

 Open access • Journal Article • DOI:10.1111/JTSA.12271

## Inference on multivariate heteroscedastic time varying random coefficient models

— [Source link](#) 

Liudas Giraitis, George Kapetanios, Tony Yates

**Institutions:** Queen Mary University of London, King's College London, University of Birmingham

**Published on:** 01 Mar 2018 - Journal of Time Series Analysis (John Wiley & Sons, Ltd)

**Topics:** Heteroscedasticity, Autoregressive model, Conditional variance, Estimator and Kernel density estimation

Related papers:

- [Inference on stochastic time-varying coefficient models](#)
- [Time Varying Structural Vector Autoregressions and Monetary Policy](#)
- [Inference of time-varying regression models](#)
- [Semiparametric Spatial Autoregressive Panel Data Model with Fixed Effects and Time-Varying Coefficients](#)
- [Maximum Likelihood Estimation for Generalized Autoregressive Score Models](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/inference-on-multivariate-heteroscedastic-time-varying-14g8511wd1>

# Inference on multivariate heteroscedastic time varying random coefficient models

L. Giraitis\*

*Queen Mary, University of London*

G. Kapetanios

*King's College, London*

T. Yates

*University of Birmingham*

July 5, 2017

## Abstract

In this paper we introduce the general setting of a multivariate time series autoregressive model with stochastic time-varying coefficients and time-varying conditional variance of the error process. This allows modeling VAR dynamics for non-stationary times series and estimation of time varying parameter processes by well-known rolling regression estimation techniques. We establish consistency, convergence rates and asymptotic normality for kernel estimators of the paths of coefficient processes and provide pointwise valid standard errors. The method is applied to a popular 7 variable data set to analyze evidence of time-variation in empirical objects of interest for the DSGE literature.

## 1 Introduction

This paper considers a general *multivariate* VAR model for non-stationary processes generated by stochastic coefficients that evolve as bounded persistent processes, such as, e.g., bounded random walks. In addition, it allows for potentially time-varying

---

\*Giraitis' research is supported by the ESRC grant RES062230790. We are grateful for comments from Fabio Canova which acted as the spur for us to explain the practical benefits of the kernel estimator more fully. We thank Kostas Theodoridis for making available his update to the Smets-Wouters dataset.

volatilities of disturbances. We establish consistency rates and asymptotic normality of kernel estimates of the paths of coefficient processes and the volatility process of the disturbances, and supplement theoretical results with Monte Carlo evidence. Finally, we use our estimation methods to characterise the dynamic evolution of the dataset used originally by Smets and Wouters (2007). We investigate changes in a variety of features of the multivariate model. These include the impact on hours worked of a technology shock, the impact on output of a monetary policy shock, and the predictability of inflation. These three objects are at the centre of a number of debates in recent work in empirical macroeconomics.

Our work extends in various important directions the contribution of Giraitis, Kapetanios, and Yates (2014). The broad purpose of the focus of Giraitis, Kapetanios, and Yates (2014) and the current paper on kernel estimation methods for stochastically-varying coefficient autoregressive models stems from the exclusive focus of the relevant literature on a Bayesian approach to the same econometric problem. Papers by Cogley and Sargent (2005) and others have popularised estimators that use the Kalman filter embedded within Gibbs sampling algorithms. Stochastic time varying coefficient models have been deployed in the study of many topics in empirical macroeconomics including: changes in inflation persistence over time, changes in the persistence of the real exchange rate, estimating the contribution of good luck versus good policy to the reduction in macroeconomic volatility during the great moderation, and changes in the response of hours worked to technology shocks.

Time varying models with deterministic parameters, known in the literature as locally stationary processes, were introduced by Priestley (1965) and Dahlhaus (1997) and extensively studied in the literature, see the comprehensive review by Dahlhaus (2012). It is well known that such parameters can be estimated using kernel estimation methods. Our class of models allows parameters to vary as persistent random processes which enables modeling non-stationary processes. There is also a large liter-

ature on existence of stationary random parameter, threshold and Markov Switching AR models, see Bougerol and Picard (1992), Cline (2007) and Francq and Zakoan (2001). For estimation and testing of structural change in time series regression models, see Kristensen (2012). Finally, the early work of Nicholls and Quinn (1982) as well as the recent work of Alj, Azrak, Ley, and Melard (2016) is of interest.

Our kernel estimation approach applied in this setting has a number of attractions as an alternative tool. First, comparable theoretical results on consistency and rates of convergence are not available for existing estimators, under the assumption that parameter processes follow a bounded random walk. Although most applications of the MCMC algorithms are superficially Bayesian, many of them stress the use of uninformative priors where possible. So, the theoretical results of this paper are not of mere academic interest. Second, our kernel estimator avoids the ‘pile up’ problem that MCMC methods are known to be prone to, which leads to an overstatement of the probability that parameters do not change, as documented by Stock and Watson (1998). Third, the kernel estimates are very fast to compute, taking seconds to produce the estimates for a VAR of large dimension, rather than days for the Bayesian competitor, with the computational gains increasing as the dimension of the VAR increases. At some point, with VARs of dimension 5 or 6, the alternative algorithm becomes entirely intractable. The difficulty of Bayesian methods stems from the need to admit only those paths of VAR coefficient processes which imply point-wise instantaneous stationarity, i.e. satisfy restriction of a bounded random walk. As the dimension of the VAR model increases, the law of large numbers makes it even harder to find draws of parameter processes that satisfy this bounding constraint. Typically, the sampling algorithm draws parameter path in one go, then accepting or rejecting as appropriate. The problem is aggravated by the fact that there is considerable persistence in many macro time series, so with the increase of VAR dimension it becomes more and more likely to find at least one time period for which restriction

condition fails. Koop and Potter (2011) modify the method so that parameters for each period are drawn separately and, therefore, rejection does not mean discarding the satisfactory draws in the parameter process accumulated up to that point.

Our kernel estimation method does not suffer from this problem. It produces a single path of point estimates directly. This provides the option of assessing ex post whether the point estimate of the coefficient process satisfies the bounding condition for the whole path. If it does, then the estimation is completed. If not, (in fact regardless of whether it does or not), our estimates can be used as the input into some subsequent Bayesian kind procedure in which the prior that, the bounding condition should hold, is imposed.

A sceptical reader might accept that our kernel estimates can handle large dimension time varying VARs, but nevertheless wonder whether we really need them? Do we forgo anything of import by confining ourselves to studying smaller dimension systems? Our application is an attempt to convince for the need of large dimension VARs. The 7 variable US data set we consider was originally used by Smets and Wouters (2007) to estimate a medium scale DSGE model with various frictions in price and wage-setting, consumption and investment. That model is widely cited and has given rise to similar models used in many central banks. It was predicated on the idea that the dynamics, the models' structural features were wired to capture, were not themselves subject to significant time variation within the sample itself. As such it is a useful laboratory to look for time-variation. For comparison, to assess the benefit got by focusing on the larger dimension VAR, we estimate a smaller, 4 variable system. We identify monetary policy and technology shocks using sign restrictions, and look for time-variation in the impulse responses to these shocks. We also compute multivariate measures of inflation predictability and assess how this has changed over the sample period.

The application turns up results that are interesting in their own right. We uncover

pronounced shifts in the impulse response of real variables like output and hours to an identified monetary policy shock that imply changes in the extent of nominal rigidity in the economy. We compute that the impulse response of real wages to a monetary policy shock has shifted too: it begins our sample period mild and negative, but in later periods we see a strong positive response. These results can loosely be interpreted as suggesting that wages became more sticky relative to prices. For example, following a contractionary monetary policy shock that leads to a recession, the more sticky wages are relative to prices, the more likely real wages are to increase. We find that the impulse response of hours worked to a technology shock was initially mild and negative, but becomes steadily larger and positive through to the present day, implying that the economy (in this regard at least) more closely resembles a frictionless real business cycle model at the end of the sample than at the beginning. Inflation predictability shows some pronounced rises and falls in our sample period, but there is no clear tendency for it to be less predictable in the 80's as seems to be the consensus in the literature.

Along some dimensions, our smaller, 4-variable system does a good job of characterising these dynamics, but along others it does not. For example, shifts in the impulse response of output and hours worked to a monetary policy shock look very similar whether through the lens of the 7 or 4 variable system. However, the 4 variable system gives a very different read on the response of hours worked to a technology shock: here there is no tendency for this impulse response to become more positive over time, so in that respect the smaller system fails to adequately describe the time-variation evident in the larger system. Also, we find that the 4 variable system overstates the fall in inflation predictability in the 1990s, obscuring the essential continuity in inflation dynamics that the larger, 7-variable VAR system finds.

Our theoretical results extend well beyond the results of Giraitis, Kapetanios, and Yates (2014) by allowing more general coefficient processes and providing more refined

multivariate version of convergence results, on top of the extension to multivariate heteroscedastic models. We evaluate the performance of our extended estimator via Monte Carlo analysis and, crucially, illustrate in detail the extended scope of our modelling toolkit by analysing a well known dataset using a realistic macroeconomic model.

The rest of the paper is structured as follows. Section 2 presents our setup and theoretical results. Sections 3 and 4 present our Monte Carlo and empirical evidence, while Section 5 concludes. Proofs are presented in the Appendix.

## 2 Theoretical considerations

This paper considers two major extensions compared to the work of Giraitis, Kapetanios, and Yates (2014). The first extension concerns the setup and estimation of a multivariate autoregressive model with time varying stochastic coefficients while the second centers on estimation of paths of the process of the time varying conditional variance of disturbances. We start by considering two versions of a multivariate dynamic autoregressive model given by

$$\mathbf{y}_t = \mathbf{\Psi}_{t-1}\mathbf{y}_{t-1} + \mathbf{u}_t, \quad t = 1, 2, \dots, n, \quad (2.1)$$

and

$$\mathbf{y}_t = \boldsymbol{\alpha}_t + \mathbf{\Psi}_{t-1}\mathbf{y}_{t-1} + \mathbf{u}_t, \quad (2.2)$$

where  $\mathbf{y}_t = (y_{1t}, \dots, y_{mt})'$ , the noise  $\mathbf{u}_t = (u_{1t}, \dots, u_{mt})'$  and  $\boldsymbol{\alpha}_t = (\alpha_{1t}, \dots, \alpha_{mt})'$  are  $m$ -dimensional vectors, and  $\mathbf{\Psi}_t = [\psi_{t,ij}]$  is  $m \times m$  matrix of (random) coefficient processes while  $E\mathbf{u}_t\mathbf{u}_s' = \mathbf{0}$ ,  $t \neq s$ . Alternatively, the model

$$\mathbf{y}_t = \boldsymbol{\mu}_t + \mathbf{\Psi}_{t-1}(\mathbf{y}_{t-1} - \boldsymbol{\mu}_{t-1}) + \mathbf{u}_t \quad (2.3)$$

aims to describe the VAR dynamic of a non-stationary time series  $\mathbf{y}_t$  around some persistent possibly non-stationary process  $\boldsymbol{\mu}_t$  (attractor), see (2.19). If VAR(1) struc-

ture is absent, the attractor  $\boldsymbol{\mu}_t$  collapses to  $\boldsymbol{y}_t$ , differently from stationary VAR(1) models where the constant  $\boldsymbol{\mu} = \boldsymbol{\mu}_t$  plays the role of the mean.

To ensure that this dynamic model generates a bounded process  $\boldsymbol{y}_t$  and to enable estimation of the model, it is important to bound the spectral norm,  $\|\boldsymbol{\Psi}_t\|_{sp} := \sup_{\|x\|=1} \|\boldsymbol{\Psi}_t x\|$ , of  $\boldsymbol{\Psi}_t$ , above by one, i.e.  $\|\boldsymbol{\Psi}_t\|_{sp} < 1$ . Here,  $\|\cdot\|$  denotes the Euclidean norm. There are a variety of ways to implement such a bounding, see for examples section 2.4. We assume that  $\boldsymbol{\Psi}_t$  has the following smoothness properties.

**Assumption 2.1.** *The random coefficients  $\boldsymbol{\Psi}_t$  are such that  $\|\boldsymbol{\Psi}_t\|_{sp} \leq r < 1$ ,  $t \geq 0$ . Moreover, as  $h \rightarrow \infty$ ,  $h = o(t)$ ,  $t \rightarrow \infty$ ,*

$$\sup_{s:|s-t|\leq h} \|\boldsymbol{\Psi}_t - \boldsymbol{\Psi}_s\|_{sp}^2 = O_p(h/t). \quad (2.4)$$

The above local stability assumption enables consistent estimation of  $\boldsymbol{\Psi}_t$ . Specific examples of  $\boldsymbol{\Psi}_t$  that are standardized unit root processes are given in Subsection 2.4. Furthermore, this condition could be generalised to

$$\sup_{s:|s-t|\leq h} \|\boldsymbol{\Psi}_t - \boldsymbol{\Psi}_s\|_{sp}^2 = O_p((h/t)^\gamma).$$

for some  $0 < \gamma < 2$  at the cost of further mathematical complexity. It is possible that the assumption  $\|\boldsymbol{\Psi}_t\|_{sp} < 1$  can be modified in a variety of ways such as, e.g., by adopting some nonstationary version of a negative Lyapunov exponent (as in Brandt (1986)), although we do not pursue such extensions in the paper.

In VAR modeling of non-stationary  $\boldsymbol{y}_t$ 's with persistent (non-stationary) parameters  $\boldsymbol{\Psi}_t$ ,  $\boldsymbol{\alpha}_t$ , attractor  $\boldsymbol{\mu}_t$  and volatility process  $\boldsymbol{H}_t$  introduced below, smoothness restriction has to be sufficiently weak, comparing to locally stationary models with deterministic parameters, see Dahlhaus (2012), which by their nature are closer to stationary processes and where  $\gamma = 2$ . The leading example of an allowable parameter process given by  $\frac{a_t}{\max_{0 \leq s \leq t} |a_s|}$  where  $a_t$  is a random walk is instructive. Its limit is heuristically a standardised Brownian motion which provides the required degree of 'smoothness' enabling consistent estimation.



The second extension of the paper allows for a martingale difference noise given by

$$\mathbf{u}_t = \mathbf{H}_{t-1}\boldsymbol{\varepsilon}_t, \quad E[\mathbf{u}_t|\mathcal{F}_{t-1}] = \mathbf{0} \quad (2.5)$$

with respect to natural filtration  $\mathcal{F}_t$ , where  $\mathbf{H}_t = \{h_{t,ij}\}$  is a  $m \times m$  time varying random volatility process, and  $\boldsymbol{\varepsilon}_t$  is a vector-valued standardized i.i.d. noise,  $E\boldsymbol{\varepsilon}_t = \mathbf{0}$ ,  $E\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t' = \mathbf{I}$ . In models (2.2)-(2.1) we set  $\mathcal{F}_t = \sigma(\boldsymbol{\varepsilon}_s, \mathbf{H}_s, \boldsymbol{\Psi}_s, \boldsymbol{\alpha}_s, s \leq t)$  while in (2.3) we define  $\mathcal{F}_t = \sigma(\boldsymbol{\varepsilon}_s, \mathbf{H}_s, \boldsymbol{\Psi}_s, \boldsymbol{\mu}_s, s \leq t)$ , respectively. This setting does not require an assumption of mutual independence of  $(\boldsymbol{\varepsilon}_s)$  and parameter processes  $(\boldsymbol{\Psi}_s)$ ,  $(\boldsymbol{\alpha}_s)$ ,  $(\boldsymbol{\mu}_s)$ .

Denote by  $\boldsymbol{\Sigma}_t = \mathbf{H}_{t-1}\mathbf{H}_{t-1}' = E[\mathbf{u}_t\mathbf{u}_t'|\mathcal{F}_{t-1}]$  the conditional variance-covariance matrix. We assume the following about  $(\mathbf{u}_t)$ .

**Assumption 2.2.** (i) In  $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, \dots, \varepsilon_{mt})'$ ,  $E\varepsilon_{i1}^4 < \infty$  and  $Ey_{i0}^4 < \infty$  for  $i = 1, \dots, m$ .

(ii) For  $t \geq 0$ ,  $Eh_{t,ij}^4 \leq C$ ; for  $1 \leq k \leq t/2$ ,  $E\|\mathbf{H}_t - \mathbf{H}_{t+k}\|_{sp}^2 \leq Ck/t$ .

(iii)  $\|\mathbf{H}_t^{-1}\|_{sp} = O_p(1)$  as  $t \rightarrow \infty$ .

Assumption 2.2 implies that  $\max_j E\|u_j^4\| < \infty$ , where  $\|\cdot\|$  denotes the Euclidean norm. For examples of  $\mathbf{H}_t$ , see section 2.4. Notice that Assumption 2.2(ii) excludes stationary conditionally heteroscedastic ARCH type volatility processes  $(\mathbf{H}_t)$  used in the volatility literature, but allows handling unconditional heterogeneity and conditional variances evolving as non-stationary persistent random processes commonly used for modelling macroeconomic volatility.<sup>1</sup> We expect that our analysis can easily accommodate weaker assumptions for  $\boldsymbol{\varepsilon}_t$  similar to those made in Giraitis, Kapetanios, and Yates (2014) for the error term. For example,  $\boldsymbol{\varepsilon}_t$  can be a martingale difference process at the expense of more intricate theoretical analysis.

---

<sup>1</sup>Foster and Nelson (1996) developed rolling regression strategies for the estimation of time varying covariances and variances that allow to accommodate ARCH type conditional heteroscedasticity.

The next theorem states structural results. In particular, we show that  $\mathbf{y}_t$  can be written as a moving average of the noise  $\mathbf{u}_j$  with time varying (random) weights  $\Pi_{t,0} := \mathbf{1}$ ,  $\Pi_{t,j} := \Psi_{t-1} \cdots \Psi_{t-j}$ ,  $1 \leq j \leq t$ , and approximated by a truncated VAR(1) process  $\mathbf{z}_t$ , see (2.9). Notice that

$$\|\Pi_{t,j}\|_{sp} \leq \|\Psi_{t-1}\|_{sp} \cdots \|\Psi_{t-j}\|_{sp} \leq r^j, \quad 1 \leq j \leq t. \quad (2.6)$$

Below equation (2.7) contains an expression of  $\mathbf{y}_t$  in terms of the parameter process  $\Psi_j$  and the noise  $\mathbf{u}_j$  while (2.9) establishes a simple approximation of (2.7) by a moving average process as  $t$  increases. This approximation is very useful for proving the theoretical results as it allows use of a linear MA representation in place of the more complex one given by (2.7).

**Theorem 2.1.** *Under Assumption 2.1, the process  $\mathbf{y}_t$  of (2.1) can be written as*

$$\mathbf{y}_t = \sum_{j=0}^{t-1} \Pi_{t,j} \mathbf{u}_{t-j} + \Pi_{t,t} \mathbf{y}_0, \quad t \geq 1, \quad (2.7)$$

$$E\|\mathbf{y}_t\|^4 \leq C, \quad t \geq 0, \quad (2.8)$$

$$\mathbf{y}_t = \mathbf{z}_t + o_p(1), \quad \mathbf{z}_t := \sum_{j=0}^{t-1} \Psi_t^j \mathbf{u}_{t-j}, \quad t \rightarrow \infty. \quad (2.9)$$

Approximation (2.9) can be improved to  $\mathbf{y}_t = \mathbf{z}_t + O_p((\log t)/t)$  as  $t \rightarrow \infty$ .

## 2.1 Estimation of VAR(1) model with no intercept

To estimate the paths  $\Psi_1, \dots, \Psi_n$  and  $\alpha_1, \dots, \alpha_n$  of the coefficient processes in (2.1) from the sample  $\mathbf{y}_1, \dots, \mathbf{y}_n$ , we use the kernel estimate

$$\hat{\Psi}_t := \left( \sum_{j=1}^n k_{tj} \mathbf{y}_j \mathbf{y}'_{j-1} \right) \left( \sum_{j=1}^n k_{tj} \mathbf{y}_{j-1} \mathbf{y}'_{j-1} \right)^{-1},$$

with the weights  $k_{tj} := K((t-j)/H_\psi)$  where  $K(x) \geq 0$ ,  $x \in \mathbb{R}$  is a bounded function and  $H_\psi$  is a bandwidth parameter such that  $H_\psi \rightarrow \infty$ ,  $H_\psi = o(n/\log n)$ . Such estimates are a simple generalisation of a rolling window estimator

$$\hat{\Psi}_t := \left( \sum_{j=t-H_\psi}^{t+H_\psi} \mathbf{y}_j \mathbf{y}'_{j-1} \right) \left( \sum_{j=t-H_\psi}^{t+H_\psi} \mathbf{y}_{j-1} \mathbf{y}'_{j-1} \right)^{-1},$$

which is a local sample correlation of  $y_t$ 's at lag 1. We assume that  $K$  is a non-negative bounded function with a piecewise bounded derivative  $\dot{K}(x)$  such that  $\int K(x)dx = 1$ . For example,

$$\begin{aligned} K(x) &= (1/2)I(|x| \leq 1), & \text{flat kernel,} \\ K(x) &= (3/4)(1 - x^2)I(|x| \leq 1), & \text{Epanechnikov kernel,} \\ K(x) &= (1/\sqrt{2\pi})e^{-x^2/2}, & \text{Gaussian kernel.} \end{aligned}$$

If  $K$  has unbounded support, we assume in addition that

$$K(x) \leq C \exp(-cx^2), \quad |\dot{K}(x)| \leq C(1 + x^2)^{-1}, \quad x \geq 0, \quad \text{for some } C > 0, \quad c > 0. \quad (2.10)$$

To estimate the conditional variance-covariance matrix  $\Sigma_t = \mathbf{H}_{t-1}\mathbf{H}'_{t-1}$ , we use the kernel estimate based on residuals  $\hat{\mathbf{u}}_j = \mathbf{y}_j - \hat{\Psi}_t \mathbf{y}_{j-1}$ ,

$$\Sigma_{\hat{\mathbf{u}},t} = L_t^{-1} \sum_{j=1}^n l_{tj} \hat{\mathbf{u}}_j \hat{\mathbf{u}}_j', \quad \text{where } l_{tj} := L\left(\frac{t-j}{H_h}\right), \quad L_t := \sum_{j=1}^n l_{tj}, \quad (2.11)$$

where  $H_h \rightarrow \infty$ ,  $H_h = o(n)$  is another bandwidth parameter, and the kernel function  $L$  obeys the same restrictions as  $K$ . For a symmetric positive definite matrix  $\mathbf{A}$  we denote by  $\mathbf{A}^{1/2}$  the unique positive definite square root of  $\mathbf{A}$ . If  $\mathbf{H}_{t-1}$  is positive definite then  $\Sigma_t^{1/2} = \mathbf{H}_{t-1}$ . Subsequently,  $\Sigma_{\hat{\mathbf{u}},t}$  is also used to evaluate standard errors in estimation of parameter of VAR model, see Remarks 2.1 and 2.2. Note that we use the slightly cumbersome notation,  $\Sigma_{\hat{\mathbf{u}},t}$ , for our estimator, to emphasise that estimation is based on  $\hat{\mathbf{u}}_j$ , since our proofs will also consider the properties of an infeasible estimator based on  $\mathbf{u}_j$  to derive the properties of  $\Sigma_{\hat{\mathbf{u}},t}$ .

Below we set  $\bar{H}_\psi = H_\psi(\log H_\psi)^{1/2}$  if  $K$  has unbounded support, and  $\bar{H}_\psi = H_\psi$  if  $K$  has bounded support. Similarly we define  $\bar{H}_h$ . Below notation  $a_n \ll b_n$  indicates that  $a_n/b_n \rightarrow 0$  as  $n \rightarrow \infty$ .

Denote  $K_t = \sum_{j=1}^n k_{tj}$ ,  $K_{2,t} = \sum_{j=1}^n k_{tj}^2$ ,  $L_{2,t} = \sum_{j=1}^n l_{tj}^2$  and set

$$\kappa_{n,\psi} := (\bar{H}_\psi/n)^{1/2} + H_\psi^{-1/2}, \quad \kappa_{n,h} := (\bar{H}_h/n)^{1/2} + H_h^{-1/2}. \quad (2.12)$$

Then, the following holds.

**Theorem 2.2.** *Let  $\mathbf{y}_1, \dots, \mathbf{y}_n$  be defined as in (2.1), and  $t = \lceil n\tau \rceil$ , where  $0 < \tau < 1$  is fixed. Suppose that Assumptions 2.1 and 2.2 hold, and  $K$  satisfies (2.10).*

(i) *Then, for  $H_\psi = o(n/\log n)$ ,  $H_h = o(n/\log n)$ ,*

$$\hat{\Psi}_t - \Psi_t = O_p(\kappa_{n,\psi}), \quad (2.13)$$

$$\Sigma_{\hat{u},t} - \Sigma_t = O_p(\kappa_{n,\psi}^2 + \kappa_{n,h}). \quad (2.14)$$

*In particular,  $\kappa_{n,\psi}^2 + \kappa_{n,h} \leq 3\kappa_{n,h}$  if  $H_h^{1/2} \leq H_\psi \leq (H_h n)^{1/2}/\log n$ .*

(ii) *In addition, if  $H_\psi \bar{H}_\psi = o(n)$ , then for any real  $m \times 1$ - vector  $\mathbf{a}$  such that  $\|\mathbf{a}\| = 1$ ,*

$$(K_t/K_{2,t})^{1/2} \mathbf{H}_{t-1}^{-1} (\hat{\Psi}_t - \Psi_t) \left( \sum_{j=1}^n k_{tj} \mathbf{y}_{j-1} \mathbf{y}'_{j-1} \right)^{1/2} \mathbf{a} \rightarrow_D \mathcal{N}(0, \mathbf{I}). \quad (2.15)$$

(iii) *In addition, if  $H_h \bar{H}_h = o(n)$  and  $H_h^{1/2} \ll H_\psi \ll n/(H_h \log n)^{1/2}$ , then*

$$(L_t/L_{2,t}^{1/2}) \mathbf{H}_{t-1}^{-1} (\Sigma_{\hat{u},t} - \Sigma_t) \mathbf{H}'_{t-1} \rightarrow_D \mathbf{Z} \quad (2.16)$$

*where the elements of  $\mathbf{Z} = (z_{ij})_{i,j=1,\dots,m}$  are independent normal variables such that  $z_{ij} \sim N(0, v_{ij}^2)$  where  $v_{ij}^2 = 1$  if  $i \neq j$  and  $v_{ii}^2 = \text{Var}(\varepsilon_{i1}^2)$ .*

**Remark 2.1.** *Theorem 2.2(ii) establishes normal approximation (2.15) for  $m$ - dimensional vector of linear combinations of the elements of the rows of the matrix  $\hat{\Psi}_t - \Psi_t$ , e.g. the vector  $\mathbf{a} = (0, \dots, i, \dots)'$  gives the  $i$ -th column of this matrix. In (2.15)  $\mathbf{H}_{t-1}^{-1}$  can be replaced by  $\Sigma_{\hat{u},t}^{-1/2}$  in view of  $\|\Sigma_{\hat{u},t}^{-1/2} - \Sigma_t^{-1/2}\| = o_p(1)$  of Lemma 6.5 (i). In normal approximation (2.16) for  $\Sigma_{\hat{u},t}$  that can be done if  $\Sigma_t^{-1/2} = \mathbf{H}_{t-1}^{-1}$  which holds if  $\mathbf{H}_{t-1}^{-1}$  is positive definite.*

*In addition, notice that in (2.15),*

$$K_t^{-1} \sum_{j=1}^n k_{tj} \mathbf{y}_{j-1} \mathbf{y}'_{j-1} = V_{\psi,t} + o_p(1), \quad V_{\psi,t} := \sum_{k=0}^{\infty} \Psi_t^k \Sigma_t \Psi_t'^k. \quad (2.17)$$

The use of estimates  $\hat{\Psi}_t$  and  $\Sigma_{\hat{u},t}$  requires choosing the bandwidth parameters  $H_\psi$  and  $H_h$ . While consistency (2.13)-(2.14) holds under minimal restrictions on  $H_\psi$  and  $H_h$ , asymptotic normality results (2.15)- (2.16) require stronger restrictions, e.g. one can set  $H_\psi = o((n/\log n)^{1/2})$  and  $H_h = H_\psi$ . Practical suggestions for bandwidth parameters are given in Section 3, while data driven optimal selection remains an open problem. For that purpose one could develop a leading-order expansion of the bias term and then similarly as in nonparametric density estimation use it together with asymptotic variance to obtain the time varying bandwidth minimising the MSE. The results (2.15)- (2.16) allow flexible choice of kernel, each of which produces a rate of convergence proportional to the square-root of the bandwidth parameter. For practical work it would be of interest to conduct additional study to determine, which kernel, e.g., flat, Gaussian or Epanechnikov, yields the smallest standard error. In our simulations, the Gaussian kernel as a rule outperforms the rolling window. The limiting distribution (2.16) involves the kurtosis of  $\varepsilon_{i1}$  that needs to be estimated.

To highlight the basic statistical properties of a multivariate VAR(1) process with no intercept, the following proposition analyzes the asymptotic behavior of the weighted sample mean  $\bar{\mathbf{y}}_t \equiv K_t^{-1} \sum_{j=1}^n k_{tj} \mathbf{y}_j$ . The latter is shown to be asymptotically negligible,  $\bar{\mathbf{y}}_t = O_p(\kappa_{n,\psi}) = o_p(1)$ . It also satisfies the multivariate Beveridge-Nelson decomposition well-known for univariate stationary linear processes, see Phillips and Solo (1992), and the normal approximation holds. The latter is the same as under deterministic parameters but involves possibly random normalisation.

**Proposition 2.1.** *Under assumptions of Theorem 2.2(i), with  $\kappa_{n,\psi}^* := (\bar{H}_\psi/n)^{1/2} + H_\psi^{-1}$ ,*

$$\bar{\mathbf{y}}_t = (I - \Psi_t)^{-1} \bar{\mathbf{u}}_t + O_p(\kappa_{n,\psi}^*) = (I - \Psi_t)^{-1} \mathbf{H}_{t-1} \bar{\boldsymbol{\varepsilon}}_t + O_p(\kappa_{n,\psi}^*) = O_p(\kappa_{n,\psi}).$$

*In addition, if  $H_\psi \bar{H}_\psi = o(n)$ , then  $(K_t/K_{2,t}^{1/2}) H_{t-1}^{-1} (I - \Psi_t) \bar{\mathbf{y}}_t \rightarrow_D \mathcal{N}(0, \mathbf{I})$ .*

Proposition 2.1 is derived in Lemma 6.4(i).

