REM WORKING PAPER SERIES

Generalised Empirical Likelihood Kernel Block Bootstrapping

Paulo M.D.C. Parente, Richard J. Smith

REM Working Paper 055-2018

November 2018

REM – Research in Economics and Mathematics
Rua Miguel Lúpi 20,
1249-078 Lisboa,
Portugal

ISSN 2184-108X

Generalised Empirical Likelihood Kernel Block Bootstrapping

Paulo M.D.C. Parente
ISEG- Lisbon School of Economics & Management, Universidade de Lisboa
REM - Research in Economics and Mathematics;
CEMAPRE- Centro de Matemática Aplicada à Previsão e Decisão Económica.
pparente@iseg.ulisboa.pt

Richard J. Smith
*c*emmap, U.C.L and I.F.S.
Faculty of Economics, University of Cambridge
Department of Economics, University of Melbourne
ONS Economic Statistics Centre of Excellence
rjs27@econ.cam.ac.uk

This Draft: October 2018

Abstract

This article unveils how the kernel block bootstrap method of Parente and Smith (2018a,2018b) can be applied to make inferences on parameters of models defined through moment restrictions. Bootstrap procedures that resort to generalised empirical likelihood implied probabilities to draw observations are also introduced. We prove the first-order asymptotic validity of bootstrapped test statistics for overidentifying moment restrictions, parametric restrictions and additional moment restrictions. Resampling methods based on such probabilities were shown to be efficient by Brown and Newey (2002). A set of simulation experiments reveals that the statistical tests based on the proposed bootstrap methods perform better than those that rely on first-order asymptotic theory.

JEL Classification: C14, C15, C32
Keywords: Bootstrap; heteroskedastic and autocorrelation consistent inference; Generalised Method of Moments; Generalised Empirical Likelihood

1 Introduction

The objective of this article is to propose new bootstrap methods for models defined through moment restrictions in the time-series context using a novel bootstrap method introduced recently by Parente and Smith (2018a, 2018b). Simultaneously, we amend some of the existent results in the related literature.

The generalized method of moments (GMM) estimator of Hansen (1982) has become one of the most popular tools in econometrics due to its applicability in different and varied situations. It can be used, for instance to estimate parameters of interest under endogeneity and measurement error. Consequently, the richness of the set of inferential statistics provided by GMM may be extremely useful to economists doing empirical work. These statistics allow to test for overidentifying moment conditions, parametric restrictions and additional moment conditions.

The performance of statistics based on GMM has been revealed to be poor in finite samples and this situation worsens in time-series data due to the presence of autocorrelation [see Newey and West (1994), Burnside and Eichenbaum (1996), Christiano and den Haan (1996) among others]. To tackle this problem several alternative approaches have been proposed in the literature, being the bootstrap among the methods that has produced better results. The bootstrap is a resampling method introduced by Efron (1979) to make inferences on parameters of interest. It can be used not only to approximate the (asymptotic) distribution of an estimator or statistic, but also to estimate its variance. From the practical standpoint it has the benefit of not requiring the application of complicated formulae and from the theoretical viewpoint it allows to obtain asymptotic refinements when the statistic of interest is smooth and asymptotically pivotal.

Bootstrap methods in the context of moment restrictions have been introduced previously by Hahn (1996) and Brown and Newey (2002) for random samples and Hall and Horowitz (1996), Andrews (2002), Inoue and Shintani (2006), Allen, et al. (2011) and Bravo and Crudu (2011) for dependent data. This literature can be divided in two strands.

Hahn (1996) proves consistency of the `i.i.d.` bootstrap distribution for GMM, but he did not consider bootstrapped test statistics based on GMM. Hall and Horowitz (1996), Andrews (2002) and Inoue and Shintani (2006) propose the use of the standard moving blocks bootstrap applied to GMM. A second line of research is followed by Brown and Newey (2002), Allen, et al. (2011) and Bravo and

Crudu (2011) who use empirical likelihood and generalised empirical likelihood implied probabilities to draw observations or blocks of data.

Hall and Horowitz (1996) suggested applying the non-overlapping blocks bootstrap method of Carlstein (1986) to GMM after centering the bootstrap moment restrictions at their sample means. They prove that this method yields asymptotic refinements not only for the bootstrapped $\mathcal{J}$ statistic of Hansen (1982), but also for the bootstrapped t statistic for testing a single parametric restriction. Andrews (2002) extends Hall and Horowitz (1996) method to overlapping moving blocks bootstrap of Künsch (1989) and Liu and Singh (1992) and the k-step bootstrap of Davidson and Mackinnon (1999). However, Hall and Horowitz (1996) and Andrews (2002) require uncorrelateness of the moment indicators after a certain number of lags. This assumption is relaxed by Inoue and Shintani (2006) in the special case of linear models estimated using instruments.

Brown and Newey (2002) in the `i.i.d.` setting mention, though without a formal proof, that the same improvements can be obtained, by using a method that they denominate empirical likelihood (EL) bootstrap. The EL bootstrap consists in first computing the empirical likelihood implied probabilities associated with each observation under a set of moment restrictions and using these probabilities to draw each observation in order to construct the bootstrap samples. Although Brown and Newey (2002) did not prove the asymptotic validity of the method, they showed heuristically that it is efficient in the sense that the difference between the finite sample distribution of a statistic and its EL bootstrap counterpart is asymptotically normal (after a proper scaling) with minimum variance. Recently the EL bootstrap method was extended to the time series context by Allen, et al. (2011) and Bravo and Crudu (2011) using a MBB procedure. Both articles suggest first computing implied probabilities for blocks of observations and use these probabilities to draw blocks in order to construct the bootstrap samples. There are some differences between these two articles. Firstly, while Allen, et al. (2011) consider EL implied probabilities, Bravo and Crudu (2011) use the generalised empirical likelihood (GEL) implied probabilities of Smith (2011). Secondly, Allen, et al. (2011) propose using both non-overlapping blocks and overlapping blocks whereas Bravo and Crudu (2011) only study the latter. Thirdly, Allen, et al. (2011) investigate the first order validity of the method for general GMM estimators and Bravo and Crudu (2011) consider only the efficient GMM estimator. Both articles address the first-order asymptotic behaviour of bootstrapped $\mathcal{J}$ statistic and bootstrapped Wald ($\mathcal{W}$) statistics tests for parametric restrictions. Finally, in the case of tests of parametric restrictions, Bravo and Crudu (2011), additionally, propose drawing bootstrap samples based on the GEL implied probabilities computed under the null hypothesis and the moment restrictions and put forward the bootstrapped Lagrange multiplier ($\mathcal{LM}$) and distance ($\mathcal{D}$) statistics in this framework.

In this article we also consider a time-series setting, but depart from the dominant paradigm of using bootstrap methods based on moving blocks and introduce an alternative to these resampling schemes based on the kernel block bootstrap (KBB) method of Parente and Smith (2018a, 2018b). The KBB method consists in transforming the data using weighted moving averages of all observations and drawing bootstrap samples with replacement from the transformed sample. This method is akin to the Tapered Block Bootstrap (TBB) method of Paparoditis and Politis (2001) in that if the kernel chosen is of bounded support the KBB method can be seen as a variant of TBB that allows the inclusion of incomplete blocks. However, KBB can be implemented also using kernels with unbounded support. In the case of the sample mean and for a particular choice of the kernel with unbounded support it allows to obtain a bootstrap variance estimator that is asymptotically equivalent to the quasi-spectral estimator of the long run variance which Andrews (1991) proved to be optimal. Additionally, the technical assumptions required by Paparoditis and Politis (2001) to prove the asymptotic validity of TBB are not satisfied by truncated kernels that are non-monotonic in the positive quadrant such as the flap-top cosine windows described in D'Antona and Ferrero (2006, p.40), while KBB can be applied using this kernel. We note however that both TBB and KBB allow the most popular truncated kernels to be used, such as the rectangular, Bartlett and Tuckey-Hanning.

We use the new method to approximate the asymptotic distribution of the $\mathcal{J}$ statistic of Hansen (1982) that allows to test for the overidentifying moment restrictions, and the trinity of test statistics (Wald, Lagrange multiplier and distance statistics, cf. Newey and McFadden 1994, section 9 and Ruud,2000, chapter 22) that permit testing parametric restrictions and additional moment conditions. We show that the first order validity of the bootstrap test for overidentifying conditions GMM estimator does not require prior centering of the bootstrap moments, this centering can be done a posteriori.

In the spirit of Brown and Newey (2002), we propose additionally to use the GEL implied probability associated with each transformed observation [Smith, 2011] to construct the bootstrap sample. We prove the first order validity of the method and corresponding test statistics. As Allen et al. (2011) and Bravo and Crudu (2011) we prove the first order validity of the bootstrapped distribution of the estimator and the bootstrapped $\mathcal{J}$ statistic, and tests for parametric restrictions and additional moment conditions.

We show in this article that the proof of consistency of the EL block bootstrap of Allen, et al. (2011) is in error in that when applied to the inefficient GMM estimator the bootstrap distribution of the latter has to be centered at the efficient GMM estimator. Hence the results stated in their Theorem 1 and 2 are invalid in general, though they hold if the weighting matrix is a consistent estimator of the inverse of the covariance matrix of the moment indicators [cf. Theorem 1 of Bravo and Crudu (2010).] Although our proof of this results applies only to the new bootstrap methods introduced in this article, the demonstration for EL block bootstrapping is analogous.

When testing for parametric restrictions and additional moment conditions the GEL implied probabilities can be computed under the null or under the maintained hypothesis. Hence, two types of KBB bootstrap methods can be used, one using the GEL implied probabilities computed under the maintained hypothesis as in Brown and Newey (2002) and Allen et al. (2011) and another based on these probabilities computed under the null as suggest in the case of parametric restrictions by Bravo and Crudu (2011). This article investigates these two types of bootstrap methods. We note that Allen, et al. (2011) in the case of EL block bootstrap actually do not present the formula of the bootstrapped Wald statistic, though their Theorem 3, which is based on theirs incorrect Theorems 1 and 2, refers to it. On the other hand, the formula for this statistic presented in Bravo and Crudu (2011) is only valid if the implied probabilities were computed under the maintained hypothesis and not under the null hypothesis, though it is presented jointly with the $\mathcal{LM}$ and $\mathcal{D}$ statistics which are obtained with the implied probabilities computed under the null. We show that the trinity of tests statistics can be computed using implied probabilities obtained under the null and under the maintained hypothesis and that they have different mathematical expressions depending on the resampling scheme chosen.

This paper is organized as follows. In the first section we introduce the KBB-method for moment restrictions. In section 2 we summarize some important results on GMM and GEL in the time-series context. The KBB method is briefly explained in section 3. In section 4 we present the first order asymptotic theory on the KBB methods computed using the following different probabilities to draw observations: uniform (standard non-parametric KBB method), the implied probabilities associated with the moment restrictions and the implied probabilities associated with the maintained hypothesis, parametric restrictions and additional moment conditions. In section 5 we present a Monte Carlo study in which we investigate the performance of the proposed bootstrap methods in finite samples. Finally section 6 concludes. The proofs of the results are given in the Appendix.

2 Framework

Let z_t, $(t = 1, ..., T)$ denote observations on a finite dimensional (strictly) stationary process $\{z_t\}_{t=1}^{\infty}$. We assume initially that the process is ergodic, but later we will require the stronger condition of mixing. Consider the moment indicator $g(z_t, \beta)$, an $m-$vector functions of the data observation z_t and the p-vector β of unknown parameters which are the object of inferential interest, where $m \geq p$. It is assumed that the true parameter vector β_0 uniquely satisfies the moment condition

$$\mathrm{E}[g(z_t, \beta_0)] = 0,$$

where $\mathrm{E}[\cdot]$ denotes expectation taken with respect to the unknown distribution of z_t.

2.1 The Generalized Method of Moments estimator

2.1.1 The Estimator

For notational simplicity we define $g_t(\beta) \equiv g(z_t, \beta)$, $(t = 1, ..., T)$, and $\hat{g}(\beta) \equiv \sum_{s=1}^{T} g_t(\beta)/T$, let also $G_t(\beta) \equiv \partial g_t(\beta)/\partial \beta'$, $(t = 1, ..., T)$, $G \equiv \mathrm{E}[G_t(\beta_0)]$ and $\Omega \equiv \lim_{n\to\infty} \mathrm{var}[\sqrt{T}\hat{g}(\beta_0)]$. Denote $\hat{W}$ a symmetric weighting matrix that converges in probability to a non-random matrix W. The GMM estimator is defined as

$$\begin{aligned} \hat{\beta} &= \arg\min_{\beta \in \mathcal{B}} \hat{Q}(\beta), \\ \hat{Q}_T(\beta) &= \hat{g}(\beta)'\hat{W}\hat{g}(\beta). \end{aligned}$$

Hansen (1982) showed that under some regularity conditions $\hat{\beta} \xrightarrow{p} \beta_0$ and

$$\sqrt{T}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, \mathrm{avar}(\hat{\beta})),$$

where $\xrightarrow{p}$ and $\xrightarrow{d}$ denote convergence in probability and distribution respectively and

$$\mathrm{avar}(\hat{\beta}) = (G'WG)^{-1}G'W\Omega WG(G'WG)^{-1}.$$

Denote $\Sigma \equiv (G'\Omega^{-1}G)^{-1}$ and $\hat{G}(\beta) = \sum_{i=1}^{T} \hat{G}_t(\beta)/T$, $\hat{G} = \hat{G}(\hat{\beta})$. Hansen (1982) proved also that the most efficient GMM estimator $\hat{\beta}^e$ is obtained when we set $W = \Omega^{-1}$ and in this case $\mathrm{avar}(\hat{\beta}^e) = \Sigma$.

We consider the following regularity conditions that are sufficient to prove consistency.

Assumption 2.1 **(i)** *The observed data are realizations of a stochastic process* $z \equiv \{z_t : \Omega \rightarrow \mathbb{R}^n, n \in \mathbb{N}, t = 1, 2, ...\}$ *on the complete probability space* $(\Omega, \mathcal{F}, P)$ *where* $\Omega = \times_{t=1}^{\infty} \mathbb{R}^k$ *and* $\mathcal{F} = \mathcal{B}(\times_{t=1}^{\infty} R^n)$ *(the Borel* $\sigma-$*field generated by the measure finite dimension product cylinders);* **(ii)** z_t *is stationary and ergodic ;* **(iii)** $g(., \beta)$ *is Borel measurable for each* $\beta \in \mathcal{B}$, $g(z_t, \beta)$ *is continuous on* $\mathcal{B}$ *for each* $z_t \in \mathcal{Z}$, **(iv)** $\mathrm{E}[\sup_{\beta \in B} \|g(z_t, \beta)\|] < \infty$, **(v)** $\mathrm{E}[g(z_t, \beta)]$ *is continuous on* $\mathcal{B}$*;* **(vi)** $\mathrm{E}[g(z_t, \beta)] = 0$ *only for* $\beta = \beta_0$, **(vii)** $\mathcal{B}$ *is compact.* **(viii)** $\hat{W} = W + o_p(1)$ *and* W *is a positive semi-definite definite matrix.*

The following theorem corresponds to Theorem 3.1 of Hall (2005, p.68)

Theorem 2.1 *Under assumption 2.1* $\hat{\beta} = \beta_0 + o_p(1)$.

