ISSN 1755-5361

University of Essex

Department of Economics

Discussion Paper Series

No. 761 February 2015

Dynamic Vector Mode Regression

Gordon C R Kemp, Paulo M D C Parente and
J M C Santo Silva

Dynamic vector mode regression*

Gordon C.R. Kemp[†] Paulo M.D.C Parente[‡] J.M.C. Santos Silva[§]

9 February 2015

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 allowing for possibly dependent data. We specifically consider the estimation of vector autoregressive conditional mode models and of structural systems of linear simultaneous equations defined by mode restrictions. The proposed estimator is easy to implement using standard software and the results of a small simulation study suggest that it is well behaved in finite samples.

JEL classification code: C13, C14, C32, C36.

Key words: Multivariate conditional mode, Multivariate density estimation, Robust regression, Simultaneous equations, Vector autoregression.

---

*We are grateful to Marcus Chambers, Davide Delle Monache, Abhimanyu Gupta, Stepana Lazarova, Myoung-jae Lee, and Martin Weale for many helpful comments and discussions. The usual disclaimer applies. Santos Silva gratefully acknowledges partial financial support from Fundação para a Ciência e Tecnologia (Programme PEst-OE/EGE/UI0491/2013).

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

[‡]Instituto Universitário de Lisboa (ISCTE-IUL), Business Research Unit (BRU-IUL). Email: pmdcp@iscte.pt.

[§]Department of Economics, University of Essex and CEMAPRE. E-mail: jmcss@essex.ac.uk.

1. INTRODUCTION

The mode is an interesting measure of location for multivariate distributions, 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.[1] Indeed, the multivariate mean is well known not to be robust and other measures of location are not easy to generalize to the multivariate case. For example, Koenker (2005, p. 272) states that the "search for a satisfactory notion of multivariate quantiles has become something of a quest for the statistical holy grail". The interest of the multivariate mode is reflected by 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. When the variates have skewed distributions and are not conditionally independent, the difference between the modal value of the pair and the pair of marginal modal (or mean) values can be substantial. Likewise, the conditional multivariate mode may also be of interest as a predictor.[2] For example, the Bank of England's quarterly *Inflation Report* presents parametrically estimated mode-based forecasts for the inflation

[1]The importance of possible outliers in a multivariate context is highlighted, for example, in Tsay, Peña, and Pankratz (2000) and Galeano, Peña, and Tsay (2006).

[2]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), who noted that, for a given level of confidence, prediction intervals constructed around the conditional mode are generally shorter than those constructed around other predictors.

and output, but it might be also 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.[3] 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.[4] We derive the asymptotic properties of the estimator allowing for possibly 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.

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 study 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.

[3] As in Lee (1989, 1993) and Kemp and Santos Silva (2012), the 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.

[4] See also the related work by Yao and Li (2014a and 2014b).

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 the results of an illustrative simulation study, and Section 5 concludes. The proofs of all theorems and other technical details are presented in an 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). However, all these systems are generally interpreted as representing conditional expectations, whereas we will consider the case in which the system defines a conditional multivariate mode.

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 $\{(Y_{t-1-j}, Z_{t-j})\}_{j=0}^{\infty}$. Also, let $\mathcal{P} = (\Omega, \mathcal{F}, P)$ 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 P is a probability measure. We are interested in the case where the conditional mode of U_t given $\mathcal{F}_{t-1}$, denoted $\text{Mode}\left(U_t|\mathcal{F}_{t-1}\right)$, 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 $\text{Mode}\left(Y_t|\mathcal{F}_{t-1}\right)$, satisfies:

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

As in the pioneering work of Lee (1989, 1993) and in Kemp and Santos Silva (2012), we obtain our estimator for the $(GK \times 1)$ vector $\alpha_0 \equiv \text{vec}\left(A_0\right)$ 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

$$\text{L}_t\left(Y_t, Z_t, \alpha_A\right) = 1 - \gamma \mathcal{K}\left(\frac{Y_t - AZ_t}{\delta_T}\right), \tag{2}$$

where $\alpha_A \equiv \text{vec}\left(A\right)$, $\mathcal{K}\left(\cdot\right)$ denotes a multivariate smooth kernel function, $\gamma = \mathcal{K}\left(0\right)^{-1}$ is a scaling constant, and δ_T is a non-stochastic strictly positive bandwidth that depends on T.[5] Notice that, as the bandwidth approaches 0, $\text{L}_t\left(Y_t, Z_t, \alpha_A\right)$ 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).[6] Therefore, as shown below, the minimizer of the expectation of $\text{L}_t\left(Y_t, Z_t, \alpha_A\right)$ will approach the conditional mode as $\delta_T \to 0$.

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

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

which leads to the estimator of interest, a multivariate version of the mode regression estimator of Kemp and Santos Silva (2012):

$$\widehat{\alpha}_T = \arg\max_{A \in \mathcal{A}} Q_T\left(\alpha_A\right). \tag{4}$$

[5] For simplicity, here we consider the same bandwidth for all equations. However, all our results hold if the scale of the bandwidth is equation specific, as in the simulations presented in Section 4.

[6] 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). As in our case, the mode also emerges as the optimal predictor when the tolerance goes to zero and therefore the step loss function approaches the 0-1 loss function.

Although many multivariate smooth kernels are available (see, for example, Scott, 1992), here we focus on the multiplicative standard normal kernel; that is, $\mathcal{K}\left(\frac{Y_t - AZ_t}{\delta_T}\right) = (2\pi)^{-G/2} \exp\left(-\frac{(Y_t - AZ_t)'(Y_t - AZ_t)}{2\delta_T^2}\right)$.[7] While this choice of kernel is not innocuous,[8] the multiplicative normal kernel has several important advantages and in particular, and in parallel with what was found by Kemp and Santos Silva (2012), this choice is attractive because it generates a loss function which has both the multivariate mode and the multivariate mean as minimizers in limiting cases.

Indeed, 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

$$\mathrm{E}\left[\exp\left(-\frac{(Y_t - AZ_t)'(Y_t - AZ_t)}{2\delta_T^2}\right)(Z_t \otimes I_G)(Y_t - AZ_t)\right] = 0. \tag{5}$$

It is clear that (5) 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 $\mathrm{E}[\mathrm{L}_t(Y_t, Z_t, \alpha_A)]$ is equivalent to estimating each equation by least squares. To put it differently, when $\mathcal{K}(\cdot)$ is the multiplicative standard normal kernel, minimizing $\mathrm{E}[\mathrm{L}_t(Y_t, Z_t, \alpha_A)]$ is equivalent to solving a set of moment conditions that estimate $\mathrm{Mode}(Y_t|\mathcal{F}_{t-1})$ when $\delta_T \to 0$, or $\mathrm{E}(Y_t|\mathcal{F}_{t-1})$ when $\delta_T \to \infty$.

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 $\mathrm{E}[\mathrm{L}_t(Y_t, Z_t, \alpha_A)]$ defines a measure of location which, in some sense,

[7] Notice, however, the our asymptotic results will be obtained under much more general conditions on the chosen kernel.

[8] As in Eddy (1980) and Romano (1988), it may be possible to obtain estimators with somewhat improved asymptotic properties by using different kernels. However, this would also require strengthening some assumptions.

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 δ_T not only determines the properties of the estimator but also, and more importantly, defines the conditional measure of central tendency that is estimated. Hence, δ_T should be chosen by the researcher and not determined by a data-driven method such as cross validation.

The moment conditions in (5) are also informative about the choice of algorithm to maximize (2). Because $Q_T(\alpha_A)$ is differentiable, it can be maximized using a Newton-type algorithm of the kind typically available in standard econometrics software. Moreover, (5) 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, (5) 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 $Q_T(\alpha_A)$.

2.2. Asymptotic results

We now consider the asymptotic properties of the estimator of $\alpha_0 = \text{vec}\,(A_0)$, which is defined by (4); the proofs of all theorems are provided in Appendix A1.

The following assumptions will be used in obtaining our results; 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., $\|M\| = [\text{trace}\,(M'M)]^{1/2}$, and use the following convention for the derivatives of a vector-valued function $F(a)$ with respect to the vector a: $F^{(1)}(a) \equiv \partial F(a)/\partial a'$, $F^{(2)}(a) \equiv \partial^2 F(a)/\partial a \partial a'$, $F^{(3)}(a) \equiv \partial \text{vec}(F^{(2)}(a))/\partial a'$.

1. *(Stationarity and Ergodicity)* $\{(Y_t, Z_t)\}_{t=-\infty}^{\infty}$ is a strictly stationary ergodic sequence of random vectors.

2. *(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 given $\mathcal{F}_{t-1}$, denoted by $f_t(\cdot|\mathcal{F}_{t-1})$, such that: *(i)* $\sup_{t,u,\omega} f_t(u|\mathcal{F}_{t-1}) < \infty$; *(ii)* $f_t(u|\mathcal{F}_{t-1}) \leq f_t(0|\mathcal{F}_{t-1})$ with equality if and only if $u = 0$; *(iii)* $f_t(u|\mathcal{F}_{t-1})$ continuous in u for all (t, ω) with $\omega \in \Omega$.

3. *(Parameter Space)* $\mathcal{A}$ is compact.

4. *(Moments I)* $E(\|Z_t\|) < \infty$.

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)* $\mathcal{K}(\cdot) : \mathbb{R}^G \to \mathbb{R}$ satisfies *(i)* $\int_{\mathbb{R}^G} \mathcal{K}(u)\, du = 1$ with $\int_{\mathbb{R}^G} |\mathcal{K}(u)|\, du < \infty$; *(ii)* $\sup_{u \in \mathbb{R}^G} |\mathcal{K}(u)| < \infty$; *(iii)* $\sup_{u \in \mathbb{R}^G} \left\|\mathcal{K}^{(1)}(u)\right\| < \infty$.

7. *(Bandwidth I)* $\{\delta_T\}_{T=1}^{\infty}$ is a sequence of finite strictly positive constants such that: *(i)* $\delta_T = o(1)$; *(ii)* $\ln(T) / \left(T\delta_T^G\right) = o(1)$.

8. *(Conditional Density II)* *(i)* $f_t(u|\mathcal{F}_{t-1})$ is three times differentiable with respect to u for all $\mathcal{F}_{t-1}$ such that $\sup_{t,u,\omega} \left\|f_t^{(j)}(u|\mathcal{F}_{t-1})\right\| < \infty$, $j = 1, 2, 3$; *(ii)* $f_t^{(2)}(0|\mathcal{F}_{t-1})$ is negative definite for all $\omega \in \Omega$.

9. *(Interior Parameter Value)* $\mathcal{A}$ has a non-empty interior, denoted $\text{int}(\mathcal{A})$ and $A_0 \in \text{int}(\mathcal{A})$.

10. *(Moments II)* $\mathrm{E}\left(\|Z_t\|^{G+4+\xi}\right) < \infty$, for some $\xi > 0$.

11. *(Kernel Function II)* *(i)* $\int_{\mathbb{R}^G} u\mathcal{K}(u)\, du = 0$; *(ii)* $\mathcal{K}(\cdot)$ is three times differentiable such that $\sup_u \left\|\mathcal{K}^{(j)}(u)\right\| < \infty$ for $j = 2, 3$; *(iii)* $\int_{\mathbb{R}^G} \left\|\mathcal{K}^{(j)}(u)\right\|^2 du < \infty$ for $j = 1, 2$; *(iv)* $\lim_{M\to\infty} \sup_{u:\|u\|\geq M} |\mathcal{K}(u)| = 0$; *(v)* $\lim_{M\to\infty} \sup_{u:\|u\|\geq M} \left\|\mathcal{K}^{(1)}(u)\right\| = 0$; *(vi)* $\int \|u\|^2 |\mathcal{K}(u)|\, du < \infty$.

12. *(Bandwidth II)* The sequence $\{\delta_T\}_{T=1}^{\infty}$ is such that: *(i)* $\frac{\ln(T)}{T\delta_T^{G+4}} = o(1)$; *(ii)* $T\delta_T^{G+6} = o(1)$.

These assumptions are remarkably similar to those in Kemp and Santos Silva (2012). Indeed, the major differences are obviously A1 and A2 and the fact that the other assumptions are adapted to take into account the multivariate nature of the problem being considered here. We note, however, that although A10 is analogous to Assumption B1 in

Kemp and Santos Silva (2012), it is potentially more restrictive in the context we are considering here. Indeed, if Z_t includes lagged values of Y_t, something that was not admitted by Kemp and Santos Silva (2012), A10 also imposes the existence of finite moments of Y_t, which otherwise is not required.

The following theorem establishes the existence of $\widehat{\alpha}_T$, the estimator of interest.

Theorem 1 (Existence) *Under Assumptions A1, A3, A6 and A7 there exists a random variable* $\widehat{\alpha}_T$ *such that:*

$$\begin{aligned} \Pr\left(\widehat{\alpha}_T \in \operatorname{vec}\left(\mathcal{A}\right)\right) &= 1, \\ \Pr\left(Q_T\left(\alpha_A\right) \leq Q_T\left(\widehat{\alpha}_T\right), \quad \forall A \in \mathcal{A}\right) &= 1, \end{aligned}$$

where:

$$\operatorname{vec}\left(\mathcal{A}\right) = \left\{\alpha : \alpha = \operatorname{vec}\left(A\right) \textit{ for } \textit{ some } A \in \mathcal{A}\right\}.$$

The consistency of $\widehat{\alpha}_T$ is established by the following theorem.

Theorem 2 (Consistency) *Under Assumptions A1–A7, then* $\widehat{\alpha}_T = \alpha_0 + o_p\left(1\right)$.

We next establish the asymptotic normality of $\widehat{\alpha}_T$ and its rate of convergence.

Theorem 3 (Asymptotic Normality I) *Under Assumptions A1–A12, defining:*

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

then:

$$\sqrt{T\delta_T^{G+2}}\left(\widehat{\alpha}_T - \alpha_0\right) \xrightarrow{d} \mathcal{N}\left(0, D_0^{-1} B_0 D_0^{-1}\right).$$

Given the restrictions imposed on the bandwidth, this result implies that $\widehat{\alpha}_T$ 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 A12).[9]

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

Theorem 4 (Consistent Asymptotic Variance Matrix) *Under Assumptions A1–A12, defining:*

$$\begin{aligned}\widehat{B}_T(\alpha_A) &\equiv (T\delta_T)^{-1}\sum_{t=1}^{T}(Z_t\otimes I_G)\,\mathcal{K}^{(1)}\left(\frac{Y_t-AZ_t}{\delta_T}\right)\mathcal{K}^{(1)}\left(\frac{Y_t-AZ_t}{\delta_T}\right)'(Z_t'\otimes I_G)\,,\\ \widehat{D}_T\left(\alpha_A\right) &\equiv \left(T\delta_T^{G+2}\right)^{-1}\sum_{t=1}^{T}(Z_t\otimes I_G)\,\mathcal{K}^{(2)}\left(\frac{Y_t-AZ_t}{\delta_T}\right)(Z_t\otimes I_G)'\,,\end{aligned}$$

then:

$$\widehat{D}_T\left(\widehat{\alpha}_T\right)^{-1}\widehat{B}_T\left(\widehat{\alpha}_T\right)\widehat{D}_T\left(\widehat{\alpha}_T\right)^{-1}=D_0^{-1}B_0D_0^{-1}+o_p\left(1\right).$$

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, which have been a centrepiece of econometrics since the very early days. In particular, we consider standard simultaneous equation systems of the form

$$Y_t'\Gamma_0+Z_t'\Psi_0=V_t',\qquad t=1,\ldots,T, \tag{6}$$

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 $\text{Mode}\left(V_t|\mathcal{F}_{t-1}\right)=0$, and Y_t, Z_t are defined as before. Additionally, we assume that Γ_0 is non-singular and note that (6) 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 (6) because of the evident simultaneity issue. However, it is possible to show

[9] As in Lee (1989, 1993), it is also possible to consider an estimator with a fixed bandwidth. Under suitable regularity conditions, such estimator is $\sqrt{T}$-consistent.

that our earlier results can be used to estimate the parameters of the reduced form of the model, which is given by

$$Y_t' = Z_t' A_0' + U_t', \tag{7}$$

with $A_0' = -\Psi_0 \Gamma_0^{-1}$ and $U_t' = V_t' \Gamma_0^{-1}$. Indeed, it is possible to show that $\text{Mode}\,(V_t|\mathcal{F}_{t-1}) = 0$ implies that $\text{Mode}\,(\Upsilon V_t|\mathcal{F}_{t-1}) = 0$ for any non-stochastic non-singular matrix Υ, and in turn this result implies that $\text{Mode}\,(U_t|\mathcal{F}_{t-1}) = 0$.[10] Hence, (7) 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 (6); these can involve only the elements of Γ_0 and Ψ_0, or also additional restrictions on the conditional distribution of V_t; we consider the two cases separately.

Restrictions on the conditional distribution of V_t are not needed when 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

$$\text{rank}((I_G \otimes A_0', I_{GK})', \Phi') = G\,(G+K)\,. \tag{8}$$

[10] Let S_t and W_t be two random vectors such that $S_t = \Upsilon W_t$, and let $f_{S_t}(s_t|\mathcal{F}_{t-1})$ and $f_{W_t}(w_t|\mathcal{F}_{t-1})$ denote the conditional density functions of S_t and W_t, respectively. Note that because $f_{S_t}(s_t|\mathcal{F}_{t-1}) = f_{W_t}\left(\Upsilon^{-1} s_t|\mathcal{F}_{t-1}\right) / \left|\det(\Upsilon)\right|$, we have that if $\text{Mode}\,(W_t|\mathcal{F}_{t-1}) = 0$, then $f_{S_t}(s_t|\mathcal{F}_{t-1}) = f_{W_t}\left(\Upsilon^{-1} s_t|\mathcal{F}_{t-1}\right) / \left|\det(\Upsilon)\right| \leq f_{W_t}(0|\mathcal{F}_{t-1}) / \left|\det(\Upsilon)\right| = f_{S_t}(0|\mathcal{F}_{t-1})$, and therefore $\text{Mode}\,(S_t|\mathcal{F}_{t-1}) = 0$. Uniqueness of the conditional mode of S_t follows from the fact that Υ is non-singular.

