Supplement to "Testing for fundamental vector moving average representations"

BIN CHEN
Department of Economics, University of Rochester

JINHO CHOI
AMRO and Bank of Korea

JUAN CARLOS ESCANCIANO
Department of Economics, Indiana University

S1. ADDITIONAL RESULTS IN MONTE CARLO EVIDENCE

S1.1 *Lag order distribution by criterion using Kilian (2001)*

As a route to determine the VAR lag length in our simulation study, we follow Kilian (2001) to obtain the finite-sample distribution of the lag order estimates for each lag order selection criterion: the Schwarz information criterion (SIC), the Hannan–Quinn criterion (HQC), and the Akaike information criterion (AIC). We run two sets of simulations where the maximum lag $p^{\max}$ is set to four or eight. Table S1 summarizes the percentage distribution of lag order estimates using the simulated data for the invertible models DGP1 and DPG2 in the main text, with the sample size of 500 generated from 1000 Monte Carlo replications. We find that each of the information-based criteria lends strong support to the first-order VAR (i.e., $p = 1$).

S1.2 *Simulation results using additional DGPs*

This subsection reports the finite-sample performances of $\hat{Q}^1(h)$ using the DGPs with non-Gaussian, conditional heteroskedastic errors as well as Gaussian errors. Given the foresight models, first, under the null hypothesis, we generate the bivariate, invertible MA representation with no foresight (5.3) using the following DGPs.

DGP-S1. No foresight model with iid shocks $\{\varepsilon_{\tau,t}, \varepsilon_{A,t}\}$, mutually independent, and distributed as Gaussian processes; in short, $\boldsymbol{\varepsilon}_t \sim$ iid $N(0, \mathbf{I}_2)$.

Bin Chen: binchen@rochester.edu
Jinho Choi: choi.jinho@amro-asia.org
Juan Carlos Escanciano: jescanci@indiana.edu

TABLE S1. Percentage distribution of lag order estimates by criterion.

		Number of Lags ($1 \leq p \leq p^{\max}$)							
		1	2	3	4	5	6	7	8
		Invertible, Student's $t(3)$ distribution (DGP1)							
$p^{\max} = 8$	SIC	**98.3**	1.1	0.3	0.1	0.1	0.1	0.0	0.0
	HQC	**94.4**	3.2	0.9	0.7	0.3	0.4	0.1	0.0
	AIC	**78.1**	9.4	4.9	2.1	2.1	1.3	0.9	1.2
$p^{\max} = 4$	SIC	**97.8**	1.1	0.7	0.4	–	–	–	–
	HQC	**94.8**	3.1	1.4	0.7	–	–	–	–
	AIC	**83.3**	9.3	5.1	2.3	–	–	–	–
		Invertible, $\chi^2(3)$ distribution (DGP2)							
$p^{\max} = 8$	SIC	**99.1**	0.9	0.0	0.0	0.0	0.0	0.0	0.0
	HQC	**96.2**	3.1	0.6	0.1	0.0	0.0	0.0	0.0
	AIC	**79.4**	9.0	4.6	3.1	1.8	0.6	1.1	0.4
$p^{\max} = 4$	SIC	**98.8**	1.2	0.0	0.0	–	–	–	–
	HQC	**96.8**	2.7	0.4	0.1	–	–	–	–
	AIC	**81.6**	11.9	4.1	2.4	–	–	–	–

Note: Note that $\alpha = 0.4$, $\beta = 0.99$, $\tau = 0.25$, $T = 500$. The boldface type indicate lag orders preferred by SIC, HQC, or AIC.

DGP-S2. No foresight model with iid shocks $\{\varepsilon_{\tau,t}, \varepsilon_{A,t}\}$, mutually independent, and distributed as generalized autoregressive conditional heteroskedastic (GARCH) $(1, 1)$ Student's $t(3)$ variable

$$\boldsymbol{\varepsilon}_t = \boldsymbol{\sigma}_t \mathbf{v}_t,$$

where $\mathbf{v}_t$ follows independent Student's t-distributions with 3 degrees of freedom, and $\boldsymbol{\sigma}_t \boldsymbol{\sigma}_t' = \mathbf{C} + \mathbf{A} \odot (\boldsymbol{\varepsilon}_{t-1} \boldsymbol{\varepsilon}_{t-1}') + \mathbf{G} \odot (\boldsymbol{\sigma}_{t-1} \boldsymbol{\sigma}_{t-1}')$, with $\mathbf{C} = \text{diag}(1, 1)$ and $\mathbf{A} = \mathbf{G} = \text{diag}(0.3, 0.3)$, which is called diagonal vectorized GARCH as in Bollerslev, Engle, and Wooldridge (1988); in short, $\boldsymbol{\varepsilon}_t \sim$ iid GARCH$(1, 1)$-$t(3)$.

DGP-S3. No foresight model with iid shocks $\{\varepsilon_{\tau,t}, \varepsilon_{A,t}\}$, mutually independent, and distributed as GARCH$(1, 1)$-standardized $\chi^2(3)$ innovations $\boldsymbol{\varepsilon}_t = \boldsymbol{\sigma}_t \mathbf{v}_t$, where $\boldsymbol{\sigma}_t$ is defined as in the DGP-S2 and $\mathbf{v}_t$ follows independent standardized chi-square distributions with 3 degrees of freedom; in short, $\boldsymbol{\varepsilon}_t \sim$ iid GARCH$(1, 1)$-$\chi^2(3)$.

Next, under the alternative, we generate the noninvertible MA representation with the tax foresight (5.4) in combination with the following DGPs.

DGP-S4. Two-period foresight model with $\boldsymbol{\varepsilon}_t \sim$ iid $N(0, \mathbf{I}_2)$.

DGP-S5. Two-period foresight model with $\boldsymbol{\varepsilon}_t \sim$ iid GARCH$(1, 1)$-$t(3)$.

DGP-S6. Two-period foresight model with $\boldsymbol{\varepsilon}_t \sim$ iid GARCH$(1, 1)$-$\chi^2(3)$.

The DGP-S1 and DGP-S4 are used to examine the effects of Gaussian errors. Our theoretical results require $\{\boldsymbol{\varepsilon}_t\}$ to be iid and, therefore, homoskedastic. However, we employ DGP-S2 and DPG-S3, and DGP-S5 and DPG-S6 to check the effects of conditional heteroskedasticity on the test's performance. For the baseline calibration, we use $\alpha = 0.4$,

TABLE S2. Empirical size of the $\hat{Q}^1$ test using additional DGPs.

	$\overline{h}$	$T = 100$			$T = 250$			$T = 500$		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
		DGP-S1: Invertible with Gaussian errors $\boldsymbol{\varepsilon}_t \sim$ iid $N(0, 1)$								
VAR(1)	10	6.8	4.6	1.9	10.5	7.2	2.8	8.6	6.0	3.4
	20	6.6	4.4	1.2	12.3	9.1	4.6	12.3	7.9	3.9
	30	6.8	3.2	0.7	11.7	8.8	3.9	12.5	8.8	4.0
	40	5.7	2.7	0.5	11.2	7.5	3.0	13.1	8.5	4.1
VAR($\widehat{p}_{\mathrm{SIC}}$)	10	6.6	3.8	1.5	8.5	5.8	2.6	9.5	7.1	3.8
	20	6.5	3.3	1.1	9.3	6.5	3.9	13.9	9.7	4.7
	30	5.7	2.5	0.4	9.0	6.4	3.4	14.7	9.8	5.4
	40	5.7	1.9	0.2	9.1	6.2	2.9	13.9	10.8	4.8
		DGP-S2: Invertible with $\boldsymbol{\varepsilon}_t \sim$ iid GARCH(1, 1)-Student's $t(3)$								
VAR(1)	10	11.0	6.5	2.7	13.1	8.9	4.2	18.0	13.3	7.1
	20	6.4	3.6	1.4	9.9	5.9	3.1	14.5	10.0	5.2
	30	4.1	2.6	0.9	8.3	5.1	2.3	12.7	8.4	3.7
	40	3.6	1.5	0.7	6.9	4.4	1.5	11.2	7.3	3.4
VAR($\widehat{p}_{\mathrm{SIC}}$)	10	10.8	6.5	1.9	16.1	11.6	5.8	24.3	19.3	10.3
	20	6.3	3.9	1.3	11.9	7.3	3.2	18.9	13.8	7.2
	30	5.0	2.8	0.6	9.4	5.9	1.7	15.7	11.1	4.8
	40	4.5	1.8	0.2	7.9	4.4	1.3	14.1	8.6	3.8
		DGP-S3: Invertible with $\boldsymbol{\varepsilon}_t \sim$ iid GARCH(1, 1)-$\chi^2(3)$								
VAR(1)	10	7.8	4.5	1.6	11.5	8.0	4.2	12.8	9.3	4.4
	20	5.9	3.4	1.2	10.9	7.4	2.7	12.6	8.6	4.9
	30	4.7	2.4	0.4	10.5	6.5	2.0	12.1	8.6	4.5
	40	4.5	1.7	0.3	9.2	5.3	1.5	11.4	7.9	3.5
VAR($\widehat{p}_{\mathrm{SIC}}$)	10	7.1	3.7	1.6	9.6	7.0	2.8	13.6	10.1	5.5
	20	6.3	3.0	0.6	8.1	6.2	2.4	13.8	9.5	4.6
	30	5.6	2.0	0.5	8.4	5.3	2.1	13.0	8.4	3.5
	40	4.4	1.7	0.2	7.9	5.2	1.3	11.5	7.2	2.4

$\beta = 0.99$, and $\tau = 0.25$, consistent with Leeper, Walker, and Yang (2013) and Forni and Gabetti (2014) to simulate the bivariate processes (5.3) and (5.4) under DGP-S1–DGP-S6.

