Online Supplement to: Change Detection and the Causal Impact of the Yield Curve*

Shuping Shi
Macquarie University

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

Stan Hurn
Queensland University of Technology

August 24, 2018

This document supplements the paper "Change Detection and the Causal Impact of the Yield Curve" by (i) providing proofs of the limit distributions of the subsample Wald and sup Wald statistics under homoskedasticity and conditional heteroskedasticity of unknown form, and (ii) reporting sensitivity analysis for the empirical application.

1 Limit Theory under Homoskedasticity

Lemma 3.1 and Proposition 3.1 are proved under Assumptions **A0** and **A2**. The proof for strictly stationary and ergodic sequences ε_t (Assumption **A1**) follows similar lines and is omitted.

*We thank the Editor and three referees for helpful comments which identified an error in the earlier version of the paper and led to many improvements. 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.

1.1 Proof of Lemma 3.1

(a) Let $\Lambda_0 = \{(f_1, f_2) : 1 \geq f_2 \geq f_1 + f_0 > 0 \text{ and } 0 \leq f_1 \leq 1 - f_0\}$ for some $f_0 > 0$. Set $T_w = \lfloor Tf_w \rfloor$ with $f_w = f_2 - f_1$ and $(f_1, f_2) \in \Lambda_0$ Write the estimation error as

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

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 $T_w^{-1} \sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \xi_t \rightarrow_{a.s.} \mathbf{0}$ by a standard martingale strong law. Define $\hat{\mathbf{Q}}_{f_1,f_2} = \frac{1}{T_w} \sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \mathbf{x}_t \mathbf{x}_t'$. Then, by a strong law for second order moments of linear processes (Phillips and Solo (1992), Theorem 3.7), we have $\hat{\mathbf{Q}}_{f_1,f_2} \rightarrow_{a.s.} \mathbf{Q} = \mathbb{E}(x_t x_t') > 0$ and then

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

so that $\hat{\pi}_{f_1,f_2} \rightarrow_{a.s.} \pi_{f_1,f_2} = \pi$ under the maintained null of constant coefficients. The convergence (1) also holds uniformly over $f_w = f_2 - f_1 \geq f_0$ by virtue of the martingale maximal inequality.

(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

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

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

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

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

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 $\delta > 0$

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

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,

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

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,

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

which ensures that the Lindeberg condition (2) holds. Then, by the martingale invariance principle for linear processes (Phillips and Solo, 1992, Theorems 3.4), for fixed $f_2 > f_1$ it follows that $T^{-1/2} \sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \xi_t \Rightarrow B(f_2) - B(f_1)$, where B is vector Brownian motion with covariance matrix $\mathbf{\Omega} \otimes \mathbf{Q}$.

The limit theory may be extended to allow for indexing by $(f_1, f_2) \in \Lambda_0$. We define the partial sum proces $X_T^0(r) = T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \xi_t \Rightarrow B(r)$ on $D[0,1]^{nk}$ equipped with the uniform topology, set $g_{f_i}\left(X_T^0\right) := X_T^0(f_i)$, and let $g_{f_1,f_2}\left(X_T^0\right) := f_w^{-1}\left(g_{f_2}\left(X_T^0\right) - g_{f_1}\left(X_T^0\right)\right)$. Since g_{f_i} is continuous for $i = 1, 2$ and $f_w \geq f_0 > 0$, so is the functional $g_{f_1,f_2}\left(X_T^0\right)$. Moreover, the functional $g_{f_1,f_2}(\cdot)$ may be interpreted as a map of the space $D[0,1]^{nk}$ onto a function that is defined on Λ_0. We may then write subsequent functionals of g_{f_1,f_2} as composed functionals of X_T^0. In particular, we may write the scaled estimation error $\sqrt{T}\left(\hat{\pi}_{f_1,f_2}-\pi_{f_1,f_2}\right)$ as

$$\begin{aligned}
\sqrt{T}\left(\hat{\pi}_{f_1,f_2}-\pi_{f_1,f_2}\right) &= \left[\mathbf{I}_n \otimes \hat{\mathbf{Q}}_{f_1,f_2}\right]^{-1}\left(\frac{T}{T_w}\frac{1}{\sqrt{T}}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\xi_t\right) \\
&= \left[\mathbf{I}_n \otimes \frac{1}{Tf_w}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\mathbf{x}_t\mathbf{x}_t'\right]^{-1}\left[g_{f_1,f_2}\left(X_T^0\right)+o_p(1)\right] \\
&= \left[\mathbf{I}_n \otimes \mathbf{Q}\right]^{-1} g_{f_1,f_2}\left(X_T^0\right)+o_p(1) \\
&=: k_{f_1,f_2}\left(X_T^0\right)+o_p(1), \qquad (3)
\end{aligned}$$

where $k_{f_1,f_2}\left(X_T^0\right)$, like $g_{f_1,f_2}\left(X_T^0\right)$, is a continuous linear functional of the process X_T^0 indexed by $(f_1,f_2)\in\Lambda_0$. By the continuous mapping theorem, it follows that $k_{f_1,f_2}\left(X_T^0\right)\Rightarrow k_{f_1,f_2}(B)$, giving

$$\begin{aligned}
\sqrt{T}\left(\hat{\pi}_{f_1,f_2}-\pi_{f_1,f_2}\right) &= \left[\mathbf{I}_n \otimes \hat{\mathbf{Q}}_{f_1,f_2}\right]^{-1}\left(\frac{1}{f_w}\frac{1}{\sqrt{T}}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\xi_t\right)\Rightarrow\left[\mathbf{I}_n \otimes \mathbf{Q}\right]^{-1} g_{f_1,f_2}(B) \\
&= k_{f_1,f_2}(B)=\left[\mathbf{I}_n \otimes \mathbf{Q}\right]^{-1}\left[\frac{B(f_2)-B(f_1)}{f_w}\right]. \qquad (4)
\end{aligned}$$

The limit in (4) 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 $N\left(\mathbf{0},\boldsymbol{\Omega}\otimes f_w^{-1}\mathbf{Q}^{-1}\right)$, so that $\sqrt{T}\left(\hat{\pi}_{f_1,f_2}-\pi_{f_1,f_2}\right)\Rightarrow N\left(\mathbf{0},\boldsymbol{\Omega}\otimes f_w^{-1}\mathbf{Q}^{-1}\right)$, as stated.