Note that condition (8) 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 can be implemented using the following two-stage procedure. First, obtain $\widehat{\alpha}_T$, $\widehat{B}_T(\widehat{\alpha}_T)$, and $\widehat{D}_T(\widehat{\alpha}_T)$ by estimating the transpose of (7) using the multivariate conditional mode estimator defined by (4). Second, estimate β_0^r by solving the following minimum distance problem:

$$\widehat{\beta}_T^r = \arg\min_{\beta^r \in \mathcal{B}^r} \left[\widehat{\alpha}_T + \text{vec}\left(\left(\Psi_r \Gamma_r^{-1}\right)'\right)\right]' \left[\widehat{Avar}(\widehat{\alpha}_T)\right]^{-1} \left[\widehat{\alpha}_T + \text{vec}\left(\left(\Psi_r \Gamma_r^{-1}\right)'\right)\right], \tag{9}$$

where $\widehat{Avar}(\widehat{\alpha}_T) = \widehat{D}_T(\widehat{\alpha}_T)^{-1} \widehat{B}_T(\widehat{\alpha}_T) \widehat{D}_T(\widehat{\alpha}_T)^{-1} \xrightarrow{p} D_0^{-1} B_0 D_0^{-1}$ and $\mathcal{B}^r$ denotes the parameter space of β^r.[11]

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

13. *(Identification)* The matrices Γ_0, A_0, and Φ are such that: *(i)* $\text{rank}(\Gamma_0) = G$; *(ii)* $\text{rank}((I_G \otimes A_0', I_{GK})', \Phi') = G(G+K)$.
14. *(Parameter Space - II)* $\mathcal{B}^r$ is compact.
15. *(Rank Condition)* $\text{rank}(C_0) = G(G+K) - \rho$.
16. *(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 A1–A7, A13 and A14:* $\widehat{\beta}_T^r \xrightarrow{p} \beta_0^r$.

[11] 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 $\widehat{\alpha}_T + \text{vec}\left(\left(\Psi_r \Gamma_r^{-1}\right)'\right) = 0$ for Γ_r and Ψ_r.

Then, theorems 1–4 imply the following result.

Theorem 6 (Asymptotic Normality II) *Under Assumptions A1–A16:*

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

There are models in which the available restrictions on Γ_0 and Ψ_0 are not enough to ensure that Assumption A13 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 (6) 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 (6) 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 and are left for future research. 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.

Without loss of generality, suppose that Γ is lower triangular so that (6) can be written as

$$\begin{aligned} y_{tg} &= \sum_{j=g+1}^{G} -\gamma_{jg} y_{tj} + \sum_{k=1}^{K} \psi_{kg} z_{tk} + v_{tg}, \quad g = 1, \ldots, G-1, \ \& \ t = 1, \ldots, T, \\ y_{tG} &= \sum_{k=1}^{K} \psi_{kG} z_{tk} + v_{tG}, \qquad t = 1, \ldots, T, \end{aligned}$$

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

$$\text{Mode}\left(y_{tG}|\mathcal{F}_{t-1}\right) = \sum_{k=1}^{K} \psi_{kG} z_{tk}. \tag{10}$$

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

$$\text{Mode}\left(y_{tg}|\mathcal{F}_{t-1}\right) = \sum_{j=g+1}^{G} -\gamma_{jg} y_{tj} + \sum_{k=1}^{K} \psi_{kg} z_{tk}, \qquad g = 1, \ldots, G-1. \tag{11}$$

Equations (10) and (11) show that in this case it is possible to estimate each equation separately by using the univariate estimator proposed in Kemp and Santos Silva (2012); Section 2 provides the asymptotic results needed for valid inference in this context.

4. SIMULATION EVIDENCE

In this section we present the results of simulation experiments illustrating the finite sample performance of the proposed estimator. In particular, in these experiments data for $t = -99, \ldots, T$ are generated from the system

$$y_{g,t} = a_{g0} + a_{g1} y_{1,t-1} + a_{g2} y_{2,t-1} + u_{g,t}, \qquad g \in \{1, 2\},$$

with $y_{g,-100} = 0$ and

$$u_{g,t} = \varepsilon_{g,t} \frac{\exp\left(h\left(y_{g,t-1} - 4\right)\right)}{\sqrt{\mathrm{Var}\left(\varepsilon_{g,t}\right)}}.$$

That is, for $h = 0$ the errors are homoskedastic with variance one, and for $h \neq 0$ the errors exhibit multiplicative heteroskedasticity.

We perform two sets of experiments. In the first one the errors $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are generated independently as the log of independent draws from gamma-distributed random variables with mean θ_g/κ_g and variance θ_g/κ_g^2, for $\theta_g, \kappa_g > 0$. As in Kemp and Santos Silva (2012), we set $\theta_g = \kappa_g$ to ensure that $\varepsilon_{g,t}$ has zero mode, and set $\kappa_1 = 5$ and $\kappa_2 = 0.05$ to generate distributions of the two errors with very different degrees of skewness. Because $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are independent, the system can be estimated either equation-by-equation or using the full-information estimator described in Subsection 2.1. In these experiments we use both estimators to gain some insight into the costs of the "curse of dimensionality" incurred when using the system estimator.

In the second set of experiments $\varepsilon_{1,t}$ is obtained as the log of independent draws from a gamma-distributed random variable with mean $\kappa_1^{-1}(\kappa_1 + 1)$ and variance $\kappa_1^{-2}(\kappa_1 + 1)$, and $\varepsilon_{2,t}$ is obtained as the product of $\exp(\varepsilon_{1,t})$ by the log of independent draws from a gamma-distributed random variable, independent of $\varepsilon_{1,t}$, with mean 1 and variance κ_2^{-1}. As shown in Appendix 2, the mode of the joint density $f(\varepsilon_{1,t}, \varepsilon_{2,t})$ is at $\varepsilon_{1,t} = \varepsilon_{2,t} = 0$, and we performed simulations with $\kappa_1 = \kappa_2 = 2$. Because in this second set of experiments $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ are not independent, the system-estimator has to be used.

Notice that, for both sets of experiments, in the homoskedastic case the conditional expectations $\mathrm{E}\left(y_{g,t}|y_{1,t-1}, y_{2,t-1}\right)$ are linear functions of the regressors with slopes a_{g1} and a_{g2}; that is, these parameters are identified by a standard vector autoregression when $h = 0$. Therefore, for the experiments with $h = 0$, it is possible to compare the efficiency of the mean and mode estimators of the slope parameters. However, in the heteroskedastic case the conditional expectations are non-linear in $y_{g,t-1}$ and therefore the parameters of interest cannot be consistently estimated using a standard linear vector autoregression.

In all experiments we set $a_{10} = a_{12} = a_{20} = a_{21} = 0$, and $a_{11} = a_{22} = 0.5$, and estimation is performed for $t = 1, \ldots, T$, with $T \in \{250, 1000, 4000, 16000\}$, and $h \in \{0.0, 0.2\}$. In each run we estimate both a standard vector autoregression and the autoregressive conditional mode model. The mode estimator was implemented using the (multivariate) iterative weighted least squares estimator described in Subsection 2.1, using equation specific smoothing parameters. In particular, the smoothing parameter for equation g, denoted $\delta_{g,T}$, is defined as $\delta_{g,T} = 1.6\text{MAD}_g T^{-d}$, where MAD_g denotes the median of the absolute deviation from the median least squares residual for equation g, and $d = 0.143$ for the univariate estimator used in the first set of experiments, and $d = 0.126$ for the bivariate estimator.[12]

Table 1 contains the mean and standard errors of the estimates of each parameter obtained in 10000 replicas of experiment 1. The results obtained when the (potentially misspecified) system is estimated by least squares (labelled "Mean") illustrate the well-known properties of the estimator in this context: the estimator is biased but consistent and its rate of convergence is $\sqrt{T}$. The results obtained both with the univariate and with the bivariate autoregressive conditional mode estimators (respectively labelled "Univariate mode" and "Bivariate mode") are reminiscent of those reported by Kemp and Santos Silva (2012) in that the slope parameters are generally estimated with little bias. As for the estimates of the intercepts, the biases are more noticeable, especially for $g = 2$ and $h = 0$, but naturally they decrease with the sample size. The results in Table 1 also show that, at least with this design, there is little to choose between the univariate and bivariate mode estimators. Indeed, the bivariate mode estimator, which uses a larger bandwidth, typically leads to slightly larger biases but to smaller standard errors, suggesting that the "curse of dimensionality" is not particularly severe in this context.

As noted above, the mean and both mode regression estimators identify the same slope parameters when $h = 0$. Therefore, for these cases, it is meaningful to compare the results obtained by mean and mode regression. As in Kemp and Santos Silva (2012), we find that for the low skewness case ($g = 1$, $\kappa_1 = 5$) the mean regression estimator has much smaller standard errors than the mode estimators, but that for the high skewness case

[12] For $G = 2$, the exponent of T has to be strictly between $\frac{-1}{6}$ and $\frac{-1}{6+G}$, and the rate of convergence improves as the exponent approaches its upper bound. The value of 1.6 as a scaling factor is inspired by Silverman's (1986, p. 48) rule-of-thumb and takes into account that for the normal distribution $\sigma = 1.4826\text{MAD}$.

Table 1: Simulation results with independent errors and $\kappa_1 = 5$ and $\kappa_2 = 0.05$

			$g=1$			$g=2$		
h	T	Regression	Const.	$y_{1,t-1}$	$y_{2,t-1}$	Const.	$y_{1,t-1}$	$y_{2,t-1}$
0.0	250	Mean	−0.226	0.488	0.000	−0.893	0.000	0.488
			(0.120)	(0.055)	(0.056)	(0.119)	(0.056)	(0.055)
		Univariate mode	−0.053	0.488	0.003	−0.301	0.001	0.502
			(0.250)	(0.115)	(0.119)	(0.087)	(0.041)	(0.041)
		Bivariate mode	−0.059	0.488	0.003	−0.324	0.000	0.501
			(0.227)	(0.104)	(0.108)	(0.087)	(0.040)	(0.040)
	1000	Mean	−0.222	0.497	0.000	−0.878	0.000	0.497
			(0.059)	(0.027)	(0.028)	(0.059)	(0.027)	(0.028)
		Univariate mode	−0.032	0.498	0.002	−0.250	0.000	0.501
			(0.153)	(0.071)	(0.073)	(0.042)	(0.020)	(0.019)
		Bivariate mode	−0.040	0.497	0.001	−0.276	0.000	0.501
			(0.133)	(0.062)	(0.063)	(0.041)	(0.019)	(0.019)
	4000	Mean	−0.220	0.499	0.000	−0.875	0.000	0.499
			(0.029)	(0.014)	(0.014)	(0.030)	(0.014)	(0.014)
		Univariate mode	−0.023	0.498	0.001	−0.207	0.000	0.500
			(0.096)	(0.045)	(0.045)	(0.021)	(0.010)	(0.010)
		Bivariate mode	−0.029	0.498	0.001	−0.237	0.000	0.500
			(0.080)	(0.037)	(0.038)	(0.020)	(0.009)	(0.010)
	16000	Mean	−0.220	0.500	0.000	−0.873	0.000	0.500
			(0.015)	(0.007)	(0.007)	(0.015)	(0.007)	(0.007)
		Univariate mode	−0.017	0.499	0.000	−0.169	0.000	0.500
			(0.062)	(0.029)	(0.029)	(0.011)	(0.005)	(0.005)
		Bivariate mode	−0.022	0.500	0.000	−0.200	0.000	0.500
			(0.050)	(0.023)	(0.023)	(0.010)	(0.005)	(0.005)
0.2	250	Mean	−0.102	0.469	0.000	−0.394	0.000	0.426
			(0.054)	(0.056)	(0.064)	(0.048)	(0.051)	(0.052)
		Univariate mode	−0.022	0.490	0.004	−0.122	0.000	0.496
			(0.115)	(0.113)	(0.133)	(0.034)	(0.037)	(0.037)
		Bivariate mode	−0.025	0.490	0.003	−0.132	0.000	0.494
			(0.104)	(0.103)	(0.121)	(0.034)	(0.036)	(0.036)
	1000	Mean	−0.100	0.478	0.000	−0.389	0.000	0.433
			(0.027)	(0.028)	(0.032)	(0.024)	(0.025)	(0.026)
		Univariate mode	−0.014	0.500	0.002	−0.100	0.000	0.498
			(0.070)	(0.069)	(0.080)	(0.016)	(0.018)	(0.017)
		Bivariate mode	−0.017	0.500	0.001	−0.111	0.000	0.497
			(0.061)	(0.060)	(0.070)	(0.016)	(0.017)	(0.017)
	4000	Mean	−0.099	0.480	0.000	−0.388	0.000	0.435
			(0.013)	(0.014)	(0.016)	(0.012)	(0.013)	(0.013)
		Univariate mode	−0.010	0.500	0.001	−0.082	0.000	0.499
			(0.044)	(0.043)	(0.050)	(0.008)	(0.009)	(0.009)
		Bivariate mode	−0.012	0.500	0.001	−0.094	0.000	0.498
			(0.037)	(0.036)	(0.042)	(0.008)	(0.009)	(0.009)
	16000	Mean	−0.099	0.481	0.000	−0.388	0.000	0.435
			(0.007)	(0.007)	(0.008)	(0.006)	(0.006)	(0.006)
		Univariate mode	−0.007	0.501	0.000	−0.066	0.000	0.500
			(0.029)	(0.028)	(0.032)	(0.004)	(0.005)	(0.004)
		Bivariate mode	−0.009	0.501	0.000	−0.079	0.000	0.499
			(0.023)	(0.023)	(0.026)	(0.004)	(0.004)	(0.004)

Table 2: Simulation results with dependent errors and $\kappa_1 = \kappa_2 = 2$

			$g = 1$			$g = 2$		
h	T	Regression	Const.	$y_{1,t-1}$	$y_{2,t-1}$	Const.	$y_{1,t-1}$	$y_{2,t-1}$
0.0	250	Mean	0.374	0.488	0.000	−0.295	0.000	0.489
			(0.080)	(0.056)	(0.057)	(0.082)	(0.057)	(0.055)
		Bivariate mode	0.167	0.489	−0.003	−0.063	−0.002	0.498
			(0.217)	(0.147)	(0.157)	(0.117)	(0.080)	(0.082)
	1000	Mean	0.367	0.497	0.000	−0.290	0.000	0.497
			(0.040)	(0.028)	(0.028)	(0.040)	(0.028)	(0.028)
		Bivariate mode	0.132	0.495	−0.003	−0.046	−0.001	0.499
			(0.140)	(0.096)	(0.098)	(0.065)	(0.046)	(0.046)
	4000	Mean	0.366	0.499	0.000	−0.288	0.000	0.499
			(0.020)	(0.014)	(0.014)	(0.020)	(0.014)	(0.014)
		Bivariate mode	0.102	0.499	0.000	−0.035	0.000	0.500
			(0.093)	(0.064)	(0.065)	(0.039)	(0.028)	(0.028)
	16000	Mean	0.366	0.500	0.000	−0.288	0.000	0.500
			(0.010)	(0.007)	(0.007)	(0.010)	(0.007)	(0.007)
		Bivariate mode	0.078	0.500	0.000	−0.026	0.000	0.500
			(0.064)	(0.044)	(0.045)	(0.025)	(0.017)	(0.017)
0.2	250	Mean	0.168	0.523	−0.001	−0.132	−0.001	0.466
			(0.036)	(0.056)	(0.071)	(0.036)	(0.049)	(0.054)
		Bivariate mode	0.068	0.513	−0.025	−0.027	−0.002	0.497
			(0.094)	(0.139)	(0.190)	(0.050)	(0.069)	(0.080)
	1000	Mean	0.166	0.532	0.000	−0.130	−0.001	0.473
			(0.018)	(0.027)	(0.035)	(0.018)	(0.024)	(0.027)
		Bivariate mode	0.052	0.514	−0.024	−0.019	−0.001	0.500
			(0.059)	(0.090)	(0.120)	(0.028)	(0.039)	(0.043)
	4000	Mean	0.165	0.534	0.000	−0.129	−0.001	0.476
			(0.009)	(0.014)	(0.017)	(0.009)	(0.012)	(0.014)
		Bivariate mode	0.040	0.514	−0.019	−0.015	0.000	0.501
			(0.039)	(0.059)	(0.081)	(0.017)	(0.023)	(0.025)
	16000	Mean	0.165	0.535	0.000	−0.129	−0.001	0.476
			(0.004)	(0.007)	(0.009)	(0.004)	(0.006)	(0.007)
		Bivariate mode	0.030	0.511	−0.017	−0.011	0.000	0.501
			(0.027)	(0.040)	(0.056)	(0.011)	(0.015)	(0.015)

($g = 2$, $\kappa_2 = 0.05$) the results of the mode estimators are slightly better than those obtained with the standard vector autoregression estimator.

Table 2 summarizes a similar set of results for experiment 2; overall, these results are in line with those obtained in experiment 1. In particular, the bivariate mode regression generally leads to estimates of the slope parameters that have little bias. However, we note that for $h = 0.2$ the estimates of a_{11} have a reasonably large bias that is slow to go down when the sample becomes larger.[13] Also, with this particular design the least

[13] To investigate this issue, we performed a set of simulations using $\delta_{g,T} = 0.8\text{MAD}_g T^{-0.126}$ and in this case this bias is very small even for $T = 250$; naturally the standard errors are larger.

squares estimator always has much lower standard errors than the mode estimator. This difference is especially noticeable for $g = 1$.

Overall, the results of the two sets of experiments are encouraging in that they suggest that the proposed mode estimators are likely to have a reasonable performance in moderately large samples.

5. CONCLUDING REMARKS

We introduce a full-system estimator of the conditional mode of a random vector, which extends the results of Kemp and Santos Silva (2012) to the multivariate case. We do this allowing for dynamic models and dependent data and, consequently, we also implicitly generalize the results of Kemp and Santos Silva (2012) to the time series context. The estimator we propose can be used in the estimation of dynamic vector mode autoregressive 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 the results of a small simulation study suggest that it is well behaved in finite samples.

Several avenues for future research are left open. For example, our results on the estimation of systems of simultaneous equations identified by restrictions on the structural parameters can be extended to cover the case where restrictions are non-linear. Also, as mentioned before, using an approach similar to that adopted by Sakata (2007) 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.

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 estimator for vector autoregressive 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.

APPENDIX

A1. Proofs

In this appendix we provide the proofs of all the theorems presented in Sections 2 and 3. 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). 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 $\|M\| = [\text{trace}(M'M)]^{1/2}$. Furthermore, we use the following convention for the derivatives of a vector-valued function $F(a)$ with respect to the vector a: $F^{(1)}(a) \equiv \partial F(a)/\partial a'$, $F^{(2)}(a) \equiv \partial^2 F(a)/\partial a \partial a'$, $F^{(3)}(a) \equiv \partial \text{vec}(F^{(2)}(a))/\partial a'$. Also, the constants L_j for $j = 0, 1, \ldots, 11$ are defined as follows:

$$L_0 = \sup_{t,u,\omega} f(u|\mathcal{F}_{t-1}) < \infty, \qquad L_1 = \sup_{u\in\mathbb{R}^G} \left\|\mathcal{K}^{(1)}(u)\right\| < \infty, \qquad L_2 = \int_{\mathbb{R}^G} \mathcal{K}(u)^2\, du < \infty,$$

which exist by Assumptions A2, A6 and A7, and:

$$\begin{aligned} L_3 &= \sup_{u\in\mathbb{R}^G} \left\|f_t^{(1)}(u|\mathcal{F}_{t-1})\right\|, \qquad L_4 = \sup_{u\in\mathbb{R}^G} \left\|f_t^{(2)}(u|\mathcal{F}_{t-1})\right\|, \qquad L_5 = \sup_{u\in\mathbb{R}^G} \left\|f_t^{(3)}(u|\mathcal{F}_{t-1})\right\|, \\ L_6 &= \sup_{u\in\mathbb{R}^G} \left\|\mathcal{K}^{(2)}(u)\right\|, \qquad L_7 = \sup_{u\in\mathbb{R}^G} \left\|\mathcal{K}^{(3)}(u)\right\|, \qquad L_8 = \int_{\mathbb{R}^G} \left\|\mathcal{K}^{(1)}(u)\right\|^2 du, \\ L_9 &= \int_{\mathbb{R}^G} \left\|\mathcal{K}^{(2)}(u)\right\|^2 du, \qquad L_{10} = \int_{\mathbb{R}^G} \|u\|^2 \, |\mathcal{K}(u)|\, du, \qquad L_{11} = E\left(\|Z_t\|^4\right), \end{aligned}$$

which exist and are finite by Assumptions A10, A8 and A11.