Table S2 reports the empirical rejection probabilities of $\hat{Q}^1(h)$ under invertible (no foresight) models under the DGP-S1–DPG-S3 at the 10%, 5%, and 1% levels. There are some size distortions for the heteroskedastic processes DGP-S2 and DPG-S3, but the overall performance is satisfactory for sample sizes typically encountered in macroeconomic applications (e.g., $T = 250$). The simulation results suggest that our characterization with iid shocks may hold more generally with mds structural shocks and confirm that our test is robust to conditional heteroskedasticity of unknown form, which might be important in applications.

Table S3 reports the empirical power of our proposed test against noninvertible (two-period foresight) models under DGP-S4–DPG-S6 at the 10%, 5%, and 1% levels. Under DGP-S4 with Gaussian errors, $\hat{Q}^1(h)$ has no power, which is consistent with our theoretical results. In contrast, for the non-Gaussian noninvertible processes DGP-S5

TABLE S3. Empirical power of the $\hat{Q}^1$ test using additional DGPs.

	$\overline{h}$	$T = 100$			$T = 250$			$T = 500$		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
DGP-S4: Noninvertible with Gaussian errors $\varepsilon_t \sim$ iid $N(0, 1)$										
VAR(1)	10	9.0	5.7	2.2	10.3	7.1	4.5	11.8	9.5	5.4
	20	8.0	4.3	1.7	13.0	9.1	5.0	15.4	11.8	6.9
	30	7.8	3.6	0.9	12.5	8.3	4.1	17.3	13.1	6.7
	40	6.5	3.1	0.8	11.4	7.2	2.9	17.8	12.7	5.7
VAR($\widehat{p}_{\text{SIC}}$)	10	6.6	4.0	1.4	8.7	6.4	3.3	11.6	9.2	4.9
	20	7.0	3.5	1.5	9.0	6.3	3.2	15.9	11.6	6.0
	30	5.8	3.3	1.2	9.2	6.2	2.6	16.1	11.7	6.8
	40	5.1	3.0	0.5	9.2	5.4	2.0	15.8	11.6	6.2
DGP-S5: Noninvertible with $\varepsilon_t \sim$ iid GARCH(1, 1)-Student's $t(3)$										
VAR(1)	10	17.2	11.5	6.0	29.7	20.7	13.0	43.3	34.7	22.9
	20	12.1	7.6	3.4	22.1	15.1	7.4	32.2	24.7	16.0
	30	9.5	5.8	1.8	18.9	11.9	5.0	27.1	21.2	11.8
	40	7.6	3.5	1.3	14.7	9.0	4.0	23.9	17.9	9.1
VAR($\widehat{p}_{\text{SIC}}$)	10	18.7	12.4	6.1	33.8	26.6	15.3	49.0	41.5	29.4
	20	11.8	7.1	2.2	24.8	17.5	9.0	40.3	32.6	20.8
	30	8.8	4.4	1.1	19.2	13.1	6.1	35.0	26.8	15.0
	40	6.7	2.7	0.5	15.6	10.3	4.2	30.3	22.5	11.4
DGP-S6: Noninvertible with $\varepsilon_t \sim$ iid GARCH(1, 1)-$\chi^2(3)$										
VAR(1)	10	32.0	23.8	13.5	70.3	64.5	52.8	96.9	94.5	90.7
	20	22.8	14.4	5.9	59.4	51.8	35.9	92.2	89.5	82.3
	30	15.5	8.8	2.9	50.2	41.3	24.0	87.4	82.4	70.7
	40	11.8	6.8	1.3	43.5	33.1	16.5	82.0	76.5	60.1
VAR($\widehat{p}_{\text{SIC}}$)	10	28.7	21.6	13.1	70.5	64.8	53.6	95.3	94.0	91.7
	20	20.5	13.3	6.9	57.6	48.9	35.7	91.7	89.6	82.3
	30	14.5	10.0	3.9	49.4	39.6	24.2	87.3	82.6	71.0
	40	11.7	6.6	2.3	42.4	32.3	16.6	81.3	75.6	61.3

and DPG-S6 under conditional heteroskedasticity, our test has nontrivial power, particularly for errors with asymmetric distributions such as the χ^2. The power increases with the sample size, as expected.

S1.3 *Simulation results for AIC*

To examine the sensitivity of the lag length criteria choice, we replicate Tables 1 and 2 in the main text using the AIC criteria for the data-driven lag length $\widehat{p}$. Tables S4 and S5 show that the overall results for $\widehat{p}_{\text{AIC}}$ are slightly inferior to those for $\widehat{p}_{\text{SIC}}$.

S1.4 *Simulation results with the degree of persistence varying*

We examine the effects of the severity in the noninvertibility problem on the finite-sample performance of our test by considering values of θ different from the baseline

TABLE S4. Empirical size of the $\hat{Q}^1$ test using alternative lag orders.

	$\overline{h}$	$T = 100$			$T = 250$			$T = 500$		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
		DGP1: Invertible with non-Gaussian errors $\varepsilon_t \sim$ iid Student's $t(3)$								
VAR(1)	10	7.9	4.7	1.7	9.9	6.7	2.8	9.7	6.3	3.4
	20	6.2	2.8	0.7	9.1	5.4	1.3	9.8	6.5	2.4
	30	3.7	1.9	0.3	8.0	3.9	0.7	8.4	5.5	2.3
	40	2.8	1.0	0.2	5.8	3.1	0.6	7.9	4.9	1.7
VAR($\hat{p}_{SIC}$)	10	7.4	4.7	1.5	9.6	7.1	3.4	10.5	7.2	3.4
	20	5.7	2.5	0.3	8.3	4.9	2.0	10.8	6.9	3.0
	30	3.6	1.4	0.0	6.5	4.0	1.1	10.0	5.9	2.2
	40	2.3	1.0	0.0	5.3	3.0	0.7	7.8	4.9	2.0
VAR($\hat{p}_{AIC}$)	10	6.1	3.8	1.0	8.1	6.0	2.9	9.9	6.8	2.8
	20	4.8	1.9	0.2	6.4	3.8	1.5	9.5	6.0	2.3
	30	3.0	1.0	0.0	5.1	3.0	0.7	8.7	5.3	1.9
	40	1.8	0.6	0.0	4.3	2.4	0.4	7.2	4.4	1.9
		DGP2: Invertible with non-Gaussian errors $\varepsilon_t \sim$ iid standardized $\chi^2(3)$								
VAR(1)	10	5.8	3.7	1.4	7.6	4.8	2.4	6.9	5.4	2.6
	20	6.3	3.3	1.2	8.2	5.3	2.1	9.2	6.4	2.9
	30	5.3	2.8	0.7	8.4	3.8	1.8	9.3	6.6	2.4
	40	3.9	1.9	0.4	7.7	3.9	1.2	9.2	6.2	2.3
VAR($\hat{p}_{SIC}$)	10	5.1	3.4	0.9	7.7	5.1	2.1	6.4	4.6	2.3
	20	5.0	2.3	0.4	9.2	5.2	2.1	8.4	5.9	2.3
	30	4.1	1.8	0.2	7.7	4.9	1.9	9.5	6.4	2.4
	40	3.1	1.2	0.1	7.0	4.1	1.1	9.8	5.6	2.7
VAR($\hat{p}_{AIC}$)	10	2.3	1.4	0.3	6.1	4.0	1.6	3.7	2.8	1.5
	20	2.2	1.0	0.0	6.9	3.6	1.7	6.1	4.1	1.3
	30	2.2	0.7	0.0	5.8	3.7	1.5	7.6	4.7	1.8
	40	1.5	0.3	0.0	5.2	3.2	0.8	7.7	4.7	2.0

value of 0.297 used in Tables 1 and 2 in the main text. Here we do this in conjunction with the sensitivity to the persistence in the process by varying the persistence parameter $\alpha \in [0.01, 0.1, 0.2, \ldots, 0.8, 0.9, 0.99]$, setting the discount factor and the steady state tax rate to $\beta = 0.99$ and $\tau = 0.25$, respectively, and computing the corresponding θ with $\theta = \alpha\beta(1 - \tau)$, leading to the values in $\theta \in [0.007, 0.074, 0.149, 0.223, 0.297, 0.371, 0.446, 0.520, 0.594, 0.668, 0.735]$.

The results varying the degree of persistence are reported in Figure S1, the right panel of which suggests that the power results under the alternative (DGP4) are not sensitive to the persistence parameter when compared with Figure 1 in the main text. We also simulated under the null (DGP2) varying the degree of persistence. The left panel of Figure S1 exhibits that large values of persistence such as $\alpha = 0.99$ (i.e., $\theta = 0.735$) do seem to have an effect on the empirical size of the test, which suggests a different asymptotic theory for nonstationary processes. Establishing this theory is beyond the scope of this study. Hence, practitioners are highly recommended to transform the data

TABLE S5. Empirical power of the $\hat{Q}^1$ test using alternative lag orders.

