UvA ECONOMETRICS

Discussion Paper: 2006/09

Expansions of GMM statistics and the bootstrap

Frank Kleibergen

www.fee.uva.nl/ke/UvA-Econometrics

Amsterdam School of Economics
Department of Quantitative Economics
Roetersstraat 11
1018 WB AMSTERDAM
The Netherlands

UvA UNIVERSITEIT VAN AMSTERDAM

Expansions of GMM statistics and the bootstrap

Frank Kleibergen*

Brown University and University of Amsterdam

August 2006

Abstract

We construct higher order expressions of (weak instrument robust) generalized method of moment (GMM) test statistics. We use these expressions to obtain Edgeworth approximations of their finite sample distributions and to show the sensitivity to instrument quality. The Edgeworth approximations show that usage of bootstrapped critical values that result from resampling the moment conditions reduces the order of the approximation error of the finite sample distribution of the weak instrument robust statistics compared to usage of the asymptotic critical values. These results even hold when the instruments are weak and thus extend previously known results on the bootstrap and the Edgeworth approximation. We illustrate the resulting reduction of the size distortions and conduct a power comparison using a panel autoregressive model of order one.

JEL classification: C11, C20, C30

1 Introduction

Two common approaches for reducing size distortions of test statistics are to Edgeworth-correct the asymptotic critical values, see *e.g.* Rothenberg (1984), and usage of bootstrapped critical values, see *e.g.* Horowitz (2001). Both of these approaches remove the approximation error of the finite sample distribution of the test statistics up to a higher order in the sample size than the limiting distribution. This leads to a reduction of the size distortion when we use critical values that result from these procedures compared to the asymptotic ones. The regularity conditions under which the Edgeworth approximation, see *e.g.* Bhattacharya and Ghosh (1978), and the bootstrap, see *e.g.* Horowitz (2001), improve the approximation of the finite sample distribution request that the hypothesized parameters are well identified. Parameters that are estimated in the generalized method of moments (GMM) with weak instruments are not well identified, see *e.g.* Staiger and Stock (1997) and Stock and Wright (2000). Thus the Edgeworth approximation and bootstrap fail to improve the approximation of the finite sample distributions of test statistics in GMM with weak instruments.

*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.econ.brown.edu/fac/Frank_Kleibergen. The research in this paper is partly funded by the NWO grant "Empirical Comparison of Economic Models".

To overcome size distortions in GMM with weak instruments, test statistics have been proposed whose limiting distributions are robust to instrument quality, see *e.g.* Stock and Wright (2000) and Kleibergen (2005). The limiting distributions of these statistics apply under more general conditions than those of the traditional GMM statistics. These limiting distributions therefore typically lead to a better approximation of the finite sample distribution than for the traditional GMM statistics. We further improve the approximation of the finite sample distribution of the weak instrument robust GMM statistics by constructing Edgeworth-corrections of the asymptotic critical values and an algorithm for obtaining bootstrapped critical values. We show that usage of either the Edgeworth-corrected critical values or the bootstrapped critical values reduces the order of the approximation error of the finite sample distribution. These improvements remain to hold when the instruments are weak and show that Edgeworth approximations can be constructed even when the parameters are weakly identified. This provides an extension to the results put forward in Bhattacharya and Ghosh (1978).

The paper is organized as follows. In the second section, we introduce GMM and define our GMM statistics of interest: the S-statistic of Stock and Wright (2000), the GMM Lagrange multiplier (LM) statistic of Kleibergen (2005), a GMM extension of Moreira's (2003) conditional likelihood ratio statistic and the GMM LM statistic of Newey and West (1987). The second section also makes the assumptions under which we derive our results. In the third section, we decompose the statistics into several components that are of a different order in the sample size. The fourth section provides an algorithm to bootstrap our statistics of interest and decomposes the bootstrapped statistics into several components that are of a different order in the bootstrap sample size. The fifth section discusses Edgeworth approximations to the finite sample distributions of the orginal statistics and their bootstrapped counterparts. The Edgeworth approximations show the higher order improvement that results from the bootstrap. The sixth section illustrates the theoretical results and conducts a simulation experiment using a panel autoregressive model of order one. It shows that usage of bootstrapped or Edgeworth-corrected critical values reduces the size distortion compared usage of critical values that stem from the limiting distribution. The sixth section also conducts a power comparison. The seventh section briefly discusses some further extensions. The eight section concludes.

Throughout the paper we use the notation: I_m is the $m \times m$ identity matrix, $P_A = A(A'A)^{-1}A'$ for a full rank $n \times m$ matrix V and $M_A = I_n - P_A$. Furthermore, "$\underset{p}{\rightarrow}$" stands for convergence in probability, "$\underset{d}{\rightarrow}$" for convergence in distribution and E is the expectation operator.

2 Generalized Method of Moments

We consider the estimation of a scalar parameter θ whose parameter region is $\mathbb{R}$ and for which the $k \times 1$ dimensional moment equation

$$\mathrm{E}(f(\theta, Y_i)) = 0, \qquad i = 1, \dots, N, \tag{1}$$

holds. We use a scalar parameter instead of a vector of parameters to simplify the analysis. We later show how the results extend to the multiple parameter case. The data vector Y_i is observed at individual/time i. The number of equations k exceeds or is equal to the number of parameters. The $k \times 1$ dimensional vector function f of θ is finite for finite values of θ, continuous and twice continuous differentiable. The unique value of θ, at which (1) holds, is equal to θ_0. To estimate θ in (1), we use Hansen's (1982) GMM.

For a data-set $(Y_i,\ i = 1, \ldots, N)$, the objective function for the continuous updating estimator (CUE) reads

$$Q(\theta) = N f_N(\theta, Y)' \hat{V}_{ff}(\theta)^{-1} f_N(\theta, Y), \tag{2}$$

with $f_N(\theta, Y) = \frac{1}{N} \sum_{t=1}^{T} f_i(\theta),\ f_i(\theta) = f(\theta, Y_i)$. The covariance matrix estimator $\hat{V}_{ff}(\theta)$ that we use in (2) is the Eicker-White covariance matrix estimator, see Eicker (1967) and White (1980). Usage of the Eicker-White covariance matrix estimator implies that we do not allow for dependence between the moments. This is done for expository purposes and we discuss how to deal with dependence between the moments later. We make extensive use of the derivative of the moment functions

$$q_N(\theta, Y) = \tfrac{1}{N} \textstyle\sum_{i=1}^{N} q_i(\theta), \tag{3}$$

with $q_i(\theta) = \frac{\partial}{\partial \theta'} f_i(\theta)$. We use the Eicker-White covariance matrix estimator as well to estimate the covariance between the moments and their derivatives. Thus we employ the covariance matrix estimators:

$$\begin{array}{rl} \hat{V}_{ff}(\theta) = & \frac{1}{N} \sum_{i=1}^{N} f_i(\theta) f_i(\theta)' - f_N(\theta, Y) f_N(\theta, Y)', \\ \hat{V}_{qf}(\theta) = & \frac{1}{N} \sum_{i=1}^{N} q_i(\theta) f_i(\theta)' - q_N(\theta, Y) f_N(\theta, Y)', \\ \hat{V}_{qq}(\theta) = & \frac{1}{N} \sum_{i=1}^{N} q_i(\theta) q_i(\theta)' - q_N(\theta, Y) q_N(\theta, Y)', \end{array} \tag{4}$$

of the covariance matrices $V_{ff}(\theta) = \mathrm{E}(\bar{f}_i(\theta)\bar{f}_i(\theta)')$, $V_{qf}(\theta) = \mathrm{E}(\bar{q}_i(\theta)\bar{f}_i(\theta)')$ and $V_{qq}(\theta) = \mathrm{E}(\bar{q}_i(\theta)\bar{q}_i(\theta)')$, where $\bar{f}_i(\theta) = f_i(\theta) - f_N(\theta, Y)$ and $\bar{q}_i(\theta) = q_i(\theta) - q_N(\theta, Y)$. Alongside the independence of moments across individuals/time, we make some other assumptions. We have grouped all assumptions in Assumption 1.

Assumption 1. *Under $H_0 : \theta = \theta_0$, the following assumptions hold jointly:*

a. *The vectors of moments and derivatives $(f_i(\theta)' \vdots q_i(\theta)')'$ are independent across individuals/time.*

b. *For $d_i(\theta_0) : d_i(\theta_0) = q_i(\theta_0) - V_{qf}(\theta_0) V_{ff}(\theta_0)^{-1} f_i(\theta_0)$, it holds that:*

$$\begin{array}{ll} \mathrm{E}(f_i(\theta_0) | d_i(\theta_0)) = & 0 \\ \mathrm{E}(f_i(\theta_0) f_i(\theta_0)' | d_i(\theta_0)) = & V_{ff}(\theta_0) \\ \mathrm{E}(f_i(\theta_0) f_i(\theta_0)' \otimes f_i(\theta_0) f_i(\theta_0)' | d_i(\theta_0)) = & \mathrm{E}(f_i(\theta_0) f_i(\theta_0)' \otimes f_i(\theta_0) f_i(\theta_0)'). \end{array} \tag{5}$$

c. *The sixth order moments of $f_i(\theta_0)$ and $d_i(\theta_0)$ are finite.*

Assumption 1a has been discussed before and justifies usage of the Eicker-White covariance matrix estimator. Assumption 1b implies that the first, second and fourth order conditional moments of $f_i(\theta_0)$ given $d_i(\theta_0)$ are equal to the unconditional moments. The covariance (central moment) between $d_i(\theta_0)$ and $f_i(\theta_0)$, $f_i(\theta_0) f_i(\theta_0)'$ and $f_i(\theta_0) f_i(\theta_0)' \otimes f_i(\theta_0) f_i(\theta_0)'$ is equal to zero because of Assumption 1b. The conditional mean and variance assumptions in Assumption 1b are necessary for the absence of any zero-th order bias in the weak instrument robust statistics. Assumption 1c is such that the N^{-2}-th order components of the statistics that we analyze have a finite mean.

Under H_0, $d_i(\theta_0)$ is an (infeasible) estimator of the derivative of $f_i(\theta)$ with respect to θ that is uncorrelated with $f_i(\theta_0)$. The zero covariance between $d_i(\theta_0)$ and $f_i(\theta_0)$ holds by construction. Assumption 1b implies that the zero covariance between $d_i(\theta_0)$ and $f_i(\theta_0)$ and the zero mean of $f_i(\theta_0)$ result from the conditional moment restriction: $\mathrm{E}(f_i(\theta_0)|d_i(\theta_0)) = 0$. Besides a conditional first moment that does not depend on $d_i(\theta_0)$, Assumption 1b implies that the conditional second and fourth moments of $f_i(\theta_0)$ do not depend on $d_i(\theta_0)$ as well. Any conditional heteroscedasticity or leptokurticy of $f_i(\theta_0)$ is thus not a function of $d_i(\theta)$. A sufficient condition for Assumption 1b to hold is that $f_i(\theta_0)$ and $d_i(\theta_0)$ are independent.

A feasible estimator of the mean of $d_i(\theta)$ is

$$\hat{D}_N(\theta, Y) = q_N(\theta, Y) - \hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1} f_N(\theta, Y). \tag{6}$$

Corollary 1. *Under $H_0 : \theta = \theta_0$ and Assumption 1, $\sqrt{N} f_N(\theta_0, Y)$ and $\sqrt{N}\left[\hat{D}_N(\theta_0, Y) - J_\theta(\theta_0)\right]$, with $J_\theta(\theta_0) = \mathrm{E}[q_i(\theta_0)]$, have independent normal limiting distributions with mean zero and covariance matrices $V_{ff}(\theta_0)$ and $V_{\theta\theta}(\theta_0) = V_{qq}(\theta_0) - V_{qf}(\theta_0)V_{ff}(\theta_0)^{-1}V_{qf}(\theta_0)'$.*

Proof. see Kleibergen (2005). ■

The asymptotic independence of $f_N(\theta_0, Y)$ and $\hat{D}_N(\theta_0, Y)$ allows for the construction of test statistics that test H_0 and whose limiting distributions are robust to zero values of $J_\theta(\theta_0)$. We use Assumption 1 to determine the higher order properties of three of these statistics and one whose limiting distribution is only valid under a non-zero value of $J_\theta(\theta_0)$.

Definition 1. Four different statistics that test $\mathrm{H}_0 : \theta = \theta_0$ are:

1. The S-statistic, see Stock and Wright (2000), which is the generalisation of the Anderson-Rubin statistic, see Anderson and Rubin (1949), towards GMM,

$$\mathrm{S}(\theta_0) = N f_N(\theta_0, Y)'\hat{V}_{ff}(\theta_0)^{-1} f_N(\theta_0, Y). \tag{7}$$

   Under H_0 and Assumption 1, the S-statistic has a $\chi^2(k)$ limiting distribution regardless of the value of $J_\theta(\theta_0)$.

2. The KLM-statistic which is a GMM-Lagrange multiplier (LM) statistic based on the CUE, see Kleibergen (2005):

$$\begin{aligned}\mathrm{KLM}(\theta_0) = \;& N f_N(\theta_0, Y)'\hat{V}_{ff}(\theta_0)^{-1}\hat{D}_N(\theta_0, Y)\left[\hat{D}_N(\theta_0, Y)'\hat{V}_{ff}(\theta_0)^{-1}\hat{D}_N(\theta_0, Y)\right]^{-1} \\ & \hat{D}_N(\theta_0, Y)'\hat{V}_{ff}(\theta_0)^{-1} f_N(\theta_0, Y).\end{aligned} \tag{8}$$

   Under H_0 and Assumption 1, the KLM-statistic has a $\chi^2(1)$ limiting distribution regardless of the value of $J_\theta(\theta_0)$.

3. The GMM-MLR-statistic which is Moreira's (2003) conditional likelihood ratio (LR) statistic applied in a GMM-setting, see Kleibergen (2005):

$$\text{GMM-MLR}(\theta_0) = \tfrac{1}{2}\left[\mathrm{S}(\theta_0) - \mathrm{r}(\theta_0) + \sqrt{(\mathrm{S}(\theta_0) + \mathrm{r}(\theta_0))^2 - 4\left[\mathrm{S}(\theta_0) - \mathrm{KLM}(\theta_0)\right]\mathrm{r}(\theta_0))}\right], \tag{9}$$

with $\mathrm{r}(\theta_0) = N\hat{D}_N(\theta_0,Y)'\left[\hat{V}_{qq}(\theta_0) - \hat{V}_{qf}(\theta_0)\hat{V}_{ff}(\theta_0)^{-1}\hat{V}_{fq}(\theta_0)\right]^{-1}\hat{D}_N(\theta_0,Y)$. Under H_0 and Assumption 1, the GMM-MLR statistic has a conditional limiting distribution given $\mathrm{r}(\theta_0)$ in which $\mathrm{KLM}(\theta_0)$ and $\mathrm{S}(\theta_0){-}\mathrm{KLM}(\theta_0)$ have independent $\chi^2(1)$ and $\chi^2(k-1)$ limiting distributions.

4. The GMM-LM statistic, see Newey and West (1987):

$$\begin{aligned}\mathrm{LM}(\theta_0) = \; & Nf_N(\theta_0,Y)'\hat{V}_{ff}(\theta_0)^{-1}q_N(\theta_0,Y)\left[q_N(\theta_0,Y)'\hat{V}_{ff}(\theta_0)^{-1}q_N(\theta_0,Y)\right]^{-1} \\ & q_N(\theta_0,Y)'\hat{V}_{ff}(\theta_0)^{-1}f_N(\theta_0,Y).\end{aligned} \tag{10}$$

Under Assumption 1 and when $J_\theta(\theta_0)$ has a non-zero value, the GMM-LM statistic has a $\chi^2(1)$ limiting distribution.

The above four statistics are commonly used to test the parameters in models that are estimated using GMM. The (conditional) limiting distributions of the S, KLM and GMM-MLR statistics are robust to weak instruments while that of the GMM-LM statistic is not. We added the GMM-LM statistic to illustrate the issues associated with GMM statistics that are not robust to weak instruments. Since these issues are similar for all non-robust statistics, we just discuss them for the one for which they are the most straightforward to obtain.

3 Higher order expansions

We construct the higher order expressions of the statistics from Definition 1 by replacing the covariance estimators involved in these statistics by Taylor expansions around their true values. For the S-statistic, $\mathrm{S}(\theta_0)$, this implies that we use a Taylor approximation of $\hat{V}_{ff}(\theta_0)$ around $V_{ff}(\theta_0)$ while for the KLM statistic, $\mathrm{KLM}(\theta)$, we use Taylor approximations of both $\hat{V}_{ff}(\theta_0)$ and $\hat{D}_N(\theta_0,Y)$ around $V_{ff}(\theta_0)$ and $D_N(\theta_0,Y) = \frac{1}{N}\sum_{i=1}^N d_i(\theta_0)$ resp.. The Taylor approximations of these estimators are stated in Lemmas 1-8 in Appendix A.

Theorem 1 states the higher order expressions of the four statistics from Definition 1. The order of the different components of the statistics results from the convergence rate of the expectation of these components given $D_N(\theta_0,Y)$. For the non-robust GMM-LM statistic (10), Theorem 1 just states some of the elements of its higher order expression. These elements suffice to show the dependence of the limiting distribution of the GMM-LM statistic on the quality of the instruments.

Theorem 1. *Under H_0 and Assumption 1, the higher order expressions of the statistics from Definition 1 read:*
1. *For the S-statistic* (7):

$$\mathrm{S}(\theta_0) = S_0 + \tfrac{1}{N}S_1 + O_p(\tfrac{1}{N^2}), \tag{11}$$

where

$$\begin{aligned} S_0 = \; & Nf_N(\theta_0,Y)'V_{ff}(\theta_0)^{-1}f_N(\theta_0,Y) \underset{d}{\rightarrow} \chi^2(k) \\ S_1 = \; & -N^2 f_N(\theta_0,Y)'V_{ff}(\theta_0)^{-1}\left[\hat{V}_{ff}(\theta_0) - V_{ff}(\theta_0)\right]V_{ff}(\theta_0)^{-1}f_N(\theta_0,Y) \\ \mathrm{E}(S_1) = \; & -\tfrac{N-1}{N}\mathrm{E}\left[\left(f_i'(\theta_0)V_{ff}(\theta_0)^{-1}f_i(\theta_0)\right)^2\right] + \tfrac{N-1}{N}\left[k^2+2k\right] + k.\end{aligned} \tag{12}$$

2. *For the KLM-statistic* (8):

$$\text{KLM}(\theta_0) = KLM_0 + \tfrac{1}{N}KLM_1 + \tfrac{1}{N\sqrt{N}}KLM_2 + O_p(\tfrac{1}{N^2}), \tag{13}$$

where

$$\begin{aligned}
KLM_0 =& \; Nf_N(\theta,Y)'V_{ff}(\theta)^{-\frac{1}{2}}P_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}f_N(\theta,Y) \\
KLM_1 =& \; -N^2 f_N(\theta,Y)'V_{ff}(\theta)^{-\frac{1}{2}}P_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right] \\
& V_{ff}(\theta)^{-\frac{1}{2}}P_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}f_N(\theta,Y) - \\
& 2N^2 f_N(\theta,Y)'V_{ff}(\theta)^{-\frac{1}{2}}P_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right] \\
& V_{ff}(\theta)^{-\frac{1}{2}}M_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}f_N(\theta,Y) \\
KLM_2 =& \; -2N^2\sqrt{N}f_N(\theta,Y)'V_{ff}(\theta)^{-\frac{1}{2}}M_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}\hat{V}_{\theta f}(\theta)V_{ff}(\theta)^{-1}f_N(\theta,Y) \\
& \left[D_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y)\right]^{-1}D_N(\theta,Y)'V_{ff}(\theta)^{-1}f_N(\theta,Y),
\end{aligned} \tag{14}$$

with $\hat{V}_{\theta f}(\theta) = \frac{1}{N}\sum_{i=1}^{N} d_i(\theta_0)f(\theta_0)' - D_N(\theta_0,Y)f_N(\theta_0,Y)'$, *and the conditional expectations of these elements given* $D_N(\theta_0,Y)$ *read:*

$$\begin{aligned}
\text{E}\left[KLM_0|D_N(\theta_0,Y)\right] =& \; 1 \\
\text{E}\left[KLM_1|D_N(\theta_0,Y)\right] =& \; -\tfrac{N-1}{N}\left\{\text{E}\left[\left(f_i(\theta_0)'V_{ff}(\theta_0)^{-\frac{1}{2}}P_{V_{ff}(\theta_0)^{-\frac{1}{2}}D_N(\theta_0,Y)}V_{ff}(\theta_0)^{-\frac{1}{2}}f_i(\theta_0)\right)^2 + \right.\right. \\
& 2f_i(\theta_0)'V_{ff}(\theta_0)^{-\frac{1}{2}}P_{V_{ff}(\theta_0)^{-\frac{1}{2}}D_N(\theta_0,Y)}V_{ff}(\theta_0)^{-\frac{1}{2}}f_i(\theta_0)f_i(\theta_0)'V_{ff}(\theta_0)^{-\frac{1}{2}} \\
& \left.\left. M_{V_{ff}(\theta_0)^{-\frac{1}{2}}D_N(\theta_0,Y)}V_{ff}(\theta_0)^{-\frac{1}{2}}f_i(\theta_0)|D_N(\theta_0,Y)\right] - 2(k+1)\right\} + \tfrac{1}{N} \\
\text{E}\left[\sqrt{N}KLM_2|D_N(\theta_0,Y)\right] =& \; 0.
\end{aligned} \tag{15}$$

3. *For the GMM-MLR statistic* (9) *given* $\text{r}(\theta_0)$:

$$\begin{aligned}
\text{GMM-MLR}(\theta_0) =& \; \tfrac{1}{2}\left[S_0 - \text{r}(\theta_0) + \sqrt{(S_0+\text{r}(\theta_0))^2 - 4\left[S_0 - KLM_0\right]\text{r}(\theta_0))}\right] + \\
& \tfrac{1}{2N}\left[1 + \tfrac{S_0-\text{r}(\theta_0)}{\sqrt{(S_0+\text{r}(\theta_0))^2-4[S_0-KLM_0]\text{r}(\theta_0))}}\right]S_1 + \\
& \tfrac{1}{N}\tfrac{\text{r}(\theta_0)}{\sqrt{(S_0+\text{r}(\theta_0))^2-4[S_0-KLM_0]\text{r}(\theta_0))}}(KLM_1 + \tfrac{1}{\sqrt{N}}KLM_2) + O_p(\tfrac{1}{N^2}).
\end{aligned} \tag{16}$$

4. *A part of the higher order expression of the GMM-LM statistic (10) reads:*

$$\begin{aligned}
\text{LM}(\theta) =& \; KLM_0 + \tfrac{1}{N}LM_1 + \tfrac{1}{N}\text{tr}\left(LM_{D_1}\left(D_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y)\right)^{-1}D_N(\theta,Y)'V_{ff}(\theta)^{-1}\right) + \\
& \tfrac{1}{N^2}LM_{D_2}\left(D_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y)\right)^{-1},
\end{aligned} \tag{17}$$

where

$$\begin{aligned}
LM_1 =& \; -2N^2 f_N(\theta,Y)'V_{ff}(\theta)^{-1}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]V_{ff}(\theta)^{-1}D_N(\theta,Y) \\
& \left(D_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y)\right)^{-1}D_N(\theta,Y)'V_{ff}(\theta)^{-1}f_N(\theta,Y) \\
LM_{D_1} =& \; 2N\left\{f_N(\theta,Y)f_N(\theta,Y)'V_{ff}(\theta)^{-1}V_{qf}(\theta)V_{ff}(\theta)^{-1}f_N(\theta,Y) - \left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]V_{ff}(\theta)^{-1}\right. \\
& \left. f_N(\theta,Y)f_N(\theta,Y)'V_{ff}(\theta)^{-1}V_{qf}(\theta)'V_{ff}(\theta)^{-1}f_N(\theta,Y)\right\} \\
LM_{D_2} =& \; N\left\{f_N(\theta,Y)'V_{ff}(\theta)^{-1}V_{qf}(\theta)'V_{ff}(\theta)^{-1}f_N(\theta,Y)\right\}^2
\end{aligned} \tag{18}$$

and the conditional expectations of these elements given $D_N(\theta_0, Y)$ read:

$$\begin{array}{rl}
\mathrm{E}(LM_1|D_N(\theta_0,Y)) = & 2 - 2\mathrm{tr}\left\{\mathrm{E}\left[f_i(\theta_0)f_i(\theta_0)'V_{ff}(\theta_0)^{-1}f_i(\theta_0)f_i(\theta_0))'\right]V_{ff}(\theta_0)^{-\frac{1}{2}}\right. \\
& \left. P_{V_{ff}(\theta_0)^{-\frac{1}{2}}D_N(\theta_0,Y)}V_{ff}(\theta_0)^{-\frac{1}{2}}\right\} \\
\mathrm{E}(LM_{D_1}|D_N(\theta_0,Y)) = & 2\mathrm{E}\left[f_i(\theta_0)f_i(\theta_0)'V_{ff}(\theta_0)^{-1}V_{qf}(\theta_0)V_{ff}(\theta_0)^{-1}f_i(\theta_0)V_{ff}(\theta_0)^{-1}|D_N(\theta_0,Y)\right] - \\
& 2\frac{N-1}{N}\left\{\mathrm{tr}(V_{ff}(\theta_0)^{-1}V_{qf}(\theta_0))\mathrm{E}\left(f_i(\theta_0)'V_{ff}(\theta_0)^{-1}f_i(\theta_0)f_i(\theta_0)'|D_N(\theta_0,Y)\right) + \right. \\
& \mathrm{E}\left(f_i(\theta_0)'V_{ff}(\theta_0)^{-1}V_{qf}(\theta_0)'V_{ff}(\theta_0)^{-1}f_i(\theta_0)f_i(\theta_0)'|D_N(\theta_0,Y)\right) + \\
& \left.\mathrm{E}\left(f_i(\theta_0)'V_{ff}(\theta_0)^{-1}V_{qf}(\theta_0)V_{ff}(\theta_0)^{-1}f_i(\theta_0)f_i(\theta_0)'|D_N(\theta_0,Y)\right)\right\} - \\
& \frac{2}{N}\mathrm{E}[f_i(\theta_0)'V_{ff}(\theta_0)^{-1}V_{qf}(\theta_0)V_{ff}(\theta_0)^{-1}f_i(\theta_0)f_i(\theta_0)'V_{ff}(\theta_0)^{-1}f_i(\theta_0)f_i(\theta_0)' \\
& |D_N(\theta_0,Y)] - \frac{2}{N}\mathrm{E}\left(f_i(\theta_0)'V_{ff}(\theta_0)^{-1}V_{qf}(\theta_0)V_{ff}(\theta_0)^{-1}f_i(\theta_0)f_i(\theta_0)'|D_N(\theta_0,Y)\right) \\
\mathrm{E}(LM_{D_2}|D_N(\theta_0,Y)) = & (N-1)\left\{\left[\mathrm{tr}(V_{ff}(\theta_0)^{-1}V_{qf}(\theta_0))\right]^2 + \mathrm{tr}(V_{ff}(\theta_0)^{-1}V_{qf}(\theta_0)V_{ff}(\theta_0)^{-1}V_{qf}(\theta_0)') + \right. \\
& \left.\mathrm{tr}(V_{ff}(\theta_0)^{-1}V_{qf}(\theta_0)V_{ff}(\theta_0)^{-1}V_{qf}(\theta_0))\right\} + \\
& \mathrm{E}\left[f_i(\theta_0)'V_{ff}(\theta_0)^{-1}V_{qf}(\theta_0)V_{ff}(\theta_0)^{-1}f(\theta_0))^2\right].
\end{array} \tag{19}$$

Proof. see Appendix B. ■

We determined the order of the different elements in the higher order expressions in Theorem 1 using the conditional expectations given $D_N(\theta_0, Y)$. To construct these conditional expectations we first construct the conditional expectation given $d_i(\theta_0)$, $i = 1, \ldots, N-1$, and $D_N(\theta_0, Y)$ and then integrate this conditional expectation with respect to $d_i(\theta_0)$, $i = 1, \ldots, N-1$. Since the conditional expectations just depend on $D_N(\theta_0, Y)$, the latter integration step does not alter the expressions. We use Assumption 1b to construct the conditional expectations given $d_i(\theta_0)$. The S-statistic does not depend on $D_N(\theta_0, Y)$ so $D_N(\theta_0, Y)$ does not influence the higher order expression of $\mathrm{S}(\theta_0)$.