1.2 Proof of Proposition 3.1

In view of (4), under the null hypothesis $\mathbf{R}\pi_{f_1,f_2}=0$ we have

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

where W is standard Brownian motion with covariance matrix $\mathbf{I}_{nk}$ and $g_{f_1,f_2}(W)=\left(W(f_2)-W(f_1)\right)/f_w$. Setting $T_w=\lfloor Tf_w\rfloor$, it follows that

$$Z_{f_2}(f_1):=\left[\mathbf{R}\left(\hat{\boldsymbol{\Omega}}_{f_1,f_2}\otimes\left(\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\mathbf{x}_t\mathbf{x}_t'\right)^{-1}\right)\mathbf{R}'\right]^{-1/2}\mathbf{R}\hat{\pi}_{f_1,f_2}$$

$$= \left[\mathbf{R} \left(\hat{\boldsymbol{\Omega}}_{f_1,f_2} \otimes \left(\frac{Tf_w}{T} \frac{1}{T_w} \sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \mathbf{x}_t \mathbf{x}_t' \right)^{-1} \right) \mathbf{R}' \right]^{-1/2} \sqrt{T} \mathbf{R} \hat{\pi}_{f_1,f_2}$$
$$= f_w^{1/2} \left[\mathbf{R} \left(\boldsymbol{\Omega} \otimes \mathbf{Q}^{-1} \right) \mathbf{R}' \right]^{-1/2} \mathbf{R} \left[\mathbf{I}_n \otimes \mathbf{Q} \right]^{-1} g_{f_1,f_2} \left(X_T^0 \right) + o_p(1)$$
$$\Rightarrow f_w^{1/2} \left[\mathbf{R} \left(\boldsymbol{\Omega} \otimes \mathbf{Q}^{-1} \right) \mathbf{R}' \right]^{-1/2} \mathbf{R} \left[\mathbf{I}_n \otimes \mathbf{Q} \right]^{-1} g_{f_1,f_2} (B) .$$

The Wald statistic has the form

$$\begin{aligned}
\mathcal{W}_{f_2}(f_1) &= Z_{f_2}(f_1)' Z_{f_2}(f_1) \\
&= f_w g_{f_1,f_2} \left(X_T^0 \right)' \left[\mathbf{I}_n \otimes \mathbf{Q} \right]^{-1\prime} \mathbf{R}' \left[\mathbf{R} \left(\boldsymbol{\Omega} \otimes \mathbf{Q}^{-1} \right) \mathbf{R}' \right]^{-1} \mathbf{R} \left[\mathbf{I}_n \otimes \mathbf{Q} \right]^{-1} g_{f_1,f_2} \left(X_T^0 \right) + o_p(1) \\
&= f_w g_{f_1,f_2} \left(X_T^0 \right)' \left[\boldsymbol{\Omega}^{-1/2} \otimes \mathbf{Q}^{-1/2} \right] \left[\boldsymbol{\Omega}^{1/2} \otimes \mathbf{Q}^{-1/2} \right] \mathbf{R}' \left[\mathbf{R} \left(\boldsymbol{\Omega} \otimes \mathbf{Q}^{-1} \right) \mathbf{R}' \right]^{-1} \mathbf{R} \left[\boldsymbol{\Omega}^{1/2} \otimes \mathbf{Q}^{-1/2} \right] \\
&\quad \times \left[\boldsymbol{\Omega}^{-1/2} \otimes \mathbf{Q}^{-1/2} \right] g_{f_1,f_2} \left(X_T^0 \right) + o_p(1) \\
&= f_w g_{f_1,f_2} \left(X_T^0 \right)' \left[\boldsymbol{\Omega}^{-1/2} \otimes \mathbf{Q}^{-1/2} \right] \mathbf{A} \left(\mathbf{A}'\mathbf{A} \right)^{-1} \mathbf{A}' \left[\boldsymbol{\Omega}^{-1/2} \otimes \mathbf{Q}^{-1/2} \right] g_{f_1,f_2} \left(X_T^0 \right) + o_p(1) \\
&=: h \left(X_T^0 \right) + o_p(1) ,
\end{aligned}$$

with the $nk \times d$ matrix $\mathbf{A} = \left[\boldsymbol{\Omega}^{1/2} \otimes \mathbf{Q}^{-1/2} \right] \mathbf{R}'$. Now for $(f_1, f_2) \in \Lambda_0$, we have $h_{f_1,f_2} \left(X_T^0 \right) \Rightarrow h_{f_1,f_2}(B)$ where

$$\begin{aligned}
h_{f_1,f_2}(B) &= f_w g_{f_1,f_2}(B)' \left[\boldsymbol{\Omega}^{-1/2} \otimes \mathbf{Q}^{-1/2} \right] \mathbf{A} \left(\mathbf{A}'\mathbf{A} \right)^{-1} \mathbf{A}' \left[\boldsymbol{\Omega}^{-1/2} \otimes \mathbf{Q}^{-1/2} \right] g_{f_1,f_2}(B) \\
&= \left(\frac{W(f_2) - W(f_1)}{f_w^{1/2}} \right)' \mathbf{A} \left(\mathbf{A}'\mathbf{A} \right)^{-1} \mathbf{A}' \left(\frac{W(f_2) - W(f_1)}{f_w^{1/2}} \right) .
\end{aligned}$$

with the $nk \times d$ matrix $\mathbf{A} = \left[\boldsymbol{\Omega}^{1/2} \otimes \mathbf{Q}^{-1/2} \right] \mathbf{R}'$.