	$\overline{h}$	$T = 100$			$T = 250$			$T = 500$		
		10%	5%	1%	10%	5%	1%	10%	5%	1%
DGP3: Noninvertible with non-Gaussian errors $\varepsilon_t \sim$ iid Student's $t(3)$										
VAR(1)	10	14.0	8.8	4.7	31.7	24.4	14.1	58.0	51.8	40.4
	20	9.6	5.5	1.7	22.7	15.4	8.2	48.4	40.5	26.0
	30	6.9	3.0	0.6	16.2	10.1	4.4	40.2	29.5	17.3
	40	4.3	1.7	0.2	13.3	7.7	2.5	31.8	21.9	11.2
VAR($\hat{p}_{SIC}$)	10	13.2	9.6	4.3	32.2	26.9	15.7	57.6	51.1	39.9
	20	8.4	5.1	1.4	25.7	17.3	7.7	47.2	38.8	26.7
	30	5.7	2.8	0.5	19.8	12.2	4.5	39.0	29.5	17.9
	40	4.2	1.2	0.1	15.3	8.7	2.6	32.4	23.0	11.6
VAR($\hat{p}_{AIC}$)	10	11.9	8.5	3.8	28.8	23.7	13.4	53.1	46.3	35.3
	20	7.0	4.3	1.0	22.2	14.6	6.2	41.8	33.7	22.9
	30	4.6	2.2	0.4	16.8	10.1	3.8	34.5	25.7	15.3
	40	3.0	1.2	0.0	13.0	6.9	2.1	28.9	20.1	9.5
DGP4: Noninvertible with non-Gaussian errors $\varepsilon_t \sim$ iid standardized $\chi^2(3)$										
VAR(1)	10	33.9	25.9	15.1	79.6	75.9	66.0	98.7	98.4	96.8
	20	23.4	14.4	7.0	70.1	62.6	46.5	96.2	95.5	91.6
	30	17.1	10.3	3.6	59.7	49.7	32.8	93.3	90.5	83.8
	40	13.2	6.7	1.4	51.1	39.2	22.8	89.6	85.2	75.0
VAR($\hat{p}_{SIC}$)	10	33.2	26.3	16.3	78.9	74.7	65.9	99.0	98.7	97.0
	20	23.5	16.3	6.6	69.5	63.0	48.5	96.2	95.0	91.4
	30	16.9	9.8	3.1	61.1	51.5	35.0	93.3	90.6	84.0
	40	11.9	6.2	1.4	53.5	42.7	24.4	89.2	84.7	76.5
VAR($\hat{p}_{AIC}$)	10	28.4	21.5	12.9	74.4	69.7	61.4	98.8	98.5	96.3
	20	17.6	11.9	4.7	63.7	56.9	43.2	95.4	93.9	89.0
	30	12.1	6.6	1.3	55.6	46.4	30.9	91.2	88.1	80.5
	40	8.4	3.6	0.5	48.2	37.5	20.3	86.8	81.7	72.5

to induce stationarity by conventional methods, for example, differencing, allowing for trends, or dividing by another variables, before applying our test.

S2. PROOFS OF THE MAIN RESULTS

Throughout this supplement, C denotes a generic positive and bounded constant that may differ from place to place.

PROOF OF THEOREM 1. If $\{\mathbf{x}_t\}$ is invertible, then $\{\mathbf{u}_t\} = \{\boldsymbol{\Phi}_0^{-1}\boldsymbol{\Theta}_0\boldsymbol{\varepsilon}_t\}$ is a mds. We prove the reciprocal: if $\{\mathbf{x}_t\}$ is noninvertible, then $\{\mathbf{u}_t\}$ is not a mds. Suppose on the contrary that $\{\mathbf{u}_t\}$ is a mds. Then we will reach a contradiction by using and extending to the multivariate case the results of Rosenblatt (2000, Section 5.4).
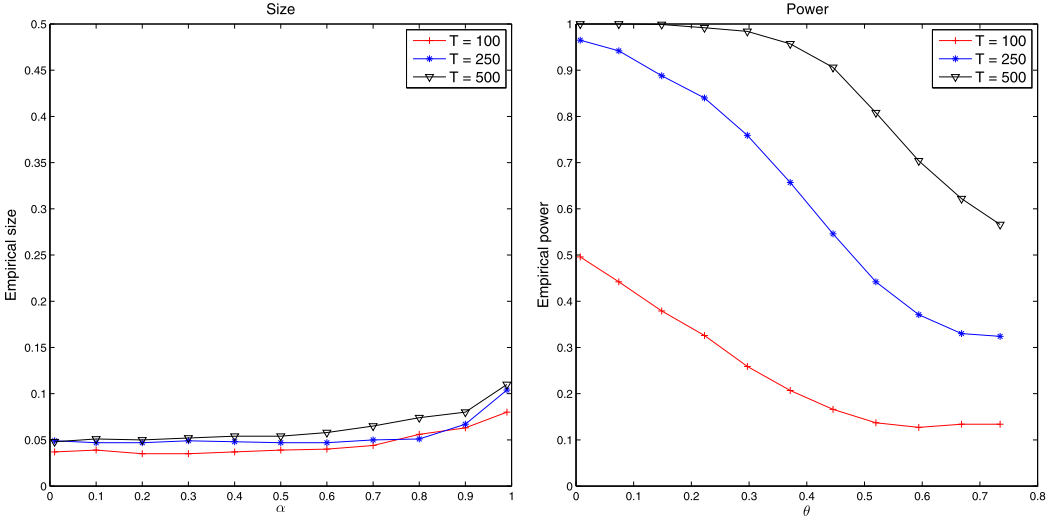
FIGURE S1. Sensitivity of $\hat{Q}^1$ test performance to the noninvertibility governing parameter (θ) under DGP2 (size) and the DGP4 (power) with the degree of persistence α varying from 0.01 to 0.99. As the value of θ becomes smaller, the problem of noninvertibility becomes increasingly serious.

Consider first the case $r = 1$ (i.e., Assumption A.1(i)). By Theorem 1 in Lippi and Reichlin (1994) and definition of $\widetilde{\varepsilon}_t = \mathbf{K}_1 \varepsilon_t$, we have

$$u_{t,1} = \frac{L - b_1}{1 - b_1 L} \widetilde{\varepsilon}_{t,1}.$$

Define the MA(1) process $y_t = (L - b_1)\widetilde{\varepsilon}_{t,1}$. Then, by Corollary 5.4.3 in Rosenblatt (2000), the process y_t has a nonlinear one-ahead predictor $E[y_t|\mathcal{F}^y_{t-1}]$. But then, since $(1 - b_1 L)u_{t,1} = y_t$,

$$E\big[u_{t,1}|\mathcal{F}^y_{t-1}\big] = b_1 E\big[u_{t-1,1}|\mathcal{F}^y_{t-1}\big] + E\big[y_t|\mathcal{F}^y_{t-1}\big].$$

The condition $|b_1| < 1$ implies that $\mathcal{F}^y_{t-1} = \mathcal{F}^{u_{t,1}}_{t-1}$, so that

$$E\big[u_{t,1}|\mathcal{F}^{u_{t,1}}_{t-1}\big] = b_1 u_{t-1,1} + E\big[y_t|\mathcal{F}^y_{t-1}\big]. \tag{1}$$

Therefore, we conclude that if $E[u_{t,1}|\mathcal{F}^{u_{t,1}}_{t-1}]$ is zero, then $E[y_t|\mathcal{F}^y_{t-1}]$ is linear, which contradicts Corollary 5.4.3 in Rosenblatt (2000). Therefore, $\{u_{t,1}\}$, and hence $\{\mathbf{u}_t\}$, is not a mds.

Consider now the case $r > 1$ (i.e., Assumption A.1(ii)). Let $\psi(\boldsymbol{\tau})$ denote the characteristic function ε_t. Define the VMA(q) noninvertible process

$$\mathbf{y}_t = \boldsymbol{\Theta}(L)\varepsilon_t.$$

The characteristic function of $\{\ldots, \mathbf{y}_{-s}, \mathbf{y}_{-s+1}, \ldots, \mathbf{y}_0\}$ is given by

$$
\eta(\ldots, \boldsymbol{\tau}_s, \boldsymbol{\tau}_{s-1}, \ldots, \boldsymbol{\tau}_0) = E\left[\exp\left(i\sum_{l=0}^{\infty}\boldsymbol{\tau}_l'\mathbf{y}_{-l}\right)\right]
$$
$$
= \prod_{k=0}^{\infty}\psi\left(\sum_{l=0}^{\infty}\boldsymbol{\tau}_l'\boldsymbol{\Theta}_{k-l}\right),
$$
(2)

where the equality follows from the independence of $\{\boldsymbol{\varepsilon}_t\}$ and a change of indexes ($k = l+j$), that is,

$$
\sum_{l=0}^{\infty}\boldsymbol{\tau}_l'\mathbf{y}_{-l} = \sum_{l=0}^{\infty}\boldsymbol{\tau}_l'\left(\sum_{j=0}^{q}\boldsymbol{\Theta}_j\boldsymbol{\varepsilon}_{-l-j}\right) = \sum_{k=0}^{\infty}\left(\sum_{l=0}^{\infty}\boldsymbol{\tau}_l'\boldsymbol{\Theta}_{k-l}\right)\boldsymbol{\varepsilon}_{-k}.
$$

Note that for a VMA(q) model $\boldsymbol{\Theta}_k = \mathbf{0}$ for $k < 0$ and $k > q$.