The conditional expectation of the different elements of the higher order expression of $\mathrm{KLM}(\theta_0) : KLM_0$, KLM_1 and KLM_2, depend on $D_N(\theta_0, Y)$ but are invariant with respect to its length. Since they therefore only depend on the direction of $D_N(\theta_0, Y)$, the order of the conditional expectations does not depend on $D_N(\theta_0, Y)$ and is as stated in Theorem 1. The equality of the unconditional expectation of $f_i(\theta_0)f_i(\theta_0)'$ and its conditional expectation given $d_i(\theta_0)$ is a necessary assumption for $\mathrm{KLM}(\theta)$ to converge to a $\chi^2(1)$ distributed random variable. If we do not make this assumption, the zero-th order element KLM_0 alters and no longer converges to a $\chi^2(1)$ distributed random variable.

The higher order elements of $\mathrm{KLM}(\theta_0) : KLM_1$ and KLM_2 result from the different covariance estimators that are involved in $\mathrm{KLM}(\theta_0) : KLM_1$ results from $\hat{V}_{ff}(\theta)$ while KLM_2 results from (the infeasible covariance matrix estimator) $\hat{V}_{\theta f}(\theta_0)$. KLM_1 is therefore comparable to the S_1 higher order element of $\mathrm{S}(\theta_0)$ which also results from $\hat{V}_{ff}(\theta)$. It is interesting to note that the higher order element that results from $\hat{V}_{\theta f}(\theta_0)$, *i.e.* KLM_2, is of a lower order than the one which results from $\hat{V}_{ff}(\theta)$, *i.e.* KLM_1. This results from Assumption 1b. Assumption 1b implies that

$$\begin{array}{ll}
\mathrm{E}\left[d_i(\theta_0)\left\{f_i(\theta_0)'V_{ff}(\theta_0)^{-1}f_i(\theta_0) - f_j(\theta_0)'V_{ff}(\theta_0)^{-1}f_j(\theta_0)\right\}|i \neq j\right] & = 0 \Leftrightarrow \\
\mathrm{E}\left(d_i(\theta_0)f_i(\theta_0)'V_{ff}(\theta_0)^{-1}f_i(\theta_0)\right) - \mathrm{E}(d_i(\theta_0))\mathrm{E}\left(f_i(\theta_0)'V_{ff}(\theta_0)^{-1}f_i(\theta_0)\right) & = 0
\end{array} \tag{20}$$

and a similar expression holds for the fourth order moment. Hence,

$$\begin{array}{l}
\left[\left(\sum_{i=1}^N d_i(\theta_0)\right)' V_{ff}(\theta_0)^{-1}\left(\sum_{i=1}^N d_i(\theta_0)\right)\right]^{-\frac{1}{2}} \\
\sum_{i=1}^N d_i(\theta_0)\left\{f_i(\theta_0)'V_{ff}(\theta_0)^{-1}f_i(\theta_0) - \frac{1}{N}\sum_{j=1,j\neq i}^N f_j(\theta_0)'V_{ff}(\theta_0)^{-1}f_j(\theta_0)\right\} \underset{p}{\to} 0.
\end{array} \tag{21}$$

In the KLM_2 higher order term, $V_{ff}(\theta)^{-\frac{1}{2}}\hat{V}_{\theta f}(\theta)V_{ff}(\theta)^{-1}f_N(\theta,Y)$ is pre-multiplied by $M_{V_{ff}(\theta_0)^{-\frac{1}{2}}D_N(\theta_0,Y)}$ which is a projection on the space orthogonal to $V_{ff}(\theta_0)^{-\frac{1}{2}}D_N(\theta_0,Y)$. Equation (21) describes the convergence behavior of $\hat{V}_{\theta f}(\theta)V_{ff}(\theta)^{-1}f_N(\theta,Y)$ and shows that it is proportional to $D_N(\theta_0,Y)=\frac{1}{N}\sum_{i=1}^N d_i(\theta_0)$. The projection matrix $M_{V_{ff}(\theta_0)^{-\frac{1}{2}}D_N(\theta_0,Y)}$ maps $V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta_0,Y)$ onto zero which explains why KLM_2 is of a lower order than KLM_1. Thus the higher order elements that result from $\hat{V}_{\theta f}(\theta_0)$ are of a lower order than those that result from $\hat{V}_{ff}(\theta_0)$.

The higher order expression of GMM-MLR(θ_0) stated in Theorem 1 is conditional on r(θ_0). It is therefore obtained using a Taylor expansion with respect to the other two components of GMM-MLR(θ_0) : S(θ_0) and KLM(θ_0). Identical to KLM(θ_0), all $\frac{1}{N}$-order components of the higher order expression of GMM-MLR(θ_0) given r(θ_0) result from $\hat{V}_{ff}(\theta_0)$.

The higher order expression of LM(θ_0) in Theorem 1 just states some of the higher order elements of LM(θ_0). We only want to show the dependence of the higher order elements of LM(θ_0) on $D_N(\theta_0,Y)$ for which we do not need the full higher order expression. We use the conditional expectation given $D_N(\theta_0,Y)$ to show this dependence. The elements of LM(θ_0) in Theorem 1 are such that both LM_{D_1} and LM_{D_2} are multiplied by a function that is not invariant with respect to the length of $D_N(\theta_0,Y)$. The conditional expectations show that LM_{D_1} is at most of order zero in N. Multiplied by $\left(D_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y)\right)^{-1}D_N(\theta,Y)'$ and divided by N, the full contribution of LM_{D_1} is at most of order $\frac{1}{\sqrt{N}}$, which occurs when $J_\theta(\theta)$ equals zero such that $\left(D_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y)\right)^{-1}D_N(\theta,Y)'$ is of order $\sqrt{N}$, so it can not alter the limiting distribution of LM(θ_0). The conditional expectation of LM_{D_2} is proportional N. When $J_\theta(\theta_0)$ equals zero, $\left(D_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y)\right)^{-1}$ is proportional to N as well so LM_{D_2} changes the limiting distribution of LM(θ) in this case since it leads to an element of zero-th order in N in the higher order expression of LM(θ_0). This explains why the limiting distribution of LM(θ_0) depends on $J_\theta(\theta_0)$. The same result can be shown for other GMM statistics that are not robust to weak instruments as well. For reasons of brevity, we refrain from showing this for these statistics.

4 Bootstraping weak instrument robust statistics

The N^{-1}-th order components of S(θ) and KLM(θ) stated in Theorem 1 both result from the estimation of $V_{ff}(\theta_0)$ and are the highest order components present in S(θ) and KLM(θ) besides the zero-th order terms. This suggests that we can obtain a bootstrap approximation of the finite sample distribution of both S(θ) and KLM(θ) by just resampling $\bar{f}_i(\theta_0)$ with replacement[1]. When we condition on r(θ_0), GMM-MLR(θ_0) only depends on S(θ) and KLM(θ) so we obtain a bootstrap approximation of the conditional finite sample distribution of GMM-MLR(θ_0) given $r(\theta_0)$ along the same lines. Theorem 1 shows that the limiting distribution of LM(θ) depends on nuisance parameters so we can not construct a bootstrap algorithm that resamples LM(θ) and that approximates the finite sample distribution of LM(θ) for all values of the nuisance parameters. We therefore do not construct a bootstrap algorithm for LM(θ).

A bootstrap algorithm to approximate the finite sample distributions of S(θ_0), KLM(θ_0) and GMM-MLR(θ_0) reads:

[1] We resample $\bar{f}_i(\theta_0)$ instead of $f_i(\theta_0)$ because the moment condition does not hold for the realized observations $f_i(\theta_0)$, $i=1,\ldots,N$, see *e.g.* Hall and Horowitz (1996).

Bootstrap algorithm:

1. Compute $\hat{D}_N(\theta_0, Y)$ and $\mathrm{r}(\theta_0)$ and set bootstrap sample size B and number of simulations I.

2. For $i = 1, \ldots, I$:

   (a) Sample $\{f_j^*(\theta_0),\ j = 1, \ldots, B\}$ independently with replacement from $\{\bar{f}_j(\theta_0),\ j = 1, \ldots, N\}$:
   $$\Pr\left[f_j^*(\theta_0) = \bar{f}_l(\theta_0)\right] = \tfrac{1}{N},\ \ l = 1, \ldots, N. \tag{22}$$

   (b) Compute:
   $$\begin{aligned} f_B^*(\theta_0, Y)_i = &\ \tfrac{1}{B}\textstyle\sum_{j=1}^{B} f_j^*(\theta_0) \text{ and} \\ V_{ff}^*(\theta_0)_i = &\ \tfrac{1}{B}\textstyle\sum_{j=1}^{B} f_j^*(\theta_0) f_j^*(\theta_0)' - f_B^*(\theta_0, Y) f_B^*(\theta_0, Y)', \end{aligned} \tag{23}$$

   from the bootstrap sample $\{f_j^*(\theta_0),\ j = 1, \ldots, B\}$.

   (c) Compute:
   $$\begin{aligned} \mathrm{S}^*(\theta_0)_i = &\ B f_B^*(\theta_0, Y)_i' V_{ff}^*(\theta_0)_i^{-1} f_B^*(\theta_0, Y)_i \\ \mathrm{KLM}^*(\theta_0)_i = &\ B f_B^*(\theta_0, Y)_i' V_{ff}^*(\theta_0)_i^{-\frac{1}{2}} P_{V_{ff}^*(\theta_0)_i^{-\frac{1}{2}} \hat{D}_N(\theta_0, Y)} V_{ff}^*(\theta_0)_i^{-\frac{1}{2}} f_B^*(\theta_0, Y)_i \\ \mathrm{GMM\text{-}MLR}^*(\theta_0)_i = &\ \tfrac{1}{2}\Big[\mathrm{S}^*(\theta_0)_i - \mathrm{r}(\theta_0) + \\ &\ \sqrt{(\mathrm{S}^*(\theta_0)_i + \mathrm{r}(\theta_0))^2 - 4\left[\mathrm{S}^*(\theta_0)_i - \mathrm{KLM}^*(\theta_0)_i\right]\mathrm{r}(\theta_0))}\Big] \end{aligned} \tag{24}$$

3. Construct the (conditional) bootstrap distributions of $\mathrm{S}(\theta_0)$, $\mathrm{KLM}(\theta_0)$ and $\mathrm{GMM\text{-}MLR}(\theta_0)$ from the sample $\{\mathrm{S}^*(\theta_0)_i,\ \mathrm{KLM}^*(\theta_0)_i,\ \mathrm{GMM\text{-}MLR}^*(\theta_0)_i,\ i = 1, \ldots, I\}$.

The bootstrap algorithm is such that $\mathrm{E}^*(f_B^*(\theta_0, Y)_i) = 0$, $\mathrm{E}^*(B f_B^*(\theta_0, Y)_i f_B^*(\theta_0, Y)_i') = \hat{V}_{ff}(\theta)$ and $\mathrm{E}^*(V_{ff}^*(\theta_0)_i) = \hat{V}_{ff}(\theta)$, where E^* is the expectation operator with respect to the resampling distribution. The above algorithm just bootstraps a sample mean and Assumption 1 is a sufficient condition for the bootstrapped sample mean to satisfy a central limit theorem. Moreira *et. al.* (2004) show the validity of the bootstrap for the linear instrumental variables regression model.

Identical to Theorem 1, higher order expressions for the bootstrap statistics can be constructed. These are stated in Theorem 2.

Theorem 2. *Under H_0 and Assumption 1, the higher order expressions of the bootstrap statistics* $\mathrm{S}^*(\theta_0)$, $\mathrm{KLM}^*(\theta_0)$ *and* $\mathrm{GMM\text{-}MLR}^*(\theta_0)$ *defined in (24) read:*
1. *For the S^*-statistic*:

$$\mathrm{S}^*(\theta_0) = S_0^* + \tfrac{1}{B} S_1^* + O_p(\tfrac{1}{B^2}), \tag{25}$$

where $f_B^*(\theta_0, Y) = \frac{1}{B}\sum_{j=1}^{B} f_j^*(\theta_0)$,

$$\begin{aligned} S_0^* = &\ B f_B^*(\theta_0, Y)' \hat{V}_{ff}(\theta_0)^{-1} f_B^*(\theta_0, Y) \\ S_1^* = &\ -B^2 f_B^*(\theta_0, Y)' \hat{V}_{ff}(\theta_0)^{-1}\left[V_{ff}^*(\theta_0) - \hat{V}_{ff}(\theta_0)\right] \hat{V}_{ff}(\theta_0)^{-1} f_B^*(\theta_0, Y) \end{aligned} \tag{26}$$

and $\mathrm{E}(S_0^*) = k$ *and* $\mathrm{E}^*(S_1^*) = -\frac{B-1}{B}\left\{\frac{1}{N}\sum_{i=1}^N \left(\bar{f}_i(\theta_0)\hat{V}_{ff}(\theta_0)^{-1}\bar{f}_i(\theta_0)\right)^2 + [k^2+2k]\right\} + k.$

2. *For the KLM*-statistic*:

$$\mathrm{KLM}^*(\theta_0) = KLM_0^* + \tfrac{1}{B}KLM_1^* + O_p(\tfrac{1}{B^2}), \tag{27}$$

where

$$\begin{aligned} KLM_0^* = \;& Bf_B^*(\theta_0,Y)'\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}P_{\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}\hat{D}_N(\theta_0,Y)}\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}f_B^*(\theta_0,Y) \\ KLM_1^* = \;& -Bf_B^*(\theta_0,Y)'\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}P_{\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}\hat{D}_N(\theta_0,Y)}\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}\left[V_{ff}^*(\theta_0)-\hat{V}_{ff}(\theta_0)\right]\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}} \\ & P_{\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}\hat{D}_N(\theta_0,Y)}\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}f_B^*(\theta_0,Y) - 2Bf_B^*(\theta_0,Y)'\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}P_{\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}\hat{D}_N(\theta_0,Y)}\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}} \\ & \left[V_{ff}^*(\theta_0)-\hat{V}_{ff}(\theta_0)\right]\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}M_{\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}\hat{D}_N(\theta_0,Y)}\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}f_B^*(\theta_0,Y) \end{aligned} \tag{28}$$

and $\mathrm{E}^*\left[KLM_0^*|\hat{D}_N(\theta_0,Y)\right] = 1,$

$$\begin{aligned} \mathrm{E}^*\left[KLM_1^*|\hat{D}_N(\theta_0,Y)\right] = \;& -\tfrac{B-1}{B}\left\{\tfrac{1}{N}\sum_{i=1}^N\left(\bar{f}_i(\theta_0)'\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}P_{\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}\hat{D}_N(\theta_0,Y)}\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}\bar{f}_i(\theta_0)\right)^2 + \right. \\ & \tfrac{2}{N}\sum_{i=1}^N\left(\bar{f}_i(\theta_0)'\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}P_{\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}\hat{D}_N(\theta_0,Y)}\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}\bar{f}_i(\theta_0)\right. \\ & \left.\left.\bar{f}_i(\theta_0)'\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}M_{\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}\hat{D}_N(\theta_0,Y)}\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}\bar{f}_i(\theta_0)\right) - 2(k+1)\right\} + \tfrac{1}{B}. \end{aligned} \tag{29}$$

3. *For the GMM-MLR*-statistic given r(θ)*:

$$\begin{aligned} \text{GMM-MLR}^*(\theta_0) = \;& \tfrac{1}{2}\left[S_0^* - \mathrm{r}(\theta_0) + \sqrt{\left(S_0^*+\mathrm{r}(\theta_0)\right)^2 - 4\left[S_0^* - KLM_0^*\right]\mathrm{r}(\theta_0))}\right] + \\ & \tfrac{1}{2B}\left[1 + \frac{S_0^*-\mathrm{r}(\theta_0)}{\sqrt{\left(S_0^*+\mathrm{r}(\theta_0)\right)^2 - 4[S(\theta_0)-KLM(\theta_0)]\mathrm{r}(\theta_0))}}\right]S_1^* + \\ & \tfrac{1}{B}\frac{\mathrm{r}(\theta_0)}{\sqrt{\left(S_0^*+\mathrm{r}(\theta_0)\right)^2 - 4\left[S_0^*-KLM_0^*\right]\mathrm{r}(\theta_0))}}KLM_1^* + O_p(\tfrac{1}{B^2}). \end{aligned}$$

Proof. see Appendix B. The results for the GMM-MLR statistic follow from Theorem 1 and Theorems 2.1 and 2.2. ■

The higher order expressions of the bootstrap statistics in Theorem 2 are identical to those of the orginal statistics in Theorem 1. They are also such that the expectations of the higher order elements converge to the same limits when B equals N and N goes to infinity. This holds since

$$\hat{D}_N(\theta_0,Y) = \; D_N(\theta_0,Y) - \hat{V}_{\theta f}(\theta_0)V_{ff}(\theta_0)^{-1}f_N(\theta_0,Y) + O_p(\tfrac{1}{N}), \tag{30}$$

which is stated in Lemma 2 in Appendix A, and $\hat{V}_{\theta f}(\theta_0)$ converges to zero such that $\hat{D}_N(\theta_0,Y)$ and $D_N(\theta_0,Y)$ have the same convergence behavior. This suggests that usage of the bootstrap distributions of the statistics to determine their significance leads to a higher order of precision in terms of the order of the sample size compared to the limiting distribution. To verify this statement we construct the Edgeworth approximation of the statistics and of their bootstrapped counterparts.

5 Edgeworth Approximations

The regularity conditions from Bhattacharya and Ghosh (1978) for constructing the Edgeworth approximation of the marginal finite sample distribution of $\mathrm{KLM}(\theta_0)$ are such that $\mathrm{E}(D_N(\theta_0, Y)) = J_\theta(\theta_0) \neq 0$. Thus the Edgeworth approximation that results from Bhattacharya and Ghosh (1978) does not allow for weak instruments. We therefore construct an Edgeworth approximation of the conditional finite sample distribution of $\mathrm{KLM}(\theta_0)$ given $D_N(\theta, Y)$ that allows for weak instruments. This Edgeworth approximation is based on the conditional characteristic function of $\mathrm{KLM}(\theta_0)$ given $D_N(\theta_0, Y)$. For this characteristic function to exist, Cramèr's condition has to hold.

Assumption 2. *Cramèr condition: for a k-dimensional vector $t \in \mathbb{R}^k$, it holds that*

$$\lim\nolimits_{\|t\|\to\infty} \sup \left| \mathrm{E}\left[\exp(it' f_j(\theta_0)) | d_i(\theta_0)\right]\right| < 1, \; j = 1, \ldots, N. \tag{31}$$

We construct the Edgeworth expansions of the S and KLM-statistics and their bootstrapped counterparts. The Edgeworth expansions are stated in Theorem 3.

Theorem 3. **A.** *Under H_0 and Assumptions 1 and 2, the Edgeworth approximations of the (conditional) finite sample distributions of $S(\theta_0)$ and $KLM(\theta_0)$ read:*
1. *For $S(\theta_0)$:*

$$\begin{aligned} \Pr\left[\mathrm{S}\left(\theta_0\right) \leq x\right] = & \; \Pr\nolimits_{\chi^2(k)}(x) - \tfrac{1}{N} \tfrac{\mathrm{E}(S_1)}{k} x p_{\chi^2(k)}(x) + O(N^{-2}) \\ = & \; \Pr\nolimits_{\chi^2(k)}\left(x - \tfrac{1}{N}\tfrac{\mathrm{E}(S_1)}{k} x\right) + O(N^{-2}), \end{aligned} \tag{32}$$

where $\Pr_{\chi^2(k)}(x)$ and $p_{\chi^2(k)}(x)$ are the distribution and density function of a $\chi^2(k)$ distributed random variable evaluated at x and $\mathrm{E}(S_1)$ is defined in Theorem 1.
2. *For $KLM(\theta_0)$:*

$$\begin{aligned} \Pr\left[\mathrm{KLM}\left(\theta_0\right) \leq x | D_N(\theta_0, Y)\right] = & \; \Pr\nolimits_{\chi^2(1)}(x) + \tfrac{1}{N}\left(ax + b\sqrt{2\pi x}\right) p_{\chi^2(1)}(x) + o(N^{-1}) \\ = & \; \Pr\nolimits_{\chi^2(1)}\left(x + \tfrac{1}{N}\left(ax + b\sqrt{2\pi x}\right)\right) + o(N^{-1}), \end{aligned} \tag{33}$$

where

$$\begin{aligned} a = & \; \tfrac{N-1}{N}\left\{\mathrm{E}\left[\left(f_i(\theta_0)' V_{ff}(\theta_0)^{-\frac{1}{2}} P_{V_{ff}(\theta_0)^{-\frac{1}{2}} D_N(\theta_0,Y)} V_{ff}(\theta_0)^{-\frac{1}{2}} f_i(\theta_0)\right)^2\right] - 4\right\} - \tfrac{1}{N} \\ b = & \; \tfrac{N-1}{N}\left\{\mathrm{E}\left[f_i(\theta_0)' V_{ff}(\theta_0)^{-\frac{1}{2}} P_{V_{ff}(\theta_0)^{-\frac{1}{2}} D_N(\theta_0,Y)} V_{ff}(\theta_0)^{-\frac{1}{2}} f_i(\theta_0)\right.\right. \\ & \left.\left. f_i(\theta_0)' V_{ff}(\theta_0)^{-\frac{1}{2}} M_{V_{ff}(\theta_0)^{-\frac{1}{2}} D_N(\theta_0,Y)} V_{ff}(\theta_0)^{-\frac{1}{2}} f_i(\theta_0)\right]\right\} - (k-1). \end{aligned} \tag{34}$$

B. *Under H_0 and Assumptions 1 and 2, the Edgeworth approximations of the finite sample distributions of bootstrapped $S(\theta_0)$ and $KLM(\theta_0)$, $S^*(\theta_0)$ and $KLM^*(\theta_0)$, read:*
1. *For $S^*(\theta_0)$:*

$$\Pr\left[\mathrm{S}^*\left(\theta_0\right) \leq x\right] = \; \Pr\nolimits_{\chi^2(k)}\left(x - \tfrac{1}{B}\tfrac{\mathrm{E}(S_1^*)}{k} x\right) + O(B^{-2}), \tag{35}$$

where $\mathrm{E}(S_1^*)$ is defined in Theorem 2.
2. *For $KLM^*(\theta_0)$:*

$$\begin{aligned} \Pr\left[\mathrm{KLM}^*\left(\theta_0\right) \leq x | \hat{D}_N(\theta_0, Y)\right] = & \; \Pr\nolimits_{\chi^2(1)}(x) + \tfrac{1}{B}\left(a^* x + b^*\sqrt{2\pi x}\right) p_{\chi^2(1)}(x) + o(B^{-1}) \\ = & \; \Pr\nolimits_{\chi^2(1)}\left(x + \tfrac{1}{B}\left(a^* x + b^*\sqrt{2\pi x}\right)\right) + o(B^{-1}), \end{aligned} \tag{36}$$

where

$$
\begin{aligned}
a^* = & \ \tfrac{B-1}{B}\left\{\tfrac{1}{N}\sum_{i=1}^N\left[\bar{f}_i(\theta_0)'\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}P_{\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}\hat{D}_N(\theta_0,Y)}\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}\bar{f}_i(\theta_0)\right]^2-4\right\}-\tfrac{1}{B}, \\
b^* = & \ \tfrac{B-1}{B}\left\{\tfrac{1}{N}\sum_{i=1}^N\left[\bar{f}_i(\theta_0)'\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}P_{\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}\hat{D}_N(\theta_0,Y)}\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}\bar{f}_i(\theta_0)\right.\right. \\
& \left.\left.\bar{f}_i(\theta_0)'\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}M_{\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}\hat{D}_N(\theta_0,Y)}\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}\bar{f}_i(\theta_0)\right]-(k-1)\right\}.
\end{aligned} \tag{37}
$$

Proof. see Appendix B. ■

Since $\mathrm{E}(S_1^*)$ converges to $\mathrm{E}(S_1)$ and a^* and b^* converge to a and b when B equals N and N goes to infinity, the Edgeworth approximations of the bootstrapped statistics converge to those of the orginal statistics when B equals N. Usage of bootstrapped critical values then leads to a higher order efficiency gain, see *e.g.* Horowitz (2001).

Corollary 2. *Under H_0, Assumptions 1 and 2 and when B equals N, the bootstrap critical values that result from $S^*(\theta_0)$ and KLM$^*(\theta_0)$ remove the approximation error of the (conditional) finite sample distribution of S(θ) and KLM(θ) up to/including the order $\frac{1}{N}$.*

Corollary 2 shows that the bootstrap can provide higher order efficiency gains even in cases when the parameter of interest is not identified. Corollary 2 therefore extends the previously known results for the bootstrap which only apply to well identified cases.

Theorem 3 only states results for the S and KLM statistics and not for the GMM-MLR statistic. Given r(θ), the GMM-MLR statistic is only a function of the S and KLM statistics. The S and KLM statistic are not asymptotically independent but the GMM-MLR statistic can as well be specified as a function of

$$
\mathrm{JKLM}(\theta_0)=\mathrm{S}(\theta_0)-\mathrm{KLM}(\theta_0), \tag{38}
$$

since

$$
\begin{aligned}
\text{GMM-MLR}(\theta_0)= & \ \tfrac{1}{2}\left[\mathrm{KLM}(\theta_0)+\mathrm{JKLM}(\theta_0)-\mathrm{r}(\theta_0)+\right., \\
& \left.\sqrt{(\mathrm{KLM}(\theta_0)+\mathrm{JKLM}(\theta_0)+\mathrm{r}(\theta_0))^2-4\mathrm{JKLM}(\theta_0)\mathrm{r}(\theta_0))}\right].
\end{aligned} \tag{39}
$$

Under H_0 and Assumption 1, the JKLM statistic converges to a $\chi^2(k-1)$ distributed random variable which is independent of the $\chi^2(1)$ random variable where the KLM statistic converges to, see Kleibergen (2005,2006a). The JKLM statistic can easily be incorporated in the bootstrap algorithm in Section 4 which using Theorem 3 can then also be shown to lead to a higher order efficiency gain for approximating the conditional finite sample distribution of JKLM(θ_0). The bootstrapped critical values of GMM-MLR(θ) that result from bootstrapping KLM(θ) and JKLM(θ) also lead to a higher order efficiency gain for the GMM-MLR statistic since this statistic is just a function of KLM(θ) and JKLM(θ), as r(θ) is fixed, and the bootstrap leads to a higher order efficiency gain for both of these statistics.

Corollary 3. *Under H_0, Assumptions 1 and 2 and when B equals N, the bootstrap critical values that result from GMM-MLR$^*(\theta_0)$ remove the approximation error of the conditional finite sample distribution of GMM-MLR(θ_0) given r(θ_0) up to/including the order $\frac{1}{N}$.*

It is not possible to construct an Edgeworth approximation of the conditional finite sample distribution of GMM-MLR(θ) given r(θ) since the analytical expression of the conditional characteristic function of GMM-MLR(θ) given r(θ) is unknown. We can therefore not proof the higher order gains from the bootstrap using the Edgeworth approximation for GMM-MLR(θ). Thus the higher order improvement of usage of bootstrap critical values for GMM-MLR(θ) can only be verified using the argument that the bootstrap leads to a higher order improvement for KLM(θ) and JKLM(θ) and GMM-MLR(θ) is, given r(θ), just a function of these two statistics.

Besides using the bootstrap to achieve higher order improvements for approximating the (conditional) finite sample distributions of S(θ) and KLM(θ), the Edgeworth approximations from Theorem 3 can be used for this purpose as well. It is interesting to note that the Edgeworth approximations of the (conditional) finite sample distributions of S(θ) and KLM(θ) are almost identical when the $f_i(\theta)$'s are normally distributed. In that case, $\mathrm{E}(S_1) = k$, $a = -1$ and $b = 0$ so the corrections of the critical value is $(1 - \frac{1}{N})x$ for both S(θ) and KLM(θ).

