12-1-2016

Change Detection and the Causal Impact of the Yield Curve

Stan Hurn

Peter C.B. Phillips

Shu-Ping Shi

CHANGE DETECTION AND THE CAUSAL IMPACT OF THE YIELD CURVE

By

Stan Hurn, Peter C. B. Phillips, and Shuping Shi

December 2016

COWLES FOUNDATION DISCUSSION PAPER NO. 2058

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Change Detection and the Causal Impact of the Yield Curve*

Stan Hurn
Queensland University of Technology

Peter C. B. Phillips
Yale University, University of Auckland,
University of Southampton & Singapore Management University

Shuping Shi
Macquarie University

Abstract

Causal relationships in econometrics are typically based on the concept of predictability and are established in terms of tests for Granger causality. These causal relationships are susceptible to change, especially during times of financial turbulence, making the real-time detection of instability an important practical issue. This paper develops a test for detecting changes in causal relationships based on a recursive rolling window, which is analogous to the procedure used in recent work on financial bubble detection. The limiting distribution of the test takes a simple form under the null hypothesis and is easy to implement in conditions of homoskedasticity, conditional heteroskedasticity and unconditional heteroskedasticity. Simulation experiments compare the efficacy of the proposed test with two other commonly used tests, the forward recursive and the rolling window tests. The results indicate that both the rolling and the recursive rolling approaches offer good finite sample performance in situations where there are one or two changes in the causal relationship over the sample period, although the performance of the rolling window algorithm seems to be the best. The testing strategies are illustrated in an empirical application that explores the causal impact of the slope of the yield curve on real economic activity in the United States over the period 1985–2013.

Keywords: Causality, Forward recursion, Hypothesis testing, Inflation, Output, Recursive rolling test, Rolling window, Yield curve

JEL classification: C12, C15, C32, G17

*This research was supported under Australian Research Council's Discovery Projects funding scheme (project number DP150101716). Phillips acknowledges support from the NSF under Grant No. SES 12-85258. Stan Hurn, Business School, Queensland University of Technology; Email: s.hurn@qut.edu.au. Peter C.B. Phillips, Yale University, University of Auckland, University of Southampton & Singapore Management University; Email: peter.phillips@yale.edu. Shuping Shi, Department of Economics, Macquarie University; Email: shuping.shi@mq.edu.au.

1 Introduction

Causality in econometrics typically relies on economic theory to justify the direction of causality between variables and to inform empirical testing of the causal hypotheses. In many situations, however, there is no relevant theoretical foundation for determining the causal ordering between variables that appear to be jointly determined. Even in celebrated cases, such as the money-income causality debate, there are difficulties in interpretation, test execution, and treatment of additional relevant variables. In these instances an empirical view of the concept of causality based on Granger (1969, 1988) has enjoyed widespread use in econometrics because of its eminent pragmatism. A variable X causes a variable Y in Granger's sense if taking into account past values of X enables better predictions to be made for Y, other things being equal. The popularity of Granger causality stems in part from the fact that it is not specific to a particular structural model but depends solely on the stochastic nature of variables, with no requirement to delimit some variables as dependent variables and others as independent variables.

It is well known that, among other things, testing for Granger causality is sensitive to the time period of estimation. The most well studied problems in this area are the money-income relationship (Stock and Watson, 1989; Thoma, 1994; Swanson, 1998; Psaradakis et al., 2005) and the energy consumption and economic output relationship (Stern, 2000, Balcilar et al., 2010, and Arora and Shi, 2015), where causal links are found in various subsamples. In view of the increasing importance of the financial sector in economic modeling, there is now a growing literature concerned with the detection of changes in patterns of systemic risk. For example, Billio et al. (2012) and Chen et al. (2013) use Granger causality to explore the causal links between banks and insurance companies and show that banks are a source of systemic risk to the rest of the system while insurers are victims of shocks. Their approach necessarily requires that crisis periods be defined exogenously. Other empirical approaches to systemic risk are similarly hampered by the need to choose sample periods appropriately (Acharya et al., 2011; Diebold and Yilmaz, 2013). These limitations point to the need for an endogenous approach to determining and dating changes in Granger causality.

Several methods have been employed in the literature to deal with the time-varying nature of causal relationships. These methods include a forward expanding window version of the Granger causality test (Thoma, 1994, and Swanson, 1998), a rolling window Granger causality test (Swanson, 1998, Balcilar et al., 2010, and Arora and Shi, 2015), and a Markov-switching

Granger causality test (Psaradakis et al., 2005). The recent literature for detecting and dating financial bubbles (Phillips and Yu, 2011; Phillips, Wu and Yu, 2011; Phillips, Shi and Yu, 2015a, 2015b; Leybourne, Kim and Taylor, 2007) recognises that in order to be useful to policymakers econometric methods for detecting changes in economic and financial structures must have at least two qualities. These qualities are a good positive detection rate, in order to ensure early and effective policy implementation, and a low false detection rate so that unnecessary interventions are avoided.

This paper proposes a new time-varying Granger causality test. The recursive method we implement was first proposed in Phillips, et al. (2015a, 2015b) for conducting real time detection of financial bubbles. The procedure involves intensive recursive calculations of the relevant test statistic, which in the current setting is a Wald test for Granger causality,[1] for all subsamples in a backward expanding sample sequence in which the final observation of all samples is the (current) observation of interest. Inference regarding the presence of Granger causality for this observation rely on the supremum taken over the values of all the test statistics in the entire recursion. As the observation of interest moves forward through the sample, the subsamples in which the recursive calculations are performed accordingly move forward and the whole sequence of calculations rolls ahead. This procedure is therefore called a recursive rolling algorithm.

Asymptotic distributions under the null hypothesis of no Granger causality are derived for the subsample Wald statistic process for forward and rolling window versions of the tests and the subsample sup Wald statistic process for the recursive rolling window procedure. Limit theory under the assumption of conditional (and hence unconditional) homoskedasticity is provided first. To take potential influences of conditional and unconditional heteroskedasticity into account, heteroskedastic consistent versions of the Wald and sup Wald statistics are proposed. The asymptotic distributions of these test statistics are then derived under the assumption of conditional heteroskedasticity of unknown form and a general form of non-stochastic time-varying unconditional heteroskedasticity. The major result for practical work that emerges from this limit theory is that the robust test statistics have the same pivotal asymptotics under homoskedasticity, conditional heteroskedasticity and unconditional heteroskedasticity.

The finite sample performance of forward, rolling and recursive rolling approaches in the context of Granger causality testing are then examined in detail. The data generating process employed in the simulations is a bivariate vector autoregressive (VAR) model, so that third

[1] In the original bubble-detection context the relevant test statistic was a right-sided unit root test.

variable causal effects are not taken into account in the present study. Under the alternative hypothesis, one or more episodes of unidirectional Granger causality are specified. In the simulation study, the means and standard deviations of the false detection proportion under the null hypothesis and the successful detection rate as well as the estimation accuracy of the causality switch-on and switch-off dates under the alternative hypothesis are reported. The false detection proportion is defined as the ratio between the number of false detections and the total number of hypotheses, while the successful detection rate is calculated as the proportion of samples finding the correct causality episode.

The simulation results suggest that both the recursive rolling and the rolling window approaches have good finite sample performance, with slightly higher false detection proportions but much higher correct detection rates than the forward expanding method. On the evidence presented here, the forward expanding window approach is identified as the least preferred method while the rolling window algorithm emerges as the most preferred option. Although the false detection rate of the rolling window test is usually slightly higher than that of the recursive rolling method, the difference is negligible. More importantly, the rolling window test provides a much higher successful detection rate and more accurate estimates of the change origination and termination dates than the recursive rolling window approach, particularly when there is more than one causal episodes in the sample.

These causality detection methods are used to investigate the causal impact of the yield curve spread on real economic activity in the United States over the period 1985 - 2013. The ability of the yield curve to predict real activity is a well-researched area in empirical macroeconomics. Some evidence of its predictive capability was first provided in the late 1980s and 1990s for various industrialized countries. The empirical literature also suggests that predictive relationships between the slope of the yield curve and macroeconomic activity have not been constant over time (Haubrich and Dombrosky,1996; Dotsey, 1998; Stock and Watson, 1999; Estrella, Rodrigues and Schich, 2003; Chauvet and Potter, 2005; Giacomini and Rossi, 2006; Benati and Goodhart, 2008; Chauvet and Senyuz, 2009; Kucko and Chinn, 2009; Hamilton, 2010). The test procedures developed in the present paper provide a natural mechanism for causal detection in which fragilities in a causal relationship can be captured through intensive subsample data analysis of the type recommended here.

The paper is organized as follows. Section 2 reviews the concept of Granger causality and describes the forward expanding window, rolling window, and the new recursive rolling Granger

causality tests. Section 3 gives the limit distributions of these test statistics under the null hypothesis of no causality and assumptions of conditional homoskedasticity, conditional heteroskedasticity and unconditional heteroskedasticity. Section 4 reports the results of simulations investigating performance characteristics of the various tests and dating strategies. In Section 5, the three procedures are used to investigate the causal impact of the yield curve spread on real economic activity in the United States over the last three decades. Section 6 is a brief conclusion. Proofs are given in the Appendices.

2 Identifying Changes in Causal Relationships

Consider the bivariate Gaussian VAR(p) model given by

$$y_{1t} = \phi_{10} + \sum_{i=1}^{p} \phi_{11i} y_{1\,t-i} + \sum_{i=1}^{p} \phi_{12i} y_{2\,t-i} + \varepsilon_{1t} \tag{1}$$

$$y_{2t} = \phi_{20} + \sum_{i=1}^{p} \phi_{21i} y_{1\,t-i} + \sum_{i=1}^{p} \phi_{22i} y_{2\,t-i} + \varepsilon_{2t}, \tag{2}$$

where y_{1t} and y_{2t} are dependent variables, p is the lag length and ε_{1t} and ε_{2t} are finite variance, martingale difference disturbances. If y_{2t} is important in predicting future values of y_{1t} over and above lags of y_{1t} alone, then y_{2t} is said to cause y_{1t} in Granger's sense, and vice versa. In equation (1), the null (non causal) hypotheses of interest are

$$\begin{array}{lll} H_0: & y_{2t} \nrightarrow y_{1t} & \phi_{121} = \phi_{122} = \phi_{123} = \cdots = \phi_{12p} = 0 \\ H_0: & y_{1t} \nrightarrow y_{2t} & \phi_{211} = \phi_{212} = \phi_{213} = \cdots = \phi_{21p} = 0\,, \end{array}$$

where the symbol $\nrightarrow$ reads "does not Granger cause".

To establish notation, the unrestricted VAR(p) may be written as

$$\mathbf{y}_t = \Phi_0 + \Phi_1 \mathbf{y}_{t-1} + \Phi_2 \mathbf{y}_{t-2} + \cdots + \Phi_p \mathbf{y}_{t-p} + \varepsilon_t, \tag{3}$$

or in multivariate regression format simply as

$$\mathbf{y}_t = \Pi \mathbf{x}_t + \varepsilon_t, \quad t = 1, ..., T \tag{4}$$

where $\mathbf{y}_t = (y_{1t}, y_{2t})'$, $\mathbf{x}_t = \left(1, \mathbf{y}'_{t-1}, \mathbf{y}'_{t-2}, \cdots, \mathbf{y}'_{t-p}\right)'$, and $\Pi_{2\times(2p+1)} = [\Phi_0, \Phi_1, \ldots, \Phi_p]$. The ordinary least squares (or unrestricted Gaussian maximum likelihood) estimator $\widehat{\Pi}$ has standard

limit theory under stationarity of the system (3) given by

$$\sqrt{T}\left(\widehat{\Pi}-\Pi\right)\xrightarrow{L}\mathrm{N}(0,\Sigma_{\Pi}), \tag{5}$$

where the variance matrix (for the row stacking of $\widehat{\Pi}$) is $\Sigma_{\Pi}=\Omega\otimes Q^{-1}$, with $\Omega=\mathbb{E}\left(\varepsilon_t\varepsilon_t'\right)$, and $Q=\mathbb{E}\left(\mathbf{x}_t\mathbf{x}_t'\right)$. In (5) and the remainder of the paper the notation $\xrightarrow{L}$ is used to signify convergence in distribution in Euclidean space. Let $\hat{\varepsilon}_t=\mathbf{y}_t-\widehat{\Pi}\mathbf{x}_t$ be the regression residuals, $\hat{\Omega}=T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t\hat{\varepsilon}_t'$ be the usual least squares estimate of the error covariance matrix Ω, and $\mathbf{X}'=[\mathbf{x}_1,...,\mathbf{x}_T]$ be the observation matrix of the regressors in (4).

The Wald test of the restrictions imposed by the null hypothesis $H_0: y_{2t}\nrightarrow y_{1t}$ has the simple form

$$W=\left[\mathbf{R}\,\mathrm{vec}(\widehat{\Pi})\right]'\left[\mathbf{R}\left(\hat{\Omega}\otimes\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)\mathbf{R}'\right]^{-1}\left[\mathbf{R}\,\mathrm{vec}(\widehat{\Pi})\right], \tag{6}$$

where $\mathrm{vec}(\widehat{\Pi})$ denotes the (row vectorized) $2\left(2p+1\right)\times 1$ coefficients of $\widehat{\Pi}$ and R is the $p\times 2(2p+1)$ selection matrix

$$\mathbf{R}=\begin{bmatrix}0&0&1&0&0&\cdots&0&0\cdots0&0\\0&0&0&0&1&\cdots&0&0\cdots0&0\\\vdots&\vdots&\vdots&\vdots&\vdots&\ddots&\vdots&\vdots\ddots\vdots&\vdots\\0&0&0&0&0&\cdots&1&0\cdots0&0\end{bmatrix}.$$

Each row of R picks one of the coefficients to set to zero under the non-causal null hypothesis. In the present case these are the p coefficients on the lagged values of y_{2t} in equation (1), $\phi_{12,1}\cdots\phi_{12,p}$. Under the null hypothesis and assumption of conditional homoskedasticity, the Wald test statistic is asymptotically χ^2_p, with degrees of freedom corresponding to the number of zero restrictions being tested.

As indicated in the introductory remarks, there is ample reason to expect causal relationships to change over the course of a time series sample. Any changes in economic policy, regulatory structure, governing institutions, or operating environments that impinge on the variables y_{1t} and y_{2t} may induce changes in causal relationships over time. In these circumstances, testing that is based on the entire sample using a statistic like (6) averages the sample information and inevitably destroys potentially valuable economic intelligence concerning the impact of changes in policy or structures. Testing for Granger casualty in exogenously defined subsamples of the data does provide useful information but does not enable the data to reveal the changes or change points. Consequently, the ultimate objective is to conduct tests that allow the change points to be determined (and hence identified) endogenously in the sample data.

Thoma (1994) and Swanson (1998) provide early attempts to isolate changes in causal relationships using forward expanding and rolling window Wald tests. Let f be the (fractional) observation of interest and f_0 be the minimum (fractional) window size required to estimate the model. Both recursive tests suggest computing a Wald statistic of the null hypothesis $H_0 : y_{2t} \nrightarrow y_{1t}$ for each observation from $[Tf_0]$ to T obtaining the full sequence of test statistics. The difference between these two procedures lies in the starting point of the regression used to calculate the Wald statistics. The ending points of the regressions (f_2) of both procedures are on the observation of interest, $f_2 = f$. For the Thoma (1994) procedure, the starting point of the regression (f_1) is fixed on the first available observation. As the observation of interest f moves forward from f_0 to 1, the regression window size expands (fractionally) from f_0 to 1 and hence is referred to as a forward expanding window Wald test. In contrast, the regression window size of the rolling procedure is a fixed constant. As the observation of interest (f and hence f_2) rolls forward from f_0 to 1, the starting point follows accordingly, maintaining a fixed distance from f_2. A significant change in causality is detected when an element of the Wald statistic sequence exceeds its corresponding critical value, so that the origination (termination) date of a change in causality is identified as the first observation whose test statistic value exceeds (goes below) its corresponding critical value.

While it is possible to use the recursive Wald statistics computed in this fashion to document any subsample instability in causal relationships, conclusions drawn on the basis of this approach may be incomplete. Drawing from the recent literature on dating multiple financial bubbles (Phillips, Shi and Yu, 2015a, 2015b), this paper proposes a test that is based on the supremum norm (sup) of a series of recursively calculated Wald statistics as follows. For each (fractional) observation of interest ($f \in [f_0, 1]$), the Wald statistics are computed for a backward expanding sample sequence. As above, the end point of the sample sequence is fixed at f. However, the starting point of the samples extends backwards from $(f - f_0)$, which is the minimum sample size to accommodate the regression, to 0. The Wald statistic obtained for each subsample regression is denoted by $W_{f_2}(f_1)$ and the sup Wald statistic is defined as

$$SW_f(f_0) = \sup \{W_{f_2}(f_1) : f_1 \in [0, f_2 - f_0], f_2 = f\}.$$

Heteroskedastic consistent Wald and sup Wald statistics are also proposed in the next section. Both the forward expanding and rolling window procedures are special cases of the new procedure

with f_1 fixed at value zero and $f_1 = f_2 - f_0$ respectively.[2] Importantly, all three procedures rely only on past information and can therefore be used for real-time monitoring. The added flexibility obtained by relaxing f_1 allows the procedure to search for the optimum starting point of the regression for each observation (in the sense of finding the largest Wald statistic). This flexibility accommodates re-initialization in the subsample to accord with (and thereby help to detect) any changes in structure or causal direction that may occur within the full sample.

Let f_e and f_f denote the origination and termination points in the causal relationship. These are estimated as the first chronological observation that respectively exceed or fall below the critical value. In a single switch case, the dating rules are giving by the following crossing times:

$$\text{Forward}: \hat{f}_e = \inf_{f\in[f_0,1]} \{f : W_f(0) > cv\} \text{ and } \hat{f}_f = \inf_{f\in[\hat{f}_e,1]} \{f : W_f(0) < cv\}, \tag{7}$$

$$\text{Rolling}: \hat{f}_e = \inf_{f\in[f_0,1]} \{f : W_f(f - f_0) > cv\} \text{ and } \hat{f}_f = \inf_{f\in[\hat{f}_e,1]} \{f : W_f(f - f_0) < cv\}, \tag{8}$$

$$\text{Recursive Rolling}: \hat{f}_e = \inf_{f\in[f_0,1]} \{f : SW_f(f_0) > scv\} \text{ and } \hat{f}_f = \inf_{f\in[\hat{f}_e,1]} \{f : SW_f(f_0) < scv\}, \tag{9}$$

where cv and scv are the corresponding critical values of the W_f and SW_f statistics. Now suppose there are multiple switches in the sample period. The origination and terminations of the ith causal relationship are denoted by f_{ie} and f_{if} for successive episodes $i = 1, 2, \ldots, I$ and the estimation of dates associated with the first episode (f_{1e} and f_{1f}) are exactly the same as those for the single switch case. For $i \geq 2$, f_{ie} and f_{if} are estimated as follows:

$$\text{Forward}: \hat{f}_{ie} = \inf_{f\in[\hat{f}_{i-1t},1]} \{f : W_f > cv\} \text{ and } \hat{f}_{if} = \inf_{f\in[\hat{f}_{ie},1]} \{f : W_f < cv\}, \tag{10}$$

$$\text{Rolling}: \hat{f}_{ie} = \inf_{f\in[\hat{f}_{i-1t},1]} \{f : W_f(f - f_0) > cv\} \text{ and } \hat{f}_{if} = \inf_{f\in[\hat{f}_{ie},1]} \{f : W_f(f - f_0) < cv\} \tag{11}$$

$$\text{Recursive Rolling}: \hat{f}_{ie} = \inf_{f\in[\hat{f}_{i-1t},1]} \{f : SW_f(f_0) > scv\} \text{ and } \hat{f}_{if} = \inf_{f\in[\hat{f}_{ie},1]} \{f : SW_f(f_0) < scv\}. \tag{12}$$

[2] It is assumed that the rolling window size equals f_0.

3 Asymptotic Distributions

The notation introduced in the previous section is now used for the general multivariate case, which allows both for changing coefficients in subsamples of the data and for changing (fractional) sample sizes in the asymptotic theory.

Let $\|\cdot\|$ denote the Euclidean norm, $\|.\|_p$ the L_p-norm so that $\|\mathbf{x}\|_p = \left(\mathbb{E}\|\mathbf{x}\|^p\right)^{1/p}$, and $\mathcal{F}_t = \sigma\{\varepsilon_t, \varepsilon_{t-1}..\}$ be the natural filtration. Consider an $n \times 1$ vector of dependent variables $\mathbf{y}_t$ whose dynamics follow a VAR(p) given by

$$\mathbf{y}_t = \Phi_0 + \Phi_1 \mathbf{y}_{t-1} + \Phi_2 \mathbf{y}_{t-2} + \cdots + \Phi_p \mathbf{y}_{t-p} + \varepsilon_t, \tag{13}$$

with constant coefficients over the subsample $t = \lfloor Tf_1 \rfloor, \ldots, \lfloor Tf_2 \rfloor$. The sample size in this regression is $T_w = [Tf_w]$ where $f_w = f_2 - f_1 \in [f_0, 1]$ for some fixed $f_0 \in (0,1)$.

Assumption (A0): *The roots of* $\left|\mathbf{I}_n - \Phi_1 z - \Phi_2 z^2 - \cdots - \Phi_p z^p\right| = 0$ *lie outside the unit circle.*

Under assumption **A0**, $\mathbf{y}_t$ has a simple moving average (linear process) representation in terms of the past history of ε_t

$$\mathbf{y}_t = \tilde{\Phi}_0 + \Psi(L)\varepsilon_t,$$

where $\Psi(L) = \left(\mathbf{I}_n - \Phi_1 L - \Phi_2 L^2 ... - \Phi_p L^p\right)^{-1} = \sum_{i=0}^{\infty} \Psi_i L^i$ with $\|\Psi_i\| < C\theta^i$ for some $\theta \in (0,1)$ and $\tilde{\Phi}_0 = \Psi(1)\Phi_0$. The model can be written in regression format as

$$\mathbf{y}_t = \Pi_{f_1,f_2} \mathbf{x}_t + \varepsilon_t, \tag{14}$$

in which $\mathbf{x}_t = \left(1, \mathbf{y}'_{t-1}, \mathbf{y}'_{t-2}, \cdots, \mathbf{y}'_{t-p}\right)'$ and $\Pi_{f_1,f_2} = [\Phi_0, \Phi_1, \ldots, \Phi_p]$.

The ordinary least squares (or Gaussian maximum likelihood with fixed initial conditions) estimator of the autoregressive coefficients is denoted by $\hat{\Pi}_{f_1,f_2}$ and defined as

$$\underset{n\times(np+1)}{\hat{\Pi}_{f_1,f_2}} = \left[\sum_{t=[Tf_1]}^{[Tf_2]} \mathbf{y}_t \mathbf{x}'_t\right] \left[\sum_{t=[Tf_1]}^{[Tf_2]} \mathbf{x}_t \mathbf{x}'_t\right]^{-1}.$$

Let $k = np + 1$ and let $\hat{\pi}_{f_1,f_2} \equiv \text{vec}\left(\hat{\Pi}_{f_1,f_2}\right)$ denote the (row vectorized) $nk \times 1$ coefficients resulting from an OLS regression of each of the elements of $\mathbf{y}_t$ on $\mathbf{x}_t$ for a sample running from

$[Tf_1]$ to $[Tf_2]$ given by

$$\hat{\pi}_{f_1,f_2} = \left[\begin{array}{cccc} \hat{\pi}_{1,f_1,f_2} & \hat{\pi}_{2,f_1,f_2} & \ldots & \hat{\pi}_{n,f_1,f_2} \end{array} \right]',$$

in which $\hat{\pi}_{i,f_1,f_2} = \left[\sum_{t=[Tf_1]}^{[Tf_2]} y_{it}\mathbf{x}_t'\right]\left[\sum_{t=[Tf_1]}^{[Tf_2]} \mathbf{x}_t\mathbf{x}_t'\right]^{-1}$. It follows that

$$\hat{\pi}_{f_1,f_2} - \pi_{f_1,f_2} = \left[\mathbf{I}_n \otimes \sum_{t=[Tf_1]}^{[Tf_2]} \mathbf{x}_t\mathbf{x}_t'\right]^{-1}\left[\sum_{t=[Tf_1]}^{[Tf_2]} \xi_t\right],$$

where π_{f_1,f_2} denotes the corresponding population coefficients and $\xi_t \equiv \varepsilon_t \otimes \mathbf{x}_t$. The corresponding estimate of the residual variance matrix $\mathbf{\Omega}$ is $\hat{\Omega}_{f_1,f_2} = T_w^{-1}\sum_{t=[Tf_1]}^{[Tf_2]} \hat{\varepsilon}_t\hat{\varepsilon}_t'$, where $\hat{\varepsilon}_t' = [\hat{\varepsilon}_{1t}, \hat{\varepsilon}_{2t}, \ldots, \hat{\varepsilon}_{nt}]$ and $\hat{\varepsilon}_{it} = y_{it} - \mathbf{x}_t'\hat{\pi}_{i,f_1,f_2}$.

The primary concern is the distribution of the Wald test for causality under the null hypothesis. In this instance, the coefficient matrix Π_{f_1,f_2} has constant coefficients for the entire sample $[f_1, f_2]$. The null hypothesis for the causality test falls in the general framework of linear hypotheses of the form $H_0 : R\pi_{f_1,f_2} = 0$, where R is a coefficient restriction matrix (of full row rank d). Given (f_1, f_2), the usual form of the Wald statistic for this null hypothesis is

$$W_{f_2}(f_1) = (\mathbf{R}\hat{\pi}_{f_1,f_2})'\left\{\mathbf{R}\left[\hat{\mathbf{\Omega}}_{f_1,f_2} \otimes \left(\sum_{t=[Tf_1]}^{[Tf_2]} \mathbf{x}_t\mathbf{x}_t'\right)^{-1}\right]\mathbf{R}'\right\}^{-1}(\mathbf{R}\hat{\pi}_{f_1,f_2}). \tag{15}$$

The heteroskedasiticity consistent version of the Wald statistic is denoted by $W_{f_2}^*(f_1)$ and is defined as

$$W_{f_2}^*(f_1) = T_w(\mathbf{R}\hat{\pi}_{f_1,f_2})'\left[\mathbf{R}\left(\hat{\mathbf{V}}_{f_1,f_2}^{-1}\hat{\mathbf{W}}_{f_1,f_2}\hat{\mathbf{V}}_{f_1,f_2}^{-1}\right)\mathbf{R}'\right]^{-1}(\mathbf{R}\hat{\pi}_{f_1,f_2}), \tag{16}$$

where $\hat{\mathbf{V}}_{f_1,f_2} \equiv \mathbf{I}_n \otimes \hat{\mathbf{Q}}_{f_1,f_2}$ with $\hat{\mathbf{Q}}_{f_1,f_2} \equiv \frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]} \mathbf{x}_t\mathbf{x}_t'$, and $\hat{\mathbf{W}}_{f_1,f_2} \equiv \frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]} \hat{\xi}_t\hat{\xi}_t'$ with $\hat{\xi}_t \equiv \hat{\varepsilon}_t \otimes \mathbf{x}_t$. The heteroskedasticity consistent sup Wald statistic is

$$SW_f^*(f_0) := \sup\left\{W_{f_2}^*(f_1) : f_1 \in [0, f_2 - f_0], f_2 = f\right\}.$$

As the fractions (f_1, f_2) vary, the statistics $W_{f_2}(f_1)$ and $W_{f_2}^*(f_1)$ are stochastic processes indexed with (f_1, f_2).

