can handle *causal chains* among more than two variables. See Lütkepohl (1993), Dufour and Renault (1998), Dufour, Pelletier, and Renault (2006), and Hill (2007). This test suffers from the adverse effect of temporal aggregation since standard VAR models require to work on a single frequency. To alleviate this problem, Ghysels, Hill, and Motegi (2013) develop a set of Granger causality tests that explicitly take advantage of data sampled at mixed frequencies. They extend Dufour, Pelletier, and Renault's (2006) VAR-based causality test using Ghysels' (2012) mixed frequency vector autoregressive (MF-VAR) models. MF-VAR models avoid temporal aggregation by stacking all observations of high frequency variables.[1]

[1]MIDAS, standing for Mi(xed) Da(ta) S(ampling), regression models have been put forward in recent work by Ghysels, Santa-Clara, and Valkanov (2004), Ghysels, Santa-Clara, and Valkanov (2006), and Andreou, Ghysels, and Kourtellos (2010). See Andreou, Ghysels, and Kourtellos (2011) and Armesto, Engemann, and Owyang

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

Based on this motivation, the present paper proposes regression-based mixed frequency Granger causality tests. Our methodology is based on Sims' (1972) two-sided regression, not on the MF-VAR framework. We combine multiple parsimonious models where the i-th model regresses a low frequency variable x_L onto the i-th high frequency lag or lead of a high frequency variable x_H for $i \in \{1, \ldots, h\}$. Let $\hat{\beta}_i$ be an estimator for the loading of the i-th high frequency lag or lead, then our test statistic basically takes the maximum among $\{\hat{\beta}_1^2, \ldots, \hat{\beta}_h^2\}$. In this sense our test can be called the *max test* for short.

While the max test statistic follows a non-standard asymptotic distribution under the null hypothesis of Granger non-causality, a simulated p-value is readily available through simulation from the null distribution. Our test is thus very easy to implement in practice.

We will show that the max test is consistent under any type of Granger causality like decaying or lagged causality. In local power analysis, we show that our test is more powerful than the Wald test based on a naïve regression model. In small sample, we show via Monte Carlo simulations that our test produces no size distortions under realistic parameterizations and it is more powerful than the naïve Wald test.

As an empirical application, we conduct a rolling window analysis on weekly interest rate spread and real GDP growth in the U.S. We get an intuitive result that the yield spread used to be a strong predictor of GDP until 1980 or around, and its predictability has declined more recently. We also find that the mixed frequency approach that works on weekly spread achieves

---

(2010) for surveys. VAR models for mixed frequency data were independently introduced by Anderson, Deistler, Felsenstein, Funovits, Zadrozny, Eichler, Chen, and Zamani (2012), Ghysels (2012), and McCracken, Owyang, and Sekhposyan (2013). An early example of related ideas appears in Friedman (1962). Foroni, Ghysels, and Marcellino (2013) provide a survey of mixed frequency VAR models and related literature.

more frequent rejections of non-causality than the conventional low frequency approach that works on aggregated quarterly spread.

This paper is structured as follows. Section 3.2 derives the max test and proves its consistency formally. In Section 3.3 we conduct local power analysis to compare the local asymptotic power of the max test and the naïve Wald test. In Section 3.4 we run Monte Carlo simulations to compare the finite sample size and power of these two tests. Section 3.5 covers the empirical application on yield spread and GDP, while Section 3.6 concludes the paper. All tables and figures are collected after Section 3.6. Proofs for all theorems as well as some theoretical details are provided in Technical Appendices B.

3.2 Methodology

This paper focuses on a bivariate case where we have a high frequency variable x_H and a low frequency variable x_L. A trivariate case should await future research since it involves an extra complexity of causality chains (see Dufour and Renault (1998) and Dufour, Pelletier, and Renault (2006)).

For each low frequency time period $\tau_L \in \mathcal{Z}$, we have m high frequency time periods. We sequentially observe $\{x_H(\tau_L, 1), \ldots, x_H(\tau_L, m), x_L(\tau_L)\}$ in a period τ_L. A simple example would be a month vs. quarter case, where $m = 3$ since one quarter has three months. $x_H(\tau_L, 1)$ is the first monthly observation of x_H in quarter τ_L, $x_H(\tau_L, 2)$ is the second, and $x_H(\tau_L, 3)$ is the third. We then observe $x_L(\tau_L)$, the quarterly observation of x_L. The assumption that $x_L(\tau_L)$ is observed after $x_H(\tau_L, m)$ is just by convention.

The ratio of sampling frequencies, m, depends on τ_L in some applications like week vs. month, where m is four *or* five. This paper postpones such a case to the future work since time-dependent m complicates our statistical theory substantially.

Section 3.2.1 discusses high-to-low causality (i.e. causality from x_H to x_L), while Section 3.2.2 discusses low-to-high causality (i.e. causality from x_L to x_H).

3.2.1 High-to-Low Granger Causality

Assume that x_L depends on q low frequency autoregressive lags of x_L as well as p high frequency lags of x_H under the true DGP:

$$\begin{aligned} x_L(\tau_L) &= \sum_{k=1}^{q} a_k x_L(\tau_L - k) + \sum_{j=1}^{p} b_j x_H(\tau_L - 1, m + 1 - j) + \epsilon_L(\tau_L), \\ \epsilon_L(\tau_L) &\overset{i.i.d.}{\sim} (0, \sigma_L^2), \quad \sigma_L^2 > 0. \end{aligned} \tag{3.2.1}$$

Relaxing the i.i.d. assumption of ϵ_L to m.d.s. should be a future task. A constant term is omitted for algebraic simplicity, but could be added without any extra complexity.

To rewrite (3.2.1) in matrix form, define $\boldsymbol{X}_L(\tau_L - 1) = [x_L(\tau_L - 1), \ldots, x_L(\tau_L - q)]'$, $\boldsymbol{X}_H^{(p)}(\tau_L - 1) = [x_H(\tau_L - 1, m + 1 - 1), \ldots, x_H(\tau_L - 1, m + 1 - p)]'$, $\boldsymbol{a} = [a_1, \ldots, a_q]'$, and $\boldsymbol{b} = [b_1, \ldots, b_p]'$. Then, (3.2.1) can be rewritten as

$$x_L(\tau_L) = \boldsymbol{X}_L(\tau_L - 1)'\boldsymbol{a} + \boldsymbol{X}_H^{(p)}(\tau_L - 1)'\boldsymbol{b} + \epsilon_L(\tau_L). \tag{3.2.2}$$

When $p > m$, we need to clarify a notation used in (3.2.1) since the second argument of x_H is no longer positive for $j = m + 1, \ldots, p$. In such a case we take another lag in the first argument accordingly. For instance, $x_H(\tau_L - 1, 0)$ is understood as $x_H(\tau_L - 2, m)$ while $x_H(\tau_L - 1, -1)$ is understood as $x_H(\tau_L - 2, m - 1)$. More generally, we can interchangeably write $x_H(\tau_L - i, j) = x_H(\tau_L, j - im)$ for $j = 1, \ldots, m$ and $i \geq 0$. Complete details of these notational conventions are given in Appendix B.1.

We state assumptions on x_H below.

Assumption 3.2.1. Assume that $\{\{x_H(\tau_L, j)\}_{j=1}^m\}_{\tau_L \in \mathcal{Z}}$ follows a covariance stationary process with mean zero, variance $\gamma_0^H > 0$, autocovariance $\gamma_k^H = E[x_H(\tau_L, j)x_H(\tau_L, j - k)]$, and autocorrelation $\rho_k^H = \gamma_k^H / \gamma_0^H$ for $k \in \mathcal{Z}$. x_H is an exogenous variable in the sense that $E[x_H(\tau_L, j)\epsilon_L(\tau_L - k)] = 0$ for any $j, k, \tau_L \in \mathcal{Z}$.

Note that the order of the autocovariance γ_k^H is in terms of *high* frequency. Assumption 3.2.1 excludes Granger causality from x_L to x_H so that we can focus on causality from x_H to x_L. In the future work this assumption should be relaxed since we usually do not know whether

x_L causes x_H or not.

We also require the stability condition of x_L.

Assumption 3.2.2. All roots of $1 - \sum_{k=1}^{q} a_k z^k = 0$ lie outside the unit circle.

Treating non-stationary processes which violate Assumption 3.2.2 should await future research. Assumptions 3.2.1 and 3.2.2 guarantee that the DGP (3.2.1) can be transformed to an MA(∞) representation with infinite lags of x_H, which will be useful for proving subsequent theorems. In particular, $\{x_L(\tau_L)\}$ turns out to be a covariance stationarity process with mean zero, variance $\gamma_0^L > 0$, and autocovariance $\gamma_k^L = E[x_L(\tau_L)x_L(\tau_L - k)]$. Note that the order of the autocovariance γ_k^L is in terms of *low* frequency. $\{\gamma_k^L\}$ and the cross-covariance structure between x_H and x_L are characterized in terms of underlying parameters $\boldsymbol{a}$, $\boldsymbol{b}$, σ_L^2, and $\{\gamma_k^H\}$ in Appendix B.2. These characterizations are not required for proving subsequent theorems, but useful for understanding our methodology in general.

As in the past literature, we say x_H does not Granger cause x_L *if and only if* $\boldsymbol{b} = \mathbf{0}_{p\times 1}$. A naïve approach for testing non-causality is to postulate a regression model:

Naïve Regression Model

$$x_L(\tau_L) = \sum_{k=1}^{q} \alpha_k x_L(\tau_L - k) + \sum_{j=1}^{h} \beta_j x_H(\tau_L - 1, m + 1 - j) + u_L(\tau_L), \quad \tau_L = 1, \ldots, T_L. \quad (3.2.3)$$

Here we are assuming that the autoregressive lag order q is known since we are primarily interested in Granger causality from x_H to x_L. Consider fitting OLS to (3.2.3) and then testing $H_0 : \beta_1 = \cdots = \beta_h = 0$. This test clearly has power approaching one if $h \geq p$, but there will be size distortions when T_L is small and p is large (see Ghysels, Hill, and Motegi, 2013). The size distortions may be deleted after bootstrapping, but then finite sample power may be quite low. If in turn $h < p$, then there may be less size distortions but power may not approach one in the presence of Granger causality at lags beyond h.

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

$$x_L(\tau_L) = \sum_{k=1}^{q} \alpha_{k,j} x_L(\tau_L - k) + \beta_j x_H(\tau_L - 1, m + 1 - j) + u_{L,j}(\tau_L) \quad \text{for } j = 1, \ldots, h. \quad (3.2.4)$$

In a matrix form, model j is rewritten as

$$x_L(\tau_L) = \underbrace{\begin{bmatrix} \boldsymbol{X}_L(\tau_L - 1)' & x_H(\tau_L - 1, m + 1 - j) \end{bmatrix}}_{\equiv \boldsymbol{x}_j(\tau_L - 1)'} \underbrace{\begin{bmatrix} \alpha_{1,j} \\ \vdots \\ \alpha_{q,j} \\ \beta_j \end{bmatrix}}_{\equiv \boldsymbol{\theta}_j} + u_{L,j}(\tau_L). \quad (3.2.5)$$

Note that model j contains q low frequency autoregressive lags of x_L as well as the *only* j-th high frequency lag of x_H. Recall from (3.2.1) that q is in terms of low frequency and p is in terms of high frequency. Hence p tends to be larger than q in practice, especially when m is large. For example, when we handle a month vs. year case and thus $m = 12$, typical values for q and p would be 1 year and 12 months, respectively. Each parsimonious model (3.2.4) therefore has much fewer parameters than the naïve model (3.2.3). This feature alleviates size distortions in small sample with large p.

We describe how to combine all h parsimonious models to get a test statistic. First, consider fitting OLS for each model (3.2.4). In general, the resulting estimator will be biased due to omitted regressors. We thus need to characterize the pseudo-true value of $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_h]'$, denoted by $\boldsymbol{\beta}^* = [\beta_1^*, \ldots, \beta_h^*]'$, in terms of underlying parameters $\boldsymbol{a}$, $\boldsymbol{b}$, σ_L^2, and γ_k^H. Stack all parameters $\boldsymbol{\theta} = [\boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_h']'$ and let $\boldsymbol{\theta}^*$ be the pseudo-true value of $\boldsymbol{\theta}$. We construct a selection matrix $\boldsymbol{R}$ such that $\boldsymbol{\beta} = \boldsymbol{R}\boldsymbol{\theta}$. Specifically, $\boldsymbol{R}$ is an $h \times (q+1)h$ matrix whose $(j, (q+1)j)$ element is 1 for $j = 1, \ldots, h$ and all others are zeros.

Theorem 3.2.1. Let Assumptions 3.2.1 and 3.2.2 hold. Then, the pseudo-true value of $\boldsymbol{\beta}$

associated with OLS is given by

$$\begin{aligned}
&\boldsymbol{\beta}^* = \boldsymbol{R}\boldsymbol{\theta}^*, \quad \boldsymbol{\theta}^* \equiv \left[\boldsymbol{\theta}_1^{*\prime}, \ldots, \boldsymbol{\theta}_h^{*\prime}\right]', \\
&\boldsymbol{\theta}_j^* \equiv \begin{bmatrix} \alpha_{1,j}^* \\ \vdots \\ \alpha_{q,j}^* \\ \beta_j^* \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_q \\ 0 \end{bmatrix} + \underbrace{\left[E\left[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{x}_j(\tau_L - 1)'\right]\right]^{-1}}_{\equiv \boldsymbol{\Gamma}_{j,j}^{-1}} \underbrace{E\left[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{X}_H^{(p)}(\tau_L - 1)'\right]}_{\equiv \boldsymbol{C}_j} \boldsymbol{b}.
\end{aligned} \tag{3.2.6}$$

Proof 3.2.1. See Appendix B.3. Analytical expressions for $E[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{x}_j(\tau_L - 1)']$ and $E[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{X}_H^{(p)}(\tau_L - 1)']$ are not required for the proof, but they are derived in Appendix B.2 for completeness.

As shown in Appendix B.2, $[E[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{x}_j(\tau_L - 1)']]^{-1}$ and $E[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{X}_H^{(p)}(\tau_L - 1)']$ are highly nonlinear functions of $\boldsymbol{b}$ in general. Hence, $\boldsymbol{\beta}^*$ and $\boldsymbol{b}$ are related with each other in a complicated fashion generally. When $q = 0$ (i.e. the DGP has no autoregressive components), $\boldsymbol{\beta}^*$ becomes a simple linear function of $\boldsymbol{b}$. To see this, note that $\boldsymbol{R} = \boldsymbol{I}_h$ and $\boldsymbol{x}_j(\tau_L - 1) = x_H(\tau_L - 1, m + 1 - j)$ when $q = 0$. Then by Assumption 3.2.1 $E[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{x}_j(\tau_L - 1)'] = \gamma_0^H$ and $E[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{X}_H^{(p)}(\tau_L - 1)'] = [\gamma_{j-1}^H, \ldots, \gamma_{j-p}^H]$. Recalling the notation $\rho_k^H = \gamma_k^H / \gamma_0^H$, we have that

$$\underbrace{\begin{bmatrix} \beta_1^* \\ \vdots \\ \beta_h^* \end{bmatrix}}_{=\boldsymbol{\beta}^*} = \underbrace{\begin{bmatrix} \rho_{1-1}^H & \cdots & \rho_{1-p}^H \\ \vdots & \ddots & \vdots \\ \rho_{h-1}^H & \cdots & \rho_{h-p}^H \end{bmatrix}}_{\equiv \boldsymbol{R}_{h,p}^H} \underbrace{\begin{bmatrix} b_1 \\ \vdots \\ b_p \end{bmatrix}}_{=\boldsymbol{b}}. \tag{3.2.7}$$

Thus, there is a linear relationship between $\boldsymbol{\beta}^*$ and $\boldsymbol{b}$ when $q = 0$.

While (3.2.7) immediately implies that $\boldsymbol{b} = \boldsymbol{0}_{p\times 1} \Rightarrow \boldsymbol{\beta}^* = \boldsymbol{0}_{h\times 1}$, it also gives us a useful insight for identifying $\boldsymbol{b}$ via $\boldsymbol{\beta}^*$. Since $\boldsymbol{R}_{h,p}^H$ is of full rank under Assumption 3.2.1, we have the following relationship.

1. If $p > h$, then $\boldsymbol{b}$ is not identified by $\boldsymbol{\beta}^*$. In particular, $\boldsymbol{b}$ may not be a null vector even if $\boldsymbol{\beta}^*$ is.

2. If $p = h$, then $\boldsymbol{b}$ is exactly identified by the formula $\boldsymbol{b} = (\boldsymbol{R}_{h,h}^H)^{-1}\boldsymbol{\beta}^*$. In particular, $\boldsymbol{b}$ is a

null vector whenever $\boldsymbol{\beta}^*$ is.

3. If $p < h$, then $\boldsymbol{b}$ is over-identified by $\boldsymbol{\beta}^*$. In particular, $\boldsymbol{b}$ is a null vector whenever $\boldsymbol{\beta}^*$ is. This result follows from the positive definiteness of $(\boldsymbol{R}_{h,p}^H)'\boldsymbol{R}_{h,p}^H$ and the condition $\boldsymbol{\beta}^{*'}\boldsymbol{\beta}^* = 0$.

Generalizing this insight to an arbitrary autoregressive lag order q, we can establish the following theorem.

Theorem 3.2.2. Let Assumptions 3.2.1 and 3.2.2 hold. We have that $\boldsymbol{b} = \mathbf{0}_{p\times 1} \Rightarrow \boldsymbol{\beta}^* = \mathbf{0}_{h\times 1}$, regardless of p and h. When $h \geq p$, the converse is also true: $\boldsymbol{\beta}^* = \mathbf{0}_{h\times 1} \Rightarrow \boldsymbol{b} = \mathbf{0}_{p\times 1}$.

Proof 3.2.2. See Appendix B.4.

We are interested in the null hypothesis of Granger non-causality, $H_0 : \boldsymbol{b} = \mathbf{0}_{p\times 1}$, and the first part of Theorem 3.2.2 implies that $\boldsymbol{\beta}^* = \mathbf{0}_{h\times 1}$ under H_0. Moreover, the second part of Theorem 3.2.2 essentially states that $\boldsymbol{\beta}^* \neq \mathbf{0}_{h\times 1}$ under a general alternative hypothesis $H_1 : \boldsymbol{b} \neq \mathbf{0}_{p\times 1}$, given $h \geq p$. Exploiting these properties, we construct a test statistic for H_0. For each parsimonious model (3.2.4) we get $\hat{\beta}_j$, the OLS estimator for β_j. Define $\hat{\boldsymbol{\beta}} = [\hat{\beta}_1, \ldots, \hat{\beta}_h]'$. The basic idea of our test, inspired by Andrews and Ploberger (1994), is to look at the maximum value among $\{\hat{\beta}_1^2, \ldots, \hat{\beta}_h^2\}$ with a certain weighting scheme.

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

$$\boldsymbol{W}_{T_L} = \begin{bmatrix} w_{T_L,1} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & w_{T_L,h} \end{bmatrix} \quad \text{and} \quad \boldsymbol{W} = \begin{bmatrix} w_1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & w_h \end{bmatrix}. \tag{3.2.8}$$

A trivial choice of $w_{T_L,j}$ is w_j, a non-random constant, but we can consider any other weighting structure as well.

We propose a test statistic:

Max Test Statistic for High-to-Low Causality

$$\mathcal{T} = \max_{1 \leq j \leq h} \left(\sqrt{T_L} w_{T_L,j} \hat{\beta}_j \right)^2. \tag{3.2.9}$$

We call this the *max test statistic* since it takes the maximum of the square of properly scaled individual OLS estimators. Theorem 3.2.3, one of our main results, derives the asymptotic distribution of $\mathcal{T}$ under H_0 and proves that our test is consistent given $h \geq p$.

Theorem 3.2.3. Let Assumptions 3.2.1 and 3.2.2 hold. (i) Under $H_0 : \boldsymbol{b} = \mathbf{0}_{p\times 1}$, we have that $\mathcal{T} \xrightarrow{d} \max_{1\leq j\leq h} \mathcal{N}_j^2$ as $T_L \to \infty$. $\boldsymbol{\mathcal{N}} = [\mathcal{N}_1, \ldots, \mathcal{N}_h]'$ is a vector-valued random variable drawn from $N(\mathbf{0}_{h\times 1}, \boldsymbol{V})$, where

$$\boldsymbol{V} = \frac{\sigma_L^2}{\gamma_0^H} \boldsymbol{W} \boldsymbol{R}_{h,h}^H \boldsymbol{W} \quad \text{with} \quad \boldsymbol{R}_{h,h}^H = \begin{bmatrix} \rho_{1-1}^H & \cdots & \rho_{1-h}^H \\ \vdots & \ddots & \vdots \\ \rho_{h-1}^H & \cdots & \rho_{h-h}^H \end{bmatrix}. \tag{3.2.10}$$

(ii) Given $h \geq p$, $\mathcal{T} \xrightarrow{p} \infty$ under a general alternative hypothesis $H_1 : \boldsymbol{b} \neq \mathbf{0}_{p\times 1}$.

Proof 3.2.3. See Appendix B.5.

Although a formal proof is provided in Appendix B.5, Theorem 3.2.3.(ii) is intuitively clear in view of Theorem 3.2.2. Equation (3.2.9) indicates that $\mathcal{T} \xrightarrow{p} \infty$ *if and only if* $\boldsymbol{\beta}^* \neq \mathbf{0}_{h\times 1}$. Theorem 3.2.2 states that, as long as $h \geq p$, nonzero $\boldsymbol{b}$ implies nonzero $\boldsymbol{\beta}^*$. Our test is therefore consistent.

If one happens to choose h that is smaller than p, there may be a certain form of causality such that the power does not approach one. To see this point, we consider a very simple example where $q = 0$, $p = 2$, and $h = 1$. Then (3.2.7) implies that $\beta_1^* = b_1 + \rho_1^H b_2$. If $b_1 = 1$ and $b_2 = -1/\rho_1^H$ for example, then $\beta_1^* = 0$. As a result $\mathcal{T} = 0$ and thus there is in fact *no* power. This example may be unlikely to occur in most economic applications since it requires $|b_1| = |\rho_1^H b_2| < |b_2|$ (i.e. the first high frequency lag of x_H should have a *smaller* impact on x_L than the second high frequency lag of x_H does). But some applications may have such a tricky Granger causality due to lagged information transmission or seasonal effects. It is thus advised to take a sufficiently large h when the possibility of lagged causality cannot be ruled out.

Another important feature of Theorem 3.2.3 is that the asymptotic covariance matrix $\boldsymbol{V}$

does not depend on $\boldsymbol{a}$ at all. This property is essentially due to the assumption that our model includes exactly q low frequency lags of x_L.

Furthermore, $\boldsymbol{V}$ does not depend on m either. This is a natural result since we are not aggregating x_H.

Even though the test statistic $\mathcal{T}$ has a non-standard limit distribution under H_0, a simulated p-value is easily available via simulation from the null distribution.

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

$$\hat{p} = (1/R)\sum_{r=1}^{R} I\left(\mathcal{T}_r > \mathcal{T}\right). \tag{3.2.11}$$

It turns out that we can compute a consistent estimator for $\boldsymbol{V}$ under H_0 although we cannot in general. Recall (3.2.10) to see this point. First, $\boldsymbol{W}_{T_L} \xrightarrow{p} \boldsymbol{W}$ by assumption. Second, γ_k^H can be consistently estimated by the sample autocovariance of x_H of order k: $\hat{\gamma}_k^H = (1/mT_L)\sum_{\tau_L=1}^{T_L}\sum_{j=1}^{m} x_H(\tau_L, j)x_H(\tau_L, j-k) \xrightarrow{p} \gamma_k^H$. Similarly, $\hat{\rho}_k^H = \hat{\gamma}_k^H/\hat{\gamma}_0^H \xrightarrow{p} \rho_k^H$. Hence, the availability of consistent $\hat{\boldsymbol{V}}$ depends entirely on the availability of consistent $\hat{\sigma}_L^2$. Since the DGP (3.2.1) reduces to a pure AR(q) process under H_0, $\hat{\sigma}_L^2$ can be calculated by fitting an AR(q) model for x_L and computing the sample variance of residuals. If we do not impose H_0 then we cannot get consistent $\hat{\sigma}_L^2$ due to the misspecification of each parsimonious model, but all we need to implement our test is a consistent estimator for $\boldsymbol{V}$ *under* H_0. Therefore, we can implement statistical inference using the asymptotic p-value approximation in (3.2.11).

While $\boldsymbol{V}$ itself does not depend on m as explained above, $\hat{\boldsymbol{V}}$ does depend on it through $\hat{\gamma}_k^H$. In fact, the precision of $\hat{\gamma}_k^H$ *improves* as m grows since the high frequency sample size mT_L gets larger. In that sense having a large m is a *favorable* situation for the max test, while it is a challenging situation for the existing mixed frequency Granger causality test (see Ghysels, Hill, and Motegi (2013)).

3.2.2 Low-to-High Granger Causality

Consider a high frequency variable x_H and a low frequency variable x_L with the ratio of sampling frequencies m. Define the mixed frequency vector

$$\boldsymbol{X}(\tau_L) = [x_H(\tau_L, 1), \ldots, x_H(\tau_L, m), x_L(\tau_L)]' \in \mathcal{R}^K$$

with $K = m + 1$. Assume that x_H and x_L follow MF-VAR(q):

$$\boldsymbol{X}(\tau_L) = \sum_{i=1}^{q} \boldsymbol{A}_i \boldsymbol{X}(\tau_L - i) + \boldsymbol{\epsilon}(\tau_L), \quad \boldsymbol{\epsilon}(\tau_L) \overset{i.i.d.}{\sim} (\mathbf{0}_{K\times 1}, \boldsymbol{\Omega}), \tag{3.2.12}$$

where

$$\boldsymbol{A}_i = \begin{bmatrix} a_{11,i} & \ldots & a_{1K,i} \\ \vdots & \ddots & \vdots \\ a_{K1,i} & \ldots & a_{KK,i} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\epsilon}(\tau_L) = \begin{bmatrix} \epsilon_H(\tau_L, 1) \\ \vdots \\ \epsilon_H(\tau_L, m) \\ \epsilon_L(\tau_L) \end{bmatrix}.$$

Relaxing the i.i.d. assumption of $\boldsymbol{\epsilon}$ to m.d.s. should be a future task.

Focusing on the last row of (3.2.12), we have that

$$x_L(\tau_L) = \sum_{k=1}^{q} a_{KK,k} x_L(\tau_L - k) + \sum_{k=1}^{q} \sum_{l=1}^{m} a_{Kl,k} x_H(\tau_L - k, l) + \epsilon_L(\tau_L), \quad \epsilon_L(\tau_L) \overset{i.i.d.}{\sim} (0, \sigma_L^2). \tag{3.2.13}$$

To test Granger causality from x_L to x_H (i.e. low-to-high causality), we naïvely consider Sims' two-sided regression model.

Naïve Regression Model

$$x_L(\tau_L) = \sum_{k=1}^{q} \alpha_{k,j} x_L(\tau_L - k) + \sum_{k=1}^{mq} \beta_{k,j} x_H(\tau_L - 1, m + 1 - k) + \sum_{j=1}^{h} \gamma_j x_H(\tau_L + 1, j) + u_{L,j}(\tau_L), \tag{3.2.14}$$

Instruments: $\{\text{all } q + mq + h \text{ regressors}, x_H(\tau_L, 1), \ldots, x_H(\tau_L, m)\}$.

Here we are assuming that the true MF-VAR lag order q is known. Besides all explanatory variables, we include m contemporaneous high frequency observations of x_H in the set of instruments in order to handle simultaneity between x_L and x_H. We test the significance of

$\gamma_1, \ldots, \gamma_h$, the parameters on the high frequency *leads* of x_H. Under the null hypothesis of low-to-high non-causality, all those parameters should be equal to zero.