The assumptions 2.2 ensure that the estimator asymptotically normal distributed.[1]

Assumption 2.2 **(i)** $\{z_t, -\infty < t < \infty\}$ *is a strong mixing process with mixing coefficients of size* $-r/(r-2)$, $r > 2$, $\mathrm{E}[\|g(z_t, \beta_0)\|^r] < \infty$, $r \geq 2$; **(ii)** $G_t(\beta)$ *exists and is continuous on* $\mathcal{B}$ *for each* $z_t \in \mathcal{Z}$ **(iii)** $rank(G) = p$; **(iv)** $\mathrm{E}[\sup_{\beta \in \mathcal{N}} \|G_t(\beta)\|] < \infty$, *where* $\mathcal{N}$ *is a neighborhood of* β_0.

The following Theorem is proven in Hansen (1982, Theorem 3.1) or Hall (2005, p. 71).

Theorem 2.2 *Under assumption 2.1 and 2.2*

$$\sqrt{T}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, \mathrm{avar}(\hat{\beta})),$$

where $\mathrm{avar}(\hat{\beta}) = (G'WG)^{-1}G'W\Omega WG(G'WG)^{-1}$.

To obtain an efficient estimator we need to estimate Ω. Numerous estimators for Ω have been proposed in the literature under different assumptions [see White (1984), Newey and West (1987), Gallant (1987), Andrews (1991), Ng and Perron (1996).] Let $\hat{\Omega} = \Omega + o_p(1)$, the efficient two-step GMM estimator is defined as

$$\begin{aligned} \hat{\beta}^e &= \arg\min_{\beta \in B} \tilde{Q}(\beta), \\ \tilde{Q}_T(\beta) &= \hat{g}(\beta)'\hat{\Omega}^{-1}\hat{g}(\beta). \end{aligned}$$

[1]These assumptions are different from those stated in Hansen (1982), but facilitate comparisions with the assumptions made later in the paper for GEL and KBB.

Overidentification tests Consider the hypothesis $\mathrm{H}_0 : \mathrm{E}[g_t(\beta_0)] = 0$ vs $\mathrm{H}_1 : \mathrm{E}[g_t(\beta_0)] \neq 0$. Hansen (1982) proposed the J statistic to test this hypothesis which is defined as

$$\mathcal{J} = n\hat{g}(\hat{\beta}^e)'\hat{\Omega}^{-1}\hat{g}(\hat{\beta}^e),$$

where $\hat{\Omega}$ is a consistent estimator of Ω . Hansen (1982, Lemma 4.2) proved the following Theorem.

Theorem 2.3 *Under assumption 2.1 and 2.2 and if $m > p$, $\mathcal{J} \xrightarrow{d} \chi^2(m-p)$.*

Specification Tests Here we consider tests for the null hypothesis

$$\mathrm{H}_0 : a(\beta_0) = 0,\ \mathrm{E}[q(z_t, \beta_0)] = 0,$$

where $q(z_t, \beta_0)$ is a $s-$vector of moment indicators and $a(\beta)$ is a $r-$vector of constraints. The alternative H_1 is $a(\beta_0) \neq 0$ and/or $\mathrm{E}[q(z_t, \beta_0)] \neq 0$.

In the context of GMM, test statistics for parametric restrictions were proposed by Newey and West (1987) and for additional moment restrictions by Newey (1985), Eichebaum et al. (1988) and Ruud (2000) [see also Smith (1997) for tests based on GEL.]

In order to introduce these statistics define $h(z_t, \beta) \equiv (g(z_t, \beta)', q(z_t, \beta)')'$, $q_t(\beta) \equiv q(z_t, \beta)$, $h_t(\beta) \equiv h(z_t, \beta)$ $(t = 1, ..., T)$, $\hat{h}\,(\beta) \equiv \sum_{t=1}^{T} h_t(\beta)/T$, $\hat{q}\,(\beta) \equiv \sum_{t=1}^{T} q_t(\beta)/T$. Let also $\Xi \equiv \lim_{T\to\infty} \mathrm{var}[\sqrt{T}\hat{h}(\beta_0)]$, $\Xi_{12} \equiv \lim_{n\to\infty} \mathrm{E}[\sum_{i=1}^{n} g_t(\beta_0)q_t(\beta_0)'/\sqrt{T}]$ and $\Xi_{22} \equiv \lim_{n\to\infty} \mathrm{E}[\sqrt{n}\hat{q}(\beta_0)']$. Denote $\hat{\Xi}$ a consistent estimator of Ξ and let $\hat{\Xi}_{12}$ and $\hat{\Xi}_{22}$ be the submatrices of $\hat{\Xi}$ that consistently estimate Ξ_{12} and Ξ_{22} respectively. Let also

$$R(\beta) \equiv \begin{pmatrix} A(\beta) & 0_{r\times s} \\ 0_{s\times r} & I_s \end{pmatrix},$$

where $A(\beta) \equiv \partial a(\beta)/\partial\beta'$ (a $r \times p$ matrix). The restricted efficient GMM estimator is defined as

$$\begin{aligned} \hat{\beta}^e_r &= \arg\min_{\beta\in\mathcal{B}_r} \bar{Q}_T(\beta), \\ \bar{Q}_T(\beta) &= \hat{h}(\beta)'\hat{\Xi}^{-1}\hat{h}(\beta), \end{aligned}$$

where $\mathcal{B}_r = \{\beta \in \mathcal{B} : a\,(\beta) = 0\}$. Let $\hat{\gamma} \equiv \hat{q}(\hat{\beta}^e) - \hat{\Xi}_{21}\hat{\Xi}_{11}^{-1}g(\hat{\beta}^e)$, $\hat{r} \equiv (a(\hat{\beta}^e)', \hat{\gamma}')'$ and $\hat{R} \equiv R(\hat{\beta}^e)$. Define also $\hat{Q}_t(\beta) \equiv \partial q_t(\beta)/\partial\beta'$, $\hat{Q}\,(\beta) \equiv \sum_{i=1}^{T} \hat{Q}_t(\beta)/T$ and $Q \equiv \mathrm{E}[\partial q_t(\beta_0)/\partial\beta']$. Let $\Psi \equiv (D'\Xi^{-1}D)^{-1}$ and $\hat{\Psi} \equiv (\hat{D}'\hat{\Xi}^{-1}\hat{D})^{-1}$where

$$D = \begin{pmatrix} G & 0_{m\times s} \\ Q & -I_s \end{pmatrix},\ \hat{D}\,(\beta) = \begin{pmatrix} \hat{G}\,(\beta) & 0_{m\times s} \\ \hat{Q}\,(\beta) & -I_s \end{pmatrix},$$

and $\hat{D} = \hat{D}(\hat{\beta}^e)$.

We consider the following versions of the Wald, score and distance statistics

$$\begin{aligned}
\mathcal{W} &= \hat{r}'(\hat{R}\hat{\Psi}\hat{R}')^{-1}\hat{r}, \\
\mathcal{S} &= T\hat{h}(\hat{\beta}_r^e)'\hat{\Xi}^{-1}\hat{D}\hat{\Psi}\hat{D}'\hat{\Xi}^{-1}\hat{h}(\hat{\beta}_r^e), \\
\mathcal{D} &= T[\hat{h}(\hat{\beta}_r^e)'\hat{\Xi}^{-1}\hat{h}(\hat{\beta}_r^e) - \hat{g}(\hat{\beta}^e)'\hat{\Omega}^{-1}\hat{g}(\hat{\beta}^e)].
\end{aligned}$$

The results of Newey and West (1987), Newey (1985), Eichebaum et al. (1988) and Ruud (2000) are summarized in the following Theorem which is proven in the Appendix for completeness.

We require the following additional assumptions to hold

Assumption 2.3 **(i)** *β_0 is the unique solution of* $\mathrm{E}[h_t(\beta)] = 0$ *and* $a(\beta) = 0$; **(ii)** *$q(.,\beta)$ is Borel measurable for each $\beta \in \mathcal{B}$ and $q_t(\beta)$ is continuous in β for each $z_t \in \mathcal{Z}$* **(iii)** *$a(\beta)$ is twice continuously differentiable on $\mathcal{B}$*, **(iv)** $\mathrm{E}[\|q(z_t,\beta_0)\|^r] < \infty$, $r \geq 2$; **(v)** *$Q_t(\beta)$ exists and is continuous on $\mathcal{B}$ for each $z_t \in \mathcal{Z}$*; **(vi)** *$rank(Q) = s$*; **(vii)** $\mathrm{E}[\sup_{\beta\in\mathcal{N}} \|Q_t(\beta)\|] < \infty$; **(viii)** *$\Xi$ is non-singular and $\hat{\Xi} = \Xi + o_p(1)$.*

Theorem 2.4 unveils the asymptotic distribution of the trinity of the test statistics.

Theorem 2.4 *Under assumptions 2.1, 2.2 and 2.3 the statistics $\mathcal{W}$, $\mathcal{S}$ and $\mathcal{D}$ are asymptotically equivalent and converge in distribution to $\chi^2(s+r)$.*

2.1.2 Generalised Empirical Likelihood

In this section we review the efficient GEL estimator for time-series proposed by Smith (2011). Consider the smoothed moments

$$g_{tT}(\beta) = \frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1} k(\frac{s}{S_T})g_t(\beta), t = 1, ..., T,$$

where the kernel function $k(\cdot)$ satisfies $\int_{-\infty}^{+\infty} k(a)da = 1$, S_T is a bandwidth parameter. Define $k_2 \equiv \int_{-\infty}^{+\infty} k(a)^2 da$.

Let $\rho(\cdot)$ be a function that is concave on its domain $\mathcal{V}$, an open interval containing zero. It is convenient to impose a normalization on $\rho(\cdot)$. Let $\rho_j(\cdot) = \partial^j\rho(\cdot)/\partial v^j$ and $\rho_j = \rho_j(0)$, $(j = 0, 1, 2, ...)$. We normalize this function so that $\rho_1 = \rho_2 = -1$. The GEL criteria for weakly dependent data was defined by Smith (2011) as

$$\hat{P}_T(\beta, \lambda) = \sum\nolimits_{t=1}^{T}[\rho(k\lambda' g_{tT}(\beta)) - \rho_0]/T,$$

where $k = 1/k_2$. The GEL estimator is

$$\hat{\beta}_{\text{GEL}} = \arg\min_{\beta \in \mathcal{B}} \sup_{\lambda \in \Lambda_T} \hat{P}_T(\beta, \lambda),$$

where Λ_T is defined below in Assumption 2.8. Let $\hat{\lambda}(\beta) = \arg\sup_{\lambda \in \Lambda_T} \hat{P}_T(\beta, \lambda)$, $\hat{\lambda} \equiv \hat{\lambda}(\hat{\beta}_{\text{GEL}})$ and $G_{tT}(\beta) \equiv \partial g_{tT}(\beta)/\partial \beta'$.

Smith (2011) defined the implied probabilities as

$$\pi_t(\beta) = \frac{\rho_1(k\hat{\lambda}(\beta)' g_{tT}(\beta))}{\sum_{t=1}^{T} \rho_1(k\hat{\lambda}(\beta)' g_{tT}(\beta))}, t = 1, ..., T.$$

Smith (2011) required the following assumptions to hold.

Assumption 2.4 *The finite dimensional stochastic process $\{z_t\}_{t=1}^{\infty}$ is stationary and strong mixing with mixing coefficients α of size $-3v/(v-1)$ for some $v > 1$.*

Remark 2.1 *The mixing coefficient condition in Assumption 2.4 guarantees that $\sum_{j=1}^{\infty} j^2 \alpha(j)^{(v-1)/v} < \infty$ is satisfied, see Andrews (1991, p.824), a condition required for the results in Smith (2011).*

Assumption 2.5 **(i)** *$S_T \to \infty$, $S_T/T^{1/2} \to 0$;* **(ii)** *$k(.) : \mathbb{R} \to [-k_{\max}, k_{\max}].k_{\max} < \infty, k(0) \neq 0, k_1 \neq 0$ and is continuous at zero at almost everywhere;* **(iii)** *$\int_{-\infty}^{\infty} \bar{k}(x)dx < \infty$ where $\bar{k}(x) = I(x \geq 0)\sup_{y \geq x}|k(y)| + I(x < 0)\sup_{y \leq x}|k(y)|$;* **(iv)** *$|K(\lambda)| \geq 0$ for all $\lambda \in \mathbb{R}$, where $K(\lambda) = (2\pi)^{-1}\int k(x)\exp(-ix\lambda)dx$.*

Assumption 2.6 *$_T \to \infty$, $S_T = O(T^{1/2-\eta})$ for some $\eta \in (0, 1/2)$;*

Assumption 2.7 **(i)** *$\beta_0 \in \mathcal{B}$ is the unique solution of* $\text{E}[g_t(\beta)] = 0$; **(ii)** *$\mathcal{B}$ is compact*; **(iii)** *$g_t(\beta)$ is continuous at each $\beta \in B$*; **(iv)** $\text{E}[\sup_{\beta \in \mathcal{B}} \|g_t(\beta)\|^{\alpha}] < \infty$ *for some* $\alpha > \max(4v, 1/\eta)$; **(v)** $\Omega(\beta)$ *is finite and p.d. for all $\beta \in \mathcal{B}$.*

Assumption 2.8 **(i)** *$\rho(\cdot)$ is twice differentiable and concave on its domain an open interval $\mathcal{V}$ containing zero, $\rho_1 = \rho_2 = -1$*; **(ii)** *$\lambda \in \Lambda_T$, where $\Lambda_T = \left\{\lambda : \|\lambda\| \leq D(T/S_T^2)^{-\zeta}\right\}$ for some $D > 0$ with* $1/2 > \zeta > 1/(2\alpha\eta)$.

Theorem 2.5 is proven in Smith (2011).

Theorem 2.5 *If Assumptions 2.4, 2.6, 2.7 and 2.8 are satisfied $\hat{\beta} \xrightarrow{p} \beta_0$ and $\hat{\lambda} \xrightarrow{p} 0$. Moreover, $\left\|\hat{\lambda}\right\| = O_p[(T/S_T^2)^{-1/2}]$ and $\left\|\hat{g}_T(\hat{\beta})\right\| = O_p(T^{-1/2})$.*

Let $H \equiv \Sigma G'\Omega^{-1}$ and $P \equiv \Omega^{-1} - \Omega^{-1}G\Sigma G'\Omega^{-1}$. The proof of asymptotic normality of Smith (2011) also required the following assumptions.

Assumption 2.9 **(i)** $\beta_0 \in int\,(\mathcal{B})$; **(ii)** $g(\cdot, \beta)$ *is differentiable in a neighborhood* $\mathcal{N}$ *of* β_0 *and* $\mathrm{E}[\sup_{\beta\in\mathcal{N}} \|G_t\,(\beta)\|^{\alpha/(\alpha-1)}] < \infty$; **(iii)** $rank(G) = p$.