3.1 Homoskedasticity

Under the assumption of homoskedasticity, the innovations are stationary, conditionally homoskedastic martingale differences satisfying either of the following two conditions.

Assumption (A1): $\{\varepsilon_t, \mathcal{F}_t\}$ *is a strictly stationary and ergodic martingale difference sequence (mds) with* $\mathbb{E}(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}) = \mathbf{\Omega}$ *a.s. and positive definite* Ω.

Assumption (A2): $\{\varepsilon_t, \mathcal{F}_t\}$ *is a covariance stationary mds with* $\mathbb{E}(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}) = \mathbf{\Omega}$ *a.s., positive definite* $\mathbf{\Omega}$, *and* $\sup_t \mathbb{E}\|\varepsilon_t\|^{4+c} < \infty$ *for some* $c > 0$.

Lemma 3.1 *Given the model (13), under assumption* ***A0*** *and* ***A1*** *or* ***A2*** *and the null (maintained) hypothesis of an unchanged coefficient matrix* $\Pi_{f_1,f_2} = \Pi$ *for all (fractional) subsamples* (f_1, f_2) *we have*

$$
\begin{aligned}
&(a)\ \ \hat{\pi}_{f_1,f_2} \to_{a.s.} \pi_{f_1,f_2} = vec(\Pi_{f_1,f_2}), \\
&(b)\ \ \hat{\mathbf{\Omega}}_{f_1,f_2} \to_{a.s.} \mathbf{\Omega}, \\
&(c)\ \sqrt{T}(\hat{\pi}_{f_1,f_2} - \pi_{f_1,f_2}) \Rightarrow [\mathbf{I}_n \otimes \mathbf{Q}]^{-1} \left[\frac{B(f_2) - B(f_1)}{f_w}\right],
\end{aligned}
$$

where B *is vector Brownian motion with covariance matrix* $\mathbf{\Omega} \otimes \mathbf{Q}$, *where* $\mathbf{Q} = \mathbb{E}(\mathbf{x}_t\mathbf{x}_t') > 0$, *and* $\hat{\pi}_{f_1,f_2}$ *and* $\hat{\mathbf{\Omega}}_{f_1,f_2}$ *are the least squares estimators of* π_{f_1,f_2} *and* $\mathbf{\Omega}$. *The finite dimensional distribution of the limit in* (c) *for fixed* f_2 *and* f_1 *is* $N\left(\mathbf{0}, \frac{1}{f_w}\mathbf{\Omega} \otimes \mathbf{Q}^{-1}\right)$.

The proof of lemma 3.1 is given in the Appendix A. From part (c) and for fixed (f_1, f_2) the asymptotic variance-covariance matrix of $\sqrt{T}(\hat{\pi}_{f_1,f_2} - \pi)$ is $f_w^{-1}\left(\mathbf{\Omega} \otimes \mathbf{Q}^{-1}\right)$ and is dependent on the fractional window size f_w. The limit in (c) may be interpreted as a linear functional of the process $B(\cdot)$.

Note that under **A2**, the limit of the matrix $\hat{\mathbf{W}}_{f_1,f_2}$ that appears in the heteroskedastic consistent Wald statistic (16) would be given by $\mathbf{\Omega} \otimes \mathbf{Q}$ and the asymptotic covariance matrix would simplify as follows

$$
\hat{\mathbf{V}}_{f_1,f_2}^{-1}\hat{\mathbf{W}}_{f_1,f_2}\hat{\mathbf{V}}_{f_1,f_2}^{-1} \to_{a.s} (\mathbf{I}_n \otimes \mathbf{Q})^{-1}(\mathbf{\Omega} \otimes \mathbf{Q})(\mathbf{I}_n \otimes \mathbf{Q})^{-1} = \mathbf{\Omega} \otimes \mathbf{Q}^{-1}.
$$

In this case, therefore, the heteroskedasitic consistent test statistics, $W_{f_2}^*(f_1)$ and $SW_f^*(f_0)$, reduce to the conventional Wald and sup Wald statistics of $W_{f_2}(f_1)$ and $SW_f(f_0)$.

Proposition 3.1 *Under **A0** and **A1** or **A2**, the null hypothesis* $\mathbf{R}\pi_{f_1,f_2} = \mathbf{0}$*, and the maintained null of an unchanged coefficient matrix* $\Pi_{f_1,f_2} = \Pi$ *for all subsamples, the subsample Wald process and sup Wald statistic converge weakly to the following limits*

$$W_{f_2}(f_1) \Rightarrow \left[\frac{W_d(f_2) - W_d(f_1)}{(f_2 - f_1)^{1/2}}\right]' \left[\frac{W_d(f_2) - W_d(f_1)}{(f_2 - f_1)^{1/2}}\right] \tag{17}$$

$$SW_f(f_0) \xrightarrow{L} \sup_{f_1 \in [0, f_2 - f_0], f_2 = f} \left[\frac{W_d(f_2) - W_d(f_1)}{(f_2 - f_1)^{1/2}}\right]' \left[\frac{W_d(f_2) - W_d(f_1)}{(f_2 - f_1)^{1/2}}\right] \tag{18}$$

$$= \sup_{f_w \in [f_0, f_2], f_2 = f} \left[\frac{W_d(f_w)' W_d(f_w)}{f_w}\right] \tag{19}$$

where W_d *is vector Brownian motion with covariance matrix* $\mathbf{I}_d$ *and* d *is the number of restrictions (the rank of* R*) under the null.*

The proof of Proposition 3.1 is given in the Appendix A. The limit process that appears in (17) is a quadratic functional of the limit process $W_d(\cdot)$. Its finite dimensional distribution for fixed f_1 and f_2 is χ^2_d, whereas the sup functional that appears in (18) and (19) involves the supremum of the continuous stochastic process $\frac{W_d(f_w)' W_d(f_w)}{f_w}$ taken over $f_w \in [f_0, f_2]$ with $f_2 = f$.

3.2 Conditional heteroskedasticity of unknown form

The conditional heteroskedasticity case requires the following additional assumption.

Assumption (A3): $\{\varepsilon_t, \mathcal{F}_t\}$ *is an mds satisfying the following conditions:*

(i) ε_t *is strongly uniformly integrable with a dominating random variable* ε *that satisifies* $\mathbb{E}\left(\|\varepsilon\|^{4+c}\right) < \infty$ *for some* $c > 0$*;*

(ii) $T^{-1}\sum_{t=1}^{T} \mathbb{E}\left(\varepsilon_t \varepsilon_t' | \mathcal{F}_{t-1}\right) \to_{a.s.} \mathbf{\Omega}$ *where* $\mathbf{\Omega}$ *is positive definite with elements* $\mathbf{\Omega} = (\Omega_{ij})$*;*

(iii) $T^{-1}\sum_{t=1}^{T} \mathbb{E}\left(\varepsilon_{i,t}^2 | \mathcal{F}_{t-1}\right) \varepsilon_{t-s} \to_{a.s.} 0$ *and* $T^{-1}\sum_{t=1}^{T} \mathbb{E}\left(\varepsilon_{i,t}\varepsilon_{j,t} | \mathcal{F}_{t-1}\right) \varepsilon_{t-s}\varepsilon_{t-s}' \to_{a.s.} \Omega_{ij}\mathbf{\Omega}$ *for* $i, j = 1, \cdots, n$ *and* $s \geq 1$.

Strong uniform integrability is commonly assumed in cases of conditional and unconditional heterogeneity (see, for instance, Phillips and Solo, (1992), Remarks 2.4(i) and 2.8 (i) and (ii)). Assumption **A3** implies that $\{\varepsilon_t\}$ is serially uncorrelated, unconditionally homoskedastic if

$\mathbb{E}(\varepsilon_t \varepsilon_t') = \mathbf{\Omega}$ for all t (and hence covariance stationary in that case), but potentially conditionally heteroskedastic. **A3** allows, among other possibilities, stable ARCH or GARCH errors. Note that **A3**(i) is equivalent to assuming that

$$\sup_t \mathbb{E} \|\varepsilon_t\|^{4+c} < \infty \text{ for some } c > 0,$$

a condition that is often used in work involving conditional and unconditional heteroskedasticity (see, for example, Boswijk et al. (2013) and Bodnar and Zabolotskyy (2011)). **A3**(iii) is required for Lemma 3.3(b), and is used by Hannan and Heyde (1972, Theorem 2), Gonçalves and Kilian (2004), and Boswijk et al. (2013).

Lemma 3.2 *Under **A0** and **A3**, for all $f_2, f_1 \in [0,1]$ and $f_2 > f_1$,*

(a) $T_w^{-1} \sum_{t=[Tf_1]}^{[Tf_2]} \varepsilon_t \to_{a.s} \mathbf{0}$,

(b) $T_w^{-1} \sum_{t=[Tf_1]}^{[Tf_2]} \varepsilon_t \varepsilon_t' \to_{a.s} \mathbf{\Omega}$,

(c) $T_w^{-1} \sum_{t=[Tf_1]}^{[Tf_2]} \varepsilon_t \varepsilon_s' \to_{a.s} \mathbf{0}$,

(d) $T_w^{-1} \sum_{t=[Tf_1]}^{[Tf_2]} \mathbf{x}_t \varepsilon_t' \to_{a.s} \mathbf{0}$,

(e) $T_w^{-1} \sum_{t=[Tf_1]}^{[Tf_2]} \mathbf{x}_t \mathbf{x}_t' \to_{a.s} \mathbf{Q}$, *where* $\mathbf{Q}$ *is defined as*

$$\mathbf{Q} \equiv \begin{bmatrix} 1 & \mathbf{1}_p' \otimes \tilde{\Phi}_0' \\ \mathbf{1}_p \otimes \tilde{\Phi}_0 & \mathbf{I}_p \otimes \tilde{\Phi}_0 \tilde{\Phi}_0' + \mathbf{\Theta} \end{bmatrix} \text{ with } \mathbf{\Theta} = \sum_{i=0}^{\infty} \begin{bmatrix} \Psi_i \mathbf{\Omega} \Psi_i' & \cdots & \Psi_{i+p-1} \mathbf{\Omega} \Psi_i' \\ \vdots & \ddots & \vdots \\ \Psi_i \mathbf{\Omega} \Psi_{i+p-1}' & \cdots & \Psi_i \mathbf{\Omega} \Psi_i' \end{bmatrix}.$$

The proof of this Lemma is in Appendix B. In view of the covariance stationarity of ε_t, Lemma 3.2 holds for all possible fixed fractions of data with $f_2, f_1 \in [0,1]$ and $f_2 > f_1$. However, this is not in general true under global covariance stationary (Davidson, 1994) or nonstationary volatility settings, where the right hand side of the statements in Lemma 3.2 may depend on f_1 and f_2.

Recalling that $\xi_t \equiv \varepsilon_t \otimes x_t$ it is now shown that $\{\xi_t\}$ obeys a martingale invariance principle as in Theorem 3 of Brown (1971), for example. This invariance result requires the two conditions stated in Lemma 3.3 below.

Lemma 3.3 *Under **A0** and **A3**, the mds $\{\xi_t, \mathcal{F}_t\}$ satisfies the following Lindeberg and stability conditions:*

(a) For every $\delta > 0$

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left\{\|\xi_t\|^2 .\mathbf{1}\left(\|\xi_t\| \geq \sqrt{T}\delta\right)|\mathcal{F}_{t-1}\right\} \xrightarrow{p} 0, \tag{20}$$

(b) $T^{-1}\sum_{t=1}^{T}\mathbb{E}\{\xi_t\xi_t'|\mathcal{F}_{t-1}\} \to_{a.s} \mathbf{W}$, where $\mathbf{W} = \left\{\mathbf{W}^{(i,j)}\right\}_{i,j\in[1,n]}$ with block partitioned elements

$$\mathbf{W}^{(i,j)} = \begin{bmatrix} \Omega_{ij} & \mathbf{1}_p' \otimes \Omega_{ij}\tilde{\Phi}_0' \\ \mathbf{1}_p \otimes \Omega_{ij}\tilde{\Phi}_0 & \mathbf{I}_p \otimes \Omega_{ij}\tilde{\Phi}_0\tilde{\Phi}_0' + \mathbf{\Xi}^{(i,j)} \end{bmatrix}$$

and

$$\mathbf{\Xi}^{(i,j)} \equiv \sum_{i=0}^{\infty}\begin{bmatrix} \Psi_i\Omega_{ij}\mathbf{\Omega}\Psi_i' & \cdots & \Psi_{i+p-1}\Omega_{ij}\mathbf{\Omega}\Psi_i' \\ \vdots & \ddots & \vdots \\ \Psi_i\Omega_{ij}\mathbf{\Omega}\Psi_{i+p-1}' & \cdots & \Psi_i\Omega_{ij}\mathbf{\Omega}\Psi_i' \end{bmatrix}$$

The proof of this lemma is in Appendix B. Under Lemma 3.3, partial sums of $\{\xi_t\}$ satisfy a martingale invariance principle, so that

$$\frac{1}{\sqrt{T}}\sum_{t=[Tf_1]}^{[Tf_2]}\xi_t \Rightarrow B(f_2) - B(f_1), \tag{21}$$

where the limit process in (21) is a linear functional of the vector Brownian motion $B(\cdot)$ with covariance matrix $\mathbf{W}$. Here and elsewhere, the notation $\Rightarrow$ is used to signify weak convergence in the Skorohod space $D[0,1]$.

Lemma 3.4 *Under **A0** and **A3**,*

(a) $\hat{\pi}_{f_1,f_2} \to_{a.s.} \pi_{f_1,f_2}$,

(b) $\hat{\mathbf{\Omega}}_{f_1,f_2} \to_{a.s.} \mathbf{\Omega}$,

(c) $\sqrt{T_w}(\hat{\pi}_{f_1,f_2} - \pi_{f_1,f_2}) \Rightarrow f_w^{-1/2}\mathbf{V}^{-1}[\mathbf{B}(f_2) - \mathbf{B}(f_1)]$, where $\mathbf{V} = \mathbf{I}_n \otimes \mathbf{Q}$ and B is vector Brownian motion with covariance matrix $\mathbf{W}$.

(d) $T_w^{-1}\sum_{t=[Tf_1]}^{[Tf_2]}\hat{\xi}_t\hat{\xi}_t' \to_{a.s.} \mathbf{W}$, where $\hat{\xi}_t \equiv \hat{\varepsilon}_t \otimes \mathbf{x}_{t-1}$.

Proposition 3.2 *Under **A0** and **A3**, the null hypothesis* $\mathbf{R}\pi_{f_1,f_2} = \mathbf{0}$*, and the maintained hypothesis of an unchanged coefficient matrix* $\Pi_{f_1,f_2} = \Pi$ *for all subsamples, the subsample heteroskedastic consistent Wald process and sup Wald statistic converge weakly to the following limits*

$$W^*_{f_2}(f_1) \Rightarrow \left[\frac{W_d(f_2) - W_d(f_1)}{(f_2 - f_1)^{1/2}}\right]' \left[\frac{W_d(f_2) - W_d(f_1)}{(f_2 - f_1)^{1/2}}\right],$$
$$SW^*_f(f_0) \xrightarrow{L} \sup_{f_w \in [f_0, f_2], f_2 = f} \left[\frac{W_d(f_w)' W_d(f_w)}{f_w}\right],$$

where W_d *is vector Brownian motion with covariance matrix* $\mathbf{I}_d$ *and* d *is the number of restrictions (the rank of* R*) under the null.*

If the presence of conditional heteroskdasticity in $\mathbf{y}_t$ is ignored in the construction of the (conventional) test statistic (15), the Wald and sup Wald statistics have non-standard asymptotic distributions as detailed in the following result.

Proposition 3.3 *Under **A0** and **A3**, the null hypothesis* $\mathbf{R}\pi_{f_1,f_2} = \mathbf{0}$*, and the maintained hypothesis of an unchanged coefficient matrix* $\Pi_{f_1,f_2} = \Pi$ *for all subsamples, the subsample Wald process and sup Wald statistic converge weakly to the following limits*

$$W_{f_2}(f_1) \Rightarrow \left[\frac{W_{nk}(f_2) - W_{nk}(f_1)}{f_w^{1/2}}\right]' \mathbf{A}\mathbf{B}^{-1}\mathbf{A}' \left[\frac{W_{nk}(f_2) - W_{nk}(f_1)}{f_w^{1/2}}\right]',$$
$$SW_{f_2}(f_0) \xrightarrow{L} \sup_{f_1 \in [0, f_2 - f_0], f_2 = f} \left[\frac{W_{nk}(f_2) - W_{nk}(f_1)}{f_w^{1/2}}\right]' \mathbf{A}\mathbf{B}^{-1}\mathbf{A}' \left[\frac{W_{nk}(f_2) - W_{nk}(f_1)}{f_w^{1/2}}\right]',$$

where W_{nk} *is vector Brownian motion with covariance matrix* $\mathbf{I}_{nk}$*,* $\mathbf{A} = \mathbf{W}^{1/2}\mathbf{V}^{-1}\mathbf{R}'$*, and* $\mathbf{B} = \mathbf{R}(\mathbf{\Omega} \otimes \mathbf{Q})\mathbf{R}'$*.*

3.3 Unconditional heteroskedasticity

Consider an array error specification of the form $\varepsilon_t := \mathbf{G}(t/T)\mathbf{u}_t$ where the matrix function $\mathbf{G}(\cdot)$ and error process $\mathbf{u}_t$ are defined below in Assumptions **A4** and **A5.** This framework involves a time evolving error variance matrix that allows for unconditional error heteroscedasticity.

Assumption (A4): *The matrix function* $\mathbf{G}(\cdot)$ *is nonstochastic, measurable and uniformly bounded on the interval* $(-\infty, 1]$ *with a finite numbers of points of discontinuity, and satisfies a*

Lipschitz condition except at points of discontinuity.

This formulation of heteroskedasticity was used in Phillips and Xu (2006) for the univariate case and Bodnar and Zabolotskyy (2011) and Boswijk et al. (2013) for the multivariate case. **A4** implies that each element of the matrix $\mathbf{G}(r) = \{g_{ij}(r)\}_{i,j=1,\ldots,n}$ is integrable on $[0,1]$ up to any finite order, $\int_0^1 |g_{ij}(r)|^m dr < \infty$ for all $m > 0$. The function $\mathbf{G}(\cdot)$ is defined for $r \in (-\infty, 1]$ since the initial conditions are in the infinite past and we make use of the infinite moving average representation of the process $\{\mathbf{y}_t\}$. Since $\{\varepsilon_t\}$ and $\{\mathbf{y}_t\}$ are triangular arrays, an additional subscript T should be included to signify the presence of an array but will be subsumed within the usual time series notation for simplicity in what follows.

Assumption (A5): $\{\mathbf{u}_t, \mathcal{F}_t\}$ *is an mds satisfying*

(i) $\mathbf{u}_t$ *is strongly uniformly integrable with dominating random variable* $\mathbf{u}$ *that satisifies* $\mathbb{E}\left(\|\mathbf{u}\|^{4+c}\right) < \infty$ *for some* $c > 0$*;*

(ii) $\mathbb{E}(\mathbf{u}_t\mathbf{u}_t'|\mathcal{F}_{t-1}) = \mathbf{I}_n$ *a.s.*

A5 implies that $\{\mathbf{u}_t\}$ is serially uncorrelated and homoskedastic (both conditionally and unconditionally) and hence covariance stationary. Note that **A5**(i) implies that $\sup_t \mathbb{E}\|\mathbf{u}_t\|^{4+c} < \infty$ for some $c > 0$. As in Phillips and Xu (2006) and Bodnar and Zabolotskyy (2011), it follows that $\mathbb{E}(\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}) = \mathbb{E}\left(\mathbf{G}(t/T)\mathbf{u}_t\mathbf{u}_t'\mathbf{G}(t/T)'|\mathcal{F}_{t-1}\right) = \mathbf{G}(t/T)\mathbf{G}(t/T)'$ and $\mathbb{E}(\varepsilon_t\varepsilon_t') = \mathbf{G}(t/T)\mathbf{G}(t/T)'$. Both conditional and unconditional variances of $\{\varepsilon_t\}$ are nonstochastic and time-varying of the form $\mathbf{G}(t/T)\mathbf{G}(t/T)'$. Unlike Phillips and Xu (2006) and Bodnar and Zabolotskyy (2011), the assumption of strong (α-)mixing $\{\mathbf{u}_t\}$ is not required.

Lemma 3.5 *Under* ***A0, A4*** *and* ***A5****,*

(a) $T_w^{-1}\sum_{t=[Tf_1]}^{[Tf_2]}\varepsilon_t \to_{a.s} \mathbf{0}$*;*

(b) $T_w^{-1}\sum_{t=[Tf_1]}^{[Tf_2]}\varepsilon_t\varepsilon_t' \to_{a.s} \mathbf{\Omega}_{f_1,f_2} \equiv \int_{f_1}^{f_2}\mathbf{G}(r)\mathbf{G}(r)'dr$;

(c) $T_w^{-1}\sum_{t=[Tf_1]}^{[Tf_2]}\left(\mathbf{y}_{t-h} - \tilde{\Phi}_0\right)\left(\mathbf{y}_{t-h-j} - \tilde{\Phi}_0\right)' \to_{a.s} \int_{f_1}^{f_2}\sum_{i=0}^{\infty}\Psi_{i+j}\mathbf{G}(r)\mathbf{G}(r)'\Psi_i'dr$;

(d) $T_w^{-1}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbf{x}_t\varepsilon_t' \to_{a.s} \mathbf{0}$;

(e) $T_w^{-1}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbf{x}_t\mathbf{x}_t' \to_{a.s} \mathbf{Q}_{f_1,f_2}$, *where the* $n\times n$ *matrix* $\mathbf{Q}_{f_1,f_2}$ *is defined as*

$$\mathbf{Q}_{f_1,f_2}=\begin{bmatrix} 1 & \mathbf{1}_p'\otimes\tilde{\Phi}_0' \\ \mathbf{1}_p\otimes\tilde{\Phi}_0 & \mathbf{I}_p\otimes\tilde{\Phi}_0\tilde{\Phi}_0'+\mathbf{\Theta}_{f_1,f_2}\end{bmatrix}$$

with

$$\mathbf{\Theta}_{f_1,f_2}:=\sum_{i=0}^{\infty}\int_{f_1}^{f_2}\begin{bmatrix} \Psi_i\mathbf{G}(r)\mathbf{G}(r)'\Psi_i' & \cdots & \Psi_{i+p-1}\mathbf{G}(r)\mathbf{G}(r)'\Psi_i' \\ \vdots & \ddots & \vdots \\ \Psi_i\mathbf{G}(r)\mathbf{G}(r)'\Psi_{i+p-1}' & \cdots & \Psi_i\mathbf{G}(r)\mathbf{G}(r)'\Psi_i'\end{bmatrix}dr.$$

The proof is given in Appendix C. Once again partial sums of ξ_t satisfy a martingale invariance principle, which is verified using the two conditions established in Lemma 3.6.

Lemma 3.6 *Under* ***A0, A4*** *and* ***A5****,* $\{\xi_t,\mathcal{F}_t\}$ *is an mds satisfying the following Lindeberg and stability conditions:*

(a) $T^{-1}\sum_{t=1}^{T}\mathbb{E}\left\{\|\xi_t\|^2.\mathbf{1}\left(\|\xi_t\|\geq\sqrt{T}\delta\right)|\mathcal{F}_{t-1}\right\}\xrightarrow{p}0$, *for all* $\delta>0$; *and*

(b) $T_w^{-1}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbb{E}\{\xi_t\xi_t'|\mathcal{F}_{t-1}\}\to_{a.s}\mathbf{W}_{f_1,f_2}$, *where* $\mathbf{W}_{f_1,f_2}=\left\{\mathbf{W}_{f_1,f_2}^{(i,j)}\right\}_{i,j\in[1,n]}$ *with block partitioned form*

$$\mathbf{W}_{f_1,f_2}^{(i,j)}=\begin{bmatrix} \int_{f_1}^{f_2}\sum_{q=1}^{n}g_{iq}(r)g_{jq}(r)dr & \mathbf{1}_p'\otimes\int_{f_1}^{f_2}\sum_{q=1}^{n}g_{iq}(r)g_{jq}(r)dr\tilde{\Phi}_0' \\ \mathbf{1}_p\otimes\int_{f_1}^{f_2}\sum_{q=1}^{n}g_{iq}(r)g_{jq}(r)dr\tilde{\Phi}_0 & \mathbf{I}_p\otimes\int_{f_1}^{f_2}\sum_{q=1}^{n}g_{iq}(r)g_{jq}(r)dr\tilde{\Phi}_0\tilde{\Phi}_0'+\mathbf{\Xi}_{f_1,f_2}^{(i,j)}\end{bmatrix},$$

and

$$\mathbf{\Xi}_{f_1,f_2}^{(i,j)}\equiv\sum_{i=0}^{\infty}\begin{bmatrix} \Psi_i\mathbf{\Lambda}_{f_1,f_2}^{(i,j)}\Psi_i' & \cdots & \Psi_{i+p-1}\mathbf{\Lambda}_{f_1,f_2}^{(i,j)}\Psi_i' \\ \vdots & \ddots & \vdots \\ \Psi_i\mathbf{\Lambda}_{f_1,f_2}^{(i,j)}\Psi_{i+p-1}' & \cdots & \Psi_i\mathbf{\Lambda}_{f_1,f_2}^{(i,j)}\Psi_i'\end{bmatrix},$$

$$\mathbf{\Lambda}_{f_1,f_2}^{(i,j)}=\int_{f_1}^{f_2}\sum_{q=1}^{n}g_{iq}(r)g_{jq}(r)\mathbf{G}(r)\mathbf{G}(r)'dr.$$

The proof is given in Appendix C. Under Lemma 3.6, partial sums of ξ_t satisfy a martingale invariance principle, so that

$$\frac{1}{\sqrt{T}}\sum_{t=[Tf_1]}^{[Tf_2]}\xi_t\to B^*(f_2)-B^*(f_1), \tag{22}$$

where B^* is vector Brownian motion with covariance matrix $\mathbf{W}_{f_1,f_2}$. Using (22) we find the limit behavior of the estimator process $\hat{\pi}_{f_1,f_2}$ and the heteroskedasticity consistent Wald statistic process $W^*_{f_2}(f_1)$.