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

Parsimonious Regression Model $j \in \{1, \ldots, h\}$

$$x_L(\tau_L) = \sum_{k=1}^{q} \alpha_{k,j} x_L(\tau_L - k) + \sum_{k=1}^{mq} \beta_{k,j} x_H(\tau_L - 1, m + 1 - k) + \gamma_j x_H(\tau_L + 1, j) + u_{L,j}(\tau_L), \tag{3.2.15}$$

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

We are combining h parsimonious regression models, and the j-th model contains the j-th high frequency lead of x_H. As in the naïve regression model (3.2.14), we include m contemporaneous high frequency observations of x_H as instruments.

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

$$\mathcal{T} = \max_{1 \leq j \leq h} \left(\sqrt{T_L} w_{T_L,j} \hat{\gamma}_j \right)^2, \tag{3.2.16}$$

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

Theorem 3.2.4. Under $H_0 : x_L \nrightarrow x_H$, it follows that $\mathcal{T} \equiv \max_{1 \leq j \leq h} (\sqrt{T_L} w_{T_L,j} \hat{\gamma}_j)^2 \xrightarrow{d} \max_{1 \leq j \leq h} \mathcal{N}_j^2$, where $\boldsymbol{\mathcal{N}} = [\mathcal{N}_1, \ldots, \mathcal{N}_h]' \sim N(\boldsymbol{0}_{h \times 1}, \boldsymbol{U})$.

Proof 3.2.4. See Appendix B.6. The covariance matrix $\boldsymbol{U}$ is derived there.

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

3.3 Local Asymptotic Power Analysis

Using a local asymptotic power framework, this section examines the asymptotic performance of the high-to-low max test.[2] Our analysis has three goals. First, we compare the local power of the max test and the local power of the Wald test based on the naïve regression model (3.2.3). It will turn out that the max test has clearly higher power than the Wald test. The difference between these two can be as large as 15-20% for some parametrizations.

Second, we investigate how the the local power of the max test evolves over the true lag order p and the number of lags considered, h. We will find that local power is maximized when h is close to p.

Third, we compare the power of the mixed frequency max test and its low frequency counterpart which works on an aggregated x_H instead of the original high frequency series. It will turn out that the mixed frequency approach can capture finer causal patterns appearing within each low frequency period.

We keep imposing Assumptions 3.2.1 and 3.2.2, and consider the same DGP (3.2.1) again. Our null hypothesis is the same as before (i.e. $H_0 : \boldsymbol{b} = \mathbf{0}_{p\times 1}$), but here we consider a local alternative hypothesis $H_1^l : \boldsymbol{b} = (1/\sqrt{T_L})\boldsymbol{\nu}$. In the literature $\boldsymbol{\nu} = [\nu_1, \ldots, \nu_p]'$ is often called the Pitman drift. Under H_1^l, the DGP is written as

$$\begin{aligned} x_L(\tau_L) &= \sum_{k=1}^{q} a_k x_L(\tau_L - k) + \sum_{j=1}^{p} \frac{\nu_j}{\sqrt{T_L}} x_H(\tau_L - 1, m + 1 - j) + \epsilon_L(\tau_L) \\ &= \boldsymbol{X}_L(\tau_L - 1)'\boldsymbol{a} + \boldsymbol{X}_H^{(p)}(\tau_L - 1)' \left(\frac{1}{\sqrt{T_L}}\boldsymbol{\nu}\right) + \epsilon_L(\tau_L), \quad \epsilon_L(\tau_L) \overset{i.i.d.}{\sim} (0, \sigma_L^2). \end{aligned} \tag{3.3.1}$$

We combine h parsimonious regression models (3.2.4) to formulate the test statistic $\mathcal{T}$ in (3.2.9). The asymptotic distribution under H_0 is already derived in Theorem 3.2.3. Here we derive the asymptotic distribution under H_1^l.

Theorem 3.3.1. Let Assumptions 3.2.1 and 3.2.2 hold. Then, we have that $\mathcal{T} \xrightarrow{d} \max_{1\le i\le h} \mathcal{M}_i^2$ as $T_L \to \infty$ under $H_1^l : \boldsymbol{b} = (1/\sqrt{T_L})\boldsymbol{\nu}$. $\boldsymbol{\mathcal{M}} = [\mathcal{M}_1, \ldots, \mathcal{M}_h]'$ is a vector-valued random variable

[2] The low-to-high case remains as a future task.

drawn from $N(\boldsymbol{\mu}, \boldsymbol{V})$, where $\boldsymbol{V}$ is defined in (3.2.10) and

$$\boldsymbol{\mu} = \boldsymbol{W}\boldsymbol{R}_{h,p}^{H}\boldsymbol{\nu} \quad \text{with} \quad \boldsymbol{R}_{h,p}^{H} = \begin{bmatrix} \rho_{1-1}^{H} & \cdots & \rho_{1-p}^{H} \\ \vdots & \ddots & \vdots \\ \rho_{h-1}^{H} & \cdots & \rho_{h-p}^{H} \end{bmatrix}. \tag{3.3.2}$$

Proof 3.3.1. See Appendix B.7.

In Theorem 3.2.3 we have shown that the asymptotic covariance matrix $\boldsymbol{V}$ does not depend on $\boldsymbol{a}$. In (3.3.2) we see that the noncentrality parameter $\boldsymbol{\mu}$ does not depend on $\boldsymbol{a}$ either. Thus, the autoregressive component of x_L does not affect local asymptotic power at all. This result comes from the assumption that the autoregressive lag order q is known and therefore each parsimonious model contains exactly q lags of x_L.

In addition, neither $\boldsymbol{\mu}$ nor $\boldsymbol{V}$ depends on m. This is an intuitive result since we are not aggregating x_H.

In the local power analysis we know all of underlying parameters, so $\boldsymbol{V}$ and $\boldsymbol{\mu}$ can be calculated easily from (3.2.10) and (3.3.2). Then Theorems 3.2.3 and 3.3.1 allow us to compute local asymptotic power for any desired (h, p) numerically. The procedure is as follows.

Step 1 Draw R_1 samples $\boldsymbol{\mathcal{N}}^{(1)}, \ldots, \boldsymbol{\mathcal{N}}^{(R_1)}$ independently from the limit distribution under H_0, $N(\mathbf{0}_{h\times 1}, \boldsymbol{V})$, and calculate a set of test statistics $\mathcal{T}_r = \max_{1\le i\le h}\left(\boldsymbol{\mathcal{N}}_i^{(r)}\right)^2$.

Step 2 Sort the test statistics $\mathcal{T}_{(1)} \le \cdots \le \mathcal{T}_{(R_1)}$ and take the $100(1-\alpha)\%$ quantile, which is a numerical approximation of the critical value associated with a nominal size α. Call that quantile d^*.

Step 3 Draw R_2 samples $\boldsymbol{\mathcal{M}}^{(1)}, \ldots, \boldsymbol{\mathcal{M}}^{(R_2)}$ independently from the limit distribution under H_1^l, $N(\boldsymbol{\mu}, \boldsymbol{V})$, and calculate another set of test statistics $\tilde{\mathcal{T}}_r = \max_{1\le i\le h}\left(\boldsymbol{\mathcal{M}}_i^{(r)}\right)^2$. Local asymptotic power $\mathcal{P}$ is given by $\mathcal{P} = (1/R_2)\sum_{r=1}^{R_2} I(\tilde{\mathcal{T}}_r > d^*)$.

Naïve Regression Model It is of interest to compare the local power of the max test and the local power of the Wald test based on the naive regression model (3.2.3):

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

Under $H_0 : \boldsymbol{b} = \mathbf{0}_{p\times 1}$, this model always includes the DGP as a special case. Under $H_1^l : \boldsymbol{b} = (1/\sqrt{T_L})\boldsymbol{\nu}$, it does not include the DGP when $h < p$. Based on the standard statistical argument, it is straightforward to derive the asymptotic distribution of the Wald statistic W with respect to H_0.

Theorem 3.3.2. Let Assumptions 3.2.1 and 3.2.2 hold. Let W be the Wald statistic with respect to $H_0 : \boldsymbol{b} = \mathbf{0}_{p\times 1}$. Then, the asymptotic distribution of W is χ_h^2 under H_0 and $\chi_h^2(\kappa)$ under H_1^l. $\chi_h^2(\kappa)$ is the noncentral chi-squared distribution with degrees of freedom h and noncentrality $\kappa = (\gamma_0^H/\sigma_L^2)\boldsymbol{\nu}'\boldsymbol{R}_{h,p}'\boldsymbol{R}_{h,h}^{-1}\boldsymbol{R}_{h,p}\boldsymbol{\nu}$.

Proof 3.3.2. This theorem can be proven by the classic argument of the Wald test and local asymptotic power literature.

As seen in Theorem 3.3.2, W has a convenient asymptotic distribution both under H_0 and H_1^l. Hence local power can be calculated by definition: $\mathcal{P} = 1 - F_1[F_0^{-1}(1-\alpha)]$, where F_0 is the asymptotic distribution under H_0 (i.e. χ_h^2) and F_1 is the asymptotic distribution under H_1^l (i.e. $\chi_h^2(\kappa)$).

Note that the autoregressive parameters $a_1, \ldots, a_q$ are not playing any role in Theorem 3.3.2. This result stems from out assumption of known q. The ratio of sampling frequency m does not play any role either.

Low Frequency Counterpart Another interesting exercise is to compare the mixed frequency max test and a conventional low frequency max test in terms of local power. The former works on the original high frequency series $\{\{x_H(\tau_L, j)\}\}$, while the latter works on its aggregated version $\{x_H(\tau_L)\}$. We consider linear aggregation scheme $x_H(\tau_L) = \sum_{j=1}^{m} \delta_j x_H(\tau_L, j)$ with $\delta_j \geq 0$ for all $j = 1, \ldots, m$ and $\sum_{j=1}^{m} \delta_j = 1$. The linear aggregation scheme is sufficiently

general for most economic applications since it includes flow sampling (i.e. $\delta_j = 1/m$ for $j = 1, \ldots, m$) and stock sampling (i.e. $\delta_j = I(j = m)$ for $j = 1, \ldots, m$) as special cases. Note that δ_j is not a parameter to estimate; it is fixed by the researcher.

Given Assumption 3.2.1, we can show that $\{x_H(\tau_L)\}$ is a covariance stationary process with mean zero and autocovariance

$$\begin{aligned}\gamma_k^{H,LF} &\equiv E[x_H(\tau_L)x_H(\tau_L - k)] = E\left[\left(\sum_{i=1}^{m}\delta_i x_H(\tau_L, i)\right)\left(\sum_{j=1}^{m}\delta_j x_H(\tau_L - k, j)\right)\right] \\ &= \sum_{i=1}^{m}\sum_{j=1}^{m}\delta_i\delta_j E[x_H(\tau_L, j)x_H(\tau_L - k, j)] = \sum_{i=1}^{m}\sum_{j=1}^{m}\delta_i\delta_j\gamma_{j-i-km}^{H}, \quad \text{for } k \in \mathcal{Z}.\end{aligned} \tag{3.3.3}$$

See Appendix B.2 for a more formal derivation.

Keeping our DGP the same, we combine the following h parsimonious models as a conventional low frequency approach:

$$x_L(\tau_L) = \sum_{k=1}^{q}\alpha_{k,j}^{LF}x_L(\tau_L - k) + \beta_j^{LF}x_H(\tau_L - j) + u_{L,j}^{LF}(\tau_L), \quad j = 1, \ldots, h \tag{3.3.4}$$

or in a matrix form

$$x_L(\tau_L) = \boldsymbol{x}_j^{LF}(\tau_L - 1)'\boldsymbol{\theta}_j^{LF} + u_{L,j}^{LF}(\tau_L)$$

with $\boldsymbol{x}_j^{LF}(\tau_L - 1) = [\boldsymbol{X}_L(\tau_L - 1)',\ x_H(\tau_L - j)]'$ and $\boldsymbol{\theta}_j^{LF} = [\alpha_{1,j}^{LF}, \ldots, \alpha_{q,j}^{LF}, \beta_j^{LF}]'$. We can use different h's between the mixed frequency test and the low frequency test, but the same notation is used here for brevity. The only difference between the mixed frequency model (3.2.4) and the low frequency model (3.3.4) is whether we include the j-th lag of the *original* x_H or the j-th lag of the *aggregated* x_H. As a result, the former involves h *high* frequency lags in terms of x_H while the latter involves h *low* frequency lags. This suggests that the former would perform better than the latter when there is a certain form of causality *within* a low frequency time period. One specific example would be a month vs. year case where the only one-month lag of x_H has a nonzero coefficient $b_1 \neq 0$.

For illustration, we present model (3.3.4) under stock sampling:

$$x_L(\tau_L) = \sum_{k=1}^{q}\alpha_{k,j}^{LF}x_L(\tau_L - k) + \beta_j^{LF}x_H(\tau_L - j, m) + u_{L,j}^{LF}(\tau_L) \tag{3.3.5}$$

and under flow sampling:

$$x_L(\tau_L) = \sum_{k=1}^{q} \alpha_{k,j}^{LF} x_L(\tau_L - k) + \beta_j^{LF} \left(\frac{1}{m} \sum_{i=1}^{m} x_H(\tau_L - j, i) \right) + u_{L,j}^{LF}(\tau_L).$$

We run OLS for (3.3.4) to get $\hat{\beta}_j^{LF}$ and then formulate a test statistic:

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

We can use different weighting schemes between the mixed frequency test and the low frequency test, but the same notation is used here for brevity.

The following theorem derives the limit distributions of $\mathcal{T}_{LF}$ under H_0 and H_1^l.

Theorem 3.3.3. Let Assumptions 3.2.1 and 3.2.2 hold. Then, (i) we have that $\mathcal{T}_{LF} \xrightarrow{d} \max_{1 \le j \le h} (\mathcal{N}_j^{LF})^2$ as $T_L \to \infty$ under $H_0 : \boldsymbol{b} = \mathbf{0}_{p \times 1}$. $\boldsymbol{\mathcal{N}}^{LF} = [\mathcal{N}_1^{LF}, \ldots, \mathcal{N}_h^{LF}]'$ is a vector-valued random variable following $N(\mathbf{0}_{h \times 1}, \boldsymbol{V}^{LF})$, where

$$\boldsymbol{V}^{LF} = \frac{\sigma_L^2}{\gamma_0^{H,LF}} \boldsymbol{W} \boldsymbol{R}^{H,LF} \boldsymbol{W} \quad \text{with} \quad \boldsymbol{R}^{H,LF} = \frac{1}{\gamma_0^{H,LF}} \begin{bmatrix} \gamma_{1-1}^{H,LF} & \cdots & \gamma_{1-h}^{H,LF} \\ \vdots & \ddots & \vdots \\ \gamma_{h-1}^{H,LF} & \cdots & \gamma_{h-h}^{H,LF} \end{bmatrix}. \tag{3.3.6}$$

(ii) We have that $\mathcal{T}_{LF} \xrightarrow{d} \max_{1 \le j \le h} (\mathcal{M}_j^{LF})^2$ under $H_1^l : \boldsymbol{b} = (1/\sqrt{T_L})\boldsymbol{\nu}$. $\boldsymbol{\mathcal{M}}^{LF} = [\mathcal{M}_1^{LF}, \ldots, \mathcal{M}_h^{LF}]'$ is a vector-valued random variable following $N(\boldsymbol{\mu}^{LF}, \boldsymbol{V}^{LF})$, where

$$\boldsymbol{\mu}^{LF} = \boldsymbol{W} \boldsymbol{\Delta} \boldsymbol{\nu} \quad \text{with} \quad \boldsymbol{\Delta} = \frac{1}{\gamma_0^{H,LF}} \begin{bmatrix} \sum_{i=1}^{m} \delta_i \gamma_{m+1-1-i+(1-1)m}^{H} & \cdots & \sum_{i=1}^{m} \delta_i \gamma_{m+1-p-i+(1-1)m}^{H} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{m} \delta_i \gamma_{m+1-1-i+(h-1)m}^{H} & \cdots & \sum_{i=1}^{m} \delta_i \gamma_{m+1-p-i+(h-1)m}^{H} \end{bmatrix}. \tag{3.3.7}$$

Proof 3.3.3. See Appendix B.8.

The two key quantities determining the local asymptotic power of the low frequency test, $\boldsymbol{\mu}^{LF}$ and $\boldsymbol{V}^{LF}$, do not depend on $\boldsymbol{a}$. This is essentially because we are assuming that the true autoregressive lag order q is known.

Recall from (3.3.3) that $\gamma_k^{H,LF}$ depends on m, so both $\boldsymbol{\mu}^{LF}$ and $\boldsymbol{V}^{LF}$ depend on m. This

result comes from the temporal aggregation of x_H.

Local asymptotic power based on the low frequency test can be computed in the same manner as the mixed frequency test. Follow Steps 1-3 right after Theorem 3.3.1.

Numerical Examples Here we evaluate the local asymptotic power of each test described above under some realistic parameterizations. For the true DGP (3.3.1), we try each of high frequency lag length $p \in \{1, \ldots, 5\}$ and consider two causality patterns:

1. Decaying Causality: $\nu_j = 0.8 - 0.1(j-1)$ for $j = 1, \ldots, p$. This is a commonly observed causal pattern where the coefficients decay gradually. For example, $\boldsymbol{\nu} = [0.8, 0.7, 0.6, 0.5, 0.4]'$ when $p = 5$.

2. Lagged Causality: $\nu_j = 2 \times I(j = p)$ for $j = 1, \ldots, p$. Only ν_p is 2 and all other coefficients are zeros. This case corresponds to seasonality or lagged response of x_L to x_H.

We assume that x_H follows a covariance stationary AR(1) process:

$$x_H(\tau_L, j) = \phi x_H(\tau_L, j-1) + \epsilon_H(\tau_L, j), \quad \epsilon_H(\tau_L, j) \overset{i.i.d.}{\sim} (0, 1),$$

in which case $\gamma_k^H = \phi^{|k|}/(1-\phi^2)$ and hence $\rho_k^H = \phi^{|k|}$ for $k \in \mathcal{Z}$. We try $\phi \in \{0.2, 0.8\}$.

We do not have to specify the autoregressive lag order q or those coefficients $a_1, \ldots, a_q$ since they will not affect the local asymptotic power of any tests discussed previously.

For the mixed frequency max test we try $h \in \{1, \ldots, 5\}$ and use equal weights: $w_j = 1/h$ for $j = 1, \ldots, h$. Given p, we expect that the local power increases as h approaches p from below.

For the Wald test based on the naïve model (3.2.3), we again try $h \in \{1, \ldots, 5\}$. It is of interest to see which of the mixed frequency max test and Wald test gets higher power for given $\{p, h\}$.

For the low frequency max test we assume $m = 12$, which can be thought of as a month vs. year case or approximately a week vs. quarter case. We try both stock sampling and flow sampling. The number of models combined is picked from $\underline{h} \in \{1, \ldots, 3\}$. We are explicitly distinguishing h and $\underline{h}$ since each lag in the mixed frequency models (3.2.4) is in terms of *high frequency* while each lag in the low frequency models (3.3.4) is in terms of *low frequency*. We use equal weights: $\underline{w}_j = 1/\underline{h}$ for $j = 1, \ldots, \underline{h}$.

Other quantities are set as follows. The error variance in the DGP is $\sigma_L^2 = 1$. The nominal size is $\alpha = 0.05$. For the max tests, the number of draws from the limit distributions is $R_0 = R_1 = 100,000$.

Table 3.1 compares the local asymptotic power of the mixed frequency max test and the Wald test. Panel A considers Decaying Causality, while Panel B considers Lagged Causality. Panels A.1 and B.1 consider $\phi = 0.2$, a relatively transitory x_H. Panels A.2 and B.2 consider $\phi = 0.8$, a relatively persistent x_H.

Since we have two causality patterns, two values for ϕ, five values for h, and five values for p, there are $2 \times 2 \times 5 \times 5 = 100$ ways to compare the max test and the Wald test in total. Remarkably, the max test has higher power than the Wald test in 92 out of the 100 slots. The difference increases in h since the Wald test gets more parameters to estimate at once. For example, the power of the max test is 57.4% while that of the Wald test is 40.7% in Panel A.2 with $(h, p) = (5, 2)$.

The two tests have the same power in six slots. The max test has lower power than the Wald test in only two slots, and the difference is as small as 0.1%; see Panel A.1 with $(h, p) = (4, 5)$ and Panel B.2 with $(h, p) = (1, 2)$.

Thus, we can conclude that the mixed frequency max test almost always gets higher power than the Wald test due to its parsimonious model specification. The difference in power increases can be as large as 15-20% when h is large, since the naïve regression model starts suffering from parameter proliferation.

We now investigate how the power of the mixed frequency max test evolves over h and p. We also compare the mixed frequency max test and its low frequency counterpart. See Table 3.2 for the results. Panel A is on the mixed frequency test, Panel B is on the low frequency test with stock sampling, and Panel C is on the low frequency test with flow sampling. Panel A.1 considers Decaying Causality, while Panel A.2 considers Lagged Causality. For each type of causality we try transitory x_H ($\phi = 0.2$) and persistent x_H ($\phi = 0.8$). The same structure applies for Panels B and C.

We first focus on Panel A, the mixed frequency case. The choice of h does not seem so important for Decaying Causality. For example, fixing $\phi = 0.2$ and $p = 3$, the local power is

16.9%, 20.1%, 19.7%, 17.5%, and 16.2% as h goes from 1 to 5. While these five values suggest that choosing too small h or too large h decreases local power, the difference does not look substantial.

The choice of h does look crucial for Lagged Causality, however. Fixing $\phi = 0.2$ and $p = 3$ again, the local power is 5.2%, 6.4%, 39.4%, 36.2%, and 33.7% as h goes from 1 to 5. These values indicate that h needs to be at least as large as p in order to achieve high power. Since x_H has small autocorrelation when $\phi = 0.2$, just including the first through $(p-1)$-th lags of x_H does not approximate the p-th, significant lag. When $\phi = 0.8$, the negative effect of including too few lags gets less severe; the local power is 57.3%, 72.5%, 88.0%, 87.3%, and 86.1% as seen in the middle column of Panel A.2.2. This is because the first through $(p-1)$-th lags of x_H approximate the p-th, significant lag fairly well. Thus, our conclusion from Panel A is that we should pick a sufficiently large $h \geq p$ when x_H has small autocorrelation and there is likely to be a lagged causality.

We now focus on Panels B and C, the low frequency cases. Recall that $m = 12$ and $p \leq 5$ in this experiment. This means that the causal effect from x_H to x_L exists only within one low frequency time period. Since the number of models combined $\underline{h}$ is in terms of *low* frequency, a reasonable conjecture is that letting $\underline{h} \geq 2$ should not improve local power. As seen in Panels B and C, this conjecture is in fact true regardless of ϕ, p, and the type of causality. We thus focus on $\underline{h} = 1$ here.

As far as Decaying Causality is concerned, the low frequency test does not perform much worse than the mixed frequency test. Let $\phi = 0.2$ and $p = 5$ for example. The mixed frequency test achieves power 21.1% when $h = 3$, while the low frequency test with stock sampling achieves 16.7% and the low frequency test with flow sampling achieves 18.0%. While the mixed frequency test has the highest power, the difference is not too large. The same pattern is observed when $\phi = 0.8$.

In the presence of Lagged Causality, the low frequency test often suffers from much lower power than the mixed frequency test. When $\phi = 0.2$, recall from the lower-triangular part of Panel A.2.1 that the mixed frequency test gets power between 33.5% and 53.7% by choosing $h \geq p$. The local power based on the low frequency test is all lower than 18.1% except for the first column of Panel B.2.1, which covers stock sampling with $p = 1$ and Lagged Causality. The

local power there is as high as 53.0%. To explain this outlier, recall that the only nonzero term under the local DGP (3.3.1) is $x_H(\tau_L - 1, m)$ when $p = 1$. Since $x_H(\tau_L - 1) = x_H(\tau_L - 1, m)$ under stock sampling, this term is exactly included in the low frequency model with stock sampling (see (3.3.5)). Except for this coincidence, local power based on the low frequency test is much lower than local power based on the mixed frequency test.

The superiority of the mixed frequency test is preserved when we focus on $\phi = 0.8$ with Lagged Causality. Let $p = 5$, then we see from Panel A.2.2 that the mixed frequency test has power 84.7%. In contrast, the low frequency test with stock sampling has 27.6% while the low frequency test with flow sampling has 73.1%. The flow-sampling model performs much better than the stock-sampling model since the former has the simple sum of $x_H(\tau_L - 1, 1), \ldots,$ $x_H(\tau_L - 1, 12)$ as a regressor, while the latter has only $x_H(\tau_L - 1, 12)$. Since $p = 5$ and $m = 12$, the only nonzero term under the local DGP is $x_H(\tau_L - 1, 8)$ and hence the flow-sampling model has a relatively relevant regressor.

We summarize the main implications from Table 3.2. First, choosing a sufficiently large $h \geq p$ is important when the mixed frequency approach is taken. This is especially true when x_H has low persistence and Lagged Causality is likely to exist. Second, the mixed frequency test with sufficiently large $h \geq p$ achieves higher power than the low frequency test in the presence of Lagged Causality. For Decaying Causality their performance is not so different, but the mixed frequency test never performs worse than the low frequency test at least.

3.4 Monte Carlo Simulations

In this section we run Monte Carlo simulations to examine the finite sample performance of the max test. Section 3.4.1 is concerned with high-to-low causality, while Section 3.4.2 is concerned with low-to-high causality.

3.4.1 High-to-Low Causality

We compare the finite sample performance of the mixed frequency high-to-low max test and the Wald test based on the naïve regression model (3.2.3).

The DGP is set as follows:

$$x_L(\tau_L) = 0.2x_L(\tau_L - 1) + \sum_{j=1}^{12} b_j x_H(\tau_L - 1, m + 1 - j) + \epsilon_L(\tau_L), \quad \epsilon_L(\tau_L) \overset{i.i.d.}{\sim} N(0,1),$$
$$x_H(\tau_L, j) = \phi x_H(\tau_L, j-1) + \epsilon_H(\tau_L, j), \quad \epsilon_H(\tau_L, j) \overset{i.i.d.}{\sim} N(0,1). \tag{3.4.1}$$

For the key coefficient $\boldsymbol{b}$, we try three cases below.

1. Non-causality: $\boldsymbol{b} = \mathbf{0}_{12\times 1}$. In this case we can check the empirical size of our tests.

2. Decaying Causality: $b_j = 0.1/j$ for $j = 1, \ldots, 12$. This is a commonly observed causal pattern where the coefficients decay gradually.

3. Lagged Causality: $b_j = 0.25 \times I(j = 12)$ for $j = 1, \ldots, 12$. Only b_{12} is 0.25 and all other coefficients are zeros. This case corresponds to seasonality or lagged response of x_L to x_H.

For the AR(1) coefficient of x_H, we try $\phi \in \{0.2, 0.8\}$. The ratio of sampling frequencies we try is $m \in \{3, 12\}$. $m = 3$ can be thought of as a month vs. quarter case, while $m = 12$ can be thought of as a week vs. quarter case approximately. The sample size in terms of low frequency is taken from $T_L \in \{40, 80, 120\}$. These values can be thought of as 40 quarters (i.e. 10 years) through 120 quarters (i.e. 30 years).

For the max test, we combine h parsimonious models (3.2.4) with $h \in \{4, 8, 12\}$. We use the equal weighting scheme, and the test statistic is computed based on 1,000 draws from the asymptotic null distribution.

For the Wald test, we postulate the naïve regression model (3.2.3) with $h \in \{4, 8, 12\}$. We use the parametric bootstrap with 499 replications and the normality assumption. Since the error ϵ_L is indeed normally distributed here, the parametric bootstrap controls the empirical size of the Wald test well. We also tried the Lagrange multiplier test and likelihood ratio test, but they turned out to be too conservative in small sample (i.e., their empirical size is way below the nominal size 5%). Hence we report the results of the Wald test only.