6 Power and size comparison for Panel AR(1)

We illustrate the size improvements from using bootstrap or Edgeworth-corrected critical values for a panel autoregressive model of order one: panel AR(1). An elaborate literature on panel autoregressive models exists, see *e.g.* Anderson and Hsiao (1981), Arellano and Bond (1991) and Arellano and Honoré (2001). For individual i at time t, the panel AR(1) model reads

$$y_{t,i} = c_i + \theta y_{t-1,i} + \varepsilon_{t,i} \qquad t = 1, \ldots, T,\ i = 1, \ldots, N. \tag{40}$$

A sufficient condition for Assumption 1 to hold is that the disturbances $\varepsilon_{t,i}$ are independently distributed with mean zero and finite sixth order moments. We take first differences to remove the individual specific effects c_i, $i = 1, \ldots, N$:

$$\Delta y_{t,i} = \theta \Delta y_{t-1,i} + \Delta \varepsilon_{t,i} \qquad t = 2, \ldots, T,\ i = 1, \ldots, N, \tag{41}$$

with $\Delta y_{t,i} = y_{t,i} - y_{t-1,i}$. Estimation of the parameter θ in (41) 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. We specify the moment equation (1) for the panel AR(1) using all two period and more lagged level values of $y_{t,i}$ as instruments, see Arellano and Bond (1991). The specification of the moment vector $f_i(\theta)$ then reads

$$f_i(\theta) = X_i \varphi_i(\theta) : \tfrac{1}{2}(T-1)(T-2) \times 1 \qquad i = 1, \ldots, N, \tag{42}$$

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

$$X_i = \begin{pmatrix} y_{1,i} & 0 \ldots 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 \ldots 0 & \begin{pmatrix} y_{1,i} \\ \vdots \\ y_{T-2,i} \end{pmatrix} \end{pmatrix} : \tfrac{1}{2}(T-1)(T-2) \times (T-2). \tag{43}$$

We use the Eicker-White covariance matrix estimator (4) with $q_i(\theta) = \frac{\partial}{\partial \theta} f_i(\theta) = X_i \Delta y_{-1,i}$ for $\Delta y_{-1,i} = (\Delta y_{2,i} \ldots \Delta y_{T-1,i})'$. Because

$$\begin{aligned} \binom{f_i(\theta)}{q_i(\theta)} &= \binom{X_i(\Delta y_i - \theta \Delta y_{-1,i})}{X_i \Delta y_{-1,i}} \\ &= \left[\begin{pmatrix} 1 & -\theta \\ 0 & 1 \end{pmatrix} \otimes I_{\frac{1}{2}(T-1)(T-2)} \right] \binom{X_i \Delta y_i}{X_i \Delta y_{-1,i}}, \end{aligned} \tag{44}$$

θ_0	N	$\mathbf{S}(\theta_0)$			$\mathbf{KLM}(\theta_0)$				$\mathbf{GMM\text{-}MLR}(\theta_0)$		$\mathbf{LM}(\theta_0)$
		A	$*$	E	A	$*$	$**$	E	A	$*$	A
0.5	50	22.7	0.7	17.4	13.2	2.0	1.9	9.9	17.5	1.3	16.6
	100	11.2	2.3	11.6	8.7	3.2	3.9	8.3	9.6	2.9	10.5
	250	7.7	5.6	7.6	6.8	4.8	5.0	6.4	7.4	4.8	6.8
0.7	50	22.5	0.5	17.5	14.1	2.2	2.3	10.7	19.2	1.0	18.8
	100	11.4	1.8	10.7	8.8	3.8	3.8	7.7	9.8	3.5	11.6
	250	7.6	5.2	7.1	6.0	5.4	5.7	5.9	7.0	5.5	7.8
0.9	50	21.7	0.5	17.8	15.8	1.6	1.9	11.0	23.9	0.2	36.3
	100	11.1	2.2	10.7	9.6	4.5	4.0	8.7	13.4	2.1	25.3
	250	8.1	6.6	7.0	7.1	5.5	6.2	7.0	9.0	6.3	13.1
0.95	50	21.5	0.1	17.2	14.6	1.6	1.7	10.1	24.5	0.1	44.1
	100	11.3	2.3	10.8	9.9	4.2	4.0	9.0	13.9	2.4	33.6
	250	8.2	5.3	7.1	7.5	6.0	6.1	7.3	10.7	6.6	25.5

Table 1: Size of the different statistics in percentages that test $\mathrm{H}_0 : \theta = \theta_0$ at the 95% significance level. A: asymptotic critical values, $*$ bootstrapped critical values, E Edgeworth-corrected critical values, $**$ bootstrap critical values where also the $q_i(\theta)$'s are resampled.

with $\Delta y_i = (\Delta y_{3,i} \dots \Delta y_{T,i})'$, and X_i (43) consists of lagged values of $y_{t,i}$,

$$\hat{V}(\theta) = \begin{pmatrix} \hat{V}_{ff}(\theta) & \hat{V}_{qf}(\theta)' \\ \hat{V}_{qf}(\theta) & \hat{V}_{qq}(\theta) \end{pmatrix} \tag{45}$$

is singular since some of the elements of $\binom{X_i \Delta y_i}{X_i \Delta y_{-1,i}}$ are identical. We therefore obtain $[\hat{V}_{qq}(\theta) - \hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\hat{V}_{fq}(\theta)]^{-1}$, that is involved in $\mathrm{r}(\theta)$ and thus in GMM-MLR(θ), from a generalized inverse of $\hat{V}(\theta)$.

The derivative of the moments, $q_i(\theta) = X_i \Delta y_{-1,i}$, is a white noise series when $\theta = 1$. The parameter θ is therefore not identified in the moment equations when it is equal to one. Weakly identified values of θ occur when θ is close to one relative to the sample size, *i.e.* when $\frac{1-\theta}{N}$ is small. This implies that the LM(θ_0) statistic from Definition 1 becomes size distorted when θ_0 is close to one relative to the sample size.

Size results We compute the size of $\mathrm{S}(\theta)$, KLM(θ), GMM-MLR(θ) using asymptotic, bootstrapped and Edgeworth-corrected 95% critical values in a simulation experiment that uses the previously discussed panel AR(1) model. To illustrate the sensitivity of the size of LM(θ) to the value of θ, we also compute the size of LM(θ) using its asymptotic critical value.

The panel AR(1) model has independent disturbances which are generated from a student t distribution with ten degrees of freedom, mean zero and variance one so Assumption 1 is satisfied. The individual specific constants c_i are specified as $c_i = (1-\theta)\mu_i$ where the μ_i's are independent realizations from a $N(0,2)$ distribution. The initial observations $y_{0,i}$ are simulated such that $y_{0,i} = \mu_i + \varepsilon_{0,i}$ where the $\varepsilon_{0,i}$'s are independent realizations of standard normal random variables. The number of simulated datasets equals one thousand.

We compute the bootstrap critical values that result from $\mathrm{S}^*(\theta)$ and KLM$^*(\theta)$ using hundred simulations of bootstrap datasets of N observations. The bootstrap datasets are obtained using the algorithm in Section 4. We also use a bootstrap which resamples $\bar{f}_i(\theta)$ and $q_i(\theta)$ jointly and constructs $f_B^{**}(\theta, Y)$ and $D_B^{**}(\theta, Y)$ from the resampled $\bar{f}_i(\theta)$ and $q_i(\theta)$'s identical to how $f_N(\theta, Y)$ and $\hat{D}_N(\theta, Y)$ are obtained from $f_i(\theta)$ and $q_i(\theta)$. To distinguish this bootstrap, which is only used for KLM(θ), from the other bootstrap, its results are indicated by $**$.

The Edgeworth-corrected critical values are computed using the Edgeworth approximations from Theorem 3 where the estimators of $\mathrm{E}(S_1)$, a and b equal $\mathrm{E}(S_1^*)$, a^* and b^* with B equal to N.

Table 1 shows the observed size of the statistics when we test at the 95% significance level in a simulation experiment that uses four different values of θ_0 : 0.5, 0.7, 0.9 and 0.95 and three different values of N : 50, 100, 250. The number of time series observations, T, is equal to 6 for all cases.

Panel 1: Power curves for $T = 6,\ N = 50$.

Fig. 1.1-1.4: Power of S(θ) when using asymptotic (solid line), bootstrapped (dashed line) and Edgeworth-corrected (dashed-dotted line) critical values.

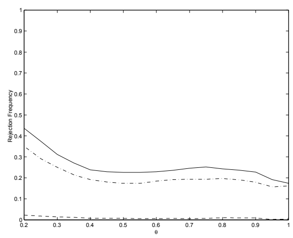
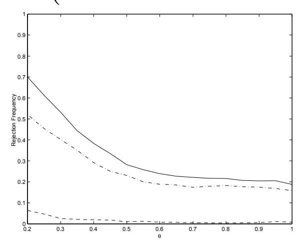
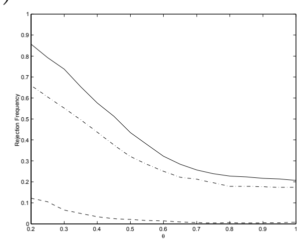
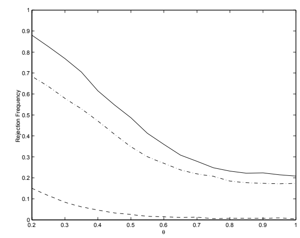

Fig 1.1: $\theta_0 = 0.5$ Fig. 1.2: $\theta_0 = 0.7$ Fig. 1.3: $\theta_0 = 0.9$ Fig. 1.4: $\theta_0 = 0.95$

Fig. 1.5-1.8: Power of KLM(θ) when using asymptotic (solid line), bootstrapped (dashed line), with resampled $q_i(\theta)$ (dotted line) and Edgeworth-corrected (dashed-dotted line) critical values.

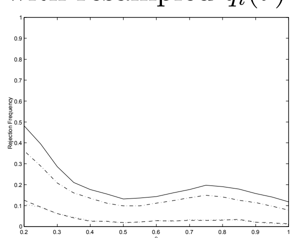
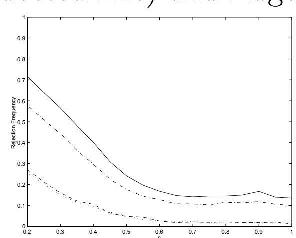
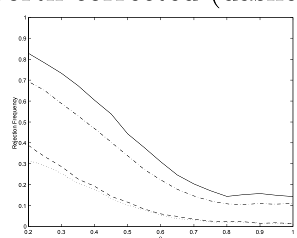
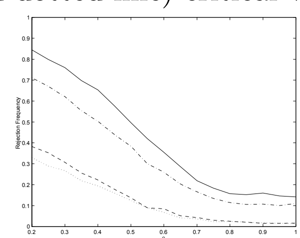

Fig 1.5: $\theta_0 = 0.5$ Fig. 1.6: $\theta_0 = 0.7$ Fig. 1.7: $\theta_0 = 0.9$ Fig. 1.8: $\theta_0 = 0.95$

Fig. 1.8-1.12: Power of GMM-MLR(θ) when using conditional asymptotic (solid line) and bootstrapped (dashed line) critical values and LM(θ) (dashed-dotted line).

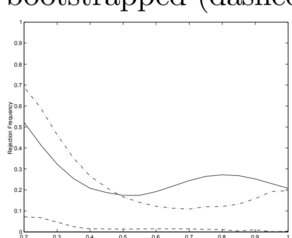
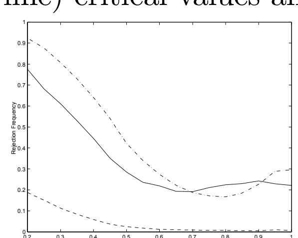
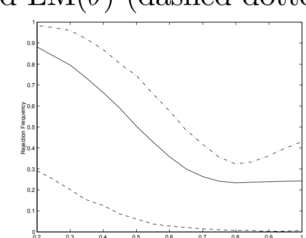
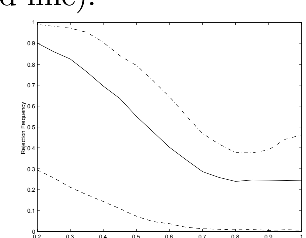

Fig 1.9: $\theta_0 = 0.5$ Fig. 1.10: $\theta_0 = 0.7$ Fig. 1.11: $\theta_0 = 0.9$ Fig. 1.12: $\theta_0 = 0.95$

The observed sizes reported in Table 1 show that usage of the critical values that stem from the asymptotic distributions leads to large size distortions in small samples. Both the Edgeworth correction and the bootstrap decrease these size distortions in all cases. The reduction of the size distortion that results from the bootstraps is, however, much larger than the one that results from the Edgeworth expansion. The reductions result from the higher order improvements from using the bootstrap and/or the Edgeworth corrections.

The observed sizes of KLM(θ) when using the critical values that stem from the bootstrap that resamples the $q_i(\theta)$'s ($**$) are almost identical to those that result from using the critical values in the bootstrap which uses $\hat{D}_N(\theta, Y)$ ($*$). This results since the resampling of the $q_i(\theta)$'s only effects size distortions which are of order $(N\sqrt{N})^{-1}$ while the resampling of the $f_i(\theta)$'s effects size distortions which are of order N^{-1}. The resampling of the $q_i(\theta)$'s is therefore of lesser importance and does not lead to any further size improvements. Thus the size distortions that result from the estimation of the covariance matrix, which is a 10×10 matrix in the simulation experiment, exceed those that result from $\hat{D}_N(\theta, Y)$.

Panel 2: Power curves for $T = 6,\ N = 100$.

Fig. 2.1-2.4: Power of S(θ) when using asymptotic (solid line), bootstrapped (dashed line) and Edgeworth-corrected (dashed-dotted line) critical values.

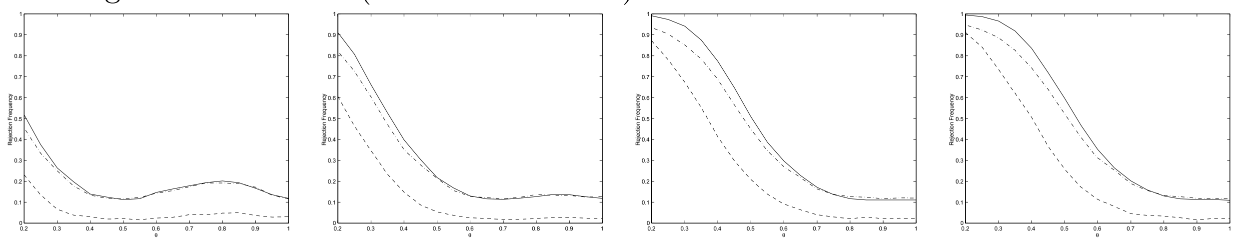

Fig 2.1: $\theta_0 = 0.5$ Fig. 2.2: $\theta_0 = 0.7$ Fig. 2.3: $\theta_0 = 0.9$ Fig. 2.4: $\theta_0 = 0.95$

Fig. 2.5-2.8: Power of KLM(θ) when using asymptotic (solid line), bootstrapped (dashed line), with resampled $q_i(\theta)$ (dotted line) and Edgeworth-corrected (dashed-dotted line) critical values.

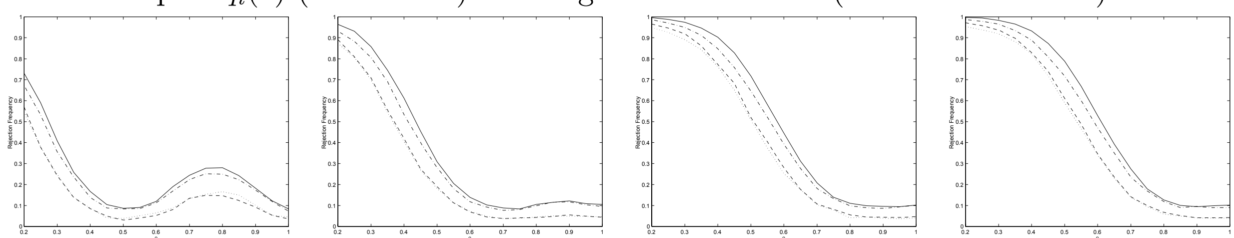

Fig 2.5: $\theta_0 = 0.5$ Fig. 2.6: $\theta_0 = 0.7$ Fig. 2.7: $\theta_0 = 0.9$ Fig. 2.8: $\theta_0 = 0.95$

Fig. 2.8-2.12: Power of GMM-MLR(θ) when using conditional asymptotic (solid line) and bootstrapped (dashed line) critical values and LM(θ) (dashed-dotted line).

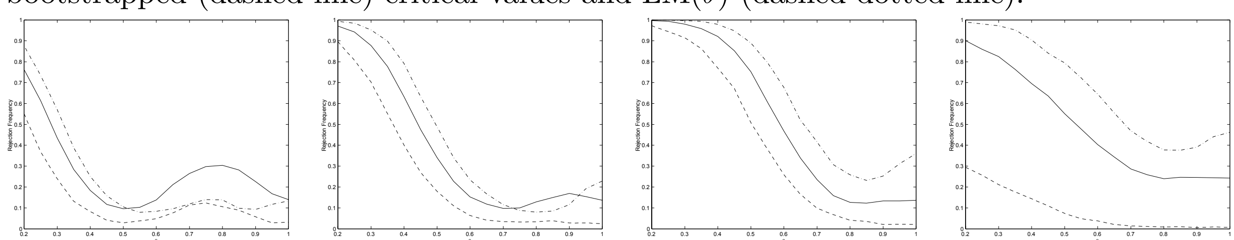

Fig 2.9: $\theta_0 = 0.5$ Fig. 2.10: $\theta_0 = 0.7$ Fig. 2.11: $\theta_0 = 0.9$ Fig. 2.12: $\theta_0 = 0.95$

The size distortions of GMM-MLR(θ) exceed those of S(θ) and KLM(θ) when we use the critical values that stem from its conditional limiting distribution. The bootstrap reduces these size distortions which, however, remain larger than for S(θ) and KLM(θ). The size distortions for KLM(θ) are in general the smallest both when using asymptotic and bootstrap critical values.

The size distortions of LM(θ) show the sensitivity of its distribution to the value of θ_0. Table 1 clearly shows that the size distortions increase when θ_0 gets closer to one. An increase of the sample size for the same value of θ_0 decreases the size distortion of LM(θ). The same results can be shown for other statistics whose distributions are sensitive to the value of θ_0, like, for example, the Wald t-statistic. For reasons of brevity, we refrain from showing these results.

Panel 3: Power curves for $T = 6,\ N = 250$.

Fig. 3.1-3.4: Power of S(θ) when using asymptotic (solid line), bootstrapped (dashed line) and Edgeworth-corrected (dashed-dotted line) critical values.

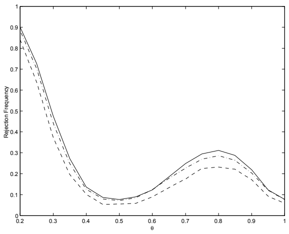

Fig 3.1: $\theta_0 = 0.5$

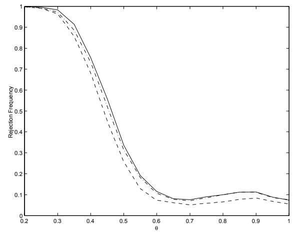

Fig. 3.2: $\theta_0 = 0.7$

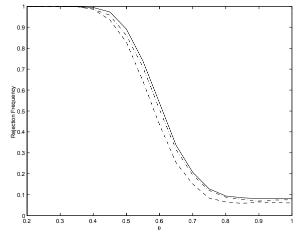

Fig. 3.3: $\theta_0 = 0.9$

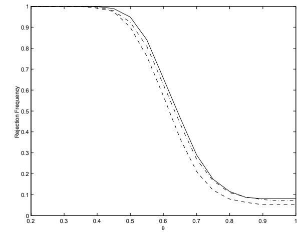

Fig. 3.4: $\theta_0 = 0.95$

Fig. 3.5-3.8: Power of KLM(θ) when using asymptotic (solid line), bootstrapped (dashed line), with resampled $q_i(\theta)$ (dotted line) and Edgeworth-corrected (dashed-dotted line) critical values.

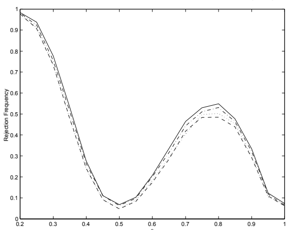

Fig 3.5: $\theta_0 = 0.5$

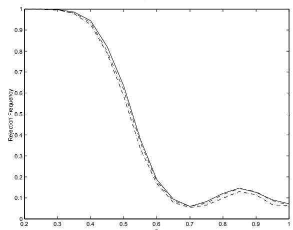

Fig. 3.6: $\theta_0 = 0.7$

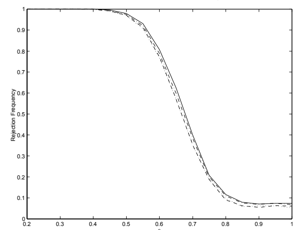

Fig. 3.7: $\theta_0 = 0.9$

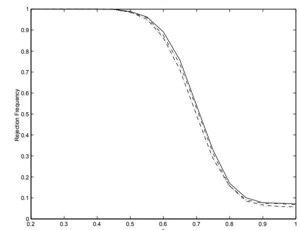

Fig. 3.8: $\theta_0 = 0.95$

Fig. 3.8-3.12: Power of GMM-MLR(θ) when using conditional asymptotic (solid line) and bootstrapped (dashed line) critical values and LM(θ) (dashed-dotted line).

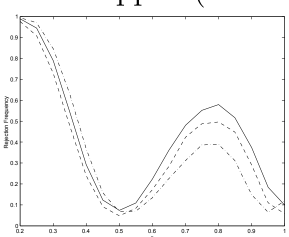

Fig 3.9: $\theta_0 = 0.5$

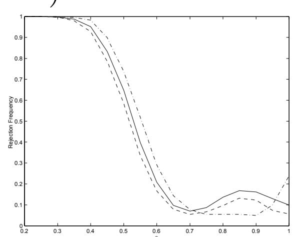

Fig. 3.10: $\theta_0 = 0.7$

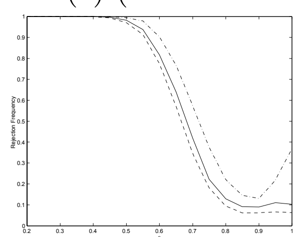

Fig. 3.11: $\theta_0 = 0.9$

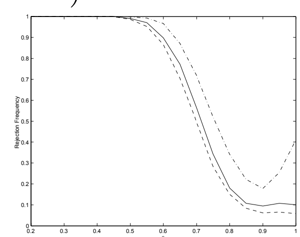

Fig. 3.12: $\theta_0 = 0.95$

Power comparison To further analyse the performance of the different statistics, we compare the power of the different statistics for the three different values of N : 50, 100 and 250, and four different values of θ_0 : 0.5, 0.7, 0.9 and 0.95, that were used in Table 1. Panel 1-3 show these different power curves. All Panels use a value of T equal to 6 while $N = 50$ in Panel 1, $N = 100$ in Panel 2 and $N = 250$ in Panel 3.

The power curves in Panel 1, all reveal the large size distortions of the different statistics when we use the asymptotic critical values. The usage of the Edgeworth-corrected critical values shifs the power curve downwards but, as already shown in Table 1, not enough to completely remove the size distortions. The usage of bootstrap critical values makes the statistics too conservative but the size distortion is much smaller than when using the Edgeworth-corrected critical values. The power curves show that the statistics have power when using the bootstrap critical values.

The power curves of KLM(θ) in Panel 1 when using the bootstrap and bootstrap with resampled $q_i(\theta)$'s are almost indistinguisable which is in line with Table 1 and shows that there is not need to resample $q_i(\theta)$.

When $\theta = 1$, the moment conditions do not identify θ and the power and size of the different statistics should coincide. This explains the peculiar shape of the power curves in Panel 1. It is interesting to see that the size distortions at $\theta = \theta_0$ are the same as at $\theta = 1$ of all statistics except LM(θ) which is the only statistic whose distribution depends on the value of θ.

Panel 1 shows that the size distortions of all statistics except LM(θ) just depend on N and not on θ_0. For LM(θ_0), Fig. 1.9-1.12 nicely show the increase of the size distortion when we increase θ_0.

In Panel 2, all size distortions have decreased compared to Panel 1. Besides that most general findings from Panel 1 appear in Panel 2 as well: bootstrap power curves lie below the power curves that result from using the asymptotic critical value, resampling $q_i(\theta)$ for the bootstrapped power curve of KLM(θ) does not change it, power and size coincide and $\theta = \theta_0$ and $\theta = 1$ except for LM(θ), size distortion of all statistics except LM(θ) does not depend on θ_0.

Panel 3 shows that the power curves that result from using a different critical value become almost indistinguishable when $N = 250$. The usage of bootstrap critical values still leads to a smaller size distortion but the size distortion from using the asymptotic critical value is rather small as well. It is interesting to see that the size distortion of LM(θ) is now small for $\theta_0 = 0.5$ but is still very large for $\theta_0 = 0.95$ which shows the sensitivity of the distribution of LM(θ) to the value of θ_0.

7 Extensions

Multiple parameters and subsets Sofar we have just been concerned with testing a parameter vector that consists of only one element. This has been done for expository purposes and the results extend to a vector of multiple parameters as well. This holds true since one of the main results of the paper that the N^{-1} approximation errors just result from the covariance matrix estimator extend towards a multiple parameter setting when Assumption 1b is extended appropriately. We refrain from showing this result since it involves a lot of additional notation as we have to use vectorization results for matrices.

The extension to tests on subsets of the parameters is less straightforward because no results on the limiting distributions of such tests without assuming that the partialled out parameters are strongly identified have been derived for GMM. Kleibergen (2006b) shows that such tests are conservative in the linear instrumental variables regression model when we use the asymptotic critical values from the limiting distributions that apply when the partialled out parameters are well identified.

Dependent observations If the moment equations are dependent, the Eicker-White covariance matrix estimator (4) is not consistent. When we can employ a consistent covariance matrix estimator with a parametric convergence rate as in Hall and Horowitz (1996), the higher order efficiency of the bootstrap should extend to the case of dependent moment equations. To account for the dependence between the moments, the block-bootstrap should be used such that similar to Hall and Horowitz (1996), a correction factor has to be incorporated in the bootstrapped statistics to take account of the difference between the covariance matrices of the block bootstrap moments and the orginal moments. In order to do so, Assumption 1 has to be adapted and further regularity conditions have to be imposed, see *e.g.* Hall and Horowitz (1996).

8 Conclusions

We show that usage of Edgeworth-corrected critical values or bootstrapped critical values lead to a higher order improvement of the approximation of the finite sample distribution of weak instrument robust GMM statistics compared to usage of critical values that stem from its limiting distribution. The bootstrapped critical values are obtained from just resampling the moments under the hypothesis of interest since the covariance matrix estimator is the sole contributor to the approximation error of the highest order. It is therefore not helpful to resample anything more than the moments.

Appendix

A. Lemma 1. *When $\theta = \theta$ and Assumption 1 holds, the higher order decomposition of $\hat{V}_{ff}(\theta)^{-1}$ reads*

$$\hat{V}_{ff}(\theta)^{-1} = V_{ff}(\theta)^{-1} - V_{ff}(\theta)^{-1}\tfrac{1}{N}\left(\sum_{i=1}^{N}\bar{f}_i(\theta)\bar{f}_i(\theta)' - V_{ff}(\theta)\right)V_{ff}(\theta)^{-1} + O_p(\tfrac{1}{N}).$$

Proof. To obtain the higher order specification of $\hat{V}_{ff}(\theta)^{-1}$, we specify it as

$$\begin{aligned}\hat{V}_{ff}(\theta)^{-1} = & \left[V_{ff}(\theta) + \left(\hat{V}_{ff}(\theta) - V_{ff}(\theta)\right)\right]^{-1} \\ = & V_{ff}(\theta)^{-1} - V_{ff}(\theta)^{-1}\tfrac{1}{N}\left(\sum_{i=1}^{N}\bar{f}_i(\theta)\bar{f}_i(\theta)' - V_{ff}(\theta)\right)V_{ff}(\theta)^{-1} + O_p(\tfrac{1}{N}),\end{aligned}$$

where the $O_p(\frac{1}{N})$ order of the remainder term results from the $\sqrt{N}$ convergence rate of the Eicker-White covariance matrix estimator. ■

Lemma 2. *When $\theta = \theta$ and Assumption 1 holds, the higher order specification of $\hat{D}_N(\theta, Y)$ reads:*

$$\hat{D}_N(\theta, Y) = D_N(\theta, Y) - \hat{V}_{\theta f}(\theta)V_{ff}(\theta_0)^{-1}f_N(\theta, Y) + O_p(\tfrac{1}{N}),$$

where $\hat{V}_{\theta f}(\theta) = \hat{V}_{qf}(\theta) - V_{qf}(\theta)V_{ff}(\theta)^{-1}\hat{V}_{ff}(\theta) = \frac{1}{N}\sum_{i=1}^{N} d_i(\theta)f_i(\theta)' - D_N(\theta, Y)f_N(\theta, Y)'$.