Smith (2011) proved the following theorem.

Theorem 2.6 *If Assumptions 2.4, 2.6, 2.7, 2.8 and 2.9 are satisfied*

$$\begin{pmatrix} T^{1/2}(\hat{\beta}_{\text{GEL}} - \beta_0) \\ T^{1/2}\hat{\lambda}/S_T \end{pmatrix} \xrightarrow{p} N(0, \mathrm{diag}(\Sigma, P)).$$

3 The kernel block bootstrap method

The idea behind the KBB method is to replace the original sample by a transformed sample and apply the `i.i.d.` bootstrap to the latter. To be more precise consider a sample of T observations, $(X_1, ..., X_T)$, on the zero mean finite dimensional stationary and strong mixing stochastic process $\{X_t\}_{t=1}^{\infty}$ with $\mathrm{E}[X_t] = 0$. Let $\bar{X} = \sum_{t=1}^{T} X_t/T$. Define the transformed variables

$$Y_{tT} = \frac{1}{S_T} \sum\nolimits_{s=t-T}^{t-1} k(\frac{s}{S_T})X_{t-s}, (t = 1, ..., T),$$

where S_T is a bandwidth parameter and $k(\cdot)$ is a kernel function standardized such that $\int_{-\infty}^{\infty} k(v)dv = 1$.

The standard KBB method consists in applying the non-parametric bootstrap for `i.i.d.` data using the transformed sample $(Y_{1T}, ..., Y_{TT})$ obtaining a bootstrap sample of size $m_T = T/S_T$, that is each bootstrap observation is drawn from $(Y_{1T}, ..., Y_{TT})$ with equal probability $1/T$. The asymptotic validity of the method was proven by Parente and Smith (2018a, 2018b).

In this article we modify the original method in that each observation is drawn with probability $\mathcal{P}[Y^*_{jT} = Y_{tT}] = p_{tT}$, for $j = 1, ..., m_T$ and $t = 1, ..., T$ where p_{tT} can depend on the data and satisfy $0 \le p_{tT} \le 1$ and $\sum_{t=1}^{T} p_{tT} = 1$. The standard KBB method of Parente and Smith (2018a, 2018b) is obtained with $p_{tT} = 1/T$ for for $j = 1, ..., m_T$ and $t = 1, ..., T$. Let $\tilde{Y} = \sum_{t=1}^{T} p_{tT}Y_{tT}$.

In order to prove that the bootstrap distribution of $\sqrt{T}(\bar{Y}^* - \tilde{Y})$ is close to the asymptotic distribution of $T^{1/2}\bar{X}$ as T goes to infinite, we required the following assumptions taken from Parente and Smith (2018a, 2018b).

Assumption 3.1 *The finite dimensional stochastic process* $\{X_t\}_{t=1}^{\infty}$ *is stationary and strong mixing with mixing coefficients* α *of size* $-3v/(v-1)$ *for some* $v > 1$*.*

Assumption 3.2 **(i)** $m_T = T/S_T$, $S_T \to \infty$, $S_T = O(T^{1/2-\eta})$ *for some* $\eta \in (0, 1/2)$*;* **(ii)** $\mathrm{E}[|X_t|^{\alpha}] < \infty$, *for some* $\alpha > \max(4v, 1/\eta)$*,* **(iii)** $\sigma^2_{\infty} \equiv \lim_{T\to\infty} \mathrm{var}[T^{1/2}\bar{X}]$ *is finite.*

Assumption 3.3 **(i)** $0 \le p_{tT} \le 1$, $\sum_{t=1}^{T} p_{tT} = 1$, $\max_{1\le t\le T} |Tp_{tT}| = o_p(1)$, **(ii)** $\sqrt{T}\tilde{Y} = O_p(1)$.

Similarly to Gonçalves and White (2004) $\mathcal{P}$ denotes the probability measure of the original time series and $\mathcal{P}^*$ that induced by the bootstrap method. For a bootstrap statistic θ^*_T we write $\theta^*_T \to 0$ prob-$\mathcal{P}^*$, prob-$\mathcal{P}$ if for any $\varepsilon > 0$ and any $\delta > 0$, $\lim_{T\to\infty} \mathcal{P}\{\mathcal{P}^*\{|\theta^*_T| > \varepsilon\} > \delta\} = 0$. We also use measures of magnitude of bootstrapped sequences as defined by Hanh (1997). Let $\xi^*_T = O^{\omega}_p(a_T)$ if ξ^*_T. when conditioned on ω is $O_p(a_T)$ and $\xi^*_T = o^{\omega}_p(a_T)$ if ξ^*_T. when conditioned on ω is $o_p(a_T)$. We write $\xi^*_T = O_B(1)$ if, for a given subsequence $\{T'\}$ there exists a further subsequence $\{T''\}$ such that $O^{\omega}_p(1)$. Similarly we write $\xi^*_T = o_B(1)$ if, for a given subsequence $\{T'\}$ there exists a further subsequence $\{T''\}$ such that $o^{\omega}_p(1)$.

The Theorem 3.1 shows that the bootstrap distribution of $\sqrt{T/k_2}(\bar{Y}^* - \tilde{Y})$ is uniformly close to the asymptotic distribution of $T^{1/2}\bar{X}$.

Theorem 3.1 *Under Assumptions 3.1-3.3 and 2.5, if* $\mathrm{E}[X_t] = 0$

$$\lim_{T\to\infty} \mathcal{P}\left\{\sup_{x\in\mathbb{R}} \left|\mathcal{P}^*\{\sqrt{T/k_2}(\bar{Y}^* - \tilde{Y}) \le x\} - \mathcal{P}\{T^{1/2}\bar{X} \le x\}\right| \ge \varepsilon\right\} = 0,$$

where $k_2 = \int_{-\infty}^{\infty} k^2(v)dv$*.*

The GEL- KBB method is obtained when $p_{tT} = \hat{\pi}_t$, where $\hat{\pi}_t = \pi_t(\hat{\beta}_{\mathrm{GEL}})$.

Lemma 3.1 *Assumption 3.3 is satisfied if* $p_{tT} = \hat{\pi}_t$*.*

4 Kernel block bootstrap methods for GMM

4.1 The standard KBB method

Consider a bootstrap sample of size m_T, $\{g^*_{tT}(\beta)\}_{t=1}^{m_T}$, drawn from $\{g_{tT}(\beta)\}_{t=1}^{T}$ and let $W^*_T = W_T + o_B(1)$, where W^*_T is positive semi-definite matrix. Define also $\hat{g}^*_T(\beta) = \sum_{s=1}^{m_T} g^*_{sT}(\beta)/m_T$ and

$$Q^*_T(\beta) = \hat{g}^*_T(\beta)' W^*_T \hat{g}^*_T(\beta).$$

To prove consistency we require the Assumption 4.1.

Assumption 4.1 **(i)** *The observed data are realizations of a stochastic process* $z \equiv \{z_t : \Omega \to \mathbb{R}^n, n \in \mathbb{N}, t = 1, 2, ...\}$ *on the complete probability space* $(\Omega, \mathcal{F}, P)$ *where* $\Omega = \times_{t=1}^{\infty} \mathbb{R}^n$ *and* $\mathcal{F} = \mathcal{B}(\times_{t=1}^{\infty} R^n)$ *(the Borel* $\sigma-$*field generated by the measure finite dimension product cylinders);* **(ii)** z_t *is stationary and ergodic ;* **(iii)** $g : \mathbb{R}^l \times B \to \mathbb{R}$ *is measurable for each* $\beta \in \mathcal{B}$*,* $\mathcal{B}$ *a compact subset of* $\mathbb{R}^p$*, and* $g(z_t, \cdot)$ *is continuous;* **(iv)** $\mathrm{E}[g(z_t, \beta)] = 0$ *only for* $\beta = \beta_0$, **(v)** $W_T = W + o_p(1)$ *and* Ω *is a positive definite matrix,* $W_T^* = W_T + o_B(1)$ **(vi)** $\mathrm{E}[\sup_{\beta \in \mathcal{B}} \|g(z_t, \beta)\|^\alpha] < \infty$ *for some* $\alpha \geq 1$*;* **(vii)** $T^{1/\alpha}/m_T = o(1)$*, where* $m_T \to \infty$.

Theorem 4.1 shows that the GMM bootstrap estimator is consistent.

Theorem 4.1 *Under assumption 4.1* $\hat{\beta}^* - \hat{\beta} \to 0$*, prob-*$\mathcal{P}^*$*, prob-*$\mathcal{P}$*.*

To prove the consistency of the bootstrap distribution of the GMM estimator we require assumption 4.2 to be satisfied.

Assumption 4.2 **(i)** *The* $(k \times 1)$ *random vectors* $\{z_t, -\infty < t < \infty\}$ *form a strictly stationary and mixing with mixing coefficients of size* $-3v/(v-1)$ *for some* $v > 1$; **(ii)** $\beta_0 \in int(\mathcal{B})$; **(iii)** $g(z_t.\beta)$ *is continuously differentiable in a neighborhood* $\mathcal{N}$ *of* β *with probability approaching one;* **(iv)** $\mathrm{E}(g(z, \beta_0)) = 0$ *and* $\mathrm{E}[\|g(z, \beta_0)\|^\alpha]$ *is finite for for some* $\alpha > \max(4v, 1/\eta)$; **(v)** $\mathrm{E}[\sup_{\beta \in \mathcal{N}} \|\partial g(z, \beta)/\partial \beta'\|^a] < \infty$ *for some* $a > 2/(1 + 2\eta)$; **(vi)** $G'WG$ *is nonsingular and* Ω *exists and is positive definite* **(vii)** $m_T = T/S_T$.

Theorem 4.2 demonstrates the consistency of the KBB distribution of the GMM estimator.

Theorem 4.2 *Under Assumptions 2.5, 4.1 and 4.2,*

$$\lim_{T \to \infty} \mathcal{P} \left\{ \sup_{x \in \mathbb{R}^k} \left| \mathcal{P}^* \{ \sqrt{\frac{T}{k_2}} (\hat{\beta}^* - \hat{\beta}) \leq x \} - \mathcal{P}\{T^{1/2}(\hat{\beta} - \beta_0) \leq x\} \right| \geq \varepsilon \right\} = 0.$$

4.1.1 Bootstrap Estimation of Ω

Hansen (1982) showed that the most efficient estimator is obtained if one sets $W = \Omega^{-1}$. We now show how to obtain consistent estimator for Ω using the bootstrap. Let $\hat{\Omega}^*(\tilde{\beta}^*) \equiv S_T \sum_{t=1}^{m_T} g_t^*(\tilde{\beta}^*) g_t^*(\tilde{\beta}^*)' / (m_T k_2)$ where $\tilde{\beta}^*$ is a bootstrap estimator of β_0 such that $\sqrt{T}(\tilde{\beta}^* - \beta_0) = O_B(1)$.

Assumption 4.3 is going to be required.

Assumption 4.3 $\mathrm{E}[\sup_{\beta\in\mathcal{N}}\|\partial g(z,\beta)/\partial\beta'\|^{2\alpha/(\alpha-1)}]<\infty$.

The desired result is given by Lemma 4.1

Lemma 4.1 *Under assumptions 2.5, 4.2* **(i)**, **(iii)**, **(iv)**, **(vi)**, **(vii)** *and 4.3 and if* $\sqrt{T}(\tilde{\beta}^{*}-\beta_0)=O_B(1)$ *we have*

$$\lim_{T\to\infty}\mathcal{P}[\mathcal{P}^{*}[\left|\hat{\Omega}^{*}(\tilde{\beta}^{*})-\Omega\right|>\varepsilon]>\delta]=0.$$

4.1.2 Testing for overidentifying restrictions

Let $\hat{\Omega}^{*}=\Omega+o_B(1)$, and let $\hat{\beta}^{e*}$ be the bootstrap GMM estimator obtained with $W_T^{*}=\hat{\Omega}^{*-1}$ and define

$$\mathcal{J}^{*}=\frac{T}{k_2}[\hat{g}^{*}(\hat{\beta}^{e*})-\hat{g}(\hat{\beta}^{e})]'\hat{\Omega}^{*-1}[\hat{g}^{*}(\hat{\beta}^{e*})-\hat{g}(\hat{\beta}^{e})].$$

The following Theorem proves the validity of the KBB- $\mathcal{J}$ test for overidentifying restrictions.

Theorem 4.3 *Under Assumptions 2.5, 4.1 and 4.2,*

$$\lim_{T\to\infty}\mathcal{P}\left\{\sup_{x\in\mathbb{R}}|\mathcal{P}^{*}\{\mathcal{J}^{*}\leq x\}-\mathcal{P}\{\mathcal{J}\leq x\}|\geq\varepsilon\right\}=0.$$

4.1.3 Bootstrap tests for parametric restrictions and additional moment conditions.

In this subsection we propose bootstrap versions of the tests for parametric restrictions and additional moment conditions. Let

$$h_{tT}(\beta)=\frac{1}{S_T}\sum_{s=t-T}^{t-1}k(\frac{s}{S_T})h_t(\beta),\ t=1,...,T$$

and consider a bootstrap sample of size m_T, $\{h_{tT}^{*}(\beta)\}_{t=1}^{m_T}$,drawn from $\{h_{tT}(\beta)\}_{t=1}^{T}$. Let $\tilde{\Omega}^{*}=\Omega+o_B(1)$ and $\hat{\Xi}_T=\Xi+o_B(1)$. Define also $\hat{h}_T^{*}(\beta)=\sum_{s=1}^{m_T}h_{sT}^{*}(\beta)/m_T$,

$$\bar{Q}_T^{*}(\beta)=\hat{h}_T^{*}(\beta)'\hat{\Xi}_T^{*-1}\hat{h}_T^{*}(\beta),$$

and

$$\hat{\beta}_r^{e*}=\arg\min_{\beta\in B_r}Q_{h,T}^{*}(\beta).$$

Let $\hat{\gamma}^{*}=\hat{q}^{*}(\hat{\beta}^{e*})-\hat{\Xi}_{21}^{*}\hat{\Xi}_{11}^{*-1}\hat{g}^{*}(\hat{\beta}^{e*})$, $r^{*}=((a(\hat{\beta}^{e*})',\hat{\gamma}^{*\prime})'$ and $\hat{R}^{*}=R(\hat{\beta}^{e*})$. Additionally, denote $\hat{Q}_t^{*}(\beta)\equiv\partial q_t^{*}(\beta)/\partial\beta'$, $\hat{Q}^{*}(\beta)\equiv\sum_{i=1}^{T}\hat{Q}_t^{*}(\beta)/T$, $\hat{\Psi}^{*}\equiv(\hat{D}^{*\prime}\hat{\Xi}^{*-1}\hat{D}^{*})^{-1}$, where

$$\hat{D}^{*}(\beta)=\begin{pmatrix}\hat{G}^{*}(\beta) & 0_{m\times s}\\ \hat{Q}^{*}(\beta) & -I_s\end{pmatrix},$$

and $\hat{D}^* = \hat{D}^*(\hat{\beta})$. We consider the following bootstrapped statistics

$$\begin{aligned}
\mathcal{W}^* &= (\frac{T}{k_2})[\hat{r}^* - \hat{r}]'[\hat{R}^*\hat{\Psi}^*\hat{R}^{*\prime}]^{-1}[\hat{r}^* - \hat{r}], \\
\mathcal{S}^* &= (\frac{T}{k_2})[\hat{h}^*(\hat{\beta}_r^{e*}) - \hat{h}(\hat{\beta}_r^e)]'\hat{\Xi}^{*-1}\hat{D}\hat{\Psi}^*\hat{D}^{*\prime}\hat{\Xi}^{*-1}[\hat{h}^*(\hat{\beta}_r^{e*}) - \hat{h}(\hat{\beta}_r^e)], \\
\mathcal{D}^* &= (\frac{T}{k_2})([\hat{h}^*(\hat{\beta}_r^{e*}) - \hat{h}(\hat{\beta}_r^e)]'\hat{\Xi}^{*-1}[\hat{h}^*(\hat{\beta}_r^{e*}) - \hat{h}(\hat{\beta}_r^e)] \\
&\quad -[\hat{g}^*(\hat{\beta}^{e*}) - \hat{g}(\hat{\beta}^e)]'\hat{\Omega}^{*-1}[\hat{g}^*(\hat{\beta}^{e*}) - \hat{g}(\hat{\beta}^e)]).
\end{aligned}$$

Hall and Horowitz (1996) considered t-statistics for tests on a single parameter for GMM using MBB and consequently these statistics seem to be new in the literature.

