Expansions of GMM statistics that indicate their properties under weak and/or many instruments and the bootstrap

Frank Kleibergen*

Preliminary and Incomplete

January 30, 2004

Abstract

We construct higher order expressions for Wald and Lagrange multiplier (LM) GMM statistics that are based on 2step and continuous updating estimators (CUE). We show that the sensitivity of the limit distribution to weak and many instruments results from superfluous elements in the higher order expansion. When the instruments are strong and their number is small, these elements are of higher order and result in higher order biases. When instruments are weak and/or their number is large, they are, however, of zero-th order and influence the limiting distributions. Edgeworth approximations do not remove the superfluous elements. The expansion of the LM-CUE statistic, which is Kleibergen's (2003) K-statistic, does not contain the superfluous higher order elements so it is robust to weak or many instruments. An Edgeworth approximation of its finite sample distribution shows that the bootstrap reduces the size distortion. We compute power curves for tests on the autocorrelation parameter in a panel autoregressive model to illustrate the consequences of the higher order.terms and the improvement that results from applying the bootstrap.

JEL classification: C11, C20, C30

1 Introduction

The finite sample distributions of Generalized Method of Moments (GMM) estimators and statistics are affected by the quality and number of instruments, see *e.g.* Hansen *et. al.* (1996) and Stock *et. al.* (2002). It has therefore become customary to conduct non-identification pre-tests

*Department of Economics, Brown University, 64 Waterman Street, Providence, RI 02912, United States and Department of Quantitative Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands, Email:Frank_Kleibergen@brown.edu. Homepage: http://www.fee.uva.nl/ke/kleibergen.htm. The research in this paper is partly funded by the NWO grant "Empirical Comparison of Economic Models".

on the parameters. Pre-testing for parameter non-identification, however, implies that all subsequent inferential procedures are conditional on the outcome of the pre-test. Stock and Yogo (2001), for example, show that 2-step GMM estimators are still considerably biased at moderate, but significant at the 95% level, values of the non-identification statistics. Inferential procedures have therefore been developped that are robust to many instruments, see *e.g.* Bekker (1994), and/or weak instruments, see *e.g.* Stock and Wright (2000), Kleibergen (2001,2003) and Moreira (2003).

We construct higher order expressions for LM and Wald GMM statistics that are based on 2step or continuous updating estimators (CUE), see Hansen *et. al.* (1996). These higher order expressions indicate the behavior of the different statistics in case of weak and/or many instruments. In case of strong identification, the 2-step Wald and LM statistics, see Hansen (1982) and Newey and West (1987a), have higher order elements that distort their limit distributions when the instruments become weak or irrelevant. Edgeworth approximations of the finite sample distributions of these statistics remain sensitive to these higher order elements. The bias caused by the higher order elements implies a further distortion of the limit distribution when the number of instruments gets large. The Wald and LM statistics that are based on the CUE do not possess these higher order elements. The limit distribution of the Wald-CUE statistic remains, however, because it uses the estimated parameter value in the covariance matrix estimator, sensitive to weak instruments. The limit distribution of the LM-CUE statistic, which is Kleibergen's (2001,2003) K-statistic, is robust to weak and/or many instruments. The absence of the higher order elements implies that the limit distribution of the K-statistic is also a better approximation of its finite sample distribution in case of appropriate identified parameters. The Edgeworth approximation shows that we can further improve upon this approximation by using the bootstrap.

Tests of misspecification hypothezes can also be based upon the robust K-statistic. The higher order expression of the resulting misspecification statistic also indicates its robustness to weak instruments when compared to misspecification statistics that are based on non-robust statistics. The robustness of this misspecification statistic again results from the improved approximation of the finite sample distribution by the limiting distribution in case of valid instruments.

The outline of the paper is as follows. The second section discusses GMM and states its assumptions. The third section constructs the higher order expressions of the 2-step Wald, LM and CUE Wald, LM statistics under different limit sequences of a GMM concentration parameter and the number of instruments. It shows that an Edgeworth approximation of the finite sample distribution of the 2-step Wald and LM statistic does not remove the sensitivity to higher order elements. The fourth section discusses misspecification statistics. The fifth section constructs an Edgeworth approximation of the finite sample distribution of the LM-CUE or K-statistic. It shows that the bootstrap improves upon the limit distribution of the K-statistic as an approximation of its finite sample distribution. The sixth section shows the consequences of the higher order expressions and the bootstrap for a size and power comparison that tests the autoregressive parameter in a panel autoregressive model of order 1. The seventh section concludes.

Throughout the paper we use the notation: $a = \text{vec}(A)$ for the column vectorization of the $n \times m$ matrix A such that for $A = (a_1 \cdots a_m)$, $\text{vec}(A) = (a_1' \cdots a_m')'$ and I_m is the $m \times m$ identity matrix. Furthermore, "$\underset{p}{\rightarrow}$" stands for convergence in probability and "$\underset{d}{\rightarrow}$" for convergence in distribution.

2 Generalized Method of Moments

We consider the estimation of the $m \times 1$ dimensional parameter vector $\theta = (\theta_1 \dots \theta_m)'$, whose parameter region is the $\mathbb{R}^m$, for which the $l \times 1$ dimensional moment equation

$$E\left[\varphi(\theta_0, Y_t)\right] = 0 \tag{1}$$

holds, with E the expectation operator. The data vector Y_t is observed for observation t. The $l \times 1$ dimensional vector function φ of θ is finite for finite values of θ, continuous and twice continuous differentiable. The specific true value of θ, at which (1) holds, is equal to θ_0. To estimate the parameter θ in (1), we use Hansen's (1982) GMM framework. We involve a k-dimensional vector of instruments X_t that is such that k_f $(= kl)$ exceeds m. The instruments are uncorrelated with $\varphi(\theta_0, Y_t)$,

$$E\left[X_t \varphi(\theta_0, Y_t)'\right] = 0. \tag{2}$$

For a data-set $(Y_t, X_t,\ t = 1, \dots, T)$, the objective function in the GMM framework reads

$$Q(\theta) = f_T(\theta, Y)'\hat{V}_{ff}(\theta)^{-1} f_T(\theta, Y), \tag{3}$$

with $f_T(\theta, Y) = \sum_{t=1}^{T} f_t(\theta)$,

$$f_t(\theta) = \mathrm{vec}\left(X_t \varphi(\theta, Y_t)'\right) = \left(\varphi(\theta, Y_t) \otimes X_t\right), \tag{4}$$

and $V_{ff}(\theta)$ is the covariance matrix of $f_T(\theta, Y)$ with $\bar{f}_t(\theta) = f_t(\theta) - E(f_t(\theta))$,

$$V_{ff}(\theta) = \lim_{T\to\infty} E\left\{\tfrac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T} \bar{f}_t(\theta)\bar{f}_j(\theta)'\right\} \tag{5}$$

while $\hat{V}_{ff}(\theta)$ is a consistent estimator of $V_{ff}(\theta_0)$,

$$\hat{V}_{ff}(\theta) \underset{p}{\to} V_{ff}(\theta_0). \tag{6}$$

To construct higher order expressions of test statistics, we make an assumption about the behavior of $f_t(\theta)$ and its derivative with respect to θ, see Kleibergen (2003).

Assumption 1. *The $k_f \times 1$ dimensional derivative of $f_t(\theta_0)$ with respect to θ_i,*

$$p_{i,t}(\theta_0) = \tfrac{\partial f_t(\theta)}{\partial \theta_i}\Big|_{\theta_0} : k_f \times 1, \qquad i = 1, \dots, m, \tag{7}$$

is such that

$$\bar{p}_{i,t}(\theta_0) = A_i \bar{q}_{i,t}(\theta_0) \tag{8}$$

with $\bar{p}_{i,t}(\theta_0) = p_{i,t}(\theta_0) - E(p_{i,t}(\theta_0))$, $q_{i,t}(\theta_0) : k_i \times 1$, $\bar{q}_{i,t}(\theta_0) = q_{i,t}(\theta_0) - E(q_{i,t}(\theta_0))$ and A_i a deterministic full-rank $k_f \times k_i$ dimensional matrix, $k_i \leq k_f$. The behavior of the sums of the martingale difference series $\bar{f}_t(\theta_0)$ $(= f_t(\theta_0))$ and $\bar{q}_t(\theta_0) = (\bar{q}_{1t}(\theta_0)' \dots \bar{q}_{mt}(\theta_0)')'$ reads

$$\tfrac{1}{\sqrt{T}} \sum_{t=1}^{T} \begin{pmatrix} \bar{f}_t(\theta_0) \\ \bar{q}_t(\theta_0) \end{pmatrix} = m_0, \tag{9}$$

where $m_0 : (k_f + k_\theta) \times 1,\ k_\theta = \sum_{i=1}^m k_i$; *and*

$$m_0 \underset{d}{\rightarrow} \begin{pmatrix} \psi_f \\ \psi_\theta \end{pmatrix} \tag{10}$$

with $\psi_f : k_f \times 1,\ \psi_\theta : k_\theta \times 1,$

$$\begin{pmatrix} \psi_f \\ \psi_\theta \end{pmatrix} \sim N(0, V(\theta)), \tag{11}$$

and

$$V(\theta) = \begin{pmatrix} V_{ff}(\theta) & V_{f\theta}(\theta) \\ V_{\theta f}(\theta) & V_{\theta\theta}(\theta) \end{pmatrix}, \tag{12}$$

with $V_{ff}(\theta) : k_f \times k_f,\ V_{\theta f}(\theta) = V_{f\theta}(\theta)' : k_\theta \times k_f,\ V_{\theta\theta}(\theta) : k_\theta \times k_\theta$, *and*

$$V(\theta) = \ \lim_{T\to\infty} E\left\{ \tfrac{1}{T} \sum_{t=1}^T \sum_{j=1}^T \begin{pmatrix} \bar{f}_t(\theta) \\ \bar{q}_t(\theta) \end{pmatrix} \begin{pmatrix} \bar{f}_j(\theta) \\ \bar{q}_j(\theta) \end{pmatrix}' \right\}. \tag{13}$$

Assumption 1 implies that $(\bar{f}_t(\theta)'\ \bar{q}_t(\theta)')'$, $t = 1, \ldots, T$, is a stationary series so Assumption 1 is a central limit theorem for stationary series. It is therefore satisfied under weak conditions for $\bar{f}_t(\theta_0)$ and $\bar{q}_t(\theta_0)$. Sufficient conditions that ensure such convergence are that: 1. the r-th moment of the absolute value of $\bar{f}_t(\theta_0)$ and $\bar{q}_{i,t}(\theta_0)$, $i = 1, \ldots, m$, is finite for some $r > 2$, 2. $V(\theta)$ is well-defined and 3. the average value of the outer-product of $(\bar{f}_t(\theta_0)'\ \bar{q}_t(\theta_0)')'$ converges in probability to $V(\theta)$, see *e.g.* White (1984).

Additional higher order terms can be added to the behavior of $(\bar{f}_t(\theta)'\ \bar{q}_t(\theta)')'$ in (9). We did not add such terms as they obstruct the construction of the higher behavior of the statistics which we conduct lateron.

The A_i matrices in Assumption 1 allow for a degenerate limit behavior of $\frac{\partial}{\partial\theta_i} f_t(\theta)$ in which case A_i is equal to zero. For more details on the specification of the A_i-matrices, we refer to Kleibergen (2003).

We use Assumption 1 to determine the convergence rate of the limit behavior of

$$\begin{array}{rl} D_T(\theta_0, Y) = & [\ p_{1,T}(\theta_0, Y) - A_1 V_{\theta f,1}(\theta_0) V_{ff}(\theta_0)^{-1} f_T(\theta_0, Y) \quad \cdots \\ & \quad p_{m,T}(\theta_0, Y) - A_m V_{\theta f,m}(\theta_0) V_{ff}(\theta_0)^{-1} f_T(\theta_0, Y)\ ], \end{array} \tag{14}$$

with $V_{\theta f,i}(\theta_0) : k_i \times k_f$, $i = 1, \ldots, m$, $V_{\theta f}(\theta_0) = (V_{\theta f,1}(\theta_0)' \ldots V_{\theta f,m}(\theta_0)')'$, $p_{i,T}(\theta_0, Y) : k_f \times 1$, $i = 1, \ldots, m$, $p_T(\theta_0, Y) = (p_{1,T}(\theta_0, Y) \ldots p_{m,T}(\theta_0, Y))$, $p_{i,T}(\theta_0, Y) = \sum_{i=1}^T p_{i,t}(\theta_0)$. We are intested in the behavior of $D_T(\theta_0, Y)$ since the derivative of $f_T(\theta, Y)' V_{ff}(\theta)^{-1} f_T(\theta, Y)$ with respect to θ equals $2D_T(\theta, Y)' V_{ff}(\theta)^{-1} f_T(\theta, Y)$, see Kleibergen (2003).

Lemma 1. *When Assumption 1 holds, the behavior of* $T^{-\frac{1}{2}(1+\nu)} D_T(\theta_0, Y)$ *is characterized by*

$$T^{-\frac{1}{2}(1+\nu)} D_T(\theta_0, Y) = D_0 + O_p(T^{-\frac{1}{2}(\nu+1)}), \tag{15}$$

where $D_0 = T^{\frac{1}{2}(1-\nu)} E\left[\frac{1}{T} \sum_{t=1}^T p_t(\theta_0)\right] + T^{-\frac{1}{2}\nu} [A_1(m_{0,\theta_1} - V_{\theta f,1}(\theta_0) V_{ff}(\theta_0)^{-1} m_{0,f}) \cdots A_m(m_{0,\theta_m} - A_m V_{\theta f,m}(\theta_0)\ V_{ff}(\theta_0)^{-1} m_{0,f})],\ m_0 = (m_{0,f}'\ m_{0,\theta}'),\ m_{0,\theta} = (m_{0,\theta_1}' \ldots m_{0,\theta_m}')'.$

Proof. results directly from Assumption 1 when we note that $p_{i,T}(\theta_0,Y)=\sum_{i=1}^{T}\bar{p}_{i,t}(\theta_0)$ and $\bar{p}_{i,t}(\theta_0)=A_i\bar{q}_{i,t}(\theta_0)$ in case A_i does not equal zero. ■

The derivative $D_T(\theta_0,Y)$ is constructed in such a manner that D_0 has a number of convenient properties which we state in the following two corollaries. One of these corollaries deals with the appropriate choice of the convergence rate ν.

Corollary 1. *When Assumption 1 holds,*

$$\text{vec}\left[T^{\frac{1}{2}}\left(T^{-\frac{1}{2}(1-\nu)}D_0-J_\theta(\theta_0)\right)\right] \quad = m_{0,\theta.f} \tag{16}$$

with

$$J_\theta(\theta_0)=\lim_{T\to\infty}E\left[\tfrac{1}{T}\sum_{t=1}^{T}p_t(\theta_0)\right], \tag{17}$$

$m_{0,\theta.f}=m_{0,\theta}-V_{\theta f}(\theta_0)V_{ff}(\theta_0)^{-1}m_{0,f}$ *and*

$$m_{0,\theta.f}\underset{d}{\to}\quad A\psi_{\theta.f} \tag{18}$$

where $A=$diag$(A_1,\ldots,A_m)$, $\psi_{\theta.f}=\psi_\theta-V_{\theta f}(\theta_0)V_{ff}(\theta_0)^{-1}\psi_f$ *and*

$$\psi_{\theta.f}\sim N(0,V_{\theta\theta.f}(\theta_0)), \tag{19}$$

with $V_{\theta\theta.f}(\theta_0)=V_{\theta\theta}(\theta_0)-V_{\theta f}(\theta_0)V_{ff}(\theta_0)^{-1}V_{f\theta}(\theta_0)$, *and* $\psi_{\theta.f}$ *is independent of* ψ_f.

Proof. see Kleibergen (2003). ■

Corollary 1 shows that $D_T(\theta_0,Y)$ is an estimator of the Jacobian $J_\theta(\theta_0)$ whose limit behavior is independent of the limit behavior of $f_T(\theta_0,Y)$.

Corollary 2. *Given* $J_\theta(\theta_0)$, *the convergence rate* ν *in Lemma 1 is such that:*

1. *For a fixed full rank value of* $J_\theta(\theta_0):\nu=1$ *so* $D_0\underset{p}{\to}J_\theta(\theta_0)$ *and*

$$D_0'V_{ff}(\theta_0)^{-1}D_0\underset{p}{\to}J_\theta(\theta_0)'V_{ff}(\theta_0)^{-1}J_\theta(\theta_0). \tag{20}$$

2. *For a weak value of* $J_\theta(\theta_0)$ *such that* $J_\theta(\theta_0)=J_{\theta,T}$, $J_{\theta,T}=\frac{1}{\sqrt{T}}C$, $C:k_f\times m$ *and* rank$(C)=m:\nu=0$, $D_0\underset{d}{\to}C+(A_1\psi_{\theta.f,1}\ldots A_m\psi_{\theta.f,m})$ *and*

$$D_0'V_{ff}(\theta_0)^{-1}D_0\underset{d}{\to}[C+(A_1\psi_{\theta.f,1}\ldots A_m\psi_{\theta.f,m})]'V_{ff}(\theta_0)^{-1}[C+(A_1\psi_{\theta.f,1}\ldots A_m\psi_{\theta.f,m})]. \tag{21}$$

3. *For a zero value of* $J_\theta(\theta_0)$: $\nu=0$, $D_0\underset{d}{\to}(A_1\psi_{\theta.f,1}\ldots A_m\psi_{\theta.f,m})$ *and*

$$D_0'V_{ff}(\theta_0)^{-1}D_0\underset{d}{\to}(A_1\psi_{\theta.f,1}\ldots A_m\psi_{\theta.f,m})'V_{ff}(\theta_0)^{-1}(A_1\psi_{\theta.f,1}\ldots A_m\psi_{\theta.f,m}). \tag{22}$$

The first case in Corollary 2 is the traditional setting of a fixed full rank value of the expected Jacobian. In this setting, GMM-estimators have normal limiting distributions, see *e.g.*Hansen (1982) and Newey and McFadden (1994), which does not result in the other two cases. Case 2 deals with weak instruments, see *e.g.* Staiger and Stock (1997) and Stock and Wright (2000), while the instruments are irrelevant in Case 3. Corollary 2 shows that the convergence rate ν is a function of the strength of the instruments, *i.e.* $\nu = 1$ in case of valid instruments while $\nu = 0$ in case of weak or irrelevant instruments. The convergence rate of $D_T(\theta_0, Y)$ in Lemma 1 therefore depends on ν. Corollary 1 shows that the limit behavior of D_0 is independent of the limit behavior of $m_{0.f}$ so the higher order expressions of statistics that test $\mathrm{H}_0 : \theta = \theta_0$ are polynomials of $T^{-\frac{1}{2}\nu}$. Rothenberg (1984) constructs the higher order properties of estimators and test statistics in the linear instrumental variables regression model as a function of the concentration parameter. The statistic $\frac{1}{T^{1+\nu}} D_T(\theta_0, Y)' V_{ff}(\theta_0)^{-1} D_T(\theta_0, Y)$ has a limit behavior that is independent of $m_{0.f}$ and is comparable to the concentration parameter in the linear instrumental variables regression model. We therefore use it to obtain higher order properties of test statistics.

We specify the test statistics as polynomials of the convergence behavior of $D_T(\theta_0, Y)$, *i.e.* $T^{-\frac{1}{2}\nu}$, see *e.g.* Nagar (1959). The derivative matrix $D_T(\theta_0, Y)$ does, however, depend on the covariance matrix $V(\theta_0)$ which is typically unknown. We therefore replace it with an estimator, $\hat{V}(\theta_0)$. We account for the unknown covariance matrix by specifying the test statistics as polynomials of the convergence behaviors of $D_T(\theta_0, Y)$ and the covariance matrix estimator $\hat{V}(\theta_0)$ on which we make the following assumption.

Assumption 2: *The convergence of the covariance matrix estimator $\hat{V}(\theta_0)$ is such that*

$$T^{\frac{1}{2}\mu}\mathrm{vec}(\hat{V}(\theta_0) - V(\theta_0)) = u_0 + O_p(T^{-\frac{1}{2}\mu}), \tag{23}$$

with μ the convergence rate of the covariance matrix estimator and u_0 (=vec(U_0)) converges to a normal distributed random variable,

$$u_0 \underset{d}{\rightarrow} \psi_u,$$

where $S'_{(m+1)k_f}\psi_u \sim N(0, W(\theta_0))$, with $S_j : j^2 \times [\frac{1}{2}j(j+1)]$ a selection matrix that selects the unique elements of the vectorization of a symmetric $j \times j$ matrix and $W(\theta_0)$ is the covariance matrix.

Assumption 2 does not specify the covariance matrix estimator and therefore allows for parametric as well as non-parametric covariance matrix estimators, see *e.g.* Andrews (1991) and Newey and West (1987b). These estimators lead to different convergence rates μ. We indicate usage of the covariance matrix estimator $\hat{V}(\theta_0)$ in the specification of $D_T(\theta_0, Y)$ by denoting it by $\hat{D}_T(\theta_0, Y)$.

3 Higher Order Properties of Statistics that test $\mathbf{H}_0 : \theta = \theta_0$.

We analyze the higher order properties of four statistics that test $\mathrm{H}_0 : \theta = \theta_0$:

1. GMM-Wald statistic evaluated at the 2-step GMM estimator, $\hat{\theta}_{2s}$, see *e.g.* Hansen (1982):[1]

$$\begin{aligned} \mathrm{W}_{2s}(\theta_0) = &\ (\hat{\theta}_{2s} - \theta_0)' \left[\tfrac{1}{T} p_T(\hat{\theta}_{2s}, Y)' \hat{V}_{ff}(\hat{\theta}_{2s})^{-1} p_T(\hat{\theta}_{2s}, Y)\right] (\hat{\theta}_{2s} - \theta_0) \\ \approx &\ \tfrac{1}{T} f_T(\theta_0, Y)' V_{ff}(\hat{\theta}_{2s})^{-1} p_T(\hat{\theta}_{2s}, Y) \left[p_T(\hat{\theta}_{2s}, Y)' V_{ff}(\hat{\theta}_{2s})^{-1} p_T(\hat{\theta}_{2s}, Y)\right]^{-1} \\ &\ p_T(\hat{\theta}_{2s}, Y)' V_{ff}(\hat{\theta}_{2s})^{-1} f_T(\theta_0, Y). \end{aligned} \tag{24}$$

2. GMM-Wald statistic evaluated at the continuous updating estimator (CUE), $\hat{\theta}_{\text{cue}}$, of Hansen *et. al.* (1996):

$$\begin{aligned} \mathrm{W}_{\text{cue}}(\theta_0) = &\ (\hat{\theta}_{\text{cue}} - \theta_0)' \left[\tfrac{1}{T} \hat{D}_T(\hat{\theta}_{\text{cue}}, Y)' V_{ff}(\hat{\theta}_{\text{cue}})^{-1} \hat{D}_T(\hat{\theta}_{\text{cue}}, Y)\right] (\hat{\theta}_{\text{cue}} - \theta_0) \\ \approx &\ \tfrac{1}{T} f_T(\theta_0, Y)' V_{ff}(\hat{\theta}_{\text{cue}})^{-1} \hat{D}_T(\hat{\theta}_{\text{cue}}, Y) \left[\hat{D}_T(\hat{\theta}_{\text{cue}}, Y)' \hat{V}_{ff}(\hat{\theta}_{\text{cue}})^{-1} \hat{D}_T(\hat{\theta}_{\text{cue}}, Y)\right]^{-1} \\ &\ \hat{D}_T(\hat{\theta}_{\text{cue}}, Y)' \hat{V}_{ff}(\hat{\theta}_{\text{cue}})^{-1} f_T(\theta_0, Y). \end{aligned} \tag{25}$$

   The first order condition for a minimal value of $Q(\theta)$ is: $\hat{D}_T(\theta, Y)' \hat{V}_{ff}(\theta)^{-1} f_T(\theta, Y) = 0$ so $\hat{D}_T(\hat{\theta}_{cue}, Y)' \hat{V}_{ff}(\hat{\theta}_{cue})^{-1} f_T(\hat{\theta}_{cue}, Y) = 0$, see Kleibergen (2001). This explains the second part of (25), which results from a Taylor approximation, that we use to obtain the higher order properties of $\mathrm{W}_{\text{cue}}(\theta_0)$.

3. GMM-Lagrange multiplier (LM) statistic, see Newey and West (1987a):

$$\begin{aligned} \mathrm{LM}(\theta_0) = &\ \tfrac{1}{T} f_T(\theta_0, Y)' \hat{V}_{ff}(\theta_0)^{-1} p_T(\theta_0, Y) \left[p_T(\theta_0, Y)' \hat{V}_{ff}(\theta_0)^{-1} p_T(\theta_0, Y)\right]^{-1} \\ &\ p_T(\theta_0, Y)' \hat{V}_{ff}(\theta_0)^{-1} f_T(\theta_0, Y). \end{aligned} \tag{26}$$

4. K-statistic, see Kleibergen (2001):

$$\begin{aligned} \mathrm{K}(\theta_0) = &\ \tfrac{1}{T} f_T(\theta_0, Y)' \hat{V}_{ff}(\theta_0)^{-1} \hat{D}_T(\theta_0, Y) \left[\hat{D}_T(\theta_0, Y)' \hat{V}_{ff}(\theta_0)^{-1} \hat{D}_T(\theta_0, Y)\right]^{-1} \\ &\ \hat{D}_T(\theta_0, Y)' \hat{V}_{ff}(\theta_0)^{-1} f_T(\theta_0, Y). \end{aligned} \tag{27}$$

$\mathrm{W}_{2s}(\theta_0)$ and $\mathrm{LM}(\theta_0)$ are a Wald and LM statistic that are based on the two-step estimator $\hat{\theta}_{2s}$ while $\mathrm{W}_{cue}(\theta_0)$ and $\mathrm{K}(\theta_0)$ are a Wald and LM statistic that are based on the CUE $\hat{\theta}_{\text{cue}}$.

Under a fixed full rank of $J_\theta(\theta_0)$, $\mathrm{W}_{2s}(\theta_0)$, $\mathrm{W}_{\text{cue}}(\theta_0)$ and $\mathrm{LM}(\theta_0)$ have a $\chi^2(m)$ zero-th order limit distribution, see *e.g.* Newey and McFadden (1994). The zero-th order limit distribution of $\mathrm{K}(\theta_0)$ is $\chi^2(m)$ regardless of the value of $J_\theta(\theta_0)$, see Kleibergen (2003). $\mathrm{W}_{2s}(\theta_0)$ and $\mathrm{LM}(\theta_0)$ are

We construct higher order expressions of $\mathrm{W}_{2s}(\theta_0)$, $\mathrm{W}_{\text{cue}}(\theta_0)$, $\mathrm{LM}(\theta_0)$ and $\mathrm{K}(\theta_0)$ as polynomials of the convergence rates of $D_T(\theta_0, Y)' V_{ff}(\theta_0)^{-1} D_T(\theta_0, Y)$ and $\hat{V}(\theta_0)$. We also consider a convergence process where the number of observations and the number of instruments jointly converge to infinity as in Bekker (1994).

[1] The second expression of $\mathrm{W}_{2s}(\theta_0)$ results from a Taylor approximation of $f_T(\theta_0, Y)$. We use this expression to obtain the higher order properties of $\mathrm{W}_{2s}(\theta_0)$.

3.1 Fixed number of instruments

Theorem 1 states the higher order expressions that result from Assumptions 1 and 2, of $\mathrm{W}_{2\mathrm{s}}(\theta_0)$, $\mathrm{W}_{\mathrm{cue}}(\theta_0)$, $\mathrm{LM}(\theta_0)$ and $\mathrm{K}(\theta_0)$ in case of a fixed number of instruments. Theorem 1 specifies the higher order expressions as functions of the parameters ν and μ that characterize the convergence rates of $D_T(\theta_0, Y)$, $T^{-\frac{1}{2}(1+\nu)}$, and $\hat{V}(\theta_0)$, $T^{-\frac{1}{2}\mu}$.

Theorem 1. *When the number of instruments k is fixed, Assumptions 1 and 2 imply higher order expressions for $W_{2s}(\theta_0)$, $W_{cue}(\theta_0)$, $LM(\theta_0)$ and $K(\theta_0)$ under $H_0 : \theta = \theta_0$ that are characterized by:*

$$\left.\begin{array}{r} \mathrm{W}_{2s}(\theta_0) \\ \mathrm{W}_{cue}(\theta_0) \\ \mathrm{LM}(\theta_0) \\ \mathrm{K}(\theta_0) \end{array}\right\} = \left\{\begin{array}{ll} n_0 + & \text{:zero-th order} \\ T^{-\frac{\nu}{2}} n_\nu + T^{-\nu} n_{2\nu} + T^{-\frac{3}{2}\nu} n_{3\nu} + & :D_T(\theta_0, Y) \\ T^{-\frac{\kappa}{2}} n_\kappa + T^{-\kappa} n_{2\kappa} + & :\hat{V}(\theta_0) \\ T^{-\frac{\nu+\kappa}{2}} n_{\nu+\kappa} + T^{-\frac{1}{2}(2\nu+\kappa)} n_{2\nu+\kappa} + T^{-\frac{1}{2}(\nu+2\kappa)} n_{\nu+2\kappa} + & \text{:mixed} \\ o_p(T^{-\frac{3}{2}\nu}), & \end{array}\right. \tag{28}$$

where:

1. *for $W_{2s}(\theta_0) : \kappa = \min(\nu, \mu)$ and*

$$\begin{array}{rcl} n_0 & = & s_0' G_0^{-1} s_0 \\ D_T(\theta_0, Y) & : & \left\{\begin{array}{rl} n_\nu = & s_0' Q_1 s_0 + s_{1\nu,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{1\nu,1} \\ n_{2\nu} = & s_{1\nu,1}' Q_1 s_0 + s_0' Q_1 s_{1\nu,1} + s_{1\nu,1}' G_0^{-1} s_{1\nu,1} \\ n_{3\nu} = & s_{1\nu,1}' Q_1 s_{1\nu,1} \end{array}\right. \\ \hat{V}(\theta_0) & : & \left\{\begin{array}{rl} n_\kappa = & s_{1\kappa,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{1\kappa,1} \\ n_{2\kappa} = & s_{1\kappa,1}' G_0^{-1} s_{1\kappa,1} \end{array}\right. \\ \text{mixed} & : & \left\{\begin{array}{rl} n_{\nu+\kappa} = & s_{1\nu,1}' G_0^{-1} s_{1\kappa,1} + s_{1\kappa,1}' G_0^{-1} s_{1\nu,1} + s_{1\kappa,1}' Q_1 s_0 + s_0' Q_1 s_{1\nu+\kappa} + \\ & (s_{\nu+\kappa,1} + s_{\nu+\kappa,2})' G_0^{-1} s_0 + s_0' G_0^{-1} (s_{\nu+\kappa,1} + s_{\nu+\kappa,2}) \\ n_{2\nu+\kappa} = & s_{1\nu,1}' Q_1 s_{1\kappa,1} + s_{1\kappa,1}' Q_1 s_{1\nu,1} + (s_{\nu+\kappa,1} + s_{\nu+\kappa,2})' G_0^{-1} s_{1\nu,1} + \\ & s_{1\nu,1}' G_0^{-1} (s_{\nu+\kappa,1} + s_{\nu+\kappa,2}) \\ n_{\nu+2\kappa} = & (s_{\nu+\kappa,1} + s_{\nu+\kappa,2})' G_0^{-1} s_{1\kappa,1} + s_{1\kappa,1}' G_0^{-1} (s_{\nu+\kappa,1} + s_{\nu+\kappa,2}) + \\ & s_{\nu+2\kappa,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{\nu+\kappa,1} + s_{1\kappa,1}' Q_1 s_{1\kappa,1}. \end{array}\right. \end{array} \tag{29}$$

2. *for $W_{cue}(\theta_0) : \kappa = \min(\nu, \mu)$, all terms that result from $D_T(\theta_0, Y) : n_\nu$, $n_{2\nu}$, $n_{3\nu}$ and $n_{2\nu+\kappa}$ are equal to zero and*

$$\begin{array}{rcl} n_0 & = & s_0' G_0^{-1} s_0 \\ \hat{V}(\theta_0) & : & \left\{\begin{array}{rl} n_\kappa = & s_0' Q_1 s_0 + s_{1\kappa,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{1\kappa,1} \\ n_{2\kappa} = & s_{1\kappa,1}' G_0^{-1} s_{1\kappa,1} + s_{1\kappa,1}' Q_1 s_0 + s_0' Q_1 s_{1\kappa,1} \end{array}\right. \\ \text{mixed} & : & \left\{\begin{array}{rl} n_{\nu+\kappa} = & s_{\nu+\kappa,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{\nu+\kappa,1} \\ n_{\nu+2\kappa} = & s_{\nu+\kappa,1}' G_0^{-1} s_{1\kappa,1} + s_{1\kappa,1}' G_0^{-1} s_{\nu+\kappa,1} + s_{\nu+2\kappa,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{\nu+2\kappa,1} + \\ & s_0' Q_1 s_{\nu+\kappa,1} + s_{\nu+\kappa,1}' Q_1 s_0. \end{array}\right. \end{array} \tag{30}$$

3. *for $LM(\theta_0)$ the elements are identical to those for $W_{2s}(\theta_0)$ in (29) but with $\kappa = \mu$.*

4. *for* $K(\theta_0)$ *the elements are identical to those for* $W_{cue}(\theta_0)$ *in (30) but with* $\kappa = \mu$ *and for all statistics:*

$$
\begin{array}{rcl}
s_0 & = & m_{0,f}' V_{ff}(\theta_0)^{-1} D_0 \\
D_T(\theta_0, Y) & : & \left\{ \begin{array}{rl} s_{1\nu,1} = & m_{0,f}' V_{ff}(\theta_0)^{-1} \{ \frac{1}{\sqrt{T}} [p_T(\hat{\theta}, Y) - \hat{D}_T(\hat{\theta}, Y)] \} \\ = & m_{0,f}' V_{ff}(\theta_0)^{-1} [A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)] \\ & [I_m \otimes V_{ff}(\theta_0)^{-1} m_{0,f}] \end{array} \right. \\
\hat{V}(\theta_0) & : & \left\{ s_{1\kappa,1} = T^{\frac{\kappa}{2}} m_{0,f}' [\hat{V}_{ff}(\theta_0)^{-1} - V_{ff}(\theta_0)^{-1}] D_0 \right. \\
\text{mixed} & : & \left\{ \begin{array}{rl} s_{\nu+\kappa,1} = & T^{\frac{1}{2}(\kappa-1)} m_{0,f}' V_{ff}(\theta_0)^{-1} [\hat{D}_T(\hat{\theta}, Y) - D_T(\theta_0, Y)] \\ = & T^{\frac{1}{2}\kappa} m_{0,f}' V_{ff}(\theta_0)^{-1} \{ [A_1 \hat{V}_{\theta f,1}(\theta_0) \cdots A_m \hat{V}_{\theta f,m}(\theta_0)][I_m \otimes \hat{V}_{ff}(\theta_0)^{-1}] \\ & -[A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)][I_m \otimes V_{ff}(\theta_0)^{-1}] \} [I_m \otimes m_{0,f}] \\ s_{\nu+\kappa,2} = & T^{\frac{1}{2}\kappa} m_{0,f}' [\hat{V}_{ff}(\theta_0)^{-1} - V_{ff}(\theta_0)^{-1}][A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)] \\ & [I_m \otimes V_{ff}(\theta_0)^{-1} m_{0,f}] \\ s_{\nu+2\kappa,1} = & T^{\frac{1}{2}(2\kappa-1)} m_{0,f}' [\hat{V}_{ff}(\theta_0)^{-1} - V_{ff}(\theta_0)^{-1}][\hat{D}_T(\theta_0, Y) - D_T(\theta_0, Y)], \end{array} \right.
\end{array}
\tag{31}
$$

with $G_0 = D_0' V_{ff}(\theta_0) D_0$ *and the expressions for the remaining G and Q matrices are given in the Appendix.*

Proof. see the Appendix. ■

The $D_T(\theta_0, Y)$, $\hat{V}(\theta_0)$ and mixed terms in Theorem 1 indicate where the higher order terms originate from. Unlike the convergence rate μ of the covariance matrix estimator, the convergence rate ν of $D_T(\theta, Y)$ is unknown. The higher order expressions in Theorem 1 therefore depend on the unknown convergence rate of the concentration parameter. The parameter κ in Theorem 1 indicates that the convergence rate of the covariance matrix estimator depends on the involved value of θ. The Wald statistics, $\mathrm{W}_{2s}(\theta_0)$ and $\mathrm{W}_{cue}(\theta_0)$, use the covariance matrix estimator at the estimated value of θ, $\hat{\theta}$. The convergence rate of the covariance matrix estimator for the Wald statistics is therefore equal to the minimum of the convergence rate of $\hat{\theta}$, $T^{\frac{1}{2}\nu}$, and $\hat{V}(\theta_0)$, $T^{\frac{1}{2}\mu}$, which we indicate by $T^{\frac{1}{2}\kappa}$ with $\kappa = \min(\mu, \nu)$. The Lagrange multiplier statistics $\mathrm{LM}(\theta_0)$ and $\mathrm{K}(\theta_0)$ use the covariance matrix estimator evaluated at θ_0. The convergence rate of the covariance matrix estimator in these statistics is therefore equal to $T^{\frac{1}{2}\mu}$. Hence, $\kappa = \mu$ for these statistics.