Proof. To obtain the higher order specification, we specify $\hat{D}_N(\theta, Y)$ as

$$\begin{aligned}&\hat{D}_N(\theta, Y) = q_N(\theta, Y) - \hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}f_N(\theta, Y) \\ &= q_N(\theta, Y) - V_{qf}(\theta)V_{ff}(\theta)^{-1}f_N(\theta, Y) + \left[V_{qf}(\theta)V_{ff}(\theta)^{-1} - \hat{V}_{qf}(\theta)\hat{V}_{ff}(\theta)^{-1}\right]f_N(\theta, Y) \\ &= D_N(\theta, Y) - \left[\hat{V}_{qf}(\theta) - V_{qf}(\theta)V_{ff}(\theta)^{-1}\hat{V}_{ff}(\theta)\right]\hat{V}_{ff}(\theta)^{-1}f_N(\theta, Y) \\ &= D_N(\theta, Y) - \hat{V}_{\theta f}(\theta)V_{ff}(\theta_0)^{-1}f_N(\theta, Y) + O_p(\tfrac{1}{N}),\end{aligned}$$

where we used Lemma 1 for the fourth equation. The order of the remainder terms results from the $\sqrt{N}$ convergence rate of the Eicker-White covariance matrix estimator. ■

Lemma 3. *When $\theta = \theta$ and Assumption 1 holds, the higher order specification of the score vector $s_N(\theta, Y) = \hat{D}_N(\theta, Y)'\hat{V}_{ff}(\theta_0)^{-1}f_N(\theta, Y)$ reads*

$$\begin{aligned}\sqrt{N}s_N(\theta, Y) = & \sqrt{N}D_N(\theta, Y)'V_{ff}(\theta)^{-1}f_N(\theta, Y) - \sqrt{N}D_N(\theta, Y)'V_{ff}(\theta)^{-1}\left(\hat{V}_{ff}(\theta) - V_{ff}(\theta)\right) \\ & V_{ff}(\theta)^{-1}f_N(\theta, Y) - \sqrt{N}f_N(\theta, Y)'V_{ff}(\theta)^{-1}\hat{V}_{\theta f}(\theta)'V_{ff}(\theta)^{-1}f_N(\theta, Y) + O_p(\tfrac{1}{N}).\end{aligned}$$

Proof. The decomposition of the score vector is as follows

$$
\begin{aligned}
&\sqrt{N}s_N(\theta,Y)\\
&= \sqrt{N}\hat{D}_N(\theta,Y)'\hat{V}_{ff}(\theta_0)^{-1}f_N(\theta,Y)\\
&= \sqrt{N}\left[D_N(\theta,Y)-\hat{V}_{\theta f}(\theta)\left\{V_{ff}(\theta)^{-1}-V_{ff}(\theta)^{-1}\left(\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right)V_{ff}(\theta)^{-1}\right\}f_N(\theta,Y)\right]'\\
&\quad\left\{V_{ff}(\theta)^{-1}-V_{ff}(\theta)^{-1}\left(\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right)V_{ff}(\theta_0)^{-1}\right\}f_N(\theta,Y)+O_p(\tfrac{1}{N})\\
&= \sqrt{N}D_N(\theta,Y)'V_{ff}(\theta)^{-1}f_N(\theta,Y)-\sqrt{N}D_N(\theta,Y)'V_{ff}(\theta)^{-1}\left(\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right)V_{ff}(\theta)^{-1}f_N(\theta,Y)-\\
&\quad\sqrt{N}f_N(\theta,Y)'V_{ff}(\theta)^{-1}\hat{V}_{\theta f}(\theta)'V_{ff}(\theta)^{-1}f_N(\theta,Y)+O_p(\tfrac{1}{N}).
\end{aligned}
$$

where the order of the remainder term results from the convergence speed of the covariance matrix estimator. ■

Lemma 4. *When $\theta=\theta$ and Assumption 1 holds, the higher order specification of* $\left[\hat{D}_N(\theta,Y)'\hat{V}_{ff}(\theta)^{-1}\hat{D}_N(\theta,Y)\right]^{-1}$ *reads:*

$$
\begin{aligned}
&\left[\hat{D}_N(\theta,Y)'\hat{V}_{ff}(\theta)^{-1}\hat{D}_N(\theta,Y)\right]^{-1}\\
&= [D_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y)]^{-1}+[D_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y)]^{-1}\\
&\quad\left\{D_N(\theta,Y)'V_{ff}(\theta)^{-1}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]V_{ff}(\theta)^{-1}D_N(\theta,Y)+\right.\\
&\quad\left.2D_N(\theta,Y)'V_{ff}(\theta)^{-1}\hat{V}_{\theta f}(\theta)V_{ff}(\theta)^{-1}f_N(\theta,Y)\right\}[D_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y)]^{-1}+O_p(\tfrac{1}{N}).
\end{aligned}
$$

Proof. To construct the higher order expression of $\left[\hat{D}_N(\theta,Y)'\hat{V}_{ff}(\theta)^{-1}\hat{D}_N(\theta,Y)\right]^{-1}$, we first construct that of $\hat{D}_N(\theta,Y)'\hat{V}_{ff}(\theta)^{-1}\hat{D}_N(\theta,Y)$:

$$
\begin{aligned}
&\hat{D}_N(\theta,Y)'\hat{V}_{ff}(\theta)^{-1}\hat{D}_N(\theta,Y)\\
&= \left[D_N(\theta,Y)-\hat{V}_{\theta f}(\theta)\left\{V_{ff}(\theta)^{-1}-V_{ff}(\theta)^{-1}\left(\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right)V_{ff}(\theta)^{-1}\right\}f_N(\theta,Y)\right]'\\
&\quad\left\{V_{ff}(\theta)^{-1}-V_{ff}(\theta)^{-1}\left(\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right)V_{ff}(\theta)^{-1}\right\}\\
&\quad\left[D_N(\theta,Y)-\hat{V}_{\theta f}(\theta)\left\{V_{ff}(\theta)^{-1}-V_{ff}(\theta)^{-1}\left(\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right)V_{ff}(\theta)^{-1}\right\}f_N(\theta,Y)\right]\\
&= D_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y)-D_N(\theta,Y)'V_{ff}(\theta)^{-1}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]V_{ff}(\theta)^{-1}D_N(\theta,Y)-\\
&\quad 2D_N(\theta,Y)'V_{ff}(\theta)^{-1}\hat{V}_{\theta f}(\theta)V_{ff}(\theta)^{-1}f_N(\theta,Y)+O_p(\tfrac{1}{N}).
\end{aligned}
$$

Using the higher order decomposition of $\hat{V}_{ff}(\theta)^{-1}$ from Lemma 1, the higher order expression for $\left[\hat{D}_N(\theta,Y)'\hat{V}_{ff}(\theta)^{-1}\hat{D}_N(\theta,Y)\right]^{-1}$ results. ■

Lemma 5. *When $\theta=\theta$ and Assumption 1 holds, the higher order specification of*

$$\sqrt{N}f_N(\theta,Y)'\hat{V}_{ff}(\theta)^{-1}q_N(\theta,Y)$$

reads:

$$
\begin{aligned}
&\sqrt{N}f_N(\theta,Y)'\hat{V}_{ff}(\theta)^{-1}q_N(\theta,Y)\\
&= \sqrt{N}f_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y)+\sqrt{N}f_N(\theta,Y)'V_{ff}(\theta)^{-1}V_{qf}(\theta)V_{ff}(\theta)^{-1}f_N(\theta,Y)-\\
&\quad\sqrt{N}f_N(\theta,Y)'V_{ff}(\theta)^{-1}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]V_{ff}(\theta)^{-1}D_N(\theta,Y)+O_p(\tfrac{1}{N}).
\end{aligned}
$$

Proof. The higher order decomposition results from the specification of $D_N(\theta,Y)$ and $\hat{V}_{ff}(\theta)^{-1}$:

$$\begin{array}{l} \sqrt{N}f_N(\theta,Y)'\hat{V}_{ff}(\theta)^{-1}q_N(\theta,Y) \\ \quad = \quad \sqrt{N}f_N(\theta,Y)'\hat{V}_{ff}(\theta)^{-1}D_N(\theta,Y)+\sqrt{N}f_N(\theta,Y)'\hat{V}_{ff}(\theta)^{-1}V_{qf}(\theta)V_{ff}(\theta)^{-1}f_N(\theta,Y) \\ \quad = \quad \sqrt{N}f_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y)+\sqrt{N}f_N(\theta,Y)'V_{ff}(\theta)^{-1}V_{qf}(\theta)V_{ff}(\theta)^{-1}f_N(\theta,Y)- \\ \qquad \sqrt{N}f_N(\theta,Y)'V_{ff}(\theta)^{-1}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]V_{ff}(\theta)^{-1}D_N(\theta,Y)- \\ \qquad \sqrt{N}f_N(\theta,Y)'\hat{V}_{ff}(\theta)^{-1}V_{qf}(\theta)V_{ff}(\theta)^{-1}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]V_{ff}(\theta)^{-1}f_N(\theta,Y)+O_p(\frac{1}{N}). \end{array}$$

■

Lemma 6. *When $\theta=\theta$ and Assumption 1 holds, the higher order specification of $\left[q_N(\theta,Y)'\hat{V}_{ff}(\theta)^{-1}q_N(\theta,Y)\right]^{-1}$ reads:*

$$\begin{array}{l} \left[q_N(\theta,Y)'\hat{V}_{ff}(\theta)^{-1}q_N(\theta,Y)\right]^{-1} \\ \quad = \quad (D_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y))^{-1}-(D_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y))^{-1} \\ \qquad \{2D_N(\theta,Y)'V_{ff}(\theta)^{-1}V_{qf}(\theta)V_{ff}(\theta)^{-1}f_N(\theta,Y)+ \\ \qquad f_N(\theta,Y)'V_{ff}(\theta)^{-1}V_{qf}(\theta)'V_{ff}(\theta)^{-1}V_{qf}(\theta)V_{ff}(\theta)^{-1}f_N(\theta,Y)-D_N(\theta,Y)'V_{ff}(\theta)^{-1} \\ \qquad \left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]V_{ff}(\theta)^{-1}D_N(\theta,Y)-2D_N(\theta,Y)'V_{ff}(\theta)^{-1}V_{qf}(\theta)V_{ff}(\theta)^{-1} \\ \qquad \left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]V_{ff}(\theta)^{-1}f_N(\theta,Y)-f_N(\theta,Y)'V_{ff}(\theta)^{-1}V_{qf}(\theta)'V_{ff}(\theta)^{-1} \\ \qquad \left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]V_{ff}(\theta)^{-1}V_{qf}(\theta)V_{ff}(\theta)^{-1}f_N(\theta,Y)\Big\}(D_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y))^{-1}+ \\ \qquad O_p(\frac{1}{N}). \end{array}$$

Proof. We obtain the higher order specification of $\left[q_N(\theta,Y)'\hat{V}_{ff}(\theta)^{-1}q_N(\theta,Y)\right]^{-1}$ from that of $q_N(\theta,Y)'\hat{V}_{ff}(\theta)^{-1}q_N(\theta,Y)$:

$$\begin{array}{l} q_N(\theta,Y)'\hat{V}_{ff}(\theta)^{-1}q_N(\theta,Y) \\ \quad = \quad [D_N(\theta,Y)+V_{qf}(\theta)V_{ff}(\theta)^{-1}f_N(\theta,Y)]'\left[V_{ff}(\theta)-V_{ff}(\theta)^{-1}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]V_{ff}(\theta)^{-1}\right] \\ \qquad [D_N(\theta,Y)+V_{qf}(\theta)V_{ff}(\theta)^{-1}f_N(\theta,Y)]+O_p(\frac{1}{N}) \\ \quad = \quad D_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y)+2D_N(\theta,Y)'V_{ff}(\theta)^{-1}V_{qf}(\theta)V_{ff}(\theta)^{-1}f_N(\theta,Y)+ \\ \qquad f_N(\theta,Y)'V_{ff}(\theta)^{-1}V_{qf}(\theta)'V_{ff}(\theta)^{-1}V_{qf}(\theta)V_{ff}(\theta)^{-1}f_N(\theta,Y)-D_N(\theta,Y)'V_{ff}(\theta)^{-1} \\ \qquad \left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]V_{ff}(\theta)^{-1}D_N(\theta,Y)-2D_N(\theta,Y)'V_{ff}(\theta)^{-1}V_{qf}(\theta)V_{ff}(\theta)^{-1} \\ \qquad \left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]V_{ff}(\theta)^{-1}f_N(\theta,Y)-f_N(\theta,Y)'V_{ff}(\theta)^{-1}V_{qf}(\theta)'V_{ff}(\theta)^{-1} \\ \qquad \left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]V_{ff}(\theta)^{-1}V_{qf}(\theta)V_{ff}(\theta)^{-1}f_N(\theta,Y)+O_p(\frac{1}{N}). \end{array}$$

■

Lemma 7. The conditional expectations of the higher order elements of KLM(θ_0) read

$$\begin{array}{rl} \mathrm{E}(KLM_j|D_N(\theta_0,Y)= & \int \mathrm{E}(KLM_j|d_1(\theta_0),\ldots,d_N(\theta_0)) \\ & p_{(d_1,\ldots,d_{N-1})}(v_1,\ldots,v_{N-1}|D_N(\theta_0,Y))dv_1\ldots dv_{N-1}, \end{array}$$

for $j=0,1,2$ and where $p_{(d_1,\ldots,d_{N-1})}(v_1,\ \ldots,v_{N-1}|D_N(\theta_0,Y))$ is the conditional density of $(d_1(\theta_0),\ldots,d_{N-1}(\theta_0))$ given $D_N(\theta_0,Y)$.

Proof.

$$
\begin{aligned}
\mathrm{E}(KLM_j|D_N(\theta_0,Y) = \;& \textstyle\iint KLM_j p_{(f_1,\ldots,f_N,d_1,\ldots,d_{N-1})}(u_1,\ldots,u_N,v_1,\ldots,v_{N-1}|D_N(\theta_0,Y)) \\
& du_1\ldots du_N dv_1\ldots dv_{N-1} \\
= \;& \textstyle\int \left[\int KLM_j p_{(f_1,\ldots,f_N)}(u_1,\ldots,u_N|d_1,\ldots,d_N)du_1\ldots du_N\right] \\
& p_{(d_1,\ldots,d_{N-1})}(v_1,\ldots,v_{N-1}|D_N(\theta_0,Y))dv_1\ldots dv_{N-1} \\
= \;& \textstyle\int \mathrm{E}(KLM_j|d_1(\theta_0),\ldots,d_N(\theta_0)) \\
& p_{(d_1,\ldots,d_{N-1})}(v_1,\ldots,v_{N-1}|D_N(\theta_0,Y))dv_1\ldots dv_{N-1},
\end{aligned}
$$

where $p_{(f_1,\ldots,f_N,d_1,\ldots,d_{N-1})}(u_1,\ldots,u_N,v_1,\ldots,v_{N-1}|D_N(\theta_0,Y))$ is the conditional density of $(f_1(\theta_0),$ $\ldots,f_N(\theta_0),d_1(\theta_0),\ldots,d_{N-1}(\theta_0))$ given $D_N(\theta_0,Y)$ and we used that conditioning on $(d_1(\theta_0),\ldots,$ $d_{N-1}(\theta_0),D_N(\theta_0,Y))$ is identical to conditioning on $(d_1(\theta_0),\ldots,d_{N-1}(\theta_0),d_N(\theta_0))$ since $D_N(\theta_0,Y)=$ $\frac{1}{N}\sum_{i=1}^N d_i(\theta_0)$. ■

Lemma 8. The conditional expectations of the higher order components of $\mathrm{S}(\theta_0)$ and $\mathrm{KLM}(\theta_0)$ are such that

$$
\begin{aligned}
\mathrm{E}(S_j) = \;& \textstyle\int S_j p_{f_N(\theta_0,Y),\hat V_{ff}(\theta_0))}(u,v)dudv \\
\mathrm{E}(KLM_j|D_N(\theta_0,Y)) = \;& \textstyle\iiint KLM_j p_{(f_N(\theta_0,Y),\hat V_{ff}(\theta_0),\hat V_{\theta f}(\theta_0))}(u,v,w|D_N(\theta_0,Y))dudvdw,
\end{aligned}
$$

for $j=0,1,2$ and where $p_{(f_N(\theta_0,Y),\hat V_{ff}(\theta_0))}(u,v)$ and $p_{(f_N(\theta_0,Y),\hat V_{ff}(\theta_0),\hat V_{\theta f}(\theta_0))}(u,v,w|D_N(\theta_0,Y))$ are the joint sampling distribution of $f_N(\theta_0,Y)$ and $\hat V_{ff}(\theta_0)$ and the conditional joint sampling distribution of $(f_N(\theta_0,Y),\hat V_{ff}(\theta_0),\hat V_{\theta f}(\theta_0))$ given $D_N(\theta_0,Y)$ in a sample of size N.

Proof.

$$
\begin{aligned}
\mathrm{E}(S_j) = \;& \textstyle\int S_j(f_N(\theta_0,Y),\hat V_{ff}(\theta_0))p_{(f_1,\ldots,f_N)}(u_1,\ldots,u_N)du_1\ldots du_N \\
= \;& \textstyle\iint S_j(f_N(\theta_0,Y),\hat V_{ff}(\theta_0))p_{(f_N(\theta_0,Y),\hat V_{ff}(\theta_0),f_1,\ldots,f_{N_j})}(u,v,u_1,\ldots,u_{N_j})dudvdu_1\ldots du_{N_j} \\
= \;& \textstyle\int p_{(f_1,\ldots,f_{N_j})}(u,v,u_1,\ldots,u_{N_j}|f_N(\theta_0,Y),\hat V_{ff}(\theta_0)) \\
& \left[\textstyle\int S_j(f_N(\theta_0,Y),\hat V_{ff}(\theta_0))p_{(f_N(\theta_0,Y),\hat V_{ff}(\theta_0))}(u,v)dudv\right]du_1\ldots du_{N_j} \\
= \;& \textstyle\int S_1 p_{(f_N(\theta_0,Y),\hat V_{ff}(\theta_0))}(u,v)dudv
\end{aligned}
$$

where $S_j(f_N(\theta_0,Y),\hat V_{ff}(\theta_0))$ indicates that S_j is a function of $(f_N(\theta_0,Y),\hat V_{ff}(\theta_0))$, $p_{(f_N(\theta_0,Y),\hat V_{ff}(\theta_0),f_1,\ldots,f_{N_j})}(u,v,u_1,\ldots,u_{N_j})$ is the joint density of $(f_N(\theta_0,Y),\hat V_{ff}(\theta_0),f_1,\ldots,f_{N_j})$ with N_j a scalar smaller than N which is such that there is an invertible relationship between $(f_N(\theta_0,Y),\hat V_{ff}(\theta_0),f_1,\ldots,f_{N_j})$ and $(f_1,\ldots,f_N)$, $p_{(f_1,\ldots,f_{N_j})}(u,v,u_1,\ldots,u_{N_j}|f_N(\theta_0,Y),\hat V_{ff}(\theta_0))$ is the conditional density of $(f_1,\ldots,f_{N_j})$ given $(f_N(\theta_0,Y),\hat V_{ff}(\theta_0))$.

$$
\begin{aligned}
& \mathrm{E}(KLM_j|D_N(\theta_0,Y)) \\
= \;& \textstyle\int KLM_j(f_N(\theta_0,Y),\hat V_{ff}(\theta_0),\hat V_{\theta f}(\theta_0)) \\
& p_{(f_1,\ldots,f_N,d_1,\ldots,d_{N-1})}(u_1,\ldots,u_N,v_1,\ldots,v_{N-1}|D_N(\theta_0,Y))du_1\ldots du_N dv_1\ldots dv_{N-1} \\
= \;& \textstyle\int KLM_j(f_N(\theta_0,Y),\hat V_{ff}(\theta_0),\hat V_{\theta f}(\theta_0)) \\
& p_{(f_N(\theta_0,Y),\hat V_{ff}(\theta_0),\hat V_{\theta f}(\theta_0),f_1,\ldots,f_{N_{j_1}},d_1,\ldots,d_{N_{j2}})}(u,v,w,u_1,\ldots,u_{N_{j1}},v_1,\ldots,v_{N_{j_2}}|D_N(\theta_0,Y)) \\
& dudvdwdu_1\ldots du_{N_{j_1}}dv_1\ldots dv_{N_{j_2}-1} \\
= \;& \textstyle\int p_{(f_1,\ldots,f_{N_{j_1}},d_1,\ldots,d_{N_{j2}})}(u,v,w,u_1,\ldots,u_{N_{j1}},v_1,\ldots,v_{N_{j_2}}|f_N(\theta_0,Y),\hat V_{ff}(\theta_0),\hat V_{\theta f}(\theta_0),D_N(\theta_0,Y)) \\
& \left[\textstyle\int KLM_j(f_N(\theta_0,Y),\hat V_{ff}(\theta_0),\hat V_{\theta f}(\theta_0))p_{(f_N(\theta_0,Y),\hat V_{ff}(\theta_0),\hat V_{\theta f}(\theta_0))}(u,v,w|D_N(\theta_0,Y))dudvdw\right] \\
& du_1\ldots du_{N_{j_1}}dv_1\ldots dv_{N_{j_2}-1} \\
= \;& \textstyle\int KLM_j(f_N(\theta_0,Y),\hat V_{ff}(\theta_0),\hat V_{\theta f}(\theta_0))p_{(f_N(\theta_0,Y),\hat V_{ff}(\theta_0),\hat V_{\theta f}(\theta_0))}(u,v,w|D_N(\theta_0,Y))dudvdw
\end{aligned}
$$

where $KLM_j(f_N(\theta_0,Y),\hat{V}_{ff}(\theta_0),\hat{V}_{\theta f}(\theta_0))$ indicates that KLM_j is a function of $(f_N(\theta_0,Y),\hat{V}_{ff}(\theta_0),$ $\hat{V}_{\theta f}(\theta_0))$, $p_{(f_1,\ldots,f_N,d_1,\ldots,d_{N-1})}(u_1,\ldots,u_N,v_1,\ldots,v_{N-1}|D_N(\theta_0,Y))$ is the conditional density of $(f_1,\ldots,f_N,d_1,\ldots,d_{N-1})$ given $D_N(\theta_0,Y)$, N_{j_1} and N_{j_2} are scalars which are such that given $D_N(\theta_0,Y)$ there is an invertible relationship between $(f_N(\theta_0,Y),\hat{V}_{ff}(\theta_0),\hat{V}_{\theta f}(\theta_0),f_1,\ldots,f_{N_{j_1}},$ $d_1,\ldots,d_{N_{j2}})$ and $(f_1,\ldots,f_N,d_1,\ldots,d_{N-1})$, $p_{(f_N(\theta_0,Y),\hat{V}_{ff}(\theta_0),\hat{V}_{\theta f}(\theta_0),f_1,\ldots,f_{N_{j_1}},d_1,\ldots,d_{N_{j2}})}(u,v,w,u_1,$ $\ldots,u_{N_{j1}},v_1,\ldots,v_{N_{j_2}}|D_N(\theta_0,Y))$ is the conditional density of $(f_N(\theta_0,Y),\hat{V}_{ff}(\theta_0),\hat{V}_{\theta f}(\theta_0),f_1,\ldots,$ $f_{N_{j_1}},d_1,\ldots,d_{N_{j2}})$ given $D_N(\theta_0,Y)$, $p_{(f_1,\ldots,f_{N_{j_1}},d_1,\ldots,d_{N_{j2}})}(u,v,w,u_1,\ldots,u_{N_{j1}},v_1,\ldots,v_{N_{j_2}}|f_N(\theta_0,Y),$ $\hat{V}_{ff}(\theta_0),\hat{V}_{\theta f}(\theta_0),D_N(\theta_0,Y))$ is the conditional density of $(f_1,\ldots,f_{N_{j_1}},d_1,\ldots,d_{N_{j2}})$ given $(f_N(\theta_0,Y),\hat{V}_{ff}(\theta_0),\hat{V}_{\theta f}(\theta_0),D_N(\theta_0,Y))$. ■

B. Proof of Theorem 1.

a. S-statistic: We use the decomposition of $\hat{V}_{ff}(\theta)^{-1}$ from Lemma 1 to obtain the higher order components of the S-statistic:

$$\begin{aligned}\mathrm{S}(\theta)= \;& Nf_N(\theta,Y)'V_{ff}(\theta)^{-1}f_N(\theta,Y)-f_N(\theta,Y)'V_{ff}(\theta)^{-1}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]V_{ff}(\theta)^{-1}f_N(\theta,Y)+\\ & O_p(\tfrac{1}{N^2}).\end{aligned}$$

We determine the order of the different components by constructing their expectations. It is directly obvious that the expectation of $Nf_N(\theta,Y)'V_{ff}(\theta)^{-1}f_N(\theta,Y)$ is equal to k since $\mathrm{E}(f_i(\theta)f_j(\theta)')=V_{ff}(\theta)$ when $i=j$ and equals zero when $i\neq j$.