Lemma 3.7 *Under **A0, A4** and **A5**, we have*

(a) $\hat{\pi}_{f_1,f_2} \to_{a.s.} \pi_{f_1,f_2}$,

(b) $T_w^{-1} \sum_{t=[Tf_1]}^{[Tf_2]} \hat{\varepsilon}_t \hat{\varepsilon}_t' \to_{a.s.} \mathbf{\Omega}_{f_1,f_2}$,

(c) $\sqrt{T_w}(\hat{\pi}_{f_1,f_2} - \pi_{f_1,f_2}) \Rightarrow f_w^{-1/2} \mathbf{V}_{f_1,f_2}^{-1} [B^*(f_2) - B^*(f_1)]$, *where* $\mathbf{V}_{f_1,f_2} = \mathbf{I}_n \otimes \mathbf{Q}_{f_1,f_2}$ *and* B^* *is vector Brownian motion with covariance matrix* $\mathbf{W}_{f_1,f_2}$.

(d) $T_w^{-1} \sum_{t=[Tf_1]}^{[Tf_2]} \hat{\xi}_t \hat{\xi}_t' \to_{a.s.} \mathbf{W}_{f_1,f_2}$, *where* $\hat{\xi}_t \equiv \hat{\varepsilon}_t \otimes \mathbf{x}_{t-1}$.

Proposition 3.4 *Under **A0, A4** and **A5**, the null hypothesis* $\mathbf{R}\pi_{f_1,f_2} = \mathbf{0}$, *and the maintained hypothesis of an unchanged coefficient matrix* $\Pi_{f_1,f_2} = \Pi$ *for all subsamples, the subsample heteroskedastic consistent Wald process and sup Wald statistic converge weakly to the following limits*

$$W^*_{f_2}(f_1) \Rightarrow \left[\frac{W_d(f_2) - W_d(f_1)}{(f_2 - f_1)^{1/2}}\right]' \left[\frac{W_d(f_2) - W_d(f_1)}{(f_2 - f_1)^{1/2}}\right]$$
$$SW^*_f(f_0) \xrightarrow{L} \sup_{f_w \in [f_0,f_2], f_2 = f} \left[\frac{W_d(f_w)' W_d(f_w)}{f_w}\right],$$

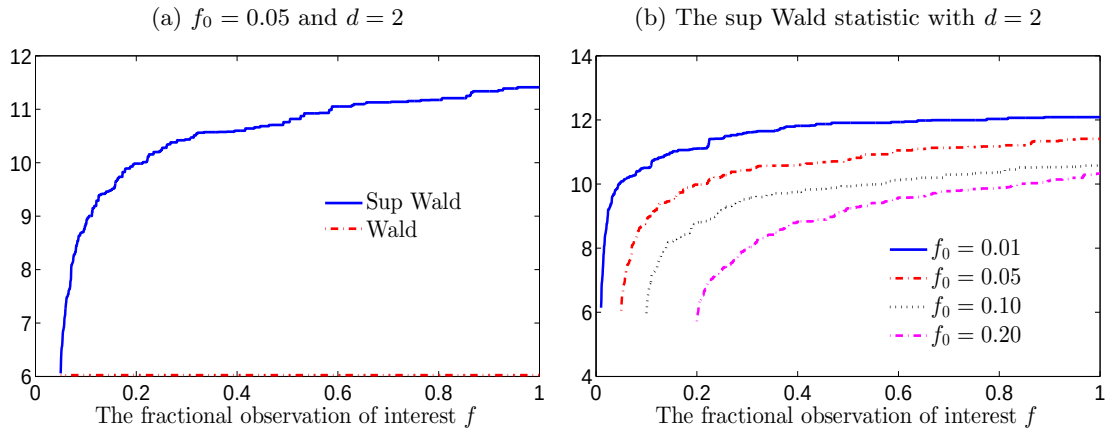
where W_d *is vector Brownian motion with covariance matrix* $\mathbf{I}_d$ *and* d *is the number of restrictions (the rank of* R*) under the null.*

The presence of nonstochastic and time-varying errors affects the limit behavior of the standard Wald statistic, which no longer has the limit (17). In consequence, use of the limit theory (19) for the sup Wald statistic may lead to invalid and distorted inference.

Proposition 3.5 *Under **A0, A4** and **A5**, the null hypothesis* $\mathbf{R}\pi_{f_1,f_2} = \mathbf{0}$, *and the maintained hypothesis of an unchanged coefficient matrix* $\Pi_{f_1,f_2} = \Pi$ *for all subsamples, the subsample heteroskedastic consistent Wald process and the sup Wald statistic have the following limits*

$$W_{f_2}(f_1) \Rightarrow \left[\frac{W_{nk}(f_2) - W_{nk}(f_1)}{f_w^{1/2}}\right]' \mathbf{A}_{f_1,f_2} \mathbf{B}_{f_1,f_2}^{-1} \mathbf{A}_{f_1,f_2}' \left[\frac{W_{nk}(f_2) - W_{nk}(f_1)}{f_w^{1/2}}\right],$$

Figure 1: Panel (a) shows the 5% asymptotic critical values of the Wald and sup Wald statistic with $d = 2$ and $f_0 = 0.05$. Panel (b) shows the 5% asymptotic critical value of the sup Wald statistic with $d = 2$ and $f_0 = \{0.01, 0.05, 0.10, 0.20\}$. These are estimated from 2,000 replications. The Wiener process is approximated by partial sums of 2000 standard normal variates.



$$SW_f(f_0) \xrightarrow{L} \sup_{f_w \in [f_0, f_2], f_2 = f} \left\{ \left[\frac{W_{nk}(f_2) - W_{nk}(f_1)}{f_w^{1/2}} \right]' \mathbf{A}_{f_1,f_2} \mathbf{B}_{f_1,f_2}^{-1} \mathbf{A}_{f_1,f_2}' \left[\frac{W_{nk}(f_2) - W_{nk}(f_1)}{f_w^{1/2}} \right] \right\},$$

where $\mathbf{A}_{f_1,f_2} = \mathbf{W}_{f_1,f_2}^{1/2} \mathbf{V}_{f_1,f_2}^{-1} \mathbf{R}'$, $\mathbf{B}_{f_1,f_2} = \mathbf{R} \left(\mathbf{\Omega}_{f_1,f_2} \otimes \mathbf{Q}_{f_1,f_2} \right) \mathbf{R}$, *and* W_{nk} *is vector Brownian motion with covariance matrix* $\mathbf{I}_{nk}$.

3.4 Simulated Asymptotic Distributions

The limit theory shows that the robust test statistics remain unchanged for all three scenarios – homoskedasticity, conditional heteroskedasticity, and unconditional heteroskedasticity. The asymptotic distributions are the same as those of the Wald process and sup Wald statistic under the assumption of homoskedasticity, given in equations (17) and (19).

Figure 1 plots the 5% standard asymptotic critical values (estimated from 2,000 replications) of the test statistics (17) and (19) against the (fractional) observation of interest f. Wiener processes are approximated by partial sums of $2,000$ standard normal variates. Panel (a) compares critical values of the Wald and sup Wald statistics with fixed values of d and f_0 ($d = 2$ and $f_0 = 0.05$). It is clear that the critical values for the sup Wald statistic are well above those of the Wald statistic, which is distributed as χ_2^2. In addition, one can see that the 5% critical value of the sup Wald statistic rises from 6.06 to 11.4 as the observation of interest f increases from

0.05 to 1. Moreover, the distribution stretches out to the right as the search range $[f_0, f]$ expands with f. Panel (b) plots the 5% asymptotic critical value of the sup Wald statistic for various minimum window sizes f_0 ($f_0 \in \{0.01, 0.05, 0.10, 0.20\}$) and for $d = 2$. It is evident smaller values of f_0 lead to larger critical values for the sup Wald statistic. This result is consistent with expectations because the search range $[f_0, f]$ widens as f_0 decreases. Although the results are not reported here, the critical values of both the Wald and sup Wald statistics increase with the value d.

4 Simulation Experiments

There is significant evidence to suggest that Wald tests, including Granger causality tests, suffer from size distortion to an extent that makes small sample considerations important in empirical work (Guilkey and Salemi, 1982; Toda and Phillips, 1993, 1994). By its very nature both the Wald test of the rolling window approach and the sup Wald test of the recursive rolling procedure involve the repeated use of small subsamples of data, thereby accentuating the importance of finite sample performance. This section therefore reports a series of simulation experiments designed to assess the finite sample characteristics of the causality tests.

The prototype model used in the simulation experiments is the bivariate VAR(1) model:

$$\text{DGP}: \begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \phi_{11} & \phi_{s_t} \\ 0 & \phi_{22} \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \tag{23}$$

where ε_{1t} and ε_{2t} are i.i.d. $N(0,1)$. Assumption **A0** requires $|\phi_{11}| < 1$ and $|\phi_{22}| < 1$. For simplicity, the causal channel from y_1 to y_2 is shut down. Parameter ϕ_{s_t} controls the strength of the causal path running from y_{2t} to y_{1t}. Under the null hypothesis of no causality, $\phi_{s_t} = 0$. Under the alternative hypothesis, causation runs from y_{2t-1} to y_{1t} for certain periods of the sample. Let s_t be a causal indicator that takes the value unity for the causal periods and zero otherwise so that the autoregressive coefficient ϕ_{s_t} is defined as $\phi_{s_t} = \phi_{12} s_t$.

The next two subsections investigate the performance of the forward expanding, rolling and recursive rolling causality tests under this DGP with different parameter settings under the null and alternative hypotheses. In these experiments the following general specifications apply:

(i) asymptotic critical values are obtained from simulating the distributions in Proposition 3.1 with 10,000 replications;

(ii) Wiener processes are approximated by partial sums of standard normal variates with $2,000$ steps;

(iii) the lag length p in the regression model is fixed at unity;

(iv) initial values of the data series (y_{11} and y_{21}) are set to unity;

(v) the rolling window test procedure uses a window length taken to be the minimum window size, f_0; and

(vi) the experiments are repeated 2,000 times for each parameter constellation.

4.1 False Detection Proportion

For all three approaches, we compare the test statistic with its corresponding critical value for each observation starting from $\lfloor Tf_0 \rfloor$ to T, so that the number of hypotheses tested, N, equals $T - \lfloor Tf_0 \rfloor + 1$. It is well known that the probability of making a Type I error rises with the number of hypotheses in a test, a phenomenon commonly referred to as *multiplicity*. Instead of examining the family wise error rate or size (probability of rejecting at least one true null hypothesis), therefore, we report the mean and standard deviation of the actual false detection proportion, which is defined as the ratio between the number of false rejections, F, and the total number of hypotheses N, given by F/N. Notice that this ratio differs from the false discovery rate promoted by Benjamini and Hockberg (1995). They define the false discovery rate as the expected value of the proportion of false discoveries among all discoveries, or $\mathbb{E}\left[F/\max\left(R,1\right)\right]$, where R is the total number of rejections. By construction, therefore, under the null hypothesis the false discovery rate takes the value of unity.

Table 1 reports the impact of the persistence parameters $\{\phi_{11}, \phi_{22}\}$ (top panel), the minimum window size f_0 (middle panel), and the sample size T (bottom panel), respectively, on the switch detection rates of the three algorithms under the null hypothesis. The top panel of Table 1 shows the effects of different parameter settings of $\{\phi_{11}, \phi_{22}\}$ with a fixed minimum window size and sample size ($f_0 = 0.24$ and $T = 100$). Summary statistics that are reported refer to the means and standard deviations (in parentheses) of the false detection proportion.

Overall, the rolling and recursive rolling window approaches have higher false detection proportion than the forward expansion approach. For example, in the top panel of Table 1, when $\{\phi_{11}, \phi_{22}\} = \{0.5, 0.8\}$, the false detection proportion is 11% using the rolling and recursive

Table 1: The mean and standard deviation (in parentheses) of the false detection proportion of the testing procedures under the null hypothesis based on the 5% asymptotic critical values. Parameter settings: $y_{11} = y_{21} = 1$ and $\phi_{12} = 0$. Calculations are based on 2,000 replications.

	Forward	Rolling	Recursive Rolling
(ϕ_{11}, ϕ_{22}): $f_0 = 0.24$ and $T = 100$			
(0.5,0.5)	0.07 (0.17)	0.10 (0.10)	0.09 (0.14)
(0.5,0.8)	0.08 (0.18)	0.11 (0.11)	0.11 (0.15)
(-0.5,0.8)	0.06 (0.16)	0.08 (0.09)	0.07 (0.13)
(0.5,-0.8)	0.05 (0.15)	0.07 (0.09)	0.06 (0.12)
f_0: $(\phi_{11}, \phi_{22}) = (0.5, 0.8)$ and $T = 100$			
0.12	0.09 (0.17)	0.19 (0.08)	0.22 (0.15)
0.24	0.07 (0.17)	0.10 (0.10)	0.09 (0.14)
0.36	0.08 (0.19)	0.09 (0.13)	0.08 (0.15)
0.48	0.07 (0.20)	0.08 (0.15)	0.06 (0.15)
T: $(\phi_{11}, \phi_{22}) = (0.5, 0.8)$ and $f_0 = 0.24$			
100	0.07 (0.17)	0.10 (0.10)	0.09 (0.14)
200	0.07 (0.17)	0.08 (0.10)	0.07 (0.13)
400	0.00 (0.04)	0.02 (0.13)	0.01 (0.09)

rolling window approaches – in contrast to 8% for the forward expanding approach. The false detection rate of the rolling window approach is slightly higher than that of the recursive rolling algorithm, except when the minimum window size is small, i.e. $f_0 = 0.12$. These results also reveal that there is a slightly greater chance of drawing false positive conclusions when y_{2t} is more persistent (witness the case where ϕ_{22} rises from 0.5 to 0.8 with ϕ_{11} fixed at 0.5). The false detection proportion appears to decline when the persistence parameters ϕ_{11} and ϕ_{22} are of different signs, showing that differing autoregressive behaviour in the two series can improve performance when the null is true.

The results reported in Table 1 show that the problem of false identification is alleviated when the number of observations included in the minimum window is increased. This can be achieved in one of two ways.

(i) When T expands from 100 to 400 with a fixed fractional window size, f_0, the false detection proportion decreases by 7%, 8% and 8% respectively for the forward, rolling and recursive rolling approaches (bottom panel).

(ii) When f_0 rises from 0.12 to 0.48, the false detection proportion reduces from 9% to 7%, from 19% to 8% and from 22% to 6% for the forward, rolling, and recursive rolling algorithms respectively (middle panel). The reduction is particularly obvious for the rolling and recursive rolling window approaches (11% and 16% reductions in the false detection proportion) where the minimum window size plays a more decisive role.

4.2 Causality Detection

The performance of the three algorithms under the alternative hypothesis is now investigated. We first consider the case when there is a single causality episode in the sample period, switching on at $\lfloor f_e T \rfloor$ and off at $\lfloor f_f T \rfloor$. Specifically, let s_t in (23) be defined as

$$s_t = \begin{cases} 1, & \text{if } \lfloor f_e T \rfloor \leq t \leq \lfloor f_f T \rfloor \\ 0, & \text{otherwise} \end{cases} .$$

Performance is evaluated from several perspectives: the successful detection rate (SDR), the mean and standard deviation (in parentheses) of the bias of the estimated fractional origination and termination dates of the switches ($\hat{f}_e - f_e$ and $\hat{f}_f - f_f$),[3], as well as the average number of switches detected. Successful detection is defined as an outcome when the estimated switch origination date falls between the true origination and termination dates, that is $f_e \leq \hat{f}_e \leq f_f$. The mean and standard deviation of the bias are calculated among those episodes that have been successfully detected.

Table 2 considers the impact of the general model parameters on test performance. Causal strength is fixed with the value $\phi_{12} = 0.8$ and causality (from $y_{2t} \to y_{1t}$) switches on in the middle of the sample ($f_e = 0.5$) and the relationship lasts for 20% of the sample with termination at $f_f = 0.7$. We vary the autoregressive parameters (ϕ_{11}, ϕ_{22}) (top panel), the minimum window size $\lfloor f_0 T \rfloor$ (middle panel), and the sample size T (bottom panel) in the simulations. Table 3 focuses on the impact of causal characteristics, namely, causal strength ϕ_{12} (top panel), causal duration, $\mathcal{D}$ (middle panel), and the location of the causal episode f_e (bottom panel).

It is apparent from the results reported in Tables 2 and 3 that the rolling window procedure has the highest successful detection rate, followed by the recursive rolling procedure. The detection rate of the forward expansion algorithm is the lowest among the three algorithms.

[3]Let *stat* denote the test statistic and *cv* be the corresponding critical values. A switch originates at period t if $stat_{t-2} < cv_{t-2}$, $stat_{t-1} < cv_{t-1}$, $stat_t > cv_t$ and $stat_{t+1} > cv_{t+1}$ and terminates at period t' if $stat_{t'-1} > cv_{t'-1}$, $stat_{t'} < cv_{t'}$, $stat_{t'+1} < cv_{t'+1}$.

Table 2: The impact of general model characteristics on test performance based on 5% asymptotic critical values. Parameter settings: $y_{11} = y_{21} = 1, \phi_{12} = 0.8, f_e = 0.5$, and $\mathcal{D} = 0.2$. Figures in parentheses are standard deviations. Calculations are based on 1,000 replications.

	Forward				Rolling				Recursive Rolling			
	SDR	$\hat{f}_e - f_e$	$\hat{f}_f - f_f$	# Switches	SDR	$\hat{f}_e - f_e$	$\hat{f}_f - f_f$	# Switches	SDR	$\hat{f}_e - f_e$	$\hat{f}_f - f_f$	# Switches
(ϕ_{11}, ϕ_{22}): $f_0 = 0.24$ and $T = 100$												
(0.5,0.5)	0.500	0.12 (0.06)	0.22 (0.11)	0.80 (0.72)	0.864	0.10 (0.05)	0.12 (0.07)	1.48 (0.80)	0.788	0.11 (0.05)	0.21 (0.11)	1.26 (0.68)
(0.5,0.8)	0.570	0.11 (0.06)	0.23 (0.11)	0.90 (0.74)	0.882	0.10 (0.05)	0.12 (0.07)	1.62 (0.85)	0.810	0.10 (0.05)	0.23 (0.10)	1.37 (0.72)
(-0.5,0.8)	0.632	0.10 (0.06)	0.25 (0.09)	0.91 (0.66)	0.884	0.10 (0.05)	0.13 (0.07)	1.45 (0.77)	0.829	0.10 (0.05)	0.24 (0.09)	1.26 (0.62)
(0.5,-0.8)	0.625	0.10 (0.06)	0.26 (0.08)	0.83 (0.59)	0.885	0.10 (0.05)	0.13 (0.06)	1.24 (0.63)	0.840	0.10 (0.05)	0.24 (0.09)	1.17 (0.55)
f_0: $\phi_{11} = 0.5, \phi_{22} = 0.8$ and $T = 100$												
0.12	0.570	0.11 (0.06)	0.23 (0.11)	1.03 (0.82)	0.909	0.08 (0.05)	0.03 (0.06)	3.18 (1.28)	0.846	0.08 (0.04)	0.20 (0.11)	2.21 (1.09)
0.24	0.570	0.11 (0.06)	0.23 (0.11)	0.90 (0.74)	0.882	0.10 (0.05)	0.12 (0.07)	1.62 (0.85)	0.810	0.10 (0.05)	0.23 (0.10)	1.37 (0.72)
0.36	0.570	0.11 (0.06)	0.23 (0.11)	0.81 (0.68)	0.774	0.11 (0.05)	0.20 (0.10)	1.19 (0.63)	0.708	0.11 (0.05)	0.23 (0.11)	1.13 (0.65)
0.48	0.570	0.11 (0.06)	0.23 (0.11)	0.75 (0.67)	0.688	0.11 (0.06)	0.22 (0.11)	1.08 (0.64)	0.595	0.11 (0.05)	0.23 (0.11)	0.99 (0.62)
T: $\phi_{11} = 0.5, \phi_{22} = 0.8$ and $f_0 = 0.24$												
$T = 100$	0.570	0.11 (0.06)	0.23 (0.11)	0.90 (0.74)	0.882	0.10 (0.05)	0.12 (0.07)	1.62 (0.85)	0.810	0.10 (0.05)	0.23 (0.10)	1.37 (0.72)
$T = 200$	0.786	0.10 (0.05)	0.27 (0.07)	1.14 (0.71)	0.917	0.08 (0.05)	0.15 (0.06)	1.61 (0.87)	0.911	0.09 (0.05)	0.27 (0.07)	1.36 (0.73)
$T = 400$	0.922	0.08 (0.04)	0.29 (0.04)	1.19 (0.68)	0.910	0.05 (0.03)	0.18 (0.04)	1.60 (1.00)	0.927	0.06 (0.03)	0.30 (0.02)	1.25 (0.71)

Table 3: The impact of causal characteristics on test performance based on 5% asymptotic critical values. Parameter settings: $y_{11} = y_{21} = 1, \phi_{11} = 0.5, \phi_{22} = 0.8, T = 100$. Figures in parentheses are standard deviations. The minimum window has 24 observations. Calculations are based on 1,000 replications.

	Forward				Rolling				Recursive Rolling			
	SDR	$\hat{f}_e - f_e$	$\hat{f}_f - f_f$	# Switches	SDR	$\hat{f}_e - f_e$	$\hat{f}_f - f_f$	# Switches	SDR	$\hat{f}_e - f_e$	$\hat{f}_f - f_f$	# Switches
Causality strength ϕ_{12}: $\mathcal{D} = 0.2, f_e = 0.5$												
$\phi_{12} = 0.2$	0.139	0.11 (0.06)	0.13 (0.15)	0.41 (0.70)	0.456	0.12 (0.06)	0.02 (0.09)	1.38 (1.00)	0.315	0.11 (0.06)	0.08 (0.13)	1.01 (0.92)
$\phi_{12} = 0.8$	0.570	0.11 (0.06)	0.23 (0.11)	0.90 (0.74)	0.882	0.10 (0.05)	0.12 (0.07)	1.62 (0.85)	0.810	0.10 (0.05)	0.23 (0.10)	1.37 (0.72)
$\phi_{12} = 1.5$	0.781	0.10 (0.05)	0.27 (0.08)	1.04 (0.60)	0.921	0.08 (0.05)	0.15 (0.06)	1.55 (0.76)	0.890	0.09 (0.05)	0.26 (0.08)	1.33 (0.61)
Causality Duration $\mathcal{D}$: $\phi_{12} = 0.8, f_e = 0.5$												
$\mathcal{D} = 0.1$	0.234	0.05 (0.03)	0.24 (0.16)	0.60 (0.81)	0.500	0.06 (0.03)	0.11 (0.08)	1.58 (0.98)	0.380	0.05 (0.03)	0.19 (0.15)	1.30 (0.92)
$\mathcal{D} = 0.2$	0.570	0.11 (0.06)	0.23 (0.11)	0.90 (0.74)	0.882	0.10 (0.05)	0.12 (0.07)	1.62 (0.85)	0.810	0.10 (0.05)	0.23 (0.10)	1.37 (0.72)
$\mathcal{D} = 0.3$	0.777	0.15 (0.08)	0.18 (0.05)	1.02 (0.58)	0.918	0.11 (0.06)	0.11 (0.06)	1.47 (0.76)	0.900	0.11 (0.06)	0.18 (0.05)	1.28 (0.61)
Causality Location f_e: $\phi_{12} = 0.8, \mathcal{D} = 0.2$												
$f_e = 0.3$	0.649	0.11 (0.06)	0.34 (0.19)	1.00 (0.79)	0.887	0.10 (0.05)	0.12 (0.07)	1.66 (0.85)	0.830	0.10 (0.05)	0.31 (0.18)	1.35 (0.78)
$f_e = 0.5$	0.570	0.11 (0.06)	0.23 (0.11)	0.90 (0.74)	0.882	0.10 (0.05)	0.12 (0.07)	1.62 (0.85)	0.810	0.10 (0.05)	0.23 (0.10)	1.37 (0.72)
$f_e = 0.7$	0.500	0.11 (0.06)	0.09 (0.02)	0.77 (0.71)	0.885	0.10 (0.05)	0.08 (0.04)	1.65 (0.84)	0.789	0.11 (0.05)	0.09 (0.03)	1.40 (0.77)

For example, from the top panel of Table 2, when $(\phi_{11}, \phi_{22}) = (0.5, 0.8)$, the SDR of the rolling procedure is, respectively, 7.2% and 31.2% higher than those of the recursive rolling and forward expanding procedures. Notice that, relative to the forward expanding procedure, the difference in SDR between the rolling and recursive procedures is much less dramatic.

There is no obvious difference in the estimation accuracy of the causal switch-on date. For example, when $(\phi_{11}, \phi_{22}) = (0.5, 0.8)$ in the top panel of Table 2, the average delay in the detection of the switch-on date is 10 to 11 observations (with a standard deviation of 5 to 6 observations) for all three procedures. Importantly, the rolling window procedure provides a much more accurate estimator for the switch-off date in the sense that the quantity $\hat{f}_f - f_f$ is of smaller magnitude and has less variance. With the same parameter settings, the average delay in the switch-off point detection is 12 observations (standard deviation of 7 observations) for the rolling procedure, as opposed to 23 observations delay (standard deviations of 10 and 11 observations) for the recursive rolling and forward expanding algorithms, respectively.

As mentioned earlier, the rolling and recursive window procedures have more significant size distortions than the forward expanding window approach. This observation is reflected in the estimated average number of switches reported in Tables 2 and 3. The true number of switches in the simulation is unity. It can be seen from Tables 2 and 3 that the rolling and recursive rolling window procedures tend to detect more causal episodes than there are. In addition, the upward bias in the estimator for the rolling window procedure is higher than that of the recursive rolling procedure. The forward expanding algorithm underestimates the number of switches when the sample size is 100 and overestimates the statistic (at a lesser magnitude than the rolling and recursive rolling procedures) when the sample size increases to 200 and 400 (bottom panel of Table 2).

Taking a closer look at Table 2, in the top panel, we see that for all three approaches the SDR increases when the persistence level of y_{2t} (ϕ_{22}) increases from 0.5 to 0.8, with $f_0 = 0.24, T = 100$ and ϕ_{11} fixed at 0.5. Successful detections are generally higher when the persistent parameters ϕ_{11} and ϕ_{22} are of different signs. No obvious difference is observed in the estimation accuracy of the switch-on and -off dates. For the middle panel, we set $(\phi_{11}, \phi_{22}) = (0.5, 0.8)$, $T = 100$, and let the minimum window size vary from 24 to 48 observations. The minimum window size does not have any impact on the correct detection rate of the forward expanding procedure. However, we observe significant reductions in SDR for the rolling and recursive rolling procedures when the minimum window size increases. As a case in point, there is a 10.8% and 10.2% drop in SDR,

respectively, for the former and the latter when f_0 rises from 0.24 to 0.36. However, these falls are not as extensive as the declines in the false detection rates (23.8% and 21.5% respectively).