We analyze the higher order expressions from Theorem 2 for both $\nu = 0$ and $\nu = 1$. We first discuss $\nu = 1$ which, as shown in Corollary 2, corresponds with the traditional case of a fixed full rank value of $J_\theta(\theta_0)$. Afterwards we discuss $\nu = 0$ which leads to a limit distribution of some of the statistics that depends on nuisance parameters.

3.1.1 Identified parameters or $\nu = 1$

When $\nu = 1$, the zero-th order term and therefore the limit distribution is the same for all statistics in Theorem 1,

$$n_0 = s_0' G_0^{-1} s_0 \underset{d}{\rightarrow} \chi^2(m). \tag{32}$$

The higher order elements in Theorem 1 effect the accuracy of the approximation of the finite sample distribution by the limit of the zero-th order element. Higher order Edgeworth approximations have therefore been proposed to obtain a more accurate approximation of the finite sample distribution, see *e.g.* Bhattacharya and Ghosh (1978), Sargan (1980), Götze and Hipp (1983), Rothenberg (1984) and Phillips and Park (1988). Under a set of regularity conditions, Rothenberg (1984) states that a statistic S whose higher order properties are characterized by

$$S = \quad s_0 + \tfrac{1}{\sqrt{T}} s_1(s_0, y_0) + \tfrac{1}{T} s_2(s_0, y_0) + o_p(\tfrac{1}{T}), \tag{33}$$

with y_0 a vector of sample moments that converges to a random variable different from the random variable where s_0 converges to, has a second order Edgeworth approximation to its finite sample distribution that reads

$$\Pr[S \leq s] \approx F\left[s - \tfrac{1}{\sqrt{T}} s_1(s) + \tfrac{1}{2T}\left\{2s_1(s)\left[\tfrac{\partial}{\partial s} s_1(s)\right] + c(s)v_1(s) + \left[\tfrac{\partial}{\partial s} v_1(s)\right] - 2s_2(s)\right\}\right], \tag{34}$$

where F is the distribution function of the limiting distribution of s_0, $c(s) = \frac{\partial}{\partial s}\log[\frac{\partial}{\partial s}F(s)]$, $s_1(s) = E_{y_0}(s_1(s_0, y_0)|s_0 = s)$, $s_2(s) = E_{y_0}(s_2(s_0, y_0)|s_0 = s)$ and $v_1(s) = \text{var}_{y_0}(s_1(s_0, y_0)|s_0 = s)$. The second order Edgeworth approximation (34) removes the approximation errors of the finite sample distribution up to the second order. Hence, the difference between the finite sample distribution and the second order Edgeworth approximation is $O_p(T^{-\frac{3}{2}})$ while the difference between the finite sample distribution and the approximation by the limit of its zero-th order element is $O_p(T^{-\frac{1}{2}})$.

When we assume that $\mu = 1$ and that the regularity conditions for the second order Edgeworth approximation are satisfied, which imply that $\nu = 1$, we can construct the second order Edgeworth approximation for the statistics in Theorem 1. For $\text{W}_{2s}(\theta_0)$ and $\text{LM}(\theta_0)$, we then need to obtain the conditional expectation of n_v, $n_{2\nu}$, n_κ, $n_{2\kappa}$ and $n_{\nu+\kappa}$ given n_0. We just show that the second order Edgeworth approximation does not perform adequately for $\text{W}_{2s}(\theta_0)$ and $\text{LM}(\theta_0)$. We only need to construct the conditional expectation of n_v and $n_{2\nu}$ for this purpose. We show that these lead to a unsatisfactory performance of the second order Edgeworth approximation. In order to construct the conditional expectations of n_ν and $n_{2\nu}$, we adapt Assumption 2.

Assumption 2*. *The limiting distribution ψ_u from Assumption 2 is independent of ψ_f.*

In order to determine the properties of the second order Edgeworth approximation, we first obtain the limit expressions of the conditional expectations of n_ν and $n_{2\nu}$ given $\rho = (D_0' V_{ff}(\theta_0)^{-1} D_0)^{-1} D_0' V_{ff}(\theta_0)^{-1}\psi_f$ so $\lim_{T\to\infty} n_0 = \rho' D_0' V_{ff}(\theta_0)^{-1} D_0 \rho$. Because of the law of iterated expectations,

$$E[\lim_{T\to\infty} n_{i\nu}|\rho' D_0' V_{ff}(\theta_0)^{-1} D_0\rho = n_0] = \quad E[E[\lim_{T\to\infty} n_{i\nu}|\rho]|\rho' D_0' V_{ff}(\theta_0)^{-1} D_0 \rho = n_0]. \tag{35}$$

Hence, $E[\lim_{T\to\infty} n_\nu|\rho]$ and $E[\lim_{T\to\infty} n_{2\nu}|\rho]$ are involved in the second order Edgeworth approximation.

Lemma 2. *When $\mu = \nu = 1$ and Assumptions 1 and 2, 2* hold, the conditional expectations of the limit expressions of n_v and $n_{2\nu}$ given $\rho = (D_0' V_{ff}(\theta_0)^{-1} D_0)^{-1} D_0' V_{ff}(\theta_0)^{-1}\psi_f = (\rho_1 \ldots \rho_m)'$ read:*[2]

$$\begin{aligned} &E[\lim_{T\to\infty} n_\nu|\rho] = \\ &\quad 3\textstyle\sum_{i=1}^m \rho_i \rho' D_0' V_{ff}(\theta_0)^{-1} A_i V_{\theta f,i}(\theta_0) V_{ff}(\theta_0)^{-1} D_0 \rho + \\ &\quad 2\textstyle\sum_{i=1}^m \rho_i \sum_{j=1}^{k_f - m}\sum_{n=1}^{k_f - m}\left\{\left[(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-1}\right]_{jn}[D_{0,\perp}' A_i V_{\theta f,i}(\theta_0) D_{0,\perp}]_{nj}\right\}, \end{aligned} \tag{36}$$

[2] We note that when $\nu = 1$, $D_0 \underset{p}{\rightarrow} J_\theta(\theta_0)$.

with $D_{0,\perp} : k_f \times (k_f - m)$, $D_{0,\perp}' D_0 \equiv 0$, $D_{0,\perp}' D_{0,\perp} \equiv I_{k_f - m}$ *and* $\left[(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-1} \right]_{jn}$ *and* $[D_{0,\perp}' A_i V_{\theta f,i}(\theta_0) D_{0,\perp}]_{jn}$ *are the jn-th elements of the respective matrix; and*

$$\begin{aligned} E[\lim_{T\to\infty} n_{2\nu}|\rho] = \; & a_1 + a_2 + a_3 + a_4 + \\ & \sum_{i=1}^{m} \sum_{j=1}^{m} [a_{ij} + b_{ij} + c_{ij} + d_{ij} + e_{ij}](D_0' V_{ff}(\theta_0)^{-1} D_0)_{ij}^{-1}, \end{aligned} \tag{37}$$

with

$$\begin{aligned} a_1 = \; & \rho' D_0' V_{ff}(\theta_0)^{-1} [A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)][I_m \otimes V_{ff}(\theta_0)^{-1} D_0 \rho](D_0' V_{ff}(\theta_0)^{-1} D_0)^{-1} \\ & D_0' V_{ff}(\theta_0)^{-1} [A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)][I_m \otimes V_{ff}(\theta_0)^{-1} D_0 \rho]\rho, \\ a_2 = \; & \mathrm{tr}([\rho \otimes D_{0,\perp}][\textstyle\sum_{i=1}^{k_f - m} (D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})_i^{-1} \rho' D_0' V_{ff}(\theta_0)^{-1} A_1 V_{\theta f,1}(\theta_0)_i \cdots \\ & \textstyle\sum_{i=1}^{k_f - m} \mathrm{tr}\{(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})_i^{-1} \rho' D_0' V_{ff}(\theta_0)^{-1} A_m V_{\theta f,m}(\theta_0)_i](D_0' V_{ff}(\theta_0)^{-1} D_0)^{-1} D_0' V_{ff}(\theta_0)^{-1} \\ & [A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)]), \\ a_3 = \; & E\{\mathrm{tr}([\rho \otimes D_{0,\perp}(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-1} D_{0,\perp}' [A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)] \\ & [I_m \otimes V_{ff}(\theta_0)^{-1} D_0 \rho](D_0' V_{ff}(\theta_0)^{-1} D_0)^{-1} D_0' V_{ff}(\theta_0)^{-1} [A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)]), \\ a_4 = \; & [\mathrm{tr}(D_{0,\perp}(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-1} D_{0,\perp}' A_1 V_{\theta f,1}(\theta_0)) \ldots \mathrm{tr}(D_{0,\perp}(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-1} D_{0,\perp}' A_m \\ & V_{\theta f,m}(\theta_0))] D_0' V_{ff}(\theta_0)^{-1} D_0)^{-1} D_0' V_{ff}(\theta_0)^{-1} [A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)][I_m \otimes V_{ff}(\theta_0)^{-1} D_0 \rho]\rho, \end{aligned}$$

and

$$\begin{aligned} a_{ij} = \; & \{[\rho' D_0' V_{ff}(\theta_0)^{-1} A_i V_{\theta f,i}(\theta_0) V_{ff}(\theta_0)^{-1} D_0 \rho][\rho' D_0' V_{ff}(\theta_0)^{-1} A_j V_{\theta f,j}(\theta_0) V_{ff}(\theta_0)^{-1} D_0 \rho] \\ b_{ij} = \; & 2[\rho' D_0' V_{ff}(\theta_0)^{-1} A_i V_{\theta f,i}(\theta_0) V_{ff}(\theta_0)^{-1} D_0 \rho] \mathrm{tr}[D_{0,\perp}(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-1} D_{0,\perp}' A_j V_{\theta f,j}(\theta_0)] \\ c_{ij} = \; & 3 \textstyle\sum_{i_1=1}^{k_f - m} [(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-\frac{1}{2}\prime} D_{0,\perp}' A_i V_{\theta f,i}(\theta_0) D_{0,\perp}(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-\frac{1}{2}}]_{i_1 i_1}^2 + \\ & 2 \textstyle\sum_{i_1=1}^{k_f - m} \sum_{j_1=1, j_1 \neq i_1}^{k_f - m} [(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-\frac{1}{2}\prime} D_{0,\perp}' A_i V_{\theta f,i}(\theta_0) D_{0,\perp}(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-\frac{1}{2}}]_{i_1 i_1} \\ & [(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-\frac{1}{2}\prime} D_{0,\perp}' A_j V_{\theta f,j}(\theta_0) D_{0,\perp}(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-\frac{1}{2}}]_{j_1 j_1} + \\ & 2 \textstyle\sum_{i_1=1}^{k_f - m} \sum_{j_1=1, j_1 \neq i_1}^{k_f - m} [(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-\frac{1}{2}\prime} D_{0,\perp}' A_i V_{\theta f,i}(\theta_0) D_{0,\perp}(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-\frac{1}{2}}]_{i_1 j_1} \\ & [(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-\frac{1}{2}\prime} D_{0,\perp}' A_j V_{\theta f,j}(\theta_0) D_{0,\perp}(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-\frac{1}{2}}]_{i_1 j_1} + \\ & 2 \textstyle\sum_{i_1=1}^{k_f - m} \sum_{j_1=1, j_1 \neq i_1}^{k_f - m} [(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-\frac{1}{2}\prime} D_{0,\perp}' A_i V_{\theta f,i}(\theta_0) D_{0,\perp}(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-\frac{1}{2}}]_{i_1 j_1} \\ & [(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-\frac{1}{2}\prime} D_{0,\perp}' A_j V_{\theta f,j}(\theta_0) D_{0,\perp}(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-\frac{1}{2}}]_{j_1 i_1} \\ d_{ij} = \; & \rho' D_0' V_{ff}(\theta_0)^{-1} A_i V_{\theta f,i}(\theta_0) D_{0,\perp}(D_{0,\perp} V_{ff}(\theta_0) D_{0,\perp})^{-1} D_{0,\perp} V_{\theta f,j}(\theta_0)' A_j' V_{ff}(\theta_0)^{-1} D_0 \rho \\ e_{ij} = \; & 2\rho' D_0' V_{ff}(\theta_0)^{-1} A_i V_{\theta f,i}(\theta_0) D_{0,\perp}(D_{0,\perp} V_{ff}(\theta_0) D_{0,\perp})^{-1} D_{0,\perp} V_{\theta f,j}(\theta_0)' A_j' V_{ff}(\theta_0)^{-1} D_0 \rho. \end{aligned}$$

Proof. see the Appendix. ■

Lemma 2 states the conditional expectation of n_v and $n_{2\nu}$ given ρ. Because $\lim_{n\to\infty} n_0 = \rho' D_0' V_{ff}(\theta_0)^{-1} D_0 \rho$, we can specify ρ as $\rho = n_0^{\frac{1}{2}} h$ with $h : m \times 1$ and $h' D_0' V_{ff}(\theta_0)^{-1} D_0 h = 1$. To obtain the conditional expectation for the second order Edgeworth approximation, the law of iterated expectations (35) then implies that we construct the expectation of the conditional expectations of n_ν and $n_{2\nu}$ from Lemma 2 with respect to h.

Corollary 3. *Lemma 2 implies that the limiting expressions of the conditional expectations of* n_ν *and* $n_{2\nu}$ *given* n_0 *read*

$$E[\lim_{T\to\infty} n_\nu | n_0] = \; E_h[E[\lim_{T\to\infty} n_\nu | \rho] = 0 \tag{38}$$

and

$$E[\lim_{T\to\infty} n_{2\nu}|n_0] = E_h[E[\lim_{T\to\infty} n_{2\nu}|\rho]] =$$
$$E_h[a_2+a_3+a_4|\rho = n_0^{\frac{1}{2}}h] + \sum_{i=1}^m \sum_{j=1}^m (D_0'V_{ff}(\theta_0)^{-1}D_0)_{ij}^{-1}\left\{c_{ij} + E[b_{ij}+d_{ij}+e_{ij}|\rho = n_0^{\frac{1}{2}}h]\right\}. \tag{39}$$

Proof. Because n_0 has a $\chi^2(m)$ limiting distribution and ρ is normally distributed with mean zero, the first and third order moments of h are zero. The expectations of a_1 and a_{ij} from Lemma 2 with respect to h are therefore equal to zero. ∎

The elements of the conditional expectation of $n_{2\nu}$ given n_0 (39) are proportional to $(D_0'V_{ff}(\theta_0)^{-1}D_0)^{-1}$ which can be estimated by $(\frac{1}{T^{1+\nu}}D_T(\theta_0,Y)'V_{ff}(\theta_0)^{-1}D_T(\theta_0,Y))^{-1}$. The second order Edgeworth approximation (34) therefore contains, for example, the second order term

$$\frac{1}{T}\sum_{i=1}^m \sum_{j=1}^m \left[\frac{1}{T^2}\hat{D}_T(\theta_0,Y)'\hat{V}_{ff}(\theta_0)^{-1}\hat{D}_T(\theta_0,Y)\right]_{ij}^{-1}\hat{c}_{ij}, \tag{40}$$

which is part of "$-\frac{1}{T}s_2(s)$" in (34) and that assumes that $\nu=1$. The assumption that $\nu=1$ is a high level assumption which we can not verify. If $\nu=0$, (40) is of order zero instead of $\frac{1}{T}$. The second order Edgeworth approximation does then not remove all second order approximation errors. The second order Edgeworth approximation thus only removes second order approximation errors when $\nu=1$. We need to assume this a priori so it does not have to hold for the analyzed data.

Alongside the sensitivity of the second order Edgeworth approximation to the value of ν also the number of instruments k_f $(=kl)$ is of importance for the accuracy of the second order Edgeworth approximation. The c_{ij} elements in (39) consist of $(k_f-m)^2$ components and are thus proportional to k_f^2. When k_f^2 is large and proportional to T, the second order term of the Edgeworth approximation (40) becomes a zero-th order term instead of $\frac{1}{T}$. The second order Edgeworth approximation does then not remove the second order approximation errors.

The sensitivity to the value of ν and the number of instruments k_f shows that a second order Edgeworth approximation does not remove the second order approximation error of the finite sample distribution of $\mathrm{W}_{2s}(\theta_0)$ and $\mathrm{LM}(\theta_0)$ in all instances. This indicates that the Edgeworth approximation will not perform satisfactorily for $\mathrm{W}_{2s}(\theta_0)$ and $\mathrm{LM}(\theta_0)$ since the improvement of the distributions depends on unknown nuisance parameters. The n_ν and $n_{2\nu}$ elements are not present in the higher order expressions of $\mathrm{W}_{\mathrm{cue}}(\theta_0)$ and $\mathrm{K}(\theta_0)$. When $\nu=1$, the quality of the approximation of the finite sample distribution of these statistics by their zero-th order element is therefore less sensitive to the number of instruments. This corresponds with Brown and Newey (1998) and Newey and Smith (2004) where it is shown that the bias of the CUE is smaller than that of the 2-step GMM estimator and is much less affected by the number of instruments. Also Donald and Newey (2000) show that the bias of the CUE is smaller than that of the 2step estimator since the CUE works like a jackknife. $\mathrm{W}_{\mathrm{cue}}(\theta_0)$ and $\mathrm{K}(\theta_0)$ are both based upon the CUE and show that the results of Brown and Newey (1998), Donald and Newey (2000) and Newey and Smith (2004) extend to such statistics. These statistics thus contain a considerable part of the corrections that the second order Edgeworth approximation of the distribution of $\mathrm{W}_{2s}(\theta_0)$ and $\mathrm{LM}(\theta_0)$ applies.

Corollary 3 is not only helpful for the analysis of the Edgeworth approximation but also shows that $n_{2\nu}$ is proportional to k_f^2. When k_f and T jointly converge to infinity and k_f^2 is proportional to T, $n_{2\nu}$ therefore becomes a zero-th order term. Hence, in order to preserve the limiting distributions of $\mathrm{W}_{2\mathrm{s}}(\theta_0)$ and $\mathrm{LM}(\theta_0)$ in limiting sequences where k_f and T jointly converge to infinity, $\lim_{T\to\infty,k_f\to\infty}\frac{k_f^2}{T}=0$ has to hold. We note that the Edgeworth approximation should remove this distortion of the zero-th order behavior of $\mathrm{W}_{2\mathrm{s}}(\theta_0)$ and $\mathrm{LM}(\theta_0)$.

3.1.2 Weak/non-identification or $\nu = 0$

The higher order elements of $\mathrm{W}_{cue}(\theta_0)$ and $\mathrm{K}(\theta_0)$ in Theorem 1 are identical when $\nu = 1$. When $\nu = 0$, the GMM estimators $\hat{\theta}_{2s}$ and $\hat{\theta}_{cue}$ converge to random variables, see *e.g.* Phillips (1989) and Stock and Wright (2000). The covariance matrix estimators involved in the Wald statistics, $\mathrm{W}_{2s}(\theta_0)$ and $\mathrm{W}_{cue}(\theta_0)$, are then evaluated at a random variable and are thus inconsistent. The convergence rate κ $(=\min(\mu,\nu))$ in Theorem 1 then equals zero for these statistics and indicates the inconsistency. The covariance matrix estimators involved in $\mathrm{LM}(\theta_0)$ and $\mathrm{K}(\theta_0)$ are evaluated at θ_0 and still converge to the true covariance matrix with convergence rate μ. The convergence rate κ in Theorem 1 is therefore equal to μ for these statistics and we can obtain the limit expression of the zero-th order term of the higher order expression when $\nu = 0$. This expression is given in Corollary 4.

Corollary 4. *For weak and zero values of $J_\theta(\theta_0)$, for which $\nu = 0$, and a fixed number of instruments, Theorem 1 implies higher order properties for $W_{2s}(\theta_0)$ and $W_{cue}(\theta_0)$ under $H_0 : \theta = \theta_0$ that are characterized by:*

$$\left.\begin{array}{r}\mathrm{W}_{2s}(\theta_0)\\ \mathrm{W}_{cue}(\theta_0)\end{array}\right\} = n_0 + n_\nu + n_\kappa + n_{\nu+\kappa} + n_{2\nu} + n_{2\kappa} + n_{2\nu+\kappa} + n_{\nu+2\kappa} + n_{3\nu},\ \textit{with}\ \kappa = 0, \tag{41}$$

for $LM(\theta_0)$:

$$\mathrm{LM}(\theta_0) = n_0 + n_\nu + n_{2\nu} + n_{3\nu} + T^{-\frac{\kappa}{2}}(n_\kappa + n_{\nu+\kappa} + n_{2\nu+\kappa}) + T^{-\kappa}(n_{2\kappa} + n_{\nu+2\kappa}),\ \textit{with}\ \kappa = \mu, \tag{42}$$

and for $K(\theta_0)$:

$$\mathrm{K}(\theta_0) = n_0 + T^{-\frac{\kappa}{2}}(n_\kappa + n_{\nu+\kappa}) + T^{-\kappa}(n_{2\kappa} + n_{\nu+2\kappa}),\ \textit{with}\ \kappa = \mu, \tag{43}$$

where the different n-elements are defined in Theorem 1. Given D_0, the limiting distribution of $LM(\theta_0)$, or limiting distribution of $n_0 + n_\nu + n_{2\nu} + n_{3\nu}$, reads

$$\begin{array}{l} LM(\theta_0) \underset{d}{\rightarrow} \psi_f' V_{ff}(\theta_0)^{-1}\{D_0 + [A_1 V_{\theta f,1}(\theta_0)\cdots A_m V_{\theta f,m}(\theta_0)](I_m \otimes V_{ff}(\theta_0)^{-1}\psi_f)\}[\{D_0 + \\ [A_1 V_{\theta f,1}(\theta_0)\cdots A_m V_{\theta f,m}(\theta_0)](I_m \otimes V_{ff}(\theta_0)^{-1}\psi_f)\}' V_{ff}(\theta_0)^{-1}\{D_0 + [A_1 V_{\theta f,1}(\theta_0)\cdots A_m V_{\theta f,m}(\theta_0)] \\ (I_m \otimes V_{ff}(\theta_0)^{-1}\psi_f)\}]^{-1}\{D_0 + [A_1 V_{\theta f,1}(\theta_0)\cdots A_m V_{\theta f,m}(\theta_0)](I_m \otimes V_{ff}(\theta_0)^{-1}\psi_f)\} V_{ff}(\theta_0)^{-1}\psi_f, \end{array} \tag{44}$$

while the limiting distribution of $K(\theta_0)$ is $\chi^2(m)$.

We do not give the expressions of the limiting distributions of $\mathrm{W}_{2s}(\theta_0)$ and $\mathrm{W}_{cue}(\theta_0)$ when $\nu = 0$. These Wald statistics involve inconsistent covariance matrix estimators, since the covariance matrix estimators are evaluated at the inconsistent estimator of θ, $\hat{\theta}$. Hence, we could only give limit expressions that involve the inconsistent estimators. The limit distribution of $\mathrm{LM}(\theta_0)$

in (44) is no longer $\chi^2(m)$ and depends on nuisance parameters. The distortion of the limit distribution of $\text{LM}(\theta_0)$, compared to its $\chi^2(m)$ limit distribution when $\nu = 1$, is caused by the higher order terms of the limit distribution when $\nu = 1$.The higher order terms of $\text{K}(\theta_0)$ when $\nu = 1$ remain higher order terms when $\nu = 0$ and do therefore not distort the limit distribution. The K-statistic is thus a higher order correction of $\text{LM}(\theta_0)$ which overcomes the change of the limit distribution of $\text{LM}(\theta_0)$ when $\nu = 0$. Unlike higher order Edgeworth corrections as in (34), the K-statistic does not involve conditional expectations of random variables.

3.2 Number of instruments that goes to infinity

When the number of instruments is proportional to the number of observations, the higher order expressions from Theorem 1 are invalid. We therefore construct higher order expressions when both the number of observations and the number of instruments jointly converge to infinity as in *e.g.* Bekker (1994). In order to do so, we make an assumption about the convergence behavior of the number of instruments k relative to that of the number of observations T.

Assumption 3. *The joint convergence of the number of instruments k and the number of observations T is such that*

$$\lim_{k,T\to\infty} \tfrac{k}{T^{\alpha}} = c, \tag{45}$$

with c a fixed finite constant.

When we construct the higher order expressions with a number of instruments that converges to infinity, we maintain the property of Assumption 1 that

$$\tfrac{1}{T^{1+\nu}} D_T(\theta_0, Y)' V_{ff}(\theta_0)^{-1} D_T(\theta_0, Y) \quad = D_0' V_{ff}(\theta_0)^{-1} D_0, \tag{46}$$

where $D_0' V_{ff}(\theta_0)^{-1} D_0$ is a finite valued random variable. It implies that $\nu \geq \alpha$ and enables us to determine the convergence rates of the different elements involved in the statistics by means of a sequential convergence scheme in which we first let T converge to infinity and afterwards k. Given a fixed value of k, we have shown in Theorem 1 that all elements converge appropriately when T goes to infinity. Lemma 6 of Phillips and Moon (1999) therefore applies and we can let T and k converge to infinity sequentially, so first T and then k.

Bekker (1994) constructs the limit distribution of the CUE in the linear instrumental variables regression model under a limit sequence where the number of instruments is proportional to the number of observations, so $\alpha = 1$, $\nu = 1$ and $J_\theta(\theta_0)' V_{ff}(\theta_0)^{-1} J_\theta(\theta_0)$ goes to a constant when k and T converge to infinity.

Theorem 2 states the higher order expressions of $\text{W}_{2s}(\theta_0)$, $\text{W}_{cue}(\theta_0)$, $\text{LM}(\theta_0)$ and $\text{K}(\theta_0)$ when the number of instruments gets large according to Assumption 3. The proof of Theorem 2 also verifies the validity of constructing the limits in a sequential manner.

Theorem 2. *When the number of instruments k converges to infinity according to Assumption 3 with $\nu \geq \alpha$, Assumptions 1 and 2 imply higher order properties of $W_{2s}(\theta_0)$, $W_{cue}(\theta_0)$, $LM(\theta_0)$*

and $K(\theta_0)$ *under* $H_0 : \theta = \theta_0$ *that are characterized by:*

$$\left.\begin{array}{r} \mathrm{W}_{2s}(\theta_0) \\ \mathrm{W}_{cue}(\theta_0) \\ \mathrm{LM}(\theta_0) \\ \mathrm{K}(\theta_0) \end{array}\right\} = \left\{\begin{array}{ll} n_0 + & : \text{zero-th order} \\ T^{-\frac{\nu-2\alpha}{2}} n_{\nu-2\alpha} + T^{-(\nu-\alpha)} n_{2(\nu-\alpha)} + & : \begin{array}{l} D_T(\theta_0, Y)+ \\ \text{instruments} \end{array} \\ T^{-(\nu-2\alpha)} n_{2(\nu-2\alpha)} + T^{-(\nu-\alpha)} n_{2(\nu-\alpha)} + & \\ T^{-\nu} n_\nu + & : D_T(\theta_0, Y) \\ T^{-\frac{\kappa}{2}} n_\kappa + T^{-\kappa} n_{2\kappa} + & : \hat{V}(\theta_0) \\ T^{-\frac{1}{2}(\nu+\kappa-2\alpha)} n_{\nu+\kappa-2\alpha} + T^{-\frac{1}{2}(\nu+\kappa)} n_{\nu+\kappa} + & \\ T^{-\frac{1}{2}(\nu+2(\kappa-\alpha))} n_{\nu+2(\kappa-\alpha)} + & \\ T^{-\frac{1}{2}(2\nu+\kappa-2\alpha)} n_{2\nu+\kappa-2\alpha} + T^{-(\nu+\kappa-\alpha)} n_{2(\nu+\kappa-\alpha)} + & : \text{mixed} \\ T^{-\frac{1}{2}\kappa+2(\nu-2\alpha)} n_{\kappa+2(\nu-2\alpha)} + T^{-(\nu+\kappa-2\alpha)} n_{2(\nu+\kappa-2\alpha))} & \end{array}\right. \quad (47)$$

where:

1. *for* $W_{2s}(\theta_0)$*:* $\kappa = \min(\mu, \nu)$*, and*

$$\begin{array}{rcl} n_0 & = & s_0' G_0^{-1} s_0 \\ \begin{array}{c} D_T(\theta_0, Y)+ \\ \text{instruments} \end{array} & : & \left\{\begin{array}{rcl} n_{\nu-2\alpha} & = & s_{\nu-2\alpha,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{\nu-2\alpha,1} \\ n_{2(\nu-2\alpha)} & = & s_{\nu-2\alpha,1}' G_0^{-1} s_{\nu-2\alpha,1} \\ n_{2(\nu-\alpha)} & = & s_{\nu-2\alpha,1}' Q_1 s_0 + s_0' Q_1 s_{\nu-2\alpha,1} \end{array}\right. \\ D_T(\theta_0, Y) & : & \left\{\ n_\nu = s_0' Q_1 s_0 \right. \\ \hat{V}(\theta_0) & : & \left\{\begin{array}{rcl} n_\kappa & = & s_{1\kappa,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{1\kappa,1} \\ n_{2\kappa} & = & s_{1\kappa,1}' G_0^{-1} s_{1\kappa,1} \end{array}\right. \\ \text{mixed} & : & \left\{\begin{array}{rcl} n_{\nu+\kappa-2\alpha} & = & (s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,2})' G_0^{-1} s_0 + s_{\nu-2\alpha,1}' G_0^{-1} s_{1\kappa,1} + \\ & & s_0' G_0^{-1} (s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,2}) + s_{1\kappa,1}' G_0^{-1} s_{\nu-2\alpha,1} \\ n_{\nu+\kappa} & = & s_{1\kappa,1}' Q_1 s_0 + s_0' Q_1 s_{1\kappa,1} \\ n_{\nu+2(\kappa-\alpha)} & = & (s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,2})' G_0^{-1} s_{1\kappa,1} + s_{\nu+2\kappa-2\alpha,1}' G_0^{-1} s_0 + \\ & & s_{1\kappa,1}' G_0^{-1} (s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,2}) + s_0' G_0^{-1} s_{\nu+2\kappa-2\alpha,1} \\ n_{2\nu+\kappa-2\alpha} & = & (s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,2})' Q_1 s_0 + s_0' Q_1 (s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,2}) \\ n_{2(\nu+\kappa-\alpha)} & = & s_{\nu+2(\kappa-\alpha),1}' Q_1 s_0 + (s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,2})' Q_1 s_{1\kappa,1} + \\ & & s_{1\kappa,1}' Q_1 (s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,2}) + s_0' Q_1 s_{\nu+2(\kappa-\alpha),1} \\ n_{\kappa+2(\nu-2\alpha)} & = & s_{\nu-2\alpha,1}' Q_1 s_{\nu-2\alpha,1} \\ n_{2(\nu+\kappa-2\alpha)} & = & (s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,2})' G_0^{-1} (s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,2}) \end{array}\right. \end{array} \quad (48)$$

2. *for* $W_{cue}(\theta_0)$*:* $\kappa = \min(\mu, \nu)$, $n_{\nu-2\alpha} = n_{2(\nu-2\alpha)} = n_v = n_{\nu+\kappa} = n_{2(\nu-\alpha)} = n_{2\nu+\kappa-2\alpha} = n_{2(\nu+\kappa-\alpha)} = n_{\kappa+2(\nu-2\alpha)} = 0,$

$$\begin{array}{rcl} n_0 & = & s_0' G_0^{-1} s_0 \\ \hat{V}(\theta_0) & : & \left\{\begin{array}{rcl} n_\kappa & = & s_0' Q_1 s_0 + s_{1\kappa,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{1\kappa,1} \\ n_{2\kappa} & = & s_{1\kappa,1}' G_0^{-1} s_{1\kappa,1} + s_{1\kappa,1}' Q_1 s_0 + s_0' Q_1 s_{1\kappa,1} \\ n_{3\kappa} & = & s_{1\kappa,1}' Q_1 s_{1\kappa,1} \end{array}\right. \\ \text{mixed} & : & \left\{\begin{array}{rcl} n_{\nu+\kappa-2\alpha} & = & s_{\nu+\kappa-2\alpha,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{\nu+\kappa-2\alpha,1} \\ n_{\nu+2(\kappa-\alpha)} & = & s_{\nu+\kappa-2\alpha,1}' G_0^{-1} s_{1\kappa,1} + s_{1\kappa,1}' G_0^{-1} s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,1}' Q_1 s_0 + \\ & & s_0' Q_1 s_{\nu+\kappa-2\alpha,1} + s_{\nu+2\kappa-2\alpha,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{\nu+2\kappa-2\alpha,1} \\ n_{2\nu+\kappa-2\alpha} & = & s_{\nu+\kappa-2\alpha,1}' G_0^{-1} s_{\nu+\kappa-2\alpha,1}. \end{array}\right. \end{array} \quad (49)$$

3. *for LM(θ_0) the elements are identical to those for $W_{2s}(\theta_0)$ in (48) but with $\kappa = \mu$.*

4. *for K(θ_0) the elements are identical to those for $W_{cue}(\theta_0)$ in (49) but with $\kappa = \mu$*

and for all statistics

$$\begin{array}{rl} s_0 = & m_{0,f}' V_{ff}(\theta_0)^{-1} D_0 \\ s_{\nu-2\alpha,1} = & \frac{1}{k} m_{0,f}' V_{ff}(\theta_0)^{-1} \{ \frac{1}{\sqrt{T}} [p_T(\hat{\theta}, Y) - \hat{D}_T(\hat{\theta}, Y)] \} \\ = & \frac{1}{k} m_{0,f}' V_{ff}(\theta_0)^{-1} [A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)][I_m \otimes V_{ff}(\theta_0)^{-1} m_{0,f}] \\ s_{1\kappa,1} = & T^{\frac{1}{2}\kappa} m_{0,f}' [\hat{V}_{ff}(\theta_0)^{-1} - V_{ff}(\theta_0)^{-1}] D_0 \\ s_{\nu+\kappa-2\alpha,1} = & \frac{T^{\frac{1}{2}\kappa}}{k} m_{0,f}' V_{ff}(\theta_0)^{-1} \{ [A_1 \hat{V}_{\theta f,1}(\theta_0) \cdots A_m \hat{V}_{\theta f,m}(\theta_0)][I_m \otimes \hat{V}_{ff}(\theta_0)^{-1}] \\ & -[A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)][I_m \otimes V_{ff}(\theta_0)^{-1}] \} [I_m \otimes m_{0,f}] \\ s_{\nu+\kappa-2\alpha,2} = & \frac{T^{\frac{1}{2}\kappa}}{k} m_{0,f}' [\hat{V}_{ff}(\theta_0)^{-1} - V_{ff}(\theta_0)^{-1}][A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)] \\ & [I_m \otimes V_{ff}(\theta_0)^{-1} m_{0,f}] \\ s_{\nu+2(\kappa-\alpha),1} = & \frac{T^{\frac{1}{2}(2\kappa-1)}}{k} m_{0,f}' [\hat{V}_{ff}(\theta_0)^{-1} - V_{ff}(\theta_0)^{-1}][\hat{D}_T(\theta_0, Y) - D_T(\theta_0, Y)], \end{array} \tag{50}$$

with $G_0 = D_0' V_{ff}(\theta_0)^{-1} D_0$ and the remaining expressions of the G and Q matrices are given in the Appendix.