The expectation of $Nf_N(\theta,Y)'V_{ff}(\theta)^{-1}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]V_{ff}(\theta)^{-1}f_N(\theta,Y)$ results from

$$\begin{aligned}
&\mathrm{E}\left[Nf_N(\theta,Y)'V_{ff}(\theta)^{-1}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]V_{ff}(\theta)^{-1}f_N(\theta,Y)\right]\\
&=\mathrm{E}\left[\tfrac{1}{N^2}\textstyle\sum_{i_1}f_{i_1}'V_{ff}^{-1}\sum_{i_2}\left(\bar{f}_{i_2}\bar{f}_{i_2}'-V_{ff}\right)V_{ff}^{-1}\sum_{i_3}f_{i_3}\right]\\
&=\mathrm{E}\left[\tfrac{1}{N^2}\textstyle\sum_{i_1}f_{i_1}'V_{ff}^{-1}\sum_{i_2}\left(f_{i_2}f_{i_2}'-V_{ff}\right)V_{ff}^{-1}\sum_{i_3}f_{i_3}\right]-\\
&\quad\mathrm{E}\left[\tfrac{1}{N^3}\textstyle\sum_{i_1}f_{i_1}'V_{ff}^{-1}\sum_{i_2}f_{i_2}\sum_{i_4}f_{i_4}'V_{ff}^{-1}\sum_{i_3}f_{i_3}\right]\\
&=\mathrm{E}\left[\tfrac{1}{N^2}\textstyle\sum_{i_1}\sum_{i_2\neq i_1}f_{i_1}'V_{ff}^{-1}\left(f_{i_2}f_{i_2}'-V_{ff}\right)V_{ff}^{-1}f_{i_1}\right]+\mathrm{E}\left[\tfrac{1}{N^2}\textstyle\sum_{i_1}f_{i_1}'V_{ff}^{-1}\left(f_{i_1}f_{i_1}'-V_{ff}\right)V_{ff}^{-1}f_{i_1}\right]-\\
&\quad\mathrm{E}\left[\tfrac{1}{N^3}\textstyle\sum_{i_1}\sum_{i_2\neq i_1}f_{i_1}'V_{ff}^{-1}f_{i_1}f_{i_2}'V_{ff}^{-1}f_{i_2}\right]-\mathrm{E}\left[\tfrac{1}{N^3}\textstyle\sum_{i_1}\sum_{i_2\neq i_1}f_{i_1}'V_{ff}^{-1}f_{i_2}f_{i_1}'V_{ff}^{-1}f_{i_2}\right]-\\
&\quad\mathrm{E}\left[\tfrac{1}{N^3}\textstyle\sum_{i_1}\sum_{i_2\neq i_1}f_{i_1}'V_{ff}^{-1}f_{i_2}f_{i_2}'V_{ff}^{-1}f_{i_1}\right]-\mathrm{E}\left[\tfrac{1}{N^3}\textstyle\sum_{i_1}f_{i_1}'V_{ff}^{-1}f_{i_1}f_{i_1}'V_{ff}^{-1}f_{i_1}\right]\\
&=\mathrm{E}\left[\tfrac{1}{N^2}\textstyle\sum_{i_1}f_{i_1}'V_{ff}^{-1}f_{i_1}f_{i_1}'V_{ff}^{-1}f_{i_1}\right]-\mathrm{E}\left[\tfrac{1}{N^2}\textstyle\sum_{i_1}f_{i_1}'V_{ff}^{-1}f_{i_1}\right]-\\
&\quad\mathrm{E}\left[\tfrac{1}{N^3}\textstyle\sum_{i_1}\sum_{i_2\neq i_1}f_{i_1}'V_{ff}^{-1}f_{i_1}f_{i_2}'V_{ff}^{-1}f_{i_2}\right]-\mathrm{E}\left[\tfrac{1}{N^3}\textstyle\sum_{i_1}\sum_{i_2\neq i_1}f_{i_1}'V_{ff}^{-1}f_{i_2}f_{i_1}'V_{ff}^{-1}f_{i_2}\right]-\\
&\quad\mathrm{E}\left[\tfrac{1}{N^3}\textstyle\sum_{i_1}\sum_{i_2\neq i_1}f_{i_1}'V_{ff}^{-1}f_{i_2}f_{i_2}'V_{ff}^{-1}f_{i_1}\right]-\mathrm{E}\left[\tfrac{1}{N^3}\textstyle\sum_{i_1}f_{i_1}'V_{ff}^{-1}f_{i_1}f_{i_1}'V_{ff}^{-1}f_{i_1}\right]\\
&=\tfrac{N-1}{N^2}\mathrm{E}\left[\left(f_i'V_{ff}^{-1}f_i\right)^2\right]-\tfrac{k}{N}-\tfrac{N-1}{N^2}\left[k^2+2k\right],
\end{aligned}$$

where $f_i=f_i(\theta)$, $\bar{f}_i=f_i(\theta)-f_N(\theta,Y)$, $V_{ff}=V_{ff}(\theta)$ and $\sum_{i_1}=\sum_{i_1=1}^{N}$. The above expression consists of second and fourth order moments of f_i. It uses the independence of f_{i_1} and f_{i_2} for $i_1\neq i_2$ and that $\mathrm{E}(f_i)=0$ which explains why no third order moments are present in the expression. The above expression shows that $Nf_N(\theta,Y)'V_{ff}(\theta)^{-1}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]V_{ff}(\theta)^{-1}f_N(\theta,Y)$ is of the order $\frac{1}{N}$ and we can specify $\mathrm{S}(\theta)$ as

$$\mathrm{S}(\theta)=S_0+\tfrac{1}{N}S_1+O_p(\tfrac{1}{N^2}),$$

where $S_0=Nf_N(\theta_0,Y)'V_{ff}(\theta_0)^{-1}f_N(\theta_0,Y)\underset{d}{\rightarrow}\chi^2(k)$, so $\mathrm{E}(S_0)=k$, and $S_1=-N^2f_N(\theta_0,Y)'$ $V_{ff}(\theta_0)^{-1}\left[\hat{V}_{ff}(\theta_0)-V_{ff}(\theta_0)\right]V_{ff}(\theta_0)^{-1}f_N(\theta_0,Y)$, $\mathrm{E}(S_1)=-\frac{N-1}{N}\mathrm{E}\left[\left(f_i'V_{ff}^{-1}f_i\right)^2\right]+k+\frac{N-1}{N}\left[k^2+2k\right].$

The order of the remainder term in the expression of $\mathrm{S}(\theta)$ results from the order of the highest order component that is left out: $f_N(\theta,Y)'V_{ff}(\theta)^{-1}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]V_{ff}(\theta)^{-1}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]$ $V_{ff}(\theta)^{-1}f_N(\theta,Y)$. To determine the order of this component, we construct its expectation:

$$
\begin{aligned}
&\mathrm{E}\left[Nf_N(\theta,Y)'V_{ff}(\theta)^{-1}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]V_{ff}(\theta)^{-1}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]V_{ff}(\theta)^{-1}f_N(\theta,Y)\right]\\
&= \mathrm{E}\left[\tfrac{1}{N^3}\textstyle\sum_{i_1} f_{i_1}'V_{ff}^{-1}\sum_{i_2}\left(\bar{f}_{i_2}\bar{f}_{i_2}'-V_{ff}\right)V_{ff}^{-1}\sum_{i_3}\left(\bar{f}_{i_3}\bar{f}_{i_3}'-V_{ff}\right)V_{ff}^{-1}\sum_{i_4} f_{i_4}\right]\\
&= \mathrm{E}\left[\tfrac{1}{N^3}\textstyle\sum_{i_1} f_{i_1}'V_{ff}^{-1}\sum_{i_2}\left(f_{i_2}f_{i_2}'-V_{ff}\right)V_{ff}^{-1}\sum_{i_3}\left(f_{i_3}f_{i_3}'-V_{ff}\right)V_{ff}^{-1}\sum_{i_4} f_{i_4}\right]-\\
&\quad 2\mathrm{E}\left[\tfrac{1}{N^4}\textstyle\sum_{i_1} f_{i_1}'V_{ff}^{-1}\sum_{i_5} f_{i_5}\sum_{i_6} f_{i_6}'V_{ff}^{-1}\sum_{i_3}\left(f_{i_3}f_{i_3}'-V_{ff}\right)V_{ff}^{-1}\sum_{i_4} f_{i_4}\right]-\\
&\quad \mathrm{E}\left[\tfrac{1}{N^5}\textstyle\sum_{i_1} f_{i_1}'V_{ff}^{-1}\sum_{i_5} f_{i_5}\sum_{i_6} f_{i_6}'V_{ff}^{-1}\sum_{i_7} f_{i_7}\sum_{i_8} f_{i_8}'V_{ff}^{-1}\sum_{i_4} f_{i_4}\right]\\
&= \mathrm{E}\left[\tfrac{1}{N^3}\textstyle\sum_{i_1} f_{i_1}'V_{ff}^{-1}\left(f_{i_1}f_{i_1}'-V_{ff}\right)V_{ff}^{-1}\left(f_{i_1}f_{i_1}'-V_{ff}\right)V_{ff}^{-1}f_{i_1}\right]+\\
&\quad \mathrm{E}\left[\tfrac{1}{N^3}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}'V_{ff}^{-1}\left(f_{i_1}f_{i_1}'-V_{ff}\right)V_{ff}^{-1}\left(f_{i_2}f_{i_2}'-V_{ff}\right)V_{ff}^{-1}f_{i_2}\right]+\\
&\quad \mathrm{E}\left[\tfrac{1}{N^3}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_2}'V_{ff}^{-1}\left(f_{i_1}f_{i_1}'-V_{ff}\right)V_{ff}^{-1}\left(f_{i_2}f_{i_2}'-V_{ff}\right)V_{ff}^{-1}f_{i_1}\right]-\\
&\quad 2\Big\{\mathrm{E}\left[\tfrac{1}{N^4}\textstyle\sum_{i_1} f_{i_1}'V_{ff}^{-1}f_{i_1}f_{i_1}'V_{ff}^{-1}\left(f_{i_1}f_{i_1}'-V_{ff}\right)V_{ff}^{-1}f_{i_1}\right]+\\
&\quad \mathrm{E}\left[\tfrac{1}{N^4}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}'V_{ff}^{-1}f_{i_1}f_{i_2}'V_{ff}^{-1}\left(f_{i_2}f_{i_2}'-V_{ff}\right)V_{ff}^{-1}f_{i_2}\right]+\\
&\quad \mathrm{E}\left[\tfrac{1}{N^4}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}'V_{ff}^{-1}f_{i_2}f_{i_1}'V_{ff}^{-1}\left(f_{i_2}f_{i_2}'-V_{ff}\right)V_{ff}^{-1}f_{i_2}\right]+\\
&\quad \mathrm{E}\left[\tfrac{1}{N^4}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}'V_{ff}^{-1}f_{i_2}f_{i_2}'V_{ff}^{-1}\left(f_{i_2}f_{i_2}'-V_{ff}\right)V_{ff}^{-1}f_{i_1}\right]+\\
&\quad \mathrm{E}\left[\tfrac{1}{N^4}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_2}'V_{ff}^{-1}f_{i_1}f_{i_1}'V_{ff}^{-1}\left(f_{i_2}f_{i_2}'-V_{ff}\right)V_{ff}^{-1}f_{i_2}\right]+\\
&\quad \mathrm{E}\left[\tfrac{1}{N^4}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_2}'V_{ff}^{-1}f_{i_1}f_{i_2}'V_{ff}^{-1}\left(f_{i_2}f_{i_2}'-V_{ff}\right)V_{ff}^{-1}f_{i_1}\right]+\\
&\quad \mathrm{E}\left[\tfrac{1}{N^4}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_2}'V_{ff}^{-1}f_{i_2}f_{i_1}'V_{ff}^{-1}\left(f_{i_2}f_{i_2}'-V_{ff}\right)V_{ff}^{-1}f_{i_1}\right]+\\
&\quad \mathrm{E}\left[\tfrac{1}{N^4}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}'V_{ff}^{-1}f_{i_1}f_{i_1}'V_{ff}^{-1}f_{i_2}f_{i_2}'V_{ff}^{-1}f_{i_2}\right]+\\
&\quad \mathrm{E}\left[\tfrac{1}{N^4}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}'V_{ff}^{-1}f_{i_1}f_{i_2}'V_{ff}^{-1}f_{i_2}f_{i_2}'V_{ff}^{-1}f_{i_1}\right]+\\
&\quad \mathrm{E}\left[\tfrac{1}{N^4}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}'V_{ff}^{-1}f_{i_2}f_{i_1}'V_{ff}^{-1}f_{i_2}f_{i_2}'V_{ff}^{-1}f_{i_1}\right]+\\
&\quad \mathrm{E}\left[\tfrac{1}{N^4}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_2}'V_{ff}^{-1}f_{i_1}f_{i_1}'V_{ff}^{-1}f_{i_2}f_{i_2}'V_{ff}^{-1}f_{i_1}\right]\Big\}+\\
&\quad \mathrm{E}\left[\tfrac{1}{N^5}\textstyle\sum_{i_1} f_{i_1}'V_{ff}^{-1}f_{i_1}f_{i_1}'V_{ff}^{-1}f_{i_1}f_{i_1}'V_{ff}^{-1}f_{i_1}\right]+\mathrm{E}\left[\tfrac{1}{N^5}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}'V_{ff}^{-1}f_{i_1}f_{i_2}'V_{ff}^{-1}f_{i_2}f_{i_2}'V_{ff}^{-1}f_{i_2}\right]+\\
&\quad \mathrm{E}\left[\tfrac{1}{N^5}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}'V_{ff}^{-1}f_{i_2}f_{i_1}'V_{ff}^{-1}f_{i_2}f_{i_2}'V_{ff}^{-1}f_{i_2}\right]+\mathrm{E}\left[\tfrac{1}{N^5}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}'V_{ff}^{-1}f_{i_2}f_{i_2}'V_{ff}^{-1}f_{i_1}f_{i_2}'V_{ff}^{-1}f_{i_2}\right]+\\
&\quad \mathrm{E}\left[\tfrac{1}{N^5}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}'V_{ff}^{-1}f_{i_2}f_{i_2}'V_{ff}^{-1}f_{i_2}f_{i_1}'V_{ff}^{-1}f_{i_2}\right]+\mathrm{E}\left[\tfrac{1}{N^5}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}'V_{ff}^{-1}f_{i_2}f_{i_2}'V_{ff}^{-1}f_{i_2}f_{i_2}'V_{ff}^{-1}f_{i_1}\right]+\\
&\quad \mathrm{E}\left[\tfrac{1}{N^5}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_2}'V_{ff}^{-1}f_{i_1}f_{i_1}'V_{ff}^{-1}f_{i_2}f_{i_2}'V_{ff}^{-1}f_{i_2}\right]+\mathrm{E}\left[\tfrac{1}{N^5}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_2}'V_{ff}^{-1}f_{i_1}f_{i_2}'V_{ff}^{-1}f_{i_1}f_{i_2}'V_{ff}^{-1}f_{i_2}\right]+\\
&\quad \mathrm{E}\left[\tfrac{1}{N^5}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_2}'V_{ff}^{-1}f_{i_1}f_{i_2}'V_{ff}^{-1}f_{i_2}f_{i_1}'V_{ff}^{-1}f_{i_2}\right]+\mathrm{E}\left[\tfrac{1}{N^5}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_2}'V_{ff}^{-1}f_{i_1}f_{i_2}'V_{ff}^{-1}f_{i_2}f_{i_2}'V_{ff}^{-1}f_{i_1}\right]+\\
&\quad \mathrm{E}\left[\tfrac{1}{N^5}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_2}'V_{ff}^{-1}f_{i_2}f_{i_1}'V_{ff}^{-1}f_{i_1}f_{i_2}'V_{ff}^{-1}f_{i_2}\right]+\mathrm{E}\left[\tfrac{1}{N^5}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_2}'V_{ff}^{-1}f_{i_2}f_{i_1}'V_{ff}^{-1}f_{i_2}f_{i_1}'V_{ff}^{-1}f_{i_2}\right]+\\
&\quad \mathrm{E}\left[\tfrac{1}{N^5}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_2}'V_{ff}^{-1}f_{i_2}f_{i_1}'V_{ff}^{-1}f_{i_2}f_{i_2}'V_{ff}^{-1}f_{i_1}\right]+\mathrm{E}\left[\tfrac{1}{N^5}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_2}'V_{ff}^{-1}f_{i_2}f_{i_2}'V_{ff}^{-1}f_{i_1}f_{i_1}'V_{ff}^{-1}f_{i_2}\right]+\\
&\quad \mathrm{E}\left[\tfrac{1}{N^5}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_2}'V_{ff}^{-1}f_{i_2}f_{i_2}'V_{ff}^{-1}f_{i_1}f_{i_2}'V_{ff}^{-1}f_{i_1}\right]+\mathrm{E}\left[\tfrac{1}{N^5}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_2}'V_{ff}^{-1}f_{i_2}f_{i_2}'V_{ff}^{-1}f_{i_2}f_{i_1}'V_{ff}^{-1}f_{i_1}\right]+\\
&\quad \mathrm{E}\left[\tfrac{1}{N^5}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}'V_{ff}^{-1}f_{i_1}f_{i_1}'V_{ff}^{-1}f_{i_2}f_{i_2}'V_{ff}^{-1}f_{i_2}\right]+\mathrm{E}\left[\tfrac{1}{N^5}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}'V_{ff}^{-1}f_{i_1}f_{i_2}'V_{ff}^{-1}f_{i_1}f_{i_2}'V_{ff}^{-1}f_{i_2}\right]+\\
&\quad \mathrm{E}\left[\tfrac{1}{N^5}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}'V_{ff}^{-1}f_{i_1}f_{i_2}'V_{ff}^{-1}f_{i_2}f_{i_1}'V_{ff}^{-1}f_{i_2}\right]+\mathrm{E}\left[\tfrac{1}{N^5}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}'V_{ff}^{-1}f_{i_1}f_{i_2}'V_{ff}^{-1}f_{i_2}f_{i_2}'V_{ff}^{-1}f_{i_1}\right]+\\
&\quad \mathrm{E}\left[\tfrac{1}{N^5}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}'V_{ff}^{-1}f_{i_2}f_{i_1}'V_{ff}^{-1}f_{i_1}f_{i_2}'V_{ff}^{-1}f_{i_2}\right]+\mathrm{E}\left[\tfrac{1}{N^5}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}'V_{ff}^{-1}f_{i_2}f_{i_1}'V_{ff}^{-1}f_{i_2}f_{i_1}'V_{ff}^{-1}f_{i_2}\right]+
\end{aligned}
$$

$$\mathrm{E}\left[\tfrac{1}{N^5}\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}'V_{ff}^{-1}f_{i_2}f_{i_1}'V_{ff}^{-1}f_{i_2}f_{i_2}'V_{ff}^{-1}f_{i_1}\right]+\mathrm{E}\left[\tfrac{1}{N^5}\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}'V_{ff}^{-1}f_{i_2}f_{i_2}'V_{ff}^{-1}f_{i_1}f_{i_1}'V_{ff}^{-1}f_{i_2}\right]+$$
$$\mathrm{E}\left[\tfrac{1}{N^5}\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}'V_{ff}^{-1}f_{i_2}f_{i_2}'V_{ff}^{-1}f_{i_1}f_{i_2}'V_{ff}^{-1}f_{i_1}\right]+\mathrm{E}\left[\tfrac{1}{N^5}\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}'V_{ff}^{-1}f_{i_2}f_{i_2}'V_{ff}^{-1}f_{i_2}f_{i_1}'V_{ff}^{-1}f_{i_1}\right]$$

The above expression consists of sixth order moments, combinations of second and fourth order moments and combinations of third and third order moments and uses that f_{i_1} and f_{i_2} are independent for $i_1 \neq i_2$ and that the first order moments of f_i are zero so no fifth order moments appear. Assumption 1c implies that the expression is finite. The maximal order of the above expression is $\frac{1}{N^2}$ which explains the order of the remainder term in the higher order specification of the S-statistic. This shows that "the double error" that results from the covariance matrix estimator has a higher convergence rate, $\frac{1}{N^2}$, than the "single error", $\frac{1}{N}$. The same holds true for the other statistics such that, since we are only interested in the low order errors of the remainder term, we leave out the double errors of the covariance matrix estimators.

b. KLM-statistic: Using the higher order components of $\left[\hat{D}_N(\theta,Y)'\hat{V}_{ff}(\theta)^{-1}\hat{D}_N(\theta,Y)\right]^{-1}$ and $s_N(\theta,Y)$ from Lemmas 3 and 4, the KLM statistic has the following higher order components:

$$\mathrm{KLM}(\theta) = \quad KLM_0 + \tfrac{1}{N}KLM_1 + \tfrac{1}{N\sqrt{N}}KLM_2 + O_p(\tfrac{1}{N^2}),$$

with

$$\begin{aligned}
KLM_0 = &\quad N f_N(\theta,Y)'V_{ff}(\theta)^{-\frac{1}{2}}P_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}f_N(\theta,Y)\\
KLM_1 = &\quad -N^2 f_N(\theta,Y)'V_{ff}(\theta)^{-\frac{1}{2}}P_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]\\
&\quad V_{ff}(\theta)^{-\frac{1}{2}}P_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}f_N(\theta,Y)-\\
&\quad 2N^2 f_N(\theta,Y)'V_{ff}(\theta)^{-\frac{1}{2}}P_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]\\
&\quad V_{ff}(\theta)^{-\frac{1}{2}}M_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}f_N(\theta,Y)\\
KLM_2 = &\quad -2N^2\sqrt{N}f_N(\theta,Y)'V_{ff}(\theta)^{-\frac{1}{2}}M_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}\hat{V}_{\theta f}(\theta)V_{ff}(\theta)^{-1}f_N(\theta,Y)\\
&\quad \left[D_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y)\right]^{-1}D_N(\theta,Y)'V_{ff}(\theta)^{-1}f_N(\theta,Y)
\end{aligned}$$

We determine the order of the four different components using their conditional expectations given $D_N(\theta,Y)$. To construct these conditional expectations, we use Lemma 7 and first construct the conditional expectations of the different components of $\mathrm{KLM}(\theta_0)$ given $d_i(\theta)$, $i=1,\ldots,N:$

1. $\mathrm{E}\left[N f_N(\theta,Y)'V_{ff}(\theta)^{-\frac{1}{2}}P_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}f_N(\theta,Y)|d_i\right]$
$$= \quad \mathrm{tr}\left\{V_{ff}^{-\frac{1}{2}}P_{V_{ff}^{-\frac{1}{2}}D_N}V_{ff}^{-\frac{1}{2}}\left(\mathrm{E}\left[\tfrac{1}{N}\textstyle\sum_{i_1}f_{i_1}f_{i_1}'|d_i\right]+\mathrm{E}\left[\tfrac{1}{N}\textstyle\sum_{i_2}\sum_{i_1\neq i_2}f_{i_2}f_{i_1}'|d_i\right]\right)\right\}=1.$$

2. $(-)\ \mathrm{E}\left[N^2 f_N(\theta,Y)'V_{ff}(\theta)^{-\frac{1}{2}}P_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]\right.$
$\left.V_{ff}(\theta)^{-\frac{1}{2}}P_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}f_N(\theta,Y)|d_i\right]$
$$\begin{aligned}
= &\quad \mathrm{E}\left[\tfrac{1}{N}\textstyle\sum_{i_1}f_{i_1}'V_{ff}^{-\frac{1}{2}}P_{V_{ff}^{-\frac{1}{2}}D_N}V_{ff}^{-\frac{1}{2}}\left(\sum_{i_2}\bar{f}_{i_2}\bar{f}_{i_2}'-V_{ff}\right)V_{ff}^{-\frac{1}{2}}P_{V_{ff}^{-\frac{1}{2}}D_N}V_{ff}^{-\frac{1}{2}}\sum_{i_3}f_{i_3}|d_i\right]\\
= &\quad \mathrm{E}\left[\tfrac{1}{N}\textstyle\sum_{i_1}f_{i_1}'V_{ff}^{-\frac{1}{2}}P_{V_{ff}^{-\frac{1}{2}}D_N}V_{ff}^{-\frac{1}{2}}\left(\sum_{i_2}f_{i_2}f_{i_2}'-V_{ff}\right)V_{ff}^{-\frac{1}{2}}P_{V_{ff}^{-\frac{1}{2}}D_N}V_{ff}^{-\frac{1}{2}}\sum_{i_3}f_{i_3}|d_i\right]-\\
&\quad \mathrm{E}\left[\tfrac{1}{N^2}\textstyle\sum_{i_1}f_{i_1}'V_{ff}^{-\frac{1}{2}}P_{V_{ff}^{-\frac{1}{2}}D_N}V_{ff}^{-\frac{1}{2}}\sum_{i_2}f_{i_2}\sum_{i_4}f_{i_4}'V_{ff}^{-\frac{1}{2}}P_{V_{ff}^{-\frac{1}{2}}D_N}V_{ff}^{-\frac{1}{2}}\sum_{i_3}f_{i_3}|d_i\right]
\end{aligned}$$

$$
\begin{aligned}
&= \mathrm{E}\left[\tfrac{1}{N}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}' V_{ff}^{-\frac{1}{2}} P_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} \left(f_{i_2} f_{i_2}' - V_{ff}\right) V_{ff}^{-\frac{1}{2}} P_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} f_{i_1} | d_i\right] + \\
&\quad \mathrm{E}\left[\tfrac{1}{N}\textstyle\sum_{i_1} f_{i_1}' V_{ff}^{-\frac{1}{2}} P_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} \left(f_{i_1} f_{i_1}' - V_{ff}\right) V_{ff}^{-\frac{1}{2}} P_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} f_{i_1} | d_i\right] - \\
&\quad \mathrm{E}\left[\tfrac{1}{N^2}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}' V_{ff}^{-\frac{1}{2}} P_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} f_{i_1} f_{i_2}' V_{ff}^{-\frac{1}{2}} P_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} f_{i_2} | d_i\right] - \\
&\quad \mathrm{E}\left[\tfrac{1}{N^2}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}' V_{ff}^{-\frac{1}{2}} P_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} f_{i_2} f_{i_1}' V_{ff}^{-\frac{1}{2}} P_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} f_{i_2} | d_i\right] - \\
&\quad \mathrm{E}\left[\tfrac{1}{N^2}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}' V_{ff}^{-\frac{1}{2}} P_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} f_{i_2} f_{i_2}' V_{ff}^{-\frac{1}{2}} P_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} f_{i_1} | d_i\right] - \\
&\quad \mathrm{E}\left[\tfrac{1}{N^2}\textstyle\sum_{i_1} f_{i_1}' V_{ff}^{-\frac{1}{2}} P_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} f_{i_1} f_{i_1}' V_{ff}^{-\frac{1}{2}} P_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} f_{i_1} | d_i\right] \\
&= \tfrac{N-1}{N}\left\{\mathrm{E}\left[\left(f_i' V_{ff}^{-\frac{1}{2}} P_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} f_i\right)^2 | d_i\right] - 4\right\} - \tfrac{1}{N}.
\end{aligned}
$$

3. $(-2\times)\ \mathrm{E}\Big[N^2 f_N(\theta,Y)' V_{ff}(\theta)^{-\frac{1}{2}} P_{V_{ff}(\theta)^{-\frac{1}{2}} D_N(\theta,Y)} V_{ff}(\theta)^{-\frac{1}{2}} \left[\hat{V}_{ff}(\theta) - V_{ff}(\theta)\right]$
$V_{ff}(\theta)^{-\frac{1}{2}} M_{V_{ff}(\theta)^{-\frac{1}{2}} D_N(\theta,Y)} V_{ff}(\theta)^{-\frac{1}{2}} f_N(\theta,Y) | d_i\Big]$

$$
\begin{aligned}
&= \mathrm{E}\left[\tfrac{1}{N}\textstyle\sum_{i_1} f_{i_1}' V_{ff}^{-\frac{1}{2}} P_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} \left(\sum_{i_2} \bar{f}_{i_2} \bar{f}_{i_2}' - V_{ff}\right) V_{ff}^{-\frac{1}{2}} M_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} \sum_{i_3} f_{i_3} | d_i\right] \\
&= \mathrm{E}\left[\tfrac{1}{N}\textstyle\sum_{i_1} f_{i_1}' V_{ff}^{-\frac{1}{2}} P_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} \left(\sum_{i_2} f_{i_2} f_{i_2}' - V_{ff}\right) V_{ff}^{-\frac{1}{2}} M_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} \sum_{i_3} f_{i_3} | d_i\right] - \\
&\quad \mathrm{E}\left[\tfrac{1}{N^2}\textstyle\sum_{i_1} f_{i_1}' V_{ff}^{-\frac{1}{2}} P_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} \sum_{i_2} f_{i_2} \sum_{i_2} f_{i_4}' V_{ff}^{-\frac{1}{2}} M_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} \sum_{i_3} f_{i_3} | d_i\right] \\
&= \mathrm{E}\left[\tfrac{1}{N}\textstyle\sum_{i_1} f_{i_1}' V_{ff}^{-\frac{1}{2}} P_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} \left(\sum_{i_2\neq i_1} f_{i_2} f_{i_2}' - V_{ff}\right) V_{ff}^{-\frac{1}{2}} M_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} f_{i_1} | d_i\right] + \\
&\quad \mathrm{E}\left[\tfrac{1}{N}\textstyle\sum_{i_1} f_{i_1}' V_{ff}^{-\frac{1}{2}} P_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} \left(f_{i_1} f_{i_1}' - V_{ff}\right) V_{ff}^{-\frac{1}{2}} M_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} f_{i_1} | d_i\right] - \\
&\quad \mathrm{E}\left[\tfrac{1}{N^2}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}' V_{ff}^{-\frac{1}{2}} P_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} f_{i_1} f_{i_2}' V_{ff}^{-\frac{1}{2}} M_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} f_{i_2} | d_i\right] - \\
&\quad \mathrm{E}\left[\tfrac{1}{N^2}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}' V_{ff}^{-\frac{1}{2}} P_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} f_{i_2} f_{i_1}' V_{ff}^{-\frac{1}{2}} M_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} f_{i_2} | d_i\right] - \\
&\quad \mathrm{E}\left[\tfrac{1}{N^2}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}' V_{ff}^{-\frac{1}{2}} P_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} f_{i_2} f_{i_2}' V_{ff}^{-\frac{1}{2}} M_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} f_{i_1} | d_i\right] - \\
&\quad \mathrm{E}\left[\tfrac{1}{N^2}\textstyle\sum_{i_1} f_{i_1}' V_{ff}^{-\frac{1}{2}} P_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} f_{i_1} f_{i_1}' V_{ff}^{-\frac{1}{2}} M_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} f_{i_1} | d_i\right] \\
&= \tfrac{N-1}{N}\left\{\mathrm{E}\left[f_i' V_{ff}^{-\frac{1}{2}} P_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} f_i f_i' V_{ff}^{-\frac{1}{2}} M_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} f_i | d_i\right] - (k-1)\right\}.
\end{aligned}
$$

