

Dynamic Vector Mode Regression

Gordon C.R. Kemp

Department of Economics, University of Essex
kempgcr@essex.ac.uk

Paulo M.D.C. Parente

ISEG-UL - Universidade de Lisboa; REM - Research in Economics and Mathematics;
CEMAPRE- Centro de Matemática Aplicada à Previsão e Decisão Económica
pparente@iseg.ulisboa.pt

J.M.C. Santos Silva

School of Economics, University of Surrey
jmcss@surrey.ac.uk

11 November 2018

Abstract

We study the semi-parametric estimation of the conditional mode of a random vector that has a continuous conditional joint density with a well-defined global mode. A novel full-system estimator is proposed and its asymptotic properties are studied. We specifically consider the estimation of vector autoregressive conditional mode models and of systems of linear simultaneous equations defined by mode restrictions. The proposed estimator is easy to implement and simulations suggest that it is reasonably behaved in finite samples. An empirical example illustrates the application of the proposed methods, including its use to obtain multi-step forecasts and to construct impulse response functions.

Key words: Impulse response functions, Multivariate conditional mode, Robust regression, Simultaneous equations, Vector autoregression.

1. INTRODUCTION

The mode is an interesting measure of location for multivariate data, not only because of its intuitively appealing interpretation, but also because it is currently the only practical multivariate measure of location that is robust in the sense that it is not sensitive to perturbations of the tails of the distribution (see Tsay, Peña, and Pankratz, 2000, and Galeano, Peña, and Tsay, 2006, for details on the importance of outliers in a multivariate context). The interest of the multivariate mode is reflected in the continued attention that it has received in the literature since the pioneering work by Konakov (1973), Samanta (1973), and Sager (1978, 1979); see, e.g., the contributions by Abraham, Biau, and Cadre (2003), Mokkadem and Pelletier (2003), Klemelä (2005), and Hsu and Wu (2013).

The attractive properties of the multivariate mode extend naturally to the conditional case, and the conditional mode of a multivariate distribution is likely to be of interest in areas such as economics that have systems of equations at their core. For instance, in a standard supply and demand system, the conditional multivariate mode will be informative about how the relevant covariates affect the modal realization of the equilibrium price-quantity pair. Key economic variables have skewed distributions and are not conditionally independent (see, e.g., Smith and Vahey, 2016) and in multivariate models involving such variables the difference between the modal value of the vector and the vector of marginal modal (or mean) values can be substantial.

The conditional multivariate mode may also be of interest as a predictor. For the univariate case, the use of the conditional mode as a predictor was emphasized by Collomb, Härdle, and Hassani (1987) and more recently by Yao and Li (2014a) and by Chen, Genovese, Tibshirani, and Wasserman (2016), who show that for a given level of confidence prediction sets based on the conditional mode can be smaller than those based on the conditional mean. An example of the use of univariate mode as a predictor can be found in the Bank of England's quarterly *Inflation Report*, which presents univariate mode-based predictions of inflation and output growth. Of course, it might also be interesting to consider a predictor based on the mode of the joint distribution of the two variates.

In economics, systems of equations are often dynamic; that is the case, for example, of the systems of simultaneous equations considered by Haavelmo (1943), and of the popular vector autoregressive models (Sims, 1980). Therefore, it is of particular interest to study the estimation of the multivariate conditional mode in a time series context, explicitly allowing for dynamic specifications and dependent data. Estimation of the univariate conditional mode allowing for dependent data was pioneered by Collomb, Härdle, and Hassani (1987). However, because in general the mode of a multivariate distribution is not the vector of the marginal modes, multivariate mode regression cannot be performed using single-equation estimators developed for the univariate case.

In this paper we consider the semi-parametric estimation of the conditional multivariate mode, or multivariate mode regression, for a random vector that has a continuous conditional joint density with a well-defined global mode. As in Lee (1989, 1993) and Kemp and Santos Silva (2012), the proposed estimator is semi-parametric in the sense that the conditional mode is specified as a parametric function but only mild assumptions are made about the conditional distribution of interest (see also the related work by Yao and Li, 2014a, and, 2014b). We develop a novel full-system conditional mode regression estimator which can be seen as a multivariate generalization of the estimator introduced by Kemp and Santos Silva (2012) and that, as far as we are aware, is the first conditional multivariate mode estimator. We derive the asymptotic properties of the estimator allowing for dependent data and therefore, as a by-product, we generalize to the time-series context both the results of Kemp and Santos Silva (2012) and previous work on unconditional multivariate mode estimation. Additionally, our results are obtained allowing for a stochastic bandwidth, which is another important extension of the work by Kemp and Santos Silva (2012).

We consider two particular cases where the methods we propose can be of interest. We start by studying the estimation of vector autoregressive conditional mode models and then consider the estimation of systems of linear simultaneous equations defined by conditional mode restrictions. In the latter case we investigate the conditions under which it is possible to identify the structural parameters of interest, both in the context of classic systems of simultaneous equations and in structural vector autoregressive models.

To illustrate the application and usefulness of the proposed methods, we include a simple empirical example which shows how to obtain mode-based multi-step forecasts and construct impulse response functions; the good performance of the mode-based forecasts is particularly noteworthy.

The remainder of the paper is organized as follows. The next section sets up the problem and presents the main results on the estimation of multivariate dynamic conditional mode models. Section 3 considers the estimation of systems of linear simultaneous equations defined by conditional mode restrictions. Section 4 presents simulation results, and Section 5 provides an illustrative empirical example. Finally, Section 6 concludes and discusses directions for future research. The proofs of all theorems are presented in an online appendix.

2. MAIN RESULTS

2.1. Model and estimator

We consider systems of the form

$$Y_t = A_0 Z_t + U_t, \tag{1}$$

where Y_t and U_t are $G \times 1$ random vectors, Z_t is a $K \times 1$ vector that can contain exogenous variables and lagged values of Y_t , and A_0 is a $G \times K$ matrix of unknown parameters such that $A_0 \in \mathcal{A}$, where \mathcal{A} is the parameter space.

Systems of the form of (1) are often used in economics. Examples include the reduced form of systems of simultaneous equations (Haavelmo, 1943), systems of seemingly unrelated equations (Zellner, 1962), and vector autoregressive models (Sims, 1980); all these systems are generally interpreted as representing conditional expectations, whereas we will consider the case in which the systems define conditional multivariate modes.

Suppose that we have a sample $\{(Y_t, Z_t)\}_{t=1}^T$ of size T from the strictly stationary ergodic sequence of random vectors $\{(Y_t, Z_t)\}_{t=-\infty}^{\infty}$, and let \mathcal{F}_{t-1} denote the σ -algebra generated by $\{(Y_{t-1-j}, Z_{t-1-j})\}_{j=0}^{\infty}$. Also, let $\mathcal{P} = (\Omega, \mathcal{F}, \mathbb{P})$ denote the underlying probability space for $\{(Y_t, Z_t)\}_{t=-\infty}^{\infty}$ where, as usual, Ω denotes the sample space, \mathcal{F} is the σ -algebra of events,

and \mathbb{P} is a probability measure. We are interested in the case where the conditional mode of U_t given \mathcal{F}_{t-1} , denoted $\text{Mode}(U_t|\mathcal{F}_{t-1})$, is equal to zero. Then, because Z_t is measurable with respect to \mathcal{F}_{t-1} for each t , the conditional mode of Y_t given \mathcal{F}_{t-1} , denoted $\text{Mode}(Y_t|\mathcal{F}_{t-1})$, satisfies:

$$\text{Mode}(Y_t|\mathcal{F}_{t-1}) = \text{Mode}(A_0 Z_t + U_t|\mathcal{F}_{t-1}) = A_0 Z_t.$$

As in the pioneering work of Lee (1989, 1993) and in Kemp and Santos Silva (2012), we obtain our estimator for A_0 as the minimizer of a loss function, with the difference being that here the loss function is multivariate. In particular, we consider a loss function of the form

$$L_T(Y_t, Z_t, A) = 1 - \varrho \mathcal{K}\left(\frac{Y_t - AZ_t}{\delta_T}\right), \quad (2)$$

where $\mathcal{K}(\cdot)$ denotes a multivariate smooth kernel function, $\varrho = \mathcal{K}(0)^{-1}$ is a scaling constant, and δ_T is a strictly positive bandwidth that depends on T . As shown below, the minimizer of the expectation of $L_T(Y_t, Z_t, A)$ will approach the conditional mode as $\delta_T \rightarrow 0$. Notice that, as the bandwidth approaches 0, $L_T(Y_t, Z_t, A)$ approaches a multivariate version of the 0-1 loss, whose expected value is minimized when the mode is used as the predictor (see, e.g., Ferguson, 1967, or Hastie, Tibshirani, and Friedman, 2009). The 0-1 loss function is often used in classification problems when the variate of interest is discrete. For continuous variables, the centre of the modal interval is the optimal predictor when the objective is to maximize the probability that the prediction is within a given tolerance of the actual realization (Ferguson, 1967, Manski, 1991). This corresponds to the use of the step loss function, a practice with a long tradition in the statistical analysis of quality control problems (e.g., Trietsch, 1999). In this case, the mode emerges as the optimal predictor when the tolerance goes to zero and therefore the step loss function approaches the 0-1 loss function.

Minimizing the sample analog of the expectation of (2) is equivalent to maximizing

$$Q_T(A) \equiv T^{-1} \sum_{t=1}^T \delta_T^{-G} \mathcal{K}\left(\frac{Y_t - AZ_t}{\delta_T}\right), \quad (3)$$

which is a multivariate version of the objective function considered by Kemp and Santos Silva (2012).

Although our asymptotic results will be obtained under much more general conditions, here we focus on the multiplicative standard normal kernel $\mathcal{K}\left(\frac{Y_t - AZ_t}{\delta_T}\right) = (2\pi)^{-G/2} \exp\left(-\frac{(Y_t - AZ_t)'(Y_t - AZ_t)}{2\delta_T^2}\right)$; see, e.g., Scott (1992) for examples of other multivariate smooth kernels. This choice of kernel is not innocuous because it may be possible to obtain estimators with somewhat improved asymptotic properties by using different kernels (see Eddy, 1980, and Romano, 1988), but the multiplicative normal kernel has several important advantages. In particular, because it is essentially quadratic around 0, it generates a loss function which has both the multivariate mode and the multivariate mean as minimizers in limiting cases (see also Kemp and Santos Silva, 2012).

Further insights into the nature of the objective function based on the normal kernel can be obtained by noting that, under the assumptions to be defined below, minimizing the expectation of (2) when $\mathcal{K}(\cdot)$ is the multiplicative normal kernel is equivalent to solving the following set of moment conditions

$$\mathbb{E}\left[\exp\left(-\frac{(Y_t - AZ_t)'(Y_t - AZ_t)}{2\delta_T^2}\right)(Z_t \otimes I_G)(Y_t - AZ_t)\right] = 0, \quad (4)$$

where I_G is an identity matrix of order G . It is clear that (4) defines a multivariate weighted least squares problem where the weights are functions of the residuals of the G equations in the system, implying that the equations cannot be estimated one-by-one. As noted earlier, this is because in general the mode of a multivariate distribution is not the vector of the marginal modes and therefore estimation of A_0 has to be performed using a full-system estimator. However, the weights approach a constant as δ_T passes to infinity and consequently, for large values of the bandwidth parameter, minimizing $\mathbb{E}[L_T(Y_t, Z_t, A)]$ is equivalent to estimating each equation by least squares. To put it differently, when $\mathcal{K}(\cdot)$ is the multiplicative standard normal kernel, minimizing $\mathbb{E}[L_T(Y_t, Z_t, A)]$ is equivalent to solving a set of moment conditions that identify $\text{Mode}(Y_t|\mathcal{F}_{t-1})$ when $\delta_T \rightarrow 0$, or $\mathbb{E}(Y_t|\mathcal{F}_{t-1})$ when $\delta_T \rightarrow \infty$.

These results show that, for our choice of kernel, minimization of (2) defines a continuum of multivariate conditional measures of central tendency of which the two polar cases have particularly interesting interpretations. For any other positive and finite choice of δ_T , minimization of $\mathbb{E}[L_T(Y_t, Z_t, A)]$ defines a measure of location which, in some sense,

is between the mean and the mode, and can be viewed as a multivariate generalization of the measure of location implicitly defined by a particular member of the class of M -estimators introduced by Huber (1973). That is, for $0 < \delta_T < \infty$ our estimator is a multivariate version of a robust M -estimator. As in Kemp and Santos Silva (2012), this has important implications for the choice of bandwidth because the bandwidth not only determines the properties of the estimator but also, and more importantly, defines the conditional measure of central tendency that is estimated.

The moment conditions in (4) are also informative about the choice of algorithm to maximize (2). Because $Q_T(A)$ is differentiable, it can be maximized using a Newton-type algorithm of the kind typically available in standard econometrics software. Moreover, (4) shows that an algorithm of this kind may be implemented as a multivariate version of the iterative reweighted least squares algorithm often used in robust regression estimation (e.g., Li, 1985, pp. 335-6). Finally, (4) also makes clear that, for large values of δ_T , (2) will have a single maximum. However, that will not be the case for small values of δ_T and therefore the researcher needs to ensure the estimates obtained correspond to the global maximum of $Q_T(A)$.

2.2. Asymptotic results

We now consider the asymptotic properties of the estimator of the parameters of the conditional mode, explicitly taking into account that in practice the bandwidth will be data dependent. In particular, we consider the properties of the estimator defined by

$$\widehat{A}_T = \arg \max_{A \in \mathcal{A}} \widehat{Q}_T(A), \quad (5)$$

where

$$\widehat{Q}_T(A) \equiv T^{-1} \sum_{t=1}^T \widehat{\delta}_T^{-G} \mathcal{K} \left(\frac{Y_t - AZ_t}{\widehat{\delta}_T} \right), \quad (6)$$

and $\widehat{\delta}_T$ is a strictly positive data-dependent bandwidth that depends on T . Throughout, we use $\|M\|$ to denote the non-negative square-root of the sum of the squares of the elements of any array M , i.e., $\|M\| = [\text{trace}(M'M)]^{1/2}$. Additionally, C is a finite positive constant, and we use the following convention for the derivatives of a vector-valued func-

tion $F(a)$ with respect to the vector a : $F^{(1)}(a) \equiv \partial F(a)/\partial a'$, $F^{(2)}(a) \equiv \partial^2 F(a)/\partial a \partial a'$, $F^{(3)}(a) \equiv \partial \text{vec}(F^{(2)}(a))/\partial a'$.

The following assumptions will be used in obtaining our results; the proofs of all theorems are provided in an online appendix.

1. (*Stationarity and Ergodicity*) $\{W_t\}_{t=-\infty}^{\infty}$ is a strictly stationary ergodic sequence of random vectors, where $W_t = (Y_t', Z_t)'$, defined on an underlying probability space $\mathcal{P} = (\Omega, \mathcal{F}, \mathbb{P})$.
2. (*Parameter Space*) \mathcal{A} is a compact subset of $\mathbb{R}^{G \times K}$ and A_0 is an element of \mathcal{A} .
3. (*Conditional Density I*) For each $-\infty < s < \infty$, let \mathcal{F}_{s-1} denote the σ -algebra generated by $\{(Y_{s-1-j}, Z_{s-j})\}_{j=0}^{\infty}$; then for each t there is a version of the conditional density function of $U_t = (Y_t - A_0 Z_t)$ given \mathcal{F}_{t-1} , denoted by $f_t(\cdot | \mathcal{F}_{t-1})$, such that: (i) $f_t(u | \mathcal{F}_{t-1}) \leq C$ for all $u \in \mathbb{R}^G$, $t = 1, 2, \dots$, and $\omega \in \Omega$; (ii) $f_t(u | \mathcal{F}_{t-1}) \leq f_t(0 | \mathcal{F}_{t-1})$ with equality if and only if $u = 0$ for all $t = 1, 2, \dots$, and $\omega \in \Omega$; and (iii) $f_t(u | \mathcal{F}_{t-1})$ is continuous in u for all $t = 1, 2, \dots$, and $\omega \in \Omega$.
4. (*Moments I*) $E(\|W_t\|) \leq C$.
5. (*No Multicollinearity*) $\Pr(AZ_t = 0) < 1$ for any fixed $A \in \mathbb{R}^{G \times K}$ such that $A \neq 0$.
6. (*Kernel Function I*) $\mathcal{K}(\cdot) : \mathbb{R}^G \rightarrow \mathbb{R}$ satisfies (i) $\int_{\mathbb{R}^G} \mathcal{K}(x) dx = 1$; $\int_{\mathbb{R}^G} |\mathcal{K}(x)| dx \leq C$; and (ii) $|\mathcal{K}(x)| \leq C$ and $\|\mathcal{K}^{(1)}(x)\| \leq C$ for all $x \in \mathbb{R}^G$.
7. (*Non-Stochastic Bandwidth Component I*) $\{\delta_T\}_{T=1}^{\infty}$ is a sequence of finite strictly positive constants such that: (i) $\delta_T = o(1)$; (ii) $\ln(T) / (T\delta_T^G) = o(1)$.
8. (*Stochastic Bandwidth Component I*) $\hat{\delta}_T = \delta_T / \hat{\gamma}_T$, where $\{\hat{\gamma}_T\}_{T=1}^{\infty}$ is a sequence of random variables defined on \mathcal{P} such that $\ln \hat{\gamma}_T = O_p(1)$.
9. (*Conditional Density II*) (i) $f_t(u | \mathcal{F}_{t-1})$ is three times differentiable with respect to u and $\left\| f_t^{(j)}(u | \mathcal{F}_{t-1}) \right\| \leq C$, $j = 1, 2, 3$, for all $u \in \mathbb{R}^G$, $t = 1, 2, \dots$, and $\omega \in \Omega$; (ii) $f_t^{(2)}(0 | \mathcal{F}_{t-1})$ is negative definite for all $\omega \in \Omega$.
10. (*Interior Parameter Value*) \mathcal{A} has a non-empty interior, denoted $\text{int}(\mathcal{A})$ and $A_0 \in \text{int}(\mathcal{A})$.
11. (*Moments II*) $E\left(\|W_t\|^{G+4+\tau}\right) \leq C$, for some $\tau > 0$.

12. (*Kernel Function II*) (i) $\int_{\mathbb{R}^G} x\mathcal{K}(x) dx = 0$; (ii) $\mathcal{K}(\cdot)$ is three times differentiable with $\|\mathcal{K}^{(j)}(x)\| \leq C$, $j = 2, 3$, for all $x \in \mathbb{R}^G$; (iii) $\int_{\mathbb{R}^G} \|\mathcal{K}^{(j)}(x)\|^2 dx \leq C$ for $j = 1, 2$; (iv) $\int \|x\| \|\mathcal{K}^{(1)}(x)\|^2 dx \leq C$; (v) $\int \|x\|^2 \|\mathcal{K}^{(2)}(x)\|^2 dx \leq C$; (vi) $\lim_{M \rightarrow \infty} \sup_{x: \|x\| \geq M} \|x\mathcal{K}(x)\| = 0$ and $\lim_{M \rightarrow \infty} \sup_{x: \|x\| \geq M} \|x\mathcal{K}^{(1)}(x)'\| = 0$.
13. (*Non-Stochastic Bandwidth Component II*) The sequence $\{\delta_T\}_{T=1}^\infty$ is such that: (i) $\frac{\ln(T)}{T\delta_T^{G+4}} = o(1)$; (ii) $T\delta_T^{G+6} = o(1)$.
14. (*Stochastic Bandwidth Component II*) $(T\delta_T^{G+2})^{1/2} \ln \hat{\gamma}_T = O_p(1)$.

These assumptions are largely similar to those in Kemp and Santos Silva (2012), but there are some notable differences. The major differences are that Assumptions 1 and 3 reflect the different nature of the problem considered in the present paper, and that Assumptions 8 and 14 explicitly allow for the use of a stochastic bandwidth of the type suggested by Silverman (1986, p. 45), something that was not done by Kemp and Santos Silva (2012). Note that although we only explicitly consider the case where the same bandwidth is used for all equations, our results extend straightforwardly to the case where different scaling factors are used for each equation, as we do in Sections 4 and 5. The remaining assumptions are adapted to take into account the multivariate nature of the problem being considered here. We note that Assumptions 4 and 11 are stronger than the corresponding assumptions in Kemp and Santos Silva (2012) in that they impose conditions on the moments of Y_t . These stronger conditions are needed to deal with the stochastic bandwidth, but they would also be implicitly imposed in the case where Z_t includes lagged values of Y_t , something that was not considered by Kemp and Santos Silva (2012).

The following theorem establishes the existence of \hat{A}_T , the estimator of interest.

Theorem 1 (Existence) *Under Assumptions 1, 2, 6, 7 and 8, there exists a measurable random variable \hat{A}_T such that:*

$$\begin{aligned} \Pr\left(\hat{A}_T \in \mathcal{A}\right) &= 1, \\ \Pr\left(\hat{Q}_T\left(\hat{A}_T\right) \geq \hat{Q}_T(A), \quad \forall A \in \mathcal{A}\right) &= 1. \end{aligned}$$

The consistency of \hat{A}_T is established by the following theorem.

Theorem 2 (Consistency) *Under Assumptions 1–8, \widehat{A}_T converges in probability to A_0 .*

We next establish the asymptotic normality of the estimator and its rate of convergence.

Theorem 3 (Asymptotic Normality) *Under Assumptions 1–14:*

$$\left(T\widehat{\delta}_T^{G+2}\right)^{1/2} (\widehat{\alpha}_T - \alpha_0) \xrightarrow{d} \mathcal{N} [0, D_0^{-1} B_0 D_0^{-1}],$$

where $\widehat{\alpha}_T = \text{vec} \left(\widehat{A}_T \right)$, $\alpha_0 = \text{vec} (A_0)$, and:

$$\begin{aligned} B_0 &= \text{E} \left[f_t (0 | \mathcal{F}_{t-1}) (Z_t \otimes I_G) \mathcal{M} (Z_t \otimes I_G)' \right], \\ D_0 &= \text{E} \left[(Z_t \otimes I_G) f_t^{(2)} (0 | \mathcal{F}_{t-1}) (Z_t \otimes I_G)' \right], \\ \mathcal{M} &= \int_{\mathbb{R}^G} \mathcal{K}^{(1)} (x) \mathcal{K}^{(1)} (x)' dx. \end{aligned}$$

Allowing $\widehat{\delta}_T$ to vanish suitably slowly, it follows from Theorem 3 that \widehat{A}_T converges at a rate that can be made arbitrarily close to $T^{\frac{2}{6+G}}$. Therefore, the estimator is affected by a form of the “curse of dimensionality” in that its rate of convergence goes down when G increases. This, of course, is a consequence of the fact that non-parametric density estimation is less “local” in high dimensions, i.e., larger bandwidths have to be used when the dimension of the problem increases (see Assumption 13). Note that, as in Lee (1989, 1993), it is also possible to consider an estimator with a fixed bandwidth; under suitably strong regularity conditions of the type considered by Lee (1989, 1993) such an estimator is \sqrt{T} -consistent for the parameters of the conditional mode.

Finally, the next theorem establishes the consistency of the usual “sandwich” covariance matrix estimator.

Theorem 4 (Consistent Asymptotic Covariance Estimation) *Under Assumptions 1–14:*

$$\widehat{\Sigma}_T = \widehat{\text{Avar}} (\widehat{\alpha}_T) = \widehat{D}_T^{-1} \widehat{B}_T \widehat{D}_T^{-1} \xrightarrow{p} \Sigma_0 = \text{Avar} (\widehat{\alpha}_T) = D_0^{-1} B_0 D_0^{-1},$$

where B_0 and D_0 are given as in Theorem 3 and:

$$\begin{aligned} \widehat{D}_T &= T^{-1} \sum_{t=1}^T \widehat{\delta}_T^{-(G+2)} (Z_t \otimes I_G) \mathcal{K}^{(2)} \left(\frac{Y_t - \widehat{A}_T Z_t}{\widehat{\delta}_T} \right) (Z_t \otimes I_G)' \xrightarrow{p} D_0, \\ \widehat{B}_T &= T^{-1} \sum_{t=1}^T \widehat{\delta}_T^{-G} (Z_t \otimes I_G) \mathcal{K}^{(1)} \left(\frac{Y_t - \widehat{A}_T Z_t}{\widehat{\delta}_T} \right) \mathcal{K}^{(1)} \left(\frac{Y_t - \widehat{A}_T Z_t}{\widehat{\delta}_T} \right)' (Z_t \otimes I_G)' \xrightarrow{p} B_0. \end{aligned}$$

3. SYSTEMS OF LINEAR SIMULTANEOUS EQUATIONS

In this section we discuss how our earlier results can be used in the context of systems of linear simultaneous equations. In particular, we consider standard simultaneous equation systems of the form

$$Y_t' \Gamma_0 + Z_t' \Psi_0 = V_t', \quad t = 1, \dots, T, \quad (7)$$

where Γ_0 and Ψ_0 are, respectively, $G \times G$ and $K \times G$ matrices of unknown structural parameters, V_t is a $G \times 1$ random vector such that $\text{Mode}(V_t | \mathcal{F}_{t-1}) = 0$, and Y_t, Z_t are defined as before. Additionally, we assume that Γ_0 is non-singular and note that (7) can represent either a classic system of simultaneous equations (Haavelmo, 1943) or a structural vector autoregressive model (Bernanke, 1986).

The method developed in the previous section cannot generally be used to directly estimate (7) because of the evident simultaneity. However, it is possible to show that our earlier results can be used to estimate the reduced form of the model, which is given by

$$Y_t' = Z_t' A_0' + U_t', \quad (8)$$

with $A_0' = -\Psi_0 \Gamma_0^{-1}$ and $U_t' = V_t' \Gamma_0^{-1}$. To see this, let U_t and V_t be two random vectors such that $U_t = \Upsilon V_t$, where Υ is a non-singular matrix, and let $f_{U_t}(u_t | \mathcal{F}_{t-1})$ and $f_{V_t}(v_t | \mathcal{F}_{t-1})$ denote the conditional density functions of U_t and V_t , respectively. Note that because $f_{U_t}(u_t | \mathcal{F}_{t-1}) = f_{V_t}(\Upsilon^{-1} u_t | \mathcal{F}_{t-1}) / |\det(\Upsilon)|$, we have that if $\text{Mode}(V_t | \mathcal{F}_{t-1}) = 0$, then $f_{U_t}(u_t | \mathcal{F}_{t-1}) = f_{V_t}(\Upsilon^{-1} u_t | \mathcal{F}_{t-1}) / |\det(\Upsilon)| \leq f_{V_t}(0 | \mathcal{F}_{t-1}) / |\det(\Upsilon)| = f_{U_t}(0 | \mathcal{F}_{t-1})$, and therefore $\text{Mode}(U_t | \mathcal{F}_{t-1}) = 0$. Uniqueness of the conditional mode of U_t follows from the fact that Υ is non-singular. Therefore, $\text{Mode}(V_t | \mathcal{F}_{t-1}) = 0$ implies that $\text{Mode}(U_t | \mathcal{F}_{t-1}) = 0$ and hence (8) is just the transpose of a system of the form of (1) and can be estimated in a similar fashion. However, typically economists are not interested in learning about A_0 and therefore it is interesting to study the conditions under which it is possible to identify Γ_0 and Ψ_0 .

Identification of the structural parameters in Γ_0 and Ψ_0 requires the researcher to be able to impose enough restrictions on (7); these can involve only the elements of Γ_0 and Ψ_0 , or also restrictions on the conditional distribution of V_t ; we consider the two cases separately.

3.1. Classic system of simultaneous equations

In the context of classic systems of simultaneous equations, it is commonly assumed that restrictions on Γ_0 and Ψ_0 are enough to ensure that the whole system is identified; Richmond (1974) provides a necessary and sufficient condition for system identification based on linear restrictions on Γ_0 and Ψ_0 .

Let $\beta_0 = (\text{vec}(\Gamma_0)', \text{vec}(\Psi_0)')$ and notice that the equality $A'_0 = -\Psi_0\Gamma_0^{-1}$ implies $A'_0\Gamma_0 + \Psi_0 = 0$, which can be vectorized as

$$(I_G \otimes A'_0, I_{GK})\beta_0 = 0.$$

Furthermore, assume that Γ_0 and Ψ_0 satisfy the additional set of m linear restrictions

$$\Phi\beta_0 = \varphi,$$

where Φ is a $m \times G(G+K)$ matrix and φ is a m -dimensional vector. Richmond (1974, Theorem 5) shows that the system is identified if and only if

$$\text{rank}((I_G \otimes A'_0, I_{GK})', \Phi') = G(G+K). \quad (9)$$

Note that condition (9) implies that $m \geq \rho \geq G^2$, where $\rho \equiv \text{rank}(\Phi)$, and the parametric restrictions $\Phi\beta_0 = \varphi$ imply a partition of β_0 into two subvectors β_0^r and $\bar{\beta}_0^r$ such that $\bar{\beta}_0^r = \Phi_r\beta_0^r + \varphi_r$, where Φ_r is a $\rho \times (G^2 + GK - \rho)$ matrix and φ_r , β_0^r , and $\bar{\beta}_0^r$ are vectors of dimensions ρ , $G(G+K) - \rho$, and ρ , respectively. Furthermore, imposing the restriction $\bar{\beta}^r = \Phi_r\beta^r + \varphi_r$ on Ψ and Γ we obtain Ψ_r and Γ_r .

For identified models, we estimate β_0^r and estimates of the remaining parameters of β_0 are obtained via the equation $\bar{\beta}_0^r = \Phi_r\beta_0^r + \varphi_r$. The estimator of β_0^r can be implemented using the following two-stage procedure. First, obtain \hat{A}_T , \hat{B}_T , and \hat{D}_T by estimating the transpose of (8) using the multivariate conditional mode estimator defined by (5). Second, estimate β_0^r by solving the following minimum distance problem:

$$\hat{\beta}_T^r = \arg \min_{\beta^r \in \mathcal{B}^r} \left[\hat{\alpha}_T + \text{vec} \left((\Psi_r \Gamma_r^{-1})' \right) \right]' \hat{\Sigma}_T^{-1} \left[\hat{\alpha}_T + \text{vec} \left((\Psi_r \Gamma_r^{-1})' \right) \right], \quad (10)$$

where $\hat{\Sigma}_T = \widehat{\text{Avar}}(\hat{\alpha}_T) = \hat{D}_T^{-1} \hat{B}_T \hat{D}_T^{-1}$, as in Theorem 4, and \mathcal{B}^r denotes the parameter space of β^r . Notice that when the system is exactly identified the minimum distance

estimator is not needed and estimates of the structural parameters can be obtained just by solving the system $\widehat{a}_T + \text{vec}((\Psi_r \Gamma_r^{-1})') = 0$ for Γ_r and Ψ_r .

The asymptotic properties of this two-stage estimator are closely related to those of \widehat{A}_T . To establish these properties we need the following additional assumptions where we use the definitions: $C(\beta^r) = \partial \text{vec}(\Psi_r \Gamma_r^{-1}) / \partial \beta^{r'}$ and $C_0 = C(\beta_0^r)$.

15. (*Identification*) The matrices Γ_0 , A_0 , and Φ are such that: (i) $\text{rank}(\Gamma_0) = G$; (ii) $\text{rank}((I_G \otimes A'_0, I_{GK})', \Phi') = G(G + K)$.
16. (*Parameter Space - II*) \mathcal{B}^r is compact.
17. (*Rank Condition*) $\text{rank}(C_0) = G(G + K) - \rho$.
18. (*Interior Parameter Value - II*) \mathcal{B}^r has a non-empty interior, denoted $\text{int}(\mathcal{B}^r)$, and $\beta_0^r \in \text{int}(\mathcal{B}^r)$.

The following result establishes the consistency of the proposed procedure.

Theorem 5 (Consistency II) *Under Assumptions 1–7, 15 and 16: $\widehat{\beta}_T^r \xrightarrow{p} \beta_0^r$.*

Then, Theorems 1–3 imply the following results.

Theorem 6 (Asymptotic Normality II) *Under Assumptions 1–18:*

$$\sqrt{T \delta_T^{G+2}} (\widehat{\beta}_T^r - \beta_0^r) \xrightarrow{d} \mathcal{N} \left(0, [C'_0 D_0 B_0^{-1} D_0 C_0]^{-1} \right).$$

Theorem 7 (Consistent Asymptotic Covariance Estimation II) *Under Assumptions 1–18:*

$$\widehat{\text{Avar}}(\widehat{\beta}_T^r) = [\widehat{C}'_T \widehat{D}_T \widehat{B}_T^{-1} \widehat{D}_T \widehat{C}_T]^{-1} \xrightarrow{p} \text{Avar}(\widehat{\beta}_T^r) = [C'_0 D_0 B_0^{-1} D_0 C_0]^{-1},$$

where \widehat{B}_T and \widehat{D}_T are given as in Theorem 4 and $\widehat{C}_T = C(\widehat{\beta}_T^r)$.

3.2. Structural vector autoregressive models

There are models in which the available restrictions on Γ_0 and Ψ_0 are not enough to ensure that Assumption 15 holds, but identification can be obtained by imposing restrictions on the conditional distribution of V_t . For example, assumptions on the conditional

distribution of V_t are heavily used in the identification of structural vector autoregressive models because in this case restrictions on Ψ_0 are generally difficult to justify. In this context, it is often assumed that the conditional covariance matrix of V_t is diagonal (see, e.g., Lütkepohl, 2005), reflecting the fact that the structural errors are “primitive”, in the sense that they do not have common causes (Bernanke, 1986).