## 2.2 Estimation of VAR(1) model with a random attractor

Next, we discuss the VAR(1) model  $\mathbf{y}_t$  that includes a persistent (random) term  $\boldsymbol{\mu}_t$ , that in a fixed coefficient VAR model plays the role of the mean. We decompose  $\mathbf{y}_t = \boldsymbol{\mu}_t + (\mathbf{y}_t - \boldsymbol{\mu}_t)$  into a persistent *attractor*  $\boldsymbol{\mu}_t$ , and the term

$$\mathbf{y}_t - \boldsymbol{\mu}_t = \boldsymbol{\Psi}_{t-1}(\mathbf{y}_{t-1} - \boldsymbol{\mu}_{t-1}) + \mathbf{u}_t, \quad t \geq 1 \quad (2.18)$$

which follows the VAR(1) process (2.1) with no intercept, below denoted by  $\hat{\mathbf{y}}_t := \mathbf{y}_t - \boldsymbol{\mu}_t$ . By (2.9),  $\mathbf{y}_t$  satisfies the following moving average approximation:

$$\mathbf{y}_t = \boldsymbol{\mu}_t + \sum_{k=0}^{t-1} \boldsymbol{\Psi}_t^k \mathbf{u}_{t-k} + o_p(1), \quad t \rightarrow \infty. \quad (2.19)$$

This model can also be written as a VAR(1) process  $\mathbf{y}_t = \boldsymbol{\alpha}_t + \boldsymbol{\Psi}_{t-1}\mathbf{y}_{t-1} + \mathbf{u}_t$  with the intercept  $\boldsymbol{\alpha}_t = \boldsymbol{\mu}_t - \boldsymbol{\Psi}_{t-1}\boldsymbol{\mu}_{t-1}$ . Although the attractor  $\boldsymbol{\mu}_t$  can be estimated, in general, it cannot be interpreted as the mean  $E\mathbf{y}_t$ , since it is a random quantity, but should be interpreted as the driving force determining the level of  $\mathbf{y}_t$ . Expression (2.19) shows that (non-stationary) process  $\mathbf{y}_t$  can be decomposed into a persistent non-stationary random attractor  $\boldsymbol{\mu}_t$  and time varying moving-average type term.

We estimate  $\boldsymbol{\mu}_t$ ,  $\boldsymbol{\Psi}_t$  and  $\boldsymbol{\alpha}_t$  by  $\hat{\boldsymbol{\mu}}_t := \bar{\mathbf{y}}_t$ ,

$$\hat{\boldsymbol{\Psi}}_t := \left( \sum_{j=1}^n k_{tj} \hat{\mathbf{y}}_{t,j} \hat{\mathbf{y}}'_{t,j-1} \right) \left( \sum_{j=1}^n k_{tj} \hat{\mathbf{y}}_{t,j-1} \hat{\mathbf{y}}'_{t,j-1} \right)^{-1}, \quad \hat{\boldsymbol{\alpha}}_t = \bar{\mathbf{y}}_t - \hat{\boldsymbol{\Psi}}_t \bar{\mathbf{y}}_t,$$

where  $\hat{\mathbf{y}}_{t,j} := \mathbf{y}_j - \bar{\mathbf{y}}_t$ . To estimate  $\boldsymbol{\Sigma}_t$ , we use the estimate  $\boldsymbol{\Sigma}_{\hat{\mathbf{u}}\hat{\mathbf{u}},t}$  of (2.11) based on residuals  $\hat{\mathbf{u}}_j = \hat{\mathbf{y}}_{t,j} - \hat{\boldsymbol{\Psi}}_t \hat{\mathbf{y}}_{t,j-1}$ .

The following assumption describes a class of permissible attractors  $\boldsymbol{\mu}_t$ .

**Assumption 2.3.**  $\boldsymbol{\mu}_t = (\mu_{1t}, \dots, \mu_{mt})'$  is such that  $\max_t E\mu_{it}^4 < \infty$ ,  $i = 1, \dots, m$  and satisfies either (i) or (ii).

(i)  $E\|\boldsymbol{\mu}_t - \boldsymbol{\mu}_{t+k}\|^2 \leq Ck/t$ ,  $1 \leq k \leq t/2$ .

(ii)  $\boldsymbol{\mu}_t - \boldsymbol{\mu}_{t+k} = \mathbf{m}(t, k) + \tilde{\mathbf{m}}(t, k)$ , where  $E\|\mathbf{m}(t, k)\|^2 \leq C(k/t)$ ,  $1 \leq k \leq h \leq t/2$ , and

$$\max_{1 \leq k \leq h} \|\tilde{\mathbf{m}}(t, k)\| = O_p((h/t)^{1/2} + h^{-1}).$$

The next theorem establishes consistency, convergence rates and asymptotic normality for the estimates.

**Theorem 2.3.** *Let  $\mathbf{y}_1, \dots, \mathbf{y}_n$  be a sample of a VAR(1) model (2.18) with an attractor,  $\boldsymbol{\mu}_t$ , and  $t = \lceil n\tau \rceil$ , where  $0 < \tau < 1$  is fixed. Assume that  $K$  and  $L$  satisfy (2.10), and Assumptions 2.1- 2.3 hold. Then, for  $H_\psi = o(n/\log n)$ ,  $H_h = o(n/\log n)$ ,*

$$\begin{aligned} \hat{\boldsymbol{\mu}}_t - \boldsymbol{\mu}_t &= O_p(\kappa_{n,\psi}), \quad \hat{\boldsymbol{\Psi}}_t - \boldsymbol{\Psi}_t = O_p(\kappa_{n,\psi}), \quad \hat{\boldsymbol{\alpha}}_t - \boldsymbol{\alpha}_t = O_p(\kappa_{n,\psi}), \\ \boldsymbol{\Sigma}_{\hat{u},t} - \boldsymbol{\Sigma}_t &= O_p(\kappa_{n,\psi}^2 + \kappa_{n,h}). \end{aligned} \quad (2.20)$$

(ii) *In addition, if  $H_\psi \bar{H}_\psi = o(n)$ , then with  $\hat{\mathbf{D}}_t = 1 + \hat{\boldsymbol{\mu}}_t' (K_t^{-1} \sum_{j=1}^n k_{tj} \hat{\mathbf{y}}_{j-1} \hat{\mathbf{y}}_{j-1}')^{-1} \hat{\boldsymbol{\mu}}_t$ ,*

$$\begin{aligned} (K_t/K_{2,t}^{1/2}) \mathbf{H}_{t-1}^{-1} (I - \boldsymbol{\Psi}_t) (\hat{\boldsymbol{\mu}}_t - \boldsymbol{\mu}_t) &\rightarrow_D \mathcal{N}(0, \mathbf{I}), \\ (K_t/K_{2,t}^{1/2}) \mathbf{H}_{t-1}^{-1} (\hat{\boldsymbol{\alpha}}_t - \boldsymbol{\alpha}_t) \hat{\mathbf{D}}_t^{-1/2} &\rightarrow_D \mathcal{N}(0, \mathbf{I}). \end{aligned} \quad (2.21)$$

Moreover for any real  $m \times 1$ - vector  $\mathbf{a}$  such that  $\|\mathbf{a}\| = 1$ ,

$$(K_t/K_{2,t}^{1/2})^{1/2} \mathbf{H}_{t-1}^{-1} (\hat{\boldsymbol{\Psi}}_t - \boldsymbol{\Psi}_t) \left( \sum_{j=1}^n k_{tj} \hat{\mathbf{y}}_{j-1} \hat{\mathbf{y}}_{j-1}' \right)^{1/2} \mathbf{a} \rightarrow_D \mathcal{N}(0, \mathbf{I}). \quad (2.22)$$

(iii) *In addition, if  $H_h \bar{H}_h = o(n)$  and  $H_h^{1/2} \ll H_\psi \ll n/(H_h \log n)^{1/2}$ , then*

$$(L_t/L_{2,t}^{1/2}) \mathbf{H}_{t-1}^{-1} (\boldsymbol{\Sigma}_{\hat{u},t} - \boldsymbol{\Sigma}_t) \mathbf{H}_{t-1}'^{-1} \rightarrow_D \mathbf{Z} \quad (2.23)$$

where  $\mathbf{Z}$  is the same as in (2.16).

**Remark 2.2.** *In normal approximations for  $\hat{\boldsymbol{\mu}}_t$ ,  $\hat{\boldsymbol{\alpha}}_t$  and  $\hat{\boldsymbol{\Psi}}_t$  of Theorem 2.3(ii)  $\mathbf{H}_{t-1}^{-1}$  can be replaced by  $\boldsymbol{\Sigma}_{\hat{u},t}^{-1/2}$  because of  $\|\boldsymbol{\Sigma}_{\hat{u},t}^{-1/2} - \boldsymbol{\Sigma}_t^{-1/2}\| = o_p(1)$  of Lemma 6.5 (ii). For  $\boldsymbol{\Sigma}_{\hat{u},t}$  in (2.23) that can be done if  $\boldsymbol{\Sigma}_t^{-1/2} = \mathbf{H}_{t-1}^{-1}$  which holds if  $\mathbf{H}_{t-1}^{-1}$  is positive definite.*

Moreover, in (2.21) and (2.22),

$$\begin{aligned} \hat{\mathbf{D}}_t &= \mathbf{D}_t + o_p(1), \quad \mathbf{D}_t := 1 + \boldsymbol{\mu}_t' \mathbf{V}_{\psi,t}^{-1} \boldsymbol{\mu}_t, \\ K_t^{-1} \sum_{j=1}^n k_{tj} \hat{\mathbf{y}}_{j-1} \hat{\mathbf{y}}_{j-1}' &= V_{\psi,t} + o_p(1), \quad V_{\psi,t} := \sum_{k=0}^{\infty} \boldsymbol{\Psi}_t^k \boldsymbol{\Sigma}_t \boldsymbol{\Psi}_t'^k. \end{aligned} \quad (2.24)$$

The contribution of the present paper consists, first, of setting-up a VAR(1) model for a non-stationary process  $\mathbf{y}_t$  decomposed into a non-stationary random attractor  $\boldsymbol{\mu}_t$  and a term  $\mathbf{y}_t - \boldsymbol{\mu}_t$  driven by VAR(1) dynamics and, second, showing that the paths of the parameter processes can be extracted using kernel estimation method. The standard errors in (2.21)-(2.23) have the same asymptotics as the rolling regression standard errors in the case of constant parameters. There are, however, two differences comparing to constant or smoothly changing deterministic parameter models. Firstly, standard errors are random processes that vary in time, and secondly, estimation of standard errors requires the use of significantly larger bandwidths than in deterministic case. Further research is needed to establish estimation results of time-varying parameters that are uniform in time  $t$ , constructing Bonferroni-type correction to the pointwise standard error bands and developing a valid test of the null hypothesis of no time variation.

We choose to use the auxiliary vector  $\mathbf{a}$ , in (2.22), instead of  $\text{vec}(\boldsymbol{\Psi}_t)$  to obtain parameter free limit distribution.

### 2.3 Estimation of VAR(1) model with an intercept

To conclude, we discuss estimation of a VAR(1) model (2.2) with intercept. As in (2.18),  $\mathbf{y}_t$  can be written as a VAR(1) model

$$\mathbf{y}_t = \boldsymbol{\Psi}_{t-1}\mathbf{y}_{t-1} + \boldsymbol{\alpha}_t + \mathbf{u}_t = \sum_{k=0}^{t-1} \Pi_{t,k}\boldsymbol{\alpha}_{t-k} + \left\{ \sum_{k=0}^{t-1} \Pi_{t,k}\mathbf{u}_{t-k} + \Pi_{t,t}\mathbf{y}_0 \right\} =: \boldsymbol{\mu}_t + \dot{\mathbf{y}}_t, \quad (2.25)$$

that includes the attractor  $\boldsymbol{\mu}_t = \sum_{k=0}^{t-1} \Pi_{t,k}\boldsymbol{\alpha}_{t-k}$  and VAR(1) process  $\dot{\mathbf{y}}_t$  with no intercept:  $\dot{\mathbf{y}}_t = \boldsymbol{\Psi}_{t-1}\dot{\mathbf{y}}_{t-1} + \mathbf{u}_t$ ,  $t \geq 1$ ,  $\dot{\mathbf{y}}_0 = \mathbf{y}_0$ . By (2.9),  $\mathbf{y}_t$  satisfies the moving average representation  $\mathbf{y}_t = \boldsymbol{\mu}_t + \sum_{k=0}^{t-1} \boldsymbol{\Psi}_t^k \mathbf{u}_{t-k} + o_p(1)$ . Notice relationships between the attractor  $\boldsymbol{\mu}_t = \sum_{k=0}^{t-1} \Pi_{t,k}\boldsymbol{\alpha}_{t-k}$  and the intercept  $\boldsymbol{\alpha}_t = \boldsymbol{\mu}_t - \boldsymbol{\Psi}_{t-1}\boldsymbol{\mu}_{t-1}$ .

The following assumption describes a class of permissible intercepts  $\boldsymbol{\alpha}_t$ , for which the corresponding attractor  $\boldsymbol{\mu}_t$  in (2.25) satisfies Assumption 2.3.



**Assumption 2.4.**  $\boldsymbol{\alpha}_t = (\alpha_{1t}, \dots, \alpha_{mt})'$  is such that  $\max_t E\alpha_{it}^4 < \infty$ , and  $E\|\boldsymbol{\alpha}_t - \boldsymbol{\alpha}_{t+k}\|^2 \leq C(k/t)$ ,  $t \geq 1$ ,  $1 \leq k < t/2$ .

**Proposition 2.2.** If  $\boldsymbol{\alpha}_t$  satisfies Assumption 2.4 and Assumptions 2.1 and 2.2 hold, then  $\boldsymbol{\mu}_t$  in (2.25) satisfies Assumption 2.3(ii).

Therefore, estimation of the VAR(1) model with an intercept reduces to that of a model with an attractor, discussed in the previous section. This completes the discussion of the theoretical properties of our estimators.

## 2.4 Examples

In general, parameter processes  $\boldsymbol{\Psi}_t$ ,  $\boldsymbol{\alpha}_t$ ,  $\boldsymbol{\mu}_t$  and  $\boldsymbol{H}_t$  of a VAR model may contain random and deterministic components and can be seen as the sum of a standardized unit root type process and an additive deterministic parameter function.

In setting the model for VAR parameter  $\boldsymbol{\Psi}_t = \{\psi_{ij,t}\}$ , one can use the restriction that mirrors the bounding of Giraitis, Kapetanios, and Yates (2014) for univariate processes:

$$\psi_{ij,t} = r_{ij} \frac{a_{ij,t}}{\max_{0 \leq s \leq t} \sum_j |a_{ij,s}|}, \quad t \geq 1, \quad i, j = 1, \dots, m, \quad (2.26)$$

for some  $r_{ij} > 0$ ,  $r_{i1} + \dots + r_{im} \leq r < 1$  and some persistent processes  $a_{ij,t}$ . It satisfies requirement  $\|\boldsymbol{\Psi}_t\|_{sp} \leq r < 1$  of Assumption 2.1. To assure validity of the second requirement (2.4) of Assumption 2.1, one can assume that for any  $i, j = 1, \dots, m$ ,

$$n^{-1/2} a_{ij, [\tau n]} \Rightarrow_{D[0,1]} W_{ij,\tau} + g_{ij}(\tau), \quad 0 \leq \tau \leq 1, \quad (2.27)$$

converges weakly in Skorokhod space  $D[0, 1]$ , where  $(W_{ij,\tau}, 0 \leq \tau \leq 1)$  is zero mean random process with finite variance,  $W_{ij,1}$  has continuous probability distribution, and  $g_{ij}(\tau)$  is a deterministic continuous bounded function. Such coefficient  $a_{ij,t}$  may contain both stochastic and deterministic parts. The popular empirical choice of  $a_{ij,t}$  in macroeconomic literature is a random walk

$$a_{ij,t} = v_1 + \dots + v_t, \quad v_t \sim IID(0, \sigma^2).$$

If  $v_1$  has  $2 + \delta$  finite moments, then (2.27) holds with a Brownian motion limit. The condition (2.27) allows modeling of an extremely wide class of random/deterministic coefficient processes  $\psi_{ij,t}$ , see Giraitis, Kapetanios, and Yates (2014). In case  $m = 1$ , (2.27) implies (2.4) of Assumption 2.1, see Lemma 5.1(iii) of Giraitis, Kapetanios, and Yates (2014). In general, validity of (2.27) for each component  $\psi_{ij,t}$  implies (2.4) for  $m \geq 2$ .

A typical example of an intercept  $\alpha_t = \{\alpha_{i,t}\}$  satisfying Assumption 2.4 is

$$\alpha_{i,t} = t^{-1/2}(v_{i1} + \dots + v_{it}) + t^{-1}(d_{i1} + \dots + d_{it}), \quad t \geq 1, \quad i = 1, \dots, m, \quad (2.28)$$

where  $v_{it}$ 's are stationary zero mean r.v.'s such that  $\sum_{k \geq 0} |E v_{ik} v_{i0}| < \infty$ ,  $E v_{i1}^4 < \infty$ , and  $d_{it}$ 's are non-random numbers,  $\max_t |d_{it}| < \infty$ . It covers the case of a deterministic constant intercept  $\alpha_{i,t} = \alpha$ , a time varying intercept  $\alpha_{i,t} = g(t/n)$ , and a purely random intercept  $\alpha_{i,t} = t^{-1/2} \sum_{j=1}^t v_{ij}$ . In the univariate case (2.28) was discussed in Example 2.2 of Giraitis, Kapetanios, and Yates (2014).

A typical example of a time varying random volatility process  $\mathbf{H}_t = \{h_{ij,t}\}$  satisfying Assumption 2.2(ii) is

$$h_{ij,t} = |t^{-1/2}(v_{ij,1} + \dots + v_{ij,t}) + t^{-1}(d_{ij,1} + \dots + d_{ij,t})| + c_{ij}, \quad t \geq 1, \quad i, j = 1, \dots, m, \quad (2.29)$$

where the stationary process  $\{v_{ij,t}\}$  and non-random  $d_{ij,t}$ 's have the same properties as  $\{v_{it}\}$  and  $d_{i,t}$ 's in (2.28), and  $c_{ij} \geq 0$  are non-random. Such  $\mathbf{H}_t$  can be deterministic as well as random. Assumption 2.2(iii),  $\|\mathbf{H}_t^{-1}\|_{sp} = O_p(1)$ , e.g. is satisfied when  $\mathbf{H}_t$  is diagonal and  $c_{ii} > 0$  in (2.29).

### 3 Monte Carlo study

In this Section, using Monte Carlo simulations, we evaluate the performance of the estimators of the time varying VAR coefficients and the time varying volatilities.

### 3.1 Results for $\Psi_t$ coefficient

The model for the  $m$ -dimensional vector process  $\mathbf{y}_t$  is a  $VAR(1)$  with intercept:

$$\mathbf{y}_t = \boldsymbol{\alpha}_t + \Psi_{t-1}\mathbf{y}_{t-1} + \boldsymbol{\varepsilon}_t, \quad t = 1, 2, \dots, n, \quad (3.1)$$

where  $\Psi_t = \mathbf{P}_t^{(ort)} \boldsymbol{\Lambda}_t \mathbf{P}_t^{(ort)'}$ , the matrix  $\mathbf{P}_t^{(ort)}$  is obtained from  $m \times m$  matrix  $\mathbf{P}_t = \{p_{t,ij}\}$  by the Gram-Schmidt orthogonalisation to the columns, and  $\boldsymbol{\Lambda}_t = \text{diag}[\lambda_{it}]$  is a diagonal matrix. We set

$$\begin{aligned} p_{ij,t} &= 0.9 a_{ij,t} / \max_{0 \leq s \leq t} |a_{ij,s}|, & a_{ij,t} &= a_{ij,t-1} + v_{ij,t}, \\ \lambda_{ii,t} &= 0.9 \tilde{a}_{i,t} / \max_{0 \leq s \leq t} \tilde{a}_{i,s}, & \tilde{a}_{i,t} &= \tilde{a}_{i,t-1} + \eta_{i,t}, \end{aligned}$$

where  $\varepsilon_{i,t}$ ,  $v_{ij,t}$  and  $\eta_{i,t}$  are independent i.i.d. standard normal variates. The above assumption implies  $\|\Psi_t\|_{sp} \leq r < 1$ . Finally, we set  $\boldsymbol{\alpha}_t = 0$  but allow the estimation of a time varying intercept in our estimation of the model. In order to get a feel of the behaviour of the time varying coefficient, as well as its estimate, we report a single replication of element  $[1, 1]$  of  $\Psi_t$  as well as its estimate for  $T = 800$ ,  $m = 2$  and  $H_\psi = T^{0.5}$ , in Figure 1.

Table 3.1 reports the average MSE of kernel estimates of all elements of  $\Psi_t$  and  $\boldsymbol{\alpha}_t$  in (3.1) based on 1000 replications for  $m = 2, 8$  and various values of the bandwidth  $H_\psi$ . The Gaussian kernel is used. Tables confirm that the estimator of  $\Psi_t$  is consistent. A good choice for the bandwidth  $H_\psi$  seems to be a value around  $n^{0.5} - n^{0.6}$ , while the dimension of the model is not a major determinant of the performance of the estimator: we see that a large model with 8 variables is as well estimated as a much smaller bivariate model or even better which is in part the result of stronger restrictions imposed on the coefficient matrix that is required to have a maximum absolute eigenvalue bounded by 1.

	$m = 2$				$m = 8$			
$H_\psi$	n= 100	200	400	800	100	200	400	800
$n^{0.2}$	0.155	0.127	0.104	0.088	0.399	0.253	0.179	0.135
$n^{0.4}$	0.058	0.041	0.029	0.021	0.074	0.047	0.032	0.023
$n^{0.5}$	0.041	0.027	0.018	0.012	0.042	0.026	0.017	0.011
$n^{0.6}$	0.032	0.020	0.014	0.009	0.027	0.016	0.010	0.006
$n^{0.8}$	0.029	0.021	0.016	0.013	0.014	0.007	0.004	0.003

Table 3.1: Average MSE for  $\widehat{\Psi}_t$ .

### 3.2 Results for stochastic volatility $h_t^2$

In this section we explore the consistency properties of the estimator for the time varying volatility  $h_t^2$  of the error term  $\mathbf{u}_t$  of the bivariate VAR(1) model used in the previous subsection, but allowing for stochastic volatility. In particular

$$\mathbf{y}_t = \boldsymbol{\alpha}_t + \boldsymbol{\Psi}_{t-1}\mathbf{y}_{t-1} + \mathbf{H}_{t-1}\boldsymbol{\varepsilon}_t, \quad u_t = h_t\varepsilon_t,$$

where  $\mathbf{H}_{t-1} = \text{diag}(h_{1t-1}, h_{2t-1})$ . We use two models for  $h_{it-1}$ ,  $i = 1, 2$ . Firstly for Model 1,  $h_{it-1} = c \exp(a_{it-1}/\sqrt{t-1})$ ,  $a_{it} = a_{it-1} + v_{it}$ ,  $\varepsilon_t$ ,  $v_{it}$  are i.i.d. standard normal noises, and  $c > 0$  is selected such that  $Eh_{it}^2 = 1/25$ . Secondly, for Model 2,  $h_{it-1} = c(a_{it-1}/\sqrt{t-1})^2$ ,  $a_{it} = a_{it-1} + v_{it}$ ,  $\varepsilon_t$ ,  $v_{it}$  are i.i.d. standard normal noises, and, again,  $c > 0$  is selected such that  $Eh_{it}^2 = 1/25$ . To estimate the time varying random volatility  $h_t^2$ , we use the estimate (2.11):

$$\hat{h}_t^2 = \left( \sum_{k=1}^n L\left(\frac{t-k}{H_h}\right) \right)^{-1} \sum_{k=1}^n L\left(\frac{t-k}{H_h}\right) \hat{u}_k^2, \quad \hat{u}_k = y_k - \psi_t y_{k-1}.$$

We consider a variety of values for the two bandwidths  $H_h \leq H_\psi$  and the sample sizes  $n = 100, 200, 400, 800$ . Results for the Gaussian kernel estimates based on 1000 replications are reported in Tables 3.2 and 3.3 for the two different models for  $h_{it-1}$ . They clearly suggest that the estimator is consistent albeit less well performing than the one relating to the VAR coefficients. That is of course expected. In general, values of  $H_\psi$  around  $n^{0.5}$ - $n^{0.6}$  perform well when combined with  $H_h$  around  $n^{0.4}$ .

$H_\psi$	$H_h$	n= 100	200	400	800	$H_\psi$	$H_h$	n= 100	200	400	800
$n^{0.4}$	$n^{0.2}$	0.038	0.035	0.034	0.028	$n^{0.7}$	$n^{0.2}$	0.039	0.037	0.038	0.030
	$n^{0.3}$	0.036	0.030	0.025	0.024		$n^{0.3}$	0.034	0.030	0.026	0.024
	$n^{0.4}$	0.035	0.028	0.025	0.024		$n^{0.4}$	0.033	0.027	0.024	0.023
	$n^{0.5}$	0.035	0.031	0.028	0.027		$n^{0.5}$	0.033	0.028	0.027	0.027
	$n^{0.6}$	0.039	0.034	0.032	0.033		$n^{0.6}$	0.035	0.033	0.030	0.030
	$n^{0.7}$	0.042	0.040	0.041	0.037		$n^{0.7}$	0.037	0.038	0.036	0.038
	$n^{0.8}$	0.045	0.044	0.044	0.048		$n^{0.8}$	0.043	0.042	0.042	0.044
$n^{0.5}$	$n^{0.2}$	0.039	0.034	0.035	0.030	$n^{0.8}$	$n^{0.2}$	0.037	0.037	0.038	0.032
	$n^{0.3}$	0.036	0.030	0.028	0.024		$n^{0.3}$	0.035	0.032	0.026	0.025
	$n^{0.4}$	0.033	0.028	0.025	0.024		$n^{0.4}$	0.034	0.029	0.025	0.022
	$n^{0.5}$	0.034	0.029	0.026	0.024		$n^{0.5}$	0.033	0.029	0.027	0.023
	$n^{0.6}$	0.037	0.034	0.034	0.031		$n^{0.6}$	0.035	0.032	0.031	0.029
	$n^{0.7}$	0.040	0.036	0.037	0.038		$n^{0.7}$	0.038	0.038	0.037	0.038
	$n^{0.8}$	0.044	0.044	0.042	0.044		$n^{0.8}$	0.042	0.043	0.042	0.045
$n^{0.6}$	$n^{0.2}$	0.038	0.036	0.036	0.029	$n^{0.9}$	$n^{0.2}$	0.042	0.038	0.039	0.032
	$n^{0.3}$	0.032	0.029	0.027	0.024		$n^{0.3}$	0.037	0.032	0.028	0.025
	$n^{0.4}$	0.033	0.028	0.023	0.022		$n^{0.4}$	0.033	0.030	0.026	0.021
	$n^{0.5}$	0.034	0.029	0.026	0.025		$n^{0.5}$	0.032	0.027	0.027	0.023
	$n^{0.6}$	0.036	0.032	0.032	0.029		$n^{0.6}$	0.035	0.032	0.028	0.032
	$n^{0.7}$	0.040	0.037	0.037	0.036		$n^{0.7}$	0.039	0.038	0.037	0.036
	$n^{0.8}$	0.042	0.043	0.043	0.043		$n^{0.8}$	0.042	0.041	0.045	0.045

Table 3.2: Average MSE for  $\hat{h}_t$ ,  $m = 2$ , Model 1.

$H_\psi$	$H_h$	n= 100	200	400	800	$H_\psi$	$H_h$	n= 100	200	400	800
$n^{0.4}$	$n^{0.2}$	0.015	0.014	0.013	0.012	$n^{0.7}$	$n^{0.2}$	0.015	0.015	0.014	0.011
	$n^{0.3}$	0.015	0.014	0.012	0.010		$n^{0.3}$	0.015	0.013	0.013	0.010
	$n^{0.4}$	0.016	0.014	0.013	0.011		$n^{0.4}$	0.014	0.013	0.012	0.011
	$n^{0.5}$	0.016	0.015	0.014	0.014		$n^{0.5}$	0.015	0.015	0.014	0.013
	$n^{0.6}$	0.019	0.017	0.017	0.016		$n^{0.6}$	0.018	0.017	0.016	0.017
	$n^{0.7}$	0.020	0.020	0.021	0.022		$n^{0.7}$	0.020	0.019	0.020	0.019
	$n^{0.8}$	0.021	0.023	0.022	0.023		$n^{0.8}$	0.021	0.022	0.022	0.023
	$n^{0.5}$	$n^{0.2}$	0.015	0.014	0.014		0.012	$n^{0.8}$	$n^{0.2}$	0.015	0.014
$n^{0.3}$		0.015	0.013	0.012	0.010	$n^{0.3}$	0.014		0.013	0.011	0.010
$n^{0.4}$		0.014	0.014	0.013	0.011	$n^{0.4}$	0.015		0.013	0.012	0.012
$n^{0.5}$		0.016	0.015	0.014	0.013	$n^{0.5}$	0.016		0.015	0.014	0.012
$n^{0.6}$		0.017	0.017	0.017	0.016	$n^{0.6}$	0.018		0.017	0.017	0.016
$n^{0.7}$		0.019	0.019	0.020	0.020	$n^{0.7}$	0.019		0.019	0.020	0.019
$n^{0.8}$		0.021	0.021	0.022	0.022	$n^{0.8}$	0.021		0.022	0.022	0.024
$n^{0.6}$		$n^{0.2}$	0.014	0.014	0.013	0.012	$n^{0.9}$		$n^{0.2}$	0.016	0.014
	$n^{0.3}$	0.015	0.013	0.012	0.010	$n^{0.3}$		0.015	0.013	0.012	0.011
	$n^{0.4}$	0.014	0.013	0.013	0.011	$n^{0.4}$		0.014	0.014	0.011	0.012
	$n^{0.5}$	0.016	0.015	0.015	0.013	$n^{0.5}$		0.016	0.014	0.014	0.013
	$n^{0.6}$	0.017	0.016	0.016	0.017	$n^{0.6}$		0.017	0.017	0.015	0.015
	$n^{0.7}$	0.020	0.019	0.020	0.019	$n^{0.7}$		0.019	0.020	0.020	0.019
	$n^{0.8}$	0.021	0.022	0.022	0.023	$n^{0.8}$		0.021	0.022	0.023	0.023

Table 3.3: Average MSE for  $\hat{h}_t$ ,  $m = 2$ , Model 2.

## 4 Empirical application

In this section, we use kernel methods to estimate a VAR model for the 7 variable Smets-Wouters data set that, as far as we know, would be intractable for standard estimation methods based on MCMC algorithms. The data are for the United States, quarterly, and the sample period runs from 1956Q4 through to 2010Q2. The time series comprise quarterly growth rates (log differences) of GDP, investment, consumption, real wages, and also the levels of hours worked, the Federal Funds rate and quarterly inflation. We use  $H_\Psi = H_h = n^{1/2}$  and a Gaussian kernel for the estimation, following our theoretical results. We allow one lag in the VAR model and note that serial correlation tests on the residuals do not reject the null hypothesis of no serial correlation at the 95% significance level. Our VAR model includes a time varying intercept. We feel this is an important element of an empirical time varying model given that the time varying intercept can capture parameter shifts that may not be easy to capture with time varying autoregressive coefficients. For comparison purposes and to assess the benefit of considering the larger dimensional VAR model, we estimate a smaller, 4 variable system, containing GDP, investment, the levels of hours worked and quarterly inflation. The choice is motivated by the fact that these variables are often used to specify small VAR models for economic analysis. We develop two themes in the application. The first is on inflation predictability, a focus of the Cogley and Sargent (2005) paper that helped popularize the use of the Bayesian method for estimating time varying VARs with empirical macroeconomists. The second is on measuring changes in the estimated impulse response to identified shocks using sign restrictions. As a first exploratory step, we report estimates of the attractor  $\mu_t$  for our data, in Figure 2. We see that, as expected, there is considerable variation in the reported estimate.