In order to show that the bootstrap distributions of these statistics are close to its asymptotic distributions the following assumptions are required.

Assumption 4.4 **(i)** *β_0 is the unique solution of* $\mathrm{E}[h_t(\beta)] = 0$ *and* $r(\beta) = 0$; $\mathrm{E}[\|h(z,\beta_0)\|^\alpha]$ *is finite;* **(ii)** *$q_t(\beta)$ is continuous in β for each $z_t \in \mathcal{Z}$;* **(iii)** *$r(\beta)$ is twice continuously differentiable on $\mathcal{B}$;* **(iv)** *$\partial q(z,\beta)/\partial\beta'$ exists and is continuous on $\mathcal{B}$ for each $z_t \in \mathcal{Z}$* **(v)** *$rank(Q) = s$;* **(vi)** $\mathrm{E}[\sup_{\beta\in\mathcal{N}} \|\partial q(z,\beta)/\partial\beta'\|^a] < \infty$; **(vii)** *$\Xi$ exists and is positive definite and $\hat{\Xi} = \Xi + o_p(1)$.*

Theorem 4.4 reveals that under Assumption 4.4 the bootstrapped trinity of test statistics is consistent to the asymptotic distributions of the statistics.

Theorem 4.4 *Under Assumptions 2.5, 4.1, 4.2,4.4*

$$\begin{aligned}
\lim_{T\to\infty} \mathcal{P}\left\{\sup_{x\in\mathbb{R}} |\mathcal{P}^*\{\mathcal{W}^* \le x\} - \mathcal{P}\{\mathcal{W} \le x\}| \ge \varepsilon\right\} &= 0, \\
\lim_{T\to\infty} \mathcal{P}\left\{\sup_{x\in\mathbb{R}} |\mathcal{P}^*\{\mathcal{S}^* \le x\} - \mathcal{P}\{\mathcal{S} \le x\}| \ge \varepsilon\right\} &= 0, \\
\lim_{T\to\infty} \mathcal{P}\left\{\sup_{x\in\mathbb{R}} |\mathcal{P}^*\{\mathcal{D}^* \le x\} - \mathcal{P}\{\mathcal{D} \le x\}| \ge \varepsilon\right\} &= 0.
\end{aligned}$$

Moreover, $\mathcal{W}^*$, $\mathcal{S}^*$ and $\mathcal{D}^*$ are asymptotically equivalent.

4.2 The generalised empirical likelihood kernel block bootstrap method

4.2.1 An efficient GMM estimator

In this sub-section we introduce a GMM-type estimator that is efficient and plays an important role in establishing the consistency of the kernel block bootstrap distribution to the asymptotic distribution of the GMM estimator. We consider the objective function

$$\tilde{Q}_T(\beta) = \tilde{g}(\beta)' W_T \tilde{g}(\beta),$$

where $\tilde{g}_T(\beta) = \sum_{t=1}^{T} g_{t,T}(\beta)\hat{\pi}_t$. Define the GMM-type estimator is defined as

$$\tilde{\beta} = \arg\min_{\beta \in B} \tilde{Q}_T(\beta).$$

where $W_T \xrightarrow{p} W$ and W is a positive semi-definite definite matrix.

We characterize now the asymptotic properties of the new estimator. Theorem 4.5 shows that this estimator is consistent for β_0.

Theorem 4.5 *Under Assumptions 2.4, 2.5, 2.6, 2.7 and 2.8* $\tilde{\beta} \xrightarrow{p} \beta_0$.

Theorem 4.6 reveals that $\tilde{\beta}$ is asymptotically equivalent to $\hat{\beta}^e$.

Theorem 4.6 *Under Assumptions 2.4, 2.5, 2.6,2.7, 2.8 and 2.9*

$$\sqrt{T}(\tilde{\beta} - \beta_0) - \sqrt{T}(\hat{\beta}^e - \beta_0) \xrightarrow{p} 0,$$
$$\sqrt{T}(\tilde{\beta} - \beta_0) \xrightarrow{D} N(0, \Sigma).$$

This theorem shows that no-matter the weighting matrix W_T we choose, we always obtain a estimator that is asymptotically equivalent to the efficient two-step GMM estimator.

4.2.2 The bootstrap method

Let $g^\star_{iT}(\beta)$,$i = 1, ..., m_T$ be obtained by drawing observations from $\{g_{tT}(\beta)\}_{t=1}^{T}$ where $\mathcal{P}(g^\star_{iT}(\beta) = g_{tT}(\beta)) = \hat{\pi}_t$, $t = 1, ..., T$. Denote $\hat{g}^\star_T(\beta) = \sum_{i=1}^{m_T} g^\star_{iT}(\beta)/m_T$. The generalised empirical likelihood kernel block bootstrap estimator (GEL-KBB) $\hat{\beta}^\star$ is defined as follows. Let

$$\hat{\beta}^\star = \arg\min_{\beta \in \mathcal{B}} \hat{g}^\star_T(\beta)' W^\star_T \hat{g}^\star_T(\beta)$$

where $W^\star_T = W_T + o_B(1)$.

Let $\mathcal{P}^\star$ be the bootstrap probability measure induced by the new resampling scheme.

Theorem 4.7 *Under Assumption 2.4, 2.5, 2.6, 2.7, 2.8 and 4.1* $\hat{\beta}^\star - \tilde{\beta} \to 0$ *prob-*$\mathcal{P}^\star$*, prob-*$\mathcal{P}$.

Assumption 4.5 $\mathrm{E}[\sup_{\beta \in \mathcal{N}} \|\partial g(z, \beta)/\partial \beta'\|^l] < \infty$ *for some* $l = \max\{\alpha/(\alpha - 1), 2/(1 + 2\eta) + \varepsilon\}$, *for some* $\varepsilon > 0$.

The following result shows consistency of the bootstrap estimator to the asymptotic distribution of $\hat{\beta}$.

Theorem 4.8 *Under Assumptions 2.4, 2.5, 2.6, 2.8 2.9, 4.2 strengthen by 4.5 ,*

$$\begin{aligned}\lim_{T\to\infty}\mathcal{P}\left\{\sup_{x\in\mathbb{R}^p}\left|\mathcal{P}^\star\{\sqrt{\frac{T}{k_2}}(\hat{\beta}^\star-\tilde{\beta})\leq x\}-\mathcal{P}\{T^{1/2}(\hat{\beta}-\beta_0)\leq x\}\right|\geq\varepsilon\right\} &= 0,\\ \lim_{T\to\infty}\mathcal{P}\left\{\sup_{x\in\mathbb{R}^p}\left|\mathcal{P}^\star\{\sqrt{\frac{T}{k_2}}(\hat{\beta}^\star-\hat{\beta}^e)\leq x\}-\mathcal{P}\{T^{1/2}(\hat{\beta}-\beta_0)\leq x\}\right|\geq\varepsilon\right\} &= 0.\end{aligned}$$

We note that $\hat{\beta}^\star$ is centered at the efficient estimator $\tilde{\beta}$, not on the inefficient $\hat{\beta}$, though the bootstrap distribution of $\sqrt{T/k_2}(\hat{\beta}^\star-\tilde{\beta})$ approximates the asymptotic distribution of the inefficient estimator $T^{1/2}(\hat{\beta}-\beta_0)$. This result is not specific of the GEL-KBB method, it also holds for the empirical likelihood moving blocks bootstrap of Allen et al. (2011) contradicting Theorems 1 and 2 of that article. Both estimators only coincide if $W=\Omega^{-1}$.

4.2.3 GEL-KBB Estimation of Ω

Let $\bar{\beta}^\star$ be a bootstrap estimator such that $\sqrt{T}(\bar{\beta}^\star-\beta_0)=O_B(1)$. We now prove consistency of the bootstrap estimator of Ω under the GEL-KBB measure, which is given by

$$\hat{\Omega}^\star(\bar{\beta}^\star)\equiv\frac{S_T}{m_Tk_2}\sum\nolimits_{t=1}^{m_T}g_t^\star(\bar{\beta}^\star)g_t^\star(\bar{\beta}^\star)'.$$

The consistency of $\hat{\Omega}^\star(\bar{\beta}^\star)$ is proven in Lemma 4.2.

Lemma 4.2 *Under Assumptions 2.4, 2.5, 2.6, 2.8 2.9, 4.2 strengthen by 4.3 if $\sqrt{T}(\bar{\beta}^\star-\beta_0)=O_B(1)$ we have*

$$\lim_{T\to\infty}\mathcal{P}[\mathcal{P}^\star[\left|\hat{\Omega}^\star(\bar{\beta}^\star)-\Omega\right|>\varepsilon]>\delta]=0.$$

4.2.4 Testing for overidentifying restrictions

Let $W_T^\star=\tilde{\Omega}^{\star-1}$ where $\tilde{\Omega}^\star=\Omega+o_B(1)$ and define $\hat{\beta}^{e\star}$ as the bootstrap GMM estimator computed with $W_T^\star=\hat{\Omega}^{\star-1}$. corresponds to the efficient estimator and let

$$\mathcal{J}^\star=\frac{T}{k_2}\hat{g}^\star(\hat{\beta}^{e\star})'\tilde{\Omega}^{\star-1}\hat{g}^\star(\hat{\beta}^{e\star}).$$

Theorem 4.9 *Under Assumptions 2.4, 2.5, 2.6, 2.8 2.9, 4.2 strengthen by 4.5*

$$\lim_{T\to\infty}\mathcal{P}\left\{\sup_{x\in\mathbb{R}}|\mathcal{P}^\star\{\mathcal{J}^\star\leq x\}-\mathcal{P}\{\mathcal{J}\leq x\}|\geq\varepsilon\right\}=0.$$

4.2.5 GEL-KBB tests for parametric restrictions and additional moment conditions under the maintained hypothesis

In this subsection we propose bootstrap versions of the tests for parametric restrictions and additional moment conditions. Consider a bootstrap sample of size m_T, $\{h^{\star}_{tT}(\beta)\}_{t=1}^{m_T}$, drawn from $\{h_{tT}(\beta)\}_{t=1}^{T}$ where $\mathcal{P}(h^{\star}_{jT}(\beta) = h_{tT}(\beta)) = \hat{\pi}_t$, $t = 1, ..., T$ and $j = 1, ..., m_T$. Let also $\hat{\Xi}^{\star} = \Xi + o_B(1)$, $\hat{h}^{\star}(\beta) = \sum_{s=1}^{m_T} h^{\star}_{sT}(\beta)/m_T$, $\tilde{h}_T(\beta) = \sum_{t=1}^{T} h_{t,T}(\beta)\hat{\pi}_t$ and $\tilde{q}(\beta) = \sum_{t=1}^{T} q_{t,T}(\beta)\hat{\pi}_t$. Consider the objective function $\bar{Q}^{\star}_T(\beta) = \hat{h}^{\star}(\beta)'\hat{\Xi}^{\star-1}\hat{h}^{\star}(\beta)$ and let

$$\hat{\beta}^{e\star}_r = \underset{\beta\in\mathcal{B}_r}{\arg\min}\bar{Q}^{\star}_T(\beta).$$

Define $\hat{\gamma}^{\star} = \hat{q}^{\star}(\hat{\beta}^{e\star}) - \hat{\Xi}^{\star}_{21}\hat{\Xi}^{\star-1}_{11}\hat{g}^{\star}(\hat{\beta}^{e\star})$, $\hat{r}^{\star} = ((a(\hat{\beta}^{e\star})', \hat{\gamma}^{\star\prime})'$, $\tilde{\gamma} = \tilde{q}(\hat{\beta}^{e})$, $\tilde{r} = ((a(\hat{\beta}^{e})', \tilde{\gamma}')'$, $\hat{R}^{\star} = R(\hat{\beta}^{e\star})$. Additionally, let $\hat{Q}^{\star}_t(\beta) \equiv \partial q^{\star}_t(\beta)/\partial\beta'$ and $\hat{Q}^{\star}(\beta) \equiv \sum_{i=1}^{T}\hat{Q}^{\star}_t(\beta)/T$. Denote also $\hat{\Psi}^{\star} \equiv (\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^{\star})^{-1}$where

$$\hat{D}^{\star}(\beta) = \begin{pmatrix} \hat{G}^{\star}(\beta) & 0_{m\times s} \\ \hat{Q}^{\star}(\beta) & -I_s \end{pmatrix},$$

and $\hat{D}^{\star} = \hat{D}^{\star}(\hat{\beta})$.We consider the following bootstrapped statistics

$$\begin{aligned}
\mathcal{W}^{\star} &= (\frac{T}{k_2})[\hat{r}^{\star} - \tilde{r}]'[\hat{R}^{\star}\hat{\Psi}^{\star}\hat{R}^{\star\prime}]^{-1}[\hat{r}^{\star} - \tilde{r}], \\
\mathcal{S}^{\star} &= (\frac{T}{k_2})[\hat{h}^{\star}(\hat{\beta}^{e\star}_r) - \tilde{h}(\hat{\beta}^{e}_r)]'\hat{\Xi}^{\star-1}\hat{D}^{\star}\hat{\Psi}^{\star}\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}[\hat{h}^{\star}(\hat{\beta}^{e\star}_r) - \tilde{h}(\hat{\beta}^{e}_r)], \\
\mathcal{D}^{\star} &= (\frac{T}{k_2})([\hat{h}^{\star}(\hat{\beta}^{e\star}_r) - \tilde{h}(\hat{\beta}^{e}_r)]'\hat{\Xi}^{\star-1}[\hat{h}^{\star}(\hat{\beta}^{e\star}_r) - \tilde{h}(\hat{\beta}^{e}_r)] - \hat{g}^{\star}(\hat{\beta}^{e\star})'\hat{\Omega}^{\star-1}\hat{g}^{\star}(\hat{\beta}^{e\star})).
\end{aligned}$$

The Wald statistic can be seen as a generalization of the bootstrapped Wald statistic of Allen at al. (2011) and Bravo and Crudu (2011) for parametric restrictions. The remaining statistics seem to be new in the bootstrap literature.