Naturally, restrictions on the conditional covariance of V_t do not help in the identification of (7) because the model does not impose any structure on the conditional moments of V_t . However, there are cases in which the stronger condition that the elements of V_t are conditionally independent can be used to identify Γ_0 and Ψ_0 . Strictly speaking the assumption that the elements of V_t are conditionally independent is much stronger than the assumption that they are conditionally uncorrelated. Nonetheless, conditional independence is very much in line with the idea that the structural errors are “primitive” and it is perhaps the most natural justification for the absence of conditional correlation. Moreover, the absence of conditional correlation is often coupled with the assumption of normally distributed errors (see, e.g., Lütkepohl, 2005), and together these assumptions imply conditional independence.

Estimation under conditional independence of the elements of V_t is particularly attractive because in this case the multivariate mode is just the vector of the marginal modes, and therefore it is possible to escape the curse of dimensionality by estimating each equation separately.

Estimation equation-by-equation of (7) under conditional independence may be possible by adapting Sargan’s (1958) approach to the estimation of models defined by conditional mode restrictions, much in the same way Sakata (2007) adapted it to the estimation of models defined by conditional median restrictions. The details of such method are, however, beyond the scope of the present paper. Nevertheless, our earlier results can easily be used in the leading case where the elements of V_t are assumed to be conditionally independent and Γ_0 is restricted to be a triangular matrix with ones on the main diagonal (this is the so-called Cholesky identification).

For example, assuming that Γ is lower triangular, (7) can be written as

$$y_{tg} = \sum_{j=g+1}^G -\gamma_{jg}y_{tj} + \sum_{k=1}^K \psi_{kg}z_{tk} + v_{tg}, \quad g = 1, \dots, G-1, \text{ \& } t = 1, \dots, T,$$

$$y_{tG} = \sum_{k=1}^K \psi_{kG}z_{tk} + v_{tG}, \quad t = 1, \dots, T,$$

where y_{ti} , z_{ti} , and v_{ti} denote the i th element of the vectors Y_t , Z_t , and V_t , and γ_{jg} and ψ_{kg} denote elements of the matrices Γ_0 and Ψ_0 .

By assumption, the mode of v_{tG} conditional on \mathcal{F}_{t-1} is zero and hence

$$\text{Mode}(y_{tG}|\mathcal{F}_{t-1}) = \sum_{k=1}^K \psi_{kG}z_{tk}. \quad (11)$$

In addition, by assumption, v_{tg} is conditionally independent of $(v_{tg+1}, \dots, v_{tG})$ given \mathcal{F}_{t-1} , with a conditional mode of 0. Hence it follows that

$$\text{Mode}(y_{tg}|\mathcal{F}_{t-1}) = \sum_{j=g+1}^G -\gamma_{jg}y_{tj} + \sum_{k=1}^K \psi_{kg}z_{tk}, \quad g = 1, \dots, G-1. \quad (12)$$

Equations (11) and (12) show that when the errors are independent and the Cholesky identification is used, it is possible to estimate each equation separately by using the univariate version of the estimator proposed in Section 2. We illustrate the use of this procedure in Subsection 5.3.

4. SIMULATION EVIDENCE

In this section we present the results of simulation experiments illustrating the finite sample performance of the proposed estimator. The core of these experiments is designed to shed light on how the performance of the estimator depends on the number of equations in the system and on the smoothing parameter. We also perform a small set of experiments illustrating the sensitivity of the estimators to the presence of an additive outlier.

In these experiments data for $t = -249, \dots, T$, with $T \in \{150, 300\}$, are generated by

$$y_{g,t} = a_{g0} + a_{g1}y_{1,t-1} + a_{g2}y_{2,t-1} + a_{g3}y_{3,t-1} + a_{g4}y_{4,t-1} + u_{g,t}, \quad g \in \{1, 2, 3, 4\},$$

with $y_{g,-250} = 0$, $a_{gg} = 0.75$, and $a_{gi} = 0$ for $g \neq i$. The errors are generated as $u_{g,t} = \epsilon_{g,t} (0.1 + 0.75u_{g,t-1}^2)^{0.5}$, with $u_{g,-250} = \epsilon_{g,-250}$, and therefore follow an ARCH-type

process (Engle, 1982). The variables $\epsilon_{g,t}$ are generated independently as draws from the log-normal distribution with parameters μ and σ , and are shifted and scaled to have zero mode and unit variance; we set $\mu = 0$ and, to generate errors with very different degrees of skewness, we run experiments with $\sigma \in \{0.3, 0.9\}$ (the skewness of the log-normal distribution goes to 0 with σ). The combination of heteroskedasticity and skewness implies that the conditional mean and conditional mode of $y_{g,t}$ can have different forms and therefore we can expect the estimates obtained by mode- and mean-based vector autoregressions to be different, especially in the high-skewness case with $\sigma = 0.9$ (note that this design is prone to generating innovation outliers).

Because the errors $\epsilon_{g,t}$ are independent across g , each equation can be estimated either by itself or as part of a system. To study the effect of G , we focus on the estimation of equation 1 and compare the results obtained when it is estimated as part of a system of dimension $G = 1, 2, 3, 4$. By doing this we can gain some insight into the costs of the curse of dimensionality incurred when using the system estimator proposed in Section 2.

The mode estimator was implemented using equation specific bandwidths with the smoothing parameter for equation g , denoted $\hat{\delta}_{g,T}$, defined as $\hat{\delta}_{g,T} = \mathbf{s} \text{MAD}_g T^{-l}$, where MAD_g denotes the median of the absolute deviation from the median least squares residual for equation g , and $l = 1.001 / (6 + G)$.

To complete the definition of $\hat{\delta}_{g,T}$ it is necessary to define the scaling factor \mathbf{s} . Using Silverman's (1986, p. 45, eq. 3.28) rule-of-thumb as a guide, and noting that for the normal distribution the standard deviation is approximately equal to 1.4826MAD , Kemp and Santos Silva (2012) used $\mathbf{s} = 1.6$ for the univariate case. In the multivariate case, however, it may be useful to do some oversmoothing to mitigate the effects of the curse of dimensionality. Therefore, to study the interplay between the effects of \mathbf{s} and G on the performance of the estimator, for each value of G we run experiments for $\mathbf{s} \in \{1.6, 3.2, 6.4, \infty\}$, with $\mathbf{s} = \infty$ corresponding to the mean-based estimator.

Table 1 contains the means and standard errors of the estimates of a_{10} and a_{11} obtained in 10000 replicas of simulation procedure; to conserve space we focus on these more interesting parameters. The results for the least squares estimator ($\mathbf{s} = \infty$) provide a benchmark against which we can compare the mode-based estimates and illustrate that

mode- and mean-based estimates can be quite different. The results obtained with the conditional mode estimators with $\mathbf{s} = 1.6$ are reminiscent of those reported by Kemp and Santos Silva (2012) in that the slope parameters are generally estimated with little bias but the biases of the intercepts are more noticeable; naturally the biases decrease as the samples grow. These results hold for all the values of G considered in these experiments.

The results with $\mathbf{s} = 1.6$ also clearly reveal the effects of the curse of dimensionality, with the precision of the estimates dropping quickly with G , especially when $\sigma = 0.9$. The effects of the curse of dimensionality, however, are much less noticeable with larger values of \mathbf{s} . In particular, for $\mathbf{s} = 6.4$ the standard error of the estimates is much less sensitive to the value of G . Increasing \mathbf{s} also generally increases the precision of the estimates for all values of G but, naturally, the larger bandwidths lead to larger biases, especially for the intercepts.

The results in Table 1 suggest that, if the interest is on the slope parameters, setting \mathbf{s} to a value around 6.4 is a sensible choice. However, if the estimate of the intercept is also relevant, for example if the objective is to use the model for prediction, setting \mathbf{s} to a smaller value may be advisable.

We next report the results of a smaller set of simulations illustrating the effect of an additive outlier on the results of mode- and mean-based vector autoregressions. In these experiments the data were generated exactly as before but then y_{11} is multiplied by 10 to generate an additive outlier; note that, unlike innovation outliers, this additive outlier is not generated by the assumed data generating process (the misplacement of the decimal point is often referred as a possible cause of outliers, see for example Rousseeuw and Leroy, 1987). Table 2 reports the mean and standard errors of the estimates of a_{10} and a_{11} obtained in 10000 replicas of simulation procedure for the cases with $G \in \{1, 4\}$ and $\mathbf{s} \in \{1.6, 6.4, \infty\}$. Comparing the results in Table 2 with the corresponding results in Table 1, it is clear that the estimates obtained with the mean-based vector autoregressions ($\mathbf{s} = \infty$) are severely affected by the presence of a single additive outlier even when $T = 300$; this is clear both in the mean and standard error of the estimates. In contrast, and as expected, the results obtained with the mode-based estimators are almost unaffected by the presence of the additive outlier.

Overall, these results are encouraging in that they suggest that the proposed mode estimator is likely to have a reasonable performance in moderately large samples. Moreover, these results also suggest that with a careful choice of the bandwidth is possible to mitigate the costs of the curse of dimensionality, at least for systems with a moderate number of equations. Finally, the simulation results illustrate the expected robustness of the mode-based estimator to the presence of outliers. The findings of these simulations will inform the choices made in the next section where we consider the use of mode-based vector autoregressions in practice.

Table 1: Simulation results

			$G = 1$		$G = 2$		$G = 3$		$G = 4$	
σ	T	s	Const.	$y_{1,t-1}$	Const.	$y_{1,t-1}$	Const.	$y_{1,t-1}$	Const.	$y_{1,t-1}$
0.3	150	∞	0.197 (0.147)	0.761 (0.096)	0.197 (0.147)	0.761 (0.096)	0.197 (0.147)	0.761 (0.096)	0.197 (0.147)	0.761 (0.096)
		6.4	0.150 (0.086)	0.728 (0.076)	0.155 (0.089)	0.729 (0.078)	0.158 (0.091)	0.730 (0.080)	0.161 (0.093)	0.731 (0.082)
		3.2	0.109 (0.088)	0.723 (0.079)	0.116 (0.094)	0.722 (0.084)	0.121 (0.100)	0.722 (0.090)	0.126 (0.105)	0.722 (0.094)
		1.6	0.064 (0.121)	0.723 (0.107)	0.071 (0.147)	0.721 (0.125)	0.078 (0.171)	0.721 (0.147)	0.087 (0.208)	0.718 (0.171)
	300	∞	0.176 (0.098)	0.789 (0.074)	0.176 (0.098)	0.789 (0.074)	0.176 (0.098)	0.789 (0.074)	0.176 (0.098)	0.789 (0.074)
		6.4	0.134 (0.055)	0.743 (0.049)	0.139 (0.057)	0.745 (0.051)	0.142 (0.058)	0.747 (0.053)	0.145 (0.060)	0.748 (0.054)
		3.2	0.092 (0.057)	0.738 (0.053)	0.099 (0.061)	0.738 (0.056)	0.104 (0.065)	0.739 (0.060)	0.108 (0.068)	0.739 (0.063)
		1.6	0.047 (0.083)	0.739 (0.077)	0.054 (0.098)	0.739 (0.091)	0.058 (0.117)	0.739 (0.105)	0.061 (0.139)	0.740 (0.122)
0.9	150	∞	0.196 (0.254)	0.825 (0.069)	0.196 (0.254)	0.825 (0.069)	0.196 (0.254)	0.825 (0.069)	0.196 (0.254)	0.825 (0.069)
		6.4	0.145 (0.064)	0.761 (0.035)	0.149 (0.067)	0.763 (0.037)	0.153 (0.071)	0.764 (0.039)	0.155 (0.074)	0.766 (0.041)
		3.2	0.108 (0.054)	0.752 (0.031)	0.113 (0.060)	0.753 (0.034)	0.116 (0.066)	0.754 (0.038)	0.119 (0.072)	0.755 (0.042)
		1.6	0.066 (0.053)	0.750 (0.034)	0.068 (0.066)	0.751 (0.042)	0.070 (0.082)	0.753 (0.052)	0.072 (0.102)	0.754 (0.064)
	300	∞	0.175 (0.142)	0.849 (0.049)	0.175 (0.142)	0.849 (0.049)	0.175 (0.142)	0.849 (0.049)	0.175 (0.142)	0.849 (0.049)
		6.4	0.136 (0.039)	0.761 (0.022)	0.141 (0.042)	0.763 (0.024)	0.145 (0.044)	0.765 (0.026)	0.147 (0.046)	0.767 (0.027)
		3.2	0.099 (0.033)	0.753 (0.019)	0.104 (0.037)	0.754 (0.022)	0.108 (0.041)	0.755 (0.024)	0.111 (0.046)	0.756 (0.027)
		1.6	0.058 (0.032)	0.751 (0.021)	0.062 (0.041)	0.752 (0.027)	0.065 (0.052)	0.753 (0.033)	0.066 (0.064)	0.754 (0.041)

Table 2: Simulation results with an additive outlier

		$G = 1$		$G = 4$		
σ	T	s	Const.	$y_{1,t-1}$	Const.	$y_{1,t-1}$
0.3	150	∞	0.400 (0.351)	0.554 (0.222)	0.400 (0.351)	0.554 (0.222)
		6.4	0.159 (0.092)	0.724 (0.084)	0.170 (0.102)	0.725 (0.093)
		1.6	0.073 (0.116)	0.720 (0.106)	0.101 (0.203)	0.707 (0.177)
	300	∞	0.301 (0.213)	0.656 (0.178)	0.301 (0.213)	0.656 (0.178)
		6.4	0.137 (0.056)	0.742 (0.052)	0.148 (0.061)	0.747 (0.057)
		1.6	0.051 (0.080)	0.739 (0.074)	0.065 (0.137)	0.737 (0.120)
0.9	150	∞	0.572 (0.925)	0.503 (0.240)	0.572 (0.925)	0.503 (0.240)
		6.4	0.162 (0.090)	0.766 (0.051)	0.172 (0.111)	0.770 (0.065)
		1.6	0.089 (0.066)	0.750 (0.045)	0.097 (0.116)	0.751 (0.081)
	300	∞	0.411 (0.319)	0.651 (0.198)	0.411 (0.319)	0.651 (0.198)
		6.4	0.144 (0.043)	0.765 (0.026)	0.155 (0.052)	0.770 (0.033)
		1.6	0.069 (0.034)	0.751 (0.020)	0.076 (0.066)	0.755 (0.041)

5. PUTTING THE MODE-BASED VAR THROUGH ITS PACES

In this section we illustrate the usefulness of multivariate mode auto-regression models in a context similar to that considered by Stock and Watson (2001) in their classic paper on vector autoregressions (VARs). Specifically, we compare mode- and mean-based VARs in a three-variable model for inflation (π), unemployment rate (u), and interest rate (R), estimated with US quarterly data from 1960:I to 2000:IV, and then compute pseudo out-of-sample forecasts from 2001:I to 2017:IV. Other examples of VAR models using quarterly data for inflation, unemployment, and interest rate include Cogley and Sargent (2001), Primiceri (2005), and Koop, Leon-Gonzalez, and Strachan (2009). All data were obtained from the St. Louis Fed’s Federal Reserve Economic Data (FRED) website. Specifically, the series used are CPIAUCNS (Consumer Price Index for All Urban Consumers: All Items, Monthly, Not Seasonally Adjusted), UNRATENSA (Civilian Unemployment Rate, Percent, Monthly, Not Seasonally Adjusted), and FEDFUNDS (Ef-

fective Federal Funds Rate, Percent, Monthly, Not Seasonally Adjusted). Quarterly values of all variables are obtained by averaging monthly values and monthly inflation is obtained as $100 \ln(P_t/P_{t-12})$ where P_t is CPIAUCNS.

5.1. Dynamics and Granger causality

Following Stock and Watson (2001), we start by comparing the results for Granger-causality tests based on mode- and mean-based VARs. As in Stock and Watson (2001), we will only use four-lag VARs and will not discuss in detail the choice of the lag length (see Lütkepohl, 2005, for a discussion of this topic).

A related issue that we do not discuss is the use of shrinkage methods to reduce possible overparameterization which can have a negative effect on the performance of VAR models, especially when used for forecasting (see, e.g., Canova, 1999). A popular approach to overcome this problem is to use Bayesian methods (see, e.g., Litterman, 1986, and Kilian and Lütkepohl, 2017, Ch. 5). Using Bayesian methods in the context of the proposed mode regression estimator, however, is not feasible because those methods are likelihood based and our estimator is non-parametric about the distribution of the errors. Alternatively, one could choose the lag length so as to optimize forecasting performance (see Canova, 1999) or we could perform the shrinkage using a form of LASSO (see, e.g., Hastie, Tibshirani and Friedman, 2009). Studying and implementing either of these approaches, however, is beyond the scope of this paper and we leave it for future research.

Table 3 presents the p-values for the Granger-causality tests for the mean-based VAR ($\mathbf{s} = \infty$) and for the mode-based VARs obtained with different values of \mathbf{s} . As expected, the results for the mean-based model are not very different from those obtained with the mode-based model when \mathbf{s} is large. However, for $\mathbf{s} = 1.6$ the results become substantially different, suggesting that mode- and mean-based models can have very different dynamic structures (the difference between the two models will depend on how the shape of the distribution of the errors varies with the regressors). For example, the mean-based model provides strong evidence that u does not Granger-cause π , whereas with the mode-based model with $\mathbf{s} = 1.6$ that null hypothesis is rejected at the usual 5% level. Of course,

the strong rejection of the null of no Granger-causality on mode- and mean-based models can also hide significantly different dynamic structures, leading to substantially different forecasts; we investigate that in the next subsection.

Table 3: P-values for Granger-causality tests (1960:I – 2000:IV)

Regressor	Mean			Mode								
	$s = \infty$			$s = 6.4$			$s = 3.2$			$s = 1.6$		
	π	u	R	π	u	R	π	u	R	π	u	R
π	—	0.00	0.02	—	0.00	0.06	—	0.00	0.06	—	0.00	0.01
u	0.76	—	0.23	0.90	—	0.57	0.64	—	0.83	0.01	—	0.23
R	0.01	0.00	—	0.00	0.00	—	0.00	0.00	—	0.00	0.00	—

5.2. Multiperiod Forecasts

The traditional literature on VAR models has considered iterated and direct multiperiod forecasts (see, e.g., Stock and Watson, 2012, pp. 678-684 for a textbook treatment of these approaches). If the VAR is correctly specified, iterated multiperiod forecasts are known to be more efficient, but direct forecasts are more robust and therefore the choice of method to use depends on the particular application at hand (see, e.g., Bhansali, 2002, and Stock and Watson, 2012). Iterated multiperiod forecasts, however, rely on the properties of the expectation operator and therefore this approach is not available unless the VAR characterizes the conditional mean. On the contrary, direct multiperiod forecasts do not depend on the properties of expectations and therefore can be used in our context.

Suppose that the purpose is to use the information available at time $t - 1$ to obtain a forecast of Y_{t-1+h} , with $h > 0$. The direct forecast can be obtained from estimates of the model

$$Y_t = A_0^h Z_{t+1-h} + U_t, \quad (13)$$

where Z_{t+1-h} can contain Y_{t-h} , its lags, and exogenous variables. Letting \hat{A}_0^h denote an estimate of A_0^h , the forecast of Y_{t-1+h} can be obtained as $\hat{Y}_{t-1+h} = \hat{A}_0^h Z_t$.

Naturally, the properties of the forecast will depend on the properties of the estimator of A_0^h and on the properties of U_t . Letting \mathcal{F}_{t-h} denote the σ -algebra generated by $\{(Y_{t-j}, Z_{t+1-j})\}_{j=h}^{\infty}$, the traditional direct multiperiod forecasts are obtained if the elements of A_0^h are estimated by least squares under the assumption that $E(U_t | \mathcal{F}_{t-h}) = 0$.

Likewise, we obtain a mode-based predictor of Y_{t-1+h} when A_0^h is estimated by the method proposed in Section 2 under the assumption that $\text{Mode}(U_t|\mathcal{F}_{t-h}) = 0$.

Using (13) to generate direct forecasts for $h = 1, \dots, H$, implies assuming that Y_t has a VAR representation for each h . This is a strong assumption that is unlikely to hold in practice (more generally, it is widely accepted that VARs are often misspecified; see, e.g. the discussion in Jordà, 2005). The assumption that the models are correctly specified is particularly unlikely to hold in the pseudo out-of-sample forecasts that we perform below because our models were used to make predictions during the recent global financial crisis and ensuing recession; likewise, at some point all models were estimated using data from this exceptional period. This suggests that the models we use may suffer from some degree of misspecification due to structural changes, but this is also often unavoidable in practice (e.g., Clements and Hendry, 1998, suggest that structural change is a major source of forecasting errors). However, autoregressive models estimated using an expanding window that spans structural breaks, as we use, have been shown to produce reasonably accurate forecasts (see Pesaran and Timmermann, 2005). More importantly, the major economic events of the early 21st century provide us with the opportunity to study the prediction ability of the different models in very different, and rapidly changing, economic scenarios. We note that the asymptotic theory we develop in this paper will not apply if the mode-based model is not correctly specified, but studying the asymptotic properties of misspecified mode regression is clearly beyond the scope of this paper.

In order to be able to compare the forecasts produced by the different models it is necessary to define a suitable loss function. Although many alternatives are available (see, e.g., Komunjer and Owyang, 2012, and Sinclair, Stekler, and Muller-Droge, 2016), the loss function we use here is simply the Euclidean distance between the forecast and the actual realization, defined as

$$d_{t+h} = (e_{1,t+h}^2 + e_{2,t+h}^2 + e_{3,t+h}^2)^{1/2},$$

where $e_{g,t+h}$ is the h -step-ahead forecasting error for equation g estimated with data up to t . The motivation for our choice of loss function is the intuitive interpretation of this measure, which is likely to make it attractive to non-experts (of course, it may be

possible to define an estimator of the multivariate measure of location that minimizes the expected value of this loss function but that is not our objective here). We could also consider a more general loss function where the forecasting error of each equation is given a different weight. This is particularly interesting if the variables are measured on different scales or have very different dispersion, which is not the case in the example we consider. Obviously, the ranking of the different forecasts will depend on the choice of loss function (see, e.g., Sinclair, Stekler, and Muller-Droge, 2016) but that is not particularly important in our context because, rather than ranking the different models, we mainly want to shed light on the different nature of the mode- and mean-based predictions.

We now go back to our example and run a simulation exercise similar to that performed by Stock and Watson (2001). In particular, mode- and mean-based models were estimated (with an expanding window) and used to produce forecasts for h periods ahead for each of the 68 observations between 2001:I and 2017:IV. We then computed the values of d_{t+h} , with $h \in \{1, 2, 4, 8\}$, for the iterated and direct forecasts obtained with the mean-based models, and for the direct forecasts produced by the mode-based models with $\mathbf{s} \in \{1.6, 3.2, 6.4\}$; Table 4 displays descriptive statistics of d_{t+h} for each of the models.

In this particular example, the iterated forecasts and the direct forecasts obtained with the mean-based models have comparable performance, except for $h = 8$ where the iterated forecasts are clearly superior to the mean-based direct forecasts. Comparing now mode- and mean-based forecasts, we see that the performance of the mode-based forecasts is very good for all horizons. Indeed, for $\mathbf{s} \in \{3.2, 6.4\}$ the mean of d_{t+h} is lower for the mode regressions than for any of the mean-based forecasts. The mode-based forecasts, however, tend to have larger standard errors, but the mode-based forecast with $\mathbf{s} = 3.2$ has lower standard error than the iterated mean-based forecasts for all cases with $h > 1$. Finally, all the quartiles of d_{t+h} tend to be closer to zero for the mode-based forecasts, suggesting that the distribution of d_{t+h} tends to have more mass close to zero than the mean-based forecasts.

The difference between the distributions of d_{t+h} for mode- and mean-based models is clearly illustrated by the top panel in Figure 1, which displays an estimate of the density of d_{t+h} for $h = 4$ for the mode-based forecasts with $\mathbf{s} = 3.2$ (solid line) and for the mean-

based forecasts (dashed line). The bottom panel of Figure 1 plots the values of d_{t+4} for the mode-based forecasts with $s = 3.2$ (solid line) and for mean-based forecasts (dashed line). This plot shows that both methods have similar values of d_{t+4} up to the financial crisis, but subsequently the mode-based model performed substantially better.

To complement this information, Figure 2 displays for each of the series being considered the four-period ahead forecasting errors ($e_{g,t+4}$) for the mode-based forecasts with $s = 3.2$ (solid line) and for the mean-based forecasts (dashed line). The individual mode-based forecasting errors need to be read with caution because what is being predicted is the mode of the joint distribution and not the marginal modes, but these plots reinforce the idea that the mode-based predictions are particularly good after the financial crisis. Figures 1 and 2 also show that the mode-based model leads to some of the largest forecasting errors

Table 4: Descriptive statistics for d_{t+h} (2001:I – 2017:IV)

				Quartiles		
	Method	Mean	SError	1st	2nd	3rd
$h = 1$	Mean	0.954	0.755	0.501	0.713	1.059
	Mode $s = 1.6$	0.953	0.821	0.421	0.658	1.187
	Mode $s = 3.2$	0.896	0.801	0.362	0.656	1.093
	Mode $s = 6.4$	0.896	0.797	0.404	0.662	1.091
$h = 2$	Mean (Iterated)	1.631	1.244	0.784	1.290	2.131
	Mean (Direct)	1.569	1.173	0.751	1.191	2.014
	Mode $s = 1.6$	1.518	1.324	0.627	1.136	2.102
	Mode $s = 3.2$	1.383	1.065	0.651	1.072	1.869
	Mode $s = 6.4$	1.518	1.191	0.612	1.206	1.944
$h = 4$	Mean (Iterated)	2.563	1.774	1.278	2.159	3.462
	Mean (Direct)	2.574	1.607	1.437	2.374	3.449
	Mode $s = 1.6$	2.644	2.296	1.091	1.764	3.653
	Mode $s = 3.2$	2.357	1.745	1.106	1.899	3.019
	Mode $s = 6.4$	2.443	1.592	1.258	2.298	3.308
$h = 8$	Mean (Iterated)	3.881	2.075	2.176	3.738	5.207
	Mean (Direct)	4.269	2.288	2.384	3.820	6.249
	Mode $s = 1.6$	3.701	2.075	1.844	3.652	4.756
	Mode $s = 3.2$	3.614	1.901	1.978	3.745	4.598
	Mode $s = 6.4$	3.742	2.108	1.758	3.636	5.010

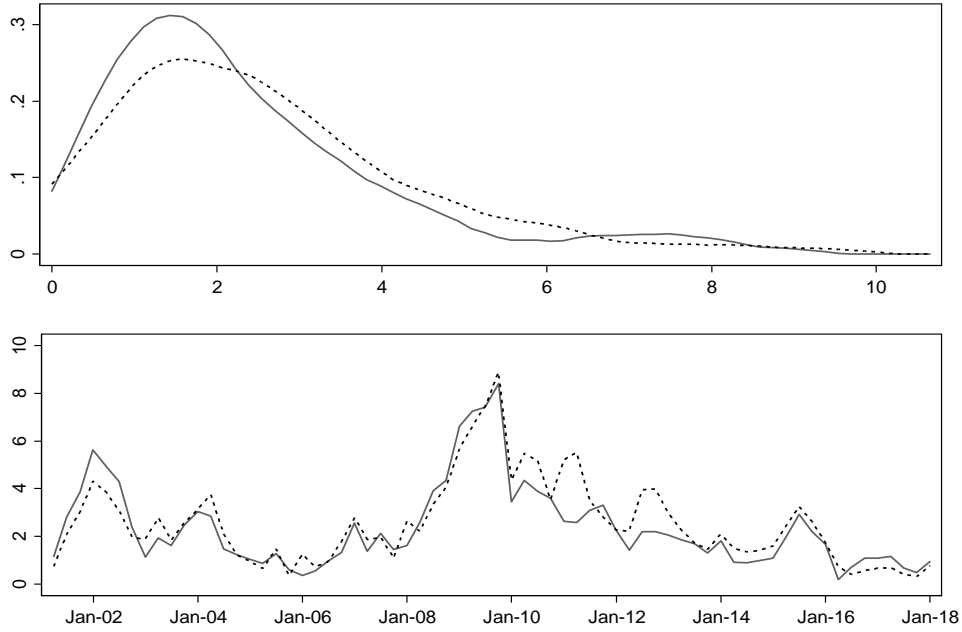


Figure 1: Top panel: estimates of the density of d_{t+4} for the mode-based forecasts with $s = 3.2$ (solid line) and for the mean-based forecasts (dashed line). Bottom panel: values of d_{t+4} for the mode-based forecasts with $s = 3.2$ (solid line) and for the mean-based forecasts (dashed line).

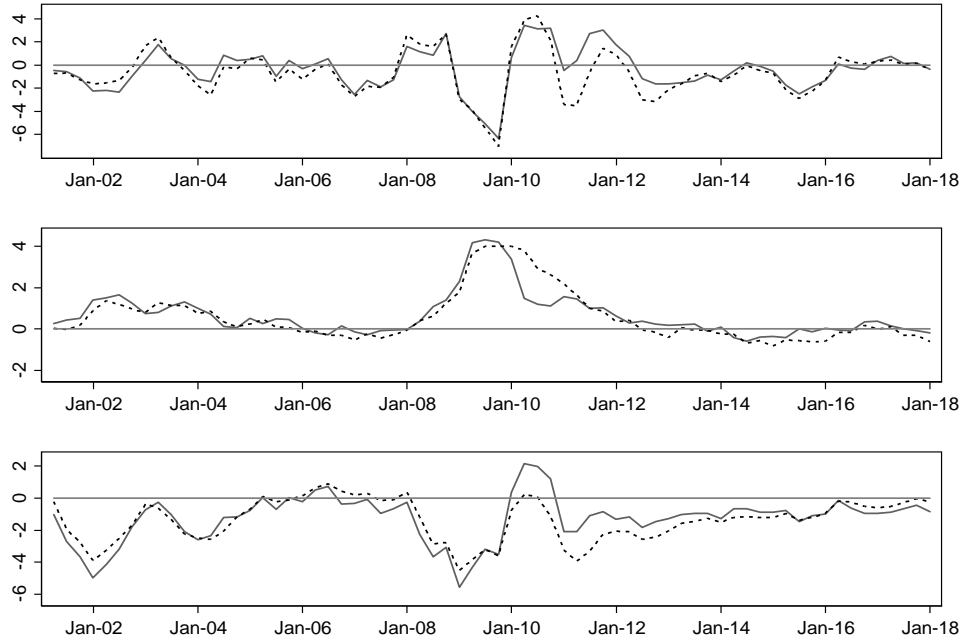


Figure 2: Four-period ahead forecast errors for inflation (top), unemployment rate (middle), and interest rate (bottom) for the mode-based forecasts with $s = 3.2$ (solid line) and for the mean-based forecasts (dashed line).

(this is reflected by their large standard errors reported in Table 4). This is not surprising because the loss function on which this estimator is based does not particularly penalize very large errors. That is, the existence of some large errors is the price to pay for having more errors close to zero, as illustrated by the top panel in Figure 1.

5.3. Impulse response functions

Impulse response functions (IRFs) are routinely used for policy analysis and their construction is closely related to the forecasting problem discussed before. In particular, just like with iterated multiperiod forecasts, the traditional IRFs obtained by inverting the VAR are difficult to interpret unless the VAR represents a conditional expectation (see Koop, Pesaran, and Potter, 1996, for other limitations of the traditional IRFs and see White, Kim, and Manganeli, 2015, for the analog of an IRF in the context of quantile-based models). However, as argued by Jordà (2005), models such as (13) provide an alternative way of obtaining IRFs. Jordà’s (2005) approach extends naturally to the case where the VAR defines a conditional mode.

To proceed it is convenient to write (13) as

$$Y_t = A_{10}^h Y_{t-h} + A_{20}^h Z_{t+1-h}^* + U_t = M(Y_t | Y_{t-h}, Z_{t+1-h}^*) + U_t, \quad (14)$$

where the role of Y_{t-h} is now explicit and Z_{t+1-h}^* contains the other variables in the model; that is $A_0^h = \begin{bmatrix} A_{10}^h & A_{20}^h \end{bmatrix}$ and $Z_{t-h+1} = \begin{bmatrix} Y_{t-h}' & Z_{t+1-h}^{*'} \end{bmatrix}'$. If (14) is estimated under the invalid assumption that $\text{Mode}(U_t | \mathcal{F}_{t-h}) = 0$, we can interpret $M(Y_t | Y_{t-h}, Z_{t+1-h}^*)$ as a “local projection” that approximates $\text{Mode}(Y_t | \mathcal{F}_{t-h})$, and we can define the Mode Impulse Response Function (MIRF) as

$$\text{MIRF}(h, t, \delta) = M(Y_t | Y_{t-h} + \delta, Z_{t+1-h}^*) - M(Y_t | Y_{t-h}, Z_{t+1-h}^*) = A_{10}^h \delta, \quad h = 0, 1, \dots \quad (15)$$

with the normalization $A_{10}^0 = I$ and $A_{20}^0 = 0$. That is, A_{10}^h gives the effect on $M(Y_t | Y_{t-h}, Z_{t+1-h}^*)$ of a shock δ to Y_{t-h} (or to U_{t-h}).