The number of Monte Carlo replications is 10,000 for the max test and 2,000 for the bootstrapped Wald test.

See Table 3.3 for the rejection frequencies of the mixed frequency max test and the Wald test based on the naïve regression model. Panel A focuses on $m = 3$, while Panel B focuses on $m = 12$.

We first focus on Non-causality to check the empirical size of our tests. The Wald test has a well-controlled size for each case due to the parametric bootstrap. The max test is also fine except for a challenging case with small T_L, small m, and large ϕ. For example, the worst empirical size of 24.3% appears when $T_L = 40$, $m = 3$, $\phi = 0.8$, and $h = 12$; see Panel A.1.2. This size distortion stems from a poor approximation of the covariance matrix $\boldsymbol{V}$ in (3.2.10). Recall that $\hat{\boldsymbol{V}}$, a consistent estimator for $\boldsymbol{V}$, is constructed from the sample moments of the high frequency process x_H. Hence, the precision of $\hat{\boldsymbol{V}}$ decreases when the high frequency sample size mT_L is small or x_H has a strong persistence. Apart from such severe combinations, the max test has a reasonable empirical size:

1. When $\phi = 0.2$, the empirical size is less than 9.7% even if $T_L = 40$ and $m = 3$ (cfr. Panel A.1.1).

2. When $T_L = 120$, the empirical size is less than 7.7% even if $\phi = 0.8$ and $m = 3$ (cfr. Panel A.1.2).

3. When $m = 12$, the empirical size is less than 8.3% even if $\phi = 0.8$ and $T_L = 40$ (cfr. Panel B.1.2).

We now compare the empirical power of the two tests. We first consider Decaying Causality. The max test has higher power than the Wald test for all slots in Panels A.2 and B.2. The difference is particularly large when $m = 12$, $\phi = 0.8$, $T_L = 80$, and $h = 12$ (cfr. Panel B.2.2). The power of the max test is 80.0% while the power of the Wald test is 50.1% there. Note that this difference is not due to size distortions since m and T_L are large (cfr. Panel B.1.2).

The exactly same goes for Lagged Causality. The max test has higher power than the Wald test for all slots in Panels A.3 and B.3. The difference is particularly large when $m = 12$, $\phi = 0.8$, $T_L = 80$, and $h = 12$ (cfr. Panel B.3.2). The power of the max test is 82.5% while the power of the Wald test is 57.0% there, and this difference is not due to size distortions.

In summary, the mixed frequency max test has higher power than the Wald test based on the naïve regression model under any plausible parameterizations. The former gets size

distortions only when T_L is small, m is small, and ϕ is large. If at least one of the three quantities is favorable, then the max test does not suffer from serious size distortions. Hence, our conclusions are twofold. First, take the Wald approach when T_L is small, m is small, and ϕ is large. Second, use the max test otherwise in order to achieve higher power than the Wald test. The improvement of power can be as large as 30% in some cases.

Finally, we discuss how the power of the max test evolves as h grows. For Decaying Causality, the power *decreases* gradually as h approaches the true lag order 12. See Panel B.2.1 with $T_L = 120$ for example. The power there is 19.5%, 15.3%, and 12.9% when h is 4, 8, and 12. This suggests that, under Decaying Causality, incorporating many lags is penalized even if they are relevant lags. An intuitive reason for this fact is that the first lag has the largest coefficient and hence the marginal benefit of including more lags is diminishing.

For Lagged Causality, the power of the max test jumps at $h = 12$. See Panel B.3.1 with $T_L = 120$ for instance. The power there is 5.7%, 5.8%, and 47.4% when h is 4, 8, and 12. This suggests that, under Lagged Causality, incorporating sufficiently many lags is crucial for getting high power.

3.4.2 Low-to-High Causality

We now consider Granger causality from x_L to x_H. Assume that the true DGP is a bivariate structural MF-VAR(1) with the ratio of sampling frequencies $m = 12$:

$$\underbrace{\begin{bmatrix} 1 & \dots & \dots & 0 & 0 \\ -0.2 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & \vdots & 0 \\ 0 & \dots & -0.2 & 1 & 0 \\ 0 & \dots & \dots & 0 & 1 \end{bmatrix}}_{\equiv \boldsymbol{N}} \underbrace{\begin{bmatrix} x_H(\tau_L, 1) \\ \vdots \\ x_H(\tau_L, 12) \\ x_L(\tau_L) \end{bmatrix}}_{\equiv \boldsymbol{X}(\tau_L)} = \underbrace{\begin{bmatrix} 0 & \dots & 0.2 & c_1 \\ \vdots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & c_m \\ 0.2/12 & \dots & 0.2/1 & 0.2 \end{bmatrix}}_{\equiv \boldsymbol{M}} \underbrace{\begin{bmatrix} x_H(\tau_L - 1, 1) \\ \vdots \\ x_H(\tau_L - 1, 12) \\ x_L(\tau_L - 1) \end{bmatrix}}_{= \boldsymbol{X}(\tau_L - 1)} + \underbrace{\begin{bmatrix} \eta_H(\tau_L, 1) \\ \vdots \\ \eta_H(\tau_L, m) \\ \eta_L(\tau_L) \end{bmatrix}}_{\equiv \boldsymbol{\eta}(\tau_L)} \tag{3.4.2}$$

with $\boldsymbol{\eta}(\tau_L) \overset{i.i.d.}{\sim} N(\mathbf{0}_{13\times 1}, \boldsymbol{I}_{13})$. There is decaying Granger causality from x_H to x_L in the sense that $x_L(\tau_L)$ depends on $\sum_{j=1}^{12}(0.2/j)x_H(\tau_L-1, m+1-j)$. The low frequency AR(1) coefficient of x_L is 0.2, while the high frequency AR(1) coefficient of x_H is 0.2 as well. x_H is also affected by past x_L, which is our main interest here. Specifically, $x_H(\tau_L, j)$ depends on $c_j x_L(\tau_L - 1)$ for $j = 1, \ldots, 12$. A key parameter vector $\boldsymbol{c} = [c_1, \ldots, c_{12}]'$ represents low-to-high causality, and we consider the following three cases.

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

2. Decaying Causality: $c_j = 0.3/j$ for $j = 1, \ldots, 12$. This is a commonly observed causal pattern where the coefficients decay gradually.

3. Lagged Causality: $c_j = 0.4 \times I(j = 12)$ for $j = 1, \ldots, 12$. Only c_{12} is 0.4 and all other coefficients are zeros. This case corresponds to seasonality or lagged response of x_H to x_L.

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

For the mixed frequency max test, we combine $h \in \{4, 8, 12\}$ parsimonious regression models, and the j-th model is specified as

$$x_L(\tau_L) = \alpha_{1,j} x_L(\tau_L - 1) + \sum_{k=1}^{p} \beta_{k,j} x_H(\tau_L - 1, m + 1 - k) + \gamma_j x_H(\tau_L + 1, j) + u_{L,j}(\tau_L), \tag{3.4.3}$$
$$\text{Instruments: } \{\text{all } p+2 \text{ regressors in model } j, x_H(\tau_L, 1), \ldots, x_H(\tau_L, m)\}.$$

The number of high frequency lags of x_H, namely p, is taken from $p \in \{4, 8, 12\}$.

For the purpose of comparison, we also formulate a low frequency counterpart to the model (3.4.3). We aggregate x_H using the linear aggregation scheme: $x_H(\tau_L) = \sum_{j=1}^{12} \delta_j x_H(\tau_L, j)$. In particular, we focus on stock sampling $\delta_j = I(j = 12)$ and flow sampling $\delta_j = 1/12$ in this simulation study. Using the aggregated x_H, we combine $\underline{h} \in \{1, 2, 3\}$ parsimonious regression

models, and the j-th model is specified as

$$x_L(\tau_L) = \alpha_{1,j}^{LF} x_L(\tau_L - 1) + \sum_{k=1}^{\underline{p}} \beta_{k,j}^{LF} x_H(\tau_L - k) + \gamma_j^{LF} x_H(\tau_L + j) + u_{L,j}^{LF}(\tau_L), \tag{3.4.4}$$

Instruments: $\left\{\text{all } \underline{p}+2 \text{ regressors in model } j, x_H(\tau_L)\right\}$.

The number of low frequency lags of x_H, namely $\underline{p}$, is taken from $\underline{p} \in \{1,2,3\}$. We can formulate the max test corresponding to the low frequency model (3.4.4) in a completely analogous fashion with the mixed frequency case.

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

See Table 3.4 for the rejection frequencies. Panel A has Non-causality, Panel B has Decaying Causality, and Panel C has Lagged Causality. For each panel we have the mixed frequency case, low frequency case with stock sampling, and low frequency case with flow sampling. We first focus on Panel A to check empirical size. Since the low frequency approach involves few parameters, the empirical size is always very close to the nominal size 5% (cfr. Panels A.2 and A.3). The mixed frequency approach involves more parameters, so there is a size distortion issue when T_L is as small as 40 (cfr. Panel A.1). The worst empirical size of 0.176 occurs when $(h, p, T_L) = (12, 12, 40)$. The empirical size converges to the nominal size 5% quickly as T_L grows to 80, however.

We now focus on Panel B: Decaying Causality. For the mixed frequency case, the empirical power is at most 0.375 when $T_L = 40$, 0.629 when $T_L = 80$, and 0.827 when $T_L = 120$ (cfr. Panel B.1). Fixing p, a larger h always produces lower power. This is reasonable since having more leads of x_H is not so informative given the decaying structure of low-to-high causality while the increased number of parameters certainly lowers power. Similarly, having larger p does not always improve power due to the decaying pattern of high-to-low causality.

Panel B.2 indicates that the low frequency test with stock sampling has absolutely no power. In contrast, Panel B.3 indicates that the low frequency test with flow sampling is in fact more powerful than the mixed frequency test. For example, the low frequency test with flow sampling at $(\underline{h}, \underline{p}, T_L) = (1, 1, 40)$ yields the rejection frequency of 0.370, while the mixed frequency test at

$(h, p, T_L) = (4, 4, 40)$ yields 0.305. This result suggests that the informational loss of switching from $\{\{x_H(\tau_L, j)\}\}$ to $\{\sum_{j=1}^{12} x_H(\tau_L, j)\}$ is relatively small when Decaying Causality is present, and thus the low frequency approach is preferred due to fewer parameters.

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

Panel C.3 shows that the low frequency test with flow sampling has very low power, which implies that the lagged causality at high frequency basis cannot be captured by $\{\sum_{j=1}^{12} x_H(\tau_L, j)\}$. In contrast, Panel C.2 indicates that the low frequency test with stock sampling is much more powerful than the mixed frequency test. For example, the low frequency test with stock sampling at $(\underline{h}, \underline{p}, T_L) = (1, 1, 40)$ yields the rejection frequency of 0.633, while the mixed frequency test at $(h, p, T_L) = (12, 4, 40)$ yields only 0.209. This is not surprising since the the low frequency test with stock sampling works on $\{x_H(\tau_L, 12)\}$, exactly relevant observations for Lagged Causality.

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

3.5 Empirical Application

In this section we use the max tests to examine the relationship between a weekly yield spread and the quarterly real GDP growth in the U.S. We are particularly interested in Granger

causality from the spread to the GDP growth (i.e. high-to-low causality). Yield spread used to be regarded as a strong predictor of business cycle, but more recent evidence questions its predictability. One well-known episode is "Greenspan's Conundrum" around 2005, when yield spread declined substantially due to constant long-term rates and increased short-term rates but the U.S. macroeconomy did not run into recession at that position. Although the U.S. economy did get a serious recession due to the subprime mortgage crisis starting December 2007, the time lag between the declined yield spread and that recession seems much larger than it used to be in the past. Based on this motivation, we investigate how Granger causality from yield spread to GDP growth evolved over past fifty years.

As a business cycle measure, the seasonally-adjusted quarterly real GDP growth is used. The data can be found at Federal Reserve Economic Data (FRED). To remove potential seasonal effects remaining after seasonal adjustment, we use percentage growth rate from previous year.

For short-term and long-term interest rates, we first download daily series of 1-year Treasury constant maturity rate and 10-year Treasury constant maturity rate at FRED. A convenient feature of these two series is that they share the identical time grid. The federal fund rate or 3-month Treasury bill rate may be a more popular proxy for short-term interest rates, but they have different time grids from the 10-year Treasury constant maturity rate. While we could directly work on the daily interest rates and the quarterly GDP, the ratio of sampling frequencies m seems too large to ensure reasonable size and power. We thus aggregate the daily series into weekly by picking the last observation in each week, recalling that interest rates are stock variables. Finally, we calculate yield spread as the difference between the weekly 10-year rate and the weekly 1-year rate.

Figure 3.1 shows the weekly 10-year rate, weekly 1-year rate, their spread, and the quarterly GDP growth from January 5, 1962 through December 31, 2013. This sample period covers 2,736 weeks or 208 quarters. The shaded areas represent recession periods defined by the National Bureau of Economic Research (NBER). Until 1980, sharp decline of the spread seemed to be immediately followed by recession. After 1980, however, we find a weaker evidence or at least there is a larger time lag between declined spread and recession.

Table 3.5 shows sample statistics of the weekly 10-year rate, weekly 1-year rate, their spread, and the quarterly real GDP growth. The 10-year rate is 1% point larger than the 1-year rate

on average. The average GDP growth is 3.151%, indicating a fairly steady growth of the U.S. economy over the past 52 years. The spread has a relatively small kurtosis of 2.750, while the GDP growth has a kurtosis of 3.543.

When we apply the mixed frequency causality test, a slightly inconvenient aspect of our data is that the number of weeks contained in each quarter is not constant. Specifically, (i) 13 quarters have 12 weeks each, (ii) 150 quarters have 13 weeks each, and (iii) 45 quarters have 14 weeks each. Since our asymptotic theory requires m to be constant, we assume $m = 13$ by making the following modification. We (i) duplicate the twelfth observation once when a quarter contains 12 weeks, (ii) do nothing when it contains 13 weeks, and (iii) cut the last observation when it contains 14 weeks. This gives us a manageable dataset with $T_L = 208$, $m = 13$, and thus $T = mT_L = 2,704$.

For high-to-low causality (i.e. causality from spread to GDP), we fit mixed frequency parsimonious models:

$$x_L(\tau_L) = \alpha_{0,j} + \sum_{k=1}^{q} \alpha_{k,j} x_L(\tau_L - k) + \beta_j x_H(\tau_L - 1, m + 1 - j) + u_{L,j}(\tau_L), \quad j = 1, \ldots, h,$$

and low frequency parsimonious models:

$$x_L(\tau_L) = \alpha_{0,j}^{LF} + \sum_{k=1}^{q} \alpha_{k,j}^{LF} x_L(\tau_L - k) + \beta_j^{LF} x_H(\tau_L - j, m) + u_{L,j}^{LF}(\tau_L), \quad j = 1, \ldots, \underline{h},$$

where we are using the fact that yield spread is a stock variable.[3]

Since our entire sample size is as large as 52 years, we implement rolling window analysis with the window width being 10 years or 20 years. For example, when the width is 10 years, the first subsample is 1962:I-1971:IV, the second one is 1962:II-1972:I, and so on. The 10-year width gives us 169 subsamples, while the 20-year width gives us 129 subsamples.

The trade-off between a small window width and a large one is that the large window is more likely to contain a structural break but it allows us to include more leads and lags in our model. For the 10-year case, we set $h = 13$, $\underline{h} = 1$, and $q = 2$. This means that we include

[3] Mixed frequency models and low frequency models for low-to-high causality (i.e. Granger causality from GDP to spread) are not presented here, but their details and the empirical results are available upon request.

13 weeks of lagged x_H and 2 quarters of lagged x_L in the mixed frequency models, while we include 1 quarter of lagged x_H and 2 quarters of lagged x_L in the low frequency models. For the 20-year case, we set $h = 26$, $\underline{h} = 2$, and $q = 2$. This means that we include 26 weeks of lagged x_H and 2 quarters of lagged x_L in the mixed frequency models, while we include 2 quarters of lagged x_H and 2 quarters of lagged x_L in the low frequency models.

Figure 3.2 plots the asymptotic p-values of the max tests with respect to the null hypothesis of high-to-low non-causality. Panel (a) is for the 10-year window width, while Panel (b) is for the 20-year width. "MF" means a mixed frequency approach which works on weekly spread and quarterly GDP, while "LF" means a low frequency approach which works on quarterly spread and quarterly GDP. The shaded area represents the 5% level.

The first half of our entire sample shows very strong rejection of non-causality in both MF and LF cases, which means that yield spread used to be a valid predictor of GDP. In Panel (a), the p-values are almost always below 5% before subsample 1982:I-1991:IV. In Panel (b), the p-values are always below 5% before subsample 1982:I-2001:IV.

After these periods the MF-based p-values start to fluctuate between a relatively narrow range $[0, 0.5]$, while the LF-based p-values start to fluctuate between a wide range $[0, 1]$. Most recent samples, including the period of Greenspan's Conundrum, have insignificant causality for both approaches. In Panel (a), the p-values are always above 5% after subsample 2001:I-2010:IV. In Panel (b), the p-values are always above 5% after subsample 1991:I-2010:IV. This result suggests that yield spread is a less valid predictor of GDP than it used to be, probably due to structural changes of U.S. economy as well as Federal Reserve Board's financial policies.

Throughout the entire sample, the MF-based p-values are always smaller than the LF-based ones. This result is consistent with Ghysels, Hill, and Motegi's (2013) Theorem 4.2 stating that high-to-low non-causality given MF information set implies high-to-low non-causality given LF information set. We can thus conclude that using weekly yield spread is more informative than using quarterly spread in terms of GDP prediction.

3.6 Conclusions

This paper proposes regression-based mixed frequency Granger causality tests by combining multiple parsimonious models where the i-th model regresses a low frequency variable x_L onto the i-th high frequency lag or lead of a high frequency variable x_H. Letting $\hat{\beta}_i$ be an estimator for the loading of the i-th lag or lead of x_H, our test statistic basically takes the maximum among $\hat{\beta}_1^2, \ldots, \hat{\beta}_h^2$. In this sense our test can be called the *max test* for short.

In local power analysis on high-to-low causality, we show that the max test is more powerful than the Wald test based on a naïve regression model which contains all relevant lags at once. The difference in power can get large up to 15-20% when h gets large.

In small sample, we show via Monte Carlo simulations that the max test produces no size distortions under realistic parameterizations and it is more powerful than the naïve Wald test. The difference in power can be as large as 30% in some cases.

As an empirical application, we run a rolling window analysis on weekly interest rate spread and quarterly real GDP growth in the U.S. We get a reasonable result that the spread used to be a valid predictor of GDP until 1980 or around, but its predictive ability declined more recently. We also find that the mixed frequency approach has more frequent rejections of non-causality from spread to GDP than the conventional low frequency approach, which suggests that taking the mixed frequency approach provides more accurate prediction of GDP based on yield spread.

Table 3.1: Local Asymptotic Power of Max Test and Wald Test (High-to-Low Causality)

This table shows the local asymptotic power of the mixed frequency max test and the Wald test based on the naïve regression model (3.2.3). It focuses on Granger causality from x_H to x_L. The max test uses an equal weighting scheme. Panel A considers Decaying Causality, where $\nu_i = 0.8 - 0.1(i-1)$ for $i = 1, \ldots, p$. Panel B considers Lagged Causality, where $\nu_i = 2 \times I(i = p)$ for $i = 1, \ldots, p$. x_H is assumed to follow AR(1) with coefficient $\phi \in \{0.2, 0.8\}$. Panels A.1 and B.1 consider $\phi = 0.2$, a relatively transitory x_H. Panels A.2 and B.2 consider $\phi = 0.8$, a relatively persistent x_H. The lag length in the DGP is $p \in \{1, \ldots, 5\}$, while the number of lags included in each model is $h \in \{1, \ldots, 5\}$. Other minor quantities are as follows: $\sigma_L^2 = 1$, $\alpha = 0.05$, $R_1 = R_2 = 100,000$.

Panel A. Decaying Causality

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

	$p=1$		$p=2$		$p=3$		$p=4$		$p=5$	
	Max	Wald	Max	Wald	Max	Wald	Max	Wald	Max	Wald
$h=1$	0.131	0.129	0.163	0.160	0.169	0.166	0.170	0.167	0.170	0.167
$h=2$	0.106	0.104	0.180	0.170	0.201	0.192	0.205	0.195	0.205	0.196
$h=3$	0.093	0.092	0.151	0.146	0.197	0.192	0.209	0.206	0.211	0.209
$h=4$	0.086	0.086	0.135	0.132	0.175	0.172	0.202	0.202	0.209	0.210
$h=5$	0.081	0.081	0.125	0.122	0.162	0.157	0.187	0.184	0.202	0.202

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

	$p=1$		$p=2$		$p=3$		$p=4$		$p=5$	
	Max	Wald	Max	Wald	Max	Wald	Max	Wald	Max	Wald
$h=1$	0.268	0.266	0.624	0.621	0.829	0.828	0.916	0.915	0.950	0.950
$h=2$	0.245	0.204	0.642	0.555	0.867	0.809	0.946	0.916	0.974	0.956
$h=3$	0.226	0.175	0.617	0.490	0.871	0.770	0.956	0.904	0.983	0.954
$h=4$	0.212	0.156	0.595	0.443	0.858	0.726	0.955	0.882	0.983	0.945
$h=5$	0.199	0.143	0.574	0.407	0.845	0.688	0.949	0.856	0.982	0.932

Panel B. Lagged Causality

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

	$p=1$		$p=2$		$p=3$		$p=4$		$p=5$	
	Max	Wald	Max	Wald	Max	Wald	Max	Wald	Max	Wald
$h=1$	0.537	0.532	0.071	0.069	0.052	0.051	0.052	0.050	0.052	0.050
$h=2$	0.443	0.431	0.443	0.431	0.064	0.063	0.052	0.051	0.051	0.050
$h=3$	0.390	0.372	0.395	0.372	0.394	0.372	0.060	0.060	0.051	0.050
$h=4$	0.359	0.333	0.361	0.333	0.362	0.333	0.358	0.333	0.059	0.058
$h=5$	0.336	0.304	0.339	0.304	0.337	0.304	0.338	0.304	0.335	0.304

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

	$p=1$		$p=2$		$p=3$		$p=4$		$p=5$	
	Max	Wald	Max	Wald	Max	Wald	Max	Wald	Max	Wald
$h=1$	0.916	0.915	0.759	0.760	0.573	0.569	0.403	0.400	0.278	0.277
$h=2$	0.895	0.856	0.895	0.856	0.725	0.663	0.531	0.465	0.369	0.313
$h=3$	0.878	0.808	0.884	0.808	0.880	0.808	0.697	0.597	0.502	0.404
$h=4$	0.862	0.767	0.872	0.767	0.873	0.767	0.863	0.767	0.673	0.548
$h=5$	0.847	0.731	0.859	0.731	0.861	0.731	0.859	0.731	0.847	0.731

Table 3.2: Local Asymptotic Power of Max Tests (High-to-Low Causality)

This table shows the local asymptotic power of the regression-based Granger causality test from x_H to x_L. We conduct the mixed frequency test in Panel A, the low frequency test with stock sampling in Panel B, and the low frequency test with flow sampling in Panel C. All of these tests use the equal weighting scheme. For each panel we consider two cases: Decaying Causality where $\nu_i = 0.8 - 0.1(i-1)$ for $i = 1, \ldots, p$, and Lagged Causality where $\nu_i = 2 \times I(i = p)$ for $i = 1, \ldots, p$. x_H is assumed to follow AR(1) with coefficient $\phi \in \{0.2, 0.8\}$. The high frequency lag length in the DGP is $p \in \{1, \ldots, 5\}$. The number of mixed frequency models combined is $h \in \{1, \ldots, 5\}$, while the number of low frequency models combined is $\underline{h} \in \{1, \ldots, 3\}$. Other quantities are as follows: $m = 12$, $\sigma_L^2 = 1$, $\alpha = 0.05$, $R_1 = R_2 = 100,000$.

Panel A. Mixed Frequency Test

A.1. Decaying Causality

	A.1.1. $\phi = 0.2$					A.1.2. $\phi = 0.8$				
	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$
$h=1$	0.131	0.163	0.169	0.170	0.170	0.268	0.624	0.829	0.916	0.950
$h=2$	0.106	0.180	0.201	0.205	0.205	0.245	0.642	0.867	0.946	0.974
$h=3$	0.093	0.151	0.197	0.209	0.211	0.226	0.617	0.871	0.956	0.983
$h=4$	0.086	0.135	0.175	0.202	0.209	0.212	0.595	0.858	0.955	0.983
$h=5$	0.081	0.125	0.162	0.187	0.202	0.199	0.574	0.845	0.949	0.982

A.2. Lagged Causality

	A.2.1 $\phi = 0.2$					A.2.2. $\phi = 0.8$				
	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$
$h=1$	0.537	0.071	0.052	0.052	0.052	0.916	0.759	0.573	0.403	0.278
$h=2$	0.443	0.443	0.064	0.052	0.051	0.895	0.895	0.725	0.531	0.369
$h=3$	0.390	0.395	0.394	0.060	0.051	0.878	0.884	0.880	0.697	0.502
$h=4$	0.359	0.361	0.362	0.358	0.059	0.862	0.872	0.873	0.863	0.673
$h=5$	0.336	0.339	0.337	0.338	0.335	0.847	0.859	0.861	0.859	0.847

Panel B. Low Frequency Test (Stock Sampling)

B.1. Decaying Causality

	B.1.1. $\phi = 0.2$					B.1.2. $\phi = 0.8$				
	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$
$\underline{h}=1$	0.128	0.160	0.166	0.167	0.167	0.265	0.618	0.824	0.913	0.949
$\underline{h}=2$	0.101	0.122	0.127	0.127	0.127	0.203	0.526	0.754	0.866	0.916
$\underline{h}=3$	0.090	0.108	0.112	0.112	0.112	0.174	0.471	0.707	0.833	0.893

B.2 Lagged Causality

	B.2.1. $\phi = 0.2$					B.2.2. $\phi = 0.8$				
	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$
$\underline{h}=1$	0.530	0.069	0.050	0.050	0.050	0.913	0.757	0.566	0.399	0.276
$\underline{h}=2$	0.436	0.062	0.050	0.049	0.049	0.865	0.674	0.477	0.325	0.230
$\underline{h}=3$	0.385	0.060	0.050	0.049	0.049	0.832	0.624	0.423	0.279	0.195

Panel C. Low Frequency Test (Flow Sampling)

C.1. Decaying Causality

	C.1.1. $\phi = 0.2$					C.1.2. $\phi = 0.8$				
	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$
$\underline{h}=1$	0.056	0.077	0.109	0.145	0.180	0.113	0.316	0.587	0.798	0.912
$\underline{h}=2$	0.054	0.067	0.088	0.112	0.137	0.091	0.245	0.493	0.722	0.864
$\underline{h}=3$	0.053	0.064	0.080	0.100	0.120	0.082	0.209	0.439	0.673	0.830

C.2. Lagged Causality

	C.2.1. $\phi = 0.2$					C.2.2. $\phi = 0.8$				
	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=1$	$p=2$	$p=3$	$p=4$	$p=5$
$\underline{h}=1$	0.093	0.109	0.112	0.113	0.113	0.452	0.564	0.645	0.699	0.731
$\underline{h}=2$	0.078	0.088	0.090	0.090	0.090	0.364	0.471	0.554	0.610	0.646
$\underline{h}=3$	0.071	0.080	0.081	0.082	0.082	0.317	0.418	0.499	0.556	0.593

Table 3.3: Rejection Frequencies of Max Test and Wald Test (High-to-Low Causality)

This table shows the rejection frequencies of the mixed frequency max test and the Wald test based on the naïve regression model. The max test uses the equal weighting scheme, and the test statistic is computed based on 1,000 draws from the asymptotic null distribution. When we implement the Wald test, the parametric bootstrap with 499 replications is employed. Panel A focuses on $m = 3$, which can be thought of as a month vs. quarter case. Panel B focuses on $m = 12$, which can be thought of as a week vs. quarter case approximately. The sample size T_L is 40, 80, or 120 quarters. For each panel we consider three cases: Non-causality, Decaying Causality, and Lagged Causality. x_H is assumed to follow AR(1) with coefficient $\phi \in \{0.2, 0.8\}$. The number of high frequency lags included in our models is $h \in \{4, 8, 12\}$. The number of Monte Carlo replications is 10,000 for the max test and 2,000 for the bootstrapped Wald test. The nominal size is 5%.