Theorem 4.10 proves consistency of the bootstrap distribution of the trinity of test statistics.

Theorem 4.10 *Under Assumptions Under Assumptions 2.4, 2.5, 2.6, 2.8 2.9, 4.2 strengthen by 4.5 and 4.4*

$$\begin{aligned}
\lim_{T\to\infty}\mathcal{P}\left\{\sup_{x\in\mathbb{R}}|\mathcal{P}^{\star}\{\mathcal{W}^{\star}\leq x\} - \mathcal{P}\{\mathcal{W}\leq x\}| \geq \varepsilon\right\} &= 0, \\
\lim_{T\to\infty}\mathcal{P}\left\{\sup_{x\in\mathbb{R}}|\mathcal{P}^{\star}\{\mathcal{S}^{\star}\leq x\} - \mathcal{P}\{\mathcal{S}\leq x\}| \geq \varepsilon\right\} &= 0, \\
\lim_{T\to\infty}\mathcal{P}\left\{\sup_{x\in\mathbb{R}}|\mathcal{P}^{\star}\{\mathcal{D}^{\star}\leq x\} - \mathcal{P}\{\mathcal{D}\leq x\}| \geq \varepsilon\right\} &= 0.
\end{aligned}$$

Moreover, $\mathcal{W}^{\star}$, $\mathcal{S}^{\star}$ *and* $\mathcal{D}^{\star}$ *are asymptotically equivalent.*

4.2.6 GEL-KBB tests for parametric restrictions and additional moment conditions under the null hypothesis

In this subsection we propose kernel block bootstrap versions of the tests for parametric restrictions and additional moment conditions that impose the null hypothesis through the generalised empirical likelihood implied probabilities similar to the method proposed by Bravo and Crudu (2011).

Before introducing the method we need to introduce the GEL criteria for weakly dependent data for additional moments which is given by

$$\bar{P}_T(\beta,\varphi)=\sum\nolimits_{t=1}^{T}[\rho(k\varphi' h_{tT}(\beta))-\rho_0]/T,$$

where $k=1/k_2$. The GEL estimator is defined as

$$\hat{\beta}_{r,\text{GEL}}=\arg\min_{\beta\in\mathcal{B}_r}\sup_{\varphi\in\Delta_T}\bar{P}_T(\beta,\varphi)$$

where Δ_T is defined below in Assumption 4.7 define also $\hat{\varphi}(\beta)=\arg\sup_{\varphi\in\Delta_T}\hat{P}_T(\beta,\varphi)$, $\hat{\varphi}_r\equiv\hat{\varphi}(\hat{\beta}_{r,\text{GEL}})$.

Consider a bootstrap sample of size m_T ,$\left\{h^\dagger_{tT}(\beta)\right\}_{t=1}^{m_T}$, drawn from $\{h_{tT}(\beta)\}_{t=1}^{T}$ where $\mathcal{P}(h^\dagger_{jT}(\beta)=h_{tT}(\beta))=\tilde{\pi}_t$, $t=1,...,T$ and $j=1,...,m_T$ where

$$\tilde{\pi}_t=\frac{\rho_1(\hat{\varphi}_r' h_{tT}(\hat{\beta}_{r,\text{GEL}}))}{\sum_{j=1}^{T}\rho_1(\hat{\varphi}_r' h_{jT}(\hat{\beta}_{r,\text{GEL}}))},t=1,...,T.$$

We consider the case that the bootstrap weighting matrix is $W^\dagger_T=\hat{\Xi}^{\dagger-1}$, where $\hat{\Xi}^\dagger=\Xi+o_B(1)$. Define $\hat{h}^\dagger_T(\beta)\equiv\frac{1}{m_T}\sum_{s=1}^{m_T}h^\dagger_{sT}(\beta)$, $\bar{Q}^\dagger_T(\beta)=\hat{h}^\dagger_T(\beta)'\hat{\Xi}^{\dagger-1}\hat{h}^\dagger_T(\beta)$and let

$$\hat{\beta}^{e\dagger}=\arg\min_{\beta\in\mathcal{B}}\bar{Q}^\dagger_T(\beta),\ \ \hat{\beta}^{e\dagger}_r=\arg\min_{\beta\in\mathcal{B}_r}\bar{Q}^\dagger_T(\beta).$$

Define $\hat{\gamma}^\dagger=\hat{q}^\dagger(\hat{\beta}^{e\dagger})-\hat{\Xi}^\dagger_{21}\hat{\Xi}^{\dagger-1}_{11}\hat{g}^\dagger(\hat{\beta}^{e\dagger})$, $\hat{r}^\dagger=((a(\hat{\beta}^{e\dagger})',\hat{\gamma}^{\dagger\prime})'$ and $\hat{R}^\dagger=R(\hat{\beta}^\dagger)$. Additionally, let us define $\hat{Q}^\dagger_t(\beta)\equiv\partial q^\dagger_t(\beta)/\partial\beta'$, $\hat{Q}^\dagger(\beta)\equiv\sum_{i=1}^{T}\hat{Q}^\dagger_t(\beta)/T$. Denote also $\hat{\Psi}^\dagger\equiv(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^\dagger)^{-1}$where

$$\hat{D}^\dagger(\beta)=\begin{pmatrix}\hat{G}^\dagger(\beta) & 0_{m\times s}\\ \hat{Q}^\dagger(\beta) & -I_s\end{pmatrix},$$

We consider the following bootstrapped statistics

$$\begin{aligned}
\mathcal{W}^\dagger &= (\frac{T}{k_2})\hat{r}^{\dagger\prime}[\hat{R}^\dagger\hat{\Psi}^\dagger\hat{R}^{\dagger\prime}]^{-1}\hat{r}^\dagger,\\
\mathcal{S}^\dagger &= (\frac{T}{k_2})\hat{h}^\dagger(\hat{\beta}^{e\dagger}_r)'\hat{\Xi}^{\dagger-1}\hat{D}^\dagger\hat{\Psi}^\dagger\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{h}^\dagger(\hat{\beta}^{e\dagger}_r),\\
\mathcal{D}^\dagger &= (\frac{T}{k_2})[\hat{h}^\dagger(\hat{\beta}^{e\dagger}_r)'\hat{\Xi}^{\dagger-1}\hat{h}^\dagger(\hat{\beta}^{e\dagger}_r)-\hat{g}^\dagger(\hat{\beta}^{e\dagger})'\tilde{\Omega}^{\dagger-1}\hat{g}^\dagger(\hat{\beta}^{e\dagger})].
\end{aligned}$$

where $\tilde{\Omega}^{\dagger} = \Omega + o_B(1)$.

Versions of the statistics $\mathcal{S}^{\dagger}$ and $\mathcal{D}^{\dagger}$ for moving blocks bootstrap and parametric restrictions were introduced previously by Bravo and Crudu (2011). The statistic $\mathcal{W}^{\dagger}$ is new.

In order to show that the bootstrap distributions of these statistics are close to its asymptotic distributions the following assumptions are required.

Assumption 4.6 **(i)** *$\beta_0 \in \mathcal{B}$ is the unique solution of* $\mathrm{E}[h_t(\beta)] = 0$; **(ii)** *$\mathcal{B}$ is compact*; **(iii)** *$h_t(\beta)$ is continuous at each $\beta \in B$*; **(iv)** $\mathrm{E}[\sup_{\beta\in\mathcal{B}} \|h_t(\beta)\|^{\alpha}] < \infty$ *for some* $\alpha > \max(4v, 1/\eta)$; **(v)** *$\Xi(\beta)$ is finite and p.d. for all $\beta \in B$.*

Assumption 4.7 *$\varphi \in \Delta_T$, where $\Delta_T = \left\{\varphi : \|\varphi\| \leq D(T/S_T^2)^{-\zeta}\right\}$, for some $D > 0$ with $1/2 > \zeta > 1/(2\alpha\eta)$.*

Assumption 4.8 **(i)** *$\beta_0 \in int(\mathcal{B})$*; **(ii)** *$h(\bullet, \beta)$ is differentiable in a neighborhood $\mathcal{N}$ of β_0 and* $\mathrm{E}[\sup_{\beta\in\mathcal{N}} \|H_t(\beta)\|^{l}] < \infty$ *where* $l = \max\{\alpha/(\alpha-1), 2/(1+2\eta)+\varepsilon\}$; **(iii)** *$rank(H) = p + q$.*

Theorem 4.11 demonstrates that the bootstrapped Wald, score and distance statistics are asymptotically valid.

Theorem 4.11 *Under Assumptions 2.5, 4.6, 4.7, 4.8, 4.2, 4.4*

$$\begin{aligned}
\lim_{T\to\infty} \mathcal{P}\left\{\sup_{x\in\mathbb{R}} \left|\mathcal{P}^{\dagger}\{\mathcal{W}^{\dagger} \leq x\} - \mathcal{P}\{\mathcal{W} \leq x\}\right| \geq \varepsilon\right\} &= 0, \\
\lim_{T\to\infty} \mathcal{P}\left\{\sup_{x\in\mathbb{R}} \left|\mathcal{P}^{\dagger}\{\mathcal{S}^{\dagger} \leq x\} - \mathcal{P}\{\mathcal{S} \leq x\}\right| \geq \varepsilon\right\} &= 0, \\
\lim_{T\to\infty} \mathcal{P}\left\{\sup_{x\in\mathbb{R}} \left|\mathcal{P}^{\dagger}\{\mathcal{D}^{\dagger} \leq x\} - \mathcal{P}\{\mathcal{D} \leq x\}\right| \geq \varepsilon\right\} &= 0.
\end{aligned}$$

Moreover, $\mathcal{W}^{\dagger}$, $\mathcal{S}^{\dagger}$ and $\mathcal{D}^{\dagger}$ are asymptotically equivalent.

5 Monte Carlo Study

In this section we present a simulation study in which we investigate the small sample properties of the proposed bootstrap methods. The model used in our study is a version of an asset-pricing model considered in the Monte Carlo study of Hall and Horowitz (1996). The moment restrictions of this model are

$$E\{\exp[\mu_s - \theta_0(x+z) + 3z] - 1\} = 0,$$
$$E\{z\exp[\mu_s - \theta_0(x+z) + 3z] - z\} = 0$$

where $\theta_0 = 3$, $\mu_s = -9s^2/2$, and x and z are scalars. The random variable x has distribution normal with mean zero and variance s^2, with $s = 0.2$ or 0.4. z is independent of x, has a marginal distribution normal with zero mean and variance s^2, and is either sampled independently from this distribution or follows an AR(1) process with first-order serial correlation coefficient $\rho_z = 0.75$.

We evaluate the performance of Hansen (1982)'s $\mathcal{J}$ test and the symmetrical t tests for the null hypothesis $H_0 : \theta_0 = 3$ with asymptotic and bootstrap critical values. The $\mathcal{J}$ statistic is computed using the two-step GMM estimator in which the weighting matrix used in the first step is the identity matrix. In the second step the long-run variance of the moment indicators is computed using the Newey-West estimator (Newey and West, 1987).[2]

We obtain the bootstrap critical values for the $\mathcal{J}$-tests and t-tests using the standard moving blocks bootstrap, the kernel blocks bootstrap (KBB) based on different kernel functions and the versions of these methods based on the Empirical Likelihood (EL) implied probabilities. KBB is computed using the truncated kernel (KKB_{TR}), the Bartlett Kernel (KKB_{BT}), the kernel that induces the quadratic-spectral kernel (KKB_{QS}) [see Smith (2011)] and the kernel version of the optimal taper of Paparoditis and Politis (2001) (KKB_{PP}). The EL implied probabilities are computed imposing the moment restrictions in the sample. In the tables of results we use the superscript EL to denote the results obtained with the bootrap method based on the implied probabilities. Although the methods were computed for the case that there is dependence in the data, we also apply the same method in the case that there is no dependence.[3]

In order to investigate whether the methods proposed are sensitive to the choice of the band-

[2]We also computed a two step GMM estimator in which the long-run variance of the moment indicators is estimated using the Andrews (1991) estimator based on the Quadratic Spectral kernel. These results are available upon request. Additionally, we investigated the performance of the tests based on the $\mathcal{J}$-statistic in which the long run variance of the moment indicators was estimated using the approach of Andrews and Monahan (1992) which requires pre-whitened series. The results obtained were not satisfactory in the Monte Carlo design considered and consequently are not presented.

[3]The quasi-Newton algorithm of MATLAB is used to compute GMM and EL hence ensuring a local optimum. The Newton method is used to locate $\hat{\lambda}(\beta)$ for given β which is required for the profile EL objective function. EL computation requires some care since the EL criterion involves the logarithm function which is undefined for negative arguments; this difficulty is avoided by employing the approach due to Owen in which logarithms are replaced by a function that is logarithmic for arguments larger than a small positive constant and quadratic below that threshold. See Owen (2001, (12.3), p.235); Note however, that this method might produce estimates that lie outside the convex hull of the data. In our study the worst case in which this problem occurred affected 1% of the replications and corresponded to the case $n = 50$, $s = 0.2$ and the truncated kernel was used. In all the remaining designs the problem only occurred in less or equal than 0.6% of the replications. Hence our results are not considerably affected by this issue.

width/block size we compute these parameters using two methods: the automatic bandwidth of Andrews (1991) based on an AR(1) model and a non-parametric version of the Andrews (1991) method based on a taper proposed by Romano and Politis (1995). These methods to compute the bandwidth were applied to the residuals obtained in the first step of the GMM problem [see Parente and Smith (2018b), section 4.3, for details]. Additionally, given that the computed automatic bandwidth $\hat{S}_T$ might induce values of $m_T = \lceil T/S_T \rceil$ larger than T or equal to 1, where $\lceil \cdot \rceil$ is the ceiling function, we replace $\hat{S}_T$ by $\hat{S}_T^* = \max\{S_T, 1\}$ and m_T by $m_T^* = \max\left\{\left\lceil T/\hat{S}_T^* \right\rceil, 2\right\}$. Consequently we have $2 \leq m_T^* \leq T$.