As in Jordà (2005), the MIRF defined by (15) gives the response to a shock to the reduced form errors. However, economists are often more interested in the responses to shocks to the structural errors. To see how those can be obtained, recall that from (8)

we have that $U_t = (\Gamma_0^{-1})' V_t$, where V_t denotes the vector of structural errors. Therefore, the effect of a shock to the structural error V_{t-h} is given by $A_{10}^h (\Gamma_0^{-1})'$ (this mimics the approach used by Jordà, 2005, see Kilian and Kim, 2011, footnote 4).

Of course, whether or not it is possible to compute the structural MIRF depends on whether Γ_0 is identified. In this application we use the standard Cholesky identification and assume that the structural model is a recursive VAR with the order π , u , R (as in Stock and Watson, 2001), and that the errors of the equations are independent. As discussed in Section 3, such model can be estimated using the system estimator defined in Section 2, or it can be estimated equation by equation, which has the advantage of avoiding the curse of dimensionality; this is the approach we use in this application.

IRFs are generally accompanied by the corresponding confidence intervals. Although asymptotic theory can be used to compute these, it is well-known that asymptotic approximations do not work well for the sample sizes of most applications, and therefore simulation-based methods are often preferred (see Kilian and Kim, 2011, and Stock and Watson, 2018, footnote 11). Obtaining confidence intervals for the MIRF is also challenging because the model defined by (13) is likely to be misspecified and the asymptotic theory provided in this paper will not be valid in this context. To avoid this problem, we use the blocks-of-blocks bootstrap described in Kilian and Kim (2011) (see also Kilian and Lütkepohl, 2017, pp. 351-353) and Efron's percentile method to compute the confidence intervals for the IRFs and MIRFs (Kim and Kilian, 2011, use the bias-corrected percentile method but note that in the case of IRFs based on local projections the bias correction improves performance only slightly). As in Kilian and Kim (2011), all the results are obtained using 2000 bootstrap replicas and blocks of size 4 for all horizons.

Figures 3 to 5 display, respectively, the mean-based (at the top) and mode-based (at the bottom) impulse response functions corresponding to 1 p.p. shock to inflation, unemployment rate, and interest rate. As in Stock and Watson (2001), the (M)IRFs are accompanied by 66% confidence intervals and are estimated with data from 1960:I to 2000:IV. As noted earlier, the parameters of the structural model were estimated equation-by-equation, and the parameters of the reduced form VAR were estimated using the estimator described

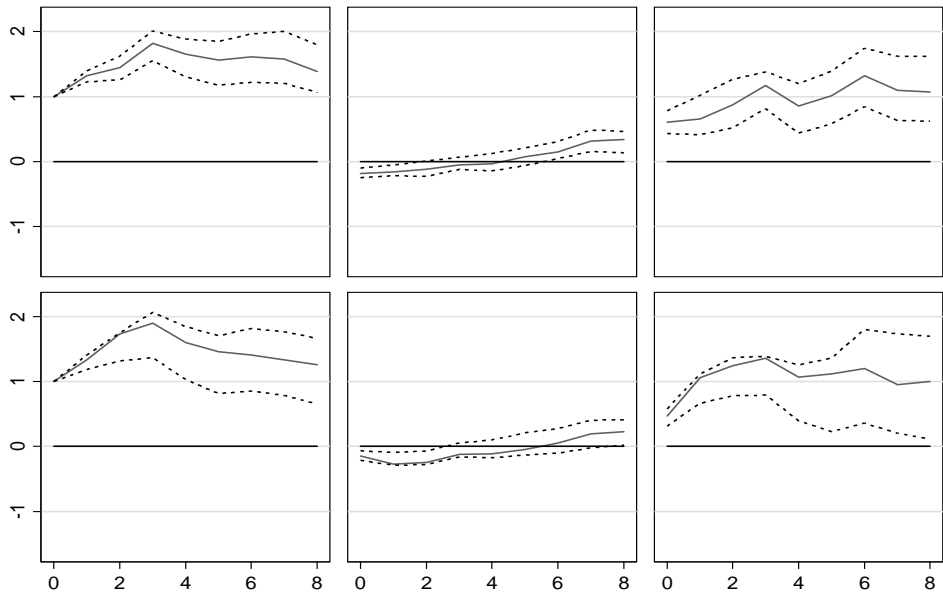


Figure 3: Impulse response functions and 66% confidence intervals (mean-based at the top and mode-based with $s = 6.4$ at the bottom) for the effect of a shock of 1 p.p. to π on (from left to right) π , u , and R .

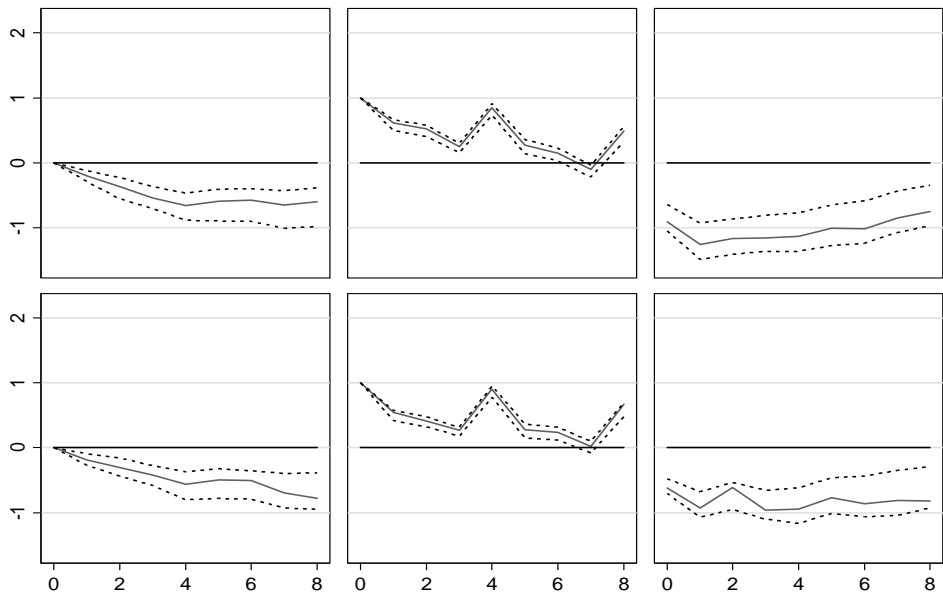


Figure 4: Impulse response functions and 66% confidence intervals (mean-based at the top and mode-based with $s = 6.4$ at the bottom) for the effect of a shock of 1 p.p. to u on (from left to right) π , u , and R .

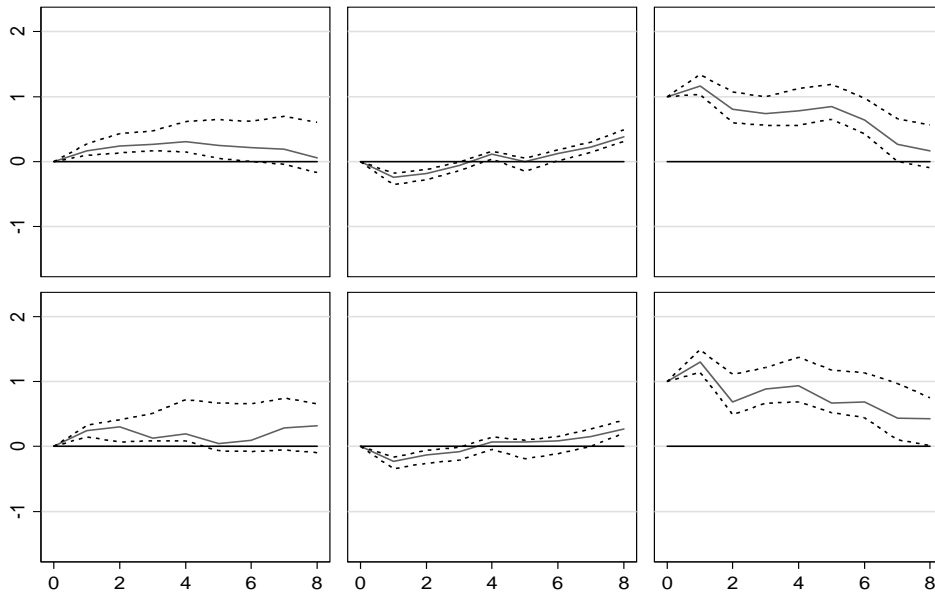


Figure 5: Impulse response functions and 66% confidence intervals (mean-based at the top and mode-based with $s = 6.4$ at the bottom) for the effect of a shock of 1 p.p. to R on (from left to right) π , u , and R .

in Section 2; given the simulation results in Section 4, and given that the (M)IRFs depend only on the slope parameters, all mode-based models were estimated with $s = 6.4$.

The overall pattern of the mode- and mean-based response functions is reassuringly similar and, with such small sample, it is not surprising to find that the confidence intervals of the two sets of impulse response functions are largely overlapping. Looking in detail, however, we can find some differences, for example in the way π and R react to a shock to π , and especially on how R reacts to a shock to u . These differences, naturally, reflect the different dynamic structures of the mode- and mean-based VARs that were revealed by their different forecasting performances.

This application illustrates how mode-based models can be used for some of the tasks VARs are more frequently used for. In the particular three-variable model we considered, mode- and mean-based VARs lead to broadly comparable Granger-causality results and influence response functions. However, the two approaches lead to substantially different pseudo out-of-sample forecasts, with the mode-based models doing particularly well in this

respect. Overall, these results suggest that mode-based models can provide an interesting complement to standard VARs in many applications.

6. CONCLUDING REMARKS

We introduce an estimator of the conditional mode of a random vector allowing for dynamic models and dependent data, thereby extending the results of Kemp and Santos Silva (2012) to the multivariate case and to the time-series context. The proposed approach can be used in the estimation of vector autoregressive conditional mode models, as well as in the estimation of some structural systems of simultaneous equations defined by conditional mode restrictions. The multivariate mode regression estimator is easy to implement using standard software, and it was found to be reasonably well behaved in small samples. An empirical example illustrating the application of the proposed methods suggests that mode- and mean-based models can have very different dynamic structures and therefore can lead to forecasts and impulse response functions with different properties.

Several avenues for future research are left open. As noted before, it would be interesting to consider using the LASSO to perform shrinkage in larger models identified by conditional mode restrictions. Moreover, our results on the estimation of systems of simultaneous equations can be extended to cover the case where restrictions are non-linear and, as mentioned before, it may be possible to develop an estimator for general structural vector autoregressive models under the assumption that the errors of the equations are conditionally independent.

It would also be interesting to study how long-run restrictions can be used to help with identification of structural models, as in Blanchard and Quah (1989). The traditional way to estimate models identified by long-run restrictions involves the estimation of the covariance of the errors, something that cannot be done in our context because the model does not impose any structure on the conditional moments. However, it may be possible to follow a similar approach by replacing the covariance matrix with the Hessian of the conditional density of the errors evaluated at the conditional mode. Indeed, this Hessian shares important characteristics of the conditional covariance matrix and, in particular,

both matrices are diagonal when the errors are conditionally jointly independent, and the relation between the Hessian of the structural and reduced form errors closely resembles the relation between the corresponding covariance matrices.

Finally, because the mode is a robust measure of location, the availability of the multivariate mode regression estimator also offers a possible alternative to several multivariate robust estimators; see, for example, the multivariate regression estimators of Rousseeuw, Van Aelst, Van Driessen, and Agulló (2004) and Agulló, Croux, and Van Aelst (2008), the estimator for VAR models introduced by Muler and Yohai (2013), and the estimators for simultaneous equations models developed by Krishnakumar and Ronchetti (1997) and Maronna and Yohai (1997). Naturally, it would be interesting to explicitly compare the properties and performance of these estimators.

ACKNOWLEDGMENTS

We are very grateful to the Editor Todd Clark and to the Associate Editor and two referees for many useful comments that led to a much more interesting paper. We also thank Omar Boye, Lorenzo Camponovo, Cristiano Cantore, Marcus Chambers, Szabolcs Deák, Davide Delle Monache, Abhimanyu Gupta, Gary Koop, Stepana Lazarova, Myoung-jae Lee, Ricardo Nunes, Ivan Petrella, Sandra Rhouma, and Martin Weale for many helpful comments and discussions. The usual disclaimer applies. Paulo Parente is grateful for the financial support of Fundação para a Ciência e a Tecnologia (Project PTDC/IIM-ECO/3097/2014).

REFERENCES

- Abraham, C., Biau, G. and Cadre, B. (2003). "Simple estimation of the mode of a multivariate density," *Canadian Journal of Statistics*, 31, 23-34.
- Agulló, J., Croux, C. and Van Aelst, S. (2008). "The multivariate least trimmed squares estimator," *Journal of Multivariate Analysis*, 99, 311-338.
- Bernanke, B.S. (1986). "Alternative explanations of the money-income correlation," *Carnegie-Rochester Conference Series on Public Policy*, 25, 49-100.

- Bhansali, R.J. (2002). "Multi-step forecasting," in Clements, M.P. and Hendry, D.F. (eds.), *A companion to economic forecasting*. Oxford: Blackwell Publishing, 206-221.
- Blanchard, O.J. and Quah, D. (1989). "The dynamic effects of aggregate demand and supply disturbances," *American Economic Review*, 79, 655-73.
- Canova, F. (1999), "Vector autoregressive models: Specification, estimation, inference and forecasting" , in Pesaran, H. and Wickens, M. (eds.), *Handbook of Applied Econometrics*, 53-110.
- Chen, Y.-C., Genovese, C.R., Tibshirani, R.J. and Wasserman, L. (2016). "Nonparametric modal regression," *The Annals of Statistics*, 44, 489-514.
- Clements, M.P. and Hendry, D.F. (1998). *Forecasting Economic Time Series*, Cambridge: CUP.
- Cogley, T. and Sargent, T. (2001). "Evolving post-WorldWar II inflation dynamics," *NBER Macroeconomic Annual*, 16, 331-373.
- Collomb, G., Härdle, W. and Hassani, S. (1987). "A note on prediction via estimation of the conditional mode function," *Journal of Statistical Planning and Inference*, 15, 227-236.
- Eddy, W. (1980). "Optimum kernel estimators of the mode," *Annals of Statistics*, 9, 870-882.
- Engle, R.F. (1982). "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, 50, 987-1007.
- Ferguson, T.S. (1967). *Mathematical statistics: A decision theoretic approach*. New York (NY): Academic Press
- Galeano, P., Peña, D. and Tsay, R.S. (2006). "Outlier detection in multivariate time series by projection pursuit," *Journal of the American Statistical Association* 101, 654-669.
- Haavelmo, T. (1943). "The statistical implications of a system of simultaneous equations," *Econometrica*, 11, 1-12.

- Hastie, T., Tibshirani, R. and Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd edition). New York (NY): Springer.
- Hsu, C.-Y. and Wu, T.-J. (2013). “Efficient estimation of the mode of continuous multivariate data,” *Computational Statistics and Data Analysis*, 63, 148-159.
- Huber, P.J. (1973). “Robust regression: Asymptotics, conjectures and Monte Carlo,” *Annals of Statistics*, 1, 799-821.
- Jordà, Ò. (2005). “Estimation and inference of impulse responses by local projections,” *American Economic Review*, 95, 161-182.
- Kemp, G.C.R. and Santos Silva, J.M.C. (2012). “Regression towards the mode,” *Journal of Econometrics*, 170, 92-101.
- Kilian, L. and Lütkepohl, H. (2017). *Structural vector autoregressive analysis*. Cambridge: Cambridge University Press.
- Kilian, L. and Kim, Y.J. (2011). “How reliable are local projection estimators of impulse responses?,” *Review of Economics and Statistics*, 93, 1460-1466.
- Klemelä, J. (2005). “Adaptive estimation of the mode of a multivariate density,” *Journal of Nonparametric Statistics*, 17, 83-105.
- Komunjer, I. and Owyang, M.T. (2012). “Multivariate forecast evaluation and rationality testing,” *The Review of Economics and Statistics*, 94, 1066-1080.
- Konakov, V. (1973). “On the asymptotic normality of the mode of multidimensional distributions,” *Theory of Probability & Its Applications*, 18, 794-799.
- Koop, G., Leon-Gonzalez, R. and Strachan, R. (2009). “On the evolution of the monetary policy transmission mechanism,” *Journal of Economic Dynamics and Control*, 33, 997-1017.
- Koop, G., Pesaran, M. and Potter, S. (1996). “Impulse response analysis in nonlinear multivariate models,” *Journal of Econometrics*, 74, 119-147.
- Krishnakumar, J. and Ronchetti, E. (1997). “Robust estimators for simultaneous equations models,” *Journal of Econometrics*, 78, 295-314.
- Lee, M.J. (1989). “Mode regression,” *Journal of Econometrics*, 42, 337-349.

- Lee, M.J. (1993). "Quadratic mode regression," *Journal of Econometrics*, 57, 1-19.
- Li, G. (1985). "Robust regression," in Hoaglin, D.C, Mosteller, F. and Tukey, J.W. (eds.), *Exploring data tables, trends and shapes*. New York (NY): John Wiley & Sons, 281-340.
- Litterman, R. (1986). "Forecasting with Bayesian vector autoregressions — Five years of experience," *Journal of Business & Economic Statistics*, 4, 25-38.
- Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Berlin: Springer.
- Manski, C.F. (1991). "Regression," *Journal of Economic Literature*, 29, 34-50.
- Maronna, R.A. and Yohai, V.J. (1997). "Robust estimation in simultaneous equations models," *Journal of Statistical Planning and Inference*, 57, 233-244.
- Mokkadem, A. and Pelletier, M. (2003). "The law of the iterated logarithm for the multivariate kernel mode estimator," *ESAIM: Probability and Statistics*, 7, 1-21.
- Muler, N. and Yohai, V.J. (2013). "Robust estimation for vector autoregressive models," *Computational Statistics & Data Analysis*, 65, 68-79.
- Pesaran, M.H. and Timmermann, A. (2005). "Small sample properties of forecasts from autoregressive models under structural breaks," *Journal of Econometrics*, 129, 183-217.
- Primiceri, G. (2005). "Time varying structural vector autoregressions and monetary policy," *Review of Economic Studies*, 72, 821-852.
- Richmond, J. (1974). "Identifiability in linear models," *Econometrica*, 42, 731-736.
- Romano, J.P. (1988). "On weak convergence and optimality of kernel density estimates of the mode," *Annals of Statistics*, 16, 629-647.
- Rousseeuw, P.J. and Leroy, A.M. (1987). *Robust regression and outlier detection*. New York: John Wiley and Sons.
- Rousseeuw, P.J., Van Aelst, S., Van Driessen, K. and Agulló, J. (2004). "Robust multivariate regression," *Technometrics*, 46, 293-305.
- Sager, T.W. (1978). "Estimation of a multivariate mode," *The Annals of Statistics*, 6, 802-812.

- Sager, T.W. (1979). "An iterative method for estimating a multivariate mode and isopleth," *Journal of the American Statistical Association*, 74, 329-339.
- Sakata, S. (2007). "Instrumental variable estimation based on conditional median restriction," *Journal of Econometrics*, 141, 350-382.
- Samanta, M. (1973). "Nonparametric estimation of the mode of a multivariate density," *South African Statistical Journal*, 7, 109-117.
- Sargan, J.D. (1958). "The estimation of economic relationships using instrumental variables," *Econometrica*, 26, 393-415.
- Scott, D.W. (1992). *Multivariate density estimation: Theory, practice, and visualization*. New York (NY): John Wiley & Sons.
- Silverman, B.W. (1986). *Density estimation for statistics and data analysis*, London: Chapman & Hall.
- Sims, C. (1980). "Macroeconomics and reality," *Econometrica*, 48 1-48.
- Sinclair, T., Stekler, H.O. and Muller-Droge, H.C. (2016). "Evaluating forecasts of a vector of variables: A German forecasting competition," *Journal of Forecasting*, 35, 493-503.
- Smith, M.S. and Vahey, S.P. (2016). "Asymmetric forecast densities for U.S. macroeconomic variables from a Gaussian copula model of cross-sectional and serial dependence," *Journal of Business & Economic Statistics*, 34, 416-434.
- Stock, J.H. and Watson, M.W. (2001). "Vector autoregressions," *Journal of Economic Perspectives*, 15, 101-115.
- Stock, J.H. and Watson, M.W. (2012). *Introduction to Econometrics* (3rd ed.), Boston (MA): Pearson.
- Stock, J.H. and Watson, M.W. (2018). "Identification and estimation of dynamic causal effects in macroeconomics using external instruments," *Economic Journal*, 128, 917-948.
- Trietsch, D. (1999). *Statistical quality control: A loss minimization approach*. Singapore: World Scientific.

- Tsay, R.S., Peña, D. and Pankratz, A.E. (2000). “Outliers in multivariate time series,” *Biometrika*, 87, 789-804.
- White, H., Kim, T.-H. and Manganelli, S. (2015). “VAR for VaR: Measuring tail dependence using multivariate regression quantiles,” *Journal of Econometrics*, 187, 169-188.
- Yao, W. and Li, L. (2014a). “A new regression model: Modal linear regression,” *Scandinavian Journal of Statistics*, 41, 656-671.
- Yao, W. and Li, L. (2014b). “Acknowledgement of priority,” *Scandinavian Journal of Statistics*, 41, 1195.
- Zellner, A. (1962). “An efficient method of estimating seemingly unrelated regression equations and tests for aggregation bias,” *Journal of the American Statistical Association*, 57, 348–368.

Appendix of Dynamic Vector Mode Regression

Gordon C.R. Kemp¹ Paulo M.D.C Parente¹ J.M.C. Santos Silva³

22 November 2018

In this appendix we provide the proofs of all the theorems presented in Sections 2 and 3. In particular, in Section A1 we prove the relevant theorems making use of a number of lemmas and Section A2 presents the lemmas and the respective proofs.

A1. PROOF OF THEOREMS

In what follows CR, CS, H, J, M, and T denote the c_r , Cauchy-Schwarz, Hölder, Jensen, Markov, and triangle inequalities respectively, and MVT denotes the mean value theorem; see Davidson (1994, pages 75, 132, 133, 138, 140, 340). In addition, ULLN refers to the uniform law of large numbers for stationary ergodic processes; see Lemma 7.2 of Hayashi (2000). Furthermore, for any array M we let $\|M\|$ denote the non-negative square-root of the sum of the squares of the elements of M . Thus, for example, if M is a matrix then $\|M\| = [\text{trace}(M'M)]^{1/2}$. We also define the set $\text{vec}(\mathcal{A}) \equiv \{y \in \mathbb{R}^{GK} | y = \text{vec}(A), A \in \mathcal{A}\}$. Finally, C denotes a finite positive constant, and we use the following convention for the derivatives of a vector-valued function $F(a)$ with respect to the vector a : $F^{(1)}(a) \equiv \partial F(a)/\partial a'$, $F^{(2)}(a) \equiv \partial^2 F(a)/\partial a \partial a'$, $F^{(3)}(a) \equiv \partial \text{vec}(F^{(2)}(a))/\partial a'$.

Proof of Theorem 1 From Assumption 6 it follows that $\mathcal{K}(\cdot)$ is continuous. From Assumptions 1, 7 and 8 it follows that δ_T is strictly positive and $\widehat{\delta}_T$ is strictly positive with probability 1 and hence that there exists a subset Ω_0 of Ω such that $\Pr(\omega \in \Omega_0) = 1$ and $\widehat{Q}_T(A; \omega)$ is a continuous function of A for all $\omega \in \Omega_0$. Since \mathcal{A} is compact, by Assumption 2, the result follows immediately by Lemma 7.1 from Hayashi (2000). ■

¹Department of Economics, University of Essex. E-mail: kempgcr@essex.ac.uk.

¹ISEG-UL - Universidade de Lisboa; REM - Research in Economics and Mathematics; CEMAPRE-Centro de Matemática Aplicada à Previsão e Decisão Económica. Email: pparente@iseg.ulisboa.pt.

³School of Economics, University of Surrey. E-mail: jmcss@surrey.ac.uk.

Proof of Theorem 2 Let $\theta = (\gamma, \text{vec}(A))'$, where $0 < \gamma < \infty$, and define:

$$J_T(\theta) = T^{-1} \sum_{t=1}^T \delta_T^{-G} \mathcal{K} \left(\frac{\gamma(Y_t - AZ_t)}{\delta_T} \right).$$

Observe that $\widehat{Q}_T(\widehat{A}_T) = \widehat{\gamma}_T^G J_T(\widehat{\theta}_T)$, where $\widehat{\theta}_T = \left(\widehat{\gamma}_T, \text{vec}(\widehat{A}_T) \right)'$ and $\widehat{\gamma}_T = \delta_T / \widehat{\delta}_T$. Lemma 5 below establishes that there exists a continuous function $Q_0(\cdot) : \mathcal{A} \mapsto \mathbb{R}$ such that:

$$\lim_{T \rightarrow \infty} \mathbb{E} [\gamma^G J_T(\theta)] = Q_0(A), \quad \forall \gamma > 0 \text{ \& } A \in \mathcal{A},$$

and that $Q_0(A)$ achieves a unique strict global max on \mathcal{A} at A_0 . Lemma 6 below establishes that for any fixed constant $\nu > 0$ then:

$$\sup_{\theta \in \Theta} |\gamma^G J_T(\theta) - Q_0(A)| = o_p(1), \tag{A.1}$$

where $\Theta = [e^{-\nu}, e^\nu] \times \text{vec}(\mathcal{A})$.

Now fix $\varepsilon > 0$; Assumption 8 implies that there exists $0 < c < \infty$ such that $\Pr(|\ln \widehat{\gamma}_T| > c) < \varepsilon$ for all T . Define:

$$\begin{aligned} \widehat{\gamma}_{T,\varepsilon} &= \begin{cases} \widehat{\gamma}_T, & |\ln \widehat{\gamma}_T| \leq c; \\ 1, & |\ln \widehat{\gamma}_T| > c; \end{cases} \\ \widehat{Q}_{T,\varepsilon}(A) &= \widehat{\gamma}_{T,\varepsilon}^G J_T(\widehat{\gamma}_{T,\varepsilon}, A), \end{aligned}$$

and observe that Equation (A.1) then implies that:

$$\sup_{A \in \mathcal{A}} |\widehat{Q}_{T,\varepsilon}(A) - Q_0(A)| = o_p(1).$$

By the same arguments used to establish the existence of \widehat{A}_T it follows that there exists a random variable $\widehat{A}_{T,\varepsilon}$ such that:

$$\begin{aligned} \Pr(\widehat{A}_{T,\varepsilon} \in \mathcal{A}) &= 1, \\ \Pr(\widehat{Q}_{T,\varepsilon}(\widehat{A}_{T,\varepsilon}) \geq \widehat{Q}_{T,\varepsilon}(A), \forall A \in \mathcal{A}) &= 1. \end{aligned}$$

Now define:

$$\widehat{A}_{T,\varepsilon}^* = \begin{cases} \widehat{A}_T, & |\ln \widehat{\gamma}_T| \leq c, \\ \widehat{A}_{T,\varepsilon}, & |\ln \widehat{\gamma}_T| > c. \end{cases}$$

Then $\widehat{A}_{T,\varepsilon}^*$ is also a random variable and satisfies:

$$\begin{aligned}\Pr\left(\widehat{A}_{T,\varepsilon}^* \in \mathcal{A}\right) &= 1, \\ \Pr\left(\widehat{Q}_{T,\varepsilon}\left(\widehat{A}_{T,\varepsilon}^*\right) \geq \widehat{Q}_{T,\varepsilon}(A), \forall A \in \mathcal{A}\right) &= 1.\end{aligned}$$

Since \mathcal{A} is compact then $\widehat{A}_{T,\varepsilon}^*$ converges in probability to A_0 by Theorem 2.1 from Newey and McFadden (1994). But:

$$\Pr\left(\left\{\left|\widehat{A}_T - A_0\right| > \varepsilon\right\} \cap \left\{\left|\widehat{A}_{T,\varepsilon}^* - A_0\right| > \varepsilon\right\}\right) \leq \Pr\left(\widehat{A}_{T,\varepsilon}^* \neq \widehat{A}_T\right),$$

and hence:

$$\Pr\left(\left|\widehat{A}_T - A_0\right| > \varepsilon\right) \leq \Pr\left(\widehat{A}_{T,\varepsilon}^* \neq \widehat{A}_T\right) + \Pr\left(\left|\widehat{A}_{T,\varepsilon}^* - A_0\right| > \varepsilon\right) \leq \varepsilon + o(1)$$

since $\widehat{A}_{T,\varepsilon}^* \neq \widehat{A}_T$ implies $|\ln \widehat{\gamma}_T| > c$ which occurs with probability less than or equal to ε . Since $\varepsilon > 0$ was arbitrary it follows that \widehat{A}_T converges in probability to A_0 . \blacksquare

Proof of Theorem 3 Theorem 2 implies that \widehat{A}_T is a consistent estimator of A_0 under these assumptions. Since $A \in \text{int}(\mathcal{A})$, by Assumption 9, it follows that:

$$\lim_{T \rightarrow \infty} \Pr\left(\left[\left.\frac{\partial \widehat{Q}_T(A)}{\partial \text{vec}(A)}\right|_{A=\widehat{A}_T}\right] = 0\right) = 1,$$

noting that $\widehat{Q}_T(A)$ is differentiable with respect to A for all A except possibly on a set of outcomes with probability zero. Noting that $\widehat{Q}_T(A) = \widehat{\gamma}_T^G J_T(\widehat{\gamma}_T, A)$, where $J_T(\gamma, A)$ is defined as in the proof of Theorem 2, it follows that $\lim_{T \rightarrow \infty} \Pr\left(R_T(\widehat{\theta}_T) = 0\right) = 1$, where:

$$R_T(\theta) = \gamma^{-1} \frac{\partial J_T(\theta)}{\partial \text{vec}(A)} = T^{-1} \sum_{t=1}^T \delta_T^{-(G+1)} (Z_t \otimes I_G) \mathcal{K}^{(1)}\left(\frac{\gamma(Y_t - AZ_t)}{\delta_T}\right),$$

$\theta = (\gamma, \text{vec}(A))'$, and $\widehat{\theta}_T = \left(\widehat{\gamma}_T, \text{vec}\left(\widehat{A}_T\right)\right)'$. A MVT expansion of $R_T(\theta)$ around $\theta_0 = (1, \text{vec}(A_0))'$ implies that:

$$R_T\left(\widehat{\theta}_T\right) = R_T\left(\theta_0\right) + \left[\left.\frac{\partial R_T(\theta)}{\partial \text{vec}(A)}\right|_{\theta=\theta_0^*}\right] \text{vec}\left(\widehat{A}_T - A_0\right) + \left[\left.\frac{\partial R_T(\theta)}{\partial \gamma}\right|_{\theta=\theta_0^*}\right] (\widehat{\gamma}_T - 1), \quad (\text{A.2})$$

where θ_T^* lies on the line segment joining $\widehat{\theta}_T$ and θ_0 .