Panel A. $m = 3$ (month vs. quarter)

A.1. Non-causality ($\boldsymbol{b} = \mathbf{0}_{12\times1}$)

	A.1.1. $\phi = 0.2$						A.1.2. $\phi = 0.8$					
	$T_L = 40$		$T_L = 80$		$T_L = 120$		$T_L = 40$		$T_L = 80$		$T_L = 120$	
	Max	Wald	Max	Wald	Max	Wald	Max	Wald	Max	Wald	Max	Wald
$h = 4$	0.063	0.041	0.061	0.045	0.066	0.046	0.125	0.043	0.070	0.043	0.056	0.048
$h = 8$	0.079	0.047	0.061	0.043	0.062	0.045	0.183	0.044	0.108	0.043	0.063	0.042
$h = 12$	0.096	0.042	0.064	0.050	0.058	0.046	0.243	0.044	0.125	0.053	0.076	0.045

A.2. Decaying Causality ($b_j = 0.1/j,\ \ j = 1, \ldots, 12$)

	A.2.1. $\phi = 0.2$						A.2.2. $\phi = 0.8$					
	$T_L = 40$		$T_L = 80$		$T_L = 120$		$T_L = 40$		$T_L = 80$		$T_L = 120$	
	Max	Wald	Max	Wald	Max	Wald	Max	Wald	Max	Wald	Max	Wald
$h = 4$	0.106	0.072	0.142	0.123	0.208	0.168	0.559	0.296	0.810	0.638	0.933	0.856
$h = 8$	0.106	0.056	0.120	0.086	0.162	0.129	0.563	0.198	0.820	0.517	0.920	0.762
$h = 12$	0.114	0.055	0.118	0.082	0.132	0.103	0.585	0.147	0.787	0.415	0.913	0.675

A.3. Lagged Causality ($b_j = 0.25 \times I(j = 12),\ \ j = 1, \ldots, 12$)

	A.3.1. $\phi = 0.2$						A.3.2. $\phi = 0.8$					
	$T_L = 40$		$T_L = 80$		$T_L = 120$		$T_L = 40$		$T_L = 80$		$T_L = 120$	
	Max	Wald	Max	Wald	Max	Wald	Max	Wald	Max	Wald	Max	Wald
$h = 4$	0.061	0.041	0.060	0.048	0.064	0.043	0.133	0.049	0.093	0.057	0.088	0.064
$h = 8$	0.075	0.034	0.062	0.042	0.063	0.048	0.226	0.055	0.227	0.093	0.228	0.141
$h = 12$	0.177	0.088	0.300	0.193	0.462	0.331	0.577	0.178	0.824	0.520	0.951	0.784

Panel B. $m = 12$ (week vs. quarter, approximately)

B.1. Non-causality ($\boldsymbol{b} = \mathbf{0}_{12\times1}$)

	B.1.1. $\phi = 0.2$						B.1.2. $\phi = 0.8$					
	$T_L = 40$		$T_L = 80$		$T_L = 120$		$T_L = 40$		$T_L = 80$		$T_L = 120$	
	Max	Wald	Max	Wald	Max	Wald	Max	Wald	Max	Wald	Max	Wald
$h = 4$	0.074	0.049	0.074	0.052	0.061	0.041	0.080	0.056	0.048	0.050	0.055	0.044
$h = 8$	0.076	0.053	0.077	0.044	0.058	0.050	0.082	0.050	0.062	0.046	0.049	0.051
$h = 12$	0.071	0.045	0.071	0.046	0.055	0.043	0.075	0.049	0.050	0.049	0.051	0.041

B.2. Decaying Causality ($b_j = 0.1/j,\ \ j = 1, \ldots, 12$)

	B.2.1. $\phi = 0.2$						B.2.2. $\phi = 0.8$					
	$T_L = 40$		$T_L = 80$		$T_L = 120$		$T_L = 40$		$T_L = 80$		$T_L = 120$	
	Max	Wald	Max	Wald	Max	Wald	Max	Wald	Max	Wald	Max	Wald
$h = 4$	0.112	0.080	0.171	0.124	0.195	0.173	0.598	0.358	0.842	0.704	0.958	0.887
$h = 8$	0.102	0.069	0.143	0.099	0.153	0.140	0.556	0.271	0.854	0.605	0.949	0.818
$h = 12$	0.089	0.062	0.119	0.084	0.129	0.112	0.468	0.194	0.800	0.501	0.934	0.743

B.3. Lagged Causality ($b_j = 0.25 \times I(j = 12),\ \ j = 1, \ldots, 12$)

	B.3.1. $\phi = 0.2$						B.3.2. $\phi = 0.8$					
	$T_L = 40$		$T_L = 80$		$T_L = 120$		$T_L = 40$		$T_L = 80$		$T_L = 120$	
	Max	Wald	Max	Wald	Max	Wald	Max	Wald	Max	Wald	Max	Wald
$h = 4$	0.075	0.052	0.076	0.049	0.057	0.041	0.095	0.061	0.076	0.067	0.094	0.068
$h = 8$	0.078	0.045	0.079	0.048	0.058	0.054	0.143	0.071	0.207	0.119	0.242	0.173
$h = 12$	0.144	0.091	0.321	0.220	0.474	0.350	0.453	0.217	0.825	0.570	0.959	0.815

Table 3.4: Rejection Frequencies of Max Test for Low-to-High Causality

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

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

Panel A.1. Mixed Frequency

	$T_L = 40$			$T_L = 80$			$T_L = 120$		
	$p = 4$	$p = 8$	$p = 12$	$p = 4$	$p = 8$	$p = 12$	$p = 4$	$p = 8$	$p = 12$
$h = 4$	0.071	0.100	0.154	0.062	0.069	0.080	0.050	0.064	0.067
$h = 8$	0.070	0.106	0.163	0.051	0.068	0.093	0.051	0.058	0.073
$h = 12$	0.071	0.105	0.176	0.053	0.069	0.091	0.051	0.062	0.077

Panel A.2. Low Frequency (Stock Sampling)

	$T_L = 40$			$T_L = 80$			$T_L = 120$		
	$\underline{p} = 1$	$\underline{p} = 2$	$\underline{p} = 3$	$\underline{p} = 1$	$\underline{p} = 2$	$\underline{p} = 3$	$\underline{p} = 1$	$\underline{p} = 2$	$\underline{p} = 3$
$\underline{h} = 1$	0.056	0.063	0.067	0.057	0.051	0.058	0.049	0.056	0.054
$\underline{h} = 2$	0.054	0.056	0.061	0.054	0.056	0.054	0.051	0.047	0.051
$\underline{h} = 3$	0.048	0.054	0.057	0.052	0.049	0.057	0.056	0.055	0.048

Panel A.3. Low Frequency (Flow Sampling)

	$T_L = 40$			$T_L = 80$			$T_L = 120$		
	$\underline{p} = 1$	$\underline{p} = 2$	$\underline{p} = 3$	$\underline{p} = 1$	$\underline{p} = 2$	$\underline{p} = 3$	$\underline{p} = 1$	$\underline{p} = 2$	$\underline{p} = 3$
$\underline{h} = 1$	0.055	0.056	0.060	0.055	0.058	0.061	0.048	0.055	0.052
$\underline{h} = 2$	0.051	0.051	0.060	0.049	0.048	0.054	0.049	0.055	0.056
$\underline{h} = 3$	0.045	0.051	0.058	0.051	0.057	0.056	0.059	0.050	0.048

Table 3.4: Rejection Frequencies of Max Test for Low-to-High Causality (Continued)

Panel B. Decaying Causality ($c_j = 0.3/j$, $j = 1, \ldots, 12$)

Panel B.1. Mixed Frequency

	$T_L = 40$			$T_L = 80$			$T_L = 120$		
	$p = 4$	$p = 8$	$p = 12$	$p = 4$	$p = 8$	$p = 12$	$p = 4$	$p = 8$	$p = 12$
$h = 4$	0.305	0.350	0.375	0.627	0.621	0.629	0.827	0.816	0.816
$h = 8$	0.224	0.270	0.342	0.503	0.520	0.538	0.735	0.740	0.744
$h = 12$	0.182	0.230	0.313	0.436	0.446	0.472	0.676	0.681	0.690

Panel B.2. Low Frequency (Stock Sampling)

	$T_L = 40$			$T_L = 80$			$T_L = 120$		
	$\underline{p} = 1$	$\underline{p} = 2$	$\underline{p} = 3$	$\underline{p} = 1$	$\underline{p} = 2$	$\underline{p} = 3$	$\underline{p} = 1$	$\underline{p} = 2$	$\underline{p} = 3$
$\underline{h} = 1$	0.057	0.067	0.065	0.059	0.062	0.055	0.068	0.071	0.071
$\underline{h} = 2$	0.050	0.061	0.065	0.051	0.055	0.058	0.060	0.062	0.062
$\underline{h} = 3$	0.055	0.057	0.061	0.054	0.061	0.061	0.052	0.054	0.056

Panel B.3. Low Frequency (Flow Sampling)

	$T_L = 40$			$T_L = 80$			$T_L = 120$		
	$\underline{p} = 1$	$\underline{p} = 2$	$\underline{p} = 3$	$\underline{p} = 1$	$\underline{p} = 2$	$\underline{p} = 3$	$\underline{p} = 1$	$\underline{p} = 2$	$\underline{p} = 3$
$\underline{h} = 1$	0.370	0.376	0.379	0.663	0.666	0.663	0.843	0.834	0.836
$\underline{h} = 2$	0.290	0.288	0.285	0.564	0.548	0.563	0.766	0.764	0.762
$\underline{h} = 3$	0.240	0.231	0.244	0.496	0.494	0.511	0.715	0.712	0.706

Panel C. Lagged Causality ($c_j = 0.4 \times I(j = 12)$, $j = 1, \ldots, 12$)

Panel C.1. Mixed Frequency

	$T_L = 40$			$T_L = 80$			$T_L = 120$		
	$p = 4$	$p = 8$	$p = 12$	$p = 4$	$p = 8$	$p = 12$	$p = 4$	$p = 8$	$p = 12$
$h = 4$	0.065	0.099	0.142	0.055	0.072	0.083	0.055	0.069	0.074
$h = 8$	0.068	0.107	0.156	0.054	0.073	0.092	0.052	0.069	0.078
$h = 12$	0.209	0.234	0.294	0.582	0.587	0.593	0.844	0.833	0.836

Panel C.2. Low Frequency (Stock Sampling)

	$T_L = 40$			$T_L = 80$			$T_L = 120$		
	$\underline{p} = 1$	$\underline{p} = 2$	$\underline{p} = 3$	$\underline{p} = 1$	$\underline{p} = 2$	$\underline{p} = 3$	$\underline{p} = 1$	$\underline{p} = 2$	$\underline{p} = 3$
$\underline{h} = 1$	0.633	0.639	0.629	0.920	0.914	0.913	0.987	0.985	0.983
$\underline{h} = 2$	0.542	0.527	0.512	0.879	0.872	0.867	0.977	0.973	0.972
$\underline{h} = 3$	0.457	0.443	0.444	0.833	0.829	0.822	0.959	0.960	0.959

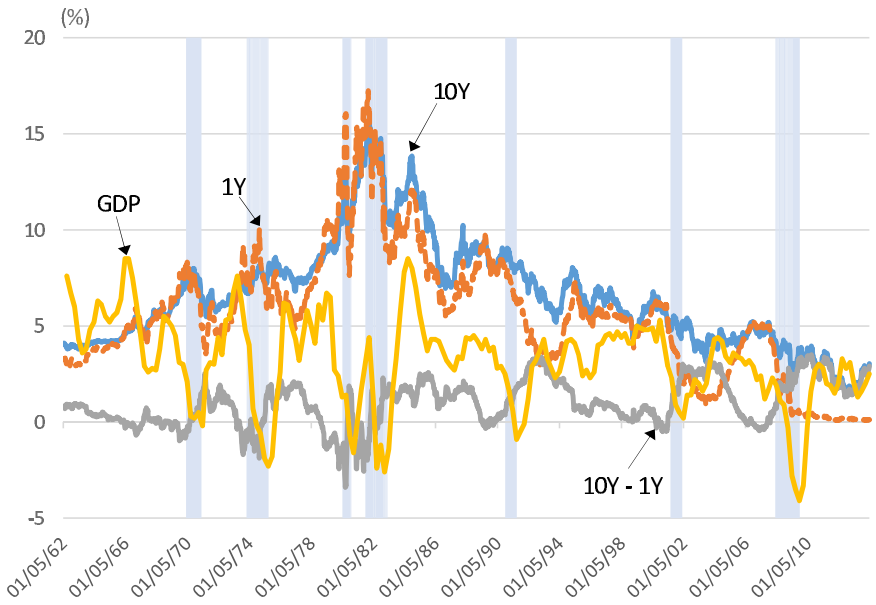
Panel C.3. Low Frequency (Flow Sampling)

	$T_L = 40$			$T_L = 80$			$T_L = 120$		
	$\underline{p} = 1$	$\underline{p} = 2$	$\underline{p} = 3$	$\underline{p} = 1$	$\underline{p} = 2$	$\underline{p} = 3$	$\underline{p} = 1$	$\underline{p} = 2$	$\underline{p} = 3$
$\underline{h} = 1$	0.087	0.096	0.102	0.133	0.126	0.134	0.184	0.189	0.177
$\underline{h} = 2$	0.075	0.086	0.088	0.115	0.112	0.114	0.146	0.148	0.156
$\underline{h} = 3$	0.072	0.070	0.076	0.094	0.096	0.098	0.121	0.128	0.129

Table 3.5: Sample Statistics of U.S. Interest Rates and Real GDP Growth

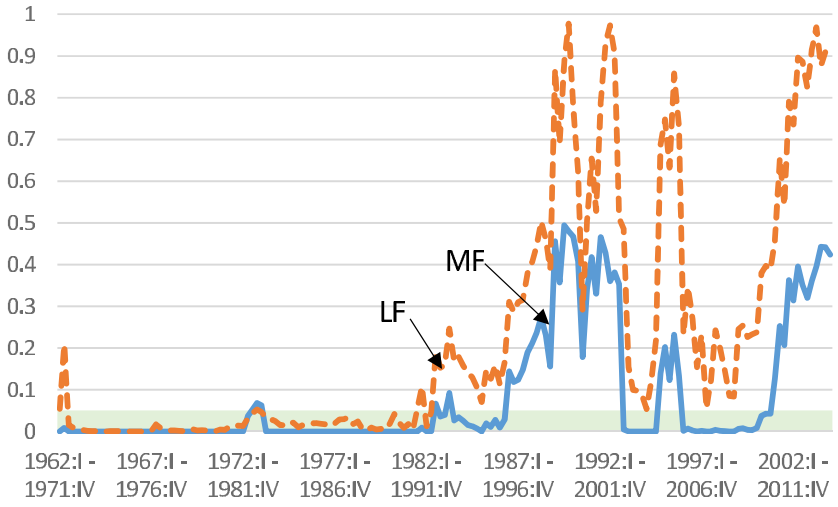
Sample statistics of weekly 10-year Treasury constant maturity rate, weekly 1-year Treasury constant maturity rate, their spread, and the quarterly real GDP growth from previous year. All these series are in terms of percentage. The sample period covers January 5, 1962 through December 31, 2013, which has 2,736 weeks or 208 quarters.

	mean	median	std. dev.	skewness	kurtosis
10-Year	6.555	6.210	2.734	0.781	3.488
1-Year	5.555	5.450	3.278	0.599	3.733
Spread	0.999	0.920	1.176	-0.120	2.750
GDP	3.151	3.250	2.349	-0.461	3.543

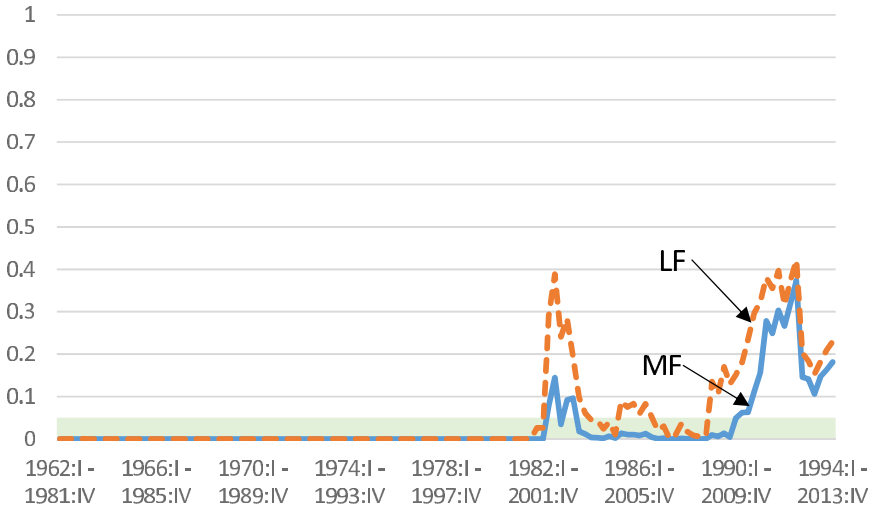


Note: This figure plots weekly 10-year Treasury constant maturity rate, weekly 1-year Treasury constant maturity rate, their spread, and the quarterly real GDP growth from previous year. The sample period covers January 5, 1962 through December 31, 2013, which has 2,736 weeks or 208 quarters. The shaded areas represent recession periods defined by the National Bureau of Economic Research (NBER).

Figure 3.1: Time Series Plot of U.S. Interest Rates and Real GDP Growth



(a) 10-Year Subsamples



(b) 20-Year Subsamples

Note: This figure plots asymptotic p-values with respect to the null hypothesis of high-to-low non-causality (i.e. non-causality from yield spread to GDP). Panel (a) is for the 10-year rolling window, while Panel (b) is for the 20-year rolling window. "MF" means a mixed frequency approach which works on weekly spread and quarterly GDP, while "LF" means a low frequency approach which works on quarterly spread and quarterly GDP. The shaded area represents the 5% level.

Figure 3.2: Asymptotic p-values for High-to-Low Non-Causality

APPENDIX A

TECHNICAL APPENDICES FOR CHAPTER 2

A.1 Asymptotic Properties of MF-VAR Parameter Estimators

In this section we derive the asymptotic distribution of the MF-VAR parameter estimators leading to the proofs of Theorems 2.2.1 and 2.2.2. We additionally present a simple consistent *almost surely* positive semi-definite estimator of the least squares asymptotic variance that we use in the simulation study and empirical application.

A.1.1 Least Squares Estimator and Asymptotic Variance

In this subsection we present the compact model that leads to the least squares estimator $\hat{\boldsymbol{B}}(h)$ of the parameter set $\boldsymbol{B}(h)$ appearing in equation (2.2.6). We then characterize the matrix components that enter into the least squares asymptotic covariance $\boldsymbol{\Sigma}_p(h) = (\boldsymbol{I}_K \otimes \boldsymbol{\Gamma}_{p,0}^{-1})\boldsymbol{D}_p(h)(\boldsymbol{I}_K \otimes \boldsymbol{\Gamma}_{p,0}^{-1})'$ appearing in the proof of Theorem 2.2.1 below. We save notation by writing $\boldsymbol{\Sigma}_p$ instead of $\boldsymbol{\Sigma}_p(h)$ throughout the appendix. We then explicitly derive the covariance matrices $\boldsymbol{\Gamma}_{p,0}$ and $\boldsymbol{D}_p(h)$. Finally, we present a simple consistent HAC estimator of $\boldsymbol{\Sigma}_p$ that satisfies the requirements of Theorem 2.2.2. The proofs of Theorems 2.2.1 and 2.2.2 are presented in Appendix A.1.2 where we explicitly verify the form of $\boldsymbol{\Sigma}_p$.

Least Squares Estimator

We require a more compact notation in order to derive the least squares estimator $\hat{\boldsymbol{B}}(h)$. Define

$$\begin{aligned}
\boldsymbol{W}_h(k) &= [\boldsymbol{X}(h), \boldsymbol{X}(1+h), \ldots, \boldsymbol{X}(T_L - k + h)]' \in \mathcal{R}^{(T_L-k+1)\times K} \\
\boldsymbol{W}(\tau_L, p) &= \left[\boldsymbol{X}(\tau_L)', \boldsymbol{X}(\tau_L - 1)', \ldots, \boldsymbol{X}(\tau_L - p + 1)'\right]' \in \mathcal{R}^{pK\times 1} \\
\overline{\boldsymbol{W}}_p(h) &= [\boldsymbol{W}(0,p), \boldsymbol{W}(1,p), \ldots, \boldsymbol{W}(T_L - h, p)]' \in \mathcal{R}^{(T_L-h+1)\times pK},
\end{aligned} \tag{A.1.1}$$

and define the error

$$\boldsymbol{u}^{(h)}(\tau_L) = \sum_{k=0}^{h-1} \boldsymbol{\Psi}_k \boldsymbol{\epsilon}(\tau_L - k) \tag{A.1.2}$$

stacked as follows:

$$\boldsymbol{U}_l(k) = \left[\boldsymbol{u}^{(h)}(l), \boldsymbol{u}^{(h)}(1+l), \ldots, \boldsymbol{u}^{(h)}(T_L - k + l)\right]' \in \mathcal{R}^{(T_L-k+1)\times K}. \tag{A.1.3}$$

Then the (p,h)-autoregression appearing in (2.2.5) has the equivalent representation

$$\boldsymbol{W}_h(h) = \overline{\boldsymbol{W}}_p(h)\boldsymbol{B}(h) + \boldsymbol{U}_h(h). \tag{A.1.4}$$

The estimator $\hat{\boldsymbol{B}}(h) = [\overline{\boldsymbol{W}}_p(h)'\overline{\boldsymbol{W}}_p(h)]^{-1}\overline{\boldsymbol{W}}_p(h)'\boldsymbol{W}_h(h)$ then follows.

Asymptotic Variance Components: Covariance Matrices

We now derive the components $\boldsymbol{\Gamma}_{p,0}$ and $\boldsymbol{D}_p(h)$ of the asymptotic variance $\boldsymbol{\Sigma}_p$. First, let $\boldsymbol{\Gamma}_{p,0}$ denote the variance matrices for $\boldsymbol{W}(\tau_L, p)$ in (A.1.1):

$$\boldsymbol{\Gamma}_{p,0} \equiv E\left[\boldsymbol{W}(\tau_L, p)\boldsymbol{W}(\tau_L, p)'\right].$$

By Assumptions 2.2.1-2.2.2 it is easily verified that $\boldsymbol{\Gamma}_{p,0}$ is positive definite. Second, by a standard first order expansion we require the long-run variance of a vectorized $\boldsymbol{W}(\tau_L, p)\boldsymbol{u}(\tau_L + h)'$, denoted

$$\boldsymbol{Y}(\tau_L+h, p) \equiv \text{vec}\left[\boldsymbol{W}(\tau_L, p)\boldsymbol{u}^{(h)}(\tau_L + h)'\right] = \left(\boldsymbol{I}_K \otimes \boldsymbol{W}(\tau_L, p)\right)\boldsymbol{u}^{(h)}(\tau_L+h) \in \mathcal{R}^{pK^2\times 1}. \tag{A.1.5}$$

Under Assumption 2.2.1 $\boldsymbol{\epsilon}(\tau_L)$ is a stationary mds with respect to $\mathcal{F}_{\tau_L}$, where

$$E\left[\boldsymbol{\epsilon}(\tau_L)\boldsymbol{\epsilon}(\tau_L)'\right] \equiv \boldsymbol{\Omega} \text{ is positive definite.}$$

Trivially, therefore, $\boldsymbol{\epsilon}(\tau_L)$ has a continuous, bounded and positive spectral density. Hence by stationarity Assumption 2.2.2, $\boldsymbol{X}(\tau_L)$ has a continuous, bounded and everywhere positive spectral density. Therefore $\{\boldsymbol{Y}(\tau_L+h, p)\}_{\tau_L}$ is a zero mean L_2-bounded stationary process with continuous, everywhere positive spectrum, and therefore auto-covariances

$$\boldsymbol{\Delta}_{p,s}(h) \equiv E\left[\boldsymbol{Y}(\tau_L + h + s, p)\boldsymbol{Y}(\tau_L + h, p)'\right]$$

that satisfy

$$\boldsymbol{\Delta}_{p,0}(h) \text{ is positive definite} \quad \forall h \geq 0$$
$$\boldsymbol{\Delta}_{p,s}(h) = \mathbf{0}_{pK^2 \times pK^2} \quad \forall s \geq h.$$

Analytical characterizations of $\boldsymbol{\Gamma}_{p,0}$ and $\boldsymbol{\Delta}_{p,s}(h)$, and a proof that $\boldsymbol{\Delta}_{p,s}(h) = \mathbf{0}_{pK^2 \times pK^2}$ $\forall s \geq h$ are presented below. The partial sum variance of $\boldsymbol{Y}(\tau_L + h, p)$ is therefore:

$$\begin{aligned} \boldsymbol{D}_{p,T_L^*}(h) &\equiv \text{Var}\left[\frac{1}{\sqrt{T_L^*}} \sum_{\tau_L=0}^{T_L^*-1} \boldsymbol{Y}(\tau_L + h, p)\right] \\ &= \boldsymbol{\Delta}_{p,0}(h) + \sum_{s=1}^{h-1}\left[1 - \frac{s}{T_L^*}\right] \times \left[\boldsymbol{\Delta}_{p,s}(h) + \boldsymbol{\Delta}_{p,s}(h)'\right] \\ &= \boldsymbol{\Delta}_{p,0}(1) \text{ if } h = 1, \end{aligned} \tag{A.1.6}$$

where $T_L^* = T_L - h + 1$. We define $\boldsymbol{D}_p(h)$ as the long-run variance of $\boldsymbol{Y}(\tau_L + h, p)$:

$$\begin{aligned} \boldsymbol{D}_p(h) &\equiv \lim_{T_L^* \to \infty} \boldsymbol{D}_{p,T_L^*}(h) = \boldsymbol{\Delta}_{p,0}(h) + \sum_{s=1}^{h-1}\left[\boldsymbol{\Delta}_{p,s}(h) + \boldsymbol{\Delta}_{p,s}(h)'\right] \\ &= \boldsymbol{\Delta}_{p,0}(1) \text{ if } h = 1. \end{aligned} \tag{A.1.7}$$

Observe that $\boldsymbol{D}_{p,T_L^*}(h)$ for T_L^* sufficiently large is positive definite, hence $\boldsymbol{D}_p(h)$ is positive definite. Simply note that by stationarity and spectral density positiveness for $\boldsymbol{X}(\tau_L)$, it follows $\boldsymbol{a}'\boldsymbol{Y}(\tau_L + h, p)\boldsymbol{a}$ is for any conformable $\boldsymbol{a} \neq \mathbf{0}$, $\boldsymbol{a}'\boldsymbol{a} = 1$, stationary and has a continuous, bounded everywhere positive spectral density $f_a(\lambda)$. Therefore $\boldsymbol{a}'\boldsymbol{D}_{p,T_L^*}(h)\boldsymbol{a} = 2\pi f_a(0) + o(1) > 0$ for T_L^* sufficiently large (see eq. (1.7) in Ibragimov (1962)).