In the bottom panel of Table 2, we increase the sample size from 100 to 400, keeping (ϕ_{11}, ϕ_{22}) and f_0 fixed at $(0.5, 0.8)$ and 0.24. It is clear from the results in this panel that for all tests, the successful detection rate and the estimation accuracy of the switch-on date increases with the sample size, whereas the estimation accuracy of the switch-off date deteriorates. Notice that the SDR of the forward expanding and recursive rolling procedures rises rapidly and exceeds that of the rolling approaches when the sample size reaches 400. Nevertheless, the SDR of all three approaches are above 90% when the sample size is larger than 400 and the difference in SDR are not dramatic.

Table 3 focuses on the characteristics of the causal relationship. For all tests, SDR increases with the strength of the causal relationship (captured by the value of ϕ_{12}). One can see that when ϕ_{12} rises from 0.2 to 1.5, the SDR increases from 13.9% to 78.1%, from 45.6% to 92.1%, and from 31.5% to 89% for the forward expanding, rolling and recursive rolling algorithms, respectively. Notice that the gap between the SDRs of the rolling and recursive rolling procedures narrows as causality strengthens. Moreover, as the causal relationship gets stronger, there is some mild improvement in the estimation accuracy of the switch-on date using those three approaches, whereas for all three tests the accuracy of the estimates of the switch-off date deteriorates. For example, for the rolling test, when ϕ_{12} rises from 0.2 to 1.5, the bias of the switch-on date reduces from 12 observations to 8 observations, while the bias of the switch-off date increases from 2 observations to 15 observations. The dramatic increase in the estimation bias of the switch-off date is mainly due to situations in which a switch is detected but the termination date of this switch is not found until the end of the sample. If this situation eventuates, a termination date of $\hat{r}_f = 1$ is imposed at the cost of significant bias in the estimates. The proportion of samples for which this occurs increases as the causal relationship gets stronger.

In the middle panel of Table 3, the causal relationship is switched on at the 50th observation and the causal episode is defined to last for 10%, 20%, and 30% of the sample, respectively. The SDR of all tests rises dramatically as the duration, D, of the causal relationship increases. The SDR increases from 50% to 91.8% (from 38% to 90%) for the rolling (recursive rolling) algorithm as the duration expands from 10 to 30 observations. Interestingly, it is also clear that the biases of the estimated origination dates also increase with longer causal duration. As for the termination dates, while the estimation accuracy improves slightly for the forward

expanding approach, no obvious change patterns are observed for the rolling and recursive rolling approaches.

The bottom panel of Table 3 reports results for the location parameter f_e, which takes the values $f_e = \{0.3, 0, 5, 0.7\}$. In the first scenario, causality is switched on at the 30th observation and lasts for 20 observations, while for the second and third scenarios causality is assumed to originate from the 50th and 70th observations, respectively, and last for the same length of time. The performance is better (with higher SDR and smaller bias in the switch-on date estimate) for the forward expanding approach and slightly better for the recursive rolling approach when the change in causality happens early in the sample. By contrast, the location of the switch does not have an obvious impact on the performance of the rolling algorithm. Notice that the bias of the termination date estimates declines significantly as the causal episode moves towards the end of the sample period. This bias is mainly due to the truncation that is imposed in the estimation. Specifically, when the causal effect terminates at the 90^{th} data point, due to the delay bias in estimation, the procedure may not detect the switch-off date until the end of the sample. In these cases, the estimated termination date is set to be the last observation of the sample, a strategy which results in a bias of 0.10 for the estimated of f_f and which reduces both the bias and the variance of the estimate.

4.2.1 Multiple Episodes

Suppose there are two switches in the sample period, where the first period of causality runs from f_{1e} to f_{1f} and the second from f_{2e} to f_{2f}. This situation is denoted as follows:

$$s_t = \begin{cases} 1, & \text{if } \lfloor f_{1e}T \rfloor \leq t \leq \lfloor f_{1f}T \rfloor \text{ and } \lfloor f_{2e}T \rfloor \leq t \leq \lfloor f_{2f}T \rfloor \\ 0, & \text{otherwise} \end{cases} .$$

The strength of the first and second episodes are denoted by ϕ_{12}^1 and ϕ_{12}^2, respectively, and the durations of the causal episodes are $\mathcal{D}_1 = f_{1f} - f_{1e}$ and $\mathcal{D}_2 = f_{2f} - f_{2e}$. The sample size is set to be 200 and the minimum window size $f_0 = 0.24$.

In the top panel of Table 4, the location of the switches is set at the 25th and 75th observations respectively ($f_{1e} = 0.25$, $f_{2e} = 0.75$) and the causality strength of both episodes is set to 0.8 ($\phi_{12}^1 = \phi_{12}^2 = 0.8$). The durations of the causal episodes are varied, using $\{\mathcal{D}_1 = 0.1, \mathcal{D}_2 = 0.1\}$, $\{\mathcal{D}_1 = 0.1, \mathcal{D}_2 = 0.2\}$ and $\{\mathcal{D}_1 = 0.2, \mathcal{D}_2 = 0.1\}$. In the bottom panel, with the causality duration fixed at $\mathcal{D}_1 = 0.1$ and $\mathcal{D}_2 = 0.1$, the causal strength of the second episode ϕ_{12}^2 is increased

Table 4: Test performance in the presence of two causal episodes based on 5% asymptotic critical values. Parameter settings: $y_{11} = y_{21} = 1, \phi_{11} = 0.5, \phi_{22} = 0.8, f_0 = 0.24, T = 200$. Figures in parentheses are standard deviations. Calculations are based on 1,000 replications.

	First Episode			Second Episode			
	SDR	$\hat{f}_{1e} - f_{1e}$	$\hat{f}_{1f} - f_{1f}$	SDR	$\hat{f}_{2e} - f_{2e}$	$\hat{f}_{2f} - f_{2f}$	# Switches
$f_{1e} = 0.25$, $f_{2e} = 0.75$, $\phi_{12}^1 = \phi_{12}^2 = 0.8$							
$\mathcal{D}_1 = 0.1$, $\mathcal{D}_2 = 0.1$							
Forward	0.515	0.05 (0.03)	0.36 (0.27)	0.378	0.05 (0.03)	0.14 (0.03)	1.44 (1.12)
Rolling	0.633	0.05 (0.03)	0.14 (0.07)	0.601	0.06 (0.03)	0.11 (0.06)	2.50 (1.06)
Recursive Rolling	0.601	0.06 (0.03)	0.32 (0.25)	0.423	0.05 (0.03)	0.13 (0.04)	1.98 (1.14)
$\mathcal{D}_1 = 0.1$, $\mathcal{D}_2 = 0.2$							
Forward	0.515	0.05 (0.03)	0.36 (0.27)	0.579	0.08 (0.05)	0.05 (0.00)	1.56 (1.01)
Rolling	0.633	0.05 (0.03)	0.14 (0.07)	0.938	0.08 (0.04)	0.05 (0.02)	2.44 (0.96)
Recursive Rolling	0.601	0.06 (0.03)	0.32 (0.25)	0.633	0.08 (0.05)	0.05 (0.01)	1.99 (1.04)
$\mathcal{D}_1 = 0.2$, $\mathcal{D}_2 = 0.1$							
Forward	0.862	0.09 (0.05)	0.47 (0.17)	0.151	0.04 (0.02)	0.14 (0.03)	1.23 (0.74)
Rolling	0.910	0.08 (0.04)	0.16 (0.06)	0.601	0.06 (0.03)	0.11 (0.06)	2.30 (0.85)
Recursive Rolling	0.924	0.08 (0.04)	0.47 (0.16)	0.125	0.04 (0.02)	0.13 (0.04)	1.26 (0.74)
$\mathcal{D}_1 = 0.1$, $\mathcal{D}_2 = 0.1$							
$f_{1e} = 0.25$, $f_{2e} = 0.75$, $\phi_{12}^1 = 0.8$, and $\phi_{12}^2 = 1.5$							
Forward	0.515	0.05 (0.03)	0.36 (0.27)	0.512	0.04 (0.03)	0.14 (0.02)	1.55 (1.05)
Rolling	0.633	0.05 (0.03)	0.14 (0.07)	0.773	0.05 (0.02)	0.13 (0.04)	2.49 (1.04)
Recursive Rolling	0.601	0.06 (0.03)	0.32 (0.25)	0.545	0.04 (0.03)	0.14 (0.03)	2.02 (1.11)
$f_{1e} = 0.25$, $f_{2e} = 0.85$, $\phi_{12}^1 = 0.8$, and $\phi_{12}^2 = 1.5$							
Forward	0.515	0.05 (0.03)	0.35 (0.26)	0.515	0.04 (0.03)	0.05 (0.00)	1.62 (1.16)
Rolling	0.633	0.05 (0.03)	0.14 (0.07)	0.786	0.05 (0.03)	0.05 (0.01)	2.47 (1.08)
Recursive Rolling	0.601	0.06 (0.03)	0.30 (0.24)	0.603	0.05 (0.03)	0.05 (0.01)	2.10 (1.13)

from 0.8 to 1.5 (first section) and the second episode is moved further towards the end of the sample period so that these two episodes are further apart, that is $f_{2e} = 0.85$ (second section).

Three general observations may be made on the results reported in Table 4. *First*, in all of the reported scenarios, the correct detection rates of the rolling procedure for both episodes are generally the highest of the three procedures, followed by the recursive rolling window procedure – except for the case of $\{\mathcal{D}_1 = 0.2, \mathcal{D}_2 = 0.1\}$ where the detection rate of the first episode of the recursive rolling window method is the highest. *Second*, location plays a decisive role in the success of the detection procedures for multiple causality episodes. This result is particular true for the forward and recursive rolling window procedures and is partially due to the low

estimation accuracy in the causal termination date. As mentioned, when using the forward and recursive rolling window approaches, the termination dates are not found until the end of the sample period for a significant proportion of replications. It is, therefore, impossible to detect the second episode of causality for those sample replications. *Third*, the correct detection rates of all procedures increase with causal strength and the distance between two episodes.

There are also a number of more specific results. For causal episodes of the same causal strength and duration, the detection rate is higher for the episode occurring first. As a case in point, when $\mathcal{D}_1 = 0.1, \mathcal{D}_2 = 0.1$, the detection rates of the first and second episodes are 63.3% and 60.1%, respectively, using the rolling window approach. The detection rates of the second episode using the forward and recursive rolling window algorithms are 37.8% and 42.3%, which are 13.7% and 17.8% lower than those for the first episode.

It is also easier for all procedures to detect episodes with longer duration. For example, when $\mathcal{D}_1 = 0.1$ and $\mathcal{D}_2 = 0.2$, the detection rates of the second episode, respectively, 6.4%, 30.5%, and 3.2%, are higher than the detection rates for the first episode for all three algorithms. Combining location and duration, it is expected that the detection rates will be lower when the duration of the second bubble is shorter than the first one. This expectation is realized when moving from the case of $\mathcal{D}_1 = 0.1$ and $\mathcal{D}_2 = 0.2$ to $\mathcal{D}_1 = 0.2$ and $\mathcal{D}_2 = 0.1$. Specifically, when $\mathcal{D}_1 = 0.2$ and $\mathcal{D}_2 = 0.1$, the detection rates of the second episode decline from 57.9% to 15.1%, from 93.8% to 60.1%, from 63.3% to 12.5%, respectively, for the forward, rolling and recursive rolling procedures. Notice that the detection rates of the second episode of the forward expanding and recursive rolling procedures for this case are around 15%. This result is partially due to the inaccuracy of these two procedures in estimating the termination date of the first episode. The average delay $\hat{f}_{1f} - f_{1f}$ of these two procedures is 0.47.

It is also obvious from the bottom panel that SDR increases with causal strength and the distance between two episodes. The successful detection rate of the second episode of the forward, rolling and recursive rolling methods rises 13.4%, 17.2% and 12.2% respectively when ϕ_{12}^2 rises from 0.8 to 1.5. When moving the second episode from the 75th observation to the 85th observation, there is a slight increase in the detection rates (0.3%, 1.3%, and 5.8%, respectively, for forward, rolling and recursive rolling approaches).

Finally, the estimated average numbers of switches for all three algorithms are reported in the last column of Table 4. The rolling window procedure tends to overestimate the number of causal episodes whereas the forward approach tends to underestimate this number.

4.3 Asymptotic versus Finite Sample Critical Values

In practical work, the residual based bootstrap method is often used to generate small sample critical values with the intent to improve finite sample performance characteristics. See, for example, Balcilar et al. (2010) and Arora and Shi (2015). The calculations for the family-wise false positive detection rate (size) and successful detection rates are repeated using tests based on 5% bootstrap critical values using 1,000 replications. Similar conclusions to those reported above are drawn with regard to the relative performance of the forward, rolling and recursive rolling algorithms. There are some reductions in both sizes and successful detection rates for all algorithms, as the finite sample critical values are generally higher than the corresponding asymptotic critical values (especially when the sample size is small).

Table 5: The differences in false detection proportion and SDRs of the testing algorithms using 5% asymptotic and bootstrap critical values. Calculations are based on 1,000 replications.

	Difference in false detection proportion			Difference in SDRs		
	Forward	Rolling	Recursive Rolling	Forward	Rolling	Recursive Rolling
100	-0.03	-0.17	-0.15	-0.05	-0.02	-0.00
200	-0.02	-0.10	-0.12	-0.03	-0.00	-0.02
400	-0.03	-0.01	-0.05	-0.00	-0.01	-0.01

For the sake of brevity, all the results are not reported here but as an example, Table 5 reports differences in the sizes and the successful detection rates between using the bootstrap and asymptotic critical values for a typical set of parameters. Specifically, $y_{11} = y_{21} = 1$ and $\phi_{11} = 0.5, , \phi_{22} = 0.8$, $\lfloor f_0 T \rfloor = 0.24$, $f_e = 0.5$ and $f_f = 0.7$, and $\phi_{12} = 0$ under the null and 0.8 under the alternative. The sample size varies from 100 to 400.[4]

Overall, the bootstrap critical values do not lead to dramatic reductions in the successful detection rates of the three algorithms and in the false positive detection rate of the forward expanding approach. However, there are substantial reductions in the false positive rates of the rolling and recursive rolling procedures when the sample size is small. For example, when $T = 100$, the false positive detection rates of the rolling and recursive rolling window procedure drop 17% and 15% respectively when replacing the asymptotic critical values with the bootstrap

[4]The residual bootstrap method is computationally intensive. For example, it takes around 380 hours to finish a simulation with 1,000 replications for the case of $T = 400$ by doing parallel computing on a 16-core high performance machine. Due to limitations in available computing power, simulations using the residual based bootstrap method for the case of $T = 1000$ are not conducted.

critical values. This reduction becomes significantly smaller and almost negligible when the sample size increases to 400. Putting these together, the results suggest a strategy of using the residual based bootstrap method for the rolling and recursive rolling window algorithms when the sample size is smaller and using the asymptotic critical values for all other cases.

5 The Predictive Power of the Slope of the Yield Curve

The slope of the yield curve (usually defined as the difference between zero-coupon interest rates on three-month Treasury bills and 10-year Treasury bonds) has traditionally been regarded as a potentially important explanatory variable in the prediction of real economic activity and inflation (Harvey, 1988). The term structure of interest rates embodies market expectations of the behaviour of the future short-term interest rate (the expectations theory) and contains a term premium component that compensates for the risk of holding longer-term securities (the liquidity premium theory). The link between the slope of the yield curve and macroeconomic activity which is founded on the expectations theory is now widely accepted, whereas the contribution of the term premium to the prediction of output growth and inflation is less well established.[5]

Empirical evidence of the ability of the slope of the yield curve to forecast macroeconomic activity, including real economic growth or recessions, was provided in the 1980s and 1990s for several countries (Stock and Watson, 1989; Estrella and Hardouvelis, 1991; Estrella and Mishkin, 1998; Dotsey, 1998; Estrella and Mishkin, 1997; Plosser and Rouwenhorst, 1994). The slope of the yield curve was also found to be a significant predictor of inflation (Mishkin, 1990a, 1990b, 1990c; Jorion and Mishkin, 1991). More recent work in the context of predicting real activity and recessions suggests that the slope of the yield curve still retains its predictive power (Estrella, 2005; Chauvet and Potter, 2002, 2002; Ang, Piazzesi, and Wei, 2006; Wright, 2006; Estrella and Trubin, 2006; Rudebusch and Williams, 2009; Kauppi and Saikonen, 2008).

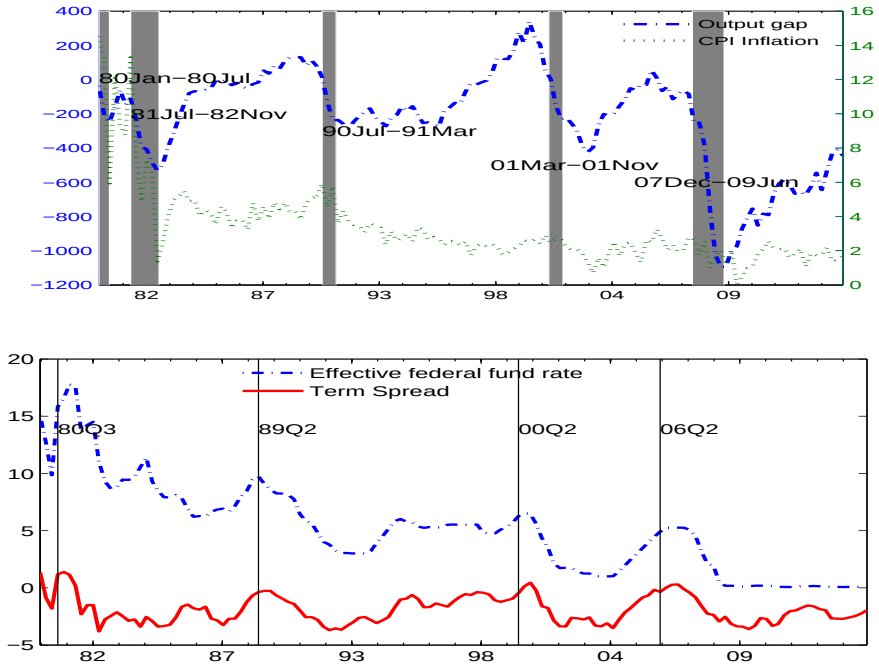
While most of the earlier literature focused on the ability of the yield curve to predict real activity, it is reasonable to conjecture that feedback effects from real activity to monetary policy and therefore to the yield curve exist (Estrella and Hardouvelis, 1991; Estrella and Mishkin, 1997; Estrella, 2005). Consequently a substantial body of empirical work in this area has been conducted in terms of VAR models (Ang and Piazzesi, 2003; Evans and Marshall, 2007; Diebold,

[5] Although Hamilton and Kim (2002) find that both components make significant contributions to forecasting real economic activity, Estrella and Wu (2008) find the opposite result, namely, that decomposing the spread into expectations and term premium components does not enhance the predictive power of the yield curve.

Rudebusch and Aruba, 2006), which provides ample precedence to support the use of VAR models to establish the direction of Granger causality in these macroeconomic relationships.

In the present application a four-variable VAR model is used to test for changes in Granger causal relationships between the slope of the yield curve and the macroeconomy. The variables included are the output gap (y_t), inflation (π_t), the monetary policy interest rate (i_t), and the yield curve spread (S_t). The data are quarterly data for the United States for the period 1980:Q1 to 2015:Q1 with $T = 141$ observations. The data are plotted in Figure 2.

Figure 2: Time series plots of the data used in the VAR model to test for changes in Granger causal relationships between the slope of the yield curve and the macroeconomy for the period 1980:Q1 to 2015:Q1 (141 observations). The output gap and inflation (right axis) are plotted in the top panel, with official NBER recession periods shaded in grey. The Federal funds rate and the slope of the yield curve are plotted in the bottom panel with vertical lines marking the generally accepted dates of the onset of an inverted yield curve.



The output gap is calculated using the official Congressional Budget Office (CBO) 2016 measure of real potential output (billions of chained 2009 dollars, not seasonally adjusted) and 2016:Q1 GDP (billions of chained 2009 dollars, seasonally adjusted annual rate) data. Inflation

is measured from the core consumer price index and calculated as quarterly log differences (multiplied by 400). The policy rate is measured using the effective Federal funds rate. Term spread is defined as the difference between the three-month treasury bill rate and the 10-year government bond rate. All macroeconomic data are either obtained quarterly or monthly from the Federal Reserve Bank of St. Louis FRED[6] where appropriate monthly observations are converted to quarterly frequency by averaging.

The variability of the inflation gap is more muted than that of the output gap. The inflation gap fluctuates around the 2% level and shows persistent decline towards the end of the sample period, consistent with the deflationary conditions prevalent in the United States economy after the Global Financial Crisis. Official NBER recession periods that coincide with the sample period, namely 1980:M01-M07,1981:M07-1982:M11,1990:M07-1991:M03, 2001:M03-M11 and 2007:M12-2009:M06 are marked in grey. Since the yield curve is typically upward sloping, the slope factor, defined as the difference between the zero-coupon interest rates on three-month Treasury bills and 10-year Treasury bonds, usually takes a negative value. Steeper yield curves are represented by lower values of the slope factor. If the yield curve becomes inverted then the slope factor will be positive and the dates of the onset of an inverted yield curve are shown by vertical lines.[7] Notable instances are in 2000 (when a recession followed) and in 2006 (when the inverted yield curve was not immediately followed by a recession). A final feature of the data in the bottom panel of Figure 2 is the settling of the effective funds rate at zero for the latter part of the sample period after 2009Q1, the so-called zero lower bound period of monetary policy.

The decision to use a four-variable VAR model means that two other factors associated with the term structure of interest rates, the level and curvature, are not included. The fact that the level of the Federal funds target rate is included in the VAR means that the level factor may be safely omitted without much loss of information. The main reason for omitting the curvature (or bow) of the yield curve is that its relationship with the macroeconomy has been hard to establish. There have been attempts to devise theoretical links between curvature and the macroeconomy (Dewachter and Lyrio, 2006; Modena, 2008; Moench, 2008) but there is little evidence to support the nature of the relationship. In view of the ambivalent evidence and the shortage of degrees of freedom in the present application with quarterly data, it was decided not to include the curvature in the VAR.

[6]Website: wwww.research.stlouisfed.org/fred2/.

[7]Note that these dates are generally established using higher frequency data on the yield curve than the quarterly data plotted here.

In estimating the VAR and implementing tests of Granger causality, the lag order is obtained using the Bayesian information criterion (BIC) with a maximum potential lag length 12. When implementing the recursive testing procedure the minimum window size is 28 observations ($f_0 = 0.2$) and this constant window size is also used for the rolling testing procedure. The 5% critical value sequences are obtained from bootstrapping with 500 replications.[8]

5.1 Yield Curve Slope to the Output Gap

Figure 3 displays the time-varying Wald test statistics for causal effects from the slope of the yield curve to the output gap. The three rows illustrate the sequences of test statistics obtained from the forward recursive, rolling window and recursive-rolling procedures, respectively, while the columns of the figure refer to the two different assumptions of the residual error term (homoskedasticity and heteroskedasticity) for the VAR. Sequences of the test statistics start from 1986:Q4 as the first 28 observations represent the minimum window size, the shaded areas are the NBER recession periods, vertical lines are the dates of the onset of an inverted yield curve and the dates of causal episodes are also shown.

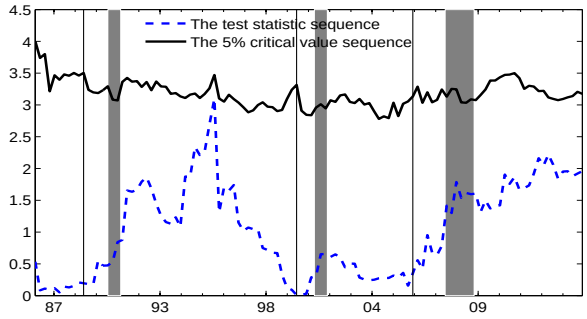
It is clearly apparent from panel (a) of Figure 3, that the test statistic of the predictive power of the slope of the yield curve for the output gap lies below its critical value at the end of the sample period in 2015:Q1. Consequently, the null hypothesis of no Granger causality from the yield curve slope to the output gap over the whole sample period cannot be rejected. This result highlights the danger of simply using Wald tests of Granger causality indiscriminately over the full sample period. Indeed, the fact that slope of the yield curve would have little predictive power towards the end of the sample is to be expected given that starting in 2009:Q1 the Federal funds rate has been constrained by the zero lower bound. The relative lack of information encoded in the slope of the yield curve during this period will have a significant influence on inference based on the entire sample.

The results for the full-sample forward recursive Wald test of causality indicate no causal relationship (or change in the causal relationship) between the slope of the yield curve and real economic activity at all over the entire sample period. This conclusion is contrary to expectations and to all existing evidence of the usefulness of the slope in predicting real economic activity. The failure of the forward recursive procedure to identify any periods of Granger causality confirms the results of the simulation exercises where this approach performed poorly by comparison with

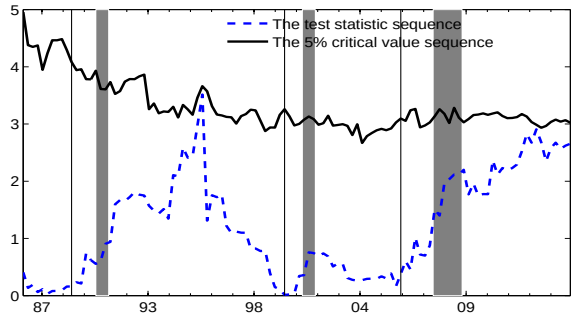
[8]The data and Matlab codes are available for download from `http://www.ncer.edu.au/data/data.jsp`.

Figure 3: Tests for Granger causality running from the yield curve slope to the output gap. Tests are obtained from a VAR model allowing for homoskedasitic errors (panels (a), (c) and (e)) and for heteroskedastic errors (panels (b), (d) and (f)). The sequence of tests for the forward recursive, rolling window and recursive-rolling procedures run from 1986:Q4 to 2015:Q1 with 28 observations for the minimum window size. The lag orders are selected using BIC with a maximum lag order of 12. The shaded areas are the NBER recession periods, the vertical lines are the dates of the onset of an inverted yield curve and causal periods are shown in text.