First, observe that:

$$(T\delta_T^{G+2})^{1/2} R_T(\theta_0) = (T\delta_T^G)^{-1/2} \sum_{t=1}^T (Z_t \otimes I_G) \mathcal{K}^{(1)} \left(\frac{U_t}{\delta_T} \right) \xrightarrow{d} \mathcal{N}(0, B_0),$$

by Lemma 14, where:

$$B_0 = \mathbb{E} \left[f_t(0|\mathcal{F}_{t-1}) (Z_t \otimes I_G) \mathcal{M} (Z_t \otimes I_G)' \right], \quad \mathcal{M} = \int_{\mathbb{R}^G} \mathcal{K}^{(1)}(s) \mathcal{K}^{(1)}(s)' ds.$$

Note that B_0 is non-singular by Lemma 7. Next observe that:

$$\left[\frac{\partial R_T(\theta)}{\partial \text{vec}(A)'} \right] = \gamma T^{-1} \sum_{t=1}^T \delta_T^{-(G+2)} (Z_t \otimes I_G) \mathcal{K}^{(2)} \left(\frac{\gamma \eta_t(A)}{\delta_T} \right) (Z_t \otimes I_G)',$$

where $\eta_t(A) = (Y_t - AZ_t)$ and it then follows from Lemma 9 that $D_T(\theta) = \gamma^{G+1} \left[\frac{\partial R_T(\theta)}{\partial \text{vec}(A)'} \right]$ converges in probability uniformly to:

$$D(\theta) = \mathbb{E} \left[(Z_t \otimes I_G) f_t^{(2)}((A - A_0) Z_t | \mathcal{F}_{t-1}) (Z_t \otimes I_G)' \right],$$

on any sufficiently small open neighborhood of θ_0 . Assumptions Assumptions 7, 8 and 13 imply that $(T\delta_T^{G+2})^{-1} = o(1)$ which combined with Assumption 14 implies that $\ln \widehat{\gamma}_T = o_p(1)$ and hence that $\widehat{\gamma}_T$ converges in probability to 1. Since \widehat{A}_T converges in probability to A_0 by Theorem 2 then $\widehat{\theta}_T$ converges in probability to θ_0 and hence θ_T^* also converges in probability to θ_0 . Then since $D(\theta)$ is continuous at θ_0 , by Lemma 8, it follows that:

$$\left[\frac{\partial R_T(\theta)}{\partial \text{vec}(A)'} \Big|_{\theta=\theta_T^*} \right] = D_0 + o_p(1),$$

where:

$$D_0 = D(\theta_0) = \mathbb{E} \left[(Z_t \otimes I_G) f_t^{(2)}(0|\mathcal{F}_{t-1}) (Z_t \otimes I_G)' \right].$$

Finally, observe that:

$$\left[\frac{\partial R_T(\theta)}{\partial \gamma} \right] = \gamma T^{-1} \sum_{t=1}^T \delta_T^{-(G+2)} (Z_t \otimes I_G) \mathcal{K}^{(2)} \left(\frac{\gamma \eta_t(A)}{\delta_T} \right) \eta_t(A).$$

It follows from Lemma 11 that $P_T(\theta) = \gamma^{G+1} \left[\frac{\partial R_T(\theta)}{\partial \gamma} \right]$ converges uniformly in probability to:

$$P(\theta) = (G+1) \mathbb{E} \left[(Z_t \otimes I_G) f_t^{(1)}((A - A_0) Z_t | \mathcal{F}_{t-1}) \right],$$

on any sufficiently small open neighborhood of θ_0 . Since θ_T^* converges in probability to $\theta_0 = (1, \text{vec}(A_0)')'$, $f_t^{(1)}(u|\mathcal{F}_{t-1})$ is uniformly bounded with $f_t^{(1)}(0|\mathcal{F}_{t-1}) = 0$, by Assumptions 3 and 9, $P(\theta)$ is continuous, by Lemma 10 and $E\{\|Z_t\|\} < \infty$, by Assumption 4, then:

$$\left[\frac{\partial R_T(\theta)}{\partial \gamma} \Big|_{\theta=\theta_T^*} \right] = o_p(1).$$

Since D_0 is non-singular by Lemma 7 it follows that:

$$h_T^{1/2} \text{vec}(\widehat{A}_T - A_0) = [D_0 + o_p(1)]^{-1} \left[\mathcal{N}(0, B_0) + o_p(1) h_T^{1/2} (\widehat{\gamma}_T - 1) + o_p(1) \right],$$

where $h_T = T\delta_T^{G+2}$. Assumptions 7, 8 13 and 14 then imply that $h_T^{1/2}(\widehat{\gamma}_T - 1) = O_p(1)$ so it follows that:

$$h_T^{1/2} \text{vec}(\widehat{A}_T - A_0) \xrightarrow{d} \mathcal{N}[0, D_0^{-1} B_0 D_0^{-1}],$$

noting that D_0 is symmetric by Lemma 7. But since $(\widehat{\delta}_T/\delta_T) = \widehat{\gamma}_T = 1 + o_p(1)$ it follows that:

$$\left(T\widehat{\delta}_T^{G+2} \right)^{1/2} \text{vec}(\widehat{A}_T - A_0) \xrightarrow{d} \mathcal{N}[0, D_0^{-1} B_0 D_0^{-1}].$$

■

Proof of Theorem 4 First, observe that $\widehat{D}_T = D_T(\widehat{\theta}_T)$ where:

$$D_T(\theta) = \gamma^{G+2} T^{-1} \sum_{t=1}^T \delta_T^{-(G+2)} (Z_t \otimes I_G) \mathcal{K}^{(2)} \left(\frac{\gamma \eta_t(A)}{\delta_T} \right) \eta_t(A),$$

where $\eta_t(A) = (Y_t - AZ_t)$ and $\pi_T = \delta_T/\gamma$. Lemma 9 implies that $D_T(\theta)$ converges in probability uniformly to:

$$D(\theta) = E \left[(Z_t \otimes I_G) f_t^{(2)}((A - A_0) Z_t | \mathcal{F}_{t-1}) (Z_t \otimes I_G)' \right],$$

on any sufficiently small open neighbourhood of θ_0 . But $D(\theta)$ is continuous at θ_0 , by Lemma 8, and $\widehat{\theta}_T$ converges in probability to $\theta_0 = (1, \text{vec}(A_0)')'$, as established in the proof of Theorem 3, and hence it follows that \widehat{D}_T converges in probability to $D_0 = D_0(\theta_0)$.

Second, define:

$$B_T(\theta) = \widehat{\gamma}_T^G T^{-1} \sum_{t=1}^T \delta_T^{-G} (Z_t \otimes I_G) \mathcal{K}^{(1)} \left(\frac{\gamma \eta_t(A)}{\delta_T} \right) \mathcal{K}^{(1)} \left(\frac{\gamma \eta_t(A)}{\delta_T} \right)' (Z_t \otimes I_G)',$$

and observe that $\widehat{B}_T = B_T(\widehat{\theta}_T)$. Lemma 13 implies that $B_T(\theta)$ converges in probability uniformly on any sufficiently small open neighbourhood of θ_0 to:

$$B(\theta) = \mathbb{E} [f_t((A - A_0) Z_t | \mathcal{F}_{t-1}) (Z_t \otimes I_G) \mathcal{M}(Z_t \otimes I_G)'],$$

where $\mathcal{M} = \int_{\mathbb{R}^G} \mathcal{K}^{(1)}(x) \mathcal{K}^{(1)}(x)' dx$. But $B(\theta)$ is continuous at θ_0 , by Lemma 12, and $\widehat{\theta}_T$ converges in probability to $\theta_0 = (1, \text{vec}(A_0))'$ and hence it follows that \widehat{B}_T converges in probability to $B_0 = B(\theta_0)$. ■

Proof of Theorem 5 Note that the function

$$Q_0(\beta^r) = [\alpha_0 + \text{vec}((\Psi_r \Gamma_r^{-1}))]' \Sigma_0^{-1} [\alpha_0 + \text{vec}((\Psi_r \Gamma_r^{-1}))]$$

where $\Sigma_0 = \text{Avar}(\widehat{\alpha}_T)$, is continuous in β^r . Since $\text{Avar}(\widehat{\alpha}_T)$ is positive definite by Lemma 7, it follows that $Q_0(\beta^r) > 0$ for any $\beta^r : \alpha_0 \neq -\text{vec}((\Psi_r \Gamma_r^{-1}))$ and $Q_0(\beta^r) = 0$ if and only if $\alpha_0 = -\text{vec}((\Psi_r \Gamma_r^{-1}))$. Hence by Assumption 15 the minimum is unique. Note also that

$$Q_T(\beta^r) = [\widehat{\alpha}_T + \text{vec}(\Psi_r \Gamma_r^{-1})]' \widehat{\Sigma}_T^{-1} [\widehat{\alpha}_T + \text{vec}(\Psi_r \Gamma_r^{-1})]$$

where $\widehat{\Sigma}_T = \widehat{\text{Avar}}(\widehat{\alpha}_T)$, converges uniformly to $Q_0(\beta^r)$. Since \mathcal{B}^r is compact by Assumption 16, all the assumptions of Theorem 2.1 of Newey and McFadden (1994) are satisfied and hence $\widehat{\beta}_T^r \xrightarrow{p} \beta_0^r$. ■

Proof of Theorem 6 Since β_0^r belongs to the interior of \mathcal{B}^r by Assumption 18, and $\widehat{\beta}_T^r \xrightarrow{p} \beta_0^r$ by Theorem 5, it follows that the first order conditions of the minimization problem (10) are satisfied with probability approaching one, yielding

$$C(\widehat{\beta}_T^r)' \widehat{\Sigma}_T^{-1} [\widehat{\alpha}_T + \text{vec}(\widehat{\Psi}_r \widehat{\Gamma}_r^{-1})] = 0$$

where, as before, $\widehat{\Sigma}_T = \widehat{\text{Avar}}(\widehat{\alpha}_T)$. Note that since $\Sigma_0 = \text{Avar}(\widehat{\alpha}_T)$ is non-singular, by Lemma 7, then $\widehat{\Sigma}_T$ is non-singular with probability tending to 1, by Theorem 4.

Now by a Taylor expansion we have

$$\text{vec}(\widehat{\Psi}_r \widehat{\Gamma}_r^{-1}) = \text{vec}(\Psi_{0,r} \Gamma_{0,r}^{-1}) + C(\widetilde{\beta}_T^r) (\widehat{\beta}_T^r - \beta_0^r),$$

where $\tilde{\beta}_T^r$ is on a line joining $\widehat{\beta}_T^r$ and β_0^r and $\Psi_{0,r}$ and $\Gamma_{0,r}$ correspond to the matrices Ψ_r , Γ_r evaluated at $\beta^r = \beta_0^r$. But $\text{vec}(\Psi_{0,r}\Gamma_{0,r}^{-1}) = -\alpha_0$ and therefore we have

$$C(\widehat{\beta}_T^r)' \widehat{\Sigma}_T^{-1} C(\tilde{\beta}_T^r) \sqrt{T\delta_T^{G+2}} (\widehat{\beta}_T^r - \beta_0^r) = -C(\widehat{\beta}_T^r)' \widehat{\Sigma}_T^{-1} \sqrt{T\delta_T^{G+2}} (\widehat{\alpha}_T - \alpha_0).$$

Now Assumptions 15, 17 and 18 imply that $C(\beta^r)$ is continuous with full rank in a neighbourhood of β_0^r . Since $\widehat{\beta}_T^r \xrightarrow{p} \beta_0^r$, by Theorem 5, it follows that $\tilde{\beta}_T^r \xrightarrow{p} \beta_0^r$ and hence that both $C(\widehat{\beta}_T^r)$ and $C(\tilde{\beta}_T^r)$ converge in probability to C_0 . The result follows from Theorems 3 and 4 together with the fact that $\text{rank}(C_0) = G(G+K) - \rho$, by Assumption 17. ■

Proof of Theorem 7 This follows immediately since $\widehat{\text{Avar}}(\widehat{\alpha}_T) = \widehat{D}_T^{-1} \widehat{B}_T \widehat{D}_T^{-1}$ converges in probability to $\text{Avar}(\widehat{\alpha}_T) = D_0^{-1} B_0 D_0^{-1}$, by Theorem 4, $\text{Avar}(\widehat{\alpha}_T) = D_0^{-1} B_0 D_0^{-1}$ is non-singular, by Lemma 7, and since $\widehat{C}_T = C(\widehat{\beta}_T^r)$ converges in probability to the full rank matrix C_0 , as argued in the proof of Theorem 6. ■

A2. LEMMAS

Lemma 1 *Suppose Assumption 2 is satisfied. Let $0 < \nu < \infty$ be a scalar constant and define $\Theta = [e^{-\nu}, e^\nu] \times \text{vec}(\mathcal{A})$; then there exists a constant $d < \infty$ such for each $T = 1, 2, \dots$, there is a finite subset Θ_T of Θ and a mapping $\bar{\theta}_T$ from Θ to Θ_T for which:*

1. $\|\theta - \bar{\theta}_T(\theta)\| \leq T^{-2}$ for all $\theta \in \Theta$, where for any $\theta_1 = (\gamma_1, \text{vec}(A_1)')'$ and $\theta_2 = (\gamma_2, \text{vec}(A_2)')'$:

$$\|\theta_1 - \theta_2\| = \{(\gamma_1 - \gamma_2)^2 + \text{vec}(A_1 - A_2)' \text{vec}(A_1 - A_2)\}^{1/2}$$

2. *the number of elements of Θ_T is less than or equal to $T^{2G(K+1)}d$.*

Proof. By construction $[e^{-\nu}, e^\nu]$ is a closed bounded interval and since $\text{vec}(\mathcal{A})$ is a compact subset of \mathbb{R}^{GK} , by Assumption 2, then Θ is a compact subset of $\mathbb{R}^{G(K+1)}$ and the result follows immediately. ■

Lemma 2 *Let $\{\theta_i\}_{i=1}^n$ be a finite subset of some non-empty set Θ and suppose that for each $i = 1, 2, \dots, n$, $\{(\zeta_t(\theta_i), \mathcal{G}_t)\}_{t=1}^\infty$ is a martingale difference sequence. In addition,*

suppose that there exists $0 < c < \infty$ such that $\Pr(|\zeta_t(\theta_i)| \leq c) = 1$ for all $i = 1, 2, \dots, n$ and $t = 1, 2, \dots$. Then for any finite positive real constants a and b and any finite positive integer constant T :

$$\Pr\left(\sup_{1 \leq i \leq n} \left| \sum_{t=1}^T \zeta_t(\theta_i) \right| \geq a\right) \leq 2n \exp\left\{\frac{-a^2/2}{ac+b}\right\} + \Pr\left(\sup_{1 \leq j \leq n} \sum_{t=1}^T \text{Var}(\zeta_t(\theta_j) | \mathcal{G}_{t-1}) > b\right).$$

Proof. For each $i = 1, \dots, n$ and $t = 1, \dots, T$ define $S_i = \left| \sum_{t=1}^T \zeta_t(\theta_i) \right|$ and $V_i = \sum_{t=1}^T \text{Var}(\zeta_t(\theta_i) | \mathcal{G}_{t-1})$. Then:

$$\begin{aligned} \Pr\left(\sup_{1 \leq i \leq n} \left| \sum_{t=1}^T \zeta_t(\theta_i) \right| \geq a\right) &= \Pr\left(\sup_{1 \leq i \leq n} S_i \geq a\right) \\ &\leq \Pr\left(\sup_{1 \leq i \leq n} S_i \geq \gamma^G a \ \& \ \sup_{1 \leq j \leq n} V_j \leq b\right) + \Pr\left(\sup_{1 \leq j \leq n} V_j > b\right) \\ &\leq \sum_{i=1}^n \Pr\left(S_i \geq a \ \& \ \sup_{1 \leq j \leq n} V_j \leq b\right) + \Pr\left(\sup_{1 \leq j \leq n} V_j > b\right) \\ &\leq \sum_{i=1}^n \Pr(S_i \geq a \ \& \ V_i \leq b) + \Pr\left(\sup_{1 \leq j \leq n} V_j > b\right). \end{aligned}$$

But by Freedman's inequality (Freedman, 1975, Theorem 1.6), it follows that:

$$\Pr(S_i \geq a \ \& \ V_i \leq b) = \Pr\left(\left| \sum_{t=1}^T \zeta_t(\theta_i) \right| \geq a \ \& \ \sum_{t=1}^T \text{Var}(\zeta_t(\theta_i) | \mathcal{G}_{t-1}) \leq b\right) \leq 2 \exp\left\{\frac{-a^2/2}{ac+b}\right\},$$

noting that $\{(-\zeta_t(\theta_i), \mathcal{G}_t)\}_{t=1}^\infty$ is also a martingale with $\text{Var}(-\zeta_t(\theta_i) | \mathcal{G}_{t-1}) = \text{Var}(\zeta_t(\theta_i) | \mathcal{G}_{t-1})$. Hence it follows that:

$$\Pr\left(\sup_{1 \leq i \leq n} \left| \sum_{t=1}^T \zeta_t(\theta_i) \right| \geq a\right) \leq 2n \exp\left\{\frac{-a^2/2}{ac+b}\right\} + \Pr\left(\sup_{1 \leq j \leq n} \sum_{t=1}^T \text{Var}(\zeta_t(\theta_j) | \mathcal{G}_{t-1}) > b\right),$$

as desired. ■

Lemma 3 Suppose that Assumptions 1 and 2 are satisfied. Let $0 < \nu < \infty$ be a scalar constant and define $\Theta = [e^{-\nu}, e^\nu] \times \text{vec}(\mathcal{A})$. In addition, let σ be a scalar constant such that $0 \leq \sigma \leq 2$, $\psi(\cdot)$ be a continuously differentiable function from \mathbb{R}^G to \mathbb{R} with $\sup_u \|\psi^{(1)}(u)\| \leq C_\psi$ for some $C_\psi < \infty$, and $m(\cdot)$ be a function from \mathbb{R}^{G+K} to \mathbb{R} such

that $|m(W_t)| \leq \kappa \|W_t\|^p$ for some $\kappa < \infty$ and $0 \leq p < \infty$. Define:

$$\begin{aligned}\Psi_T(\theta) &= T^{-1} \sum_{t=1}^T \delta_T^{-(G+\sigma)} \psi \left(\frac{\eta_t(A)}{\pi_T} \right) m(W_t), \\ \Psi_T^e(\theta) &= T^{-1} \sum_{t=1}^T \delta_T^{-(G+\sigma)} \mathbb{E} \left[\psi \left(\frac{\eta_t(A)}{\pi_T} \right) m(W_t) \mid \mathcal{F}_{t-1} \right]\end{aligned}$$

where $\pi_T = \delta_T/\gamma$, $\eta_t(A) = (Y_t - AZ_t)$ and $\theta = (\gamma, \text{vec}(A))'$, where $\{\delta_T\}_{T=1}^\infty$ is a sequence of finite positive constants. If $\{\delta_T\}_{T=1}^\infty$ satisfies $\delta_T = o(1)$ and $T^{-2}\delta_T^{-(G+\sigma+1)} = o(1)$ and $\mathbb{E} \{ \|W_t\|^{p+1} \} < \infty$ then:

$$\begin{aligned}\sup_{\theta \in \Theta} |\Psi_T(\theta) - \Psi_T(\bar{\theta}_T(\theta))| &= o_p(1), \\ \sup_{\theta \in \Theta} |\Psi_T^e(\theta) - \Psi_T^e(\bar{\theta}_T(\theta))| &= o_p(1),\end{aligned}$$

where $\bar{\theta}_T(\cdot)$ is characterized as in Lemma 1.

Proof. Define:

$$h_{tT}(\theta) = T^{-1} \delta_T^{-(G+\sigma)} \psi \left(\frac{\gamma \eta_t(A)}{\delta_T} \right) m(W_t)$$

so $\Psi_T(\theta) = \sum_{t=1}^T h_{tT}(\theta)$. By MVT:

$$\begin{aligned}h_{tT}(\theta) - h_{tT}(\bar{\theta}_T(\theta)) &= T^{-1} \delta_T^{-(G+\sigma)} \left[\psi \left(\frac{\gamma \eta_t(A)}{\delta_T} \right) - \psi \left(\frac{\bar{\gamma}_T(\theta) \eta_t(\bar{A}_T(\theta))}{\delta_T} \right) \right] m(W_t) \\ &= T^{-1} \delta_T^{-(G+\sigma)} \psi^{(1)} \left(\frac{\gamma_T^* \eta_t(A_T^*)}{\delta_T} \right)' \left(\frac{\gamma \eta_t(A) - \bar{\gamma}_T(\theta) \eta_t(\bar{A}_T(\theta))}{\delta_T} \right) m(W_t),\end{aligned}$$

where $(\gamma_T^*, \text{vec}(A_T^*))'$ lies on the line segment joining $(\gamma, \text{vec}(A))'$ and $(\bar{\gamma}_T(\theta), \text{vec}(\bar{A}_T(\theta)))'$. Hence it follows by CS that:

$$|h_{tT}(\theta) - h_{tT}(\bar{\theta}_T(\theta))| \leq C_\psi T^{-1} \delta_T^{-(G+\sigma+1)} \|\gamma \eta_t(A) - \bar{\gamma}_T(\theta) \eta_t(\bar{A}_T(\theta))\| |m(W_t)|.$$

Now:

$$\begin{aligned}\gamma \eta_t(A) - \bar{\gamma}_T(\theta) \eta_t(\bar{A}_T(\theta)) &= (\gamma - \bar{\gamma}_T(\theta)) Y_t - [\gamma A - \bar{\gamma}_T(\theta) \bar{A}_T(\theta)] Z_t \\ &= (\gamma - \bar{\gamma}_T(\theta)) U_t - [\gamma (A - \bar{A}_T(\theta)) + (\gamma - \bar{\gamma}_T(\theta)) \bar{A}_T(\theta)] Z_t,\end{aligned}$$

and thus:

$$\|\gamma\eta_t(A) - \bar{\gamma}_T(\theta)\eta_t(\bar{A}_T(\theta))\| \leq \|\theta - \bar{\theta}_T(\theta)\| \|Y_t\| + (e^\nu + d_A) \|\theta - \bar{\theta}_T(\theta)\| \|Z_t\|,$$

where $d_A = \sup_{A \in \mathcal{A}} \|A\| < \infty$, which implies that:

$$\sup_{\theta \in \Theta} |h_{tT}(\theta) - h_{tT}(\bar{\theta}_T(\theta))| \leq C_\psi T^{-3} \delta_T^{-(G+\sigma+1)} \{\|Y_t\| + (e^\nu + d_A) \|Z_t\|\} |m(W_t)|.$$

In turn this implies that:

$$\begin{aligned} \sup_{\theta \in \Theta} |\Psi_T(\theta) - \Psi_T(\bar{\theta}_T(\theta))| &\leq \sum_{t=1}^T \left\{ \sup_{\theta \in \Theta} |h_{tT}(\theta) - h_{tT}(\bar{\theta}_T(\theta))| \right\} \\ &\leq C_\psi T^{-3} \delta_T^{-(G+\sigma+1)} \sum_{t=1}^T \{\|Y_t\| + (e^\nu + d_A) \|Z_t\|\} |m(W_t)| \end{aligned}$$

and hence that:

$$\begin{aligned} \mathbb{E} \left\{ \sup_{\theta \in \Theta} |\Psi_T(\theta) - \Psi_T(\bar{\theta}_T(\theta))| \right\} &\leq C_\psi T^{-3} \delta_T^{-(G+\sigma+1)} \mathbb{E} \left[\sum_{t=1}^T \{\|Y_t\| + (e^\nu + d_A) \|Z_t\|\} |m(W_t)| \right] \\ &\leq C_\psi \kappa T^{-3} \delta_T^{-(G+\sigma+1)} (1 + e^\nu + d_A) \mathbb{E} \left[\sum_{t=1}^T \|W_t\|^{p+1} \right] \\ &= O \left[T^{-2} \delta_T^{-(G+\sigma+1)} \right] = o(1), \end{aligned}$$

since $T^{-2} \delta_T^{-(G+\sigma+1)} = o(1)$, $\mathbb{E} \{\|W_t\|^{p+1}\} < \infty$, and the data are stationary. Then M implies that:

$$\sup_{\theta \in \Theta} |\Psi_T(\theta) - \Psi_T(\bar{\theta}_T(\theta))| = o_p(1).$$

Now define $h_{tT}^e(\theta) = \mathbb{E}[h_{tT}(\theta) | \mathcal{F}_{t-1}]$, so $\Psi_T^e(\theta) = \sum_{t=1}^T h_{tT}^e(\theta)$, and observe that:

$$\begin{aligned} |h_{tT}^e(\theta) - h_{tT}^e(\bar{\theta}_T(\theta))| &= |\mathbb{E}[h_{tT}(\theta) | \mathcal{F}_{t-1}] - \mathbb{E}[h_{tT}(\bar{\theta}_T(\theta)) | \mathcal{F}_{t-1}]| \\ &\leq \mathbb{E} \{ |h_{tT}(\theta) - h_{tT}(\bar{\theta}_T(\theta))| | \mathcal{F}_{t-1} \}, \end{aligned}$$

and thus:

$$\begin{aligned} |\Psi_T^e(\theta) - \Psi_T^e(\bar{\theta}_T(\theta))| &\leq \sum_{t=1}^T |h_{tT}^e(\theta) - h_{tT}^e(\bar{\theta}_T(\theta))| \\ &\leq \sum_{t=1}^T \mathbb{E} \{ |h_{tT}(\theta) - h_{tT}(\bar{\theta}_T(\theta))| | \mathcal{F}_{t-1} \}, \end{aligned}$$

by J, which implies that:

$$\sup_{\theta \in \Theta} |\Psi_T^e(\theta) - \Psi_T^e(\bar{\theta}_T(\theta))| \leq \sum_{t=1}^T \mathbb{E} \left\{ \sup_{\theta \in \Theta} |h_{tT}(\theta) - h_{tT}(\bar{\theta}_T(\theta))| \mid \mathcal{F}_{t-1} \right\}.$$

But then:

$$\begin{aligned} \mathbb{E} \left\{ \sup_{\theta \in \Theta} |\Psi_T^e(\theta) - \Psi_T^e(\bar{\theta}_T(\theta))| \right\} &\leq \sum_{t=1}^T \mathbb{E} \left\{ \sup_{\theta \in \Theta} |h_{tT}(\theta) - h_{tT}(\bar{\theta}_T(\theta))| \right\} \\ &\leq C_\psi T^{-3} \delta_T^{-(G+\sigma+1)} \sum_{t=1}^T \mathbb{E} [\{\|Y_t\| + (e^\nu + d_A) \|Z_t\|\} |m(W_t)|] \\ &= o(1), \end{aligned}$$

and hence by M:

$$\sup_{\theta \in \Theta} |\Psi_T^e(\theta) - \Psi_T^e(\bar{\theta}_T(\theta))| = o_p(1).$$

■

Lemma 4 *Suppose that Assumptions 1 and 2 are satisfied. Let $0 < \nu < \infty$ be a scalar constant and define $\Theta = [e^{-\nu}, e^\nu] \times \text{vec}(\mathcal{A})$. In addition, let σ be a scalar constant such that $0 \leq \sigma \leq 2$, $\psi(\cdot)$ be a continuously differentiable function \mathbb{R}^G to \mathbb{R} , such that $\sup_u |\psi(u)| < \infty$ and $\sup_u \|\psi^{(1)}(u)\| < \infty$, and $m(\cdot)$ be a function from \mathbb{R}^{G+K} to \mathbb{R} such that $\mathbb{E}\{|m(W_t)|^r\} < \infty$ for some $r > (G + \sigma + 2)/2$. Define:*

$$\begin{aligned} \Psi_T^\dagger(\theta) &= T^{-1} \sum_{t=1}^T \delta_T^{-(G+\sigma)} \psi \left(\frac{\gamma \eta_t(A)}{\delta_T} \right) m_T^\dagger(W_t), \\ \Psi_T^{\dagger,e}(\theta) &= T^{-1} \sum_{t=1}^T \delta_T^{-(G+\sigma)} \mathbb{E} \left[\psi \left(\frac{\gamma \eta_t(A)}{\delta_T} \right) m_T^\dagger(W_t) \mid \mathcal{F}_{t-1} \right], \\ m_T^\dagger(w) &= m(w) \chi \{ |m(w)| > \delta_T^{-2} \}, \end{aligned}$$

where $\eta_t(A) = (Y_t - AZ_t)$, $\theta = (\gamma, \text{vec}(A)')'$ and $\chi(\cdot)$ is the indicator function, and where $\{\delta_T\}_{T=1}^\infty$ is a sequence of strictly positive constants such that $\delta_T = o(1)$; then:

$$\sup_{\theta \in \Theta} |\Psi_T^\dagger(\theta)| = o_p(1), \quad \sup_{\theta \in \Theta} |\Psi_T^{\dagger,e}(\theta)| = o_p(1).$$

Proof. By assumption there exists $C_\psi < \infty$ such that $\sup_u |\psi(u)| \leq C_\psi$ and $\sup_u \|\psi^{(1)}(u)\| \leq C_\psi$. Define:

$$h_{tT}^\dagger(\theta) = T^{-1} \delta_T^{-(G+\sigma)} \psi \left(\frac{\gamma \eta_t(A)}{\delta_T} \right) m_T^\dagger(W_t), \quad h_{tT}^{\dagger,e}(\theta) = \mathbb{E} \left\{ h_{tT}^\dagger(\theta) \mid \mathcal{F}_{t-1} \right\},$$

so:

$$\Psi_T^\dagger(\theta) = \sum_{t=1}^T h_{tT}^\dagger(\theta), \quad \Psi_T^{\dagger,e}(\theta) = \sum_{t=1}^T h_{tT}^{\dagger,e}(\theta).$$

From T it follows that:

$$\left| \Psi_T^\dagger(\theta) \right| \leq \sum_{t=1}^T \left| h_{tT}^\dagger(\theta) \right| \leq C_\psi T^{-1} \sum_{t=1}^T \delta_T^{-(G+\sigma)} \left| m_T^\dagger(W_t) \right|,$$

and hence that:

$$\sup_{\theta \in \Theta} \left| \Psi_T^\dagger(\theta) \right| \leq C_\psi \delta_T^{-(G+\sigma)} T^{-1} \sum_{t=1}^T \left\{ \sup_{\theta \in \Theta} \left| m_T^\dagger(W_t) \right| \right\},$$

which in turn implies that:

$$\mathbb{E} \left\{ \sup_{\theta \in \Theta} \left| \Psi_T^\dagger(\theta) \right| \right\} \leq C_\psi \delta_T^{-(G+\sigma)} \mathbb{E} \left\{ \sup_{\theta \in \Theta} \left| m_T^\dagger(W_t) \right| \right\}.$$

Now for any random variable X with $\mathbb{E}(|X|^s) < \infty$ for some $s > 1$ and any constant $c > 0$ then it follows by H and M that:

$$\mathbb{E}(|X| \chi \{|X| > c\}) \leq c^{-(s-1)} \mathbb{E}(|X|^s).$$

Setting $X = \sup_{\theta \in \Theta} |m(W_t; \theta)|$ then it follows that:

$$\begin{aligned} \sup_{\theta \in \Theta} \left| m_T^\dagger(W_t) \right| &= \sup_{\theta \in \Theta} |m(W_t) \chi \{|m(W_t)| > \delta_T^{-2}\}| \\ &= \sup_{\theta \in \Theta} |m(W_t)| \chi \left\{ \sup_{\theta \in \Theta} |m(W_t)| > \delta_T^{-2} \right\} = |X| \chi \{|X| > \delta_T^{-2}\}. \end{aligned}$$

Since $r > 1$ this then implies that:

$$\mathbb{E} \left\{ \sup_{\theta \in \Theta} \left| m_T^\dagger(W_t) \right| \right\} \leq \delta_T^{2(r-1)} \mathbb{E} \left\{ \sup_{\theta \in \Theta} |m(W_t)|^r \right\} \leq \delta_T^{2(r-1)} C_r,$$

where $C_r = \mathbb{E} \left\{ \sup_{\theta \in \Theta} |m(W_t)|^r \right\}$, and hence that:

$$\mathbb{E} \left\{ \sup_{\theta \in \Theta} \left| \Psi_T^\dagger(\theta) \right| \right\} \leq \delta_T^{2r-2-G-\sigma} C_\psi C_r = o(1),$$

noting that $r > (G + \sigma + 2)/2$. This then implies that $\sup_{\theta \in \Theta} \left| \Psi_T^\dagger(\theta) \right| = o_p(1)$ by M.

Next, observe that:

$$\mathbb{E} \left\{ \sup_{\theta \in \Theta} \left| h_{tT}^{\dagger,e}(\theta) \right| \right\} \leq \mathbb{E} \left\{ \sup_{\theta \in \Theta} \left| \mathbb{E} \left[h_{tT}^\dagger(\theta) \mid \mathcal{F}_{t-1} \right] \right| \right\} \leq \mathbb{E} \left\{ \sup_{\theta \in \Theta} \left| h_{tT}^\dagger(\theta) \right| \right\},$$

by J, and hence that:

$$\mathbb{E} \left\{ \sup_{\theta \in \Theta} \left| \Psi_T^{\dagger, e}(\theta) \right| \right\} \leq \sum_{t=1}^T \mathbb{E} \left\{ \sup_{\theta \in \Theta} \left| h_{tT}^{\dagger, e}(\theta) \right| \right\} \leq \sum_{t=1}^T \mathbb{E} \left\{ \sup_{\theta \in \Theta} \left| h_{tT}^{\dagger}(\theta) \right| \right\} = o(1),$$

by J, which implies that $\sup_{\theta \in \Theta} \left| \Psi_T^{\dagger, e}(\theta) \right| = o_p(1)$ by M. ■

Lemma 5 *Suppose that Assumptions 1–7 are satisfied. Let $\theta = (\gamma, \text{vec}(A)')'$ where $0 < \gamma < \infty$ and $A \in \mathcal{A}$, and define $J_T(\theta) = T^{-1} \sum_{t=1}^T \delta_T^{-G} \mathcal{K} \left(\frac{\gamma(Y_t - AZ_t)}{\delta_T} \right)$. Then:*

$$\lim_{T \rightarrow \infty} \mathbb{E} [\gamma^G J_T(\theta)] = Q_0(A) = \mathbb{E} [f_t((A - A_0) Z_t | \mathcal{F}_{t-1})],$$

$Q_0(\cdot)$ is continuous on \mathcal{A} , and $Q_0(A) \leq Q_0(A_0)$ for all $A \in \mathcal{A}$ with equality if and only if $A = A_0$.