Proof. see the Appendix. ■

When $\alpha = 0$, the higher order expressions in Theorem 2 are identical to those in Theorem 1 that were constructed for a fixed number of instruments. An important difference with the elements of the higher order expressions in Theorem 1 results from the convergence of

$$s_{\nu-2\alpha,1} = \frac{1}{k} m_{0,f}' V_{ff}(\theta_0)^{-1} [A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)][I_m \otimes V_{ff}(\theta_0)^{-1} m_{0,f}]. \tag{51}$$

When k and T converge to infinity,

$$s_{\nu-2\alpha,1} \underset{p}{\rightarrow} \omega(\theta_0), \tag{52}$$

where $\omega(\theta_0) = (\omega_1(\theta_0) \ldots \omega_m(\theta_0))$ and

$$\omega_i(\theta_0) = \lim_{k\rightarrow\infty} \frac{1}{k} \text{tr}(V_{ff}(\theta_0)^{-1} A_i V_{\theta f,i}(\theta_0)). \tag{53}$$

The convergence of $s_{\nu-2\alpha,1}$ towards a constant implies that ν needs to exceed 2α for the $\chi^2(m)$ limiting distribution to remain valid for $\text{W}_{2s}(\theta_0)$ and $\text{LM}(\theta_0)$. Otherwise, $\text{W}_{2s}(\theta_0)$ and $\text{LM}(\theta_0)$ converge to $\lim_{k,T\rightarrow\infty} n_{2(\nu-2\alpha)} = \omega(\theta_0)' G_0^{-1} \omega(\theta_0)$ because $2(\nu-2\alpha) < (\nu-2\alpha)$ when $\nu < 2\alpha$. This sensitivity to the number of instruments of $\text{W}_{2s}(\theta_0)$ and $\text{LM}(\theta_0)$ is also indicated by Corollary 3 where the conditional expectation of higher order elements of $\text{W}_{2s}(\theta_0)$ and $\text{LM}(\theta_0)$ depends on the number of instruments. Theorem 2 further emphasizes this sensitivity to the number of instruments of $\text{W}_{2s}(\theta_0)$ and $\text{LM}(\theta_0)$. Even for values of ν that correspond with a well-identified θ_0, $\nu \geq 1$, the limiting distributions of $\text{W}_{2s}(\theta_0)$ and $\text{LM}(\theta_0)$ can be affected by the number of instruments.

Theorem 2 assumes that $\nu \geq \alpha$. The number of instruments can therefore affect the limiting distribution of $\text{W}_{cue}(\theta_0)$ and $\text{K}(\theta_0)$ when $\alpha \geq \mu$. Corollary 5 states these distortions for a stylized setting with $\nu = \alpha = \mu = 1$ and which corresponds with Bekker (1994).

Corollary 5. *When $\nu = \alpha = \mu = 1$, the higher order expressions of $W_{cue}(\theta_0)$ and $K(\theta_0)$ that result from Theorem 2 read*

$$\left.\begin{array}{r} \mathrm{W}_{cue}(\theta_0) \\ \mathrm{K}(\theta_0) \end{array}\right\} = \quad n_0 + n_{\nu+\kappa-2\alpha} + n_{2(\nu+\kappa-2\alpha)} + T^{-\frac{1}{2}}(n_\kappa + n_{\nu+2(\kappa-\alpha)}) + T^{-1}n_{2\kappa} + o_p(T^{-1}) \quad (54)$$

where the expressions of the n-elements are stated in Theorem 2.

Corollary 5 shows that additional zero-th order elements, *i.e.* $n_{\nu+\kappa-2\alpha} + n_{2(\nu+\kappa-2\alpha)}$, appear when $\nu = \alpha = \mu = 1$. Both $n_{\nu+\kappa-2\alpha}$ and $n_{2(\nu+\kappa-2\alpha)}$ consists of, alongside s_0, $s_{\nu+\kappa-2\alpha,1}$. We therefore state the limiting distribution of $s_{\nu+\kappa-2\alpha,1}$ in Lemma 3.

Lemma 3. *When k and T converge to infinity, and Assumption 1, 2, 2* and 3 hold, the convergence of $s_{\nu+\kappa-2\alpha,1}$ defined in Theorem 3 is characterized by*

$$s_{\nu+\kappa-2\alpha,1} \underset{d}{\rightarrow} \lambda, \quad (55)$$

where $\lambda \sim N(0, \Sigma(\theta_0))$ and independent of ψ_f with $\Sigma(\theta_0) = \{\sigma_{ij}(\theta_0)\}_{i,j=1,\ldots,m}$ and

$$\begin{array}{rl} \sigma_{ij}(\theta_0) = & \lim_{k\rightarrow\infty}\left[\frac{1}{\sqrt{k}}\psi_f' V_{ff}(\theta_0)^{-1} \otimes \frac{1}{\sqrt{k}}\psi_f' V_{ff}(\theta_0)^{-1}A_i\right]' \bar{W}_{ij}(\theta_0) \\ & \left[\frac{1}{\sqrt{k}}\psi_f' V_{ff}(\theta_0)^{-1} \otimes \frac{1}{\sqrt{k}}\psi_f' V_{ff}(\theta_0)^{-1}A_j\right] \end{array} \quad (56)$$

with

$$\bar{W}_{ij}(\theta_0) = \quad \lim_{T\rightarrow\infty} E[\mathrm{vec}(U_{\theta f,i} - V_{\theta f,i}(\theta_0)V_{ff}(\theta_0)^{-1}U_{ff})\mathrm{vec}(U_{\theta f,j} - V_{\theta f,j}(\theta_0)V_{ff}(\theta_0)^{-1}U_{ff})'], \quad (57)$$

which expression results from Assumption 2.

Proof. see the Appendix. ■

Lemma 3 indicates that the zero-th order term from Corollary 5 does not have a $\chi^2(m)$ limiting distribution when $\nu = \mu = \alpha = 1$. We can account for the distortion of the $\chi^2(m)$ limiting distribution by including an estimate of $\Sigma(\theta_0)$ in the covariance matrix estimators involved in $\mathrm{W}_{cue}(\theta_0)$ and $\mathrm{K}(\theta_0)$. Bekker (1994) proposes such a covariance matrix estimator for $\mathrm{W}_{cue}(\theta_0)$ in the linear instrumental variables regression model for a limit sequence with $\nu = \mu = \alpha = 1$.

The elements $\sigma_{ij}(\theta_0)$ (56), that we need to incorporate in $\mathrm{W}_{cue}(\theta_0)$ and $\mathrm{K}(\theta_0)$ to preserve their $\chi^2(m)$ limit distributions in a limit sequence with $\nu = \mu = \alpha = 1$, are of order $T^{2\alpha+\mu}$ ($= k^2T^\mu$). In case k is fixed, so $\alpha = 0$, $\nu = 0$ and $\mu = 1$, this order equals T and is identical to the convergence rate of $D_T(\theta_0, Y)'V_{ff}(\theta_0)^{-1}D_T(\theta_0, Y)$. The robustness of the limiting distribution of $\mathrm{W}_{cue}(\theta_0)$ and $\mathrm{K}(\theta_0)$ to limit sequences where $\nu = \mu = \alpha = 1$ comes therefore at the price of non-robustness of the limiting distribution of $\mathrm{W}_{cue}(\theta_0)$ and $\mathrm{K}(\theta_0)$ to limit sequences where $\nu = \alpha = 0$, see also Bekker and Kleibergen (2003). The limiting distribution of $\mathrm{W}_{cue}(\theta_0)$ is non-robust to such limit sequences but the limiting distribution of $\mathrm{K}(\theta_0)$ is robust to these limit sequences. Hence, robustifying $\mathrm{K}(\theta_0)$ to allow for $\nu = \mu = \alpha = 1$ means losing the robustness to $\nu = \alpha = 0$. Without adapting the covariance matrix estimator, the limiting distribution of $\mathrm{K}(\theta_0)$ remains $\chi^2(m)$ when $\mu > \alpha$.

4 Higher Order Properties of Statistics that test H_e : $E(f_t(\theta)) = 0$.

Alongside tests of hypothezes specified on the parameter θ, like $\mathrm{H}_0 : \theta = \theta_0$, it is customary to test whether Assumption 1 holds so the model is not misspecified at θ_0: $\mathrm{H}_e : E(f_t(\theta_0)) = 0$ or to conduct a joint test of H_0 and H_e. For the latter kind of joint hypothezes, we can use the objective function evaluated at θ_0, which is Stock and Wright's (2000) S-statistic:

$$\mathrm{S}(\theta_0) = \tfrac{1}{T} f_T(\theta_0, Y)' \hat{V}_{ff}(\theta_0)^{-1} f_T(\theta_0, Y). \tag{58}$$

Under H_0 and H_e, $\mathrm{S}(\theta_0)$ has a $\chi^2(k_f)$ limit distribution regardless of the value of $J_\theta(\theta_0)$.

To obtain the elements of $\mathrm{S}(\theta_0)$ that test H_e, we can use a J-statistic, see *e.g.* Hansen (1982), that results from substracting one of the statistics $\mathrm{W}_{2s}(\theta_0)$, $\mathrm{W}_{cue}(\theta_0)$, $\mathrm{LM}(\theta_0)$ or $\mathrm{K}(\theta_0)$ from $\mathrm{S}(\theta_0)$:

$$\begin{array}{rl} \mathrm{J}_{2s}(\theta_0) = & \mathrm{S}(\theta_0) - \mathrm{W}_{2s}(\theta_0) \\ \mathrm{J}_{cue}(\theta_0) = & \mathrm{S}(\theta_0) - \mathrm{W}_{cue}(\theta_0) \\ \mathrm{J}_{LM}(\theta_0) = & \mathrm{S}(\theta_0) - \mathrm{LM}(\theta_0) \\ \mathrm{J}_{K}(\theta_0) = & \mathrm{S}(\theta_0) - \mathrm{K}(\theta_0). \end{array} \tag{59}$$

Under H_0 and H_e, all J-statistics in (59) have $\chi^2(k_f - m)$ limiting distributions when $J_\theta(\theta_0)$ has a fixed full rank value. Only $\mathrm{J}_K(\theta_0)$ has a $\chi^2(k_f - m)$ limiting distribution for any value of $J_\theta(\theta_0)$, see Kleibergen (2003,2002b). The J-statistics that are commonly used, *i.e.* $\mathrm{J}_{2s}(\hat{\theta}_{2s})$ and $\mathrm{J}_{cue}(\hat{\theta}_{cue})$, only have a $\chi^2(k_f - m)$ limiting distribution under H_e when $J_\theta(\theta_0)$ has a fixed full rank value. Theorem 3 states the higher order expressions of the S and J-statistics for a fixed number of instruments. Because the S and J-statistics have limiting distributions that depend on the number of instruments, we do not construct their higher order expressions in a limit sequence where the number of instruments and the number of observations jointly converge to infinity.

Theorem 3. *Assumptions 1, 2 and Theorem 1 imply higher order expressions for the S-statistic (58) and J-statistics (59) that read:*

$$S(\theta_0) = \quad n_0 + n_{0,\perp} + T^{-\frac{\mu}{2}} w_\mu + o_p(T^{-\frac{\mu}{2}}), \tag{60}$$

with $w_\mu = T^{\frac{\mu}{2}} m'_{0.f}[\hat{V}_{ff}(\theta_0)^{-1} - V_{ff}(\theta_0)^{-1}]m_{0.f}$, $n_0 = m'_{0.f} V_{ff}(\theta_0)^{-1} D_0 (D'_0 V_{ff}(\theta_0)^{-1} D_0)^{-1} D'_0 V_{ff}(\theta_0)^{-1} m_{0.f}$ *and* $n_{0,\perp} = m'_{0.f} D_{0,\perp}(D'_{0,\perp} V_{ff}(\theta_0) D_{0,\perp})^{-1} D'_{0,\perp} m_{0.f}$, *where* $D_{0,\perp} : k_f \times (k_f - m)$, $D'_{0,\perp} D_0 \equiv 0$, $D'_{0,\perp} D_{0,\perp} \equiv I_{k_f - m}$; *and*

$$\left. \begin{array}{r} \mathrm{J}_{2s}(\theta_0) \\ \mathrm{J}_{cue}(\theta_0) \\ \mathrm{J}_{LM}(\theta_0) \\ \mathrm{J}_{\mathrm{K}}(\theta_0) \end{array} \right\} = n_{0,\perp} + T^{-\frac{\mu}{2}} w_\mu - \left\{ \begin{array}{l} T^{-\frac{\nu}{2}} n_\nu + T^{-\frac{\kappa}{2}} n_\kappa + T^{-\frac{\nu+\kappa}{2}} n_{\nu+\kappa} + T^{-\nu} n_{2\nu} + T^{-\kappa} n_{2\kappa} + \\ T^{-\frac{1}{2}(2\nu+\kappa)} n_{2\nu+\kappa} + T^{-\frac{1}{2}(\nu+2\kappa)} n_{\nu+2\kappa} + T^{-\frac{3}{2}\nu} n_{3\nu} + o_p(T^{-\frac{3}{2}\nu}), \end{array} \right. \tag{61}$$

where the specification of the different n-elements for a specific statistic is given in Theorem 1.

Proof. see the Appendix. ∎

Theorem 3 shows that the J-statistics (59) possess similar higher order properties as the statistics whose properties are stated in Theorem 1. Since

$$n_{0,\perp} \underset{d}{\rightarrow} \chi^2(k_f - m),$$

all J-statistics converge to a $\chi^2(k_f - m)$ distributed random variable when $\nu = 1$ but only $\mathrm{J_K}(\theta_0)$ converges to such a random variable when $\nu = 0$. Identical to the statistics in Theorem 1, the distortion of the limit distribution when $\nu = 0$ results from elements that are of higher order when $\nu = 1$. These elements are not present amongst the higher order elements of $\mathrm{J}_K(\theta_0)$ and can therefore not alter the limit distribution of $\mathrm{J}_K(\theta_0)$ when ν becomes equal to zero. We conclude from Theorems 1 and 3 that the statistics whose higher order properties do not depend on ν, *i.e.* $\mathrm{S}(\theta_0)$, $\mathrm{K}(\theta_0)$ and $\mathrm{J}_K(\theta_0)$, are also optimal from a higher order perspective since they posses less and "smaller", in a bias or variance sense, higher order elements.

The higher order properties of the commonly used J-statistics, $\mathrm{J}_{2s}(\hat{\theta}_{2s})$ and $\mathrm{J}_{cue}(\hat{\theta}_{cue})$, are similar to those of $\mathrm{J}_{2s}(\theta_0)$ and $\mathrm{J}_{cue}(\theta_0)$ in Theorem 3. A $\chi^2(k_f - m)$ limiting distribution is therefore only valid for these statistics when $\nu = 1$ and thus for full rank values of $J_\theta(\theta_0)$. Because $\mathrm{J}_{2s}(\hat{\theta}_{2s})$ results from $\mathrm{W}_{2s}(\theta_0)$ that can be severly biased when the number of instruments and/or the correlation is large, we also for other reasons have to be careful when using $\mathrm{J}_{2s}(\hat{\theta}_{2s})$.

Theorems 1 and 3 show that the limiting distributions of $\mathrm{K}(\theta_0)$ and $\mathrm{J}_K(\theta_0)$ are robust to the value of ν. Since $\mathrm{K}(\theta_0)$ is a score or Lagrange multiplier statistic, it suffers from a spurious power decline around values of θ where the objective function is maximal or has an inflexion point. The J-statistic $\mathrm{J}_K(\theta_0)$ has discriminatory power at these values of θ and is since its limiting distribution is independently distributed from $\mathrm{K}(\theta_0)$ ideally suited to be combined with $\mathrm{K}(\theta_0)$, see Kleibergen (2003,2002b). These statistics can be combined in a unconditional or in a conditional manner. A unconditional manner implies that we use fixed significance levels for $\mathrm{K}(\theta_0)$ and $\mathrm{J}_K(\theta_0)$, α_K and α_{J_K}, that add up to the significance level α by which we want to test, $\alpha = \alpha_K + \alpha_{J_K} - \alpha_K\alpha_{J_K} \approx \alpha_K + \alpha_{J_K}$. A conditional manner implies that we use an additional independently distributed statistic to combine $\mathrm{K}(\theta_0)$ and $\mathrm{J}_K(\theta_0)$. The conditional likelihood ratio statistic of Moreira (2003) in the linear instrumental variables regression model with $m = 1$ operates in such manner. Its conditional limiting distribution is the sum of the limiting distributions of $\mathrm{K}(\theta_0)$ and a weighted value of $\mathrm{J_K}(\theta_0)$. It uses $D_T(\theta_0, Y)'V_{ff}(\theta)^{-1}D_T(\theta_0, Y)$ as the independently distributed conditioning statistic. When $D_T(\theta_0, Y)'V_{ff}(\theta)^{-1}D_T(\theta_0, Y)$ is large, the conditional limiting distribution is identical to that of $\mathrm{K}(\theta_0)$ while it resembles $\mathrm{K}(\theta_0)+\mathrm{J}_K(\theta_0)$ $(=\mathrm{S}(\theta_0))$ when $D_T(\theta_0, Y)'V_{ff}(\theta)^{-1}D_T(\theta_0, Y)$ is small. Because we can only approximate Moreira's conditional likelihood ratio statistic in GMM, see Kleibergen (2003,2002b), we refrain from constructing its higher order properties.

5 Bootstrapping robust statistics

Theorems 1 and 3 show that the zero-th order elements of several GMM-statistics depend on the value of ν. For these statistics we can not use the bootstrap to approximate the finite sample distribution. The zero-th order elements of $\mathrm{K}(\theta_0)$, $\mathrm{J_K}(\theta_0)$ and $\mathrm{S}(\theta_0)$ do not depend on ν. We construct the Edgeworth approximations of the finite sample distribution of these statistics to determine if we can improve the approximation of the finite sample distribution by using the bootstrap, see *e.g.* Horowitz (2001). For reasons of brevity, we only discuss the case of independent moments which we reflect in Assumption 4. The case of dependent moment would make the bootstrap algorithms more involved and imply that we should use parametric or block-bootstrap procedures, see *e.g.* Horowitz (2001).

Assumption 4. *The moments* $(f_t(\theta_0)'\ q_t(\theta_0)')$ *and* $(f_j(\theta_0)'\ q_j(\theta_0)')$ *are independent for* $t \neq j$.

5.1 Edgeworth Approximations of finite sample distributions of robust statistics and their bootstraps

Assumption 1 implies that $E(f_t(\theta_0)) = 0$. Since Assumption 1 has to hold for the empirical distribution that we use to obtain the bootstrap distributions of the statistics, we recenter the realizations of $f_t(\theta_0)$ such that $\hat{E}(\bar{f}_t(\theta_0)) = 0$, with $\bar{f}_t(\theta_0) = f_t(\theta_0) - \frac{1}{T}\sum_{t=1}^T f_t(\theta_0)$ and where $\hat{E}$ indicates that the expectation is taken with respect to the empirical distribution, see *e.g.* Hall and Horowitz (1996). The empirical distribution of $(f_t(\theta_0)'q_t(\theta_0)')')$ then becomes: $\hat{F}(x) = \frac{1}{T}\sum_{t=1}^T I\{(\bar{f}_t(\theta_0)'\ q_t(\theta_0)')' \leq x\}$ with $I(.)$ the indicator function. Because of Assumption 4, the Glivenko-Cantelli Theorem implies that the empirical distribution of $((f_t(\theta_0)'\ q_t(\theta_0)')'$ converges to the true distribution of $((f_t(\theta_0)'\ q_t(\theta_0)')'$ when T goes to infinity.

Corollary 6. For random drawings with replacement $(\tilde{f}_t(\theta_0)'\ \tilde{q}_t(\theta_0))'$ from the empirical distribution $\hat{F}(x) = \frac{1}{T}\sum_{t=1}^T I\{(\bar{f}_t(\theta_0)'\ q_t(\theta_0)')' \leq x\}$,it holds that:

1. $\hat{E}[\tilde{f}_t(\theta_0)] = 0$.

2. $\hat{E}[\binom{\tilde{f}_t(\theta_0)}{\tilde{q}_t(\theta_0)}\binom{\tilde{f}_t(\theta_0)}{\tilde{q}_t(\theta_0)}'] = \hat{V}(\theta_0)$ so $\hat{E}[\tilde{f}_t(\theta_0)\tilde{f}_t(\theta_0)'] = \hat{V}_{ff}(\theta_0)$, with $\hat{V}(\theta_0)$ the covariance matrix estimator that results from $\{(f_t(\theta_0)'q_t(\theta_0)')',\ t = 1, \ldots, T\}$.

3. The higher order behavior for the sum of $(\tilde{f}_t(\theta_0)'\ \tilde{q}_t(\theta_0)')'$ is characterized by:

$$\frac{1}{\sqrt{T}}\sum_{t=1}^T \begin{pmatrix} \tilde{f}_t(\theta_0) \\ \tilde{q}_t(\theta_0) - \hat{E}(q_t(\theta_0)) \end{pmatrix} = \tilde{m}_0 + O_p(\tfrac{1}{\sqrt{T}}), \tag{62}$$

   where $\hat{E}(q_t(\theta_0)) = \frac{1}{T}\sum_{t=1}^T q_t(\theta_0)$ is such that the i-th column of $\hat{D}_T(\theta_0, Y)$ equals $A_i[\hat{E}(q_t(\theta_0)) - \hat{V}_{\theta f,i}(\theta_0)^{-1}\hat{V}_{ff}(\theta_0)^{-1}f_T(\theta_0, Y)]$, in case A_i is not a zero matrix, and

$$\tilde{m}_0 \underset{d}{\rightarrow} \begin{pmatrix} \phi_f \\ \phi_\theta \end{pmatrix} \quad \text{with} \quad \begin{pmatrix} \phi_f \\ \phi_\theta \end{pmatrix} \sim N(0, \hat{V}(\theta_0)). \tag{63}$$

Proof. Results directly from the independence of the $(f_t(\theta_0)'q_t(\theta_0)')')$'s and their finite variance. ■

Corollary 6 allows us to construct higher order expressions for the bootstrapped K, J_K and S-statistics. The higher order expressions can be used to obtain the Edgeworth approximations of the finite sample distributions of the (bootstrapped) K, J_K and S-statistics. Theorem 4 states the Edgeworth approximation of the standard and bootstrapped K-statistic. This Edgeworth approximation can be used as well to obtain the Edgeworth approximation of the finite sample distributions of the J_K and S statistics.

Theorem 4. *The 2κ-th order Edgeworth approximation of the distribution of $K(\theta_0)$ and bootstrapped $K(\theta_0)$, $\tilde{K}(\theta_0)$, read*

$$\begin{array}{l}\Pr[\mathrm{K}(\theta_0) \leq s] = F_{s_0}(s) - T^{-\frac{1}{2}\kappa} f_{s_0}(s)[n_\kappa(s) - T^{-\frac{1}{2}\kappa} n_{2\kappa}(s) - O_p(T^{-\kappa})], \\ \Pr[\tilde{\mathrm{K}}(\theta_0) \leq s] = F_{s_0}(s) - T^{-\frac{1}{2}\kappa} f_{s_0}(s)[\tilde{n}_\kappa(s) - T^{-\frac{1}{2}\kappa} \tilde{n}_{2\kappa}(s) - O_p(T^{-\kappa})],\end{array} \tag{64}$$

where $f_{s_0}(s)$, $F_{s_0}(s)$ are the density, distribution function of the $\chi^2(m)$ distribution and $n_\kappa(s)$, $n_{2\kappa}(s)$, $\tilde{n}_\kappa(s)$ and $\tilde{n}_{2\kappa}(s)$ are defined in Appendix F..

Proof. see Appendix F. ■

The Edgeworth approximations in Theorem 4 show that the difference between the limiting distribution and the finite sample distribution is of order $O(T^{-\frac{1}{2}\kappa})$. When Assumption 2* holds, which implies independence between the limiting distributions of $\hat{V}(\theta_0)$ and $(f_t(\theta_0)'\ q_t(\theta_0)')'$), $n_\kappa(s)$ and $\tilde{n}_\kappa(s)$ are equal to zero and the limiting distribution is accurate up to order $O(T^{-\kappa})$ as an approximation of the finite sample distribution. To show the improved accuracy that results from the bootstrap, we substract $\Pr[\tilde{\mathrm{K}}(\theta_0) \leq s]$ from $\Pr[\mathrm{K}(\theta_0) \leq s]$, see *e.g.* Horowitz (2001):

$$\Pr[\mathrm{K}(\theta_0) \leq s] - \Pr[\tilde{\mathrm{K}}(\theta_0) \leq s] = T^{-\frac{1}{2}\kappa} f_{s_0}(s)[n_\kappa(s) - \tilde{n}_\kappa(s) - T^{-\frac{1}{2}\kappa}(n_{2\kappa}(s) - \tilde{n}_{2\kappa}(s)) - O_p(T^{-\kappa})]. \tag{65}$$

The Edgeworth approximations of $\mathrm{K}(\theta_0)$ and $\tilde{\mathrm{K}}(\theta_0)$ are with respect to the unconditional and empirical distribution of $(f_t(\theta_0)'\ q_t(\theta_0)')$. Since the empirical distribution converges to the unconditional distribution, all higher order elements of the Edgeworth approximation of the distribution of $\tilde{\mathrm{K}}(\theta_0)$ converge to their respective higher order element in the Edgeworth approximation of the distribution of $\mathrm{K}(\theta_0)$. Hence, $\tilde{n}_{2\kappa}(s) = n_{2\kappa}(s) + O_p(T^{-\frac{1}{2}\kappa})$ and $\tilde{n}_\kappa(s) = n_\kappa(s) + O_p(T^{-\frac{1}{2}\kappa})$.

Corollary 7. *Given Theorem 4, the approximations of the finite sample distribution of $K(\theta_0)$ by the distribution of its bootstrap $\tilde{K}(\theta_0)$ read:*

1. When Assumpion 2 holds such that $n_\kappa(s) = \tilde{n}_\kappa(s) = 0$:*

$$\Pr[\mathrm{K}(\theta_0) \leq s] = \Pr[\tilde{\mathrm{K}}(\theta_0) \leq s] + O_p(T^{-\frac{3}{2}\kappa}), \tag{66}$$

2. When $n_\kappa(s) \neq 0$, $\tilde{n}_\kappa(s) \neq 0$:

$$\Pr[\mathrm{K}(\theta_0) \leq s] = \Pr[\tilde{\mathrm{K}}(\theta_0) \leq s] + O_p(T^{-\kappa}). \tag{67}$$

Corollary 7 shows that the bootstrap leads to an improved approximation of the finite sample distribution of $\mathrm{K}(\theta_0)$ both when $n_\kappa(s) = \tilde{n}_\kappa(s) = 0$ and when $n_\kappa(s) \neq 0$, $\tilde{n}_\kappa(s) \neq 0$. The improvement of the approximation of the finite sample distribution is valid in all cases of $J_\theta(\theta_0)$.

In an identical manner as outlined above for $\mathrm{K}(\theta_0)$, it is possible to obtain 2κ-th order Edgeworth approximations to the finite sample distributions of $\mathrm{J}_K(\theta_0)$ and $\mathrm{S}(\theta_0)$. Hence, also for these statistics the bootstrap leads to an improvement of the approximation of the finite sample distribution.

5.2 Bootstrap Algorithms

The bootstrap algorithm to obtain the distribution of $\mathrm{K}(\theta_0)$ can be specified by:

1. Obtain bootstrap sample $\{[\tilde{f}_t(\theta_0)' \; \tilde{q}_t(\theta_0)']', t = 1, \ldots, T^*\}$ by drawing from $\{(\bar{f}_t(\theta_0)' \; q_t(\theta_0)')'$, $t = 1, \ldots, T\}$ with replacement.
2. Construct: $\tilde{V}(\theta_0)$, $\tilde{D}_T(\theta_0, Y)$ and $\tilde{f}_T(\theta_0, Y)$ from the bootstrap sample $\{[\tilde{f}_t(\theta_0)' \; \tilde{q}_t(\theta_0)']'$, $t = 1, \ldots, T^*\}$.
3. Compute:

$$\begin{aligned}\tilde{\mathrm{K}}(\theta_0) = \; & \tfrac{1}{T^*} \tilde{f}_T(\theta_0, Y)' \tilde{V}_{ff}(\theta_0)^{-1} \tilde{D}_T(\theta_0, Y) [\tilde{D}_T(\theta_0, Y)' \tilde{V}_{ff}(\theta_0)^{-1} \\ & \tilde{D}_T(\theta_0, Y)]^{-1} \tilde{D}_T(\theta_0, Y)' \tilde{V}_{ff}(\theta_0)^{-1} \tilde{f}_T(\theta_0, Y).\end{aligned}$$

We illustrate the performance of the bootstrap algorithm in Section 6 for a dynamic panel data model. When $[f_t(\theta)' \; q_t(\theta)']'$, $t = 1, \ldots, T$, are dependent, we can use a parametric bootstrap that results from a statistical model that incorporates the dependence or use block-bootstrap algorithms, see *e.g.* Hall and Horowitz (1996). By drawing blocks of the appropriate length, we incorporate the dependence of $[f_t(\bar{\theta}_0)' \; q_t(\theta_{cue})']'$, $t = 1, \ldots, T$, into the bootstrap. The bootstrap algoritms for $\mathrm{J_K}(\theta_0)$ and $\mathrm{S}(\theta_0)$ are identical to the bootstrap algorithm for $\mathrm{K}(\theta_0)$ and only differ with respect to the computed statistic.