Next, we define the sup functional

$$h^* \left(h_{f_1,f_2} \left(X_T^0 \right) \right) = \sup_{f_w = f_2 - f_1 \geq f_0} h_{f_1,f_2} \left(X_T^0 \right) ,$$

which maps functions defined on Λ_0 onto $\mathbb{R}$. By virtue of the continuity of the map h^*, the continuity of h_{f_1,f_2}, and the weak convergence of the process $X_T^0(r) = T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \xi_t \Rightarrow B(r)$ on $D[0,1]^{nk}$, the continuous mapping theorem gives the weak convergence

$$h^* \left(h_{f_1,f_2} \left(X_T^0 \right) \right) \Rightarrow \sup_{f_w = f_2 - f_1 \geq f_0} h_{f_1,f_2}(B)$$

$$= \sup_{f_w = f_2 - f_1 \geq f_0} \left(\frac{W(f_2) - W(f_1)}{f_w^{1/2}} \right)' \mathbf{A} \left(\mathbf{A}'\mathbf{A} \right)^{-1} \mathbf{A}' \left(\frac{W(f_2) - W(f_1)}{f_w^{1/2}} \right).$$

Finally, we note that $\mathcal{W}_{f_2}(f_1) = h_{f_1,f_2}\left(X_T^0\right) + o_p(1)$ by virtue of the convergences $\hat{\mathbf{\Omega}}_{f_1,f_2} \to_p \mathbf{\Omega}$ and $T_w^{-1} \sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \mathbf{x}_t \mathbf{x}_t' \to_p \mathbf{Q}$, which hold uniformly over $(f_1, f_2) \in \Lambda_0$, so that

$$\begin{aligned}
\sup_{f_w = f_2 - f_1 \geq f_0} \mathcal{W}_{f_2}(f_1) &= h^*\left(h_{f_1,f_2}\left(X_T^0\right)\right) + o_p(1) = \sup_{f_w = f_2 - f_1 \in [f_0, f_2]} h_{f_1,f_2}\left(X_T^0\right) + o_p(1) \\
&\Rightarrow \sup_{f_w = f_2 - f_1 \geq f_0} \left(\frac{W(f_2) - W(f_1)}{f_w^{1/2}} \right)' \mathbf{A} \left(\mathbf{A}'\mathbf{A} \right)^{-1} \mathbf{A}' \left(\frac{W(f_2) - W(f_1)}{f_w^{1/2}} \right) \\
&=^d \sup_{f_w = f_2 - f_1 \geq f_0} \left(\frac{W_d(f_2) - W_d(f_1)}{f_w^{1/2}} \right)' \left(\frac{W_d(f_2) - W_d(f_1)}{f_w^{1/2}} \right),
\end{aligned}$$

where $=_d$ denotes equivalence in distribution and W_d is standard Brownian motion with covariance matrix $\mathbf{I}_d$ and d is the dimension of the restriction matrix $\mathbf{R}$. Thus, the sup Wald statistic satisfies the stated limit theory

$$\mathcal{SW}_f(f_0) \Rightarrow \sup_{(f_1,f_2) \in \Lambda_0; f_2 = f} \left[\frac{W_d(f_2) - W_d(f_1)}{f_w^{1/2}} \right]' \left[\frac{W_d(f_2) - W_d(f_1)}{f_w^{1/2}} \right],$$

by virtue of continuous mapping and the weak convergence $X_T^0(\cdot) \Rightarrow B(\cdot)$.

2 Limit Theory under Conditional Heteroskedasticity of Unknown Form

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

2.1 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 $h \geq 0, z > 0$

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

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

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

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

For (d), by construction

$$\frac{1}{T_w}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor} \mathbf{x}_{t-1}\varepsilon_t' = \frac{1}{T_w}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor} \begin{bmatrix} \varepsilon_t & \varepsilon_t\mathbf{y}_{t-1}' & \cdots & \varepsilon_t\mathbf{y}_{t-p}' \end{bmatrix}'.$$

and, from (a), $T_w^{-1}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor} \varepsilon_t \rightarrow_{a.s} \mathbf{0}$. 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 $T_w^{-1}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor} \mathbf{y}_{t-h}\varepsilon_t' \rightarrow_{a.s} \mathbf{0}$, giving the required $T_w^{-1}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor} \mathbf{x}_{t-1}\varepsilon_t' \rightarrow_{a.s} \mathbf{0}$.

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

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

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

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

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

by results in (a), and

$$\frac{1}{T_w}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\left(\mathbf{y}_{t-h} - \tilde{\Phi}_0\right)\left(\mathbf{y}_{t-h-j} - \tilde{\Phi}_0\right)' = \frac{1}{T_w}\sum_{t=\lfloor Tf_1\rfloor}^{\lfloor Tf_2\rfloor}\left(\sum_{i=0}^{\infty}\Psi_i\varepsilon_{t-h-i}\right)\left(\sum_{i=0}^{\infty}\Psi_i\varepsilon_{t-h-j-i}\right)'$$

$$\rightarrow_{a.s} \sum_{i=0}^{\infty} \Psi_{i+j} \mathbf{\Omega} \Psi_i',$$

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

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

giving

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

with

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

2.2 Proof of Lemma 3.3

(a) We show that the following conditional Lindeberg condition holds for all $\delta > 0$:

$$\frac{1}{T} \sum_{t=1}^{T} \mathbb{E} \left[\|\xi_t\|^2 \mathbf{1} \left(\|\xi_t\| \geq \sqrt{T}\delta \right) | \mathcal{F}_{t-1} \right] \xrightarrow{p} 0. \tag{5}$$

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,

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

for some constant $K < \infty$ as $T \rightarrow \infty$ 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 (5) holds by L_1 convergence.