We now explicitly characterize the covariance matrices $\boldsymbol{\Gamma}_{p,0} \equiv E[\boldsymbol{W}(\tau_L, p)\boldsymbol{W}(\tau_L, p)']$ and $\boldsymbol{\Delta}_{p,s}(h) \equiv E[\boldsymbol{Y}(\tau_L + h + s, p)\boldsymbol{Y}(\tau_L + h, p)']$. Denote the auto-covariances of $\boldsymbol{X}(\tau_L)$ as

$$\boldsymbol{\Upsilon}_s = [\upsilon_{ij,s}]_{i,j=1}^K \equiv E\left[\boldsymbol{X}(\tau_L + s)\boldsymbol{X}(\tau_L)'\right] = \begin{cases} \sum_{k=0}^{\infty} \boldsymbol{\Psi}_{s+k}\boldsymbol{\Omega}\boldsymbol{\Psi}_k' & \text{if } s \geq 0 \\ \boldsymbol{\Upsilon}_{-s}' & \text{if } s < 0, \end{cases} \tag{A.1.8}$$

where $\boldsymbol{\Psi}_k$ is defined by the moving average representation (2.2.4). In view of $|\boldsymbol{\Psi}_k| = O(\rho^h)$ for $\rho \in (0,1)$, and $||\boldsymbol{H}(\tau_L)||_{2+\delta} \in (0,\infty)$, it follows $||\boldsymbol{\Omega}|| < \infty$ and therefore $\sum_{s=-\infty}^{\infty} |\upsilon_{ij,s}| < \infty$ for any i, j. The process $\{\boldsymbol{W}(\tau_L, p)\}_{\tau_L}$ defined by (A.1.1) therefore has auto-covariances

$$\boldsymbol{\Gamma}_{p,s} \equiv E\left[\boldsymbol{W}(\tau_L + s, p)\boldsymbol{W}(\tau_L, p)'\right] = \begin{bmatrix} \boldsymbol{\Upsilon}_s & \boldsymbol{\Upsilon}_{s+1} & \cdots & \boldsymbol{\Upsilon}_{s+p-1} \\ \boldsymbol{\Upsilon}_{s-1} & \boldsymbol{\Upsilon}_s & \cdots & \boldsymbol{\Upsilon}_{s+p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Upsilon}_{s-p+1} & \boldsymbol{\Upsilon}_{s-p+2} & \cdots & \boldsymbol{\Upsilon}_s \end{bmatrix}. \tag{A.1.9}$$

Further, $\boldsymbol{u}^{(h)}(\tau_L)$ has auto-covariances

$$\boldsymbol{Q}_s(h) \equiv E\left[\boldsymbol{u}^{(h)}(\tau_L + s)\boldsymbol{u}^{(h)}(\tau_L)'\right] = \begin{cases} \sum_{k=0}^{h-s-1} \boldsymbol{\Psi}_{s+k}\boldsymbol{\Omega}\boldsymbol{\Psi}_k' & \text{if } 0 \le s < h \\ \boldsymbol{Q}_{-s}(h)' & \text{if } -h < s < 0 \\ \boldsymbol{0}_{K\times K} & \text{if } |s| \ge h. \end{cases} \tag{A.1.10}$$

Using (A.1.10) and $\boldsymbol{Y}(\tau_L + h, p) \equiv (\boldsymbol{I}_K \otimes \boldsymbol{W}(\tau_L, p))\boldsymbol{u}^{(h)}(\tau_L + h)$, the auto-covariances $\boldsymbol{\Delta}_{p,s}(h)$ of $\boldsymbol{Y}(\tau_L + h, p)$ can now be fully characterized:

$$\boldsymbol{\Delta}_{p,s}(h) \equiv E[\boldsymbol{Y}(\tau_L + h + s, p)\boldsymbol{Y}(\tau_L + h, p)'] = \begin{cases} \boldsymbol{Q}_0(h) \otimes \boldsymbol{\Gamma}_{p,0} & \text{if } s = 0 \\ \boldsymbol{\Delta}_{p,-s}(h)' & \text{if } -h < s < 0 \\ \boldsymbol{0}_{pK^2\times pK^2} & \text{if } |s| \ge h. \end{cases} \tag{A.1.11}$$

Note that $\boldsymbol{Y}(\tau_L + h, p)$ is serially uncorrelated at lag $|s| \ge h$, although in general we cannot say $\boldsymbol{Y}(\tau_L + h, p)$ is $h-1$ dependent. Evidently a convenient expression for $\boldsymbol{\Delta}_{p,s}(h)$ does not exist when $s \in \{1, \ldots, h-1\}$.

We now prove $\boldsymbol{\Delta}_{p,s}(h) = \boldsymbol{0}_{pK^2\times pK^2}$ for $|s| \ge h$. Assume without loss of generality that $s \ge h$. Equation (A.1.5) and the definition of $\boldsymbol{\Delta}_{p,s}(h)$ imply that

$$\boldsymbol{\Delta}_{p,s}(h) = E\left[(\boldsymbol{I}_K \otimes \boldsymbol{W}(\tau_L + s, p))\, \boldsymbol{u}^{(h)}(\tau_L + s + h)\boldsymbol{u}^{(h)}(\tau_L + h)'\, (\boldsymbol{I}_K \otimes \boldsymbol{W}(\tau_L, p)')\right]. \tag{A.1.12}$$

Let $I(\tau_L + s) = \sigma\{\boldsymbol{\epsilon}(\tau) | \tau \le \tau_L + s\}$. Note that $\boldsymbol{W}(\tau_L, p)$, $\boldsymbol{W}(\tau_L + s, p)$, and $\boldsymbol{u}^{(h)}(\tau_L + h)$ are all

known at period τ_L+s, while $\boldsymbol{u}^{(h)}(\tau_L+s+h)$ depends only on $\{\boldsymbol{\epsilon}(\tau_L+s+1),\ldots,\boldsymbol{\epsilon}(\tau_L+s+h)\}$ and therefore $E[\boldsymbol{u}^{(h)}(\tau_L+s+h)|I(\tau_L+s)] = \sum_{k=0}^{h-1}\boldsymbol{\Psi}_k E[\boldsymbol{\epsilon}(\tau_L+s+h-k)|I(\tau_L+s)] = \mathbf{0}$ by the mds Assumption 2.2.1. We can thus get the desired result by applying the law of iterated expectations to (A.1.12). Similarly, $\boldsymbol{\Delta}_{p,0}(h) = \boldsymbol{Q}_0(h)\otimes\boldsymbol{\Gamma}_{p,0}$ can be shown by applying the law of iterated expectations given $I(\tau_L)$ to (A.1.12).

Example 5 ($h=1$): It is useful to derive the least squares asymptotic variance $\boldsymbol{\Sigma}_p = (\boldsymbol{I}_K\otimes\boldsymbol{\Gamma}_{p,0}^{-1})\boldsymbol{D}_p(h)(\boldsymbol{I}_K\otimes\boldsymbol{\Gamma}_{p,0}^{-1})'$ for the case $h=1$. Use (A.1.8) and (A.1.9) to deduce $\boldsymbol{\Gamma}_{p,0} = \boldsymbol{\Upsilon}_0 = \sum_{k=0}^{\infty}\boldsymbol{\Psi}_k\boldsymbol{\Omega}\boldsymbol{\Psi}_k'$. Next, use (A.1.7) and (A.1.11) to deduce $\boldsymbol{D}_p(1) = \boldsymbol{\Delta}_{p,0}(1) = \boldsymbol{Q}_0(1)\otimes\boldsymbol{\Gamma}_{p,0}$, hence by (A.1.9) and (A.1.10) it follows $\boldsymbol{D}_p(1) = \boldsymbol{\Omega}\otimes\boldsymbol{\Gamma}_{p,0} = \boldsymbol{\Omega}\otimes\sum_{k=0}^{\infty}\boldsymbol{\Psi}_k\boldsymbol{\Omega}\boldsymbol{\Psi}_k'$. Kronecker product properties therefore imply $\boldsymbol{\Sigma}_p$ is identically $\boldsymbol{\Omega}\otimes\boldsymbol{\Gamma}_{p,0}^{-1} = \boldsymbol{\Omega}\otimes\boldsymbol{\Upsilon}_0^{-1} = \boldsymbol{\Omega}\otimes(\sum_{k=0}^{\infty}\boldsymbol{\Psi}_k\boldsymbol{\Omega}\boldsymbol{\Psi}_k')^{-1}$.

Consistent and Almost Surely Positive Semi-Definite HAC Estimator

We need only estimate the components of $\boldsymbol{\Sigma}_p = (\boldsymbol{I}_K\otimes\boldsymbol{\Gamma}_{p,0}^{-1})\boldsymbol{D}_p(h)(\boldsymbol{I}_K\otimes\boldsymbol{\Gamma}_{p,0}^{-1})'$. A natural estimator of $\boldsymbol{\Gamma}_{p,0}$ is the sample conjugate:

$$\hat{\boldsymbol{\Gamma}}_{p,0} = \frac{1}{T_L^*}\sum_{\tau_L=0}^{T_L^*-1}\boldsymbol{W}(\tau_L,p)\boldsymbol{W}(\tau_L,p)'.$$

Under Assumptions 2.2.1-2.2.2 $\hat{\boldsymbol{\Gamma}}_{p,0}$ is *almost surely* positive definite.

Turning to the long-run variance $\boldsymbol{D}_p(h)$, denote the least squares residual $\hat{\boldsymbol{U}}_h(h) \equiv \boldsymbol{W}_h(h) - \overline{\boldsymbol{W}}_p(h)\hat{\boldsymbol{B}}(h)$ for model (A.1.4) and the resulting residual $\hat{\boldsymbol{u}}^{(h)}(\tau_L) \equiv \boldsymbol{X}(\tau_L) - \sum_{k=1}^{p}\hat{\boldsymbol{A}}_k^{(h)}\boldsymbol{X}(\tau_L - h + 1 - k)$ computed from (A.1.3). Now compute the sample version of $\boldsymbol{Y}(\tau_L+h,p)$ defined in (A.1.5),

$$\hat{\boldsymbol{Y}}(\tau_L+h,p) = \text{vec}\left[\boldsymbol{W}(\tau_L,p)\hat{\boldsymbol{u}}^{(h)}(\tau_L+h)'\right],$$

and compute

$$\hat{\boldsymbol{\Delta}}_{p,s}(h) = \frac{1}{T_L^*}\sum_{\tau_L=s}^{T_L^*-1}\hat{\boldsymbol{Y}}(\tau_L+h,p)\hat{\boldsymbol{Y}}(\tau_L+h-s,p)'.$$

If $h=1$ then from (A.1.6) the estimator of $\boldsymbol{D}_p(h)$ need only be $\hat{\boldsymbol{D}}_{p,T_L^*}(1) = \hat{\boldsymbol{\Delta}}_{p,0}(1)$. Otherwise, a naïve estimator of $\boldsymbol{D}_p(h)$ simply substitutes $\hat{\boldsymbol{\Delta}}_{p,s}(h)$ for $\boldsymbol{\Delta}_{p,s}(h)$ in the right-hand side of (A.1.6), but it is well-known that such an estimator may not be positive semi-definite unless

$h = 1$.

We therefore exploit Newey and West's (1987) Bartlett kernel-based HAC estimator which ensures *almost sure* positive semi-definiteness for any $T_L^* \geq 1$ (see Newey and West (1987) and Andrews (1991)):[1]

$$\hat{\boldsymbol{D}}_{p,T_L^*}(h) = \hat{\boldsymbol{\Delta}}_{p,0}(h) + \sum_{s=1}^{n_{T_L^*}-1} \left(1 - \frac{s}{n_{T_L^*}}\right) (\hat{\boldsymbol{\Delta}}_{p,s}(h) + \hat{\boldsymbol{\Delta}}_{p,s}(h)') \tag{A.1.13}$$

with bandwidth $n_{T_L^*}$: $h \leq n_{T_L^*} \leq T_L^*$, $n_{T_L^*} \to \infty$ and $n_{T_L^*} = o(T_L^*)$. Intuitively since $\boldsymbol{Y}(\tau_L, p)$ is serially uncorrelated for all lags above $h-1$, and $\hat{\boldsymbol{\Delta}}_{p,s}(h) = 1/T_L^* \sum_{\tau_L=s}^{T_L^*-1} \boldsymbol{Y}(\tau_L + h, p)\boldsymbol{Y}(\tau_L + h - s, p)' + o_p(1)$ is easily verified, we only need $h-1$ lags, that is $\hat{\boldsymbol{\Delta}}_{p,0}(h) + \sum_{s=1}^{h-1}(1 - s/n_{T_L^*})(\hat{\boldsymbol{\Delta}}_{p,s}(h) + \hat{\boldsymbol{\Delta}}_{p,s}(h)')$ is a valid estimator in place of (A.1.13). But this estimator also need not be positive semi-definite in small samples.

Our proposed estimator of $\boldsymbol{\Sigma}_p$ is therefore

$$\hat{\boldsymbol{\Sigma}}_p = \left(\boldsymbol{I}_K \otimes \hat{\boldsymbol{\Gamma}}_{p,0}^{-1}\right) \times \hat{\boldsymbol{D}}_{p,T_L^*}(h) \times \left(\boldsymbol{I}_K \otimes \hat{\boldsymbol{\Gamma}}_{p,0}^{-1}\right). \tag{A.1.14}$$

Almost sure positive definiteness of $\hat{\boldsymbol{\Gamma}}_{p,0}$ and positive semi-definiteness of $\hat{\boldsymbol{D}}_{p,T_L^*}(h)$ imply $\hat{\boldsymbol{\Sigma}}_p$ is *almost surely* positive semi-definite. Consistency can be shown given stronger moment and mixing conditions.

Assumption A.1.1. For some $\delta > 0$ let $||\boldsymbol{\epsilon}(\tau_L)||^{4+\delta} < \infty$ and the mixing coefficients α_h of $\boldsymbol{X}(\tau_L)$ satisfy $\alpha_h = O(h^{-(4+\delta)\backslash\delta})$.

Lemma A.1.1. Under Assumptions 2.2.1-2.2.2 and A.1.1 $\hat{\boldsymbol{\Sigma}}_p$ is *almost surely* positive semi-definite for any $T_L^* \geq 1$, and $\hat{\boldsymbol{\Sigma}}_p \xrightarrow{p} \boldsymbol{\Sigma}_p$ where $\boldsymbol{\Sigma}_p$ is positive definite.

Proof. *Almost sure* positive semi-definiteness of $\hat{\boldsymbol{\Sigma}}_p$ follows from *almost sure* positive definiteness of $\hat{\boldsymbol{\Gamma}}_{p,0}$ under Assumptions 2.2.1 - 2.2.2, and *almost sure* positive semi-definiteness of $\hat{\boldsymbol{D}}_{p,T_L^*}(h)$ by Theorem 1 in Newey and West (1987). Further $\hat{\boldsymbol{\Gamma}}_{p,0} \xrightarrow{p} \boldsymbol{\Gamma}_{p,0}$ by the ergodic theorem given stationarity, ergodicity, and square integrability under Assumptions 2.2.1-2.2.3.

[1]There is a large choice of valid kernels, including Parzen and Tukey-Hanning. See de Jong and Davidson (2000).

Since $\boldsymbol{\Sigma}_p = (\boldsymbol{I}_K \otimes \boldsymbol{\Gamma}_{p,0}^{-1}) \times \boldsymbol{D}_p(h) \times (\boldsymbol{I}_K \otimes \boldsymbol{\Gamma}_{p,0}^{-1})$ it therefore suffices to show $\hat{\boldsymbol{D}}_{p,T_L^*}(h) \xrightarrow{p} \boldsymbol{D}_p(h)$. The latter follows from Theorem 2.2 in de Jong and Davidson (2000) [JD] if we verify their Assumptions 1-4.

First, the Bartlett kernel satisfies JD's Assumption 1. Second, JD's Assumptions 2 and 3 hold since $n_{T_L^*} \to \infty$ as $T_L^* \to \infty$, $n_{T_L^*} = o(T_L^*)$, and by Assumptions 2.2.1-2.2.3 and A.1.1 and given measurability, $\{1/\sqrt{T_L^*}\boldsymbol{Y}(\tau_L + h, p) : 1 \leq \tau_L \leq T_L^*\}_{T_L^* \geq 1}$ forms an $L_{2+\delta}$-bounded α-mixing triangular array with coefficients $\alpha_h = O(h^{(4+\delta)\backslash\delta})$.[2]

Finally, in order to verify JD's Assumption 4, define the regression error function $\boldsymbol{u}^{(h)}(\tau_L, \tilde{\boldsymbol{B}}) \equiv \boldsymbol{X}(\tau_L) - \sum_{k=1}^{p} \tilde{\boldsymbol{A}}_k \boldsymbol{X}(\tau_L - h + 1 - k)$ for any conformable $\tilde{\boldsymbol{A}}_k$ where $\tilde{\boldsymbol{B}} \equiv [\tilde{\boldsymbol{A}}_1, \ldots, \tilde{\boldsymbol{A}}_p]'$, and $\boldsymbol{Y}(\tau_L + h, p, \tilde{\boldsymbol{B}}) = (\boldsymbol{I}_K \otimes \boldsymbol{W}(\tau_L, p))\boldsymbol{u}^{(h)}(\tau_L + h, \tilde{\boldsymbol{B}})$. Now define $\boldsymbol{Z}(\tau_L + h, p, \tilde{\boldsymbol{B}}) \equiv \boldsymbol{Y}(\tau_L + h, p, \tilde{\boldsymbol{B}})/\sqrt{T_L^*}$ and note $\hat{\boldsymbol{Y}}(\tau_L + h, p) = \boldsymbol{Y}(\tau_L + h, p, \hat{\boldsymbol{B}}(h))$. In order to match JD's standardization, we work with $\boldsymbol{Z}(\tau_L + h, p, \tilde{\boldsymbol{B}})$. Assumption 4 consists of three parts, (a)-(c), with a scale factor κ_n that is simply $\boldsymbol{I}_{pK^2}$ in our case. Part (a) applies since $\hat{\boldsymbol{B}}(h)$ is $\sqrt{T_L^*}$-convergent by Theorem 2.2.1. Next, (b) applies since under our assumptions and by model linearity $1/\sqrt{T_L^*}\sum_{\tau_L=0}^{T_L^*-1} E[(\partial/\partial\tilde{\boldsymbol{B}})\boldsymbol{Z}(\tau_L + h, p, \tilde{\boldsymbol{B}})]$ is trivially continuous at $\boldsymbol{B}(h)$ uniformly in T_L^*. Finally, (c) involves a uniform LLN for $\boldsymbol{DZ}(\tau_L + h, p, \tilde{\boldsymbol{B}}) \equiv (\partial/\partial\tilde{\boldsymbol{B}})\boldsymbol{Z}(\tau_L + h, p, \tilde{\boldsymbol{B}})$. The latter is not a function of $\tilde{\boldsymbol{B}}$ in view of linearity (i.e. $\boldsymbol{DZ}(\tau_L + h, p, \tilde{\boldsymbol{B}}) = \boldsymbol{DZ}(\tau_L + h, p)$), hence a uniform LLN reduces to a pointwise LLN which holds by the ergodic theorem given stationarity, ergodicity, and integrability of $\boldsymbol{DZ}(\tau_L + h, p)$ which follows from stationarity and square integrability of $\boldsymbol{\epsilon}(\tau_L)$. $\mathcal{QED}$.

A.1.2 Proof of Theorems 2.2.1 and 2.2.2

Recall $\boldsymbol{D}_{p,T_L^*}(h) \equiv \text{Var}[1/\sqrt{T_L^*}\sum_{\tau_L=0}^{T_L^*-1} \boldsymbol{Y}(\tau_L + h, p)]$ in (A.1.6) and $\boldsymbol{D}_p(h) \equiv \lim_{T_L^* \to \infty} \boldsymbol{D}_{p,T_L^*}(h)$. The proof of Theorem 2.2.1 exploits the following central limit theorem.

Lemma A.1.2. Under Assumptions 2.2.1-2.2.3 $1/\sqrt{T_L^*}\sum_{\tau_L=0}^{T_L^*-1} \boldsymbol{Y}(\tau_L + h, p) \xrightarrow{d} N(\boldsymbol{0}_{pK^2 \times 1}, \boldsymbol{D}_p(h))$ where $\boldsymbol{D}_p(h)$ is positive definite.

Proof. By the Cramér-Wold theorem it is necessary and sufficient to show $1/\sqrt{T_L^*}\sum_{\tau_L=0}^{T_L^*-1}$

[2] See Chapter 17 in Davidson (1994) for verification that geometric strong mixing satisfies the Near Epoch Dependence property in de Jong and Davidson's (2000) Assumption 2.

$\boldsymbol{a}'\boldsymbol{Y}(\tau_L+h,p)\xrightarrow{d} N(0,\boldsymbol{a}'\boldsymbol{D}_p(h)\boldsymbol{a})$ for any conformable $\boldsymbol{a}$, $\boldsymbol{a}'\boldsymbol{a}=1$. By construction, measurability and Assumptions 2.2.1-2.2.3 it follows $\{\boldsymbol{a}'\boldsymbol{Y}(\tau_L+h,p)\}_{\tau_L}$ is a zero mean, $L_{2+\delta}$-bounded α-mixing process with coefficients that satisfy $\sum_{h=0}^{\infty}\alpha_{2^h}<\infty$. Further, by the discussion following (A.1.7) both $\boldsymbol{D}_{p,T_L^*}(h)$ for sufficiently large T_L^* and $\boldsymbol{D}_p(h)\equiv\lim_{T_L^*\to\infty}\boldsymbol{D}_{p,T_L^*}(h)$ are positive definite. Therefore $1/\sqrt{T_L^*}\sum_{\tau_L=0}^{T_L^*-1}\boldsymbol{a}'\boldsymbol{Y}(\tau_L+h,p)/(\boldsymbol{a}'\boldsymbol{D}_{p,T_L^*}(h)\boldsymbol{a})\xrightarrow{d}N(0,1)$ by Theorem 2.2 in Ibragimov (1975). Since $\boldsymbol{a}'\boldsymbol{D}_{p,T_L^*}(h)\boldsymbol{a}\to\boldsymbol{a}'\boldsymbol{D}_p(h)\boldsymbol{a}>0$ the claim now follows from Cramér's Theorem. $\mathcal{QED}$.

We now prove Theorems 2.2.1 and 2.2.2. By the construction of $\hat{\boldsymbol{B}}(h)$, $\hat{\boldsymbol{\Gamma}}_{p,0}$ and $\boldsymbol{Y}(\tau_L+h,p)$ it follows

$$\begin{aligned}\sqrt{T_L^*}\text{vec}\left[\hat{\boldsymbol{B}}(h)-\boldsymbol{B}(h)\right]&=\sqrt{T_L^*}\text{vec}\left[\left(\overline{\boldsymbol{W}}_p(h)'\overline{\boldsymbol{W}}_p(h)\right)^{-1}\overline{\boldsymbol{W}}_p(h)'\boldsymbol{U}_h(h)\right]\\&=\left[\boldsymbol{I}_K\otimes\left(\frac{1}{T_L^*}\overline{\boldsymbol{W}}_p(h)'\overline{\boldsymbol{W}}_p(h)\right)^{-1}\right]\times\text{vec}\left[\frac{1}{\sqrt{T_L^*}}\overline{\boldsymbol{W}}_p(h)'\boldsymbol{U}_h(h)\right]\\&=\left[\boldsymbol{I}_K\otimes\hat{\boldsymbol{\Gamma}}_{p,0}^{-1}\right]\times\frac{1}{\sqrt{T_L^*}}\sum_{\tau_L=0}^{T_L^*-1}\boldsymbol{Y}(\tau_L+h,p).\end{aligned}$$

Note that $\hat{\boldsymbol{\Gamma}}_{p,0}=1/T_L^*\sum_{\tau_L=0}^{T_L^*-1}\boldsymbol{W}(\tau_L,p)\boldsymbol{W}(\tau_L,p)'\xrightarrow{p}E[\boldsymbol{W}(\tau_L,p)\boldsymbol{W}(\tau_L,p)']=\boldsymbol{\Gamma}_{p,0}$ in view of stationarity, ergodicity and square integrability of $\boldsymbol{W}(\tau_L,p)$. Further, $\boldsymbol{D}_{p,T_L^*}(h)=\text{Var}[1/\sqrt{T_L^*}\sum_{\tau_L=0}^{T_L^*-1}\boldsymbol{Y}(\tau_L+h,p)]\to\boldsymbol{D}_p(h)$. Now use

$$\boldsymbol{\Sigma}_p\equiv(\boldsymbol{I}_K\otimes\boldsymbol{\Gamma}_{p,0}^{-1})\times\boldsymbol{D}_p(h)\times(\boldsymbol{I}_K\otimes\boldsymbol{\Gamma}_{p,0}^{-1})',$$

combined with Lemma A.1.2, and Slutsky's and Cramér's Theorems to deduce $\sqrt{T_L^*}\text{vec}[\hat{\boldsymbol{B}}(h)-\boldsymbol{B}(h)]\xrightarrow{d}N(\boldsymbol{0}_{pK^2\times1},\boldsymbol{\Sigma}_p)$. Finally, $\boldsymbol{\Sigma}_p$ is positive definite given the positive definiteness of $\boldsymbol{\Gamma}_{p,0}$ and $\boldsymbol{D}_p(h)$ as discussed in Appendix A.1.1. This proves Theorem 2.2.1.

The proof of Theorem 2.2.2 follows instantly from Theorem 2.2.1, the assumption $\hat{\boldsymbol{\Sigma}}_p\xrightarrow{p}\boldsymbol{\Sigma}_p$, and the mapping theorem.

A.2 Proof of Theorem 2.4.1

In view of Theorem 1 in Lütkepohl (1984) it suffices to show that $\boldsymbol{X}(\tau_L)$ and $\underline{\boldsymbol{X}}(\tau_L)$ are linear transformations of a VAR process. Lütkepohl (1984) defines a VAR process as having a vector white noise error term, hence any subsequent VAR process need only have a second order stationary and serially uncorrelated error. Define $mK^* \times 1$ vectors:

$$\overline{\boldsymbol{X}}(\tau_L) = [\boldsymbol{X}(\tau_L, 1)', \ldots, \boldsymbol{X}(\tau_L, m)']' \quad \text{and} \quad \overline{\boldsymbol{\eta}}(\tau_L) = [\boldsymbol{\eta}(\tau_L, 1)', \ldots, \boldsymbol{\eta}(\tau_L, m)']'.$$

We first show that $\{\overline{\boldsymbol{X}}(\tau_L)\}$ follows a VAR(s) process with $s = \lceil p/m \rceil$, the smallest integer not smaller than p/m. We then prove the claim.