On the other hand, it is well known that

$$
\eta_{\boldsymbol{\tau}_0}(\ldots, \boldsymbol{\tau}_s, \ldots, \boldsymbol{\tau}_1, \mathbf{0}) \equiv \left.\frac{\partial\eta(\ldots, \boldsymbol{\tau}_s, \ldots, \boldsymbol{\tau}_1, \boldsymbol{\tau}_0)}{\partial\boldsymbol{\tau}_0}\right|_{\boldsymbol{\tau}_0=\mathbf{0}}
$$
$$
= E\left[i\mathbf{y}_0\exp\left(i\sum_{l=1}^{\infty}\boldsymbol{\tau}_l'\mathbf{y}_{-l}\right)\right]
$$
(3)
$$
= E\left[iE[\mathbf{y}_0|\mathcal{F}_{-1}^y]\exp\left(i\sum_{l=1}^{\infty}\boldsymbol{\tau}_l'\mathbf{y}_{-l}\right)\right].
$$

Similarly, the joint characteristic function of $\{\ldots, \mathbf{y}_{-s}, \mathbf{y}_{-s+1}, \ldots, \mathbf{y}_{-1}\}$ is given by

$$
\tilde{\eta}(\ldots, \boldsymbol{\tau}_s, \ldots, \boldsymbol{\tau}_1) = \prod_{k=0}^{\infty}\psi\left(\sum_{l=1}^{\infty}\boldsymbol{\tau}_l'\boldsymbol{\Theta}_{k-l}\right).
$$

It then follows from (2) that

$$
\left.\frac{\partial\log\eta(\ldots, \boldsymbol{\tau}_s, \ldots, \boldsymbol{\tau}_1, \boldsymbol{\tau}_0)}{\partial\boldsymbol{\tau}_0}\right|_{\boldsymbol{\tau}_0=\mathbf{0}} =: \frac{\eta_{\boldsymbol{\tau}_0}(\ldots, \boldsymbol{\tau}_s, \ldots, \boldsymbol{\tau}_1, \mathbf{0})}{\tilde{\eta}(\ldots, \boldsymbol{\tau}_s, \ldots, \boldsymbol{\tau}_1)}
$$
$$
= \sum_{k=0}^{\infty}\boldsymbol{\Theta}_k\mathbf{h}\left(\sum_{l=1}^{\infty}\boldsymbol{\tau}_l'\boldsymbol{\Theta}_{k-l}\right)
$$
(4)

for all $\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \ldots$, where

$$
\mathbf{h}(\boldsymbol{\tau}) = \frac{\partial\log\psi(\boldsymbol{\tau})}{\partial\boldsymbol{\tau}}.
$$

Also, for all $j \geq 1$,

$$\frac{\partial \log \tilde{\eta}(\ldots, \boldsymbol{\tau}_s, \ldots, \boldsymbol{\tau}_1)}{\partial \boldsymbol{\tau}_j} =: \frac{\eta_{\boldsymbol{\tau}_j}(\ldots, \boldsymbol{\tau}_s, \ldots, \boldsymbol{\tau}_1)}{\tilde{\eta}(\ldots, \boldsymbol{\tau}_s, \ldots, \boldsymbol{\tau}_1)}$$

$$= \sum_{k=0}^{\infty} \boldsymbol{\Theta}_{k-j} \mathbf{h}\left(\sum_{l=1}^{\infty} \boldsymbol{\tau}_l' \boldsymbol{\Theta}_{k-l}\right). \tag{5}$$

If the best predictor $E[\mathbf{y}_0 | \mathcal{F}_{-1}^y]$ is linear, that is,

$$E[\mathbf{y}_0 | \mathcal{F}_{-1}^y] = \sum_{j=1}^{\infty} \mathbf{b}_j \mathbf{y}_{-j},$$

then from (3) we must have

$$\eta_{\boldsymbol{\tau}_0}(\ldots, \boldsymbol{\tau}_s, \ldots, \boldsymbol{\tau}_1, \mathbf{0}) = \sum_{j=1}^{\infty} \mathbf{b}_j \eta_{\boldsymbol{\tau}_j}(\ldots, \boldsymbol{\tau}_s, \ldots, \boldsymbol{\tau}_1).$$

Using (4) and (5) we conclude

$$\sum_{k=0}^{\infty} \boldsymbol{\Gamma}_k \mathbf{h}\left(\sum_{l=1}^{\infty} \boldsymbol{\tau}_l' \boldsymbol{\Theta}_{k-l}\right) = 0, \tag{6}$$

where

$$\boldsymbol{\Gamma}_k := \boldsymbol{\Theta}_k - \sum_{j=1}^{\infty} \mathbf{b}_j \boldsymbol{\Theta}_{k-j}.$$

Let $h_j(\boldsymbol{\tau})$ be the jth component of $\mathbf{h}(\boldsymbol{\tau})$, and let $\gamma_{k,ij}$ be the ijth component of the matrix $\boldsymbol{\Gamma}_k$. Then, with this notation, (6) is written as

$$0 = \sum_{k=0}^{\infty} \sum_{j=0}^{d} \gamma_{k,ij} h_j\left(\sum_{l=1}^{\infty} \boldsymbol{\tau}_l' \boldsymbol{\Theta}_{k-l}\right) \quad \text{for each } i = 1, 2, \ldots, d.$$

Then differentiating this equation with respect to $\boldsymbol{\tau}_{l_1}$ and $\boldsymbol{\tau}_{l_2}$ and evaluating at $\boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \cdots = 0$, we obtain for each $i = 1, 2, \ldots, d$,

$$\mathbf{0} = \sum_{k=0}^{\infty} \sum_{j=0}^{d} \gamma_{k,ij} \frac{\partial^2 h_j\left(\sum_{l=1}^{\infty} \boldsymbol{\tau}_l' \boldsymbol{\Theta}_{k-l}\right)}{\partial \boldsymbol{\tau}_{l_1} \partial \boldsymbol{\tau}_{l_2}'} \tag{7}$$

$$= \sum_{j=0}^{d} \mathbf{H}_j \sum_{k=0}^{\infty} \gamma_{k,ij} \boldsymbol{\Theta}_{k-l_1}' \boldsymbol{\Theta}_{k-l_2}.$$

By the linear independence of the matrices $\mathbf{H}_j$ we obtain

$$\mathbf{0} = \sum_{k=0}^{\infty} \gamma_{k,ij} \boldsymbol{\Theta}'_{k-l_1} \boldsymbol{\Theta}_{k-l_2} \tag{8}$$

for any $l_1, l_2 = 1, 2, \ldots$.

We now relate $\boldsymbol{\Gamma}_k$ to the Wold innovations. Note that the Wold innovations satisfy

$$\mathbf{u}_0 = \widetilde{\boldsymbol{\Theta}}^{-1}(L)\mathbf{y}_0$$

$$= \left(\mathbf{I}_d - \sum_{j=1}^{\infty} \mathbf{b}_j L^j \right) \mathbf{y}_0.$$

Therefore, with $\boldsymbol{\Gamma}(L) = \sum_{j=0}^{\infty} \boldsymbol{\Gamma}_j L^j$, and since $\boldsymbol{\Gamma}_k = \boldsymbol{\Theta}_k - \sum_{j=1}^{\infty} \mathbf{b}_j \boldsymbol{\Theta}_{k-j}$, we obtain

$$\boldsymbol{\Gamma}(L) = \widetilde{\boldsymbol{\Theta}}^{-1}(L)\boldsymbol{\Theta}(L), \tag{9}$$

which is the Blaschke matrix $\mathbf{A}(L) = \widetilde{\boldsymbol{\Theta}}^{-1}(L)\boldsymbol{\Theta}(L)$. Since there is at least one noninvertible root we have that $\boldsymbol{\Gamma}_k \neq 0$ for some $k > 0$.

Now, applying equation (8) with $l_1 = r - q$ and $l_2 = r$, for some $r > q$, we get

$$\mathbf{0} = \boldsymbol{\Theta}'_q \boldsymbol{\Theta}_0 \gamma_{r,ij}.$$

Since $\boldsymbol{\Theta}'_q \boldsymbol{\Theta}_0 \neq 0$, we have shown that $\boldsymbol{\Gamma}_r = 0$ for $r > q$.

On the other hand, from (9), we can show that for $k > 0$,

$$\boldsymbol{\Gamma}_k = \sum_{j=1}^{r} \boldsymbol{\alpha}_j b_j^k, \tag{10}$$

where b_j are the noninvertible roots and the $\{\boldsymbol{\alpha}_j\}_{j=1}^r$ are $d \times d$ matrices. To see this, first consider the case $r = 1$ and use Theorem 1 in Lippi and Reichlin (1994) to conclude

$$\boldsymbol{\Gamma}(L) = \boldsymbol{R}(b_1, L)\mathbf{K}_1,$$

where $\mathbf{K}_1$ is an orthogonal matrix, $|b_1| < 1$ is the noninvertible root, and

$$\boldsymbol{R}(\alpha, L) = \begin{pmatrix} \dfrac{L - \alpha}{1 - \alpha L} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{d-1} \end{pmatrix}.$$

Then simple calculations verify (10) with

$$\boldsymbol{\alpha}_1 = \begin{pmatrix} \dfrac{1 + b_1}{b_1} & \dfrac{1 + b_1}{b_1} \\ \dfrac{1}{1 - b_1} & \dfrac{1}{1 - b_1} \end{pmatrix} \mathbf{K}_1.$$

For $r > 1$ the expansion (10) can be shown by induction on the number of noninvertible roots r and using Theorem 1 in Lippi and Reichlin (1994).

However, from $\boldsymbol{\Gamma}_r = 0$ for $r > q$, we conclude that all $\boldsymbol{\alpha}_j$s in (10) must be zero. But this implies that $\boldsymbol{\Gamma}_k = 0$ for all $k > 0$, which contradicts (9) (i.e., that $\boldsymbol{\Gamma}_k \neq 0$ for some $k > 0$). Therefore, we conclude that $E[\mathbf{y}_0|\mathcal{F}^y_{-1}]$ must be nonlinear. Finally, since the VAR filter is causal and $\boldsymbol{\Phi}(L)\mathbf{x}_t = \mathbf{y}_t$, we argue as in (1) to show that $E[\mathbf{x}_0|\mathcal{F}^x_{-1}]$ must be nonlinear. The last nonlinearity implies that the Wold innovations are not a mds.   □

Henceforth, we let $\tilde{Q}^1(h)$ be defined in the same way as $\hat{Q}^1(h)$ in (3.9) with $\hat{\mathbf{u}}_t$ replaced by $\mathbf{u}_t$.