## 4.1 Changes in inflation predictability

Following Cogley and Sargent (2005), the predictability  $P_t^j$  of the  $i$ th variable of the vector of observables  $Y$  at horizon  $j$  and time  $t$ , is defined as

$$P_t^j = 1 - \frac{e_i' \sum_{k=0}^{j-1} (\Psi_t^k) \Sigma_t (\Psi_t^k)' e_i}{e_i' \sum_{k=0}^{\infty} (\Psi_t^k) \Sigma_t (\Psi_t^k)' e_i},$$

where  $e_i$  is a selector matrix with a 1 on the  $i$ th row and zeros elsewhere. To provide intuition, consider a univariate model for inflation  $\pi_t = \rho_t \pi_{t-1} + u_t$ . Then, predictability at horizon 2 is given by  $P_t^2 = \rho_t^2$ , which in univariate studies is often labelled as squared ‘persistence’. So, the measure  $P_t^j$  is a multivariate counterpart to persistence.

Inflation predictability has been of interest for many reasons. Firstly, changes in predictability can be thought of as a way of characterising changes in macroeconomic performance. The underlying context is some macroeconomic model in which ideal monetary policy stabilises inflation perfectly up to some unforecastable error, in which case there is no inflation predictability. A fall in predictability came to be accepted as a main feature of the Great Moderation period, arising out of more effective monetary policy. Another reason for interest in this concept relates to what one might infer about the state of economy from strong variability in predictability. Predictability can arise from hard-wired features of private agents in a DSGE model like indexation, habits or investment adjustment costs. It can also be produced by monetary or fiscal policy. However, the more one observes variation in predictability, the less one is likely to conclude that it is the product of private agent behavioural features, which are supposed to be time variant. If one needs time-variation in behavioural features of a model to capture changes in macro dynamics, then, staying true to the spirit of microfounded model building, one probably has a mis-specified model.

We compute the Cogley-Sargent measure of inflation predictability for our 7 and 4 variable systems and plot them side by side in Figure 1.

The l.h.s. chart for the 7 variable system of Figure 1 reveals that there have been



some pronounced fluctuations in predictability, particularly in the last ten to fifteen years. But yet there is no clear delineation between pre- and post-Volcker episodes, emphasised in the early work by Cogley and Sargent (2005). Interestingly, this feature seems to be more evident in the 4 variable system, where one can see a pronounced fall in predictability in the early 1980s coincident with the Volcker appointment, and another very pronounced fall in the early 2000s, reversed by the end of the sample. Based on the 7 variable model we would say that the 4 variable model overstates the changes in predictability. In so far, as changes in monetary policy were adduced to be the cause of these changes in predictability, the 7 variable model would again suggest more evidence of continuity in monetary policy than is evident from the 4 variable system.

## 4.2 Shifts in the impulse response functions to identified monetary policy and technology shocks

Another way to investigate the nature of the process  $\mathbf{y}_t$  via the time varying VAR(1) model is to see what it implies about the changing impulse response functions to identified structural shocks  $\boldsymbol{\varepsilon}_t$  where the structural shocks are defined by  $\mathbf{u}_t = \mathbf{B}_t \boldsymbol{\varepsilon}_t$  or equivalently  $\boldsymbol{\varepsilon}_t = \mathbf{B}_t^{-1} \mathbf{u}_t$  where  $\boldsymbol{\Sigma}_t = \mathbf{B}_t \mathbf{B}_t'$ . Once  $\mathbf{B}_t$  is known and an estimator of it, denoted by  $\hat{\mathbf{B}}_t$ , is available then impulse responses  $\mathbf{R}(j; t)$ , at horizon  $j \geq 0$ , are computed as  $\hat{\mathbf{R}}(j; t) \equiv \hat{\boldsymbol{\Psi}}_t^j \hat{\mathbf{B}}_t$ . Before proceeding to discuss the operational details of constructing impulse responses it is important to address a conceptual question on what this impulse response is. In particular, standard impulse response is the change in  $\mathbf{y}_t - \boldsymbol{\mu}$  to a shock  $\mathbf{u}_{t-h}$  for some  $h \geq 0$ , where  $\boldsymbol{\mu}$  is the mean of  $\mathbf{y}_t$ . In our case, we will be considering the response of  $\mathbf{y}_t - \boldsymbol{\mu}_t$  to the shock  $\mathbf{u}_{t-h}$ , rather than  $\mathbf{y}_{t-h}$ . The shock  $\mathbf{u}_{t-h}$  could indeed affect  $\boldsymbol{\mu}_t$  but we abstract from that. It is well known that  $\mathbf{B}_t$  is not defined uniquely since for any factorisation  $\boldsymbol{\Sigma}_t = \mathbf{B}_t \mathbf{B}_t'$  and for any nonsingular orthogonal matrix  $\mathbf{Q}_t$ , it holds that  $\boldsymbol{\Sigma}_t = \mathbf{B}_t \mathbf{Q}_t \mathbf{Q}_t' \mathbf{B}_t'$ . In order to proceed, we use sign restrictions to construct impulse responses. Sign restrictions

operate by considering orthogonal rotations  $\mathbf{Q}_t = \mathbf{Q}_t(\boldsymbol{\theta})$ , parametrized by a set of parameters  $\boldsymbol{\theta}$ , such that elements of  $\hat{\mathbf{R}}(j; t) \equiv \hat{\Psi}_t^j \hat{\mathbf{B}}_t \mathbf{Q}_t$ , for particular  $j$ , have a sign (either positive or negative) chosen based on economic theory. They define the moving-average representation  $\mathbf{y}_t = \hat{\boldsymbol{\mu}}_t + \sum_{j=0}^{t-1} \hat{\mathbf{R}}(j; t) \boldsymbol{\varepsilon}_{t-j} + o_p(1)$ . The intention is to choose sign restrictions (4.1) that feature in different business cycle models (see, e.g., Peersman and Straub (2009)). So once  $\mathbf{Q}_t(\boldsymbol{\theta})$  are generated, those that lead to impulses which satisfy these sign restrictions are kept and the final responses are obtained by averaging over all responses obtained which satisfy the restrictions. Then, the main question is how to generate  $\mathbf{Q}_t(\boldsymbol{\theta})$  such that a good span of the space of orthogonal rotations is obtained. Usually this is done by considering Givens rotations. We consider a more general method that, in our view, provides a better span of the space of orthogonal rotations.

The search through the grid of parameters,  $\boldsymbol{\theta}$ , for possible rotations  $\mathbf{Q}_t$  is undertaken, with  $\mathbf{Q}_t$  parameterised as the product

$$\mathbf{Q}_t(\boldsymbol{\theta}) = \prod_{i=1}^{n-1} \prod_{j=i+1}^n \mathbf{I}_{ij}^n(\theta_{ij}), \quad 0 \leq \theta_{ij} \leq \pi/2,$$

of Givens rotations:

$$\mathbf{I}_{pq}^n(\theta) = \begin{pmatrix} 1 & \dots & 0 & \dots & 0 & \dots & 0 \\ \dots & \ddots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \cos(\theta) & \dots & -\sin(\theta) & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \sin(\theta) & \dots & \cos(\theta) & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & 0 & \dots & 0 & \dots & 1 \end{pmatrix}, \quad 0 \leq \theta \leq \pi/2,$$

where  $p < q$  denote the positions on the diagonal taken by  $\cos(\theta)$ .  $\theta$  is the  $n(n-1)/2$  vector containing all the scalar  $\theta_{ij}$ . These matrices respect the desired property that  $\mathbf{I}_{pq}^n(\theta) \mathbf{I}_{pq}^{n'}(\theta) = \mathbf{I}$ . It is obvious that

$$\mathbf{Q}_t(\boldsymbol{\theta}) \mathbf{Q}_t(\boldsymbol{\theta})' = \mathbf{I}.$$

We find it useful to search through multiple products of such rotations. This increases the number of acceptable rotations, which may be beneficial when imposing many restrictions, restrictions at multiple horizons, or repeat the analysis at every point in time as we do.

In our 7-variable VAR, among seven shocks,  $\varepsilon_{1t}, \dots, \varepsilon_{7t}$ , we identify four shocks: monetary policy, technology, labour supply and demand shocks, using sign restrictions set out in the table below. Blank cells indicate that the responses are left free. Restrictions are imposed on impact,  $R_{0,t}$ , only.

	$\Delta c$	$\Delta i$	$\Delta y$	$h$	$\pi$	$\Delta w/p$	$r$	
monetary policy	-	-	-		-			+
technology		+	+		-			(4.1)
labour supply			+	+	-	-		
demand	+		+		+	+	+	

In the table above,  $\Delta$  denotes log difference,  $c$  consumption,  $i$ , investment,  $y$  GDP,  $h$  hours,  $\pi$  inflation,  $w/p$  real wages and  $r$  the Federal Funds Rate. Interpreting the signs above we have that a contractionary monetary policy shock  $\varepsilon_{i_1t}$  (that raises nominal interest rates) is taken to be one that causes inflation, GDP, consumption and investment all to fall, on impact. A positive technology shock  $\varepsilon_{i_2t}$  is supposed to raise output and investment on impact, and to lower inflation. Importantly, we leave the response of hours worked free, so we can comment on whether time-variation is related to the question of the sign of this impulse response. A positive labour supply shock  $\varepsilon_{i_3t}$  is taken to be one that reduces real wages and inflation and increases hours worked and output. A demand shock  $\varepsilon_{i_4t}$  is identified as one that causes monetary policy to tighten, but despite that induces an increase in real wages, inflation, output and consumption. We concern ourselves only with the monetary policy and technology shocks, but identify others to improve the precision of our analysis.

Our identification scheme for the technology shock warrants some comment. Early papers in the debate by Gali (1999) and Christiano, Eichenbaum, and Vigfusson

(2003) used a long run restriction and identified technology shocks as the only thing that should contribute to long run changes in labour productivity. We adopt a different strategy, akin to that used by Peersman and Straub (2009), but with two differences. First, we leave the response of technology to real wages free. We leave it free because we found we could do without it: in our results we find that on impact of the technology shock, real wages rise. A second difference is that we include investment and restrict its response to be positive, while Peersman and Straub (2009) leave investment out of their set of observables.

We also identify monetary policy and technology shocks from the 4 variable VAR, so that we can gauge whether inference about changing impulse responses is affected by limiting the dimension of the VAR. Relevant sign restrictions are shown below.

	$\Delta y$	$\pi$	$\Delta i$	$h$
monetary policy	-	-	+	
technology	+	-		

The first object we look at is the impulse response of output  $\Delta y$  to the monetary policy shock. This gives us an overall impression of the strength of nominal rigidities in the economy: with flexible prices and wages, a monetary policy shock would have no effect on output at all. Chart4 shows the impulse response of output to a monetary policy shock, under the assumption that the reduced form VAR has time varying volatilities, which means that the size of the shock hitting the economy is potentially different at different points in time, so that we can nevertheless compare like with like. The impulse responses are normalised so as to deliver a 25 basis points increase on impact of the shock in the central bank policy rate  $r$ .

The magnitude of the response of output, measured by the absolute distance of the impulse response away from the zero plane, has fluctuated a lot. The cleanest indicator of the degree of nominal rigidities is the response on impact effect of the monetary policy shock. Looking at the l.h.s. panel of Figure 2 which records the results for the 7 variable VAR, we might deduce that the economy evolved toward

being a more flex-price economy (the impact of the shock falls from the beginning of the sample). Also the tendency for the economy to oscillate thereafter reduces, suggestive of the fact that corrective monetary policy responses improved. The 4 variable VAR shows some differences in these patterns. For a start, for most of the period the 4 variable VAR understates the impact of the monetary policy shock. Second, although the fall in the impact is captured by this smaller system, the 4 variable VAR shows the impact growing again towards the end of the sample, a change that is much less evident in the 7 variable system.

We turn next to the impulse response of hours worked to a technology shock. To set the scene, and give a flavour of the importance of this object, it is useful to recap briefly on the empirical macro-literature on this topic. The influential work by Gali (1999) contested the real business cycle view that technology shocks were a key driver of the business cycle, which had been at the forefront of the debate since the seminal work of Kydland and Prescott (1982). He identified technology shocks as being the only thing that gave rise to long run changes in labour productivity. He estimated that these shocks caused hours worked to *fall*. This implies that either the RBC account of business cycles is incorrect, or at least incomplete; or the model has to be modified to change the sign of the response of hours worked: the sticky price version of the RBC model does this, or some other shock was responsible for the major part of business cycles. Christiano, Eichenbaum, and Vigfusson (2003) made things look better for the RBC model by noting that whether hours rise or fall depends on whether the per capita hours work variable enters the VAR in levels or rates of change. Peersman and Straub (2009), whose sign restriction identification scheme most closely resembles ours, found that hours worked rise. An interesting comparison for our purposes is with Gali and Gambetti (2009). They estimate a 2-variable time varying VAR involving hours and labour productivity for the US using estimation based on MCMC algorithms. Technology shocks, as in the original Gali

(1999) paper, identified to be the only shocks that explain labour productivity in the long run. They find that the impulse response of hours worked to a technology shock is always negative, but gets less so over time.

Figure 3 shows our impulse response functions, from the VAR with time varying reduced form volatilities, and for both the 4 and 7 variable VARs. To compare like with like, we normalise the impulse responses here so that on impact the technology shock has a 25 basis points effect on output, imposed to be positive in the sign restrictions identification scheme.<sup>2</sup>

We find clear evidence of time-variation in impulse responses in the 7 variable system. The impulse of hours worked to our identified technology shock starts out small and negative, showing the economy resembling a sticky price version of the RBC model with fairly weak propagation, and evolves steadily to the most recent period to be large, hump-shaped, and positive, looking much more like a traditional flex-price real business cycle model. (Note the echo here with the evolution in the impulse response of output to a monetary policy shock, which also shows some evidence of the economy becoming more like a flex-price economy in the first half of the sample). This mirrors the Peersman and Straub (2009) results, which use an identification scheme most similar to ours. The results of Gali and Gambetti (2009) are comparable in the sense that they also find that the economy comes to resemble the RBC economy more, but his impulse response starts out markedly negative, and finishes the sample period marginally negative, whereas ours becomes markedly positive. The 4-variable system fails to pick up the tendency for the impulse response of hours to become more positive. There are some marked fluctuations in the 4 variable impulse response, but

---

<sup>2</sup>For the monetary policy shock, one can uniquely and fairly uncontroversially normalise the impulse responses over time so that they have the same effect on impact on the policy rate, thereby eliminating variability in the IRF due to time variation in the volatilities. When identifying time variation in changes in a tax instrument, normalisation could be achieved by isolating the impact of the shock on some tax rate. However, for shocks to economic primitives like technology there is no unique way to normalise the impulse responses, and doing it using output seems reasonable.

no clear pattern emerges, as indicated in the 7 variable system.

## 5 Concluding remarks

Giraitis, Kapetanios, and Yates (2014) offered an alternative, kernel-based method for estimating simple autoregressive models with stochastic time varying coefficients, in which parameters are assumed to be bounded random walks. This paper extends the theoretical results to the case of more realistic, multivariate models with time-varying volatilities, and shows through Monte-Carlo evidence that the theoretical results translate into good small-sample performance.

We also apply the estimation method to a 7 variable VAR with time-varying volatilities estimated on the US Smets-Wouters data set. We compute time-variation in three objects that have been the focus of attention in recent research in empirical macroeconomics. We find some evidence that the impulse response of output to a monetary policy shock has gotten smaller, as though the economy has become more like a flexible price economy. It appears that the impulse response of hours worked to a technology shock has become steadily more positive through the sample period, also implying that the economy has become more like a flex-price RBC economy. We also find fluctuations in inflation predictability, but no clear evidence that predictability fell in post-Volcker period, as has been advocated by previous work. Our results suggest some advantage in using the 7 variable system, since estimates of a 4 variable system contradict those from the larger system in a number of respects. To begin with, the 4-variable system understates both the impact of and changes in the impact of the monetary policy shock on output. Second, the 4-variable system fails to record the slow increase in the response of hours worked to a technology shock. Third, the 4 variable system shows a much more marked difference pre- and post-Volcker in inflation predictability, when the 7 variable system was suggestive of more continuity in this respect.

## References

- ABADIR, K. M., W. DISTASO, L. GIRAITIS, AND H. KOUL (2014): “Asymptotic normality for weighted sums of linear processes,” *Econometric Theory*, 30, 252–284.
- ALJ, A., R. AZRAK, C. LEY, AND G. MELARD (2016): “Asymptotic Properties of QML Estimators for VARMA Models with Time-Dependent Coefficients,” *Working paper*.
- BOUGEROL, P., AND R. PICARD (1992): “Strict stationarity of generalized autoregressive processes,” *Annals of Probability*, 20, 1714–1730.
- BRANDT, A. (1986): “The stochastic equation  $Y_{n+1} = A_n Y_n + B_n$  with stationary coefficients,” *Advances in Applied Probability*, 18, 211–220.
- CHRISTIANO, L. J., M. EICHENBAUM, AND R. VIGFUSSON (2003): “What happens after a technology shock?,” NBER Working Papers 9819, National Bureau of Economic Research, Inc.
- CLINE, D. (2007): “Evaluating the Lyapounov exponent and existence of moments for threshold AR-ARCH models,” *Journal of Time Series Analysis*, 28, 241–260.
- COGLEY, T., AND T. J. SARGENT (2005): “Drift and volatilities: monetary policies and outcomes in the post WWII U.S,” *Review of Economic Dynamics*, 8(2), 262–302.
- DAHLHAUS, R. (1997): “Fitting time series models to nonstationary processes,” *Annals of Statistics*, 25, 1–37.
- (2012): “Locally stationary processes,” *Handbook of Statistics*, 30.
- DAVIES, R. B. (1973): “Asymptotic inference in stationary Gaussian time series,” *Advances in Applied Probability*, 5, 469–497.



- FOSTER, D. P., AND D. B. NELSON (1996): “Continuous record asymptotics for rolling sample variance estimators,” *Econometrica*, 64, 139–174.
- FRANCO, C., AND J. ZAKOAN (2001): “Stationary of Multivariate Markov-Switching ARMA Models,” *Journal of Econometrics*, 102, 339–364.
- GALI, J. (1999): “Technology, employment, and the business cycle: do technology shocks explain aggregate fluctuations?,” *American Economic Review*, 89(1), 249–271.
- GALI, J., AND L. GAMBETTI (2009): “On the sources of the great moderation,” *American Economic Journal: Macroeconomics*, 1(1), 26–57.
- GIRAITIS, L., G. KAPETANIOS, AND T. YATES (2014): “Inference on stochastic time-varying coefficient models,” *Journal of Econometrics*, 179, 46–65.
- HALL, P., AND C. C. HEYDE (1980): *Martingale Limit Theory and Its Application*. Academic Press.
- KOOP, G., AND S. M. POTTER (2011): “Time varying VARs with inequality restrictions,” *Journal of Economic Dynamics and Control*, 35(7), 1126–1138.
- KRISTENSEN, D. (2012): “Nonparametric detection and estimation of structural change,” *Econometrics Journal*, 15, 420–461.
- KYDLAND, F. E., AND E. C. PRESCOTT (1982): “Time to build and aggregate fluctuations,” *Econometrica*, 50(6), 1345–70.
- NICHOLLS, D. F., AND B. G. QUINN (1982): *Random Coefficient Autoregressive Models*. Springer-Verlag.
- PEERSMAN, G., AND R. STRAUB (2009): “Technology shocks and robust sign restrictions in a Euro area SVAR,” *International Economic Review*, 50(3), 727–750.

- PHILLIPS, P. C. B., AND V. SOLO (1992): “Asymptotics for linear processes,” *Annals of Statistics*, 20, 971–1001.
- PRIESTLEY, M. (1965): “Evolutionary spectra and nonstationary processes,” *Journal of Royal Statistical Society, Series B*, 27, 204–37.
- SMETS, F., AND R. WOUTERS (2007): “Shocks and frictions in US business cycles: A Bayesian DSGE approach,” *American Economic Review*, 97(3), 586–606.
- STOCK, J. H., AND M. W. WATSON (1998): “Median unbiased estimation of co-efficient variance in a time-varying parameter model,” *Journal of the American Statistical Association*, 93(441), pp. 349–358.

Figure 1: True and Estimated element  $[1, 1]$  of  $\Psi_t$  for  $T = 800$ ,  $m = 2$  and  $H_\psi = T^{0.5}$

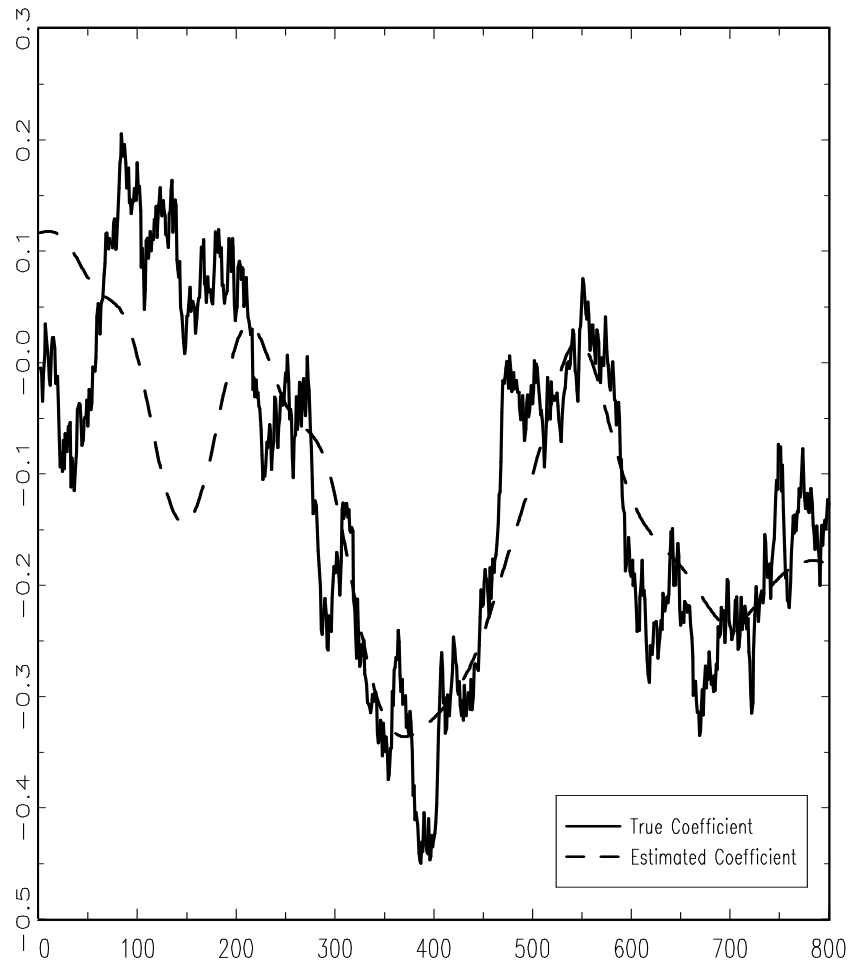


Figure 2: Estimated attractor ( $\hat{\mu}_t$ ) for the growth rates of GDP, investment, consumption, real wages, and the levels of hours worked, the Federal Funds rate and quarterly inflation

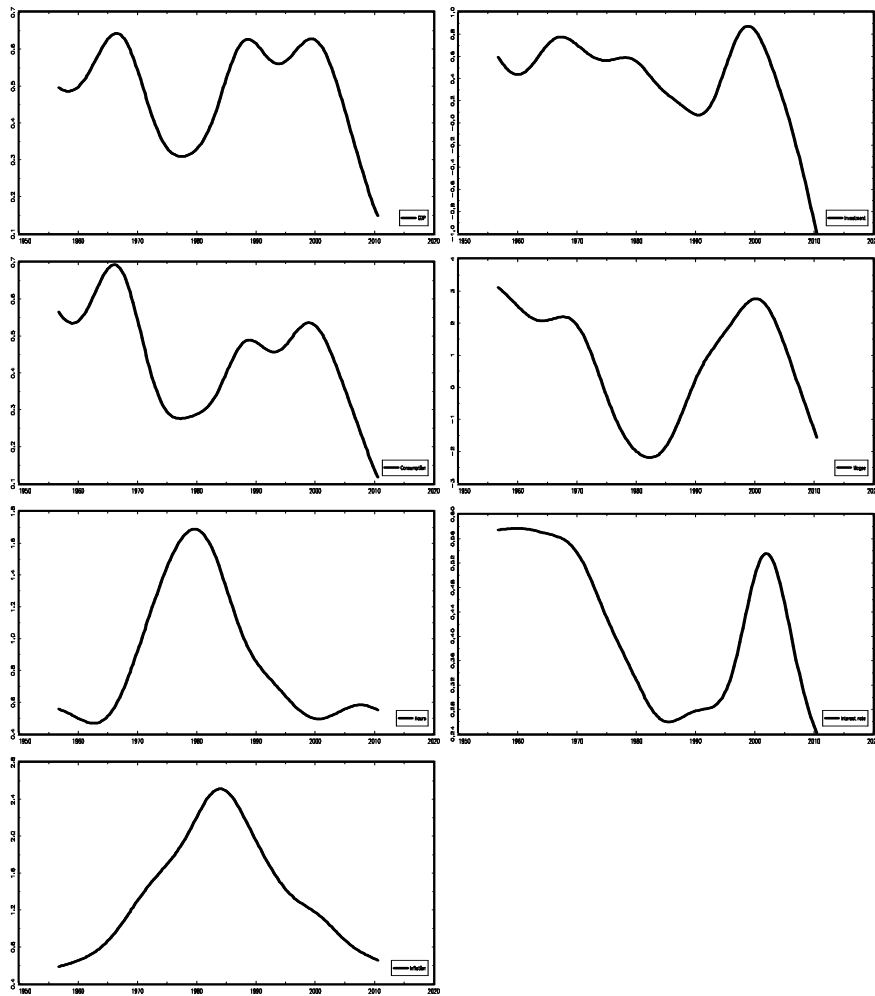


Figure 3: Inflation Predictability

Predictability of inflation (7 variable VAR)

Predictability of inflation (4 variable VAR)

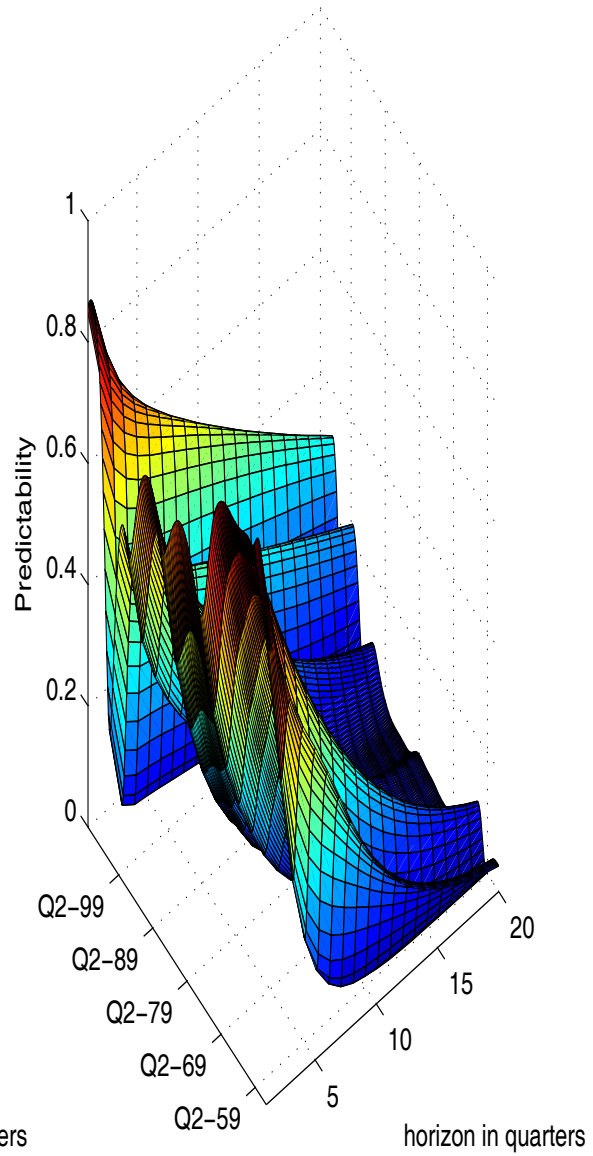
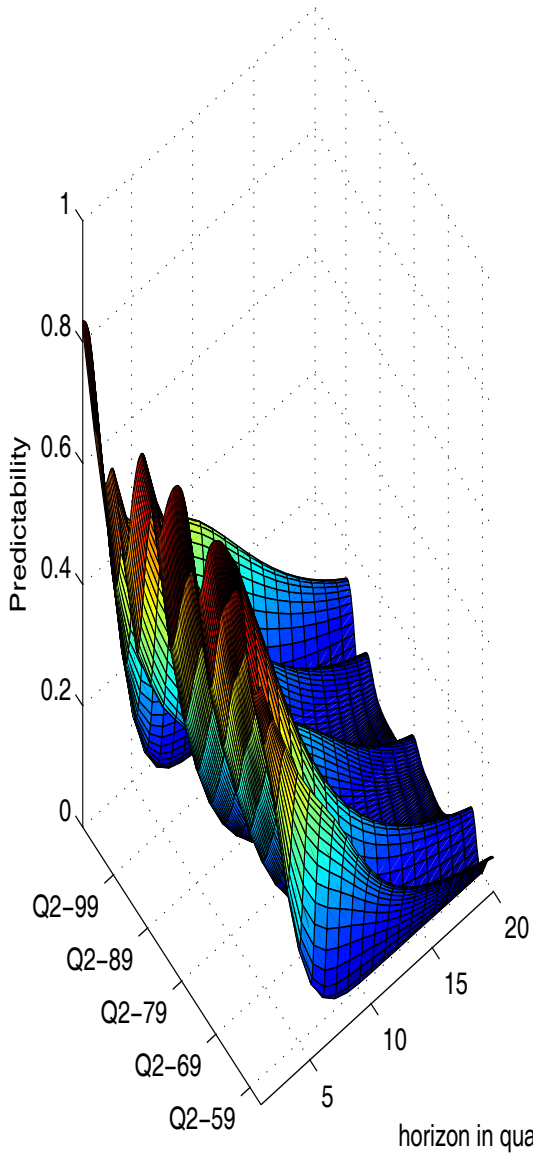


Figure 4: Monetary Policy Shock

IRF of output to a mon pol shock (7 variable VAR)

IRF of output to a mon pol shock (4 variable VAR)