Proof. First, fix $0 < \gamma < \infty$. Then for each $t = 1, \dots, T$ define:

$$\begin{aligned} h_{tT}(\theta) &= (T\pi_T^G)^{-1} \mathcal{K} \left(\frac{\eta_t(A)}{\pi_T} \right) = (T\pi_T^G)^{-1} \mathcal{K} \left(\frac{U_t - (A - A_0) Z_t}{\pi_T} \right), \\ h_{tT}^e(\theta) &= \mathbb{E} [h_{tT}(\theta) | \mathcal{F}_{t-1}] = T^{-1} \int_{\mathbb{R}^G} \pi_T^{-G} \mathcal{K} \left(\frac{u - (A - A_0) Z_t}{\pi_T} \right) f_t(u | \mathcal{F}_{t-1}) du \end{aligned}$$

where $\eta_t(A) = (Y_t - AZ_t)$ and $\pi_T = \delta_T / \gamma$, and note that:

$$h_{tT}^e(\theta) = T^{-1} \int_{\mathbb{R}^G} \mathcal{K}(x) f_t((A - A_0) Z_t + \pi_T x | \mathcal{F}_{t-1}) dx,$$

by transformation of variables from u to $x = \pi_T^{-1} [u - (A - A_0) Z_t]$. In addition, define:

$$Q_{t0}^e(A, \pi) = \int_{\mathbb{R}^G} \mathcal{K}(x) f_t((A - A_0) Z_t + \pi x | \mathcal{F}_{t-1}) dx,$$

so $h_{tT}^e(\theta) = T^{-1} Q_{t0}^e(A, \pi_T)$. Assumption 1 implies that:

$$\begin{aligned} \mathbb{E} [\gamma^G J_T(\theta)] &= \mathbb{E} \left[\gamma^G T^{-1} \sum_{t=1}^T \delta_T^{-G} \mathcal{K} \left(\frac{\gamma(Y_t - AZ_t)}{\delta_T} \right) \right] \\ &= \mathbb{E} [T h_{tT}(\theta)] = \mathbb{E} [T h_{tT}^e(\theta)] \\ &= \mathbb{E} \left[\int_{\mathbb{R}^G} \mathcal{K}(x) f_t((A - A_0) Z_t + \pi_T x | \mathcal{F}_{t-1}) dx \right], \end{aligned}$$

for all $A \in \mathcal{A}$. But $Q_{t0}^e(A, \pi)$ is continuous in $(\text{vec}(A)', \pi)'$ for all (t, ω) by dominated convergence since $\sup_u |f_t(u | \mathcal{F}_{t-1})| < \infty$ for all (u, t, ω) , by Assumption 3, $f_t(u | \mathcal{F}_{t-1})$

is continuous in u , by Assumption 3, and $\int_{\mathbb{R}^G} |\mathcal{K}(x)| dx \leq C$, by Assumption 6. Hence dominated convergence implies that:

$$\begin{aligned} \lim_{T \rightarrow \infty} \mathbb{E} [\gamma^G J_T(\theta)] &= \lim_{T \rightarrow \infty} \mathbb{E} [Q_{t_0}^e(A, \delta_T)] \\ &= \mathbb{E} \left[\int_{\mathbb{R}^G} \mathcal{K}(x) f_t((A - A_0) Z_t | \mathcal{F}_{t-1}) dx \right] \\ &= \mathbb{E} [f_t((A - A_0) Z_t | \mathcal{F}_{t-1})] = Q_0(A), \end{aligned}$$

since $\int_{\mathbb{R}^G} \mathcal{K}(x) dx = 1$, by Assumption 6, and $\delta_T = o(1)$, by Assumption 7, which implies that $\pi_T = o(1)$, and also since that $Q_0(A)$ is continuous in A , which follows from the continuity of $\mathcal{K}(\cdot)$ by Assumption 6.

Finally, by Assumption 3 then for any A :

$$f_t((A - A_0) Z_t | \mathcal{F}_{t-1}) \leq f_t(0 | \mathcal{F}_{t-1}), \quad \forall \omega \in \Omega,$$

while by Assumption 5 it follows that for any $A \neq A_0$ there exists a set $S \in \mathcal{F}_{t-1}$ with $\mathbb{P}(S) > 0$ such that:

$$f_t((A - A_0) Z_t | \mathcal{F}_{t-1}) < f_t(0 | \mathcal{F}_{t-1}), \quad \forall \omega \in S,$$

and hence it follows that for all $A \neq A_0$:

$$Q_0(A) < Q_0(A_0).$$

Thus $Q_0(A)$ achieves a unique strict global maximum over $A \in \mathcal{A}$ at $A = A_0$, as desired.

■

Lemma 6 *Suppose that Assumptions 1–8 are satisfied. Let $0 < \nu < \infty$ be a scalar constant and define $\Theta = [e^{-\nu}, e^\nu] \times \text{vec}(\mathcal{A})$; then:*

$$\sup_{\theta \in \Theta} |\gamma^G J_T(\theta) - Q_0(A)| = o_p(1),$$

where $\theta = (\gamma, \text{vec}(A)')'$, $J_T(\cdot)$ is defined as in Theorem 2 and $Q_0(\cdot)$ is given as in Lemma 5.

Proof. Let Θ_T and $\bar{\theta}_T(\cdot)$ be characterized as in Lemma 1. Then:

$$\begin{aligned} \gamma^G J_T(\theta) - Q_0(A) &= \gamma^G [J_T(\theta) - J_T(\bar{\theta}_T(\theta))] + \gamma^G [J_T(\bar{\theta}_T(\theta)) - J_T^e(\bar{\theta}_T(\theta))] \\ &\quad + \gamma^G [J_T^e(\bar{\theta}_T(\theta)) - J_T^e(\theta)] + [\gamma^G J_T^e(\theta) - Q_0(A)], \end{aligned} \quad (\text{A.3})$$

where:

$$J_T(\theta) = T^{-1} \sum_{t=1}^T \delta_T^{-G} \mathcal{K} \left(\frac{\eta_t(A)}{\pi_T} \right).$$

$$J_T^e(\theta) = T^{-1} \sum_{t=1}^T \delta_T^{-G} \mathbf{E} \left[\mathcal{K} \left(\frac{\eta_t(A)}{\pi_T} \right) \middle| \mathcal{F}_{t-1} \right],$$

where $\pi_T = \delta_T/\gamma$ and $\eta_t(A) = (Y_t - AZ_t)$. Equation (A.3) implies by T that:

$$\sup_{\theta \in \Theta} |\gamma^G J_T(\theta) - Q_0(A)| \leq M_{1,T} + M_{2,T} + M_{3,T} + M_{4,T}, \quad (\text{A.4})$$

where:

$$M_{1,T} = \sup_{\theta \in \Theta} \gamma^G |J_T(\theta) - J_T(\bar{\theta}_T(\theta))|, \quad M_{2,T} = \sup_{\theta \in \Theta} \gamma^G |J_T(\bar{\theta}_T(\theta)) - J_T^e(\bar{\theta}_T(\theta))|$$

$$M_{3,T} = \sup_{\theta \in \Theta} \gamma^G |J_T^e(\bar{\theta}_T(\theta)) - J_T^e(\theta)|, \quad M_{4,T} = \sup_{\theta \in \Theta} |\gamma^G J_T^e(\theta) - Q_0(A)|$$

Observe that $J_T(\theta)$ can be re-written in the form:

$$J_T(\theta) = T^{-1} \sum_{t=1}^T \delta_T^{-(G+\sigma)} \psi \left(\frac{\eta_t(A)}{\pi_T} \right) m(W_t),$$

with $\sigma = 0$, $\psi(x) = \mathcal{K}(x)$ for all x , and $m(W_t) = 1 = \|W_t\|^0$ for all (t, ω) . Since $\sup_u |\mathcal{K}^{(1)}(u)| < \infty$, by Assumption 6, $\mathbf{E}\{\|W_t\|\} < \infty$, by Assumption 4, and $T^{-2} \delta_T^{-(G+1)} = o(1)$, by Assumption 7, then $J_T(\theta)$ satisfies the requirements on $\Psi_T(\theta)$ for Lemma 3 to apply and hence:

$$\sup_{\theta \in \Theta} |J_T(\theta) - J_T(\bar{\theta}_T(\theta))| = o_p(1), \quad \sup_{\theta \in \Theta} |J_T^e(\theta) - J_T^e(\bar{\theta}_T(\theta))| = o_p(1).$$

Since $[e^{-c}, e^c]$ is compact it then follows that $M_{1,T}$ and $M_{3,T}$ are both $o_p(1)$.

Now fix T and for each $\theta \in \Theta$ define:

$$h_{tT}(\theta) = T^{-1} \delta_T^{-(G+\sigma)} \mathcal{K} \left(\frac{\eta_t(A)}{\pi_T} \right), \quad \zeta_{tT}(\theta) = h_{tT}(\theta) - \mathbf{E}[h_{tT}(\theta) | \mathcal{F}_{t-1}],$$

with $\sigma = 0$ so $J_T(\theta) - J_T^e(\theta) = \sum_{t=1}^T \zeta_{tT}(\theta)$. Then $\{(\zeta_{tT}(\theta), \mathcal{F}_t)\}_{t=1}^\infty$ is a martingale difference sequence for any $\theta \in \Theta$ and T , and by Assumption 6:

$$|\zeta_t(\theta)| \leq 2T^{-1} \delta_T^{-G} C,$$

for all θ and t . In addition:

$$\text{Var}(\zeta_t(\theta) | \mathcal{F}_{t-1}) \leq \mathbb{E}[h_t(\theta)^2 | \mathcal{F}_{t-1}] = T^{-2} \delta_T^{-2G} \mathbb{E} \left[\mathcal{K} \left(\frac{\eta_t(A)}{\pi_T} \right)^2 | \mathcal{F}_{t-1} \right],$$

so:

$$\begin{aligned} \mathbb{E} \left[\mathcal{K} \left(\frac{\eta_t(A)}{\pi_T} \right)^2 | \mathcal{F}_{t-1} \right] &= \int_{\mathbb{R}^G} \mathcal{K} \left(\frac{u - (A - A_0) Z_t}{\pi_T} \right)^2 f_t(u | \mathcal{F}_{t-1}) du \\ &= \pi_T^G \int_{\mathbb{R}^G} \mathcal{K}(x)^2 f_t((A - A_0) Z_t + \pi_T x | \mathcal{F}_{t-1}) dx \\ &\leq \delta_T^G e^{G\nu} C^3, \end{aligned}$$

by transformation of variables from u to $s = \frac{u - (A - A_0) Z_t}{\pi_T}$, noting that $\sup_{\gamma \in [e^{-\nu}, e^\nu]} |\gamma| = e^{G\nu}$, that $\int_{\mathbb{R}^G} \mathcal{K}(x)^2 dx \leq C^2$, by Assumption 6, and that $\sup_u f_t(u | \mathcal{F}_{t-1}) \leq C$, by Assumption 3. It follows that:

$$\sum_{t=1}^T \text{Var}(\zeta_t(\theta) | \mathcal{F}_{t-1}) \leq T^{-1} \delta_T^{-G} e^{G\nu} C^3.$$

Lemma 1 implies that the number of elements of Θ_T is less than or equal to $T^{2G(K+1)}d$. Hence, by setting $a = \eta$, $b = T^{-1} \delta_T^{-G} e^{G\nu} C^3$ and $c = 2T^{-1} \delta_T^{-G} C$, Lemma 2 implies that for any fixed $\eta > 0$:

$$\begin{aligned} \Pr \left(\sup_{\theta \in \Theta_T} \left| \sum_{t=1}^T \zeta_{tT}(\theta) \right| \geq \eta \right) &\leq 2T^{2G(K+1)}d \exp \left\{ \frac{-\eta^2/2}{2T^{-1} \delta_T^{-G} C \eta + T^{-1} \delta_T^{-G} e^{G\nu} C^3} \right\} \\ &= 2d \exp \left\{ \left[2G(K+1) \frac{\ln T}{(T \delta_T^G)} - \frac{\eta^2}{4C\eta + 2e^{G\nu} C^3} \right] (T \delta_T^G) \right\} \end{aligned}$$

By Assumption 7, $\ln T / (T \delta_T^G) = o(1)$ which implies that $(T \delta_T^G)^{-1} = o(1)$ and since $J_T(\theta) - J_T^e(\theta) = \sum_{t=1}^T \zeta_{tT}(\theta)$ it follows that:

$$\lim_{T \rightarrow \infty} \Pr \left(\sup_{\theta \in \Theta} |J_T(\bar{\theta}_T(\theta)) - J_T^e(\bar{\theta}_T(\theta))| \geq \eta \right) = 0,$$

for any fixed $\eta > 0$ and hence that:

$$\sup_{\theta \in \Theta} |J_T(\bar{\theta}_T(\theta)) - J_T^e(\bar{\theta}_T(\theta))| = o_p(1).$$

Since $[e^{-\nu}, e^\nu]$ is compact this implies that $M_{2,T}$ is $o_p(1)$.

Last, define $h_{tT}^e(\theta) = \mathbb{E}[h_{tT}(\theta) | \mathcal{F}_{t-1}]$ and observe that $h_{tT}^e(\theta) = \xi_t(\pi_T, A)$, where:

$$\xi_t(\pi, A) = \int_{\mathbb{R}^G} f_t((A - A_0) Z_t + \pi x | \mathcal{F}_{t-1}) \mathcal{K}(x) dx.$$

Assumptions 3 and 6 imply that $\xi_t(\pi, A)$ is continuous in $(\pi, \text{vec}(A))'$, by dominated convergence, and is uniformly bounded from above in absolute value by the non-stochastic term $C \int_{\mathbb{R}^G} |\mathcal{K}(x)| dx < \infty$. Assumption 1 then implies that $\{\xi_t(\pi, A)\}_{t=-\infty}^{\infty}$ is strictly stationary and ergodic and hence it follows by ULLN that $T^{-1} \sum_{t=1}^T \xi_t(\pi, A)$ converges uniformly in probability to $E[\xi_t(\pi, A)]$ on any compact set of values of $(\pi, \text{vec}(A))'$. Since $\gamma^G J_T^e(\theta) = T^{-1} \sum_{t=1}^T \xi_t(\gamma^{-1} \delta_T, A)$ and $\delta_T = o(1)$ it follows that:

$$\sup_{\theta \in \Theta} |\gamma^G J_T^e(\theta) - E[\xi_t(0, A)]| = o_p(1).$$

But:

$$E[\xi_t(0, A)] = E[f_t((A - A_0) Z_t | \mathcal{F}_{t-1})] = Q_0(A),$$

and hence $M_{4,T}$ is $o_p(1)$.

Since $M_{1,T}, \dots, M_{4,T}$ are all $o_p(1)$ it then follows from Equation (A.4) that:

$$\sup_{\theta \in \Theta} |\gamma^G J_T(\theta) - Q_0(A)| = o_p(1).$$

■

Lemma 7 *Under Assumptions 1–13, B_0 is symmetric positive definite and D_0 is symmetric negative definite.*

Proof. First, it is clear that \mathcal{M} exists, by Assumption 12, and that \mathcal{M} is symmetric, by construction. Now, for any fixed $(G \times 1)$ vector $c_1 \neq 0$:

$$\begin{aligned} c_1' \mathcal{M} c_1 &= \int_{\mathbb{R}^G} c_1' \mathcal{K}^{(1)}(x) \mathcal{K}^{(1)}(x)' c_1 dx \\ &= \int_{\mathbb{R}^G} (\mathcal{K}^{(1)}(x)' c_1)^2 dx. \end{aligned}$$

Clearly, $c_1' \mathcal{M} c_1 \geq 0$ for all c_1 with equality if and only if $\mathcal{K}^{(1)}(x)' c_1 = 0$ for almost all x . Since $\mathcal{K}(\cdot)$ is twice differentiable it follows that $\mathcal{K}^{(1)}(\cdot)$ is continuous; hence it follows that $\mathcal{K}^{(1)}(x)' c_1 = 0$ for almost all x if and only if $\mathcal{K}^{(1)}(x)' c_1 = 0$ for all x . Now since $c_1 \neq 0$ we can construct a non-singular $(G \times G)$ matrix \mathcal{C} whose first column is given by c_1 and then define $\tilde{\mathcal{K}}_{\mathcal{C}}(\cdot) : \mathbb{R}^G \rightarrow \mathbb{R}$ such that:

$$\tilde{\mathcal{K}}_{\mathcal{C}}(x) = \mathcal{K}(\mathcal{C}x).$$

It then follows that:

$$\frac{\partial \tilde{\mathcal{K}}_{\mathcal{C}}(x)}{\partial x_1} = \frac{d\mathcal{K}(\mathcal{C}x)}{dx_1} = [\mathcal{K}^{(1)}(\mathcal{C}x)]' \frac{\partial(\mathcal{C}x)}{\partial x_1} = [\mathcal{K}^{(1)}(\mathcal{C}x)]' c_1,$$

since $\mathcal{C}x = \sum_{j=1}^G c_j x_j$ where c_j is the j th column of \mathcal{C} , and hence $c_1' \mathcal{M} c_1 = 0$ if and only if $\frac{\partial \tilde{\mathcal{K}}_{\mathcal{C}}(x)}{\partial x_1} = 0$ for all x . But since \mathcal{C} is non-singular then it follows from Assumption 12 that:

$$\lim_{M \rightarrow \infty} \sup_{x: \|x\| \geq M} \left| \tilde{\mathcal{K}}_{\mathcal{C}}(x) \right| = 0.$$

Consequently $\frac{\partial \tilde{\mathcal{K}}_{\mathcal{C}}(x)}{\partial x_1} = 0$ can only be true for all x if $\tilde{\mathcal{K}}_{\mathcal{C}}(x) = 0$ for all x and hence $\mathcal{K}(x) = 0$ for all x which contradicts Assumption 6. Thus there is no $c_1 \neq 0$ such that $\mathcal{K}^{(1)}(x)' c_1 = 0$ for almost all x and hence there is no $c_1 \neq 0$ such that $c_1' \mathcal{M} c_1 = 0$. It follows that $c_1' \mathcal{M} c_1 > 0$ for all $c_1 \neq 0$ and hence \mathcal{M} must be a symmetric positive definite matrix.

Second, since \mathcal{M} is symmetric, as shown above, then B_0 is also symmetric. Now, fix $A \neq 0$; then:

$$\begin{aligned} \text{vec}(A)' B_0 \text{vec}(A) &= \text{E} \left[f_t(0|\mathcal{F}_{t-1}) \text{vec}(A)' (Z_t \otimes I_G) \mathcal{M} (Z_t \otimes I_G)' \text{vec}(A) \right] \\ &= \text{E} \left[f_t(0|\mathcal{F}_{t-1}) \text{vec}(AZ_t)' \mathcal{M} \text{vec}(AZ_t) \right] \geq 0, \end{aligned}$$

since $\Pr(f_t(0|\mathcal{F}_{t-1}) > 0) = 1$, by Assumption 3, and since \mathcal{M} is positive definite, as established above. In addition, Assumption 5 implies that $\Pr(\text{vec}(AZ_t) = 0) < 1$. Together these imply that:

$$\Pr(f_t(0|\mathcal{F}_{t-1}) \text{vec}(AZ_t)' \mathcal{M} \text{vec}(AZ_t) = 0) < 1,$$

and hence that $\text{vec}(A)' B_0 \text{vec}(A) > 0$ for all $A \neq 0$ which in turns implies that B_0 is positive definite.

Third, since $f_t(u|\mathcal{F}_{t-1})$ is three times differentiable for all $u \in \mathbb{R}^G$, by Assumption 9, then it follows that $f_t^{(2)}(0|\mathcal{F}_{t-1})$ is symmetric and hence that D_0 is also symmetric. Furthermore, $f_t^{(2)}(0|\mathcal{F}_{t-1})$ is negative definite, by Assumption 9. Next, fix $A \in \mathbb{R}^G \times \mathbb{R}^K$ such that $A \neq 0$; then:

$$\begin{aligned} \text{vec}(A)' D_0 \text{vec}(A) &= \text{E} \left[\text{vec}(A)' (Z_t \otimes I_G) f_t^{(2)}(0|\mathcal{F}_{t-1}) (Z_t \otimes I_G)' \text{vec}(A) \right] \\ &= \text{E} \left[\text{vec}(AZ_t)' f_t^{(2)}(0|\mathcal{F}_{t-1}) \text{vec}(AZ_t) \right]. \end{aligned}$$

Since $\Pr(\text{vec}(AZ_t) = 0) < 1$, by Assumption 5, then:

$$\Pr\left(\text{vec}(AZ_t)' f_t^{(2)}(0 | \mathcal{F}_{t-1}) \text{vec}(AZ_t) = 0\right) < 1,$$

and hence $\text{vec}(A)' D_0 \text{vec}(A) < 0$ for all $A \neq 0$ which in turn implies that D_0 is negative definite. ■

Lemma 8 *Suppose that Assumptions 1–14 are satisfied. Let $\theta = (\gamma, A)$ where $0 < \gamma < \infty$ and $A \in \mathcal{A}$. Then:*

$$\lim_{T \rightarrow \infty} \mathbb{E}[D_T(\theta)] = D(\theta) = \mathbb{E}\left[(Z_t \otimes I_G) f_t^{(2)}((A - A_0) Z_t | \mathcal{F}_{t-1}) (Z_t \otimes I_G)'\right],$$

where $D_T(\theta) = \gamma^{G+1} \left[\frac{\partial R_T(\theta)}{\partial \text{vec}(A)'} \right]$ and $R_T(\cdot)$ is defined as in the proof of Theorem 3. In addition, $D(\theta)$ is continuous.

Proof. For any fixed $0 < \gamma < \infty$ define $\pi_T = \delta_T / \gamma$. Observe that:

$$\left[\frac{\partial R_T(\theta)}{\partial \text{vec}(A)'} \right] = \gamma^{T-1} \sum_{t=1}^T \delta_T^{-(G+2)} (Z_t \otimes I_G) \mathcal{K}^{(2)} \left(\frac{\eta_t(A)}{\pi_T} \right) (Z_t \otimes I_G)',$$

where $\eta_t(A) = (Y_t - AZ_t)$, so by stationarity and the law of iterated expectations it follows that:

$$\begin{aligned} \mathbb{E}[D_T(\theta)] &= \pi_T^{-(G+2)} T^{-1} \sum_{t=1}^T \mathbb{E}\left[(Z_t \otimes I_G) \mathcal{K}^{(2)} \left(\frac{\eta_t(A)}{\pi_T} \right) (Z_t \otimes I_G)'\right] \\ &= \pi_T^{-(G+2)} \mathbb{E}\left\{(Z_t \otimes I_G) \mathbb{E}\left[\mathcal{K}^{(2)} \left(\frac{\eta_t(A)}{\pi_T} \right) | \mathcal{F}_{t-1} \right] (Z_t \otimes I_G)'\right\}. \end{aligned}$$

Using transformation of variables and repeated integration by parts it follows from Assumptions 3, 9 and 12:

$$\mathbb{E}\left[\mathcal{K}^{(2)} \left(\frac{\eta_t(A)}{\pi_T} \right) | \mathcal{F}_{t-1} \right] = \pi_T^{(G+2)} \int_{\mathbb{R}^G} \mathcal{K}(x) f_t^{(2)}((A - A_0) Z_t + \pi_T x | \mathcal{F}_{t-1}) dx,$$

where $\pi_T = \delta_T / \gamma$, and hence that:

$$\mathbb{E}[D_T(\theta)] = \mathbb{E}\left[(Z_t \otimes I_G) \int_{\mathbb{R}^G} \mathcal{K}(x) f_t^{(2)}((A - A_0) Z_t + \pi_T x | \mathcal{F}_{t-1}) dx (Z_t \otimes I_G)'\right].$$

Assumptions 3, 6 and 7 then imply by dominated convergence that:

$$D(\theta) = \lim_{T \rightarrow \infty} \mathbb{E}[D_T(\theta)] = \mathbb{E}\left[(Z_t \otimes I_G)' f_t^{(2)}((A - A_0) Z_t | \mathcal{F}_{t-1}) (Z_t \otimes I_G)'\right],$$

and also that $D(\theta)$ is continuous. ■

Lemma 9 Suppose that Assumptions 1–14 are satisfied. Let $0 < \nu < \infty$ be a scalar constant and define $\Theta = [e^{-\nu}, e^{\nu}] \times \text{vec}(\mathcal{A})$; then:

$$\sup_{\theta \in \Theta} \|D_T(\theta) - D(\theta)\| = o_p(1),$$

where $\theta = (\gamma, \text{vec}(A)')'$, and where $D_T(\theta)$ and $D(\theta)$ are given as in Lemma 8.

Proof. Let $\pi_T = \delta_T/\gamma$ and observe that:

$$D_T(\theta) = T^{-1} \sum_{t=1}^T \pi_T^{-(G+2)} (Z_t \otimes I_G) \mathcal{K}^{(2)} \left(\frac{\eta_t(A)}{\pi_T} \right) (Z_t \otimes I_G)',$$

so each element of $D_T(\theta)$ can be expressed as a linear combination with fixed coefficients (that do not depend on T or θ) of terms of the form:

$$T^{-1} \sum_{t=1}^T \pi_T^{-(G+2)} \mathcal{K}_{rs}^{(2)} \left(\frac{\eta_t(A)}{\pi_T} \right) Z_{it} Z_{jt},$$

where $\mathcal{K}_{rs}^{(2)}(\cdot)$ is the (r, s) th element of $\mathcal{K}_{rs}^{(2)}(\cdot)$, and $1 \leq i, j \leq K$ and $1 \leq r, s \leq G$. By the same line of proof used to establish Lemma 8 it follows that $D(\theta)$ is the corresponding linear combination of terms of the form:

$$\mathbb{E} \left[f_{t,rs}^{(2)}((A - A_0) Z_t | \mathcal{F}_{t-1}) Z_{it} Z_{jt} \right],$$

where $f_{t,rs}^{(2)}(\cdot | \mathcal{F}_{t-1})$ is the (r, s) th element of $f_t^{(2)}(\cdot | \mathcal{F}_{t-1})$. Since $[e^{-\nu}, e^{\nu}]$ is compact it suffices to establish that for every fixed $1 \leq i, j \leq K$ and $1 \leq r, s \leq G$:

$$\sup_{\theta \in \Theta} \left| T^{-1} \sum_{t=1}^T \delta_T^{-(G+2)} \mathcal{K}_{rs}^{(2)} \left(\frac{\eta_t(A)}{\pi_T} \right) Z_{it} Z_{jt} - \gamma^{-(G+2)} \mathbb{E} \left[f_{t,rs}^{(2)}((A - A_0) Z_t | \mathcal{F}_{t-1}) Z_{it} Z_{jt} \right] \right| = o_p(1). \quad (\text{A.5})$$

Now fix $1 \leq i, j \leq K$ and $1 \leq r, s \leq G$ and define:

$$\begin{aligned} h_{tT}(\theta) &= T^{-1} \delta_T^{-(G+2)} \mathcal{K}_{rs}^{(2)} \left(\frac{\eta_t(A)}{\pi_T} \right) Z_{it} Z_{jt}, \\ h_{tT}^\dagger(\theta) &= T^{-1} \delta_T^{-(G+2)} \mathcal{K}_{rs}^{(2)} \left(\frac{\eta_t(A)}{\pi_T} \right) Z_{it} Z_{jt} \chi(|Z_{it} Z_{jt}| \geq \delta_T^{-2}), \end{aligned}$$

and then define:

$$\begin{aligned} h_{tT}^e(\theta) &= \mathbb{E}[h_{tT}(\theta) | \mathcal{F}_{t-1}], & h_{tT}^{\dagger,e}(\theta) &= \mathbb{E}[h_{tT}^\dagger(\theta) | \mathcal{F}_{t-1}], \\ h_{tT}^*(\theta) &= h_{tT}(\theta) - h_{tT}^\dagger(\theta), & h_{tT}^{*,e}(\theta) &= \mathbb{E}[h_{tT}^*(\theta) | \mathcal{F}_{t-1}] \end{aligned}$$

and:

$$\begin{aligned}
H_T(\theta) &= \sum_{t=1}^T h_{tT}(\theta), & H_T^\dagger(\theta) &= \sum_{t=1}^T h_{tT}^\dagger(\theta), & H_T^*(\theta) &= \sum_{t=1}^T h_{tT}^*(\theta), \\
H_T^e(\theta) &= \sum_{t=1}^T h_{tT}^e(\theta), & H_T^{\dagger,e}(\theta) &= \sum_{t=1}^T h_{tT}^{\dagger,e}(\theta), & H_T^{*,e}(\theta) &= \sum_{t=1}^T h_{tT}^{*,e}(\theta), \\
H_0(\theta) &= \lim_{T \rightarrow \infty} \mathbb{E}[H_T(\theta)] = \lim_{T \rightarrow \infty} \mathbb{E}[H_T^e(\theta)].
\end{aligned}$$

In addition, let Θ_T and $\bar{\theta}_T(\cdot)$ be characterized as in Lemma 1. Then:

$$\begin{aligned}
H_T(\theta) - H_0(\theta) &= [H_T(\theta) - H_T(\bar{\theta}_T(\theta))] + H_T^\dagger(\bar{\theta}_T(\theta)) + [H_T^*(\bar{\theta}_T(\theta)) - H_T^{*,e}(\bar{\theta}_T(\theta))] \\
&\quad - H_T^{\dagger,e}(\bar{\theta}_T(\theta)) + [H_T^e(\bar{\theta}_T(\theta)) - H_T^e(\theta)] + [H_T^e(\theta) - H_0(\theta)],
\end{aligned}$$

and hence by T it follows that:

$$\sup_{\theta \in \Theta} |\gamma H_T(\theta) - \gamma H_0(\theta)| \leq M_{1,T} + M_{2,T} + M_{3,T} + M_{4,T} + M_{5,T} + M_{6,T}, \quad (\text{A.6})$$

where:

$$\begin{aligned}
M_{1,T} &= \sup_{\theta \in \Theta} \gamma |H_T(\theta) - H_T(\bar{\theta}_T(\theta))|, & M_{2,t} &= \sup_{\theta \in \Theta_T} \gamma |H_T^\dagger(\theta)|, \\
M_{3,T} &= \sup_{\theta \in \Theta_T} \gamma |H_T^*(\theta) - H_T^{*,e}(\theta)|, & M_{4,t} &= \sup_{\theta \in \Theta_T} \gamma |H_T^{\dagger,e}(\theta)|, \\
M_{5,T} &= \sup_{\theta \in \Theta} \gamma |H_T^e(\bar{\theta}_T(\theta)) - H_T^e(\theta)|, & M_{6,t} &= \sup_{\theta \in \Theta} \gamma |H_T^e(\theta) - H_0(\theta)|.
\end{aligned}$$

First, observe that $H_T(\theta)$ and $H_T^e(\theta)$ can be expressed in the form:

$$\begin{aligned}
H_T(\theta) &= T^{-1} \sum_{t=1}^T \delta_T^{-(G+\sigma)} \psi \left(\frac{\eta_t(A)}{\pi_T} \right) m(W_t), \\
H_T^e(\theta) &= T^{-1} \sum_{t=1}^T \delta_T^{-(G+\sigma)} \mathbb{E} \left[\psi \left(\frac{\eta_t(A)}{\pi_T} \right) m(W_t) \mid \mathcal{F}_{t-1} \right],
\end{aligned}$$

where $\sigma = 2$, $\psi(x) = \mathcal{K}_{rs}^{(2)}(x)$ and $m(W_t) = Z_{it}Z_{jt}$. Since $\sup_x \|\mathcal{K}_{rs}^{(3)}(x)\| < \infty$, by Assumption 12, $|Z_{it}Z_{jt}| \leq \|W_t\|^2$ by CS, $\mathbb{E}\{\|W_t\|^3\} < \infty$, by Assumption 11, and $\delta_T = o(1)$, $T^{-2}\delta_T^{-(G+3)} = o(1)$, by Assumption 13, then it follows that $M_{1,T}$ and $M_{5,T}$ are both $o_p(1)$ by Lemma 3.

Second, observe that $H_T^\dagger(\theta)$ and $H_T^{\dagger,e}(\theta)$ then can be written as:

$$\begin{aligned} H_T^\dagger(\theta) &= T^{-1} \sum_{t=1}^T \delta_T^{-(G+\sigma)} \psi\left(\frac{\eta_t(A)}{\pi_T}\right) m_T^\dagger(W_t), \\ H_T^{\dagger,e}(\theta) &= T^{-1} \sum_{t=1}^T \delta_T^{-(G+\sigma)} \mathbb{E}\left[\psi\left(\frac{\eta_t(A)}{\pi_T}\right) m_T^\dagger(W_t) \mid \mathcal{F}_{t-1}\right], \end{aligned}$$

where $m_T^\dagger(W_t) = m(W_t) \chi(|m(W_t)| > \delta_T^{-2})$. Fix r such that :

$$(G+4) < 2r \leq (G+4+\tau),$$

and note that this then implies $r > (G+\sigma+2)/2$ since $\sigma = 2$ so:

$$\mathbb{E}\{|m(W_t)|^r\} \leq \mathbb{E}\{\|W_t\|^{2r}\} \leq \mathbb{E}\{\|W_t\|^{G+4+\tau}\} < \infty,$$

by Assumption 11. Since $\sup_x \|\mathcal{K}_{rs}^{(3)}(x)\| < \infty$, by Assumption 12, and $\delta_T = o(1)$, by Assumption 7, then it follows that $M_{2,T}$ and $M_{4,T}$ are both $o_p(1)$ by Lemma 4, noting that $\Theta_T \subseteq \Theta$.