We can find in the literature different bootstrap symmetric t-tests. Hall (1992, see sections 3.5, 3.6 and 3.12) considers the two-sided symmetric percentile t-test, the two sided equal-tailed t-test. Here we report only the results on the former method because it provided the best results in our study. Additionally, because our objective is to compare the performance of several different bootstrap tests we present, for succinctness, only the results on computed using the 5% nominal level.[4]

Table 1 reports the empirical rejection rates of the Hansen (1982)'s $\mathscr{J}$ test. The results obtained reveal that the $\mathscr{J}$ test based on asymptotic critical values are slightly undersized for $s = 0.2$ and they become to some extent oversized for $s = 0.4$. Note that in the latter case the rejection frequencies do not get closer to the nominal size when the sample size increases from 50 to 100.[5] The tests based on standard KBB and MBB critical values are considerably undersized. The tests based on the empirical likelihood versions of the bootstrap methods although are undersized for $s = 0.2$, yield empirical rejection rates closer to the nominal size for $s = 0.4$.

Table 2 presents the results on the t-tests for the hypothesis $H_0 : \theta_0 = 3$. The empirical rejection rates of the t-tests based on the asymptotic critical values are considerably larger than the nominal rate. On the other hand, the performance of the t-tests based on the critical values obtained with MBB and KBB are noticeably better than those based on the asymptotic critical values. However, the t-tests based on the taper of Paparoditis and Politis (2000) are undersized. The empirical-likelihood versions of these t-tests, in general are slightly oversized, apart from the case in which the kernel version of the taper of Paparoditis and Politis (2001).

Overall the results obtained with both methods to compute the automatic bandwidth are very similar

[4]The results on 1% and 10% nominal level were also computed and are available upon request.

[5]Note that these results are different to those reported by Hall and Horowitz (1995), specially in the case $s = 0.4$, though they computed the GMM estimator using a different weighting matrix.

Table 1: Empirical rejection rates of the J-tests with asymptotic and bootstrap critical values at 5% level

	Politis and Romano								Andrews							
n	50				100				50				100			
ρ_z	0		0.75		0		0.75		0		0.75		0		0.75	
s	0.2	0.4	0.2	0.4	0.2	0.4	0.2	0.4	0.2	0.4	0.2	0.4	0.2	0.4	0.2	0.4
ASYMP	2.4	7.9	3.8	5.6	3.1	9.2	2.9	7.5	3.3	8.0	3.5	7.0	3.6	8.8	4.5	7.7
KBB_{TR}	0.3	1.3	0.6	1.5	0.3	2.4	0.7	2.6	0.5	2.1	0.6	1.9	0.7	2.8	1.1	2.1
$\text{KBB}_{\text{TR}}^{\text{EL}}$	1.2	5.2	1.9	3.1	2.4	7.1	2.4	5.9	1.6	5.0	1.6	3.8	2.8	7.2	3.4	5.7
KBB_{BT}	0.1	0.4	0.1	0.6	0.5	0.8	0.6	1.1	0.3	0.5	0.3	0.8	0.7	0.9	0.8	1.2
$\text{KBB}_{\text{BT}}^{\text{EL}}$	0.8	4.4	1.3	2.9	1.5	6.4	1.7	5.2	0.7	4.6	1.3	3.7	2.2	6.9	2.5	4.7
KBB_{PP}	0.0	0.0	0.1	0.0	0.2	0.3	0.5	0.6	0.2	0.2	0.2	0.4	0.6	0.4	0.7	0.7
$\text{KBB}_{\text{PP}}^{\text{EL}}$	0.5	3.6	1.8	2.6	1.0	5.1	1.5	4.1	0.6	3.6	1.5	3.3	1.3	5.4	2.1	3.6
KBB_{QS}	0.1	0.2	0.1	0.3	0.3	0.8	0.7	0.9	0.3	0.1	0.2	0.8	0.7	0.6	0.8	1.2
$\text{KBB}_{\text{QS}}^{\text{EL}}$	0.6	3.3	1.5	2.0	1.5	6.3	1.7	4.6	1.0	3.9	1.3	2.9	1.9	6.1	2.7	3.9
MBB_{BT}	0.2	0.1	0.2	0.1	0.3	0.5	0.5	0.7	0.4	0.2	0.3	0.6	0.8	0.6	0.7	0.9
MBB^{EL}	0.6	4.0	1.3	2.4	1.2	6.2	1.6	4.5	0.5	3.9	1.1	3.3	1.6	6.0	2.3	4.0

which may indicate that the proposed methods are robust to the choice of this parameter.

6 Conclusion

In this article we put forward new bootstrap methods for models defined through moment restrictions for time series data that build on the kernel block bootstrap method of Parente and Smith (2018a, 2018b). These methods approximate the asymptotic distributions of tests for overidentifying conditions, parametric restrictions and additional moment restrictions. We consider methods that impose the null hypothesis, methods that impose the maintained hypothesis and methods that do not impose any restriction in the way the bootstrap samples are generated. We prove the first-order validity of the methods generalizing and correcting the work of Allen et al. (2011) and Bravo and Crudu (2011). A simulation study reveals that the proposed methods perform well in practice.

Appendix: Proofs

Throughout the Appendix, C and Δ will denote generic positive constants that may be different in different uses, and C, M, and T the Chebyshev, Markov, and triangle inequalities respectively. We use the same notation of Gonçalves and White (2004). For a bootstrap statistic $W_T^*(.,\omega)$ we write $W_T^*(.,\omega) \to 0$ $prob-\mathcal{P}^*$, $prob-\mathcal{P}$ if for any $\varepsilon > 0$ and any $\delta > 0$, $\lim_{T\to\infty} \mathcal{P}[\mathcal{P}_{T,\omega}^*[|W_T^*(\lambda,\omega)| > \varepsilon] > \delta] = 0$.

A.1 Proofs of the results in subsection 2.1.1

Proof of Theorem 2.4: As Tauchen (1985) and Ruud (2000) we recast the test for H_0 as a test for parametric restrictions $q_t^a(\beta,\gamma) \equiv q_t(\beta) - \gamma$ and construct the moment indicators $h_t^a(\beta,\gamma) \equiv (g_t(\beta)', q_t^a(\beta,\gamma)')$. Under the null hypothesis $\gamma = 0$, $a(\beta_0) = 0$ thus we have the model $\text{E}(h_t^a(\beta_0, 0)) = 0$ and $a(\beta_0) = 0$. Define $\theta = (\beta', \gamma')'$ and $\hat{h}^a(\theta) = \sum_{t=1}^T h_t^a(\beta,\gamma)/T$.

Table 2: Empirical rejection rates of the t-tests with asymptotic and bootstrap critical values at 5% level

	Politis and Romano								Andrews							
n	50				100				50				100			
ρ_z	0		0.75		0		0.75		0		0.75		0		0.75	
s	0.2	0.4	0.2	0.4	0.2	0.4	0.2	0.4	0.2	0.4	0.2	0.4	0.2	0.4	0.2	0.4
ASYMP	24.3	25.8	22.2	25.7	18.3	20.4	19.5	19.9	22.3	26.0	20.0	25.9	18.9	20.1	17.4	20.0
KBB_{TR}	4.4	6.6	6.0	7.9	4.6	6.1	5.9	6.6	4.4	7.1	4.7	6.6	5.5	6.3	5.8	6.9
$\text{KBB}_{\text{TR}}^{\text{EL}}$	6.9	8.7	7.5	9.3	6.6	7.8	7.3	7.8	6.5	8.4	6.2	8.2	7.1	6.9	6.8	7.5
KBB_{BT}	4.1	4.4	4.8	6.5	3.6	3.9	5.1	4.5	3.9	4.2	4.1	5.2	3.8	3.8	4.3	5.2
$\text{KBB}_{\text{BT}}^{\text{EL}}$	6.0	6.4	6.8	8.5	5.5	5.7	6.9	6.0	6.5	5.3	5.7	6.9	7.6	4.9	6.2	5.8
KBB_{PP}	2.9	2.5	3.4	4.1	2.8	2.5	3.3	3.2	2.9	2.9	2.9	4.0	3.0	2.3	3.0	3.1
$\text{KBB}_{\text{PP}}^{\text{EL}}$	4.4	4.0	4.5	5.5	4.0	4.3	5.2	4.3	4.4	3.1	3.7	5.1	6.0	3.1	4.6	3.7
KBB_{QS}	4.4	4.5	5.1	6.4	3.6	4.3	5.2	4.8	4.0	4.8	4.3	5.4	4.0	4.2	4.2	5.1
$\text{KBB}_{\text{QS}}^{\text{EL}}$	6.6	6.6	6.8	8.9	5.8	6.4	7.0	7.0	6.9	6.3	6.2	7.9	7.7	5.8	6.6	7.0
MBB	3.6	3.3	4.4	5.2	3.0	2.9	4.6	4.0	3.6	3.5	3.9	4.6	3.4	3.2	3.8	4.1
MBB^{EL}	5.8	5.3	5.6	7.4	5.1	5.0	6.6	4.9	5.7	4.2	5.5	5.9	6.7	4.6	6.1	5.2

Define $r(\theta) = (a(\beta)', \gamma')$ and the unrestricted GMM objective function

$$\hat{Q}^a(\theta) = \hat{h}^a(\theta)'\hat{\Xi}^{-1}\hat{h}^a(\theta).$$

Consider the GMM estimator

$$\hat{\theta}^e = \underset{\theta\in\Theta}{\arg\min}\hat{Q}^a(\theta).$$

As pointed out by Ruud (2000, p. 574-575) the sub-vectors of $\hat{\theta}$ are

$$\begin{aligned}\hat{\beta}^e &= \underset{\beta\in\mathcal{B}}{\arg\min}\hat{g}(\beta)'\hat{\Omega}\hat{g}(\beta),\\ \hat{\gamma} &= \hat{q}(\hat{\beta}) - \hat{\Xi}_{21}\hat{\Xi}_{11}^{-1}\hat{g}(\hat{\beta}).\end{aligned}$$

We note that by Theorem 2.1 $\hat{\beta}^e = \beta_0 + o_p(1)$ also as $\hat{\Xi} = \Xi + o_p(1)$ and Ξ_{11} is invertible we have by a UWL that $\hat{\gamma} = o_p(1)$ as $\mathrm{E}(h_t^a(\beta_0, 0)) = 0$ under the regularity conditions of the Theorem 2.1. and

$$\sqrt{T}(\hat{\theta}^e - \theta_0) \xrightarrow{d} N(0, \Lambda)$$

by Theorem 2.2 as $\beta_0 \in int(\mathcal{B})$ and $0 \in int(\mathbb{R}) = \mathbb{R}$ where $\Lambda = (D'\Xi^{-1}D)$.

Furthermore using the usual arguments based on first order conditions we have

$$\sqrt{T}\begin{pmatrix}\hat{\beta}^e - \beta_0\\ \hat{\gamma}\end{pmatrix} = -[D'\Xi^{-1}D]^{-1}D'\Xi^{-1}\sqrt{T}\hat{h}^a(\beta_0, 0) + o_p(1).$$

Thus by a Taylor expansion we have under H_0

$$\begin{aligned}\sqrt{T}\begin{pmatrix}a(\hat{\beta}^e)\\ \hat{\gamma}\end{pmatrix} &= -R(\ddot{\theta})[D'\Xi^{-1}D]^{-1}D'\Xi^{-1}\sqrt{T}\hat{h}^a(\beta_0, 0) + o_p(1)\\ &= -R[D'\Xi^{-1}D]^{-1}D'\Xi^{-1}\sqrt{T}\hat{h}^a(\beta_0, 0) + o_p(1)\end{aligned}$$

where $\ddot{\theta}$ is in a line between $(\hat{\beta}^{e\prime}, \hat{\gamma}')'$ and 0. Hence

$$\begin{aligned}\mathcal{W} &= T\begin{pmatrix}a(\hat{\beta}^e)\\ \hat{\gamma}\end{pmatrix}'\left[\hat{R}(\hat{D}'\hat{\Xi}^{-1}\hat{D})^{-1}\hat{R}\right]^{-1}\begin{pmatrix}a(\hat{\beta}^e)\\ \hat{\gamma}\end{pmatrix}\\ &= \sqrt{T}\hat{h}^a(\beta_0, 0)'K\sqrt{T}\hat{h}^a(\beta_0, 0) + o_p(1),\end{aligned}$$

as $\hat{D} = D + o_p(1)$, $\hat{\Xi} = \Xi + o_p(1)$, $\hat{R} = R + o_p(1)$, $\sqrt{T}\hat{h}(\beta_0, 0) = O_p(1)$ and where

$$K \equiv \Xi^{-1}D[D'\Xi^{-1}D]^{-1}R'\left[R(D'\Xi^{-1}D)^{-1}R\right]^{-1}R[D'\Xi^{-1}D]^{-1}D'\Xi^{-1}.$$

Note that $\Xi K\Xi K\Xi = \Xi K\Xi$ and $tr(K\Xi) = s + r$.Thus by Theorem 9.2.1 of Rao and Mitra(1971) It follows that $\mathcal{W}\xrightarrow{d}\chi^2(r + s)$.

We consider now the $\mathcal{LM}$ statistic

$$\mathcal{LM} = T\hat{h}(\hat{\theta}_r^e)'\hat{\Xi}^{-1}\hat{D}(\hat{D}'\hat{\Xi}^{-1}\hat{D})^{-1}\hat{D}'\hat{\Xi}^{-1}\hat{h}(\hat{\theta}_r^e).$$

Note that the restricted GMM estimator solves

$$\hat{\theta}^e_r = \arg\min_{\theta\in\Theta_r} \hat{h}^a(\theta)'\hat{\Xi}^{-1}\hat{h}^a(\theta),$$

where $\Theta_r = \{(\gamma', \beta') \in \Theta : a(\beta) = 0, \gamma = 0\}$. We note that since Θ is compact, Θ_r is compact. Note that $\hat{\theta}^e_r = (\hat{\beta}^{e\prime}_r, 0')'$ and $\hat{\beta}^e_r$ is consistent by Theorem 2.1.

We derive now the distribution of the restricted estimator. The Lagrangian is

$$L = \hat{h}^a(\theta)'\hat{\Xi}^{-1}\hat{h}^a(\theta) - \lambda' r(\theta)$$

and the first order conditions are

$$\begin{aligned} \hat{D}_r'\hat{\Xi}^{-1}\hat{h}^a(\hat{\theta}^e_r) - R(\hat{\theta}^e_r)\lambda &= 0, \\ r(\hat{\theta}^e_r) &= 0, \end{aligned}$$

where $\hat{D}_r = \hat{D}(\hat{\beta}^e_r)$. Thus by the usual arguments we have

$$\sqrt{T}(\hat{\theta}^e_r - \theta_0) = -\left(\Lambda - \Lambda R'(R\Lambda R)^{-1} R\Lambda\right) D'\Xi^{-1}\sqrt{T}\hat{h}^a(\beta_0, 0) + o_p(1).$$

where $\Lambda = [D'\Xi^{-1}D]^{-1}$. By a Taylor expansion

$$\begin{aligned} \sqrt{T}\hat{h}^a(\hat{\theta}^e_r) &= \sqrt{T}\hat{h}^a(\beta_0, 0) + D\sqrt{T}(\hat{\theta}^e_r - \theta_0) \\ &= [I_{m+s} - \left(D\Lambda D'\Xi^{-1} - D\Lambda R'(R\Lambda R)^{-1} R\Lambda D'\Xi^{-1}\right]\sqrt{T}\hat{h}(\beta_0, 0) + o_p(1). \end{aligned} \quad \text{(A.1)}$$

Thus

$$\begin{aligned} \mathcal{LM} &= T\hat{h}^a(\hat{\beta}^e_r, 0)'\hat{\Xi}^{-1}\hat{D}(\hat{D}'\hat{\Xi}^{-1}\hat{D})^{-1}\hat{D}'\hat{\Xi}^{-1}\hat{h}(\hat{\beta}^e_r, 0) \\ &= T\hat{h}^a(\hat{\theta}^e_r)'\hat{\Xi}^{-1}\hat{D}(\hat{D}'\hat{\Xi}^{-1}\hat{D})^{-1}\hat{D}'\hat{\Xi}^{-1}\hat{h}^a(\hat{\theta}^e_r) \\ &= \sqrt{T}\hat{h}^a(\beta_0, 0)'K\sqrt{T}\hat{h}^a(\beta_0, 0) + o_p(1) \end{aligned}$$

as $\hat{D} = D + o_p(1)$ and $\hat{\Xi} = \Xi + o_p(1)$. Thus $\mathcal{LM}$ is asymptotically equivalent to $\mathcal{W}$.