(b) The stability condition involves the convergences

$$\frac{1}{T}\sum_{t=1}^{T}\xi_t\xi_t', \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left\{\xi_t\xi_t'|\mathcal{F}_{t-1}\right\} \rightarrow_{a.s} \mathbf{\Sigma}. \tag{6}$$

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

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left(\xi_t\xi_t'|\mathcal{F}_{t-1}\right) = \mathbf{\Sigma}, \tag{7}$$

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 $T^{-1}\sum_{t=1}^{T}\left\{\varepsilon_{1,t}^2 - \mathbb{E}\left(\varepsilon_{1,t}^2|\mathcal{F}_{t-1}\right)\right\} \rightarrow_{a.s} 0$ and from Lemma 3.2(b) $T^{-1}\sum_{t=1}^{T}\varepsilon_{1,t}^2 \rightarrow_{a.s} \Omega_{11}$, with $T^{-1}\sum_{t=1}^{T}\mathbb{E}\left(\varepsilon_{1,t}^2|\mathcal{F}_{t-1}\right) \rightarrow_{a.s} \Omega_{11}$ 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

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

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

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

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

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\varepsilon_{1,t}^{2}\left(\mathbf{y}_{t-h}-\tilde{\Phi}_0\right)\left(\mathbf{y}_{t-h-j}-\tilde{\Phi}_0\right)'|\mathcal{F}_{t-1}\right]\\
&=\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left(\varepsilon_{1,t}^{2}|\mathcal{F}_{t-1}\right)\left(\sum_{i=0}^{\infty}\Psi_i\varepsilon_{t-h-i}\right)\left(\sum_{i=0}^{\infty}\Psi_i\varepsilon_{t-h-j-i}\right)\\
&=\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left(\varepsilon_{1,t}^{2}|\mathcal{F}_{t-1}\right)\sum_{i=0}^{\infty}\Psi_{i+j}\varepsilon_{t-h-j-i}\varepsilon_{t-h-j-i}'\Psi_i'+o_p(1)\times\mathbf{11}'\\
&=\sum_{i=0}^{\infty}\Psi_{i+j}\left[\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left(\varepsilon_{1,t}^{2}|\mathcal{F}_{t-1}\right)\varepsilon_{t-h-j-i}\varepsilon_{t-h-j-i}'\right]\Psi_i'+o_p(1)\times\mathbf{11}'\\
&\rightarrow_{a.s}\sum_{i=0}^{\infty}\Psi_{i+j}\Gamma_{h+j+i}^{(1,1)}\Psi_i',
\end{aligned}$$

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

$$\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\varepsilon_{1,t}^{2}\mathbf{y}_{t-h}\mathbf{y}_{t-h-j}'|\mathcal{F}_{t-1}\right]\rightarrow_{a.s}\left[\Omega_{11}\tilde{\Phi}_0\tilde{\Phi}_0'+\sum_{i=0}^{\infty}\Psi_{i+j}\Gamma_{h+j+i}^{(1,1)}\Psi_i'\right].$$

Therefore

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left(\varepsilon_{1,t}^{2}\mathbf{x}_t\mathbf{x}_t'|\mathcal{F}_{t-1}\right)\\
&\rightarrow_{a.s}\begin{bmatrix}
\Omega_{11} & \Omega_{11}\tilde{\Phi}_0' & \cdots & \Omega_{11}\tilde{\Phi}_0'\\
\Omega_{11}\tilde{\Phi}_0 & \Omega_{11}\tilde{\Phi}_0\tilde{\Phi}_0'+\sum_{i=0}^{\infty}\Psi_i\Gamma_{h+j+i}^{(1,1)}\Psi_i' & \cdots & \Omega_{11}\tilde{\Phi}_0\tilde{\Phi}_0'+\sum_{i=0}^{\infty}\Psi_{i+p-1}\Gamma_{h+j+i}^{(1,1)}\Psi_i'\\
\vdots & \vdots & \ddots & \vdots\\
\Omega_{11}\tilde{\Phi}_0 & \Omega_{11}\tilde{\Phi}_0\tilde{\Phi}_0'+\sum_{i=0}^{\infty}\Psi_i\Gamma_{h+j+i}^{(1,1)}\Psi_{i+p-1}' & \cdots & \Omega_{11}\tilde{\Phi}_0\tilde{\Phi}_0'+\sum_{i=0}^{\infty}\Psi_i\Gamma_{h+j+i}^{(1,1)}\Psi_i'
\end{bmatrix},
\end{aligned}$$

with similar calculations for the other components of the matrix partition, leading to the stability condition (7), with $\mathbf{\Sigma}=\left\{\mathbf{\Sigma}^{(i,j)}\right\}_{i,j\in[1,n]}$ defined in terms of the component matrix partitions

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

and

$$\mathbf{\Xi}^{(i,j)} \equiv \sum_{i=0}^{\infty} \left[\begin{array}{ccc} \Psi_i \Gamma_{h+j+i}^{(i,j)} \Psi_i' & \cdots & \Psi_{i+p-1} \Gamma_{h+j+i}^{(i,j)} \Psi_i' \\ \vdots & \ddots & \vdots \\ \Psi_i \Gamma_{h+j+i}^{(i,j)} \Psi_{i+p-1}' & \cdots & \Psi_i \Gamma_{h+j+i}^{(i,j)} \Psi_i' \end{array} \right].$$