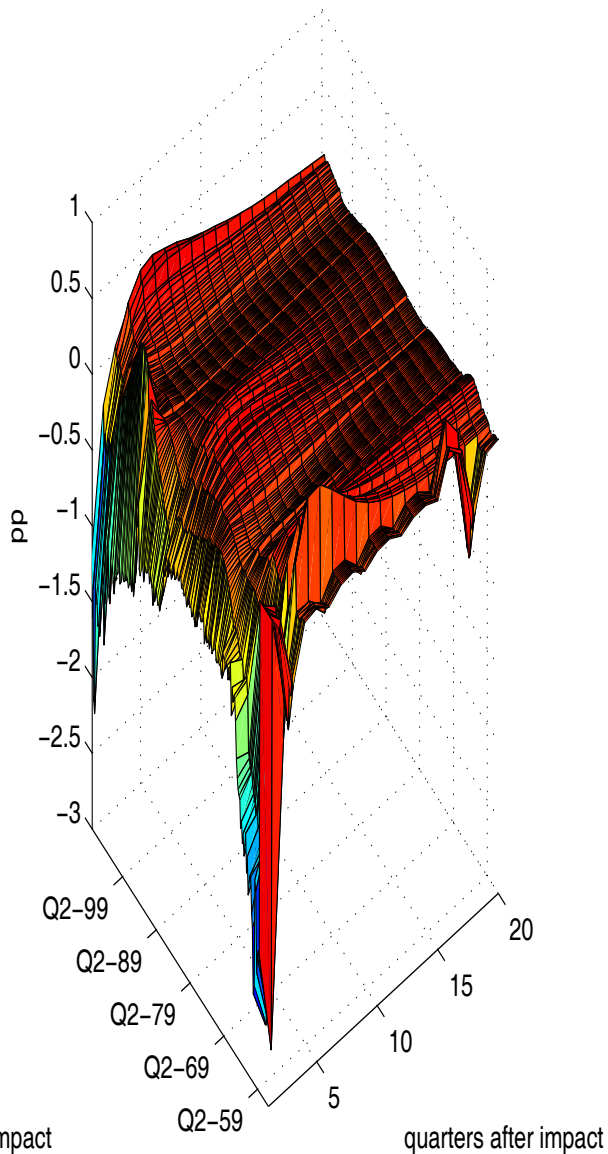
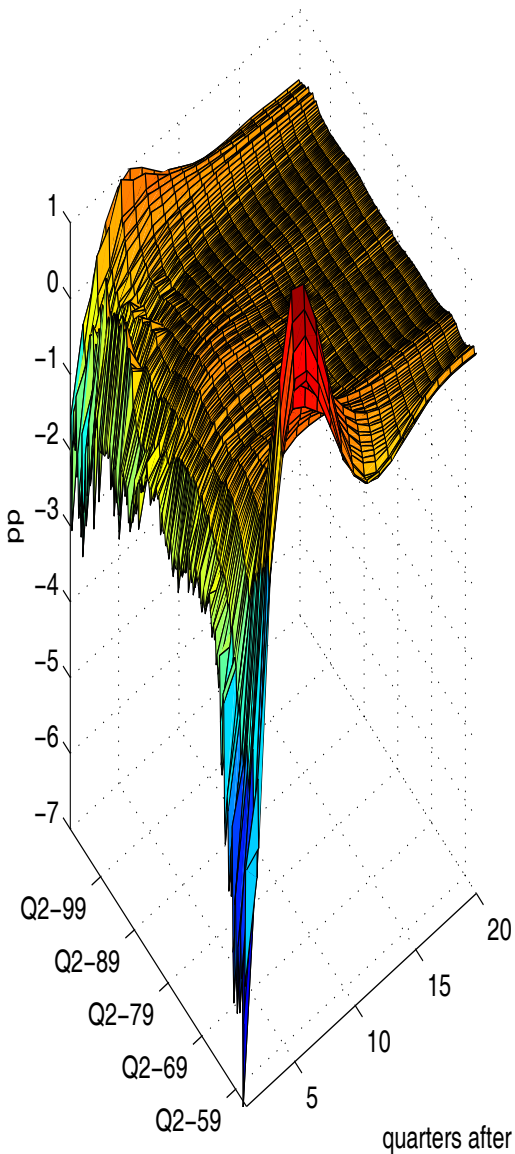
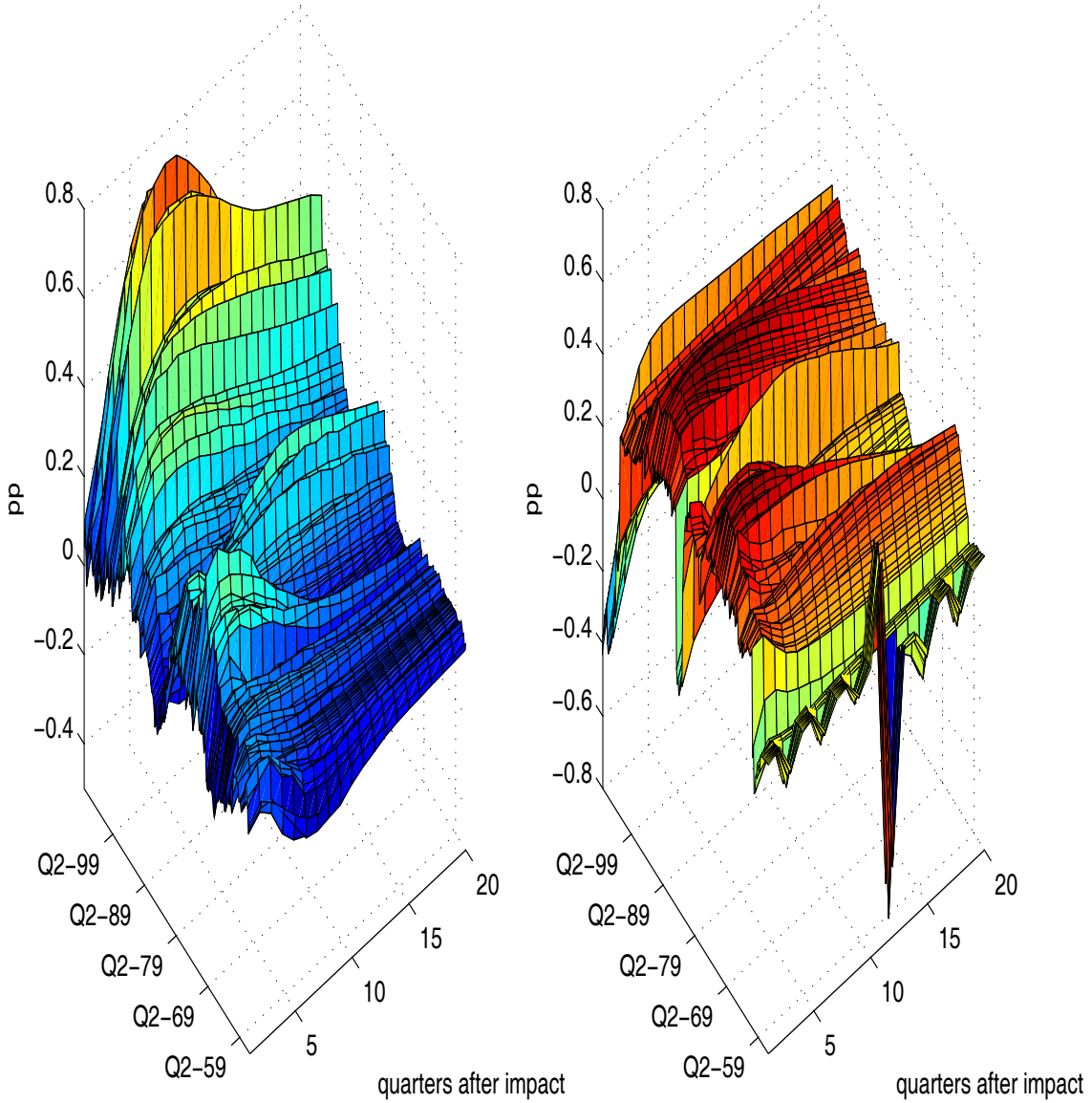


Figure 5: Technology Shock

IRF of hours to a tech shock, (7 variable VAR)

IRF of hours to a tech shock (4 variable VAR)



## 6 Online Appendix

### 6.1 Proof of Theorems 2.1-2.3

In the proof we use repeatedly the following properties of the spectral  $\|\cdot\|_{sp}$  and Euclidean  $\|\cdot\|$  norms of matrices:  $\|AB\|_{sp} \leq \|A\|_{sp}\|B\|_{sp}$ ,  $\|AB\| \leq \|A\|_{sp}\|B\|$ , and for a vector  $\mathbf{a}$  it holds  $\|\mathbf{a}\|_{sp} = \|\mathbf{a}\|$ .

Recall notation  $\bar{H} = H$  when  $K$  has finite support, and  $\bar{H} = H \log^{1/2} H$  when  $K$  has infinite support. In addition to  $\kappa_{n,\psi}$  of (2.12), we define

$$\kappa_{n,\psi}^* := (\bar{H}_\psi/n)^{1/2} + H_\psi^{-1}, \quad \kappa_{n,h}^* := (\bar{H}_h/n)^{1/2} + H_h^{-1}.$$

We will use the following property of the weights  $k_{tj}$  (see (6.16) in Giraitis, Kapetanios, and Yates (2014)): for  $t = \lceil \tau n \rceil$  ( $0 < \tau < 1$ ) there exists  $b > 0$  such that, as  $H \rightarrow \infty$ ,

$$\begin{aligned} \sum_{1 \leq j \leq n, |t-j| \geq b\bar{H}} k_{tj} &= o(1), \\ \sum_{j=1}^n k_{tj} &\sim H, \quad \sum_{j=1}^n k_{tj}^2 \sim \beta H, \quad \beta > 0. \end{aligned} \tag{6.1}$$

**Proof of Theorem 2.1.** Part (2.7) follows applying recursion (2.1). To show (2.8), note that by Assumptions 2.1-2.2,  $\|\Pi_{t,j}\|_{sp} \leq r^j$ ,  $E\|\mathbf{y}_0\|^4 < \infty$  and  $\max_{j \geq 0} E\|\mathbf{u}_j\|^4 < \infty$ , which together with (2.7) implies

$$\begin{aligned} \|\mathbf{y}_t\| &\leq \sum_{j=0}^{t-1} \|\Pi_{t,j}\|_{sp} \|\mathbf{u}_{t-j}\| + \|\Pi_{t,t}\|_{sp} \|\mathbf{y}_0\| \leq \sum_{j=0}^{t-1} r^j \|\mathbf{u}_{t-j}\| + r^t \|\mathbf{y}_0\|, \\ E\|\mathbf{y}_t\|^4 &\leq (\max_{j \geq 0} E\|\mathbf{u}_j\|^4 + E\|\mathbf{y}_0\|^4) (\sum_{j=0}^{\infty} r^j)^4 < \infty. \end{aligned}$$

To show (2.9), use  $|a_1 \cdots a_j - b_1 \cdots b_j| = |(a_1 - b_1)a_2 \cdots a_j + b_1(a_2 - b_2)a_3 \cdots a_j + \cdots + b_1 \cdots b_{j-1}(a_j - b_j)| \leq j \max_{i=1, \dots, j} |a_i - b_i| a^{j-1}$ , if  $|a_i| \leq a$  and  $|b_i| \leq a$ , to obtain

$$\|\Pi_{t,j} - \Psi_t^j\|_{sp} \leq j r^{j-1} \max_{i=1, \dots, j} \|\Psi_{t-i} - \Psi_t\|_{sp}. \tag{6.2}$$

Denote

$$R_{t,h} := \max_{s: |s-t| \leq h} \|\Psi_t - \Psi_s\|_{sp}, \tag{6.3}$$



where  $h \rightarrow \infty$  and  $h = o(t)$ . Then, by (2.7),

$$\begin{aligned} \|\mathbf{y}_t - \mathbf{z}_t\| &\leq \sum_{j=1}^{t-1} \|\Pi_{t,j} - \Psi_t^j\|_{sp} \|\mathbf{u}_{t-j}\| + \|\Pi_{t,t}\|_{sp} \|\mathbf{y}_0\| \\ &\leq R_{t,h} \sum_{j=1}^h j r^{j-1} \|\mathbf{u}_{t-j}\| + \{2 \sum_{j=h+1}^{t-1} j r^j \|\mathbf{u}_{t-j}\| + r^t \|\mathbf{y}_0\|\} =: R_{t,h} z_{t,h,1} + z_{t,h,2} = o_p(1), \end{aligned}$$

because  $R_{t,h} = O_p(\sqrt{h/t}) = o_p(1)$  by Assumption 2.1, whereas  $E z_{t,h,1} \leq \max_j E \|\mathbf{u}_j\| \sum_{j=1}^{\infty} j r^{j-1} < \infty$  and  $E z_{t,h,2} \leq C \sum_{j=h+1}^{\infty} j r^j \rightarrow 0$  as  $h \rightarrow \infty$  implies  $z_{t,h,1} = O_p(1)$  and  $z_{t,h,2} = o_p(1)$ .  $\square$

**Proof of Theorem 2.2.** (i) First we prove (2.13). Set  $\mathbf{V}_{yy,t} := K_t^{-1} \sum_{j=1}^n k_{tj} \mathbf{y}_{j-1} \mathbf{y}'_{j-1}$ ,

$$\mathbf{S}_{yy,t} := K_t^{-1} \sum_{j=1}^n k_{tj} \mathbf{y}_j \mathbf{y}'_{j-1}, \quad \mathbf{S}_{uy,t} := K_t^{-1} \sum_{j=1}^n k_{tj} \mathbf{u}_j \mathbf{y}'_{j-1}, \quad \Sigma_{uu,t} := L_t^{-1} \sum_{j=1}^n l_{tj} \mathbf{u}_j \mathbf{u}'_j. \quad (6.4)$$

Then,  $\widehat{\Psi}_t = \mathbf{S}_{yy,t} \mathbf{V}_{yy,t}^{-1}$ . We will show that

$$\|\widehat{\Psi}_t - \Psi_t - \mathbf{S}_{uy,t} \mathbf{V}_{yy,t}^{-1}\|_{sp} = O_p(\kappa_{n,\psi}^*). \quad (6.5)$$

Then  $\|\widehat{\Psi}_t - \Psi_t\| \leq \|\widehat{\Psi}_t - \Psi_t - \mathbf{S}_{uy,t} \mathbf{V}_{yy,t}^{-1}\|_{sp} + \|\mathbf{S}_{uy,t}\|_{sp} \|\mathbf{V}_{yy,t}^{-1}\|_{sp} = O_p(\kappa_{n,\psi}^*) + O_p(\kappa_{n,\psi}) = O_p(\kappa_{n,\psi})$  by (6.5), noting that  $\|\mathbf{S}_{uy,t}\|_{sp} = O_p(H_\psi^{-1/2})$  by Lemma 6.2(i), and  $\|\mathbf{V}_{yy,t}^{-1}\|_{sp} = O_p(1)$  by Lemma 6.1(vi), which proves (2.13). Since by (2.1),

$$\widehat{\Psi}_t - \Psi_t = (\mathbf{S}_{uy,t} + \mathbf{r}_t) \mathbf{V}_{yy,t}^{-1}, \quad \mathbf{r}_t := K_t^{-1} \sum_{j=1}^n k_{tj} (\Psi_{j-1} - \Psi_t) \mathbf{y}_{j-1} \mathbf{y}'_{j-1}, \quad (6.6)$$

(6.5) follows from

$$\|\mathbf{r}_t\|_{sp} = O_p((\bar{H}_\psi/n)^{1/2} + H_\psi^{-1}). \quad (6.7)$$

To show (6.7), select  $h = b\bar{H}_\psi$  such that (6.1) holds. By Assumptions 2.1 and 2.2,  $R_{t,h}$  in (6.3) satisfies  $R_{t,h} = O_p((h/t)^{1/2}) = O_p((\bar{H}_\psi/n)^{1/2})$ ,  $\|\Psi_{t-j} - \Psi_t\|_{sp} \leq \|\Psi_{t-j}\|_{sp} + \|\Psi_t\|_{sp} \leq 2r$ ,  $E\|u_1\|^2 < \infty$ . Therefore,

$$\|\mathbf{r}_t\|_{sp} \leq K_t^{-1} \sum_{j=1}^n k_{tj} \|\Psi_{t-j} - \Psi_t\|_{sp} \|\mathbf{y}_{j-1}\|^2 \leq R_{t,h} q_{n,1} + q_{n,2},$$

where  $q_{n,1} := K_t^{-1} \sum_{j=1}^h k_{tj} \|\mathbf{y}_{j-1}\|^2$  and  $q_{n,2} := 2rK_t^{-1} \sum_{|j-t|>h}^n k_{tj} \|\mathbf{y}_{j-1}\|^2$ .

Observe that  $Eq_{n,1} \leq CK_t^{-1} \sum_{j=1}^h k_{tj} \leq C$  and  $Eq_{n,2} \leq CK_t^{-1} \sum_{j:|t-j|\geq h} k_{tj} = H_\psi^{-1}O(1) = O(H_\psi^{-1})$  by (6.1) which implies  $q_{n,1} = O_p(1)$  and  $q_{n,2} = O_p(H_\psi^{-1})$ . This proves (6.7):  $\|\mathbf{r}_t\|_{sp} = O_p((h/t)^{1/2})O_p(1) + O_P(H_\psi^{-1}) = O_P((\bar{H}_\psi/n)^{1/2} + H_\psi^{-1})$ , which completes the proof of (6.7) and (2.13).

(ii) *Proof of (2.15).* Denote by  $T_{n,t} := (K_t/K_{2,t})^{1/2} \mathbf{H}_{t-1}^{-1} (\hat{\Psi}_t - \Psi_t) (\sum_{j=1}^n k_{tj} \mathbf{y}_{j-1} \mathbf{y}'_{j-1})^{1/2} \mathbf{a}$  the l.h.s. of (2.15). By (6.6), one can write

$$T_{n,t} = (K_t/K_{2,t}^{1/2}) \mathbf{H}_{t-1}^{-1} \mathbf{S}_{uy,t} \mathbf{V}_{yy,t}^{-1/2} \mathbf{a} + (K_t/K_{2,t}^{1/2}) \mathbf{H}_{t-1}^{-1} \mathbf{r}_t \mathbf{V}_{yy,t}^{-1/2} \mathbf{a} := T_{n,t;1} + T_{n,t;2}.$$

By Lemma 6.2(ii),  $T_{n,t;1} \rightarrow_D \mathcal{N}(0, \mathbf{I})$ . It remains to show  $\|T_{n,t;2}\|_{sp} \rightarrow_p 0$ . By (6.1),  $K_t/K_{2,t}^{1/2} = O(H_\psi^{1/2})$ . Hence,  $\|T_{n,t;2}\|_{sp} \leq CH_\psi^{1/2} \|\mathbf{H}_{t-1}^{-1}\|_{sp} \|\mathbf{r}_t\|_{sp} \|\mathbf{V}_{yy,t}^{-1/2}\|_{sp}$ , where  $\|\mathbf{H}_{t-1}^{-1}\|_{sp} = O_p(1)$  by Assumption 2.2(iii),  $\|\mathbf{r}_t\|_{sp}$  satisfies (6.7) and  $\|\mathbf{V}_{yy,t}^{-1/2}\|_{sp} = O_p(1)$  by Lemma 6.1(vi). So,  $\|T_{n,t;2}\|_{sp} \leq CH_\psi^{1/2} O_p((\bar{H}_\psi/n)^{1/2} + H_\psi^{-1}) = o_p(1)$  for  $H_\psi \bar{H}_\psi = o(n)$ , which completes the proof of (2.15).

(iii) *Proof of (2.14).* Use

$$\Sigma_{\hat{u}\hat{u},t} - \Sigma_t = (\Sigma_{\hat{u}\hat{u},t} - \Sigma_{uu,t}) + (\Sigma_{uu,t} - \Sigma_t), \quad (6.8)$$

to bound  $\|\Sigma_{\hat{u}\hat{u},t} - \Sigma_t\|_{sp} \leq \|\Sigma_{\hat{u}\hat{u},t} - \Sigma_{uu,t}\|_{sp} + \|\Sigma_{uu,t} - \Sigma_t\|_{sp}$ . By Lemma 6.5(i),  $\|\Sigma_{\hat{u}\hat{u},t} - \Sigma_{uu,t}\|_{sp} = O_p(\kappa_{n,\psi}^2 + \kappa_{n,h})$ , while by (6.32) of Lemma 6.3,  $\|\Sigma_{uu,t} - \Sigma_t\|_{sp} = O_p(\kappa_{n,h})$ , which yields (2.14):  $\|\Sigma_{\hat{u}\hat{u},t} - \Sigma_t\|_{sp} = O_p(\kappa_{n,\psi}^2 + \kappa_{n,h})$ .

To complete the proof, observe that  $\kappa_{n,\psi}^2 \leq 2((\bar{H}_\psi/n) + H_\psi^{-1}) \leq 2\kappa_{n,h} = 2((\bar{H}_h/n)^{1/2} + H_h^{-1/2})$  under assumption  $H_h^{1/2} \leq H_\psi \leq (H_h n / \log^2 n)^{1/2}$ , because  $H_\psi^{-1} \leq H_h^{-1/2}$  when  $H_h^{1/2} \leq H_\psi$ , and  $(\bar{H}_\psi/n) \leq (\bar{H}_h/n)^{1/2}$  since  $(\bar{H}_\psi/n)/(\bar{H}_h/n)^{1/2} \leq \log n (H_\psi/n)/(H_h/n)^{1/2} \leq 1$  for  $H_\psi \leq (H_h n)^{1/2} / \log n$ .

*Proof of (2.16).* Use (6.8), to write

$$(L_t/L_{2,t}^{1/2}) \mathbf{H}_{t-1}^{-1} (\Sigma_{\hat{u}\hat{u},t} - \Sigma_t) \mathbf{H}'_{t-1}^{-1} = q_{n,t} + r_{n,t},$$

where  $q_{n,t} := (L_t/L_{2,t}^{1/2})\mathbf{H}_{t-1}^{-1}(\boldsymbol{\Sigma}_{uu,t} - \boldsymbol{\Sigma}_t)\mathbf{H}'_{t-1}$  and  $r_{n,t} := (L_t/L_{2,t}^{1/2})\mathbf{H}_{t-1}^{-1}(\boldsymbol{\Sigma}_{\hat{u}\hat{u},t} - \boldsymbol{\Sigma}_{uu,t})\mathbf{H}'_{t-1}$ . Notice that assumption  $H_h\bar{H}_h = o(n)$  implies  $H_h = o(n^{1/2})$ . Thus, by Lemma 6.3(ii),  $q_{n,t} \rightarrow_D \mathbf{Z}$ . It remains to show that  $\|r_{n,t}\|_{sp} \rightarrow_p 0$ . By (6.44),  $\|\boldsymbol{\Sigma}_{\hat{u}\hat{u},t} - \boldsymbol{\Sigma}_{uu,t}\|_{sp} = O_p(\kappa_{n,\psi}^2 + (\bar{H}_h/n)^{1/2} + H_h^{-1}) = o(H_h^{-1/2})$  because  $(\bar{H}_h/n)^{1/2} = o(H_h^{-1/2})$  under assumption  $H_h\bar{H}_h = o(n)$ , and  $\kappa_{n,\psi}^2 \leq 2((\bar{H}_\psi/n) + H_\psi^{-1}) = o(H_h^{-1/2})$  under assumption  $H_h^{1/2} \ll H_\psi \ll n/(H_h \log n)^{1/2}$ . Recall that by (6.1),  $L_t/L_{2,t}^{1/2} = O(H_h^{1/2})$ , and  $\|\mathbf{H}_{t-1}^{-1}\|_{sp} = O_p(1)$  by Assumption (2.2)(iii). Therefore,  $\|r_{n,t}\|_{sp} \leq (L_t/L_{2,t}^{1/2})\|\mathbf{H}_{t-1}^{-1}\|_{sp}^2\|r_{n,t}\|_{sp} = O(H_h^{1/2})o_p(H_h^{-1/2}) = o_p(1)$  which completes the proof of (2.16) and of the theorem.  $\square$

**Proof of Theorem 2.3.** *Proof of the claims about  $\bar{\mathbf{y}}_t$ .* Note that

$$\|\bar{\mathbf{y}}_t - \boldsymbol{\mu}_t\| \leq \|\bar{\mathbf{y}}_t - \boldsymbol{\mu}_t - \bar{\dot{\mathbf{y}}}_t\| + \|\bar{\dot{\mathbf{y}}}_t\| = O_p(\kappa_{n,\psi}) \quad (6.9)$$

since  $\|\bar{\mathbf{y}}_t - \boldsymbol{\mu}_t - \bar{\dot{\mathbf{y}}}_t\| = O_p(\kappa_{n,\psi}^*)$  by Lemma 6.4(ii), and  $\|\bar{\dot{\mathbf{y}}}_t\| \equiv \|K_t^{-1} \sum_{j=1}^n k_{tj} \dot{\mathbf{y}}_j\| = O_p(\kappa_{n,\psi})$  by (6.38) of Lemma 6.4, which proves (2.20).

To show (2.21), write

$$\begin{aligned} (K_t/K_{2,t}^{1/2})\mathbf{H}_{t-1}^{-1}(I - \boldsymbol{\Psi}_t)(\hat{\boldsymbol{\mu}}_t - \boldsymbol{\mu}_t) &= (K_t/K_{2,t}^{1/2})\mathbf{H}_{t-1}^{-1}(I - \boldsymbol{\Psi}_t)\bar{\dot{\mathbf{y}}}_t \\ &+ (K_t/K_{2,t}^{1/2})\mathbf{H}_{t-1}^{-1}(I - \boldsymbol{\Psi}_t)(\hat{\boldsymbol{\mu}}_t - \boldsymbol{\mu}_t - \bar{\dot{\mathbf{y}}}_t) =: v_{n,1} + v_{n,2}. \end{aligned}$$

By (6.39) of Lemma 6.4,  $v_{n,1} \rightarrow_D \mathcal{N}(0, I)$ . Thus, to prove (2.21), it remains to show that  $\|v_{n,2}\| = o_p(1)$ . Indeed,  $\|v_{n,2}\| \leq (K_t/K_{2,t}^{1/2})\|\mathbf{H}_{t-1}^{-1}\|_{sp}\|I - \boldsymbol{\Psi}_t\|_{sp}\|\hat{\boldsymbol{\mu}}_t - \boldsymbol{\mu}_t - \bar{\dot{\mathbf{y}}}_t\| = O(H_\psi^{1/2})O_p(\kappa_{n,\psi}^*)$ , because  $(K_t/K_{2,t}^{1/2}) = O(H_\psi^{1/2})$  by (6.1),  $\|\mathbf{H}_{t-1}^{-1}\| = O_p(1)$  by Assumption 2.2(iii),  $\|I - \boldsymbol{\Psi}_t\|_{sp} \leq 1 + \|\boldsymbol{\Psi}_t\|_{sp} \leq 1 + r$  by Assumption 2.1, and  $\|\hat{\boldsymbol{\mu}}_t - \boldsymbol{\mu}_t - \bar{\dot{\mathbf{y}}}_t\| = O_p(\kappa_{n,\psi}^*)$  by Lemma 6.4(ii). Since  $\kappa_{n,\psi}^* = o(H_\psi^{-1/2})$  for  $H_\psi\bar{H}_\psi = o(n)$ , this implies  $\|v_{n,2}\| = o_p(1)$ .

*Proof of the claims about  $\hat{\boldsymbol{\Psi}}_t$ .* Using notation (6.4), write  $\hat{\boldsymbol{\Psi}}_t \equiv \hat{\boldsymbol{\Psi}}_{\hat{\mathbf{y}}\hat{\mathbf{y}},t} = S_{\hat{\mathbf{y}}\hat{\mathbf{y}},t}V_{\hat{\mathbf{y}}\hat{\mathbf{y}},t}^{-1}$ .

Notice that  $\hat{\boldsymbol{\Psi}}_{\hat{\mathbf{y}}\hat{\mathbf{y}},t} = S_{\hat{\mathbf{y}}\hat{\mathbf{y}},t}V_{\hat{\mathbf{y}}\hat{\mathbf{y}},t}^{-1}$  is an estimate of  $\boldsymbol{\Psi}_t$  of a VAR(1) model (2.18) for  $\hat{\mathbf{y}}_j$  with no intercept. Therefore it has consistency property (2.13) and satisfies

asymptotic normality (2.15) of Theorem 2.2. Write

$$\hat{\Psi}_{\hat{y}\hat{y},t} - \Psi_t = (\hat{\Psi}_{\dot{y}\dot{y},t} - \Psi_t) + (\hat{\Psi}_{\hat{y}\hat{y},t} - \hat{\Psi}_{\dot{y}\dot{y},t}). \quad (6.10)$$

To prove consistency claim (2.20) of Theorem 2.3, bound  $\|\hat{\Psi}_{\hat{y}\hat{y},t} - \Psi_t\|_{sp} \leq \|\hat{\Psi}_{\dot{y}\dot{y},t} - \Psi_t\|_{sp} + \|\hat{\Psi}_{\hat{y}\hat{y},t} - \hat{\Psi}_{\dot{y}\dot{y},t}\|_{sp}$ , where  $\|\hat{\Psi}_{\dot{y}\dot{y},t} - \Psi_t\|_{sp} = O_p(\kappa_{n,\psi})$  by (2.13) of Theorem 2.2, and

$$j_n := \|\hat{\Psi}_{\hat{y}\hat{y},t} - \hat{\Psi}_{\dot{y}\dot{y},t}\|_{sp} = O_p(\kappa_{n,\psi}^*). \quad (6.11)$$

Indeed,  $j_n = \|S_{\hat{y}\hat{y},t}V_{\hat{y}\hat{y},t}^{-1} - S_{\dot{y}\dot{y},t}V_{\dot{y}\dot{y},t}^{-1}\|_{sp} \leq \|S_{\hat{y}\hat{y},t} - S_{\dot{y}\dot{y},t}\|_{sp}\|V_{\hat{y}\hat{y},t}^{-1}\|_{sp} + \|S_{\dot{y}\dot{y},t}\|_{sp}\|V_{\hat{y}\hat{y},t}^{-1} - V_{\dot{y}\dot{y},t}^{-1}\|_{sp} = O_p(\kappa_{n,\psi}^*)$  because by Lemma 6.5(iii),  $\|S_{\hat{y}\hat{y},t} - S_{\dot{y}\dot{y},t}\|_{sp} = O_p(\kappa_{n,\psi}^*)$ ,  $\|V_{\hat{y}\hat{y},t}^{-1}\|_{sp} = O_p(1)$  and  $\|V_{\hat{y}\hat{y},t}^{-1} - V_{\dot{y}\dot{y},t}^{-1}\|_{sp} = O_p(\kappa_{n,\psi}^*)$ , while  $\|S_{\dot{y}\dot{y},t}\|_{sp} = O_p(1)$  because  $E\|S_{\dot{y}\dot{y},t}\|_{sp} \leq K_t^{-1} \sum_{j=1}^n k_{tl} E\|\dot{\mathbf{y}}_j\| \|\dot{\mathbf{y}}_{j-1}\| \leq CK_t^{-1} \sum_{j=1}^n k_{tl} = C$  in view of (2.8). So,  $j_n = O_p(\kappa_{n,\psi}^*)$  and  $\|\hat{\Psi}_{\hat{y}\hat{y},t} - \Psi_t\|_{sp} = O_p(\kappa_{n,\psi}) + O_p(\kappa_{n,\psi}^*) = O_p(\kappa_{n,\psi})$ , which proves (2.20).