Now we consider the distance statistic

$$\begin{aligned} \mathcal{D} &= T[\hat{h}(\hat{\beta}^e_r)'\hat{\Xi}^{-1}\hat{h}(\hat{\beta}^e_r) - \hat{g}(\hat{\beta}^e)'\hat{\Omega}^{-1}\hat{g}(\hat{\beta}^e)] \\ &= T[\hat{h}^a(\hat{\theta}^e_r)'\hat{\Xi}^{-1}\hat{h}^a(\hat{\theta}^e_r) - \hat{h}^a(\hat{\theta}^e)'\hat{\Xi}^{-1}\hat{h}^a(\hat{\theta}^e)]. \end{aligned}$$

It follows from replacing $\sqrt{T}\hat{h}^a(\hat{\theta}^e_r)$ by $(A.1)$ and $\sqrt{T}\hat{h}^a(\hat{\theta}^e)$ by

$$\begin{aligned} \sqrt{T}\hat{h}^a(\hat{\theta}^e) &= \sqrt{T}\hat{h}^a(\beta_0, 0) + D\sqrt{T}(\hat{\theta}^e - \theta_0) + o_p(1) \\ &= \sqrt{T}\hat{h}^a(\beta_0, 0) - D[D'\Xi^{-1}D]^{-1}D'\Xi^{-1}\sqrt{T}\hat{h}^a(\beta_0, 0) + o_p(1) \\ &= [I_{m+s} - D[D'\Xi^{-1}D]^{-1}D'\Xi^{-1}]\sqrt{T}\hat{h}^a(\beta_0, 0) + o_p(1). \end{aligned}$$

Thus as $\sqrt{T}\hat{h}^a(\beta_0, 0) = O_p(1)$ we have

$$\mathcal{D} = \sqrt{T}\hat{h}^a(\beta_0, 0)'K\sqrt{T}\hat{h}^a(\beta_0, 0) + o_p(1)$$

and the result follows. ■

A.2 Auxiliary results on Generalised Empirical Likelihood

A.2.1 Unrestricted models

The following Lemma corresponds to a version of Lemma A.1 of Ramalho and Smith (2011) for weakly dependent data.

Lemma A.1 *If Assumptions 2.4, 2.6, 2.7 and 2.8 are satisfied, then* $T\hat{\pi}_t = 1 + o_p(1)$ *and*

$$T^{1/2}(\hat{\pi}_t - 1/T) = \frac{S_T}{T}\hat{g}'_{tT}\frac{T^{1/2}}{S_T}\hat{\lambda}(1/k_2 + o_p(1)) + O_p(\frac{S_T}{T}).$$

uniformly $t = 1, ..., T$.

Proof: The proof of the is contained in the proof of Theorem 3.1 of Smith (2004,p. A.11). ■

Let $w_{tT} = \frac{1}{S_T}\sum_{s=t-T}^{t-1} k(\frac{s}{S_T})w_t$, $t = 1, ..., T$, $\tilde{w} = \sum_{t=1}^{T}\hat{\pi}_t w_{tT}$, $\hat{w} = \sum_{t=1}^{T} w_t/T$.

Assumption A.1 **(i)** *The random vectors* $\{(w_t, z_t), -\infty < t < \infty\}$ *form a strictly stationary and mixing with mixing coefficients of size* $-3v/(v-1)$ *for some* $v > 1$, **(ii)** $\mathrm{E}[w_t] = 0$, $\mathrm{E}[\|w_t\|^{\alpha}] < \infty$, *for some* $\alpha > \max(4v, 1/\eta)$ *and* $\Upsilon = \lim_{T\to\infty} \mathrm{var}[T^{1/2}\hat{w}]$ *is finite and p.d.*

The following Lemma corresponds to a simplified version of Theorem 3.1 of Smith (2011).

Lemma A.2 *Under assumptions 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, and A.1*

$$\sqrt{T}\tilde{w} = T^{-1/2}\sum\nolimits_{t=1}^{T} w_t - B_0 P T^{1/2}\hat{g}(\beta_0) + o_p(1),$$

where $B_0 = \sum_{s=-\infty}^{\infty} \mathrm{E}[w_t g_{t-s}(\beta_0)']$. *Additionally if* $w_t = g(z_t, \beta_0)$ *we have*

$$\sqrt{T}\tilde{w} = [G\Sigma G'\Omega^{-1}]T^{1/2}\hat{g}(\beta_0) + o_p(1).$$

Proof: Note that by Lemma A.1

$$\begin{aligned} T^{1/2}\tilde{w} &= T^{1/2}\sum\nolimits_{t=1}^{T}\hat{\pi}_t w_{tT} \\ &= T^{1/2}\sum\nolimits_{t=1}^{T} w_{tT}/T + \sum\nolimits_{t=1}^{T}[\frac{S_T}{T}\hat{g}_{tT}'\frac{T^{1/2}}{S_T}\hat{\lambda}(k+o_p(1)) + O_p(\frac{S_T}{T})]w_{tT} \\ &= T^{1/2}\sum\nolimits_{t=1}^{T} w_{tT} + \frac{S_T}{T}\sum\nolimits_{t=1}^{T} w_{tT}\hat{g}_{tT}'\frac{T^{1/2}}{S_T}\hat{\lambda}(k+o_p(1)) + \sum\nolimits_{t=1}^{T} w_{tT}O_p(\frac{S_T}{T})]. \end{aligned}$$

Now by Smith (2011) Proof of Theorem 2.3 (see expression B.2, p A.11) we have

$$\frac{T^{1/2}}{S_T}\hat{\lambda} = -T^{1/2}P\hat{g}_T(\beta_0) + o_p(1).$$

Thus

$$\begin{aligned} T^{1/2}\tilde{w} &= T^{-1/2}\sum_{t=1}^{T} w_{tT} + \frac{S_T}{T}\sum_{t=1}^{T} w_{tT}\hat{g}_{tT}'[-T^{1/2}P\hat{g}_T(\beta_0) + o_p(1)](k+o_p(1)) \\ &\quad + \sum_{t=1}^{T} w_{tT}O_p(\frac{S_T}{T})], \end{aligned}$$

where $\hat{g}_T(\beta) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{S_T}\sum_{s=t-T}^{t-1} k\left((s/S_T)\, g_{t-s}(\beta)\right.$, and $\hat{g}_{tT} = \hat{g}_{tT}(\hat{\beta})$. Now note that as in Lemma A.2 of Smith (2011) we have

$$\begin{aligned} T^{-1/2}\sum_{t=1}^{T} w_{tT} &= T^{-1/2}\sum\nolimits_{t=1}^{T} w_t + O_p(T^{-1/2}), \\ T^{1/2}\hat{g}_T(\beta) &= T^{1/2}\sum\nolimits_{t=1}^{T} g_t(\beta_0) + O_p(T^{-1/2}). \end{aligned}$$

and

$$\begin{aligned} \sum_{t=1}^{T} w_{tT}O_p(\frac{S_T}{T}) &= O_p(\frac{S_T}{T^{1/2}})[T^{-1/2}\sum_{t=1}^{T} w_t + O_p(T^{-1/2})] \\ &= o_p(1). \end{aligned}$$

By similar arguments of the proof of Lemma A3 of Smith (2011) we have

$$\frac{S_T}{T}k\sum_{t=1}^{T} w_{tT}\hat{g}_{tT}' = B_0 + o_p(1),$$

where $B_0 = \sum_{s=-\infty}^{\infty}\mathrm{E}[w_t g_{t-s}(\beta_0)']$.

$$\sqrt{T}\tilde{w} = T^{-1/2}\sum_{t=1}^{T} w_t - B_0 P T^{1/2}\hat{g}(\beta_0) + o_p(1).$$

Now note that if $w_t = g(z_t, \beta_0)$ we have $T^{-1/2}\sum_{t=1}^{T} w_t = T^{1/2}\hat{g}(\beta_0)$ and $B_0 = \Omega$ and hence

$$\begin{aligned} T^{-1/2}\sum_{t=1}^{T} w_t - B_0 P T^{1/2}\hat{g}(\beta_0) &= T^{1/2}\hat{g}(\beta_0) - \Omega[\Omega^{-1} - \Omega^{-1}G\Sigma G'\Omega^{-1}]T^{1/2}\hat{g}(\beta_0) \\ &= G\Sigma G'\Omega^{-1}T^{1/2}\hat{g}(\beta_0). \end{aligned}$$

■

Proof of Theorem 4.5: Note that by CS

$$\left|\tilde{Q}_T(\beta) - \hat{Q}(\beta)\right| \leq \|\tilde{g}(\beta) - \hat{g}(\beta)\|^2\,\|W_T\|\,.$$

Note that by T

$$\sup_{\beta\in B}\|\tilde{g}(\beta)-\hat{g}(\beta)\| \le \sup_{\beta\in B}\|\tilde{g}(\beta)-\mathrm{E}[g(z_t,\beta]]\| + \sup_{\beta\in B}\|\hat{g}(\beta)-\mathrm{E}[g(z_t,\beta]\|.$$

Also by a UWL

$$\sup_{\beta\in B}\|\hat{g}(\beta)-\mathrm{E}[g(z_t,\beta]\| = o_p(1).$$

Now

$$\begin{aligned}\sup_{\beta\in B}\|\tilde{g}(\beta)-\mathrm{E}[g(z_t,\beta]]\| &= \sup_{\beta\in B}\|\tilde{g}(\beta)-\hat{g}_T(\beta)\| + \sup_{\beta\in B}\|\hat{g}(\beta)-\mathrm{E}[g(z_t,\beta]\| \\ &\le \max_{1\le t\le T}|Tp_{tT}-1|\sup_{\beta\in B}\|\hat{g}_T(\beta)\|+o_p(1) \\ &= o_p(1),\end{aligned}$$

by $\max_{1\le t\le T}|Tp_{tT}-1| = 1+o_B(1)$ and an UWL. Hence $\left|\tilde{Q}_T(\beta)-\hat{Q}(\beta)\right| = o_p(1)$ as $\|W_T-W\|=o_p(1)$. Thus the result follows by Theorem 2.1. ■

Proof of Theorem 4.6: The first order criteria yield $\sqrt{T}\tilde{G}_T'W_T\tilde{g}_T(\tilde{\beta})=0$ where $\tilde{G}_T\equiv\partial\tilde{g}_T(\tilde{\beta})/\partial\beta'$. Hence by a Taylor expansion

$$\sqrt{T}\hat{G}_T'W_T\tilde{g}_T(\beta_0)+\hat{G}_T'W_T\ddot{G}_T\sqrt{T}(\tilde{\beta}-\beta_0)=0,$$

where $\ddot{G}_T\equiv\partial\tilde{g}_T(\ddot{\beta})/\partial\beta'$ where $\ddot{\beta}$ is in a line joining $\tilde{\beta}$ and β_0. Solving for $\sqrt{T}(\tilde{\beta}-\beta_0)$ we obtain

$$\begin{aligned}\sqrt{T}(\tilde{\beta}-\beta_0) &= -(\hat{G}_T'W_T\ddot{G}_T^*)^{-1}\sqrt{T}\hat{G}_T'W_T\tilde{g}_T(\beta_0) \\ &= -(\hat{G}_T'W_T\ddot{G}_T)^{-1}\hat{G}_T'W_T\{[G\Sigma G'\Omega^{-1}]T^{1/2}\hat{g}(\beta_0)+o_p(1)\}\end{aligned} \tag{A.2}$$

by Lemma A.2. But by Lemma A.1 of Smith (2011) we have $\hat{G}_T=G+o_p(1)$, $\ddot{G}_T=G+o_p(1)$. And since $W_T=W+o_p(1)$ and $T^{1/2}\hat{g}(\beta_0)=O_p(1)$. Thus

$$\begin{aligned}\sqrt{T}(\tilde{\beta}-\beta_0) &= -(G'WG)^{-1}G'WG\Sigma G'\Omega^{-1}T^{1/2}\hat{g}(\beta_0)+o_p(1) \\ &= \Sigma G'\Omega^{-1}T^{1/2}\hat{g}(\beta_0)+o_p(1)\end{aligned}$$

which corresponds to the asymptotic representation of the efficient GMM estimator (see for instance Hall, 2005, p. 70 eq 3.26 with $W_T=\Omega^{-1}$) ■

A.2.2 Restricted models

For notational convenience we now define the restricted GEL estimator in a slightly different but equivalent manner to what is done in sub-section 4.2.6. Let

$$\begin{aligned}\bar{P}_n(\theta,\varphi) &= \frac{1}{T}\sum\nolimits_{t=1}^{T}[\rho([\varphi'h_{tT}^a(\theta)]/k_2)-\rho_0], \\ P_n(\beta,\varphi) &= \bar{P}_n\left((\beta',0')',\varphi\right)=\frac{1}{T}\sum\nolimits_{t=1}^{T}[\rho([\varphi'h_{tT}(\beta)]/k_2)-\rho_0], \\ \tilde{P}_n(\theta,\varphi,\mu) &= \frac{1}{T}\sum\nolimits_{t=1}^{T}[\rho([\varphi'h_{tT}^a(\theta)+\mu'r(\theta)]/k_2)-\rho_0],\end{aligned}$$

where $\theta=(\beta',\gamma')'$, $h^a(z_t,\theta)=(g(z_t,\beta)',q(z_t,\beta)'-\gamma')'$, $h_{tT}^a(\theta)=\frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1}k(\frac{s}{S_T})h^a(z_t,\theta)$, $t=1,...,T$. Let $\Theta_r=\left\{\theta=(\beta',\gamma')':a(\beta)=0,\gamma=0\right\}$, thus $\Theta_r=\mathcal{B}_r\times\{0\}$. Let $\Delta_T=\left\{\varphi:\|\varphi\|\le D(T/S_T^2)^{-\zeta}\right\}$.