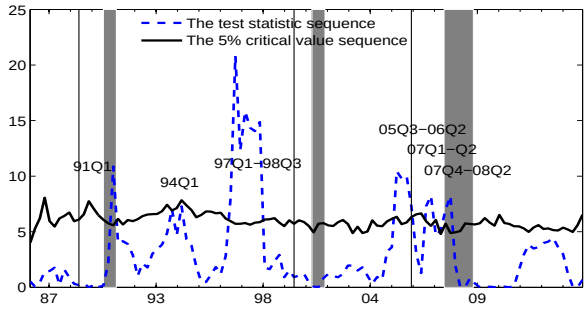
(a) Forward - Homoskedasticity



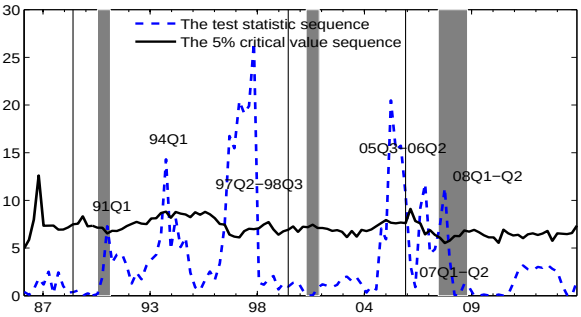
(b) Forward - Heteroskedasticity



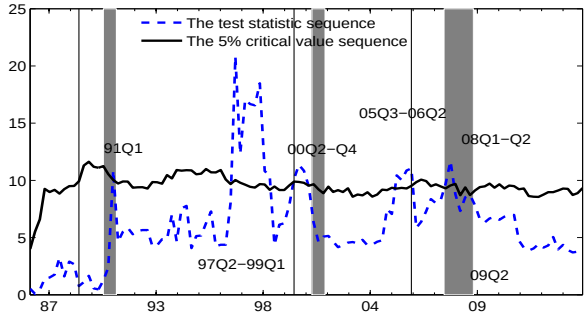
(c) Rolling - Homoskedasticity



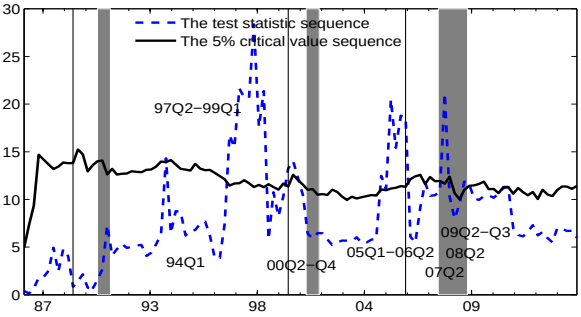
(d) Rolling - Heteroskedasticity



(e) Recursive Rolling - Homoskedasticity



(f) Recursive Rolling - Heteroskedasticity



the rolling window and recursive-rolling procedures.

Indeed, panels (c) and (e) of Figure 3 show a very different picture from that of an unequivocal failure to reject the null hypothesis of no predictability. Instead a dynamic picture of the evolution of Granger causal relationships between the slope of the yield curve and the output gap is revealed. The two major periods of predictability that are detected in these tests are over 1997 - 1999 and 2005 - 2008 although there is some disagreement between the rolling and the recursive rolling tests about the exact dates of occurrence of these two episodes. The rolling window algorithm suggests the predictability of the yield curve slope to output gap in the period of 1997:Q1 - 1998:Q3, while the recursive rolling approach detects the origination (termination) of the causality one (two) quarter(s) later than the rolling method. For the latter period, the recursive rolling window tests suggest two sub-periods (2005:Q3-2006:Q2 and 2008:Q1-Q2), while this division is less pronounced in the case of the recursive rolling procedure (last episode starts in 2007:Q4 with an additional short episode in 2007:Q1-Q2). It is also interesting to note that these patterns in causality are robust to heterogeneity of various forms in the VAR errors.

Additionally, both methods detect two short causality episodes: 1991Q1 and 1994Q1. The episode in 1994 can be observed from the heteroskedastic consistent results in (d) and (f). Lastly, the recursive rolling methods identifies an additional period of causality in 2000:Q2-Q4 which is not found by the rolling approach.

The results from panels (d) and (e) in Figure 3 are consistent with some existing evidence. In particular, the test sequence findings corroborate two general conclusions in the literature. First, Dotsey (1998) argues that, in contrast to previous periods, the information content of the slope of the term structure is not statistically significant for predicting output between the beginning of 1985 and the middle of 1997. Here, only two single-observation causal episodes (1991:Q1 and 1994:Q1) are observed. Second, Kucko and Chinn (2009) find that the overall predictive ability of the yield slope decreased after 1998 (although their measure of real activity is industrial production rather than the output gap). A quiet period between 1998 and 2005 is observed in this data and there is then a re-bound in the causal relationship in the second half of 2005. The causality from the yield curve spread to output gap lasts until 2008:Q2 -beginning of the 2008 recession.

This second period of predictability, ending in 2008, appears to have led to a spate of recent empirical findings that have claimed the slope of the yield curve still provides information about output (Estrella, 2005; Chauvet and Potter, 2002, 2005; Ang, Piazzesi, and Wei, 2006; Wright,

2006; Estrella and Trubin, 2006). Later studies, by Rudebusch and Williams (2009) and Kauppi and Saikonen (2008), reached a similar conclusion and use sample periods that end in 2006, just as the predictive power of the slope appears to be on the wane. Nonetheless, a finding of significant predictive ability over the entire sample period used in these studies is consistent with the subsample results discovered here.

Finally, it is worth noting that the test results from the rolling window approach shows the value of the Wald test of causality from the slope of the yield curve to economic activity reach rock bottom around 2009Q1, corresponding exactly to the start of the zero lower bound era of monetary policy. This result accords with expectations and confirms that the rolling window method provides the most dynamic sequence of test statistics. The rolling recursive test results do not show this severe drop-off but they do show a steady decline toward the end of the sample period. A slight rebound in the values in the sequence of the Wald statistics occurs around the 2010Q3 which may be attributed to the second round of quantitative easing in the United States. The effect on the slope of the term structure is evident in Figure 2, but the values of the test statistics remain well below their critical values.

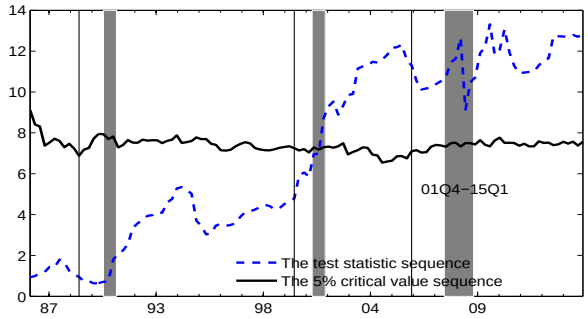
5.2 Output to the Yield Curve Slope

Figure 4 displays the time-varying Wald test statistics for causal effects running from the output gap to the slope of the yield curve. The rows of the figure again display sequences of tests obtained from forward recursive, rolling window and recursive-rolling procedures, while the columns refer to the different assumptions on the errors (homoskedasticity and heteroskedasticity) of the VAR. Sequences of test statistics start from 1986:Q4 as the first 28 observations represent the minimum window size.

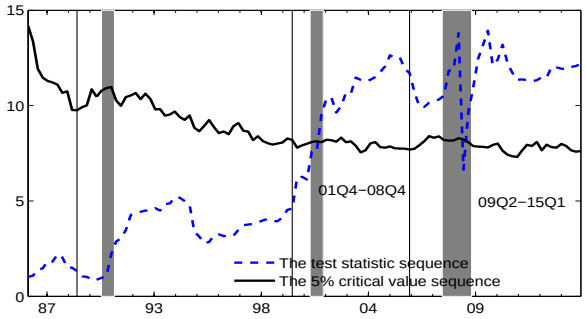
Both the rolling and recursive rolling procedures detect a causality episode during the 1990 recession, which is not found by the forward expanding procedure. The rolling procedure identifies two sub-periods (1990:Q1-1991:Q3 and 1993:Q1-1994:Q4), while the recursive rolling sup Wald sequence indicates a longer causal episode running from 1990:Q1 to 1995:Q1. Additionally, both the rolling and recursive rolling procedures finds a causality episode in the second half of 1990s, respectively, 1997:Q3-1998:Q1 and 1996:Q2-1999:Q1 and 1999:Q3. Notice that the termination dates estimated by the recursive rolling method are consistently later than those from the rolling window method, a pattern that accords with the simulation evidence presented previously.

Figure 4: Tests for Granger causality running from the yield curve slope to the inflation gap. Tests are obtained from a VAR model allowing for homoskedasitic errors (panels (a), (c) and (e)) and for heteroskedastic errors (panels (b), (d) and (f)). The sequence of tests for the forward recursive, rolling window and recursive-rolling procedures run from 1991:Q1 to 2013:Q4 with 22 observations for the minimum window size and a fixed lag order 2.

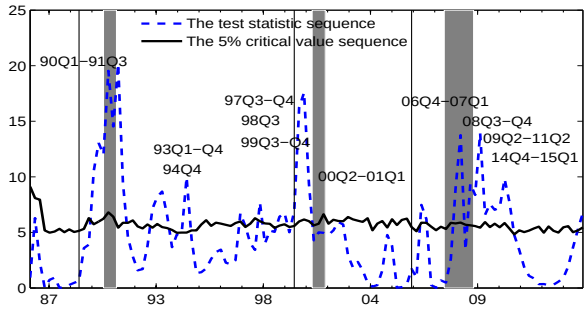
(a) Forward - Homoskedasticity



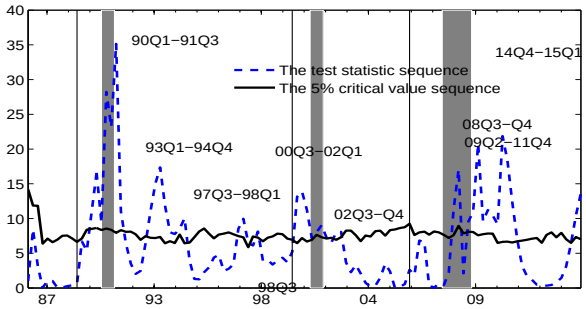
(b) Forward - Heteroskedasticity



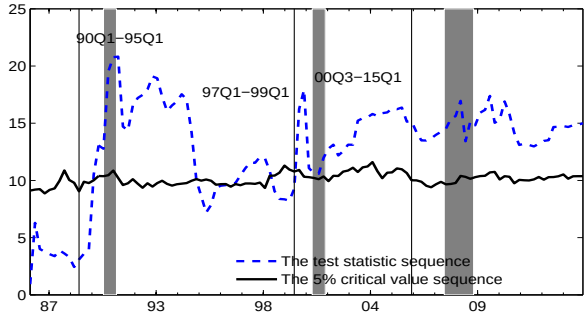
(c) Rolling - Homoskedasticity



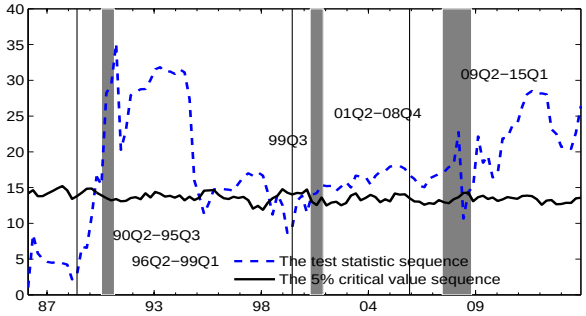
(d) Rolling - Heteroskedasticity



(e) Recursive Rolling - Homoskedasticity



(f) Recursive Rolling - Heteroskedasticity



All three procedures detect a change in causality in early 2000s. From panel (b), (d), and (f), the warning signal provided by the rolling algorithm for this change is earlier (2000:Q3) than that of the forward expanding and recursive rolling methods (2001:Q4 and 2001:Q2). The change in causality in the early 2000s appears to be a robust feature of the data and is also consistent with the results reported by Ang and Piazzesi (2003) who argue that macroeconomic factors are able to help explain movements in bonds yields primarily at the shorter maturities. There is, however, disagreement between the testing algorithms about the termination date of this causal episode. The rolling method identifies a termination date of 2002:Q4 (with a one-quarter break in 2002:Q2), whereas both the forward expanding and the recursive rolling procedures suggest the causal relationship lasts until the end of the sample period (with a short break in 2009:Q1). This result is consistent with the simulation evidence that the forward and recursive rolling procedures may fail to find causality termination dates when the strength of causality is strong.

The rolling window algorithm detects several additional short durations episodes from 2002:Q4 onwards, which include two episodes during and after the 2008 recession (2008:Q3-Q4 and 2009:Q2–2011:Q4) and one at the end of the sample period (2014:Q4-2015:Q1). This pattern is consistent with the simulations where the forward and recursive rolling window approaches were found to have low power in detecting later causality episodes.

6 Conclusion

This paper introduces a recursive rolling testing procedure to detect and date changes in Granger causal relationships which use sequences the supremum norm of Wald statstics. Variants of the test that are robust to departures from homoskedasticity are also examined. Asymptotic distributions of the tests are obtained and are shown to have simple forms that are amenable to computation for purpose of providing critical values. The recursive rolling procedure is compared to simple recursive testing and to tests based on a rolling window. The simulation findings suggest that the recursive rolling and the rolling window procedures are generally to be preferred to the simple forward recursive testing approach.

These tests are used to investigate causal relationships between the slope of the yield curve and the output gap using United States data over 1985-2013. The empirical application builds on earlier findings in the literature concerning bidirectional causal effects between the yield slope

on output. The results are consistent with much of the earlier literature but their most striking feature is that fact that causal relationships show considerable sensitivity to the subsample period. In sum, the various approaches to testing for causality reveal how endogenous detection of switches in causality gives useful insights about the trajectory of the macroeconomic impact of the yield curve slope and also issues a strong warning about the indiscriminate application of the tests over arbitrarily chosen subsamples.

References

Acharya, V.V., Drechsler, I. and Schnabl, P. (2011), A PYRRIC victory? - Bank bailouts and sovereign credit risk, *NBER Working Paper #17136,*.

Ang, A. and Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables, *Journal of Monetary Economics*, **50**, 745–787.

Ang, A., Piazzesi, M. and Wei, M. (2006). What does the yield curve tell us about GDP growth?, *Journal of Econometrics*, **131**, 359–403.

Arora, V., and Shi, S. (2015). Energy consumption and economic growth in the United States. Forthcoming *Applied Economics*.

Aruoba, S. B., Diebold, F. X. and Scotti, C. (2009). Real-time measurement of business conditions, *Journal of Business and Economic Statistics*, **27**, 417–427.

Balcilar, M., Ozdemir, Z. A., and Arslanturk, Y. (2010). Economic growth and energy consumption causal nexus viewed through a bootstrap rolling window. *Energy Economics*, **32**, 1398–1410.

Benati, L. and Goodhart, C. (2008). Investigating time-variation in the marginal predictive power of the yield spread, *Journal of Economic Dynamics and Control*, **32**, 1236–1272.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, **57**, 289–300.

Bianchi, F., Mumtaz, H. and Surico, P. (2009). Dynamics of the term structure of UK interest rates, *Working Paper #363*, Bank of England.

Billio, M., Getmansky, M., Lo, A. W. and Pelizzon, L. (2012), Econometric measures of connectedness and systemic risk in the finance and insurance sectors, *Journal of Financial Economics*, **104**, 535–559.

Bodnar, T., and Zabolotskyy, T. (2011). Estimation and inference of the vector autoregressive process under heteroscedasticity. *Theory of Probability and Mathematical Statistics*, **83**, 27–45.

Boswijk, H.P.,Cavaliere, G., Rahbek, A., and Taylor, A.M.R. (2013). Inference on co-integration parameters in heteroskedastic vector autoregressions. *Working Paper*, Institute of Economics, University of Copenhagen.

Brown, B.M. (1971). Martingale central limit theorems. *The Annals of Mathmatical Statistics*, **42**, 59-66.

Chauvet, M. and S. Potter (2002). Predicting recessions: Evidence from the yield curve in the presence of structural breaks, *Economics Letters*, **77**, 245–253.

Chauvet, M., and S. Potter (2005). Forecasting recessions using the yield curve, *Journal of Forecasting*, **24**, 77–103.

Chauvet, M. and Senyuz, Z. (2009). A joint dynamic bi-factor model of the yield curve and the economy as a predictor of business cycles, *Unpublished Manuscript*.

Chen, H., Cummins, J.D., Viswanathan, K.S. and Weiss, M.A. (2013), Systemic risk and the interconnectedness between banks and insurers: An econometric analysis, *Journal of Risk and Insurance*, **81**, 1 – 29.

Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians: An Introduction for Econometricians*, Oxford University Press: Oxford.

Dewachter, H. and Lyrio, M. (2006). Macro factors and the term structure of interest rates, *Journal of Money, Credit, and Banking*, **38**, 119–140.

Diebold, F., Rudebusch, G. D. and Aruoba, S. B. (2006), The macroeconomy and the yield curve: a dynamic latent factor approach, *Journal of Econometrics*, **131**, 309–338.

Diebold, F.X. and Yilmaz, K. (2013), Measuring the dynamics of global business cycle connectedness, *Working Paper #13-070*, Penn Institute for Economic Research, Department of Economics, University of Pennsylvania.

Dotsey, M. (1998). The predictive content of the interest rate term spread for future economic growth, *Federal Reserve Bank of Richmond Economic Quarterly*, **84**, 31–51.

Estrella, A. (2005). Why does the yield curve predict output and inflation, *The Economic Journal*, **115**, 722-744.

Estrella, A., and Hardouvelis, G.A. (1991). The term structure as a predictor of real economic activity, *The Journal of Finance,* **46**, 555-576.

Estrella, A., and Mishkin, F.S. (1997). The term structure of interest rates and its role in monetary policy for the European Central Bank, *European Economic Review*, **41**, 1375-1401.

Estrella, A., and Mishkin, F.S. (1998). Predicting U.S. recessions: Financial variables as leading indicators, *The Review of Economics and Statistics*, **80**, 45-61.

Estrella, A., Rodrigues, A.P., and Schich, S. (2003). How stable Is the predictive power of the yield curve? Evidence from Germany and the United States, *The Review of Economics and Statistics*, **85**, 629-644.

Estrella, A. and Trubin, M. (2006), The yield curve as a leading indicator: Some practical issues, *Federal Reserve Bank of New York Current Issues in Economics and Finance*, **12**, 1–7.

Estrella, A. and Wu, H. (2008). Term premiums and the predictability of recessions, *Upublished Manuscript.*

Evans, C. L. and Marshall, D. A. (2007). Economic determinants of the nominal treasury yield curve, *Journal of Monetary Economics*, **54**, 1986–2003.

Giacomini, R. and Rossi, B. (2006). How stable is the forecasting performance of the yield curve for output growth, *Oxford Bulletin of Economics and Statistics*, **68**, 783–795.

Gonçalves, S., and Kilian, L. (2004). Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. *Journal of Econometrics*, **123**, 89-120.

Granger, C.W.J. (1969), Investigating causal relations by econometric models and cross-spectral methods, *Econometrica*, **37**, 424–438.

Granger, C.W.J. (1988), Some recent development in a concept of causality, *Journal of Econometrics*, **39**, 199–211.

Guilkey, D. K. and Salemi, M. K. (1982). Small sample properties of three tests for Granger-causal ordering in a bivariate stochastic system. *Review of Economics and Statistics*, **64**, 668–680.

Hall, P. and Heyde, C.C. (1980) *Martingale limit theory and its application*, Academic Press: London.

Hamilton, J. D. (2010). Calling recessions in real time, *NBER Working Paper #16162.*

Hamilton, J. D. and Kim, D. H. (2002). Reexamination of the predictability of economic activity using the yield spread, *Journal of Money, Credit and Banking*, **34**, 340–360.

Hannan, E. J., and Heyde, C. C. (1972). On Limit Theorems for Quadratic Functions of Discrete Time Series. *The Annals of Mathematical Statistics*, **43**, 2058–2066.

Harvey, C. (1988). The real term structure and consumption growth *Journal of Financial Economics*, **22**, 305–333.

Haubrich, J.G. and Dombrosky, A.M. (1996). Predicting real growth using the yield curve *Federal Reserve Bank of Cleveland Economic Review*, **32**, 26–34.

Jorion, P. and Mishkin, F. S. (1991) A multicountry comparison of term-structure forecasts at long horizons, *Journal of Financial Economics*, **29**, 59–80.

Kauppi, H. and Saikkonen, P. (2008), Predicting U.S. recessions with dynamic binary response models, *Review of Economics and Statistics*, **90**, 777–791.

Kucko, K. and Chinn, M. D. (2009), The predictive power of the yield curve across countries and time, *Unpublished Manuscript.*

Leybourne, S., Kim, T-H., Taylor, A.M.R. (2007), Detecting multiple changes in persistence, *Studies in Nonlinear Dynamics and Econometrics*, **11**, 1–34.

Mishkin, F. S. (1990a), What does the term structure tell us about future inflation, *Journal of Monetary Economics*, **25**, 77–95.

Mishkin, F. S. (1990b), The information in the longer maturity term structure about future inflation, *The Quarterly Journal of Economics*, **442**, 815–828.

Mishkin, F. S. (1990c), A multi-country study of the information in the shorter maturity term structure about future inflation, *Journal of International Money and Finance*, **10**, 2–22.

Modena, M. (2008). An empirical analysis of the curvature factor of the term structure of interest rates, *MPRA Discussion Paper # 11597.*,

Moench, E. (2008). Term structure surprises: The predictive content of curvature, level, and slope, *Unpublished Manuscript*

Phillips, P.C.B., and Solo, V. (1992). Asymptotics for linear processes. *The Annals of Statistics*, **20**, 971–1001.

Phillips, P.C.B., Wu, Y. and Yu J. (2011). Explosive behaviour in the 1990s Nasdaq: When did exuberance escalate asset values? *International Economic Review*, **52**, 201–226.

Phillips, P.C.B. and Yu, J. (2011). Dating the timeline of financial bubbles during the Subprime Crisis. *Quantitative Economics*, **2**, 455–491.

Phillips, P.C.B., Shi, S. and Yu, J. (2015a). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International Economic Review*, forthcoming.

Phillips, P.C.B., Shi, S. and Yu, J. (2015b). Testing for Multiple bubbles: Limit theory of real time detectors. *International Economic Review*, forthcoming.

Phillips, P.C.B. and Xu, K. (2006). Inference in autoregression under heteroskedasticity. *Journal of Time Series Analysis,* **27**, 289–308.

Plosser, C. I., and Rouwenhorst, K. G. (1994), International term structures and real economic growth, *Journal of Monetary Economics*, **33**, 133–155.

Psaradakis, Z., Ravn, M. O. and Sola, M. (2005). Markov switching causality and the money-output relationship. *Journal of Applied Econometrics*, **20**, 665–683.

Rudebusch, G. D. and Williams, J. C. (2009). Forecasting recessions: The puzzle of the enduring power of the yield curve, *Journal of Business and Economics Statistics*, **27**, 492–503.

Stock, J. H. and Watson, M. W. (1989). Interpreting the evidence on money-income causality. *Journal of Econometrics*, **40**, 161–181.

Stock, J. H., and Watson, M. W. (1999). Business cycle fluctuations in U.S. macroeconomic time-series, in J. Taylor and M. Woodford (Eds), *Handbook of Macroeconomics*, Vol. 1, 3–64.

Wright, J. (2006). The yield curve and predicting recessions, *Finance and Economics Discussion Series*, Federal Reserve Board.

Stern, D. I. (2000). A multivariate cointegration analysis of the role of energy in the US macroeconomy. *Energy Economics*, **22**, 267-283

Swanson, N. R. (1998). Money and output viewed through a rolling window. *Journal of Monetary Economics*, **41**, 455–474

Thoma, M. A. (1994). Subsample instability and asymmetries in money-income causality. *Journal of Econometrics*, **64**, 279–306

Toda, H.Y. and P. C. B. Phillips (1993). "Vector autoregressions and causality, *Econometrica*, **61**, 1367–1393

Toda, H.Y. and Phillips, P.C.B. (1994). Vector autoregression and causality: a theoretical overview and simulation study, *Econometric Reviews*, **13**, 259–285.

A Appendix A: Limit Theory Under Assumption A1 and A2

Lemma 3.1 and Proposition 3.1 is first proved under Assumptions **A0** and **A2**. The proof for strictly stationary and ergodic sequences ε_t (Assumption **A1**) is standard and therefore omitted.

A.1 Proof of Lemma 3.1

(a) Write the estimation error as

$$\hat{\pi}_{f_1,f_2} - \pi_{f_1,f_2} = \left[\mathbf{I}_n \otimes \sum_{t=[Tf_1]}^{[Tf_2]} \mathbf{x}_t \mathbf{x}_t' \right]^{-1} \left[\sum_{t=[Tf_1]}^{[Tf_2]} \xi_t \right],$$

and, under **A2**, $\{\xi_t, \mathcal{F}_t\}$ is a covariance stationary mds with $\mathbb{E}(\xi_t|\mathcal{F}_{t-1}) = \mathbf{0}$ and $\sup_t \mathbb{E}\left(\|\xi_t\|^2\right) < \infty$, so that $T_w^{-1}\sum_{t=[Tf_1]}^{[Tf_2]} \xi_t \to_{a.s.} \mathbf{0}$ by a standard martingale strong law. Define $\hat{\mathbf{Q}}_{f_1,f_2} = \frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]} \mathbf{x}_t\mathbf{x}_t'$. Then, by a strong law for second order moments of linear processes (Phillips and Solo, 1992, Theorem 3.7), we have $\hat{\mathbf{Q}}_{f_1,f_2} \to_{a.s.} \mathbf{Q} = \mathbb{E}(x_t x_t') > 0$ and then

$$\hat{\pi}_{f_1,f_2} - \pi_{f_1,f_2} = \left[\mathbf{I}_n \otimes \hat{\mathbf{Q}}_{f_1,f_2}\right]^{-1} \left(\frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]} \xi_t\right) \to_{a.s.} \mathbf{0},$$

so that $\hat{\pi}_{f_1,f_2} \to_{a.s.} \pi_{f_1,f_2} = \pi$ under the maintained null of constant coefficients.

(b) Because $\hat{\varepsilon}_t = \varepsilon_t - \left(\hat{\pi}'_{f_1,f_2} - \pi'_{f_1,f_2}\right)(\mathbf{I}_n \otimes \mathbf{x}_t)$, we have

$$\hat{\mathbf{\Omega}}_{f_1,f_2} = \frac{1}{\lfloor Tf_w \rfloor}\sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \varepsilon_t\varepsilon_t' - \frac{2}{\lfloor Tf_w \rfloor}\sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \varepsilon_t \left(\mathbf{I}_n \otimes \mathbf{x}_t\right)' \left(\hat{\pi}_{f_1,f_2} - \pi_{f_1,f_2}\right)$$
$$+ \frac{1}{\lfloor Tf_w \rfloor}\sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \left(\hat{\pi}'_{f_1,f_2} - \pi'_{f_1,f_2}\right)\left(\mathbf{I}_n \otimes \mathbf{x}_t\mathbf{x}_t'\right)\left(\hat{\pi}_{f_1,f_2} - \pi_{f_1,f_2}\right) \xrightarrow{p} \mathbf{\Omega},$$

since $\frac{1}{\lfloor Tf_w \rfloor}\sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \varepsilon_t\varepsilon_t' \to_{a.s.} \mathbf{\Omega}$, $\hat{\pi}_{f_1,f_2} \to_{a.s.} \pi_{f_1,f_2}$, $\frac{1}{\lfloor Tf_w \rfloor}\sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \xi_t \to_{a.s.} \mathbf{0}$, and $\hat{\mathbf{Q}}_{f_1,f_2} \to_{a.s.} \mathbf{Q} > \mathbf{0}$.