2.3 Proof of Lemma 3.4

(a) By definition and using Lemma 3.2 (iv) and (v), we have

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

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

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

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

(c) We follow a similar composite functional argument as that used in Lemma 3.1(c). In particular, the scaled and centred estimation error process can be written in the following form

$$\begin{aligned} \sqrt{T_w} \left(\hat{\pi}_{f_1,f_2} - \pi_{f_1,f_2} \right) &= \left[\mathbf{I}_n \otimes \frac{1}{T_w} \sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \mathbf{x}_t \mathbf{x}_t' \right]^{-1} \left[\frac{\sqrt{T}}{\sqrt{T_w}} \frac{1}{\sqrt{T}} \sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \xi_t \right] \\ = & \left[\mathbf{I}_n \otimes \frac{1}{Tf_w} \sum_{t=\lfloor Tf_1 \rfloor}^{\lfloor Tf_2 \rfloor} \mathbf{x}_t \mathbf{x}_t' \right]^{-1} g_{f_1,f_2} \left(X_T^0 \right) = \left[\mathbf{I}_n \otimes \mathbf{Q} \right]^{-1} g_{f_1,f_2} \left(X_T^0 \right) + o_p(1) \\ = & \; k_{f_1,f_2} \left(X_T^0 \right) + o_p(1) \\ \Rightarrow & \; k_{f_1,f_2} (B) = \left[\mathbf{I}_n \otimes \mathbf{Q} \right]^{-1} \left[\frac{B(f_2) - B(f_1)}{f_w^{1/2}} \right], \end{aligned}$$

where $g_{f_1,f_2}\left(X_T^0\right)$ and $k_{f_1,f_2}\left(X_T^0\right)$ are continuous linear functionals of the process X_T^0 indexed by $(f_1, f_2) \in \Lambda_0$ with the property that $k_{f_1,f_2}\left(X_T^0\right) \Rightarrow k_{f_1,f_2}(B)$, where B is vector Brownian motion with covariance matrix $\mathbf{\Sigma}$. The finite dimensional distribution for fixed (f_1, f_2) is $\sqrt{T_w}\left(\hat{\pi}_{f_1,f_2} - \pi_{f_1,f_2}\right) \xrightarrow{L} N\left(0, \mathbf{V}^{-1}\mathbf{\Sigma}\mathbf{V}^{-1}\right)$, where $\mathbf{V} = \mathbf{I}_n \otimes \mathbf{Q}$.

(d) By definition

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

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

2.4 Proof of Proposition 3.3 Under the Assumption of Conditional Heteroskedasticity

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

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

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

$$Z_{f_2}^*(f_1) := \left[\mathbf{R}\left(\hat{\mathbf{V}}_{f_1,f_2}^{-1}\hat{\mathbf{\Sigma}}_{f_1,f_2}\hat{\mathbf{V}}_{f_1,f_2}^{-1}\right)\mathbf{R}'\right]^{-1/2}\left(\sqrt{T_w}\mathbf{R}\hat{\pi}_{f_1,f_2}\right) \tag{8}$$
$$\Rightarrow f_w^{-1/2}\left[\mathbf{R}\left(\mathbf{V}^{-1}\mathbf{\Sigma}\mathbf{V}^{-1}\right)\mathbf{R}'\right]^{-1/2}\mathbf{R}\mathbf{V}^{-1}\mathbf{\Sigma}^{1/2}\left[W(f_2) - W(f_1)\right].$$

Observe that the Wald statistic process

$$\mathcal{W}_{f_2}^*(f_1) = Z_{f_2}^*(f_1)'\, Z_{f_2}^*(f_1)$$

$$\Rightarrow f_w^{-1}\left[W\left(f_2\right)-W\left(f_1\right)\right]'\mathbf{A}\left(\mathbf{A}'\mathbf{A}\right)^{-1}\mathbf{A}'\left[W\left(f_2\right)-W\left(f_1\right)\right]$$
$$=^d f_w^{-1}\left[W_d\left(f_2\right)-W_d\left(f_1\right)\right]'\left[W_d\left(f_2\right)-W_d\left(f_1\right)\right],$$

with $\mathbf{A}=\mathbf{\Sigma}^{1/2}\mathbf{V}^{-1}\mathbf{R}'$, whose finite dimensional distribution for fixed (f_1,f_2) is χ_d^2, and where W_d is vector Brownian motion with covariance matrix $\mathbf{I}_d$, as earlier. It follows by continuous mapping that as $T\to\infty$

$$\mathcal{SW}_{f_2}^{*}\left(f_0\right)\Rightarrow\sup_{(f_1,f_2)\in\Lambda_0;,f_2=f}\left[\frac{W_d\left(f_2\right)-W_d\left(f_1\right)}{f_w^{1/2}}\right]'\left[\frac{W_d\left(f_2\right)-W_d\left(f_1\right)}{f_w^{1/2}}\right]$$

This completes the proof of Proposition 3.3 under the assumptions **A0** and **A3**.

The proof of Proposition 3.2 under the assumptions **A0** and **A3** (without the use of a heteroskedastic consistent statistic) follows in the same manner, with the quantity $\hat{\mathbf{V}}_{f_1,f_2}^{-1}\hat{\mathbf{\Sigma}}_{f_1,f_2}\hat{\mathbf{V}}_{f_1,f_2}^{-1}$ in (8) replaced by the quantity $\left(\hat{\mathbf{\Omega}}_{f_1,f_2}\otimes\hat{\mathbf{Q}}_{f_1,f_2}\right)^{-1}$. The details are omitted.

3 Appendix: Robustness Checks

3.1 Minimum Window Size and Critical Values

We conduct sensitivity analysis to check the robustness of the heteroskedastic consistent results to the selection of the minimum window size and critical values. Figure 1 is for the causal relationship running from the yield curve slope to the growth rate of industrial production and Figure 2 is for causality running from the growth rate of industrial production to the yield curve slope. The minimum window size is $f_0 = 0.25$ (instead of $f_0 = 0.20$) in the first column (i.e., (a), (c), and (e)), In the second column (i.e., (b), (d) and (f)), we control the empirical sizes over a two-year (instead of three-year) period.