4. $(-2\times)\ \mathrm{E}\Big[N^2 f_N(\theta,Y)' V_{ff}(\theta)^{-\frac{1}{2}} M_{V_{ff}(\theta)^{-\frac{1}{2}} D_N(\theta,Y)} V_{ff}(\theta)^{-\frac{1}{2}} \hat{V}_{\theta f}(\theta) V_{ff}(\theta)^{-1}$
$f_N(\theta,Y) \left[D_N(\theta,Y)' V_{ff}(\theta)^{-1} D_N(\theta,Y)\right]^{-1} D_N(\theta,Y)' V_{ff}(\theta)^{-1} f_N(\theta,Y) | d_i\Big]$

$$
\begin{aligned}
&= \mathrm{E}\Big[\tfrac{1}{N}\textstyle\sum_{i_1} f_{i_1}' V_{ff}^{-\frac{1}{2}} M_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} \sum_{i_2} \bar{d}_{i_2} \bar{f}_{i_2}' V_{ff}^{-1} \sum_{i_3} f_{i_3} \left[D_N' V_{ff}^{-1} D_N\right]^{-1} \\
&\quad D_N' V_{ff}^{-1} \textstyle\sum_{i_4} f_{i_4} | d_i\Big] \\
&= \mathrm{E}\Big[\tfrac{1}{N}\textstyle\sum_{i_1} f_{i_1}' V_{ff}^{-\frac{1}{2}} M_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} \sum_{i_2} \bar{d}_{i_2} \bar{f}_{i_2}' V_{ff}^{-1} \sum_{i_3} f_{i_3} \left[D_N' V_{ff}^{-1} D_N\right]^{-1} \\
&\quad D_N' V_{ff}^{-1} \textstyle\sum_{i_4} f_{i_4} | d_i\Big] - \mathrm{E}\Big[\tfrac{1}{N^2}\sum_{i_1} f_{i_1}' V_{ff}^{-\frac{1}{2}} M_{V_{ff}^{-\frac{1}{2}} D_N} V_{ff}^{-\frac{1}{2}} \sum_{i_2} d_{i_2} \\
&\quad \textstyle\sum_{i_5} f_{i_5}' V_{ff}^{-1} \sum_{i_3} f_{i_3} \left[D_N' V_{ff}^{-1} D_N\right]^{-1} D_N' V_{ff}^{-1} \sum_{i_4} f_{i_4} | d_i\Big]
\end{aligned}
$$

$$
\begin{aligned}
= \; & \mathrm{E}\left[\tfrac{1}{N}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}'V_{ff}^{-\frac{1}{2}}M_{V_{ff}^{-\frac{1}{2}}D_N}V_{ff}^{-\frac{1}{2}}d_{i_1}f_{i_1}'V_{ff}^{-1}f_{i_2}\left[D_N'V_{ff}^{-1}D_N\right]^{-1}\right.\\
& \left.\left[D_N'V_{ff}^{-1}D_N\right]^{-1}D_N'V_{ff}^{-1}f_{i_2}|d_i\right] + \mathrm{E}\left[\tfrac{1}{N}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}'V_{ff}^{-\frac{1}{2}}M_{V_{ff}^{-\frac{1}{2}}D_N}V_{ff}^{-\frac{1}{2}}\right.\\
& \left. d_{i_2}f_{i_2}'V_{ff}^{-1}f_{i_2}\left[D_N'V_{ff}^{-1}D_N\right]^{-1}\left[D_N'V_{ff}^{-1}D_N\right]^{-1}D_N'V_{ff}^{-1}f_{i_1}|d_i\right]+\\
& \mathrm{E}\left[\tfrac{1}{N}\textstyle\sum_{i_1}\sum_{i_2\neq i_1} f_{i_1}'V_{ff}^{-\frac{1}{2}}M_{V_{ff}^{-\frac{1}{2}}D_N}V_{ff}^{-\frac{1}{2}}d_{i_2}f_{i_2}'V_{ff}^{-1}f_{i_1}\left[D_N'V_{ff}^{-1}D_N\right]^{-1}\right.\\
& \left.\left[D_N'V_{ff}^{-1}D_N\right]^{-1}D_N'V_{ff}^{-1}f_{i_2}|d_i\right] + \mathrm{E}\left[\tfrac{1}{N}\textstyle\sum_{i_1} f_{i_1}'V_{ff}^{-\frac{1}{2}}M_{V_{ff}^{-\frac{1}{2}}D_N}V_{ff}^{-\frac{1}{2}}d_{i_1}f_{i_1}'\right.\\
& \left. V_{ff}^{-1}f_{i_1}\left[D_N'V_{ff}^{-1}D_N\right]^{-1}D_N'V_{ff}^{-1}f_{i_1}|d_i\right]\\
= \; & \mathrm{E}\left[\tfrac{1}{N}\textstyle\sum_{i_1} d_{i_1}'V_{ff}^{-\frac{1}{2}}M_{V_{ff}^{-\frac{1}{2}}D_N}V_{ff}^{-\frac{1}{2}}f_{i_1}f_{i_1}'V_{ff}^{-1}f_{i_1}f_{i_1}'V_{ff}^{-1}D_N\left[D_N'V_{ff}^{-1}D_N\right]^{-1}|d_i\right]\\
= \; & 0,
\end{aligned}
$$

since $M_{V_{ff}^{-\frac{1}{2}}D_N}V_{ff}^{-\frac{1}{2}}\sum_{i_2}d_{i_2}=0$ and where $D_N=D_N(\theta,Y)$.

Given $D_N(\theta_0,Y)$, the above conditional expectations do not depend on $(d_1(\theta_0),\ldots,d_{N-1}(\theta_0))$ such that the conditional expectations of the different higher order elements of $\mathrm{KLM}(\theta_0)$ given $D_N(\theta_0,Y)$ read

$$
\begin{aligned}
\mathrm{E}\left[KLM_0|D_N(\theta_0,Y)\right] = \; & 1\\
\mathrm{E}\left[KLM_1|D_N(\theta_0,Y)\right] = \; & -\tfrac{N-1}{N}\left\{\mathrm{E}\left[\left(f_i'V_{ff}^{-\frac{1}{2}}P_{V_{ff}^{-\frac{1}{2}}D_N}V_{ff}^{-\frac{1}{2}}f_i\right)^2|D_N\right]-2(k+1)+\right.\\
& \left. 2\mathrm{E}\left[f_i'V_{ff}^{-\frac{1}{2}}P_{V_{ff}^{-\frac{1}{2}}D_N}V_{ff}^{-\frac{1}{2}}f_if_i'V_{ff}^{-\frac{1}{2}}M_{V_{ff}^{-\frac{1}{2}}D_N}V_{ff}^{-\frac{1}{2}}f_i|D_N\right]\right\}+\tfrac{1}{N}\\
\mathrm{E}\left[\sqrt{N}KLM_2|D_N(\theta_0,Y)\right] = \; & 0.
\end{aligned}
$$

c. GMM-MLR statistic. To expand the expression of the GMM-MLR statistic, we use a Taylor approximation of $\mathrm{S}(\theta)$ around S_0 and $\mathrm{KLM}(\theta_0)$ around KLM_0 :

$$
\begin{aligned}
\text{GMM-MLR}(\theta_0) = \; & \tfrac{1}{2}\left[S_0-\mathrm{r}(\theta_0)+\sqrt{(S_0+\mathrm{r}(\theta_0))^2-4\left[S_0-KLM_0\right]\mathrm{r}(\theta_0))}\right]+\\
& \tfrac{1}{2N}\left[1+\tfrac{S_0-\mathrm{r}(\theta_0)}{\sqrt{(S_0+\mathrm{r}(\theta_0))^2-4[S(\theta_0)-KLM(\theta_0)]\mathrm{r}(\theta_0))}}\right]S_1+\\
& \tfrac{1}{N}\tfrac{\mathrm{r}(\theta_0)}{\sqrt{(S_0+\mathrm{r}(\theta_0))^2-4[S(\theta_0)-KLM(\theta_0)]\mathrm{r}(\theta_0))}}(KLM_1+\tfrac{1}{\sqrt{N}}KLM_2)+O_p(\tfrac{1}{N^2}).
\end{aligned}
$$

d. LM statistic. We just show that the higher order components of $\mathrm{LM}(\theta)$ depend on $D_N(\theta,Y)$. We therefore only construct a higher order decomposition which we uses just the $(D_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y))^{-1}$-element from the higher order specification of $\left[q_N(\theta,Y)'\hat{V}_{ff}(\theta)^{-1}q_N(\theta,Y)\right]^{-1}$ in Lemma 6.

$$
\begin{aligned}
& \mathrm{LM}(\theta)=Nf_N(\theta,Y)'\hat{V}_{ff}(\theta)^{-\frac{1}{2}}P_{\hat{V}_{ff}(\theta)^{-\frac{1}{2}}q_N(\theta,Y)}\hat{V}_{ff}(\theta)^{-\frac{1}{2}}f_N(\theta,Y)\\
\approx \; & N\left[f_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y)+f_N(\theta,Y)'V_{ff}(\theta)^{-1}V_{qf}(\theta)V_{ff}(\theta)^{-1}f_N(\theta,Y)-\right.\\
& \left. f_N(\theta,Y)'V_{ff}(\theta)^{-1}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]V_{ff}(\theta)^{-1}D_N(\theta,Y)\right]\left(D_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y)\right)^{-1}\\
& \left[f_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y)+f_N(\theta,Y)'V_{ff}(\theta)^{-1}V_{qf}(\theta)V_{ff}(\theta)^{-1}f_N(\theta,Y)-\right.\\
& \left. f_N(\theta,Y)'V_{ff}(\theta)^{-1}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]V_{ff}(\theta)^{-1}D_N(\theta,Y)\right]'
\end{aligned}
$$

$$
\begin{aligned}
= \;& Nf_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y)\left(D_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y)\right)^{-1}D_N(\theta,Y)'V_{ff}(\theta)^{-1}f_N(\theta,Y)+\\
& 2Nf_N(\theta,Y)'V_{ff}(\theta)^{-1}V_{qf}(\theta)V_{ff}(\theta)^{-1}f_N(\theta,Y)\left(D_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y)\right)^{-1}\\
& D_N(\theta,Y)'V_{ff}(\theta)^{-1}f_N(\theta,Y)+Nf_N(\theta,Y)'V_{ff}(\theta)^{-1}V_{qf}(\theta)V_{ff}(\theta)^{-1}f_N(\theta,Y)\\
& \left(D_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y)\right)^{-1}f_N(\theta,Y)'V_{ff}(\theta)^{-1}V_{qf}(\theta)'V_{ff}(\theta)^{-1}f_N(\theta,Y)-\\
& 2Nf_N(\theta,Y)'V_{ff}(\theta)^{-1}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]V_{ff}(\theta)^{-1}D_N(\theta,Y)\left(D_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y)\right)^{-1}\\
& \left[D_N(\theta,Y)'V_{ff}(\theta)^{-1}f_N(\theta,Y)+f_N(\theta,Y)'V_{ff}(\theta)^{-1}V_{qf}(\theta)'V_{ff}(\theta)^{-1}f_N(\theta,Y)\right]+\\
& Nf_N(\theta,Y)'V_{ff}(\theta)^{-1}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]V_{ff}(\theta)^{-1}D_N(\theta,Y)\left(D_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y)\right)^{-1}\\
& D_N(\theta,Y)'V_{ff}(\theta)^{-1}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]V_{ff}(\theta)^{-1}f_N(\theta,Y)
\end{aligned}
$$

We use the analog of Lemma 7 to obtain the conditional expectation of LM(θ) given $D_N(\theta,Y)$. We therefore first construct the conditional expectation of the first five elements of the approximation of LM(θ) given d_i.

1. $\mathrm{E}\left[Nf_N(\theta,Y)'V_{ff}(\theta)^{-\frac{1}{2}}P_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}f_N(\theta,Y)|d_i\right]=1.$

2. $(2\times)$ $\mathrm{E}\Big[Nf_N(\theta,Y)'V_{ff}(\theta)^{-1}V_{qf}(\theta)V_{ff}(\theta)^{-1}f_N(\theta,Y)\left(D_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y)\right)^{-1}$ $D_N(\theta,Y)'V_{ff}(\theta)^{-1}f_N(\theta,Y)|d_i\Big]$

$$
\begin{aligned}
&= \mathrm{E}\left[\tfrac{1}{N}\textstyle\sum_{i_1}f_{i_1}'V_{ff}^{-1}V_{qf}V_{ff}^{-1}\sum_{i_2}f_{i_2}\left[\sum_{i_3}d_{i_3}'V_{ff}^{-1}\sum_{i_4}d_{i_4}\right]^{-1}\sum_{i_5}d_{i_5}'V_{ff}^{-1}\sum_{i_6}f_{i_6}|d_i\right]\\
&= \mathrm{E}\left[\tfrac{1}{N}\textstyle\sum_{i_1}f_{i_1}'V_{ff}^{-1}V_{qf}V_{ff}^{-1}f_{i_1}\left[\sum_{i_3}d_{i_3}'V_{ff}^{-1}\sum_{i_4}d_{i_4}\right]^{-1}\sum_{i_5}d_{i_5}'V_{ff}^{-1}f_{i_1}|d_i\right]\\
&= \mathrm{tr}\left\{\mathrm{E}\left[f_{i_1}f_{i_1}'V_{ff}^{-1}V_{qf}V_{ff}^{-1}f_{i_1}\left[\textstyle\sum_{i_3}d_{i_3}'V_{ff}^{-1}\sum_{i_4}d_{i_4}\right]^{-1}\sum_{i_5}d_{i_5}'V_{ff}^{-1}|d_i\right]\right\},
\end{aligned}
$$

where $V_{qf}=V_{qf}(\theta)$.

3. $\mathrm{E}\Big[Nf_N(\theta,Y)'V_{ff}(\theta)^{-1}V_{qf}(\theta)V_{ff}(\theta)^{-1}f_N(\theta,Y)\left(D_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y)\right)^{-1}$ $f_N(\theta,Y)'V_{ff}(\theta)^{-1}V_{qf}(\theta)'V_{ff}(\theta)^{-1}f_N(\theta,Y)|d_i\Big]$

$$
\begin{aligned}
&= \mathrm{E}\left[\tfrac{1}{N}\textstyle\sum_{i_1}f_{i_1}'V_{ff}^{-1}V_{qf}V_{ff}^{-1}\sum_{i_2}f_{i_2}\left[\sum_{i_3}d_{i_3}'V_{ff}^{-1}\sum_{i_4}d_{i_4}\right]^{-1}\sum_{i_5}f_{i_5}'V_{ff}^{-1}V_{qf}'V_{ff}^{-1}\sum_{i_6}f_{i_6}|d_i\right]\\
&= \mathrm{E}\left[\tfrac{1}{N}\textstyle\sum_{i_1}f_{i_1}'V_{ff}^{-1}V_{qf}V_{ff}^{-1}f_{i_1}\left[\sum_{i_3}d_{i_3}'V_{ff}^{-1}\sum_{i_4}d_{i_4}\right]^{-1}\sum_{i_5\neq i_1}f_{i_5}'V_{ff}^{-1}V_{qf}'V_{ff}^{-1}f_{i_5}|d_i\right]+\\
&\quad \mathrm{E}\left[\tfrac{1}{N}\textstyle\sum_{i_1}f_{i_1}'V_{ff}^{-1}V_{qf}V_{ff}^{-1}\sum_{i_2\neq i_1}f_{i_2}\left[\sum_{i_3}d_{i_3}'V_{ff}^{-1}\sum_{i_4}d_{i_4}\right]^{-1}f_{i_2}'V_{ff}^{-1}V_{qf}'V_{ff}^{-1}f_{i_1}|d_i\right]+\\
&\quad \mathrm{E}\left[\tfrac{1}{N}\textstyle\sum_{i_1}f_{i_1}'V_{ff}^{-1}V_{qf}V_{ff}^{-1}\sum_{i_2\neq i_1}f_{i_2}\left[\sum_{i_3}d_{i_3}'V_{ff}^{-1}\sum_{i_4}d_{i_4}\right]^{-1}f_{i_1}'V_{ff}^{-1}V_{qf}'V_{ff}^{-1}f_{i_2}|d_i\right]+\\
&\quad \mathrm{E}\left[\tfrac{1}{N}\textstyle\sum_{i_1}f_{i_1}'V_{ff}^{-1}V_{qf}V_{ff}^{-1}f_{i_1}\left[\sum_{i_3}d_{i_3}'V_{ff}^{-1}\sum_{i_4}d_{i_4}\right]^{-1}f_{i_1}'V_{ff}^{-1}V_{qf}'V_{ff}^{-1}f_{i_1}|d_i\right]\\
&= (N-1)\left\{\left[\mathrm{tr}(V_{ff}^{-1}V_{qf})\right]^2+\mathrm{tr}(V_{ff}^{-1}V_{qf}V_{ff}^{-1}V_{qf}')+\mathrm{tr}(V_{ff}^{-1}V_{qf}V_{ff}^{-1}V_{qf})\right\}\left[\textstyle\sum_{i_3}d_{i_3}'V_{ff}^{-1}\sum_{i_4}d_{i_4}\right]^{-1}+\\
&\quad \mathrm{E}\left[(f_i'V_{ff}V_{qf}V_{ff}f_i)^2\right]\left[\textstyle\sum_{i_3}d_{i_3}'V_{ff}^{-1}\sum_{i_4}d_{i_4}\right]^{-1}
\end{aligned}
$$

4. $(2\times)$ $\mathrm{E}\Big[Nf_N(\theta,Y)'V_{ff}(\theta)^{-1}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]V_{ff}(\theta)^{-1}D_N(\theta,Y)$ $\left(D_N(\theta,Y)'V_{ff}(\theta)^{-1}D_N(\theta,Y)\right)^{-1}D_N(\theta,Y)'V_{ff}(\theta)^{-1}f_N(\theta,Y)|d_i\Big]$

$$
\begin{aligned}
&= \mathrm{E}\Big[\tfrac{1}{N^2}\textstyle\sum_{i_1}f_{i_1}'V_{ff}^{-1}\sum_{i_2}\left(f_{i_2}f_{i_2}'-V_{ff}\right)V_{ff}^{-1}\sum_{i_3}d_{i_3}\left[\sum_{i_4}d_{i_4}'V_{ff}^{-1}\sum_{i_5}d_{i_5}\right]^{-1}\\
&\quad \textstyle\sum_{i_6}d_{i_6}'V_{ff}^{-1}\sum_{i_7}f_{i_7}|d_i\Big]\\
&= \mathrm{E}\Big[\tfrac{1}{N^2}\textstyle\sum_{i_1}f_{i_1}'V_{ff}^{-1}\sum_{i_2}\left(f_{i_2}f_{i_2}'-V_{ff}\right)V_{ff}^{-1}\sum_{i_3}d_{i_3}\left[\sum_{i_4}d_{i_4}'V_{ff}^{-1}\sum_{i_5}d_{i_5}\right]^{-1}\\
&\quad \textstyle\sum_{i_6}d_{i_6}'V_{ff}^{-1}f_{i_1}|d_i\Big]+\mathrm{E}\Big[\tfrac{1}{N^2}\sum_{i_1}f_{i_1}'V_{ff}^{-1}\left(f_{i_1}f_{i_1}'-V_{ff}\right)V_{ff}^{-1}\sum_{i_3}d_{i_3}\\
&\quad \left[\textstyle\sum_{i_4}d_{i_4}'V_{ff}^{-1}\sum_{i_5}d_{i_5}\right]^{-1}\sum_{i_6}d_{i_6}'V_{ff}^{-1}f_{i_1}|d_i\Big]
\end{aligned}
$$

$$\begin{aligned}
&= \mathrm{E}\left[\tfrac{1}{N^2}\textstyle\sum_{i_1} f_{i_1}' V_{ff}^{-1} f_{i_1} f_{i_1}' V_{ff}^{-1} \sum_{i_3} d_{i_3} \left[\sum_{i_4} d_{i_4}' V_{ff}^{-1} \sum_{i_5} d_{i_5}\right]^{-1} \sum_{i_6} d_{i_6}' V_{ff}^{-1} f_{i_1} | d_i\right] - \\
&\quad \mathrm{E}\left[\tfrac{1}{N^2}\textstyle\sum_{i_1} f_{i_1}' V_{ff}^{-1} \sum_{i_3} d_{i_3} \left[\sum_{i_4} d_{i_4}' V_{ff}^{-1} \sum_{i_5} d_{i_5}\right]^{-1} \sum_{i_6} d_{i_6}' V_{ff}^{-1} f_{i_1} | d_i\right] \\
&= \tfrac{1}{N}\mathrm{tr}\left[\mathrm{E}\left(f_i f_i' V_{ff}^{-1} f_i f_i'\right) V_{ff}^{-1} \textstyle\sum_{i_3} d_{i_3} \left[\sum_{i_4} d_{i_4}' V_{ff}^{-1} \sum_{i_5} d_{i_5}\right]^{-1} \sum_{i_6} d_{i_6}' V_{ff}^{-1}\right] - \\
&\quad \tfrac{1}{N}\mathrm{tr}\left[\mathrm{E}\left(f_i f_i'\right) V_{ff}^{-1} \textstyle\sum_{i_3} d_{i_3} \left[\sum_{i_4} d_{i_4}' V_{ff}^{-1} \sum_{i_5} d_{i_5}\right]^{-1} \sum_{i_6} d_{i_6}' V_{ff}^{-1}\right] \\
&= \tfrac{1}{N}\mathrm{tr}\left[\mathrm{E}\left(f_i f_i' V_{ff}^{-1} f_i f_i'\right) V_{ff}^{-1} \textstyle\sum_{i_3} d_{i_3} \left[\sum_{i_4} d_{i_4}' V_{ff}^{-1} \sum_{i_5} d_{i_5}\right]^{-1} \sum_{i_6} d_{i_6}' V_{ff}^{-1}\right] - \tfrac{1}{N}.
\end{aligned}$$

5. $(2\times)$ $\mathrm{E}\Big[N f_N(\theta,Y)' V_{ff}(\theta)^{-1}\left[\hat{V}_{ff}(\theta) - V_{ff}(\theta)\right] V_{ff}(\theta)^{-1} D_N(\theta,Y)$
$\left(D_N(\theta,Y)' V_{ff}(\theta)^{-1} D_N(\theta,Y)\right)^{-1} f_N(\theta,Y)' V_{ff}(\theta)^{-1} V_{qf}(\theta)' V_{ff}(\theta)^{-1} f_N(\theta,Y) | d_i\Big]$

$$\begin{aligned}
&= \mathrm{E}\Big[\tfrac{1}{N^2}\textstyle\sum_{i_1} f_{i_1}' V_{ff}^{-1} \sum_{i_2}\left(f_{i_2} f_{i_2}' - V_{ff}\right) V_{ff}^{-1} \sum_{i_3} d_{i_3}\left[\sum_{i_4} d_{i_4}' V_{ff}^{-1} \sum_{i_5} d_{i_5}\right]^{-1} \\
&\quad \textstyle\sum_{i_6} f_{i_6}' V_{ff}^{-1} V_{qf} V_{ff}^{-1} \sum_{i_7} f_{i_7} | d_i\Big] \\
&= \mathrm{E}\Big[\tfrac{1}{N^2}\textstyle\sum_{i_1} f_{i_1}' V_{ff}^{-1} \sum_{i_2\neq i_1}\left(f_{i_2} f_{i_2}' - V_{ff}\right) V_{ff}^{-1} \sum_{i_3} d_{i_3}\left[\sum_{i_4} d_{i_4}' V_{ff}^{-1} \sum_{i_5} d_{i_5}\right]^{-1} \\
&\quad f_{i_1}' V_{ff}^{-1} V_{qf} V_{ff}^{-1} f_{i_1} | d_i\Big] + \mathrm{E}\Big[\tfrac{1}{N^2}\textstyle\sum_{i_1} f_{i_1}' V_{ff}^{-1}\left(f_{i_1} f_{i_1}' - V_{ff}\right) V_{ff}^{-1} \\
&\quad \textstyle\sum_{i_3} d_{i_3}\left[\sum_{i_4} d_{i_4}' V_{ff}^{-1} \sum_{i_5} d_{i_5}\right]^{-1} \sum_{i_6\neq i_1} f_{i_6}' V_{ff}^{-1} V_{qf} V_{ff}^{-1} f_{i_6} | d_i\Big] + \\
&\quad \mathrm{E}\Big[\tfrac{1}{N^2}\textstyle\sum_{i_1} f_{i_1}' V_{ff}^{-1} \sum_{i_2\neq i_1}\left(f_{i_2} f_{i_2}' - V_{ff}\right) V_{ff}^{-1} \sum_{i_3} d_{i_3}\left[\sum_{i_4} d_{i_4}' V_{ff}^{-1} \sum_{i_5} d_{i_5}\right]^{-1} \\
&\quad f_{i_2}' V_{ff}^{-1} V_{qf} V_{ff}^{-1} f_{i_1} | d_i\Big] + \mathrm{E}\Big[\tfrac{1}{N^2}\textstyle\sum_{i_1} f_{i_1}' V_{ff}^{-1} \sum_{i_2\neq i_1}\left(f_{i_2} f_{i_2}' - V_{ff}\right) V_{ff}^{-1} \\
&\quad \textstyle\sum_{i_3} d_{i_3}\left[\sum_{i_4} d_{i_4}' V_{ff}^{-1} \sum_{i_5} d_{i_5}\right]^{-1} f_{i_1}' V_{ff}^{-1} V_{qf} V_{ff}^{-1} f_{i_2} | d_i\Big] + \\
&\quad \mathrm{E}\Big[\tfrac{1}{N^2}\textstyle\sum_{i_1} f_{i_1}' V_{ff}^{-1}\left(f_{i_1} f_{i_1}' - V_{ff}\right) V_{ff}^{-1} \sum_{i_3} d_{i_3}\left[\sum_{i_4} d_{i_4}' V_{ff}^{-1} \sum_{i_5} d_{i_5}\right]^{-1} \\
&\quad f_{i_1}' V_{ff}^{-1} V_{qf} V_{ff}^{-1} f_{i_1} | d_i\Big] \\
&= \tfrac{N-1}{N}\Big\{\mathrm{E}\left(f_i' V_{ff}^{-1} f_i f_i'\right) V_{ff}^{-1} \textstyle\sum_{i_3} d_{i_3}\left[\sum_{i_4} d_{i_4}' V_{ff}^{-1} \sum_{i_5} d_{i_5}\right]^{-1} \mathrm{E}(f_{i_6}' V_{ff}^{-1} V_{qf} V_{ff}^{-1} f_{i_6}) + \\
&\quad \mathrm{tr}\left[V_{ff}^{-1} \textstyle\sum_{i_3} d_{i_3}\left[\sum_{i_4} d_{i_4}' V_{ff}^{-1} \sum_{i_5} d_{i_5}\right]^{-1} \mathrm{E}\left(f_{i_2}' V_{ff}^{-1} V_{qf} V_{ff}^{-1} f_{i_2} f_{i_2}'\right)\right] + \\
&\quad \mathrm{tr}\left[V_{ff}^{-1} \textstyle\sum_{i_3} d_{i_3}\left[\sum_{i_4} d_{i_4}' V_{ff}^{-1} \sum_{i_5} d_{i_5}\right]^{-1} \mathrm{E}\left(f_{i_2}' V_{ff}^{-1} V_{qf}' V_{ff}^{-1} f_{i_2} f_{i_2}'\right)\right]\Big\} + \\
&\quad \mathrm{E}\left[\tfrac{1}{N^2}\textstyle\sum_{i_1} f_{i_1}' V_{ff}^{-1} f_{i_1} f_{i_1}' V_{ff}^{-1} \sum_{i_3} d_{i_3}\left[\sum_{i_4} d_{i_4}' V_{ff}^{-1} \sum_{i_5} d_{i_5}\right]^{-1} f_{i_1}' V_{ff}^{-1} V_{qf} V_{ff}^{-1} f_{i_1} | d_i\right] - \\
&\quad \mathrm{E}\left[\tfrac{1}{N^2}\textstyle\sum_{i_1} f_{i_1}' V_{ff}^{-1} \sum_{i_3} d_{i_3}\left[\sum_{i_4} d_{i_4}' V_{ff}^{-1} \sum_{i_5} d_{i_5}\right]^{-1} f_{i_1}' V_{ff}^{-1} V_{qf} V_{ff}^{-1} f_{i_1} | d_i\right] \\
&= \tfrac{N-1}{N}\Big\{\left[\mathrm{tr}(V_{ff}^{-1} V_{qf})\mathrm{E}\left(f_i' V_{ff}^{-1} f_i f_i'\right) + \mathrm{E}\left(f_{i_2}' V_{ff}^{-1} V_{qf}' V_{ff}^{-1} f_{i_2} f_{i_2}'\right) + \mathrm{E}\left(f_{i_2}' V_{ff}^{-1} V_{qf} V_{ff}^{-1} f_{i_2} f_{i_2}'\right)\right] \\
&\quad V_{ff}^{-1} \textstyle\sum_{i_3} d_{i_3}\left[\sum_{i_4} d_{i_4}' V_{ff}^{-1} \sum_{i_5} d_{i_5}\right]^{-1}\Big\} + \\
&\quad \tfrac{1}{N}\mathrm{E}\left[f_i' V_{ff}^{-1} V_{qf} V_{ff}^{-1} f_i f_i' V_{ff}^{-1} f_i f_i' | d_i\right] V_{ff}^{-1} \textstyle\sum_{i_3} d_{i_3}\left[\sum_{i_4} d_{i_4}' V_{ff}^{-1} \sum_{i_5} d_{i_5}\right]^{-1} + \\
&\quad \tfrac{1}{N}\mathrm{E}\left[f_i' V_{ff}^{-1} V_{qf} V_{ff}^{-1} f_i f_i' | d_i\right] V_{ff}^{-1} \textstyle\sum_{i_3} d_{i_3}\left[\sum_{i_4} d_{i_4}' V_{ff}^{-1} \sum_{i_5} d_{i_5}\right]^{-1}
\end{aligned}$$