6 Power comparison for Panel AR(1) Model

Panel AR(1) model. We compare power curves of statistics that test a hypothesis on the autoregressive parameter of a panel autoregressive model of order 1 (AR(1)). For $\mathrm{K}(\theta_0)$, we use both critical values that result from the limiting distribution of its zero-th order element and from the bootstrap from Section 5. An elaborate literature on panel autoregressive (AR) models exists, see *e.g.* Anderson and Hsiao (1981), Arellano and Bond (1991) and Arellano and Honoré (2001). In panel data models the cross-section dimension N exceeds the time series dimension T. In line with the literature on panel data models, we therefore indicate the sample size by N. In the previous sections, the sample size was indicated by T.

For individual n at time t, the panel AR(1) model reads

$$y_{t,n} = c_n + \theta y_{t-1,n} + \varepsilon_{t,n} \qquad t = 1, \ldots, T, \; n = 1, \ldots, N. \tag{68}$$

The disturbances $\varepsilon_{t,n}$ are assumed to be independent with mean zero. We take first differences to remove individual specific constants:

$$\Delta y_{t,n} = \theta \Delta y_{t-1,n} + \Delta \varepsilon_{t,n} \qquad t = 2, \ldots, T, \; n = 1, \ldots, N, \tag{69}$$

with $\Delta y_{t,n} = y_{t,n} - y_{t-1,n}$. Estimation of the parameter θ in (69) by means of least squares leads to a biased estimator in samples with a finite value of T, see *e.g.* Nickel (1981). We therefore estimate it using GMM. The moment equation (1) for the panel AR(1) reads

$$E(\varphi(\theta, y_{t,n})) = E(\Delta\varepsilon_{t,n}) = E(\Delta y_{t,n} - \theta \Delta y_{t-1,n}) = 0 \qquad t = 2, \ldots, T, \; n = 1, \ldots, N. \tag{70}$$

A common choice of the instruments is to use all two period and more lagged level values of $y_{t,n}$, *i.e.* $X_{t,n} = (y_{t-2,n} \cdots y_{1,n})'$, see *e.g.* Arellano and Bond (1991). This leads to the specification of the moment equation $f_n(\theta)$,

$$f_n(\theta) = X_n \varphi_n(\theta) : \tfrac{1}{2}(T-1)(T-2) \times 1 \qquad n = 1, \ldots, N, \tag{71}$$

with $\varphi_n(\theta) = (\Delta y_{3,n} - \theta \Delta y_{2,n} \ldots \Delta y_{T,n} - \theta \Delta y_{T-1,n})'$ and

$$X_n = \begin{pmatrix} y_{1,n} & 0\ldots0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0\ldots0 & \begin{pmatrix} y_{1,n} \\ \vdots \\ y_{T-2,n} \end{pmatrix} \end{pmatrix} : \tfrac{1}{2}(T-1)(T-2) \times (T-2). \tag{72}$$

Besides the independence of $\varepsilon_{t,n}$ and finite fourth order moments, we make no assumptions about the covariance structure of $\varepsilon_{t,n}$. We therefore use White's (1980) covariance matrix estimator:

$$\hat{V}_{ff}(\theta) = \tfrac{1}{N}\sum_{n=1}^{N} \bar{f}_n(\theta)\bar{f}_n(\theta)' : \tfrac{1}{2}(T-1)(T-2) \times \tfrac{1}{2}(T-1)(T-2). \tag{73}$$

We also use White's (1980) non-parametric covariance matrix estimator for $\hat{V}_{\theta f}(\theta)$ which is involved in $\mathrm{K}(\theta_0)$:

$$\hat{V}_{\theta f}(\theta) = \tfrac{1}{N}\sum_{n=1}^{N} \left[\tfrac{\partial}{\partial\theta}\bar{f}_n(\theta)\right]\bar{f}_n(\theta)' : \tfrac{1}{2}(T-1)(T-2) \times \tfrac{1}{2}(T-1)(T-2). \tag{74}$$

The derivative $\frac{\partial}{\partial\theta}\varphi_n(\theta) = -(\Delta y_{2,n} \ldots \Delta y_{T-1,n})'$ is a white noise series when $\theta = 1$. The parameter θ is therefore not identified when it is equal to one. Weak identified values of θ occur when it is close to one relative to the sample size, *i.e.* when $\frac{1-\theta}{N}$ is small. It implies that the statistics in Theorem 1 whose zero-th order elements depend on ν become size distorted when θ_0 is close to one relative to the sample size. We analyze this by computing power curves for the different statistics for various values of θ_0 and N.

Power comparison. We use the moment equations and covariance matrix estimators for the panel AR(1) model to conduct a size and power comparison of the different statistics discussed previously. We therefore compute power curves for $\mathrm{W}_{2s}(\theta_0)$, $\mathrm{W}_{cue}(\theta_0)$, $\mathrm{LM}(\theta_0)$ and $\mathrm{K}(\theta_0)$ that test $\mathrm{H}_0 : \theta = \theta_0$ with the covariance matrix estimators (73)-(74) and a 95% asymptotic critical value that results from the limiting distribution of the zero-th order term. We also compute the power curve of $\mathrm{K}(\theta_0)$ when we use the 95% critical value that results from the bootstrap algorithm from Section 5.

We compute power curves of the different statistics using a data generating process that has independent disturbances $\varepsilon_{t,n}$ which are generated from a student t distribution with 10 degrees of freedom and mean zero and variance one. The individual specific constant terms c_n are specified as $c_n = (1-\theta)\mu_n$ where the μ_n's are independent realizations from a $N(0,2)$ distribution. The initial observations $y_{0,n}$ are simulated such that $y_{0,n} = \mu_n + \varepsilon_{0,n}$ where the $\varepsilon_{0,n}$'s are independent realizations of standard normal random variables. The bootstrap critical values are computed using 99 bootstrap realizations from the empirical distribution for each simulated dataset. The number of simulated datasets equals 1000. Panel 1 shows the power curves when $N = 50$, Panel 2 when $N = 100$ and Panel 3 when $N = 250$. The number of time periods is equal to six in all three panels, $T = 6$. All three panels contain the power curves for hypothezes that test for four different values of θ : 0.5, 0.7, 0.9 and 0.95.

Panel 1: Power curves of $\mathrm{W}_{2s}(\theta_0)$ (solid with stars), $\mathrm{W}_{cue}(\theta_0)$ (solid with plusses), $\mathrm{LM}(\theta_0)$ (dashed), $\mathrm{K}(\theta_0)$ (solid) that test $\mathrm{H}_0 : \theta = \theta_0$ with 95% significance using asymptotic critical value and bootstrap critical values (dashed-dotted) for $\mathrm{K}(\theta_0)$, $T = 6$, $N = 50$.

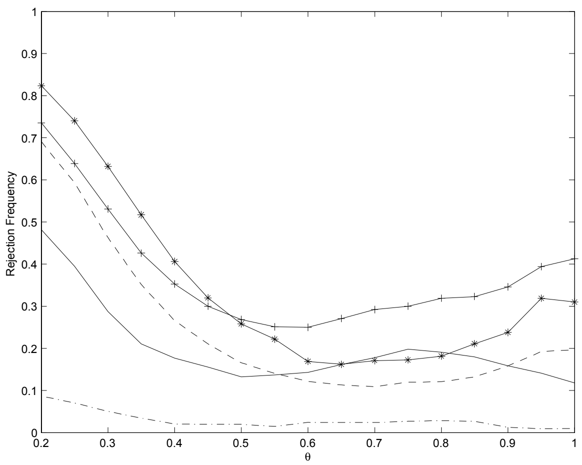

Figure 1.1: $\theta_0 = 0.5$

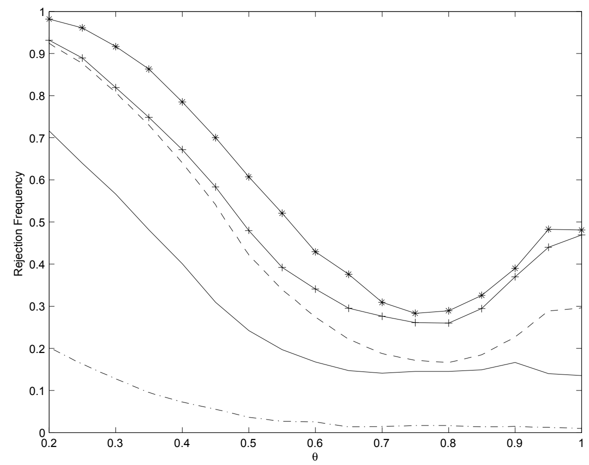

Figure 1.2: $\theta_0 = 0.7$

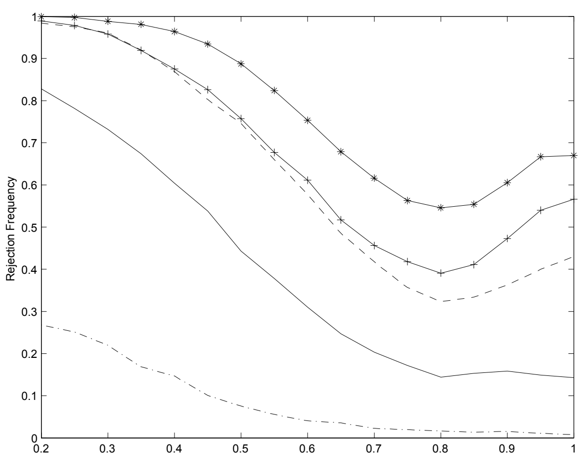

Figure 1.3: $\theta_0 = 0.9$

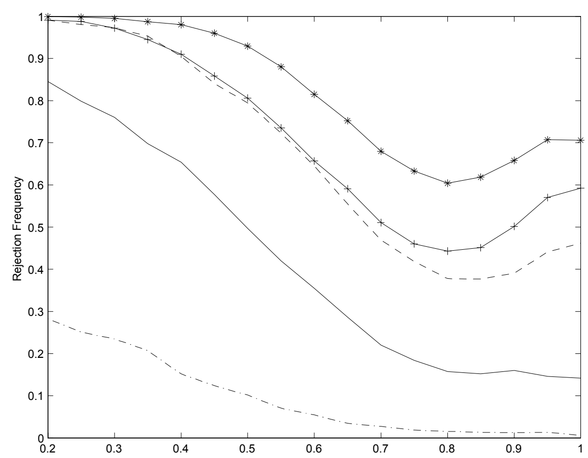

Figure 1.4: $\theta_0 = 0.95$

Panel 1 shows the power curves for data sets with $T = 6$ and $N = 50$. All statistics in Panel 1 are size distorted. The size distortion is clearly the smallest for $\mathrm{K}(\theta_0)$ both when we use asymptotic or bootstrap critical values. Panel 1 also shows that the size distortion for $\mathrm{K}(\theta_0)$ is more or less independent of θ_0 both when we use asymptotic or bootstrap critical values. As expected from the higher order expressions, the size distortion of $\mathrm{W}_{2s}(\theta_0)$, $\mathrm{W}_{\mathrm{cue}}(\theta_0)$ and $\mathrm{LM}(\theta_0)$ rises when θ_0 increases. The size distortion of $\mathrm{W}_{2s}(\theta_0)$ exceeds that of $\mathrm{W}_{\mathrm{cue}}(\theta_0)$ which is in accordance with the higher order expressions from Theorem 1. The size distortion of $\mathrm{LM}(\theta_0)$ is smaller than that of $\mathrm{W}_{2s}(\theta_0)$ and $\mathrm{W}_{\mathrm{cue}}(\theta_0)$. This indicates that a considerable part of the size distortion results from the covariance matrix estimator $\hat{V}(\theta)$ (73). $\mathrm{LM}(\theta_0)$ evaluates $\hat{V}(\theta)$ at θ_0 while $\mathrm{W}_{2s}(\theta_0)$ and $\mathrm{W}_{cue}(\theta_0)$ evaluate it at $\hat{\theta}_{2s}$ and $\hat{\theta}_{cue}$ resp.. Hence, a large part of the size distortion results from evaluating $\hat{V}(\theta)$ at an estimate of θ instead of the true value, see also Bond and Windmeijer (2003).

Since θ is not identified in the moment equations when θ equals one, the power and size of the different statistics should coincide when θ equals one. This property holds for $\mathrm{K}(\theta_0)$ both when we use asymptotic or bootstrap critical values. The power of $\mathrm{K}(\theta_0)$ when θ equals one is similar for all the different values of θ_0 that are considered in Panel 1. The power of $\mathrm{W}_{2s}(\theta_0)$, $\mathrm{W}_{\mathrm{cue}}(\theta_0)$ and $\mathrm{LM}(\theta_0)$ at $\theta = 1$ clearly depends on the value of θ_0.

Panel 2: Power curves of $\mathrm{W}_{2s}(\theta_0)$ (solid with stars), $\mathrm{W}_{cue}(\theta_0)$ (solid with plusses), $\mathrm{LM}(\theta_0)$ (dashed), $\mathrm{K}(\theta_0)$ (solid) that test $\mathrm{H}_0 : \theta = \theta_0$ with 95% significance using asymptotic critical value and bootstrap critical values (dashed-dotted) for $\mathrm{K}(\theta_0)$, $T = 6$, $N = 100$.

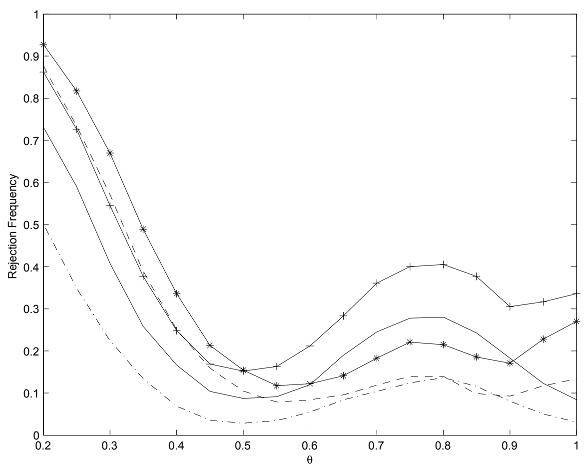

Figure 2.1: $\theta_0 = 0.5$

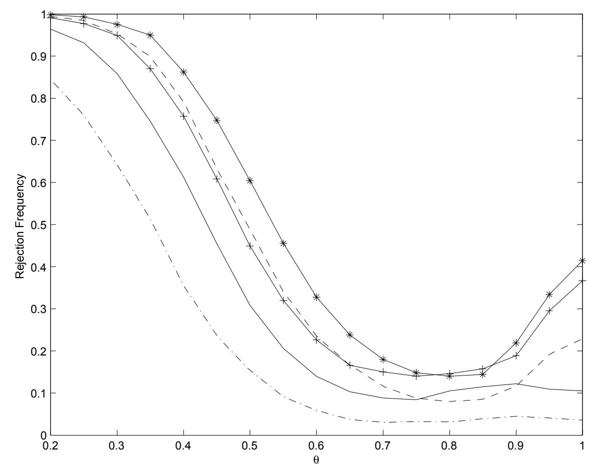

Figure 2.2: $\theta_0 = 0.7$

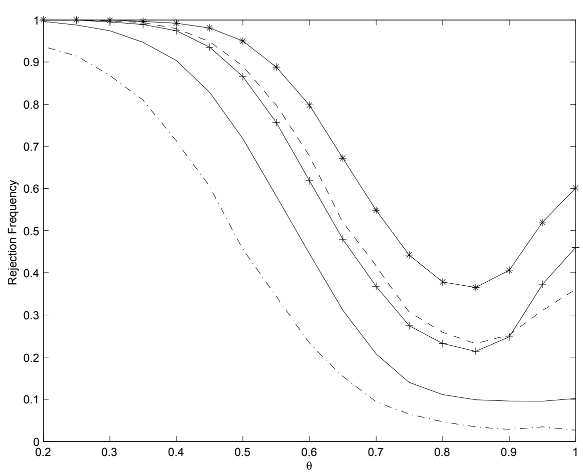

Figure 2.3: $\theta_0 = 0.9$

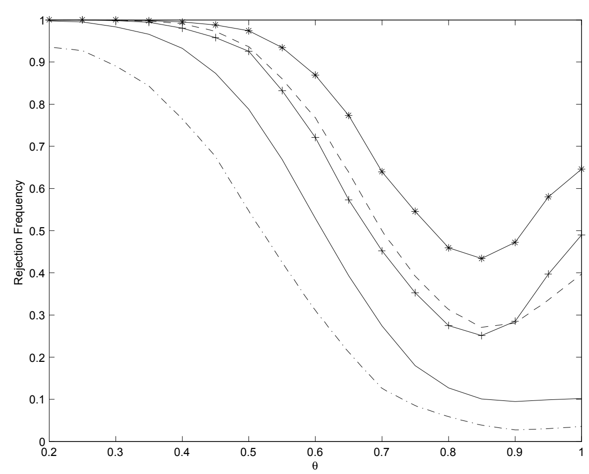

Figure 2.4: $\theta_0 = 0.95$

Panel 2 shows that the size distortions of the different statistics are reduced when $N = 100$ compared to $N = 50$ and that the power has increased. The size distortion of $\mathrm{K}(\theta_0)$ when we use a bootstrap critical value is clearly smaller than the size distortion that results from using the asymptotic critical value. The size distortion is the smallest for $\mathrm{K}(\theta_0)$ also when we use the asymptotic critical value. The size distortion of $\mathrm{W}_{2s}(\theta_0)$, $\mathrm{W}_{\mathrm{cue}}(\theta_0)$ and $\mathrm{LM}(\theta_0)$ is an increasing function of θ_0 as expected from Theorem 1.

The power of $\mathrm{K}(\theta_0)$ when θ equals one in Panel 2 is similar to its size both when we use asymptotic or bootstrap critical values. This is clearly not the case for any of the other statistics and the power when θ equals one clearly depends on θ_0.

Panel 3 shows that the size distortion of $\mathrm{K}(\theta_0)$ with bootstrap critical values has become neglibible when $N = 250$. There is still some size distortion when we use $\mathrm{K}(\theta_0)$ with the asymptotic critical value. For $\mathrm{W}_{2s}(\theta_0)$, $\mathrm{W}_{\mathrm{cue}}(\theta_0)$ and $\mathrm{LM}(\theta_0)$, it holds that their size distortion is small when θ_0 equals 0.5 and 0.7 but is still considerable for larger values of θ_0, *i.e.* 0.9 and 0.95, as expected from Theorem 1. Also the power of these statistics when θ equals one is not equal to the size.

Panel 3: Power curves of $\mathrm{W}_{2s}(\theta_0)$ (solid with stars), $\mathrm{W}_{cue}(\theta_0)$ (solid with plusses), $\mathrm{LM}(\theta_0)$ (dashed), $\mathrm{K}(\theta_0)$ (solid) that test $\mathrm{H}_0 : \theta = \theta_0$ with 95% significance using asymptotic critical value and bootstrap critical values (dashed-dotted) for $\mathrm{K}(\theta_0)$, $T = 6$, $N = 250$.

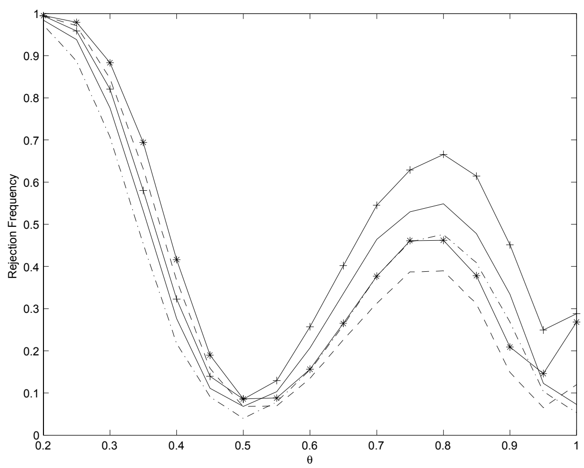

Figure 3.1: $\theta_0 = 0.5$

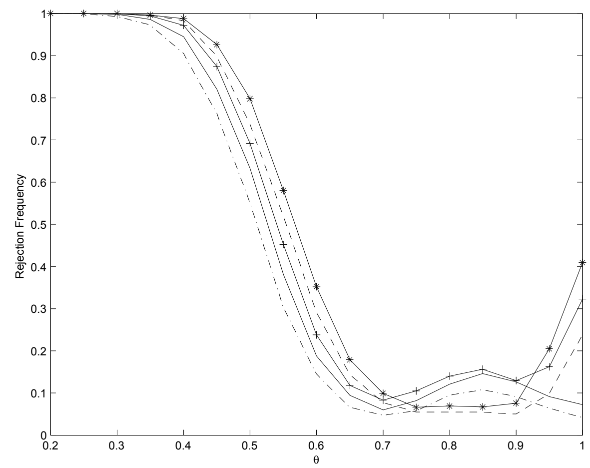

Figure 3.2: $\theta_0 = 0.7$

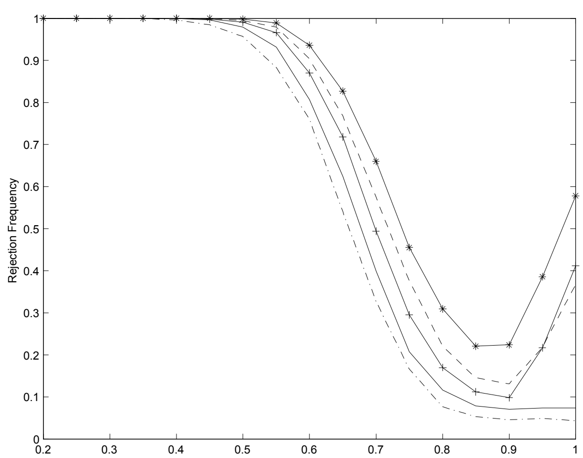

Figure 3.3: $\theta_0 = 0.9$

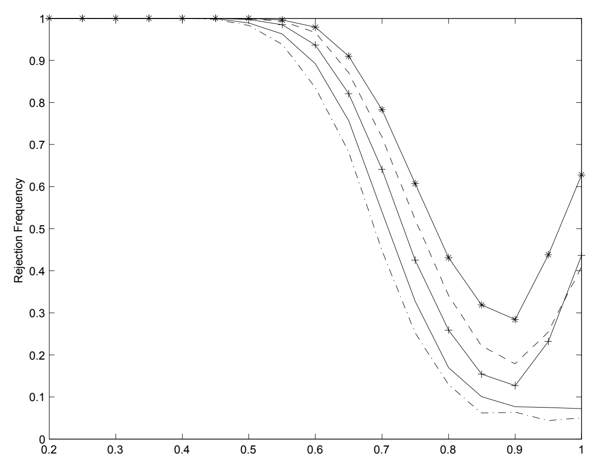

Figure 3.4: $\theta_0 = 0.95$

Panels 1-3 confirm the theoretical findings from Sections 2-5. Panels 1-3 show that the finite sample distributions of $\mathrm{W}_{2s}(\theta_0)$, $\mathrm{W}_{\mathrm{cue}}(\theta_0)$ and $\mathrm{LM}(\theta_0)$ converge at a slower rate than the distribution of $\mathrm{K}(\theta_0)$. This holds especially at larger values of θ_0. Because the moment equations do not identify θ when it is equal to one, the concentration parameter has a different convergence rate for larger values of θ_0. Theorem 1 shows that this leads to a slower convergence of the finite sample distribution because of the additional lower order terms in the higher order expressions of $\mathrm{W}_{2s}(\theta_0)$, $\mathrm{W}_{\mathrm{cue}}(\theta_0)$ and $\mathrm{LM}(\theta_0)$ compared to $\mathrm{K}(\theta_0)$. This explains the slower convergence of the finite sample distributions of $\mathrm{W}_{2s}(\theta_0)$, $\mathrm{W}_{\mathrm{cue}}(\theta_0)$ and $\mathrm{LM}(\theta_0)$ towards the limiting distribution compared to $\mathrm{K}(\theta_0)$. Panels 1-3 also shows that the bootstrap improves the approximation of the finite sample distribution of $\mathrm{K}(\theta_0)$ compared to its limiting distribution as stated in Corollary 7.

7 Conclusions

Appendix

A. Proof of Theorem 1.

We construct the higher order properties of the statistics (1, 2, 3 and 4) in a sequence of steps. First, we obtain the higher order properties of the score vectors involved in the different statistics, step a. Secondly, we obtain the higher order properties of the inverse of the covariance matrix, step b. We combine the different elements of the score vectors and the covariance matrix to obtain the higher order properties of the statistics, step c.

1a. Higher order properties of $f_T(\theta_0,Y)'V_{ff}(\hat{\theta}_{2s})^{-1}p_T(\hat{\theta}_{2s},Y)$ **used in** $\mathbf{W}_{2s}(\theta_0)$**.** To obtain the higher order elements for the 2-step Wald-statistic, we use that

$$(\hat{\theta}-\theta_0)\approx \quad [p_T(\hat{\theta}_{2s},Y)'\hat{V}_{ff}(\hat{\theta}_{2s})^{-1}p_T(\hat{\theta}_{2s},Y)]^{-1}p_T(\hat{\theta}_{2s},Y)'\hat{V}_{ff}(\hat{\theta}_{2s})^{-1}f_T(\theta_0,Y).$$

We specify $p_T(\hat{\theta}_{2s},Y)$ as

$$p_T(\hat{\theta}_{2s},Y)=D_T(\theta_0,Y)+p_T(\hat{\theta}_{2s},Y)-\hat{D}_T(\hat{\theta}_{2s},Y)+\hat{D}_T(\hat{\theta}_{2s},Y)-D_T(\theta_0,Y),$$

with

$$\begin{aligned}&p_T(\hat{\theta}_{2s},Y)-\hat{D}_T(\hat{\theta}_{2s},Y)\\&=\left[A_1\hat{V}_{\theta f,1}(\hat{\theta}_{2s})\hat{V}_{ff}(\hat{\theta}_{2s})^{-1}f_T(\hat{\theta}_{2s},Y)\cdots A_m\hat{V}_{\theta f,m}(\hat{\theta}_{2s})\hat{V}_{ff}(\hat{\theta}_{2s})^{-1}f_T(\hat{\theta}_{2s},Y)\right]\\&=[A_1\hat{V}_{\theta f,1}(\hat{\theta}_{2s})\cdots A_m\hat{V}_{\theta f,m}(\hat{\theta}_{2s})][I_m\otimes\hat{V}_{ff}(\hat{\theta}_{2s})^{-1}f_T(\hat{\theta}_{2s},Y)]\end{aligned}$$

and $\hat{V}(\hat{\theta}_{2s})=V(\theta_0)+[\hat{V}(\hat{\theta}_{2s})-\hat{V}(\theta_0)]+[\hat{V}(\theta_0)-V(\theta_0)]=V(\theta_0)+[\hat{V}(\hat{\theta}_{2s})-V(\hat{\theta}_{2s})]+[V(\hat{\theta}_{2s})-V(\theta_0)]$. The convergence rate of $\hat{V}_{ff}(\hat{\theta})-V_{ff}(\theta_0)$ and $\hat{D}_T(\hat{\theta}_{2s},Y)-D_T(\theta_0,Y)$ is therefore equal to $T^{-\frac{1}{2}\kappa}$, with $\kappa=\min(\mu,\nu)$.

Using Assumption 1, $\frac{1}{\sqrt{T}}f_T(\theta_0,Y)'\hat{V}_{ff}(\hat{\theta}_{2s})^{-1}[T^{-\frac{1}{2}(1+\nu)}p_T(\hat{\theta}_{2s},Y)]$ then reads

$$\begin{aligned}\tfrac{1}{\sqrt{T}}f_T(\theta_0,Y)'\hat{V}_{ff}(\hat{\theta}_{2s})^{-1}[T^{-\frac{1}{2}(1+\nu)}p_T(\hat{\theta}_{2s},Y)]=\;& s_0+T^{-\frac{\nu}{2}}s_{1\nu,1}+T^{-\frac{\kappa}{2}}s_{1\kappa,1}+\\&T^{-\frac{1}{2}(\nu+\kappa)}(s_{\nu+\kappa,1}+s_{\nu+\kappa,2})+T^{-\frac{1}{2}(\nu+2\kappa)}s_{\nu+2\kappa,1}+o_p(T^{-\frac{1}{2}}),\end{aligned}$$

with $\kappa=\min(\mu,\nu)$ and

$$\begin{aligned}s_0=\;& m_{0,f}'V_{ff}(\theta_0)^{-1}D_0\\ s_{1\nu,1}=\;& m_{0,f}'V_{ff}(\theta_0)^{-1}\{\tfrac{1}{\sqrt{T}}[p_T(\hat{\theta},Y)-\hat{D}_T(\hat{\theta},Y)]\}\\ =\;& m_{0,f}'V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}m_{0,f}]\\ s_{1\kappa,1}=\;& T^{\frac{\kappa}{2}}m_{0,f}'[\hat{V}_{ff}(\theta_0)^{-1}-V_{ff}(\theta_0)^{-1}]D_0\\ s_{\nu+\kappa,1}=\;& T^{\frac{1}{2}(\kappa-1)}m_{0,f}'V_{ff}(\theta_0)^{-1}[\hat{D}_T(\hat{\theta},Y)-D_T(\theta_0,Y)]\\ =\;& T^{\frac{1}{2}\kappa}m_{0,f}'V_{ff}(\theta_0)^{-1}\{[A_1\hat{V}_{\theta f,1}(\theta_0)\cdots A_m\hat{V}_{\theta f,m}(\theta_0)][I_m\otimes\hat{V}_{ff}(\theta_0)^{-1}]\\ &-[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}]\}[I_m\otimes m_{0,f}]\\ s_{\nu+\kappa,2}=\;& T^{\frac{1}{2}\kappa}m_{0,f}'[\hat{V}_{ff}(\theta_0)^{-1}-V_{ff}(\theta_0)^{-1}][A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}m_{0,f}]\\ s_{\nu+2\kappa,1}=\;& T^{\frac{1}{2}(2\kappa-1)}m_{0,f}'[\hat{V}_{ff}(\theta_0)^{-1}-V_{ff}(\theta_0)^{-1}][\hat{D}_T(\theta_0,Y)-D_T(\theta_0,Y)],\end{aligned}$$

We also used that

$$m_{0,f}'V_{ff}(\theta_0)^{-1}[\hat{D}_T(\hat{\theta},Y)-D_T(\theta_0,Y)]=$$
$$m_{0,f}'V_{ff}(\theta_0)^{-1}[p_T(\hat{\theta},Y)-p_T(\theta_0,Y)]+m_{0,f}'V_{ff}(\theta_0)^{-1}\{[A_1V_{\theta f,1}(\hat{\theta})\cdots A_mV_{\theta f,m}(\hat{\theta})][I_m\otimes V_{ff}(\hat{\theta})^{-1}]-$$
$$[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}]\}[I_m\otimes V_{ff}(\theta_0)^{-1}m_{0,f}].$$

The convergence of $\hat{p}_T(\hat{\theta},Y)-p_T(\theta_0,Y)$ is of order $T^{\frac{1}{2}\nu}$ and therefore $T^{-\frac{1}{2}(1+\nu)}[\hat{p}_T(\hat{\theta},Y)-p_T(\theta_0,Y)]$ is of a lower order, $T^{-(1+v)}$, than the other elements. We have therefore left it out. The convergence rate of $m_{0,f}'V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}m_{0,f}]$ results from

$$E[\lim_{T\to\infty}\tfrac{1}{T}(f_T(\theta_0,Y)'\otimes f_T(\theta_0,Y)')]=\ \text{vec}[E(\lim_{T\to\infty}\tfrac{1}{T}f_T(\theta_0,Y)f_T(\theta_0,Y)']=\text{vec}[V_{ff}(\theta_0)],$$

so we obtain the expression for the limiting expectation:

$$\begin{aligned}&E\{\lim_{T\to\infty}\tfrac{1}{T}f_T(\theta_0,Y)'V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}f_T(\theta_0,Y)]\}\\&=[\{I_m\otimes\{E[\lim_{T\to\infty}\tfrac{1}{T}f_T(\theta_0,Y)'\otimes f_T(\theta_0,Y)']\}\}\\&\quad\text{vec}[V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)V_{ff}(\theta_0)^{-1}\cdots A_mV_{\theta f,m}(\theta_0)V_{ff}(\theta_0)^{-1}]]]'\\&=\{I_m\otimes\text{vec}[V_{ff}(\theta_0)]'\}\text{vec}[V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)V_{ff}(\theta_0)^{-1}\cdots A_mV_{\theta f,m}(\theta_0)V_{ff}(\theta_0)^{-1}]]\\&=\text{vec}[V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)]],\end{aligned}$$

which shows the appropriate convergence rate.

2a. Higher order properties of $f_T(\theta_0,Y)'\hat{V}_{ff}(\hat{\theta}_{cue})^{-1}\hat{D}_T(\hat{\theta}_{cue},Y)$ used in $\mathbf{W}_{cue}(\theta_0)$.

$$\tfrac{1}{\sqrt{T}}f_T(\theta_0,Y)'\hat{V}_{ff}(\hat{\theta}_{cue})^{-1}[T^{-\frac{1}{2}(1+\nu)}\hat{D}_T(\hat{\theta}_{cue},Y)]=\ s_0+T^{-\frac{\kappa}{2}}s_{1\kappa,1}+T^{-\frac{1}{2}(\nu+\kappa)}s_{\nu+\kappa,1}+T^{-\frac{1}{2}(\nu+2\kappa)}s_{\nu+2\kappa,1}+o_p(T^{-\frac{1}{2}}),$$

with $\kappa=\min(\nu,\mu)$.

3a. Higher order properties of $f_T(\theta_0,Y)'V_{ff}(\theta_0)^{-1}p_T(\theta_0,Y)$ used in LM(θ_0).

$$\tfrac{1}{\sqrt{T}}f_T(\theta_0,Y)'\hat{V}_{ff}(\theta_0)^{-1}[T^{-\frac{1}{2}(1+\nu)}p_T(\theta_0,Y)]=\ s_0+T^{-\frac{\nu}{2}}s_{1\nu,1}+T^{-\frac{\kappa}{2}}s_{1\kappa,1}+T^{-\frac{1}{2}(\nu+\kappa)}(s_{\nu+\kappa,1}+s_{\nu+\kappa,2})+T^{-\frac{1}{2}(\nu+2\kappa)}s_{\nu+2\kappa,1}+o_p(T^{-\frac{1}{2}}),$$

with $\kappa=\mu$ since we evaluate all elements in θ_0 only.