(c) Under **A2** the martingale conditional variance satisfies the strong law

$$\frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]} \mathbb{E}\left(\xi_t\xi_t'|\mathcal{F}_{t-1}\right) = \mathbf{\Omega} \otimes \frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]} \mathbf{x}_t\mathbf{x}_t' \to_{a.s.} \mathbf{\Omega} \otimes \mathbf{Q} > \mathbf{0},$$

so that the stability condition for the martingale CLT is satisfied (Phillips and Solo, 1992, Theorem 3.4). Next, the conditional Lindeberg condition is shown to hold hold, so that for every $\delta > 0$

$$\frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]} \mathbb{E}\left\{\|\xi_t\|^2 .\mathbf{1}\left(\|\xi_t\| \geq \sqrt{T_w}\delta\right)|\mathcal{F}_{t-1}\right\} \xrightarrow{p} 0. \tag{24}$$

Let $A_T = \left\{\xi_t : \|\xi_t\| \geq \sqrt{T_w}\delta\right\}$. For some $\alpha \in (0, c/2)$ if follows that

$$\mathbb{E}\left[\|\xi_t\|^2 \mathbf{1}\left(\|\xi_t\| \geq \sqrt{T_w}\delta\right)\right] = \int_{A_T} \|\xi_t\|^2 dP \leq \frac{1}{\left(\sqrt{T_w}\delta\right)^\alpha}\int_{A_T} \|\xi_t\|^{2+\alpha} dP$$

Hence,

$$\mathbb{E}\left[\frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbb{E}\left\{\|\xi_t\|^2\mathbf{1}\left(\|\xi_t\|\geq\sqrt{T_w}\delta\right)|\mathcal{F}_{t-1}\right\}\right]$$
$$=\frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbb{E}\left\{\|\xi_t\|^2\mathbf{1}\left(\|\xi_t\|\geq\sqrt{T_w}\delta\right)\right\}$$
$$\leq T_w^{-\alpha/2}\delta^{-\alpha}\sup_t\mathbb{E}\left[\|\xi_t\|^{2+\alpha}\right]\leq T_w^{-\alpha/2}\delta^{-\alpha/2}K\sup_t\mathbb{E}\|\varepsilon_t\|^{4+2\alpha}\to 0$$

for some constant $K<\infty$ as $T\to\infty$ since

$$\mathbb{E}\|\xi_t\|^{2+\alpha}=\mathbb{E}\|\varepsilon_t\otimes\mathbf{x}_t\|^{2+\alpha}=\mathbb{E}\left(\|\varepsilon_t\|^{2+\alpha}\|\mathbf{x}_t\|^{2+\alpha}\right)$$
$$\leq K\sup_t\mathbb{E}\|\varepsilon_t\|^{4+2\alpha}<\infty,$$

in view of **A2**. Hence,

$$\frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbb{E}\left\{\|\xi_t\|^2.\mathbf{1}\left(\|\xi_t\|\geq\sqrt{T_w}\delta\right)|\mathcal{F}_{t-1}\right\}\xrightarrow{L_1}0,$$

which ensures that the Lindeberg condition (24) holds.

By the martingale invariance principle for linear processes (Phillips and Solo, 1992, Theorems 3.4), for $f_2>f_1$ it follows that

$$\frac{1}{\sqrt{T}}\sum_{t=[Tf_1]}^{[Tf_2]}\xi_t\Rightarrow B(f_2)-B(f_1),$$

where B is vector Brownian motion with covariance matrix $\mathbf{\Omega}\otimes\mathbf{Q}$. Rewriting $\sqrt{T}(\hat{\pi}_{f_1,f_2}-\pi_{f_1,f_2})$ as

$$\sqrt{T}(\hat{\pi}_{f_1,f_2}-\pi_{f_1,f_2})=\left[\mathbf{I}_n\otimes\hat{\mathbf{Q}}_{f_1,f_2}\right]^{-1}\left(\frac{T}{T_w}\frac{1}{\sqrt{T}}\sum_{t=[Tf_1]}^{[Tf_2]}\xi_t\right)\Rightarrow[\mathbf{I}_n\otimes\mathbf{Q}]^{-1}\left[\frac{B(f_2)-B(f_1)}{f_w}\right],\tag{25}$$

The limit in (25) may be interpreted as a linear functional of the limit process $B(\cdot)$, whose finite dimensional distribution for fixed f_1 and f_2 is simply $N\left(\mathbf{0},\mathbf{\Omega}\otimes f_w^{-1}\mathbf{Q}^{-1}\right)$, so that $\sqrt{T}(\hat{\pi}_{f_1,f_2}-\pi_{f_1,f_2})\xrightarrow{L}N\left(\mathbf{0},\mathbf{\Omega}\otimes f_w^{-1}\mathbf{Q}^{-1}\right)$, as stated.

A.2 Proof of Proposition 3.1

In view of (25), under the null hypothesis

$$\sqrt{T}\mathbf{R}\hat{\pi}_{f_1,f_2} \Rightarrow \mathbf{R}\left[\mathbf{I}_n \otimes \mathbf{Q}\right]^{-1}\left[\frac{\mathbf{B}\left(f_2\right)-\mathbf{B}\left(f_1\right)}{f_w}\right] = \mathbf{R}\left[\mathbf{\Omega}^{1/2}\otimes \mathbf{Q}^{-1/2}\right]\left[\frac{W_{nk}\left(f_2\right)-W_{nk}\left(f_1\right)}{f_w}\right],$$

where W_{nk} is vector standard Brownian motion with covariance matrix $\mathbf{I}_{nk}$. It follows that

$$\begin{aligned}
Z_{f_2}\left(f_1\right) &:= \left[\mathbf{R}\left(\hat{\mathbf{\Omega}}_{f_1,f_2}\otimes\left(\sum_{t=[Tf_1]}^{[Tf_2]}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}\right)\mathbf{R}'\right]^{-1/2}\mathbf{R}\hat{\pi}_{f_1,f_2} \\
&= \left[\mathbf{R}\left(\hat{\mathbf{\Omega}}_{f_1,f_2}\otimes\left(\frac{T_w}{T}\frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}\right)\mathbf{R}'\right]^{-1/2}\sqrt{T}\mathbf{R}\hat{\pi}_{f_1,f_2} \\
&\Rightarrow f_w^{1/2}\left[\mathbf{R}\left(\mathbf{\Omega}\otimes\mathbf{Q}^{-1}\right)\mathbf{R}'\right]^{-1/2}\mathbf{R}\left[\mathbf{I}_n\otimes\mathbf{Q}\right]^{-1}\left[\frac{B\left(f_2\right)-B\left(f_1\right)}{f_w}\right] \\
&= \left[\mathbf{R}\left(\mathbf{\Omega}\otimes\mathbf{Q}^{-1}\right)\mathbf{R}'\right]^{-1/2}\mathbf{R}\left[\mathbf{\Omega}^{1/2}\otimes\mathbf{Q}^{-1/2}\right]\left[\frac{W_{nk}\left(f_2\right)-W_{nk}\left(f_1\right)}{f_w}\right],
\end{aligned}$$

whose finite dimensional distribution for fixed f_1 and f_2 is $N\left(\mathbf{0},\mathbf{I}_d\right)$. Next, observe that the Wald statistic

$$\begin{aligned}
&W_{f_2}\left(f_1\right) = Z_{f_2}\left(f_1\right)' Z_{f_2}\left(f_1\right) \\
&\Rightarrow \left[\frac{W_{nk}\left(f_2\right)-W_{nk}\left(f_1\right)}{f_w^{1/2}}\right]'\left[\mathbf{\Omega}^{1/2}\otimes\mathbf{Q}^{-1/2}\right]\mathbf{R}'\left[\mathbf{R}\left(\mathbf{\Omega}\otimes\mathbf{Q}^{-1}\right)\mathbf{R}'\right]^{-1}\mathbf{R}\left[\mathbf{\Omega}^{1/2}\otimes\mathbf{Q}^{-1/2}\right]\left[\frac{W_{nk}\left(f_2\right)-W_{nk}\left(f_1\right)}{f_w^{1/2}}\right] \\
&= \left[\frac{W_{nk}\left(f_2\right)-W_{nk}\left(f_1\right)}{f_w^{1/2}}\right]'\mathbf{A}\left(\mathbf{A}'\mathbf{A}\right)^{-1}\mathbf{A}'\left[\frac{W_{nk}\left(f_2\right)-W_{nk}\left(f_1\right)}{f_w^{1/2}}\right], \text{ with } \mathbf{A}_{nk\times d} = \left[\mathbf{\Omega}^{1/2}\otimes\mathbf{Q}^{-1/2}\right]\mathbf{R}', \\
&=^d \left[\frac{W_d\left(f_2\right)-W_d\left(f_1\right)}{f_w^{1/2}}\right]'\left[\frac{W_d\left(f_2\right)-W_d\left(f_1\right)}{f_w^{1/2}}\right] \qquad (26)
\end{aligned}$$

which is a quadratic functional of the limit process $W_d\left(\cdot\right)$. The finite dimensional distribution of (26) for fixed f_1 and f_2 is χ^2_d. It follows by continuous mapping that as $T\to\infty$

$$SW_{f_2}\left(f_0\right) \xrightarrow{L} \sup_{f_1\in[0,f_2-f_0],f_2=f}\left[\frac{W_d\left(f_2\right)-W_d\left(f_1\right)}{f_w^{1/2}}\right]'\left[\frac{W_d\left(f_2\right)-W_d\left(f_1\right)}{f_w^{1/2}}\right],$$

where W_d is vector Brownian motion with covariance matrix $\mathbf{I}_d$.

B Appendix B: Limit Theory Under Assumption A3

This section provides proofs of Lemma 3.2, 3.3 and 3.4 and Proposition 3.2 and 3.3 under **A0** and **A3**.

B.1 Proof of Lemma 3.2

The proof of (a) follows directly from the strong law of large number for martingales (Hall and Heyde, 1980, theorem 2.19) under **A3**(i).

For the proof of (b) and (c), it is shown that for all $h \geq 0, z > 0$

$$\begin{aligned} P\left(\left\|\varepsilon_t \varepsilon_{t-h}'\right\| \geq z\right) &= P\left(\|\varepsilon_t\| \left\|\varepsilon_{t-h}'\right\| \geq z\right) \\ &\leq P\left(\|\varepsilon_t\| \geq z^{1/2}\right) + P\left(\|\varepsilon_{t-h}\| \geq z^{1/2}\right) \\ &\leq 2\gamma P\left(\|\varepsilon\|^2 \geq z\right). \end{aligned}$$

The last inequality follows by uniform integrability because $P\left(\|\varepsilon_t\| \geq z\right) \leq \gamma P\left(\|\varepsilon\| \geq z\right)$ for each $z \geq 0$, $t \geq 1$ and for some constant γ under **A3**(i). Therefore, from the martingale strong law

$$\frac{1}{T_w} \sum_{t=[Tf_1]}^{[Tf_2]} \varepsilon_t \varepsilon_t' \to_{a.s} \mathbf{\Omega} \text{ and } \frac{1}{T_w} \sum_{t=[Tf_1]}^{[Tf_2]} \varepsilon_t \varepsilon_s' \to_{a.s} 0 \text{ for } s \neq t.$$

See also Remarks 2.8(i) and (ii) of Phillips and Solo (1992).

For (d), by construction

$$\frac{1}{T_w} \sum_{t=[Tf_1]}^{[Tf_2]} \mathbf{x}_{t-1} \varepsilon_t' = \frac{1}{T_w} \sum_{t=[Tf_1]}^{[Tf_2]} \left[\begin{array}{cccc} \varepsilon_t & \varepsilon_t \mathbf{y}_{t-1}' & \cdots & \varepsilon_t \mathbf{y}_{t-p}' \end{array} \right]'.$$

and, from (a), $T_w^{-1} \sum_{t=[Tf_1]}^{[Tf_2]} \varepsilon_t \to_{a.s} \mathbf{0}$. Next consider the product $\mathbf{y}_{t-h} \varepsilon_t'$ with $1 \leq h \leq p$. Since

$$\mathbf{y}_{t-h} \varepsilon_t' = \left[\tilde{\Phi}_0 + \sum_{i=0}^{\infty} \Psi_i \varepsilon_{t-h-i} \right] \varepsilon_t' = \tilde{\Phi}_0 \varepsilon_t' + \sum_{i=0}^{\infty} \Psi_i \varepsilon_{t-h-i} \varepsilon_t',$$

it follows from absolute summability that $\sum_{i=0}^{\infty} \|\Psi_i\| < \infty$ and results (a) and (c), that $T_w^{-1} \sum_{t=[Tf_1]}^{[Tf_2]} \mathbf{y}_{t-h} \varepsilon_t' \to_{a.s} \mathbf{0}$, giving the required $T_w^{-1} \sum_{t=[Tf_1]}^{[Tf_2]} \mathbf{x}_{t-1} \varepsilon_t' \to_{a.s} \mathbf{0}$.

For (e), note that typical block elements of $\mathbf{x}_t\mathbf{x}_t'$ have the form $\mathbf{y}_{t-h}\mathbf{y}_{t-h-j}'$ and $\mathbf{y}_{t-h}$, so it suffices to calculate the limits of the following sample moments

$$(i)\ \frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbf{y}_{t-h},\ \text{where } 1\le h\le p;$$

$$(ii)\ \frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbf{y}_{t-h}\mathbf{y}_{t-h-j}',\ \text{where } 1\le h\le p \text{ and } 1\le j\le p-h.$$

Since $\mathbf{y}_{t-h}-\tilde{\Phi}_0=\sum_{i=0}^{\infty}\Psi_i\varepsilon_{t-h-i}$ and $\sum_{i=0}^{\infty}\|\Psi_i\|<\infty$ by virtue of **A0**, it follows that

$$\frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\left(\mathbf{y}_{t-h}-\tilde{\Phi}_0\right)=\frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\sum_{i=0}^{\infty}\Psi_i\varepsilon_{t-h-i}=\sum_{i=0}^{\infty}\Psi_i\left(\frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\varepsilon_{t-h-i}\right)\to_{a.s}0,$$

by results in (a), and

$$\begin{aligned}\frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\left(\mathbf{y}_{t-h}-\tilde{\Phi}_0\right)\left(\mathbf{y}_{t-h-j}-\tilde{\Phi}_0\right)'&=\frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\left(\sum_{i=0}^{\infty}\Psi_i\varepsilon_{t-h-i}\right)\left(\sum_{i=0}^{\infty}\Psi_i\varepsilon_{t-h-j-i}\right)'\\&\to_{a.s}\sum_{i=0}^{\infty}\Psi_{i+j}\mathbf{\Omega}\Psi_i',\end{aligned}$$

by results in (b) and (c). Hence,

$$T_w^{-1}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbf{y}_{t-h}\to_{a.s}\tilde{\Phi}_0,\ T_w^{-1}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbf{y}_{t-h}\mathbf{y}_{t-h-j}'\to_{a.s}\tilde{\Phi}_0\tilde{\Phi}_0'+\sum_{i=0}^{\infty}\Psi_{i+j}\mathbf{\Omega}\Psi_i',$$

giving

$$T_w^{-1}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbf{x}_{t-1}\mathbf{x}_{t-1}'\to_{a.s}\mathbf{Q}\equiv\begin{bmatrix}1 & \mathbf{1}_p'\otimes\tilde{\Phi}_0'\\ \mathbf{1}_p\otimes\tilde{\Phi}_0 & \mathbf{I}_p\otimes\tilde{\Phi}_0\tilde{\Phi}_0'+\mathbf{\Theta}\end{bmatrix},$$

with

$$\mathbf{\Theta}=\sum_{i=0}^{\infty}\begin{bmatrix}\Psi_i\mathbf{\Omega}\Psi_i' & \cdots & \Psi_{i+p-1}\mathbf{\Omega}\Psi_i'\\ \vdots & \ddots & \vdots\\ \Psi_i\mathbf{\Omega}\Psi_{i+p-1}' & \cdots & \Psi_i\mathbf{\Omega}\Psi_i'\end{bmatrix}.$$

B.2 Proof of Lemma 3.3

(a) The following conditional Lindeberg condition holds for all $\delta > 0$:

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\|\xi_t\|^2 .\mathbf{1}\left(\|\xi_t\| \geq \sqrt{T}\delta\right)|\mathcal{F}_{t-1}\right] \xrightarrow{p} 0. \tag{27}$$

Let $A_T = \left\{\xi_t : \|\xi_t\| \geq \sqrt{T}\delta\right\}$. For some $\alpha \in (0, c/2)$

$$\mathbb{E}\left[\|\xi_t\|^2 \mathbf{1}\left(\|\xi_t\| \geq \sqrt{T}\delta\right)\right] = \int_{A_T}\|\xi_t\|^2 dP \leq \frac{1}{\left(\sqrt{T}\delta\right)^{\alpha}}\int_{A_T}\|\xi_t\|^{2+\alpha} dP \leq \frac{1}{\left(\sqrt{T}\delta\right)^{\alpha}}\mathbb{E}\left(\|\xi_t\|^{2+\alpha}\right).$$

Hence,

$$\begin{aligned}
&\mathbb{E}\left\{\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\|\xi_t\|^2 \mathbf{1}\left(\|\xi_t\| \geq \sqrt{T}\delta\right)|\mathcal{F}_{t-1}\right]\right\} \\
&= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\|\xi_t\|^2 \mathbf{1}\left(\|\xi_t\| \geq \sqrt{T}\delta\right)\right] \\
&\leq T^{-\alpha/2}\delta^{-\alpha}\sup_t \mathbb{E}\left(\|\xi_t\|^{2+\alpha}\right) \leq T^{-\alpha/2}\delta^{-\alpha}K\sup_t \mathbb{E}\left(\|\varepsilon_t\|^{4+2\alpha}\right) \to 0,
\end{aligned}$$

for some constant $K < \infty$ as $T \to \infty$ since

$$\mathbb{E}\|\xi_t\|^{2+\alpha} = \mathbb{E}\|\varepsilon_t \otimes \mathbf{x}_t\|^{2+\alpha} \leq \mathbb{E}\left(\|\varepsilon_t\|^{2+\alpha}\|\mathbf{x}_t\|^{2+\alpha}\right) \leq K\mathbb{E}\|\varepsilon\|^{4+2\alpha} < \infty,$$

in view of **A3**(i) and the stability condition **A0** which ensures that $\|\mathbf{x}_t\| \leq A\sum_{i=0}^{\infty}\theta^i\|\varepsilon_{t-i}\|$ for some constant A and $|\theta| < 1$. Then (27) holds by L_1 convergence.

(b) The stability condition involves the convergences

$$\frac{1}{T}\sum_{t=1}^{T}\xi_t\xi_t', \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left\{\xi_t\xi_t'|\mathcal{F}_{t-1}\right\} \to_{a.s} \mathbf{W}. \tag{28}$$

By **A3**(i) and **A0,** it follows that $\mathbb{E}\left\{\|\xi_t\xi_t'\|^{1+\delta}\right\} = \mathbb{E}\left\{\|\varepsilon_t\varepsilon_t'\|^{1+\delta}\|\mathbf{x}_t\mathbf{x}_t'\|^{1+\delta}\right\} \leq K\mathbb{E}\|\varepsilon\|^{4+4\delta} < \infty$ for some finite $K > 0$ and $\delta < c/4$. Then, by the martingale strong law (Hall and Heyde, 1980,

theorem 2.19) we have $T^{-1}\sum_{t=1}^{T}\{\xi_t\xi_t'| - \mathbb{E}(\xi_t\xi_t'|\mathcal{F}_{t-1})\} \to_{a.s} 0$, where the limit

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left(\xi_t\xi_t'|\mathcal{F}_{t-1}\right) = W, \tag{29}$$

may be obtained by an explicit calculation using **A3**(ii) and (iii). By definition

$$\xi_t\xi_t' = \varepsilon_t\varepsilon_t' \otimes \mathbf{x}_t\mathbf{x}_t' = \begin{bmatrix} \varepsilon_{1,t}^2\mathbf{x}_t\mathbf{x}_t' & \cdots & \varepsilon_{1,t}\varepsilon_{n,t}\mathbf{x}_t\mathbf{x}_t' \\ \vdots & \ddots & \vdots \\ \varepsilon_{1,t}\varepsilon_{n,t}\mathbf{x}_t\mathbf{x}_t' & \cdots & \varepsilon_{n,t}^2\mathbf{x}_t\mathbf{x}_t' \end{bmatrix},$$

and therefore $\lim_{T\to\infty} T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(\varepsilon_{1,t}^2\mathbf{x}_t\mathbf{x}_t'|\mathcal{F}_{t-1}\right)$. The other limits can be computed in the same way. The leading block submatrix of $\xi_t\xi_t'$ is

$$\varepsilon_{1,t}^2\mathbf{x}_t\mathbf{x}_t' = \begin{bmatrix} \varepsilon_{1,t}^2 & \varepsilon_{1,t}^2\mathbf{y}_{t-1}' & \cdots & \varepsilon_{1,t}^2\mathbf{y}_{t-p}' \\ \varepsilon_{1,t}^2\mathbf{y}_{t-1} & \varepsilon_{1,t}^2\mathbf{y}_{t-1}\mathbf{y}_{t-1}' & \cdots & \varepsilon_{1,t}^2\mathbf{y}_{t-1}\mathbf{y}_{t-p}' \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{1,t}^2\mathbf{y}_{t-p} & \varepsilon_{1,t}^2\mathbf{y}_{t-p}\mathbf{y}_{t-1}' & \cdots & \varepsilon_{1,t}^2\mathbf{y}_{t-p}\mathbf{y}_{t-p}' \end{bmatrix}.$$

First, by the same martingale strong law $T^{-1}\sum_{t=1}^{T}\left\{\varepsilon_{1,t}^2 - \mathbb{E}\left(\varepsilon_{1,t}^2|\mathcal{F}_{t-1}\right)\right\} \to_{a.s} 0$ and from Lemma 3.2(b) $T^{-1}\sum_{t=1}^{T}\varepsilon_{1,t}^2 \to_{a.s} \Omega_{11}$, with $T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(\varepsilon_{1,t}^2|\mathcal{F}_{t-1}\right) \to_{a.s} \Omega_{11}$ from **A3**(ii). To obtain the limit of $T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(\varepsilon_{1,t}^2\mathbf{y}_{t-1}|\mathcal{F}_{t-1}\right)$, note that

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\varepsilon_{1,t}^2\left(\mathbf{y}_{t-1} - \tilde{\Phi}_0\right)|\mathcal{F}_{t-1}\right] = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left(\varepsilon_{1,t}^2|\mathcal{F}_{t-1}\right)\left(\mathbf{y}_{t-1} - \tilde{\Phi}_0\right) \\
&= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left(\varepsilon_{1,t}^2|\mathcal{F}_{t-1}\right)\sum_{i=0}^{\infty}\Psi_i\varepsilon_{t-1-i} = \sum_{i=0}^{\infty}\Psi_i\left[\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left(\varepsilon_{1,t}^2|\mathcal{F}_{t-1}\right)\varepsilon_{t-1-i}\right] \\
&\to_{a.s} 0,
\end{aligned}$$

from Assumption **A3**(iii) and **A0**. It follows that

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\varepsilon_{1,t}^2\mathbf{y}_{t-1}|\mathcal{F}_{t-1}\right] \to_{a.s} \Omega_{11}\tilde{\Phi}_0 \text{ and } \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left(\varepsilon_{1,t}^2\mathbf{y}_{t-1}'|\mathcal{F}_{t-1}\right) \to_{a.s} \Omega_{11}\tilde{\Phi}_0'.$$

Similarly, to obtain the limit of $T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(\varepsilon_{1,t}^{2}\mathbf{y}_{t-h}\mathbf{y}_{t-h-j}'|\mathcal{F}_{t-1}\right)$, observe that

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\varepsilon_{1,t}^{2}\left(\mathbf{y}_{t-h}-\tilde{\Phi}_{0}\right)\left(\mathbf{y}_{t-h-j}-\tilde{\Phi}_{0}\right)'|\mathcal{F}_{t-1}\right]\\
&=\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left(\varepsilon_{1,t}^{2}|\mathcal{F}_{t-1}\right)\left(\sum_{i=0}^{\infty}\Psi_{i}\varepsilon_{t-h-i}\right)\left(\sum_{i=0}^{\infty}\Psi_{i}\varepsilon_{t-h-j-i}\right)\\
&=\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left(\varepsilon_{1,t}^{2}|\mathcal{F}_{t-1}\right)\sum_{i=0}^{\infty}\Psi_{i+j}\varepsilon_{t-h-j-i}\varepsilon_{t-h-j-i}'\Psi_{i}'+o_{p}\left(1\right)\times\mathbf{11}'\\
&=\sum_{i=0}^{\infty}\Psi_{i+j}\left[\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left(\varepsilon_{1,t}^{2}|\mathcal{F}_{t-1}\right)\varepsilon_{t-h-j-i}\varepsilon_{t-h-j-i}'\right]\Psi_{i}'+o_{p}\left(1\right)\times\mathbf{11}'\\
&\rightarrow_{a.s}\sum_{i=0}^{\infty}\Psi_{i+j}\Omega_{11}\mathbf{\Omega}\Psi_{i}',
\end{aligned}$$

from Assumption **A3**(iii) and **A0**. It may be deduced that

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\varepsilon_{1,t}^{2}\mathbf{y}_{t-h}\mathbf{y}_{t-h-j}'|\mathcal{F}_{t-1}\right]\rightarrow_{a.s}\left[\Omega_{11}\tilde{\Phi}_{0}\tilde{\Phi}_{0}'+\sum_{i=0}^{\infty}\Psi_{i+j}\Omega_{11}\mathbf{\Omega}\Psi_{i}'\right].$$

Therefore

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left(\varepsilon_{1,t}^{2}\mathbf{x}_{t}\mathbf{x}_{t}'|\mathcal{F}_{t-1}\right)\\
&\rightarrow_{a.s}\left[\begin{array}{cccc}
\Omega_{11} & \Omega_{11}\tilde{\Phi}_{0}' & \cdots & \Omega_{11}\tilde{\Phi}_{0}'\\
\Omega_{11}\tilde{\Phi}_{0} & \Omega_{11}\tilde{\Phi}_{0}\tilde{\Phi}_{0}'+\sum_{i=0}^{\infty}\Psi_{i}\Omega_{11}\mathbf{\Omega}\Psi_{i}' & \cdots & \Omega_{11}\tilde{\Phi}_{0}\tilde{\Phi}_{0}'+\sum_{i=0}^{\infty}\Psi_{i+p-1}\Omega_{11}\mathbf{\Omega}\Psi_{i}'\\
\vdots & \vdots & \ddots & \vdots\\
\Omega_{11}\tilde{\Phi}_{0} & \Omega_{11}\tilde{\Phi}_{0}\tilde{\Phi}_{0}'+\sum_{i=0}^{\infty}\Psi_{i}\Omega_{11}\mathbf{\Omega}\Psi_{i+p-1}' & \cdots & \Omega_{11}\tilde{\Phi}_{0}\tilde{\Phi}_{0}'+\sum_{i=0}^{\infty}\Psi_{i}\Omega_{11}\mathbf{\Omega}\Psi_{i}'
\end{array}\right],
\end{aligned}$$

with similar calculations for the other components of the matrix partition, leading to the stability condition (29), with $\mathbf{W}=\left\{\mathbf{W}^{(i,j)}\right\}_{i,j\in[1,n]}$ defined in terms of the component matrix partitions

$$\mathbf{W}^{(i,j)}=\left[\begin{array}{cc}
\Omega_{ij} & \mathbf{1}_{p}'\otimes\Omega_{ij}\tilde{\Phi}_{0}'\\
\mathbf{1}_{p}\otimes\Omega_{ij}\tilde{\Phi}_{0} & \mathbf{I}_{p}\otimes\Omega_{ij}\tilde{\Phi}_{0}\tilde{\Phi}_{0}'+\mathbf{\Xi}^{(i,j)}
\end{array}\right],$$

and

$$\mathbf{\Xi}^{(i,j)} \equiv \sum_{i=0}^{\infty} \begin{bmatrix} \Psi_i \Omega_{ij} \mathbf{\Omega} \Psi_i' & \cdots & \Psi_{i+p-1} \Omega_{ij} \mathbf{\Omega} \Psi_i' \\ \vdots & \ddots & \vdots \\ \Psi_i \Omega_{ij} \mathbf{\Omega} \Psi_{i+p-1}' & \cdots & \Psi_i \Omega_{ij} \mathbf{\Omega} \Psi_i' \end{bmatrix}.$$

B.3 Proof of Lemma 3.4

(a) By definition

$$\hat{\pi}_{f_1,f_2} - \pi_{f_1,f_2} = \left[\mathbf{I}_n \otimes \frac{1}{T_w} \sum_{t=[Tf_1]}^{[Tf_2]} \mathbf{x}_t \mathbf{x}_t' \right]^{-1} \left[\frac{\sqrt{T}}{T_w} \frac{1}{\sqrt{T}} \sum_{t=[Tf_1]}^{[Tf_2]} \xi_t \right] \to_{a.s.} 0,$$

from Lemma 3.2(e) and (21).