The HF-VAR(p) process in (2.4.1) satisfies:

$$\boldsymbol{N}\overline{\boldsymbol{X}}(\tau_L) = \sum_{k=1}^{s} \boldsymbol{M}_k \overline{\boldsymbol{X}}(\tau_L - k) + \overline{\boldsymbol{\eta}}(\tau_L), \tag{A.2.1}$$

where

$$\boldsymbol{N} = \begin{bmatrix} \boldsymbol{I}_{K^*} & \boldsymbol{0}_{K^* \times K^*} & \ldots & \boldsymbol{0}_{K^* \times K^*} \\ -\boldsymbol{\Phi}_1 & \boldsymbol{I}_{K^*} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \boldsymbol{0}_{K^* \times K^*} \\ -\boldsymbol{\Phi}_{m-1} & \ldots & -\boldsymbol{\Phi}_1 & \boldsymbol{I}_{K^*} \end{bmatrix} \text{ and } \boldsymbol{M}_k = \begin{bmatrix} \boldsymbol{\Phi}_{km} & \boldsymbol{\Phi}_{km-1} & \ldots & \boldsymbol{\Phi}_{(k-1)m+1} \\ \boldsymbol{\Phi}_{km+1} & \boldsymbol{\Phi}_{km} & \ldots & \boldsymbol{\Phi}_{(k-1)m+2} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Phi}_{(k+1)m-1} & \boldsymbol{\Phi}_{(k+1)m-2} & \ldots & \boldsymbol{\Phi}_{km} \end{bmatrix}$$

for $k = 1, \ldots, s$. It is understood that $\boldsymbol{\Phi}_k = \boldsymbol{0}_{K^* \times K^*}$ whenever $k > p$. We have:

$$\boldsymbol{N}^{-1} = \begin{bmatrix} \boldsymbol{N}_1 & \boldsymbol{0}_{K^* \times K^*} & \ldots & \boldsymbol{0}_{K^* \times K^*} \\ \boldsymbol{N}_2 & \boldsymbol{N}_1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \boldsymbol{0}_{K^* \times K^*} \\ \boldsymbol{N}_m & \ldots & \boldsymbol{N}_2 & \boldsymbol{N}_1 \end{bmatrix},$$

where $\boldsymbol{N}_1 = \boldsymbol{I}_{K^*}$ and $\boldsymbol{N}_k = \sum_{l=1}^{k-1} \boldsymbol{\Phi}_{k-l} \boldsymbol{N}_l$ for $k = 2, \ldots, m$. Using this property, (A.2.1) can be rewritten as:

$$\overline{\mathcal{A}}(\mathcal{L}_L)\overline{\boldsymbol{X}}(\tau_L) = \overline{\boldsymbol{\epsilon}}(\tau_L),$$

where $\mathcal{L}_L$ is the low frequency lag operator, $\overline{\mathcal{A}}(\mathcal{L}_L) = \boldsymbol{I}_{mK^*} - \sum_{k=1}^{s} \overline{\boldsymbol{A}}_k \mathcal{L}_L^k$, $\overline{\boldsymbol{A}}_k = \boldsymbol{N}^{-1}\boldsymbol{M}_k$,

and $\overline{\boldsymbol{\epsilon}}(\tau_L) = \boldsymbol{N}^{-1}\overline{\boldsymbol{\eta}}(\tau_L)$ is second order stationary and serially uncorrelated by the stationary martingale difference property of $\overline{\boldsymbol{\eta}}(\tau_L)$. Hence, $\{\overline{\boldsymbol{X}}(\tau_L)\}$ follows a VAR(s) process.

Now consider $\boldsymbol{X}(\tau_L)$ and $\underline{\boldsymbol{X}}(\tau_L)$. Recall the generic aggregation schemes (2.2.1) detailed in Section 2.2 with selection vector $\boldsymbol{w}$. Define $\boldsymbol{H} = [\boldsymbol{I}_{K_H}, \boldsymbol{0}_{K_H \times K_L}]$, $\boldsymbol{L} = [\boldsymbol{0}_{K_L \times K_H}, \boldsymbol{I}_{K_L}]$, $\boldsymbol{F}_{H\to M} = [\boldsymbol{I}_m \otimes \boldsymbol{H}', \boldsymbol{w} \otimes \boldsymbol{L}']'$, and

$$\boldsymbol{F}_{M\to L} = \begin{bmatrix} \boldsymbol{w}' \otimes \boldsymbol{I}_{K_H} & \boldsymbol{0}_{K_H \times K_L} \\ \boldsymbol{0}_{K_L \times mK_H} & \boldsymbol{I}_{K_L} \end{bmatrix}.$$

Observe that $\boldsymbol{X}(\tau_L)$ and $\underline{\boldsymbol{X}}(\tau_L)$ are finite order linear transformations of $\overline{\boldsymbol{X}}(\tau_L)$: $\boldsymbol{X}(\tau_L) = \boldsymbol{F}_{H\to M}\overline{\boldsymbol{X}}(\tau_L)$ and $\underline{\boldsymbol{X}}(\tau_L) = \boldsymbol{F}_{H\to L}\overline{\boldsymbol{X}}(\tau_L)$, where $\boldsymbol{F}_{H\to L} = \boldsymbol{F}_{M\to L}\boldsymbol{F}_{H\to M} = [\boldsymbol{w} \otimes \boldsymbol{H}', \boldsymbol{w} \otimes \boldsymbol{L}']'$. Moreover, in view of the transformation being a finite order, if $\overline{\boldsymbol{X}}(\tau_L)$ is stationary then so are $\boldsymbol{X}(\tau_L)$ and $\underline{\boldsymbol{X}}(\tau_L)$.

A.3 Proof of Theorem 2.4.2

We prove only part (ii) since part (i) is similar or even simpler. Recall that the high frequency reference information set at time t is expressed as $\overline{\mathcal{I}}(t)$ and the mapping between single time index t and double time indices (τ_L, k) is that $t = m(\tau_L - 1) + k$. Also recall our notation that $\tilde{\boldsymbol{x}}_H(\tau_L) = [\tilde{\boldsymbol{x}}_{H,1}(\tau_L)', \ldots, \tilde{\boldsymbol{x}}_{H,K_H}(\tau_L)']'$ and $\tilde{\boldsymbol{x}}_{H,i}(\tau_L) = [x_{H,i}(\tau_L, 1), \ldots, x_{H,i}(\tau_L, m)]'$. We have that:

$$\begin{aligned} P[\tilde{\boldsymbol{x}}_H(\tau_L + 1) \,|\, \mathcal{I}(\tau_L)] &= P\left[P[\tilde{\boldsymbol{x}}_H(\tau_L + 1) \,|\, \overline{\mathcal{I}}(m\tau_L)] \,|\, \mathcal{I}(\tau_L)\right] \\ &= P\left[P[\tilde{\boldsymbol{x}}_H(\tau_L + 1) \,|\, \overline{\mathcal{I}}_{(L)}(m\tau_L)] \,|\, \mathcal{I}(\tau_L)\right] \\ &= P\left[P[\tilde{\boldsymbol{x}}_H(\tau_L + 1) \,|\, \mathcal{I}_{(L)}(\tau_L)] \,|\, \mathcal{I}(\tau_L)\right] \\ &= P[\tilde{\boldsymbol{x}}_H(\tau_L + 1) \,|\, \mathcal{I}_{(L)}(\tau_L)]. \end{aligned}$$

The first equality follows from the law of iterated projections for orthogonal projections on a Hilbert space; the second from the linear aggregation scheme and the assumption that $\boldsymbol{x}_L \nrightarrow \boldsymbol{x}_H \,|\, \overline{\mathcal{I}}$; and the third holds because $\overline{\mathcal{I}}_{(L)}(m\tau_L) = \mathcal{I}_{(L)}(\tau_L)$. Hence $\boldsymbol{x}_L \nrightarrow \boldsymbol{x}_H \,|\, \mathcal{I}$ as claimed.

A.4 Proof of Theorem 2.4.3

We prove part (i) only since parts (ii)-(iv) are analogous. The following two cases complete part (i):

Case 1 (low $\nrightarrow$ low). Suppose that x_{L,j_1} does not cause x_{L,j_2} up to high frequency horizon m given $\overline{\mathcal{I}}$ (i.e., $x_{L,j_1} \nrightarrow_{(m)} x_{L,j_2} | \, \overline{\mathcal{I}}$). Then, $\mathbf{\Phi}_{LL,1}^{[k]}(j_2, j_1) = 0$ for any $k \in \{1, \ldots, m\}$ and hence x_{L,j_1} does not cause x_{L,j_2} at horizon 1 given $\mathcal{I}$ (i.e., $x_{L,j_1} \nrightarrow x_{L,j_2} | \, \mathcal{I}$) in view of (2.4.4). The converse does not necessarily hold; a simple counter-example is that $K_H = 1$, $K_L = 2$, $m = 2$, $(j_1, j_2) = (1, 2)$, and

$$\mathbf{\Phi}_1 = \begin{bmatrix} \phi_{HH} & 0.3 & \phi_{HL} \\ \phi_{LH} & 0.2 & \phi_{LL} \\ -0.1 & 0.1 & 0.1 \end{bmatrix},$$

where ϕ_{HH}, ϕ_{HL}, ϕ_{LH}, and ϕ_{LL} are arbitrary coefficients. It is evident that $\mathbf{\Phi}_{LL,1}(2, 1) = 0.1$ and $\mathbf{\Phi}_{LL,1}^{[2]}(2, 1) = 0$. The former denies that $x_{L,j_1} \nrightarrow_{(m)} x_{L,j_2} | \, \overline{\mathcal{I}}$, while the latter implies that $x_{L,j_1} \nrightarrow x_{L,j_2} | \, \mathcal{I}$.

Suppose now that $x_{L,j_1} \nrightarrow x_{L,j_2} | \, \mathcal{I}$. Then, $\mathbf{\Phi}_{LL,1}^{[m]}(j_2, j_1) = 0$ and hence $x_{L,j_1} \nrightarrow x_{L,j_2} | \, \underline{\mathcal{I}}$ in view of (2.4.9). The converse is also true.

Case 2 (high $\nrightarrow$ low). Suppose that $x_{H,i_1} \nrightarrow_{(m)} x_{L,j_1} | \, \overline{\mathcal{I}}$. Then, $\mathbf{\Phi}_{LH,1}^{[k]}(j_1, i_1) = 0$ for any $k \in \{1, \ldots, m\}$ and hence $x_{H,i_1} \nrightarrow x_{L,j_1} | \, \mathcal{I}$. The converse does not necessarily hold.

Suppose now that $x_{H,i_1} \nrightarrow x_{L,j_1} | \, \mathcal{I}$. Then, $\mathbf{\Phi}_{LH,1}^{[m]}(j_1, i_1) = 0$ and hence $x_{H,i_1} \nrightarrow x_{L,j_1} | \, \underline{\mathcal{I}}$. The converse is also true.

APPENDIX B

TECHNICAL APPENDICES FOR CHAPTER 3

B.1 Double Time Indices

Throughout this paper we consider a low frequency variable x_L and a high frequency variable x_H. The low frequency variable has a single time index $x_L(\tau_L)$ for $\tau_L \in \mathcal{Z}$ as in the usual time series literature. The high frequency variable, on the other hand, has two time indices $x_H(\tau_L, j)$ for $\tau_L \in \mathcal{Z}$ and $j \in \{1, \ldots, m\}$.

When we derive time series properties of x_H, it is useful to introduce a notational convention that allows the second argument of x_H to be any integer. For example, it is understood that $x_H(\tau_L, 0) = x_H(\tau_L - 1, m)$, $x_H(\tau_L, -1) = x_H(\tau_L - 1, m-1)$, and $x_H(\tau_L, m+1) = x_H(\tau_L + 1, 1)$. In general, we can introduce the following notation without any confusion:

High Frequency Simplification

$$x_H(\tau_L, j) = \begin{cases} x_H\left(\tau_L - \left\lceil \frac{1-j}{m} \right\rceil, m\left\lceil \frac{1-j}{m} \right\rceil + j\right) & \text{if} \quad j \leq 0, \\ x_H\left(\tau_L + \left\lfloor \frac{j-1}{m} \right\rfloor, j - m\left\lfloor \frac{j-1}{m} \right\rfloor\right) & \text{if} \quad j \geq m+1. \end{cases} \tag{B.1.1}$$

$\lceil x \rceil$ is the smallest integer not smaller than x, while $\lfloor x \rfloor$ is the largest integer not larger than x. We call (B.1.1) the *high frequency simplification* in the sense that any integer put in the second argument of x_H can be transformed to a natural number between 1 and m by modifying the first argument appropriately. In fact, we can verify that $m\left\lceil \frac{1-j}{m} \right\rceil + j \in \{1, \ldots, m\}$ when $j \leq 0$, and $j - m\left\lfloor \frac{j-1}{m} \right\rfloor \in \{1, \ldots, m\}$ when $j \geq m+1$.

Since the high frequency simplification allows both arguments of x_H to be any integer, we can verify the following relationship.

Low Frequency Simplification

$$x_H(\tau_L - i, j) = x_H(\tau_L, j - im), \quad \forall i, j, \tau_L \in \mathcal{Z}. \tag{B.1.2}$$

We call (B.1.2) the *low frequency simplification* in the sense that any lag or lead i put in the first

argument of x_H can be deleted by modifying the second argument appropriately. As a result the second argument may become non-positive or larger than m, but such a case is covered by (B.1.1).

B.2 Autocovariance Structures of x_L and x_H

All asymptotic results shown in this paper are based on the autocovariance structures of x_L and x_H as well as the cross-covariance structure between x_L and x_H. This section derives those properties, exploiting the notational convention given in Appendix B.1. Section B.2.1 has some basic results based on the DGP (3.2.1), which is replicated below for convenience. Section B.2.2 focuses on some important covariances associated with mixed frequency models, while Section B.2.3 focuses on their low frequency counterparts. The difference between the mixed frequency models and the low frequency models is whether we work on the original x_H or an aggregated x_H.

B.2.1 Preliminaries

We assume that x_L follows the DGP (3.2.1):

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

or in matrix form (3.2.2):

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

with $\boldsymbol{X}_L(\tau_L - 1) = [x_L(\tau_L - 1), \ldots, x_L(\tau_L - q)]'$, $\boldsymbol{X}_H^{(p)}(\tau_L - 1) = [x_H(\tau_L - 1, m + 1 - 1), \ldots, x_H(\tau_L - 1, m + 1 - p)]'$, $\boldsymbol{a} = [a_1, \ldots, a_q]'$, and $\boldsymbol{b} = [b_1, \ldots, b_p]'$. We impose Assumptions 3.2.1 and 3.2.2.

First, We use the low frequency simplification (B.1.2) to express the autocovariance of x_H in full generality:

$$E\left[x_H(\tau_L - i_1, m + 1 - j_1)x_H(\tau_L - i_2, m + 1 - j_2)\right] = \gamma^H_{j_2 - j_1 + (i_2 - i_1)m}, \quad \forall i_1, i_2, j_1, j_2, \tau_L \in \mathcal{Z}. \quad \text{(B.2.1)}$$

Equation (B.2.1) can be shown by observing that there are $|j_2 - j_1 + (i_2 - i_1)m|$ high frequency time periods between $x_H(\tau_L - i_1, m + 1 - j_1)$ and $x_H(\tau_L - i_2, m + 1 - j_2)$.

Next we derive the autocovariance structure of x_L. The first step is to transform the DGP into MA(∞) with infinite lags of x_H. Using the *low frequency* lag operator L, the DGP can be rewritten as $a(L)x_L(\tau_L) = \sum_{j=1}^{p} b_j x_H(\tau_L - 1, m+1-j) + \epsilon_L(\tau_L)$, where $a(L) = 1 - \sum_{k=1}^{q} a_k L^k$. The corresponding MA($\infty$) representation should be that

$$x_L(\tau_L) = \psi(L) \left\{ \sum_{j=1}^{p} b_j x_H(\tau_L - 1, m + 1 - j) + \epsilon_L(\tau_L) \right\}, \tag{B.2.2}$$

where $\psi(L) = \sum_{i=0}^{\infty} \psi_i L^i$. It must be the case that $a(L)\psi(L) = 1$ so that we can recover the original DGP starting from (B.2.2). This condition implies that

$$\psi_k = I(k > 0) \sum_{l=1}^{q} a_l \psi_{k-l} + I(k = 0), \quad \forall k \in \mathcal{Z}. \tag{B.2.3}$$

Besides (B.2.3), there are three useful properties for deriving the autocovariance of x_L. First, (B.2.1) implies that $E[\psi(L)x_H(\tau_L - 1, m + 1 - i) \times \psi(L)x_H(\tau_L - 1 - k, m + 1 - j)] = \sum_{l=1}^{\infty} \sum_{s=1}^{\infty} \psi_{l-1}\psi_{s-1}\gamma^H_{j-i+(s+k-l)m}$. Second, we have that $E[\psi(L)\epsilon_L(\tau_L) \times \psi(L)\epsilon_L(\tau_L - k)] = \sigma_L^2 \sum_{s=0}^{\infty} \psi_{k+s}\psi_s$ since $\epsilon_L(\tau_L) \overset{i.i.d.}{\sim} (0, \sigma_L^2)$ by assumption. Third, Assumption 3.2.1 ensures that $E[x_H(\tau_L, j)\epsilon_L(\tau_L - k)] = 0$ for any $j, k, \tau_L \in \mathcal{Z}$. These properties and (B.2.2) imply that $\{x_L(\tau_L)\}$ is a covariance stationary process with mean zero and autocovariance

$$\begin{aligned} \gamma_k^L &\equiv E[x_L(\tau_L)x_L(\tau_L - k)] \\ &= \sigma_L^2 \sum_{s=0}^{\infty} \psi_{k+s}\psi_s + \sum_{i=1}^{p} \sum_{j=1}^{p} \sum_{l=1}^{\infty} \sum_{s=1}^{\infty} b_i b_j \psi_{l-1}\psi_{s-1}\gamma^H_{j-i+(s+k-l)m}, \quad \forall k \in \mathcal{Z}. \end{aligned} \tag{B.2.4}$$

B.2.2 Mixed Frequency Models

We derive some covariance terms associated with our model (3.2.4):

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

In a matrix form, model j is rewritten as in (3.2.5):

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

As suggested in Theorem 3.2.1, key quantities in the subsequent proofs will be $E[\boldsymbol{x}_i(\tau_L - 1)\boldsymbol{x}_j(\tau_L-1)']$ and $E[\boldsymbol{x}_j(\tau_L-1)\boldsymbol{X}_H^{(p)}(\tau_L-1)']$. The first quantity, the covariance between all regressors in model i and all regressors in model j, is characterized as follows. Using (B.2.1), we get that $E[x_H(\tau_L-k, m+1-i)\times\psi(L)x_H(\tau_L-1-s, m+1-j)] = \sum_{l=1}^{\infty}\psi_{l-1}\gamma^H_{j-i+(l+s-k)m}$. This result and (B.2.2) imply that

$$\begin{aligned} c_{k,i,s} &\equiv E[x_H(\tau_L-k, m+1-i)x_L(\tau_L-s)] \\ &= \sum_{j=1}^{p}\sum_{l=1}^{\infty} b_j\psi_{l-1}\gamma^H_{j-i+(l+s-k)m}, \quad \forall k,i,s \in \mathcal{Z}. \end{aligned} \tag{B.2.5}$$

One trivial but useful property is that $c_{k,i,s} = c_{k',i,s'}$ whenever $s-k = s'-k'$. This fact suggests that we could drop either subscript k or subscript s from $c_{k,i,s}$ without loss of generality, but we would rather keep the three subscripts since this is often easier to understand when we deal with various lag orders of x_L and x_H below.

Based on (B.2.5), we have that

$$\begin{aligned} \boldsymbol{\Gamma}_{i,j} &\equiv E\left[\boldsymbol{x}_i(\tau_L-1)\boldsymbol{x}_j(\tau_L-1)'\right] \\ &= \begin{bmatrix} E[\boldsymbol{X}_L(\tau_L-1)\boldsymbol{X}_L(\tau_L-1)'] & E[\boldsymbol{X}_L(\tau_L-1)x_H(\tau_L-1, m+1-j)] \\ E[x_H(\tau_L-1, m+1-i)\boldsymbol{X}_L(\tau_L-1)'] & E[x_H(\tau_L-1, m+1-i)x_H(\tau_L-1, m+1-j)] \end{bmatrix} \\ &= \begin{bmatrix} \gamma^L_{1-1} & \cdots & \gamma^L_{1-q} & c_{1,j,1} \\ \vdots & \ddots & \vdots & \vdots \\ \gamma^L_{q-1} & \cdots & \gamma^L_{q-q} & c_{1,j,q} \\ c_{1,i,1} & \cdots & c_{1,i,q} & \gamma^H_{i-j} \end{bmatrix}, \quad \text{for } i,j \in \{1,\ldots,h\}. \end{aligned} \tag{B.2.6}$$

The third equality follows from (B.2.4) and (B.2.5). While $\boldsymbol{\Gamma}_{i,j}$ is neither symmetric nor non-singular in general, it is a symmetric non-singular matrix when $i=j$. $\boldsymbol{\Gamma}^{-1}_{j,j}$ can be obtained by

applying the well-known formula of block matrix inversion.

The second key quantity, the covariance between all regressors in model j and p high frequency lags of x_H, is characterized as follows.

$$
\begin{aligned}
&\boldsymbol{C}_j \equiv E[\boldsymbol{x}_j(\tau_L-1)\boldsymbol{X}_H^{(p)}(\tau_L-1)'] \\
&= \begin{bmatrix}
E[x_L(\tau_L-1)x_H(\tau_L-1,m+1-1)] & \dots & E[x_L(\tau_L-1)x_H(\tau_L-1,m+1-p)] \\
\vdots & \ddots & \vdots \\
E[x_L(\tau_L-q)x_H(\tau_L-1,m+1-1)] & \dots & E[x_L(\tau_L-q)x_H(\tau_L-1,m+1-p)] \\
E[x_H(\tau_L-1,m+1-j)x_H(\tau_L-1,m+1-1)] & \dots & E[x_H(\tau_L-1,m+1-j)x_H(\tau_L-1,m+1-p)]
\end{bmatrix} \\
&= \begin{bmatrix}
c_{1,1,1} & \dots & c_{1,p,1} \\
\vdots & \ddots & \vdots \\
c_{1,1,q} & \dots & c_{1,p,q} \\
\gamma_{j-1}^H & \dots & \gamma_{j-p}^H
\end{bmatrix}, \quad \text{for } j \in \{1,\dots,h\}.
\end{aligned}
\tag{B.2.7}
$$

The last equality follows from (B.2.5).

B.2.3 Low Frequency Models

It is of interest to see how better the mixed frequency model performs than its low frequency counterpart. The former works on $\{\{x_H(\tau_L,j)\}\}$, while the latter works on its aggregated version $\{x_H(\tau_L)\}$. We consider linear aggregation scheme $x_H(\tau_L) = \sum_{j=1}^m \delta_j x_H(\tau_L,j)$ with $\delta_j \geq 0$ for all $j = 1,\dots,m$ and $\sum_{j=1}^m \delta_j = 1$. The linear aggregation scheme includes flow sampling (i.e. $\delta_j = 1/m$ for $j = 1,\dots,m$) and stock sampling (i.e. $\delta_j = I(j=m)$ for $j = 1,\dots,m$) as special cases.

We first deduce the autocovariance structure of $\{x_H(\tau_L)\}$. It is a covariance stationary process with mean zero and autocovariance

$$
\begin{aligned}
\gamma_k^{H,LF} &\equiv E[x_H(\tau_L)x_H(\tau_L-k)] = E\left[\left(\sum_{i=1}^m \delta_i x_H(\tau_L,i)\right)\left(\sum_{j=1}^m \delta_j x_H(\tau_L-k,j)\right)\right] \\
&= \sum_{i=1}^m\sum_{j=1}^m \delta_i\delta_j E[x_H(\tau_L,j)x_H(\tau_L-k,j)] = \sum_{i=1}^m\sum_{j=1}^m \delta_i\delta_j\gamma_{j-i-km}^H, \quad \text{for } k \in \mathcal{Z}.
\end{aligned}
\tag{B.2.8}
$$

The last equality of (B.2.8) follows from (B.2.1).

We will also need the cross-covariance between the original x_H and its aggregated version.

Using (B.2.1) again, it is straightforward to show that

$$E[x_H(\tau_L-k)x_H(\tau_L-s,m+1-j)]=\sum_{i=1}^{m}\delta_i\gamma^H_{m+1-j-i+(k-s)m},\quad \forall k,s,j\in\mathcal{Z}. \tag{B.2.9}$$

Next we consider the cross-covariance between the aggregated x_H and x_L. We have that

$$\begin{aligned}
c^{LF}_{i,s} &\equiv E[x_H(\tau_L-i)x_L(\tau_L-s)]\\
&= E\left[\left(\sum_{l=1}^{m}\delta_l x_H(\tau_L-i,l)\right)\left(\sum_{j=1}^{p}b_j\psi(L)x_H(\tau_L-1-s,m+1-j)+\psi(L)\epsilon_L(\tau_L-s)\right)\right]\\
&= \sum_{l=1}^{m}\sum_{j=1}^{p}\delta_l b_j E[x_H(\tau_L-i,l)\times\psi(L)x_H(\tau_L-1-s,m+1-j)]\\
&= \sum_{l=1}^{m}\sum_{j=1}^{p}\delta_l b_j E[x_H(\tau_L-i,l)\times\sum_{k=0}^{\infty}\psi_k x_H(\tau_L-1-k-s,m+1-j)]\\
&= \sum_{l=1}^{m}\sum_{j=1}^{p}\sum_{k=0}^{\infty}\delta_l b_j\psi_k\gamma^H_{j+l-1-(s+k-2-i)m},\quad \text{for } i,s\in\mathcal{Z}.
\end{aligned} \tag{B.2.10}$$

The second equality of (B.2.10) follows from (B.2.2), the third equality follows from the independence assumption between x_H and ϵ_L, and the last equality follows from (B.2.1).

We now consider the low frequency parsimonious model (3.3.4):

$$x_L(\tau_L)=\sum_{k=1}^{q}\alpha^{LF}_{k,j}x_L(\tau_L-k)+\beta^{LF}_j x_H(\tau_L-j)+u^{LF}_{L,j}(\tau_L),\quad j=1,\ldots,h$$

or in a matrix form

$$x_L(\tau_L)=\boldsymbol{x}^{LF}_j(\tau_L-1)'\boldsymbol{\theta}^{LF}_j+u^{LF}_{L,j}(\tau_L)$$

with $\boldsymbol{x}^{LF}_j(\tau_L-1)=[\boldsymbol{X}_L(\tau_L-1)',\ x_H(\tau_L-j)]'$ and $\boldsymbol{\theta}^{LF}_j=[\alpha^{LF}_{1,j},\ldots,\alpha^{LF}_{q,j},\beta^{LF}_j]'$. There are two quantities which will play an important role in the low frequency model. The first one is $E[\boldsymbol{x}^{LF}_i(\tau_L-1)\boldsymbol{x}^{LF}_j(\tau_L-1)']$, the covariance between all regressors in model i and all regressors

in model j. It is easy to verify that

$$\begin{aligned}\boldsymbol{\Gamma}_{i,j}^{LF} &\equiv E\left[\boldsymbol{x}_i^{LF}(\tau_L-1)\boldsymbol{x}_j^{LF}(\tau_L-1)'\right] = \begin{bmatrix} E[\boldsymbol{X}_L(\tau_L-1)\boldsymbol{X}_L(\tau_L-1)'] & E[\boldsymbol{X}_L(\tau_L-1)x_H(\tau_L-j)] \\ E[x_H(\tau_L-i)\boldsymbol{X}_L(\tau_L-1)'] & E[x_H(\tau_L-i)x_H(\tau_L-j)] \end{bmatrix} \\ &= \begin{bmatrix} \gamma_{1-1}^{L} & \cdots & \gamma_{1-q}^{L} & c_{j,1}^{LF} \\ \vdots & \ddots & \vdots & \vdots \\ \gamma_{q-1}^{L} & \cdots & \gamma_{q-q}^{L} & c_{j,q}^{LF} \\ c_{i,1}^{LF} & \cdots & c_{i,q}^{LF} & \gamma_{i-j}^{H,LF} \end{bmatrix}, \quad \text{for } i,j \in \{1,\ldots,h\}.\end{aligned} \tag{B.2.11}$$

The last equality of (B.2.11) is a simple implication of (B.2.4), (B.2.8), and (B.2.10). While $\boldsymbol{\Gamma}_{i,j}^{LF}$ is neither symmetric nor non-singular in general, it is a symmetric non-singular matrix when $i=j$. The inverse matrix can be obtained by applying the well-known formula of block matrix inversion.