4a. Higher order properties of $f_T(\theta_0,Y)'V_{ff}(\theta_0)^{-1}D_T(\theta_0,Y)$ used in K(θ_0).

$$\tfrac{1}{\sqrt{T}}f_T(\theta_0,Y)'\hat{V}_{ff}(\theta_0)^{-1}[T^{-\frac{1}{2}(1+\nu)}D_T(\theta_0,Y)]=\ s_0+T^{-\frac{\kappa}{2}}s_{1\kappa,1}+T^{-\frac{1}{2}(\nu+\kappa)}s_{\nu+\kappa,1}+T^{-\frac{1}{2}(\nu+2\kappa)}s_{\nu+2\kappa,1},$$

with $\kappa=\mu$ since we evaluate all elements in θ_0 only.

1b. Higher order properties of $p_T(\hat{\theta},Y)'\hat{V}_{ff}(\hat{\theta})^{-1}p_T(\hat{\theta},Y)$ used in $\mathbf{W}_{2s}(\theta_0)$.

$$\begin{aligned}&[T^{-\frac{1}{2}(1+\nu)}p_T(\hat{\theta}_{2s},Y)]'\hat{V}_{ff}(\hat{\theta}_{2s})^{-1}[T^{-\frac{1}{2}(1+\nu)}p_T(\hat{\theta}_{2s},Y)]=G_0+T^{-\frac{\nu}{2}}G_{1\nu,1}+T^{-\frac{\kappa}{2}}G_{1\kappa,1}+\\&T^{-\nu}G_{2\nu,1}+T^{-\frac{1}{2}(\nu+\kappa)}(G_{\nu+\kappa,1}+G_{\nu+\kappa,2})+T^{-\frac{1}{2}(2\nu+\kappa)}(G_{2\nu+\kappa,1}+G_{2\nu+\kappa,2})+\\&T^{-\frac{1}{2}(\nu+2\kappa)}G_{\nu+2\kappa,1}+T^{-(\nu+\kappa)}(G_{2(\nu+\kappa),1}+G_{2(\nu+\kappa),2})+T^{-\frac{2\nu+3\kappa}{2}}G_{2\nu+3\kappa,1}+O_p(T^{-\frac{1}{2}(2\nu+1)}),\end{aligned}$$

with $\kappa = \min(\nu, \mu)$ and

$$\begin{aligned}
G_0 = & \; D_0' V_{ff}(\theta_0)^{-1} D_0 \\
G_{1\nu,1} = & \; D_0' V_{ff}(\theta_0)^{-1} [A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)][I_m \otimes V_{ff}(\theta_0)^{-1} m_{0,f}] + \\
& [I_m \otimes V_{ff}(\theta_0)^{-1} m_{0,f}]' [A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)]' V_{ff}(\theta_0)^{-1} D_0 \\
G_{1\kappa,1} = & \; T^{\frac{\kappa}{2}} D_0' [\hat{V}_{ff}(\theta_0)^{-1} - V_{ff}(\theta_0)^{-1}] D_0 \\
G_{2\nu,1} = & \; [I_m \otimes V_{ff}(\theta_0)^{-1} m_{0,f}]' [A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)]' V_{ff}(\theta_0)^{-1} \\
& [A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)][I_m \otimes V_{ff}(\theta_0)^{-1} m_{0,f}] \\
G_{\nu+\kappa,1} = & \; T^{\frac{\kappa}{2}} D_0' [\hat{V}_{ff}(\theta_0)^{-1} - V_{ff}(\theta_0)^{-1}][A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)][I_m \otimes V_{ff}(\theta_0)^{-1} m_{0,f}] + \\
& T^{\frac{\kappa}{2}} [I_m \otimes V_{ff}(\theta_0)^{-1} m_{0,f}]' [A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)]' [\hat{V}_{ff}(\theta_0)^{-1} - V_{ff}(\theta_0)^{-1}] D_0 \\
G_{\nu+\kappa,2} = & \; T^{\frac{1}{2}(\kappa-1)} [\hat{D}_T(\theta_0, Y) - D_T(\theta_0, Y)]' V_{ff}(\theta_0)^{-1} D_0 + \\
& T^{\frac{1}{2}(\kappa-1)} D_0' V_{ff}(\theta_0)^{-1} [\hat{D}_T(\theta_0, Y) - D_T(\theta_0, Y)] \\
G_{2\nu+\kappa,1} = & \; T^{\frac{1}{2}\kappa} [I_m \otimes V_{ff}(\theta_0)^{-1} m_{0,f}]' [A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)]' [\hat{V}_{ff}(\theta_0)^{-1} - V_{ff}(\theta_0)^{-1}] \\
& [A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)][I_m \otimes V_{ff}(\theta_0)^{-1} m_{0,f}]
\end{aligned}$$

$$\begin{aligned}
G_{2\nu+\kappa,2} = & \; T^{\frac{1}{2}(\kappa-1)} [\hat{D}_T(\theta_0, Y) - D_T(\theta_0, Y)]' V_{ff}(\theta_0)^{-1} [A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)] \\
& [I_m \otimes V_{ff}(\theta_0)^{-1} m_{0,f}] + T^{\frac{1}{2}(\kappa-1)} [I_m \otimes V_{ff}(\theta_0)^{-1} m_{0,f}]' [A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)]' \\
& V_{ff}(\theta_0)^{-1} [\hat{D}_T(\theta_0, Y) - D_T(\theta_0, Y)] \\
G_{\nu+2\kappa,3} = & \; T^{\frac{1}{2}(2\kappa-1)} [\hat{D}_T(\theta_0, Y) - D_T(\theta_0, Y)]' [\hat{V}_{ff}(\theta_0)^{-1} - V_{ff}(\theta_0)^{-1}] D_0 + \\
& T^{\frac{1}{2}(2\kappa-1)} D_0' [\hat{V}_{ff}(\theta_0)^{-1} - V_{ff}(\theta_0)^{-1}][\hat{D}_T(\theta_0, Y) - D_T(\theta_0, Y)] \\
G_{2\nu+2\kappa,1} = & \; T^{\frac{1}{2}(2\kappa-1)} [\hat{D}_T(\theta_0, Y) - D_T(\theta_0, Y)]' [\hat{V}_{ff}(\theta_0)^{-1} - V_{ff}(\theta_0)^{-1}][A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)] \\
& [I_m \otimes V_{ff}(\theta_0)^{-1} m_{0,f}] + T^{\frac{1}{2}(2\kappa-1)} [I_m \otimes V_{ff}(\theta_0)^{-1} m_{0,f}]' [A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)]' \\
& [\hat{V}_{ff}(\theta_0)^{-1} - V_{ff}(\theta_0)^{-1}][\hat{D}_T(\theta_0, Y) - D_T(\theta_0, Y)] \\
G_{2\nu+2\kappa,2} = & \; T^{(\kappa-1)} [\hat{D}_T(\theta_0, Y) - D_T(\theta_0, Y)]' V_{ff}(\theta_0)^{-1} [\hat{D}_T(\theta_0, Y) - D_T(\theta_0, Y)] \\
G_{2\nu+3\kappa,1} = & \; T^{\frac{1}{2}(3\kappa-2)} [D_T(\hat{\theta}, Y) - D_T(\theta_0, Y)]' [V_{ff}(\hat{\theta})^{-1} - V_{ff}(\theta_0)^{-1}][D_T(\hat{\theta}, Y) - D_T(\theta_0, Y)]
\end{aligned}$$

Hence,

$$T^{(1+\nu)} [p_T(\hat{\theta}, Y)' V_{ff}(\theta_0)^{-1} p_T(\hat{\theta}, Y)]^{-1} = G_0^{-1} + T^{-\frac{\nu}{2}} Q_1,$$

with

$$Q_1 = -G_0^{-1} [(G_{1\nu,1} + T^{-\frac{\nu}{2}} H)^{-1} + T^{-\frac{\nu}{2}} G_0^{-1}]^{-1} G_0^{-1},$$

where $H = T^{-\frac{\kappa-2\nu}{2}} G_{1\kappa,1} + G_{2\nu,1} + T^{-\frac{1}{2}(\kappa-\nu)} (G_{\nu+\kappa,1} + G_{\nu+\kappa,2}) + T^{-\frac{1}{2}\kappa} (G_{2\nu+\kappa,1} + G_{2\nu+\kappa,2}) + T^{-\frac{1}{2}(2\kappa-\nu)} G_{\nu+2\kappa,1} + T^{-\kappa} (G_{2(\nu+\kappa),1} + G_{2(\nu+\kappa),2}) + T^{-\frac{3\kappa}{2}} G_{2\nu+3\kappa,1}$.

2b. Higher order properties of $p_T(\hat{\theta}, Y)' V_{ff}(\hat{\theta})^{-1} p_T(\hat{\theta}, Y)$ used in $\mathbf{W}_{cue}(\theta_0)$.

$$\begin{aligned}
& [T^{-\frac{1}{2}(1+\nu)} p_T(\hat{\theta}, Y)]' \hat{V}_{ff}(\hat{\theta})^{-1} [T^{-\frac{1}{2}(1+\nu)} p_T(\hat{\theta}, Y)] = G_0 + T^{-\frac{\kappa}{2}} G_{1\kappa,1} + T^{-\frac{1}{2}(\nu+\kappa)} G_{\nu+\kappa,2} + \\
& T^{-\frac{1}{2}(\nu+2\kappa)} G_{\nu+2\kappa,1} + T^{-(\nu+\kappa)} G_{2(\nu+\kappa),2} + T^{-\frac{1}{2}(2\nu+3\kappa)} G_{2\nu+3\kappa,1} + O_p(T^{-\frac{1}{2}(2\nu+1)}),
\end{aligned}$$

with $\kappa = \min(\nu, \mu)$. Hence,

$$T^{(1+\nu)} [p_T(\hat{\theta}, Y)' V_{ff}(\hat{\theta})^{-1} p_T(\hat{\theta}, Y)]^{-1} = G_0^{-1} + T^{-\frac{\kappa}{2}} Q_1,$$

with

$$Q_1 = \; -G_0^{-1} [(G_{1\kappa,1} + T^{-\frac{\kappa}{2}} H)^{-1} + T^{-\frac{\kappa}{2}} G_0^{-1}]^{-1} G_0^{-1},$$

where $H=T^{-\frac{1}{2}\nu}G_{\nu+\kappa,2}+T^{-\kappa}G_{\nu+2\kappa,1}+T^{-\frac{1}{2}(\nu+2\kappa)}G_{2(\nu+\kappa),2}+T^{-\frac{1}{2}(\nu+3\kappa)}G_{2\nu+3\kappa,1}$.

3b. Higher order properties of $p_T(\theta_0,Y)'\hat{V}_{ff}(\theta_0)^{-1}p_T(\theta_0,Y)$ used in LM(θ_0).

$$[T^{-\frac{1}{2}(1+\nu)}p_T(\theta_0,Y)]'\hat{V}_{ff}(\theta_0)^{-1}[T^{-\frac{1}{2}(1+\nu)}p_T(\theta_0,Y)]=G_0+T^{-\frac{\nu}{2}}G_{1\nu,1}+ T^{-\frac{\kappa}{2}}G_{1\kappa,1}+T^{-\nu}G_{2\nu,1}+T^{-\frac{1}{2}(\nu+\kappa)}(G_{\nu+\kappa,1}+G_{\nu+\kappa,2})+T^{-\frac{1}{2}(2\nu+\kappa)}(G_{2\nu+\kappa,1}+G_{\nu+\kappa,2})+ T^{-\frac{1}{2}(\nu+2\kappa)}G_{\nu+2\kappa,1}+T^{-(\nu+\kappa)}(G_{2(\nu+\kappa),1}+G_{2(\nu+\kappa),2})+T^{-\frac{1}{2}(2\nu+3\kappa)}G_{2\nu+3\kappa,1}+O_p(T^{-\frac{1}{2}(2\nu+1)}),$$

with $\kappa=\mu$. Hence,

$$T^{(1+\nu)}[p_T(\theta_0,Y)'\hat{V}_{ff}(\theta_0)^{-1}p_T(\theta_0,Y)]^{-1}=G_0^{-1}+T^{-\frac{\nu}{2}}Q_1,$$

with

$$Q_1=\ -G_0^{-1}[(G_{1\kappa,1}+T^{-\frac{\nu}{2}}H)^{-1}+T^{-\frac{\nu}{2}}G_0^{-1}]^{-1}G_0^{-1},$$

where $H=T^{-\frac{\kappa-2\nu}{2}}G_{1\kappa,1}+G_{2\nu,1}+T^{-\frac{1}{2}(\kappa-\nu)}(G_{\nu+\kappa,1}+G_{\nu+\kappa,2})+T^{-\frac{1}{2}\kappa}(G_{2\nu+\kappa,1}+G_{\nu+\kappa,2})+T^{-\frac{1}{2}(2\kappa-\nu)}G_{\nu+2\kappa,1}+T^{-\kappa}(G_{2(\nu+\kappa),1}+G_{2(\nu+\kappa),2})+T^{-\frac{3\kappa}{2}}G_{2\nu+3\kappa,1}$,

4b. Higher order properties of $\hat{D}_T(\theta_0,Y)'\hat{V}_{ff}(\theta_0)^{-1}\hat{D}_T(\theta_0,Y)$ used in K(θ_0).

$$[T^{-\frac{1}{2}(1+\nu)}\hat{D}_T(\theta_0,Y)]'\hat{V}_{ff}(\theta_0)^{-1}[T^{-\frac{1}{2}(1+\nu)}\hat{D}_T(\theta_0,Y)]=G_0+T^{-\frac{\kappa}{2}}G_{1\kappa,1}+T^{-\frac{1}{2}(\nu+\kappa)}G_{\nu+\kappa,2}+T^{-\frac{1}{2}(\nu+2\kappa)}G_{\nu+2\kappa,1}+ T^{-(\nu+\kappa)}G_{2(\nu+\kappa),2}+T^{-\frac{1}{2}(2\nu+3\kappa)}G_{2\nu+3\kappa,1}+O_p(T^{-\frac{1}{2}(2\nu+1)}),$$

with $\kappa=\mu$. Hence,

$$T^{(1+\nu)}\left[\hat{D}_T(\theta_0,Y)'\hat{V}_{ff}(\theta_0)^{-1}\hat{D}_T(\theta_0,Y)\right]^{-1}=G_0^{-1}+T^{-\frac{\kappa}{2}}Q_1,$$

with

$$Q_1=\ -G_0^{-1}[(G_{1\kappa,1}+T^{-\frac{\kappa}{2}}H)^{-1}+T^{-\frac{\kappa}{2}}G_0^{-1}]^{-1}G_0^{-1},$$

where $H=T^{-\frac{1}{2}(\nu-\kappa)}G_{\nu+\kappa,2}+T^{-\frac{1}{2}\nu}G_{\nu+2\kappa,1}+T^{-\nu}G_{2(\nu+\kappa),2}+T^{-(\nu+\frac{1}{2}\kappa)}G_{2\nu+3\kappa,1}$.

1c. The higher order components of $\mathrm{W}_{2s}(\theta_0)$ that result from m_0 in Assumption 1 can be specified as:

$$\begin{aligned}\mathrm{W}_{2s}(\theta_0)=\ & n_0+T^{-\frac{\nu}{2}}n_\nu+T^{-\frac{\kappa}{2}}n_\kappa+T^{-\frac{\nu+\kappa}{2}}n_{\nu+\kappa}+T^{-\nu}n_{2\nu}+T^{-\kappa}n_{2\kappa}\\ & +T^{-\frac{1}{2}(2\nu+\kappa)}n_{2\nu+\kappa}+T^{-\frac{1}{2}(\nu+2\kappa)}n_{\nu+2\kappa}+O_p(T^{-\frac{3}{2}\nu})\end{aligned}$$

with

$$\begin{aligned}n_0=\ & s_0'G_0^{-1}s_0\\ n_\nu=\ & s_0'Q_1s_0+s_{1\nu,1}'G_0^{-1}s_0+s_0'G_0^{-1}s_{1\nu,1}\\ n_\kappa=\ & s_{1\kappa,1}'G_0^{-1}s_0+s_0'G_0^{-1}s_{1\kappa,1}\\ n_{\nu+\kappa}=\ & s_{1\nu,1}'G_0^{-1}s_{1\kappa,1}+s_{1\kappa,1}'G_0^{-1}s_{1\nu,1}+s_{1\kappa,1}'Q_1s_0+s_0'Q_1s_{1\nu+\kappa}+\\ & (s_{\nu+\kappa,1}+s_{\nu+\kappa,2})'G_0^{-1}s_0+s_0'G_0^{-1}(s_{\nu+\kappa,1}+s_{\nu+\kappa,2})\\ n_{2\nu}=\ & s_{1\nu,1}'Q_1s_0+s_0'Q_1s_{1\nu,1}+s_{1\nu,1}'G_0^{-1}s_{1\nu,1}\\ n_{2\kappa}=\ & s_{1\nu,\kappa}'G_0^{-1}s_{1\kappa,1}\\ n_{2\nu+\kappa}=\ & s_{1\nu,\kappa}'Q_1s_{1\kappa,1}+s_{1\kappa,1}'Q_1s_{1\nu,1}+(s_{\nu+\kappa,1}+s_{\nu+\kappa,2})'G_0^{-1}s_{1\nu,1}+s_{1\nu,1}'G_0^{-1}(s_{\nu+\kappa,1}+s_{\nu+\kappa,2})\\ n_{\nu+2\kappa}=\ & (s_{\nu+\kappa,1}+s_{\nu+\kappa,2})'G_0^{-1}s_{1\kappa,1}+s_{1\kappa,1}'G_0^{-1}(s_{\nu+\kappa,1}+s_{\nu+\kappa,2})+s_{\nu+2\kappa,1}'G_0^{-1}s_0+\\ & s_0'G_0^{-1}s_{\nu+\kappa,1}+s_{1\kappa,1}'Q_1s_{1\kappa,1}\end{aligned}$$

and $\kappa = \min(\nu, \mu)$.

2c. The higher order components of $\mathrm{W}_{cue}(\theta_0)$ that result from m_0 in Assumption 1 can be specified as:

$$\mathrm{W}_{cue}(\theta_0) = \quad n_0 + T^{-\frac{\kappa}{2}} n_\kappa + T^{-\frac{\nu+\kappa}{2}} n_{\nu+\kappa} + T^{-\kappa} n_{2\kappa} + T^{-\frac{\nu+2\kappa}{2}} n_{\nu+2\kappa} + T^{-\frac{3}{2}\kappa} n_{3\kappa} + o_p(T^{-\frac{3\kappa}{2}})$$

with

$$\begin{aligned}
n_0 =& \quad s_0' G_0^{-1} s_0 \\
n_\kappa =& \quad s_0' Q_1 s_0 + s_{1\kappa,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{1\kappa,1} \\
n_{\nu+\kappa} =& \quad s_{\nu+\kappa,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{\nu+\kappa,1} \\
n_{2\kappa} =& \quad s_{1\kappa,1}' G_0^{-1} s_{1\kappa,1} + s_{1\kappa,1}' Q_1 s_0 + s_0' Q_1 s_{1\kappa,1} \\
n_{\nu+2\kappa} =& \quad s_{\nu+\kappa,1}' G_0^{-1} s_{1\kappa,1} + s_{1\kappa,1}' G_0^{-1} s_{\nu+\kappa,1} + s_0' G_0^{-1} s_{\nu+2\kappa,1} \\
& \quad s_0' Q_1 s_{\nu+\kappa,1} + s_{\nu+\kappa,1}' Q_1 s_0 + s_{\nu+2\kappa,1}' G_0^{-1} s_0 \\
n_{3\kappa} =& \quad s_{1\kappa,1}' Q_1 s_{1\kappa,1}
\end{aligned}$$

and $\kappa = \min(\mu, \nu)$.

3c. The higher order components of $\mathrm{LM}(\theta_0)$ that result from m_0 in Assumption 1 can be specified as:

$$\begin{aligned}
\mathrm{LM}(\theta_0) =& \quad n_0 + T^{-\frac{\nu}{2}} n_\nu + T^{-\frac{\kappa}{2}} n_\kappa + T^{-\frac{\nu+\kappa}{2}} n_{\nu+\kappa} + T^{-\nu} n_{2\nu} + T^{-\kappa} n_{2\kappa} \\
& + T^{-\frac{1}{2}(2\nu+\kappa)} n_{2\nu+\kappa} + T^{-\frac{1}{2}(\nu+2\kappa)} n_{\nu+2\kappa} + O_p(T^{-\frac{3}{2}\nu})
\end{aligned}$$

with $\kappa = \mu$ and

$$\begin{aligned}
n_0 =& \quad s_0' G_0^{-1} s_0 \\
n_\nu =& \quad s_0' Q_1 s_0 + s_{1\nu,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{1\nu,1} \\
n_\kappa =& \quad s_{1\kappa,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{1\kappa,1} \\
n_{\nu+\kappa} =& \quad s_{1\nu,1}' G_0^{-1} s_{1\kappa,1} + s_{1\kappa,1}' G_0^{-1} s_{1\nu,1} + s_{1\kappa,1}' Q_1 s_0 + s_0' Q_1 s_{1\kappa,1} + \\
& \quad s_{\nu+\kappa,2}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{\nu+\kappa,2} \\
n_{2\nu} =& \quad s_{1\nu,1}' G_0^{-1} s_{1\nu,1} + s_{1\nu,1}' Q_1 s_0 + s_0' Q_1 s_{1\nu,1} \\
n_{2\kappa} =& \quad s_{1\kappa,1}' G_0^{-1} s_{1\kappa,1} \\
n_{2\nu+\kappa} =& \quad s_{1\kappa,1}' Q_1 s_{1\nu,1} + s_{1\nu,1}' Q_1 s_{1\kappa,1} + s_{\nu+\kappa,2}' Q_1 s_0 + s_{\nu+\kappa,2}' G_0^{-1} s_{\nu,1} + s_0' Q_1 s_{\nu+\kappa,2} + s_{1\nu,1}' G_0^{-1} s_{\nu+\kappa,2} \\
n_{\nu+2\kappa} =& \quad s_{\nu+2\kappa,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{\nu+2\kappa,1} + s_{1\kappa,1}' Q_1 s_{1\kappa,1} + s_{\nu+\kappa,2}' G_0^{-1} s_{1\kappa,1} + s_{1\kappa,1}' G_0^{-1} s_{\nu+\kappa,2}
\end{aligned}$$

4c. The higher order components of $\mathrm{K}(\theta_0)$ that result from m_0 in Assumption 1 can be specified as:

$$\mathrm{K}(\theta_0) = \quad n_0 + T^{-\frac{\kappa}{2}} n_\kappa + T^{-\frac{\nu+\kappa}{2}} n_{\nu+\kappa} + T^{-\kappa} n_{2\kappa} + T^{-\frac{\nu+2\kappa}{2}} n_{\nu+2\kappa} + T^{-\frac{3}{2}\kappa} n_{3\kappa} + o_p(T^{-\frac{3\kappa}{2}})$$

with $\kappa = \mu$ and

$$\begin{aligned}
n_0 =& \quad s_0' G_0^{-1} s_0 \\
n_\kappa =& \quad s_{1\kappa,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{1\kappa,1} + s_0' Q_1 s_0 \\
n_{\nu+\kappa} =& \quad s_{\nu+\kappa,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{\nu+\kappa,1} \\
n_{2\kappa} =& \quad s_{1\kappa,1}' G_0^{-1} s_{1\kappa,1} + s_{1\kappa,1}' Q_1 s_0 + s_0' Q_1 s_{1\kappa,1} \\
n_{\nu+2\kappa} =& \quad s_{\nu+2\kappa,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{\nu+2\kappa,1} + s_{\nu+\kappa,1}' G_0^{-1} s_{1\kappa,1} + s_{1\kappa,1}' G_0^{-1} s_{\nu+\kappa,2} + \\
& \quad s_0' Q_1 s_{\nu+\kappa,1} + s_{\nu+\kappa,1}' Q_1 s_0 \\
n_{3\kappa} =& \quad s_{1\kappa,1}' Q_1 s_{1\kappa,1}.
\end{aligned}$$

B. Lemma 2. We construct the conditional expectation of the limit expressions of n_ν and $n_{2\nu}$ given ρ when the number of observations converges to infinity. We begin with n_v which consists of two parts: $s'_{1\nu,1}G_0^{-1}s_0$ and $s'_0Q_1s_0$:

$\mathbf{s}'_{1\nu,1}\mathbf{G}_0^{-1}\mathbf{s}_0$: We specify ψ_f as

$$V_{ff}(\theta_0)^{-\frac{1}{2}}\psi_f = \quad V_{ff}(\theta_0)^{-\frac{1}{2}}D_0\rho + V_{ff}(\theta_0)^{\frac{1}{2}}D_{0,\perp}\lambda,$$

with $\rho = (D'_0V_{ff}(\theta_0)^{-1}D_0)^{-1}D'_0V_{ff}(\theta_0)^{-1}\psi_f$ and $\lambda = (D'_{0,\perp}V_{ff}(\theta_0)D_{0,\perp})^{-1}D'_{0,\perp}\psi_f$ and $D_{0,\perp} : k_f \times (k_f - m)$, $D'_{0,\perp}D_0 \equiv 0$, $D'_{0,\perp}D_{0,\perp} \equiv I_{k_f-m}$ so ρ and λ are independent and $\rho \sim N(0, (D'_0V_{ff}(\theta_0)^{-1}D_0)^{-1})$, $\lambda \sim N(0, (D'_{0,\perp}V_{ff}(\theta_0)D_{0,\perp})^{-1})$. This implies that $\lim_{T\to\infty}s'_{1\nu,1}G_0^{-1}s_0$ can be specified as:

$$\begin{aligned}
&\lim_{T\to\infty} s'_{1\nu,1}G_0^{-1}s_0 \\
&= \psi'_fV_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m \otimes V_{ff}(\theta_0)^{-1}\psi_f]\rho \\
&= [\lambda'D'_{0,\perp} + \rho'D'_0V_{ff}(\theta_0)^{-1}][A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][\rho \otimes \{V_{ff}(\theta_0)^{-1}D_0\rho + D_{0,\perp}\lambda\}] \\
&= \text{tr}\{[\rho \otimes \{V_{ff}(\theta_0)^{-1}D_0\rho + D_{0,\perp}\lambda\}][\lambda'D'_{0,\perp} + \rho'D'_0V_{ff}(\theta_0)^{-1}][A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)]\},
\end{aligned}$$

since $[I_m \otimes d]c = [d'c_1 \dots d'c_m]' = [c \otimes d]$ with c and d $m \times 1$ and $k_f \times 1$ vectors. Because

$$\begin{aligned}
&E\{[\rho \otimes \{V_{ff}(\theta_0)^{-1}D_0\rho + D_{0,\perp}\lambda\}][\lambda'D_{0,\perp} + \rho'D'_0V_{ff}(\theta_0)^{-1}]|\rho\} \\
&= E\{\rho \otimes V_{ff}(\theta_0)^{-1}D_0\rho\lambda'D_{0,\perp}|\rho\} + E\{\rho \otimes V_{ff}(\theta_0)^{-1}D_0\rho\rho'D'_0V_{ff}(\theta_0)^{-1}|\rho\}+ \\
&\quad E\{\rho \otimes D_{0,\perp}\lambda\lambda'D_{0,\perp}|\rho\} + E\{\rho \otimes D_{0,\perp}\lambda\rho'D'_0V_{ff}(\theta_0)^{-1}|\rho\} \\
&= E\{\rho \otimes V_{ff}(\theta_0)^{-1}D_0\rho\rho'D'_0V_{ff}(\theta_0)^{-1}|\rho\} + E\{\rho \otimes D_{0,\perp}\lambda\lambda'D'_{0,\perp}|\rho\} \\
&= [\rho \otimes V_{ff}(\theta_0)^{-1}D_0\rho\rho'D'_0V_{ff}(\theta_0)^{-1}] + [\rho \otimes D_{0,\perp}(D'_{0,\perp}V_{ff}(\theta_0)D_{0,\perp})^{-1}D'_{0,\perp}]
\end{aligned}$$

where we used that $E(\lambda) = 0$, $E(\rho|\rho) = \rho$ and $E(\lambda\lambda') = (D'_{0,\perp}V_{ff}(\theta_0)D_{0,\perp})^{-1}$, we obtain that

$$\begin{aligned}
&E[\lim_{T\to\infty} s'_{1\nu,1}G_0^{-1}s_0|\rho] \\
&= \text{tr}\{[\rho \otimes V_{ff}(\theta_0)^{-1}D_0\rho\rho'D'_0V_{ff}(\theta_0)^{-1}][A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)]\}+ \\
&\quad\quad \text{tr}\{[\rho \otimes D_{0,\perp}(D'_{0,\perp}V_{ff}(\theta_0)D_{0,\perp})^{-1}D'_{0,\perp}][A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)]\} \\
&= \textstyle\sum_{i=1}^m \text{tr}\{\rho_iV_{ff}(\theta_0)^{-1}D_0\rho\rho'D'_0V_{ff}(\theta_0)^{-1}A_iV_{\theta f,i}(\theta_0)\}+ \\
&\quad\quad \textstyle\sum_{i=1}^m \text{tr}\{\rho_iD_{0,\perp}(D'_{0,\perp}V_{ff}(\theta_0)D_{0,\perp})^{-1}D'_{0,\perp}A_iV_{\theta f,i}(\theta_0)\} \\
&= \textstyle\sum_{i=1}^m \rho_i\rho'D'_0V_{ff}(\theta_0)^{-1}A_iV_{\theta f,i}(\theta_0)V_{ff}(\theta_0)^{-1}D_0\rho+ \\
&\quad\quad \textstyle\sum_{i=1}^m \rho_i \sum_{j=1}^{k_f-m}\sum_{n=1}^{k_f-m} \left[(D'_{0,\perp}V_{ff}(\theta_0)D_{0,\perp})^{-1}\right]_{jn} [D'_{0,\perp}A_iV_{\theta f,i}(\theta_0)D_{0,\perp}]_{nj},
\end{aligned}$$

where $\left[(D'_{0,\perp}V_{ff}(\theta_0)D_{0,\perp})^{-1}\right]_{jn}$ and $[D'_{0,\perp}A_iV_{\theta f,i}(\theta_0)D_{0,\perp}]_{jn}$ are the jn-th element of the respective matrix.