The five elements of the approximation of $\mathrm{LM}(\theta)$ are such that we can specify it as

$$\begin{aligned}
\mathrm{LM}(\theta) = \;& LM_0 + \tfrac{1}{N} LM_1 + \mathrm{tr}\left(\tfrac{1}{N} LM_{D_1}\left(D_N(\theta,Y)' V_{ff}(\theta)^{-1} D_N(\theta,Y)\right)^{-1} D_N(\theta,Y)' V_{ff}(\theta)^{-1}\right) + \\
& \tfrac{1}{N^2} LM_{D_2}\left(D_N(\theta,Y)' V_{ff}(\theta)^{-1} D_N(\theta,Y)\right)^{-1},
\end{aligned}$$

where

$$\begin{aligned}
\mathrm{E}(LM_0 | d_i) &= 1 \\
\mathrm{E}(LM_1 | d_i) &= 2 - 2\mathrm{tr}\left[\mathrm{E}\left(f_i f_i' V_{ff}^{-1} f_i f_i'\right) V_{ff}^{-1} \textstyle\sum_{i_3} d_{i_3}\left[\sum_{i_4} d_{i_4}' V_{ff}^{-1} \sum_{i_5} d_{i_5}\right]^{-1} \sum_{i_6} d_{i_6}' V_{ff}^{-1}\right]
\end{aligned}$$

$$
\begin{aligned}
\mathrm{E}(LM_{D_1}|d_i) = \; & 2\mathrm{E}\left[f_{i_1}f_{i_1}'V_{ff}^{-1}V_{qf}V_{ff}^{-1}f_{i_1}|d_i\right] - 2\tfrac{N-1}{N}\left\{\mathrm{tr}(V_{ff}^{-1}V_{qf})\mathrm{E}\left(f_i'V_{ff}^{-1}f_if_i'\right) + \right. \\
& \mathrm{E}\left(f_{i_2}'V_{ff}^{-1}V_{qf}'V_{ff}^{-1}f_{i_2}f_{i_2}'\right) + \left.\mathrm{E}\left(f_{i_2}'V_{ff}^{-1}V_{qf}V_{ff}^{-1}f_{i_2}f_{i_2}'\right)\right\} - \\
& \tfrac{2}{N}\mathrm{E}\left[f_i'V_{ff}^{-1}V_{qf}V_{ff}^{-1}f_if_i'V_{ff}^{-1}f_if_i'|d_i\right] - \tfrac{2}{N}\mathrm{E}\left[f_i'V_{ff}^{-1}V_{qf}V_{ff}^{-1}f_if_i'|d_i\right] \\
\mathrm{E}(LM_{D_2}|d_i) = \; & (N-1)\left\{\left[\mathrm{tr}(V_{ff}^{-1}V_{qf})\right]^2 + \mathrm{tr}(V_{ff}^{-1}V_{qf}V_{ff}^{-1}V_{qf}') + \mathrm{tr}(V_{ff}^{-1}V_{qf}V_{ff}^{-1}V_{qf})\right\} + \\
& \mathrm{E}\left[(f_i'V_{ff}V_{qf}V_{ff}f_i)^2\right].
\end{aligned}
$$

All these conditional expectations are identical to the conditional expectations given $D_N(\theta, Y)$ since they only depend on d_i through $D_N(\theta, Y)$.

B. Proof of Theorem 2.

a. S*-statistic: The higher order components of the bootstrapped S-statistic, $\mathrm{S}^*(\theta)$, read:

$$
\begin{aligned}
\mathrm{S}^*(\theta) = \; & Bf_B^*(\theta,Y)'\hat{V}_{ff}(\theta)^{-1}f_B^*(\theta,Y) - \\
& Bf_B^*(\theta,Y)'\hat{V}_{ff}(\theta)^{-1}\left[V_{ff}^*(\theta) - \hat{V}_{ff}(\theta)\right]\hat{V}_{ff}(\theta)^{-1}f_B^*(\theta,Y) + O_p(\tfrac{1}{B^2}).
\end{aligned}
$$

The expectation of the first component with respect to the sampling distribution of f_i^*, E^*, results from

$$
\begin{aligned}
& \mathrm{E}^*\left[Bf_B^*(\theta,Y)'\hat{V}_{ff}(\theta)^{-1}f_B^*(\theta,Y)\right] = \mathrm{E}^*\left[\tfrac{1}{B}\mathrm{tr}\left(\hat{V}_{ff}^{-1}\textstyle\sum_{i_1}f_{i_1}^*\sum_{i_2}f_{i_2}^{*\prime}\right)\right] \\
= \; & \mathrm{E}^*\left[\tfrac{1}{B}\mathrm{tr}\left(\hat{V}_{ff}^{-1}\textstyle\sum_{i_1}f_{i_1}^*f_{i_1}^{*\prime}\right)\right] + \mathrm{E}^*\left[\tfrac{1}{B}\mathrm{tr}\left(\hat{V}_{ff}^{-1}\textstyle\sum_{i_1}\sum_{i_2\neq i_1}f_{i_1}^*f_{i_2}^{*\prime}\right)\right] \\
= \; & \tfrac{1}{B}\mathrm{tr}\left(\hat{V}_{ff}^{-1}\textstyle\sum_{i_1}\mathrm{E}^*(f_{i_1}^*f_{i_1}^{*\prime})\right) + \tfrac{1}{B}\mathrm{tr}\left(\hat{V}_{ff}^{-1}\textstyle\sum_{i_1}\mathrm{E}^*(f_{i_1}^*)\sum_{i_2\neq i_1}\mathrm{E}^*(f_{i_2}^*)'\right) \\
= \; & \tfrac{1}{B}\mathrm{tr}\left(\hat{V}_{ff}^{-1}\textstyle\sum_{i_1=1}^{B}\tfrac{1}{N}\sum_{i=1}^{N}\bar{f}_i\bar{f}_i'\right) + \tfrac{1}{B}\mathrm{tr}\left(\hat{V}_{ff}^{-1}\textstyle\sum_{i_1=1}^{B}\tfrac{1}{N}\sum_{i=1}^{N}\bar{f}_i\left[\sum_{i_2\neq i_1}\mathrm{E}^*(f_{i_2}^*)'\right]\right) = k,
\end{aligned}
$$

since $\mathrm{E}^*(f_{i_1}^*) = \frac{1}{N}\sum_i^N \bar{f}_i = 0$ and $\mathrm{E}^*(f_{i_1}^*f_{i_1}^{*\prime}) = \frac{1}{N}\sum_{j1=1}^N \bar{f}_i\bar{f}_i' = \hat{V}_{ff}$. The expectation of $Bf_B^*(\theta,Y)'\hat{V}_{ff}(\theta)^{-1}\left[V_{ff}^*(\theta) - \hat{V}_{ff}(\theta)\right]\hat{V}_{ff}(\theta)^{-1}f_B^*(\theta,Y)$ results from

$$
\begin{aligned}
& \mathrm{E}^*\left[Bf_B^*(\theta,Y)'\hat{V}_{ff}(\theta)^{-1}\left[V_{ff}^*(\theta) - \hat{V}_{ff}(\theta)\right]\hat{V}_{ff}(\theta)^{-1}f_B^*(\theta,Y)\right] \\
= \; & \mathrm{E}^*\left[\tfrac{1}{B^2}\textstyle\sum_{i_1}f_{i_1}^{*\prime}\hat{V}_{ff}^{-1}\sum_{i_2}\left(f_{i_2}^*f_{i_2}^{*\prime} - \hat{V}_{ff}\right)\hat{V}_{ff}^{-1}\sum_{i_3}f_{i_3}^*\right] - \\
& \mathrm{E}^*\left[\tfrac{1}{B^3}\textstyle\sum_{i_1}f_{i_1}^{*\prime}V_{ff}^{-1}\sum_{i_2}f_{i_2}^*\sum_{i_4}f_{i_4}^{*\prime}V_{ff}^{-1}\sum_{i_3}f_{i_3}^*\right] \\
= \; & \mathrm{E}^*\left[\tfrac{1}{B^2}\textstyle\sum_{i_1}f_{i_1}^{*\prime}\hat{V}_{ff}^{-1}\sum_{i_2\neq i_1}\left(f_{i_2}^*f_{i_2}^{*\prime} - \hat{V}_{ff}\right)\hat{V}_{ff}^{-1}f_{i_1}^*\right] + \mathrm{E}^*\left[\tfrac{1}{B^2}\textstyle\sum_{i_1}f_{i_1}^{*\prime}\hat{V}_{ff}^{-1}\left(f_{i_1}^*f_{i_1}^{*\prime} - \hat{V}_{ff}\right)\hat{V}_{ff}^{-1}f_{i_1}^*\right] - \\
& \mathrm{E}^*\left[\tfrac{1}{B^3}\textstyle\sum_{i_1}\sum_{i_2\neq i_1}f_{i_1}^{*\prime}\hat{V}_{ff}^{-1}f_{i_1}^*f_{i_2}^{*\prime}\hat{V}_{ff}^{-1}f_{i_2}^*\right] - \mathrm{E}^*\left[\tfrac{1}{B^3}\textstyle\sum_{i_1}\sum_{i_2\neq i_1}f_{i_1}^{*\prime}\hat{V}_{ff}^{-1}f_{i_2}^*f_{i_1}^{*\prime}\hat{V}_{ff}^{-1}f_{i_2}^*\right] - \\
& \mathrm{E}^*\left[\tfrac{1}{B^3}\textstyle\sum_{i_1}\sum_{i_2\neq i_1}f_{i_1}^{*\prime}\hat{V}_{ff}^{-1}f_{i_2}^*f_{i_2}^{*\prime}\hat{V}_{ff}^{-1}f_{i_1}^*\right] - \mathrm{E}^*\left[\tfrac{1}{B^3}\textstyle\sum_{i_1}f_{i_1}^{*\prime}\hat{V}_{ff}^{-1}f_{i_1}^*f_{i_1}^{*\prime}\hat{V}_{ff}^{-1}f_{i_1}^*\right] \\
= \; & \mathrm{E}^*\left[\tfrac{1}{B^2}\textstyle\sum_{i_1}\left(f_{i_1}^{*\prime}\hat{V}_{ff}^{-1}f_{i_1}^*\right)^2\right] - \mathrm{E}^*\left[\tfrac{1}{B^2}\textstyle\sum_{i_1}f_{i_1}^{*\prime}\hat{V}_{ff}^{-1}f_{i_1}^*\right] - \tfrac{B-1}{B^2}\left[k^2+2k\right] - \\
& \mathrm{E}^*\left[\tfrac{1}{B^2}\left(f_i^{*\prime}\hat{V}_{ff}^{-1}f_i^*\right)^2\right] \\
= \; & \tfrac{B-1}{B^2}\mathrm{E}^*\left[\left(f_i^{*\prime}\hat{V}_{ff}^{-1}f_i^*\right)^2\right] - \tfrac{k}{B} - \tfrac{B-1}{B^2}\left[k^2+2k\right] \\
= \; & \tfrac{B-1}{B^2}\tfrac{1}{N}\textstyle\sum_{i=1}^N\left(\bar{f}_i'\hat{V}_{ff}^{-1}\bar{f}_i\right)^2 - \tfrac{k}{B} - \tfrac{B-1}{B^2}\left[k^2+2k\right].
\end{aligned}
$$

Hence,

$$
\mathrm{S}^*(\theta) = S_0^* + \tfrac{1}{B}S_1^* + O_p(\tfrac{1}{B^2}),
$$

where $S_0^* = Bf_B^*(\theta,Y)'\hat{V}_{ff}(\theta)^{-1}f_B^*(\theta,Y)$, $\mathrm{E}(S_0^*) = k$, and $S_1^* = -Bf_B^*(\theta,Y)'\hat{V}_{ff}(\theta)^{-1}\left[V_{ff}^*(\theta) - \hat{V}_{ff}(\theta)\right]$ $\hat{V}_{ff}(\theta)^{-1}f_B^*(\theta,Y)$, $\mathrm{E}^*(S_1) = -\frac{B-1}{B}\frac{1}{N}\sum_{i=1}^N \left(\bar{f}_i'\hat{V}_{ff}^{-1}\bar{f}_i\right)^2 + k + \frac{B-1}{B}\left[k^2+2k\right]$.

b. KLM*-statistic: The bootstrapped KLM statistic, KLM*(θ), has the following higher order components:

$$\mathrm{KLM}^*(\theta) = KLM_0^* + \tfrac{1}{B}KLM_1^* + O_p(\tfrac{1}{B^2}),$$

where

$$\begin{aligned}
KLM_0^* = {} & Bf_B^*(\theta,Y)'\hat{V}_{ff}(\theta)^{-\frac{1}{2}}P_{\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{D}_N(\theta,Y)}\hat{V}_{ff}(\theta)^{-\frac{1}{2}}f_B^*(\theta,Y)\\
KLM_1^* = {} & -Bf_B^*(\theta,Y)'\hat{V}_{ff}(\theta)^{-\frac{1}{2}}P_{\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{D}_N(\theta,Y)}\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\left[V_{ff}^*(\theta) - \hat{V}_{ff}(\theta)\right]\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\\
& P_{\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{D}_N(\theta,Y)}\hat{V}_{ff}(\theta)^{-\frac{1}{2}}f_B^*(\theta,Y) - 2Bf_B^*(\theta,Y)'\hat{V}_{ff}(\theta)^{-\frac{1}{2}}P_{\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{D}_N(\theta,Y)}\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\\
& \left[V_{ff}^*(\theta) - \hat{V}_{ff}(\theta)\right]\hat{V}_{ff}(\theta)^{-\frac{1}{2}}M_{\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{D}_N(\theta,Y)}\hat{V}_{ff}(\theta)^{-\frac{1}{2}}f_B^*(\theta,Y).
\end{aligned}$$

We determine the order of the three different components using their expected values with respect to the resampling distribution.

$$\begin{aligned}
& 1.\ \mathrm{E}^*\left[Bf_B^*(\theta,Y)'\hat{V}_{ff}(\theta)^{-\frac{1}{2}}P_{\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{D}_N(\theta,Y)}\hat{V}_{ff}(\theta)^{-\frac{1}{2}}f_B^*(\theta,Y)\right]\\
= {} & \mathrm{E}^*\left[\tfrac{1}{B}\textstyle\sum_{i_1} f_{i_1}^{*\prime}\hat{V}_{ff}^{-\frac{1}{2}}P_{\hat{V}_{ff}^{-\frac{1}{2}}\hat{D}_N}\hat{V}_{ff}^{-\frac{1}{2}}\sum_{i_2} f_{i_2}^*\right]\\
= {} & \mathrm{tr}\left\{\hat{V}_{ff}^{-\frac{1}{2}}P_{\hat{V}_{ff}^{-\frac{1}{2}}\hat{D}_N}\hat{V}_{ff}^{-\frac{1}{2}}\tfrac{1}{B}\textstyle\sum_{i_1}\mathrm{E}^*\left(f_{i_1}^*f_{i_1}^*\right)'\right\} +\\
& \mathrm{tr}\left\{\hat{V}_{ff}^{-\frac{1}{2}}P_{\hat{V}_{ff}^{-\frac{1}{2}}\hat{D}_N}\hat{V}_{ff}^{-\frac{1}{2}}\tfrac{1}{B}\textstyle\sum_{i_6}\mathrm{E}^*\left(f_{i_6}^*\right)\sum_{i_1\neq i_6}\mathrm{E}^*\left(f_{i_1}^*\right)'\right\} = 1.
\end{aligned}$$

$$\begin{aligned}
& 2.\ \mathrm{E}^*\left[Bf_B^*(\theta,Y)'\hat{V}_{ff}(\theta)^{-\frac{1}{2}}P_{\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{D}_N(\theta,Y)}\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\left[V_{ff}^*(\theta) - \hat{V}_{ff}(\theta)\right]\right.\\
& \left.\hat{V}_{ff}(\theta)^{-\frac{1}{2}}P_{\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{D}_N(\theta,Y)}\hat{V}_{ff}(\theta)f_B^*(\theta,Y)\right]\\
= {} & \mathrm{E}^*\left[\tfrac{1}{B^2}\textstyle\sum_{i_1} f_{i_1}^{*\prime}\hat{V}_{ff}^{-1}A\hat{V}_{ff}^{-1}\left[\sum_{i_6}\left(f_{i_6}^*f_{i_6}^{*\prime} - \hat{V}_{ff}\right)\right]\hat{V}_{ff}^{-1}A\hat{V}_{ff}^{-1}\sum_{i_{11}} f_{i_{11}}^*\right] -\\
& \mathrm{E}^*\left[\tfrac{1}{B^3}\textstyle\sum_{i_1} f_{i_1}^{*\prime}\hat{V}_{ff}^{-1}A\hat{V}_{ff}^{-1}\left[\sum_{i_{12}} f_{i_{12}}^*\sum_{i_{13}} f_{i_{13}}^{*\prime}\right]\hat{V}_{ff}^{-1}A\hat{V}_{ff}^{-1}\sum_{i_{11}} f_{i_{11}}^*\right]\\
= {} & \mathrm{E}^*\left[\tfrac{1}{B^2}\mathrm{tr}\left\{\textstyle\sum_{i_1}\sum_{i_6\neq i_1}\hat{V}_{ff}^{-1}A\hat{V}_{ff}^{-1}\left[\left(f_{i_6}^*f_{i_6}^{*\prime} - \hat{V}_{ff}\right)\right]\hat{V}_{ff}^{-1}A\hat{V}_{ff}^{-1}f_{i_1}^*f_{i_1}^{*\prime}\right\}\right] +\\
& \mathrm{E}^*\left[\tfrac{1}{B^2}\mathrm{tr}\left\{\textstyle\sum_{i_1}\hat{V}_{ff}^{-1}A\hat{V}_{ff}^{-1}\left[\left(f_{i_1}^*f_{i_1}^{*\prime} - \hat{V}_{ff}\right)\right]\hat{V}_{ff}^{-1}A\hat{V}_{ff}^{-1}f_{i_1}^*f_{i_1}^{*\prime}\right\}\right] -\\
& \mathrm{E}^*\left[\tfrac{1}{B^3}\textstyle\sum_{i_1}\sum_{i_{12}}\sum_{i_{13}}\sum_{i_{11}} f_{i_1}^{*\prime}\hat{V}_{ff}^{-1}A\hat{V}_{ff}^{-1}f_{i_{12}}^*f_{i_{13}}^{*\prime}\hat{V}_{ff}^{-1}A\hat{V}_{ff}^{-1}f_{i_{11}}^*\right]\\
= {} & \mathrm{E}^*\left[\tfrac{1}{B^2}\left\{\textstyle\sum_{i_1}\left(f_{i_1}^{*\prime}\hat{V}_{ff}^{-1}A\hat{V}_{ff}^{-1}f_{i_1}^*\right)^2\right\}\right] - \mathrm{E}^*\left[\tfrac{1}{B^2}\mathrm{tr}\left\{\textstyle\sum_{i_1}\hat{V}_{ff}^{-1}A\hat{V}_{ff}^{-1}A\hat{V}_{ff}^{-1}f_{i_1}^*f_{i_1}^{*\prime}\right\}\right] -\\
& \mathrm{E}^*\left[\tfrac{1}{B^3}\textstyle\sum_{i_1} f_{i_1}^{*\prime}\hat{V}_{ff}^{-1}A\hat{V}_{ff}^{-1}f_{i_1}^*f_{i_1}^{*\prime}\hat{V}_{ff}^{-1}A\hat{V}_{ff}^{-1}f_{i_1}^*\right] -\\
& \mathrm{E}^*\left[\tfrac{1}{B^3}\textstyle\sum_{i_1}\sum_{i_{12}\neq i_1} f_{i_1}^{*\prime}\hat{V}_{ff}^{-1}A\hat{V}_{ff}^{-1}f_{i_{12}}^*f_{i_1}^{*\prime}\hat{V}_{ff}^{-1}A\hat{V}_{ff}^{-1}f_{i_{12}}^*\right] -\\
& \mathrm{E}^*\left[\tfrac{1}{B^3}\textstyle\sum_{i_1}\sum_{i_{12}\neq i_1} f_{i_1}^{*\prime}\hat{V}_{ff}^{-1}A\hat{V}_{ff}^{-1}f_{i_1}^*f_{i_{12}}^{*\prime}\hat{V}_{ff}^{-1}A\hat{V}_{ff}^{-1}f_{i_{12}}^*\right] -\\
& \mathrm{E}^*\left[\tfrac{1}{B^3}\textstyle\sum_{i_1}\sum_{i_{12}\neq i_1} f_{i_1}^{*\prime}\hat{V}_{ff}^{-1}A\hat{V}_{ff}^{-1}f_{i_{12}}^*f_{i_{12}}^{*\prime}\hat{V}_{ff}^{-1}A\hat{V}_{ff}^{-1}f_{i_1}^*\right]\\
= {} & \tfrac{B-1}{B^2}\left[\tfrac{1}{N}\textstyle\sum_{i=1}^N\left(\bar{f}_i'\hat{V}_{ff}^{-\frac{1}{2}}P_{\hat{V}_{ff}^{-\frac{1}{2}}\hat{D}_N(\theta_0,Y)}\hat{V}_{ff}^{-\frac{1}{2}}\bar{f}_i\right)^2 - 4\right] - \tfrac{1}{B^2}
\end{aligned}$$

with $A=\sum_{i_7}\hat{d}_{i_7}\left[\sum_{i_8}\hat{d}_{i_8}'\hat{V}_{ff}^{-1}\sum_{i_9}\hat{d}_{i_9}\right]^{-1}\sum_{i_{10}}\hat{d}_{i_{10}}'$, $\hat{d}_i=q_i(\theta_0)-\hat{V}_{\theta f}(\theta_0)\hat{V}_{ff}(\theta_0)^{-1}f_i(\theta_0)$.