Proof of Theorem 1 Since $\mathcal{K}(\cdot)$ is differentiable, by Assumption A6(iii), it follows that $\mathcal{K}(\cdot)$ is continuous. Furthermore, since δ_T is finite and strictly positive for all T, by Assumption A7(i), it follows that $Q_T(A)$ is continuous with respect to A, except possibly on a subset of Ω with probability zero, and is a random variable with respect to $\mathcal{P}$ for every $A \in \mathcal{A}$, by Assumption A1. Since $\mathcal{A}$ is compact, by Assumption A3, then by Lemma 7.1 from Hayashi (2000) it follows that there exists a random variable $\widehat{\alpha}_T$ with respect to $\mathcal{P}$ such that $\Pr(\widehat{\alpha}_T \in \text{vec}(\mathcal{A})) = 1$ and that $Q_T(\alpha_A) \leq Q_T(\widehat{\alpha}_T)$ for all $A \in \mathcal{A}$ except possibly on a subset of Ω with probability 0, as desired. $\square$

Proof of Theorem 2 Lemma 3 below establishes that there exists a continuous function $Q_0(\cdot) : \text{vec}(\mathcal{A}) \to \mathbb{R}$ such that:

$$\lim_{T\to\infty} E\left[Q_T(\alpha_A)\right] = Q_0(\alpha_A), \qquad \forall A \in \mathcal{A},$$

and such that $Q_0(\alpha_A)$ achieves a unique strict global max on $\text{vec}(\mathcal{A})$ at $\alpha_A = \alpha_0 = \text{vec}(A_0)$. Lemma 4 below establishes that:

$$\sup_{A\in\mathcal{A}} \left|Q_T(\alpha_A) - Q_0(\alpha_A)\right| = o_p(1).$$

Since $\mathcal{A}$ is compact, by Assumption 3, and hence so too is $\text{vec}(\mathcal{A})$, then by Theorem 2.1 from Newey and McFadden (1994) it follows that the maximization estimator $\widehat{\alpha}_T$ from Theorem 1 converges in probability to α_0 as desired. □

Proof of Theorem 3 Since A_0 belongs to the interior of $\mathcal{A}$, by Assumption A9, it follows that a_0 belongs to the interior of $\text{vec}(\mathcal{A})$. Then, since $\widehat{\alpha}_T$ is a consistent estimator of α_0, by Theorem 2, it follows that:

$$\lim_{T\to\infty} \Pr\left(\left[\left.\frac{\partial Q_T(\alpha_A)}{\partial \alpha_A}\right|_{\alpha_A=\widehat{\alpha}_T}\right] = 0\right) = 1,$$

and hence that:

$$-\left(T\delta_T^{G+1}\right)^{-1}\sum_{t=1}^{T}(Z_t\otimes I_G)\,\mathcal{K}^{(1)}\left(\frac{Y_t-\widehat{A}_T Z_t}{\delta_T}\right) = o_p(1),$$

where $\text{vec}\left(\widehat{A}_T\right) = \widehat{\alpha}_T$. Now by a Taylor expansion around α_0 we have:

$$\begin{aligned} o_p(1) &= \left(T\delta_T^{G+1}\right)^{-1}\sum_{t=1}^{T}(Z_t\otimes I_G)\,\mathcal{K}^{(1)}\left(\frac{U_t}{\delta_T}\right) \\ &\quad -\left(T\delta_T^{G+2}\right)^{-1}\sum_{t=1}^{T}(Z_t\otimes I_G)\,\mathcal{K}^{(2)}\left(\frac{Y_t-\widehat{A}_T^{*} Z_t}{\delta_T}\right)(Z_t\otimes I_G)'\left(\widehat{\alpha}_T-\alpha_0\right), \end{aligned}$$

for some $\widehat{A}_T^{*}$ such that $\widehat{\alpha}_T^{*} = \text{vec}\left(\widehat{A}_T^{*}\right)$ lies on the line segment joining $\widehat{\alpha}_T$ and α_0. Multiplying both sides by $\sqrt{T\delta_T^{G+2}}$ gives:

$$\begin{aligned} o_p(1) &= \left(T\delta_T^{G}\right)^{-1/2}\sum_{t=1}^{T}(Z_t\otimes I_G)\,\mathcal{K}^{(1)}\left(\frac{U_t}{\delta_T}\right) \\ &\quad -\left[\left(T\delta_T^{G+2}\right)^{-1}\sum_{t=1}^{T}(Z_t\otimes I_G)\,\mathcal{K}^{(2)}\left(\frac{Y_t-\widehat{A}_T^{*} Z_t}{\delta_T}\right)(Z_t\otimes I_G)'\right]\sqrt{T\delta_T^{G+2}}\left(\widehat{\alpha}_T-\alpha_0\right). \end{aligned}$$

But:

$$\left(T\delta_T^{G+2}\right)^{-1}\sum_{t=1}^{T}\left(Z_t\otimes I_G\right)\mathcal{K}^{(2)}\left(\frac{Y_t-\widehat{A}_T^{*}Z_t}{\delta_T}\right)\left(Z_t\otimes I_G\right)'=D_0+o_p\left(1\right),$$

by Lemma 6, since $\widehat{\alpha}_T^{*}=\alpha_0+o_p\left(1\right)$ by Theorem 2. In addition:

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

by Lemma 7. Since B_0 and D_0 are symmetric and non-singular, by Lemma 5, then the desired result follows immediately. □

Proof of Theorem 4 First, observe that:

$$T^{-1}\delta_T^{-(G+2)}\sum_{t=1}^{T}\left(Z_t\otimes I_G\right)\mathcal{K}^{(2)}\left(\frac{Y_t-\widehat{A}_TZ_t}{\delta_T}\right)\left(Z_t\otimes I_G\right)'=D_0+o_p\left(1\right),$$

by Lemma 6 since $\widehat{\alpha}_T=\alpha_0+o_p\left(1\right)$ by Theorem 2. But D_0 is invertible, by Lemma 5, so $\widehat{D}_T\left(\alpha_A\right)^{-1}=D_0^{-1}+o_p\left(1\right)$.

Second, observe that:

$$\left(T\delta_T\right)^{-1}\sum_{t=1}^{T}\left(Z_t\otimes I_G\right)\mathcal{K}^{(1)}\left(\frac{Y_t-AZ_t}{\delta_T}\right)\mathcal{K}^{(1)}\left(\frac{Y_t-\widehat{A}_TZ_t}{\delta_T}\right)'\left(Z_t'\otimes I_G\right)=B_0+o_p\left(1\right)$$

by Lemma 8 since $\widehat{\alpha}_T=\alpha_0+o_p\left(1\right)$ by Theorem 2, and thus $\widehat{B}_T\left(\alpha_A\right)=B_0+o_p\left(1\right)$. The desired result then follows immediately. □

Lemma 1 *Let $\{X_t,\mathcal{G}_t\}_{t=0}^{\infty}$ be a martingale difference sequence with* $\Pr\left(|X_t|\leq c\right)=1$ *for some constant $c<\infty$; then for any constants $a,b>0$ and $n\in\mathbb{N}$:*

$$\Pr\left(\left|\sum_{t=1}^{n}X_t\right|\geq a\;\&\;\sum_{t=1}^{n}Var\left(X_t|\mathcal{G}_{t-1}\right)\leq b\right)\leq 2\exp\left(-\frac{a^2/2}{ac+b^2}\right).$$

Proof. This follows immediately from Freedman's inequality, see Freedman (1975, Proposition 2.1), noting that $\{\left(-X_t\right),\mathcal{G}_t\}$ is also a martingale with $Var\left(-X_t|\mathcal{G}_{t-1}\right)=V_t$. □

Lemma 2 *For any collection of pairs of random variables $\{(X_i,Y_i)\}_{i=1}^{n}$, where n is a finite constant, and any constants a and b then:*

$$\Pr\left(\sup_{1\leq i\leq n}X_i\geq a\right)\leq\sum_{i=1}^{n}\Pr\left(X_i\geq a\;\&\;Y_i\leq b\right)+\Pr\left(\sup_{1\leq i\leq n}Y_i>b\right).$$

Proof. Observe that:

$$\Pr\left(\sup_{1\le i\le n} X_i \ge a\right) \le \Pr\left(\sup_{1\le i\le n} X_i \ge a \,\&\, \sup_{1\le j\le n} Y_j \le b\right) + \Pr\left(\sup_{1\le i\le n} Y_i > b\right).$$

But $\sup_{1\le i\le n} X_i \ge a$ implies that there exists at least one $i \in \{1, 2, \dots, n\}$ such that $X_i \ge a$ so it follows that:

$$\Pr\left(\sup_{1\le i\le n} X_i \ge a \,\&\, \sup_{1\le j\le n} Y_j \le b\right) \le \sum_{i=1}^{n} \Pr\left(X_i \ge a \,\&\, \sup_{1\le j\le n} Y_j \le b\right),$$

while $\sup_{1\le j\le n} Y_j \le b$ means that $Y_j \le b$ for all $j \in \{1, 2, \dots, n\}$ so it follows that:

$$\Pr\left(X_i \ge a \,\&\, \sup_{1\le j\le n} Y_j \le b\right) \le \Pr\left(X_i \ge a \,\&\, Y_i \le b\right),$$

and hence:

$$\Pr\left(\sup_{1\le i\le n} X_i \ge a \,\&\, \sup_{1\le j\le n} Y_j \le b\right) \le \sum_{i=1}^{n} \Pr\left(X_i \ge a \,\&\, Y_i \le b\right).$$

□

Lemma 3 *Under Assumptions A1–A7, then there exists a function $Q_0(\cdot) : vec(\mathcal{A}) \to \mathbb{R}$ such that: (i) $Q_0(\alpha_A) = \lim_{T\to\infty} E[Q_T(\alpha_A)]$, for all $A \in \mathcal{A}$; (ii) $Q_0(\cdot)$ is continuous on $vec(\mathcal{A})$; and (iii) $Q_0(\alpha) \le Q_0(\alpha_0)$ for all $\alpha \in \mathcal{A}$ with equality if and only if $\alpha = \alpha_0$.*

Proof. First, for each t and each $T = 1, 2, \dots,$ define:

$$\begin{aligned} q_{tT}(\alpha_A) &\equiv \left(T\delta_T^G\right)^{-1} \mathcal{K}\left(\frac{Y_t - AZ_t}{\delta_T}\right), \\ q_{tT}^e(\alpha_A) &\equiv E\left[q_{tT}(\alpha_A) | \mathcal{F}_{t-1}\right] = \int_{\mathbb{R}^G} \delta_T^{-G} \mathcal{K}\left(\frac{u - (A - A_0) Z_t}{\delta_T}\right) f_t(u|\mathcal{F}_{t-1})\, du \\ &= T^{-1} \int_{\mathbb{R}^G} \mathcal{K}(s) f_t\left((A - A_0) Z_t + \delta_T s | \mathcal{F}_{t-1}\right) ds, \end{aligned}$$

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

$$Q_{t0}^e(\alpha_A, \delta) \equiv \int_{\mathbb{R}^G} \mathcal{K}(s) f_t\left((A - A_0) Z_t + \delta s | \mathcal{F}_{t-1}\right) ds,$$

so $q_{tT}^e(\alpha_A) = T^{-1} Q_{t0}^e(\alpha_A, \delta_T)$. Assumption A1 implies that:

$$\begin{aligned} E\left[Q_T(\alpha_A)\right] &= E\left[T q_{tT}(\alpha_A)\right] = E\left[T q_{tT}^e(\alpha_A)\right] = E\left[Q_{t0}^e(\alpha_A, \delta_T)\right] \\ &= \int_{\Omega} \int_{\mathbb{R}^G} \mathcal{K}(s) f_t\left((A - A_0) Z_t + \delta s | \mathcal{F}_{t-1}\right) ds dP(\omega), \end{aligned}$$

for all $A \in \mathcal{A}$. But $Q_{t0}^{e}(\alpha_A, \delta)$ is continuous in (α_A, δ) for all (t, ω) by dominated convergence since $|f_t(u|\mathcal{F}_{t-1}) \leq L_0|$ for all (u, t, ω), by Assumption A2, $[(A - A_0) Z_t + \delta s]$ is continuous in (α_A, δ) for all ω, and $\int_{\mathbb{R}^G} |\mathcal{K}(s)| \, ds < \infty$, by Assumption A6. Hence dominated convergence implies that:

$$
\begin{aligned}
\lim_{T \to \infty} E[Q_T(\alpha_A)] &= \lim_{T \to \infty} E[Q_{t0}^{e}(\alpha_A, \delta_T)] \\
&= \int_{\Omega} \int_{\mathbb{R}^G} \mathcal{K}(s) f_t((A - A_0) Z_t | \mathcal{F}_{t-1}) \, ds dP(\omega) \\
&= \int_{\Omega} f_t((A - A_0) Z_t | \mathcal{F}_{t-1}) \, dP(\omega) = Q_0(\alpha_A),
\end{aligned}
$$

since $\int_{\mathbb{R}^G} \mathcal{K}(s) \, ds = 1$, by Assumption A6, and $\delta_T = o(1)$, by Assumption A7, and also that $Q_0(\alpha_A)$ is continuous in α_A.

Second, by Assumption A2 then for any A:

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

while by Assumption A5 it follows that for any $A \neq A_0$ then exists a set $S \in \mathcal{F}_{t-1}$ with $P(S) > 0$ such that:

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

and hence it follows that for all $\alpha_A \neq \alpha_0$:

$$
Q_0(\alpha_A) < Q_0(0).
$$

Thus $Q_0(\alpha_A)$ achieves a unique strict global maximum over $\alpha_A \in \operatorname{vec}(\mathcal{A})$ at $\alpha_A = \alpha_0$, as desired. □

Lemma 4 *Under Assumptions A1–A7:*

$$
\sup_{A \in \mathcal{A}} |Q_T(\alpha_A) - Q_0(\alpha_A)| \leq o_p(1),
$$

where $Q_0(\cdot)$ is characterized as in Lemma 3.

Proof. Since $\mathcal{A}$ is compact, by Assumption 3, then there exists a constant $0 < J_1 < \infty$ such that for each $T = 1, 2, \ldots,$ then there exist $\mathcal{A}_{1T} \subset \mathcal{A}$ and a function $\bar{A}_{1T}(\cdot)$:

$\mathcal{A} \to \mathcal{A}_{1T}$ such that the number of elements in $\mathcal{A}_{1T}$ is less than or equal to $J_1 T^{(G+1)K}$ and $\sup_{A \in \mathcal{A}} \left\| A - \bar{A}_{1T}(A) \right\| \leq T^{-(G+1)/G}$. In addition, define the function $\bar{\alpha}_{1T}(\cdot)$ by $\bar{\alpha}_{1T}(\alpha_A) \equiv \text{vec}\left(\bar{A}_{1T}(A)\right)$ for all $A \in \mathcal{A}$.

Now for each $A \in \mathcal{A}$ define:

$$
\begin{aligned}
q_{tT}^{e}(\alpha_A) &\equiv E\left[q_{tT}(\alpha_A) | \mathcal{F}_{t-1}\right], \qquad Q_T^{e}(\alpha_A) \equiv \sum_{t=1}^{T} q_{tT}^{e}(\alpha_A), \\
q_{tT}^{*}(\alpha_A) &\equiv q_{tT}(\alpha_A) - q_{tT}^{e}(\alpha_A), \qquad Q_T^{*}(\alpha_A) \equiv \sum_{t=1}^{T} q_{tT}^{*}(\alpha_A),
\end{aligned}
$$

so for any given $A \in \mathcal{A}$:

$$
Q_T(\alpha_A) - Q_0(\alpha_A) = B_{1T}(\alpha_A) + B_{2T}(\bar{\alpha}_{1T}(\alpha_A)) + B_{3T}(\alpha_A) + B_{4T}(\alpha_A),
$$

where:

$$
\begin{aligned}
B_{1T}(\alpha_A) &\equiv Q_T(\alpha_A) - Q_T(\bar{\alpha}_{1T}(\alpha_A)), \\
B_{2T}(\alpha_A) &\equiv Q_T(\alpha_A) - Q_T^{e}(\alpha_A) = Q_T^{*}(\alpha_A), \\
B_{3T}(\alpha_A) &\equiv Q_T^{e}(\bar{\alpha}_{1T}(\alpha_A)) - Q_T^{e}(\alpha_A), \\
B_{4T}(\alpha_A) &\equiv Q_T^{e}(\alpha_A) - Q_0(\alpha_A),
\end{aligned}
$$

and hence:

$$
\begin{aligned}
\sup_{A \in \mathcal{A}} |Q_n(\alpha_A) - Q_0(\alpha_A)| \leq\ & \sup_{A \in \mathcal{A}} |B_{1T}(\alpha_A)| + \sup_{A \in \mathcal{A}_{1T}} |B_{2T}(\alpha_A)| \\
& + \sup_{A \in \mathcal{A}} |B_{3T}(\alpha_A)| + \sup_{A \in \mathcal{A}} |B_{4T}(\alpha_A)|.
\end{aligned}
$$

In order to establish the desired result, it suffices to establish that each of the terms on the right-hand-side of the above equation is $o_p(1)$.