$\mathbf{s}'_0\mathbf{Q}_1\mathbf{s}_0$: We assume that $\nu = \mu = 1$,

$$\begin{aligned}
&\lim_{T\to\infty} Q_1 = \lim_{T\to\infty} G_0^{-1}(G_{1\nu,1} + G_{1\kappa,1})G_0^{-1} = (D'_0V_{ff}(\theta_0)^{-1}D_0)^{-1}[D'_0V_{ff}(\theta_0)^{-1} \\
&[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m \otimes \{V_{ff}(\theta_0)^{-1}D_0\rho + D_{0,\perp}\lambda\}] + [I_m \otimes \{V_{ff}(\theta_0)^{-1}D_0\rho + D_{0,\perp}\lambda\}]' \\
&[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)]'V_{ff}(\theta_0)^{-1}D_0 + D'_0\Psi_uD_0](D'_0V_{ff}(\theta_0)^{-1}D_0)^{-1}
\end{aligned}$$

and

$$\begin{aligned}
&\lim_{T\to\infty} s'_0Q_1s_0 = \\
&\rho'\{[D'_0V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m \otimes \{V_{ff}(\theta_0)^{-1}D_0\rho + D_{0,\perp}\lambda\}]+ \\
&[I_m \otimes \{V_{ff}(\theta_0)^{-1}D_0\rho + D_{0,\perp}\lambda\}]'[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)]'V_{ff}(\theta_0)^{-1}D_0 + D'_0\Psi_uD_0\}\rho \\
&= \rho'D'_0V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\{V_{ff}(\theta_0)^{-1}D_0\rho + D_{0,\perp}\lambda\}\cdots A_mV_{\theta f,m}(\theta_0)\{V_{ff}(\theta_0)^{-1}D_0\rho + D_{0,\perp}\lambda\}]\rho+ \\
&\rho'[A_1V_{\theta f,1}(\theta_0)\{V_{ff}(\theta_0)^{-1}D_0\rho + D_{0,\perp}\lambda\}\cdots A_mV_{\theta f,m}(\theta_0)\{V_{ff}(\theta_0)^{-1}D_0\rho + D_{0,\perp}\lambda\}]'V_{ff}(\theta_0)^{-1}D_0\rho+ \\
&\rho'D'_0\Psi_uD_0\rho.
\end{aligned}$$

The conditional expectation of $\rho' D_0' \Psi_u D_0 \rho_0$ given ρ equals zero because, by Assumption 2, Ψ_u is independent of ψ_f. The conditional expectation of the remaining part of $s_0' Q_1 s_0$,

$$
\begin{aligned}
&E[\rho'\{D_0' V_{ff}(\theta_0)^{-1}[A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)][I_m \otimes \{V_{ff}(\theta_0)^{-1} D_0 \rho + D_{0,\perp}\lambda\}]\}\rho|\rho] \\
&= \rho' D_0' V_{ff}(\theta_0)^{-1}[A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)][\rho \otimes V_{ff}(\theta_0)^{-1} D_0 \rho] \\
&= \textstyle\sum_{i=1}^m \rho_i \rho' D_0' V_{ff}(\theta_0)^{-1} A_i V_{\theta f,i}(\theta_0) V_{ff}(\theta_0)^{-1} D_0 \rho,
\end{aligned}
$$

which we have shown for the expression of $s_{1\nu,1}' G_0^{-1} s_0$ so

$$
\lim_{T\to\infty} E[s_0' Q_1 s_0|\rho] = \textstyle\sum_{i=1}^m \rho_i \rho' D_0' V_{ff}(\theta_0)^{-1} A_i V_{\theta f,i}(\theta_0) V_{ff}(\theta_0)^{-1} D_0 \rho
$$

and

$$
\begin{aligned}
\lim_{T\to\infty} E[n_\nu|\rho] = \ & 3 \textstyle\sum_{i=1}^m \rho_i \rho' D_0' V_{ff}(\theta_0)^{-1} A_i V_{\theta f,i}(\theta_0) V_{ff}(\theta_0)^{-1} D_0 \rho + \\
& 2 \textstyle\sum_{i=1}^m \rho_i \sum_{j=1}^{k_f - m} \sum_{n=1}^{k_f - m} \left[(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-1} \right]_{jn} [D_{0,\perp}' A_i V_{\theta f,i}(\theta_0) D_{0,\perp}]_{nj}.
\end{aligned}
$$

$n_{2\nu}$ consists of $s_{1\nu,1}' Q_1 s_0$ and $s_{1\nu,1}' G_0^{-1} s_{1\nu,1}$. We construct the limit expressions of the conditional expectations of both of these expressions given ρ.
$\mathbf{s}_{1\nu,1}' \mathbf{Q}_1 \mathbf{s}_0$. We assume that $\nu = \mu = 1$,

$$
\begin{aligned}
&\lim_{T\to\infty} Q_1 = \lim_{T\to\infty} G_0^{-1}(G_{1\nu,1} + G_{1\kappa,1}) G_0^{-1} \\
&= (D_0' V_{ff}(\theta_0)^{-1} D_0)^{-1}[D_0' V_{ff}(\theta_0)^{-1}[A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)][I_m \otimes \{V_{ff}(\theta_0)^{-1} D_0 \rho + D_{0,\perp}\lambda\}] \\
&+ [I_m \otimes \{V_{ff}(\theta_0)^{-1} D_0 \rho + D_{0,\perp}\lambda\}]'[A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)]' V_{ff}(\theta_0)^{-1} D_0 + \\
&D_0' \Psi_u D_0](D_0' V_{ff}(\theta_0)^{-1} D_0)^{-1}
\end{aligned}
$$

so

$$
\begin{aligned}
&\lim_{T\to\infty} s_{1\nu,1}' Q_1 s_0 = \\
&= \psi_f' V_{ff}(\theta_0)^{-1}[A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)][I_m \otimes V_{ff}(\theta_0)^{-1}\psi_f](D_0' V_{ff}(\theta_0)^{-1} D_0)^{-1}\{D_0' V_{ff}(\theta_0)^{-1} \\
&[A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)][I_m \otimes V_{ff}(\theta_0)^{-1}\psi_f] + [I_m \otimes V_{ff}(\theta_0)^{-1}\psi_f]' \\
&[A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)]' V_{ff}(\theta_0)^{-1} D_0 + D_0' \Psi_u D_0\}\rho.
\end{aligned}
$$

Because of the independence of Ψ_u and ρ, the conditional expectation of the part of $s_{1\nu,1}' Q_1 s_0$ that contains Ψ_u equals zero and can be left aside. We construct the conditional expectation of the remaining two parts of $s'_{1\nu,1}Q_1s_0$ given ρ :

$$\begin{array}{l} E\{\psi_f'V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}\psi_f](D_0'V_{ff}(\theta_0)^{-1}D_0)^{-1} \\ D_0'V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}\psi_f]\rho|\rho\} \\ = E\{\rho'D_0'V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}\psi_f](D_0'V_{ff}(\theta_0)^{-1}D_0)^{-1} \\ D_0'V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}\psi_f]\rho|\rho\}+ \\ E\{\lambda'D_{0,\perp}'[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}\psi_f](D_0'V_{ff}(\theta_0)^{-1}D_0)^{-1}D_0'V_{ff}(\theta_0)^{-1} \\ [A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}\psi_f]\rho|\rho\} \\ = E\{\rho'D_0'V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}D_0\rho](D_0'V_{ff}(\theta_0)^{-1}D_0)^{-1} \\ D_0'V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}\psi_f]\rho|\rho\}+E\{\rho'D_0'V_{ff}(\theta_0)^{-1} \\ [A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes D_{0,\perp}\lambda](D_0'V_{ff}(\theta_0)^{-1}D_0)^{-1}D_0'V_{ff}(\theta_0)^{-1} \\ [A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}\psi_f]\rho|\rho\}+E\{\lambda'D_{0,\perp}'[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)] \\ [I_m\otimes V_{ff}(\theta_0)^{-1}D_0\rho](D_0'V_{ff}(\theta_0)^{-1}D_0)^{-1}D_0'V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)] \\ [I_m\otimes V_{ff}(\theta_0)^{-1}\psi_f]\rho|\rho\}+E\{\lambda'D_{0,\perp}'[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes D_{0,\perp}\lambda] \\ (D_0'V_{ff}(\theta_0)^{-1}D_0)^{-1}D_0'V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}\psi_f]\rho|\rho\} \\ = \rho'D_0'V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}D_0\rho](D_0'V_{ff}(\theta_0)^{-1}D_0)^{-1} \\ D_0'V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}D_0\rho]\rho+E\{\rho'D_0'V_{ff}(\theta_0)^{-1} \\ [A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes D_{0,\perp}\lambda](D_0'V_{ff}(\theta_0)^{-1}D_0)^{-1}D_0'V_{ff}(\theta_0)^{-1} \\ [A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes D_{0,\perp}\lambda]\rho|\rho\}+E\{\lambda'D_{0,\perp}'[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)] \\ [I_m\otimes V_{ff}(\theta_0)^{-1}D_0\rho](D_0'V_{ff}(\theta_0)^{-1}D_0)^{-1}D_0'V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)] \\ [I_m\otimes D_{0,\perp}\lambda]\rho|\rho\}+E\{\lambda'D_{0,\perp}'[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes D_{0,\perp}\lambda] \\ (D_0'V_{ff}(\theta_0)^{-1}D_0)^{-1}D_0'V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}D_0\rho]\rho|\rho\} \\ = a_1+a_2+a_3+a_4, \end{array}$$

because all other elements contain first and third order moments of λ which are equal to zero. The expressions for different a-terms read:

$$\begin{array}{l} a_1=\rho'D_0'V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}D_0\rho](D_0'V_{ff}(\theta_0)^{-1}D_0)^{-1} \\ D_0'V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}D_0\rho]\rho, \end{array}$$

$$\begin{array}{l} a_2=E\{\rho'D_0'V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes D_{0,\perp}\lambda](D_0'V_{ff}(\theta_0)^{-1}D_0)^{-1} \\ D_0'V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][\rho\otimes D_{0,\perp}\lambda]|\rho\} \\ =E\{\mathrm{tr}([\rho\otimes D_{0,\perp}\lambda][\rho'D_0'V_{ff}(\theta_0)^{-1}A_1V_{\theta f,1}(\theta_0)D_{0,\perp}\lambda\cdots\rho'D_0'V_{ff}(\theta_0)^{-1}A_mV_{\theta f,m}(\theta_0)D_{0,\perp}\lambda] \\ (D_0'V_{ff}(\theta_0)^{-1}D_0)^{-1}D_0'V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)])|\rho\} \\ =E\{\mathrm{tr}([\rho\otimes D_{0,\perp}][\lambda\rho'D_0'V_{ff}(\theta_0)^{-1}A_1V_{\theta f,1}(\theta_0)D_{0,\perp}\lambda\cdots\lambda\rho'D_0'V_{ff}(\theta_0)^{-1}A_mV_{\theta f,m}(\theta_0)D_{0,\perp}\lambda] \\ (D_0'V_{ff}(\theta_0)^{-1}D_0)^{-1}D_0'V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)])|\rho\} \\ =\mathrm{tr}([\rho\otimes D_{0,\perp}][\sum_{i=1}^{k_f-m}(D_{0,\perp}V_{ff}(\theta_0)D_{0,\perp})_i^{-1}\rho'D_0'V_{ff}(\theta_0)^{-1}A_1V_{\theta f,1}(\theta_0)_i\cdots \\ \sum_{i=1}^{k_f-m}\mathrm{tr}\{(D_{0,\perp}V_{ff}(\theta_0)D_{0,\perp})_i^{-1}\rho'D_0'V_{ff}(\theta_0)^{-1}A_mV_{\theta f,m}(\theta_0)_i](D_0'V_{ff}(\theta_0)^{-1}D_0)^{-1}D_0'V_{ff}(\theta_0)^{-1} \\ [A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)]]) \end{array}$$

since

$$E(\lambda b'\lambda)=\quad E(\lambda\sum_{i=1}^{k_f-m}b_i\lambda_i)=\sum_{i=1}^{k_f-m}(D_{0,\perp}V_{ff}(\theta_0)D_{0,\perp})_i^{-1}b_i,$$

with b_i the i-th element of the $(k_f-m)\times 1$ vector b and $(D_{0,\perp}V_{ff}(\theta_0)D_{0,\perp})_i^{-1}$ the i-th column of

$(D_{0,\perp}V_{ff}(\theta_0)D_{0,\perp})^{-1}$.

$$\begin{array}{l} a_3 = E\{\lambda' D_{0,\perp}'[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}D_0\rho](D_0'V_{ff}(\theta_0)^{-1}D_0)^{-1} \\ D_0'V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes D_{0,\perp}\lambda]\rho|\rho\} = E\{\mathrm{tr}([\rho\otimes D_{0,\perp}\lambda\lambda' D_{0,\perp}' \\ [A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}D_0\rho](D_0'V_{ff}(\theta_0)^{-1}D_0)^{-1}D_0'V_{ff}(\theta_0)^{-1} \\ [A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)])|\rho\} = E\{\mathrm{tr}([\rho\otimes D_{0,\perp}(D_{0,\perp}V_{ff}(\theta_0)D_{0,\perp})^{-1}D_{0,\perp}'[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)] \\ [I_m\otimes V_{ff}(\theta_0)^{-1}D_0\rho](D_0'V_{ff}(\theta_0)^{-1}D_0)^{-1}D_0'V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)]), \end{array}$$

and

$$\begin{array}{l} a_4 = E\{\lambda' D_{0,\perp}'[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes D_{0,\perp}\lambda](D_0'V_{ff}(\theta_0)^{-1}D_0)^{-1} \\ D_0'V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}D_0\rho]\rho|\rho\} = \\ E\{[\lambda' D_{0,\perp}'A_1V_{\theta f,1}(\theta_0)D_{0,\perp}\lambda\cdots\lambda' D_{0,\perp}'A_mV_{\theta f,m}(\theta_0)D_{0,\perp}\lambda](D_0'V_{ff}(\theta_0)^{-1}D_0)^{-1}D_0'V_{ff}(\theta_0)^{-1} \\ [A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}D_0\rho]\rho|\rho\} = \\ [\mathrm{tr}(D_{0,\perp}(D_{0,\perp}V_{ff}(\theta_0)D_{0,\perp})^{-1}D_{0,\perp}'A_1V_{\theta f,1}(\theta_0))\ldots\mathrm{tr}(D_{0,\perp}(D_{0,\perp}V_{ff}(\theta_0)D_{0,\perp})^{-1}D_{0,\perp}'A_mV_{\theta f,m}(\theta_0))] \\ (D_0'V_{ff}(\theta_0)^{-1}D_0)^{-1}D_0'V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}D_0\rho]\rho, \end{array}$$

so $\lim_{T\to\infty} E(s_{1\nu,1}'Q_1s_0|\rho) = a_1+a_2+a_3+a_4$.
$\mathbf{s}_{1\nu,1}'\mathbf{G}_0^{-1}\mathbf{s}_{1\nu,1}$:

$$\begin{array}{l} \lim_{T\to\infty}\mathbf{s}_{1\nu,1}'\mathbf{G}_0^{-1}\mathbf{s}_{1\nu,1} = E\{\psi_f'V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}\psi_f] \\ (D_0'V_{ff}(\theta_0)^{-1}D_0)^{-1}[I_m\otimes V_{ff}(\theta_0)^{-1}\psi_f]'[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)]'V_{ff}(\theta_0)^{-1}\psi_f|\rho\}. \end{array}$$

We construct the conditional expectation given ρ by substituting $V_{ff}(\theta_0)^{-\frac{1}{2}}\psi_f = V_{ff}(\theta_0)^{-\frac{1}{2}}D_0\rho + V_{ff}(\theta_0)^{\frac{1}{2}}D_{0,\perp}\lambda$,

$$\begin{array}{l} \{\psi_f'V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}\psi_f](D_0'V_{ff}(\theta_0)^{-1}D_0)^{-1} \\ [I_m\otimes V_{ff}(\theta_0)^{-1}\psi_f]'[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)]'V_{ff}(\theta_0)^{-1}\psi_f|\rho\} = \\ E\{[\psi_f'V_{ff}(\theta_0)^{-1}A_1V_{\theta f,1}(\theta_0)V_{ff}(\theta_0)^{-1}\psi_f\cdots\psi_f'V_{ff}(\theta_0)^{-1}A_mV_{\theta f,m}(\theta_0)V_{ff}(\theta_0)^{-1}\psi_f](D_0'V_{ff}(\theta_0)^{-1}D_0)^{-1} \\ [\psi_f'V_{ff}(\theta_0)^{-1}A_1V_{\theta f,1}(\theta_0)V_{ff}(\theta_0)^{-1}\psi_f\cdots\psi_f'V_{ff}(\theta_0)^{-1}A_mV_{\theta f,m}(\theta_0)V_{ff}(\theta_0)^{-1}\psi_f]'|\rho\} = \\ = \sum_{i=1}^m\sum_{j=1}^m E\{[\psi_f'V_{ff}(\theta_0)^{-1}A_iV_{\theta f,i}(\theta_0)V_{ff}(\theta_0)^{-1}\psi_f][\psi_f'V_{ff}(\theta_0)^{-1}A_jV_{\theta f,j}(\theta_0)V_{ff}(\theta_0)^{-1}\psi_f]|\rho\} \\ (D_0'V_{ff}(\theta_0)^{-1}D_0)_{ij}^{-1} \\ = \sum_{i=1}^m\sum_{j=1}^m E\{[(\rho' D_0'V_{ff}(\theta_0)^{-1}+\lambda' D_{0,\perp}')A_iV_{\theta f,i}(\theta_0)(V_{ff}(\theta_0)^{-1}D_0\rho+D_{0,\perp}\lambda)] \\ [(\rho' D_0'V_{ff}(\theta_0)^{-1}+\lambda' D_{0,\perp}')A_jV_{\theta f,j}(\theta_0)(V_{ff}(\theta_0)^{-1}D_0\rho+D_{0,\perp}\lambda)]|\rho\}(D_0'V_{ff}(\theta_0)^{-1}D_0)_{ij}^{-1} \\ = \sum_{i=1}^m\sum_{j=1}^m\{[\rho' D_0'V_{ff}(\theta_0)^{-1}A_iV_{\theta f,i}(\theta_0)V_{ff}(\theta_0)^{-1}D_0\rho][\rho' D_0'V_{ff}(\theta_0)^{-1}A_jV_{\theta f,j}(\theta_0)V_{ff}(\theta_0)^{-1}D_0\rho]+ \\ 2[\rho' D_0'V_{ff}(\theta_0)^{-1}A_iV_{\theta f,i}(\theta_0)V_{ff}(\theta_0)^{-1}D_0\rho]\mathrm{tr}[D_{0,\perp}(D_{0,\perp}'V_{ff}(\theta_0)D_{0,\perp})^{-1}D_{0,\perp}'A_jV_{\theta f,j}(\theta_0)]+ \\ E\{\mathrm{tr}[D_{0,\perp}\lambda\lambda' D_{0,\perp}'A_iV_{\theta f,i}(\theta_0)]\mathrm{tr}[D_{0,\perp}\lambda\lambda' D_{0,\perp}'A_jV_{\theta f,j}(\theta_0)]+ \\ [\rho' D_0'V_{ff}(\theta_0)^{-1}A_iV_{\theta f,i}(\theta_0)D_{0,\perp}\lambda][\rho' D_0'V_{ff}(\theta_0)^{-1}A_jV_{\theta f,j}(\theta_0)D_{0,\perp}\lambda]+2[\rho' D_0'V_{ff}(\theta_0)^{-1}A_iV_{\theta f,i}(\theta_0)D_{0,\perp}\lambda] \\ [\lambda' D_{0,\perp}'A_jV_{\theta f,j}(\theta_0)V_{ff}(\theta_0)^{-1}D_0\rho]|\rho\}\}(D_0'V_{ff}(\theta_0)^{-1}D_0)_{ij}^{-1} \\ = \sum_{i=1}^m\sum_{j=1}^m[a_{ij}+b_{ij}+c_{ij}+d_{ij}+e_{ij}](D_0'V_{ff}(\theta_0)^{-1}D_0)_{ij}^{-1}, \end{array}$$

with

$$
\begin{aligned}
a_{ij} = & \ \{[\rho' D_0' V_{ff}(\theta_0)^{-1} A_i V_{\theta f,i}(\theta_0) V_{ff}(\theta_0)^{-1} D_0 \rho][\rho' D_0' V_{ff}(\theta_0)^{-1} A_j V_{\theta f,j}(\theta_0) V_{ff}(\theta_0)^{-1} D_0 \rho] \\
b_{ij} = & \ 2[\rho' D_0' V_{ff}(\theta_0)^{-1} A_i V_{\theta f,i}(\theta_0) V_{ff}(\theta_0)^{-1} D_0 \rho] \mathrm{tr}[D_{0,\perp}(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-1} D_{0,\perp}' A_j V_{\theta f,j}(\theta_0)] \\
c_{ij} = & \ 3 \textstyle\sum_{i_1=1}^{k_f-m} [(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-\frac{1}{2}\prime} D_{0,\perp}' A_i V_{\theta f,i}(\theta_0) D_{0,\perp} (D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-\frac{1}{2}}]_{i_1 i_1}^2 + \\
& 2 \textstyle\sum_{i_1=1}^{k_f-m} \sum_{j_1=1, j_1 \neq i_1}^{k_f-m} [(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-\frac{1}{2}\prime} D_{0,\perp}' A_i V_{\theta f,i}(\theta_0) D_{0,\perp} (D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-\frac{1}{2}}]_{i_1 i_1} \\
& [(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-\frac{1}{2}\prime} D_{0,\perp}' A_j V_{\theta f,j}(\theta_0) D_{0,\perp} (D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-\frac{1}{2}}]_{j_1 j_1} + \\
& 2 \textstyle\sum_{i_1=1}^{k_f-m} \sum_{j_1=1, j_1 \neq i_1}^{k_f-m} [(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-\frac{1}{2}\prime} D_{0,\perp}' A_i V_{\theta f,i}(\theta_0) D_{0,\perp} (D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-\frac{1}{2}}]_{i_1 j_1} \\
& [(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-\frac{1}{2}\prime} D_{0,\perp}' A_j V_{\theta f,j}(\theta_0) D_{0,\perp} (D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-\frac{1}{2}}]_{i_1 j_1} + \\
& 2 \textstyle\sum_{i_1=1}^{k_f-m} \sum_{j_1=1, j_1 \neq i_1}^{k_f-m} [(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-\frac{1}{2}\prime} D_{0,\perp}' A_i V_{\theta f,i}(\theta_0) D_{0,\perp} (D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-\frac{1}{2}}]_{i_1 j_1} \\
& [(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-\frac{1}{2}\prime} D_{0,\perp}' A_j V_{\theta f,j}(\theta_0) D_{0,\perp} (D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-\frac{1}{2}}]_{j_1 i_1} \\
d_{ij} = & \ \rho' D_0' V_{ff}(\theta_0)^{-1} A_i V_{\theta f,i}(\theta_0) D_{0,\perp} (D_{0,\perp} V_{ff}(\theta_0) D_{0,\perp})^{-1} D_{0,\perp} V_{\theta f,j}(\theta_0)' A_j' V_{ff}(\theta_0)^{-1} D_0 \rho \\
e_{ij} = & \ 2\rho' D_0' V_{ff}(\theta_0)^{-1} A_i V_{\theta f,i}(\theta_0) D_{0,\perp} (D_{0,\perp} V_{ff}(\theta_0) D_{0,\perp})^{-1} D_{0,\perp} V_{\theta f,j}(\theta_0)' A_j' V_{ff}(\theta_0)^{-1} D_0 \rho.
\end{aligned}
$$

since all first and third order moments with respect to λ are equal to zero and

$$
\begin{aligned}
& E\{\mathrm{tr}[D_{0,\perp}\lambda\lambda' D_{0,\perp}' A_i V_{\theta f,i}(\theta_0)] \mathrm{tr}[D_{0,\perp}\lambda\lambda' D_{0,\perp}' A_j V_{\theta f,j}(\theta_0)] | \rho\} = \\
& E\{\lambda' D_{0,\perp}' A_i V_{\theta f,i}(\theta_0) D_{0,\perp} \lambda\lambda' D_{0,\perp}' A_j V_{\theta f,j}(\theta_0) D_{0,\perp} \lambda\} = \\
& E\{\zeta' (D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{\frac{1}{2}\prime} D_{0,\perp}' A_i V_{\theta f,i}(\theta_0) D_{0,\perp} (D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{\frac{1}{2}} \zeta \\
& \zeta' (D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{\frac{1}{2}\prime} D_{0,\perp}' A_j V_{\theta f,j}(\theta_0) D_{0,\perp} (D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{\frac{1}{2}} \zeta\} = \\
& \textstyle\sum_{i_1=1}^{k_f-m} \sum_{i_2=1}^{k_f-m} \sum_{j_1=1}^{k_f-m} \sum_{j_2=1}^{k_f-m} \zeta_{i_1}\zeta_{i_2}\zeta_{j_1}\zeta_{j_2} [(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{\frac{1}{2}\prime} D_{0,\perp}' A_i V_{\theta f,i}(\theta_0) D_{0,\perp} \\
& (D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{\frac{1}{2}}]_{i_1 i_2} [(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{\frac{1}{2}\prime} D_{0,\perp}' A_j V_{\theta f,j}(\theta_0) D_{0,\perp} (D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{\frac{1}{2}}]_{j_1 j_2} = \\
& 3 \textstyle\sum_{i_1=1}^{k_f-m} [(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{\frac{1}{2}\prime} D_{0,\perp}' A_i V_{\theta f,i}(\theta_0) D_{0,\perp} (D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{\frac{1}{2}}]_{i_1 i_1}^2 + \\
& 2 \textstyle\sum_{i_1=1}^{k_f-m} \sum_{j_1=1, j_1 \neq i_1}^{k_f-m} [(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{\frac{1}{2}\prime} D_{0,\perp}' A_i V_{\theta f,i}(\theta_0) D_{0,\perp} (D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{\frac{1}{2}}]_{i_1 i_1} \\
& [(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{\frac{1}{2}\prime} D_{0,\perp}' A_j V_{\theta f,j}(\theta_0) D_{0,\perp} (D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{\frac{1}{2}}]_{j_1 j_1} + \\
& 2 \textstyle\sum_{i_1=1}^{k_f-m} \sum_{j_1=1, j_1 \neq i_1}^{k_f-m} [(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{\frac{1}{2}\prime} D_{0,\perp}' A_i V_{\theta f,i}(\theta_0) D_{0,\perp} (D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{\frac{1}{2}}]_{i_1 j_1} \\
& [(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{\frac{1}{2}\prime} D_{0,\perp}' A_j V_{\theta f,j}(\theta_0) D_{0,\perp} (D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{\frac{1}{2}}]_{i_1 j_1} + \\
& 2 \textstyle\sum_{i_1=1}^{k_f-m} \sum_{j_1=1, j_1 \neq i_1}^{k_f-m} [(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{\frac{1}{2}\prime} D_{0,\perp}' A_i V_{\theta f,i}(\theta_0) D_{0,\perp} (D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{\frac{1}{2}}]_{i_1 j_1} \\
& [(D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{\frac{1}{2}\prime} D_{0,\perp}' A_j V_{\theta f,j}(\theta_0) D_{0,\perp} (D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{\frac{1}{2}}]_{j_1 i_1},
\end{aligned}
$$

where we used that $\zeta = (D_{0,\perp}' V_{ff}(\theta_0) D_{0,\perp})^{-\frac{1}{2}} \lambda \sim N(0, I_{k-m})$. Only second and fourth order moments of the same elements of ζ are therefore non-zero.

$$
\begin{aligned}
& E\{[\rho' D_0' V_{ff}(\theta_0)^{-1} A_i V_{\theta f,i}(\theta_0) D_{0,\perp} \lambda][\rho' D_0' V_{ff}(\theta_0)^{-1} A_j V_{\theta f,j}(\theta_0) D_{0,\perp} \lambda] | \rho\} \\
& = E\{[\rho' D_0' V_{ff}(\theta_0)^{-1} A_i V_{\theta f,i}(\theta_0) D_{0,\perp} \lambda][\lambda' D_{0,\perp} V_{\theta f,j}(\theta_0)' A_j' V_{ff}(\theta_0)^{-1} D_0 \rho | \rho\} \\
& = \rho' D_0' V_{ff}(\theta_0)^{-1} A_i V_{\theta f,i}(\theta_0) D_{0,\perp} (D_{0,\perp} V_{ff}(\theta_0) D_{0,\perp})^{-1} D_{0,\perp} V_{\theta f,j}(\theta_0)' A_j' V_{ff}(\theta_0)^{-1} D_0 \rho
\end{aligned}
$$

and

$$
\begin{aligned}
& E\{[\rho' D_0' V_{ff}(\theta_0)^{-1} A_i V_{\theta f,i}(\theta_0) D_{0,\perp} \lambda][\lambda' D_{0,\perp}' A_j V_{\theta f,j}(\theta_0) V_{ff}(\theta_0)^{-1} D_0 \rho] | \rho\} = \\
& = \rho' D_0' V_{ff}(\theta_0)^{-1} A_i V_{\theta f,i}(\theta_0) D_{0,\perp} (D_{0,\perp} V_{ff}(\theta_0) D_{0,\perp})^{-1} D_{0,\perp} V_{\theta f,j}(\theta_0)' A_j' V_{ff}(\theta_0)^{-1} D_0 \rho.
\end{aligned}
$$

The conditional expectation of $n_{2\nu}$ given ρ therefore reads:

$$
E[\lim_{T\to\infty} n_{2\nu} | \rho] = \ a_1 + a_2 + a_3 + a_4 + \textstyle\sum_{i=1}^{m} \sum_{j=1}^{m} [a_{ij} + b_{ij} + c_{ij} + d_{ij} + e_{ij}].
$$

C. Proof of Theorem 2.

In order to apply Lemma 6 from Phillips and Moon (1999), we verify the conditions for Lemma 6 to hold. When k is fixed,

$$\tfrac{1}{k}m_{0.f}'V_{ff}(\theta)^{-1}m_{0.f} \underset{d}{\to} F(1,k)$$

where $F(1,k)$ indicates a Fisher distributed random variable with 1 and k degrees of freedom. The convergence of $\frac{1}{k}m_{0.f}'V_{ff}(\theta)^{-1}m_{0.f}$ is identical for all values of k and is therefore uniform such that Lemma 6 from Phillips and Moon (1999) implies that

$$\tfrac{1}{k}m_{0.f}'V_{ff}(\theta)^{-1}m_{0.f} \underset{p}{\to} 1,$$

when k converges to infinity. Also when k is fixed,

$$m_{0.f}'V_{ff}(\theta)^{-1}\left[T^{-\frac{1}{2}(1+\nu)}D_T(\theta_0,Y)\right] \underset{d}{\to} N(0,D_0'V_{ff}(\theta_0)^{-1}D_0),$$

since $T^{-(1+\nu)}D_T(\theta_0,Y)'V_{ff}(\theta_0)^{-1}D_T(\theta_0,Y) \underset{d}{\to} D_0'V_{ff}(\theta_0)^{-1}D_0$. Since we assume that $T^{-(1+\nu)}D_T(\theta_0,Y)'V_{ff}(\theta_0$ $D_0'V_{ff}(\theta_0)^{-1}D_0$ which is finite, $m_{0.f}'V_{ff}(\theta)^{-1}\left[T^{-\frac{1}{2}(1+\nu)}D_T(\theta_0,Y)\right]$ converges also uniformly when k converges to infinity and we can apply sequential limits.

1a. $\mathbf{W}_{2s}(\theta_0)$. Because $m_{0,f}$ is stochastically bounded and converges to ψ_f when T goes to infinity, the results of Lemma 6 of Phillips and Moon (1999) apply and we can let T and k converge to infinity sequentially, so first T and then k. We construct the order of the different elements of $\mathrm{W}_{2s}(\theta_0)$ when T and k jointly converge to infinity for which we assume that $D_T(\theta_0,Y)'V_{ff}(\theta_0)^{-1}D_T(\theta_0,Y)$ is of the order $T^{1+\nu}$ and $D_T(\theta_0,Y)'V_{ff}(\theta_0)^{-1}f_T(\theta_0,Y)$ is of order $T^{\frac{1}{2}(2+\nu)}$.

When k goes to infinity proportional to T^α and $\nu \geq \alpha$,

$$\begin{array}{l}\frac{1}{\sqrt{T}}f_T(\theta_0,Y)'V_{ff}(\hat{\theta}_{2s})^{-1}[T^{-\frac{1}{2}(1+\nu)}p_T(\hat{\theta}_{2s},Y)] = s_0 + T^{-\frac{\nu-2\alpha}{2}}s_{\nu-2\alpha,1} + T^{-\frac{\kappa}{2}}s_{1\kappa,1}+ \\ T^{-\frac{1}{2}(\nu+\kappa-2\alpha)}(s_{\nu+\kappa-2\alpha,1}+s_{\nu+\kappa-2\alpha,2}) + T^{-\frac{1}{2}(\nu+2(\kappa-\alpha))}s_{\nu+2(\kappa-\alpha),1},\end{array}$$

with $\kappa=\min(\nu,\mu)$ and

$$\begin{array}{rcl}s_0 &=& m_{0,f}'V_{ff}(\theta_0)^{-1}D_0 \\ s_{\nu-2\alpha,1} &=& m_{0,f}'V_{ff}(\theta_0)^{-1}\{\frac{1}{\sqrt{T}}[p_T(\hat{\theta},Y)-\hat{D}_T(\hat{\theta},Y)]\} \\ &=& \frac{1}{k}m_{0,f}'V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}m_{0,f}] \\ s_{1\kappa,1} &=& T^{\frac{1}{2}\kappa}m_{0,f}'[V_{ff}(\hat{\theta})^{-1}-V_{ff}(\theta_0)^{-1}]D_0 \\ s_{\nu+\kappa-2\alpha,1} &=& \frac{T^{\frac{1}{2}\kappa}}{k}m_{0,f}'V_{ff}(\theta_0)^{-1}\{[A_1\hat{V}_{\theta f,1}(\theta_0)\cdots A_m\hat{V}_{\theta f,m}(\theta_0)][I_m\otimes\hat{V}_{ff}(\theta_0)^{-1}] \\ && -[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}]\}[I_m\otimes m_{0,f}] \\ s_{\nu+\kappa-2\alpha,2} &=& \frac{T^{\frac{1}{2}\kappa}}{k}m_{0,f}'[\hat{V}_{ff}(\theta_0)^{-1}-V_{ff}(\theta_0)^{-1}][A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}m_{0,f}] \\ s_{\nu+2(\kappa-\alpha),1} &=& \frac{T^{\frac{1}{2}(2\kappa-1)}}{k}m_{0,f}'[\hat{V}_{ff}(\theta_0)^{-1}-V_{ff}(\theta_0)^{-1}][\hat{D}_T(\theta_0,Y)-D_T(\theta_0,Y)],\end{array}$$

which we obtained by fixing the convergence rate of $f_T(\theta_0,Y)'V_{ff}(\theta_0)^{-1}D_T(\theta_0,Y)$ to $T^{-\frac{1}{2}(2+\nu)}$, with $\nu\geq\alpha$, and use the results that $\frac{1}{k}m_{0,f}'V_{ff}(\theta_0)^{-1}m_{0,f}\underset{p}{\to} l$.

2a. $\mathbf{W}_{cue}(\theta_0)$.

$$\frac{1}{\sqrt{T}}f_T(\theta_0,Y)'\hat{V}_{ff}(\hat{\theta})^{-1}[T^{-\frac{1}{2}(1+\nu)}D_T(\hat{\theta},Y)] = s_0 + T^{-\frac{\kappa}{2}}s_{1\kappa,1} + T^{-\frac{1}{2}(\nu+\kappa-2\alpha)}s_{\nu+\kappa-2\alpha,1} + T^{-\frac{1}{2}(\nu+2(\kappa-\alpha))}s_{\nu+2(\kappa-\alpha),1},$$

with $\kappa = \min(\nu,\mu) \geq \alpha$.

3a. $\mathbf{LM}(\theta_0)$.

$$\frac{1}{\sqrt{T}}f_T(\theta_0,Y)'\hat{V}_{ff}(\theta_0)^{-1}[T^{-\frac{1}{2}(1+\nu)}p_T(\theta_0,Y)] = s_0 + T^{-\frac{\nu-2\alpha}{2}}s_{\nu-2\alpha,1} + T^{-\frac{\kappa}{2}}s_{1\kappa,1} + T^{-\frac{1}{2}(\nu+\kappa-2\alpha)}(s_{\nu+\kappa-2\alpha,1}+s_{\nu+\kappa-2\alpha,2}) + T^{-\frac{1}{2}(\nu+2(\kappa-\alpha))}s_{\nu+2(\kappa-\alpha),1},$$

with $\kappa = \mu$.

4a. $\mathbf{K}(\theta_0)$.

$$T^{-\frac{1}{2}(1+\alpha)}f_T(\theta_0,Y)'\hat{V}_{ff}(\theta_0)^{-1}[T^{-\frac{1}{2}(1+\nu)}D_T(\theta_0,Y)] = s_0 + T^{-\frac{\kappa}{2}}s_{1\kappa,1} + T^{-\frac{1}{2}(\nu+\kappa-2\alpha)}s_{\nu+\kappa-2\alpha,1} + T^{-\frac{1}{2}(\nu+2(\kappa-\alpha))}s_{\nu+2(\kappa-\alpha),1},$$

with $\kappa = \mu$.