The proof of Theorem A.1 is very similar to, but simpler than, that of Theorem 2 below, and hence it is omitted.

The proof of Theorem A.2 is very similar to, but simpler than, that of Theorem 3 below, and hence it is omitted.

PROOF OF THEOREM 2. The proof of Theorem 2 consists of the proofs of Theorems A.3 and A.4 below.

THEOREM A.3. *Under the conditions of Theorem 2, $\hat{Q}^1(h) - \tilde{Q}^1(h) \overset{p}{\to} 0$.*

THEOREM A.4. *Under the conditions of Theorem 2, $\tilde{Q}^1(h) \overset{d}{\to} N(0, 1)$.*

PROOF OF THEOREM A.3. Put $T_j := T - |j|$, and let $\gamma^1_j(\mathbf{b})$ be defined in the same way as $\hat{\gamma}^1_j(\mathbf{b})$ in (3.5) from the main text, with $\hat{\mathbf{u}}_t$ replaced by $\mathbf{u}_t$. To show $\hat{Q}^1(h) - \tilde{Q}^1(h) \overset{p}{\to} 0$, it suffices to show

$$h^{-\frac{1}{2}} \int \sum_{j=1}^{T-1} k^2(j/h) T_j \big[\|\hat{\gamma}^1_j(\mathbf{b})\|^2 - \|\tilde{\gamma}^1_j(\mathbf{b})\|^2 \big] dW(\mathbf{b}) \overset{p}{\to} 0, \tag{11}$$

$\hat{C}^1(h) - \tilde{C}^1(h) = O_P(T^{-\frac{1}{2}})$, and $\hat{D}^1(h) - \tilde{D}^1(h) = o_p(1)$, where $\tilde{C}^1(h)$ and $\tilde{D}^1(h)$ are defined in the same way as $\hat{C}^1(h)$ and $\hat{D}^1(h)$ in (3.10) in the main text, with $\hat{\mathbf{u}}_t$ replaced by $\mathbf{u}_t$. For space, we focus on the proof of (11); the proofs for $\hat{C}^1(h) - \tilde{C}^1(h) = O_P(T^{-\frac{1}{2}})$ and $\hat{D}^1(h) - \tilde{D}^1(h) = o_p(1)$ are straightforward.

Noting that

$$\|\hat{\gamma}^1_j(\mathbf{b})\|^2 - \|\tilde{\gamma}^1_j(\mathbf{b})\|^2 = \sum_{m=1}^{d} \big[|\hat{\gamma}^1_{j,m}(\mathbf{b})|^2 - |\tilde{\gamma}^1_{j,m}(\mathbf{b})|^2 \big],$$

where $\hat{\gamma}^1_{j,m}(\mathbf{b})$ and $\tilde{\gamma}^1_{j,m}(\mathbf{b})$ are the mth element of $\hat{\gamma}^1_j(\mathbf{b})$ and $\tilde{\gamma}^1_j(\mathbf{b})$, respectively. Hence it would be sufficient to show that

$$h^{-\frac{1}{2}} \int \sum_{j=1}^{T-1} k^2(j/h) T_j \big[|\hat{\gamma}^1_{j,m}(\mathbf{b})|^2 - |\tilde{\gamma}^1_{j,m}(\mathbf{b})|^2 \big] dW(\mathbf{b}) \overset{p}{\to} 0.$$

We first decompose

$$\int \sum_{j=1}^{T-1} k^2(j/h) T_j \big[|\hat{\gamma}^1_{j,m}(\mathbf{b})|^2 - |\tilde{\gamma}^1_{j,m}(\mathbf{b})|^2 \big] dW(\mathbf{b}) = \hat{A}_1 + 2\,\mathrm{Re}(\hat{A}_2), \tag{12}$$

where

$$\hat{A}_1 = \int \sum_{j=1}^{T-1} k^2(j/h) T_j |\hat{\gamma}^1_{j,m}(\mathbf{b}) - \tilde{\gamma}^1_{j,m}(\mathbf{b})|^2 \, dW(\mathbf{b}),$$

$$\hat{A}_2 = \int \sum_{j=1}^{T-1} k^2(j/h) T_j [\hat{\gamma}^1_{j,m}(\mathbf{b}) - \tilde{\gamma}^1_{j,m}(\mathbf{b})] \tilde{\gamma}^{1*}_{j,m}(\mathbf{b}) \, dW(\mathbf{b}),$$

where $\mathrm{Re}(\hat{A}_2)$ is the real part of $\hat{A}_2$ and $\tilde{\gamma}^{1*}_{j,m}(\mathbf{b})$ is the complex conjugate of $\tilde{\gamma}^1_{j,m}(\mathbf{b})$. Then (11) follows from Propositions A.1 and A.2 below, and $h \to \infty$ as $T \to \infty$. $\square$

PROPOSITION A.1. *Under the conditions of Theorem 1, $\hat{A}_1 = O_P(1)$.*

PROPOSITION A.2. *Under the conditions of Theorem 1, $h^{-\frac{1}{2}} \hat{A}_2 \xrightarrow{P} 0$.*

PROOF OF PROPOSITION A.1. Put $\hat{\delta}_t(\mathbf{b}) = e^{i\mathbf{b}'\hat{\mathbf{u}}_t} - e^{i\mathbf{b}'\mathbf{u}_t}$, $\psi_t(\mathbf{b}) := e^{i\mathbf{b}'\mathbf{u}_t} - \varphi(\mathbf{b})$, and $\varphi(\mathbf{b}) := E(e^{i\mathbf{b}'\mathbf{u}_t})$. Then straightforward algebra yields that for $j > 0$,

$$\hat{\gamma}^1_{j,m}(\mathbf{b}) - \tilde{\gamma}^1_{j,m}(\mathbf{b})$$

$$= iT_j^{-1} \sum_{t=j+1}^{T} (\hat{u}_{t,m} - u_{t,m})\hat{\delta}_{t-j}(\mathbf{b})$$

$$- i\left[T_j^{-1} \sum_{t=j+1}^{T} (\hat{u}_{t,m} - u_{t,m}) \right]\left[T_j^{-1} \sum_{t=j+1}^{T} \hat{\delta}_{t-j}(\mathbf{b}) \right]$$

$$+ iT_j^{-1} \sum_{t=j+1}^{T} u_{t,m}\hat{\delta}_{t-j}(\mathbf{b}) - i\left(T_j^{-1} \sum_{t=j+1}^{T} u_{t,m} \right)\left[T_j^{-1} \sum_{t=j+1}^{T} \hat{\delta}_{t-j}(\mathbf{b}) \right] \qquad (13)$$

$$+ iT_j^{-1} \sum_{t=j+1}^{T} (\hat{u}_{t,m} - u_{t,m})\psi_{t-j}(\mathbf{b})$$

$$- i\left[T_j^{-1} \sum_{t=j+1}^{T} (\hat{u}_{t,m} - u_{t,m}) \right]\left[T_j^{-1} \sum_{t=j+1}^{T} \psi_{t-j}(\mathbf{b}) \right]$$

$$= \hat{B}_{1j,m}(\mathbf{b}) - \hat{B}_{2j,m}(\mathbf{b}) + \hat{B}_{3j,m}(\mathbf{b}) - \hat{B}_{4j,m}(\mathbf{b}) + \hat{B}_{5j,m}(\mathbf{b}) - \hat{B}_{6j,m}(\mathbf{b}), \quad \text{say.}$$

It follows that $\hat{A}_1 \le 2^5 \sum_{a=1}^{6} \sum_{j=1}^{T-1} k^2(j/h) T_j \int |\hat{B}_{aj,m}(\mathbf{b})|^2 \, dW(\mathbf{b})$. Proposition A.1 follows from Lemmas A.1–A.6 below, and $h/T \to 0$. $\square$

LEMMA A.1. *We have $\sum_{j=1}^{T-1} k^2(j/h) T_j \int |\hat{B}_{1j,m}(\mathbf{b})|^2 \, dW(\mathbf{b}) = O_P(h/T)$.*

LEMMA A.2. *We have $\sum_{j=1}^{T-1} k^2(j/h) T_j \int |\hat{B}_{2j,m}(\mathbf{b})|^2 \, dW(\mathbf{b}) = O_P(h/T)$.*

LEMMA A.3. *We have $\sum_{j=1}^{T-1} k^2(j/h) T_j \int |\hat{B}_{3j,m}(\mathbf{b})|^2 \, dW(\mathbf{b}) = O_P(h/T)$.*

LEMMA A.4. *We have* $\sum_{j=1}^{T-1} k^2(j/h) T_j \int |\hat{B}_{4j,m}(\mathbf{b})|^2 \, dW(\mathbf{b}) = O_P(h/T)$.

LEMMA A.5. *We have* $\sum_{j=1}^{T-1} k^2(j/h) T_j \int |\hat{B}_{5j,m}(\mathbf{b})|^2 \, dW(\mathbf{b}) = O_P(1)$.

LEMMA A.6. *We have* $\sum_{j=1}^{T-1} k^2(j/h) T_j \int |\hat{B}_{6j,m}(\mathbf{b})|^2 \, dW(\mathbf{b}) = O_P(h/T)$.