First, observe that it follows from T that:

$$
\sup_{A \in \mathcal{A}} |B_{1T}(\alpha_A)| \leq \sum_{t=1}^{T} \sup_{A \in \mathcal{A}} |q_{tT}(\alpha_A) - q_{tT}(\bar{\alpha}_{1T}(\alpha_A))|,
$$

Now for any $A, A^{\dagger} \in \mathcal{A}$ it follows from MVT that:

$$
q_{tT}(\alpha_A) - q_{tT}(\alpha_{A^{\dagger}}) = -T^{-1}\delta_T^{-(G+1)} \left[\mathcal{K}^{(1)}\left(\frac{Y_t - A^{*} Z_t}{\delta_T}\right)\right]' (Z_t' \otimes I_G)(\alpha_A - \alpha_{A^{\dagger}}),
$$

for some A^* such that $\alpha_{A^*} = \mathrm{vec}\,(A^*)$ lies on the line segment joining α_A and $\alpha_{A^\dagger}$. Then applying CS gives:

$$\begin{aligned}
|q_{tT}(\alpha_A) - q_{tT}(\alpha_{A^\dagger})| &\leq G^{1/2} T^{-1} \delta_T^{-(G+1)} \left\| \mathcal{K}^{(1)} \left(\frac{Y_t - A_{1T}^*(A) Z_t}{\delta_T} \right) \right\| \|Z_t\| \, \|\alpha_A - \alpha_{A^\dagger}\| \\
&\leq G^{1/2} L_2 T^{-1} \delta_T^{-(G+1)} \|Z_t\| \, \|\alpha_A - \alpha_{A^\dagger}\|
\end{aligned}$$

since $\|Z_t \otimes I_G\| = G^{1/2} \|Z_t\|$ and since $\left\|\mathcal{K}^{(1)}(u)\right\| \leq L_2 < \infty$ for all $u \in \mathbb{R}^G$, by Assumption A6. Hence it follows that:

$$\sup_{A \in \mathcal{A}} |B_{1T}(\alpha_A)| \leq L_2 G^{1/2} T^{-(G+1)/G} \delta_T^{-(G+1)} \left(T^{-1} \sum_{t=1}^{T} \|Z_t\| \right).$$

But since $\{Z_t\}_{t=-\infty}^{\infty}$ is strictly stationary and ergodic, by Assumption A1, and since $E(\|Z_t\|) < \infty$, by Assumption A4, then $T^{-1} \sum_{t=1}^{T} \|Z_t\| = O_p(1)$ and hence:

$$\sup_{A \in \mathcal{A}} |B_{1T}(\alpha_A)| \leq L_2 G^{1/2} \left(T \delta_T^G\right)^{-(G+1)/G} O_p(1) = o_p(1),$$

since $\left(T \delta_T^G\right)^{-1} = o(1)$, by Assumption A7.

In addition, it follow from T that:

$$\sup_{A \in \mathcal{A}} |B_{3T}(\alpha_A)| \leq \sum_{t=1}^{T} \sup_{A \in \mathcal{A}} |q_{tT}^e(\bar{\alpha}_{1T}(\alpha_A)) - q_{tT}^e(\alpha_A)|.$$

But for any $A, A^\dagger \in \mathcal{A}$ then by J:

$$\begin{aligned}
|q_{tT}^e(\alpha_A) - q_{tT}^e(\alpha_{A^\dagger})| &= |E[(q_{tT}(\alpha_A) - q_{tT}(\alpha_{A^\dagger})) | \mathcal{F}_{t-1}]| \\
&\leq E[\{|q_{tT}(\alpha_A) - q_{tT}(\alpha_{A^\dagger})|\} | \mathcal{F}_{t-1}] \\
&\leq G^{1/2} L_2 T^{-1} \delta_T^{-(G+1)} E\{\|Z_t\|\} \|\alpha_A - \alpha_{A^\dagger}\|,
\end{aligned}$$

and hence:

$$\sup_{A \in \mathcal{A}} |B_{3T}(\alpha_A)| \leq L_2 G^{1/2} \left(T \delta_T^G\right)^{-(G+1)/G} \left(T^{-1} \sum_{t=1}^{T} \|Z_t\| \right) = o_p(1),$$

since $\left(T \delta_T^G\right)^{-1} = o(1)$, by Assumption A7.

Second, observe that for any $A \in \mathcal{A}$ and $T = 1, 2, \ldots,$ then by construction $\{(q_{tT}^*(\alpha_A), \mathcal{F}_t)\}_{t=-\infty}^{\infty}$ is a martingale difference sequence since: (a) $\mathcal{F}_{t-1} \subseteq \mathcal{F}_t$ for all

t; (b) $q_{tT}^{*}(\alpha_A)$ is measurable with respect to $\mathcal{F}_t$ for all t; (c) $E[q_{tT}^{*}(\alpha_A)|\mathcal{F}_{t-1}] = 0$ for all t; and (d) $E[q_{tT}^{*}(\alpha_A)] = 0$ for all t. Next, observe that since $|q_{tT}(\alpha_A)| \leq L_0 T^{-1}\delta_T^{-G} < \infty$ then $|q_{tT}^{*}(\alpha_A)| \leq 2L_0 T^{-1}\delta_T^{-G} < \infty$. Now define:

$$q_{tT}^{v}(\alpha_A) \equiv Var[q_{tT}^{*}(\alpha_A)|\mathcal{F}_{t-1}], \qquad Q_T^{v}(\alpha_A) \equiv \sum_{t=1}^{T} q_{tT}^{v}(\alpha_A),$$

so:

$$\begin{aligned} q_{tT}^{v}(\alpha_A) &= Var[q_{tT}(\alpha_A)|\mathcal{F}_{t-1}] \leq E\left[(q_{tT}(\alpha_A))^2|\mathcal{F}_{t-1}\right] \\ &= T^{-2}\int_{\mathbb{R}^G} \delta_T^{-2G}\mathcal{K}\left(\frac{u-(A-A_0)Z_t}{\delta_T}\right)^2 f_t(u|\mathcal{F}_{t-1})\,du \\ &= T^{-2}\int_{\mathbb{R}^G} \delta_T^{-G}\mathcal{K}(s)^2 f_t((A-A_0)Z_t+\delta_T s|\mathcal{F}_{t-1})\,ds \\ &\leq T^{-2}L_0\delta_T^{-G}\int_{\mathbb{R}^G}\mathcal{K}(s)^2\,ds = L_0L_2T^{-2}\delta_T^{-G}, \end{aligned}$$

since $f_t(u|\mathcal{F}_{t-1}) \leq L_0 < \infty$ for all t and u, by Assumption A2, which implies that:

$$\Pr\left(Q_T^{v}(\alpha_A) \leq L_0L_2T^{-1}\delta_T^{-G}\right) = 1.$$

Since $B_{2T}(\alpha_A) = \sum_{t=1}^{T} q_{tT}^{*}(\alpha_A)$, it follows from Lemmas 1 and 2 that for any fixed $\varepsilon > 0$:

$$\begin{aligned} \Pr\left(\left|\sup_{A\in\mathcal{A}_{1T}} B_{2T}(\alpha_A)\right| \geq \varepsilon\right) &= \sum_{A\in\mathcal{A}_{1T}} \Pr\left(|Q_T^{*}(\alpha_A)| \geq T\varepsilon \,\&\, Q_T^{v}(\alpha_A) \leq TL_2L_9\delta_T^{-G}\right) \\ &\leq 2J_1T^{(G+1)K}\exp\left(-\frac{\varepsilon^2/2}{2L_0T^{-1}\delta_T^{-G}\varepsilon + L_0L_2T^{-1}\delta_T^{-G}}\right) \\ &= 2J_1T^{(G+1)K}\exp\left(-\frac{\varepsilon^2T\delta_T^{G}}{4L_0\varepsilon+2L_0L_2}\right) = o(1), \end{aligned}$$

since $\ln(T)/\left(T\delta_T^{G}\right) = o(1)$ by Assumption A6. Since $\varepsilon > 0$ was arbitrary this implies that:

$$\sup_{A\in\mathcal{A}_{1T}} |B_{2T}(\alpha_A)| = o_p(1).$$

Third, observe that for any $A \in \mathcal{A}$:

$$\begin{aligned} B_{4T}(\alpha_A) &= \left[\sum_{t=1}^{T} q_{tT}^{e}(\alpha_A)\right] - Q_0(\alpha_A) \\ &= T^{-1}\sum_{t=1}^{T}\left\{Q_{t0}^{e}(\alpha_A,\delta_T) - E[Q_{t0}^{e}(\alpha_A,0)]\right\}, \end{aligned}$$

where $Q^e_{t0}(\alpha_A, \delta)$ is defined as in the proof of Lemma 3. Now for any $A \in \mathcal{A}$ and $\delta \in \mathbb{R}$, the sequence $\{Q^e_{t0}(A, \delta)\}_{t=-\infty}^{\infty}$ is strictly stationary and ergodic, by Assumption A1. Also, $|Q^e_{t0}(A, \delta)| \leq L_1 L_2 < \infty$, by Assumptions A2 and A6, so it follows from the uniform law of large numbers for stationary ergodic processes that:

$$\sup_{A \in \mathcal{A}} |B_{4T}(\alpha_A)| = \sup_{A \in \mathcal{A}} \left| T^{-1} \sum_{t=1}^{T} Q^e_{t0}(\alpha_A, \delta_T) - E\left[Q^e_{t0}(\alpha_A, 0)\right] \right| = o_p(1),$$

since $\delta_T = o(1)$, by Assumption A7, since $Q^e_{t0}(\alpha_A, \delta)$ is continuous in (α_A, δ) for all (t, ω) as established in the proof of Lemma 3, and since $\mathcal{A}$ is compact, by Assumption A3. □

Lemma 5 *Under Assumptions A1–A12, B_0 is symmetric positive definite and D_0 is symmetric negative definite.*

Proof. First, it is clear that $\mathcal{M}$ exists, by Assumption A11, and that $\mathcal{M}$ is symmetric, by construction. Now, for any fixed $(G \times 1)$ vector $c_1 \neq 0$:

$$\begin{aligned} c_1' \mathcal{M} c_1 &= \int_{\mathbb{R}^G} c_1' \mathcal{K}^{(1)}(u) \mathcal{K}^{(1)}(u)' c_1 \, du \\ &= \int_{\mathbb{R}^G} \left(\mathcal{K}^{(1)}(u)' c_1\right)^2 du. \end{aligned}$$

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

$$\widetilde{\mathcal{K}}_C(s) = \mathcal{K}(Cs).$$

It then follows that:

$$\frac{\partial \widetilde{\mathcal{K}}_C(s)}{\partial s_1} = \frac{d\mathcal{K}(Cs)}{ds_1} = \left[\mathcal{K}^{(1)}(Cs)\right]' \frac{\partial (C'S)}{\partial s_1} = \left[\mathcal{K}^{(1)}(Cs)\right]' c_1,$$

since $Cs = \sum_{j=1}^{G} c_j s_j$ where c_j is the j-th column of C, and hence $c_1' \mathcal{M} c_1 = 0$ if and only if $\frac{\partial \widetilde{\mathcal{K}}_C(s)}{\partial s_1} = 0$ for all s. But since C is non-singular then it follows from Assumption A11 that:

$$\lim_{M \to \infty} \sup_{s : \|s\| \geq M} \left| \widetilde{K}_C(s) \right| = 0.$$

Consequently $\frac{\partial \widetilde{\mathcal{K}}_C(s)}{\partial s_1} = 0$ can only be true for all s if $\widetilde{K}_C(s) = 0$ for all s and hence $\mathcal{K}(u) = 0$ for all u which contradicts Assumption 11. Thus there is no $c_1 \neq 0$ such that $\mathcal{K}^{(1)}(u)' c_1 = 0$ for almost all u 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 A2, and since $\mathcal{M}$ is positive definite, as established above. In addition, Assumption A5 implies that $\Pr(\text{vec}(AZ_t) = 0) < 1$. Together these imply that:

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

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 A8, 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 A8. Next, fix $A \in \mathbb{R}^G \times \mathbb{R}^K$ such that $A \neq 0$; then:

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

Then since $\Pr(\text{vec}(AZ_t) = 0) < 1$, by Assumption A5, 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 6 *Under Assumptions A1–A12, define:*

$$\begin{aligned} \widehat{D}_T(\alpha_A) &\equiv T^{-1}\delta_T^{-(G+2)}\sum_{t=1}^{T}(Z_t\otimes I_G)\,\mathcal{K}^{(2)}\left(\frac{Y_t-AZ_t}{\delta_T}\right)(Z_t\otimes I_G)', \\ D(\alpha_A) &\equiv E\left[(Z_t\otimes I_G)\,f_t^{(2)}\left((A-A_0)\,Z_t|\mathcal{F}_{t-1}\right)(Z_t\otimes I_G)'\right], \end{aligned}$$

for all $A\in\mathcal{A}$, where $\alpha_A=vec(A)$; then:

$$\sup_{A\in\mathcal{A}}\left\|\widehat{D}_T(\alpha_A)-D(\alpha_A)\right\|=o_p(1).$$

Proof. Since $\mathcal{A}$ is compact, by Assumption A3, then there is a constant $J_2<\infty$ such that for each $T=1,2...$, we can find $\mathcal{A}_{2T}\subset\mathcal{A}$ and a function $\bar{A}_{2T}(\cdot):A\to\mathcal{A}_{2T}$ such that the number of elements of $\mathcal{A}_{2T}$ is less than or equal to J_2T^{2GK} and that $\left\|A-\bar{A}_{2T}(A)\right\|\leq T^{-2}$ for all $A\in\mathcal{A}$. We then define the function $\bar{\alpha}_{2T}(\cdot)$ by $\bar{\alpha}_{2T}(\alpha_A)\equiv\text{vec}\left(\bar{A}_{2T}(A)\right)$ for all $A\in\mathcal{A}$.

From Lemma 3 it follows that there exists $Q_0(\cdot):\text{vec}(\mathcal{A})\to\mathbb{R}$ such that:

$$Q_0(\alpha_A)=\lim_{T\to\infty}E\left[Q_T(\alpha_A)\right],\qquad\forall A\in\mathcal{A},$$

and from the proof of Lemma 3 it follows that:

$$Q_0(\alpha_A)=E\left[f_t\left((A-A_0)\,Z_t|\mathcal{F}_{t-1}\right)\right].$$

Clearly $E\left[(Z_t\otimes I_G)(Z_t\otimes I_G)'\right]$ is finite for all $G\geq1$, by Assumption A10. In addition, since $f_t^{(j)}(u|\mathcal{F}_{t-1})$ is continuous in u for all $\omega\in\Omega$ and uniformly bounded from above for $j=0,1,2$, by Assumption A8 then we can interchange the order of derivatives. Therefore $\frac{\partial^2Q_0(\alpha_A)}{\partial\alpha_A\partial\alpha_A'}$ is well defined and is given by:

$$\frac{\partial^2Q_0(\alpha_A)}{\partial\alpha_A\partial\alpha_A'}=E\left[(Z_t\otimes I_G)\,f_t^{(2)}\left((A-A_0)\,Z_t|\mathcal{F}_{t-1}\right)(Z_t\otimes I_G)'\right]=D(\alpha_A).$$

Next, observe that:

$$\begin{aligned} \frac{\partial^2Q_T(\alpha_A)}{\partial\alpha_A\partial\alpha_A'} &= T^{-1}\sum_{t=1}^{T}\frac{\partial q_{tT}(\alpha_A)}{\partial\alpha_A\partial\alpha_A'} \\ &= T^{-1}\delta_T^{-(G+2)}\sum_{t=1}^{T}(Z_t\otimes I_G)\,\mathcal{K}^{(2)}\left(\frac{Y_t-AZ_t}{\delta_T}\right)(Z_t\otimes I_G)'=\widehat{D}_T(\alpha_A). \end{aligned}$$

Then fix any $\lambda \in \mathbb{R}^{GK}$ such that $\lambda \neq 0$ and define:

$$\begin{aligned}
H_0(\alpha) &\equiv \lambda' D(\alpha_A)\lambda, \\
h_{tT}(\alpha_A) &\equiv T^{-1}\lambda'\left(\frac{\partial q_{tT}(\alpha_A)}{\partial\alpha_A\partial\alpha_A'}\right)\lambda \\
&= T^{-1}\delta_T^{-(G+2)}\lambda'(Z_t\otimes I_G)\mathcal{K}^{(2)}\left(\frac{Y_t - AZ_t}{\delta_T}\right)(Z_t\otimes I_G)'\lambda \\
H_T(\alpha_A) &\equiv \sum_{t=1}^{T} h_{tT}(\alpha_A) = \lambda'\widehat{D}_T(\alpha_A)\lambda, \\
h_{tT}^e(\alpha_A) &\equiv E\left[h_{tT}(\alpha_A)|\mathcal{F}_{t-1}\right], \\
H_T^e(\alpha_A) &\equiv \sum_{t=1}^{T} h_{tT}^e(\alpha_A).
\end{aligned}$$

In addition, for any $A \in \mathcal{A}$ define:

$$\begin{aligned}
C_{1T}(\alpha_A) &\equiv \left[H_T(\alpha_A) - H_T(\bar{\alpha}_{2T}(\alpha_A))\right], \\
C_{2T}(\alpha_A) &\equiv \left[H_T(\alpha_A) - H_T^e(\alpha_A)\right], \\
C_{3T}(\alpha_A) &\equiv \left[H_T^e(\bar{\alpha}_{2T}(\alpha_A)) - H_T^e(\alpha_A)\right], \\
C_{4T}(\alpha_A) &\equiv \left[H_T^e(\alpha_A) - H_0(\alpha_A)\right],
\end{aligned}$$

and observe that:

$$H_T(\alpha_A) - H_0(\alpha_A) = C_{1T}(\alpha_A) + C_{2T}(\bar{\alpha}_{2T}(\alpha_A)) + C_{3T}(\alpha_A) + C_{4T}(\alpha_A),$$

and hence:

$$\begin{aligned}
\sup_{A\in\mathcal{A}}|H_T(\alpha_A) - H_0(\alpha_A)| \leq\ & \sup_{A\in\mathcal{A}}|C_{1T}(\alpha_A)| + \sup_{A\in\mathcal{A}_{2T}}|C_{2T}(\alpha_A)| \\
& + \sup_{A\in\mathcal{A}_{2T}}|C_{3T}(\alpha_A)| + \sup_{A\in\mathcal{A}}|C_{4T}(\alpha_A)|
\end{aligned}$$

In order to establish the desired result, it then suffices to establish that each of the terms on the right-hand-side of the above equation is $o_p(1)$.