(b) Using $\hat{\varepsilon}_t = \varepsilon_t - \left(\hat{\pi}_{f_1,f_2}' - \pi_{f_1,f_2}' \right) (\mathbf{I}_n \otimes \mathbf{x}_t)$, it follows that

$$\begin{aligned} \frac{1}{T_w} \sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \hat{\varepsilon}_t \hat{\varepsilon}_t' &= \frac{1}{T_w} \sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \varepsilon_t \varepsilon_t' - \frac{2}{T_w} \sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \varepsilon_t \left(\mathbf{I} \otimes \mathbf{x}_t \right)' \left(\hat{\pi}_{f_1,f_2} - \pi_{f_1,f_2} \right) \\ &+ \frac{1}{T_w} \sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \left(\hat{\pi}_{f_1,f_2}' - \pi_{f_1,f_2}' \right) \left(\mathbf{I} \otimes \mathbf{x}_t \mathbf{x}_t' \right) \left(\hat{\pi}_{f_1,f_2} - \pi_{f_1,f_2} \right) \to_{a.s.} \mathbf{\Omega}, \end{aligned}$$

since $T_w^{-1} \sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \varepsilon_t \varepsilon_t' \to_{a.s.} \mathbf{\Omega}$ from Lemma 3.2(b), $\hat{\pi}_{f_1,f_2} \to_{a.s.} \pi_{f_1,f_2}, T^{-1} \sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \xi_t \to_{a.s.} 0$, and $T_w^{-1} \sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \mathbf{x}_t \mathbf{x}_t' \to_{a.s.} \mathbf{Q} > 0$.

(c) The scaled and centred estimation error process is

$$\begin{aligned} \sqrt{T_w} \left(\hat{\pi}_{f_1,f_2} - \pi_{f_1,f_2} \right) &= \left[\mathbf{I}_n \otimes \frac{1}{T_w} \sum_{t=[Tf_1]}^{[Tf_2]} \mathbf{x}_t \mathbf{x}_t' \right]^{-1} \left[\frac{\sqrt{T}}{\sqrt{T_w}} \frac{1}{\sqrt{T}} \sum_{t=[Tf_1]}^{[Tf_2]} \xi_t \right] \\ &\Rightarrow f_w^{-1/2} \mathbf{V}^{-1} \left[\mathbf{B}(f_2) - \mathbf{B}(f_1) \right], \end{aligned}$$

whose finite dimensional distribution for fixed (f_1, f_2) is $\sqrt{T_w} \left(\hat{\pi}_{f_1,f_2} - \pi_{f_1,f_2} \right) \xrightarrow{L} N\left(0, \mathbf{V}^{-1} \mathbf{W} \mathbf{V}^{-1}\right)$, where $\mathbf{V} = \mathbf{I}_n \otimes \mathbf{Q}$.

(d) By definition

$$\begin{aligned}
&\frac{1}{T_w}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\hat{\xi}_t\hat{\xi}_t' = \frac{1}{T_w}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\left(\hat{\varepsilon}_t\hat{\varepsilon}_t'\otimes\mathbf{x}_t\mathbf{x}_t'\right)\\
&= \frac{1}{T_w}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\left[\varepsilon_t-\left(\hat{\pi}_{f_1,f_2}'-\pi_{f_1,f_2}'\right)\left(\mathbf{I}_n\otimes\mathbf{x}_t\right)\right]\left[\varepsilon_t-\left(\hat{\pi}_{f_1,f_2}'-\pi_{f_1,f_2}'\right)\left(\mathbf{I}_n\otimes\mathbf{x}_t\right)\right]'\otimes\mathbf{x}_t\mathbf{x}_t'\\
&= \frac{1}{T_w}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\varepsilon_t\varepsilon_t'\otimes\mathbf{x}_t\mathbf{x}_t' - \frac{2}{T_w}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\left[\left(\varepsilon_t\mathbf{I}_n\otimes\varepsilon_t\mathbf{x}_t'\right)\left(\hat{\pi}_{f_1,f_2}-\pi_{f_1,f_2}\right)\otimes\mathbf{x}_t\mathbf{x}_t'\right]\\
&+ \frac{1}{T_w}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\left(\hat{\pi}_{f_1,f_2}'-\pi_{f_1,f_2}'\right)\left(\mathbf{I}\otimes\mathbf{x}_t\mathbf{x}_t'\right)\left(\hat{\pi}_{f_1,f_2}-\pi_{f_1,f_2}\right)\otimes\mathbf{x}_t\mathbf{x}_t'\\
&= \frac{1}{T_w}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\xi_t\xi_t' + o_p(1)\to_{a.s}\mathbf{W}.
\end{aligned}$$

from Lemma 3.2(d) and (e), Lemma 3.4(a), and Lemma 3.3(b).

B.4 Proof of Proposition 3.2

In view of Lemma 3.4(c), under the null hypothesis

$$\begin{aligned}
\sqrt{T_w}\mathbf{R}\hat{\pi}_{f_1,f_2} &\Rightarrow f_w^{-1/2}\mathbf{R}\mathbf{V}^{-1}\left[B(f_2)-B(f_1)\right]\\
&= f_w^{-1/2}\mathbf{R}\mathbf{V}^{-1}\mathbf{W}^{1/2}\left[W_{nk}(f_2)-W_{nk}(f_1)\right],
\end{aligned}$$

where W_{nk} is vector standard Brownian motion with covariance matrix $\mathbf{I}_{nk}$. It follows that

$$\begin{aligned}
Z_{f_2}^*(f_1) &:= \left[\mathbf{R}\left(\hat{\mathbf{V}}_{f_1,f_2}^{-1}\hat{\mathbf{W}}_{f_1,f_2}\hat{\mathbf{V}}_{f_1,f_2}^{-1}\right)\mathbf{R}'\right]^{-1/2}\left(\sqrt{T_w}\mathbf{R}\hat{\pi}_{f_1,f_2}\right)\\
&\Rightarrow f_w^{-1/2}\left[\mathbf{R}\left(\mathbf{V}^{-1}\mathbf{W}\mathbf{V}^{-1}\right)\mathbf{R}'\right]^{-1/2}\mathbf{R}\mathbf{V}^{-1}\mathbf{W}^{1/2}\left[W_{nk}(f_2)-W_{nk}(f_1)\right].
\end{aligned}$$

Observe that the Wald statistic process

$$\begin{aligned}
W_{f_2}^*(f_1) &= Z_{f_2}^*(f_1)'Z_{f_2}^*(f_1)\\
&\Rightarrow f_w^{-1}\left[W_{nk}(f_2)-W_{nk}(f_1)\right]'\mathbf{A}\left(\mathbf{A}'\mathbf{A}\right)^{-1}\mathbf{A}'\left[W_{nk}(f_2)-W_{nk}(f_1)\right]\\
&=^d f_w^{-1}\left[W_d(f_2)-W_d(f_1)\right]'\left[W_d(f_2)-W_d(f_1)\right],
\end{aligned}$$

with $\mathbf{A} = W^{1/2}\mathbf{V}^{-1}\mathbf{R}'$, whose finite dimensional distribution for fixed (f_1, f_2) is χ^2_d. It follows by continuous mapping that as $T \to \infty$

$$\begin{aligned} SW^*_{f_2}(f_0) &\xrightarrow{L} \sup_{f_1 \in [0, f_2 - f_0], f_2 = f} \left[\frac{W_d(f_2) - W_d(f_1)}{f_w^{1/2}}\right]' \left[\frac{W_d(f_2) - W_d(f_1)}{f_w^{1/2}}\right] \\ &= \sup_{f_w \in [f_0, f_2], f_2 = f} \left[\frac{W_d(f_w)' W_d(f_w)}{f_w}\right], \end{aligned}$$

where W_d is vector Brownian motion with covariance matrix $\mathbf{I}_d$.

B.5 Proof of Proposition 3.3

In view of Lemma 3.4(c), under the null hypothesis

$$\begin{aligned} \sqrt{T_w}\mathbf{R}\hat{\pi}_{f_1,f_2} &\Rightarrow f_w^{-1/2}\mathbf{R}\mathbf{V}^{-1}\left[B(f_2) - B(f_1)\right] \\ &= f_w^{-1/2}\mathbf{R}\mathbf{V}^{-1}\mathbf{W}^{1/2}\left[W_{nk}(f_2) - W_{nk}(f_1)\right], \end{aligned}$$

where W_{nk} is vector standard Brownian motion with covariance matrix $\mathbf{I}_{nk}$. It follows that

$$\begin{aligned} Z_{f_2}(f_1) &:= \left[\mathbf{R}\left(\hat{\mathbf{\Omega}}_{f_1,f_2} \otimes \hat{\mathbf{Q}}_{f_1,f_2}\right)^{-1}\mathbf{R}'\right]^{-1/2}\left(\sqrt{T_w}\mathbf{R}\hat{\pi}_{f_1,f_2}\right) \\ &\Rightarrow f_w^{-1/2}\left[\mathbf{R}\left(\mathbf{\Omega} \otimes \mathbf{Q}\right)^{-1}\mathbf{R}'\right]^{-1/2}\mathbf{R}\mathbf{V}^{-1}\mathbf{W}^{1/2}\left[W_{nk}(f_2) - W_{nk}(f_1)\right]. \end{aligned}$$

Next, observe that the Wald statistic process

$$\begin{aligned} W_{f_2}(f_1) &= Z_{f_2}(f_1)' Z_{f_2}(f_1) \\ &\Rightarrow \left[\frac{W_{nk}(f_2) - W_{nk}(f_1)}{f_w^{1/2}}\right]' \mathbf{A}\mathbf{B}^{-1}\mathbf{A}' \left[\frac{W_{nk}(f_2) - W_{nk}(f_1)}{f_w^{1/2}}\right], \end{aligned}$$

with $\mathbf{A} = W^{1/2}\mathbf{V}^{-1}\mathbf{R}'$ and $\mathbf{B} = \mathbf{R}\left(\mathbf{\Omega} \otimes \mathbf{Q}\right)\mathbf{R}'$. It follows by continuous mapping that as $T \to \infty$

$$SW^0_{f_2}(f_0) \xrightarrow{L} \sup_{f_1 \in [0, f_2 - f_0], f_2 = f} \left[\frac{W_{nk}(f_2) - W_{nk}(f_1)}{f_w^{1/2}}\right]' \mathbf{A}\mathbf{B}^{-1}\mathbf{A}' \left[\frac{W_{nk}(f_2) - W_{nk}(f_1)}{f_w^{1/2}}\right],$$

giving the required result.

C Appendix C: Limit Theory Under Assumptions A4 and A5

This Appendix contains the proofs of Lemma 3.5, 3.6 and 3.7 and Proposition 3.4 allowing for unconditional heterogeneity in the errors under **A0**, **A4** and **A5**.

C.1 Proof of Lemma 3.5

(a) Under **A4**, by the martingale strong law and covariance stationarity of $\{\mathbf{u}_t\}$, it follows that

$$\frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbf{u}_t \to_{a.s.} \mathbb{E}\left(\mathbf{u}_t\right)=\mathbf{0} \tag{30}$$

$$\frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbf{u}_t\mathbf{u}_t' \to_{a.s.} \mathbb{E}\left(\mathbf{u}_t\mathbf{u}_t'\right)=\mathbf{I}_n \tag{31}$$

$$\frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbf{u}_t\mathbf{u}_s' \to_{a.s.} \mathbb{E}\left(\mathbf{u}_t\mathbf{u}_s'\right)=\mathbf{0} \text{ for } s\neq t, \tag{32}$$

where the results hold for every subsample involving sample fractions $f_1, f_2 \in [0,1]$ with $f_1 < f_2$. Further, since $\mathbf{G}(\cdot)$ is uniformly bounded and $\mathbf{u}_t$ is strongly uniformly integrable under **A5** , for small $\delta > 0$, it follows that

$$\sup_t \mathbb{E}\left\|\varepsilon_t\right\|^{1+\delta} \leq \sup_t \left\|\mathbf{G}\left(t/T\right)\right\|^{1+\delta} \sup_t \mathbb{E}\left\|\mathbf{u}_t\right\|^{1+\delta} < \infty$$

and then $T_w^{-1}\sum_{t=[Tf_1]}^{[Tf_2]}\varepsilon_t \to_{a.s.} \mathbf{0}$, again by the martingale strong law and for all fractions $f_1 < f_2$.

(b) The subsample second moment matrix of ε_t satisfies

$$\begin{aligned}
&\frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\varepsilon_t\varepsilon_t' = \frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbf{G}\left(t/T\right)\mathbf{u}_t\mathbf{u}_t'\mathbf{G}\left(t/T\right)' \\
&= \frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbf{G}\left(t/T\right)\mathbb{E}\left(\mathbf{u}_t\mathbf{u}_t'\right)\mathbf{G}\left(t/T\right)' + \frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbf{G}\left(t/T\right)\left\{\mathbf{u}_t\mathbf{u}_t' - \mathbb{E}\left(\mathbf{u}_t\mathbf{u}_t'\right)\right\}\mathbf{G}\left(t/T\right)' \\
&\to_{a.s.} \int_{f_1}^{f_2}\mathbf{G}\left(r\right)\mathbf{G}\left(r\right)'dr,
\end{aligned}$$

since $\mathbb{E}\left(\mathbf{u}_t\mathbf{u}_t'\right) = \mathbf{I}_n$, $T_w^{-1}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbf{G}\left(t/T\right)\mathbf{G}\left(t/T\right)' = \int_{[Tf_1]/T}^{([Tf_2]+1)/T}\mathbf{G}\left(r\right)\mathbf{G}\left(r\right)'dr + o\left(1\right) \to$

$\int_{f_1}^{f_2} \mathbf{G}(r)\,\mathbf{G}(r)'\,dr$, and

$$\frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]} \mathbf{G}(t/T)\left\{\mathbf{u}_t\mathbf{u}_t' - \mathbf{I}_n\right\}\mathbf{G}(t/T)' \to_{a.s.} 0,$$

again by the martingale strong law because

$$\sup_t \mathbb{E}\left\|\mathbf{G}(t/T)\left\{\mathbf{u}_t\mathbf{u}_t' - \mathbf{I}_n\right\}\mathbf{G}(t/T)\right\|^{1+\delta} \leq \sup_t \|\mathbf{G}(t/T)\|^{2+2\delta} \sup_t \mathbb{E}\left\|\mathbf{u}_t\mathbf{u}_t' - \mathbf{I}_n\right\|^{1+\delta} < \infty,$$

for all small $\delta > 0$ in view of **A5(i)** and the strong uniform integrability of $\|\mathbf{u}_t\|^4$.

(*c*) Using $\mathbb{E}\left(\mathbf{u}_{t-h-j-i}\mathbf{u}_{t-h-j-q}'\right) = \mathbf{I}_n\delta_{iq}$ where $\delta_{iq} = \mathbf{1}\{i=q\}$ is the Kronecker delta, so that

$$\begin{aligned}
&\frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\left(\mathbf{y}_{t-h} - \tilde{\Phi}_0\right)\left(\mathbf{y}_{t-h-j} - \tilde{\Phi}_0\right)' = \frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\left(\sum_{i=0}^{\infty}\Psi_i\varepsilon_{t-h-i}\right)\left(\sum_{q=0}^{\infty}\Psi_q\varepsilon_{t-h-j-q}\right)' \\
&= \frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\sum_{i=0}^{\infty}\Psi_{i+j}\mathbf{G}\left(\frac{t-h-j-i}{T}\right)\mathbf{G}\left(\frac{t-h-j-i}{T}\right)'\Psi_i' \\
&+ \frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\sum_{i,q=0}^{\infty}\Psi_{i+j}\mathbf{G}\left(\frac{t-h-j-i}{T}\right)\left\{\mathbf{u}_{t-h-j-i}\mathbf{u}_{t-h-j-q}' - \mathbf{I}_n\delta_{iq}\right\}\mathbf{G}\left(\frac{t-h-j-i}{T}\right)'\Psi_q' \\
&= \frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\sum_{i=0}^{\infty}\Psi_{i+j}\mathbf{G}\left(\frac{t-h-j-i}{T}\right)\mathbf{G}\left(\frac{t-h-j-i}{T}\right)'\Psi_i' + o_{a.s.}(1) \\
&= \int_{[Tf_1]/T}^{([Tf_2]+1)/T}\sum_{i=0}^{\infty}\Psi_{i+j}\mathbf{G}\left(\frac{[rT]-h-j-i}{T}\right)\mathbf{G}\left(\frac{[rT]-h-j-i}{T}\right)'\Psi_i' dr + o_{a.s.}(1),
\end{aligned}$$

since $\mathbf{G}$ is uniformly bounded, $\sum_{i,q=0}^{\infty}\|\Psi_{i+j}\|\left\|\Psi_q'\right\| < \infty$ uniformly in j by virtue of **A0,** and by the martingale strong law

$$\frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\left\{\mathbf{u}_{t-h-j-i}\mathbf{u}_{t-h-j-q}' - \mathbf{I}_n\delta_{iq}\right\} \to_{a.s.} 0.$$

Further,

$$\sum_{i=0}^{\infty}\Psi_{i+j}\mathbf{G}\left(\frac{[rT]-h-j-i}{T}\right)\mathbf{G}\left(\frac{[rT]-h-j-i}{T}\right)'\Psi_i'$$

$$= \sum_{i=0}^{S} \Psi_{i+j} \mathbf{G}\left(\frac{[rT]-h-j-i}{T}\right) \mathbf{G}\left(\frac{[rT]-h-j-i}{T}\right)' \Psi_i' \tag{33}$$

$$+ \sum_{i=S+1}^{\infty} \Psi_{i+j} \mathbf{G}\left(\frac{[rT]-h-j-i}{T}\right) \mathbf{G}\left(\frac{[rT]-h-j-i}{T}\right)' \Psi_i' \tag{34}$$

where $S > 0$ such that $\frac{S}{T} + \frac{1}{S} \to 0$. The first term (33) satisfies

$$\sum_{i=0}^{S} \Psi_{i+j} \mathbf{G}\left(\frac{[rT]-h-j-i}{T}\right) \mathbf{G}\left(\frac{[rT]-h-j-i}{T}\right)' \Psi_i' \to_{a.s} \sum_{i=0}^{\infty} \Psi_{i+j} \mathbf{G}(r) \mathbf{G}(r)' \Psi_i',$$

and the second term (34) tends to zero because $\|\Psi_i\| < C\theta^i$ with $|\theta| < 1$, which gives

$$\sum_{i=S+1}^{\infty} \left\| \Psi_{i+j} \mathbf{G}\left(\frac{[rT]-h-j-i}{T}\right) \mathbf{G}\left(\frac{[rT]-h-j-i}{T}\right)' \Psi_i' \right\| \leq C^2 \theta^j \sum_{i=S+1}^{\infty} \theta^{2i} \to 0,$$

as $T, S \to \infty$. Thus,

$$\sum_{i=0}^{\infty} \Psi_{i+j} \mathbf{G}\left(\frac{[rT]-h-j-i}{T}\right) \mathbf{G}\left(\frac{[rT]-h-j-i}{T}\right)' \Psi_i' \to_{a.s} \sum_{i=0}^{\infty} \Psi_{i+j} \mathbf{G}(r) \mathbf{G}(r)' \Psi_i',$$

and hence

$$\frac{1}{T_w} \sum_{t=[Tf_1]}^{[Tf_2]} \left(\mathbf{y}_{t-h} - \tilde{\Phi}_0\right) \left(\mathbf{y}_{t-h-j} - \tilde{\Phi}_0\right)' \to_{a.s} \int_{f_1}^{f_2} \sum_{i=0}^{\infty} \Psi_{i+j} \mathbf{G}(r) \mathbf{G}(r)' \Psi_i' dr,$$

as stated.

(c) It is required to show that $T_w^{-1} \sum_{t=[Tf_1]}^{[Tf_2]} \mathbf{x}_t \varepsilon_t \to_{a.s} \mathbf{0}$. It suffices to show that $T_w^{-1} \sum_{t=[Tf_1]}^{[Tf_2]} \varepsilon_t \to_{a.s} \mathbf{0}$, which holds by (a), and $T_w^{-1} \sum_{t=[Tf_1]}^{[Tf_2]} \mathbf{y}_{t-h} \varepsilon_t \to_{a.s} \mathbf{0}$ for $1 \leq h \leq p$. Under **A0,** $\mathbf{y}_{t-h} \varepsilon_t = \left(\tilde{\Phi}_0 + \sum_{i=0}^{\infty} \Psi_i \varepsilon_{t-h-i}\right) \varepsilon_t$ and $T_w^{-1} \sum_{t=[Tf_1]}^{[Tf_2]} \tilde{\Phi}_0 \varepsilon_t \to_{a.s} \mathbf{0}$ holds by (a). Next,

$$\frac{1}{T_w} \sum_{t=[Tf_1]}^{[Tf_2]} \sum_{i=0}^{\infty} \Psi_i \varepsilon_{t-h-i} \varepsilon_t = \sum_{i=0}^{\infty} \Psi_i \left(T_w^{-1} \sum_{t=[Tf_1]}^{[Tf_2]} \varepsilon_{t-h-i} \varepsilon_t \right) \to_{a.s.} \mathbf{0},$$

by the martingale strong law since $\varepsilon_t = \mathbf{G}(t/T) \mathbf{u}_t$, $\mathbf{G}$ is uniformly bounded, $h \geq 1$, and $\mathbf{u}_{t-h-i} \mathbf{u}_t$ is strongly uniformly integrable with dominating random variable $\mathbf{u}$ satisfying $\mathbb{E} \|\mathbf{u}_t\|^{4+c} < \infty$ by **A5(i)**. It follows that $T_w^{-1} \sum_{t=[Tf_1]}^{[Tf_2]} \mathbf{x}_{t-1} \varepsilon_t' \to_{a.s} \mathbf{0}$. as required.