PROOF OF LEMMA A.1. By the Cauchy–Schwarz inequality, we have

$$
\left| \hat{B}_{1j,m}(\mathbf{b}) \right|^2 \leq \left[T_j^{-1} \sum_{t=j+1}^{T} (\hat{u}_{t,m} - u_{t,m})^2 \right] \left[T_j^{-1} \sum_{t=j+1}^{T} |\hat{\delta}_{t-j}(\mathbf{b})|^2 \right]
$$

$$
\leq \|\mathbf{b}\|^2 \left[T_j^{-1} \sum_{t=j+1}^{T} (\hat{u}_{t,m} - u_{t,m})^2 \right]^2.
$$

It follows that

$$
\sum_{j=1}^{T-1} k^2(j/h) T_j \int \left| \hat{B}_{1j,m}(\mathbf{b}) \right|^2 dW(\mathbf{b})
$$

$$
\leq \left[\sum_{j=1}^{T-1} a_T(j) \right] \left[\sum_{t=j+1}^{T} (\hat{u}_{t,m} - u_{t,m})^2 \right]^2 \int \|\mathbf{b}\|^2 \, dW(\mathbf{b}) = O_p(h/T),
$$

where $a_T(j) = k^2(j/h) T_j^{-1}$ and we have used the fact that

$$
\sum_{j=1}^{T-1} a_T(j) = O(h/T) \tag{14}
$$

given $h = cn^\lambda$ for $\lambda \in (0, 1)$, as shown in Hong (1999, (A.15), p. 1213), and

$$
\left\| \mathbf{u}_t(\boldsymbol{\theta}) - \hat{\boldsymbol{u}}_t(\boldsymbol{\theta}) \right\|^2 \leq C \sum_{k=t}^{\infty} \rho^k \|\mathbf{x}_{t-k}\|^2 + C\rho^t \|\bar{\mathbf{x}}_0\|^2, \tag{15}
$$

where $\mathbf{u}_t(\boldsymbol{\theta})$ and $\hat{\mathbf{u}}_t(\boldsymbol{\theta})$ are residuals based on the infeasible information set $\{\mathbf{x}_T, \mathbf{x}_{T-1}, \ldots\}$ and the observed information set $\{\mathbf{x}_T, \mathbf{x}_{T-1}, \ldots, \mathbf{x}_1, \bar{\mathbf{x}}_0\}$, respectively, $0 < \rho < 1$, C and ρ are constants independent of the parameter θ,

$$
E \sup_{\theta \in \Theta} \left\| \mathbf{u}_t(\boldsymbol{\theta}) - \hat{\mathbf{u}}_t(\boldsymbol{\theta}) \right\|^2 = O(\rho^t), \tag{16}
$$

where $O(\cdot)$ holds uniformly in all t, following Boubacar Mainassara and Francq (2011), and

$$
\sum_{t=1}^{T} \left\| \mathbf{u}_t(\hat{\boldsymbol{\theta}}) - \mathbf{u}_t(\boldsymbol{\theta}_0) \right\|^2 \leq T \|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|^2 T^{-1} \sum_{t=1}^{T} \sup_{\theta \in \Theta} \left\| \frac{\partial \mathbf{u}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\|^2 = O_p(1). \tag{17}
$$

The proof of Lemma A.2 is similar to the proof of Lemma A.1.

PROOF OF LEMMA A.3. A second-order Taylor series expansion yields

$$T_j\big|\hat{B}_{3j,m}(\mathbf{b})\big|$$

$$\leq \sum_{t=j+1}^{T} \big|u_{t,m}\big[e^{i\mathbf{b}'\hat{\mathbf{u}}_{t-j}} - e^{i\mathbf{b}'\mathbf{u}_{t-j}(\hat{\boldsymbol{\theta}})}\big]\big|^2 + \sum_{t=j+1}^{T} \|\mathbf{b}\|\|u_{t,m}A_t\|$$

$$+ \sum_{t=j+1}^{T} |u_{t,m}|\|\mathbf{b}\|^2\big\|\mathbf{u}_{t-j}(\hat{\boldsymbol{\theta}}) - \mathbf{u}_{t-j}\big\|^2 \tag{18}$$

$$+ \|\mathbf{b}\|\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|\bigg\|\sum_{t=j+1}^{T} u_{t,m}\bigg[\frac{\partial \mathbf{u}_{t-j}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\bigg]' e^{i\mathbf{b}'\mathbf{u}_{t-j}}\bigg\|,$$

where $A_t = (A_{t1}, \ldots, A_{td})'$,

$$A_{tj} = \frac{1}{2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)' \frac{\partial^2 \mathbf{u}_{tj}(\bar{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}\,\partial \boldsymbol{\theta}'}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0),$$

and $\bar{\boldsymbol{\theta}}$ is between $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$.

It follows that

$$\sum_{j=1}^{T-1} k^2(j/h)T_j \int \big|\hat{B}_{3j,m}(\mathbf{b})\big|^2 dW(\mathbf{b})$$

$$\leq 8\bigg[\sum_{t=1}^{T} |u_{t,m}|\big\|\hat{\mathbf{u}}_{t-j} - \mathbf{u}_{t-j}(\hat{\boldsymbol{\theta}})\big\|\bigg]^2 \sum_{j=1}^{T-1} k^2(j/h)T_j^{-1} \int \|\mathbf{b}\|^2 dW(\mathbf{b})$$

$$+ 8\bigg[\sum_{t=1}^{T} \|A_t\|^2\bigg]\bigg[\sum_{t=1}^{T} u_{t,m}^2\bigg]\sum_{j=1}^{T-1} k^2(j/h)T_j^{-1} \int \|\mathbf{b}\|^2 dW(\mathbf{b})$$

$$+ 8\big\|(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\big\|^4\bigg[\sum_{t=1}^{T} u_{t,m}^2\bigg]\bigg\{\sum_{t=1}^{T}\bigg[\sup_{\theta\in\Theta_0}\bigg\|\frac{\partial \mathbf{u}_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\bigg\|\bigg]^4\bigg\}$$

$$\times \sum_{j=1}^{T-1} k^2(j/h)T_j^{-1} \int \|\mathbf{b}\|^2 dW(\mathbf{b})$$

$$+ 8\big\|(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\big\|^2 \sum_{j=1}^{T-1} k^2(j/h)T_j^{-1}$$

$$\times \int \bigg\|\sum_{t=j+1}^{T} u_{t,m}\bigg[\frac{\partial \mathbf{u}_{t-j}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\bigg]' e^{i\mathbf{b}'\mathbf{u}_{t-j}}\bigg\|\|\mathbf{b}\|^2 dW(\mathbf{b})$$

$$= O_p(h/T),$$

where we made use of the fact that $T_j^{-1} E \| \sum_{t=j+1}^{T} u_{t,m} [\frac{\partial \mathbf{u}_{t-j}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}]' e^{i\mathbf{b}' \mathbf{u}_{t-j}} \|^2 = O(1)$ because $\mathbf{u}_t$ is a mds under $\mathbb{H}_0$. We also made use of the fact that

$$\sum_{t=1}^{T} \sup_{\theta \in \Theta} \| \mathbf{u}_t(\boldsymbol{\theta}) - \hat{\mathbf{u}}_t(\boldsymbol{\theta}) \| \leq \sum_{t=1}^{T} \left(C \sum_{k=t}^{\infty} \rho^k \| \mathbf{x}_{t-k} \| + C\rho^t \| \bar{\mathbf{x}}_0 \| \right) = O_P(1) \tag{19}$$

and

$$\frac{1}{T} \sum_{t=1}^{T} \sup_{\theta \in \Theta} \left\| \frac{\partial^2 \mathbf{u}_{tj}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \, \partial \boldsymbol{\theta}'} \right\|^2 \leq \frac{1}{T} \sum_{t=1}^{T} \left(C + \sum_{k=1}^{\infty} \rho_2^k \| \mathbf{x}_{t-k} \|^2 \right) = O_P(1), \tag{20}$$

where $0 < \rho, \rho_2 < 1$ and $j = 1, \dots, d$. □

PROOF OF LEMMA A.4. By the Cauchy–Schwarz inequality, we have

$$\sum_{j=1}^{T-1} k^2(j/h) T_j \int \left| \hat{B}_{4j,m}(\mathbf{b}) \right|^2 dW(\mathbf{b})$$
$$\leq \sum_{j=1}^{T-1} k^2(j/h) \int \left(T_j^{-1} \sum_{t=j+1}^{T} u_{t,m} \right)^2 \sum_{t=1}^{T} \| \hat{\mathbf{u}}_t - \mathbf{u}_t \|^2 \| \mathbf{b} \|^2 \, dW(\mathbf{b}) = O(h/T), \tag{21}$$

where we have used the fact that $E T_j^{-1} (\sum_{t=j+1}^{T} u_{t,m})^2 = O(1)$. □

PROOF OF LEMMA A.5. By a second-order Taylor series expansion, we have

$$\sum_{j=1}^{T-1} k^2(j/h) T_j \int \left| \hat{B}_{5j,m}(\mathbf{b}) \right|^2 dW(\mathbf{b})$$

$$\leq 8 \| (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \|^2 \sum_{j=1}^{T-1} a_T(j) \int \left\| \sum_{t=j+1}^{T} \psi_{t-j}(\mathbf{b}) \left[\frac{\partial u_{t,m}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right] \right\|^2 dW(\mathbf{b})$$

$$+ 8 \| (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \|^4 \left[\sum_{t=1}^{T} \sup_{\theta \in \Theta} \left\| \frac{\partial^2 u_{tm}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \, \partial \boldsymbol{\theta}'} \right\| \right]^2 \sum_{j=1}^{T-1} a_T(j) \int dW(\mathbf{b}) \tag{22}$$

$$+ 8 \left[\sum_{t=1}^{T} |\hat{u}_{t,m} - u_{t,m}(\hat{\boldsymbol{\theta}})| \right]^2 \sum_{j=1}^{T-1} a_T(j) \int dW(\mathbf{b})$$

$$= O_p(1) + O_P(h/T) + O_P(h/T),$$

where we have used (17), (19), and (20). □

The proof of Lemma A.6 is analogous to that of Lemma A.4.