The second key quantity is $E[\boldsymbol{x}_j^{LF}(\tau_L-1)\boldsymbol{X}_H^{(p)}(\tau_L-1)']$, the covariance between all regressors in model j and p high frequency lags of x_H. Using (B.2.5) and (B.2.9), it is trivial to see that

$$\begin{aligned}\boldsymbol{C}_j^{LF} &\equiv E[\boldsymbol{x}_j^{LF}(\tau_L-1)\boldsymbol{X}_H^{(p)}(\tau_L-1)'] \\ &= \begin{bmatrix} E[x_L(\tau_L-1)x_H(\tau_L-1,m+1-1)] & \ldots & E[x_L(\tau_L-1)x_H(\tau_L-1,m+1-p)] \\ \vdots & \ddots & \vdots \\ E[x_L(\tau_L-q)x_H(\tau_L-1,m+1-1)] & \ldots & E[x_L(\tau_L-q)x_H(\tau_L-1,m+1-p)] \\ E[x_H(\tau_L-j)x_H(\tau_L-1,m+1-1)] & \ldots & E[x_H(\tau_L-j)x_H(\tau_L-1,m+1-p)] \end{bmatrix} \\ &= \begin{bmatrix} c_{1,1,1} & \cdots & c_{1,p,1} \\ \vdots & \ddots & \vdots \\ c_{1,1,q} & \cdots & c_{1,p,q} \\ \sum_{i=1}^{m}\delta_i\gamma_{m+1-1-i+(j-1)m}^{H} & \cdots & \sum_{i=1}^{m}\delta_i\gamma_{m+1-p-i+(j-1)m}^{H} \end{bmatrix}, \quad \text{for } j \in \{1,\ldots,h\}.\end{aligned} \tag{B.2.12}$$

B.3 Proof of Theorem 3.2.1

Recall that the mixed frequency model j given in (3.2.4) is written as

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

or in matrix form

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

The moment condition with respect to OLS is that $E[\boldsymbol{x}_j(\tau_L - 1)u_{L,j}(\tau_L)] = \mathbf{0}_{(q+1)\times 1}$, so the pseudo-true value of $\boldsymbol{\theta}_j$, denoted by $\boldsymbol{\theta}_j^*$, is as follows:

$$\boldsymbol{\theta}_j^* = \left[E\left[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{x}_j(\tau_L - 1)'\right]\right]^{-1} E\left[\boldsymbol{x}_j(\tau_L - 1)x_L(\tau_L)\right]. \tag{B.3.1}$$

Recall that the DGP in matrix form is

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

Substituting this into (B.3.1), we get

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

where the second equality holds from the i.i.d. assumption of ϵ_L. We have by construction that

$$E\left[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{X}_L(\tau_L - 1)'\right] = E\left[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{x}_j(\tau_L - 1)'\right]\begin{bmatrix}\boldsymbol{I}_q \\ \mathbf{0}_{1\times q}\end{bmatrix}.$$

Using this, we obtain

$$\boldsymbol{\theta}_j^* \equiv \begin{bmatrix}\alpha_{1,j}^* \\ \vdots \\ \alpha_{q,j}^* \\ \beta_j^*\end{bmatrix} = \begin{bmatrix}a_1 \\ \vdots \\ a_q \\ 0\end{bmatrix} + \left[E\left[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{x}_j(\tau_L - 1)'\right]\right]^{-1} E\left[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{X}_H^{(p)}(\tau_L - 1)'\right]\boldsymbol{b}. \tag{B.3.2}$$

Recall from (B.2.6) that $[E[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{x}_j(\tau_L - 1)']]^{-1}$ is already quantified as $\boldsymbol{\Gamma}_{j,j}^{-1}$. $E[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{X}_H^{(p)}(\tau_L - 1)']$ is also quantified as $\boldsymbol{C}_j$ in (B.2.7). Hence, (B.3.2) provides a complete characterization of $\boldsymbol{\theta}_j^*$.

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

B.4 Proof of Theorem 3.2.2

We first show that $\boldsymbol{b} = \mathbf{0}_{p\times 1} \Rightarrow \boldsymbol{\beta}^* = \mathbf{0}_{h\times 1}$. Assume that $\boldsymbol{b} = \mathbf{0}_{p\times 1}$, then (B.3.2) implies that $\beta_j^* = 0$ for any $j = 1, \ldots, h$. We thus have that $\boldsymbol{\beta}^* = \mathbf{0}_{h\times 1}$.

We now show that $\boldsymbol{\beta}^* = \mathbf{0}_{h\times 1} \Rightarrow \boldsymbol{b} = \mathbf{0}_{p\times 1}$, assuming that $h \geq p$. We pick the last row of (B.3.2). As seen from (B.2.6), the lower left block of $[E[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{x}_j(\tau_L - 1)']]^{-1}$ is

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

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

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

Since we are assuming that $\beta_j^* = 0$, the last row of (B.3.2) is given by $n_j^{-1}\boldsymbol{d}_j'\boldsymbol{b} = 0$, where

$$\begin{aligned} \boldsymbol{d}_j \equiv & E\left[\boldsymbol{X}_H^{(p)}(\tau_L - 1)x_H(\tau_L - 1, m + 1 - j)\right] \\ & - E\left[\boldsymbol{X}_H^{(p)}(\tau_L - 1)\boldsymbol{X}_L(\tau_L - 1)'\right]\left[E\left[\boldsymbol{X}_L(\tau_L - 1)\boldsymbol{X}_L(\tau_L - 1)'\right]\right]^{-1} E\left[\boldsymbol{X}_L(\tau_L - 1)x_H(\tau_L - 1, m + 1 - j)\right]. \end{aligned} \tag{B.4.1}$$

Since n_j is a nonzero finite scalar for any $j = 1, \ldots, h$ by the non-singularity of $E[\boldsymbol{x}_j(\tau_L - 1)\boldsymbol{x}_j(\tau_L - 1)']$, it has to be the case that $\boldsymbol{d}_j'\boldsymbol{b} = 0$. Stacking these h equations, we have that

$$\underbrace{\begin{bmatrix} \boldsymbol{d}_1' \\ \vdots \\ \boldsymbol{d}_h' \end{bmatrix}}_{\equiv \boldsymbol{D}} \boldsymbol{b} = \mathbf{0}_{h\times 1} \quad \text{and hence} \quad \boldsymbol{b}'\boldsymbol{D}'\boldsymbol{D}\boldsymbol{b} = 0.$$

To conclude that $\boldsymbol{b} = \mathbf{0}_{p\times 1}$, it is sufficient to show that $\boldsymbol{D}'\boldsymbol{D}$ is positive definite. Hence it is sufficient to show that $\boldsymbol{D}$ is of full column rank p. Since we are assuming that $h \geq p$, we only have to show that $\boldsymbol{D}_p \equiv [\boldsymbol{d}_1, \ldots, \boldsymbol{d}_p]'$, the first p rows of $\boldsymbol{D}$, is of full column rank p or equivalently non-singular. Equation (B.4.1) implies that

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

Now define

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

which is trivially non-singular since $\gamma_0^H > 0$ and $\sigma_L^2 > 0$ by assumption. Evidently, $\boldsymbol{D}_p$ is the Schur complement of $\boldsymbol{\Delta}$ with respect to $E[\boldsymbol{X}_L(\tau_L-1)\boldsymbol{X}_L(\tau_L-1)']$. Thus, by the classic argument of partitioned matrix inversion, $\boldsymbol{D}_p$ is non-singular as desired.

B.5 Proof of Theorem 3.2.3

Recall that the mixed frequency model j is:

$$x_L(\tau_L) = \boldsymbol{x}_j(\tau_L-1)'\boldsymbol{\theta}_j + u_{L,j}(\tau_L), \quad j = 1,\ldots,h,$$

where $\boldsymbol{\theta}_j = [\alpha_{1,j},\ldots,\alpha_{q,j},\beta_j]'$. We collect all parameters across the h models as $\boldsymbol{\theta} = [\boldsymbol{\theta}_1',\ldots,\boldsymbol{\theta}_h']'$.

Deriving the asymptotic distribution of our test statistic $\mathcal{T} = \max_{1\le j\le h}\left(\sqrt{T_L}w_{T_L,j}\hat{\beta}_j\right)^2$ under $H_0 : \boldsymbol{b} = \boldsymbol{0}_{p\times 1}$ will turn out to be almost identical to deriving the asymptotic distribution of $\sqrt{T_L}\hat{\boldsymbol{\beta}}$ under H_0. Working on $\sqrt{T_L}\hat{\boldsymbol{\beta}}$ directly is rather cumbersome, so we work on $\boldsymbol{R}\times\sqrt{T_L}(\hat{\boldsymbol{\theta}}-\bar{\boldsymbol{\theta}}_0)$, where the selection matrix $\boldsymbol{R}$ is such that $\hat{\boldsymbol{\beta}} = \boldsymbol{R}\hat{\boldsymbol{\theta}}$ as in the last part of Appendix B.3. Note that $\bar{\boldsymbol{\theta}}_0$, a hypothesized value for the pseudo-true value of $\boldsymbol{\theta}$, can be arbitrarily chosen as long as $\boldsymbol{R}\bar{\boldsymbol{\theta}}_0 = \boldsymbol{0}_{h\times 1}$. This condition guarantees that $\sqrt{T_L}\hat{\boldsymbol{\beta}} = \boldsymbol{R}\times\sqrt{T_L}(\hat{\boldsymbol{\theta}}-\bar{\boldsymbol{\theta}}_0)$. The most convenient choice satisfying this condition is $\bar{\boldsymbol{\theta}}_0 = \boldsymbol{\iota}_h\otimes\boldsymbol{\theta}_0$ with $\boldsymbol{\theta}_0 = [a_1,\ldots,a_q,0]'$, where $\boldsymbol{\iota}_h$ is an $h\times 1$ vector of ones. $\boldsymbol{\theta}_0$ is a hypothesized value for $\boldsymbol{\theta}_j$, all parameters in model j. Although it contains unknown quantities $a_1,\ldots,a_q$, it does not violate our theory since the last element of $\boldsymbol{\theta}_0$ is 0 and hence $\boldsymbol{R}\bar{\boldsymbol{\theta}}_0 = \boldsymbol{0}_{h\times 1}$.

We first derive the asymptotic distribution of $\sqrt{T_L}(\boldsymbol{\theta}_j-\boldsymbol{\theta}_0)$ under H_0. By the construction of $\boldsymbol{\theta}_0$, the DGP is written as $x_L(\tau_L) = \boldsymbol{x}_j(\tau_L-1)'\boldsymbol{\theta}_0 + \epsilon_L(\tau_L)$ under H_0. Using this, we have

that

$$
\begin{aligned}
&\sqrt{T_L}(\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_0) \\
&\quad = \sqrt{T_L}\left[\sum_{\tau_L=1}^{T_L} \boldsymbol{x}_j(\tau_L-1)\boldsymbol{x}_j(\tau_L-1)'\right]^{-1} \sum_{\tau_L=1}^{T_L} \boldsymbol{x}_j(\tau_L-1)x_L(\tau_L) - \sqrt{T_L}\boldsymbol{\theta}_0 \\
&\quad = \sqrt{T_L}\left[\sum_{\tau_L=1}^{T_L} \boldsymbol{x}_j(\tau_L-1)\boldsymbol{x}_j(\tau_L-1)'\right]^{-1} \sum_{\tau_L=1}^{T_L} \boldsymbol{x}_j(\tau_L-1)[\boldsymbol{x}_j(\tau_L-1)'\boldsymbol{\theta}_0 + \epsilon_L(\tau_L)] - \sqrt{T_L}\boldsymbol{\theta}_0 \\
&\quad = \sqrt{T_L}\left[\sum_{\tau_L=1}^{T_L} \boldsymbol{x}_j(\tau_L-1)\boldsymbol{x}_j(\tau_L-1)'\right]^{-1} \sum_{\tau_L=1}^{T_L} \boldsymbol{x}_j(\tau_L-1)\epsilon_L(\tau_L) \\
&\quad = \left[E[\boldsymbol{x}_j(\tau_L-1)\boldsymbol{x}_j(\tau_L-1)']\right]^{-1} \frac{1}{\sqrt{T_L}} \sum_{\tau_L=1}^{T_L} \boldsymbol{x}_j(\tau_L-1)\epsilon_L(\tau_L) + o_p(1), \\
&\quad = \boldsymbol{\Gamma}_{j,j}^{-1} \frac{1}{\sqrt{T_L}} \sum_{\tau_L=1}^{T_L} \boldsymbol{x}_j(\tau_L-1)\epsilon_L(\tau_L) + o_p(1),
\end{aligned}
\tag{B.5.1}
$$

where the last equality follows just by definition in (B.2.6). Using (B.5.1), we now deduce the asymptotic distribution of $\sqrt{T_L}(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_0)$. To rely on the Cramer-Wold theorem, we define a $(q+1)h \times 1$ nonzero vector $\boldsymbol{\lambda} = [\boldsymbol{\lambda}_1', \ldots, \boldsymbol{\lambda}_h']'$ and consider $\boldsymbol{\lambda}' \times \sqrt{T_L}(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_0)$. We have that

$$
\begin{aligned}
\boldsymbol{\lambda}' \times \sqrt{T_L}(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_0) &= \sum_{j=1}^{h} \boldsymbol{\lambda}_j' \times \sqrt{T_L}(\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_0) \\
&= \sum_{j=1}^{h} \boldsymbol{\lambda}_j' \left\{ \boldsymbol{\Gamma}_{j,j}^{-1} \frac{1}{\sqrt{T_L}} \sum_{\tau_L=1}^{T_L} \boldsymbol{x}_j(\tau_L-1)\epsilon_L(\tau_L) \right\} + o_p(1) \\
&= \frac{1}{\sqrt{T_L}} \sum_{\tau_L=1}^{T_L} \underbrace{\left\{ \sum_{j=1}^{h} \boldsymbol{\lambda}_j' \boldsymbol{\Gamma}_{j,j}^{-1} \boldsymbol{x}_j(\tau_L-1) \right\}}_{\equiv X(\tau_L-1, \boldsymbol{\lambda})} \epsilon_L(\tau_L) + o_p(1),
\end{aligned}
\tag{B.5.2}
$$

where the second equality follows from (B.5.1).

Recall the definition in (B.2.6) that $\boldsymbol{\Gamma}_{j,i} = E[\boldsymbol{x}_j(\tau_L-1)\boldsymbol{x}_i(\tau_L-1)']$. Using this, we have that

$$
E\left[X(\tau_L-1, \boldsymbol{\lambda})^2\right] = \sum_{j=1}^{h}\sum_{i=1}^{h} \boldsymbol{\lambda}_j' \underbrace{\boldsymbol{\Gamma}_{j,j}^{-1}\boldsymbol{\Gamma}_{j,i}\boldsymbol{\Gamma}_{i,i}^{-1}}_{\equiv \boldsymbol{\Sigma}_{j,i}} \boldsymbol{\lambda}_i = \boldsymbol{\lambda}'\boldsymbol{\Sigma}\boldsymbol{\lambda}, \tag{B.5.3}
$$

where

$$
\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{1,1} & \ldots & \boldsymbol{\Sigma}_{1,h} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{h,1} & \ldots & \boldsymbol{\Sigma}_{h,h} \end{bmatrix}.
$$

Using (B.5.3), we apply a central limit theorem to (B.5.2) in order to obtain that $\boldsymbol{\lambda}' \times \sqrt{T_L}$ $(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_0) \xrightarrow{d} N(0, \boldsymbol{\lambda}'(\sigma_L^2 \boldsymbol{\Sigma})\ \boldsymbol{\lambda})$.

By the Cramer-Wold theorem, we get that $\sqrt{T_L}(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_0) \xrightarrow{d} N(\mathbf{0}_{(q+1)h\times 1}, \sigma_L^2 \boldsymbol{\Sigma})$. Hence,

$$\sqrt{T_L}\boldsymbol{W}_{T_L}\hat{\boldsymbol{\beta}} = \sqrt{T_L}\boldsymbol{W}_{T_L}\boldsymbol{R}(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_0) \xrightarrow{d} N(\mathbf{0}_{h\times 1}, \underbrace{\sigma_L^2 \boldsymbol{W}\boldsymbol{R}\boldsymbol{\Sigma}\boldsymbol{R}'\boldsymbol{W}}_{\equiv \boldsymbol{V}}). \tag{B.5.4}$$

Recall that our test statistic is given by $\mathcal{T} = \max_{1\le j\le h}\left(\sqrt{T_L} w_{T_L,j}\hat{\beta}_j\right)^2$. Hence we have that $\mathcal{T} \xrightarrow{d} \max_{1\le j\le h}\mathcal{N}_j^2$, where $\boldsymbol{\mathcal{N}} = [\mathcal{N}_1, \ldots, \mathcal{N}_h]'$ is a vector-valued random variable drawn from $N(\mathbf{0}_{h\times 1}, \boldsymbol{V})$.

Further, we can simplify $\boldsymbol{V}$ substantially by imposing H_0. Under H_0, x_L and x_H are independent and thus $\boldsymbol{\Gamma}_{j,i}$ becomes block diagonal:

$$\boldsymbol{\Gamma}_{j,i} = \begin{bmatrix} E[\boldsymbol{X}_L(\tau_L - 1)\boldsymbol{X}_L(\tau_L - 1)'] & \mathbf{0}_{q\times 1} \\ \mathbf{0}_{1\times q} & E[x_H(\tau_L - 1, m + 1 - j)x_H(\tau_L - 1, m + 1 - i)] \end{bmatrix}. \tag{B.5.5}$$

Using this, we get that

$$\begin{aligned}\boldsymbol{\Sigma}_{j,i} = \boldsymbol{\Gamma}_{j,j}^{-1}\boldsymbol{\Gamma}_{j,i}\boldsymbol{\Gamma}_{i,i}^{-1} &= \begin{bmatrix} [E[\boldsymbol{X}_L(\tau_L - 1)\boldsymbol{X}_L(\tau_L - 1)']]^{-1} & \mathbf{0}_{q\times 1} \\ \mathbf{0}_{1\times q} & \frac{E[x_H(\tau_L-1,m+1-j)x_H(\tau_L-1,m+1-i)]}{E[x_H(\tau_L-1,m+1-j)^2]E[x_H(\tau_L-1,m+1-j)^2]} \end{bmatrix} \\ &= \begin{bmatrix} [E[\boldsymbol{X}_L(\tau_L - 1)\boldsymbol{X}_L(\tau_L - 1)']]^{-1} & \mathbf{0}_{q\times 1} \\ \mathbf{0}_{1\times q} & \rho_{i-j}^H/\gamma_0^H \end{bmatrix}.\end{aligned}$$

hence, we have by the construction of $\boldsymbol{R}$ that

$$\boldsymbol{R}\boldsymbol{\Sigma}\boldsymbol{R}' = \frac{1}{\gamma_0^H}\begin{bmatrix} \rho_{1-1}^H & \cdots & \rho_{1-h}^H \\ \vdots & \ddots & \vdots \\ \rho_{h-1}^H & \cdots & \rho_{h-h}^H \end{bmatrix} \equiv \frac{1}{\gamma_0^H}\boldsymbol{R}_{h,h}^H.$$

Thus, the asymptotic covariance matrix is written as

$$\boldsymbol{V} = \frac{\sigma_L^2}{\gamma_0^H}\boldsymbol{W}\boldsymbol{R}_{h,h}^H\boldsymbol{W}. \tag{B.5.6}$$

Theorem 3.2.3.(ii) is straightforward to show. By the construction of $\mathcal{T}$ and $\hat{\beta}_j \xrightarrow{p} \beta_j^*$, we have that $\mathcal{T} \xrightarrow{p} \infty \Leftrightarrow \boldsymbol{\beta}^* \neq \mathbf{0}_{h\times 1}$. Given $h \geq p$, Theorem 3.2.2 ensures that $\boldsymbol{b} \neq \mathbf{0}_{p\times 1} \Rightarrow \boldsymbol{\beta}^* \neq \mathbf{0}_{h\times 1}$. Therefore, the test statistic $\mathcal{T}$ diverges in probability under a general alternative hypothesis $H_1 : \boldsymbol{b} \neq \mathbf{0}_{p\times 1}$.

B.6 Proof of Theorem 3.2.4

Recall the parsimonious regression models (3.2.15):

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

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

To rewrite them in a matrix form, define

$$\underbrace{\bar{\boldsymbol{x}}_j(\tau_L)}_{n\times 1} = \begin{bmatrix} x_L(\tau_L - 1) \\ \vdots \\ x_L(\tau_L - q) \\ x_H(\tau_L - 1, m + 1 - 1) \\ \vdots \\ x_H(\tau_L - 1, m + 1 - mq) \\ x_H(\tau_L + 1, j) \end{bmatrix}, \quad \underbrace{\boldsymbol{\theta}_j}_{n\times 1} = \begin{bmatrix} \alpha_{1,j} \\ \vdots \\ \alpha_{q,j} \\ \beta_{1,j} \\ \vdots \\ \beta_{mq,j} \\ \gamma_j \end{bmatrix}, \text{ and } \underbrace{\boldsymbol{z}_j(\tau_L)}_{(n+m)\times 1} = \begin{bmatrix} \bar{\boldsymbol{x}}_j(\tau_L) \\ x_H(\tau_L, 1) \\ \vdots \\ x_H(\tau_L, m) \end{bmatrix},$$

where $n = q + mq + 1$. $\bar{\boldsymbol{x}}_j(\tau_L)$ is a vector of all explanatory variables while $\boldsymbol{\theta}_j$ is a vector of all parameters in model j. $\boldsymbol{z}_j(\tau_L)$ is a vector of instruments consisting of all n explanatory variables and m contemporaneous high frequency observations of x_H.

Using these notations, model (3.2.15) can be rewritten as

$$x_L(\tau_L) = \bar{\boldsymbol{x}}_j(\tau_L)'\boldsymbol{\theta}_j + u_{L,j}(\tau_L) \text{ with instruments } \boldsymbol{z}_j(\tau_L), \quad j = 1, \ldots, h.$$

To derive the GIVE for $\boldsymbol{\theta}_j$, define sample moments

$$\underbrace{\hat{\boldsymbol{S}}_j}_{(n+m)\times n} = \frac{1}{T_L}\sum_{\tau_L=1}^{T_L} \boldsymbol{z}_j(\tau_L)\bar{\boldsymbol{x}}_j(\tau_L)', \qquad \underbrace{\hat{\boldsymbol{s}}_j}_{(n+m)\times 1} = \frac{1}{T_L}\sum_{\tau_L=1}^{T_L} \boldsymbol{z}_j(\tau_L)x_L(\tau_L),$$
$$\underbrace{\hat{\boldsymbol{\Sigma}}_j}_{(n+m)\times(n+m)} = \frac{1}{T_L}\sum_{\tau_L=1}^{T_L} \boldsymbol{z}_j(\tau_L)\boldsymbol{z}_j(\tau_L)'.$$

Using these matrices, the GIVE for $\boldsymbol{\theta}_j$ is given by

$$\underbrace{\hat{\boldsymbol{\theta}}_j}_{n\times 1} = \left(\hat{\boldsymbol{S}}_j'\hat{\boldsymbol{\Sigma}}_j^{-1}\hat{\boldsymbol{S}}_j\right)^{-1}\hat{\boldsymbol{S}}_j'\hat{\boldsymbol{\Sigma}}_j^{-1}\hat{\boldsymbol{s}}_j.$$

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

$$\boldsymbol{\theta}_{0,j} = [\alpha_{1,j}^*, \ldots, \alpha_{q,j}^*, \beta_{1,j}^*, \ldots, \beta_{mq,j}^*, 0]', \tag{B.6.1}$$

where the asterisk signifies the pseudo-true value. We do not know the pseudo-true values of α's and β's in practice, but that does not matter since we are only interested in the zero hypothesis with respect to γ_j. Eq. (B.6.1) is the most convenient choice of a hypothesized value in terms of mathematical derivation.

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

$$\hat{\boldsymbol{s}}_j = \frac{1}{T_L}\sum_{\tau_L=1}^{T_L} \boldsymbol{z}_j(\tau_L)\left[\bar{\boldsymbol{x}}_j(\tau_L)'\boldsymbol{\theta}_{0,j} + \epsilon_L(\tau_L)\right] = \hat{\boldsymbol{S}}_j\boldsymbol{\theta}_{0,j} + \frac{1}{T_L}\sum_{\tau_L=1}^{T_L} \boldsymbol{z}_j(\tau_L)\epsilon_L(\tau_L)$$

and thus

$$\sqrt{T_L}(\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_{0,j}) = \left(\hat{\boldsymbol{S}}_j'\hat{\boldsymbol{\Sigma}}_j^{-1}\hat{\boldsymbol{S}}_j\right)^{-1}\hat{\boldsymbol{S}}_j'\hat{\boldsymbol{\Sigma}}_j^{-1} \times \frac{1}{\sqrt{T_L}}\sum_{\tau_L=1}^{T_L} \boldsymbol{z}_j(\tau_L)\epsilon_L(\tau_L), \quad j = 1, \ldots, h. \tag{B.6.2}$$

We have that

$$\hat{\boldsymbol{S}}_j \xrightarrow{p} E[\boldsymbol{z}_j(\tau_L)\bar{\boldsymbol{x}}_j(\tau_L)'] \equiv \boldsymbol{S}_j \quad \text{and} \quad \hat{\boldsymbol{\Sigma}}_j \xrightarrow{p} E[\boldsymbol{z}_j(\tau_L)\boldsymbol{z}_j(\tau_L)'] \equiv \boldsymbol{\Sigma}_j. \tag{B.6.3}$$

Using (B.6.3), we apply the Cramér-Wold theorem to (B.6.2) in order to combine all h parsimonious regression models. To this end, define $\boldsymbol{\lambda} = [\boldsymbol{\lambda}_1', \ldots, \boldsymbol{\lambda}_h']' \in \mathcal{R}^{nh}$ as well as

$$\underbrace{\hat{\boldsymbol{\theta}}}_{nh\times 1} = \begin{bmatrix} \hat{\boldsymbol{\theta}}_1 \\ \vdots \\ \hat{\boldsymbol{\theta}}_h \end{bmatrix} \quad \text{and} \quad \underbrace{\boldsymbol{\theta}_0}_{nh\times 1} = \begin{bmatrix} \boldsymbol{\theta}_{0,1} \\ \vdots \\ \boldsymbol{\theta}_{0,h} \end{bmatrix}.$$

Then we have that

$$\begin{aligned} \boldsymbol{\lambda}'\sqrt{T_L}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) &= \sum_{j=1}^{h} \boldsymbol{\lambda}_j'\sqrt{T_L}(\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_{0,j}) \\ &= \sum_{j=1}^{h} \boldsymbol{\lambda}_j' \left[\left(\hat{\boldsymbol{S}}_j' \hat{\boldsymbol{\Sigma}}_j^{-1} \hat{\boldsymbol{S}}_j \right)^{-1} \hat{\boldsymbol{S}}_j' \hat{\boldsymbol{\Sigma}}_j^{-1} \times \frac{1}{\sqrt{T_L}} \sum_{\tau_L=1}^{T_L} \boldsymbol{z}_j(\tau_L)\epsilon_L(\tau_L) \right] \\ &= \frac{1}{\sqrt{T_L}} \sum_{\tau_L=1}^{T_L} \underbrace{\left\{ \sum_{j=1}^{h} \boldsymbol{\lambda}_j' \left(\boldsymbol{S}_j' \boldsymbol{\Sigma}_j^{-1} \boldsymbol{S}_j \right)^{-1} \boldsymbol{S}_j' \boldsymbol{\Sigma}_j^{-1} \boldsymbol{z}_j(\tau_L) \right\}}_{\equiv Z(\tau_L)} \epsilon_L(\tau_L) + o_p(1). \end{aligned} \tag{B.6.4}$$

Define

$$\underbrace{\boldsymbol{\Sigma}_{j,i}}_{(n+m)\times(n+m)} = E\left[\boldsymbol{z}_j(\tau_L) \boldsymbol{z}_i(\tau_L)' \right],$$

then we have that

$$\begin{aligned} E\left[Z(\tau_L)^2 \right] &= \sum_{j=1}^{h} \sum_{i=1}^{h} \boldsymbol{\lambda}_j' \underbrace{\left(\boldsymbol{S}_j' \boldsymbol{\Sigma}_j^{-1} \boldsymbol{S}_j \right)^{-1} \boldsymbol{S}_j' \boldsymbol{\Sigma}_j^{-1} \boldsymbol{\Sigma}_{j,i} \boldsymbol{\Sigma}_i^{-1} \boldsymbol{S}_i \left(\boldsymbol{S}_i' \boldsymbol{\Sigma}_i^{-1} \boldsymbol{S}_i \right)^{-1}}_{\equiv \boldsymbol{\Psi}_{j,i}:\ n\times n} \boldsymbol{\lambda}_i \\ &= \boldsymbol{\lambda}' \boldsymbol{\Psi} \boldsymbol{\lambda}, \end{aligned} \tag{B.6.5}$$

where

$$\underbrace{\boldsymbol{\Psi}}_{nh\times nh} = \begin{bmatrix} \boldsymbol{\Psi}_{1,1} & \ldots & \boldsymbol{\Psi}_{1,h} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Psi}_{h,1} & \ldots & \boldsymbol{\Psi}_{h,h} \end{bmatrix}.$$

Applying a central limit theorem to (B.6.4) using (B.6.5), we get that $\boldsymbol{\lambda}'\sqrt{T_L}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) \xrightarrow{d} N(0, \boldsymbol{\lambda}'(\sigma_L^2\boldsymbol{\Psi})\boldsymbol{\lambda})$. Then by the Cramér-Wold theorem, we obtain that

$$\sqrt{T_L}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}_{nh\times 1}, \sigma_L^2\boldsymbol{\Psi}). \tag{B.6.6}$$

Define

$$\underbrace{\hat{\boldsymbol{\gamma}}}_{h\times 1} = \begin{bmatrix} \hat{\gamma}_1 \\ \vdots \\ \hat{\gamma}_h \end{bmatrix}, \ \underbrace{\boldsymbol{R}}_{h\times nh} = \begin{bmatrix} \mathbf{0}_{1\times(n-1)} & 1 & \dots & \mathbf{0}_{1\times(n-1)} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_{1\times(n-1)} & 0 & \dots & \mathbf{0}_{1\times(n-1)} & 1 \end{bmatrix}, \text{ and } \underbrace{\boldsymbol{W}}_{h\times h} = \begin{bmatrix} w_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & w_h \end{bmatrix}. \tag{B.6.7}$$

$\boldsymbol{R}$ is a selection matrix choosing γ's out of the entire parameter vector $\boldsymbol{\theta}$, while $\boldsymbol{W}$ is a diagonal matrix having the L^2 limit of the weighting scheme $\boldsymbol{w}_{T_L}$. Equations (B.6.6) and (B.6.7) imply that

$$\sqrt{T_L}\boldsymbol{W}\hat{\boldsymbol{\gamma}} = \boldsymbol{W}\boldsymbol{R}\times\sqrt{T_L}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0) \xrightarrow{d} N(\mathbf{0}_{h\times 1}, \underbrace{\sigma_L^2\boldsymbol{W}\boldsymbol{R}\boldsymbol{\Psi}\boldsymbol{R}'\boldsymbol{W}}_{\equiv\boldsymbol{U}})$$

under $H_0 : x_L \nrightarrow x_H$.