Now set $\zeta_{tT}(\theta) = h_{tT}^*(\theta) - h_{tT}^{*,e}(\theta)$ which implies that:

$$\zeta_{tT}(\theta) = T^{-1} \delta_T^{-(G+2)} \left\{ \psi\left(\frac{\eta_t(A)}{\pi_T}\right) m_T^*(W_t) - \mathbb{E}\left[\psi\left(\frac{\eta_t(A)}{\pi_T}\right) m_T^*(W_t) \mid \mathcal{F}_{t-1}\right] \right\}.$$

Since $|\psi(x)| = |\mathcal{K}_{rs}^{(2)}(x)| \leq C$ for all x , by Assumption 12, and $|m_T^*(W_t)| \leq \delta_T^{-2}$ for all W_t , then:

$$|\zeta_{tT}(\theta)| \leq 2T^{-1} \delta_T^{-(G+4)} C.$$

Now:

$$\begin{aligned} \text{Var}(\zeta_{tT}(\theta) \mid \mathcal{F}_{t-1}) &\leq \mathbb{E}[h_{tT}^*(\theta)^2 \mid \mathcal{F}_{t-1}] \leq \mathbb{E}[h_{tT}(\theta)^2 \mid \mathcal{F}_{t-1}] \\ &= T^{-2} \delta_T^{-(2G+4)} \mathbb{E}\left[\mathcal{K}_{rs}^{(2)}\left(\frac{\eta_t(A)}{\pi_T}\right)^2 \mid \mathcal{F}_{t-1}\right] (Z_{it} Z_{jt})^2, \end{aligned}$$

and by transformation of variables:

$$\begin{aligned} \mathbb{E}\left[\mathcal{K}_{rs}^{(2)}\left(\frac{\eta_t(A)}{\pi_T}\right)^2 \mid \mathcal{F}_{t-1}\right] &= \int_{\mathbb{R}^G} \mathcal{K}_{rs}^{(2)}\left(\frac{u - AZ_t}{\pi_T}\right)^2 f_t(u \mid \mathcal{F}_{t-1}) du \\ &= \pi_T^G \int_{\mathbb{R}^G} \mathcal{K}_{rs}^{(2)}(x)^2 f_t((A - A_0)Z_t + \pi_T x \mid \mathcal{F}_{t-1}) dx \leq \gamma^{-G} \delta_T^G C^2, \end{aligned}$$

since $\sup_u f_t(u|\mathcal{F}_{t-1}) \leq C$, by Assumption 3, so:

$$\text{Var}(\zeta_{tT}(\theta) | \mathcal{F}_{t-1}) \leq \gamma^{-G} T^{-2} \delta_T^{-(G+4)} C^2 (Z_{it} Z_{jt})^2 \leq T^{-2} \delta_T^{-(G+4)} e^{G\nu} C^2 \|Z_t\|^2,$$

by CS, since $\gamma^{-G} \leq e^{G\nu}$ for all $\gamma \in [e^{-\nu}, e^\nu]$, and thus:

$$\sup_{\theta \in \Theta_T} \sum_{t=1}^T \text{Var}(\zeta_{tT}(\theta) | \mathcal{F}_{t-1}) \leq T^{-1} \delta_T^{-(G+4)} e^{G\nu} C^2 \left[T^{-1} \sum_{t=1}^T \|Z_t\|^2 \right].$$

Now $\mathbb{E} \{ \|Z_t\|^2 \} \leq \mu_4 = \mathbb{E} \{ \|W_t\|^4 \} < \infty$, by Assumption 11, and hence:

$$\sup_{\theta \in \Theta_T} \sum_{t=1}^T \text{Var}(\zeta_{tT}(\theta) | \mathcal{F}_{t-1}) \leq T^{-1} \delta_T^{-(G+4)} e^{G\nu} C^2 [\mu_4 + o_p(1)],$$

by the ergodic theorem. This implies that:

$$\Pr \left(\sup_{\theta \in \Theta_T} \sum_{t=1}^T \text{Var}(\zeta_{tT}(\theta) | \mathcal{F}_{t-1}) > 2T^{-1} \delta_T^{-(G+4)} \bar{C}_1 \right) = o_p(1),$$

where $\bar{C}_1 = e^{G\nu} C^2 \mu_4$. In addition, Lemma 1 implies that the number of elements of Θ_T is less than or equal to $T^{2G(K+1)}d$. Hence, by setting $a = \eta$, $b = 2T^{-1} \delta_T^{-(G+4)} \bar{C}_1$ and $c = 2T^{-1} \delta_T^{-(G+4)} C$, Lemma 2 implies that for any fixed $\eta > 0$:

$$\begin{aligned} \Pr \left(\sup_{\theta \in \Theta_T} \left| \sum_{t=1}^T \zeta_{tT}(\theta) \right| \geq \eta \right) &\leq 2T^{2G(K+1)}d \exp \left\{ \frac{-\eta^2 T \delta_T^{(G+4)}}{4C\eta + 4\bar{C}_1} \right\} + o(1) \\ &= 2d \exp \left\{ \left[2G(K+1) \frac{\ln T}{T \delta_T^{(G+4)}} - \frac{\eta^2}{4C\eta + 4\bar{C}_1} \right] T \delta_T^{(G+4)} \right\} + o(1) \\ &= o(1), \end{aligned}$$

since $\frac{\ln T}{T \delta_T^{G+4}} = o(1)$ and $T^{-1} \delta_T^{-(G+4)} = o(1)$ by Assumption 13. But:

$$M_{3,T} = \sup_{\theta \in \Theta_T} |H_T^*(\theta) - H_T^{*,e}(\theta)| = \sup_{\theta \in \Theta_T} \left| \sum_{t=1}^T \zeta_{tT}(\theta) \right|,$$

and thus $M_{3,T} = o_p(1)$.

Last, observe that:

$$\begin{aligned} h_T^e(\theta) &= T^{-1} \delta_T^{-(G+2)} \mathbb{E} \left[\mathcal{K}_{rs}^{(2)} \left(\frac{\gamma \eta_t(A)}{\delta_T} \right) Z_{it} Z_{jt} | \mathcal{F}_{t-1} \right] \\ &= T^{-1} \delta_T^{-(G+2)} \mathbb{E} \left[\mathcal{K}_{rs}^{(2)} \left(\frac{\gamma \eta_t(A)}{\delta_T} \right) | \mathcal{F}_{t-1} \right] Z_{it} Z_{jt}, \end{aligned}$$

and thus:

$$H_T^e(\theta) = T^{-1} \sum_{t=1}^T Z_{it} Z_{jt} \delta_T^{-(G+2)} \mathbb{E} \left[\mathcal{K}_{rs}^{(2)} \left(\frac{\gamma \eta_t(A)}{\delta_T} \right) | \mathcal{F}_{t-1} \right]. \quad (\text{A.7})$$

Using transformation of variables and repeated integration by parts it follows from Assumptions 3, 9 and 12 that:

$$\mathbb{E} \left[\mathcal{K}_{rs}^{(2)} \left(\frac{\eta_t(A)}{\pi_T} \right) | \mathcal{F}_{t-1} \right] = \pi_T^{(G+2)} \int_{\mathbb{R}^G} \mathcal{K}(x) f_{t,rs}^{(2)}((A - A_0) Z_t + \pi_T x | \mathcal{F}_{t-1}) dx,$$

so:

$$H_T^e(\theta) = \gamma^{-(G+2)} T^{-1} \sum_{t=1}^T Z_{it} Z_{jt} \int_{\mathbb{R}^G} \mathcal{K}(x) f_{t,rs}^{(2)}((A - A_0) Z_t + \pi_T x | \mathcal{F}_{t-1}) dx. \quad (\text{A.8})$$

By Assumptions 6 and 9 then $Z_{it} Z_{jt} \int_{\mathbb{R}^G} \mathcal{K}(s) f_{t,rs}^{(2)}((A - A_0) Z_t + \pi x | \mathcal{F}_{t-1}) dx$ is a continuous function of A and π and is uniformly bounded in absolute value by $C^2 |Z_{it} Z_{jt}|$. Since $\mathbb{E}\{|Z_{it} Z_{jt}|\} < \infty$, by Assumption 11, then for any fixed $0 < \bar{\pi} < \infty$ it follows by ULLN that:

$$T^{-1} \sum_{t=1}^T Z_{it} Z_{jt} \int_{\mathbb{R}^G} \mathcal{K}(x) f_{t,rs}^{(2)}((A - A_0) Z_t + \pi x | \mathcal{F}_{t-1}) dx \quad (\text{A.9})$$

converges uniformly in probability to:

$$\mathbb{E} \left[Z_{it} Z_{jt} \int_{\mathbb{R}^G} \mathcal{K}(x) f_{t,rs}^{(2)}((A - A_0) Z_t + \pi x | \mathcal{F}_{t-1}) dx \right], \quad (\text{A.10})$$

over $(\pi, \text{vec}(A)')' \in [-\bar{\pi}, \bar{\pi}] \times \text{vec}(\mathcal{A})$. Since $\delta_T = o(1)$ and since $\theta \in \Theta$ requires that $\gamma \in [e^{-\nu}, e^\nu]$ then $\sup_{\theta \in \Theta} |\pi_T| = o(1)$ and hence this implies that $H_T^e(\theta)$ converges uniformly in probability to:

$$H_0(\theta) = \gamma^{-(G+2)} \mathbb{E} \left[f_{t,rs}^{(2)}((A - A_0) Z_t | \mathcal{F}_{t-1}) Z_{it} Z_{jt} \right] \quad (\text{A.11})$$

over $\theta \in \Theta$ and thus that $M_{6,T} = o_p(1)$. Combining this with Equation (A.6) and the results shown earlier that $M_{1,T}, \dots, M_{5,T}$ are all $o_p(1)$ this then implies that Equation (A.5) holds. This in turn establishes the desired result. ■

Lemma 10 *Suppose that Assumptions 1–14 are satisfied. Let $\theta = (\gamma, \text{vec}(A)')'$ where $0 < \gamma < \infty$ and $A \in \mathcal{A}$. Then:*

$$\lim_{T \rightarrow \infty} \mathbb{E}[P_T(\theta)] = P(\theta) = (G+1) \mathbb{E} \left[(Z_t \otimes I_G) f_t^{(1)}((A - A_0) Z_t | \mathcal{F}_{t-1}) \right],$$

where $P_T(\theta) = \gamma^{G+1} \left[\frac{\partial R_T(\theta)}{\partial \gamma} \right]$ and $R_T(\cdot)$ is defined as in the proof of Theorem 3. In addition, $P(A)$ is continuous

Proof. Fix $\gamma > 0$ and $A \in \mathcal{A}$. Then observe that:

$$P_T(\theta) = \gamma^{G+1} \left[\frac{\partial R_T(\theta)}{\partial \gamma} \right] = T^{-1} \sum_{t=1}^T \pi_T^{-(G+2)} (Z_t \otimes I_G) \mathcal{K}^{(2)} \left(\frac{\eta_t(A)}{\pi_T} \right) \eta_t(A),$$

where $\pi_T = \delta_T/\gamma$ and $\eta_t(A) = (Y_t - AZ_t)$, so by stationarity it follows that:

$$\begin{aligned} \mathbb{E}[P_T(\theta)] &= \pi_T^{-(G+2)} \mathbb{E} \left\{ (Z_t \otimes I_G) \mathcal{K}^{(2)} \left(\frac{\eta_t(A)}{\pi_T} \right) \eta_t(A) \right\} \\ &= \pi_T^{-(G+2)} \mathbb{E} \left\{ (Z_t \otimes I_G) \mathbb{E} \left[\mathcal{K}^{(2)} \left(\frac{\eta_t(A)}{\pi_T} \right) \eta_t(A) \mid \mathcal{F}_{t-1} \right] \right\}. \end{aligned} \quad (\text{A.12})$$

Note that since γ is fixed then $\{\pi_T\}_{t=1}^\infty$ satisfies the same conditions as $\{\delta_T\}_{t=1}^\infty$.

Then observe that:

$$\begin{aligned} \mathbb{E} \left[\mathcal{K}^{(2)} \left(\frac{\eta_t(A)}{\pi_T} \right) \eta_t(A) \mid \mathcal{F}_{t-1} \right] &= \mathbb{E} \left[\mathcal{K}^{(2)} \left(\frac{U_t - (A - A_0) Z_t}{\pi_T} \right) (U_t - (A - A_0) Z_t) \mid \mathcal{F}_{t-1} \right] \\ &= \int_{\mathbb{R}^G} \mathcal{K}^{(2)} \left(\frac{u - \lambda_t}{\pi_T} \right) (u - \lambda_t) f_t(u \mid \mathcal{F}_{t-1}) du, \end{aligned}$$

where $\lambda_t = (A - A_0) Z_t$. Hence for each $i = 1, \dots, G$, the i th element of $\mathbb{E} \left[\mathcal{K}^{(2)} \left(\frac{\eta_t(A)}{\pi_T} \right) \eta_t(A) \mid \mathcal{F}_{t-1} \right]$ can be expressed as:

$$\left(\mathbb{E} \left[\mathcal{K}^{(2)} \left(\frac{\eta_t(A)}{\pi_T} \right) \eta_t(A) \mid \mathcal{F}_{t-1} \right] \right)_i = \sum_{j=1}^G \mathbb{E} \left[\mathcal{K}_{ij}^{(2)} \left(\frac{\eta_t(A)}{\pi_T} \right) \eta_{t,j}(A) \mid \mathcal{F}_{t-1} \right],$$

where $\eta_{t,j}(A)$ is the j th element of $\eta_t(A)$. Now:

$$\begin{aligned} \mathbb{E} \left[\mathcal{K}_{ij}^{(2)} \left(\frac{\eta_t(A)}{\pi_T} \right) \eta_{t,j}(A) \mid \mathcal{F}_{t-1} \right] &= \int_{\mathbb{R}^G} \mathcal{K}_{ij}^{(2)} \left(\frac{u - \lambda_t}{\pi_T} \right) (u_j - \lambda_{t,j}) f_t(u \mid \mathcal{F}_{t-1}) du \\ &= \int_{\mathbb{R}^G} \mathcal{K}_{ij}^{(2)} \left(\frac{u - \lambda_t}{\pi_T} \right) (u_j - \lambda_{t,j}) f_t(u \mid \mathcal{F}_{t-1}) du_i du_{-i}, \end{aligned}$$

where $\lambda_{t,j}$ is the j th element of λ_t and u_{-i} consists of all the elements of u aside from u_i .

Using integration by parts we have that:

$$\begin{aligned} \int_{\mathbb{R}} \mathcal{K}_{ij}^{(2)} \left(\frac{u - \lambda_t}{\pi_T} \right) (u_j - \lambda_{t,j}) f_t(u \mid \mathcal{F}_{t-1}) du_i &= -\pi_T \int_{\mathbb{R}} \mathcal{K}_j^{(1)} \left(\frac{u - \lambda_t}{\pi_T} \right) \\ &\quad \times \frac{\partial \{(u_j - \lambda_{t,j}) f_t(u \mid \mathcal{F}_{t-1})\}}{\partial u_i} du_i, \end{aligned}$$

noting that:

$$\left[\pi_T \mathcal{K}_j^{(1)} \left(\frac{u - \lambda_t}{\pi_T} \right) (u_j - \lambda_{t,j}) f_t(u | \mathcal{F}_{t-1}) \right]_{u_i = -\infty}^{u_i = \infty} = 0,$$

since:

$$\left| \mathcal{K}_j^{(1)} \left(\frac{u - \lambda_t}{\pi_T} \right) (u_j - \lambda_{t,j}) f_t(u | \mathcal{F}_{t-1}) \right| \leq L_0 \left| \mathcal{K}_j^{(1)} \left(\frac{u - \lambda_t}{\pi_T} \right) (u_j - \lambda_{t,j}) \right|,$$

by Assumption 3, and since for all fixed u_{-j} :

$$\lim_{u_i \rightarrow \pm\infty} \left| \mathcal{K}_j^{(1)} \left(\frac{u - \lambda_t}{\pi_T} \right) (u_j - \lambda_{t,j}) \right| = 0,$$

by Assumption 12. Hence:

$$\begin{aligned} \int_{\mathbb{R}^G} \mathcal{K}_{ij}^{(2)} \left(\frac{u - \lambda_t}{\pi_T} \right) (u_j - \lambda_{t,j}) f_t(u | \mathcal{F}_{t-1}) du &= -\pi_T \int_{\mathbb{R}^G} \mathcal{K}_j^{(1)} \left(\frac{u - \lambda_t}{\pi_T} \right) \\ &\quad \times \frac{\partial \{(u_j - \lambda_{t,j}) f_t(u | \mathcal{F}_{t-1})\}}{\partial u_i} du_i du_{-i} \\ &= -\pi_T \int_{\mathbb{R}^G} \mathcal{K}_j^{(1)} \left(\frac{u - \lambda_t}{\pi_T} \right) \\ &\quad \times \frac{\partial \{(u_j - \lambda_{t,j}) f_t(u | \mathcal{F}_{t-1})\}}{\partial u_i} du_j du_{-j}. \end{aligned}$$

Repeating integration by parts gives:

$$\begin{aligned} \int_{\mathbb{R}^G} \mathcal{K}_j^{(1)} \left(\frac{u - \lambda_t}{\pi_T} \right) \frac{\partial \{(u_j - \lambda_{t,j}) f_t(u | \mathcal{F}_{t-1})\}}{\partial u_i} du_j &= -\pi_T \int_{\mathbb{R}} \mathcal{K} \left(\frac{u - \lambda_t}{\pi_T} \right) \\ &\quad \times \frac{\partial^2 \{(u_j - \lambda_{t,j}) f_t(u | \mathcal{F}_{t-1})\}}{\partial u_i u_j} du_j, \end{aligned}$$

noting that:

$$\left[\pi_T \mathcal{K}_j^{(1)} \left(\frac{u - \lambda_t}{\pi_T} \right) \frac{\partial \{(u_j - \lambda_{t,j}) f_t(u | \mathcal{F}_{t-1})\}}{\partial u_i} \right]_{u_j = -\infty}^{u_j = \infty} = 0,$$

since:

$$\frac{\partial \{(u_j - \lambda_{t,j}) f_t(u | \mathcal{F}_{t-1})\}}{\partial u_i} = \chi_{ij} f_t(u | \mathcal{F}_{t-1}) + (u_j - \lambda_{t,j}) f_{t,i}^{(1)}(u | \mathcal{F}_{t-1}),$$

where $\chi_{ij} = 1$ if $i = j$ and $\chi_{ij} = 0$ if $i \neq r$, so:

$$\begin{aligned} \left| \mathcal{K}_j^{(1)} \left(\frac{u - \lambda_t}{\pi_T} \right) \frac{\partial \{(u_j - \lambda_{t,j}) f_t(u | \mathcal{F}_{t-1})\}}{\partial u_i} \right| &\leq \chi_{ij} C \left| \mathcal{K}_j^{(1)} \left(\frac{u - \lambda_t}{\pi_T} \right) \right| \\ &\quad + C \left| (u_j - \lambda_{t,j}) \mathcal{K}_j^{(1)} \left(\frac{u - \lambda_t}{\pi_T} \right) \right|, \end{aligned}$$

by Assumptions 3 and 9, and hence:

$$\lim_{u_i \rightarrow \pm\infty} \left| \mathcal{K}_j^{(1)} \left(\frac{u - \lambda_t}{\pi_T} \right) \left[\chi_{ij} f_t(u|\mathcal{F}_{t-1}) + (u_j - \lambda_{t,j}) f_{t,i}^{(1)}(u|\mathcal{F}_{t-1}) \right] \right| = 0,$$

by Assumption 12. Hence it follows that:

$$\begin{aligned} \int_{\mathbb{R}^G} \mathcal{K}_{ij}^{(2)} \left(\frac{u - \lambda_t}{\pi_T} \right) (u_j - \lambda_{t,j}) f_t(u|\mathcal{F}_{t-1}) du &= \pi_T^2 \int_{\mathbb{R}^G} \mathcal{K} \left(\frac{u - \lambda_t}{\pi_T} \right) \\ &\quad \times \frac{\partial^2 \{(u_j - \lambda_{t,j}) f_t(u|\mathcal{F}_{t-1})\}}{\partial u_i u_j} du. \end{aligned}$$

Now:

$$\frac{\partial^2 \{(u_j - \lambda_{t,j}) f_t(u|\mathcal{F}_{t-1})\}}{\partial u_i u_j} = \chi_{ij} f_{t,j}^{(1)}(u|\mathcal{F}_{t-1}) + f_{t,i}^{(1)}(u|\mathcal{F}_{t-1}) + (u_j - \lambda_{t,j}) f_{t,ij}^{(2)}(u|\mathcal{F}_{t-1}),$$

and thus:

$$\begin{aligned} \mathbb{E} \left[\mathcal{K}_{ij}^{(2)} \left(\frac{\eta_t(A)}{\pi_T} \right) \eta_{t,j}(A) | \mathcal{F}_{t-1} \right] &= \pi_T^2 \int_{\mathbb{R}^G} \mathcal{K}_{ij}^{(2)} \left(\frac{u - \lambda_t}{\pi_T} \right) (u_j - \lambda_{t,j}) f_t(u|\mathcal{F}_{t-1}) du \\ &= \chi_{ij} \pi_T^2 \int_{\mathbb{R}^G} \mathcal{K} \left(\frac{u - \lambda_t}{\pi_T} \right) f_{t,j}^{(1)}(u|\mathcal{F}_{t-1}) du \\ &\quad + \pi_T^2 \int_{\mathbb{R}^G} \mathcal{K} \left(\frac{u - \lambda_t}{\pi_T} \right) f_{t,i}^{(1)}(u|\mathcal{F}_{t-1}) du \\ &\quad + \pi_T^2 \int_{\mathbb{R}^G} \mathcal{K} \left(\frac{u - \lambda_t}{\pi_T} \right) f_{t,ij}^{(2)}(u|\mathcal{F}_{t-1}) (u_j - \lambda_{t,j}) du. \end{aligned} \tag{A.13}$$

By transformation of variables from u to $x = \frac{u - \lambda_t}{\pi_T}$ it follows that:

$$\begin{aligned} \int_{\mathbb{R}^G} \mathcal{K} \left(\frac{u - \lambda_t}{\pi_T} \right) f_{t,j}^{(1)}(u|\mathcal{F}_{t-1}) du &= \pi_T^G \int_{\mathbb{R}^G} \mathcal{K}(x) f_{t,j}^{(1)}(\lambda_t + \pi_T x | \mathcal{F}_{t-1}) dx, \\ \int_{\mathbb{R}^G} \mathcal{K} \left(\frac{u - \lambda_t}{\pi_T} \right) f_{t,ij}^{(2)}(u|\mathcal{F}_{t-1}) (u_j - \lambda_{t,j}) du &= \pi_T^{G+1} \int_{\mathbb{R}^G} \mathcal{K}(x) f_{t,ij}^{(2)}(\lambda_t + \pi_T x | \mathcal{F}_{t-1}) x_j dx, \end{aligned}$$

and hence that:

$$\begin{aligned} \mathbb{E} \left[\mathcal{K}_{ij}^{(2)} \left(\frac{\eta_t(A)}{\pi_T} \right) \eta_{t,j}(A) | \mathcal{F}_{t-1} \right] &= \chi_{ij} \pi_T^{G+2} \int_{\mathbb{R}^G} \mathcal{K}(x) f_{t,j}^{(1)}(\lambda_t + \pi_T x | \mathcal{F}_{t-1}) dx \\ &\quad + \pi_T^{G+2} \int_{\mathbb{R}^G} \mathcal{K}(x) f_{t,i}^{(1)}(\lambda_t + \pi_T x | \mathcal{F}_{t-1}) dx \\ &\quad + \pi_T^{G+3} \int_{\mathbb{R}^G} \mathcal{K}(x) f_{t,ij}^{(2)}(\lambda_t + \pi_T x | \mathcal{F}_{t-1}) x_j dx. \end{aligned} \tag{A.14}$$

Substituting this into Equation (A.12) then implies that:

$$\begin{aligned} \mathbb{E}[P_T(\theta)] &= (G+1) \mathbb{E} \left\{ (Z_t \otimes I_G) \int_{\mathbb{R}^G} \mathcal{K}(x) f_t^{(1)}((A - A_0) Z_t + \pi_T x | \mathcal{F}_{t-1}) dx \right\} \\ &\quad + \pi_T \mathbb{E} \left\{ (Z_t \otimes I_G) \int_{\mathbb{R}^G} \mathcal{K}(x) f_t^{(2)}((A - A_0) Z_t + \pi_T x | \mathcal{F}_{t-1}) x dx \right\}. \end{aligned}$$

It follows by dominated convergence that:

$$\lim_{T \rightarrow \infty} \mathbb{E}[P_T(\theta)] = P(A) = (G+1) \mathbb{E} \left\{ (Z_t \otimes I_G) f_t^{(1)}((A - A_0) Z_t | \mathcal{F}_{t-1}) \right\}, \quad (\text{A.15})$$

noting that $\mathbb{E}\{\|Z_t\|\}$ is finite by Assumption 4, $\int_{\mathbb{R}^G} |\mathcal{K}(x)| dx$ and $\int_{\mathbb{R}^G} \|x\mathcal{K}(x)\| dx$ are finite by Assumptions 6 and 12, and $\|f_t^{(1)}(x | \mathcal{F}_{t-1})\|$ and $\|f_t^{(2)}(x | \mathcal{F}_{t-1})\|$ are uniformly bounded by Assumption 9, and also that $P(A)$ is continuous in A , since $\|f_t^{(1)}(x | \mathcal{F}_{t-1})\|$ is continuous in x by Assumption 9. ■

Lemma 11 *Suppose that Assumptions 1–14 are satisfied. Let $0 < \nu < \infty$ be a scalar constant and $\Theta = [e^{-\nu}, e^\nu] \times \text{vec}(\mathcal{A})$; then:*

$$\sup_{\theta \in \Theta} \|P_T(\theta) - P(\theta)\| = o_p(1).$$

where $\theta = (\gamma, \text{vec}(A)')'$ and where $P_T(\theta)$ and $P(\theta)$ are defined as in Lemma 10.

Proof. Observe that:

$$P_T(\theta) = \gamma^{G+1} \frac{\partial R_T(\theta)}{\partial \gamma} = \gamma^{G+2} (T\delta_T^{G+2})^{-1} \sum_{t=1}^T (Z_t \otimes I_G) \mathcal{K}^{(2)} \left(\frac{\gamma \eta_t(A)}{\delta_T} \right) \eta_t(A),$$

where $\eta_t(A) = (Y_t - AZ_t)$. Hence the $((i-1)G + j)$ th element of $P_T(\theta)$ is given by:

$$v_{T,ij}(\theta) = \gamma^{G+2} (T\delta_T^{G+2})^{-1} \sum_{t=1}^T Z_{it} \sum_{s=1}^G \mathcal{K}_{js}^{(2)} \left(\frac{\gamma \eta_t(A)}{\delta_T} \right) \eta_{s,t}(A),$$

where $\eta_{s,t}(A)$ denotes the s th element of $\eta_t(A)$. In addition, as established by Lemma 10:

$$\lim_{T \rightarrow \infty} \mathbb{E}[P_T(\theta)] = P(\theta) = (G+1) \mathbb{E} \left\{ (Z_t \otimes I_G) \int_{\mathbb{R}^G} \mathcal{K}(x) f_t^{(1)}((A - A_0) Z_t | \mathcal{F}_{t-1}) dx \right\},$$

so the $((i-1)G + j)$ th element of $P(\theta)$ is given by:

$$v_{0,ij}(\theta) = (G+1) \mathbb{E} \left\{ Z_{it} \int_{\mathbb{R}^G} \mathcal{K}(x) f_{t,j}^{(1)}((A - A_0) Z_t | \mathcal{F}_{t-1}) dx \right\},$$

where $f_{t,j}^{(1)}(\cdot|\mathcal{F}_{t-1})$ denotes the j th element of $f_{t,j}^{(1)}(\cdot|\mathcal{F}_{t-1})$. Hence it suffices to establish that:

$$\sup_{\theta \in \Theta} \|v_{T,ij}(\theta) - v_{0,ij}(\theta)\| = o_p(1),$$

for all $i = 1, \dots, K$ and $j = 1, \dots, G$.

Next, observe that:

$$\begin{aligned} v_{T,ij}(\theta) &= \gamma^{G+2} (T\delta_T^{G+2})^{-1} \sum_{t=1}^T Z_{it} \sum_{s=1}^G \mathcal{K}_{js}^{(2)} \left(\frac{\gamma\eta_t(A)}{\delta_T} \right) \eta_{s,t}(A) \\ &= \gamma^{G+2} \sum_{s=1}^G \sum_{t=1}^T (T\delta_T^{G+2})^{-1} \mathcal{K}_{js}^{(2)} \left(\frac{\gamma\eta_t(A)}{\delta_T} \right) Z_{it} Y_{st} \\ &\quad - \gamma^{G+2} \sum_{s=1}^G \sum_{k=1}^K a_{sk} \sum_{t=1}^T (T\delta_T^{G+2})^{-1} \mathcal{K}_{js}^{(2)} \left(\frac{\gamma\eta_t(A)}{\delta_T} \right) Z_{it} Z_{kt}, \end{aligned}$$

where a_{sk} is the (s, k) th element of A . Now, in the proof of Lemma 9 we established that terms of the form $\sum_{t=1}^T (T\delta_T^{G+2})^{-1} \mathcal{K}_{rs}^{(2)} \left(\frac{\eta_t(A)}{\pi_T} \right) Z_{it} Z_{kt}$ converged in probability uniformly in θ over Θ to $\mathbb{E} \left[\gamma^{-(G+2)} f_{t,rs}^{(2)}((A - A_0) Z_t | \mathcal{F}_{t-1}) Z_{it} Z_{kt} \right]$; see Equation (A.5). Since Θ is compact then it suffices to establish that terms of the form $\sum_{t=1}^T (T\delta_T^{G+2})^{-1} \mathcal{K}_{js}^{(2)} \left(\frac{\eta_t(A)}{\pi_T} \right) Z_{it} Y_{st}$ converge in probability uniformly in θ over Θ .

Now fix $1 \leq i \leq K$ and $1 \leq j, s \leq G$ and define:

$$\begin{aligned} h_{iT}(\theta) &= (T\delta_T^{G+2})^{-1} \mathcal{K}_{js}^{(2)} \left(\frac{\eta_t(A)}{\pi_T} \right) Z_{it} Y_{st}, \\ h_{iT}^\dagger(\theta) &= (T\delta_T^{G+2})^{-1} \mathcal{K}_{js}^{(2)} \left(\frac{\eta_t(A)}{\pi_T} \right) Z_{it} Y_{st} \chi(|Z_{it} Y_{st}| \geq \delta_T^{-2}). \end{aligned}$$

Then define $h_{iT}^*(\theta)$, $h_{iT}^e(\theta)$, $h_{iT}^{\dagger,e}(\theta)$, $h_{iT}^{*,e}(\theta)$, $H_T(\theta)$, $H_T^\dagger(\theta)$, $H_T^*(\theta)$, $H_T^e(\theta)$, $H_T^{\dagger,e}(\theta)$, $H_T^{*,e}(\theta)$, $H_0(\theta)$, $M_{1,T}, \dots, M_{6,T}$ in the same way in relation to $h_{iT}(\theta)$ and $h_{iT}^e(\theta)$ as given in the proof of Lemma 9. Then:

$$\sup_{\theta \in \Theta} |H_T(\theta) - H_0(\theta)| \leq M_{1,T} + M_{2,T} + M_{3,T} + M_{4,T} + M_{5,T} + M_{6,T}. \quad (\text{A.16})$$

First, observe that $H_T(\theta)$ and $H_T^e(\theta)$ can be written in the form:

$$\begin{aligned} H_T(\theta) &= T^{-1} \sum_{t=1}^T \delta_T^{-(G+\sigma)\psi} \left(\frac{\eta_t(A)}{\pi_T} \right) m(W_t), \\ H_T^e(\theta) &= T^{-1} \sum_{t=1}^T \delta_T^{-(G+\sigma)} \mathbb{E} \left[\psi \left(\frac{\eta_t(A)}{\pi_T} \right) m(W_t) | \mathcal{F}_{t-1} \right], \end{aligned}$$

where $\sigma = 2$, $\psi(x) = \mathcal{K}_{js}^{(2)}(x)$ and $m(W_t) = Z_{it}Y_{st}$. Since $\sup_x \|\mathcal{K}^{(3)}(x)\| < \infty$, by Assumption 12, $|Z_{it}Y_{st}| \leq \|W_t\|^2$ by CS, $\mathbb{E}\{\|W_t\|^3\} < \infty$, by Assumption 11, and $\delta_t = o(1)$, $T^{-2}\delta_T^{-(G+3)} = o(1)$, by Assumption 13, then it follows that $M_{1,T}$ and $M_{5,T}$ are both $o_p(1)$ by Lemma 3.