To show the asymptotic normality (2.22) of Theorem 2.3, denote by  $i_n$  the l.h.s. of (2.22). Then,  $i_n = (K_t/K_{2,t})^{1/2} \mathbf{H}_{t-1}^{-1} (\hat{\Psi}_{\dot{y}\dot{y},t} - \Psi_t) (\sum_{j=1}^n k_{tj} \hat{\mathbf{y}}_{j-1} \hat{\mathbf{y}}'_{j-1})^{1/2} \mathbf{a} + (K_t/K_{2,t})^{1/2} \mathbf{H}_{t-1}^{-1} (\hat{\Psi}_{\hat{y}\hat{y},t} - \hat{\Psi}_{\dot{y}\dot{y},t}) (\sum_{j=1}^n k_{tj} \hat{\mathbf{y}}_{j-1} \hat{\mathbf{y}}'_{j-1})^{1/2} \mathbf{a} =: i_{n,1} + i_{n,2}$ , where  $i_{n,1} \rightarrow_D \mathcal{N}(0, \mathbf{I})$  by (2.15) of Theorem 2.2. To complete the proof it remains to show  $\|i_{n,2}\| = o_p(1)$ . Bound  $\|i_{n,2}\| \leq (K_t/K_{2,t}^{1/2}) \|\mathbf{H}_{t-1}^{-1}\| \|\hat{\Psi}_{\hat{y}\hat{y},t} - \hat{\Psi}_{\dot{y}\dot{y},t}\|_{sp} \|V_{\hat{y}\hat{y},t}^{1/2}\|_{sp} \|\mathbf{a}\|$ , where  $(K_t/K_{2,t}^{1/2}) = O(H_\psi^{1/2})$  by (6.1),  $\|\hat{\Psi}_{\hat{y}\hat{y},t} - \hat{\Psi}_{\dot{y}\dot{y},t}\|_{sp} = O_p(\kappa_{n,\psi}^*)$  by (6.11),  $\|\mathbf{H}_{t-1}^{-1}\| = O_p(1)$  by Assumption 2.2(iii) and  $\|V_{\hat{y}\hat{y},t}^{1/2}\|_{sp} = \|V_{\dot{y}\dot{y},t}\|_{sp}^{1/2} = O_p(1)$  by Lemma 6.5(iii). Hence,  $i_{n,2} = O(H_\psi^{1/2}) O_p(\kappa_{n,\psi}^*) = o_p(1)$  since  $\kappa_{n,\psi}^* = o(H_\psi^{-1/2})$  when  $H_\psi \bar{H}_\psi = o(n)$ . This proves (2.22).

*Proof of the claims about  $\hat{\alpha}_t$ .* Use relations  $\hat{\alpha}_t = \bar{\mathbf{y}}_t - \hat{\Psi}_t \bar{\mathbf{y}}_t$  and  $\alpha_t = \boldsymbol{\mu}_t - \Psi_{t-1} \boldsymbol{\mu}_{t-1}$  to decompose

$$\begin{aligned} \hat{\alpha}_t - \alpha_t &= (I - \hat{\Psi}_t) \bar{\mathbf{y}}_t - (\boldsymbol{\mu}_t - \Psi_{t-1} \boldsymbol{\mu}_{t-1}) = \{(I - \Psi_t)(\bar{\mathbf{y}}_t - \boldsymbol{\mu}_t) - (\hat{\Psi}_t - \Psi_t) \boldsymbol{\mu}_t\} \\ &\quad - \{(\hat{\Psi}_t - \Psi_t)(\bar{\mathbf{y}}_t - \boldsymbol{\mu}_t) + (\Psi_t \boldsymbol{\mu}_t - \Psi_{t-1} \boldsymbol{\mu}_{t-1})\} =: a_{n,1} - a_{n,2}. \end{aligned} \quad (6.12)$$

To show consistency (2.20), bound  $\|\hat{\boldsymbol{\alpha}}_t - \boldsymbol{\alpha}_t\| \leq \|a_{n,1}\| + \|a_{n,2}\|$ , where  $\|a_{n,1}\| \leq \|I - \boldsymbol{\Psi}_t\|_{sp} \|\bar{\boldsymbol{y}}_t - \boldsymbol{\mu}_t\| + \|\hat{\boldsymbol{\Psi}}_t - \boldsymbol{\Psi}_t\|_{sp} \|\boldsymbol{\mu}_t\| = O_p(\kappa_{n,\psi})$  because  $\|I - \boldsymbol{\Psi}_t\|_{sp} \leq 1 + \|\boldsymbol{\Psi}_t\|_{sp} \leq 1 + r$  by Assumption 2.1,  $\|\bar{\boldsymbol{y}}_t - \boldsymbol{\mu}_t\| = O_p(\kappa_{n,\psi})$  and  $\|\hat{\boldsymbol{\Psi}}_t - \boldsymbol{\Psi}_t\|_{sp} = O_p(\kappa_{n,\psi})$  by Theorem 2.3(i), and  $\|\boldsymbol{\mu}_t\| = O_p(1)$  by Assumption 2.3. In turn,  $\|a_{n,2}\| = O_p(\kappa_{n,\psi})$  follows from

$$\|a_{n,2}\| = o_p(\kappa_{n,\psi}). \quad (6.13)$$

To verify the latter, bound  $\|a_{n,2}\| \leq \|\hat{\boldsymbol{\Psi}}_t - \boldsymbol{\Psi}_t\|_{sp} \|\bar{\boldsymbol{y}}_t - \boldsymbol{\mu}_t\| + \{\|\boldsymbol{\Psi}_t - \boldsymbol{\Psi}_{t-1}\|_{sp} \|\boldsymbol{\mu}_t\| + \|\boldsymbol{\Psi}_{t-1}\|_{sp} \|\boldsymbol{\mu}_t - \boldsymbol{\mu}_{t-1}\|\} =: s_{n,1} + s_{n,2}$  where  $s_{n,1} = O_p(\kappa_{n,\psi}^2)$  by Theorem 2.3(i), while Assumptions 2.2 and 2.3 imply that  $s_{n,2} = o_p((H_\psi/n)^{1/2}) = o_p(\kappa_{n,\psi})$ . Hence  $\|a_{n,2}\| = O_p(\kappa_{n,\psi}^2) + o_p(\kappa_{n,\psi}) = o_p(\kappa_{n,\psi})$  completing the proof of (6.13) and verifying consistency claim (2.20) for  $\hat{\boldsymbol{\alpha}}_t$ .

To show asymptotic normality (2.21), use (6.12) and notation  $\hat{\boldsymbol{D}}_t$  of Remark 2.2, to write  $v_n := (K_t/K_{2,t}^{1/2}) \boldsymbol{H}_{t-1}^{-1} (\hat{\boldsymbol{\alpha}}_t - \boldsymbol{\alpha}_t) \hat{\boldsymbol{D}}_t^{-1/2} = (K_t/K_{2,t}^{1/2}) \boldsymbol{H}_{t-1}^{-1} a_{n,1} \hat{\boldsymbol{D}}_t^{-1/2} - (K_t/K_{2,t}^{1/2}) \boldsymbol{H}_{t-1}^{-1} a_{n,2} \hat{\boldsymbol{D}}_t^{-1/2} =: v_{n,1} - v_{n,2}$ , where  $v_{n,1} \rightarrow_D \mathcal{N}(0, \boldsymbol{I})$  by (6.40) of Lemma 6.4. Therefore, to verify (2.21) it remains to show

$$\|v_{n,2}\| = o_p(1). \quad (6.14)$$

Bound  $\|v_{n,2}\| \leq (K_t/K_{2,t}^{1/2}) \|\boldsymbol{H}_{t-1}^{-1}\|_{sp} \|a_{n,2}\| \|\hat{\boldsymbol{D}}_t^{-1/2}\|$ , where  $K_t/K_{2,t}^{1/2} = O(H_\psi^{1/2})$  by (6.1),  $\|\boldsymbol{H}_{t-1}^{-1}\|_{sp} = O_p(1)$  by Assumption 2.2(iii),  $\|a_{n,2}\| = o_p(\kappa_{n,\psi})$  by (6.13) and  $\|\hat{\boldsymbol{D}}_t^{-1/2}\| = O_p(1)$  by Lemma 6.5(iv). Hence,  $\|v_{n,2}\| = O(H_\psi^{1/2}) o_p(\kappa_{n,\psi}) = o_p(1)$  since  $\kappa_{n,\psi} = O(H_\psi^{-1/2})$  under assumption  $H_\psi \bar{H}_\psi = o(n)$ . This completes the proof of (6.14) and (2.21).

*Proof of the claims about  $\boldsymbol{\Sigma}_{\hat{u}\hat{u},t}$ .* To show (2.20), bound  $\|\boldsymbol{\Sigma}_{\hat{u}\hat{u},t} - \boldsymbol{\Sigma}_t\|_{sp} \leq \|\boldsymbol{\Sigma}_{\hat{u}\hat{u},t} - \boldsymbol{\Sigma}_{uu,t}\|_{sp} + \|\boldsymbol{\Sigma}_{uu,t} - \boldsymbol{\Sigma}_t\|_{sp} = O_p(\kappa_{n,\psi}^2 + \kappa_{n,h}) + O_p(\kappa_{n,h})$  by Lemma 6.5(ii) and Lemma 6.3(i), which proves (2.20).

To show (2.23), write  $(L_t/L_{2,t}^{1/2}) \boldsymbol{H}_{t-1}^{-1} (\boldsymbol{\Sigma}_{\hat{u}\hat{u},t} - \boldsymbol{\Sigma}_t) \boldsymbol{H}'_{t-1} = (L_t/L_{2,t}^{1/2}) \boldsymbol{H}_{t-1}^{-1} (\boldsymbol{\Sigma}_{uu,t} - \boldsymbol{\Sigma}_t) \boldsymbol{H}'_{t-1} + (L_t/L_{2,t}^{1/2}) \boldsymbol{H}_{t-1}^{-1} (\boldsymbol{\Sigma}_{\hat{u}\hat{u},t} - \boldsymbol{\Sigma}_{uu,t}) \boldsymbol{H}'_{t-1} =: q_{n,1} + q_{n,2}$ . By Lemma 6.3(ii),  $q_{n,1} \rightarrow_D \boldsymbol{Z}$ , while  $\|q_{n,2}\|_{sp} = o_p(1)$ . Indeed,  $\|q_{n,2}\|_{sp} \leq C H_h^{1/2} \|\boldsymbol{H}_{t-1}^{-1}\|_{sp}^2 \|\boldsymbol{\Sigma}_{\hat{u}\hat{u},t} -$

$\Sigma_{uu,t}\|_{sp}$  where  $L_t/L_{2,t}^{1/2} = O(H_h^{1/2})$ ,  $\|\mathbf{H}_{t-1}^{-1}\|_{sp} = O_p(1)$  and  $\|\Sigma_{\hat{u}\hat{u},t} - \Sigma_{uu,t}\|_{sp} = O_p(\kappa_{n,\psi}^2 + (\bar{H}_h/n)^{1/2} + H_h^{-1})$  by Lemma 6.5(ii). Hence,  $\|q_{n,2}\|_{sp} = O(H_h^{1/2})O_p(\kappa_{n,\psi}^2 + (\bar{H}_h/n)^{1/2} + H_h^{-1}) = o_p(1)$ , because  $H_h\bar{H}_h = o(n)$  implies  $(\bar{H}_h/n)^{1/2} = o_p(H_h^{-1/2})$ , while assumption  $H_h^{1/2} \ll H_\psi \ll n/(H_h \log n)^{1/2}$  ensures that  $\kappa_{n,\psi}^2 = o_p(H_h^{-1/2})$ . This completes the proof of (2.23) and the theorem.  $\square$

**Proof of Proposition 2.2.** By (2.25),  $\boldsymbol{\mu}_t = \sum_{j=0}^{t-1} \Pi_{t,j} \boldsymbol{\alpha}_{t-j}$ . Write

$$\begin{aligned} \boldsymbol{\mu}_{t+k} - \boldsymbol{\mu}_t &= \sum_{j=0}^{t+k-1} \Pi_{t+k,j} \boldsymbol{\alpha}_{t+k-j} - \sum_{j=0}^{t-1} \Pi_{t,j} \boldsymbol{\alpha}_{t-j} = \left\{ \sum_{j=0}^{t-1} (\Pi_{t+k,j} - \Pi_{t,j}) \boldsymbol{\alpha}_{t-j} \right\} \\ &+ \left\{ \sum_{j=0}^{t-1} \Pi_{t+k,j} (\boldsymbol{\alpha}_{t+k-j} - \boldsymbol{\alpha}_{t-j}) + \sum_{j=t}^{t+k-1} \Pi_{t+k,j} \boldsymbol{\alpha}_{t+k-j} \right\} := \tilde{m}(t, k) + m(t, k). \end{aligned}$$

It remains to show that  $\tilde{m}(t, k)$  and  $m(t, k)$  satisfy conditions of Assumption 2.3. Recall that  $\|\Pi_{t+k,j}\|_{sp} \leq r^j$  by (2.6), and  $\max_j E\|\boldsymbol{\alpha}_j\| < \infty$  by Assumption 2.3.

$$\text{Bound } \|\tilde{m}(t, k)\| \leq \sum_{j=0}^h \|\Pi_{t+k,j} - \Pi_{t,j}\|_{sp} \|\boldsymbol{\alpha}_{t-j}\| + \sum_{j=h+1}^{t-1} r^j \|\boldsymbol{\alpha}_{t-j}\|.$$

By same argument as in the proof of (6.2), for  $k \leq h$ ,  $\max_{j=1, \dots, h} \|\Pi_{t+k,j} - \Pi_{t,j}\|_{sp} \leq Cj r^{j-1} \max_{s:|s-t| \leq 3h} \|\Psi_t - \Psi_s\|_{sp} =: Cj r^{j-1} R_{t,3h}$ . Therefore,  $\|\tilde{m}(t, k)\| \leq C R_{t,3h} \sum_{j=1}^h j r^{j-1} \|\boldsymbol{\alpha}_{t-j}\| + 2 \sum_{j=h+1}^{t-1} r^j \|\boldsymbol{\alpha}_{t-j}\| =: C R_{t,3h} s_{n,h,1} + s_{n,h,2}$ . By (2.4),  $R_{t,3h} = O_p((h/t)^{1/2})$ , while  $E s_{n,h,1} \leq C \sum_{j=1}^{\infty} j r^{j-1} < \infty$  implies  $s_{n,h,1} = O_p(1)$  and  $E s_{n,h,2} \leq C \sum_{j=h+1}^{\infty} r^j \leq C r^h \leq C h^{-1}$  implies  $s_{n,h,2} = O_p(h^{-1})$ . Hence  $\max_{k \leq h} \|\tilde{m}(t, k)\| = O_p(R_{t,3h} + h^{-1}) = O_p((h/t)^{1/2} + h^{-1})$ , Therefore  $m(t, k)$  is as in Assumption 2.3.

Bound  $\|m(t, k)\| \leq \sum_{j=0}^{t-1} r^j \|\boldsymbol{\alpha}_{t+k-j} - \boldsymbol{\alpha}_{t-j}\| + \sum_{j=t}^{t+k-1} r^j \|\boldsymbol{\alpha}_{t+k-j}\|$ . By Assumption 2.3,  $E\|\boldsymbol{\alpha}_{t+k-j} - \boldsymbol{\alpha}_{t-j}\|^2 \leq C(k/(t-j)) \leq C(k/t)$  for  $j \leq t/2$ . Thus,

$$\begin{aligned} E\|m(t, k)\|^2 &\leq 2 \left( \left( \sum_{j=0}^{t/2} r^j E\|\boldsymbol{\alpha}_{t+k-j} - \boldsymbol{\alpha}_{t-j}\| \right)^2 + \left( \sum_{j=t/2}^{t-1} r^j E\|\boldsymbol{\alpha}_{t+k-j} - \boldsymbol{\alpha}_{t-j}\| \right)^2 \right) \\ &+ \left( \sum_{j=t}^{t+k-1} r^j E\|\boldsymbol{\alpha}_{t+k-j}\| \right)^2 \leq C \left( (k/t) \left( \sum_{j=0}^{t/2} r^j \right)^2 + \left( \sum_{j=t/2}^{\infty} r^j \right)^2 \right) \\ &\leq C((k/t) + r^t) \leq C(k/t), \end{aligned}$$

since  $r^t \leq C t^{-1}$ ,  $t \geq 0$  for  $0 < r < 1$ . Therefore  $m(t, k)$  satisfies condition of Assumption 2.3.  $\square$

## 6.2 Auxiliary lemmas

This section contains auxiliary lemmas used to prove Theorems 2.1-2.3.

**Lemma 6.1.** *Under assumptions of Theorem 2.2(i), with  $\mathbf{V}_{\psi,t}$  as in (2.17),*

$$\begin{aligned}
(i) \quad & \|\mathbf{V}_{yy,t} - \mathbf{V}_{\psi,t}\|_{sp} = o_p(1), \quad (ii) \quad \|\mathbf{V}_{\psi,t}^{-1}\|_{sp} = O_p(1), \quad (6.15) \\
(iii) \quad & \|\mathbf{V}_{yy,t}^{-1} - \mathbf{V}_{\psi,t}^{-1}\|_{sp} = o_p(1); \quad (iv) \quad \|\mathbf{V}_{yy,t}^{-1/2} - \mathbf{V}_{\psi,t-h}^{-1/2}\|_{sp} = o_p(1), \quad h = o(t); \\
(v) \quad & \|\mathbf{V}_{yy,t}^{-1} - \mathbf{V}_{\psi,t-h}^{-1}\|_{sp} = o_p(1), \quad h = o(t), \\
(vi) \quad & \|\mathbf{V}_{yy,t}^{-1}\|_{sp} = O_p(1), \quad \|\mathbf{V}_{yy,t}^{-1/2}\|_{sp} = O_p(1), \quad \|\mathbf{V}_{yy,t}\|_{sp} = O_p(1).
\end{aligned}$$

**Proof.** First we show (6.15)(i). Write  $\mathbf{V}_{yy,t} = \Psi_t \mathbf{V}_{yy,t} \Psi_t + \Sigma_t + r_t$ , where  $r_t := (\Sigma_{uu,t}^* - \Sigma_t) + (\mathbf{V}_{yy,t} - \Psi_t \mathbf{V}_{yy,t} \Psi_t - \Sigma_{uu,t}^*)$ , where  $\Sigma_{uu,t}^* := K_t^{-1} \sum_{j=1}^n k_{tj} \mathbf{u}_j \mathbf{u}_j'$  is computed using weights  $k_{tj}$ . (Recall that in  $\Sigma_{uu,t}$  of (6.4) the weights  $l_{tj}$  are used.)

By recursion,

$$\mathbf{V}_{yy,t} = \Psi_t^p \mathbf{V}_{yy,t} \Psi_t'^p + \sum_{k=0}^{p-1} \Psi_t^k \Sigma_t \Psi_t'^k + \sum_{k=0}^{p-1} \Psi_t^k r_t \Psi_t'^k.$$

Recall that  $\|\Psi_t\|_{sp} \leq r < 1$  by Assumption 2.1. Hence, for  $p \geq 1$ ,

$$\begin{aligned}
\|\mathbf{V}_{yy,t} - \mathbf{V}_{\psi,t}\|_{sp} & \leq \|\Psi_t^p \mathbf{V}_{yy,t} \Psi_t'^p\|_{sp} + \sum_{k=p}^{\infty} \|\Psi_t^k \Sigma_t \Psi_t'^k\|_{sp} + \|\sum_{k=0}^{p-1} \Psi_t^k r_t \Psi_t'^k\|_{sp} \\
& \leq \|\Psi_t^p\|_{sp}^2 \|\mathbf{V}_{yy,t}\|_{sp} + \sum_{k=p}^{\infty} \|\Psi_t\|_{sp}^{2k} \|\Sigma_t\|_{sp} + \sum_{k=0}^{p-1} \|\Psi_t\|_{sp}^{2k} \|r_t\|_{sp} \\
& \leq r^{2p} \|\mathbf{V}_{yy,t}\|_{sp} + \|\Sigma_t\|_{sp} \sum_{k=p}^{\infty} r^{2k} + \|r_t\|_{sp} \sum_{k=0}^{p-1} r^{2k} \\
& \leq r^{2p} (\|\mathbf{V}_{yy,t}\|_{sp} + (1 - r^2)^{-1} \|\Sigma_t\|_{sp}) + (1 - r^2)^{-1} \|r_t\|_{sp}.
\end{aligned}$$

Notice that  $\|\Sigma_t\|_{sp} \leq \|\mathbf{H}_{t-1}\|_{sp}^2 = O_p(1)$  by Assumption 2.2(ii), and  $E\|\mathbf{V}_{yy,t}\|_{sp} \leq L_t^{-1} \sum_{j=1}^n k_{tj} E\|\mathbf{y}_{j-1}\|^2 \leq CK_t^{-1} \sum_{j=1}^n k_{tj} \leq C$  by (2.8). This implies (6.15)(i):  $\|\mathbf{V}_{yy,t} - \mathbf{V}_{\psi,t}\|_{sp} = O_p(1)(r^{2p} + \|r_t\|_{sp}) = o_p(1)$  because of  $r^{2p} \rightarrow 0$  as  $p \rightarrow \infty$  and

$$\|r_t\|_{sp} = o_p(1). \quad (6.16)$$

To verify (6.16), we bound  $\|r_t\|_{sp} \leq \|\Sigma_{uu,t}^* - \Sigma_t\|_{sp} + \|\mathbf{V}_{yy,t} - \Psi_t \mathbf{V}_{yy,t} \Psi_t - \Sigma_{uu,t}^*\|_{sp} =: r_{t,1} + r_{t,2}$  and show that  $\|r_{t,i}\|_{sp} = o_p(1)$ ,  $i = 1, 2$ . By (6.32),  $r_{t,1} = O_p((\bar{H}_\psi/n)^{1/2} + H_\psi^{-1/2}) = o_p(1)$ . To evaluate  $r_{t,2}$ , use (2.1) to write  $\mathbf{y}_j \mathbf{y}'_j = (\Psi_{j-1} \mathbf{y}_{j-1} + \mathbf{u}_j)(\mathbf{y}'_{j-1} \Psi'_{j-1} + \mathbf{u}'_j)$ , to obtain  $\mathbf{y}_{j-1} \mathbf{y}'_{j-1} - \Psi_t \mathbf{y}_{j-1} \mathbf{y}'_{j-1} \Psi'_t - \mathbf{u}_j \mathbf{u}'_j = \{\mathbf{y}_{j-1} \mathbf{y}'_{j-1} - \mathbf{y}_j \mathbf{y}'_j\} + \{\Psi_{j-1} \mathbf{y}_{j-1} \mathbf{u}'_j + \mathbf{u}_j \mathbf{y}'_{j-1} \Psi'_{j-1}\} + \{\Psi_{j-1} \mathbf{y}_{j-1} \mathbf{y}'_{j-1} \Psi'_{j-1} - \Psi_t \mathbf{y}_{j-1} \mathbf{y}'_{j-1} \Psi'_t\} =: z_{tj,1} + z_{tj,2} + z_{tj,3}$ . Use this to write

$$\begin{aligned} \mathbf{V}_{yy,t} - \Psi_t \mathbf{V}_{yy,t} \Psi_t - \Sigma_{uu,t} &= K_t^{-1} \sum_{j=1}^n k_{tj} (\mathbf{y}_{j-1} \mathbf{y}'_{j-1} - \Psi_t \mathbf{y}_{j-1} \mathbf{y}'_{j-1} \Psi'_t - \mathbf{u}_j \mathbf{u}'_j) \\ &= T_{n,1} + T_{n,2} + T_{n,3}, \quad T_{n,i} = H_t^{-1} \sum_{j=1}^n k_{tj} z_{tj,i}, \quad i = 1, 2, 3. \end{aligned}$$

Hence,  $\|r_{t,2}\|_{sp} \leq \|T_{n,1}\|_{sp} + \|T_{n,2}\|_{sp} + \|T_{n,3}\|_{sp}$ , and it remains to verify that

$$\|T_{n,i}\|_{sp} = o_p(1), \quad i = 1, 2, 3. \quad (6.17)$$

Observe that  $\|T_{n,1}\|_{sp} = K_t^{-1} \|\sum_{j=1}^n (k_{tj} - k_{t,j-1}) \mathbf{y}_{j-1} \mathbf{y}'_{j-1} - k_{tn} \mathbf{y}_n \mathbf{y}'_n + k_{t0} \mathbf{y}_0 \mathbf{y}'_0\|_{sp} \leq K_t^{-1} \sum_{j=1}^n |k_{tj} - k_{t,j-1}| \|\mathbf{y}_{j-1}\|^2 + k_{tn} \|\mathbf{y}_n\|^2 + k_{t0} \|\mathbf{y}_0\|^2$ . By Assumption 2.2(i) and (2.8),  $\max_{j \geq 0} E \|\mathbf{y}_j\|^2 < \infty$ ; by (6.1)  $K_t^{-1} \leq H_\psi^{-1}$ , while by (2.10) and the mean value theorem,  $|k_{tj} - k_{t,j-1}| \leq C H_\psi^{-1} (1 + (j/H_\psi)^2)^{-1}$ . Hence,

$$\begin{aligned} E \|T_{n,1}\|_{sp} &\leq C K_t^{-1} (\sum_{j=1}^n |k_{tj} - k_{t,j-1}| + C) \\ &\leq C H_\psi^{-2} \sum_{j=1}^n (1 + (j/H_\psi)^2)^{-1} + O(H_\psi^{-1}) = O(H_\psi^{-1}), \end{aligned}$$

since  $\sum_{j=1}^n (\dots)^{-1} \leq \sum_{j=1}^{H_\psi} 1 + H_\psi^2 \sum_{j=H_\psi}^\infty j^{-2} \leq C H_\psi^{-1}$ , which implies (6.17) for  $i = 1$ .

Since  $\mathbf{y}_j = \mathbf{H}_{j-1} \boldsymbol{\varepsilon}_j$  where  $\boldsymbol{\varepsilon}_j$  i.i.d. random vectors, it is easy to verify that  $\|T_{n,2}\|_{sp} = O_p(H_\psi^{-1/2}) = o_p(1)$ , see e.g. proof of (6.23). Finally, to bound  $\|T_{n,3}\|_{sp}$  note that  $\|z_{tj,3}\|_{sp} = \|(\Psi_{j-1} - \Psi_t) \mathbf{y}_{j-1} \mathbf{y}'_{j-1} \Psi'_{j-1} + \Psi_t \mathbf{y}_{j-1} \mathbf{y}'_{j-1} (\Psi'_{j-1} - \Psi'_t) + (\Psi_{j-1} - \Psi_t) \mathbf{y}_{j-1} \mathbf{y}'_{j-1} (\Psi'_{j-1} - \Psi'_t)\|_{sp} \leq 2 \|\Psi_{j-1} - \Psi_t\|_{sp} \|\Psi_t\|_{sp} \|\mathbf{y}_{j-1}\|^2 + \|\Psi_{j-1} - \Psi_t\|_{sp}^2 \|\mathbf{y}_{j-1}\|^2 \leq 4r \|\Psi_{j-1} - \Psi_t\|_{sp} \|\mathbf{y}_{j-1}\|^2$ . Therefore, by the same argument as in the proof of (6.7), it follows that  $\|T_{n,3}\|_{sp} = O_p((\bar{H}_\psi/n)^{1/2} + H_\psi^{-1}) = o_p(1)$ , which completes the proof of (6.17) and the claim (6.15)(i).



To verify (6.15)(ii), we show that

$$\|\mathbf{V}_{\psi,t}^{-1}\|_{sp} \leq \|\boldsymbol{\Sigma}_t^{-1}\|_{sp} = O_p(1). \quad (6.18)$$

Recall that  $\boldsymbol{\Sigma}_t = \mathbf{H}_{t-1} \mathbf{H}'_{t-1}$ , and  $\|\mathbf{H}_{t-1}^{-1}\|_{sp} = O_p(1)$  by Assumption 2.2(iii), which implies  $\|\boldsymbol{\Sigma}_t^{-1}\|_{sp} = \|\mathbf{H}_{t-1}^{-1}\|_{sp}^2 = O_p(1)$ . Moreover,  $\|\boldsymbol{\Sigma}_t^{-1}\|_{sp} = \lambda_{\min,\Sigma}^{-1}$  and  $\|\mathbf{V}_{\psi,t}^{-1}\|_{sp} = \lambda_{\min,\psi}^{-1}$  where  $\lambda_{\min,\Sigma} = \inf_{\|\mathbf{x}\|=1} \mathbf{x}' \boldsymbol{\Sigma}_t \mathbf{x}$  and  $\lambda_{\min,\psi} = \inf_{\|\mathbf{x}\|=1} \mathbf{x}' \mathbf{V}_{\psi,t} \mathbf{x}$  are the smallest eigenvalues of  $\boldsymbol{\Sigma}_t$  and  $\mathbf{V}_{\psi,t}$ , respectively. To prove (6.18) it remains to show that  $\lambda_{\min,\psi} \geq \lambda_{\min,\Sigma}$ . For any real  $m$ -dimensional vector  $\mathbf{x}$ ,  $\|\mathbf{x}\| = 1$ ,

$$\mathbf{x}' \mathbf{V}_{\psi,t} \mathbf{x} = \sum_{k=0}^{\infty} (\mathbf{x}' \boldsymbol{\Psi}_t^k) \boldsymbol{\Sigma}_t (\mathbf{x}' \boldsymbol{\Psi}_t^k)' \geq \lambda_{\min,\Sigma} \sum_{k=0}^{\infty} \|\mathbf{x}' \boldsymbol{\Psi}_t^k\|^2 \geq \lambda_{\min,\Sigma},$$

which implies  $\lambda_{\min,\psi} \geq \lambda_{\min,\Sigma}$  completing the proof of (6.15)(ii).

To prove (6.15)(iii), write  $\mathbf{V}_{yy,t} = \mathbf{V}_{\psi,t}(I + \Delta_t)$ ,  $\Delta_t := \mathbf{V}_{\psi,t}^{-1}(\mathbf{V}_{yy,t} - \mathbf{V}_{\psi,t})$ . Notice that  $\|\Delta_t\|_{sp} \leq \|\mathbf{V}_{\psi,t}^{-1}\|_{sp} \|\mathbf{V}_{yy,t} - \mathbf{V}_{\psi,t}\|_{sp} = o_p(1)$  by (6.15)(ii) and (i). Therefore  $\mathbf{V}_{yy,t}^{-1} - \mathbf{V}_{\psi,t}^{-1} = \mathbf{V}_{\psi,t}^{-1}((I + \Delta_t)^{-1} - I)$ , and  $\|\mathbf{V}_{yy,t}^{-1} - \mathbf{V}_{\psi,t}^{-1}\|_{sp} \leq \|\mathbf{V}_{\psi,t}^{-1}\|_{sp} \|(I + \Delta_t)^{-1} - I\|_{sp} = o_p(1)$  since  $\|\mathbf{V}_{\psi,t}^{-1}\|_{sp} = O_p(1)$  by (6.15)(ii) and  $\|(I + \Delta_t)^{-1} - I\|_{sp} = O_p(\|\Delta_t\|_{sp}/(1 - \|\Delta_t\|_{sp})) = o_p(1)$ . This completes the proof of (iii).