B.7 Proof of Theorem 3.3.1

This proof is identical to the proof for Theorem 3.2.3 except for that we impose $H_1^l : \boldsymbol{b} = (1/\sqrt{T_L})\boldsymbol{\nu}$ instead of $H_0 : \boldsymbol{b} = \mathbf{0}_{p\times 1}$ when we derive (B.5.1). Recall (3.3.1), the DGP under H_1^l:

$$\begin{aligned} x_L(\tau_L) &= \sum_{k=1}^{q} a_k x_L(\tau_L - k) + \sum_{j=1}^{p} \frac{\nu_j}{\sqrt{T_L}} x_H(\tau_L - 1, m+1-j) + \epsilon_L(\tau_L) \\ &= \boldsymbol{X}_L(\tau_L-1)'\boldsymbol{a} + \boldsymbol{X}_H^{(p)}(\tau_L-1)'\left(\frac{1}{\sqrt{T_L}}\boldsymbol{\nu}\right) + \epsilon_L(\tau_L) \\ &= \boldsymbol{x}_j(\tau_L-1)'\begin{bmatrix} \boldsymbol{I}_q \\ \mathbf{0}_{1\times q} \end{bmatrix}\boldsymbol{a} + \boldsymbol{X}_H^{(p)}(\tau_L-1)'\left(\frac{1}{\sqrt{T_L}}\boldsymbol{\nu}\right) + \epsilon_L(\tau_L) \\ &= \boldsymbol{x}_j(\tau_L-1)'\boldsymbol{\theta}_0 + \boldsymbol{X}_H^{(p)}(\tau_L-1)'\left(\frac{1}{\sqrt{T_L}}\boldsymbol{\nu}\right) + \epsilon_L(\tau_L). \end{aligned}$$

Based on this equation, (B.5.1) should be modified as follows.

$$
\begin{aligned}
\sqrt{T_L}(\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_0) =& \sqrt{T_L}\left[\sum_{\tau_L=1}^{T_L} \boldsymbol{x}_j(\tau_L-1)\boldsymbol{x}_j(\tau_L-1)'\right]^{-1} \sum_{\tau_L=1}^{T_L} \boldsymbol{x}_j(\tau_L-1)x_L(\tau_L) - \sqrt{T_L}\boldsymbol{\theta}_0 \\
=& \sqrt{T_L}\left[\sum_{\tau_L=1}^{T_L} \boldsymbol{x}_j(\tau_L-1)\boldsymbol{x}_j(\tau_L-1)'\right]^{-1} \\
& \times \sum_{\tau_L=1}^{T_L} \boldsymbol{x}_j(\tau_L-1)\left[\boldsymbol{x}_j(\tau_L-1)'\boldsymbol{\theta}_0 + \boldsymbol{X}_H^{(p)}(\tau_L-1)'\left(\frac{1}{\sqrt{T_L}}\boldsymbol{\nu}\right) + \epsilon_L(\tau_L)\right] - \sqrt{T_L}\boldsymbol{\theta}_0 \\
=& \left[\sum_{\tau_L=1}^{T_L} \boldsymbol{x}_j(\tau_L-1)\boldsymbol{x}_j(\tau_L-1)'\right]^{-1}\left[\sum_{\tau_L=1}^{T_L} \boldsymbol{x}_j(\tau_L-1)\boldsymbol{X}_H^{(p)}(\tau_L-1)'\right]\boldsymbol{\nu} \\
& + \sqrt{T_L}\left[\sum_{\tau_L=1}^{T_L} \boldsymbol{x}_j(\tau_L-1)\boldsymbol{x}_j(\tau_L-1)'\right]^{-1} \sum_{\tau_L=1}^{T_L} \boldsymbol{x}_j(\tau_L-1)\epsilon_L(\tau_L) \\
=& \left[E[\boldsymbol{x}_j(\tau_L-1)\boldsymbol{x}_j(\tau_L-1)']\right]^{-1} E[\boldsymbol{x}_j(\tau_L-1)\boldsymbol{X}_H^{(p)}(\tau_L-1)']\boldsymbol{\nu} \\
& + \left[E[\boldsymbol{x}_j(\tau_L-1)\boldsymbol{x}_j(\tau_L-1)']\right]^{-1} \frac{1}{\sqrt{T_L}} \sum_{\tau_L=1}^{T_L} \boldsymbol{x}_j(\tau_L-1)\epsilon_L(\tau_L) + o_p(1), \\
=& \boldsymbol{\Gamma}_{j,j}^{-1}\boldsymbol{C}_j\boldsymbol{\nu} + \boldsymbol{\Gamma}_{j,j}^{-1}\frac{1}{\sqrt{T_L}} \sum_{\tau_L=1}^{T_L} \boldsymbol{x}_j(\tau_L-1)\epsilon_L(\tau_L) + o_p(1),
\end{aligned}
\tag{B.7.1}
$$

where the last equality follows simply from the definitions in (B.2.6) and (B.2.7).

Repeating (B.5.2), we get

$$
\begin{aligned}
\boldsymbol{\lambda}' \times \sqrt{T_L}(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_0) &= \sum_{j=1}^{h} \boldsymbol{\lambda}_j'\boldsymbol{\Gamma}_{j,j}^{-1}\boldsymbol{C}_j\boldsymbol{\nu} + \frac{1}{\sqrt{T_L}}\sum_{j=1}^{h} X(\tau_L-1,\lambda)\epsilon_L(\tau_L) + o_p(1) \\
&\xrightarrow{d} N\left(\boldsymbol{\lambda}'\boldsymbol{u},\, \boldsymbol{\lambda}'(\sigma_L^2\boldsymbol{\Sigma})\boldsymbol{\lambda}\right),
\end{aligned}
$$

where

$$
\boldsymbol{u} \equiv \begin{bmatrix} \boldsymbol{\Gamma}_{1,1}^{-1}\boldsymbol{C}_1 \\ \vdots \\ \boldsymbol{\Gamma}_{h,h}^{-1}\boldsymbol{C}_h \end{bmatrix}\boldsymbol{\nu}.
$$

By the Cramer-Wold theorem, we have that $\sqrt{T_L}(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_0) \xrightarrow{d} N(\boldsymbol{u},\, \sigma_L^2\boldsymbol{\Sigma})$.

Now repeat (B.5.4) to get

$$
\sqrt{T_L}\boldsymbol{W}_{T_L}\hat{\boldsymbol{\beta}} = \sqrt{T_L}\boldsymbol{W}_{T_L}\boldsymbol{R}(\hat{\boldsymbol{\theta}} - \bar{\boldsymbol{\theta}}_0) \xrightarrow{d} N(\underbrace{\boldsymbol{W}\boldsymbol{R}\boldsymbol{u}}_{\equiv\boldsymbol{\mu}},\, \underbrace{\sigma_L^2\boldsymbol{W}\boldsymbol{R}\boldsymbol{\Sigma}\boldsymbol{R}'\boldsymbol{W}}_{=\boldsymbol{V}}). \tag{B.7.2}
$$

Recall that our test statistic is given by $\mathcal{T} = \max_{1\leq j\leq h}\left(\sqrt{T_L}w_{T_L,j}\hat{\beta}_j\right)^2$. Hence we have that $\mathcal{T} \xrightarrow{d} \max_{1\leq j\leq h}\mathcal{M}_j^2$, where $\boldsymbol{\mathcal{M}} = [\mathcal{M}_1,\ldots,\mathcal{M}_h]'$ is a vector-valued random variable drawn from $N(\boldsymbol{\mu},\boldsymbol{V})$.

Furthermore, we can simplify $\boldsymbol{\mu}$ and $\boldsymbol{V}$ by imposing H_1^l. Under H_1^l, x_L and x_H are asymptotically independent and thus $\boldsymbol{\Gamma}_{i,j}$ converges to the block diagonal matrix in (B.5.5).[1] Hence, the exactly same simplification as in Appendix B.5 applies for $\boldsymbol{V}$ and we get (B.5.6) here as well. Similarly, it is asymptotically the case that

$$\boldsymbol{C}_j \equiv E[\boldsymbol{x}_j(\tau_L-1)\boldsymbol{X}_H^{(p)}(\tau_L-1)'] = \begin{bmatrix}\mathbf{0}_{q\times 1} & \cdots & \mathbf{0}_{q\times 1}\\ \gamma_{j-1}^H & \cdots & \gamma_{j-p}^H\end{bmatrix} \quad \text{and thus} \quad \boldsymbol{\Gamma}_{j,j}^{-1}\boldsymbol{C}_j = \begin{bmatrix}\mathbf{0}_{q\times 1} & \cdots & \mathbf{0}_{q\times 1}\\ \rho_{j-1}^H & \cdots & \rho_{j-p}^H\end{bmatrix}.$$

By the construction of $\boldsymbol{R}$, we get that

$$\boldsymbol{Ru} \equiv \boldsymbol{R}\begin{bmatrix}\boldsymbol{\Gamma}_{1,1}^{-1}\boldsymbol{C}_1\\ \vdots \\ \boldsymbol{\Gamma}_{h,h}^{-1}\boldsymbol{C}_h\end{bmatrix}\boldsymbol{\nu} = \underbrace{\begin{bmatrix}\rho_{1-1}^H & \cdots & \rho_{1-p}^H\\ \vdots & \ddots & \vdots \\ \rho_{h-1}^H & \cdots & \rho_{h-p}^H\end{bmatrix}}_{\equiv \boldsymbol{R}_{h,p}^H}\boldsymbol{\nu}.$$

Thus, we can conclude that $\boldsymbol{\mu} = \boldsymbol{W}\boldsymbol{R}_{h,p}^H\boldsymbol{\nu}$.

B.8 Proof of Theorem 3.3.3

Recall that the low frequency model is given by

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

or in a matrix form

$$x_L(\tau_L) = \boldsymbol{x}_j^{LF}(\tau_L-1)'\boldsymbol{\theta}_j^{LF} + u_{L,j}^{LF}(\tau_L)$$

with $\boldsymbol{x}_j^{LF}(\tau_L-1) = [\boldsymbol{X}_L(\tau_L-1)',\ x_H(\tau_L-j)]'$ and $\boldsymbol{\theta}_j^{LF} = [\alpha_{1,j}^{LF},\ldots,\alpha_{q,j}^{LF},\beta_j^{LF}]'$.

We first derive the asymptotic distribution of $\mathcal{T}_{LF} = \max_{1\leq j\leq h}(\sqrt{T_L}w_{T_L,j}\hat{\beta}_j^{LF})^2$ under H_1^l : $\boldsymbol{b} = (1/\sqrt{T_L})\boldsymbol{\nu}$. The derivation is identical to Appendix B.7 with the only difference being that

[1] If we want to verify this point algebraically, we can refer to (B.2.5) and (B.2.6) and impose $b_j = \nu_j/\sqrt{T_L} \to 0$.

we work on $\boldsymbol{x}_j^{LF}(\tau_L-1)$ instead of $\boldsymbol{x}_j(\tau_L-1)$. As a result, $\boldsymbol{\Gamma}_{i,j}$ and $\boldsymbol{C}_j$ in Appendix B.7 should be replaced with $\boldsymbol{\Gamma}_{i,j}^{LF} \equiv E[\boldsymbol{x}_i^{LF}(\tau_L-1)\boldsymbol{x}_j^{LF}(\tau_L-1)']$ and $\boldsymbol{C}_j^{LF} \equiv E[\boldsymbol{x}_j^{LF}(\tau_L-1)\boldsymbol{X}_H^{(p)}(\tau_L-1)']$. Similarly, $\boldsymbol{\mu}$ and $\boldsymbol{V}$ in (B.7.2) should be replaced with $\boldsymbol{\mu}^{LF}$ and $\boldsymbol{V}^{LF}$, where

$$\boldsymbol{\mu}^{LF} = \boldsymbol{W}\boldsymbol{R} \times \begin{bmatrix} \boldsymbol{\Gamma}_{1,1}^{LF}\boldsymbol{C}_1^{LF} \\ \vdots \\ \boldsymbol{\Gamma}_{h,h}^{LF}\boldsymbol{C}_h^{LF} \end{bmatrix} \boldsymbol{\nu} \quad \text{and} \quad \boldsymbol{V}^{LF} = \sigma_L^2 \boldsymbol{W}\boldsymbol{R}\boldsymbol{\Sigma}^{LF}\boldsymbol{R}'\boldsymbol{W} \tag{B.8.1}$$

with

$$\boldsymbol{\Sigma}^{LF} = \begin{bmatrix} \boldsymbol{\Sigma}_{1,1}^{LF} & \dots & \boldsymbol{\Sigma}_{1,h}^{LF} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{h,1}^{LF} & \dots & \boldsymbol{\Sigma}_{h,h}^{LF} \end{bmatrix}, \quad \boldsymbol{\Sigma}_{j,i}^{LF} = \left(\boldsymbol{\Gamma}_{j,j}^{LF}\right)^{-1} \boldsymbol{\Gamma}_{j,i}^{LF} \left(\boldsymbol{\Gamma}_{i,i}^{LF}\right)^{-1}.$$

Using (B.8.1), we can deduce in the same manner as before that $\mathcal{T}_{LF} \xrightarrow{d} \max_{1\le j\le h}(\mathcal{M}_j^{LF})^2$ under H_1^l. $\boldsymbol{\mathcal{M}}^{LF} = [\mathcal{M}_1^{LF}, \dots, \mathcal{M}_h^{LF}]'$ is a vector-valued random variable following $N(\boldsymbol{\mu}^{LF}, \boldsymbol{V}^{LF})$.

As in Appendix B.7, we can simplify $\boldsymbol{\mu}^{LF}$ and $\boldsymbol{V}^{LF}$ by imposing H_1^l. Since the aggregated x_H and x_L are asymptotically independent, $\boldsymbol{\Gamma}_{i,j}^{LF}$ converges to a block diagonal matrix:[2]

$$\boldsymbol{\Gamma}_{i,j}^{LF} \to \begin{bmatrix} E[\boldsymbol{X}_L(\tau_L-1)\boldsymbol{X}_L(\tau_L-1)'] & \boldsymbol{0}_{q\times 1} \\ \boldsymbol{0}_{1\times q} & E[x_H(\tau_L-i)x_H(\tau_L-j)] \end{bmatrix} \equiv \begin{bmatrix} E[\boldsymbol{X}_L(\tau_L-1)\boldsymbol{X}_L(\tau_L-1)'] & \boldsymbol{0}_{q\times 1} \\ \boldsymbol{0}_{1\times q} & \gamma_{i-j}^{H,LF} \end{bmatrix}.$$

Note that $\gamma_{i-j}^{H,LF}$ is characterized by underlying parameters in (B.2.8). Similarly, by (B.2.12) it is asymptotically the case that

$$\begin{aligned} \boldsymbol{C}_j^{LF} &\equiv E[\boldsymbol{x}_j^{LF}(\tau_L-1)\boldsymbol{X}_H^{(p)}(\tau_L-1)'] \\ &= \begin{bmatrix} \boldsymbol{0}_{q\times 1} & \dots & \boldsymbol{0}_{q\times 1} \\ E[x_H(\tau_L-j)x_H(\tau_L-1, m+1-1)] & \dots & E[x_H(\tau_L-j)x_H(\tau_L-1, m+1-p)] \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{0}_{q\times 1} & \dots & \boldsymbol{0}_{q\times 1} \\ \sum_{i=1}^m \delta_i \gamma_{m+1-1-i+(j-1)m}^H & \dots & \sum_{i=1}^m \delta_i \gamma_{m+1-p-i+(j-1)m}^H \end{bmatrix}. \end{aligned}$$

[2] If we want to see this algebraically, we can refer to (B.2.10) and (B.2.11) and impose $b_j = \nu_j/\sqrt{T_L} \to 0$.

Using these simplified $\boldsymbol{\Gamma}_{i,j}^{LF}$ and $\boldsymbol{C}_{j}^{LF}$, we can conclude that

$$\boldsymbol{\mu}^{LF} = \boldsymbol{W}\boldsymbol{\Delta}\boldsymbol{\nu} \quad \text{and} \quad \boldsymbol{V}^{LF} = \frac{\sigma_L^2}{\gamma_0^{H,LF}}\boldsymbol{W}\boldsymbol{R}^{H,LF}\boldsymbol{W},$$

where

$$\boldsymbol{\Delta} = \frac{1}{\gamma_0^{H,LF}} \begin{bmatrix} \sum_{i=1}^{m} \delta_i \gamma_{m+1-1-i+(1-1)m}^{H} & \cdots & \sum_{i=1}^{m} \delta_i \gamma_{m+1-p-i+(1-1)m}^{H} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{m} \delta_i \gamma_{m+1-1-i+(h-1)m}^{H} & \cdots & \sum_{i=1}^{m} \delta_i \gamma_{m+1-p-i+(h-1)m}^{H} \end{bmatrix}$$

and

$$\boldsymbol{R}^{H,LF} = \frac{1}{\gamma_0^{H,LF}} \begin{bmatrix} \gamma_{1-1}^{H,LF} & \cdots & \gamma_{1-h}^{H,LF} \\ \vdots & \ddots & \vdots \\ \gamma_{h-1}^{H,LF} & \cdots & \gamma_{h-h}^{H,LF} \end{bmatrix}.$$

The derivation of these formulas is analogous to Appendix B.7.

We now consider the asymptotic null distribution. Since the DGP under H_0 is identical to the DGP under H_1^l with $\boldsymbol{\nu} = \boldsymbol{0}_{p\times 1}$, it is trivial to show that $\mathcal{T}_{LF} \xrightarrow{d} \max_{1\leq j\leq h} (\mathcal{N}_j^{LF})^2$ under H_0. $\boldsymbol{\mathcal{N}}^{LF} = [\mathcal{N}_1^{LF}, \ldots, \mathcal{N}_h^{LF}]'$ is a vector-valued random variable following $N(\boldsymbol{0}_{h\times 1}, \boldsymbol{V}^{LF})$.

BIBLIOGRAPHY

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.

Anderson, B. D. O., M. Deistler, E. Felsenstein, B. Funovits, P. Zadrozny, M. Eichler, W. Chen, and M. Zamani (2012): "Identifiability of Regular and Singular Multivariate Autoregressive Models from Mixed Frequency Data," in *51st Conference on Decision and Control*, pp. 184–189, Maui, HI. IEEE Control Systems Society.

Andreou, E., E. Ghysels, and A. Kourtellos (2010): "Regression Models with Mixed Sampling Frequencies," *Journal of Econometrics*, 158, 246–261.

——— (2011): "Forecasting with Mixed-Frequency Data," in *Oxford Handbook of Economic Forecasting*, ed. by M. Clements, and D. Hendry, pp. 225–245.

Andrews, D. W. K. (1991): "Heteroscedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–85.

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

Armesto, M., K. Engemann, and M. Owyang (2010): "Forecasting with Mixed Frequencies," *Federal Reserve Bank of St. Louis Review*, 92, 521–536.

Boussama, F. (1998): "Ergodicity, Mixing and Estimation in GARCH Models," Ph.D. thesis, University of Paris.

Boussama, F., F. Fuchs, and R. Stelzer (2011): "Stationarity and Geometric Ergodicity of BEKK Multivariate GARCH Models," *Stochastic Processes and their Applications*, 121, 2331–2360.

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

Caporale, G. M., N. Pittis, and N. Spagnolo (2006): "Volatility Transmission and Financial Crises," *Journal of Economics and Finance*, 30, 376–390.

Carrasco, M., and X. Chen (2002): "Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models," *Econometric Theory*, 18, 17–39.

Chauvet, M., T. Götz, and A. Hecq (2013): "Realized Volatility and Business Cycle Fluctuations: A Mixed-Frequency VAR Approach," Working paper, University of California Riverside and Maastricht University.

Davidson, J. (1994): *Stochastic Limit Theory*. New York: Oxford University Press.

de Jong, R., and J. Davidson (2000): "Consistency of Kernel Estimators of Heteroscedastic and Autocorrelated Covariance Matrices," *Econometrica*, 68, 407–423.

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

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

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

Felsenstein, E., B. Funovits, M. Deistler, B. Anderson, M. Zamani, and W. Chen (2013): "Regular and Singular Multivariate Autoregressive Models and Mixed Frequency Data: Identifiability and Consistent Estimation," Working paper, Vienna University of Technology.

Florens, J. P., and M. Mouchart (1982): "A Note on Noncausality," *Econometrica*, 50, 583–591.

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

Foroni, C., M. Marcellino, and C. Schumacher (2013): "U-MIDAS: MIDAS Regressions with Unrestricted Lag Polynomials," *Journal of the Royal Statistical Society*, Series A (fortcoming).

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

Ghysels, E. (2012): "Macroeconomics and the Reality of Mixed Frequency Data," Working paper, University of North Carolina at Chapel Hill.

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

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

——— (2005): "There is a Risk-Return Trade-off After All," *Journal of Financial Economics*, 76, 509–548.

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

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

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

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

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

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

Hall, P., and C. C. Heyde (1980): *Martingale limit theory and its application*. Academic Press.

Hamilton, J. (2008): "Oil and the Macroeconomy," in *The New Palgrave Dictionary of Economics*, ed. by S. N. Durlauf, and L. E. Blume, vol. 6, pp. 172–175. Basingstoke, Hampshire: New York: Palgrave Macmillan, second edn.

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

Ibragimov, I. (1962): "Some Limit Theorems for Stationary Processes," *Theory of Probability and its Applications*, 7, 349–382.

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

King, M., E. Sentana, and S. Wadhwani (1994): "Volatility and Links between National Stock Markets," *Econometrica*, 62, 901–933.

Kuzin, V., M. Marcellino, and C. Schumacher (2011): "MIDAS versus Mixed-Frequency VAR: Nowcasting GDP in the Euro Area," *International Journal of Forecasting*, 27, 529–542.

Lewis, R., and G. Reinsel (1985): "Prediction of Multivariate Time Series by Autoregressive Model Fitting," *Journal of Multivariate Analysis*, 16, 393–411.

Lütkepohl, H. (1984): "Linear Transformation of Vector ARMA Processes," *Journal of Econometrics*, 26, 283–293.

——— (1987): *Forecasting Aggregated Vector ARMA Processes*. Berlin: Springer-Verlag.

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

Lütkepohl, H., and D. Poskitt (1996): "Testing for Causation Using Infinite Order Vector Autoregressive Processes," *Econometric Theory*, 12, 61–87.

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

McCracken, M., M. Owyang, and T. Sekhposyan (2013): "Real-Time Forecasting with a Large Bayesian Block Model," Discussion Paper, Federal Reserve Bank of St. Louis and Bank of Canada.

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

McLeish, D. L. (1974): "Dependent Central Limit Theorems and Invariance Principles," *Annals of Probability*, 2, 620–628.

Meitz, M., and P. Saikkonen (2008): "Stability of Nonlinear AR-GARCH Models," *Journal of Time Series Analysis*, 29, 453–475.

Nelson, D. B. (1990): "Stationarity and Persistence in the GARCH(1,1) Model," *Econometric Theory*, 6, 318–334.

Newey, W., and K. West (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

——— (1994): "Automatic Lag Selection in Covariance Matrix Estimation," *Review of Economic Studies*, 61, 631–653.

Payne, J. (2010): "Survey of the International Evidence on the Causal Relationship between Energy Consumption and Growth," *Journal of Economic Studies*, 37, 53–95.

Petersen, K. (1983): *Ergodic Theory*. Cambridge Univ. Press.

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

Rosenblatt, M. (1956): "A Central Limit Theorem and a Strong Mixing Condition," *Proceedings of the National Academy of Sciences USA*, 42, 43–47.

Saikkonen, P., and H. Lütkepohl (1996): "Infinite-Order Cointegrated Vector Autoregressive Processes: Estimation and Inference," *Econometric Theory*, 12, 814–844.

Salamaliki, P., and I. Venetis (2013): "Energy Consumption and Real GDP in G-7: Multi-Horizon Causality Testing in the Presence of Capital Stock," *Energy Economics*, forthcoming.

Silvestrini, A., and D. Veredas (2008): "Temporal Aggregation of Univariate and Multivariate Time Series Models: A Survey," *Journal of Economic Surveys*, 22, 458–497.

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

——— (1980): "Macroeconomics and Reality," *Econometrica*, 48, 1–48.

Yamamoto, T., and E. Kurozumi (2006): "Tests for Long-Run Granger Non-Causality in Cointegrated Systems," *Journal of Time Series Analysis*, 27, 703–723.

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.