First, for any $A, A^{\dagger} \in \mathcal{A}$ it follows by MVT that:

$$\begin{aligned}
h_{tT}(\alpha_A) - h_{tT}(\alpha_{A^{\dagger}}) &= T^{-1}\delta_T^{-(G+2)}\text{vec}\left[(Z_t\otimes I_G)'\lambda\lambda(Z_t'\otimes I_G)\right]' \\
&\quad\times\text{vec}\left[\mathcal{K}^{(2)}\left(\frac{Y_t - AZ_t}{\delta_T}\right) - \mathcal{K}^{(2)}\left(\frac{Y_t - A^{\dagger}Z_t}{\delta_T}\right)\right] \\
&= T^{-1}\delta_T^{-(G+3)}\text{vec}\left[(Z_t\otimes I_G)'\lambda\lambda(Z_t'\otimes I_G)\right]' \\
&\quad\times\left\{\left.\frac{\partial\text{vec}\left[\mathcal{K}^{(3)}(s)\right]}{\partial s'}\right|_{s=\delta_T^{-1}(Y_t - A^{*}Z_t)}\right\}(Z_t'\otimes I_G)\left(\alpha_A - \alpha_A^{\dagger}\right),
\end{aligned}$$

for some A^* such that $\alpha_{A^*} = \text{vec}\,(A^*)$ lies on the line segment joining α_A and $\alpha_{A^\dagger}$, where $\mathcal{K}_i^{(3)}(u) \equiv \partial \mathcal{K}^{(2)}(u)/\partial u_i$ and where $\alpha_{A,i}$ and $\alpha_{A^\dagger,i}$ are the i-th elements of α_A and $\alpha_{A^\dagger}$ respectively. Hence it follows by CS that:

$$\begin{aligned} |h_{tT}(\alpha_A) - h_{tT}(\alpha_{A^\dagger})| &\leq G^{3/2} T^{-1} \delta_T^{-(G+3)} \|\lambda\|^3 \|Z_t\|^3 \left\| \mathcal{K}^{(3)}\left(\frac{Y_t - A^* Z_t}{\delta_T}\right)\right\| \|\alpha_A - \alpha_{A^\dagger}\| \\ &\leq G^{3/2} L_7 \|\lambda\|^3 T^{-1} \delta_T^{-(G+3)} \|Z_t\|^3 \|\alpha_A - \alpha_{A^\dagger}\|, \end{aligned}$$

since $\left\|\mathcal{K}^{(3)}(u)\right\| \leq L_7 < \infty$, by Assumption A11. But for any $A \in \mathcal{A}$, it follows by T that:

$$\begin{aligned} |C_{1T}(\alpha_A)| &\leq \sum_{t=1}^{T} |h_{tT}(\alpha_A) - h_{tT}(\bar{\alpha}_{2T}(\alpha_A))| \\ &\leq G^{3/2} L_7 \|\lambda\|^3 T^{-1} \delta_T^{-(G+3)} \|\alpha_A - \bar{\alpha}_{2T}(\alpha_A)\| \left(\sum_{t=1}^{T} \|Z_t\|^3\right), \end{aligned}$$

and thus:

$$\begin{aligned} \sup_{A \in \mathcal{A}} |C_{1T}(\alpha_A)| &\leq L_7 G^{3/2} \|\lambda\|^3 T^{-1} \delta_T^{-(G+3)} \left(\sum_{t=1}^{T} \|Z_t\|^3\right) \\ &= L_7 G^{3/2} \|\lambda\|^3 \left[\frac{\delta_T^{(G+5)/2}}{T \delta_T^{G+4}}\right]^2 \left(T^{-1} \sum_{t=1}^{T} \|Z_t\|^3\right), \end{aligned}$$

since $\|\alpha_A - \bar{\alpha}_{2T}(\alpha_A)\| \leq T^{-2}$. But, $E\left(\|Z_t\|^3\right) < \infty$, by Assumption A10, and since $\{Z_t\}_{t=-\infty}^{\infty}$ is strictly stationary and ergodic, by Assumption 1, then it follows that $T^{-1} \sum_{t=1}^{T} \|Z_t\|^3 = O_p(1)$ by the ergodic theorem and hence that:

$$\sup_{A \in \mathcal{A}} |C_{1T}(\alpha_A)| = o_p(1).$$

since $\delta_T = o(1)$, by Assumption A6, and $\left(T \delta_T^{G+4}\right)^{-1} = o(1)$, by Assumption A12.

In addition, for any $A, A^\dagger \in \mathcal{A}$ it follows by J that:

$$\begin{aligned} |h_{tT}^e(\alpha_A) - h_{tT}^e(\alpha_{A^\dagger})| &= |E[h_{tT}(\alpha_A) - h_{tT}(\alpha_{A^\dagger}) | \mathcal{F}_{t-1}]| \\ &\leq E[\{|h_{tT}(\alpha_A) - h_{tT}(\alpha_{A^\dagger})|\} | \mathcal{F}_{t-1}] \\ &\leq G^{3/2} L_7 \|\lambda\|^3 T^{-1} \delta_T^{-(G+3)} \|Z_t\|^3 \|\alpha_A - \alpha_{A^\dagger}\|. \end{aligned}$$

But for any $A \in \mathcal{A}$ it follows from T that:

$$\begin{aligned} |C_{3T}(\alpha_A)| &= |H_T^e(\alpha_A) - H_T^e(\bar{\alpha}_{2T}(\alpha_A))| \leq \sum_{t=1}^{T} |h_{tT}^e(\alpha_A) - h_{tT}^e(\bar{\alpha}_{2T}(\alpha_A))| \\ &\leq G^{3/2} L_7 \|\lambda\|^3 T^{-1} \delta_T^{-(G+3)} \|\alpha_A - \bar{\alpha}_{2T}(\alpha_A)\| \left(\sum_{t=1}^{T} \|Z_t\|^3\right), \end{aligned}$$

so:

$$\sup_{A\in\mathcal{A}}\left|C_{3T}\left(\alpha_A\right)\right|\leq L_7 G^{3/2}\left\|\lambda\right\|^3 T^{-1}\delta_T^{-(G+3)}\left(\sum_{t=1}^{T}\left\|Z_t\right\|^3\right),$$

and hence it follows that:

$$\sup_{A\in\mathcal{A}}\left|C_{3T}\left(\alpha_A\right)\right|=o_p\left(1\right).$$

Second, define:

$$\begin{aligned}
h_{tT}^{*}\left(\alpha_A\right) &\equiv h_{tT}\left(\alpha_A\right)-h_{tT}^{e}\left(\alpha_A\right),\\
h_{tT,1}\left(\alpha_A\right) &\equiv \chi\left(\left\|\left(Z_t'\otimes I_G\right)\lambda\right\|^2\leq\delta_T^{-2}\right)h_{tT}\left(\alpha_A\right),\\
h_{tT,2}\left(\alpha_A\right) &\equiv \chi\left(\left\|\left(Z_t'\otimes I_G\right)\lambda\right\|^2>\delta_T^{-2}\right)h_{tT}\left(\alpha_A\right),\\
h_{tT,j}^{e}\left(\alpha_A\right) &\equiv E\left[h_{tT,j}\left(\alpha_A\right)\left|\mathcal{F}_{t-1}\right.\right],\qquad j=1,2\\
h_{tT,j}^{*}\left(\alpha_A\right) &\equiv h_{tT,j}\left(\alpha_A\right)-h_{tT,j}^{e}\left(\alpha_A\right),\qquad j=1,2,\\
H_{T,j}^{*}\left(\alpha_A\right) &\equiv \sum_{t=1}^{T}h_{tT,j}^{*}\left(\alpha_A\right),\qquad j=1,2,
\end{aligned}$$

where $\chi\left(\cdot\right)$ denotes the indicator function; this implies:

$$h_{tT}^{*}\left(\alpha_A\right)=h_{tT,1}^{*}\left(\alpha_A\right)+h_{tT,2}^{*}\left(\alpha_A\right),$$

and that

$$C_{2T}\left(\alpha_A\right)\equiv\left[H_T\left(\alpha_A\right)-H_T^{e}\left(\alpha_A\right)\right]=H_{T,1}^{*}\left(\alpha_A\right)+H_{T,2}^{*}\left(\alpha_A\right),$$

so that:

$$\sup_{A\in\mathcal{A}_{2T}}\left|C_{2T}\left(\alpha_A\right)\right|\leq\sup_{A\in\mathcal{A}_{2T}}\left|H_{T,1}^{*}\left(\alpha_A\right)\right|+\sup_{A\in\mathcal{A}_{2T}}\left|H_{T,2}^{*}\left(\alpha_A\right)\right|.$$

Now for any $A\in\mathcal{A}$ and $T=1,2,\ldots$, then by construction, $\left\{\left(h_{tT,1}^{*}\left(\alpha_A\right),\mathcal{F}_t\right)\right\}_{t=-\infty}^{\infty}$ is a martingale difference sequence. In addition:

$$\left|h_{1T,1}\left(\alpha_A\right)\right|\leq L_6 T^{-1}\delta_T^{-(G+2)}\delta_T^{-2},$$

since $\left\|K^{(2)}\left(s\right)\right\|\leq L_6$ for all u, by Assumption A11, and hence it follows from T and J that $\left|h_{tT,1}^{*}\left(\alpha_A\right)\right|\leq 2L_6T^{-1}\delta_T^{-(G+4)}<\infty$. Define:

$$h_{tT,1}^{v}\left(\alpha_A\right)\equiv Var\left[h_{tT,1}^{*}\left(\alpha_A\right)\left|\mathcal{F}_{t-1}\right.\right],\qquad H_{T,1}^{v}\left(\alpha_A\right)=\sum_{t=1}^{T}h_{tT,1}^{v}\left(\alpha_A\right),$$

and observe that since $E\left[h^*_{tT,1}|\mathcal{F}_{t-1}\right] = E\left[h^*_{tT,2}|\mathcal{F}_{t-1}\right] = 0$ and $E\left[h^*_{tT,1}h^*_{tT,2}|\mathcal{F}_{t-1}\right] = 0$ then $Cov\left[h^*_{tT,1}, h^*_{tT,2}|\mathcal{F}_{t-1}\right] = 0$ and hence:

$$\begin{aligned} h^v_{tT,1}(\alpha_A) &\leq Var\left[h^*_{tT}(\alpha_A)|\mathcal{F}_{t-1}\right] = Var\left[h_{tT}(\alpha_A)|\mathcal{F}_{t-1}\right] \leq E\left[h_{tT}(\alpha_A)^2|\mathcal{F}_{t-1}\right] \\ &\leq T^{-2}\delta_T^{-(2G+4)}\left\|(Z_t'\otimes I_G)\lambda\right\|^4 \int_{\mathbb{R}^G}\left\|\mathcal{K}^{(2)}\left(\frac{u-(A-A_0)Z_t}{\delta_T}\right)\right\|^2 f_t(u|\mathcal{F}_{t-1})\,du \\ &= T^{-2}\delta_T^{-(2G+4)}\left\|(Z_t'\otimes I_G)\lambda\right\|^4 \int_{\mathbb{R}^G}\left\|\mathcal{K}^{(2)}(s)\right\|^2 f_t((A-A_0)Z_t+\delta_T s|\mathcal{F}_{t-1})\,ds \\ &\leq G^2 L_0 L_9 \|\lambda\|^4 T^{-2}\delta_T^{-(G+4)}\|Z_t\|^4. \end{aligned}$$

noting that $f_t(u|\mathcal{F}_{t-1}) \leq L_0$ for all (t,u,ω), by Assumption A2, and $\int_{\mathbb{R}^G}\left\|\mathcal{K}^{(2)}(s)\right\|^2 ds = L_9$, by Assumption A11. But then:

$$H^v_{T,1}(\alpha_A) \leq G^2 L_0 L_9 \|\lambda\|^4 T^{-1}\delta_T^{-(G+4)}\left(T^{-1}\sum_{t=1}^T \|Z_t\|^4\right),$$

and since the right-hand-side of the above equation does not depend on α_A then:

$$\sup_{A\in\mathcal{A}_{2T}} H^v_{T,1}(\alpha_A) \leq G^2 L_0 L_9 \|\lambda\|^4 T^{-1}\delta_T^{-(G+4)}\left(T^{-1}\sum_{t=1}^T \|Z_t\|^4\right).$$

Now since the $\{Z_t\}$ are stationary and ergodic, by Assumption A1, and since $\mathrm{E}\left(\|Z_t\|^4\right) = L_{11} < \infty$, by Assumption 10, then:

$$\Pr\left(T^{-1}\sum_{t=1}^T \|Z_t\|^4 > 2L_{11}\right) = o(1),$$

by the ergodic theorem, and setting $b_T = 2G^2 L_0 L_9 L_{11}\|\lambda\|^4 T^{-1}\delta_T^{-(G+4)}$ then:

$$\Pr\left(\sup_{A\in\mathcal{A}_{2T}} H^v_{T,1}(\alpha_A) > b_T\right) = o(1),$$

and hence for any fixed $\varepsilon > 0$ it follows by Lemma 2 that:

$$\Pr\left(\sup_{A\in\mathcal{A}_{2T}}\left|H^*_{T,1}(\alpha_A)\right| \geq \varepsilon\right) \leq \sum_{A\in\mathcal{A}_{2T}} \Pr\left(\left|H^*_{T,1}(\alpha_A)\right| \geq \varepsilon \,\&\, H^v_{T,1}(\alpha_A) \leq b_T\right) + o(1),$$

since $\mathcal{A}_{2T}$ has a finite number of elements. Then form Lemma 1 it follows that:

$$\begin{aligned} &\Pr\left(\left|H^*_{T,1}(\alpha_A)\right| \geq \varepsilon \,\&\, H^v_{T,1}(\alpha_A) \leq b_T\right) \\ &\quad\leq \exp\left(-\frac{\varepsilon^2/2}{2L_6 T^{-1}\delta_T^{-(G+4)}\varepsilon + 2G^2 L_0 L_9 L_{11}\|\lambda\|^4 T^{-1}\delta_T^{-(G+4)}}\right) \\ &\quad= \exp\left(-\frac{T\delta_T^{(G+4)}\varepsilon^2}{4L_6\varepsilon + 4G^2 L_0 L_9 L_{11}\|\lambda\|^4}\right), \end{aligned}$$

and thus:

$$\begin{aligned}\Pr\left(\sup_{A\in\mathcal{A}_{2T}}\left|H^{*}_{T,1}(\alpha_A)\right|\geq\varepsilon\right) &\leq 2J_2T^{2GK}\exp\left(-\frac{T\delta_T^{(G+4)}\varepsilon^2}{4L_6\varepsilon+4G^2L_0L_9L_{11}\left\|\lambda\right\|^4}\right)+o(1)\\ &= o(1),\end{aligned}$$

since $\ln(T)/\left(T\delta_T^{G+4}\right)=o(1)$ by Assumption A12. Since $\varepsilon>0$ was arbitrary this implies that:

$$\sup_{A\in\mathcal{A}_{2T}}\left|H^{*}_{T,1}(\alpha_A)\right|=0.$$

In addition observe that since:

$$h_{tT,2}(\alpha_A)=h_{tT}(\alpha_A)\chi\left(\left\|(Z_t'\otimes I_G)\lambda\right\|^2>\delta_T^{-2}\right),$$

then:

$$\left|h_{tT,2}(\alpha_A)\right|\leq L_6T^{-1}\delta_T^{-(G+2)}\left\|(Z_t'\otimes I_G)\lambda\right\|^2\chi\left(\left\|(Z_t'\otimes I_G)\lambda\right\|^2>\delta_T^{-2}\right),$$

and hence by J and T:

$$\left|h^{*}_{tT,2}(\alpha_A)\right|\leq 2L_6T^{-1}\delta_T^{-(G+2)}\left\|(Z_t'\otimes I_G)\lambda\right\|^2\chi\left(\left\|(Z_t'\otimes I_G)\lambda\right\|^2>\delta_T^{-2}\right).$$

In addition, since the right-hand-side of the above inequality does not depend on A then:

$$\sup_{A\in A_{2T}}\left|h^{*}_{tT,2}(\alpha_A)\right|\leq 2L_6T^{-1}\delta_T^{-(G+2)}\left\|(Z_t'\otimes I_G)\lambda\right\|^2\chi\left(\left\|(Z_t'\otimes I_G)\lambda\right\|^2>\delta_T^{-2}\right),$$

and hence:

$$\begin{aligned}E\left(\sup_{A\in A_{2T}}\left|H^{*}_{T,2}(\alpha_A)\right|\right) &\leq \sum_{t=1}^{T}E\left\{\sup_{A\in A_{2T}}\left|h^{*}_{tT,2}(\alpha_A)\right|\right\}\\ &\leq 2L_6\delta_T^{-(G+2)}E\left\{\left\|(Z_t'\otimes I_G)\lambda\right\|^2\chi\left(\left\|(Z_t'\otimes I_G)\lambda\right\|^2>\delta_T^{-2}\right)\right\}.\end{aligned}$$

Now for any random variable X, constant $r>1$ such that $E(|X|^r)<\infty$, and constant $c>0$ then it follows by H and M that:

$$E\{|X|\chi(|X|>c)\}\leq c^{-(r-1)}E(|X|^r),$$

so setting $X=\left\|(Z_t'\otimes I_G)\lambda\right\|^2$, $r=(G+4+\xi)/2$, where ξ is given in Assumption A10, and $c=\delta_T^{-2}$ then:

$$E\left\{\left\|(Z_t'\otimes I_G)\lambda\right\|^2\chi\left(\left\|(Z_t'\otimes I_G)\lambda\right\|^2>\delta_T^{-2}\right)\right\}\leq\delta_T^{2(r-1)}E\left(\left\|(Z_t'\otimes I_G)\lambda\right\|^{2r}\right),$$

so:

$$E\left(\sup_{A\in A_{2T}}\left|H_{T,2}^{*}\left(\alpha_{A}\right)\right|\right)\leq 2L_{6}\delta_{T}^{-(G+2)}\delta_{T}^{2(r-1)}E\left(\left\|\left(Z_{t}'\otimes I_{G}\right)\lambda\right\|^{2r}\right)=o\left(1\right),$$

since $2\left(r-1\right)-\left(G+2\right)=2r-G-4=\xi>0$ and $E\left(\left\|\left(Z_{t}'\otimes I_{G}\right)\lambda\right\|^{2r}\right)<\infty$, by Assumption A10, and $\delta_{T}=o\left(1\right)$, by Assumption A7. It follows from M that $\sup_{A\in\mathcal{A}_{2T}}\left|H_{T,2}^{*}\left(\alpha_{A}\right)\right|=o_{p}\left(1\right)$, and since $\sup_{A\in\mathcal{A}_{2T}}\left|H_{T,1}^{*}\left(\alpha_{A}\right)\right|=o_{p}\left(1\right)$, as shown above, then:

$$\sup_{A\in\mathcal{A}_{2T}}\left|C_{2T}\left(\alpha_{A}\right)\right|=o_{p}\left(1\right),$$

Third, we have that:

$$H_{T}^{e}\left(\alpha_{A}\right)=T^{-1}\delta_{T}^{-(G+2)}\sum_{t=1}^{T}\lambda'\left(Z_{t}\otimes I_{G}\right)E\left[\mathcal{K}^{(2)}\left(\frac{Y_{t}-AZ_{t}}{\delta_{T}}\right)|\mathcal{F}_{t-1}\right]\left(Z_{t}'\otimes I_{G}\right)\lambda.$$

Now define:

$$\begin{aligned}\mathcal{K}_{i}^{(1)}\left(u\right)&\equiv\frac{\partial\mathcal{K}\left(u\right)}{\partial u_{i}},\qquad\mathcal{K}_{ij}^{(2)}\left(u\right)\equiv\frac{\partial^{2}\mathcal{K}\left(u\right)}{\partial u_{i}\partial u_{j}},\\ f_{i,t}^{(1)}\left(u|\mathcal{F}_{t-1}\right)&\equiv\frac{\partial f_{t}\left(u|\mathcal{F}_{t-1}\right)}{\partial u_{i}},\qquad f_{ij,t}^{(2)}\left(u|\mathcal{F}_{t-1}\right)\equiv\frac{\partial^{2}f_{t}\left(u|\mathcal{F}_{t-1}\right)}{\partial u_{i}\partial u_{j}}.\end{aligned}$$

and let u_{-i} denote the vector consisting of the elements of u other than u_{i}; then

$$\mathrm{E}\left[\mathcal{K}_{ij}^{(2)}\left(\frac{Y_{t}-AZ_{t}}{\delta_{T}}\right)|\mathcal{F}_{t-1}\right]=\int_{\mathbb{R}^{G-1}}\int_{\mathbb{R}}\mathcal{K}_{ij}^{(2)}\left(\frac{u-\left(A-A_{0}\right)Z_{t}}{\delta_{T}}\right)f_{t}\left(u|\mathcal{F}_{t-1}\right)du_{i}du_{-i}.$$