3. (2×) $\mathrm{E}^*\left[Bf_B^*(\theta,Y)'\hat{V}_{ff}(\theta)^{-\frac{1}{2}}P_{\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{D}_N(\theta,Y)}\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\left[V_{ff}^*(\theta)-\hat{V}_{ff}(\theta)\right]\right.$
$\left.\hat{V}_{ff}(\theta)^{-\frac{1}{2}}M_{\hat{V}_{ff}(\theta)^{-\frac{1}{2}}\hat{D}_N(\theta,Y)}\hat{V}_{ff}(\theta)^{-\frac{1}{2}}f_B^*(\theta,Y)\right]$

$$
\begin{aligned}
&= \mathrm{E}^*\left[\tfrac{1}{B^2}\textstyle\sum_{i_1}f_{i_1}^{*\prime}\hat{V}_{ff}^{-\frac{1}{2}}P_{\hat{V}_{ff}^{-\frac{1}{2}}\hat{D}_N}\hat{V}_{ff}^{-\frac{1}{2}}\sum_{i_2}\left[f_{i_2}^*f_{i_2}^{*\prime}-\hat{V}_{ff}\right]\hat{V}_{ff}^{-\frac{1}{2}}M_{\hat{V}_{ff}^{-\frac{1}{2}}\hat{D}_N}\hat{V}_{ff}^{-\frac{1}{2}}\sum_{i_3}f_{i_3}^*\right]-\\
&\quad \mathrm{E}^*\left[\tfrac{1}{B^3}\textstyle\sum_{i_1}f_{i_1}^{*\prime}\hat{V}_{ff}^{-\frac{1}{2}}P_{\hat{V}_{ff}^{-\frac{1}{2}}\hat{D}_N}\hat{V}_{ff}^{-\frac{1}{2}}\sum_{i_2}f_{i_2}^*\sum_{i_4}f_{i_4}^{*\prime}\hat{V}_{ff}^{-\frac{1}{2}}M_{\hat{V}_{ff}^{-\frac{1}{2}}\hat{D}_N}\hat{V}_{ff}^{-\frac{1}{2}}\sum_{i_3}f_{i_3}^*\right]\\
&= \mathrm{E}^*\left[\tfrac{1}{B^2}\textstyle\sum_{i_1}\sum_{i_2\neq i_1}f_{i_1}^{*\prime}\hat{V}_{ff}^{-\frac{1}{2}}P_{\hat{V}_{ff}^{-\frac{1}{2}}\hat{D}_N}\hat{V}_{ff}^{-\frac{1}{2}}\left[f_{i_2}^*f_{i_2}^{*\prime}-\hat{V}_{ff}\right]\hat{V}_{ff}^{-\frac{1}{2}}M_{\hat{V}_{ff}^{-\frac{1}{2}}\hat{D}_N}\hat{V}_{ff}^{-\frac{1}{2}}f_{i_1}^*\right]+\\
&\quad \mathrm{E}^*\left[\tfrac{1}{B^2}\textstyle\sum_{i_1}f_{i_1}^{*\prime}\hat{V}_{ff}^{-\frac{1}{2}}P_{\hat{V}_{ff}^{-\frac{1}{2}}\hat{D}_N}\hat{V}_{ff}^{-\frac{1}{2}}\left[f_{i_1}^*f_{i_1}^{*\prime}-\hat{V}_{ff}\right]\hat{V}_{ff}^{-\frac{1}{2}}M_{\hat{V}_{ff}^{-\frac{1}{2}}\hat{D}_N}\hat{V}_{ff}^{-\frac{1}{2}}f_{i_1}^*\right]-\\
&\quad \mathrm{E}^*\left[\tfrac{1}{B^3}\textstyle\sum_{i_1}\sum_{i_2\neq i_1}f_{i_1}^{*\prime}\hat{V}_{ff}^{-\frac{1}{2}}P_{\hat{V}_{ff}^{-\frac{1}{2}}\hat{D}_N}\hat{V}_{ff}^{-\frac{1}{2}}f_{i_2}^*f_{i_2}^{*\prime}\hat{V}_{ff}^{-\frac{1}{2}}M_{\hat{V}_{ff}^{-\frac{1}{2}}\hat{D}_N}\hat{V}_{ff}^{-\frac{1}{2}}f_{i_1}^*\right]-\\
&\quad \mathrm{E}^*\left[\tfrac{1}{B^3}\textstyle\sum_{i_1}\sum_{i_2\neq i_1}f_{i_1}^{*\prime}\hat{V}_{ff}^{-\frac{1}{2}}P_{\hat{V}_{ff}^{-\frac{1}{2}}\hat{D}_N}\hat{V}_{ff}^{-\frac{1}{2}}f_{i_2}^*f_{i_1}^{*\prime}\hat{V}_{ff}^{-\frac{1}{2}}M_{\hat{V}_{ff}^{-\frac{1}{2}}\hat{D}_N}\hat{V}_{ff}^{-\frac{1}{2}}f_{i_2}^*\right]-\\
&\quad \mathrm{E}^*\left[\tfrac{1}{B^3}\textstyle\sum_{i_1}\sum_{i_2\neq i_1}f_{i_1}^{*\prime}\hat{V}_{ff}^{-\frac{1}{2}}P_{\hat{V}_{ff}^{-\frac{1}{2}}\hat{D}_N}\hat{V}_{ff}^{-\frac{1}{2}}f_{i_1}^*f_{i_2}^{*\prime}\hat{V}_{ff}^{-\frac{1}{2}}M_{\hat{V}_{ff}^{-\frac{1}{2}}\hat{D}_N}\hat{V}_{ff}^{-\frac{1}{2}}f_{i_2}^*\right]-\\
&\quad \mathrm{E}^*\left[\tfrac{1}{B^3}\textstyle\sum_{i_1}f_{i_1}^{*\prime}\hat{V}_{ff}^{-\frac{1}{2}}P_{\hat{V}_{ff}^{-\frac{1}{2}}\hat{D}_N}\hat{V}_{ff}^{-\frac{1}{2}}f_{i_1}^*f_{i_1}^{*\prime}\hat{V}_{ff}^{-\frac{1}{2}}M_{\hat{V}_{ff}^{-\frac{1}{2}}\hat{D}_N}\hat{V}_{ff}^{-\frac{1}{2}}f_{i_1}^*\right]-\\
&= \tfrac{B-1}{B^2}\left[\tfrac{1}{N}\textstyle\sum_{i=1}^N\left(\bar{f}_i'\hat{V}_{ff}^{-\frac{1}{2}}P_{\hat{V}_{ff}^{-\frac{1}{2}}\hat{D}_N}\hat{V}_{ff}^{-\frac{1}{2}}\bar{f}_i\bar{f}_i'\hat{V}_{ff}^{-\frac{1}{2}}M_{\hat{V}_{ff}^{-\frac{1}{2}}\hat{D}_N}\hat{V}_{ff}^{-\frac{1}{2}}\bar{f}_i\right)-(k-1)\right]
\end{aligned}
$$

so $\mathrm{E}^*\left[KLM_0^*\right]=1$ and

$$
\begin{aligned}
\mathrm{E}^*\left[KLM_1^*\right]= &-\tfrac{B-1}{B}\left\{\tfrac{1}{N}\textstyle\sum_{i=1}^N\left(\bar{f}_i'\hat{V}_{ff}^{-\frac{1}{2}}P_{\hat{V}_{ff}^{-\frac{1}{2}}\hat{D}_N(\theta_0,Y)}\hat{V}_{ff}^{-\frac{1}{2}}\bar{f}_i\right)^2+\right.\\
&\left.\tfrac{2}{N}\textstyle\sum_{i=1}^N\left(\bar{f}_i'\hat{V}_{ff}^{-\frac{1}{2}}P_{\hat{V}_{ff}^{-\frac{1}{2}}\hat{D}_N(\theta_0,Y)}\hat{V}_{ff}^{-\frac{1}{2}}\bar{f}_i\bar{f}_i'\hat{V}_{ff}^{-\frac{1}{2}}M_{\hat{V}_{ff}^{-\frac{1}{2}}\hat{D}_N(\theta_0,Y)}\hat{V}_{ff}^{-\frac{1}{2}}\bar{f}_i\right)-2(k+1)\right\}+\tfrac{1}{B}.
\end{aligned}
$$

Proof of Theorem 3.

A.1. For the Edgeworth approximation of the distribution of $\mathrm{S}(\theta)$, we construct the characteristic function of $\mathrm{S}(\theta_0)$:

$$
\begin{aligned}
&\mathrm{cf}_{\mathrm{S}(\theta)}(t)\\
&= \int\exp\left(it\mathrm{S}(\theta)\right)p_{(\hat{f}_N(\theta,Y),\hat{V}_{ff}(\theta))}(u,v)dudv\\
&= \int\exp\left(it\left[S_0+\tfrac{1}{N}S_1+\tfrac{1}{N^2}S_2\right]\right)p_{(\hat{f}_N(\theta,Y),\hat{V}_{ff}(\theta))}(u,v)dudv+o(N^{-2})\\
&= \int\exp\left(itS_0\right)\exp\left(\tfrac{it}{N}S_1+\tfrac{it}{N^2}S_2\right)p_{(\hat{f}_N(\theta,Y),\hat{V}_{ff}(\theta))}(u,v)dudv+o(N^{-2})\\
&= \int\left[1+\tfrac{it}{N}S_1\right]\exp\left(itS_0\right)p_{(\hat{f}_N(\theta,Y),\hat{V}_{ff}(\theta))}(u,v)dudv+O(N^{-2}),
\end{aligned}
$$

where $p_{(\hat{f}_N(\theta,Y),\hat{V}_{ff}(\theta))}(u,v)$ is the joint density of $\hat{f}_N(\theta,Y)$ and $\hat{V}_{ff}(\theta)$ for a sample of size N. The last equation results from a Taylor approximation of $\exp\left(\frac{it}{N}S_1+\frac{it}{N^2}S_2\right)$. The order of the error term results from the Mean Value Theorem and the convergence rates of $\left(\frac{it}{N}S_1\right)^2$ and $\frac{it}{N^2}S_2$.

Lemma 8 states that the expected value of S_1 stated in Theorem 1 is identical to its expected value with respect to the joint distribution of $\hat{f}_N(\theta,Y)$ and $\hat{V}_{ff}(\theta)$, $\mathrm{E}(S_1)$. We use the limiting distribution of $\sqrt{N}\hat{f}_N(\theta,Y)$ to obtain the remaining part of the characteristic function. Since this limiting distribution is $N(0,V_{ff}(\theta))$, the addition of $\exp(itS_0)$ alters the variance to $\frac{1}{1-2it}V_{ff}(\theta)$. Since $\hat{f}_N(\theta,Y)$ is transformed to $\sqrt{(1-2it)}\hat{f}_N(\theta,Y)$, the characteristic function that results reads

$$\mathrm{cf}_{\mathrm{S}(\theta)}(t) = \quad (1-2it)^{-\frac{1}{2}k}\left[1+\tfrac{1}{N}\tfrac{it}{(1-2it)}\mathrm{E}(S_1)+O(N^{-2})\right],$$

where $(1-2it)^{-\frac{1}{2}k}$ results from the Jacobian of the transformation from $\sqrt{N}\hat{f}_N(\theta,Y)$ to $\sqrt{N(1-2it)}\hat{f}_N(\theta,Y)$ and we used that

$$S_1 = -\tfrac{N^2}{1-2it}\left[\sqrt{(1-2it)}\hat{f}_N(\theta,Y)\right]' V_{ff}(\theta)^{-1}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]V_{ff}(\theta)^{-1}\left[\sqrt{(1-2it)}\hat{f}_N(\theta,Y)\right].$$

We now use Theorem 2.4 from Phillips and Park (1988) to obtain the Edgeworth approximation from the above characteristic function:

$$\begin{aligned}\Pr\left[\mathrm{S}\left(\theta_0\right)\leq x\right] = &\ \Pr_{\chi^2(k)}(x)-\tfrac{1}{N}\tfrac{\mathrm{E}(S_1)}{k}xp_{\chi^2(k)}(x)+O(N^{-2})\\ = &\ \Pr_{\chi^2(k)}\left(x-\tfrac{1}{N}\tfrac{\mathrm{E}(S_1)}{k}x\right)+O(N^{-2}),\end{aligned}$$

where $\Pr_{\chi^2(k)}(x)$ and $p_{\chi^2(k)}(x)$ are the distribution and density function of a $\chi^2(k)$ distributed random variable evaluated at x. The Edgeworth approximation thus amounts to dividing the standard $\chi^2(k)$ critical values by $\left(1-\frac{1}{N}\frac{\mathrm{E}(S_1)}{k}\right)$.

A.2. For the Edgeworth approximation of the distribution of KLM(θ), we construct the conditional characteristic function of KLM(θ_0) given $D_N(\theta,Y)$:

$$\begin{aligned}&\mathrm{cf}_{\mathrm{KLM}(\theta)}(t|D_N(\theta,Y))\\ &= \int\exp\left(it\mathrm{KLM}(\theta)\right)p_{(\hat{f}_N(\theta,Y),\hat{V}_{ff}(\theta),\hat{V}_{\theta f}(\theta))}(u,v,w|D_N(\theta,Y))dudvdw\\ &= \int\exp\left(it\left[KLM_0+\tfrac{1}{N}KLM_1+\tfrac{1}{N\sqrt{N}}KLM_2\right]\right)\\ &\qquad p_{(\hat{f}_N(\theta,Y),\hat{V}_{ff}(\theta),\hat{V}_{\theta f}(\theta))}(u,v,w|D_N(\theta,Y))dudvdw+o(\tfrac{1}{N\sqrt{N}})\\ &= \int\exp\left(itKLM_0\right)\exp\left(\tfrac{it}{N}KLM_1+\tfrac{it}{N\sqrt{N}}KLM_2\right)\\ &\qquad p_{(\hat{f}_N(\theta,Y),\hat{V}_{ff}(\theta),\hat{V}_{\theta f}(\theta))}(u,v,w|D_N(\theta,Y))dudvdw+o(\tfrac{1}{N\sqrt{N}})\\ &= \int\left[1+\tfrac{it}{N}KLM_1+\tfrac{it}{N\sqrt{N}}KLM_2\right]\exp\left(itKLM_0\right)\\ &\qquad p_{(\hat{f}_N(\theta,Y),\hat{V}_{ff}(\theta),\hat{V}_{\theta f}(\theta))}(u,v,w|D_N(\theta,Y))dudvdw+o(\tfrac{1}{N\sqrt{N}}),\end{aligned}$$

where $p_{(\hat{f}_N(\theta,Y),\hat{V}_{ff}(\theta),\hat{V}_{\theta f}(\theta))}(u,v,w|D_N(\theta,Y))$ is the conditional density of $(\hat{f}_N(\theta,Y),\hat{V}_{ff}(\theta),\hat{V}_{\theta f}(\theta))$ given $D_N(\theta,Y)$. To obtain the characteristic function of KLM(θ), we note that the exponent term of the limiting distribution of $\sqrt{N}\hat{f}_N(\theta,Y)$ reads

$$\exp\left[-\tfrac{1}{2}N\hat{f}_N(\theta,Y)'V_{ff}(\theta)^{-1}\hat{f}_N(\theta,Y)\right].$$

Combining this exponent term with $\exp(itKLM_0)$, the exponent term becomes

$$
\begin{aligned}
&\exp\left[itKLM_0 - \tfrac{1}{2}N\hat{f}_N(\theta,Y)'V_{ff}(\theta)^{-1}\hat{f}_N(\theta,Y)\right] \\
= \ &\exp\left[itNf_N(\theta,Y)'V_{ff}(\theta)^{-\frac{1}{2}}P_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}f_N(\theta,Y)- \right.\\
&\left.\tfrac{1}{2}Nf_N(\theta,Y)'V_{ff}(\theta)^{-\frac{1}{2}}\left(P_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)} + M_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}\right)V_{ff}(\theta)^{-\frac{1}{2}}f_N(\theta,Y)\right] \\
= \ &\exp\left[-\tfrac{1}{2}(1-2it)\,Nf_N(\theta,Y)'V_{ff}(\theta)^{-\frac{1}{2}}P_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}f_N(\theta,Y)- \right.\\
&\left.\tfrac{1}{2}Nf_N(\theta,Y)'V_{ff}(\theta)^{-\frac{1}{2}}M_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}f_N(\theta,Y)\right] \\
= \ &\exp\left[-\tfrac{1}{2}(1-2it)\,KLM_0- \right.\\
&\left.\tfrac{1}{2}Nf_N(\theta,Y)'V_{ff}(\theta)^{-\frac{1}{2}}M_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}f_N(\theta,Y)\right].
\end{aligned}
$$

The KLM_1 higher order component of KLM(θ) contains elements of both components,

$$
P_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}f_N(\theta,Y) \text{ and } M_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}f_N(\theta,Y),
$$

of this exponent term:

$$
\begin{aligned}
&KLM_1 \\
&= \ -N^2f_N(\theta,Y)'V_{ff}(\theta)^{-\frac{1}{2}}P_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right] \\
&\quad V_{ff}(\theta)^{-\frac{1}{2}}P_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}f_N(\theta,Y)- \\
&\quad 2N^2f_N(\theta,Y)'V_{ff}(\theta)^{-\frac{1}{2}}P_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right] \\
&\quad V_{ff}(\theta)^{-\frac{1}{2}}M_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}f_N(\theta,Y)
\end{aligned}
$$

and

$$
\begin{aligned}
a = \ &\mathrm{E}\left[N^2f_N(\theta,Y)'V_{ff}(\theta)^{-\frac{1}{2}}P_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]\right. \\
&\left.V_{ff}(\theta)^{-\frac{1}{2}}P_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}f_N(\theta,Y)|d_i\right] \\
= \ &\tfrac{N-1}{N}\left\{\mathrm{E}\left[\left(f_i'V_{ff}^{-\frac{1}{2}}P_{V_{ff}^{-\frac{1}{2}}D_N}V_{ff}^{-\frac{1}{2}}f_i\right)^2|d_i\right]-4\right\}-\tfrac{1}{N}, \\
b = \ &\mathrm{E}\left[N^2f_N(\theta,Y)'V_{ff}(\theta)^{-\frac{1}{2}}P_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}\left[\hat{V}_{ff}(\theta)-V_{ff}(\theta)\right]\right. \\
&\left.V_{ff}(\theta)^{-\frac{1}{2}}M_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}f_N(\theta,Y)|d_i\right] \\
= \ &\tfrac{N-1}{N}\left\{\mathrm{E}\left[f_i'V_{ff}^{-\frac{1}{2}}P_{V_{ff}^{-\frac{1}{2}}D_N}V_{ff}^{-\frac{1}{2}}f_if_i'V_{ff}^{-\frac{1}{2}}M_{V_{ff}^{-\frac{1}{2}}D_N}V_{ff}^{-\frac{1}{2}}f_i|d_i\right]-(k-1)\right\}.
\end{aligned}
$$

The expression of the exponent term shows that the elements of $P_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}f_N(\theta,Y)$ are multiplied by $\sqrt{1-2it}$ while those of $M_{V_{ff}(\theta)^{-\frac{1}{2}}D_N(\theta,Y)}V_{ff}(\theta)^{-\frac{1}{2}}f_N(\theta,Y)$ remain unaltered. We take account of these transformations when we construct the characteristic function of KLM(θ) given $D_N(\theta_0,Y)$ for which we use the limiting distribution of $\sqrt{N}\hat{f}_N(\theta_0,Y)$:

$$
\begin{aligned}
&\mathrm{cf}_{\mathrm{KLM}(\theta)}(t|D_N(\theta_0,Y)) = \int\left[1+\tfrac{it}{N}KLM_1+\tfrac{it}{N\sqrt{N}}KLM_2\right]\exp\left(itKLM_0\right) \\
&\qquad p_{(\hat{f}_N(\theta,Y),\hat{V}_{ff}(\theta),\hat{V}_{\theta f}(\theta))}(u,v,w|D_N(\theta,Y))dudvdw + o(\tfrac{1}{N\sqrt{N}}) \\
&= \ (1-2it)^{-\frac{1}{2}}\left[1-\tfrac{1}{N}\tfrac{it}{(1-2it)}a-\tfrac{2}{N}\tfrac{it}{\sqrt{1-2it}}b+O(\tfrac{1}{N\sqrt{N}})\right],
\end{aligned}
$$

where $(1-2it)^{-\frac{1}{2}}$ results from the Jacobian of the transformation from KLM_0 to $(1-2it)KLM_0$ and we used Lemma 8 to obtain the expected values with respect to the joint distribution of $(\hat{f}_N(\theta,Y),\hat{V}_{ff}(\theta),\hat{V}_{\theta f}(\theta))$ from Theorem 1. Using Phillips and Park (1988), we obtain the Edgeworth approximation from the characteristic function:

$$\begin{aligned}\Pr\left[\mathrm{KLM}\left(\theta_0\right)\leq x|D_N(\theta_0,Y)\right] = &\ \Pr{}_{\chi^2(1)}(x)+\tfrac{1}{N}\left(ax+\sqrt{2\pi}\mathrm{E}bx^{\frac{1}{2}}\right)p_{\chi^2(1)}(x)+O(N^{-2})\\ = &\ \Pr{}_{\chi^2(1)}\left(x+\tfrac{1}{N}\left(ax+\sqrt{2\pi}bx^{\frac{1}{2}}\right)\right)+o(N^{-2}).\end{aligned}$$

The Edgeworth approximation is obtained by using the Fourier transform of the characteristic function:

$$p(x)=\tfrac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{cf}_x(t)\exp\left[-itx\right]dt.$$

It is well known that $(1-2it)^{-\frac{1}{2}}$ is the characteristic function of the $\chi^2(1)$ distribution. The above Fourier transform shows that

$$\frac{\partial p_{\chi^2(k)}(x)}{\partial x}=\frac{\partial}{\partial x}\left[\tfrac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{cf}_{\chi^2(k)}(t)\exp\left[-itx\right]dt\right]=\tfrac{1}{2\pi}\int_{-\infty}^{\infty}-\frac{it}{(1-2it)^{\frac{k}{2}}}\exp\left[-itx\right]dt$$

so the Fourier transform of $\frac{it}{(1-2it)}$ is $-\frac{\partial p_{\chi^2(2)}(x)}{\partial x}$ and that of $\frac{it}{(1-2it)^{\frac{3}{2}}}$ is $\frac{\partial p_{\chi^2(3)}(x)}{\partial x}$. The density function that results from the characteristic function is therefore:

$$p_{\mathrm{KLM}(\theta_0)}(x|D_N(\theta_0,Y))=\ p_{\chi^2(1)}(x)+\tfrac{1}{N}\left[a\frac{\partial p_{\chi^2(3)}(x)}{\partial x}+2b\frac{\partial p_{\chi^2(2)}(x)}{\partial x}\right]+O(\tfrac{1}{N\sqrt{N}}).$$

When we integrate the density function to obtain the distribution function we obtain:

$$\Pr\left[\mathrm{KLM}\left(\theta_0\right)\leq x|D_N(\theta_0,Y)\right]=\ \Pr{}_{\chi^2(1)}(x)+\tfrac{1}{N}\left(ap_{\chi^2(3)}(x)+2bp_{\chi^2(2)}(x)\right)+O(\tfrac{1}{N\sqrt{N}}).$$

The density function of the $\chi^2(1)$ random variable is such that $p_{\chi^2(2)}(x)=\frac{\Gamma(\frac{1}{2})}{\Gamma(1)\sqrt{2}}\sqrt{x}p_{\chi^2(1)}(x)=\sqrt{\frac{\pi}{2}}\sqrt{x}p_{\chi^2(1)}(x)$ since $\Gamma(1)=1$ and $\Gamma(\frac{1}{2})=\sqrt{\pi}$. Similarly, $p_{\chi^2(3)}(x)=\frac{\Gamma(\frac{1}{2})}{\Gamma(\frac{3}{2})2}xp_{\chi^2(1)}(x)=\frac{\Gamma(\frac{1}{2})}{\frac{1}{2}\Gamma(\frac{1}{2})2}xp_{\chi^2(1)}(x)=xp_{\chi^2(1)}(x)$ since $\Gamma(\frac{3}{2})=\frac{1}{2}\Gamma(\frac{1}{2})$. Inserting these expressions, we obtain that

$$\begin{aligned}\Pr\left[\mathrm{KLM}\left(\theta_0\right)\leq x|D_N(\theta_0,Y)\right]= &\ P_{\chi^2(1)}(x)+\tfrac{1}{N}\left(ax+\sqrt{2\pi}b\sqrt{x}\right)p_{\chi^2(1)}(x)+O(\tfrac{1}{N\sqrt{N}})\\ = &\ P_{\chi^2(1)}\left(x+\tfrac{1}{N}\left(ax+\sqrt{2\pi}b\sqrt{x}\right)\right)+O(\tfrac{1}{N\sqrt{N}}).\end{aligned}$$

Thus the Edgeworth approximation alters a $\chi^2(1)$ critical value x towards $x-\frac{1}{N}\left(ax+\sqrt{2\pi}b\sqrt{x}\right)$.

B.1. For the Edgeworth approximation of the distribution of $\mathrm{S}^*(\theta)$, we construct the characteristic function of $\mathrm{S}^*(\theta_0)$:

$$\begin{aligned}\mathrm{cf}_{\mathrm{S}^*(\theta)}(t) &= \int\exp\left(it\mathrm{S}^*(\theta)\right)p_{(f_B^*(\theta,Y),V_{ff}^*(\theta))}(u,v)dudv\\ &= \int\exp\left(it\left[S_0^*+\tfrac{1}{N}S_1^*\right]\right)p_{(f_B^*(\theta,Y),V_{ff}^*(\theta))}(u,v)dudv+o(B^{-1})\\ &= \int\left[1+\tfrac{it}{B}S_1^*\right]\exp\left(itS_0^*\right)p_{(f_B^*(\theta,Y),V_{ff}^*(\theta))}(u,v)dudv+o(B^{-1}),\end{aligned}$$

where $p_{(f_B^*(\theta,Y),V_{ff}^*(\theta))}(u,v)$ is the joint density of $f_B^*(\theta,Y)$ and $V_{ff}^*(\theta)$ for a bootstrap sample of size B. The order of the error term results from the Mean Value Theorem and the convergence rates of $\left(\frac{it}{B}S_1^*\right)^2$. The addition of $\exp(itS_0^*)$ alters the variance of the limiting distribution of $\sqrt{B}f_B^*(\theta,Y)$ so it becomes $N(0,\frac{1}{1-2it}\hat{V}_{ff}(\theta))$. The characteristic function then becomes

$$\mathrm{cf}_{\mathrm{S}^*(\theta)}(t)=\ (1-2it)^{-\frac{1}{2}k}\left[1+\tfrac{1}{B}\tfrac{it}{(1-2it)}\mathrm{E}(S_1^*)+o(B^{-1})\right],$$

where $\mathrm{E}(S_1^*)$ results from the higher order specification of $\mathrm{S}^*(\theta)$ in Theorem 2, such that the Edgeworth approximation reads:

$$\Pr\left[\mathrm{S}^*\left(\theta_0\right) \leq x\right] = \Pr_{\chi^2(k)}\left(x - \tfrac{1}{B}\tfrac{\mathrm{E}(S_1^*)}{k}x\right) + o(B^{-1}).$$

B.2. For the Edgeworth approximation of the distribution of $\mathrm{KLM}^*(\theta)$, we construct the characteristic function of $\mathrm{KLM}^*(\theta_0)$ given $\hat{D}_N(\theta_0, Y)$:

$$\begin{aligned}
&\mathrm{cf}_{\mathrm{KLM}^*(\theta)}(t|\hat{D}_N(\theta_0,Y)) \\
&= \int \exp\left(it\mathrm{KLM}^*(\theta)\right) p_{(f_B^*(\theta,Y),V_{ff}^*(\theta))}(u,v)dudv \\
&= \int \exp\left(it\left[KLM_0^* + \tfrac{1}{B}KLM_1^*\right]\right) p_{(f_B^*(\theta,Y),V_{ff}^*(\theta))}(u,v)dudv + o(B^{-1}) \\
&= \int \left[1 + \tfrac{it}{B}KLM_1^*\right] \exp\left(itKLM_0^*\right) p_{(f_B^*(\theta,Y),V_{ff}^*(\theta))}(u,v)dudv + o(B^{-1}).
\end{aligned}$$

Identical to $\mathrm{KLM}(\theta)$, the exponent term $\exp(itKLM_0^*)$ alters the exponent term of the limiting distribution such that the characteristic function of $\mathrm{KLM}^*(\theta)$ reads

$$\begin{aligned}
&\mathrm{cf}_{\mathrm{KLM}^*(\theta)}(t|\hat{D}_N(\theta_0,Y)) \\
&= \int \left[1 + \tfrac{it}{B}KLM_1^*\right] \exp\left(itKLM_0^*\right) p_{(f_B^*(\theta,Y),V_{ff}^*(\theta))}(u,v)dudv + o(B^{-1}) \\
&= (1-2it)^{-\frac{1}{2}}\left[1 - \tfrac{1}{B}\tfrac{it}{(1-2it)}\mathrm{E}(a^*) - \tfrac{2}{B}\tfrac{it}{\sqrt{1-2it}}\mathrm{E}(b^*) + o(B^{-1})\right],
\end{aligned}$$

where

$$\begin{aligned}
a^* &= \tfrac{B-1}{B}\left\{\tfrac{1}{N}\sum_{i=1}^N\left[\bar{f}(\theta_0)_i'\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}P_{\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}\hat{D}_N(\theta_0,Y)}\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}\bar{f}(\theta_0)\right]^2 - 4\right\} - \tfrac{1}{B}, \\
b^* &= \tfrac{B-1}{B}\left\{\tfrac{1}{N}\sum_{i=1}^N\left[\bar{f}(\theta_0)_i'\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}P_{\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}\hat{D}_N(\theta_0,Y)}\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}\bar{f}(\theta_0)\right.\right. \\
&\quad \left.\left.\bar{f}(\theta_0)_i'\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}M_{\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}\hat{D}_N(\theta_0,Y)}\hat{V}_{ff}(\theta_0)^{-\frac{1}{2}}\bar{f}(\theta_0)\right] - (k-1)\right\}.
\end{aligned}$$

The Edgeworth approximation of $\mathrm{KLM}^*(\theta_0)$ then reads

$$\begin{aligned}
\Pr\left[\mathrm{KLM}^*\left(\theta_0\right) \leq x|\hat{D}_N(\theta_0,Y)\right] &= \Pr_{\chi^2(1)}(x) + \tfrac{1}{B}\left(a^*x + \sqrt{2\pi}b^*x^{\frac{1}{2}}\right)p_{\chi^2(1)}(x) + o(B^{-1}) \\
&= \Pr_{\chi^2(1)}\left(x + \tfrac{1}{B}\left(a^*x + \sqrt{2\pi}b^*x^{\frac{1}{2}}\right)\right) + o(B^{-1}).
\end{aligned}$$

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] Anderson, T.W. and H. Rubin. Estimators of the Parameters of a Single Equation in a Complete Set of Stochastic Equations. *The Annals of Mathematical Statistics*, **21**:570–582, (1949).

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

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

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

[8] 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.

[9] Kleibergen, F. Testing Parameters in GMM without assuming that they are identified. *Econometrica*, **73**:1103–1124, 2005.

[10] Kleibergen, F. Generalizing weak instrument robust IV statistics towards multiple parameters, unrestricted covariance matrices and identification statistics. *Journal of Econometrics*, 2006a. Forthcoming.

[11] Kleibergen, F. Subset Statistics in the Linear IV Regression Model. 2006b. Technical Report, Brown University.

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

[13] Moreira, M.J., J.R. Porter and G. Suarez. Bootstrap and Higher-Order Expansion Validity When Instruments May Be Weak. 2004. Working Paper, Harvard University.

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

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

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

[17] 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.

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

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

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