To show (6.15)(iv), bound  $\|\mathbf{V}_{yy,t}^{-1/2} - \mathbf{V}_{\psi,t-h}^{-1/2}\|_{sp} \leq \|\mathbf{V}_{yy,t}^{-1/2} - \mathbf{V}_{\psi,t}^{-1/2}\|_{sp} + \|\mathbf{V}_{\psi,t}^{-1/2} - \mathbf{V}_{\psi,t-h}^{-1/2}\|_{sp}$ . It remains to show that

$$(a) \|\mathbf{V}_{yy,t}^{-1/2} - \mathbf{V}_{\psi,t}^{-1/2}\|_{sp} = o_p(1), \quad (b) \|\mathbf{V}_{\psi,t}^{-1/2} - \mathbf{V}_{\psi,t-h}^{-1/2}\|_{sp} = o_p(1). \quad (6.19)$$

To prove (a), we use the following results (xii) and (xi) of Davies (1973) p.496: for any positive definite symmetric matrices  $A$  and  $B$ , such that  $\|A - B\|_{sp} \|A^{-1}\|_{sp} \leq \varepsilon < 1/2$ , it holds

$$\|A^{1/2} - B^{1/2}\|_{sp} \|A^{-1/2}\|_{sp} \leq \varepsilon. \quad (6.20)$$

Moreover,  $\|A^{1/2}\|_{sp} = \|A\|_{sp}^{1/2}$ , which implies  $\|A^{-1/2}\|_{sp} = 1/\|A\|_{sp}^{1/2}$ .

The matrixes  $A := \mathbf{V}_{\psi,t}$  and  $B := \mathbf{V}_{yy,t}$  are positive definite and have property  $\|A - B\|_{sp} = o_p(1)$  by (6.15)(i) and  $\|A^{-1}\|_{sp} = O_p(1)$  by Lemma 6.1(ii). Therefore,

by (6.20)  $\|\mathbf{V}_{yy,t}^{1/2} - \mathbf{V}_{\psi,t}^{1/2}\|_{sp} \|\mathbf{V}_{\psi,t}^{-1/2}\|_{sp} = o_p(1)$ . Thus,

$$\begin{aligned} \|\mathbf{V}_{yy,t}^{-1/2} - \mathbf{V}_{\psi,t}^{-1/2}\|_{sp} &= \|\mathbf{V}_{yy,t}^{-1/2}(\mathbf{V}_{\psi,t}^{1/2} - \mathbf{V}_{yy,t}^{1/2})\mathbf{V}_{\psi,t}^{-1/2}\|_{sp} \\ &\leq \|\mathbf{V}_{yy,t}^{-1/2}\|_{sp} \|\mathbf{V}_{\psi,t}^{1/2} - \mathbf{V}_{yy,t}^{1/2}\|_{sp} \|\mathbf{V}_{\psi,t}^{-1/2}\|_{sp} = \|\mathbf{V}_{yy,t}^{-1/2}\|_{sp} o_p(1) = o_p(1), \end{aligned} \quad (6.21)$$

since  $\|\mathbf{V}_{yy,t}^{-1/2}\|_{sp} = \|\mathbf{V}_{yy,t}^{-1}\|_{sp}^{1/2} = O_p(1)$  by Lemma 6.1(vi), shown below.

The claim (b) can be verified applying the same argument as in the proof of (a), noting that  $\|\mathbf{V}_{\psi,t}^{-1/2}\|_{sp} = O_p(1)$  and  $\|\mathbf{V}_{\psi,t-h}^{-1/2}\|_{sp} = O_p(1)$  by Lemma (6.1)(ii), and verifying that  $\|\mathbf{V}_{\psi,t} - \mathbf{V}_{\psi,t-h}\|_{sp} = o_p(1)$ . To show the latter, observe that  $\|\mathbf{V}_{\psi,t} - \mathbf{V}_{\psi,t_0}\|_{sp} \leq \sum_{k=0}^{\infty} \|\Psi_t^k \Sigma_t \Psi_t'^k - \Psi_{t_0}^k \Sigma_{t_0} \Psi_{t_0}'^k\|_{sp}$  where  $t_0 = t - h$ . Using  $\|\Psi_t\|_{sp} \leq r < 1$  of Assumption 2.1 together with (6.2) type bound gives

$$\begin{aligned} \|\mathbf{V}_{\psi,t} - \mathbf{V}_{\psi,t_0}\|_{sp} &\leq (\sum_{k=1}^{\infty} k r^{2k-1}) \|\Psi_t - \Psi_{t_0}\|_{sp} (\|\Sigma_t\|_{sp} + \|\Sigma_{t_0}\|_{sp}) \\ &\quad + (\sum_{k=0}^{\infty} r^{2k}) \|\Sigma_t - \Sigma_{t_0}\|_{sp} = o_p(1), \end{aligned} \quad (6.22)$$

since  $\|\Psi_t - \Psi_{t_0}\|_{sp} = o_p(1)$  by Assumption 2.1, whereas by Assumption 2.2(ii),  $\|\Sigma_t\|_{sp} = O_p(1)$ ,  $\|\Sigma_{t_0}\|_{sp} = O_p(1)$  and  $\|\Sigma_t - \Sigma_{t_0}\|_{sp} = o_p(1)$ . This proves (b) and completes the proof of (6.15)(iv).

To show (6.15)(v), bound  $\|\mathbf{V}_{yy,t}^{-1} - \mathbf{V}_{\psi,t-h}^{-1}\|_{sp} \leq \|\mathbf{V}_{yy,t}^{-1} - \mathbf{V}_{\psi,t}^{-1}\|_{sp} + \|\mathbf{V}_{\psi,t}^{-1} - \mathbf{V}_{\psi,t-h}^{-1}\|_{sp} = o_p(1)$ , where  $\|\mathbf{V}_{yy,t}^{-1} - \mathbf{V}_{\psi,t}^{-1}\|_{sp} = o_p(1)$  by Lemma 6.1(iii), and  $\|\mathbf{V}_{\psi,t}^{-1} - \mathbf{V}_{\psi,t-h}^{-1}\|_{sp} \leq \|\mathbf{V}_{\psi,t-h}^{-1}\|_{sp} \|\mathbf{V}_{\psi,t} - \mathbf{V}_{\psi,t-h}\|_{sp} \|\mathbf{V}_{\psi,t}^{-1}\|_{sp} = o_p(1)$  by Lemma 6.1(ii) and (6.22).

To show (6.15)(vi), notice that  $\|\mathbf{V}_{yy,t}^{-1}\|_{sp} \leq \|\mathbf{V}_{yy,t}^{-1} - \mathbf{V}_{\psi,t}^{-1}\|_{sp} + \|\mathbf{V}_{\psi,t}^{-1}\|_{sp} = O_p(1)$  by (6.15)(ii)-(iii), which also implies  $\|\mathbf{V}_{yy,t}^{-1/2}\|_{sp} = \|\mathbf{V}_{yy,t}^{-1}\|_{sp}^{1/2} = O_p(1)$ . Finally,  $\|\mathbf{V}_{yy,t}\|_{sp} \leq \|\mathbf{V}_{yy,t} - \mathbf{V}_{\psi,t}\|_{sp} + \|\mathbf{V}_{\psi,t}\|_{sp} = O_p(1)$  by (6.1)(i) noting that  $\|\mathbf{V}_{\psi,t}\|_{sp} \leq \sum_{k=0}^{\infty} \|\Psi_t^k \Sigma_t \Psi_t'^k\|_{sp} \leq \|\Sigma_t\|_{sp} \sum_{k=0}^{\infty} r^{2k} = O_p(1)$ . This completes the proof of the lemma.  $\square$

**Lemma 6.2.** *Let  $\mathbf{S}_{yy,t}$  be as in (6.6).*

(i) Under assumptions of Theorem 2.2(i),

$$(a) \quad \|\mathbf{S}_{uy,t}\| = O_p(H_\psi^{-1/2}), \quad (b) \quad \|K_t^{-1} \sum_{j=1}^n k_{tj} \mathbf{u}_j\| = O_p(H_\psi^{-1/2}). \quad (6.23)$$

(ii) Under assumptions of Theorem 2.2(ii), for any  $m \times 1$  vector  $\mathbf{a}$  such that  $\|\mathbf{a}\| = 1$ ,

$$(K_t/K_{2,t}^{1/2}) \mathbf{H}_{t-1}^{-1} \mathbf{S}_{uy,t} \mathbf{V}_{yy,t}^{-1/2} \mathbf{a} \rightarrow_D \mathcal{N}(0, \mathbf{I}). \quad (6.24)$$

**Proof.** (i) To show (6.23)(a) note that by definition (2.5),  $E(\mathbf{u}'_j \mathbf{u}_k | \mathbf{y}_{j-1}, \mathbf{y}_{k-1}) = 0$  if  $j \neq k$ . Thus,

$$\begin{aligned} E\|\mathbf{S}_{uy,t}\|^2 &= K_t^{-2} \text{trace}(E(\mathbf{S}'_{uy,t} \mathbf{S}_{uy,t})) = K_t^{-2} \text{trace}(E \sum_{j=1}^n k_{tj}^2 \mathbf{y}_{j-1} \mathbf{u}'_j \mathbf{u}_j \mathbf{y}'_{j-1}) \\ &\leq K_t^{-2} \sum_{j=1}^n k_{tj}^2 E(\|\mathbf{y}_{j-1}\|^2 \|\mathbf{u}_j\|^2) \leq CK_t^{-2} \sum_{j=1}^n k_{tj}^2 \leq CH_\psi^{-1} \end{aligned}$$

by (6.1), since  $\max_{j \geq 0} (E(\|\mathbf{y}_j\|^4) + E\|\mathbf{u}_j\|^4) < \infty$  by (2.8) and Assumption 2.2. The same argument implies (6.23)(b).

(ii) In view of the Cramér-Wold device, it suffices to verify that for any  $m \times 1$  vector  $b$ , the scalar random variable  $p_{n,t} := (K_t/K_{2,t}^{1/2}) b' \mathbf{H}_{t-1}^{-1} \mathbf{S}_{uy,t} \mathbf{V}_{yy,t}^{-1/2} \mathbf{a}$  has property

$$p_{n,t} \rightarrow_D \mathcal{N}(0, b'b).$$

Set  $t_0 = t - h$ , where  $h = b\bar{H}_\psi$  is such that (6.1) holds. With  $\mathbf{V}_{\psi,t_0} = \sum_{k=0}^{\infty} \Psi_{t_0}^k \Sigma_{t_0} \Psi_{t_0}'^k$  defined as in Remark 2.1 and  $\mathbf{S}_{\varepsilon y,t} = K_t^{-1} \sum_{j=1}^n k_{tj} \boldsymbol{\varepsilon}_j \mathbf{y}'_{j-1}$ , write

$$\begin{aligned} p_{n,t} &= (K_t/K_{2,t}^{1/2}) b' \mathbf{S}_{\varepsilon y,t} \mathbf{V}_{\psi,t_0}^{-1/2} \mathbf{a} \\ &\quad + (K_t/K_{2,t}^{1/2}) b' [(\mathbf{H}_{t-1}^{-1} \mathbf{S}_{uy,t} - \mathbf{S}_{\varepsilon y,t}) \mathbf{V}_{\psi,t_0}^{-1/2} + \mathbf{S}_{uy,t} (\mathbf{V}_{yy,t}^{-1/2} - \mathbf{V}_{\psi,t_0}^{-1/2})] \mathbf{a} =: p_{n,t;1} + p_{n,t;2}. \end{aligned}$$

It suffices to show that

$$(a) \quad p_{n,t;1} \rightarrow_D \mathcal{N}(0, b'b), \quad (b) \quad p_{n,t;2} \rightarrow_p 0. \quad (6.25)$$

*Proof of (6.25)(a).* Setting  $\xi_{tj} := K_{2,t}^{-1/2} b' \boldsymbol{\varepsilon}_j \mathbf{y}'_{j-1} \mathbf{V}_{\psi,t_0}^{-1/2} \mathbf{a}$ , write

$$p_{n,t;1} = \sum_{j=1}^n k_{tj} \xi_{tj} = \sum_{|j-t| \leq h} k_{tj} \xi_{tj} + \sum_{|j-t| \geq h} k_{tj} \xi_{tj} =: q_{n,1} + q_{n,2}.$$

Observe that  $|q_{n,2}| \leq \|b\| \|\mathbf{V}_{\psi,t_0}^{-1/2}\|_{sp}^2 K_{2,t}^{-1/2} \sum_{|j-t| \geq h} k_{tj} \|\boldsymbol{\varepsilon}_j \mathbf{y}_{j-1}\| = o_p(1)$  since

$$\|\mathbf{V}_{\psi,t_0}^{-1/2}\|_{sp} = \|\mathbf{V}_{\psi,t_0}^{-1}\|_{sp}^{1/2} = O_p(1) \quad (6.26)$$

by Lemma 6.1(ii), and  $E \sum_{|j-t| \geq h} k_{tj} \|\boldsymbol{\varepsilon}_j \mathbf{y}_{j-1}\| \leq C \sum_{|j-t| \geq h} k_{tj} = o(1)$  by (6.1). It remains to show  $q_{n,1} \rightarrow_D \mathcal{N}(0, b'b)$ . In view of (2.5) and definition of  $\mathbf{V}_{\psi,t_0}^{-1}$ ,  $q_{n,1}$  is a sum of martingale differences  $\xi_{tj}$ ,  $E[\xi_{tj} | \mathcal{F}_{j-1}] = 0$ ,  $|j-t| \leq h$ . By the central limit theorem for martingale differences (see Corollary 3.1 in Hall and Heyde (1980)) it suffices to show that for any  $\epsilon > 0$ ,

$$j_n := \sum_{|t-j| < h} k_{tj}^2 E[\xi_{tj}^2 | \mathcal{F}_{j-1}] \rightarrow_p b'b, \quad (6.27)$$

$$v_{n,\epsilon} := \sum_{|t-j| < h} E[k_{tj}^2 \xi_{tj}^2 I(k_{tj}^2 \xi_{tj}^2 > \epsilon) | \mathcal{F}_{j-1}] \rightarrow 0. \quad (6.28)$$

By definition  $\mathbf{V}_{\psi,t_0}^{-1/2}$  is  $\mathcal{F}_{t_0}$  measurable. Therefore

$$E[\xi_{tj}^2 | \mathcal{F}_{j-1}] = K_{2,t}^{-1} E(b' \boldsymbol{\varepsilon}_j)^2 \mathbf{a}' \mathbf{V}_{\psi,t_0}^{-1/2} \mathbf{y}_{j-1} \mathbf{y}'_{j-1} \mathbf{V}_{\psi,t_0}^{-1/2} \mathbf{a} \text{ for } j > t - h = t_0$$

where  $E(b' \boldsymbol{\varepsilon}_j)^2 = \|b\|^2$ . Setting  $\tilde{V}_{yy,t} := K_{2,t}^{-1} \sum_{|t-j| < h} k_{tj}^2 \mathbf{y}_{j-1} \mathbf{y}'_{j-1}$ , we obtain

$$j_n := \|b\|^2 \mathbf{a}' \mathbf{V}_{\psi,t_0}^{-1/2} \tilde{V}_{yy,t} \mathbf{V}_{\psi,t_0}^{-1/2} \mathbf{a} = \|b\|^2 + r_n,$$

where  $r_n = \|b\|^2 \mathbf{a}' \mathbf{V}_{\psi,t_0}^{-1/2} (\tilde{V}_{yy,t} - \mathbf{V}_{\psi,t_0}) \mathbf{V}_{\psi,t_0}^{-1/2} \mathbf{a}$ . To prove (6.27) it remains to show that  $r_n \rightarrow_p 0$ . Note that  $|r_n| \leq \|b\|^2 \|\mathbf{V}_{\psi,t_0}^{-1/2}\|_{sp}^2 \|\tilde{V}_{yy,t} - \mathbf{V}_{\psi,t_0}\|_{sp} = O_p(1) \|\tilde{V}_{yy,t} - \mathbf{V}_{\psi,t_0}\|_{sp}$ . Hence, it remains to check that

$$\|\tilde{V}_{yy,t} - \mathbf{V}_{\psi,t_0}\|_{sp} = o_p(1). \quad (6.29)$$

Bound  $\|\tilde{V}_{yy,t} - \mathbf{V}_{\psi,t_0}\|_{sp} \leq \|K_{2,t}^{-1} \sum_{j=1}^n k_{tj}^2 \mathbf{y}_{j-1} \mathbf{y}'_{j-1} - \mathbf{V}_{\psi,t}\|_{sp} + \|\mathbf{V}_{\psi,t} - \mathbf{V}_{\psi,t_0}\|_{sp} + \|K_{2,t}^{-1} \sum_{|j-k| > h} k_{tj}^2 \mathbf{y}_{j-1} \mathbf{y}'_{j-1}\|_{sp} = i_{n,1} + i_{n,2} + i_{n,3} = o_p(1)$  where  $i_{n,1} = o_p(1)$  by Lemma 6.1(i),  $i_{n,2} = o_p(1)$  by (6.22) and  $i_{n,3} = o_p(1)$ , because  $E i_{n,2} \leq K_{2,t}^{-1} \sum_{|j-k| > h} k_{tj}^2 E \|\mathbf{y}_{j-1}\|^2 \leq C K_{2,t}^{-1} \sum_{|j-k| > h} k_{tj}^2 = o(1)$  by (6.1).

To show (6.28), bound  $v_{n,\epsilon} \leq \sum_{|t-j| < h} \epsilon^{-1} E[k_{tj}^4 \xi_{tj}^4 | \mathcal{F}_{j-1}]$ . Note that  $k_{tj}$  is bounded uniformly in  $t, j$ , and

$$E[\xi_{tj}^4 | \mathcal{F}_{j-1}] = K_{2,t}^{-2} E(b' \boldsymbol{\varepsilon}_j)^4 (\mathbf{y}'_{j-1} \mathbf{V}_{\psi,t_0}^{-1/2} \mathbf{a})^4$$

$$\leq CK_{2,t}^{-2} \|\mathbf{y}'_{j-1}\|^4 \|\mathbf{V}_{\psi,t_0}^{-1/2}\|_{sp}^4 \leq CK_{2,t}^{-2} \|\mathbf{y}'_{j-1}\|^4 O_p(1),$$

by (6.26) and because  $E(b'\boldsymbol{\varepsilon}_j)^4 \leq \|b\|^4 E\|\boldsymbol{\varepsilon}_j\|^4 \leq C$ . Hence,

$$v_{n,\epsilon} \leq O_p(1) K_{2,t}^{-2} \sum_{|t-j|<h} k_{tj}^4 \|\mathbf{y}'_{j-1}\|^4 = o_p(1)$$

since  $E[K_{2,t}^{-2} \sum_{|t-j|<h} k_{tj}^4 \|\mathbf{y}'_{j-1}\|^4] \leq CK_{2,t}^{-2} \sum_{j=1}^n k_{tj}^2 \leq CK_{2,t}^{-1} \rightarrow 0$  by (2.8) and (6.1).

This completes the proof of (6.28) and (6.25)(a).

*Proof of (6.25)(b).* By (6.1),  $K_t/K_{2,t}^{1/2} = O(H_\psi^{1/2})$ . Therefore

$$|p_{n,t;2}| \leq CH_\psi^{1/2} (\|\mathbf{H}_{t-1}^{-1} \mathbf{S}_{uy,t} - \mathbf{S}_{\epsilon y,t}\|_{sp} \|\mathbf{V}_{\psi,t_0}^{-1/2}\|_{sp} + \|\mathbf{S}_{uy,t}\|_{sp} \|\mathbf{V}_{yy,t}^{-1/2} - \mathbf{V}_{\psi,t_0}^{-1/2}\|_{sp}).$$

We will show below that

$$\|\mathbf{H}_{t-1}^{-1} \mathbf{S}_{uy,t} - \mathbf{S}_{\epsilon y,t}\|_{sp} = o_p(H_\psi^{-1/2}), \quad (6.30)$$

while  $\|\mathbf{V}_{\psi,t_0}^{-1/2}\|_{sp} = O_p(1)$  by (6.26),  $\|\mathbf{S}_{uy,t}\|_{sp} = O_p(H_\psi^{-1/2})$  by Lemma 6.2(i), and  $\|\mathbf{V}_{yy,t}^{-1/2} - \mathbf{V}_{\psi,t_0}^{-1/2}\|_{sp} = o_p(1)$  by Lemma 6.1(iv). Hence,  $|p_{n,t;2}| \leq CH_\psi^{1/2} (o_p(H_\psi^{-1/2}) + O_p(H_\psi^{-1/2}) o_p(1)) = o_p(1)$  proving (6.25)(b).

To show (6.30), use  $\mathbf{H}_{t-1}^{-1} \mathbf{u}_j = \mathbf{H}_{t-1}^{-1} \mathbf{H}_{j-1} \boldsymbol{\varepsilon}_j = \boldsymbol{\varepsilon}_j + \mathbf{H}_{t-1}^{-1} (\mathbf{H}_{j-1} - \mathbf{H}_{t-1}) \boldsymbol{\varepsilon}_j$  to obtain  $\mathbf{H}_{t-1}^{-1} \mathbf{S}_{uy,t} - \mathbf{S}_{\epsilon y,t} = \mathbf{H}_{t-1}^{-1} K_t^{-1} \sum_{j=1}^n k_{tj} (\mathbf{H}_{j-1} - \mathbf{H}_{t-1}) \boldsymbol{\varepsilon}_j y_{j-1} =: \mathbf{H}_{t-1}^{-1} q_{n,t}$ . Thus,  $\|\mathbf{H}_{t-1}^{-1} \mathbf{S}_{uy,t} - \mathbf{S}_{\epsilon y,t}\|_{sp} \leq \|\mathbf{H}_{t-1}^{-1}\|_{sp} \|q_{n,t}\|_{sp}$  where  $\|\mathbf{H}_{t-1}^{-1}\|_{sp} = O_p(1)$  by Assumption 2.2(iii), so that it remains to verify that

$$\|q_{n,t}\|_{sp} = o_p(H_\psi^{-1/2}).$$

Observe that

$$\begin{aligned} E\|(\mathbf{H}_{j-1} - \mathbf{H}_{t-1}) \boldsymbol{\varepsilon}_j y_{j-1}\|_{sp} &\leq (E\|\mathbf{H}_{j-1} - \mathbf{H}_{t-1}\|_{sp}^2)^{1/2} (E\|\boldsymbol{\varepsilon}_j\|^2 \|y_{j-1}\|^2)^{1/2} \\ &\leq C (E\|\mathbf{H}_{j-1} - \mathbf{H}_{t-1}\|_{sp}^2)^{1/2} \end{aligned}$$

since by Assumption 2.2(i) and (2.8),  $E[\|\boldsymbol{\varepsilon}_j\|^2 \|y_{j-1}\|^2] = E\|\boldsymbol{\varepsilon}_j\|^2 E\|y_{j-1}\|^2 \leq C$  for all  $j$ . Moreover, by Assumption 2.2(ii), for  $|t-j| \leq t/2$ ,  $E\|\mathbf{H}_{j-1} - \mathbf{H}_{t-1}\|_{sp}^2 \leq C|t-j|/t$ . Therefore,

$$E\|q_{n,t}\|_{sp} = K_t^{-1} \sum_{j=1}^n k_{tj} E\|(\mathbf{H}_{j-1} - \mathbf{H}_{t-1}) \boldsymbol{\varepsilon}_j y_{j-1}\|_{sp} \quad (6.31)$$

$$\begin{aligned}
&\leq C(H_\psi/t)^{1/2}H_\psi^{-1}\sum_{|t-j|\leq t/2}k_{tj}(|t-j|/H_\psi)^{1/2}+CH_\psi^{-1}\sum_{|t-j|>t/2}k_{tj} \\
&\leq C((H_\psi/n)^{1/2}+H_\psi^{-1}),
\end{aligned}$$

by (2.10) and (6.1), bearing in mind that  $t = \lceil \tau n \rceil$ . Together with assumption  $H_\psi \bar{H}_\psi = o(n)$  this implies  $\|q_{n,t}\|_{sp} = O_p((H_\psi/n)^{1/2} + H_\psi^{-1}) = o_p(H_\psi^{-1/2})$ , completing the proof of (6.30) and (6.25)(b).  $\square$

**Lemma 6.3.** *Under assumptions of Theorem 2.2(i),  $\Sigma_{uu,t}$  of (6.4) satisfies*

$$(i) \quad \|\Sigma_{uu,t} - \Sigma_t\|_{sp} = O_p(\kappa_{n,h}), \quad (6.32)$$

$$(ii) \quad (L_t/L_{2,t}^{1/2})\mathbf{H}_{t-1}^{-1}(\Sigma_{uu,t} - \Sigma_t)\mathbf{H}'_{t-1} \rightarrow_D \mathbf{Z} \quad \text{if } H_h = o(n^{1/2}) \quad (6.33)$$

with  $\mathbf{Z}$  as in Theorem 2.2(ii).

**Proof.** First we show (6.32). Recall  $\Sigma_t = \mathbf{H}_{t-1}\mathbf{H}'_{t-1}$  and  $E\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}'_j = I$ . Then,  $\Sigma_{\varepsilon\varepsilon,t} := L_t^{-1}\sum_{j=1}^n l_{tj}\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}'_j$  has property  $\mathbf{H}_{t-1}E[\Sigma_{\varepsilon\varepsilon,t}]\mathbf{H}'_{t-1} = \Sigma_t$ . Therefore,

$$\Sigma_{uu,t} - \Sigma_t = p_{n,t} + \delta_{n,t}, \quad (6.34)$$

$$p_{n,t} := \mathbf{H}_{t-1}(\Sigma_{\varepsilon\varepsilon,t} - E[\Sigma_{\varepsilon\varepsilon,t}])\mathbf{H}'_{t-1}; \quad \delta_{n,t} = L_t^{-1}\sum_{j=1}^n l_{tj}\xi_{tj},$$

where  $\xi_{tj} := \mathbf{u}_j\mathbf{u}'_j - \mathbf{H}_{t-1}\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}'_j\mathbf{H}'_{t-1} = \mathbf{H}_{j-1}\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}'_j\mathbf{H}'_{j-1} - \mathbf{H}_{t-1}\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}'_j\mathbf{H}'_{t-1}$ . Then  $\|\Sigma_{uu,t} - \Sigma_t\|_{sp} \leq \|p_{n,t}\|_{sp} + \|\delta_{n,t}\|_{sp}$ . We will show that

$$(i) \quad \|p_{n,t}\|_{sp} = O_p(H_h^{-1/2}), \quad (ii) \quad \|\delta_{n,t}\|_{sp} = O_p((H_h/n)^{1/2} + H_h^{-1}), \quad (6.35)$$

which implies (6.32).

The claim (6.35)(i) follows from  $\|p_{n,t}\|_{sp} \leq \|\mathbf{H}_{t-1}\|_{sp}^2\|\Sigma_{\varepsilon\varepsilon,t} - E[\Sigma_{\varepsilon\varepsilon,t}]\|_{sp} = O_p(H_h^{-1/2})$  noting that  $\|\mathbf{H}_{t-1}\|_{sp} = O_p(1)$  by Assumption 2.2 (ii), and because for independent  $\boldsymbol{\varepsilon}_j$ 's it holds  $E\|\Sigma_{\varepsilon\varepsilon,t} - E[\Sigma_{\varepsilon\varepsilon,t}]\|_{sp}^2 \leq CL_t^{-2}\sum_{j=1}^n l_{tj}^2 \leq CH_h^{-1}$  by (6.1), which implies  $\|\Sigma_{\varepsilon\varepsilon,t} - E[\Sigma_{\varepsilon\varepsilon,t}]\|_{sp} = O_p(H_h^{-1/2})$ .

To prove (6.35)(ii), observe that

$$\|\xi_{tj}\|_{sp} \leq \|(\mathbf{H}_{j-1} - \mathbf{H}_{t-1})\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}'_j\mathbf{H}'_{j-1}\|_{sp} + \|\mathbf{H}_{t-1}\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}'_j(\mathbf{H}_{j-1} - \mathbf{H}_{t-1})'\|_{sp}$$

$$\begin{aligned}
&\leq \|\mathbf{H}_{j-1} - \mathbf{H}_{t-1}\|_{sp} \|\mathbf{H}_{j-1}\|_{sp} \|\boldsymbol{\varepsilon}_j\|^2 + \|\mathbf{H}_{j-1} - \mathbf{H}_{t-1}\|_{sp} \|\mathbf{H}_{t-1}\|_{sp} \|\boldsymbol{\varepsilon}_j\|^2. \\
&\leq (\|\mathbf{H}_t\|_{sp} + 1) \|\mathbf{H}_{j-1} - \mathbf{H}_{t-1}\|_{sp} (\|\mathbf{H}_{j-1}\|_{sp} + 1) \|\boldsymbol{\varepsilon}_j\|^2.
\end{aligned}$$

By Assumption 2.2(ii),  $\|\mathbf{H}_t\|_{sp} = O_p(1)$ , and  $E[\|\mathbf{H}_{j-1} - \mathbf{H}_{t-1}\|_{sp} (\|\mathbf{H}_{j-1}\|_{sp} + 1) \|\boldsymbol{\varepsilon}_j\|^2] \leq (E\|\mathbf{H}_{j-1} - \mathbf{H}_{t-1}\|_{sp}^2)^{1/2} (E(1 + \|\mathbf{H}_{j-1}\|_{sp})^2)^{1/2} E\|\boldsymbol{\varepsilon}_j\|^2 \leq C(E\|\mathbf{H}_{j-1} - \mathbf{H}_{t-1}\|_{sp}^2)^{1/2}$ . Hence (6.35)(ii) follows by the same argument as in the proof of (6.31).

*Proof of (6.33).* Using (6.34), write

$$(L_t/L_{2,t}^{1/2})\mathbf{H}_{t-1}^{-1}(\boldsymbol{\Sigma}_{uu,t} - \boldsymbol{\Sigma}_t)\mathbf{H}'_{t-1} =: q_{n,1} + q_{n,2},$$

where  $q_{n,1} = (L_t/L_{2,t}^{1/2})(\boldsymbol{\Sigma}_{\varepsilon\varepsilon,t} - E[\boldsymbol{\Sigma}_{\varepsilon\varepsilon,t}])$  and  $q_{n,2} := (L_t/L_{2,t}^{1/2})\mathbf{H}_{t-1}^{-1}\delta_{n,t}\mathbf{H}'_{t-1}$ . To prove (6.33) it suffices to show that

$$(i) \quad q_{n,1} \rightarrow_D \mathbf{Z}, \quad (ii) \quad \|q_{n,2}\|_{sp} \rightarrow_p 0. \quad (6.36)$$

To verify (6.36)(i), observe that  $q_{n,1}$  is a  $m \times m$  matrix

$$q_{n,1} = L_{2,t}^{-1/2} \sum_{j=1}^n l_{tj} \boldsymbol{\varepsilon}_j \boldsymbol{\varepsilon}'_j = (z_{n,ik})_{i,k=1,\dots,m}$$

with a typical  $(i, k)$ -th element  $z_{n,ik} = L_{2,t}^{-1/2} \sum_{j=1}^n l_{tj} (\varepsilon_{ij} \varepsilon_{kj} - E[\varepsilon_{ij} \varepsilon_{kj}])$ . Since  $\boldsymbol{\varepsilon}_j = (\varepsilon_{1j}, \dots, \varepsilon_{mj})'$  is a vector of independent i.i.d. noises  $\{\varepsilon_{ij}\}$ ,  $i = 1, \dots, m$ , the elements  $z_{n,ik}$ 's for different  $(i, k)$ 's such that  $i \leq k$  are independent. Thus, to prove (i), it suffices to show that for each  $(i, k)$ ,

$$z_{n,ik} \rightarrow N(0, v_{ik}^2). \quad (6.37)$$

Write  $z_{n,ik} = \sum_{j=1}^n \theta_{nj} \zeta_j$ ,  $\theta_{nj} := L_{2,t}^{-1/2} l_{tj}$ ,  $\zeta_j := \varepsilon_{ij} \varepsilon_{kj} - E[\varepsilon_{ij} \varepsilon_{kj}]$  as a weighted sum of i.i.d. random variables  $\zeta_j$  with  $E\zeta_j = 0$  and  $E\zeta_j^2 = v_{ik}^2$ . Notice that  $\theta_{nj}$  satisfy  $\sum_{j=1}^n \theta_{nj}^2 = 1$  and  $\max_{j=1,\dots,n} |\theta_{nj}| \leq CL_{2,t}^{-1/2} \rightarrow 0$ . Thus,  $\zeta_j$  and  $\theta_{nj}$  satisfy sufficient conditions for asymptotic normality (6.37) of the sum  $z_{n,ik}$ , see e.g. Lemma 2.1 in Abadir, Distaso, Giraitis, and Koul (2014).