Using repeated integration by parts we have that:

$$\begin{aligned}&\int_{\mathbb{R}}\mathcal{K}_{ij}^{(2)}\left(\frac{u-\left(A-A_{0}\right)Z_{t}}{\delta_{T}}\right)f_{t}\left(u|\mathcal{F}_{t-1}\right)du_{i}\\ &\quad=\left[\delta_{T}\mathcal{K}_{j}^{(1)}\left(\frac{u-\left(A-A_{0}\right)Z_{t}}{\delta_{T}}\right)f_{t}\left(u|\mathcal{F}_{t-1}\right)\right]_{u_{i}=-\infty}^{u_{i}=\infty}\\ &\qquad-\int_{\mathbb{R}}\delta_{T}\mathcal{K}_{j}^{(1)}\left(\frac{u-\left(A-A_{0}\right)Z_{t}}{\delta_{T}}\right)f_{i,t}^{(1)}\left(u|\mathcal{F}_{t-1}\right)du_{i}.\end{aligned}$$

But:

$$\left|\mathcal{K}_{j}^{(1)}\left(\frac{u-\left(A-A_{0}\right)Z_{t}}{\delta_{T}}\right)f_{t}\left(u|\mathcal{F}_{t-1}\right)\right|\leq L_{0}\left|\mathcal{K}_{j}^{(1)}\left(\frac{u-\left(A-A_{0}\right)Z_{t}}{\delta_{T}}\right)\right|,$$

by Assumption A2, and for all fixed u_{-i}:

$$\lim_{u_{i}\to\pm\infty}\left|\mathcal{K}_{j}^{(1)}\left(\frac{u-\left(A-A_{0}\right)Z_{t}}{\delta_{T}}\right)\right|=0,$$

by Assumption A11. Hence:

$$\left[\delta_T \mathcal{K}_j^{(1)}\left(\frac{u-(A-A_0)Z_t}{\delta_T}\right) f_t(u|\mathcal{F}_{t-1})\right]_{u_i=-\infty}^{u_i=\infty}=0,$$

so:

$$\begin{aligned}
\mathrm{E}\left[\mathcal{K}_{ij}^{(2)}\left(\frac{Y_t-AZ_t}{\delta_T}\right)|\mathcal{F}_{t-1}\right] &= -\delta_T \int_{\mathbb{R}^{G-1}}\int_{\mathbb{R}} \mathcal{K}_j^{(1)}\left(\frac{u-(A-A_0)Z_t}{\delta_T}\right) f_{i,t}^{(1)}(u|\mathcal{F}_{t-1})\,du_i du_{-i} \\
&= -\delta_T \int_{\mathbb{R}^{G-1}}\int_{\mathbb{R}} \mathcal{K}_j^{(1)}\left(\frac{u-(A-A_0)Z_t}{\delta_T}\right) f_{i,t}^{(1)}(u|\mathcal{F}_{t-1})\,du_j du_{-j}.
\end{aligned}$$

Using integration by parts again we have that:

$$\begin{aligned}
&\int_{\mathbb{R}} \mathcal{K}_j^{(1)}\left(\frac{u-(A-A_0)Z_t}{\delta_T}\right) f_{i,t}^{(1)}(u|\mathcal{F}_{t-1})\,du_j \\
&\quad= \left[\delta_T \mathcal{K}\left(\frac{u-(A-A_0)Z_t}{\delta_T}\right) f_{i,t}^{(1)}(u|\mathcal{F}_{t-1})\right]_{u_j=-\infty}^{u_j=\infty} \\
&\qquad - \int_{\mathbb{R}} \delta_T \mathcal{K}\left(\frac{u-(A-A_0)Z_t}{\delta_T}\right) f_{ij,t}^{(2)}(u|\mathcal{F}_{t-1})\,du_j.
\end{aligned}$$

But:

$$\left|\mathcal{K}\left(\frac{u-(A-A_0)Z_t}{\delta_T}\right) f_{i,t}^{(1)}(u|\mathcal{F}_{t-1})\right| \le L_3 \left|\mathcal{K}\left(\frac{u-(A-A_0)Z_t}{\delta_T}\right)\right|,$$

by Assumption A8, and for all fixed u_{-j}:

$$\lim_{u_j\to\pm\infty}\left|\mathcal{K}\left(\frac{u-(A-A_0)Z_t}{\delta_T}\right)\right|=0,$$

by Assumption A11. Hence:

$$\left[\mathcal{K}\left(\frac{u-(A-A_0)Z_t}{\delta_T}\right) f_{i,t}^{(1)}(u|\mathcal{F}_{t-1})\right]_{u_i=-\infty}^{u_i=\infty}=0,$$

so:

$$\begin{aligned}
E\left[\mathcal{K}_{ij}^{(2)}\left(\frac{Y_t-AZ_t}{\delta_T}\right)|\mathcal{F}_{t-1}\right] &= \delta_T^2 \int_{\mathbb{R}^{G-1}}\int_{\mathbb{R}} \mathcal{K}\left(\frac{u-(A-A_0)Z_t}{\delta_T}\right) f_{ij,t}^{(2)}(u|\mathcal{F}_{t-1})\,du_j du_{-j} \\
&= \delta_T^2 \int_{\mathbb{R}^G} \mathcal{K}\left(\frac{u-(A-A_0)Z_t}{\delta_T}\right) f_{ij,t}^{(2)}(u|\mathcal{F}_{t-1})\,du \\
&= \delta_T^{G+2} \int_{\mathbb{R}^G} \mathcal{K}(s)\, f_{ij,t}^{(2)}((A-A_0)Z_t+\delta_T s|\mathcal{F}_{t-1})\,ds
\end{aligned}$$

and hence that:

$$E\left[\mathcal{K}^{(2)}\left(\frac{Y_t-AZ_t}{\delta_T}\right)|\mathcal{F}_{t-1}\right] = \delta_T^{G+2}\int_{\mathbb{R}^G} \mathcal{K}(s)\, f_t^{(2)}((A-A_0)Z_t+\delta_T s|\mathcal{F}_{t-1})\,ds.$$

If we now define:

$$H^e_{t0}(\alpha_A,\delta) \equiv \lambda' (Z_t \otimes I_G) \int_{\mathbb{R}^G} \mathcal{K}(s) f^{(2)}_t ((A-A_0) Z_t + \delta_T s|\mathcal{F}_{t-1}) \, ds \, (Z'_t \otimes I_G) \lambda,$$

then:

$$h^e_{tT}(\alpha_A) = H^e_{t0}(\alpha_A,\delta_T), \qquad H_0(\alpha_A) = E\left[H^e_{t0}(\alpha_A,0)\right],$$

while for any $A \in \mathcal{A}$ and $\delta \in \mathbb{R}$, the sequence $\{H^e_{t0}(\alpha_A,\delta)\}$ is strictly stationary and ergodic, by Assumption A1. Then by CS:

$$\left|H^e_{t0}(\alpha_A,\delta)\right| \leq T^{-1} L_4 G \|\lambda\|^2 \|Z_t\|^2 \left(\int_{\mathbb{R}^G} |\mathcal{K}(s)| \, ds\right) = \zeta_t,$$

since $\left\|f^{(2)}_t(u|\mathcal{F}_{t-1})\right\| \leq L_4$ for all (t,u,ω), by Assumption A8. In addition, for any (t,ω) then $H^e_{t0}(\alpha_A,\delta)$ is continuous with respect to (α_A,δ) since $f^{(2)}_t(u|\mathcal{F}_{t-1})$ is uniformly bounded and continuous, by Assumption A8, and since $\int_{\mathbb{R}^G} |\mathcal{K}(s)| \, ds < \infty$, by Assumption A11. But $E\left(\|Z_t\|^2\right) < \infty$, by Assumption A10, so $E(\zeta_t) < \infty$ and hence it follows by the uniform law of large numbers for strictly stationary ergodic process that:

$$\begin{aligned} \sup_{A\in\mathcal{A}} |C_{4T}(\alpha_A)| &= \sup_{A\in\mathcal{A}} \left|H^e_T(\alpha_A) - H_0(\alpha_A)\right| \\ = \sup_{A\in\mathcal{A}} &\left|T^{-1}\sum_{t=1}^{T} H^e_{t0}(\alpha_A,\delta_T) - E\left[H^e_{t0}(\alpha_A,0)\right]\right| = o_p(1), \end{aligned}$$

since $\delta_T = o(1)$ by Assumption A6 and $\mathcal{A}$ is compact by Assumption A3. □

Lemma 7 *Under Assumptions A1–A12:*

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

Proof. Define:

$$\begin{aligned} g_{tT} &\equiv \left(T\delta^G_T\right)^{-1/2} (Z_t \otimes I_G) \mathcal{K}^{(1)}\left(\frac{U_t}{\delta_T}\right), \\ g^e_{tT} &\equiv E\left[g_{tT}|\mathcal{F}_{t-1}\right], \\ g^*_{tT} &\equiv g_{tT} - g^e_{tT}, \end{aligned}$$

so that:

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

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, as in the proof of Lemma 6, define $\mathcal{K}_i^{(1)}(u) \equiv \partial \mathcal{K}(u)/\partial u_i$ for $i = 1, ..., G$, and then define:

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

so that $g_{tT}^{e} = (g_{1tT}^{e}, ..., g_{GtT}^{e})'$, 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:

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

where $f_{i,t}^{(1)}(u|\mathcal{F}_{t-1}) \equiv \partial f_t(u|\mathcal{F}_{t-1})/\partial u_i$, as in the proof of Lemma 6. 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 A2(i), and for all fixed u_{-i}:

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

by Assumption A11. 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:

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

so:

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

Thus we have that:

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

where:

$$\psi_{itT} \equiv \int_{\mathbb{R}^G} \mathcal{K}(s) f_{i,t}^{(1)} (\delta_T s | \mathcal{F}_{t-1}) \, ds.$$

Now a second order Taylor series expansion of $f_{i,t}^{(1)} (\delta s | \mathcal{F}_{t-1})$ around $s = 0$ gives:

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

for some $0 \leq \lambda \leq 1$, where $f_{ij,t}^{(2)} (u | \mathcal{F}_{t-1})$ denotes the (i,j)-th element of $f_t^{(2)} (u | \mathcal{F}_{t-1})$ and $f_{ijk,t}^{(3)} (u | \mathcal{F}_{t-1})$ denotes the (i,j,k)-th element of $f_t^{(2)} (u | \mathcal{F}_{t-1})$. But $f_{i,t}^{(1)} (0 | \mathcal{F}_{t-1}) = 0$, by Assumption A2, and $\int_{\mathbb{R}^G} s \mathcal{K}(s) \, ds = 0$, by Assumption A6. Hence:

$$\psi_{itT} = \left(\frac{1}{2} \right) \delta_T^2 \sum_{j=1}^{G} \sum_{k=1}^{G} \int_{\mathbb{R}^G} s_j s_k \mathcal{K}(s) \, f_{ijk,t}^{(3)} (\lambda \delta_T s | \mathcal{F}_{t-1}) \, ds,$$

where λ may vary with s, and by CS it follows that:

$$|\psi_{itT}| \leq \left(\frac{1}{2} \right) \delta_T^2 \sum_{j=1}^{G} \sum_{k=1}^{G} \int_{\mathbb{R}^G} |s_j s_k| \; |\mathcal{K}(s)| \; \left| f_{ijk,t}^{(3)} (\lambda \delta_T s | \mathcal{F}_{t-1}) \right| ds.$$

Since $\left\| f_t^{(3)} (u | \mathcal{F}_{t-1}) \right\| \leq L_5$ for all (u, t, ω), by Assumption A8, and since $\int_{\mathbb{R}^G} \|s\|^2 |\mathcal{K}| \, ds \leq L_{10} < \infty$, by Assumption A11, it follows that:

$$|\psi_{itT}| \leq \frac{G^2 L_3 L_{10}}{2} \delta_T^2,$$

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

$$\sum_{t=1}^{T} g_{itT}^{e} = \frac{G^2 L_3 L_{10}}{2} T^{1/2} \delta_T^{(G/2)+1} \delta_T^2 O_p(1) = O_p \left[\left(T \delta_T^{G+6} \right)^{1/2} \right] = o_p(1),$$

by Assumption A12. This then implies that $\sum_{t=1}^{T} g_{tT}^{e} = o_p(1)$.

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

$$z_{tT} \equiv \lambda' g^*_{tT}, \qquad \sigma^2_{tT} \equiv Var\left(z_{tT}\right), \qquad \Sigma_T \equiv \sum_{t=1}^{T} \sigma^2_{tT}, \qquad \eta_{tT} \equiv \frac{z_{tT}}{\sqrt{\Sigma_T}}.$$

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

$$\begin{aligned} \sigma^2_{tT} &= E\left(z^2_{tT}\right) = E\left[\left(\lambda' g_{tT}\right)^2 - 2\left(\lambda' g_{tT}\right)\left(\lambda' g^e_{tT}\right) + \left(\lambda' g^e_{tT}\right)^2\right] \\ &= E\left[\left(\lambda' g_{tT}\right)^2\right] - E\left[\left(\lambda' g^e_{tT}\right)^2\right], \end{aligned}$$

so:

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

Then:

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

so by the law of iterated expectations it follows that:

$$E\left(g_{tT} g'_{tT}\right) = T^{-1} E\left[\left(Z_t \otimes I_G\right) \Gamma^e_{tT} \left(Z'_t \otimes I_G\right)\right],$$

where:

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

Then by Assumption A1 we have that:

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

by dominated convergence, since $E\left(\|Z_t\|^2\right)<\infty$, by Assumption A10, and since:

$$0\leq\int_{\mathbb{R}^G}\mathcal{K}^{(1)}(s)\,\mathcal{K}^{(1)}(s)'\,f_t(\delta_T s|\mathcal{F}_{t-1})\,du\leq L_0\int_{\mathbb{R}^G}\mathcal{K}^{(1)}(s)\,\mathcal{K}^{(1)}(s)'\,du,$$

in the positive semi-definite sense, by Assumptions A2 and A11. Thus:

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

In addition, from above we have that:

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

and hence by CS it follows that:

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

Now:

$$\begin{aligned}\left\|\int_{\mathbb{R}^G}\mathcal{K}(s)\,f_t^{(1)}(\delta_T s|\mathcal{F}_{t-1})\,ds\right\|^2 &= \sum_{i=i}^{G}\left(\int_{\mathbb{R}^G}\mathcal{K}(s)\,f_{i,t}^{(1)}(\delta_T s|\mathcal{F}_{t-1})\,ds\right)^2\\ &= \sum_{i=1}^{G}\psi_{itT}^2\leq\frac{G^3L_3L_{10}}{2}\delta_T^2,\end{aligned}$$

using the bounds on ψ_{itT} from above. In addition, $\left\|\lambda'\left(Z_t\otimes I_G\right)\right\|^2\leq G\left\|\lambda\right\|^2\left\|Z_t\right\|^2$, so we have that:

$$\begin{aligned}\sum_{t=1}^{T}\left(\lambda' g_{tT}^{e}\right)^2 &\leq \left(T\delta_T^G\right)^{-1}\delta_T^{2G+2}\left(\sum_{t=1}^{T}\left\|\lambda'\left(Z_t\otimes I_G\right)\right\|^2\right)\left(\sum_{i=1}^{G}\psi_{itT}^2\right)\\ &\leq T\left(T\delta_T^G\right)^{-1}\delta_T^{2G+2}G\left(\frac{G^3L_3L_{10}}{2}\delta_T^2\right)^2G\left\|\lambda\right\|^2\left(T^{-1}\sum_{t=1}^{T}\left\|Z_t\right\|^2\right),\end{aligned}$$

and hence that:

$$\begin{aligned}E\left[\sum_{t=1}^{T}\left(\lambda' g_{tT}^{e}\right)^2\right] &\leq T\left(T\delta_T^G\right)^{-1}\delta_T^{2G+6}\left(\frac{G^8L_3^2L_{10}}{4}\right)\left\|\lambda\right\|^2E\left(T^{-1}\sum_{t=1}^{T}\left\|Z_t\right\|^2\right)\\ &= O\left(\delta_T^{G+6}\right)=o(1)\end{aligned}$$

since $E\left(T^{-1}\sum_{t=1}^{T}\left\|Z_t\right\|^2\right)=E\left(\left\|Z_t\right\|^2\right)<\infty$, by Assumptions 1 and 10, and since $\delta_T^{G+6}=o(1)$, by assumption A12. Hence we have that:

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

and since B_0 is non-singular, by Lemma 5 above, then Σ_T is positive for all T sufficiently large and hence $\sum_{t=1}^{T} Var(\eta_{tT}) = 1$ for all T sufficiently large.