PROOF OF PROPOSITION A.2. Given the decomposition in (13), we have

$$\left|\left[\hat{\gamma}_{j,m}^1(\mathbf{b}) - \tilde{\gamma}_{j,m}^1(\mathbf{b})\right]\tilde{\gamma}_{j,m}^1(\mathbf{b})^*\right| \le \sum_{a=1}^6 \left|\hat{B}_{aj,m}(\mathbf{b})\right|\left|\tilde{\gamma}_{j,m}(\mathbf{b})\right|,\tag{23}$$

where the $\hat{B}_{aj,m}(\mathbf{b})$ are defined in (13).

We first consider $a = 5$. By the triangular inequality, we have

$$\sum_{j=1}^{T-1} k^2(j/h)T_j \int \left|\hat{B}_{5j,m}(\mathbf{b})\right|\left|\tilde{\gamma}_{j,m}^1(\mathbf{b})\right| dW(\mathbf{b})$$

$$\le C\left[\sum_{t=1}^T |\hat{u}_{t,m} - u_{t,m}(\hat{\boldsymbol{\theta}})|\right]\sum_{j=1}^{T-1} k^2(j/h)\int |\tilde{\gamma}_{j,m}^1(\mathbf{b})| dW(\mathbf{b})$$

$$+ C\|(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\|\sum_{j=1}^{T-1} k^2(j/h)$$

$$\times \int \left\|T_j^{-1}\sum_{t=j+1}^T \psi_{t-j}(\mathbf{b})\left[\frac{\partial u_{t,m}(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right]'\right\||\tilde{\gamma}_{j,m}^1(\mathbf{b})| dW(\mathbf{b})$$

$$+ \left[\sum_{t=1}^T \|A_t\|\right]\sum_{j=1}^{T-1} k^2(j/h)\int |\tilde{\gamma}_{j,m}^1(\mathbf{b})| dW(\mathbf{b})$$

$$= O_P(h/T^{1/2}) + O_P(1 + h/T^{1/2}) + O_P(h/T^{1/2}) = o_P(h^{1/2}).$$

For $a = 1, 2, 3, 4, 6$, we have, by the Cauchy–Schwarz inequality,

$$\sum_{j=1}^{T-1} k^2(j/h)T_j \int \left|\hat{B}_{aj,m}(\mathbf{b})\right|\left|\tilde{\gamma}_{j,m}^1(\mathbf{b})\right| dW(\mathbf{b})$$

$$\le \left[\sum_{j=1}^{T-1} k^2(j/h)T_j \int \left|\hat{B}_{aj,m}(\mathbf{b})\right|^2 dW(\mathbf{b})\right]^{\frac{1}{2}}$$

$$\times \left[\sum_{j=1}^{T-1} k^2(j/h)T_j \int |\tilde{\gamma}_{j,m}^1(\mathbf{b})|^2 dW(\mathbf{b})\right]^{\frac{1}{2}}$$

$$= O_P(h^{\frac{1}{2}}/T^{\frac{1}{2}})O_P(h^{\frac{1}{2}}) = o_P(h^{\frac{1}{2}}),$$

given Lemmas A.1–A.4 and A.6.   □

PROOF OF THEOREM A.4. Let $q = p^{1+\frac{1}{4b-2}}(\ln^2 T)^{\frac{1}{2b-1}}$. We shall show Propositions A.3 and A.4 below.

PROPOSITION A.3. *Under the conditions of Theorem 2,*

$$h^{-\frac{1}{2}} \sum_{j=1}^{T-1} k^2(j/h) T_j \iint \left\| \tilde{\gamma}_j^1(\mathbf{b}) \right\|^2 dW(\mathbf{b}) = p^{-1/2} \tilde{C} + p^{-1/2} \tilde{V}_q + o_P(1),$$

where $\tilde{C} = \sum_{m=1}^{d} \tilde{C}_m$, $\tilde{V}_q = \sum_{m=1}^{d} \tilde{V}_{q,m}$,

$$\tilde{C}_m = \sum_{j=1}^{T-1} k^2(j/h) T_j^{-1} \sum_{t=j+1}^{T} u_{t,m}^2 \int \left| \psi_{t-j}(\mathbf{b}) \right|^2 dW(\mathbf{b})$$

and

$$\tilde{V}_{q,m} = \sum_{t=2q+2}^{T} u_{t,m} \sum_{j=1}^{q} k^2(j/h) T_j^{-1} \int \psi_{t-j}(\mathbf{b}) \left[\sum_{s=1}^{t-2q-1} u_{s,m} \psi_{s-j}^*(\mathbf{b}) \right] dW(\mathbf{b}).$$

PROPOSITION A.4. *Under the conditions of Theorem 2, $[\tilde{D}^1(h)]^{-1/2} \tilde{V}_q \xrightarrow{d} N(0,1)$.*

$\square$

PROOF OF PROPOSITION A.3. We note that

$$h^{-\frac{1}{2}} \int \sum_{j=1}^{T-1} k^2(j/h) T_j \left\| \tilde{\gamma}_j^1(\mathbf{b}) \right\|^2 dW(\mathbf{b})$$

$$= h^{-\frac{1}{2}} \sum_{m=1}^{d} \int \sum_{j=1}^{T-1} k^2(j/h) T_j \left| \tilde{\gamma}_{j,m}^1(\mathbf{b}) \right|^2 dW(\mathbf{b}).$$

Hence it suffices to show that

$$h^{-\frac{1}{2}} \sum_{m=1}^{d} \int \sum_{j=1}^{T-1} k^2(j/h) T_j \left| \tilde{\gamma}_{j,m}^1(\mathbf{b}) \right|^2 dW(\mathbf{b}) \tag{24}$$

$$= p^{-1/2} \tilde{C}_m + p^{-1/2} \tilde{V}_{q,m} + o_P(1).$$

The proof of (24) is similar to that of Proposition A.3 in Chen and Hong (2011) and is omitted for space. $\square$

PROOF OF PROPOSITION A.4. We rewrite $\tilde{V}_q = \sum_{t=2q+2}^{T} V_q(t)$, where

$$V_q(t) = \sum_{m=1}^{d} V_{q,m}(t),$$

$$V_{q,m}(t) = u_{t,m} \sum_{j=1}^{q} \int a_T(j) \psi_{t-j}(\mathbf{b}) \mathbf{H}_{j,t-2q-1,m}(\mathbf{b}) dW(\mathbf{b}),$$

and $\mathbf{H}_{j,t-2q-1,m}(\mathbf{b}) = \sum_{s=j+1}^{t-2q-1} u_{s,m} \psi_{s-j}^*(\mathbf{b})$. Following Chen and Hong (2011), we apply Brown's (1971) martingale limit theorem, which states $\mathrm{var}(2\,\mathrm{Re}\,\tilde{V}_q)^{-\frac{1}{2}} 2\,\mathrm{Re}\,\tilde{V}_q \overset{d}{\to} N(0,1)$ if

$$\mathrm{var}(2\,\mathrm{Re}\,\tilde{V}_q)^{-1} \sum_{t=2q+2}^{T} \left[2\,\mathrm{Re}\,V_q(t)\right]^2 \mathbf{1}\!\left[|2\,\mathrm{Re}\,V_q(t)| > \eta \cdot \mathrm{var}(2\,\mathrm{Re}\,\tilde{V}_q)^{\frac{1}{2}}\right] \to 0 \quad \forall \eta > 0,$$

$$\mathrm{var}(2\,\mathrm{Re}\,\tilde{V}_q)^{-1} \sum_{t=2q+2}^{T} E\!\left\{\left[2\,\mathrm{Re}\,V_q(t)\right]^2 | \mathcal{I}_{t-1}\right\} \overset{p}{\to} 1.$$

First, we compute $\mathrm{var}(2\,\mathrm{Re}\,\tilde{V}_q)$. By the mds property of $u_{t,m}$ under $\mathbb{H}_0$, we have

$$E(\tilde{V}_q^2) = \sum_{t=2q+2}^{T} E\left[\sum_{m=1}^{d} u_{t,m} \sum_{j=1}^{q} \int a_T(j)\psi_{t-j}(\mathbf{b}) \sum_{s=j+1}^{t-2q-1} u_{s,m}\psi_{s-j}^*(\mathbf{b})\, dW(\mathbf{b})\right]^2$$

$$= \sum_{j=1}^{q}\sum_{l=1}^{q}\sum_{m=1}^{d}\sum_{n=1}^{d} a_T(j)a_T(l)$$

$$\times \int\int \sum_{t=2q+2}^{T}\sum_{s=j+1}^{t-2q-1} E\big[u_{t,m}u_{t,n}\psi_{t-j}(\mathbf{b}_1)\psi_{t-l}(\mathbf{b}_2)\big]$$

$$\times E\big[u_{s,m}u_{s,n}\psi_{s-j}^*(\mathbf{b}_1)\psi_{s-l}^*(\mathbf{b}_2)\big]\, dW(\mathbf{b}_1)\, dW(\mathbf{b}_2)$$

$$+ \sum_{j=1}^{q}\sum_{l=1}^{q}\sum_{m=1}^{d}\sum_{n=1}^{d} a_T(j)a_T(l)$$

$$\times \int\int \sum_{t=2q+2}^{T}\sum_{s=j+1}^{t-2q-1} \mathrm{cov}\big[u_{t,m}u_{t,n}\psi_{t-j}(\mathbf{b}_1)\psi_{t-l}(\mathbf{b}_2),$$

$$u_{s,m}u_{s,n}\psi_{s-j}^*(\mathbf{b}_1)\psi_{s-l}^*(\mathbf{b}_2)\big]\, dW(\mathbf{b}_1)\, dW(\mathbf{b}_2)$$