(d) It is known that

$$\mathbf{x}_t\mathbf{x}_t' = \begin{bmatrix} 1 & \mathbf{y}_{t-1}' & \cdots & \mathbf{y}_{t-p}' \\ \mathbf{y}_{t-1} & \mathbf{y}_{t-1}\mathbf{y}_{t-1}' & \cdots & \mathbf{y}_{t-1}\mathbf{y}_{t-p}' \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{y}_{t-p} & \mathbf{y}_{t-p}\mathbf{y}_{t-1}' & \cdots & \mathbf{y}_{t-p}\mathbf{y}_{t-p}' \end{bmatrix}.$$

In order to prove the statement, the following limits for $1 \leq h \leq p$ must be calculated:

$$(i)\ \lim_{T\to\infty} \frac{1}{T_w} \sum_{t=[Tf_1]}^{[Tf_2]} \mathbf{y}_{t-h}; (ii)\ \lim_{T\to\infty} \frac{1}{T_w} \sum_{t=[Tf_1]}^{[Tf_2]} \mathbf{y}_{t-h}\mathbf{y}_{t-h-j}',\ \text{for } 1 \leq j \leq p-h.$$

Since $\mathbf{y}_{t-h} - \tilde{\Phi}_0 = \sum_{i=0}^{\infty} \Psi_i \varepsilon_{t-h-i}$, it follows that

$$\frac{1}{T_w} \sum_{t=[Tf_1]}^{[Tf_2]} \left(\mathbf{y}_{t-h} - \tilde{\Phi}_0\right) = \frac{1}{T_w} \sum_{t=[Tf_1]}^{[Tf_2]} \sum_{i=0}^{\infty} \Psi_i \varepsilon_{t-h-i} = \sum_{i=0}^{\infty} \Psi_i \left(\frac{1}{T_w} \sum_{t=[Tf_1]}^{[Tf_2]} \varepsilon_{t-h-i} \right) \to_{a.s} \mathbf{0}$$

by (a); and from (c),

$$\frac{1}{T_w} \sum_{t=[Tf_1]}^{[Tf_2]} \left(\mathbf{y}_{t-h} - \tilde{\Phi}_0\right) \left(\mathbf{y}_{t-h-j} - \tilde{\Phi}_0\right)' \to_{a.s} \int_{f_1}^{f_2} \sum_{i=0}^{\infty} \Psi_{i+j} \mathbf{G}(r) \mathbf{G}(r)' \Psi_i' dr.$$

Thus

$$\frac{1}{T_w} \sum_{t=[Tf_1]}^{[Tf_2]} \mathbf{y}_{t-h} \to_{a.s} \tilde{\Phi}_0,\ \frac{1}{T_w} \sum_{t=[Tf_1]}^{[Tf_2]} \mathbf{y}_{t-h}\mathbf{y}_{t-h-j}' \to_{a.s} \tilde{\Phi}_0\tilde{\Phi}_0' + \int_{f_1}^{f_2} \sum_{i=0}^{\infty} \Psi_{i+j} \mathbf{G}(r) \mathbf{G}(r)' \Psi_i' dr,$$

so that

$$\frac{1}{T_w} \sum_{t=[Tf_1]}^{[Tf_2]} \mathbf{x}_t\mathbf{x}_t' \to_{a.s} \mathbf{Q}_{f_1,f_2} = \begin{bmatrix} 1 & \mathbf{1}_p' \otimes \tilde{\Phi}_0' \\ \mathbf{1}_p \otimes \tilde{\Phi}_0 & \mathbf{I}_p \otimes \tilde{\Phi}_0\tilde{\Phi}_0' + \mathbf{\Theta}_{f_1,f_2} \end{bmatrix},$$

where

$$\mathbf{\Theta}_{f_1,f_2} \equiv \int_{f_1}^{f_2} \sum_{i=0}^{\infty} \begin{bmatrix} \Psi_i \mathbf{G}(r) \mathbf{G}(r)' \Psi_i' & \cdots & \Psi_{i+p-1} \mathbf{G}(r) \mathbf{G}(r)' \Psi_i' \\ \vdots & \ddots & \vdots \\ \Psi_i \mathbf{G}(r) \mathbf{G}(r)' \Psi_{i+p-1}' & \cdots & \Psi_i \mathbf{G}(r) \mathbf{G}(r)' \Psi_i' \end{bmatrix} dr.$$

C.2 Proof of Lemma 3.6

(a) As in the proof of lemma 3.3, the procedure is to show that the conditional Lindeberg conditon holds

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\left\|\xi_t\right\|^2.\mathbf{1}\left(\left\|\xi_t\right\|\geq\sqrt{T}\delta\right)|\mathcal{F}_{t-1}\right]\xrightarrow{p}0 \text{ for all } \delta>0. \tag{35}$$

Let $A_T=\left\{\xi_t:\left\|\xi_t\right\|\geq\sqrt{T}\delta\right\}$. For some $\alpha\in(0,c/2)$

$$\mathbb{E}\left[\left\|\xi_t\right\|^2\mathbf{1}\left(\left\|\xi_t\right\|\geq\sqrt{T}\delta\right)\right]=\int_{A_T}\left\|\xi_t\right\|^2dP\leq\frac{1}{\left(\sqrt{T}\delta\right)^{\alpha}}\int_{A_T}\left\|\xi_t\right\|^{2+\alpha}dP\leq\frac{1}{\left(\sqrt{T}\delta\right)^{\alpha}}\mathbb{E}\left(\left\|\xi_t\right\|^{2+\alpha}\right).$$

Hence,

$$\begin{aligned}
&\mathbb{E}\left\{\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\left\|\xi_t\right\|^2\mathbf{1}\left(\left\|\xi_t\right\|\geq\sqrt{T}\delta\right)|\mathcal{F}_{t-1}\right]\right\}\\
&=\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\left\|\xi_t\right\|^2\mathbf{1}\left(\left\|\xi_t\right\|\geq\sqrt{T}\delta\right)\right]\\
&\leq T^{-\alpha/2}\delta^{-\alpha}\sup_t\mathbb{E}\left(\left\|\xi_t\right\|^{2+\alpha}\right)\leq T^{-\alpha/2}\delta^{-\alpha}K\sup_t\mathbb{E}\left(\left\|\varepsilon_t\right\|^{4+2\alpha}\right)\to 0,
\end{aligned}$$

for some constant $K<\infty$ as $T\to\infty$ since

$$\mathbb{E}\left\|\xi_t\right\|^{2+\alpha}=\mathbb{E}\left\|\varepsilon_t\otimes\mathbf{x}_{t-1}\right\|^{2+\alpha}=\mathbb{E}\left(\left\|\varepsilon_t\right\|^{2+\alpha}\left\|\mathbf{x}_{t-1}\right\|^{2+\alpha}\right)\leq K\sup_t\left\|\mathbf{G}\left(r\right)\right\|^{4+2\alpha}\mathbb{E}\left\|\mathbf{u}\right\|^{4+2\alpha}<\infty,$$

in view of **A5**(i) and (ii). Hence, (35) follows.

(b) Again the proof of Lemma 3.3 is followed. The stability condition involves the convergence

$$\frac{1}{T}\sum_{t=1}^{T}\xi_t\xi_t',\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left\{\xi_t\xi_t'|\mathcal{F}_{t-1}\right\}\to_{a.s}\mathbf{W}.$$

By **A5**(i) and **A0**, it follows that $\mathbb{E}\left\{\left\|\xi_t\xi_t'\right\|^{1+\delta}\right\}=\mathbb{E}\left\{\left\|\xi_t\right\|^{2+2\delta}\right\}=\mathbb{E}\left\{\left\|\varepsilon_t\varepsilon_t'\right\|^{2+2\delta}\left\|\mathbf{x}_t\mathbf{x}_t'\right\|^{2+2\delta}\right\}\leq K\mathbb{E}\left\|\varepsilon\right\|^{4+4\delta}<\infty$ for some finite $K>0$ and $\delta<c/4$. Then, by the martingale strong law (Hall and Heyde, 1980, theorem 2.19) it follows that $T^{-1}\sum_{t=1}^{T}\left\{\xi_t\xi_t'|-\mathbb{E}\left(\xi_t\xi_t'|\mathcal{F}_{t-1}\right)\right\}\to_{a.s}0$, where the limit

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left(\xi_t\xi_t'|\mathcal{F}_{t-1}\right)=W,$$

may be obtained by an explicit calculation using **A5**(ii) and (iii). Again, by definition,

$$\xi_t\xi_t' = \left(\varepsilon_t\varepsilon_t'\right) \otimes \left(\mathbf{x}_t\mathbf{x}_t'\right) = \left[\begin{array}{ccc} \varepsilon_{1,t}^2\mathbf{x}_t\mathbf{x}_t' & \cdots & \varepsilon_{1,t}\varepsilon_{n,t}\mathbf{x}_t\mathbf{x}_t' \\ \vdots & \ddots & \vdots \\ \varepsilon_{1,t}\varepsilon_{n,t}\mathbf{x}_t\mathbf{x}_t' & \cdots & \varepsilon_{n,t}^2\mathbf{x}_t\mathbf{x}_t' \end{array}\right]_{nk\times nk}.$$

The limit $\lim_{T\to\infty}\frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbb{E}\left(\varepsilon_{1,t}^2\mathbf{x}_{t-1}\mathbf{x}_{t-1}'|\mathcal{F}_{t-1}\right)$ is calculated and the other limits in probability are computed in the same way. It then follows that

$$\varepsilon_{1,t}^2\mathbf{x}_t\mathbf{x}_t' = \left[\begin{array}{ccc} \varepsilon_{1,t}^2 & \varepsilon_{1,t}^2\mathbf{y}_{t-1}' & \cdots & \varepsilon_{1,t}^2\mathbf{y}_{t-p}' \\ \varepsilon_{1,t}^2\mathbf{y}_{t-1} & \varepsilon_{1,t}^2\mathbf{y}_{t-1}\mathbf{y}_{t-1}' & \cdots & \varepsilon_{1,t}^2\mathbf{y}_{t-1}\mathbf{y}_{t-p}' \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_{1,t}^2\mathbf{y}_{t-p} & \varepsilon_{1,t}^2\mathbf{y}_{t-p}\mathbf{y}_{t-1}' & \cdots & \varepsilon_{1,t}^2\mathbf{y}_{t-p}\mathbf{y}_{t-p}' \end{array}\right]_{k\times k}.$$

Write $\mathbf{G} = (g_{iq})$ and then from lemma 3.5(b)

$$\begin{aligned} &\frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbb{E}\left(\varepsilon_{1,t}^2|\mathcal{F}_{t-1}\right) = \frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbb{E}\left[\left(\sum_{q=1}^{n}g_{1q}\left(t/T\right)u_{jt}\right)^2|\mathcal{F}_{t-1}\right] \\ &= \frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\sum_{q=1}^{n}g_{1q}^2\left(t/T\right)\mathbb{E}\left(u_{qt}^2|\mathcal{F}_{t-1}\right) = \int_{[Tf_1]/T}^{([Tf_2]+1)/T}\sum_{q=1}^{n}g_{1q}^2\left(r\right)dr \\ &\to_{a.s.}\int_{f_1}^{f_2}\sum_{q=1}^{n}g_{1q}^2\left(r\right)dr. \end{aligned}$$

Next,

$$\begin{aligned} &\frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbb{E}\left[\varepsilon_{1,t}^2\left(\mathbf{y}_{t-h}-\tilde{\Phi}_0\right)|\mathcal{F}_{t-1}\right] = \frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbb{E}\left(\varepsilon_{1,t}^2|\mathcal{F}_{t-1}\right)\left(\mathbf{y}_{t-h}-\tilde{\Phi}_0\right) \\ &= \sum_{i=0}^{\infty}\Psi_i\left[\frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbb{E}\left(\varepsilon_{1,t}^2|\mathcal{F}_{t-1}\right)\varepsilon_{t-h-i}\right]\to_{a.s}0, \end{aligned}$$

because $\sum_{i=0}^{\infty}\|\Psi_i\|<\infty$ and

$$\frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbb{E}\left(\varepsilon_{1,t}^2|\mathcal{F}_{t-1}\right)\varepsilon_{t-h-i} = \sum_{j=1}^{n}\frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}g_{1j}^2\left(t/T\right)\mathbf{G}\left(\frac{t-h-i}{T}\right)\mathbf{u}_{t-h-i}\to_{a.s}\mathbf{0},$$

by the martingale strong law using **A5**(i) and uniform boundedness of the elements of **G**. It

follows that

$$\frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbb{E}\left[\varepsilon_{1,t}^2\mathbf{y}_{t-h}|\mathcal{F}_{t-1}\right]\rightarrow_{a.s}\int_{f_1}^{f_2}\sum_{q=1}^{n}g_{1q}^2\left(r\right)dr\tilde{\Phi}_0.$$

To evaluate $\lim_{T\rightarrow\infty}\frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbb{E}\left(\varepsilon_{1,t}^2\mathbf{y}_{t-h}\mathbf{y}_{t-h-j}'|\mathcal{F}_{t-1}\right)$, consider

$$\begin{aligned}
&\frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbb{E}\left[\varepsilon_{1,t}^2\left(\mathbf{y}_{t-h}-\tilde{\Phi}_0\right)\left(\mathbf{y}_{t-h-j}-\tilde{\Phi}_0\right)'|\mathcal{F}_{t-1}\right]\\
&=\sum_{q=1}^{n}g_{1q}^2\left(r\right)\left[\frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\left(\mathbf{y}_{t-h}-\tilde{\Phi}_0\right)\left(\mathbf{y}_{t-h-j}-\tilde{\Phi}_0\right)'\right]\\
&\rightarrow_{a.s}\sum_{i=0}^{\infty}\Psi_{i+j}\int_{f_1}^{f_2}\sum_{q=1}^{n}g_{1q}^2\left(r\right)\mathbf{G}\left(r\right)\mathbf{G}\left(r\right)'dr\Psi_i',
\end{aligned}$$

which follows from Lemma 3.5(c). Thus,

$$\frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbb{E}\left[\varepsilon_{1,t}^2\mathbf{y}_{t-h}\mathbf{y}_{t-h-j}'|\mathcal{F}_{t-1}\right]\rightarrow_{a.s}\int_{f_1}^{f_2}\sum_{q=1}^{n}g_{1q}^2\left(r\right)dr\tilde{\Phi}_0\tilde{\Phi}_0'+\sum_{i=0}^{\infty}\Psi_{i+j}\int_{f_1}^{f_2}\sum_{q=1}^{n}g_{1q}^2\left(r\right)\mathbf{G}\left(r\right)\mathbf{G}\left(r\right)'dr\Psi_i'.$$

and hence $T_w^{-1}\sum_{t=[Tf_1]}^{[Tf_2]}\xi_t\xi_t'\rightarrow_{a.s}\mathbf{W}_{f_1,f_2}$, where $\mathbf{W}_{f_1,f_2}=\left\{\mathbf{W}_{f_1,f_2}^{(i,j)}\right\}_{i,j\in[1,n]}$ with

$$\mathbf{W}_{f_1,f_2}^{(i,j)}=\begin{bmatrix}\int_{f_1}^{f_2}\sum_{q=1}^{n}g_{iq}\left(r\right)g_{jq}\left(r\right)dr & \mathbf{1}_p'\otimes\int_{f_1}^{f_2}\sum_{q=1}^{n}g_{iq}\left(r\right)g_{jq}\left(r\right)dr\tilde{\Phi}_0'\\ \mathbf{1}_p\otimes\int_{f_1}^{f_2}\sum_{q=1}^{n}g_{iq}\left(r\right)g_{jq}\left(r\right)dr\tilde{\Phi}_0 & \mathbf{I}_p\otimes\int_{f_1}^{f_2}\sum_{q=1}^{n}g_{iq}\left(r\right)g_{jq}\left(r\right)dr\tilde{\Phi}_0\tilde{\Phi}_0'+\mathbf{\Xi}_{f_1,f_2}^{(i,j)}\end{bmatrix},$$

and

$$\begin{aligned}
\mathbf{\Xi}_{f_1,f_2}^{(i,j)}&\equiv\sum_{i=0}^{\infty}\begin{bmatrix}\Psi_i\mathbf{\Lambda}_{f_1,f_2}^{(i,j)}\Psi_i' & \cdots & \Psi_{i+p-1}\mathbf{\Lambda}_{f_1,f_2}^{(i,j)}\Psi_i'\\ \vdots & \ddots & \vdots\\ \Psi_i\mathbf{\Lambda}_{f_1,f_2}^{(i,j)}\Psi_{i+p-1}' & \cdots & \Psi_i\mathbf{\Lambda}_{f_1,f_2}^{(i,j)}\Psi_i'\end{bmatrix},\\
\mathbf{\Lambda}_{f_1,f_2}^{(i,j)}&=\int_{f_1}^{f_2}\sum_{q=1}^{n}g_{iq}\left(r\right)g_{jq}\left(r\right)\mathbf{G}\left(r\right)\mathbf{G}\left(r\right)'dr.
\end{aligned}$$

C.3 Proof of Lemma 3.7

(a) As shown earlier,

$$\hat{\pi}_{f_1,f_2} - \pi_{f_1,f_2} = \left[\mathbf{I}_n \otimes \frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]} \mathbf{x}_t\mathbf{x}_t'\right]^{-1}\left[\frac{\sqrt{T}}{T_w}\frac{1}{\sqrt{T}}\sum_{t=[Tf_1]}^{[Tf_2]}\xi_t\right] \to_{a.s.} \mathbf{0},$$

using Lemma 3.5(e) and (21).

(b) Using $\hat{\varepsilon}_t = \varepsilon_t - \left(\hat{\pi}'_{f_1,f_2} - \pi'_{f_1,f_2}\right)(\mathbf{I}_n \otimes \mathbf{x}_t)$, it follows that

$$\frac{1}{T_w}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\hat{\varepsilon}_t\hat{\varepsilon}_t' = \frac{1}{T_w}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\varepsilon_t\varepsilon_t' - \frac{2}{T_w}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\varepsilon_t\left(\mathbf{I}\otimes\mathbf{x}_t\right)'\left(\hat{\pi}_{f_1,f_2} - \pi_{f_1,f_2}\right)$$
$$+\frac{1}{T_w}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\left(\hat{\pi}'_{f_1,f_2} - \pi'_{f_1,f_2}\right)\left(\mathbf{I}\otimes\mathbf{x}_t\mathbf{x}_t'\right)\left(\hat{\pi}_{f_1,f_2} - \pi_{f_1,f_2}\right) \to_{a.s.} \mathbf{\Omega}_{f_1,f_2},$$

since $T_w^{-1}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\varepsilon_t\varepsilon_t' \to^{a.s} \mathbf{\Omega}_{f_1,f_2}$ from Lemma 3.5(b), $\hat{\pi}_{f_1,f_2} \to_{a.s.} \pi_{f_1,f_2}, T^{-1}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\xi_t \to_{a.s.}$ $\mathbf{0}$, and $T_w^{-1}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\mathbf{x}_t\mathbf{x}_t' \to_{a.s.} \mathbf{Q}_{f_1,f_2} > 0$.

(c) Write the centred and scaled process $\sqrt{T_w}\left(\hat{\pi}_{f_1,f_2} - \pi_{f_1,f_2}\right)$ as

$$\left[\mathbf{I}_n \otimes \frac{1}{T_w}\sum_{t=[Tf_1]}^{[Tf_2]}\mathbf{x}_t\mathbf{x}_t'\right]^{-1}\left[\frac{\sqrt{T}}{\sqrt{T_w}}\frac{1}{\sqrt{T}}\sum_{t=[Tf_1]}^{[Tf_2]}\xi_t\right] \Rightarrow f_w^{-1/2}\mathbf{V}_{f_1,f_2}^{-1}\left[B^*(f_2) - B^*(f_1)\right],$$

whose finite dimensional distribution for fixed (f_1, f_2) is $\sqrt{T_w}\left(\hat{\pi}_{f_1,f_2} - \pi_{f_1,f_2}\right) \xrightarrow{L} N\left(0, \mathbf{V}_{f_1,f_2}^{-1}\mathbf{W}_{f_1,f_2}\mathbf{V}_{f_1,f_2}^{-1}\right)$, where $\mathbf{V}_{f_1,f_2} = \mathbf{I}_n \otimes \mathbf{Q}_{f_1,f_2}$.

(d) By definition,

$$\frac{1}{T_w}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\hat{\xi}_t\hat{\xi}_t' = \frac{1}{T_w}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\left(\hat{\varepsilon}_t\hat{\varepsilon}_t' \otimes \mathbf{x}_t\mathbf{x}_t'\right)$$
$$= \frac{1}{T_w}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\left[\varepsilon_t - \left(\hat{\pi}'_{f_1,f_2} - \pi'_{f_1,f_2}\right)(\mathbf{I}_n\otimes\mathbf{x}_t)\right]\left[\varepsilon_t - \left(\hat{\pi}'_{f_1,f_2} - \pi'_{f_1,f_2}\right)(\mathbf{I}_n\otimes\mathbf{x}_t)\right]' \otimes \mathbf{x}_t\mathbf{x}_t'$$
$$= \frac{1}{T_w}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\varepsilon_t\varepsilon_t' \otimes \mathbf{x}_t\mathbf{x}_t' - \frac{2}{T_w}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\left[\left(\varepsilon_t\mathbf{I}_n \otimes \varepsilon_t\mathbf{x}_t'\right)\left(\hat{\pi}_{f_1,f_2} - \pi_{f_1,f_2}\right) \otimes \mathbf{x}_t\mathbf{x}_t'\right]$$

$$+\frac{1}{T_w}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\left(\hat{\pi}'_{f_1,f_2}-\pi'_{f_1,f_2}\right)\left(\mathbf{I}\otimes\mathbf{x}_t\mathbf{x}_t'\right)\left(\hat{\pi}_{f_1,f_2}-\pi_{f_1,f_2}\right)\otimes\mathbf{x}_t\mathbf{x}_t'$$
$$=\frac{1}{T_w}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\xi_t\xi_t'+o_p(1)\,\mathbf{1}\mathbf{1}'\to_{a.s}\mathbf{W}_{f_1,f_2},$$

from Lemma 3.5(d) and (e), Lemma 3.7(a), and Lemma 3.6(b).

C.4 Proof of Proposition 4

In view of Lemma 3.7(c), under the null hypothesis

$$\begin{aligned}\sqrt{T_w}\mathbf{R}\hat{\pi}_{f_1,f_2}&\Rightarrow f_w^{-1/2}\mathbf{R}\mathbf{V}_{f_1,f_2}^{-1}\left[B^*(f_2)-B^*(f_1)\right]\\&=f_w^{-1/2}\mathbf{R}\mathbf{V}_{f_1,f_2}^{-1}\mathbf{W}_{f_1,f_2}^{1/2}\left[W_{nk}(f_2)-W_{nk}(f_1)\right],\end{aligned}$$

where $\mathbf{B}^*$ is vector Brownian motion with covariance matrix $\mathbf{W}_{f_1,f_2}$ and W_{nk} is vector standard Brownian motion with covariance matrix $\mathbf{I}_{nk}$. It follows that

$$\begin{aligned}Z_{f_2}^*(f_1)&:=\left[\mathbf{R}\left(\hat{\mathbf{V}}_{f_1,f_2}^{-1}\hat{\mathbf{W}}_{f_1,f_2}\hat{\mathbf{V}}_{f_1,f_2}^{-1}\right)\mathbf{R}'\right]^{-1/2}\left(\sqrt{T_w}\mathbf{R}\hat{\pi}_{f_1,f_2}\right)\\&\Rightarrow f_w^{-1/2}\left[\mathbf{R}\left(\mathbf{V}_{f_1,f_2}^{-1}\mathbf{W}_{f_1,f_2}\mathbf{V}_{f_1,f_2}^{-1}\right)\mathbf{R}'\right]^{-1/2}\mathbf{R}\mathbf{V}_{f_1,f_2}^{-1}\mathbf{W}_{f_1,f_2}^{1/2}\left[W_{nk}(f_2)-W_{nk}(f_1)\right].\end{aligned}$$

The Wald statistic process is

$$\begin{aligned}W_{f_2}^*(f_1)&=Z_{f_2}^*(f_1)'\,Z_{f_2}^*(f_1)\\&\Rightarrow f_w^{-1}\left[W_{nk}(f_2)-W_{nk}(f_1)\right]'\mathbf{A}_{f_1,f_2}\left(\mathbf{A}'_{f_1,f_2}\mathbf{A}_{f_1,f_2}\right)^{-1}\mathbf{A}'_{f_1,f_2}\left[W_{nk}(f_2)-W_{nk}(f_1)\right]\\&=^d f_w^{-1}\left[W_d(f_2)-W_d(f_1)\right]'\left[W_d(f_2)-W_d(f_1)\right],\end{aligned}$$

with $\mathbf{A}_{f_1,f_2}=\mathbf{W}_{f_1,f_2}^{1/2}\mathbf{V}_{f_1,f_2}^{-1}\mathbf{R}'$, whose finite dimensional distribution for fixed f_1 and f_2 is χ_d^2. It follows by continuous mapping that as $T\to\infty$

$$SW_{f_2}^*(f_0)\xrightarrow{L}\sup_{f_1\in[0,f_2-f_0],f_2=f}\left[\frac{W_d(f_w)'\,W_d(f_w)}{f_w}\right],$$

where W_d is vector standard Brownian motion with covariance matrix $\mathbf{I}_d$.

C.5 Proof of Proposition 5

In view of Lemma 3.7(c), under the null hypothesis the limit process is given by

$$\begin{aligned}\sqrt{T_w}\mathbf{R}\hat{\pi}_{f_1,f_2} &\Rightarrow f_w^{-1/2}\mathbf{R}\mathbf{V}_{f_1,f_2}^{-1}\left[B^*\left(f_2\right)-B^*\left(f_1\right)\right] \\ &= f_w^{-1/2}\mathbf{R}\mathbf{V}_{f_1,f_2}^{-1}\mathbf{W}_{f_1,f_2}^{1/2}\left[W_{nk}\left(f_2\right)-W_{nk}\left(f_1\right)\right],\end{aligned}$$

where $\mathbf{B}^*$ is vector Brownian motion with covariance matrix $\mathbf{W}_{f_1,f_2}$ and W_{nk} is vector standard Brownian motion with covariance matrix $\mathbf{I}_{nk}$. It follows that

$$\begin{aligned}Z_{f_2}\left(f_1\right) &:= \left[\mathbf{R}\left(\hat{\mathbf{\Omega}}_{f_1,f_2}\otimes\hat{\mathbf{Q}}_{f_1,f_2}\right)^{-1}\mathbf{R}'\right]^{-1/2}\left(\sqrt{T_w}\mathbf{R}\hat{\pi}_{f_1,f_2}\right) \\ &\Rightarrow f_w^{-1/2}\left[\mathbf{R}\left(\mathbf{\Omega}_{f_1,f_2}\otimes\mathbf{Q}_{f_1,f_2}\right)^{-1}\mathbf{R}'\right]^{-1/2}\mathbf{R}\mathbf{V}_{f_1,f_2}^{-1}\mathbf{W}_{f_1,f_2}^{1/2}\left[W_{nk}\left(f_2\right)-W_{nk}\left(f_1\right)\right].\end{aligned}$$

The Wald statistic process

$$\begin{aligned}W_{f_2}\left(f_1\right) &= Z_{f_2}\left(f_1\right)' Z_{f_2}\left(f_1\right) \\ &\Rightarrow f_w^{-1}\left[W_{nk}\left(f_2\right)-W_{nk}\left(f_1\right)\right]'\mathbf{A}_{f_1,f_2}\mathbf{B}_{f_1,f_2}^{-1}\mathbf{A}_{f_1,f_2}'\left[W_{nk}\left(f_2\right)-W_{nk}\left(f_1\right)\right],\end{aligned}$$

with $\mathbf{A}_{f_1,f_2} = \mathbf{W}_{f_1,f_2}^{1/2}\mathbf{V}_{f_1,f_2}^{-1}\mathbf{R}'$ and $\mathbf{B}_{f_1,f_2} = \mathbf{R}\left(\mathbf{\Omega}_{f_1,f_2}\otimes\mathbf{Q}_{f_1,f_2}\right)\mathbf{R}$. It follows by continuous mapping that as $T\to\infty$

$$SW_{f_2}\left(f_0\right)\xrightarrow{L}\sup_{f_1\in[0,f_2-f_0],f_2=f}\left\{\left[\frac{W_{nk}\left(f_2\right)-W_{nk}\left(f_1\right)}{f_w^{1/2}}\right]'\mathbf{A}_{f_1,f_2}\mathbf{B}_{f_1,f_2}^{-1}\mathbf{A}_{f_1,f_2}'\left[\frac{W_{nk}\left(f_2\right)-W_{nk}\left(f_1\right)}{f_w^{1/2}}\right]\right\}.$$