1b. $\mathbf{W}_{2s}(\theta_0)$.

$$[T^{-\frac{1}{2}(1+\nu)}p_T(\hat{\theta}_{2s},Y)]'V_{ff}(\hat{\theta}_{2s})^{-1}[T^{-\frac{1}{2}(1+\nu)}p_T(\hat{\theta}_{2s},Y)] = G_0 + T^{-\frac{\nu}{2}}G_{1\nu,1} + T^{-\frac{\kappa}{2}}G_{1\kappa,1} + T^{-(\nu-\alpha)}G_{2(\nu-\alpha),1} + T^{-\frac{1}{2}(\nu+\kappa)}(G_{\nu+\kappa,1}+G_{\nu+\kappa,2}) + T^{-\frac{2\nu+\kappa-2\alpha}{2}}(G_{2\nu+\kappa-2\alpha,1}+G_{2\nu+\kappa-2\alpha,2}) + T^{-\frac{\nu+2\kappa}{2}}G_{\nu+2\kappa,1} + T^{-(\nu+\kappa-\alpha)}(G_{2(\nu+\kappa-\alpha),1}+G_{2(\nu+\kappa-\alpha),2}) + T^{-\frac{1}{2}(2\nu+3\kappa-\alpha)}G_{2\nu+3\kappa-2\alpha,1}$$

with $\kappa = \min(\nu,\mu)$ and

$$\begin{aligned}
G_0 =\ & D_0'V_{ff}(\theta_0)^{-1}D_0 \\
G_{1\nu,1} =\ & D_0'V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}m_{0,f}]+ \\
& [I_m\otimes V_{ff}(\theta_0)^{-1}m_{0,f}]'[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)]'V_{ff}(\theta_0)^{-1}D_0 \\
G_{1\kappa,1} =\ & T^{\frac{\kappa}{2}}D_0'[V_{ff}(\hat{\theta})^{-1}-V_{ff}(\theta_0)^{-1}]D_0 \\
G_{2(\nu-\alpha),1} =\ & \tfrac{1}{k}[I_m\otimes V_{ff}(\theta_0)^{-1}m_{0,f}]'[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)]'V_{ff}(\theta_0)^{-1} \\
& [A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}m_{0,f}] \\
G_{\nu+\kappa,1} =\ & T^{\frac{\kappa}{2}}D_0'[V_{ff}(\hat{\theta})^{-1}-V_{ff}(\theta_0)^{-1}][A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}m_{0,f}]+ \\
& T^{\frac{\kappa}{2}}[I_m\otimes V_{ff}(\theta_0)^{-1}m_{0,f}]'[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)]'V_{ff}(\theta_0)^{-1}D_0 \\
G_{\nu+\kappa,2} =\ & T^{\frac{1}{2}(\kappa-1)}[D_T(\hat{\theta},Y)-D_T(\theta_0,Y)]'V_{ff}(\theta_0)^{-1}D_0+ \\
& T^{\frac{1}{2}(\nu-1)}D_0'V_{ff}(\theta_0)^{-1}[D_T(\hat{\theta},Y)-D_T(\theta_0,Y)]
\end{aligned}$$

$$\begin{aligned}
G_{2\nu+\kappa-2\alpha,1} =\ & \tfrac{1}{k}T^{\frac{1}{2}\kappa}[I_m\otimes V_{ff}(\theta_0)^{-1}m_{0,f}]'[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)]' \\
& [V_{ff}(\hat{\theta})^{-1}-V_{ff}(\theta_0)^{-1}][A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}m_{0,f}] \\
G_{2\nu+\kappa-2\alpha,2} =\ & \tfrac{1}{k}T^{\frac{1}{2}(\kappa-1)}[D_T(\hat{\theta},Y)-D_T(\theta_0,Y)]'V_{ff}(\theta_0)^{-1}[A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)] \\
& [I_m\otimes V_{ff}(\theta_0)^{-1}m_{0,f}] + \tfrac{1}{k}T^{\frac{1}{2}(\kappa-1)}[I_m\otimes V_{ff}(\theta_0)^{-1}m_{0,f}]' \\
& [A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)]'V_{ff}(\theta_0)^{-1}[D_T(\hat{\theta},Y)-D_T(\theta_0,Y)] \\
G_{\nu+2\kappa,1} =\ & T^{\frac{1}{2}(2\kappa-1)}[D_T(\hat{\theta},Y)-D_T(\theta_0,Y)]'[V_{ff}(\hat{\theta})^{-1}-V_{ff}(\theta_0)^{-1}]D_0+ \\
& T^{\frac{1}{2}(2\kappa-1)}D_0'[V_{ff}(\hat{\theta})^{-1}-V_{ff}(\theta_0)^{-1}][D_T(\hat{\theta},Y)-D_T(\theta_0,Y)] \\
G_{2(\nu+\kappa-\alpha),1} =\ & \tfrac{1}{k}T^{\frac{1}{2}(2\kappa-1)}[D_T(\hat{\theta},Y)-D_T(\theta_0,Y)]'[V_{ff}(\hat{\theta})^{-1}-V_{ff}(\theta_0)^{-1}] \\
& [A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)][I_m\otimes V_{ff}(\theta_0)^{-1}m_{0,f}] + T^{\frac{1}{2}(2\kappa-1)}[I_m\otimes V_{ff}(\theta_0)^{-1}m_{0,f}]' \\
& [A_1V_{\theta f,1}(\theta_0)\cdots A_mV_{\theta f,m}(\theta_0)]'[V_{ff}(\hat{\theta})^{-1}-V_{ff}(\theta_0)^{-1}][D_T(\hat{\theta},Y)-D_T(\theta_0,Y)] \\
G_{2(\nu+\kappa-\alpha),2} =\ & \tfrac{1}{k}T^{(\kappa-1)}[D_T(\hat{\theta},Y)-D_T(\theta_0,Y)]'V_{ff}(\theta_0)^{-1}[D_T(\hat{\theta},Y)-D_T(\theta_0,Y)] \\
G_{2\nu+3\kappa-2\alpha,1} =\ & \tfrac{1}{k}T^{\frac{1}{2}(3\kappa-2\alpha)}[D_T(\hat{\theta},Y)-D_T(\theta_0,Y)]'[V_{ff}(\hat{\theta})^{-1}-V_{ff}(\theta_0)^{-1}][D_T(\hat{\theta},Y)-D_T(\theta_0,Y)]
\end{aligned}$$

Hence,

$$T^{(1+\nu)}[p_T(\hat\theta,Y)'V_{ff}(\hat\theta)^{-1}p_T(\hat\theta,Y)]^{-1}=G_0^{-1}+T^{-\frac{\nu}{2}}Q_1,$$

with

$$Q_1=\ \ -G_0^{-1}[(G_{1\nu,1}+H)^{-1}+T^{-\frac{\nu}{2}}G_0^{-1}]^{-1}G_0^{-1}$$

where $H=T^{-\frac{\kappa-\nu}{2}}G_{1\kappa,1}+T^{-(\frac{1}{2}\nu-\alpha)}G_{2(\nu-\alpha),1}+T^{-\frac{1}{2}\kappa}(G_{\nu+\kappa,1}+G_{\nu+\kappa,2})+T^{-\frac{\nu+\kappa-2\alpha}{2}}(G_{2\nu+\kappa-2\alpha,1}+G_{2\nu+\kappa-2\alpha,2})+T^{-\kappa}G_{\nu+2\kappa,1}+T^{-(\frac{1}{2}\nu+\kappa-\alpha)}(G_{2(\nu+\kappa-\alpha),1}+G_{2(\nu+\kappa-\alpha),2})+T^{-\frac{1}{2}(\nu+3\kappa-\alpha)}G_{2\nu+3\kappa-2\alpha,1}$.

2b. $\mathbf{W}_{cue}(\theta_0)$.

$$[T^{-\frac{1}{2}(1+\nu)}\hat D_T(\hat\theta,Y)]'V_{ff}(\hat\theta)^{-1}[T^{-\frac{1}{2}(1+\nu)}\hat D_T(\hat\theta,Y)]=G_0+T^{-\frac{\kappa}{2}}G_{1\kappa,1}+\\ T^{-\frac{1}{2}(\nu+\kappa)}G_{\nu+\kappa,2}+T^{-\frac{\nu+2\kappa}{2}}G_{\nu+2\kappa,1}+T^{-(\nu+\kappa-\alpha)}G_{2(\nu+\kappa-\alpha),1}+T^{-\frac{1}{2}(2\nu+3\kappa-\alpha)}G_{2\nu+3\kappa-2\alpha,1},$$

with $\kappa=\min(\mu,\nu)$. Hence,

$$T^{(1+\nu)}[\hat D_T(\hat\theta,Y)'V_{ff}(\hat\theta)^{-1}\hat D_T(\hat\theta,Y)]^{-1}=G_0^{-1}+T^{-\frac{\kappa}{2}}Q_1,$$

with

$$Q_1=\ \ -G_0^{-1}[G_{1\kappa,1}+H)^{-1}+T^{-\frac{\nu}{2}}G_0^{-1}]^{-1}G_0^{-1}$$

where $H=T^{-\frac{1}{2}\nu}G_{\nu+\kappa,2}+T^{-\frac{\nu+\kappa}{2}}G_{\nu+2\kappa,1}+T^{-(\nu+\frac{1}{2}\kappa-\alpha)}G_{2(\nu+\kappa-\alpha),1}+T^{-\frac{1}{2}(2\nu+2\kappa-\alpha)}G_{2\nu+3\kappa-2\alpha,1}$.

3b. LM(θ_0):

$$[T^{-\frac{1}{2}(1+\nu)}p_T(\theta_0,Y)]'\hat V_{ff}(\theta_0)^{-1}[T^{-\frac{1}{2}(1+\nu)}p_T(\theta_0,Y)]=G_0+T^{-\frac{\nu}{2}}G_{1\nu,1}+T^{-\frac{\kappa}{2}}G_{1\kappa,1}+\\ T^{-(\nu-\alpha)}G_{2(\nu-\alpha),1}+T^{-\frac{1}{2}(\nu+\kappa)}(G_{\nu+\kappa,1}+G_{\nu+\kappa,2})+T^{-\frac{2\nu+\kappa-2\alpha}{2}}(G_{2\nu+\kappa-2\alpha,1}+G_{2\nu+\kappa-2\alpha,2})+\\ T^{-\frac{\nu+2\kappa}{2}}G_{\nu+2\kappa,1}+T^{-(\nu+\kappa-\alpha)}(G_{2(\nu+\kappa-\alpha),1}+G_{2(\nu+\kappa-\alpha),2})+T^{-\frac{1}{2}(2\nu+3\kappa-\alpha)}G_{2\nu+3\kappa-2\alpha,1}$$

with $\kappa=\mu$. Hence,

$$T^{(1+\nu)}[p_T(\theta_0,Y)'\hat V_{ff}(\theta_0)^{-1}p_T(\theta_0,Y)]^{-1}=G_0^{-1}+T^{-\frac{\nu}{2}}Q_1,$$

with

$$Q_1=\ \ -G_0^{-1}[(G_{1\nu,1}+H)^{-1}+T^{-\frac{\nu}{2}}G_0^{-1}]^{-1}G_0^{-1},$$

with $H=T^{-\frac{\kappa-\nu}{2}}G_{1\kappa,1}+T^{-(\frac{1}{2}\nu-\alpha)}G_{2(\nu-\alpha),1}+T^{-\frac{1}{2}\kappa}(G_{\nu+\kappa,1}+G_{\nu+\kappa,2})+T^{-\frac{\nu+\kappa-2\alpha}{2}}(G_{2\nu+\kappa-2\alpha,1}+G_{2\nu+\kappa-2\alpha,2})+T^{-\kappa}G_{\nu+2\kappa,1}+T^{-(\frac{1}{2}\nu+\kappa-\alpha)}(G_{2(\nu+\kappa-\alpha),1}+G_{2(\nu+\kappa-\alpha),2})+T^{-\frac{1}{2}(\nu+3\kappa-\alpha)}G_{2\nu+3\kappa-2\alpha,1}$.

4b. K(θ_0):

$$[T^{-\frac{1}{2}(1+\nu)}\hat D_T(\theta_0,Y)]'\hat V_{ff}(\theta_0)^{-1}[T^{-\frac{1}{2}(1+\nu)}\hat D_T(\theta_0,Y)]=G_0+T^{-\frac{\kappa}{2}}G_{1\kappa,1}+\\ T^{-\frac{1}{2}(\nu+\kappa)}G_{\nu+\kappa,2}+T^{-\frac{\nu+2\kappa}{2}}G_{\nu+2\kappa,1}+T^{-(\nu+\kappa-\alpha)}G_{2(\nu+\kappa-\alpha),2}+T^{-\frac{1}{2}(2\nu+3\kappa-\alpha)}G_{2\nu+3\kappa-2\alpha,1},$$

with $\kappa=\mu$. Hence,

$$T^{(1+\nu)}\left[\hat D_T(\theta_0,Y)'\hat V_{ff}(\theta_0)^{-1}\hat D_T(\theta_0,Y)\right]^{-1}=G_0^{-1}+T^{-\frac{\kappa}{2}}Q_1$$

and

$$Q_1=\ \ -G_0^{-1}[(G_{1\nu,1}+H)^{-1}+T^{-\frac{\nu}{2}}G_0^{-1}]^{-1}G_0^{-1},$$

with $H = T^{-\frac{1}{2}\nu}G_{\nu+\kappa,2} + T^{-\frac{\nu+\kappa}{2}}G_{\nu+2\kappa,1} + T^{-(\nu+\frac{1}{2}\kappa-\alpha)}G_{2(\nu+\kappa-\alpha),2} + T^{-\frac{1}{2}(2\nu+2\kappa-\alpha)}G_{2\nu+3\kappa-2\alpha,1}$.

1c. $\mathbf{W}_{2s}(\theta_0)$:

$$\begin{aligned}\mathrm{W}_{2s}(\theta_0) = \; & n_0 + T^{-\frac{\nu-2\alpha}{2}}n_{\nu-2\alpha} + T^{-(\nu-\alpha)}n_{2(\nu-\alpha)} + T^{-(\nu-2\alpha)}n_{2(\nu-2\alpha)} + T^{-\frac{\nu}{2}}n_\nu + T^{-\frac{\kappa}{2}}n_\kappa + T^{-\kappa}n_{2\kappa} + \\ & T^{-\frac{1}{2}(\nu+\kappa-2\alpha)}n_{\nu+\kappa-2\alpha} + T^{-\frac{1}{2}(\nu+\kappa)}n_{\nu+\kappa} + T^{-(\nu-\alpha)}n_{2(\nu-\alpha)} + T^{-\frac{1}{2}(\nu+2(\kappa-\alpha))}n_{\nu+2(\kappa-\alpha)} + \\ & T^{-\frac{1}{2}(2\nu+\kappa-2\alpha)}n_{2\nu+\kappa-2\alpha} + T^{-(\nu+\kappa-\alpha)}n_{2(\nu+\kappa-\alpha)} + T^{-\frac{1}{2}\kappa+2(\nu-2\alpha)}n_{\kappa+2(\nu-2\alpha)} + T^{-(\nu+\kappa-2\alpha)}n_{2(\nu+\kappa-2\alpha)},\end{aligned}$$

with $\kappa = \min(\mu,\nu)$,

$$\begin{aligned}n_0 = \; & s_0' G_0^{-1} s_0 \\ n_{\nu-2\alpha} = \; & s_{\nu-2\alpha,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{\nu-2\alpha,1} \\ n_{2(\nu-2\alpha)} = \; & s_{\nu-2\alpha,1}' G_0^{-1} s_{\nu-2\alpha,1} \\ n_\nu = \; & s_0' Q_1 s_0 \\ n_\kappa = \; & s_{1\kappa,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{1\kappa,1} \\ n_{\nu+\kappa-2\alpha} = \; & (s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,2})' G_0^{-1} s_0 + s_0' G_0^{-1}(s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,2}) + \\ & s_{\nu-2\alpha,1}' G_0^{-1} s_{1\kappa,1} + s_{1\kappa,1}' G_0^{-1} s_{\nu-2\alpha,1} \\ n_{2\kappa} = \; & s_{1\kappa,1}' G_0^{-1} s_{1\kappa,1} \\ n_{\nu+\kappa} = \; & s_{1\kappa,1}' Q_1 s_0 + s_0' Q_1 s_{1\kappa,1} \\ n_{2(\nu-\alpha)} = \; & s_{\nu-2\alpha,1}' Q_1 s_0 + s_0' Q_1 s_{\nu-2\alpha,1} \\ n_{\nu+2(\kappa-\alpha)} = \; & (s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,2})' G_0^{-1} s_{1\kappa,1} + s_{1\kappa,1}' G_0^{-1}(s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,2}) + \\ & s_{\nu+2\kappa-2\alpha,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{\nu+2\kappa-2\alpha,1} \\ n_{2\nu+\kappa-2\alpha} = \; & (s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,2})' Q_1 s_0 + s_0' Q_1 (s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,2}) \\ n_{2(\nu+\kappa-\alpha)} = \; & s_{\nu+2(\kappa-\alpha),1}' Q_1 s_0 + s_0' Q_1 s_{\nu+2(\kappa-\alpha),1} + (s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,2})' Q_1 s_{1\kappa,1} + \\ & s_{1\kappa,1}' Q_1 (s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,2}) \\ n_{\kappa+2(\nu-2\alpha)} = \; & s_{\nu-2\alpha,1}' Q_1 s_{\nu-2\alpha,1} \\ n_{2(\nu+\kappa-2\alpha)} = \; & (s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,2})' G_0^{-1} (s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,2})\end{aligned}$$

2c. $\mathbf{W}_{cue}(\theta_0)$:

$$\begin{aligned}\mathrm{W}_{cue}(\theta_0) = \; & n_0 + T^{-\frac{\kappa}{2}}n_\kappa + T^{-\kappa}n_{2\kappa} + T^{-\frac{3}{2}\kappa}n_{3\kappa} + T^{-\frac{1}{2}(\nu+\kappa-2\alpha)}n_{\nu+\kappa-2\alpha} + \\ & T^{-\frac{1}{2}(\nu+\kappa-2\alpha)}n_{\nu+\kappa-2\alpha} + T^{-\frac{1}{2}(\nu+2(\kappa-\alpha))}n_{\nu+2(\kappa-\alpha)}\end{aligned}$$

with $\kappa = \min(\mu,\nu)$,

$$\begin{aligned}n_0 = \; & s_0' G_0^{-1} s_0 \\ n_\kappa = \; & s_0' Q_1 s_0 + s_{1\kappa,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{1\kappa,1} \\ n_{2\kappa} = \; & s_{1\kappa,1}' G_0^{-1} s_{1\kappa,1} + s_{1\kappa,1}' Q_1 s_0 + s_0' Q_1 s_{1\kappa,1} \\ n_{3\kappa} = \; & s_{1\kappa,1}' Q_1 s_{1\kappa,1} \\ n_{\nu+\kappa-2\alpha} = \; & s_{\nu+\kappa-2\alpha,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{\nu+\kappa-2\alpha,1} \\ n_{\nu+2(\kappa-\alpha)} = \; & s_{\nu+\kappa-2\alpha,1}' G_0^{-1} s_{1\kappa,1} + s_{1\kappa,1}' G_0^{-1} s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,1}' Q_1 s_0 + \\ & s_0' Q_1 s_{\nu+\kappa-2\alpha,1} + s_{\nu+2\kappa-2\alpha,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{\nu+2\kappa-2\alpha,1} \\ n_{2(\nu+\kappa-2\alpha)} = \; & s_{\nu+\kappa-2\alpha,1}' G_0^{-1} s_{\nu+\kappa-2\alpha,1}\end{aligned}$$

3c. $\mathbf{LM}(\theta_0)$:

$$\begin{aligned}\mathrm{LM}(\theta_0) = \; & n_0 + T^{-\frac{\nu-2\alpha}{2}}n_{\nu-2\alpha} + T^{-(\nu-\alpha)}n_{2(\nu-\alpha)} + T^{-(\nu-2\alpha)}n_{2(\nu-2\alpha)} + T^{-\frac{\nu}{2}}n_\nu + T^{-\frac{\kappa}{2}}n_\kappa + T^{-\kappa}n_{2\kappa} + \\ & T^{-\frac{1}{2}(\nu+\kappa-2\alpha)}n_{\nu+\kappa-2\alpha} + T^{-\frac{1}{2}(\nu+\kappa)}n_{\nu+\kappa} + T^{-(\nu-\alpha)}n_{2(\nu-\alpha)} + T^{-\frac{1}{2}(\nu+2(\kappa-\alpha))}n_{\nu+2(\kappa-\alpha)} + \\ & T^{-\frac{1}{2}(2\nu+\kappa-2\alpha)}n_{2\nu+\kappa-2\alpha} + T^{-(\nu+\kappa-\alpha)}n_{2(\nu+\kappa-\alpha)} + T^{-\frac{1}{2}\kappa+2(\nu-2\alpha)}n_{\kappa+2(\nu-2\alpha)} + o_p(T^{-(\nu+\kappa-\alpha)})\end{aligned}$$

with $\kappa = \mu$,

$$\begin{aligned}
n_0 =& \; s_0' G_0^{-1} s_0 \\
n_{\nu-2\alpha} =& \; s_{\nu-2\alpha,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{\nu-2\alpha,1} \\
n_{2(\nu-2\alpha)} =& \; s_{\nu-2\alpha,1}' G_0^{-1} s_{\nu-2\alpha,1} \\
n_\nu =& \; s_0' Q_1 s_0 \\
n_\kappa =& \; s_{1\kappa,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{1\kappa,1}
\end{aligned}$$

$$\begin{aligned}
n_{\nu+\kappa-2\alpha} =& \; (s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,2})' G_0^{-1} s_0 + s_0' G_0^{-1} (s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,2}) + \\
& s_{\nu-2\alpha,1}' G_0^{-1} s_{1\kappa,1} + s_{1\kappa,1}' G_0^{-1} s_{\nu-2\alpha,1} \\
n_{2\kappa} =& \; s_{1\kappa,1}' G_0^{-1} s_{1\kappa,1} \\
n_{\nu+\kappa} =& \; s_{1\kappa,1}' Q_1 s_0 + s_0' Q_1 s_{1\kappa,1} \\
n_{2(\nu-\alpha)} =& \; s_{\nu-2\alpha,1}' Q_1 s_0 + s_0' Q_1 s_{\nu-2\alpha,1} \\
n_{\nu+2(\kappa-\alpha)} =& \; (s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,2})' G_0^{-1} s_{1\kappa,1} + s_{1\kappa,1}' G_0^{-1} (s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,2}) + \\
& s_{\nu+2\kappa-2\alpha,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{\nu+2\kappa-2\alpha,1} \\
n_{2\nu+\kappa-2\alpha} =& \; (s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,2})' Q_1 s_0 + s_0' Q_1 (s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,2}) \\
n_{2(\nu+\kappa-\alpha)} =& \; s_{\nu+2(\kappa-\alpha),1}' Q_1 s_0 + s_0' Q_1 s_{\nu+2(\kappa-\alpha),1} + (s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,2})' Q_1 s_{1\kappa,1} + \\
& s_{1\kappa,1}' Q_1 (s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,2}) \\
n_{\kappa+2(\nu-2\alpha)} =& \; s_{\nu-2\alpha,1}' Q_1 s_{\nu-2\alpha,1}
\end{aligned}$$

4c. $\mathbf{K}(\theta_0)$:

$$\begin{aligned}
\mathrm{K}(\theta_0) =& \; n_0 + T^{-\frac{\kappa}{2}} n_\kappa + T^{-\kappa} n_{2\kappa} + T^{-\frac{3}{2}\kappa} n_{3\kappa} + T^{-\frac{1}{2}(\nu+\kappa-2\alpha)} n_{\nu+\kappa-2\alpha} + \\
& T^{-\frac{1}{2}(\nu+\kappa-2\alpha)} n_{\nu+\kappa-2\alpha} + T^{-\frac{1}{2}(\nu+2(\kappa-\alpha))} n_{\nu+2(\kappa-\alpha)}
\end{aligned}$$

with $\kappa = \mu$,

$$\begin{aligned}
n_0 =& \; s_0' G_0^{-1} s_0 \\
n_\kappa =& \; s_0' Q_1 s_0 + s_{1\kappa,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{1\kappa,1} \\
n_{2\kappa} =& \; s_{1\kappa,1}' G_0^{-1} s_{1\kappa,1} + s_{1\kappa,1}' Q_1 s_0 + s_0' Q_1 s_{1\kappa,1} \\
n_{3\kappa} =& \; s_{1\kappa,1}' Q_1 s_{1\kappa,1} \\
n_{\nu+\kappa-2\alpha} =& \; s_{\nu+\kappa-2\alpha,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{\nu+\kappa-2\alpha,1} \\
n_{\nu+2(\kappa-\alpha)} =& \; s_{\nu+\kappa-2\alpha,1}' G_0^{-1} s_{1\kappa,1} + s_{1\kappa,1}' G_0^{-1} s_{\nu+\kappa-2\alpha,1} + s_{\nu+\kappa-2\alpha,1}' Q_1 s_0 + \\
& s_0' Q_1 s_{\nu+\kappa-2\alpha,1} + s_{\nu+2\kappa-2\alpha,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{\nu+2\kappa-2\alpha,1}.
\end{aligned}$$

D. Lemma 3. Convergence of

$$\begin{aligned}
s_{\nu+\kappa-2\alpha,1} =& \; \tfrac{T^{\frac{1}{2}\kappa}}{k} m_{0,f}' V_{ff}(\theta_0)^{-1} \{ [A_1 \hat{V}_{\theta f,1}(\theta_0) \cdots A_m \hat{V}_{\theta f,m}(\theta_0)][I_m \otimes \hat{V}_{ff}(\theta_0)^{-1}] \\
& -[A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)][I_m \otimes V_{ff}(\theta_0)^{-1}] \} [I_m \otimes m_{0,f}] \\
=& \; (s_{\nu+\kappa-2\alpha,1,1} \dots s_{\nu+\kappa-2\alpha,1,m})
\end{aligned}$$

with

$$\begin{aligned}
s_{\nu+\kappa-2\alpha,1,i} =& \; \tfrac{T^{\frac{1}{2}\kappa}}{k} m_{0,f}' V_{ff}(\theta_0)^{-1} A_i [\hat{V}_{\theta f,i}(\theta_0) V_{ff}()^{-1} - V_{\theta f,i}(\theta_0) V_{ff}(\theta_0)^{-1}] m_{0,f} \\
=& \; \tfrac{T^{\frac{1}{2}\kappa}}{k} m_{0,f}' V_{ff}(\theta_0)^{-1} A_i [\hat{V}_{\theta f,i}(\theta_0) - V_{\theta f,i}(\theta_0)] \hat{V}_{ff}(\theta_0)^{-1} m_{0,f} - \\
& \tfrac{T^{\frac{1}{2}\kappa}}{k} m_{0,f}' V_{ff}(\theta_0)^{-1} A_i V_{\theta f,i}(\theta_0) V_{ff}(\theta_0)^{-1} [\hat{V}_{ff}(\theta_0) - V_{ff}(\theta_0)] \hat{V}_{ff}(\theta_0)^{-1} m_{0,f} \\
=& \; \tfrac{1}{k} m_{0,f}' V_{ff}(\theta_0)^{-1} A_i [U_{\theta f,i} - V_{\theta f,i}(\theta_0) V_{ff}(\theta_0)^{-1} U_{ff}] V_{ff}(\theta_0)^{-1} m_{0,f} + o_p(1) \\
=& \; \tfrac{1}{k} m_{0,f}' V_{ff}(\theta_0)^{-1} A_i U_{\theta.f,i} V_{ff}(\theta_0)^{-1} m_{0,f} + o_p(1) \\
=& \; (\tfrac{1}{\sqrt{k}} m_{0,f}' V_{ff}(\theta_0)^{-1} \otimes \tfrac{1}{\sqrt{k}} m_{0,f}' V_{ff}(\theta_0)^{-1} A_i) \mathrm{vec}(U_{\theta.f,i}) + o_p(1).
\end{aligned}$$

where $U_{\theta.f,i} = U_{\theta f,i} - V_{\theta f,i}(\theta_0)U_{ff}$. Because of Assumption 2*,

$$\begin{aligned} s_{\nu+\kappa-2\alpha,1,i} = & \ (\tfrac{1}{\sqrt{k}} m'_{0,f} V_{ff}(\theta_0)^{-1} \otimes \tfrac{1}{\sqrt{k}} m'_{0,f} V_{ff}(\theta_0)^{-1} A_i) \text{vec}(U_{\theta.f,i}) + o_p(1) \\ \underset{d}{\rightarrow} & \ \lambda_i, \end{aligned}$$

with $\lambda_i \sim N(0, \sigma_{ii}(\theta_0)$ and $\sigma_{ij}(\theta_0) = \lim_{k\to\infty}(\frac{1}{\sqrt{k}} m'_{0,f} V_{ff}(\theta_0)^{-1} \otimes \frac{1}{\sqrt{k}} m'_{0,f} V_{ff}(\theta_0)^{-1} A_i)' \bar{W}_{ij}(\theta_0)$ $(\frac{1}{\sqrt{k}} m'_{0,f} V_{ff}(\theta_0)^{-1} \otimes \frac{1}{\sqrt{k}} m'_{0,f} V_{ff}(\theta_0)^{-1} A_j)$, where

$$\begin{aligned} \bar{W}_{ij}(\theta_0) = & \ \lim_{T\to\infty} E[\text{vec}(U_{\theta f,i} - V_{\theta f,i}(\theta_0) V_{ff}(\theta_0)^{-1} U_{ff}) \text{vec}(U_{\theta f,j} - V_{\theta f,j}(\theta_0) V_{ff}(\theta_0)^{-1} U_{ff})'], \\ = & \ E[(\psi_{u,\theta_i f} - V_{\theta f,i}(\theta_0) V_{ff}(\theta_0)^{-1} S_{k_f(m+1)} \psi_{u,ff})(\psi_{u,\theta_i f} - V_{\theta f,i}(\theta_0) V_{ff}(\theta_0)^{-1} S_{k_f(m+1)} \psi_{u,ff})'], \end{aligned}$$

which expression can be further constructed using Assumption 2*.

E. Proof of Theorem 3. The convergence of $S(\theta_0)$ is characterized by

$$S(\theta_0) = \ w_0 + T^{-\frac{\mu}{2}} w_\mu + o_p(T^{-\frac{\mu}{2}}),$$

with $w_0 = m'_{0.f} V_{ff}(\theta_0)^{-1} m_{0.f}$, $w_\mu = T^{\frac{\mu}{2}} m'_{0.f}[\hat{V}_{ff}(\theta_0)^{-1} - V_{ff}(\theta_0)^{-1}] m_{0.f}$. By decomposing $V_{ff}(\theta_0)^{-\frac{1}{2}} m_{0.f}$ as

$$V_{ff}(\theta_0)^{-\frac{1}{2}} m_{0.f} = \ V_{ff}(\theta_0)^{-\frac{1}{2}} D_0 \rho_0 + V_{ff}(\theta_0)^{\frac{1}{2}} D_{0,\perp} \lambda_0$$

with $\rho_0 = (D'_0 V_{ff}(\theta_0)^{-1} D_0)^{-1} D'_0 V_{ff}(\theta_0)^{-1} m_{0.f}$ and $\lambda_0 = (D'_{0,\perp} V_{ff}(\theta_0) D_{0,\perp})^{-1} D'_{0,\perp} m_{0.f}$ and $D_{0,\perp}$: $k_f \times (k_f - m)$, $D'_{0,\perp} D_0 \equiv 0$, $D'_{0,\perp} D_{0,\perp} \equiv I_{k_f - m}$; we can specify the higher order properties of $\text{S}(\theta_0)$ also by

$$S(\theta_0) = \ n_0 + n_{0,\perp} + T^{-\frac{\mu}{2}} w_\mu + o_p(T^{-\frac{\mu}{2}}),$$

with

$$\begin{aligned} n_0 = & \ m'_{0.f} V_{ff}(\theta_0)^{-1} D_0 (D'_0 V_{ff}(\theta_0)^{-1} D_0)^{-1} D'_0 V_{ff}(\theta_0)^{-1} m_{0.f}, \\ n_{0,\perp} = & \ m'_{0.f} D_{0,\perp} (D'_{0,\perp} V_{ff}(\theta_0) D_{0,\perp})^{-1} D'_{0,\perp} m_{0.f}. \end{aligned}$$

The higher order properties of the J-statistics result from substracting the higher order properties from Theorem 1 from the S-statistic. Consequently,

$$\left.\begin{array}{r} \text{J}_{2s}(\theta_0) \\ \text{J}_{cue}(\theta_0) \\ \text{J}_{LM}(\theta_0) \\ \text{J}_{\text{K}}(\theta_0) \end{array}\right\} = n_{0,\perp} + T^{-\frac{\mu}{2}} w_\mu - \left\{ \begin{array}{l} T^{-\frac{\nu}{2}} n_\nu + T^{-\frac{\kappa}{2}} n_\kappa + T^{-\frac{\nu+\kappa}{2}} n_{\nu+\kappa} + T^{-\nu} n_{2\nu} + T^{-\kappa} n_{2\kappa} + \\ T^{-\frac{1}{2}(2\nu+\kappa)} n_{2\nu+\kappa} + T^{-\frac{1}{2}(\nu+2\kappa)} n_{\nu+2\kappa} + T^{-\frac{3}{2}\nu} n_{3\nu} + o_p(T^{-\frac{3}{2}\nu}), \end{array}\right.$$

where all the components for the respective statistics are defined in Theorem 1.