Second, observe that:

$$\sum\nolimits_{t=1}^{T} z_{tT}^2 - \lambda' \mathrm{B}_0 \lambda = W_{1,T} + W_{2,T},$$

where:

$$\begin{aligned} W_{1,T} &\equiv \sum\nolimits_{t=1}^{T} z_{tT}^2 - \sum\nolimits_{t=1}^{T} E\left(z_{tT}^2 | \mathcal{F}_{t-1}\right), \\ W_{2,T} &\equiv \sum\nolimits_{t=1}^{T} E\left(z_{tT}^2 | \mathcal{F}_{t-1}\right) - \lambda' \mathrm{B}_0 \lambda. \end{aligned}$$

Now define:

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

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), then for any $1 < p \leq 2$:

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

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

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

so:

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

But by the law of iterated expectations:

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

while by CS and Assumption A2:

$$\begin{aligned}
&E\left[\left(\left|\lambda'\left(Z_t\otimes I_G\right)\mathcal{K}^{(1)}\left(\frac{Y_t-A_0Z_t}{\delta_T}\right)\right|^{2p}\right)|\mathcal{F}_{t-1}\right]\\
&\quad\leq \left\|\lambda'\left(Z_t\otimes I_G\right)\right\|^{2p}\int_{\mathbb{R}^G}\left\|\mathcal{K}^{(1)}\left(\frac{u}{\delta_T}\right)\right\|^{2p} f_t\left(u|\mathcal{F}_{t-1}\right)du\\
&\quad\leq G^p\left\|\lambda\right\|^{2p}\left\|Z_t\right\|^{2p}\delta_T^G\left(\int_{\mathbb{R}^G}\left\|\mathcal{K}^{(1)}\left(s\right)\right\|^{2p} f_t\left(\delta_T s|\mathcal{F}_{t-1}\right)ds\right)\\
&\quad\leq G^pL_0\left\|\lambda\right\|^{2p}\left\|Z_t\right\|^{2p}\delta_T^G\left(\int_{\mathbb{R}^G}\left\|\mathcal{K}^{(1)}\left(s\right)\right\|^{2p}ds\right).
\end{aligned}$$

Thus:

$$E\left(\left|W_{1,T}\right|^p\right)\leq 2^{p+1}G^pL_0\left\|\lambda\right\|^{2p}\left(T\delta_T^G\right)^{-p+1}\left(\int_{\mathbb{R}^G}\left\|\mathcal{K}^{(1)}\left(s\right)\right\|^{2p}ds\right)E\left(T^{-1}\sum_{t=1}^{T}\left\|Z_t\right\|^{2p}\right).$$

Now since $\left\|\mathcal{K}^{(1)}\left(s\right)\right\|$ is uniformly bounded, by Assumption A6, and $\int_{\mathbb{R}^G}\left\|\mathcal{K}^{(1)}\left(s\right)\right\|ds<\infty$, by Assumption A11, then $\left(\int_{\mathbb{R}^G}\left\|\mathcal{K}^{(1)}\left(s\right)\right\|^{2p}ds\right)<\infty$. In addition, since $1<p\leq 2$ then $E\left(\left\|Z_t\right\|^{2p}\right)<\infty$, by Assumption A10, and hence $E\left(T^{-1}\sum_{t=1}^{T}\left\|Z_t\right\|^{2p}\right)=O_p\left(1\right)$ by Assumption A1 and the ergodic theorem. In addition, $T\delta_T^G\to\infty$, by Assumption A12, and hence $\left(T\delta_T^G\right)^{-p+1}=o\left(1\right)$. But then $E\left(\left|W_{1,T}\right|^p\right)=o\left(1\right)$ so $W_{1,T}=o_p\left(1\right)$ since L_p convergence implies convergence in probability.

Next, observe that:

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

and hence:

$$W_{2,T}=\sum_{t=1}^{T}E\left[\left(\lambda' g_{tT}\right)^2|\mathcal{F}_{t-1}\right]-\sum_{t=1}^{T}\left(\lambda' g_{tT}^e\right)^2-\lambda'\mathrm{B}_0\lambda.$$

Now, observe that:

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

and define $\widehat{B}_{T,0}\equiv\widehat{B}_T\left(\alpha_0\right)$, where $\widehat{B}_T\left(\cdot\right)$ is the same as in the statement of Theorem 4. Then CS and T imply that:

$$
\begin{aligned}
&E\left(\left|\lambda'\left[\sum\nolimits_{t=1}^{T} E\left(g_{tT}g_{tT}'|\mathcal{F}_{t-1}\right)\right]\lambda - \lambda'\widehat{B}_{T,0}\lambda\right|\right) \\
&\quad \leq G\left\|\lambda\right\|^{2} T^{-1}\sum_{t=1}^{T} E\left(\left\|Z_t\right\|^2 \int_{\mathbb{R}^G}\left\|\mathcal{K}^{(1)}\left(s\right)\right\|^2 \left|f_t\left(\delta_T s|\mathcal{F}_{t-1}\right) - f_t\left(0|\mathcal{F}_{t-1}\right)\right| ds\right) \\
&\quad = G\left\|\lambda\right\|^{2}\int_{\Omega}\int_{\mathbb{R}^G}\left\|Z_t\right\|^2 \left\|\mathcal{K}^{(1)}\left(s\right)\right\|^2 \left|f_t\left(\delta_T s|\mathcal{F}_{t-1}\right) - f_t\left(0|\mathcal{F}_{t-1}\right)\right| ds\, dP\left(\omega\right) \\
&\quad = o\left(1\right),
\end{aligned}
$$

by dominated convergence, since $f_t\left(u|\mathcal{F}_{t-1}\right)$ is continuous in u and uniformly bounded for all (t,ω), by Assumption A2, and since:

$$
\int_{\Omega}\int_{\mathbb{R}^G}\left\|Z_t\right\|^2 \left\|\mathcal{K}^{(1)}\left(s\right)\right\|^2 ds\, dP\left(\omega\right) \leq L_8 E\left(\left\|Z_t\right\|^2\right) < \infty,
$$

by Assumptions A10 and A11. It follows by M that:

$$
\lambda'\left[\sum\nolimits_{t=1}^{T} E\left(g_{tT}g_{tT}'|\mathcal{F}_{t-1}\right)\right]\lambda - \lambda'\widehat{B}_{T,0}\lambda = o_p\left(1\right),
$$

and since $\widehat{B}_{T,0}$ converges in probability to B_0, by Assumptions A1 and 10 and the ergodic theorem, then:

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

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

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

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

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

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

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

Now:

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

so by CS and the law of iterated expectations:

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

Now:

$$\begin{aligned} E\left\{\left\|\mathcal{K}^{(1)}\left(\frac{U_t}{\delta_T}\right)\right\|^{2p} |\mathcal{F}_{t-1}\right\} &= \int_{\mathbb{R}^G} \left\|\mathcal{K}^{(1)}\left(\frac{u}{\delta_T}\right)\right\|^{2p} f_t\left(u|\mathcal{F}_{t-1}\right) du \\ &= \int_{\mathbb{R}^G} \delta_T^G \left\|\mathcal{K}^{(1)}\left(s\right)\right\|^{2p} f_t\left(\delta_T s|\mathcal{F}_{t-1}\right) ds \\ &\leq L_0 L_1^{2(p-1)} L_8 \delta_T^G, \end{aligned}$$

since $0 \leq f_t\left(u|\mathcal{F}_{t-1}\right) \leq L_0$ for all u, by Assumption A2, $\left\|\mathcal{K}^{(1)}\left(s\right)\right\| \leq L_1$ for all s, by Assumption A6, and $\int_{\mathbb{R}^G} \left\|\mathcal{K}^{(1)}\left(s\right)\right\|^2 ds = L_8$, by Assumption A11. Therefore we have that:

$$\begin{aligned} \Pr\left\{\max_{1\leq t\leq T} \left|z_{tT}\right| > \varepsilon\right\} &= \varepsilon^{-2p} 2^{2p} T \left(T\delta_T^G\right)^{-p} E\left(\left\|\lambda' \left(Z_t \otimes I_G\right)\right\|^{2p}\right) L_0 L_1^{2(p-1)} L_8 \delta_T^G \\ &\leq O\left[\left(T\delta_T^G\right)^{-(p-1)}\right] = o\left(1\right), \end{aligned}$$

since $E\left(\left\|\lambda' \left(Z_t \otimes I_G\right)\right\|^{2p}\right) < \infty$ for all $1 < p \leq 2$, by Assumption A10, and $T\delta_T^G \to \infty$, by Assumption A12.

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

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

□

Lemma 8 *Under Assumptions A1–A12, define:*

$$B\left(\alpha_A\right) \equiv E\left[f_t\left(\left(A - A_0\right) Z_t|\mathcal{F}_{t-1}\right) \left(Z_t \otimes I_G\right) \mathcal{M} \left(Z_t \otimes I_G\right)'\right],$$

then:

$$\sup_{A\in\mathcal{A}} \left\|\widehat{B}_T\left(\alpha_A\right) - B\left(\alpha_A\right)\right\| = o_p\left(1\right).$$

Proof. In the proof of this Lemma we use the same constant $J_2 < \infty$, sequence $\{\mathcal{A}_{2T}\}$ of subsets of $\mathcal{A}$, and sequences $\left\{\bar{A}_{2T}(\cdot)\right\}$ and $\{\bar{\alpha}_{2T}(\cdot)\}$ of functions as used in the proof of Lemma 6.

Now, fix $\lambda \neq 0$ and for any $A \in \mathcal{A}$ define:

$$\begin{aligned}
R(\alpha_A) &\equiv \lambda' B(\alpha_A)\lambda, \\
r_{tT}(\alpha_A) &\equiv \left(T\delta_T^G\right)^{-1}\left[\lambda'(Z_t \otimes I_G)\mathcal{K}^{(1)}\left(\frac{Y_t - AZ_t}{\delta_T}\right)\right]^2, \\
R_T(\alpha_A) &\equiv \lambda'\widehat{B}_T(\alpha_A)\lambda = \sum_{t=1}^{T} r_{tT}(\alpha_A), \\
r_{tT}^e(\alpha_A) &\equiv E\left[r_{tT}(\alpha_A)|\mathcal{F}_{t-1}\right], \qquad R_T^e(\alpha_A) \equiv \sum_{t=1}^{T} r_{tT}^e(\alpha_A).
\end{aligned}$$

In addition, for any $A \in \mathcal{A}$ define:

$$\begin{aligned}
S_{1T}(\alpha_A) &\equiv R_T(\alpha_A) - R_T(\bar{\alpha}_{2T}(\alpha_A)), \\
S_{2T}(\alpha_A) &\equiv R_T(\alpha_A) - R_T^e(\alpha_A), \\
S_{3T}(\alpha_A) &\equiv R_T^e(\bar{\alpha}_{2T}(\alpha_A)) - R_T^e(\alpha_A), \\
S_{4T}(\alpha_A) &\equiv R_T^e(\alpha_A) - R(\alpha_A),
\end{aligned}$$

and observe that:

$$R_T(\alpha_A) - R(\alpha_A) = S_{1T}(\alpha_A) + S_{2T}(\bar{\alpha}_{2T}(\alpha_A)) + S_{3T}(\alpha_A) + S_{4T}(\alpha_A),$$

and hence:

$$\begin{aligned}
\sup_{A\in\mathcal{A}}|R_T(\alpha_A) - R(\alpha_A)| &= \sup_{A\in\mathcal{A}}|S_{1T}(\alpha_A)| + \sup_{A\in\mathcal{A}_{2T}}|S_{2T}(\alpha_A)| \\
&\quad + \sup_{A\in\mathcal{A}}|S_{3T}(\alpha_A)| + \sup_{A\in\mathcal{A}}|S_{4T}(\alpha_A)|.
\end{aligned}$$

In order to establish the desired result it suffices to establish that each of the terms on the right-hand-side of the above equation is $o_p(1)$.

First, observe that by T:

$$|S_{1T}(\alpha_A)| \leq \sum_{t=1}^{T}|r_{tT}(\alpha_A) - r_{tT}(\bar{\alpha}_{2T}(\alpha_A))|.$$

Now for any $A, A^{\dagger} \in \mathcal{A}$ then it follows by MVT that:

$$\begin{aligned} r_{tT}(\alpha_A) - r_{tT}(\alpha_{A^{\dagger}}) &= 2T^{-1}\delta_T^{-G}\delta_T^{-1}\left[\lambda'(Z_t \otimes I_G)\mathcal{K}^{(1)}\left(\frac{Y_t - A^* Z_t}{\delta_T}\right)\right] \\ &\quad \times \left[\lambda'(Z_t \otimes I_G)\mathcal{K}^{(2)}\left(\frac{Y_t - A^* Z_t}{\delta_T}\right)(Z_t' \otimes I_G)(\alpha_A - \alpha_{A^{\dagger}})\right], \end{aligned}$$

and hence it follows by CS that:

$$|r_{tT}(\alpha_A) - r_{tT}(\alpha_{A^{\dagger}})| \leq 2G^{3/2}L_1 L_6 T^{-1}\delta_T^{-(G+1)}\|\lambda\|^2\|Z_t\|^3\|\alpha_A - \alpha_{A^{\dagger}}\|,$$

since $\left\|\mathcal{K}^{(1)}(u)\right\| \leq L_1$ and $\left\|\mathcal{K}^{(2)}(u)\right\| \leq L_6$ for all u. But then since $\|\alpha_A - \bar{\alpha}_{2T}(\alpha_A)\| \leq T^{-2}$ for all $A \in \mathcal{A}$ it follows that:

$$|r_{tT}(\alpha_A) - r_{tT}(\bar{\alpha}_{2T}(\alpha_A))| \leq 2G^{3/2}L_1 L_6 T^{-3}\delta_T^{-(G+1)}\|\lambda\|^2\|Z_t\|^3,$$

and hence that:

$$\sup_{A\in\mathcal{A}}|S_{1T}(\alpha_A)| \leq 2G^{3/2}L_1 L_6 T^{-2}\delta_T^{-(G+1)}\|\lambda\|^2\left(T^{-1}\sum_{t=1}^{T}\|Z_t\|^3\right).$$

Now $T^{-1}\sum_{t=1}^{T}\|Z_t\|^3 = O_p(1)$, by Assumptions A1 and A10 and the ergodic theorem, while $T^{-2}\delta_T^{-(G+1)} = \left(T\delta_T^{(G+1)/2}\right)^{-2} = o(1)$, by Assumption A12. Hence it follows that:

$$\sup_{A\in\mathcal{A}}|S_{1T}(\alpha_A)| = o_p(1).$$

In addition, it follows from T that:

$$|S_{3T}(\alpha_A)| \leq \sum_{t=1}^{T}|r_{tT}^e(\alpha_A) - r_{tT}^e(\bar{\alpha}_{2T}(\alpha_A))|.$$

But for any $A, A^{\dagger} \in \mathcal{A}$ then by J:

$$\begin{aligned} |r_{tT}^e(\alpha_A) - r_{tT}^e(\alpha_{A^{\dagger}})| &= |E\{[r_{tT}(\alpha_A) - r_{tT}(\alpha_{A^{\dagger}})]|\mathcal{F}_{t-1}\}| \\ &\leq E[\{|r_{tT}(\alpha_A) - r_{tT}(\alpha_{A^{\dagger}})|\}|\mathcal{F}_{t-1}] \\ &\leq 2G^{3/2}L_1 L_6 T^{-3}\delta_T^{-(G+1)}\|\lambda\|^2\|Z_t\|^3, \end{aligned}$$

and hence:

$$\sup_{A\in\mathcal{A}}|S_{3T}(\alpha_A)| \leq 2G^{3/2}L_1 L_6 T^{-2}\delta_T^{-(G+1)}\|\lambda\|^2\left(T^{-1}\sum_{t=1}^{T}\|Z_t\|^3\right) = o_p(1).$$

Second, define:

$$\begin{aligned}
r^*_{tT}(\alpha_A) &\equiv r_{tT}(\alpha_A) - r^e_{tT}(\alpha_A), \\
r_{tT,1}(\alpha_A) &\equiv \chi\left(\left\|(Z'_t \otimes I_G)\lambda\right\|^2 \leq \delta_T^{-2}\right) r_{tT}, \\
r_{tT,2}(\alpha_A) &\equiv \chi\left(\left\|(Z'_t \otimes I_G)\lambda\right\|^2 > \delta_T^{-2}\right) r_{tT}, \\
r^e_{tT,j}(\alpha_A) &\equiv E\left[r_{tT,j}(\alpha_A)|\mathcal{F}_{t-1}\right], \qquad j = 1,2, \\
r^*_{tT,j}(\alpha_A) &\equiv r_{tT,j}(\alpha_A) - r^e_{tT,j}(\alpha_A), \qquad j = 1,2, \\
R^*_{tT,j}(\alpha) &\equiv \sum_{t=1}^{T} r^*_{tT,j}(\alpha_A), \qquad j = 1,2,
\end{aligned}$$

which implies that $r^*_{tT}(\alpha_A) = r^*_{tT,1}(\alpha_A) + r^*_{tT,2}(\alpha_A)$ so that:

$$S_{2T}(\alpha_A) \equiv \left[R_T(\alpha_A) - R^e_T(\alpha_A)\right] = R^*_{T,1}(\alpha_A) + R^*_{T,2}(\alpha_A),$$

and hence that:

$$\sup_{A \in \mathcal{A}_{2T}} |S_{2T}(\alpha_A)| \leq \sup_{A \in \mathcal{A}_{2T}} \left|R^*_{T,1}(\alpha_A)\right| + \sup_{A \in \mathcal{A}_{2T}} \left|R^*_{T,2}(\alpha_A)\right|.$$

Now for any $A \in \mathcal{A}$ and $T = 1, 2, \ldots$, then by construction $\left\{\left(r^*_{tT,1}(\alpha_A), \mathcal{F}_t\right)\right\}_{t=-\infty}^{\infty}$ is a martingale difference sequence. In addition:

$$|r_{1T,1}(\alpha_A)| \leq L_1 T^{-1} \delta_T^{-G} \delta_T^{-2},$$

since $\left\|K^{(1)}(u)\right\| \leq L_1$ for all u, by Assumption A6, so it follows from T and J that $\left|r^*_{tT,1}(\alpha_A)\right| \leq 2L_1 T^{-1} \delta_T^{-G} \delta_T^{-2}$. Now define:

$$r^v_{tT,1}(\alpha_A) \equiv Var\left[r^*_{tT,1}(\alpha_A)|\mathcal{F}_{t-1}\right], \qquad R^v_{T,1}(\alpha_A) \equiv \sum_{t=1}^{T} r^v_{tT,1}(\alpha_A),$$

and observe that since:

$$E\left[r^*_{tT,1}(\alpha_A)|\mathcal{F}_{t-1}\right] = E\left[r^*_{tT,2}(\alpha_A)|\mathcal{F}_{t-1}\right] = 0,$$

and $E\left[r^*_{tT,1}(\alpha_A) r^*_{tT,2}(\alpha_A)|\mathcal{F}_{t-1}\right] = 0$ then $Cov\left[r^*_{tT,1}(\alpha_A), r^*_{tT,2}(\alpha_A)|\mathcal{F}_{t-1}\right] = 0$ so by CS:

$$\begin{aligned}
r^v_{tT,1}(\alpha_A) &\leq Var\left[r^*_{tT}(\alpha_A)|\mathcal{F}_{t-1}\right] \leq E\left[r^2_{tT}(\alpha_A)|\mathcal{F}_{t-1}\right] \\
&\leq G^2 \left(T\delta_T^G\right)^{-2} \|\lambda\|^4 \|Z_t\|^4 \int_{\mathbb{R}^G} \left\|\mathcal{K}^{(1)}\left(\frac{u - (A - A_0) Z_t}{\delta_T}\right)\right\|^4 f_t(u|\mathcal{F}_{t-1})\, du \\
&= G^2 T^{-2} \delta_T^{-G} \|\lambda\|^4 \|Z_t\|^4 \int_{\mathbb{R}^G} \left\|\mathcal{K}^{(1)}(s)\right\|^4 f_t((A - A_0) Z_t + \delta_T s|\mathcal{F}_{t-1})\, du \\
&\leq G^2 L_0 L_1^2 L_8 T^{-2} \delta_T^{-G} \|\lambda\|^4 \|Z_t\|^4,
\end{aligned}$$

since $f_t(u|\mathcal{F}_{t-1}) \leq L_0$ for all (u,t,ω), by Assumption A2, $\left\|K^{(1)}(u)\right\| \leq L_1$ for all u, by Assumption A6, and $\int_{\mathbb{R}^G} \left\|K^{(1)}(u)\right\|^2 du \leq L_8$, by Assumption A11. Hence it follows that:

$$R_{T,1}^{v}(\alpha_A) \leq G^2 L_0 L_1^2 L_8 T^{-1} \delta_T^{-G} \|\lambda\|^4 \left(T^{-1} \sum_{t=1}^{T} \|Z_t\|^4\right),$$

and since the right-hand-side of the above equation does not depend on α_A then:

$$\sup_{A\in\mathcal{A}_{2T}} R_{T,1}^{v}(\alpha_A) \leq G^2 L_0 L_1^2 L_8 T^{-1} \delta_T^{-G} \|\lambda\|^4 \left(T^{-1} \sum_{t=1}^{T} \|Z_t\|^4\right).$$