As evident in the graphs, when the minimum window size increases to 0.25, all three procedures do not detect any episode of causality from the yield curve spread to industrial production (Figure 1). This is in contrast to the findings in the main text with $f_0 = 0.2$, where the recursive evolving procedure finds evidence of causality in the late 1990s and after the subprime mortgage crisis in 2009. For causality running from the growth rate of industrial production to the output gap, the recursive evolving procedure detects an additional episode in 1995-1998. This is consistent with our expectation that the optimal minimum window size is episode specific. As the minimum window increases, we detect less episodes of causality in the former and more in the latter. When we control size over a two-year period, the overall pattern of the test results remains, although there are some small discrepancies in the exact start and end dates of the first episode.

3.2 Output Gap

Next, we use the output gap as a proxy for real economic activity. The output gap is calculated using the official Congressional Budget Office (CBO) measure of real potential output[1] (billions of chained 2009 dollars, not seasonally adjusted) and GDP (billions of chained 2009 dollars, seasonally adjusted annual rate) data. Inflation is measured from the core consumer price index and calculated as log differences (multiplied by 400). Data are downloaded from the Federal Reserve Bank of St. Louis FRED at the quarterly frequency. The data start from 1980 to the first quarter of 2015 ($T = 141$).

Figure 3 plots quaterly data on the output gap (left axis) and inflation (right axis). The two measures of real economic activity, namely real GDP and industrial production, have an important feature in common because both tend to fall sharply during recessions. There are, however, several noticeable differences in these measures. Industrial production, being a higher frequency monthly series, shows more evidence of heteroskedasticity. Also note that after the 2008-2009 recession, the growth rate of industrial production rebounds quite quickly and is relatively stable until the end of the sample. By contrast the output gap is more persistent, with actual output only narrowing the gap to potential output quite slowly. The quarterly federal funds rate and the slope of the yield curve show similar dynamic pattern as in Figure 1b and therefore omitted.

We calculate the heteroskedastic-consistent tests of Granger causality. In estimating the VAR and implementing tests of Granger causality, the lag order is assumed the same for all subsamples and selected using the Bayesian information criteria (BIC) for the whole sample

[1] Real potential GDP is the CBO's estimate of the output the economy would produce with a high rate of utilization of its capital and labor resources.

period with a maximum potential lag length 12. The selected lag order is two. The minimum window size is $f_0 = 0.2$, containing 28 observations. The critical values are obtained from bootstrapping with 499 replications. The empirical size is 5% and is controlled over a three-year period.

The forward procedure does not detect any period of causality in both directions. For potential causality running from the yield curve slope to output gap, both the rolling and recursive evolving algorithms identify one episode over the sample period, i.e. 1998Q3. In testing for reverse causality, the rolling procedure finds no evidence of causality over the entire sample period, whereas the recursive evolving algorithm suggest the existence of causal effects over the period 1991 - 1995 and in 2000. The first episode runs from 1991:Q1 to 1995:Q1 (with a break in 1991:Q3) and the second episode starts from the third quarter of 2000 and terminates at the end of the year. The change in causality in 2000 is also consistent with related research based on data around that time period to the effect that macroeconomic factors are important determinants of movements in bond yields primarily at the shorter maturities (Ang and Piazzesi, 2003). The differences in the behaviour of the output gap and the growth rate of industrial production have already been discussed, so the disparity in the conclusions between quarterly and monthly measures of economic activity is perhaps expected.

References

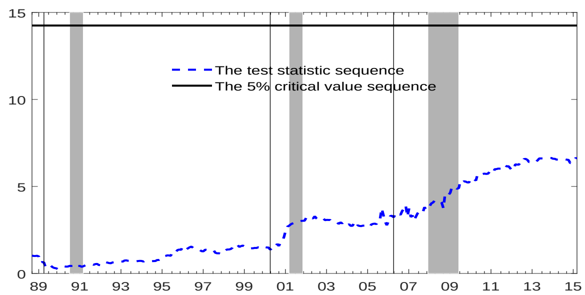
Ang, A. and M. Piazzesi (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics 50*(4), 745–787.

Hall, P. and C. Heyde (1980). Martingale limit theory and its applications.

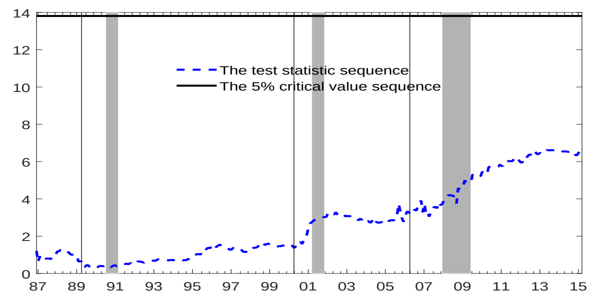
Phillips, P. C. B. and V. Solo (1992). Asymptotics for linear processes. *The Annals of Statistics 20*(2), 971–1001.

Figure 1: The heteroskedastic-consistent tests for Granger causality running from the yield curve slope to the industrial production. Tests are obtained from a VAR model with a minimum window size of $f_0 = 0.25$ in the first column and $f_0 = 0.20$ in the second column. The empirical size is controlled over a three-year period in the first column and over a two-year period in the second column. The lag order is 3 as for analysis in the main text.