To verify (6.36)(ii), bound  $\|q_{n,2}\|_{sp} \leq (L_t/L_{2,t}^{1/2})\|\mathbf{H}_{t-1}^{-1}\|_{sp}^2 \|\delta_{n,t}\|_{sp}$ , where  $L_t/L_{2,t}^{1/2} = O(H_h^{1/2})$  by (6.1),  $\|\mathbf{H}_{t-1}^{-1}\|_{sp} = O_p(1)$  by Assumption 2.2(iii), and by (6.35),  $\|\delta_{n,t}\|_{sp} =$

$O_p((H_h/n)^{1/2} + H_h^{-1}) = o_p(H_h^{-1/2})$  under assumption  $H_h = o(n^{1/2})$ . Hence,  $\|q_{n,2}\|_{sp} = O(H_h^{1/2})o_p(H_h^{-1/2}) = o_p(1)$ , which completes the proof of (6.36)(ii), (6.33) and the lemma.  $\square$

**Lemma 6.4.** (i) Under assumptions of Theorem 2.2(i),

$$\bar{\mathbf{y}}_t = (I - \Psi_t)^{-1}\bar{\mathbf{u}}_t + O_p(\kappa_{n,\psi}^*) = (I - \Psi_t)^{-1}\mathbf{H}_{t-1}\bar{\boldsymbol{\varepsilon}}_t + O_p(\kappa_{n,\psi}^*) = O_p(\kappa_{n,\psi}). \quad (6.38)$$

In addition, if  $H_\psi \bar{H}_\psi = o(n)$ , then

$$(K_t/K_{2,t}^{1/2})H_{t-1}^{-1}(I - \Psi_t)\bar{\mathbf{y}}_t \rightarrow_D \mathcal{N}(0, \mathbf{I}). \quad (6.39)$$

(ii) Under assumptions of Theorem 2.3(i),  $\|\bar{\mathbf{y}}_t - \boldsymbol{\mu}_t - \bar{\boldsymbol{\gamma}}_t\| = O_p(\kappa_{n,\psi}^*)$ .

(iii) Under assumptions of Theorem 2.3(ii), with  $\widehat{\mathbf{D}}_t$  as in Remark 2.2,

$$(K_t/K_{2,t}^{1/2})\mathbf{H}_{t-1}^{-1}((I - \Psi_t)(\bar{\mathbf{y}}_t - \boldsymbol{\mu}_t) - (\widehat{\Psi}_t - \Psi_t)\boldsymbol{\mu}_t)\widehat{\mathbf{D}}_t^{-1/2} \rightarrow_D \mathcal{N}(0, \mathbf{I}). \quad (6.40)$$

**Proof** (i) *Proof of (6.38).* By (2.7),  $\mathbf{y}_j = \sum_{i=0}^{j-1} \Pi_{j,i}\mathbf{u}_{j-i} + \Pi_{j,j}\mathbf{y}_0$ . Thus,  $\bar{\mathbf{y}}_j = K_t^{-1} \sum_{j=1}^n k_{tj}\mathbf{y}_j = K_t^{-1} \sum_{j=1}^n k_{tj} \sum_{i=0}^{j-1} \Pi_{j,i}\mathbf{u}_{j-i} + K_t^{-1} \sum_{j=1}^n k_{tj}\Pi_{j,j}\mathbf{y}_0 =: s_{n,1} + s_{n,2}$ .

We show that

$$(a) s_{n,2} = O_p(\kappa_{n,\psi}^*), \quad (b) s_{n,1} = (I - \Psi_t)^{-1}\bar{\mathbf{u}}_t + O_p(\kappa_{n,\psi}^*), \quad (6.41)$$

which implies the first claim of (6.38),  $\mathbf{y}_j = (I - \Psi_t)^{-1}\bar{\mathbf{u}}_t + O_p(\kappa_{n,\psi}^*)$ .

To show (a), use  $\|\Pi_{j,i}\|_{sp} \leq r^i$  of (2.6), to obtain  $\|s_{n,2}\| \leq \|\mathbf{y}_0\|K_t^{-1} \sum_{i=0}^{j-1} \|\Pi_{j,i}\| \leq \|\mathbf{y}_0\|K_t^{-1} \sum_{i=0}^{\infty} r^i = O_p(H_\psi^{-1})$  since  $\|\mathbf{y}_0\| = O_p(1)$  by Assumption 2.2(i), and  $K_t^{-1} = O_p(H_\psi^{-1})$  by (6.1).

To show (b), recall that  $t = [\tau n]$ . Select  $h = b\bar{H}_\psi$  that satisfies (6.1). Write

$$\begin{aligned} s_{n,1} &= K_t^{-1} \sum_{i=0}^{n-1} \sum_{j=1}^{n-i} k_{t,j+i}\Pi_{j+i,i}\mathbf{u}_j = K_t^{-1} \sum_{i=0}^h \sum_{|j-t|\leq h} k_{t,j}\Psi_t^i\mathbf{u}_j \\ &+ \{K_t^{-1} \sum_{i=0}^h \sum_{|j-t|\leq h} (k_{t,j+i}\Pi_{j+i,i} - k_{t,j}\Psi_t^i)\mathbf{u}_j\} + \{K_t^{-1} \sum_{i=0}^h \sum_{|j-t|>h} k_{t,j+i}\Pi_{j+i,i}\mathbf{u}_j\} \\ &+ K_t^{-1} \sum_{i=h+1}^{n-1} \sum_{j=1}^{n-i} k_{t,j+i}\Pi_{j+i,i}\mathbf{u}_j =: q_{n,1} + q_{n,2} + q_{n,3}. \end{aligned}$$



We verify (b) showing that

$$q_{n,1} = (I - \Psi_t)^{-1} \bar{\mathbf{u}}_t + O_p(\kappa_{n,\psi}^*), \quad q_{n,2} = O_p(\kappa_{n,\psi}^*), \quad q_{n,3} = O_p(\kappa_{n,\psi}^*). \quad (6.42)$$

To evaluate  $q_{n,1}$ , observe that  $q_{n,1} - (I - \Psi_t)^{-1} \bar{\mathbf{u}}_t = (\sum_{i=0}^h \Psi_t^i) K_t^{-1} \sum_{|j-t| \leq h} k_{t,j} \mathbf{u}_j - (\sum_{i=0}^{\infty} \Psi_t^i) K_t^{-1} \sum_{j=1}^n k_{t,j} \mathbf{u}_j$ . Then,

$$\begin{aligned} \|q_{n,1} - (I - \Psi_t)^{-1} \bar{\mathbf{u}}_t\| &\leq \|\sum_{i=h+1}^{\infty} \Psi_t^i\|_{sp} \times K_t^{-1} \sum_{|j-t| \leq h} k_{t,j} \|\mathbf{u}_j\| \\ &+ \|\sum_{i=0}^{\infty} \Psi_t^i\|_{sp} \times K_t^{-1} \sum_{|j-t| > h} k_{t,j} \|\mathbf{u}_j\| = a_{n,1} b_{n,1} + a_{n,2} b_{n,2}. \end{aligned}$$

Observe that  $a_{n,1} \leq \sum_{i=h+1}^{\infty} r^i \leq r^h (1-r)^{-1} = O(h^{-1}) = O(H_\psi^{-1})$ , and  $a_{n,2} \leq \sum_{i=0}^{\infty} r^i = (1-r)^{-1}$ , while  $b_{n,1} = O_p(1)$  because  $E b_{n,1} \leq C K_t^{-1} \sum_{|j-t| \leq h} k_{t,j} = C$ , and  $b_{n,2} = O_p(H_\psi^{-1})$  because  $E b_{n,2} \leq C \times K_t^{-1} \sum_{|j-t| > h} k_{t,j} = O_p(H_\psi^{-1})$  by (6.1), which proves the claim (6.42) about  $q_{n,1}$ .

To bound  $q_{n,2}$ , use  $\|k_{t,j+i} \Pi_{j+i,i} - k_{t,j} \Psi_t^i\| \leq |k_{t,j+i} - k_{t,j}| \|\Pi_{j+i,i}\|_{sp} + k_{t,j} \|\Pi_{j+i,i} - \Psi_t^i\|_{sp}$ , where by (2.10) and the mean value theorem,  $|k_{t,j+i} - k_{t,j}| \leq C(i/H_\psi) k_{t,j}^*$ ,  $k_{t,j}^* := (1 + (j/H_\psi)^2)^{-1}$ ,  $\|\Pi_{j+i,i}\|_{sp} \leq r^i$ , whereas for  $|j-t| \leq h$ , and  $i \leq h$ , arguing as in the proof of (6.2), it follows that  $\|\Pi_{j+i,i} - \Psi_t^i\|_{sp} \leq i r^{i-1} \max_{s:|s-t| \leq 2h} \|\Psi_s - \Psi_t\|_{sp} = i r^{i-1} R_{t,2h}$ . In addition,  $R_{t,2h} = O_p((h/t)^{1/2}) = O_p((\bar{H}_\psi/n)^{1/2})$  by assumption (2.4). Hence,  $\|k_{t,j+i} \Pi_{j+i,i} - k_{t,j} \Psi_t^i\| \leq (C H_\psi^{-1} + R_{t,2h}) i r^{i-1} (k_{t,j}^* + k_{t,j})$ , and

$$\|q_{n,2}\| \leq (C H_\psi^{-1} + R_{t,2h}) (\sum_{i=0}^h i r^{i-1}) T_n, \quad T_n := K_t^{-1} \sum_{|j-t| \leq h} (k_{t,j}^* + k_{t,j}) \|\mathbf{u}_j\|.$$

Since  $E T_n \leq C K_t^{-1} \sum_{|j-t| \leq h} (k_{t,j}^* + k_{t,j}) \leq C$ , this implies  $T_n = O_p(1)$ , and  $q_{n,2} = O_p((\bar{H}_\psi/n)^{1/2} + H_\psi^{-1}) = O_p(\kappa_{n,\psi}^*)$ , which proves (6.42) for  $q_{n,2}$ .

Finally, to bound  $q_{n,3}$ , note that  $\|\Pi_{j+i,i} \mathbf{u}_j\| \leq \|\Pi_{j+i,i}\|_{sp} \|\mathbf{u}_j\| \leq r^i \|\mathbf{u}_j\|$ , and therefore  $E q_{n,3} \leq C (\sum_{i=h+1}^{n-1} r^i) K_t^{-1} \sum_{j=1}^{n-i} k_{t,j+i} = O(r^h) = O(H_\psi^{-1})$ , which implies (6.42). This completes the proof of the first claim of (6.38):  $\bar{\mathbf{y}}_t = (I - \Psi_t)^{-1} \bar{\mathbf{u}}_t + O_p(\kappa_{n,\psi}^*)$ . To show the second claim  $\bar{\mathbf{y}}_t = (I - \Psi_t)^{-1} \mathbf{H}_{t-1} \bar{\boldsymbol{\varepsilon}}_t + O_p(\kappa_{n,\psi}^*)$ , it suffices to verify that  $q_n := (I - \Psi_t)^{-1} (\bar{\mathbf{u}}_t - \mathbf{H}_{t-1} \bar{\boldsymbol{\varepsilon}}_t) = O_p(\kappa_{n,\psi}^*)$ . Since  $\|\Psi_t\|_{sp} \leq r < 1$ , then

$\|(I - \Psi_t)^{-1}\|_{sp} \leq 1/(1 - \|\Psi_t\|_{sp}) \leq 1/(1 - r)$ . Therefore,

$$\begin{aligned} \|q_n\| &\leq \|(I - \Psi_t)^{-1}\|_{sp} \|\bar{\mathbf{u}}_t - \mathbf{H}_{t-1} \bar{\boldsymbol{\varepsilon}}_t\| \leq CK_t^{-1} \sum_{j=1}^n k_{tj} \|(\mathbf{u}_j - \mathbf{H}_{t-1}) \boldsymbol{\varepsilon}_j\| \\ &\leq CK_t^{-1} \sum_{j=1}^n k_{tj} \|\mathbf{H}_{j-1} - \mathbf{H}_{t-1}\|_{sp} \|\boldsymbol{\varepsilon}_j\| = O_p(\kappa_{n,\psi}^*) \end{aligned}$$

by the same argument as in the proof of (6.31).

*Proof of (6.39).* By (6.38),  $(K_t/K_{2,t}^{1/2})\mathbf{H}_{t-1}^{-1}(I - \Psi_t)\bar{\mathbf{y}}_t = (K_t/K_{2,t}^{1/2})\bar{\boldsymbol{\varepsilon}}_t + \delta_n$ , where

$$\begin{aligned} \|\delta_n\| &= \|(K_t/K_{2,t}^{1/2})\mathbf{H}_{t-1}^{-1}(I - \Psi_t)O_p(\kappa_{n,\psi}^*)\| \\ &\leq (K_t/K_{2,t}^{1/2})\|\mathbf{H}_{t-1}^{-1}\|_{sp} \|I - \Psi_t\| O_p(\kappa_{n,\psi}^*) = O(H_\psi^{1/2})O_p(\kappa_{n,\psi}^*) = o_p(1), \end{aligned}$$

because  $\kappa_{n,\psi}^* = o(H_\psi^{-1/2})$  for  $H_\psi \bar{H}_\psi = o(n)$ . It remains to show that  $(K_t/K_{2,t}^{1/2})\bar{\boldsymbol{\varepsilon}}_t = K_{2,t}^{-1/2} \sum_{j=1}^n k_{tj} \boldsymbol{\varepsilon}_j \rightarrow_D \mathcal{N}(0, \mathbf{I})$  which follows by the same argument as proving (6.37).

(ii) By definition,  $\mathbf{y}_j = \boldsymbol{\mu}_j + \dot{\mathbf{y}}_j$  where  $\dot{\mathbf{y}}_j$  is a VAR(1) process with no intercept. Then  $\bar{\mathbf{y}}_t - \boldsymbol{\mu}_t - \bar{\dot{\mathbf{y}}}_t = \bar{\boldsymbol{\mu}}_t - \boldsymbol{\mu}_t = K_t^{-1} \sum_{j=1}^n k_{tj} (\boldsymbol{\mu}_j - \boldsymbol{\mu}_t)$ . Thus  $\|\bar{\mathbf{y}}_t - \boldsymbol{\mu}_t - \bar{\dot{\mathbf{y}}}_t\| = O_p(\kappa_{n,\psi}^*)$  follows using Assumption 2.3, combining arguments used in the proof of (6.7) and (6.31).

(iii) To prove the asymptotic normality (6.40), denote

$$q_n := (K_t/K_{2,t}^{1/2})\mathbf{H}_{t-1}^{-1}((I - \Psi_t)(\bar{\mathbf{y}}_t - \boldsymbol{\mu}_t) - (\hat{\Psi}_t - \Psi_t)\boldsymbol{\mu}_t) \hat{\mathbf{D}}_t^{-1/2}.$$

By Lemma 6.4(ii) and (6.38),

$$\begin{aligned} \|\mathbf{H}_{t-1}^{-1}(I - \Psi_t)(\bar{\mathbf{y}}_t - \boldsymbol{\mu}_t) - \bar{\boldsymbol{\varepsilon}}_t\| &\leq \|\mathbf{H}_{t-1}^{-1}(I - \Psi_t)\|_{sp} \|\bar{\mathbf{y}}_t - \boldsymbol{\mu}_t - \bar{\dot{\mathbf{y}}}_t\| + \|\mathbf{H}_{t-1}^{-1}(I - \Psi_t)\bar{\dot{\mathbf{y}}}_t - \bar{\boldsymbol{\varepsilon}}_t\| \\ &\leq \|\mathbf{H}_{t-1}^{-1}\|_{sp} \|I - \Psi_t\|_{sp} (\|\bar{\mathbf{y}}_t - \boldsymbol{\mu}_t - \bar{\dot{\mathbf{y}}}_t\| + O_p(\kappa_{n,\psi}^*)) = O_p(\kappa_{n,\psi}^*). \end{aligned}$$

Using notation  $\hat{\Psi}_t \equiv \hat{\Psi}_{\hat{y}\hat{y},t} = S_{\hat{y}\hat{y},t} V_{\hat{y}\hat{y},t}^{-1}$  and  $\hat{\Psi}_{\dot{y}\dot{y},t} = S_{\dot{y}\dot{y},t} V_{\dot{y}\dot{y},t}^{-1}$  of the proof of Theorem 2.3(ii), bound

$$\begin{aligned} \|\mathbf{H}_{t-1}^{-1}(\hat{\Psi}_t - \Psi_t) - \mathbf{S}_{\varepsilon\dot{y},t} V_{\dot{y}\dot{y},t}^{-1}\|_{sp} &\leq \|\mathbf{H}_{t-1}^{-1}\|_{sp} \|\hat{\Psi}_{\hat{y}\hat{y},t} - \hat{\Psi}_{\dot{y}\dot{y},t}\|_{sp} \\ &\quad + \|\mathbf{H}_{t-1}^{-1}\|_{sp} \|\hat{\Psi}_{\dot{y}\dot{y},t} - \Psi_t - \mathbf{S}_{u\dot{y},t} V_{\dot{y}\dot{y},t}^{-1}\|_{sp} + \|\mathbf{H}_{t-1}^{-1} \mathbf{S}_{u\dot{y},t} - \mathbf{S}_{\varepsilon\dot{y},t}\|_{sp} \|V_{\dot{y}\dot{y},t}^{-1}\|_{sp} = O_p(\kappa_{n,\psi}^*), \end{aligned}$$

because  $\|\mathbf{H}_{t-1}^{-1}\|_{sp} = O_p(1)$  by Assumption 2.2,  $\|\hat{\Psi}_{\dot{y},t} - \hat{\Psi}_{\dot{y},t}^*\|_{sp} = O_p(\kappa_{n,\psi}^*)$  by (6.11),  $\|\hat{\Psi}_{\dot{y},t} - \Psi_t - \mathbf{S}_{u\dot{y},t} V_{\dot{y},t}^{-1}\|_{sp} = O_p(\kappa_{n,\psi}^*)$  by (6.5),  $\|\mathbf{H}_{t-1}^{-1} \mathbf{S}_{u\dot{y},t} - \mathbf{S}_{\varepsilon\dot{y},t}\|_{sp} = O_p(\kappa_{n,\psi}^*)$  by (6.30) and  $\|V_{\dot{y},t}^{-1}\|_{sp} = O_p(1)$  by Lemma 6.1(vi).

In addition, observe that assumption  $H_\psi \bar{H}_\psi = o(n)$  implies  $\kappa_{n,\psi}^* = o(H_\psi^{-1/2})$ ,  $K_t/K_{2,t}^{1/2} = O_p(H_\psi^{1/2})$  by (6.1),  $\|\boldsymbol{\mu}_t\| = O_p(1)$  by Assumption 2.3 and  $|\hat{\mathbf{D}}_t^{-1/2}| = O_p(1)$  by Lemma 6.5(iv). The above bounds imply

$$q_n = (K_t/K_{2,t}^{1/2})(\bar{\boldsymbol{\varepsilon}}_t - \mathbf{S}_{\varepsilon\dot{y},t} V_{\dot{y},t}^{-1} \boldsymbol{\mu}_t) \hat{\mathbf{D}}_t^{-1/2} + o_p(1) =: p_n + o_p(1).$$

Let  $t_0 = t - h$ , where  $h = b\bar{H}_\psi$  is such that (6.1) holds. Setting  $\mathbf{h}_{t_0} = V_{\psi,t_0}^{-1} \boldsymbol{\mu}_{t_0}$ ,  $\mathbf{D}_{t_0} = 1 + \boldsymbol{\mu}'_{t_0} \mathbf{V}_{\psi,t_0}^{-1} \boldsymbol{\mu}_{t_0}$  write

$$\begin{aligned} p_n &= (K_t/K_{2,t}^{1/2})(\bar{\boldsymbol{\varepsilon}}_t - \mathbf{S}_{\varepsilon\dot{y},t} \mathbf{h}_{t_0}) \mathbf{D}_{t_0}^{-1/2} \\ &+ \left\{ (K_t/K_{2,t}^{1/2}) \bar{\boldsymbol{\varepsilon}}_t (\hat{\mathbf{D}}_t^{-1/2} - \mathbf{D}_{t_0}^{-1/2}) - (K_t/K_{2,t}^{1/2}) \mathbf{S}_{\varepsilon\dot{y},t} (V_{\dot{y},t}^{-1} \boldsymbol{\mu}_t \hat{\mathbf{D}}_t^{-1/2} - \mathbf{h}_{t_0} \mathbf{D}_{t_0}^{-1/2}) \right\} + o_p(1) \\ &=: p_{n,1} + p_{n,2} + o_p(1). \end{aligned}$$

In view of the Cramér-Wold device, it suffices to verify that for any  $m \times 1$  vector  $b$ , the scalar random variable  $d_n := b' p_{n,1}$  satisfies

$$(a) \quad b' p_{n,1} \rightarrow \mathcal{N}(0, \|b\|^2), \quad (b) \quad \|p_{n,2}\| = o_p(1). \quad (6.43)$$

*Proof of (6.43)(a).* Letting  $\xi_{tj} := K_{2,t}^{-1/2} (b' \boldsymbol{\varepsilon}_j) (1 - \dot{\mathbf{y}}'_{j-1} \mathbf{h}_{t_0}) \mathbf{D}_{t_0}^{-1/2}$ , write  $p_{n,1} = \sum_{j=1}^n k_{tj} \xi_{tj}$ . Observe that  $\|\mathbf{h}_{t_0}\| \leq \|V_{\psi,t_0}^{-1}\|_{sp} \|\boldsymbol{\mu}_{t_0}\| = O_p(1)$  since  $\|V_{\psi,t_0}^{-1}\|_{sp} = O_p(1)$  by Lemma 6.1(ii),  $\|\boldsymbol{\mu}_{t_0}\| = O_p(1)$  by Assumption 2.3, while  $|\mathbf{D}_{t_0}^{-1/2}| = O_p(1)$  by Lemma 6.5(iv). Hence, by the same argument as in the proof of (6.25)(a) it suffices to show that for the above  $\xi_{tj}$ 's,  $j_n$  satisfies (6.27) and  $v_{n,\varepsilon}$  satisfies (6.28). Since  $\mathbf{h}_{t_0}$  and  $\mathbf{D}_{t_0}^{-1/2}$  are  $\mathcal{F}_{t_0}$  measurable then, for  $j > t - h = t_0$ ,

$$E[\xi_{tj}^2 | \mathcal{F}_{j-1}] = K_{2,t}^{-1} E(b' \boldsymbol{\varepsilon}_j)^2 \mathbf{D}_{t_0}^{-1/2} (1 - \dot{\mathbf{y}}'_{j-1} \mathbf{h}_{t_0})' (1 - \dot{\mathbf{y}}'_{j-1} \mathbf{h}_{t_0}) \mathbf{D}_{t_0}^{-1/2}.$$

Hence, with  $\tilde{V}_{yy,t} := K_{2,t}^{-1} \sum_{|t-j|<h} k_{tj}^2 \dot{\mathbf{y}}_{j-1} \dot{\mathbf{y}}'_{j-1}$  and  $S_{\dot{y},t} := K_{2,t}^{-1} \sum_{|t-j|<h} k_{tj}^2 \dot{\mathbf{y}}_{j-1}$ , write

$$j_n := \|b\|^2 \mathbf{D}_{t_0}^{-1/2} Q_{n,t} \mathbf{D}_{t_0}^{-1/2}, \quad Q_{n,t} := K_{2,t}^{-1} \sum_{|t-j|<h} k_{tj}^2 (1 - \dot{\mathbf{y}}'_{j-1} \mathbf{h}_{t_0})' (1 - \dot{\mathbf{y}}'_{j-1} \mathbf{h}_{t_0}),$$

$$Q_{n,t} = K_{2,t}^{-1} \sum_{|j-t| \leq h} k_{tj}^2 - \{S'_{\dot{y},t} \mathbf{h}_{t_0} + \mathbf{h}'_{t_0} S_{\dot{y},t}\} + \mathbf{h}'_{t_0} \tilde{V}_{\dot{y}\dot{y},t} \mathbf{h}_{t_0} =: s_{n,1} + s_{n,2} + s_{n,3}.$$

Since  $s_{n,1} = 1 + o(1)$  by (6.1),  $|s_{n,2}| \leq 2\|S_{\dot{y},t}\| \|\mathbf{h}_{t_0}\| = o_P(1)$  because  $\|S_{\dot{y},t}\| = O_p(\kappa_{n,\psi}) = o_p(1)$  by (6.38), and  $\tilde{V}_{\dot{y}\dot{y},t} = V_{\psi,t_0} + o_p(1)$  by (6.29), we obtain

$$\begin{aligned} j_n &:= \|b\|^2 \mathbf{D}_{t_0}^{-1/2} (1 + \mathbf{h}'_{t_0} V_{\psi,t_0} \mathbf{h}_{t_0}) \mathbf{D}_{t_0}^{-1/2} + o_p(1) \\ &= \|b\|^2 \mathbf{D}_{t_0}^{-1/2} (1 + \boldsymbol{\mu}'_{t_0} V_{\psi,t_0}^{-1} \boldsymbol{\mu}_{t_0}) \mathbf{D}_{t_0}^{-1/2} + o_p(1) = \|b\|^2 + o_p(1), \end{aligned}$$

proving (6.27) for  $j_n$ .

To verify (6.28) for  $v_{n,\epsilon}$ , observe that  $E[k_{tj}^4 \xi_{tj}^4 | \mathcal{F}_{j-1}] = K_{2,t}^{-2} E(b' \boldsymbol{\epsilon}_j)^4 (1 - \dot{\mathbf{y}}'_{j-1} \mathbf{h}_{t_0})^4 \mathbf{D}_{t_0}^{-2} \leq CK_{2,t}^{-2} (1 + \|\dot{\mathbf{y}}_{j-1}\|^4 \|\mathbf{h}_{t_0}\|^4) \mathbf{D}_{t_0}^{-2}$  where  $\|\mathbf{h}_{t_0}\|^4 \leq \|V_{\psi,t_0}^{-1}\|_{sp}^4 \|\boldsymbol{\mu}_{t_0}\|^4 = O_p(1)$  by Lemma 6.1(ii) and Assumption 2.3. Hence (6.28) follows by the same argument as in the proof of Lemma 6.2(ii).

*Proof of (6.43)(b). Bound*

$$\begin{aligned} \|p_{n,2}\| &\leq (K_t/K_{2,t}^{1/2}) \left( \|\bar{\boldsymbol{\epsilon}}_t\| \|\widehat{\mathbf{D}}_t^{-1/2} - \mathbf{D}_{t_0}^{-1/2}\| + \|\mathbf{S}_{\varepsilon\dot{y},t}\|_{sp} \|V_{\dot{y}\dot{y},t}^{-1} \boldsymbol{\mu}_t \widehat{\mathbf{D}}_t^{-1/2} - \mathbf{h}_{t_0} \mathbf{D}_{t_0}^{-1/2}\| \right) \\ &\leq o_p(1) + O_p(1) \|V_{\dot{y}\dot{y},t}^{-1} \boldsymbol{\mu}_t \widehat{\mathbf{D}}_t^{-1/2} - \mathbf{h}_{t_0} \mathbf{D}_{t_0}^{-1/2}\|, \end{aligned}$$

since  $K_t/K_{2,t}^{1/2} \leq CH_\psi^{1/2}$  by (6.1),  $\|\bar{\boldsymbol{\epsilon}}_t\| = O_p(H_\psi^{-1/2})$  by (6.38),  $|\widehat{\mathbf{D}}_t^{-1/2} - \mathbf{D}_{t_0}^{-1/2}| = o_p(1)$  by Lemma 6.5(iv), and  $\|\mathbf{S}_{\varepsilon\dot{y},t}\|_{sp} = O_p(H_\psi^{-1/2})$  by (6.23). Notice that

$$\begin{aligned} \|V_{\dot{y}\dot{y},t}^{-1} \hat{\boldsymbol{\mu}}_t \widehat{\mathbf{D}}_t^{-1/2} - \mathbf{h}_{t_0} \mathbf{D}_{t_0}^{-1/2}\| &\leq \|V_{\dot{y}\dot{y},t}^{-1} - V_{\psi,t_0}^{-1}\|_{sp} \|\boldsymbol{\mu}_t\| |\widehat{\mathbf{D}}_t^{-1/2}| \\ &\quad + \|V_{\psi,t_0}^{-1}\|_{sp} \|\hat{\boldsymbol{\mu}}_t - \boldsymbol{\mu}_{t_0}\| |\widehat{\mathbf{D}}_t^{-1/2}| + \|V_{\psi,t_0}^{-1}\|_{sp} \|\boldsymbol{\mu}_{t_0}\| |\widehat{\mathbf{D}}_t^{-1/2} - \mathbf{D}_{t_0}^{-1/2}| = o_p(1). \end{aligned}$$

The latter follows noting that  $\|V_{\dot{y}\dot{y},t}^{-1} - V_{\psi,t_0}^{-1}\|_{sp} = o_p(1)$  by Lemma 6.1(v);  $\|\boldsymbol{\mu}_t\| = O_p(1)$ ,  $\|\boldsymbol{\mu}_{t_0}\| = O_p(1)$  and  $\|\hat{\boldsymbol{\mu}}_t - \boldsymbol{\mu}_{t_0}\| \leq \|\hat{\boldsymbol{\mu}}_t - \boldsymbol{\mu}_t\| + \|\boldsymbol{\mu}_t - \boldsymbol{\mu}_{t_0}\| = o_p(1)$  by Theorem 2.3(ii) and Assumption 2.3;  $\|V_{\psi,t_0}^{-1}\|_{sp} = O_p(1)$  by Lemma 6.1(ii),  $|\widehat{\mathbf{D}}_t^{-1/2}| = O_p(1)$  by Lemma 6.5(iv), and  $|\widehat{\mathbf{D}}_t^{-1/2} - \mathbf{D}_{t_0}^{-1/2}| = o_p(1)$  by Lemma 6.5(iv). This proves that  $\|p_{n,2}\| = o_p(1)$  and completed the proof of the lemma.  $\square$

In Lemma 6.5 we use notation  $\mathbf{S}_{yy,t}$ ,  $\mathbf{V}_{yy,t}$ ,  $\boldsymbol{\Sigma}_{uu,t}$  of (6.4) and  $\dot{\mathbf{y}}_j$  of (2.25).