Second, observe that $H_T^\dagger(\theta)$ and $H_T^{\dagger,e}(\theta)$ can be written in the form:

$$\begin{aligned} H_T^\dagger(\theta) &= T^{-1} \sum_{t=1}^T \delta_T^{-(G+\sigma)} \psi\left(\frac{\eta_t(A)}{\pi_T}\right) m_T^\dagger(W_t), \\ H_T^{\dagger,e}(\theta) &= T^{-1} \sum_{t=1}^T \delta_T^{-(G+\sigma)} \mathbb{E}\left[\psi\left(\frac{\eta_t(A)}{\pi_T}\right) m_T^\dagger(W_t) \mid \mathcal{F}_{t-1}\right], \end{aligned}$$

where $m_T^\dagger(W_t) = m(W_t) \chi(|m(W_t)| > \delta_T^{-2})$. Fix r such that :

$$(G+4) < 2r \leq (G+4+\tau),$$

and note that this then implies $r > (G+\sigma+2)/2$, since $\sigma = 2$, and:

$$\mathbb{E}\{|m(W_t)|^r\} \leq \mathbb{E}\{\|W_t\|^{2r}\} \leq \mathbb{E}\{\|W_t\|^{G+4+\tau}\} < \infty,$$

by Assumption 11. Since $\sup_x \|\mathcal{K}^{(3)}(x)\| < \infty$, by Assumption 12, and $\delta_T = o(1)$, by Assumption 7, then it follows that $M_{2,T}$ and $M_{4,T}$ are both $o_p(1)$ by Lemma 4, noting that $\Theta_T \subseteq \Theta$.

Third, set $\zeta_{tT}(\theta) = h_{tT}^*(\theta) - h_{tT}^{*,e}(\theta)$ which implies that:

$$\zeta_{tT}(\theta) = T^{-1} \delta_T^{-(G+2)} \left\{ \psi\left(\frac{\eta_t(A)}{\pi_T}\right) m_T^*(W_t) - \mathbb{E}\left[\psi\left(\frac{\eta_t(A)}{\pi_T}\right) m_T^*(W_t) \mid \mathcal{F}_{t-1}\right] \right\}.$$

Since $|\psi(u)| = |\mathcal{K}_{rs}^{(2)}(u)| \leq C$ for all u , by Assumption 12, and $|m_T^*(w)| \leq \delta_T^{-2}$ for all w then:

$$|\zeta_{tT}(\theta)| \leq 2T^{-1} \delta_T^{-(G+4)} C.$$

Now:

$$\begin{aligned} \text{Var}[\zeta_{tT}(\theta) \mid \mathcal{F}_{t-1}] &\leq \mathbb{E}[h_{tT}^*(\theta)^2 \mid \mathcal{F}_{t-1}] \leq \mathbb{E}[h_{tT}(\theta)^2 \mid \mathcal{F}_{t-1}] \\ &= T^{-2} \delta_T^{-(2G+4)} \mathbb{E}\left[\mathcal{K}_{rs}^{(2)}\left(\frac{\eta_t(A)}{\pi_T}\right)^2 Y_{st}^2 \mid \mathcal{F}_{t-1}\right] Z_{it}^2, \end{aligned}$$

where $\pi_T = \delta_T / \gamma$. Then by transformation of variables and CS:

$$\begin{aligned}
\mathbb{E} \left[\mathcal{K}_{rs}^{(2)} \left(\frac{\eta_t(A)}{\pi_T} \right)^2 Y_{st}^2 | \mathcal{F}_{t-1} \right] &= \int_{\mathbb{R}^G} \mathcal{K}_{rs}^{(2)} \left(\frac{u - (A - A_0) Z_t}{\pi_T} \right)^2 (e'_s u + e'_s A_0 Z_t)^2 f_t(u | \mathcal{F}_{t-1}) du \\
&= \pi_T^G \int_{\mathbb{R}^G} \mathcal{K}_{rs}^{(2)}(x)^2 (e'_s A Z_t + \pi_T e'_s x)^2 \\
&\quad \times f_t((A - A_0) Z_t + \pi_T x | \mathcal{F}_{t-1}) dx \\
&\leq 2C \pi_T^G \int_{\mathbb{R}^G} \mathcal{K}_{rs}^{(2)}(x)^2 [\|A\|^2 \|Z_t\|^2 + \pi_T^2 \|x\|^2] dx \\
&\leq 2C^2 \pi_T^2 [d_{\mathcal{A}}^2 \|Z_t\|^2 + \pi_T^2 C],
\end{aligned}$$

where $d_{\mathcal{A}} = \sup_{A \in \mathcal{A}} \|A\|$ and e_s is a $(G \times 1)$ vector whose s th element is 1 and whose other elements are all 0. Then:

$$\text{Var} [\zeta_{tT}(\theta) | \mathcal{F}_{t-1}] \leq 2\gamma^{-G} T^{-2} \delta_T^{-(G+4)} C^2 [d_{\mathcal{A}}^2 \|Z_t\|^4 + \pi_T^2 C \|Z_t\|^2],$$

and so:

$$\sup_{\theta \in \Theta_T} \sum_{t=1}^T \text{Var} [\zeta_{tT}(\theta) | \mathcal{F}_{t-1}] \leq 2T^{-1} \delta_T^{-(G+4)} e^{G\nu} C^2 \left[T^{-1} \sum_{t=1}^T \{d_{\mathcal{A}}^2 \|Z_t\|^4 + \pi_T^2 C \|Z_t\|^2\} \right].$$

Set $\mu_4 = \mathbb{E} \{\|W_t\|^4\}$, which is finite by Assumption 11; then:

$$\mathbb{E} \{\|Z_t\|^4\} \leq \mathbb{E} \{\|W_t\|^4\} \leq \mu_4, \quad \mathbb{E} \{\|Z_t\|^2\} \leq \mathbb{E} \{\|W_t\|^2\} \leq \mu_4^{1/2},$$

and hence:

$$\sup_{\theta \in \Theta_T} \sum_{t=1}^T \text{Var} [\zeta_{tT}(\theta) | \mathcal{F}_{t-1}] \leq 2T^{-1} \delta_T^{-(G+4)} e^{G\nu} C^2 [d_{\mathcal{A}}^2 \mu_4^{1/2} + o_p(1)],$$

by the ergodic theorem and noting that $\pi_T^2 = o(1)$. Hence:

$$\Pr \left(\sup_{\theta \in \Theta_T} \sum_{t=1}^T \text{Var} [\zeta_{tT}(\theta) | \mathcal{F}_{t-1}] > 4T^{-1} \delta_T^{-(G+4)} \bar{C}_2 \right) = o_p(1),$$

where $\bar{C}_2 = e^{G\nu} C^2 d_{\mathcal{A}}^2 \mu_4$. In addition, Lemma 1 implies that the number of elements of Θ_T is less than or equal to $T^{2G(K+1)} d$. Hence, setting $a = \eta$, $b = 4T^{-1} \delta_T^{-(G+4)} \bar{C}_2$ and $c = 2T^{-1} \delta_T^{-(G+4)} C$, it follows from Lemma 2 that for any fixed $\eta > 0$ that:

$$\begin{aligned}
\Pr \left(\sup_{\theta \in \Theta_T} \left| \sum_{t=1}^T \zeta_{tT}(\theta) \right| \geq \eta \right) &\leq 2T^{2G(K+1)} d \exp \left\{ \frac{-\eta^2 T \delta_T^{(G+4)}}{4C\eta + 8\bar{C}_2} \right\} + o(1) \\
&= 2d \exp \left\{ \left[2G(K+1) \frac{\ln T}{T \delta_T^{(G+4)}} - \frac{\eta^2}{4C\eta + 8\bar{C}_2} \right] T \delta_T^{(G+4)} \right\} + o(1), \\
&= o(1),
\end{aligned}$$

since $\frac{\ln T}{T\delta_T^{G+4}} = o(1)$ and $T^{-1}\delta_T^{-(G+4)} = o(1)$ by Assumption 13. But:

$$M_{3,T} = \sup_{\theta \in \Theta_T} |H_T^*(\theta) - H_T^{*,e}(\theta)| = \sup_{\theta \in \Theta_T} \left| \sum_{t=1}^T \zeta_{tT}(\theta) \right|,$$

and thus $M_{3,T} = o_p(1)$.

Last, observe that:

$$\begin{aligned} h_{tT}^e(\theta) &= T^{-1}\delta_T^{-(G+2)} \mathbb{E} \left[\mathcal{K}_{rs}^{(2)} \left(\frac{\gamma\eta_t(A)}{\delta_T} \right) Y_{st} | \mathcal{F}_{t-1} \right] Z_{it} \\ &= T^{-1}\delta_T^{-(G+2)} \mathbb{E} \left[\mathcal{K}_{rs}^{(2)} \left(\frac{\gamma\eta_t(A)}{\delta_T} \right) \eta_{t,s}(A) | \mathcal{F}_{t-1} \right] Z_{it} \\ &\quad + T^{-1}\delta_T^{-(G+2)} \mathbb{E} \left[\mathcal{K}_{rs}^{(2)} \left(\frac{\gamma\eta_t(A)}{\delta_T} \right) | \mathcal{F}_{t-1} \right] Z_{it} e'_s A Z_t, \end{aligned}$$

and hence:

$$\begin{aligned} H_T^e(\theta) &= T^{-1} \sum_{t=1}^T \delta_T^{-(G+2)} \mathbb{E} \left[\mathcal{K}_{rs}^{(2)} \left(\frac{\gamma\eta_t(A)}{\delta_T} \right) \eta_{t,s}(A) | \mathcal{F}_{t-1} \right] Z_{it} \\ &\quad + T^{-1} \sum_{t=1}^T \delta_T^{-(G+2)} \mathbb{E} \left[\mathcal{K}_{rs}^{(2)} \left(\frac{\gamma\eta_t(A)}{\delta_T} \right) | \mathcal{F}_{t-1} \right] Z_{it} e'_s A Z_t. \end{aligned}$$

The proof of Lemma 9 established that:

$$T^{-1} \sum_{t=1}^T Z_{it} Z_{jt} \delta_T^{-(G+2)} \mathbb{E} \left[\mathcal{K}_{rs}^{(2)} \left(\frac{\gamma\eta_t(A)}{\delta_T} \right) | \mathcal{F}_{t-1} \right]$$

converges uniformly in probability to:

$$\gamma^{-(G+2)} \mathbb{E} \left[f_{t,rs}^{(2)}((A - A_0) Z_t | \mathcal{F}_{t-1}) Z_{it} Z_{jt} \right],$$

over Θ ; see Equations (A.7) and (A.11). Hence:

$$T^{-1} \sum_{t=1}^T \delta_T^{-(G+2)} \mathbb{E} \left[\mathcal{K}_{rs}^{(2)} \left(\frac{\eta_t(A)}{\pi_T} \right) | \mathcal{F}_{t-1} \right] Z_{it} e'_s A Z_t,$$

converges uniformly in probability to:

$$\gamma^{-(G+2)} \mathbb{E} \left[f_{t,rs}^{(2)}((A - A_0) Z_t | \mathcal{F}_{t-1}) Z_{it} e'_s A Z_t \right].$$

The proof of Lemma 10 established that:

$$\begin{aligned} \mathbb{E} \left[\mathcal{K}_{rs}^{(2)} \left(\frac{\eta_t(A)}{\pi_T} \right) \eta_{t,s}(A) | \mathcal{F}_{t-1} \right] &= \chi_{rs} \pi_T^G \int_{\mathbb{R}^G} \mathcal{K}(s) f_{t,s}^{(1)}(\lambda_t + \pi_T x | \mathcal{F}_{t-1}) dx \\ &\quad + \pi_T^G \int_{\mathbb{R}^G} \mathcal{K}(s) f_{t,r}^{(1)}(\lambda_t + \pi_T x | \mathcal{F}_{t-1}) dx \\ &\quad + \pi_T^{G+1} \int_{\mathbb{R}^G} \mathcal{K}(s) f_{t,rs}^{(2)}(\lambda_t + \pi_T x | \mathcal{F}_{t-1}) x_j dx, \end{aligned}$$

see Equation (A.14), and thus:

$$\begin{aligned}
& T^{-1} \sum_{t=1}^T \delta_T^{-(G+2)} \mathbb{E} \left[\mathcal{K}_{rs}^{(2)} \left(\frac{\eta_t(A)}{\pi_T} \right) \eta_{t,s}(A) | \mathcal{F}_{t-1} \right] Z_{it} \\
&= \chi_{rs} \gamma^G T^{-1} \sum_{t=1}^T \delta_T^{-2} \int_{\mathbb{R}^G} \mathcal{K}(s) f_{t,s}^{(1)}(\lambda_t + \pi_T x | \mathcal{F}_{t-1}) dx Z_{it} \\
&\quad + \gamma^G T^{-1} \sum_{t=1}^T \delta_T^{-2} \int_{\mathbb{R}^G} \mathcal{K}(s) f_{t,r}^{(1)}(\lambda_t + \pi_T x | \mathcal{F}_{t-1}) dx Z_{it} \\
&\quad + \gamma^{G+1} T^{-1} \sum_{t=1}^T \delta_T^{-1} \int_{\mathbb{R}^G} \mathcal{K}(s) f_{t,rs}^{(2)}(\lambda_t + \pi_T x | \mathcal{F}_{t-1}) x_j dx Z_{it}.
\end{aligned}$$

By ULLN and noting that $\delta_T = o(1)$ it follows that:

$$T^{-1} \sum_{t=1}^T \delta_T^{-(G+2)} \mathbb{E} \left[\mathcal{K}_{rs}^{(2)} \left(\frac{\eta_t(A)}{\pi_T} \right) \eta_{s,t}(A) | \mathcal{F}_{t-1} \right] Z_{it},$$

converges in probability uniformly over Θ to:

$$(G+1) \gamma^{-G} \mathbb{E} \left[f_t^{(1)}((A - A_0) Z_t | \mathcal{F}_{t-1}) Z_{it} \right],$$

and hence that $H_T^e(\theta)$ converges in probability uniformly over Θ to:

$$\begin{aligned}
H_0(\theta) &= (G+1) \gamma^{-G} \mathbb{E} \left[f_t^{(1)}((A - A_0) Z_t | \mathcal{F}_{t-1}) Z_{it} \right] \\
&\quad + \gamma^{-(G+2)} \mathbb{E} \left[f_{t,rs}^{(2)}((A - A_0) Z_t | \mathcal{F}_{t-1}) Z_{it} e'_s A Z_t \right],
\end{aligned}$$

and thus $M_{6,T} = o_p(1)$. Combined with Equation (A.16) and the earlier results that $M_{1,T}, \dots, M_{5,T}$ are $o_p(1)$ this then implies that:

$$\sup_{\theta \in \Theta} |H_T(\theta) - H_0(\theta)| = o_p(1),$$

and hence that:

$$\sup_{\theta \in \Theta} |v_{T,ij}(\theta) - v_{0,j}(\theta)| = o_p(1),$$

which in turn implies that:

$$\sup_{\theta \in \Theta} \left\| P_T(\theta) - (G+1) \mathbb{E} \left[(Z_t \otimes I_G) f_t^{(1)}((A - A_0) Z_t | \mathcal{F}_{t-1}) \right] \right\| = o_p(1).$$

■

Lemma 12 *Suppose that Assumptions 1–14 are satisfied. Let $\theta = (\gamma, \text{vec}(A))'$ where $0 < \gamma < \infty$ and $A \in \mathcal{A}$. Then:*

$$\lim_{T \rightarrow \infty} \mathbb{E} [B_T(\theta)] = B(\theta),$$

and:

$$\begin{aligned}
B_T(\theta) &= \gamma^G T^{-1} \sum_{t=1}^T \delta_T^{-G} (Z_t \otimes I_G) \mathcal{K}^{(1)} \left(\frac{\gamma(Y_t - AZ_t)}{\delta_T} \right) \mathcal{K}^{(1)} \left(\frac{\gamma(Y_t - AZ_t)}{\delta_T} \right)' (Z_t \otimes I_G)', \\
B(\theta) &= \mathbb{E} [f_t((A - A_0) Z_t | \mathcal{F}_{t-1}) (Z_t \otimes I_G) \mathcal{M} (Z_t \otimes I_G)'],
\end{aligned}$$

where $\mathcal{M} = \int_{\mathbb{R}^G} \mathcal{K}^{(1)}(x) \mathcal{K}^{(1)}(x)' dx$. In addition $B(\theta)$ is continuous.

Proof. Fix $0 < \gamma < \infty$ and define $\pi_T = \delta_T/\gamma$. Then each element of $E[B_T(\theta)]$ can be expressed as a sum of terms of the form:

$$\begin{aligned}
& T^{-1} \sum_{t=1}^T \pi_T^{-G} \mathbb{E} \left[\mathcal{K}_r^{(1)} \left(\frac{\eta_t(A)}{\pi_T} \right) \mathcal{K}_s^{(1)} \left(\frac{\eta_t(A)}{\pi_T} \right) Z_{it} Z_{jt} \right] \\
&= T^{-1} \sum_{t=1}^T \pi_T^{-G} \mathbb{E} \left[\mathbb{E} \left\{ \mathcal{K}_r^{(1)} \left(\frac{\eta_t(A)}{\pi_T} \right) \mathcal{K}_s^{(1)} \left(\frac{\eta_t(A)}{\pi_T} \right) \middle| \mathcal{F}_{t-1} \right\} Z_{it} Z_{jt} \right]
\end{aligned}$$

where $1 \leq i, j \leq K$ and $1 \leq r, s \leq G$. Now:

$$\begin{aligned}
\mathbb{E} \left\{ \mathcal{K}_r^{(1)} \left(\frac{\eta_t(A)}{\pi_T} \right) \mathcal{K}_s^{(1)} \left(\frac{\eta_t(A)}{\pi_T} \right) \middle| \mathcal{F}_{t-1} \right\} &= \int_{\mathbb{R}^G} \mathcal{K}_r^{(1)} \left(\frac{u - \lambda_t}{\pi_T} \right) \mathcal{K}_s^{(1)} \left(\frac{u - \lambda_t}{\pi_T} \right) \\
&\quad \times f_t(u | \mathcal{F}_{t-1}) du \\
&= \pi_T^G \int_{\mathbb{R}^G} \mathcal{K}_r^{(1)}(x) \mathcal{K}_s^{(1)}(x) \\
&\quad \times f_t(\lambda_t + \pi_T x | \mathcal{F}_{t-1}) dx, \quad (\text{A.17})
\end{aligned}$$

where $\lambda_t = (A - A_0) Z_t$ and thus:

$$\begin{aligned}
& T^{-1} \sum_{t=1}^T \pi_T^{-G} \mathbb{E} \left[\mathcal{K}_r^{(1)} \left(\frac{\eta_t(A)}{\pi_T} \right) \mathcal{K}_s^{(1)} \left(\frac{\eta_t(A)}{\pi_T} \right) Z_{it} Z_{jt} \right] \\
&= T^{-1} \sum_{t=1}^T \mathbb{E} \left[\int_{\mathbb{R}^G} \mathcal{K}_r^{(1)}(x) \mathcal{K}_s^{(1)}(x) f_t((A - A_0) Z_t + \gamma^{-1} \delta_T x | \mathcal{F}_{t-1}) dx Z_{it} Z_{jt} \right].
\end{aligned}$$

Since $\mathcal{K}^{(1)}(x)$ and $f_t(u | \mathcal{F}_{t-1})$ are uniformly bounded in absolute value, by Assumptions 3 and 12, $f_t(u | \mathcal{F}_{t-1})$ is continuous, by Assumption 3, Z_t has finite second moments, by Assumption 11, and $\delta_T = o(1)$, it follows by dominated convergence that:

$$\begin{aligned}
& \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \pi_T^{-G} \mathbb{E} \left[\mathcal{K}_r^{(1)} \left(\frac{\eta_t(A)}{\pi_T} \right) \mathcal{K}_s^{(1)} \left(\frac{\eta_t(A)}{\pi_T} \right) Z_{it} Z_{jt} \right] \\
&= \mathbb{E} \left[\int_{\mathbb{R}^G} \mathcal{K}_r^{(1)}(x) \mathcal{K}_s^{(1)}(x) f_t((A - A_0) Z_t | \mathcal{F}_{t-1}) dx Z_{it} Z_{jt} \right] \\
&= \mathcal{M}_{rs} \mathbb{E} [f_t((A - A_0) Z_t | \mathcal{F}_{t-1}) Z_{it} Z_{jt}],
\end{aligned}$$

where \mathcal{M}_{rs} is the (r, s) th element of \mathcal{M} and hence that:

$$\lim_{T \rightarrow \infty} B_T(\theta) = B(\theta) = \mathbb{E} \left[f_t((A - A_0) Z_t | \mathcal{F}_{t-1}) (Z_t \otimes I_G) \mathcal{M} (Z_t \otimes I_G)' \right].$$

■

Lemma 13 *Suppose that Assumptions 1–14 are satisfied. Let $0 < \nu < \infty$ be a scalar constant and $\Theta = [e^{-\nu}, e^\nu] \times \text{vec}(\mathcal{A})$; then:*

$$\sup_{\theta \in \Theta} \|B_T(\theta) - B(\theta)\| = o_p(1).$$

where $\theta = (\gamma, \text{vec}(A)')'$ and where $B_T(\theta)$ and $B(\theta)$ are given as in Lemma 12.

Proof. Observe that each element of $B_T(\theta)$ can be written as sum of terms of the form:

$$\gamma^G T^{-1} \sum_{t=1}^T \delta_T^{-G} \mathcal{K}_r^{(1)} \left(\frac{\eta_t(A)}{\delta_T} \right) \mathcal{K}_s^{(1)} \left(\frac{\eta_t(A)}{\delta_T} \right) Z_{it} Z_{jt},$$

where $1 \leq i, j \leq K$ and $1 \leq r, s \leq G$. Define:

$$\begin{aligned} h_{tT}(\theta) &= (T\delta_T^G)^{-1} \mathcal{K}_r^{(1)} \left(\frac{\eta_t(A)}{\delta_T} \right) \mathcal{K}_s^{(1)} \left(\frac{\eta_t(A)}{\delta_T} \right) Z_{it} Z_{jt}, \\ h_{tT}^\dagger(\theta) &= (T\delta_T^G)^{-1} \mathcal{K}_r^{(1)} \left(\frac{\eta_t(A)}{\delta_T} \right) \mathcal{K}_s^{(1)} \left(\frac{\eta_t(A)}{\delta_T} \right) Z_{it} Z_{jt} \chi(|Z_{it} Z_{jt}| \geq \delta_T^{-2}), \end{aligned}$$

and then define $h_{tT}^*(\theta)$, $h_{tT}^e(\theta)$, $h_{tT}^{\dagger,e}(\theta)$, $h_{tT}^{*,e}(\theta)$, $H_T(\theta)$, $H_T^\dagger(\theta)$, $H_T^*(\theta)$, $H_T^e(\theta)$, $H_T^{\dagger,e}(\theta)$, $H_T^{*,e}(\theta)$, $H_0(\theta)$, $M_{1,T}, \dots, M_{6,T}$ in the same way in relation to $h_{tT}(\theta)$ and $h_{tT}^e(\theta)$ as given in the proof of Lemma 9. Then:

$$\sup_{\theta \in \Theta} |H_T(\theta) - H_0(\theta)| \leq M_{1,T} + M_{2,T} + M_{3,T} + M_{4,T} + M_{5,T} + M_{6,T}. \quad (\text{A.18})$$

First, observe that $H_T(\theta)$ and $H_T^e(\theta)$ can be written in the form:

$$\begin{aligned} H_T(\theta) &= T^{-1} \sum_{t=1}^T \delta_T^{-(G+\sigma)} \psi \left(\frac{\gamma \eta_t(A)}{\delta_T} \right) m(W_t), \\ H_T^e(\theta) &= T^{-1} \sum_{t=1}^T \delta_T^{-(G+\sigma)} \mathbb{E} \left[\psi \left(\frac{\gamma \eta_t(A)}{\delta_T} \right) m(W_t) | \mathcal{F}_{t-1} \right], \end{aligned}$$

where $\sigma = 0$, $\psi(x) = \mathcal{K}_r^{(1)}(x) \mathcal{K}_s^{(1)}(x)$ and $m(W_t) = Z_{it} Z_{jt}$. Now:

$$\frac{\partial \psi(x)}{\partial x_l} = \mathcal{K}_{rl}^{(1)}(x) \mathcal{K}_s^{(1)}(x) + \mathcal{K}_r^{(1)}(x) \mathcal{K}_{sl}^{(1)}(x),$$

and since $\sup_x \|\mathcal{K}^{(1)}(x)\| < \infty$ and $\sup_x \|\mathcal{K}^{(2)}(x)\| < \infty$, by Assumption 12, it follows that $\sup_x \left\| \frac{\partial \psi(x)}{\partial x} \right\| < \infty$. In addition, $|Z_{it}Z_{jt}| \leq \|W_t\|^2$ by CS, $\mathbb{E}\{\|W_t\|^3\} < \infty$, by Assumption 11, and $\delta_t = o(1)$, $T^{-2}\delta_T^{-(G+3)} = o(1)$, by Assumption 13. Hence it follows that $M_{1,T}$ and $M_{5,T}$ are both $o_p(1)$ by Lemma 3.

Second, observe that $H_T^\dagger(\theta)$ and $H_T^{\dagger,e}(\theta)$ can be written in the form:

$$\begin{aligned} H_T^\dagger(\theta) &= T^{-1} \sum_{t=1}^T \delta_T^{-(G+\sigma)} \psi\left(\frac{\gamma\eta_t(A)}{\delta_T}\right) m_T^\dagger(W_t), \\ H_T^{\dagger,e}(\theta) &= T^{-1} \sum_{t=1}^T \delta_T^{-(G+\sigma)} \mathbb{E}\left[\psi\left(\frac{\gamma\eta_t(A)}{\delta_T}\right) m_T^\dagger(W_t) \mid \mathcal{F}_{t-1}\right], \end{aligned}$$

where $m_T^\dagger(W_t) = m(W_t) \chi(|m(W_t)| > \delta_T^{-2})$. Fix r such that:

$$(G+2) < 2r \leq (G+2+\tau),$$

and note that this then implies $r > (G+\sigma+2)/2$ since $\sigma = 0$ and hence that:

$$\mathbb{E}\{|m(W_t)|^r\} \leq \mathbb{E}\{\|W_t\|^{2r}\} \leq \mathbb{E}\{\|W_t\|^{G+2+\tau}\} < \infty,$$

by Assumption 11. Since $\sup_x \left\| \frac{\partial \psi(x)}{\partial x} \right\| < \infty$ and $\delta_T = o(1)$, by Assumption 7, then it follows that $M_{2,T}$ and $M_{4,T}$ are both $o_p(1)$ by Lemma 4, noting that $\Theta_T \subseteq \Theta$.

Third, set $\zeta_{tT}(\theta) = h_{tT}^*(\theta) - h_{tT}^{*,e}(\theta)$ which implies that:

$$\zeta_{tT}(\theta) = T^{-1} \delta_T^{-G} \left\{ \psi\left(\frac{\gamma\eta_t(A)}{\delta_T}\right) m_T^*(W_t) - \mathbb{E}\left[\psi\left(\frac{\gamma\eta_t(A)}{\delta_T}\right) m_T^*(W_t) \mid \mathcal{F}_{t-1}\right] \right\}.$$

Since $|\psi(u)| = |\mathcal{K}^{(1)}(u)| \leq C^2$ for all u , by Assumption 12, and $|m_T^*(w)| \leq \delta_T^{-2}$ for all w then:

$$|\zeta_{tT}(\theta)| \leq 2T^{-1} \delta_T^{-(G+2)} C^2.$$

Now:

$$\begin{aligned} \text{Var}[\zeta_{tT}(\theta) \mid \mathcal{F}_{t-1}] &\leq \mathbb{E}[h_{tT}^*(\theta)^2 \mid \mathcal{F}_{t-1}] \leq \mathbb{E}[h_{tT}(\theta)^2 \mid \mathcal{F}_{t-1}] \\ &= T^{-2} \delta_T^{-2G} \mathbb{E}\left[\mathcal{K}_r^{(1)}\left(\frac{\eta_t(A)}{\pi_T}\right)^2 \mathcal{K}_s^{(1)}\left(\frac{\eta_t(A)}{\pi_T}\right)^2 \mid \mathcal{F}_{t-1}\right] Z_{it}^2 Z_{jt}^2, \end{aligned}$$

where $\pi_T = \delta_T/\gamma$. Then by transformation of variables:

$$\begin{aligned} & \mathbb{E} \left[\mathcal{K}_r^{(1)} \left(\frac{\eta_t(A)}{\pi_T} \right)^2 \mathcal{K}_s^{(1)} \left(\frac{\eta_t(A)}{\pi_T} \right)^2 \middle| \mathcal{F}_{t-1} \right] \\ &= \int_{\mathbb{R}^G} \mathcal{K}_r^{(1)} \left(\frac{u - (A - A_0) Z_t}{\pi_T} \right)^2 \mathcal{K}_s^{(1)} \left(\frac{u - (A - A_0) Z_t}{\pi_T} \right)^2 f_t(u | \mathcal{F}_{t-1}) du \\ &= \pi_T^G \int_{\mathbb{R}^G} \mathcal{K}_r^{(1)}(x)^2 \mathcal{K}_s^{(1)}(x)^2 f_t((A - A_0) Z_t + \pi_T x | \mathcal{F}_{t-1}) dx \leq \pi_T^G C^4 \end{aligned}$$

by Assumptions 3, 6 and 12, and thus:

$$\text{Var} [\zeta_{tT}(\theta) | \mathcal{F}_{t-1}] \leq \gamma^{-G} T^{-2} \delta_T^{-G} C^4 Z_{it}^2 Z_{jt}^2,$$

so:

$$\sup_{\theta \in \Theta_T} \sum_{t=1}^T \text{Var} [\zeta_{tT}(\theta) | \mathcal{F}_{t-1}] \leq T^{-1} \delta_T^{-G} e^{G\nu} C^4 \left[T^{-1} \sum_{t=1}^T Z_{it}^2 Z_{jt}^2 \right].$$

Set $\mu_4 = \mathbb{E} \{ \|W_t\|^4 \}$, which is finite by Assumption 11; then:

$$\mathbb{E} \{ \|Z_t\|^4 \} \leq \mathbb{E} \{ \|W_t\|^4 \} = \mu_4,$$

and hence:

$$\sup_{\theta \in \Theta_T} \sum_{t=1}^T \text{Var} [\zeta_{tT}(\theta) | \mathcal{F}_{t-1}] \leq T^{-1} \delta_T^{-G} e^{G\nu} C^4 [\mu_4 + o_p(1)],$$

by the ergodic theorem and noting that $\pi_T^2 = o(1)$. Hence:

$$\Pr \left(\sup_{\theta \in \Theta_T} \sum_{t=1}^T \text{Var} [\zeta_{tT}(\theta) | \mathcal{F}_{t-1}] > 2T^{-1} \delta_T^{-(G+4)} \bar{C}_3 \right) = o_p(1),$$

where $\bar{C}_3 = e^{G\nu} C^4 \mu_4$. In addition, Lemma 1 implies that the number of elements of Θ_T is less than or equal to $T^{2G(K+1)}d$. Hence, by setting $a = \eta$, $b = 2T^{-1} \delta_T^{-G} \bar{C}_3$ and $c = 2T^{-1} \delta_T^{-(G+2)} C^2$, Lemma 2 implies that for any fixed $\eta > 0$:

$$\begin{aligned} \Pr \left(\sup_{\theta \in \Theta_T} \left| \sum_{t=1}^T \zeta_{tT}(\theta) \right| \geq \eta \right) &\leq 2T^{2G(K+1)}d \exp \left\{ \frac{-\eta^2 T \delta_T^{G+2}}{4C^2 \eta + 4\bar{C}_3} \right\} + o(1) \\ &= 2d \exp \left\{ \left[2G(K+1) \frac{\ln T}{T \delta_T^{G+2}} - \frac{\eta^2}{4C^2 \eta + 4\bar{C}_3} \right] T \delta_T^{G+2} \right\} + o(1), \\ &= o(1), \end{aligned}$$

since $\frac{\ln T}{T \delta_T^{G+2}} = o(1)$ and $T^{-1} \delta_T^{-(G+2)} = o(1)$ by Assumption 13. But:

$$M_{3,T} = \sup_{\theta \in \Theta_T} |H_T^*(\theta) - H_T^{*,e}(\theta)| = \sup_{\theta \in \Theta_T} \left| \sum_{t=1}^T \zeta_{tT}(\theta) \right|,$$

and thus $M_{3,T} = o_p(1)$.