Now, since the $\{Z_t\}$ are strictly stationary and ergodic, by Assumption A1, and since $E\left(\|Z_t\|^4\right) = L_{11}$, by Assumption A10, then setting $b_T = 2G^2 L_0 L_1^2 L_9 L_{11} \|\lambda\|^4 T^{-1}\delta_T^{-G}$ it follows that:

$$\Pr\left(\sup_{A\in\mathcal{A}_{2T}} R_{T,1}^{v}(\alpha_A) > b_T\right) = o(1),$$

and hence for any fixed $\varepsilon > 0$ it follows by Lemma 2 that:

$$\Pr\left(\sup_{A\in\mathcal{A}_{2T}} \left|R_{T,1}^{*}(\alpha_A)\right| \geq \varepsilon\right) \leq \sum_{A\in\mathcal{A}_{2T}} \Pr\left(\left|R_{T,1}^{*}(\alpha_A)\right| \geq \varepsilon \,\&\, R_{T,1}^{v}(\alpha_A) \leq b_T\right) = o(1).$$

Then from Lemma 1 it follows that:

$$\begin{aligned}
&\Pr\left(\left|R_{T,1}^{*}(\alpha_A)\right| \geq \varepsilon \,\&\, R_{T,1}^{v}(\alpha_A) \leq b_T\right) \\
&\quad \leq \exp\left(-\frac{\varepsilon^2/2}{\varepsilon 2 L_1 T^{-1}\delta_T^{-G}\delta_T^{-2} + 2G^2 L_0 L_1^2 L_8 L_{11} \|\lambda\|^4 T^{-1}\delta_T^{-G}}\right) \\
&\quad = \exp\left(-\frac{\varepsilon^2 T \delta_T^{G+2}}{4\left(L_1\varepsilon + G^2 L_0 L_1^2 L_8 L_{11} \|\lambda\|^4 \delta_T^2\right)}\right),
\end{aligned}$$

and hence that:

$$\begin{aligned}
\Pr\left(\sup_{A\in\mathcal{A}_{2T}} \left|R_{T,1}^{*}(\alpha_A)\right| \geq \varepsilon\right) &\leq 2J_2 T^{2GK} \exp\left(-\frac{\varepsilon^2 T \delta_T^{G+2}}{4\left(L_1\varepsilon + G^2 L_0 L_1^2 L_8 L_{11} \|\lambda\|^4 \delta_T^2\right)}\right) + o(1) \\
&= o(1),
\end{aligned}$$

since $\ln(T)/\left(T\delta_T^{G+2}\right) = \delta_T^2 \ln(T)/\left(T\delta_T^{G+4}\right) = o(1)$, by Assumptions A7 and A11. Since $\varepsilon > 0$ was arbitrary it follows that:

$$\sup_{A\in\mathcal{A}_{2T}} \left|R_{T,1}^{*}(\alpha_A)\right| = o_p(1).$$

In addition, observe that since:

$$r_{tT,2}(\alpha_A) \equiv \chi\left(\|(Z_t' \otimes I_G)\lambda\|^2 > \delta_T^{-2}\right) r_{tT},$$

then by CS:

$$|r_{tT,2}(\alpha_A)| \leq L_1^2 \left(T\delta_T^G\right)^{-1} \|(Z_t' \otimes I_G)\lambda\|^2 \chi\left(\|(Z_t' \otimes I_G)\lambda\|^2 > \delta_T^{-2}\right),$$

and hence by J:

$$\left|r_{tT,2}^*(\alpha_A)\right| \leq 2L_1^2 \left(T\delta_T^G\right)^{-1} \|(Z_t' \otimes I_G)\lambda\|^2 \chi\left(\|(Z_t' \otimes I_G)\lambda\|^2 > \delta_T^{-2}\right).$$

In addition, since the right-hand-side of the above inequality does not depend on A then:

$$\sup_{A \in \mathcal{A}_{2T}} \left|r_{tT,2}^*(\alpha_A)\right| \leq 2L_1^2 \left(T\delta_T^G\right)^{-1} \|(Z_t' \otimes I_G)\lambda\|^2 \chi\left(\|(Z_t' \otimes I_G)\lambda\|^2 > \delta_T^{-2}\right),$$

and hence:

$$\begin{aligned} E\left(\sup_{A \in \mathcal{A}_{2T}} \left|R_{T,2}^*(\alpha_A)\right|\right) &\leq \sum_{t=1}^T E\left\{\sup_{A \in \mathcal{A}_{2T}} \left|r_{tT,2}^*(\alpha_A)\right|\right\} \\ &\leq 2L_1^2 \delta_T^{-G} E\left\{\|(Z_t' \otimes I_G)\lambda\|^2 \chi\left(\|(Z_t' \otimes I_G)\lambda\|^2 > \delta_T^{-2}\right)\right\}. \end{aligned}$$

Now, as shown in the proof of Lemma 6 above, if we set $r = (G + 4 + \xi)/2$ then:

$$E\left\{\|(Z_t' \otimes I_G)\lambda\|^2 \chi\left(\|(Z_t' \otimes I_G)\lambda\|^2 > \delta_T^{-2}\right)\right\} \leq \delta_T^{2(r-1)} E\left(\|(Z_t' \otimes I_G)\lambda\|^{2r}\right),$$

which implies that:

$$E\left(\sup_{A \in A_{2T}} \left|R_{T,2}^*(\alpha_A)\right|\right) \leq 2L_1^2 \delta_T^{-G} \delta_T^{2(r-1)} E\left(\|(Z_t' \otimes I_G)\lambda\|^{2r}\right) = o(1),$$

since $2(r-1) - G = 2r - G - 2 = 2 + \xi > 0$ and $E\left(\|(Z_t' \otimes I_G)\lambda\|^{2r}\right) < \infty$, by Assumption A10, and $\delta_T = o(1)$, by Assumption A7. It follows from M that $\sup_{A \in \mathcal{A}_{2T}} \left|R_{T,2}^*(\alpha_A)\right| = o_p(1)$, and since $\sup_{A \in \mathcal{A}_{2T}} \left|R_{T,1}^*(\alpha_A)\right| = o_p(1)$, as shown above, then:

$$\sup_{A \in \mathcal{A}_{2T}} |S_{2T}(\alpha_A)| = o_p(1).$$

Third, observe that $R_T^e(\alpha_A) = \sum_{t=1}^T r_{tT}^e(\alpha_A)$, where

$$\begin{aligned} r_{tT}^e(\alpha_A) &= E\left\{T^{-1}\delta_T^{-G}\left[\lambda'(Z_t \otimes I_G)\mathcal{K}^{(1)}\left(\frac{Y_t - AZ_t}{\delta_T}\right)\right]^2 |\mathcal{F}_{t-1}\right\} \\ &= T^{-1}\delta_T^{-G} \int_{\mathbb{R}^G} \left[\lambda'(Z_t \otimes I_G)\mathcal{K}^{(1)}\left(\frac{Y_t - AZ_t}{\delta_T}\right)\right]^2 f_t(u|\mathcal{F}_{t-1})\,du \\ &= T^{-1} \int_{\mathbb{R}^G} \left[\lambda'(Z_t \otimes I_G)\mathcal{K}^{(1)}(s)\right]^2 f_t((A - A_0)Z_t + \delta_T s|\mathcal{F}_{t-1})\,ds. \end{aligned}$$

If we now define:

$$R^e_{t0}(\alpha_A, \delta) \equiv \int_{\mathbb{R}^G} \left[\lambda'(Z_t \otimes I_G)\mathcal{K}^{(1)}(s)\right]^2 f_t((A - A_0) Z_t + \delta s|\mathcal{F}_{t-1})\, ds.$$

then $R^e_T(\alpha_A) = \sum_{t=1}^T R^e_{t0}(\alpha_A, \delta)$ and $R_0(\alpha_A) = E[R^e_{t0}(\alpha_A)]$. But for any $A \in \mathcal{A}$ and δ it follows that $\{R^e_{t0}(\alpha_A, \delta)\}_{t=-\infty}^{\infty}$ is strictly stationary and ergodic, by Assumption A1. In addition, it follows by CS that:

$$|R^e_{t0}(\alpha_A, \delta)| \leq G L_0 L_8 \|\lambda\|^2 \|Z_t\|^2,$$

since $f_t(u|\mathcal{F}_{t-1}) \leq L_0$, by Assumption A2, and $R^e_{t0}(\alpha_A, \delta)$ is continuous in (α_A, δ), since $f_t(u|\mathcal{F}_{t-1})$ is continuous in u for all (t, u, ω), by Assumption A2, $E\left\{\|Z_t\|^2\right\} < \infty$, by Assumption A10 and $\int_{\mathbb{R}^G} \left\|\mathcal{K}^{(1)}(s)\right\| ds = L_{10}$, by Assumption A11. Furthermore, $R^e_{t0}(\alpha_A, \delta)$ is continuous in (α_A, δ). Hence, it follows by the uniform law of large numbers for strictly stationary ergodic process that:

$$\sup_{A\in\mathcal{A}} |S_{4T}(\alpha_A)| = \sup_{A\in\mathcal{A}} \left| T^{-1} \sum_{t=1}^{T} R^e_{t0}(\alpha_A, \delta_T) - E[R^e_{t0}(\alpha_A, \delta)] \right| = o_p(1),$$

since $\delta_T = o(1)$, by Assumption A6, and $\mathcal{A}$ is compact, by Assumption A3. □

Proof of Theorem 5 Note that the function

$$Q_0(\beta^r) = \left[\alpha_0 + \mathrm{vec}\left(\left(\Psi_r \Gamma_r^{-1}\right)\right)\right]' \left[D_0^{-1} B_0 D_0^{-1}\right]^{-1} \left[\alpha_0 + \mathrm{vec}\left(\left(\Psi_r \Gamma_r^{-1}\right)\right)\right]$$

is continuous in β^r. Since $D_0^{-1} B_0 D_0^{-1}$ is positive definite by Lemma 5, it follows that $Q_0(\beta^r) > 0$ for any $\beta^r : \alpha_0 \neq -\mathrm{vec}((\Psi_r \Gamma_r^{-1}))$ and $Q_0(\beta^r) = 0$ if and only if $\alpha_0 = -\mathrm{vec}((\Psi_r \Gamma_r^{-1}))$. Hence by Assumption A13 the minimum is unique. Note also that $Q_T(\beta^r) = [\widehat{\alpha}_T + \mathrm{vec}((\Psi_r \Gamma_r^{-1}))]' \left[\widehat{D}_T(\widehat{\alpha}_T)^{-1} \widehat{B}_T(\widehat{\alpha}_T) \widehat{D}_T(\widehat{\alpha}_T)^{-1}\right]^{-1} [\widehat{\alpha}_T + \mathrm{vec}((\Psi_r \Gamma_r^{-1}))]$ converges uniformly to $Q_0(\beta^r)$. Since $\mathcal{B}^r$ is compact by Assumption 14, all the assumptions of Theorem 2.1 of Newey and McFadden (1994) are satisfied and consequently $\widehat{\beta}^r_T \xrightarrow{p} \beta^r_0$.
□

Proof of Theorem 6 Since β^r_0 belongs to the interior of $\mathcal{B}^r$ by Assumption 16, and $\widehat{\beta}^r_T \xrightarrow{p} \beta^r_0$ by Theorem 5, it follows that the first order conditions of the minimization problem (9) are satisfied with probability approaching one, yielding

$$C\left(\widehat{\beta}_T^r\right)'\left[\widehat{Avar}\left(\widehat{\alpha}_T\right)\right]^{-1}\left[\widehat{\alpha}_T+\operatorname{vec}\left(\hat{\Psi}_r\hat{\Gamma}_r^{-1}\right)\right]=0$$

where, as before, $\widehat{Avar}\left(\widehat{\alpha}_T\right)=\widehat{D}_T\left(\widehat{\alpha}_T\right)^{-1}\widehat{B}_T\left(\widehat{\alpha}_T\right)\widehat{D}_T\left(\widehat{\alpha}_T\right)^{-1}$. Now by a Taylor expansion we have

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

where $\widetilde{\beta}_T^r$ 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 $\operatorname{vec}\left(\Psi_{0,r}\Gamma_{0,r}^{-1}\right)=-\alpha_0$ and therefore we have

$$C\left(\widehat{\beta}_T^r\right)'\left[\widehat{Avar}\left(\widehat{\alpha}_T\right)\right]^{-1}C\left(\widehat{\beta}_T^r\right)\sqrt{T\delta_T^{G+2}}\left(\widehat{\beta}_T^r-\beta_0^r\right)=-C\left(\widehat{\beta}_T^r\right)'\left[\widehat{Avar}\left(\widehat{\alpha}_T\right)\right]^{-1}\sqrt{T\delta_T^{G+2}}(\widehat{\alpha}_T-\alpha_0).$$

The result follows from Theorem 3, $C\left(\widehat{\beta}_T^r\right)=C_0+o_p(1)$ and the fact that $\operatorname{rank}(C_0)=G\left(G+K\right)-\rho$ by Assumption 15. □

A2. Bivariate distribution with zero mode

In this appendix we provide details on the bivariate distribution used in the second set of experiments presented in Section 4.

Let ξ_1 and ξ_2 be two independent gamma-distributed random variables with $\mathrm{E}\left(\xi_g\right)=\theta_g/\kappa_g$ and $\mathrm{Var}\left(\xi_g\right)=\theta_g/\kappa_g^2$, for $g\in\{1,2\}$. Then, the joint density of $f\left(\ln\xi_1,\ln\xi_2\right)$ is

$$f\left(\ln\xi_1,\ln\xi_2\right)=\frac{\kappa_1^{\theta_1}\kappa_2^{\theta_2}}{\Gamma\left(\theta_1\right)\Gamma\left(\theta_2\right)}\exp\left(\theta_1\ln\xi_1+\theta_2\ln\xi_2-\kappa_1\xi_1-\kappa_2\xi_2\right).$$

Next, consider the random variables ε_1 and ε_2 obtained as $\varepsilon_1=\ln\xi_1$ and $\varepsilon_2=\xi_1\ln\xi_2$. The usual results on change of variables lead to

$$f\left(\varepsilon_1,\varepsilon_2\right)=\frac{\kappa_1^{\theta_1}\kappa_2^{\theta_2}}{\Gamma\left(\theta_1\right)\Gamma\left(\theta_2\right)}\exp\left(\left(\theta_1-1\right)\varepsilon_1+\theta_2\frac{\varepsilon_2}{\exp\left(\varepsilon_1\right)}-\kappa_1\exp\left(\varepsilon_1\right)-\kappa_2\exp\left(\frac{\varepsilon_2}{\exp\left(\varepsilon_1\right)}\right)\right).$$

To find the values of the parameters for which $f\left(\varepsilon_1,\varepsilon_2\right)$ has mode at $\varepsilon_1=\varepsilon_2=0$, we take the derivatives of $f\left(\varepsilon_1,\varepsilon_2\right)$ with respect to ε_1 and ε_2 and evaluate them at $\varepsilon_1=\varepsilon_2=0$:

$$\begin{aligned}\left.\frac{\partial f\left(\varepsilon_1,\varepsilon_2\right)}{\partial\varepsilon_1}\right|_{\varepsilon_1=\varepsilon_2=0}&=\frac{-\kappa_1^{\theta_1}\kappa_2^{\theta_2}e^{-\kappa_1-\kappa_2}}{\Gamma\left(\theta_1\right)\Gamma\left(\theta_2\right)}\left(\kappa_1-\theta_1+1\right),\\\left.\frac{\partial f\left(\varepsilon_1,\varepsilon_2\right)}{\partial\varepsilon_2}\right|_{\varepsilon_1=\varepsilon_2=0}&=\frac{-\kappa_1^{\theta_1}\kappa_2^{\theta_2}e^{-\kappa_1-\kappa_2}}{\Gamma\left(\theta_1\right)\Gamma\left(\theta_2\right)}\left(\kappa_2-\theta_2\right).\end{aligned}$$

These results show that $f(\varepsilon_1, \varepsilon_2)$ will have mode at $\varepsilon_1 = \varepsilon_2 = 0$ when $\theta_1 = \kappa_1 + 1$ and $\theta_2 = \kappa_2$. Additionally, it is possible to show that $\mathrm{E}(\varepsilon_1) = (\psi_0(\theta_1) - \ln(\kappa_1))$ and $\mathrm{Var}(\varepsilon_1) = \psi_1(\theta_1)$, and that $\mathrm{E}(\varepsilon_2) = \theta_1(\psi_0(\theta_2) - \ln(\kappa_2))/\kappa_1$ and $\mathrm{Var}(\varepsilon_2) = \theta_1\left(\psi_1(\theta_2)(\theta_1 + 1) + (\psi_0(\theta_2) - \ln(\kappa_2))^2\right)/\kappa_1^2$, where $\psi_0(\cdot)$ and $\psi_1(\cdot)$ denote, respectively, the digamma and trigamma functions.

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.

Bernanke, B.S. (1986). "Alternative explanations of the money-income correlation," *Carnegie-Rochester Conference Series on Public Policy*, 25, 49-100.

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.

Davidson, J.E.H. (1994). *Stochastic limit theory.* Oxford: Oxford University Press.

Eddy, W. (1980). "Optimum kernel estimators of the mode," *Annals of Statistics*, 9, 870-882.

Ferguson, T.S. (1967). *Mathematical statistics: A decision theoretic approach.* New York (NY): Academic Press.

Freedman, D.A. (1975). "On tail probabilities for martingales," *Annals of Probability*, 3, 100-118.

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.

Hayashi, F. (2000). *Econometrics.* Princeton (NJ): Princeton University Press.

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.

Kemp, G.C.R. and Santos Silva, J.M.C. (2012). "Regression towards the mode," *Journal of Econometrics*, 170, 92-101.

Klemelä, J. (2005). "Adaptive estimation of the mode of a multivariate density," *Journal of Nonparametric Statistics*, 17, 83-105.

Koenker, R. (2005). *Quantile regression*, New York (NY): Cambridge University Press.

Konakov, V. (1973). "On the asymptotic normality of the mode of multidimensional distributions," *Theory of Probability & Its Applications*, 18, 794-799.

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.

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.

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.

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.

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.

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.

von Bahr, B. and Esseen, C.-G. (1965). "Inequalities for the rth absolute moment of a sum of random variables, $1 \leqq r \leqq 2$," *The Annals of Mathematical Statistics*, 36, 299-303.

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.