**Lemma 6.5.** (i) Under assumptions of Theorem 2.2(i)

$$(a) \|\Sigma_{\hat{u}\hat{u},t} - \Sigma_{uu,t}\|_{sp} = O_p(\kappa_{n,\psi}^2 + \kappa_{n,\psi}^*), \quad (b) \|\Sigma_{\hat{u}\hat{u},t}^{-1/2} - \Sigma_t^{-1/2}\|_{sp} = o_p(1). \quad (6.44)$$

(ii) Under assumptions of Theorem 2.3(i),

$$(a) \|\Sigma_{\hat{u}\hat{u},t} - \Sigma_{uu,t}\|_{sp} = O_p(\kappa_{n,\psi}^2 + \kappa_{n,\psi}^*), \quad (b) \|\Sigma_{\hat{u}\hat{u},t}^{-1/2} - \Sigma_t^{-1/2}\|_{sp} = o_p(1). \quad (6.45)$$

(iii) Under assumptions of Theorem 2.3(i),

$$(a) \|\mathbf{V}_{\hat{y}\hat{y},t}^{-1} - \mathbf{V}_{\hat{y}\hat{y},t}^{-1}\|_{sp} = O_p(\kappa_{n,\psi}^*), \quad (b) \|\mathbf{V}_{\hat{y}\hat{y},t}^{-1/2} - \mathbf{V}_{\hat{y}\hat{y},t}^{-1/2}\|_{sp} = O_p(\kappa_{n,\psi}^*), \quad (6.46)$$

$$(c) \|\mathbf{S}_{\hat{y}\hat{y},t} - \mathbf{S}_{yy,t}\|_{sp} = O_p(\kappa_{n,\psi}^*), \quad (d) \|\mathbf{V}_{\hat{y}\hat{y},t}^{-1}\|_{sp} = O_p(1), \quad \|\mathbf{V}_{\hat{y}\hat{y},t}\|_{sp} = O_p(1)$$

(iv) Under assumptions of Theorem 2.3(i), with  $\widehat{\mathbf{D}}_t$  and  $\mathbf{D}_t$  as in Remark 2.2,

$$(a) \mathbf{D}_t^{-1/2} = O_p(1); \quad (b) \widehat{\mathbf{D}}_t^{-1/2} = O_p(1); \quad (c) |\widehat{\mathbf{D}}_t^{-1/2} - \mathbf{D}_{t-h}^{-1/2}| = o_p(1), \quad h = o(t).$$

**Proof.** (i) To verify (6.44)(a), write  $\Sigma_{\hat{u}\hat{u},t} - \Sigma_{uu,t} = L_t^{-1} \sum_{j=1}^n l_{tj}(\hat{\mathbf{u}}_j \hat{\mathbf{u}}_j' - \mathbf{u}_j \mathbf{u}_j')$ . By definition  $\hat{\mathbf{u}}_j = \mathbf{y}_j - \widehat{\Psi}_t \mathbf{y}_{j-1} = (\Psi_{j-1} - \widehat{\Psi}_t) \mathbf{y}_{j-1} + \mathbf{u}_j$ . Hence  $\hat{\mathbf{u}}_j \hat{\mathbf{u}}_j' - \mathbf{u}_j \mathbf{u}_j' = (\Psi_{j-1} - \widehat{\Psi}_t) \mathbf{y}_{j-1} \mathbf{u}_j' + \mathbf{u}_j \mathbf{y}_{j-1}' (\Psi_{j-1} - \widehat{\Psi}_t)' + (\Psi_{j-1} - \widehat{\Psi}_t) \mathbf{y}_{j-1} \mathbf{y}_{j-1}' (\Psi_{j-1} - \widehat{\Psi}_t)'$ . Use equality  $\Psi_{j-1} - \widehat{\Psi}_t = (\Psi_{j-1} - \Psi_t) + (\Psi_t - \widehat{\Psi}_t)$  and the bound  $\|(\Psi_{j-1} - \widehat{\Psi}_t) \mathbf{y}_{j-1} \mathbf{y}_{j-1}' (\Psi_{j-1} - \widehat{\Psi}_t)'\|_{sp} \leq (\|\Psi_{j-1} - \Psi_t\|_{sp} + \|\widehat{\Psi}_t - \Psi_t\|_{sp})^2 \|\mathbf{y}_{j-1}\|^2 \leq 2(\|\Psi_{j-1} - \Psi_t\|_{sp}^2 + \|\widehat{\Psi}_t - \Psi_t\|_{sp}^2) \|\mathbf{y}_{j-1}\|^2$ , to obtain

$$\begin{aligned} \|\Sigma_{\hat{u}\hat{u},t} - \Sigma_{uu,t}\|_{sp} &\leq 2L_t^{-1} \sum_{j=1}^n l_{tj} \|\mathbf{u}_j \mathbf{y}_{j-1}' (\Psi_{j-1} - \Psi_t)'\|_{sp} \\ &\quad + 2\|\Psi_t - \widehat{\Psi}_t\|_{sp} \|L_t^{-1} \sum_{j=1}^n l_{tj} \mathbf{u}_j \mathbf{y}_{j-1}'\|_{sp} + 2\|\Psi_t - \widehat{\Psi}_t\|_{sp}^2 L_t^{-1} \sum_{j=1}^n l_{tj} \|\mathbf{y}_{j-1}\|^2 \\ &\quad + 2L_t^{-1} \sum_{j=1}^n l_{tj} \|\Psi_{j-1} - \Psi_t\|_{sp}^2 \|\mathbf{y}_{j-1}\|^2 \\ &:= 2(q_{n,1} + \|\Psi_t - \widehat{\Psi}_t\|_{sp} q_{n,2} + \|\Psi_t - \widehat{\Psi}_t\|_{sp}^2 q_{n,3} + q_{n,4}). \end{aligned}$$

By Theorem 2.2(i),  $\|\Psi_t - \widehat{\Psi}_t\|_{sp} = O_p(\kappa_{n,\psi})$ ; by Lemma 6.2(i),  $q_{n,2} = O_p(H_h^{-1/2})$ ; by the same argument as in the proof of (6.7), it follows that  $q_{n,i} = O_p((\bar{H}_h/n)^{1/2} + H_h^{-1})$  for  $i = 1$  and  $i = 4$ , and  $q_{n,3} = O_p(1)$  because  $Eq_{n,3} \leq CL_t^{-1} \sum_{j=1}^n l_{tj} = C$  by (2.8). So,

$\|\Sigma_{\hat{u}\hat{u},t} - \Sigma_{uu,t}\|_{sp} = O_p(\kappa_{n,\psi}^2 + \kappa_{n,\psi} H_h^{-1/2} + (\bar{H}_h/n)^{1/2} + H_h^{-1}) = O_p(\kappa_{n,\psi}^2 + (\bar{H}_h/n)^{1/2} + H_h^{-1})$ , proving (6.44)(a).

To show (6.44)(b), we will check

$$(j) \|\Sigma_{\hat{u}\hat{u},t} - \Sigma_t\|_{sp} = o_p(1), \quad (jj) \|\Sigma_t^{-1/2}\|_{sp} = O_p(1), \quad (jjj) \|\Sigma_{\hat{u}\hat{u},t}^{-1/2}\|_{sp} = O_p(1), \quad (6.47)$$

which implies (6.44)(b) arguing as in the proof of (6.21). Firstly,  $\|\Sigma_{\hat{u}\hat{u},t} - \Sigma_t\|_{sp} \leq \|\Sigma_{\hat{u}\hat{u},t} - \Sigma_{uu,t}\|_{sp} + \|\Sigma_{uu,t} - \Sigma_t\|_{sp} = o_p(1)$  by (6.44)(a) and (6.32). Next,  $\|\Sigma_t^{-1/2}\|_{sp} \leq \|\mathbf{H}_t^{-1}\|_{sp} = O_p(1)$  by Assumption 2.3(iii), while  $\|\Sigma_{\hat{u}\hat{u},t}^{-1}\|_{sp} \leq \|\Sigma_{uu,t}^{-1}\|_{sp} \|(1 + \Sigma_{uu,t}^{-1}(\Sigma_{\hat{u}\hat{u},t} - \Sigma_{uu,t})^{-1})\|_{sp} \leq O_p(1/(1 - \|\Sigma_{uu,t}^{-1}\|_{sp} \|\Sigma_{\hat{u}\hat{u},t} - \Sigma_{uu,t}\|_{sp})) = O_p(1)$  implies  $\|\Sigma_{\hat{u}\hat{u},t}^{-1/2}\|_{sp} = \|\Sigma_{\hat{u}\hat{u},t}^{-1}\|_{sp}^{1/2} = O_p(1)$ . This completes the proof of (6.44)(b).

(ii) *Proof of (6.45)(a)*. By (2.18),  $\mathbf{y}_j = \boldsymbol{\mu}_j + \dot{\mathbf{y}}_j$  where  $\dot{\mathbf{y}}_j$  is a VAR(1) process with no intercept, so that  $\dot{\mathbf{y}}_j - \Psi_{j-1}\dot{\mathbf{y}}_{j-1} = \mathbf{u}_j$ . By definition,  $\hat{\mathbf{u}}_j = \hat{\mathbf{y}}_j - \hat{\Psi}_t \hat{\mathbf{y}}_{j-1}$  where  $\hat{\mathbf{y}}_j = \mathbf{y}_j - \bar{\mathbf{y}}_t$ . Hence,

$$\begin{aligned} \hat{\mathbf{u}}_j - \mathbf{u}_j &= \hat{\mathbf{y}}_j - \hat{\Psi}_t \hat{\mathbf{y}}_{j-1} - (\dot{\mathbf{y}}_j - \Psi_{j-1} \dot{\mathbf{y}}_{j-1}) = (\hat{\mathbf{y}}_j - \dot{\mathbf{y}}_j) - (\hat{\Psi}_t \hat{\mathbf{y}}_{j-1} - \Psi_{j-1} \dot{\mathbf{y}}_{j-1}) \\ &= (\hat{\mathbf{y}}_j - \dot{\mathbf{y}}_j) - \hat{\Psi}_t (\hat{\mathbf{y}}_{j-1} - \dot{\mathbf{y}}_{j-1}) + (\Psi_{j-1} - \hat{\Psi}_t) \dot{\mathbf{y}}_{j-1}. \end{aligned}$$

Since  $\hat{\mathbf{y}}_j - \dot{\mathbf{y}}_j = \boldsymbol{\mu}_j - \bar{\mathbf{y}}_t$ , then

$$\begin{aligned} \hat{\mathbf{u}}_j - \mathbf{u}_j &= (\boldsymbol{\mu}_j - \bar{\mathbf{y}}_t) - \hat{\Psi}_t (\boldsymbol{\mu}_{j-1} - \bar{\mathbf{y}}_t) + (\Psi_{j-1} - \hat{\Psi}_t) \dot{\mathbf{y}}_{j-1} \\ &=: D_{tj} = D_{tj,1} + D_{t,2} + D_{t,3} \dot{\mathbf{y}}_{j-1}, \quad \text{where} \\ D_{tj,1} &:= (\boldsymbol{\mu}_j - \boldsymbol{\mu}_t) - \hat{\Psi}_t (\boldsymbol{\mu}_{j-1} - \boldsymbol{\mu}_t) + (\Psi_{j-1} - \Psi_t) \dot{\mathbf{y}}_{j-1}, \\ D_{t,2} &:= (\boldsymbol{\mu}_t - \bar{\mathbf{y}}_t) - \hat{\Psi}_t (\boldsymbol{\mu}_t - \bar{\mathbf{y}}_t), \quad D_{t,3} := \Psi_t - \hat{\Psi}_t. \end{aligned}$$

Then,

$$\begin{aligned} \hat{\mathbf{u}}_j \hat{\mathbf{u}}_j' - \mathbf{u}_j \mathbf{u}_j' &= \mathbf{u}_j D_{tj}' + D_{tj} \mathbf{u}_j' + D_{tj} D_{tj}' \\ &= (\mathbf{u}_j D_{t,2}' + D_{t,2} \mathbf{u}_j') + (\mathbf{u}_j \dot{\mathbf{y}}_{j-1}' D_{t,3}' + D_{t,3} \dot{\mathbf{y}}_{j-1} \mathbf{u}_j') + (\mathbf{u}_j D_{tj,1}' + D_{tj,1} \mathbf{u}_j') + D_{tj} D_{tj}'. \end{aligned}$$

Hence

$$\begin{aligned}
\|\Sigma_{\hat{u}\hat{u},t} - \Sigma_{uu,t}\|_{sp} &= \|L_t^{-1} \sum_{j=1}^n l_{tj}(\hat{\mathbf{u}}_j \hat{\mathbf{u}}_j' - \mathbf{u}_j \mathbf{u}_j')\|_{sp} \leq 2\|D_{t,2}\|_{sp} \|L_t^{-1} \sum_{j=1}^n l_{tj} \mathbf{u}_j\| \\
&\quad + 2\|D_{t,3}\|_{sp} \|L_t^{-1} \sum_{j=1}^n l_{tj} \mathbf{u}_j \dot{\mathbf{y}}_{j-1}'\|_{sp} + 2L_t^{-1} \sum_{j=1}^n l_{tj} \|D_{tj,1}\|_{sp} \|\dot{\mathbf{u}}_j\| + L_t^{-1} \sum_{j=1}^n l_{tj} \|D_{tj}\|_{sp}^2. \\
&=: s_{n,1} + s_{n,2} + s_{n,3} + s_{n,4}.
\end{aligned}$$

It remains to show that

$$s_{n,i} = O_p(\kappa_{n,\psi}^2 + (\bar{H}_h/n)^{1/2} + H_h^{-1}), \quad i = 1, \dots, 4. \quad (6.48)$$

Observe that

$$\|D_{t,i}\|_{sp} = O_p(\kappa_{n,\psi}), \quad i = 2, 3, \quad (6.49)$$

which follows from  $\|D_{t,2}\|_{sp} \leq \|\boldsymbol{\mu}_t - \bar{\mathbf{y}}_t\|_{sp} + \|\widehat{\boldsymbol{\Psi}}_t\|_{sp} \|\boldsymbol{\mu}_t - \bar{\mathbf{y}}_t\|_{sp} = O_p(\kappa_{n,\psi})$  and  $\|D_{t,3}\|_{sp} \leq \|\boldsymbol{\Psi}_t - \widehat{\boldsymbol{\Psi}}_t\|_{sp}$ , applying the first two claims of Theorem (2.3)(i). In addition, by Lemma 6.2(i),  $\|L_t^{-1} \sum_{j=1}^n l_{tj} \mathbf{u}_j \dot{\mathbf{y}}_{j-1}'\|_{sp} = O_p(H_h^{-1/2})$  and  $\|L_t^{-1} \sum_{j=1}^n l_{tj} \mathbf{u}_j\| = O_p(H_h^{-1/2})$ . Hence  $s_{n,1} + s_{n,2} = O_p(\kappa_{n,\psi} H_h^{-1/2}) = O_p(\kappa_{n,\psi}^2 + H_h^{-2})$  satisfies (6.48).

To bound  $s_{n,3}$ , observe that  $\|D_{tj,1}\|_{sp} \leq \|\boldsymbol{\mu}_j - \boldsymbol{\mu}_t\|_{sp} + \|\widehat{\boldsymbol{\Psi}}_t\|_{sp} \|\boldsymbol{\mu}_{j-1} - \boldsymbol{\mu}_t\|_{sp} + \|\boldsymbol{\Psi}_{j-1} - \boldsymbol{\Psi}_t\|_{sp} \|\dot{\mathbf{y}}_{j-1}\|$ . Since  $\|\widehat{\boldsymbol{\Psi}}_t\|_{sp} = O_p(1)$ , then  $s_{n,3} \leq O_p(1) L_t^{-1} \sum_{j=1}^n l_{tj} (\|\boldsymbol{\mu}_j - \boldsymbol{\mu}_t\|_{sp} + \|\boldsymbol{\mu}_{j-1} - \boldsymbol{\mu}_t\|_{sp} + \|\boldsymbol{\Psi}_{j-1} - \boldsymbol{\Psi}_t\|_{sp}) \|\dot{\mathbf{y}}_{j-1}\| (1 + \|\dot{\mathbf{y}}_{j-1}\|) = O_p((\bar{H}_h/n)^{1/2} + H_h^{-1})$ , which follows using Assumption 2.3(i)-(ii) and Assumption 2.1 about  $\boldsymbol{\mu}_j$  and  $\boldsymbol{\Psi}_j$ , combining arguments used in the proof of (6.7) and (6.31).

To estimate  $s_{n,4}$ , bound  $\|D_{tj}\|^2 \leq 3(\|D_{tj,1}\|^2 + \|D_{t,2}\|^2 + \|D_{t,3}\|^2 \|\dot{\mathbf{y}}_{j-1}\|^2)$ . Thus,

$$\begin{aligned}
s_{n,4} &\leq 3L_t^{-1} \sum_{j=1}^n l_{tj} \|D_{tj,1}\|^2 + 3\{\|D_{t,2}\|^2 + \|D_{t,3}\|^2\} L_t^{-1} \sum_{j=1}^n l_{tj} (1 + \|\dot{\mathbf{y}}_{j-1}\|^2) \\
&=: d_{n,1} + (\|D_{t,2}\|^2 + \|D_{t,3}\|^2) d_{n,2} = d_{n,1} + O_p(\kappa_{n,\psi}^2) d_{n,2}
\end{aligned}$$

by (6.49). The same argument as used above to bound  $s_{n,3}$  implies  $d_{n,1} = O_p((\bar{H}_h/n)^{1/2} + H_h^{-1})$ , while by (2.8),  $E d_{n,2} = E L_t^{-1} \sum_{j=1}^n l_{tj} (1 + \|\dot{\mathbf{y}}_{j-1}\|^2) \leq C L_t^{-1} \sum_{j=1}^n l_{tj} = C$ , which yields  $d_{n,2} = O_p(1)$ . Hence,  $s_{n,4} = O_p(\kappa_{n,\psi}^2 + (\bar{H}_h/n)^{1/2} + H_h^{-1})$ . This completes the proof of (6.48) and (6.45)(a)

To show (6.45)(b), similarly as proving (6.44)(b), it suffices to check validity of (6.47). In this case, (6.47)(j) follows from (6.45) and (6.32), while (6.47)(jj)-(jjj) hold by the same argument as in the proof of (6.44)(b). This completes the proof of Lemma 6.5(ii).

(iii) First we show that

$$\|\mathbf{V}_{\hat{y}\hat{y},t} - \mathbf{V}_{\dot{y}\dot{y},t}\|_{sp} = O_p(\kappa_{n,\psi}^*). \quad (6.50)$$

Since  $\hat{\mathbf{y}}_j - \dot{\mathbf{y}}_j = \boldsymbol{\mu}_j - \bar{\mathbf{y}}_t = (\boldsymbol{\mu}_j - \boldsymbol{\mu}_t) + (\boldsymbol{\mu}_t - \bar{\mathbf{y}}_t)$ , then  $\hat{\mathbf{y}}_j \hat{\mathbf{y}}_j' - \mathbf{y}_j \mathbf{y}_j' = (\boldsymbol{\mu}_j - \bar{\mathbf{y}}_t) \dot{\mathbf{y}}_j' + \dot{\mathbf{y}}_j (\boldsymbol{\mu}_j - \bar{\mathbf{y}}_t)' + (\boldsymbol{\mu}_j - \bar{\mathbf{y}}_t) (\boldsymbol{\mu}_j - \bar{\mathbf{y}}_t)'$ . Then,

$$\begin{aligned} \|\mathbf{V}_{\hat{y}\hat{y},t} - \mathbf{V}_{\dot{y}\dot{y},t}\|_{sp} &= \|K_t^{-1} \sum_{j=1}^n k_{tj} (\hat{\mathbf{y}}_{j-1} \hat{\mathbf{y}}_{j-1}' - \mathbf{y}_{j-1} \mathbf{y}_{j-1}')\|_{sp} \\ &\leq 2K_t^{-1} \sum_{j=1}^n k_{tj} \|\boldsymbol{\mu}_{j-1} - \boldsymbol{\mu}_t\| \|\dot{\mathbf{y}}_j\| + 2\|\boldsymbol{\mu}_t - \bar{\mathbf{y}}_t\| \|K_t^{-1} \sum_{j=1}^n k_{tj} \dot{\mathbf{y}}_j\| \\ &\quad + K_t^{-1} \sum_{j=1}^n k_{tj} \|\boldsymbol{\mu}_{j-1} - \bar{\mathbf{y}}_t\|^2 =: 2p_{n,1} + 2p_{n,2} + p_{n,3}. \end{aligned}$$

It remains to show that

$$p_{n,i} = O_p(\kappa_{n,\psi}^*), \quad i = 1, 2, 3. \quad (6.51)$$

For  $i = 1$ , using Assumption 2.3 about  $\boldsymbol{\mu}_j$ , (6.51) follows combining arguments used in the proof of (6.7) and (6.31).

For  $i = 2$ , note that  $\|\boldsymbol{\mu}_t - \bar{\mathbf{y}}_t\| = O_p(\kappa_{n,\psi})$  by (6.9) and  $\|\bar{\dot{\mathbf{y}}}_t\| \equiv \|K_t^{-1} \sum_{j=1}^n k_{tj} \dot{\mathbf{y}}_j\| = O_p(\kappa_{n,\psi})$  by (6.38) of Lemma 6.4, which implies  $p_{n,2} = O_p(\kappa_{n,\psi}^2) = O_p(\kappa_{n,\psi}^*)$ .

For  $i = 3$ , bound  $\|\boldsymbol{\mu}_j - \bar{\mathbf{y}}_t\|^2 = \|(\boldsymbol{\mu}_j - \boldsymbol{\mu}_t) + (\boldsymbol{\mu}_t - \bar{\mathbf{y}}_t)\|^2 \leq 2\|\boldsymbol{\mu}_j - \boldsymbol{\mu}_t\|^2 + 2\|\boldsymbol{\mu}_t - \bar{\mathbf{y}}_t\|^2 = 2\|\boldsymbol{\mu}_j - \boldsymbol{\mu}_t\|^2 + O_p(\kappa_{n,\psi}^*)$ . Then  $p_{n,3} \leq K_t^{-1} \sum_{j=1}^n k_{tj} \|\boldsymbol{\mu}_{j-1} - \boldsymbol{\mu}_t\|^2 + O_p(\kappa_{n,\psi}^*) = O_p(\kappa_{n,\psi}^*)$  by the same argument in the case  $i = 1$ . This completes the proof of (6.50). (Observe that the same argument implies (6.46)(c) of the lemma.)

To show (6.46)(a), write  $\mathbf{V}_{\hat{y}\hat{y},t} = \mathbf{V}_{\dot{y}\dot{y},t}(I + \Delta)$ ,  $\Delta = \mathbf{V}_{\dot{y}\dot{y},t}^{-1}(\mathbf{V}_{\hat{y}\hat{y},t} - \mathbf{V}_{\dot{y}\dot{y},t})$ . Observe that  $\|\Delta\|_{sp} \leq \|\mathbf{V}_{\dot{y}\dot{y},t}^{-1}\|_{sp} \|\mathbf{V}_{\hat{y}\hat{y},t} - \mathbf{V}_{\dot{y}\dot{y},t}\|_{sp} = O_p(\kappa_{n,\psi}^*) = o_p(1)$  by (6.50), and



$\|\mathbf{V}_{\hat{y}\hat{y},t}^{-1}\|_{sp} = O_p(1)$  of Lemma 6.1(vi). Then,  $\|\mathbf{V}_{\hat{y}\hat{y},t}^{-1} - \mathbf{V}_{\dot{y}\dot{y},t}^{-1}\|_{sp} \leq \|\mathbf{V}_{\dot{y}\dot{y},t}^{-1}\|_{sp} \|(I + \Delta)^{-1} - I\|_{sp} = O_p(\|\Delta\|_{sp}/(1 - \|\Delta\|_{sp})) = O_p(\kappa_{n,\psi}^*)$ , which proves (a). In addition,  $\|\mathbf{V}_{\hat{y}\hat{y},t}^{-1}\|_{sp} \leq \|\mathbf{V}_{\dot{y}\dot{y},t}^{-1}\|_{sp} \|(I + \Delta)^{-1}\|_s = O_p((1 - \|\Delta\|_{sp})^{-1}) = O_p(1)$ , while  $\|\mathbf{V}_{\hat{y}\hat{y},t}\|_{sp} \leq \|\mathbf{V}_{\hat{y}\hat{y},t} - \mathbf{V}_{\dot{y}\dot{y},t}\|_{sp} + \|\mathbf{V}_{\dot{y}\dot{y},t}\|_{sp} = O_P(1)$  by (6.50) and (6.15)(vi), which proves (d).

Finally, (b) follows from (6.50) and the bounds  $\|\mathbf{V}_{\hat{y}\hat{y},t}^{-1}\|_{sp} = O_P(1)$ ,  $\|\mathbf{V}_{\dot{y}\dot{y},t}^{-1}\|_{sp} = O_p(1)$ , arguing as in the proof of (6.21).

*Proof of (iv).* Observe that  $\widehat{\mathbf{D}}_t$   $\mathbf{D}_t$  are scalars.

(a) The matrix  $\mathbf{V}_{\psi,t}$  is symmetric and positive definite. Thus, it has positive eigenvalues and its spectral decomposition implies that  $\mathbf{V}_{\psi,t}^{-1}$  has positive eigenvalues and is positive definite. Hence with probability tending to 1,  $\mathbf{D}_t = 1 + \boldsymbol{\mu}'_t \mathbf{V}_{\psi,t}^{-1} \boldsymbol{\mu}_t \geq 1 + o_p(1)$  which yields (iv)(a),  $\mathbf{D}_t^{-1/2} = O_p(1)$ .

To show (b), notice that

$$\widehat{\mathbf{D}}_t - \mathbf{D}_t = o_p(1). \quad (6.52)$$

Indeed,  $|\widehat{\mathbf{D}}_t - \mathbf{D}_t| = |\hat{\boldsymbol{\mu}}'_t \widehat{\mathbf{V}}_{\hat{y}\hat{y},t}^{-1} \hat{\boldsymbol{\mu}}_t - \boldsymbol{\mu}'_t \mathbf{V}_{\psi,t}^{-1} \boldsymbol{\mu}_t| \leq \|\hat{\boldsymbol{\mu}}_t - \boldsymbol{\mu}_t\| \|\widehat{\mathbf{V}}_{\hat{y}\hat{y},t}^{-1}\|_{sp} \|\hat{\boldsymbol{\mu}}_t\| + \|\boldsymbol{\mu}_t\| \|\widehat{\mathbf{V}}_{\hat{y}\hat{y},t}^{-1} - \mathbf{V}_{\psi,t}^{-1}\|_{sp} \|\hat{\boldsymbol{\mu}}_t\| + \|\boldsymbol{\mu}_t\| \|\mathbf{V}_{\psi,t}^{-1}\|_{sp} \|\hat{\boldsymbol{\mu}}_t - \boldsymbol{\mu}_t\| = o_p(1)$ , since  $\|\boldsymbol{\mu}_t\| = O_p(1)$ ,  $\|\hat{\boldsymbol{\mu}}_t\| = O_p(1)$  and  $\|\hat{\boldsymbol{\mu}}_t - \boldsymbol{\mu}_t\| = o_p(1)$  by Assumption 2.3 and Theorem 2.3(ii);  $\|\mathbf{V}_{\hat{y}\hat{y},t}^{-1} - \mathbf{V}_{\psi,t}^{-1}\|_{sp} \leq \|\mathbf{V}_{\hat{y}\hat{y},t}^{-1} - \mathbf{V}_{\dot{y}\dot{y},t}^{-1}\|_{sp} + \|\mathbf{V}_{\dot{y}\dot{y},t}^{-1} - \mathbf{V}_{\psi,t}^{-1}\|_{sp} = o_p(1)$  by (6.46)(a) and Lemma 6.1(iii),  $\|\widehat{\mathbf{V}}_{\hat{y}\hat{y},t}^{-1}\|_{sp} = O_p(1)$  by (6.46)(d) and  $\|\mathbf{V}_{\psi,t}^{-1}\|_{sp} = O_p(1)$  by Lemma 6.1(ii). 6.1(iii)

Write  $\widehat{\mathbf{D}}_t = \mathbf{D}_t(I + \Delta)$ ,  $\Delta := \mathbf{D}_t^{-1}(\widehat{\mathbf{D}}_t - \mathbf{D}_t)$ . Then,  $|\Delta| \leq |\mathbf{D}_t^{-1}| |\widehat{\mathbf{D}}_t - \mathbf{D}_t| = o_p(1)$  by (a) and (6.52), which implies (iv)(b):  $\widehat{\mathbf{D}}_t^{-1/2} = \mathbf{D}_t^{-1/2}(I + \Delta)^{-1/2} = O_p(1)$ .

To show (c), use the bound  $\mathbf{D}_t \geq 1 + o_p(1)$  we showed proving (a), to obtain  $\widehat{\mathbf{D}}_t = \mathbf{D}_t + (\widehat{\mathbf{D}}_t - \mathbf{D}_t) = \mathbf{D}_t + o_p(1) \geq 1 + o_p(1)$ . This together with (6.52) implies  $\widehat{\mathbf{D}}_t^{-1/2} - \mathbf{D}_t^{-1/2} = (\widehat{\mathbf{D}}_t - \mathbf{D}_t)(\widehat{\mathbf{D}}_t^{-1/2} + \mathbf{D}_t^{-1/2})^{-1} = (\widehat{\mathbf{D}}_t - \mathbf{D}_t)O_p(1)$ . It remains to notice that  $\widehat{\mathbf{D}}_t - \mathbf{D}_t = o_p(1)$  by (6.52), while  $|\mathbf{D}_t - \mathbf{D}_{t-h}| = |\boldsymbol{\mu}'_t \mathbf{V}_{\psi,t}^{-1} \boldsymbol{\mu}_t - \boldsymbol{\mu}'_{t-h} \mathbf{V}_{\psi,t-h}^{-1} \boldsymbol{\mu}_{t-h}| \leq \|\boldsymbol{\mu}_t - \boldsymbol{\mu}_{t-h}\| \|\mathbf{V}_{\psi,t}^{-1}\|_{sp} \|\boldsymbol{\mu}_t\| + \|\boldsymbol{\mu}_{t-h}\| \|\mathbf{V}_{\psi,t}^{-1} - \mathbf{V}_{\psi,t-h}^{-1}\|_{sp} \|\boldsymbol{\mu}_t\| + \|\boldsymbol{\mu}_{t-h}\| \|\mathbf{V}_{\psi,t-h}^{-1}\|_{sp} \|\boldsymbol{\mu}_t - \boldsymbol{\mu}_{t-h}\|$

$\|\boldsymbol{\mu}_{t-h}\| = o_p(1)$ , since  $\|\boldsymbol{\mu}_t\| = O_p(1)$ ,  $\|\boldsymbol{\mu}_{t-h}\| = O_p(1)$  and  $\|\boldsymbol{\mu}_t - \boldsymbol{\mu}_{t-h}\| = o_p(1)$  by Assumption 2.3;  $\|\mathbf{V}_{\psi,t}^{-1} - \mathbf{V}_{\psi,t-h}^{-1}\|_{sp} = o_p(1)$  by (6.19), and  $\|\mathbf{V}_{\psi,t-h}^{-1}\|_{sp} = O_p(1)$ ,  $\|\mathbf{V}_{\psi,t-h}^{-1}\|_{sp} = O_p(1)$  by Lemma 6.1(ii). This completes the proof of (c) and the lemma.  $\square$