Last, observe that:

$$H_T^e(\theta) = T^{-1} \sum_{t=1}^T \delta_T^{-G} \mathbf{E} \left[\mathcal{K}_r^{(1)} \left(\frac{\gamma \eta_t(A)}{\delta_T} \right) \mathcal{K}_s^{(1)} \left(\frac{\gamma \eta_t(A)}{\delta_T} \right) | \mathcal{F}_{t-1} \right] Z_{it} Z_{jt}.$$

The proof of Lemma 12 established that:

$$\mathbf{E} \left\{ \mathcal{K}_r^{(1)} \left(\frac{\eta_t(A)}{\pi_T} \right) \mathcal{K}_s^{(1)} \left(\frac{\eta_t(A)}{\pi_T} \right) | \mathcal{F}_{t-1} \right\} = \pi_T^G \int_{\mathbb{R}^G} \mathcal{K}_r^{(1)}(x) \mathcal{K}_s^{(1)}(x) \times f_t((A - A_0) Z_t + \pi_T x | \mathcal{F}_{t-1}) dx,$$

where $\pi_T = \delta_T/\gamma$; see Equation (A.17). By Assumptions 3 and 12 it follows that $Z_{it} Z_{jt} \int_{\mathbb{R}^G} \mathcal{K}_r^{(1)}(x) \mathcal{K}_s^{(1)}(x) f_t((A - A_0) Z_t + \pi x | \mathcal{F}_{t-1}) dx$ is a continuous function of A and π and is bounded in absolute value by $C^2 |Z_{it} Z_{jt}|$. Since $\mathbf{E} \{|Z_{it} Z_{jt}|\} < \infty$, by Assumption 11, then for any fixed $0 < \bar{\pi} < \infty$ it follows by ULLN that:

$$T^{-1} \sum_{t=1}^T Z_{it} Z_{jt} \int_{\mathbb{R}^G} \mathcal{K}_r^{(1)}(x) \mathcal{K}_s^{(1)}(x) f_t((A - A_0) Z_t + \pi x | \mathcal{F}_{t-1}) dx,$$

converges uniformly in probability to:

$$\mathbf{E} \left[Z_{it} Z_{jt} \int_{\mathbb{R}^G} \mathcal{K}_r^{(1)}(x) \mathcal{K}_s^{(1)}(x) f_t((A - A_0) Z_t + \pi x | \mathcal{F}_{t-1}) dx \right],$$

over $(\text{vec}(A)', \pi)' \in \text{vec}(\mathcal{A}) \times [-\bar{\pi}, \bar{\pi}]$. Since $\delta_T = o(1)$ and since $\theta \in \Theta$ requires $\gamma \in [e^{-\nu}, e^\nu]$ then $\sup_{\theta \in \Theta} |\pi_T| = o(1)$ and hence it follows that $H_T^e(\theta)$ converges uniformly in probability to:

$$H_0(\theta) = \gamma^{-G} \mathbf{E} \left[Z_{it} Z_{jt} \int_{\mathbb{R}^G} \mathcal{K}_r^{(1)}(x) \mathcal{K}_s^{(1)}(x) f_t((A - A_0) Z_t | \mathcal{F}_{t-1}) dx \right],$$

over $\theta \in \Theta$ and thus that $M_{6,T} = o(1)$. Together with Equation A.18 and the results shown earlier that $M_{1,T}, \dots, M_{5,tT}$ are all $o_p(1)$ this implies that:

$$\sup_{\theta \in \Theta} |H_T(\theta) - H_0(\theta)| = o_p(1),$$

and hence that:

$$\sup_{\theta \in \Theta} |B_T(\theta) - B(\theta)| = o_p(1).$$

■

Lemma 14 *Under Assumptions 1–14:*

$$(T\delta_T^G)^{-1/2} \sum_{t=1}^T (Z_t \otimes I_G) \mathcal{K}^{(1)} \left(\frac{U_t}{\delta_T} \right) \xrightarrow{d} \mathcal{N}(0, B_0).$$

Proof. Define:

$$g_{tT} = (T\delta_T^G)^{-1/2} (Z_t \otimes I_G) \mathcal{K}^{(1)} \left(\frac{U_t}{\delta_T} \right), \quad g_{tT}^e = \mathbb{E}[g_{tT} | \mathcal{F}_{t-1}], \quad g_{tT}^* = g_{tT} - g_{tT}^e,$$

so that:

$$(T\delta_T^G)^{-1/2} \sum_{t=1}^T (Z_t \otimes I_G) \mathcal{K}^{(1)} \left(\frac{U_t}{\delta_T} \right) = \sum_{t=1}^T g_{tT}^e + \sum_{t=1}^T g_{tT}^*.$$

Then to establish the desired result it is sufficient to establish that $\sum_{t=1}^T g_{tT}^e = o_p(1)$ and

that $\sum_{t=1}^T g_{tT}^* \xrightarrow{d} \mathcal{N}(0, B_0)$.

First, for each $i = 1, \dots, G$ define the i th element of g_{tT}^e to be:

$$\begin{aligned} g_{tT,i}^e &= \mathbb{E} \left[(T\delta_T^G)^{-1/2} \mathcal{K}_i^{(1)} \left(\frac{Y_t - A_0 Z_t}{\delta_T} \right) Z_t | \mathcal{F}_{t-1} \right] \\ &= (T\delta_T^G)^{-1/2} \left[\int_{\mathbb{R}^{G-1}} \int_{\mathbb{R}} \mathcal{K}_i^{(1)} \left(\frac{u}{\delta_T} \right) f_t(u | \mathcal{F}_{t-1}) du_i du_{-i} \right] Z_t, \end{aligned}$$

where u_i denotes the i th element of u and u_{-i} denotes the vector consisting of the elements of u other than u_i . Using integration by parts it follows that:

$$\begin{aligned} \int_{\mathbb{R}} \mathcal{K}_i^{(1)} \left(\frac{u}{\delta_T} \right) f_t(u | \mathcal{F}_{t-1}) du_i &= \left[\delta_T \mathcal{K} \left(\frac{u}{\delta_T} \right) f_t(u | \mathcal{F}_{t-1}) \right]_{u_i=-\infty}^{u_i=\infty} \\ &\quad - \delta_T \int_{\mathbb{R}} \mathcal{K} \left(\frac{u}{\delta_T} \right) f_{t,i}^{(1)}(u | \mathcal{F}_{t-1}) du_i. \end{aligned}$$

But:

$$\left| \delta_T \mathcal{K} \left(\frac{u}{\delta_T} \right) f_t(u | \mathcal{F}_{t-1}) \right| \leq L_0 \delta_T \left| \mathcal{K} \left(\frac{u}{\delta_T} \right) \right|,$$

by Assumption 3, and for all fixed u_{-i} :

$$\lim_{u_i \rightarrow \pm\infty} \left| \mathcal{K}_j \left(\frac{u}{\delta_T} \right) \right| = 0,$$

by Assumption 12. Hence:

$$\left[\delta_T \mathcal{K} \left(\frac{u}{\delta_T} \right) f_t(u | \mathcal{F}_{t-1}) \right]_{u_i=-\infty}^{u_i=\infty} = 0,$$

and thus:

$$\int_{\mathbb{R}} \mathcal{K}_i^{(1)} \left(\frac{u}{\delta_T} \right) f_t(u|\mathcal{F}_{t-1}) du_i = -\delta_T \int_{\mathbb{R}} \mathcal{K} \left(\frac{u}{\delta_T} \right) f_{t,i}^{(1)}(u|\mathcal{F}_{t-1}) du_i$$

so:

$$\begin{aligned} g_{i,tT}^e &= -(T\delta_T^G)^{-1/2} Z_t \delta_T \int_{\mathbb{R}^G} \mathcal{K} \left(\frac{u}{\delta_T} \right) f_{t,i}^{(1)}(u|\mathcal{F}_{t-1}) du \\ &= -(T\delta_T^G)^{-1/2} \delta_T^{G+1} Z_t \int_{\mathbb{R}^G} \mathcal{K}(x) f_{t,i}^{(1)}(\delta_T x|\mathcal{F}_{t-1}) dx. \end{aligned}$$

Thus we have that:

$$\sum_{t=1}^T g_{itT}^e = T^{1/2} \delta_T^{(G/2)+1} \left(T^{-1} \sum_{t=1}^T Z_t \right) \psi_{tT,i},$$

where:

$$\psi_{tT,i} = \int_{\mathbb{R}^G} \mathcal{K}(x) f_{t,i}^{(1)}(\delta_T x|\mathcal{F}_{t-1}) dx.$$

A second order Taylor series expansion of $f_{t,i}^{(1)}(\delta_T x|\mathcal{F}_{t-1})$ around $x = 0$ then gives:

$$\begin{aligned} f_{t,i}^{(1)}(\delta_T x|\mathcal{F}_{t-1}) &= f_{t,i}^{(1)}(0|\mathcal{F}_{t-1}) + \delta_T \sum_{j=1}^G x_j f_{t,ij}^{(2)}(0|\mathcal{F}_{t-1}) \\ &\quad + \left(\frac{1}{2} \right) \delta_T^2 \sum_{j=1}^G \sum_{k=1}^G x_j x_k f_{t,ijk}^{(3)}(\vartheta \delta_T x|\mathcal{F}_{t-1}), \end{aligned}$$

for some $0 \leq \vartheta \leq 1$, where $f_{t,ij}^{(2)}(u|\mathcal{F}_{t-1})$ denotes the (i, j) th element of $f_t^{(2)}(u|\mathcal{F}_{t-1})$ and $f_{t,ijk}^{(3)}(u|\mathcal{F}_{t-1})$ denotes the (i, j, k) th element of $f_t^{(3)}(u|\mathcal{F}_{t-1})$. But $f_{t,i}^{(1)}(0|\mathcal{F}_{t-1}) = 0$, by Assumption 3, and $\int_{\mathbb{R}^G} x \mathcal{K}(x) ds = 0$, by Assumption 12. Hence:

$$\psi_{tT,i} = \left(\frac{1}{2} \right) \delta_T^2 \sum_{j=1}^G \sum_{k=1}^G \int_{\mathbb{R}^G} x_j x_k \mathcal{K}(x) f_{t,ijk}^{(3)}(\vartheta \delta_T x|\mathcal{F}_{t-1}) dx,$$

where ϑ may vary with x , and by CS it follows that:

$$|\psi_{tT,i}| \leq \left(\frac{1}{2} \right) \delta_T^2 \sum_{j=1}^G \sum_{k=1}^G \int_{\mathbb{R}^G} |x_j x_k \mathcal{K}(x) f_{t,ijk}^{(3)}(\vartheta \delta_T x|\mathcal{F}_{t-1})| dx.$$

Since $\sup_u \left\| f_t^{(3)}(u|\mathcal{F}_{t-1}) \right\| \leq C$, by Assumption 9, and since $\int_{\mathbb{R}^G} \|x\|^2 |\mathcal{K}(x)| dx \leq C$, by Assumption 12, it follows that:

$$|\psi_{tT,i}| \leq \frac{G^2 C^2}{2} \delta_T^2, \tag{A.19}$$

and since $T^{-1} \sum_{t=1}^T Z_t = O_p(1)$, by the ergodic theorem, then:

$$\sum_{t=1}^T g_{itT}^e = \frac{G^2 C^2}{2} T^{1/2} \delta_T^{(G/2)+1} \delta_T^2 O_p(1) = O_p \left[(T \delta_T^{G+6})^{1/2} \right] = o_p(1),$$

by Assumption 13. This then implies that $\sum_{t=1}^T g_{itT}^e = o_p(1)$.

Second, fix $\lambda \neq 0$ and define:

$$z_{itT} = \lambda' g_{itT}^*, \quad \sigma_{itT}^2 = \text{Var}(z_{itT}), \quad \Sigma_T = \sum_{t=1}^T \sigma_{itT}^2, \quad \eta_{itT} = \frac{z_{itT}}{\sqrt{\Sigma_T}}.$$

By construction $\{(z_{itT}, \mathcal{F}_t)\}_{t=-\infty}^{\infty}$ is a martingale difference array since $z_{itT} = (\lambda' g_{itT}) - E(\lambda' g_{itT} | \mathcal{F}_{t-1})$. Theorem 24.3 of Davidson (1994) implies that $\sum_{t=1}^T \eta_{itT}$ converges in distribution to a standard normal provided that (a) $\sum_{t=1}^T \text{Var}(\eta_{itT}) = 1$ for all T , (b) $\sum_{t=1}^T \eta_{itT}^2 \xrightarrow{p} 1$, and (c) $\max_{1 \leq t \leq T} |\eta_{itT}| = o_p(1)$. If there exists $0 < \Sigma_0 < \infty$ such that $\Sigma_T \rightarrow \Sigma_0$ as $T \rightarrow \infty$ then these conditions are satisfied provided that (b') $\sum_{t=1}^T z_{itT}^2 \xrightarrow{p} \Sigma_0$ and (c') $\max_{1 \leq t \leq T} |z_{itT}| = o_p(1)$, in which case it follows that $\sum_{t=1}^T z_{itT}$ converges in distribution to a $\mathcal{N}(0, \Sigma_0)$. Now observe that:

$$\begin{aligned} \sigma_{itT}^2 &= E(z_{itT}^2) = E \left[(\lambda' g_{itT})^2 - 2(\lambda' g_{itT})(\lambda' g_{itT}^e) + (\lambda' g_{itT}^e)^2 \right] \\ &= E \left[(\lambda' g_{itT})^2 \right] - E \left[(\lambda' g_{itT}^e)^2 \right], \end{aligned}$$

so:

$$\Sigma_T = E \left[\sum_{t=1}^T (\lambda' g_{itT})^2 \right] - E \left[\sum_{t=1}^T (\lambda' g_{itT}^e)^2 \right].$$

Then:

$$E \left[\sum_{t=1}^T (\lambda' g_{itT})^2 \right] = \sum_{t=1}^T \lambda' E(g_{itT} g_{itT}') \lambda,$$

so by the law of iterated expectations it follows that:

$$E(g_{itT} g_{itT}') = T^{-1} E[(Z_t \otimes I_G) \Gamma_{itT}^e (Z_t' \otimes I_G)],$$

where:

$$\begin{aligned} \Gamma_{itT}^e &= \delta_T^{-G} E \left\{ \left[\mathcal{K}^{(1)} \left(\frac{U_t}{\delta_T} \right) \mathcal{K}^{(1)} \left(\frac{U_t}{\delta_T} \right)' \right] | \mathcal{F}_{t-1} \right\} \\ &= \delta_T^{-G} \int_{\mathbb{R}^G} \mathcal{K}^{(1)} \left(\frac{u}{\delta_T} \right) \mathcal{K}^{(1)} \left(\frac{u}{\delta_T} \right)' f_t(u | \mathcal{F}_{t-1}) du \\ &= \int_{\mathbb{R}^G} \mathcal{K}^{(1)}(x) \mathcal{K}^{(1)}(x)' f_t(\delta_T x | \mathcal{F}_{t-1}) dx. \end{aligned}$$

Then by Assumption 1 we have that:

$$\begin{aligned} \mathbb{E} \left(\sum_{t=1}^T g_{tT} g'_{tT} \right) &= \mathbb{E} \left\{ (Z_t \otimes I_G) \left[\int_{\mathbb{R}^G} \mathcal{K}^{(1)}(x) \mathcal{K}^{(1)}(x)' f_t(\delta_T x | \mathcal{F}_{t-1}) dx \right] (Z_t' \otimes I_G) \right\} \\ &\longrightarrow \mathbb{E} [f_t(0 | \mathcal{F}_{t-1}) (Z_t \otimes I_G) \mathcal{M} (Z_t' \otimes I_G)] = B_0, \end{aligned}$$

by dominated convergence, since $E(\|Z_t\|^2) < \infty$, by Assumption 11, and since:

$$0 \leq \int_{\mathbb{R}^G} \mathcal{K}^{(1)}(x) \mathcal{K}^{(1)}(x)' f_t(\delta_T x | \mathcal{F}_{t-1}) dx \leq C \int_{\mathbb{R}^G} \mathcal{K}^{(1)}(x) \mathcal{K}^{(1)}(x)' dx,$$

in the positive semi-definite sense, by Assumptions 3 and 12. Thus:

$$\mathbb{E} \left[\sum_{t=1}^T (\lambda' g_{tT})^2 \right] = \lambda' B_0 \lambda + o(1).$$

In addition, from above we have that:

$$(\lambda' g_{tT}^e) = - (T \delta_T^G)^{-1/2} \delta_T^{G+1} \lambda' (Z_t \otimes I_G) \int_{\mathbb{R}^G} \mathcal{K}(x) f_t^{(1)}(\delta_T x | \mathcal{F}_{t-1}) dx,$$

and hence by CS it follows that:

$$(\lambda' g_{tT}^e)^2 \leq (T \delta_T^G)^{-1} \delta_T^{2G+2} \|\lambda' (Z_t \otimes I_G)\|^2 \left\| \int_{\mathbb{R}^G} \mathcal{K}(x) f_t^{(1)}(\delta_T x | \mathcal{F}_{t-1}) dx \right\|^2.$$

Now:

$$\begin{aligned} \left\| \int_{\mathbb{R}^G} \mathcal{K}(s) f_t^{(1)}(\delta_T s | \mathcal{F}_{t-1}) ds \right\|^2 &= \sum_{i=1}^G \left(\int_{\mathbb{R}^G} \mathcal{K}(s) f_{t,i}^{(1)}(\delta_T s | \mathcal{F}_{t-1}) ds \right)^2 \\ &= \sum_{i=1}^G \psi_{tT,i}^2 \leq G \left(\frac{G^2 C^2}{2} \delta_T^2 \right)^2, \end{aligned}$$

by Equation (A.19). In addition, $\|\lambda' (Z_t \otimes I_G)\|^2 \leq G \|\lambda\|^2 \|Z_t\|^2$, so we have that:

$$\begin{aligned} \sum_{t=1}^T (\lambda' g_{tT}^e)^2 &\leq (T \delta_T^G)^{-1} \delta_T^{2G+2} \left(\sum_{i=1}^G \psi_{tT,i}^2 \right) \left(\sum_{t=1}^T \|\lambda' (Z_t \otimes I_G)\|^2 \right) \\ &\leq T (T \delta_T^G)^{-1} \delta_T^{2G+2} G \left(\frac{G^2 C^2}{2} \delta_T^2 \right)^2 G \|\lambda\|^2 \left(T^{-1} \sum_{t=1}^T \|Z_t\|^2 \right), \end{aligned}$$

and hence that:

$$\begin{aligned} \mathbb{E} \left[\sum_{t=1}^T (\lambda' g_{tT}^e)^2 \right] &\leq T (T \delta_T^G)^{-1} \delta_T^{2G+6} \left(\frac{G^6 C^4}{4} \right) \|\lambda\|^2 \mathbb{E} \left(T^{-1} \sum_{t=1}^T \|Z_t\|^2 \right) \\ &= O(\delta_T^{G+6}) = o(1), \end{aligned}$$

since $E\left(T^{-1} \sum_{t=1}^T \|Z_t\|^2\right) = E(\|Z_t\|^2) < \infty$, by Assumptions 1 and 11, and since $\delta_T^{G+6} = o(1)$, by assumption 13. Hence we have that:

$$\Sigma_T = E \left[\sum_{t=1}^T (\lambda' g_{tT})^2 \right] - E \left[\sum_{t=1}^T (\lambda' g_{tT}^e)^2 \right] = \lambda' B_0 \lambda + o(1),$$

and since B_0 is non-singular, by Lemma 7 above, then Σ_T is positive for all T sufficiently large and hence $\sum_{t=1}^T \text{Var}(\eta_{tT}) = 1$ for all T sufficiently large.

Third, observe that:

$$\sum_{t=1}^T z_{tT}^2 - \lambda' B_0 \lambda = W_{1,T} + W_{2,T},$$

where:

$$\begin{aligned} W_{1,T} &= \sum_{t=1}^T z_{tT}^2 - \sum_{t=1}^T E(z_{tT}^2 | \mathcal{F}_{t-1}), \\ W_{2,T} &= \sum_{t=1}^T E(z_{tT}^2 | \mathcal{F}_{t-1}) - \lambda' B_0 \lambda. \end{aligned}$$

Now define:

$$\phi_{tT} = (\lambda' g_{tT}^*)^2 - E[(\lambda' g_{tT}^*)^2 | \mathcal{F}_{t-1}] = z_{tT}^2 - E(z_{tT}^2 | \mathcal{F}_{t-1}),$$

so $W_{1,T} = \sum_{t=1}^T \phi_{tT}$, and observe that $\{(\phi_{tT}, \mathcal{F}_t)\}$ is a martingale difference array. By the von Bahr-Esseen inequality, see von Bahr and Esseen (1965), for any $1 < p \leq 2$:

$$E \left(\left| \sum_{t=1}^T \phi_{tT} \right|^p \right) \leq 2 \sum_{t=1}^T E(|\phi_{tT}|^p),$$

and by CR, J and the law of iterated expectations:

$$\begin{aligned} E(|\phi_{tT}|^p) &= E \left(\left| (\lambda' g_{tT}^*)^2 - E[(\lambda' g_{tT}^*)^2 | \mathcal{F}_{t-1}] \right|^p \right) \\ &\leq 2^{p-1} \left[E(|\lambda' g_{tT}^*|^{2p}) + E \left\{ \left| E[(\lambda' g_{tT}^*)^2 | \mathcal{F}_{t-1}] \right|^p \right\} \right] \\ &\leq 2^{p-1} E(|\lambda' g_{tT}^*|^{2p}) + 2^{p-1} E \left\{ E \left[\left| (\lambda' g_{tT}^*)^2 \right|^p | \mathcal{F}_{t-1} \right] \right\} \\ &= 2^p E(|\lambda' g_{tT}^*|^{2p}), \end{aligned}$$

so:

$$E \left(\left| \sum_{t=1}^T \phi_{tT} \right|^{2p} \right) \leq 2^{p+1} \sum_{t=1}^T E(|\lambda' g_{tT}^*|^{2p}).$$

But then by the law of iterated expectations:

$$\begin{aligned} \mathbb{E} \left(|\lambda' g_{tT}^*|^{2p} \right) &= \mathbb{E} \left(\left| \lambda' T^{-1/2} \delta_T^{-G/2} (Z_t \otimes I_G) \mathcal{K}^{(1)} \left(\frac{Y_t - A_0 Z_t}{\delta_T} \right) \right|^{2p} \right) \\ &= (T \delta_T^G)^{-p} \mathbb{E} \left\{ \mathbb{E} \left[\left(\left| \lambda' (Z_t \otimes I_G) \mathcal{K}^{(1)} \left(\frac{Y_t - A_0 Z_t}{\delta_T} \right) \right|^{2p} \right) \middle| \mathcal{F}_{t-1} \right] \right\}, \end{aligned}$$

while by CS and Assumption 3:

$$\begin{aligned} &\mathbb{E} \left[\left(\left| \lambda' (Z_t \otimes I_G) \mathcal{K}^{(1)} \left(\frac{Y_t - A_0 Z_t}{\delta_T} \right) \right|^{2p} \right) \middle| \mathcal{F}_{t-1} \right] \\ &\leq \|\lambda' (Z_t \otimes I_G)\|^{2p} \int_{\mathbb{R}^G} \left\| \mathcal{K}^{(1)} \left(\frac{u}{\delta_T} \right) \right\|^{2p} f_t(u | \mathcal{F}_{t-1}) du \\ &\leq G^p \|\lambda\|^{2p} \|Z_t\|^{2p} \delta_T^G \left(\int_{\mathbb{R}^G} \|\mathcal{K}^{(1)}(x)\|^{2p} f_t(\delta_T x | \mathcal{F}_{t-1}) dx \right) \\ &\leq G^p L_0 \|\lambda\|^{2p} \|Z_t\|^{2p} \delta_T^G \left(\int_{\mathbb{R}^G} \|\mathcal{K}^{(1)}(x)\|^{2p} dx \right). \end{aligned}$$

Thus:

$$\mathbb{E} (|W_{1,T}|^p) \leq 2^{p+1} G^p L_0 \|\lambda\|^{2p} (T \delta_T^G)^{-p+1} \left(\int_{\mathbb{R}^G} \|\mathcal{K}^{(1)}(s)\|^{2p} ds \right) \mathbb{E} \left(T^{-1} \sum_{t=1}^T \|Z_t\|^{2p} \right).$$

Now since $\sup_x \|\mathcal{K}^{(1)}(x)\| \leq C$, by Assumption 6, and $\int_{\mathbb{R}^G} \|\mathcal{K}^{(1)}(x)\| dx < \infty$, by Assumption 12, then $\int_{\mathbb{R}^G} \|\mathcal{K}^{(1)}(x)\|^{2p} dx < \infty$. In addition, since $1 < p \leq 2$ then $E(\|Z_t\|^{2p}) < \infty$, by Assumption 11, and hence $E\left(T^{-1} \sum_{t=1}^T \|Z_t\|^{2p}\right) = O_p(1)$ by Assumption 1 and the ergodic theorem. In addition, $T \delta_T^G \rightarrow \infty$, by Assumption 13, and hence $(T \delta_T^G)^{-p+1} = o(1)$. But then $E(|W_{1,T}|^p) = o(1)$ so $W_{1,T} = o_p(1)$ since L_p convergence implies convergence in probability.

Fourth, observe that:

$$\begin{aligned} \mathbb{E} (z_{tT}^2 | \mathcal{F}_{t-1}) &= \mathbb{E} \left\{ [\lambda' (g_{tT} - g_{tT}^e)]^2 \middle| \mathcal{F}_{t-1} \right\} \\ &= \mathbb{E} \left[(\lambda' g_{tT})^2 \middle| \mathcal{F}_{t-1} \right] - (\lambda' g_{tT}^e)^2, \end{aligned}$$

and hence:

$$W_{2,T} = \sum_{t=1}^T \mathbb{E} \left[(\lambda' g_{tT})^2 \middle| \mathcal{F}_{t-1} \right] - \sum_{t=1}^T (\lambda' g_{tT}^e)^2 - \lambda' B_0 \lambda.$$

Now, observe that:

$$\mathbb{E} (g_{tT} g_{tT}' | \mathcal{F}_{t-1}) = T^{-1} (Z_t \otimes I_G) \left[\int_{\mathbb{R}^G} \mathcal{K}^{(1)}(x) \mathcal{K}^{(1)}(x)' f_t(\delta_T s | \mathcal{F}_{t-1}) dx \right] (Z_t' \otimes I_G),$$

and define $B_{T,0} = B_T(\alpha_0)$, where $B_T(\cdot)$ is the same as in the statement of Lemma 12. Then CS and T imply that:

$$\begin{aligned}
& \mathbb{E} \left(\left| \lambda' \left[\sum_{t=1}^T \mathbb{E}(g_{tT} g'_{tT} | \mathcal{F}_{t-1}) \right] \lambda - \lambda' B_{T,0} \lambda \right| \right) \\
& \leq G \|\lambda\|^2 \mathbb{E} \left[\int_{\mathbb{R}^G} \|Z_t\|^2 \|\mathcal{K}^{(1)}(x)\|^2 |f_t(\delta_T x | \mathcal{F}_{t-1}) - f_t(0 | \mathcal{F}_{t-1})| dx \right] \\
& \leq G \|\lambda\|^2 \delta_T \mathbb{E} \left[\int_{\mathbb{R}^G} \|Z_t\|^2 \|x\| \|\mathcal{K}^{(1)}(x)\|^2 \left\| f_t^{(1)}(\vartheta \delta_T x | \mathcal{F}_{t-1}) \right\| dx \right] \\
& \leq G \|\lambda\|^2 \delta_T C^2 \mathbb{E} \{ \|Z_t\|^2 \} = o(1)
\end{aligned}$$

where $0 \leq \vartheta \leq 1$ by MVT, since $\left\| \sup_u f_t^{(1)}(u | \mathcal{F}_{t-1}) \right\| \leq C$, by Assumption 9, $\int_{\mathbb{R}^G} \|x\| \|\mathcal{K}^{(1)}(x)\|^2 dx \leq C$, by Assumption 12, and $E\{\|Z_t\|^2\} < \infty$, by Assumption 11. It follows by M that:

$$\lambda' \left[\sum_{t=1}^T \mathbb{E}(g_{tT} g'_{tT} | \mathcal{F}_{t-1}) \right] \lambda - \lambda' B_{T,0} \lambda = o_p(1),$$

and since $B_{T,0}$ converges in probability to B_0 , by Lemma 13, then:

$$W_{2T} = \lambda' \left[\sum_{t=1}^T \mathbb{E}(g_{tT} g'_{tT} | \mathcal{F}_{t-1}) \right] \lambda - \lambda' B_0 \lambda = o_p(1).$$

Since $W_{1T} = o_p(1)$ from earlier it follows that:

$$\sum_{t=1}^T z_{tT}^2 = \lambda' B_0 \lambda + o_p(1).$$

Fifth, note that for any $p > 1$ such that $E(|z_{tT}|^{2p}) < \infty$ for all (t, T) then by M:

$$\begin{aligned}
\Pr \left\{ \max_{1 \leq t \leq T} |z_{tT}| > \varepsilon \right\} &= \Pr \left\{ \max_{1 \leq t \leq T} |z_{tT}|^{2p} > \varepsilon^{2p} \right\} \leq \sum_{t=1}^T \Pr \{ |z_{tT}|^{2p} > \varepsilon^{2p} \} \\
&\leq \varepsilon^{-2p} \sum_{t=1}^T \mathbb{E} (|z_{tT}|^{2p}).
\end{aligned}$$

Now by CR, J and the law of iterated expectations:

$$\begin{aligned}
\mathbb{E} (|z_{tT}|^{2p}) &= \mathbb{E} \left(|\lambda' (g_{tT} - g_{tT}^e)|^{2p} \right) = \mathbb{E} \left(|(\lambda' g_{tT}) - (\lambda' g_{tT}^e)|^{2p} \right) \\
&\leq 2^{2p-1} \left[\mathbb{E} \left(|\lambda' g_{tT}|^{2p} \right) + \mathbb{E} \left(|\mathbb{E}(\lambda' g_{tT} | \mathcal{F}_{t-1})|^{2p} \right) \right] \\
&= 2^{2p-1} \left[\mathbb{E} \left(|\lambda' g_{tT}|^{2p} \right) + \mathbb{E} \left\{ \mathbb{E} \left(|\lambda' g_{tT}|^{2p} \right) | \mathcal{F}_{t-1} \right\} \right] = 2^p \mathbb{E} \left(|\lambda' g_{tT}|^{2p} \right).
\end{aligned}$$

But:

$$\lambda' g_{tT} = (T\delta_T^G)^{-1/2} \lambda' (Z_t \otimes I_G) \mathcal{K}^{(1)} \left(\frac{U_t}{\delta_T} \right),$$

so by CS and the law of iterated expectations:

$$\begin{aligned} \mathbb{E} \left(|\lambda' g_{tT}|^{2p} \right) &\leq (T\delta_T^G)^{-p} \mathbb{E} \left\{ \|\lambda' (Z_t \otimes I_G)\|^{2p} \left\| \mathcal{K}^{(1)} \left(\frac{U_t}{\delta_T} \right) \right\|^{2p} \right\} \\ &= (T\delta_T^G)^{-p} \mathbb{E} \left[\|\lambda' (Z_t \otimes I_G)\|^{2p} \mathbb{E} \left\{ \left\| \mathcal{K}^{(1)} \left(\frac{U_t}{\delta_T} \right) \right\|^{2p} \middle| \mathcal{F}_{t-1} \right\} \right]. \end{aligned}$$

Now:

$$\begin{aligned} \mathbb{E} \left\{ \left\| \mathcal{K}^{(1)} \left(\frac{U_t}{\delta_T} \right) \right\|^{2p} \middle| \mathcal{F}_{t-1} \right\} &= \int_{\mathbb{R}^G} \left\| \mathcal{K}^{(1)} \left(\frac{u}{\delta_T} \right) \right\|^{2p} f_t(u | \mathcal{F}_{t-1}) du \\ &= \int_{\mathbb{R}^G} \delta_T^G \|\mathcal{K}^{(1)}(s)\|^{2p} f_t(\delta_T s | \mathcal{F}_{t-1}) ds \\ &\leq C^{2p} \delta_T^G, \end{aligned}$$

since $\sup_u f_t(u | \mathcal{F}_{t-1}) \leq C$ for all u , by Assumption 3, $\sup_x \|\mathcal{K}^{(1)}(x)\| \leq C$, by Assumption 6, and $\int_{\mathbb{R}^G} \|\mathcal{K}^{(1)}(x)\|^2 dx \leq C$, by Assumption 12. Therefore we have that:

$$\begin{aligned} \Pr \left\{ \max_{1 \leq t \leq T} |z_{tT}| > \varepsilon \right\} &= \varepsilon^{-2p} 2^{2p} T (T\delta_T^G)^{-p} \mathbb{E} \left(\|\lambda' (Z_t \otimes I_G)\|^{2p} \right) C^{2p} \delta_T^G \\ &\leq O \left[(T\delta_T^G)^{-(p-1)} \right] = o(1), \end{aligned}$$

since $E \left(\|\lambda' (Z_t \otimes I_G)\|^{2p} \right) < \infty$ for all $1 < p \leq 2$, by Assumption 11, and $T\delta_T^G \rightarrow \infty$, by Assumption 13.

This then establishes that $\sum_{t=1}^T z_{tT} \xrightarrow{d} \mathcal{N}(0, \lambda' B_0 \lambda)$ for all fixed λ and hence we have that:

$$\sum_{t=1}^T g_{tT} \xrightarrow{d} \mathcal{N}(0, B_0).$$

■

REFERENCES

- Davidson, J.E.H. (1994). *Stochastic limit theory*. Oxford: Oxford University Press.
- Freedman, D.A. (1975). "On tail probabilities for martingales," *Annals of Probability*, 3, 100-118.
- Hayashi, F. (2000). *Econometrics*. Princeton (NJ): Princeton University Press.

- Newey, W.K. and McFadden, D. (1994). "Large sample estimation and hypothesis testing," in Engle, R.F. and McFadden, D. (eds.), *Handbook of Econometrics*, Vol. 4, Ch. 36, 2111-2245, Amsterdam: Elsevier.
- von Bahr, B. and Esseen, C.-G. (1965). "Inequalities for the r th absolute moment of a sum of random variables, $1 \leq r \leq 2$," *The Annals of Mathematical Statistics*, 36, 299-303.