F. Proof of Theorem 4. When we condition on $\frac{1}{\sqrt{T}} \hat{D}_T(\theta_0, Y) = \frac{1}{\sqrt{T}} D_T(\theta_0, Y) + \frac{1}{\sqrt{T}} \left[\hat{D}_T(\theta_0, Y) \ - D_T(\theta_0, Y\right.$ the higher order expansion of $\text{K}(\theta_0)$ can be specified as

$$\text{K}(\theta_0) = n_0 + T^{-\frac{\kappa}{2}}(n_\kappa + n_{\nu+\kappa}) + T^{-\kappa}(n_{2\kappa} + n_{\nu+2\kappa}) + o_p(T^{-\nu}),$$

where

$$\begin{array}{rcl} n_0 & = & s_0' G_0^{-1} s_0 \\ \hat{V}(\theta_0) & : & \left\{ \begin{array}{rl} n_\kappa = & s_0' Q_1 s_0 + s_{1\kappa,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{1\kappa,1} \\ n_{2\kappa} = & s_{1\kappa,1}' G_0^{-1} s_{1\kappa,1} + s_{1\kappa,1}' Q_1 s_0 + s_0' Q_1 s_{1\kappa,1} \end{array} \right. \\ \text{mixed} & : & \left\{ \begin{array}{rl} n_{\nu+\kappa} = & s_{\nu+\kappa,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{\nu+\kappa,1} \\ n_{\nu+2\kappa} = & s_{\nu+\kappa,1}' G_0^{-1} s_{1\kappa,1} + s_{1\kappa,1}' G_0^{-1} s_{\nu+\kappa,1} + s_{\nu+2\kappa,1}' G_0^{-1} s_0 + s_0' G_0^{-1} s_{\nu+2\kappa,1} + \\ & s_0' Q_1 s_{\nu+\kappa,1} + s_{\nu+\kappa,1}' Q_1 s_0. \end{array} \right. \end{array}$$

and

$$\begin{array}{rcl} \text{0-th order} & : & \left\{ s_0 = m_{0,f}' V_{ff}(\theta_0)^{-1} \left[\frac{1}{\sqrt{T}} D_T(\theta_0, Y) \right] \right. \\ \hat{V}(\theta_0) & : & \left\{ s_{1\kappa,1} = \ T^{\frac{\kappa}{2}} m_{0,f}' [\hat{V}_{ff}(\theta_0)^{-1} - V_{ff}(\theta_0)^{-1}] \left[\frac{1}{\sqrt{T}} D_T(\theta_0, Y) \right] \right. \\ \text{mixed} & : & \left\{ \begin{array}{rl} s_{\nu+\kappa,1} = & T^{\frac{1}{2}\kappa} m_{0,f}' V_{ff}(\theta_0)^{-1} \left\{ \frac{1}{\sqrt{T}} [\hat{D}_T(\theta_0, Y) - D_T(\theta_0, Y)] \right\} \\ = & T^{\frac{1}{2}\kappa} m_{0,f}' V_{ff}(\theta_0)^{-1} \{ [A_1 \hat{V}_{\theta f,1}(\theta_0) \cdots A_m \hat{V}_{\theta f,m}(\theta_0)][I_m \otimes \hat{V}_{ff}(\theta_0)^{-1}] \\ & -[A_1 V_{\theta f,1}(\theta_0) \cdots A_m V_{\theta f,m}(\theta_0)][I_m \otimes V_{ff}(\theta_0)^{-1}] \} [I_m \otimes m_{0.f}] \\ s_{\nu+2\kappa,1} = & T^{\kappa} m_{0,f}' [\hat{V}_{ff}(\theta_0)^{-1} - V_{ff}(\theta_0)^{-1}] \left\{ \frac{1}{\sqrt{T}} [\hat{D}_T(\theta_0, Y) - D_T(\theta_0, Y)] \right\}, \end{array} \right. \end{array}$$

with

$$Q_1 = \ -G_0^{-1} [(G_{1\kappa,1} + \tilde{G}_{\nu+\kappa,2} + T^{-\frac{\kappa}{2}} H)^{-1} + T^{-\frac{\kappa}{2}} G_0^{-1}]^{-1} G_0^{-1},$$

where $H = G_{\nu+2\kappa,1} + G_{2(\nu+\kappa),2} + T^{-\frac{1}{2}\kappa} G_{2\nu+3\kappa,1}$ and

$$\begin{array}{rl} G_0 = & \left[\frac{1}{\sqrt{T}} D_T(\theta_0, Y) \right]' V_{ff}(\theta_0)^{-1} \left[\frac{1}{\sqrt{T}} D_T(\theta_0, Y) \right] \\ G_{1\kappa,1} = & T^{\frac{\kappa}{2}} \left[\frac{1}{\sqrt{T}} D_T(\theta_0, Y) \right]' [\hat{V}_{ff}(\theta_0)^{-1} - V_{ff}(\theta_0)^{-1}] \left[\frac{1}{\sqrt{T}} D_T(\theta_0, Y) \right] \\ G_{\nu+\kappa,2} = & T^{\frac{1}{2}\kappa} \left\{ \frac{1}{\sqrt{T}} [\hat{D}_T(\theta_0, Y) - D_T(\theta_0, Y)] \right\} V_{ff}(\theta_0)^{-1} \left[\frac{1}{\sqrt{T}} D_T(\theta_0, Y) \right] + \\ & T^{\frac{1}{2}\kappa} \left[\frac{1}{\sqrt{T}} D_T(\theta_0, Y) \right]' V_{ff}(\theta_0)^{-1} \left\{ \frac{1}{\sqrt{T}} [\hat{D}_T(\theta_0, Y) - D_T(\theta_0, Y)] \right\} \\ G_{\nu+2\kappa,1} = & T^{\kappa} \left\{ \frac{1}{\sqrt{T}} [\hat{D}_T(\theta_0, Y) - D_T(\theta_0, Y)] \right\}' [\hat{V}_{ff}(\theta_0)^{-1} - V_{ff}(\theta_0)^{-1}] \left[\frac{1}{\sqrt{T}} D_T(\theta_0, Y) \right] + \\ & T^{\kappa} \left[\frac{1}{\sqrt{T}} D_T(\theta_0, Y) \right]' [\hat{V}_{ff}(\theta_0)^{-1} - V_{ff}(\theta_0)^{-1}] \left\{ \frac{1}{\sqrt{T}} [\hat{D}_T(\theta_0, Y) - D_T(\theta_0, Y)] \right\} \\ G_{2(\nu+\kappa),2} = & T^{\kappa} \left\{ \frac{1}{\sqrt{T}} [\hat{D}_T(\theta_0, Y) - D_T(\theta_0, Y)] \right\}' V_{ff}(\theta_0)^{-1} \left\{ \frac{1}{\sqrt{T}} [\hat{D}_T(\theta_0, Y) - D_T(\theta_0, Y)] \right\} \\ G_{2\nu+3\kappa,1} = & T^{\frac{3}{2}\kappa} \left\{ \frac{1}{\sqrt{T}} [\hat{D}_T(\theta_0, Y) - D_T(\theta_0, Y)] \right\}' [\hat{V}_{ff}(\theta_0)^{-1} - V_{ff}(\theta_0)^{-1}] \\ & \left\{ \frac{1}{\sqrt{T}} [\hat{D}_T(\theta_0, Y) - D_T(\theta_0, Y)] \right\}. \end{array}$$

The higher order approximation of K(θ_0) conditional of $\frac{1}{\sqrt{T}} \hat{D}_T(\theta_0, Y)$ can be used to construct the higher order approximation of the bootstrap realizations of K(θ). The bootstrap samples are obtained by independent draws with replacement $((\tilde{f}_t(\theta)' \ \tilde{q}_t(\theta_0))$ from the population $\{(\bar{f}_t(\theta_0)' \ q_t(\theta_0)),\ t = 1, \ldots, T\}$.

The expressions of the bootstrap realizations of $\mathrm{K}(\theta_0)$ read:

$$\begin{aligned}\tilde{\mathrm{K}}(\theta_0) = & \tfrac{1}{T}\tilde{f}_T(\theta_0,Y)'\tilde{V}_{ff}(\theta_0)^{-1}\tilde{D}_T(\theta_0,Y)\left[\tilde{D}_T(\theta_0,Y)'\tilde{V}_{ff}(\theta_0)^{-1}\tilde{D}_T(\theta_0,Y)\right]^{-1} \\ & \tilde{D}_T(\theta_0,Y)'\tilde{V}_{ff}(\theta_0)^{-1}\tilde{f}_T(\theta_0,Y),\end{aligned}$$

with $\tilde{f}_T(\theta_0,Y)=\sum_{t=1}^T\tilde{f}_T(\theta_0)$,

$$\begin{aligned}\tilde{D}_T(\theta_0,Y) = & [\textstyle\sum_{t=1}^T\tilde{p}_{1,t}(\theta_0)-A_1\tilde{V}_{\theta f,1}(\theta_0)\tilde{V}_{ff}(\theta_0)^{-1}\tilde{f}_T(\theta_0,Y)\ldots \\ & \textstyle\sum_{t=1}^T\tilde{p}_{m,t}(\theta_0)-A_m\tilde{V}_{\theta f,m}(\theta_0)\tilde{V}_{ff}(\theta_0)^{-1}\tilde{f}_T(\theta_0,Y)]\end{aligned}$$

and $\tilde{V}_{\theta f}(\theta_0)$ and $\tilde{V}_{ff}(\theta_0)$ are covariance matrix estimators that are based on the bootstrap realizations $((\tilde{f}_t(\theta_0)'\ \tilde{q}_t(\theta_0)')',\ t=1,\ldots,T)$. Because $\tilde{V}_{\theta f,i}(\theta_0)\underset{p}{\to}\hat{V}_{\theta f,i}(\theta_0)$, $\tilde{V}_{ff}(\theta_0)\underset{p}{\to}\hat{V}_{ff}(\theta_0)$, we obtain that

$$\tfrac{1}{\sqrt{T}}\left[\textstyle\sum_{t=1}^T\tilde{q}_{i,t}(\theta_0)-\tilde{V}_{\theta f,i}(\theta_0)\tilde{V}_{ff}(\theta_0)^{-1}\tilde{f}_T(\theta_0,Y)-\hat{E}(q_{i,t}(\theta_0))\right]=\ m_{0.f,i}+O_p(\tfrac{1}{\sqrt{T}}),$$

with $m_{0.f}=(m_{0.f,1}\ldots m_{0.f,m})$,

$$m_{0.f}\underset{d}{\to}\phi_{\theta.f},$$

with $\phi_{\theta.f}\sim N(0,\hat{V}_{\theta\theta.f}(\theta_0))$, $\hat{V}_{\theta\theta.f}(\theta_0)=\hat{V}_{\theta\theta}(\theta_0)-\hat{V}_{\theta f}(\theta_0)\hat{V}_{ff}(\theta_0)^{-1}\hat{V}_{\theta f}(\theta_0)$, and independent from ϕ_f.

Given $\tilde{D}_T(\theta_0,Y)$, the higher order expansion of the bootstrapped $\tilde{\mathrm{K}}(\theta_0)$ directly results from the higher order expansion of $\mathrm{K}(\theta_0)$:

$$\tilde{\mathrm{K}}(\theta_0)=\tilde{n}_0+T^{-\frac{\kappa}{2}}(\tilde{n}_\kappa+\tilde{n}_{\nu+\kappa})+T^{-\kappa}(\tilde{n}_{2\kappa}+\tilde{n}_{\nu+2\kappa})+o_p(T^{-\nu}),$$

where

$$\begin{aligned}\tilde{n}_0 &= \tilde{s}_0'\tilde{G}_0^{-1}\tilde{s}_0 \\ \tilde{V}(\theta_0)\ &:\ \begin{cases}\tilde{n}_\kappa = & \tilde{s}_0'\tilde{Q}_1\tilde{s}_0+\tilde{s}_{1\kappa,1}'\tilde{G}_0^{-1}\tilde{s}_0+\tilde{s}_0'\tilde{G}_0^{-1}\tilde{s}_{1\kappa,1}\\ \tilde{n}_{2\kappa} = & \tilde{s}_{1\kappa,1}'\tilde{G}_0^{-1}\tilde{s}_{1\kappa,1}+\tilde{s}_{1\kappa,1}'\tilde{Q}_1\tilde{s}_0+\tilde{s}_0'\tilde{Q}_1\tilde{s}_{1\kappa,1}\end{cases}\\ \text{mixed}\ &:\ \begin{cases}\tilde{n}_{\nu+\kappa} = & \tilde{s}_{\nu+\kappa,1}'\tilde{G}_0^{-1}\tilde{s}_0+\tilde{s}_0'\tilde{G}_0^{-1}\tilde{s}_{\nu+\kappa,1}\\ \tilde{n}_{\nu+2\kappa} = & \tilde{s}_{\nu+\kappa,1}'\tilde{G}_0^{-1}\tilde{s}_{1\kappa,1}+\tilde{s}_{1\kappa,1}'\tilde{G}_0^{-1}\tilde{s}_{\nu+\kappa,1}+\tilde{s}_{\nu+2\kappa,1}'\tilde{G}_0^{-1}\tilde{s}_0+\tilde{s}_0'\tilde{G}_0^{-1}\tilde{s}_{\nu+2\kappa,1}+\\ & \tilde{s}_0'\tilde{Q}_1\tilde{s}_{\nu+\kappa,1}+\tilde{s}_{\nu+\kappa,1}'\tilde{Q}_1\tilde{s}_0.\end{cases}\end{aligned}$$

and

$$\begin{aligned}\text{0-th order}\ &:\ \left\{\ \tilde{s}_0{=}\tilde{m}_{0,f}'\hat{V}_{ff}(\theta_0)^{-1}\left[\tfrac{1}{\sqrt{T}}\hat{D}_T(\theta_0,Y)\right]\right.\\ \hat{V}(\theta_0)\ &:\ \left\{\ \tilde{s}_{1\kappa,1} = T^{\frac{\kappa}{2}}\tilde{m}_{0,f}'[\tilde{V}_{ff}(\theta_0)^{-1}-\hat{V}_{ff}(\theta_0)^{-1}]\left[\tfrac{1}{\sqrt{T}}\hat{D}_T(\theta_0,Y)\right]\right.\\ \text{mixed}\ &:\ \begin{cases}\tilde{s}_{\nu+\kappa,1} = & T^{\frac{1}{2}\kappa}\tilde{m}_{0,f}'\hat{V}_{ff}(\theta_0)^{-1}\left\{\tfrac{1}{\sqrt{T}}[\tilde{D}_T(\theta_0,Y)-\hat{D}_T(\theta_0,Y)]\right\}\\ = & T^{\frac{1}{2}\kappa}\tilde{m}_{0,f}'\hat{V}_{ff}(\theta_0)^{-1}\{[A_1\tilde{V}_{\theta f,1}(\theta_0)\cdots A_m\tilde{V}_{\theta f,m}(\theta_0)][I_m\otimes\tilde{V}_{ff}(\theta_0)^{-1}]\\ & -[A_1\hat{V}_{\theta f,1}(\theta_0)\cdots A_m\hat{V}_{\theta f,m}(\theta_0)][I_m\otimes\hat{V}_{ff}(\theta_0)^{-1}]\}[I_m\otimes\tilde{m}_{0.f}]\\ \tilde{s}_{\nu+2\kappa,1} = & T^{\kappa}\tilde{m}_{0,f}'[\tilde{V}_{ff}(\theta_0)^{-1}-\hat{V}_{ff}(\theta_0)^{-1}]\left\{\tfrac{1}{\sqrt{T}}[\tilde{D}_T(\theta_0,Y)-\hat{D}_T(\theta_0,Y)]\right\},\end{cases}\end{aligned}$$

with

$$Q_1 = \quad -\tilde{G}_0^{-1}[(\tilde{G}_{1\kappa,1} + \tilde{G}_{\nu+\kappa,2} + T^{-\frac{\kappa}{2}}\tilde{H})^{-1} + T^{-\frac{\kappa}{2}}\tilde{G}_0^{-1}]^{-1}\tilde{G}_0^{-1},$$

where $H = \tilde{G}_{\nu+2\kappa,1} + \tilde{G}_{2(\nu+\kappa),2} + T^{-\frac{1}{2}\kappa}\tilde{G}_{2\nu+3\kappa,1}$ and

$$\begin{aligned}
\tilde{G}_0 = & \left[\tfrac{1}{\sqrt{T}}\hat{D}_T(\theta_0, Y)\right]' \hat{V}_{ff}(\theta_0)^{-1}\left[\tfrac{1}{\sqrt{T}}\hat{D}_T(\theta_0, Y)\right] \\
\tilde{G}_{1\kappa,1} = & T^{\frac{\kappa}{2}}\left[\tfrac{1}{\sqrt{T}}\hat{D}_T(\theta_0, Y)\right]' [\tilde{V}_{ff}(\theta_0)^{-1} - \hat{V}_{ff}(\theta_0)^{-1}]\left[\tfrac{1}{\sqrt{T}}\hat{D}_T(\theta_0, Y)\right] \\
\tilde{G}_{\nu+\kappa,2} = & T^{\frac{1}{2}\kappa}\left\{\tfrac{1}{\sqrt{T}}[\tilde{D}_T(\theta_0, Y) - \hat{D}_T(\theta_0, Y)]\right\}\hat{V}_{ff}(\theta_0)^{-1}\left[\tfrac{1}{\sqrt{T}}\hat{D}_T(\theta_0, Y)\right] + \\
& T^{\frac{1}{2}\kappa}\left[\tfrac{1}{\sqrt{T}}\hat{D}_T(\theta_0, Y)\right]' \hat{V}_{ff}(\theta_0)^{-1}\left\{\tfrac{1}{\sqrt{T}}[\tilde{D}_T(\theta_0, Y) - \hat{D}_T(\theta_0, Y)]\right\} \\
\tilde{G}_{\nu+2\kappa,1} = & T^{\kappa}\left\{\tfrac{1}{\sqrt{T}}[\tilde{D}_T(\theta_0, Y) - \hat{D}_T(\theta_0, Y)]\right\}' [\tilde{V}_{ff}(\theta_0)^{-1} - \hat{V}_{ff}(\theta_0)^{-1}]\left[\tfrac{1}{\sqrt{T}}\hat{D}_T(\theta_0, Y)\right] + \\
& T^{\kappa}\left[\tfrac{1}{\sqrt{T}}\hat{D}_T(\theta_0, Y)\right]' [\tilde{V}_{ff}(\theta_0)^{-1} - \hat{V}_{ff}(\theta_0)^{-1}]\left\{\tfrac{1}{\sqrt{T}}[\tilde{D}_T(\theta_0, Y) - \hat{D}_T(\theta_0, Y)]\right\} \\
\tilde{G}_{2(\nu+\kappa),2} = & T^{\kappa}\left\{\tfrac{1}{\sqrt{T}}[\tilde{D}_T(\theta_0, Y) - \hat{D}_T(\theta_0, Y)]\right\}' \hat{V}_{ff}(\theta_0)^{-1}\left\{\tfrac{1}{\sqrt{T}}[\tilde{D}_T(\theta_0, Y) - \hat{D}_T(\theta_0, Y)]\right\} \\
\tilde{G}_{2\nu+3\kappa,1} = & T^{\frac{3}{2}\kappa}\left\{\tfrac{1}{\sqrt{T}}[\tilde{D}_T(\theta_0, Y) - \hat{D}_T(\theta_0, Y)]\right\}' [\tilde{V}_{ff}(\theta_0)^{-1} - \hat{V}_{ff}(\theta_0)^{-1}] \\
& \left\{\tfrac{1}{\sqrt{T}}[\tilde{D}_T(\theta_0, Y) - \hat{D}_T(\theta_0, Y)]\right\}.
\end{aligned}$$

Identical to the limit behavior of the zero-th order term of the higher order approximation of $\mathrm{K}(\theta_0)$, the zero-th order term of $\tilde{\mathrm{K}}(\theta_{\mathrm{cue}})$ converges to a $\chi^2(m)$ distributed random variable for all possible values of $\hat{E}(q(\theta_0))$ whenever the number of instruments is fixed.

The higher order approximations of $\mathrm{K}(\theta_0)$ and $\tilde{\mathrm{K}}(\theta_{\mathrm{cue}})$ can be used to obtain the 2κ-th order Edgeworth approximations of the distribution of these statistics by using, see Rothenberg (1984),

$$\begin{aligned}
\Pr[\mathrm{K}(\theta_0) & \leq s] = F_{n_0}[s - T^{-\frac{1}{2}\kappa}n_\kappa(s) + T^{-\kappa}n_{2\kappa}(s) + O_p(T^{-\frac{3}{2}\kappa})], \\
\Pr[\tilde{\mathrm{K}}(\theta_0) & \leq s] = F_{\tilde{n}_0}[s - T^{-\frac{1}{2}\kappa}\tilde{n}_\kappa(s) + T^{-\kappa}\tilde{n}_{2\kappa}(s) + O_p(T^{-\frac{3}{2}\kappa})],
\end{aligned}$$

with

$$\begin{aligned}
n_\kappa(s) = & \lim\nolimits_{T\to\infty} E(n_\kappa + n_{\nu+\kappa}|n_0 = s) \\
n_{2\kappa}(s) = & \lim\nolimits_{T\to\infty}\left\{\tfrac{1}{2}\{2n_k(s)\left[\tfrac{\partial}{\partial s}n_k(s)\right] + \left[\tfrac{\partial}{\partial s}\log(f_{n_0}(s))\right] var(n_\kappa + n_{\nu+\kappa}|n_0 = s) + \right. \\
& \left.\left[\tfrac{\partial}{\partial s}\mathrm{var}(n_\kappa + n_{\nu+\kappa}|n_0 = s)\right] - 2E(n_{2\kappa} + n_{\nu+2\kappa}|n_0 = s)\}\right\} \\
\tilde{n}_\kappa(s) = & \lim\nolimits_{T\to\infty} E(\tilde{n}_\kappa + \tilde{n}_{\nu+\kappa}|\tilde{n}_0 = s), \\
\tilde{n}_{2\kappa}(s) = & \lim\nolimits_{T\to\infty}\left\{\tfrac{1}{2}\{2\tilde{n}_k(s)\left[\tfrac{\partial}{\partial s}\tilde{n}_k(s)\right] + \left[\tfrac{\partial}{\partial s}\log(f_{\tilde{n}_0}(s))\right] var(\tilde{n}_\kappa + \tilde{n}_{\nu+\kappa}|\tilde{n}_0 = s) + \right. \\
& \left.\left[\tfrac{\partial}{\partial s}\mathrm{var}(\tilde{n}_\kappa + \tilde{n}_{\nu+\kappa}|\tilde{n}_0 = s)\right] - 2E(\tilde{n}_{2\kappa} + \tilde{n}_{\nu+2\kappa}|\tilde{n}_0 = s)\}\right\},
\end{aligned}$$

and F_{n_0}, f_{n_0}, $F_{\tilde{n}_0}$, $f_{\tilde{n}_0}$ are the distribution and density function that belong to the limit behavior of n_0 and $\tilde{n}_0$ resp., *i.e.* the distribution and density function of a $\chi^2(m)$ distributed random variable.

Depending on if $n_\kappa(s) = 0$ or not, which is identical to independence of the limit behavior of $T^{-\frac{1}{2}\kappa}(\hat{V}(\theta_0) - V(\theta_0))$ and m_0, we can construct two different Taylor approximations of the distribution of $\mathrm{K}(\theta_0)$ and $\tilde{\mathrm{K}}(\theta_0)$:

1. $n_\kappa(s) = 0$:

$$\begin{aligned}
\Pr[\mathrm{K}(\theta_0) & \leq s] = F_{n_0}(s) + T^{-\kappa}f_{n_0}(s)[n_{2\kappa}(s) + O_p(T^{-\frac{1}{2}\kappa})], \\
\Pr[\tilde{\mathrm{K}}(\theta_0) & \leq s] = F_{n_0}(s) + T^{-\kappa}f_{n_0}(s)[\tilde{n}_{2\kappa}(s) + O_p(T^{-\frac{1}{2}\kappa})],
\end{aligned}$$

2. $n_\kappa(s) \neq 0$:

$$\Pr[\mathrm{K}(\theta_0) \leq s] = F_{s_0}(s) - T^{-\frac{1}{2}\kappa} f_{s_0}(s)[n_\kappa(s) - T^{-\frac{1}{2}\kappa} n_{2\kappa}(s) - O_p(T^{-\kappa})],$$
$$\Pr[\tilde{\mathrm{K}}(\theta_0) \leq s] = F_{s_0}(s) - T^{-\frac{1}{2}\kappa} f_{s_0}(s)[\tilde{n}_\kappa(s) - T^{-\frac{1}{2}\kappa} \tilde{n}_{2\kappa}(s) - O_p(T^{-\kappa})],$$

where we used that $F_{s_0}(s)$ and $F_{\tilde{s}_0}(s)$, $f_{n_0}(s)$ and $f_{\tilde{n}_0}(s)$ are identical.

We subtract $\Pr[\tilde{\mathrm{K}}(\theta_0) \leq s]$ from $\Pr[\mathrm{K}(\theta_0) \leq s]$ to obtain:

1. $n_\kappa(s) = 0$:

$$\Pr[\mathrm{K}(\theta_0) \leq s] - \Pr[\tilde{\mathrm{K}}(\theta_0) \leq s] = T^{-\kappa} f_{n_0}(s)[n_{2\kappa}(s) - \tilde{n}_{2\kappa}(s) + O_p(T^{-\frac{1}{2}\kappa})],$$

2. $n_\kappa(s) \neq 0$:

$$\Pr[\mathrm{K}(\theta_0) \leq s] - \Pr[\tilde{\mathrm{K}}(\theta_0) \leq s] = T^{-\frac{1}{2}\kappa} f_{s_0}(s)[n_\kappa(s) - \tilde{n}_\kappa(s) - T^{-\frac{1}{2}\kappa}(n_{2\kappa}(s) - \tilde{n}_{2\kappa}(s)) - O_p(T^{-\kappa})].$$

The Edgeworth approximation of the finite sample distribution show that the limit distribution of $\tilde{\mathrm{K}}(\theta_0)$ converges to the limit distribution of $\mathrm{K}(\theta_0)$ in all instances. Because the empirical distribution

The empirical counterpart of $\mathrm{H}_0 : \theta = \theta_0$ for which $E(D_T(\theta_0, Y)' V_{ff}(\theta_0)^{-1} f_T(\theta_0, Y)) = E(\lim_{T \to \infty}[\sqrt{T} J_\theta(\theta_0) + (A_1 m_{\theta_1.f} \dots A_m m_{\theta_m.f})]' V_{ff}(\theta_0)^{-1} m_f) = 0$ is $\mathrm{H}_0^* : \theta = \theta_0$ for which $\hat{E}(\hat{D}_T(\theta_0, Y)' \hat{V}_{ff}(\theta_0)^{-1} f_T(\theta_0, Y)) = 0$. The empirical distribution of $(f_t(\theta_0)' \; q_t(\theta_0)')$ is therefore $\{(\tilde{f}_t(\theta_0)' \; \tilde{q}_t(\theta_0)), t = 1, \dots, T\}$ and converges to the unconditional distribution of $(f_t(\theta_0)' \; q_t(\theta_0)')$. The Edgeworth approximations of $\mathrm{K}(\theta_0)$ and $\tilde{\mathrm{K}}(\theta_0)$ are with respect to the unconditional and empirical distribution of $(f_t(\theta_0)' \; q_t(\theta_0)')$. All higher order elements of the Edgeworth approximation of the distribution of $\tilde{\mathrm{K}}(\theta_0)$ therefore converge to the respective higher order elements of the Edgeworth approximation of the distribution of $\mathrm{K}(\theta_0)$ and their converge speed is $T^{-\frac{1}{2}\kappa}$. Hence, $\tilde{n}_{2\kappa}(s) = n_{2\kappa}(s) + O_p(T^{-\frac{1}{2}\kappa})$ and $\tilde{n}_\kappa(s) = n_\kappa(s) + O_p(T^{-\frac{1}{2}\kappa})$ and we obtain the approximations of the finite sample distribution of $\mathrm{K}(\theta_0)$:

1. $n_\kappa(s) = 0$:

$$\Pr[\mathrm{K}(\theta_0) \leq s] - \Pr[\tilde{\mathrm{K}}(\theta_0) \leq s] = O_p(T^{-\frac{3}{2}\kappa}),$$

2. $n_\kappa(s) \neq 0$:

$$\Pr[\mathrm{K}(\theta_0) \leq s] - \Pr[\tilde{\mathrm{K}}(\theta_0) \leq s] = O_p(T^{-\kappa}),$$

which shows the improved approximation of the finite sample distribution of $\mathrm{K}(\theta_0)$ that results from the bootstrap compared to the approximation that results from the limiting distribution.

References

[1] Anderson, T.W. and C. Hsiao. Estimation of Dynamic Models with Error Components. *Journal of the American Statistical Association*, **76**:598–606, 1981.

[2] Andrews, D.W.K. Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica*, **59**:817–858, 1991.

[3] Arellano, M. and B. Honoré. Panel Data Models: Some Recent Developments. In J.L. Heckman and E. Leamer, editors, *Handbook of Econometrics*, volume **5**, chapter 53. Elsevier North-Holland, 2001.

[4] Arellano, M. and S. Bond. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *Review of Economic Studies*, **58**:277–297, 1991.

[5] Bekker, P.A. Alternative Approximations to the Distributions of Instrumental Variable Estimators. *Econometrica*, **62**:657–681, 1994.

[6] Bekker, P.A. and F. Kleibergen. Finite-Sample Instrumental Variables Inference using an Asymptotically Pivotal Statistic. *Econometric Theory*, 2003. Forthcoming.

[7] Bhattacharya, R.N. and J.K. Ghosh. On the Validity of the Formal Edgeworth Expansion. *Annals of Statistics*, **6**:434–451, 1978.

[8] Brown, B.W. and W.K. Newey. Efficient Semiparametric Estimation of Expectations. *Econometrica*, **66**:453–464, 1998.

[9] Donald, S.G. and W.K. Newey. A Jackknife Interpretation of the Continuous Updating Estimator. *Economics Letters*, **67**:239–243, 2000.

[10] Götze, F. and C. Hipp. Asymptotic Expansions for Sums of Weakly Dependent Random Vectors. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **64**:211–239, 1983.

[11] Hall, P. and J.L. Horowitz. Bootstrap critical values for tests based on generalized-method-of-moments estimators. *Econometrica*, **64**:891–916, 1996.

[12] Hansen, L.P. Large Sample Properties of Generalized Method Moments Estimators. *Econometrica*, **50**:1029–1054, 1982.

[13] Hansen, L.P., J. Heaton and A. Yaron. Finite Sample Properties of Some Alternative GMM Estimators. *Journal of Business and Economic Statistics*, **14**:262–280, 1996.

[14] Horowitz, J.L. The Bootstrap. In J.L. Heckman and E. Leamer, editors, *Handbook of Econometrics*, volume 5, chapter 52, pages 3159–3228. Elsevier Science B.V., 2002.

[15] Kleibergen, F. Testing Parameters in GMM without assuming that they are identified. Tinbergen Institute Discussion Paper TI 2001-67/4, 2001.

[16] Kleibergen, F. Pivotal Statistics for testing Structural Parameters in Instrumental Variables Regression. *Econometrica*, **70**:1781–1803, 2002.

[17] Kleibergen, F. Two Independent Pivotal Statistics that Test Location and Misspecification and add-up to the Anderson-Rubin Statistic. *Tinbergen Institute Discussion Paper*, TI 2002-064/4, 2002.

[18] Moreira, M.J.,. A Conditional Likelihood Ratio Test for Structural Models. *Econometrica*, **71**:1027–1048, 2003.

[19] Nagar, A.L. The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations. *Econometrica*, **27**:573–595, 1959.

[20] Newey, W.K. and D. McFadden. Large Sample Estimation and Hypothesis Testing. In R. Engle and D. McFadden, editor, *Handbook of Econometrics, Volume 4*, chapter 36, pages 2113–2148. Elsevier Science B.V., 1994.

[21] Newey, W.K. and K.D. West. A Simple Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. *Econometrica*, **55**:703–708, 1987.

[22] Newey, W.K. and K.D. West. Hypothesis Testing with Efficient Method of Moments Estimation. *International Economic Review*, **28**:777–787, 1987.

[23] Newey, W.K. and R.J. Smith. Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators. *Econometrica*, **72**:219–255, 2004.

[24] Nickell, S.J. Biases in dynamic models with fixed effects. *Econometrica*, **49**:1417–1426, 1981.

[25] Phillips, P.C.B. Partially Identified Econometric Models. *Econometric Theory*, **5**:181–240, (1989).

[26] Phillips, P.C.B. and J.Y. Park. On the Formulation of Wald Tests of Nonlinear Restrictions. *Econometrica*, **56**:1065–1083, 1988.

[27] Phillips, P.C.B. and R. Moon. Linear Regression Limit Theory for Nonstationary Panel Data. *Econometrica*, **67**:1057–1112, 1999.

[28] Rothenberg, T.J. Approximating the Distributions of Econometric Estimators and Test Statistics. In Z. Griliches and M.D. Intrilligator, editor, *Handbook of Econometrics, Volume 2*, chapter 15, pages 881–935. Elsevier Science B.V., 1984.

[29] Sargan, J.D. Some Approxmations to the Distributions of Econometric Criteria which are Asymptotically distributed as Chi-Squared. *Econometrica*, **48**:1107–1138, 1980.

[30] Staiger, D. and J.H. Stock. Instrumental Variables Regression with Weak Instruments. *Econometrica*, **65**:557–586, 1997.

[31] Stock, J.H. and J.H. Wright. GMM with Weak Identification. *Econometrica*, **68**:1055–1096, 2000.

[32] Stock, J.H. and M. Yogo. *Testing for Weak Instruments in Linear IV Regression*. 2001. Prepared for Festschrift in honor of Thomas Rothenberg.

[33] Stock, J.H., J.H. Wright and M. Yogo. A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments. *Journal of Business and Economic Statistics*, **20**:518–529, 2002.

[34] White, H. A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroscedasticity. *Econometrica*, **48**:817–838, 1980.

[35] White, H. *Asymptotic Theory for Econometricians.* Academic Press, 1984.