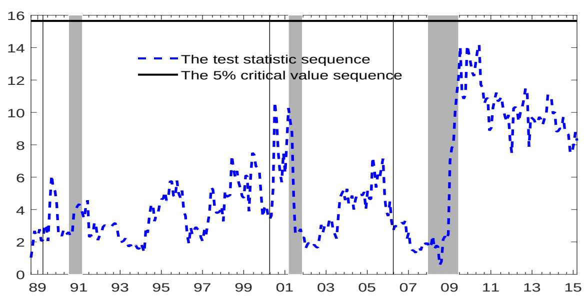
(a) Forward - $f_0 = 0.25$



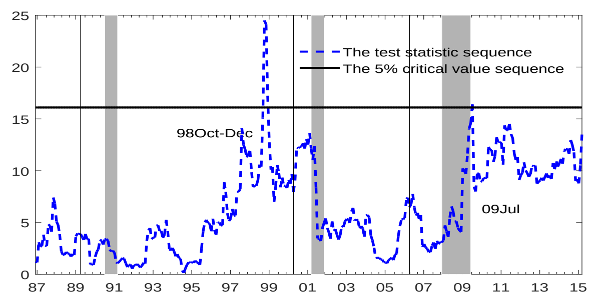
(b) Forward - Size (two years)



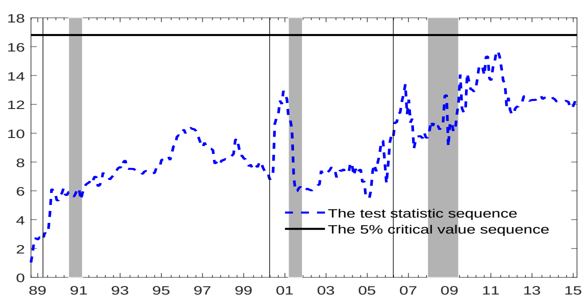
(c) Rolling - $f_0 = 0.25$



(d) Rolling - Size (two years)



(e) Recursive Rolling - $f_0 = 0.25$



(f) Recursive Rolling - Size (two years)

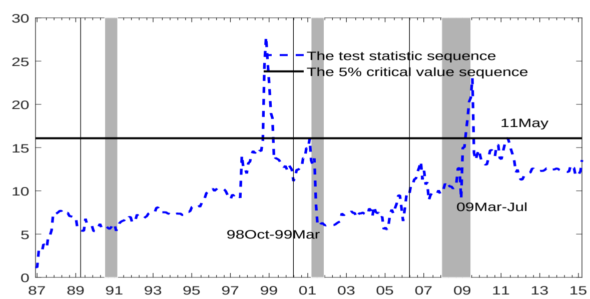
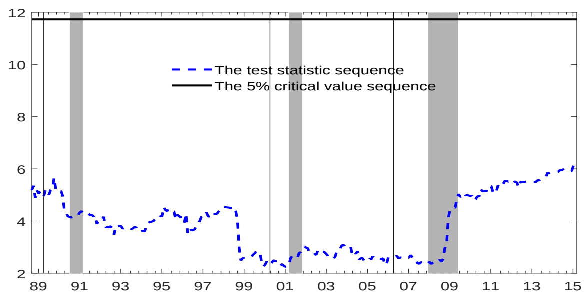
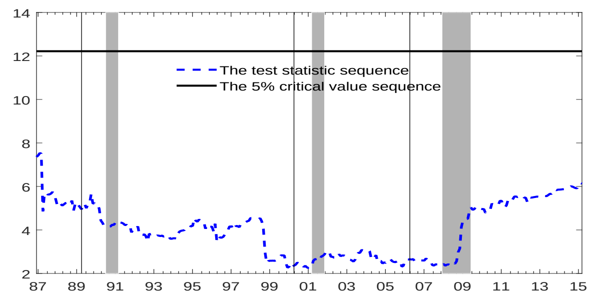


Figure 2: The heteroskedastic-consistent tests for Granger causality running from the industrial production to the yield curve slope. Tests are obtained from a VAR model with a minimum window size of $f_0 = 0.25$ in the first column and $f_0 = 0.20$ in the second column. The empirical size is controlled over a three-year period in the first column and over a two-year period in the second column. The lag order is 3 as for analysis in the main text.

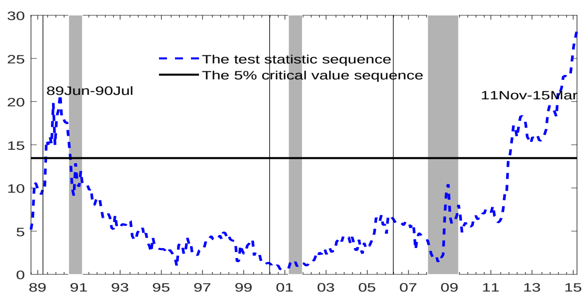
(a) Forward - $f_0 = 0.25$



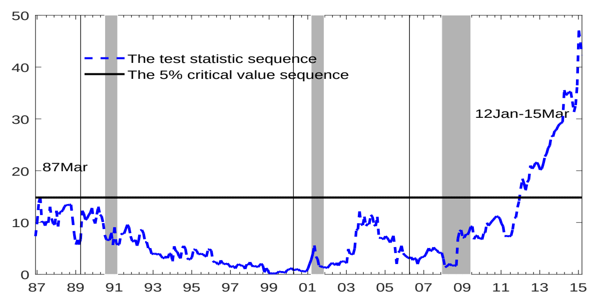
(b) Forward - Size (two years)



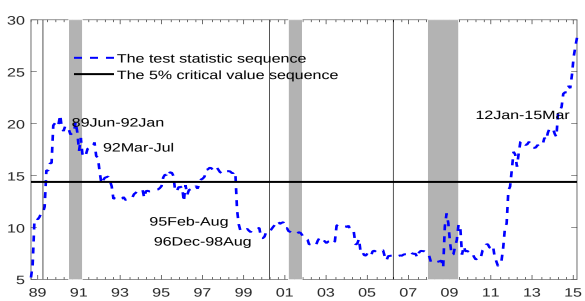
(c) Rolling - $f_0 = 0.25$



(d) Rolling - Size (two years)