$$+ 2\sum_{j=1}^{q}\sum_{m=1}^{d}\sum_{n=1}^{d} a_T^2(j)$$

$$\times \int\int \sum_{t=2q+2}^{T}\sum_{s_1=j+1}^{t-2q-1}\sum_{s_2=j+1}^{s_1-1} E\big[u_{t,m}u_{t,n}\psi_{t-j}(\mathbf{b}_1)\psi_{t-j}(\mathbf{b}_2)$$

$$\times u_{s_1,m}u_{s_2,n}\psi_{s_1-j}^*(\mathbf{b}_1)\psi_{s_2-j}^*(\mathbf{b}_2)\big]\, dW(\mathbf{b}_1)\, dW(\mathbf{b}_2)$$

$$+ 4\sum_{j=2}^{q}\sum_{l=1}^{j-1}\sum_{m=1}^{d}\sum_{n=1}^{d} a_T(j)a_T(l)$$

$$\times \int\int \sum_{t=2q+2}^{T}\sum_{s_1=j+1}^{t-2q-1}\sum_{s_2=l+1}^{s_1-1} E\big[u_{t,m}u_{t,n}\psi_{t-j}(\mathbf{b}_1)\psi_{t-l}(\mathbf{b}_2)$$

$$\times\, u_{s_1,m} u_{s_2,n} \psi_{s_1-j}^*(\mathbf{b}_1) \psi_{s_2-l}^*(\mathbf{b}_2) \big] \, dW(\mathbf{b}_1) \, dW(\mathbf{b}_2)$$

$$= \frac{1}{2} \sum_{j=1}^{q} \sum_{l=1}^{q} \sum_{m=1}^{d} \sum_{n=1}^{d} k^2(j/p) k^2(l/p)$$

$$\times \int \int \big| E\big[u_{j+l,m} u_{j+l,n} \psi_l(\mathbf{b}_1) \psi_j(\mathbf{b}_2)\big]\big|^2 \, dW(\mathbf{b}_1) \, dW(\mathbf{b}_2) + o_P(h).$$

Similarly, we can obtain

$$E\big(\tilde{V}_q^*\big)^2 = E|\tilde{V}_q|^2$$

$$= \frac{1}{2} \sum_{j=1}^{q} \sum_{l=1}^{q} \sum_{m=1}^{d} \sum_{n=1}^{d} k^2(j/p) k^2(l/p)$$

$$\times \int \int \big| E\big[u_{j+l,m} u_{j+l,n} \psi_l(\mathbf{b}_1) \psi_j(\mathbf{b}_2)\big]\big|^2 \, dW(\mathbf{b}_1) \, dW(\mathbf{b}_2) \big[1 + o(1)\big].$$

Hence,

$$\mathrm{var}(2\,\mathrm{Re}\,\tilde{V}_q) = E\big(\tilde{V}_q^2\big) + E\big(\tilde{V}_q^*\big)^2 + 2E|\tilde{V}_q|^2$$

$$= 2 \sum_{j=1}^{q} \sum_{l=1}^{q} \sum_{m=1}^{d} \sum_{n=1}^{d} k^2(j/p) k^2(l/p)$$

$$\times \int \int \big| E\big[u_{j+l,m} u_{j+l,n} \psi_l(\mathbf{b}_1) \psi_j(\mathbf{b}_2)\big]\big|^2 \, dW(\mathbf{b}_1) \, dW(\mathbf{b}_2) \big[1 + o(1)\big].$$

The remaining proof of Proposition A.4 is similar to that of Proposition A.4 in Chen and Hong (2011). □

□

PROOF OF THEOREM 3. The proof of Theorem 3 consists of the proofs of Theorems A.5 and A.6 below. □

THEOREM A.5. *Under the conditions of Theorem 3, $(h^{\frac{1}{2}}/T)[\hat{Q}^1(h) - \tilde{Q}^1(h)] \xrightarrow{p} 0$.*

THEOREM A.6. *Under the conditions of Theorem 3,*

$$\frac{h^{\frac{1}{2}}}{T} \hat{Q}^1(h) \xrightarrow{p} \frac{1}{\sqrt{D^1}} \int \int_{-\pi}^{\pi} \big\| \hat{f}^1(\lambda, \mathbf{b}) - \hat{f}_0^1(\lambda, \mathbf{b}) \big\|^2 \, d\lambda \, dW(\mathbf{b}).$$

PROOF OF THEOREM A.5. It suffices to show that

$$T^{-1} \int \sum_{j=1}^{T-1} k^2(j/h) T_j \big[\big\| \hat{\gamma}_{j,m}(\mathbf{b}) \big\|^2 - \big\| \tilde{\gamma}_{j,m}(\mathbf{b}) \big\|^2 \big] \, dW(\mathbf{b}) \xrightarrow{p} 0, \tag{25}$$

$$m = 1, \ldots, d,$$

$\hat{C}^1(h) - \tilde{C}^1(h) = O_P(1)$, and $\hat{D}^1(h) - \tilde{D}^1(h) = o_P(1)$. We focus on the proof of (25) as the proofs for $\hat{C}^1(h) - \tilde{C}^1(h) = O_P(1)$, and $\hat{D}^1(h) - \tilde{D}^1(h) = o_P(1)$ are straightforward. From (13), the Cauchy–Schwarz inequality, and the fact that

$$T^{-1} \int \sum_{j=1}^{T-1} k^2(j/h) T_j \|\tilde{\gamma}_{j,m}(\mathbf{b})\|^2 \, dW(\mathbf{b}) = O_P(1)$$

as is implied by Theorem A.6 (the proof of Theorem A.6 does not depend on Theorem A.5), it suffices to show that $T^{-1}\hat{A}_1 \xrightarrow{P} 0$, where $\hat{A}_1$ is defined as in (12). Given (13), we shall show that

$$T^{-1} \int \sum_{j=1}^{T-1} k^2(j/h) T_j |\hat{B}_{aj,m}(\mathbf{b})|^2 \, dW(\mathbf{b}) \xrightarrow{P} 0,$$

$a = 1, \ldots, 6$, and $m = 1, \ldots, d$. We first consider $a = 1$. By the Cauchy–Schwarz inequality, we have

$$T^{-1} \int \sum_{j=1}^{T-1} k^2(j/h) T_j |\hat{B}_{1j,m}(\mathbf{b})|^2 \, dW(\mathbf{b})$$

$$\leq \left[\sum_{t=1}^{T} (\hat{u}_{t,m} - u_{t,m})^2 \right] \sum_{j=1}^{T-1} a_T(j) \int \left[T_j^{-1} \sum_{t=j+1}^{T} |\hat{\delta}_{t-j}(\mathbf{b})|^2 \right] dW(\mathbf{b}) = O_P(h/T),$$

where we have used the fact that $|\hat{\delta}_{t-j}(\mathbf{b})| \leq 2$. The proof for $a = 2$ is similar.

For $a = 3$, we have

$$T^{-1} \int \sum_{j=1}^{T-1} k^2(j/h) T_j |\hat{B}_{3j,m}(\mathbf{b})|^2 \, dW(\mathbf{b})$$

$$\leq \left(T^{-1} \sum_{t=1}^{T} u_{t,m}^2 \right) \sum_{j=1}^{T-1} k^2(j/h) \int \left[T_j^{-1} \sum_{t=j+1}^{T} |\hat{\delta}_{t-j}(\mathbf{b})|^2 \right] dW(\mathbf{b})$$

$$\leq \left(T^{-1} \sum_{t=1}^{T} u_{t,m}^2 \right) \left[\sum_{t=1}^{T} \|\hat{u}_t - u_t\|^2 \right] \sum_{j=1}^{T-1} a_T(j) \int \|\mathbf{b}\|^2 \, dW(\mathbf{b}) = O_P(h/T).$$

The proof for $a = 4, 5, 6$ is similar to that for $a = 3$. This completes the proof for Theorem A.5.   □

The proof of Theorem A.6 is a straightforward extension for that of Hong (1999, Proof of Theorem 5), for the case $(m, l) = (1, 0)$.

REFERENCES

Bollerslev, T., R. F. Engle, and J. M. Wooldridge (1988), "A capital asset pricing model with time varying covariances." *Journal of Political Economy*, 96, 116–131. [2]

Boubacar Mainassara, Y. and C. Francq (2011), "Estimating structural VARMA models with uncorrelated but non-independent error terms." *Journal of Multivariate Analysis*, 102, 496–505. [13]

Brown, B. M. (1971), "Martingale central limit theorems." *The Annals of Mathematical Statistics*, 42, 59–66. [18]

Chen, B. and Y. Hong (2011), "Generalized spectral testing for multivariate continuous-time models." *Journal of Econometrics*, 164, 268–293. [17, 18, 19]

Forni, M. and L. Gabetti (2014), "Sufficient information in structural VARs." *Journal of Monetary Economics*, 66, 124–136. [3]

Hong, Y. (1999), "Hypothesis testing in time series via the empirical characteristic function: A generalized spectral density approach." *Journal of the American Statistical Association*, 94, 1201–1220. [13, 20]

Kilian, L. (2001), "Impulse response analysis in vector autoregressions with unknown lag order." *Journal of Forecasting*, 20, 161–179. [1]

Leeper, E. M., T. B. Walker, and S.-C. S. Yang (2013), "Fiscal foresight and information flows." *Econometrica*, 81, 1115–1145. [3]

Lippi, M. and L. Reichlin (1994), "VAR analysis, non-fundamental representations, Blaschke matrices." *Journal of Econometrics*, 63, 307–325. [7, 10]

Rosenblatt, M. (2000), *Gaussian and Non-Gaussian Linear Time Series and Random Fields*. Springer-Verlag, New York. [6, 7]