(e) Recursive Rolling - $f_0 = 0.25$



(f) Recursive Rolling - Size (two years)

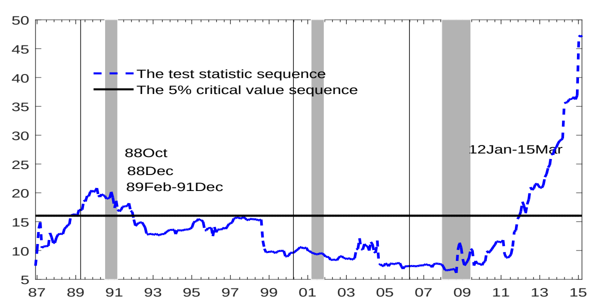


Figure 3: Time series plots of the output gap and inflation in the United States. Also shown are official NBER recession periods shaded in grey, namely, 1980:M01-M07,1981:M07-1982:M11,1990:M07-1991:M03, 2001:M03-M11 and 2007:M12-2009:M06.

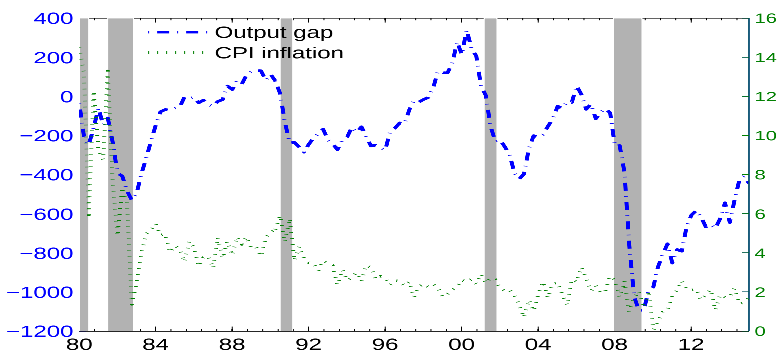
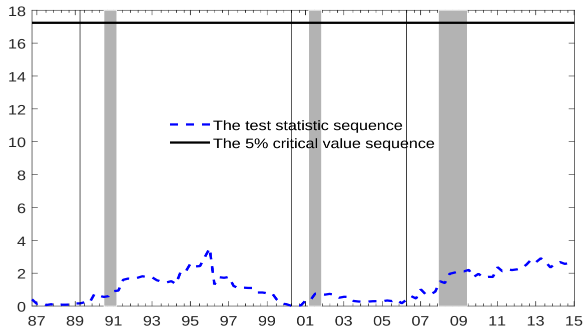
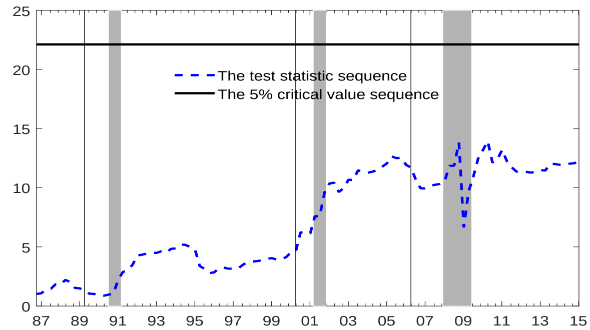


Figure 4: The heteroskedastic-consistent tests for Granger causality running from the yield curve slope to output gap in the first column (i.e., (a),(c), and (e)) and from output gap to the yield curve slope in the second column (i.e., (b), (d) and (f)). The shaded areas are the NBER recession periods, the vertical lines are the dates of the onset of an inverted yield curve and causal periods are shown in text.

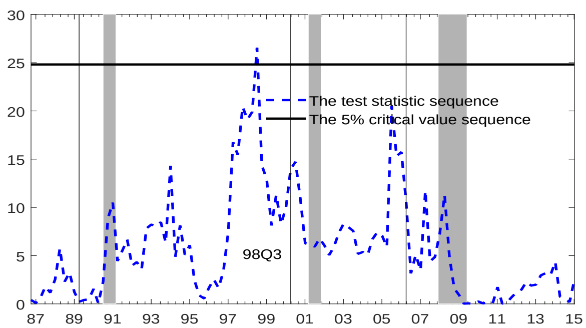
(a) Forward - from the yield curve slope to output gap



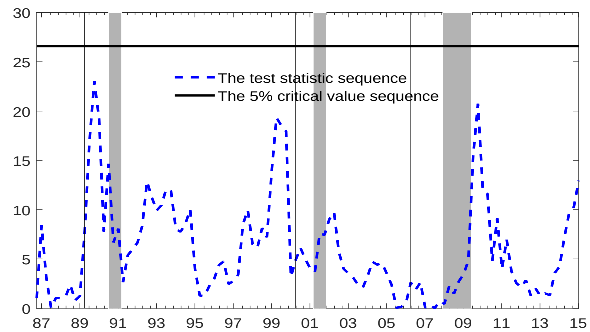
(b) Forward - from output gap to the yield curve slope



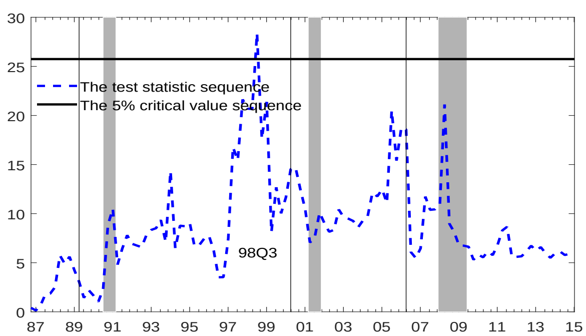
(c) Rolling - from the yield curve slope to output gap



(d) Rolling - from output gap to the yield curve slope



(e) Recursive Evolving - from the yield curve slope to output gap



(f) Recursive Evolving - from output gap to the yield curve slope