Let

$$(\varphi(\theta)',\mu(\theta)')=\arg\max_{\varphi\in\Delta_T,\mu\in\mathbb{R}^s}\tilde{P}_n(\theta,\varphi,\mu).$$

Note that $\varphi(\theta)$ can also be defined as

$$\varphi(\theta)=\arg\max_{\varphi\in\Delta_T}\bar{P}_n(\theta,\varphi),\ \theta\in\Theta_r$$

and

$$\begin{aligned}\hat{\theta}_r &= \arg\min_{\theta\in\Theta_r}\bar{P}_n(\theta,\varphi(\theta)) \\ &= \arg\min_{\theta\in\Theta_r}\tilde{P}_n(\theta,\varphi(\theta),\mu(\theta))\end{aligned}$$

and let $\hat{\varphi}_r=\varphi(\hat{\theta}_r)$, $\hat{\mu}_r=\mu(\hat{\theta}_r)$.

We note that $\hat{\theta}_r=S_1\hat{\beta}_r$ where

$$\hat{\beta}_r=\arg\min_{\beta\in\mathcal{B}_r}\sup_{\varphi\in\Delta_T}P_n(\beta,\varphi)$$

and S_1 is a matrix such that $S_1\hat{\beta}_r=(\hat{\beta}_r',0_{s\times1}')'$.

The following Theorem provides a convenient asymptotic representation of the restricted GEL estimator and corresponding Lagrange multiplier.

Theorem A.1 *If Assumptions 2.4, 2.6, 4.6, 4.7 and 4.8 are satisfied* $\hat{\beta}_r \xrightarrow{p} \beta_0$ *and* $\hat{\varphi}_r \xrightarrow{p} 0, \hat{\mu}_r \xrightarrow{p} 0$. *Moreover,* $\|\hat{\varphi}_r\| = O_p[(T/S_T^2)^{-1/2}], \|\hat{\mu}_r\| = O_p[(T/S_T^2)^{-1/2}]$,

$$\begin{aligned} S_1\sqrt{T}\left(\hat{\beta}_r - \beta_0\right) &= -[\Lambda - \Lambda R'\left[R\Lambda R'\right]^{-1} R\Lambda]D'\Xi^{-1}\sqrt{T}\hat{h}_T(\beta_0) + o_p(1), \\ \frac{\sqrt{T}\hat{\varphi}_r}{S_T} &= -P_r\sqrt{T}\hat{h}_T(\beta_0) + o_p(1), \end{aligned}$$

where $P_r = \Xi^{-1} - \Xi^{-1}DS_1[\Lambda - \Lambda R'\left[R\Lambda R'\right]^{-1} R\Lambda]D'\Xi^{-1}$.

Proof: Let $k = 1/k_2$. The first order conditions are

$$\begin{aligned} k\frac{1}{T}\sum\nolimits_{t=1}^{T} \rho_1(k(\hat{\varphi}_r' h_{tT}(\hat{\beta}_r)))h_{tT}^a(\hat{\beta}_r, 0) &= 0, \quad \text{(A.3)} \\ k\frac{1}{T}\sum\nolimits_{t=1}^{T} \rho_1(k(\hat{\varphi}_r' h_{tT}(\hat{\beta}_r)))r(\hat{\beta}_r, 0) &= 0, \\ k\frac{1}{T}\sum\nolimits_{t=1}^{T} \rho_1(k(\hat{\varphi}_r' h_{tT}(\hat{\beta}_r)))(D_{tT}(\hat{\beta}_r)'\hat{\varphi}_r + R(\hat{\beta}_r)'\hat{\mu} &= 0, \end{aligned}$$

where $D_{tT}(\theta) = \partial h_{tT}^a(\theta)/\partial\theta'$ and $R(\theta) = 0$. Note that $r(\hat{\beta}_r, 0) = 0$. Similarly to Theorem 2.5 we have $\hat{\beta}_r \to \beta_0$, $\hat{\varphi}_r \xrightarrow{p} 0$ and $\|\hat{\varphi}_r\| = O_p[(T/S_T^2)^{-1/2}]$. Therefore
$\max_{1\le i\le T}\left|\rho_1(k(\hat{\varphi}_r' h_{tT}^a(\hat{\beta}_r, 0))) + 1\right| \xrightarrow{p} 0$. Thus

$$-k[1 + o_p(1)]\frac{1}{T}\sum\nolimits_{t=1}^{T} D_{tT}(\hat{\beta}_r, 0)'\hat{\varphi} + R(\hat{\beta}_r, 0)'\hat{\mu}_r = 0$$

as $\frac{1}{T}\sum_{t=1}^{T} D_{tT}(\hat{\beta}_r, 0)' \xrightarrow{p} D$ by a UWL and $\hat{\varphi}_r \xrightarrow{p} 0$ and we have

$$(R + o_p(1))\hat{\mu}_r = o_p(1)$$

and consequently as $rank(R) = r + s$ we must have $\hat{\mu} \xrightarrow{p} 0$.

Note also that $\|\hat{\varphi}_r\| = O_p[(T/S_T^2)^{-1/2}]$ hence

$$(R + o_p(1))\hat{\mu}_r = O_p[(T/S_T^2)^{-1/2}]$$

and consequently $\hat{\mu}_r = O_p[(T/S_T^2)^{-1/2}]$.

Now a first order Taylor expansion of the lfs of $(A.3)$ around $\varphi_r = 0$ gives

$$-k\frac{1}{T}\sum\nolimits_{t=1}^{T} h_{tT}(\hat{\beta}_r) + \frac{1}{T}\sum\nolimits_{t=1}^{T} \rho_2(k(\tilde{\varphi}_r' h_{tT}^a(\hat{\beta}_r, 0)))h_{tT}(\hat{\beta}_r)h_{tT}(\hat{\beta}_r)'\hat{\varphi}_r = 0, \quad \text{(A.4)}$$

where $\tilde{\varphi}_r$ is in a line joining $\hat{\varphi}_r$ and 0. Now note that by a Taylor expansion we have

$$h_{tT}(\hat{\beta}_r) = h_{tT}(\beta_0) + D_{tT}(\tilde{\beta}_r)S_1\left(\hat{\beta}_r - \beta_0\right), \quad \text{(A.5)}$$

where $\tilde{\beta}_r$ lies in a line joining $\hat{\beta}_r$ and β_0. Replacing $(A.5)$ in $(A.4)$ yields

$$\begin{aligned} &-k\frac{1}{T}\sum\nolimits_{t=1}^{T} h_{tT}(\beta_0) - k\frac{1}{T}\sum\nolimits_{t=1}^{T} D_{tT}(\tilde{\beta}_r)S_1\left(\hat{\beta}_r - \beta_0\right) \\ &+\frac{S_T}{T}\sum\nolimits_{t=1}^{T} \rho_2(k(\tilde{\varphi}_r' h_{tT}(\hat{\beta}_r)))h_{tT}(\hat{\beta}_r)h_{tT}(\hat{\beta}_r)'\frac{\hat{\varphi}_r}{S_T} = 0. \end{aligned}$$

Now as $\hat{\mu}_r' = O_p(S_T/\sqrt{T})$, $\hat{\varphi}_r = O_p(S_T/\sqrt{T})$, $\sqrt{T}(\hat{\beta}_r - \beta_0) = O_p(1)$ (which is a consequence of the fact that $\left\|h_T(\hat{\beta}_r)\right\| = O_p(T^{-1/2})$ by Theorem 2.2 of Smith (2011) and assumption (4.8)) and $\max_{1\le i\le T}\left|\rho_2(k(\tilde{\varphi}_r' h_{tT}(\hat{\beta}_r))) + 1\right|$ we have by a UWL and continuity of $R(\beta)$

$$\begin{aligned} D'\frac{\sqrt{T}}{S_T}\hat{\varphi}_r + R'\frac{\sqrt{T}}{S_T}\hat{\mu}_r &= o_p(1) \\ \sqrt{T}\hat{h}_T(\beta_0) + DS_1\sqrt{T}\left(\hat{\beta}_r - \beta_0\right) + \Xi\frac{\sqrt{T}\hat{\varphi}_r}{S_T} &= o_p(1), \\ R\sqrt{T}S_1(\hat{\beta}_r - \beta_0) &= o_p(1). \end{aligned}$$

Multiply the first equation by $R\Lambda$, where $\Lambda = [D'\Xi^{-1}D]^{-1}$, and solving for $\hat{\mu}_r$ we obtain

$$\frac{\sqrt{T}}{S_T}\hat{\mu}_r = -\left[R\Lambda R'\right]^{-1} R\Lambda D'\frac{\sqrt{T}}{S_T}\hat{\varphi}_r + o_p(1).$$

Replacing it in the first equation yields

$$D'\frac{\sqrt{T}}{S_T}\hat{\varphi}_r - R'\left[R\Lambda R'\right]^{-1} R\Lambda D'\frac{\sqrt{T}}{S_T}\hat{\varphi}_r = o_p(1)$$

and multiplying both sides by Λ we have

$$\Lambda D'\frac{\sqrt{T}}{S_T}\hat{\varphi}_r-\Lambda R'\left[R\Lambda R'\right]^{-1}R\Lambda D'\frac{\sqrt{T}}{S_T}\hat{\varphi}_r=o_p(1),$$

which is equivalent to

$$[\Lambda-\Lambda R'\left[R\Lambda R'\right]^{-1}R\Lambda]D'\frac{\sqrt{T}}{S_T}\hat{\varphi}_r=o_p(1).$$

Consider now

$$\sqrt{T}\hat{h}_T(\beta_0)+DS_1\sqrt{T}\left(\hat{\beta}_r-\beta_0\right)+\Xi\frac{\sqrt{T}\hat{\varphi}_r}{S_T}=o_p(1).$$

Multiplying both sides by $[\Lambda-\Lambda R'\left[R\Lambda R'\right]^{-1}R\Lambda]D'\Xi^{-1}$ we obtain

$$\begin{gathered}[\Lambda-\Lambda R'\left[R\Lambda R'\right]^{-1}]R\Lambda]D'\Xi^{-1}\sqrt{T}\hat{h}_T(\beta_0)\\ +[\Lambda-\Lambda R'\left[R\Lambda R'\right]^{-1}R\Lambda]D'\Xi^{-1}DS_1\sqrt{T}\left(\hat{\beta}_r-\beta_0\right)=o_p(1).\end{gathered}$$

Now

$$R\Lambda(D)'\Xi^{-1}DS_1\sqrt{T}\left(\hat{\beta}_r-\beta_0\right)=RS_1\sqrt{T}\left(\hat{\beta}_r-\beta_0\right)=R\sqrt{T}\left(\hat{\beta}_r-\beta_0\right)=o_p(1).$$

Hence we have

$$S_1\sqrt{T}\left(\hat{\beta}_r-\beta_0\right)=-[\Lambda-\Lambda R'\left[R\Lambda R'\right]^{-1}R\Lambda]D'\Xi^{-1}\sqrt{T}\hat{h}_T(\beta_0)+o_p(1). \tag{A.6}$$

Additionally note that

$$\sqrt{T}\hat{h}_T(\beta_0)+DS_1\sqrt{T}\left(\hat{\beta}_r-\beta_0\right)+\Xi\frac{\sqrt{T}\hat{\varphi}_r}{S_T}=o_p(1),$$

and replacing $(A.6)$ in this equation and solving for $\sqrt{T}\hat{\varphi}_r/S_T$ yields

$$\begin{aligned}\frac{\sqrt{T}\hat{\varphi}_r}{S_T}&=-\Xi^{-1}[\sqrt{T}\hat{h}_T(\beta_0)+DS_1\sqrt{T}\left(\hat{\beta}_r-\beta_0\right)]\\&=-\Xi^{-1}[\sqrt{T}\hat{h}_T(\beta_0)-DS_1[\Lambda-\Lambda R'\left[R\Lambda R'\right]^{-1}R\Lambda](D)'\Xi^{-1}\sqrt{T}\hat{h}_T(\beta_0).]\\&=[-\Xi^{-1}+\Xi^{-1}DS_1[\Lambda-\Lambda R'\left[R\Lambda R'\right]^{-1}R\Lambda]D'\Xi^{-1}]\sqrt{T}\hat{h}_T(\beta_0).\end{aligned}$$

■

Let

$$\tilde{\pi}_t=\frac{\rho_1([\hat{\varphi}_r'h_{tT}(\hat{\theta}_r)]/k_2)}{\sum_{t=1}^T\rho_1([\hat{\varphi}_r'h_{tT}(\hat{\theta}_r)]/k_2)},t=1,...,T.$$

Lemma A.3 *If Assumptions 2.4, 2.6, 2.7 and 2.8 are satisfied, then* $T\tilde{\pi}_t=1+o_p(1)$ *and*

$$T^{1/2}(\tilde{\pi}_t-1/T)=\frac{S_T}{T}\hat{h}_{tT}'\frac{T^{1/2}}{S_T}\hat{\varphi}_r(1/k_2+o_p(1))+O_p(\frac{S_T}{T}).$$

uniformly $t=1,...,T.$

Proof: This is similar to the proof of Lemma A.1 ■
Let $\tilde{w}_r=\sum_{t=1}^T\tilde{\pi}_t w_{tT}$.

Lemma A.4 *Under assumptions 2.4, 2.6, 4.6, 4.7 and 4.8 and A.1*

$$\sqrt{T}\tilde{w}_r=T^{-1/2}\sum_{t=1}^T w_t-J_0P_rT^{1/2}\hat{h}(\beta_0)+o_p(1),$$

where $J_0=\sum_{s=-\infty}^{\infty}\mathrm{E}[w_th_{t-s}(\beta_0)']$, $P_r=\Xi^{-1}-\Xi^{-1}DS_1[\Lambda-\Lambda R'\left[R\Lambda R'\right]^{-1}R\Lambda]D'\Xi^{-1}$. *Additionally if* $w_t=h(z_t,\beta_0)$ *we have*

$$\sqrt{T}\tilde{w}_r=DS_1[\Lambda-\Lambda R'\left[R\Lambda R'\right]^{-1}R\Lambda]D'\Xi^{-1}]\sqrt{T}\hat{h}(\beta_0)+o_p(1).$$

Proof: Note that by Lemma A.3

$$\begin{aligned}T^{1/2}\tilde{w}_r&=T^{1/2}\sum_{t=1}^T\tilde{\pi}_t w_{tT}\\&=T^{1/2}\sum_{t=1}^T w_{tT}/T+\sum_{t=1}^T[\frac{S_T}{T}\hat{h}_{tT}'\frac{T^{1/2}}{S_T}\hat{\varphi}_r(k+o_p(1))+O_p(\frac{S_T}{T})]w_{tT}\\&=T^{1/2}\sum_{t=1}^T w_{tT}+\frac{S_T}{T}\sum_{t=1}^T w_{tT}\hat{h}_{tT}'\frac{T^{1/2}}{S_T}\hat{\varphi}_r(k+o_p(1))+\sum_{t=1}^T w_{tT}O_p(\frac{S_T}{T})].\end{aligned}$$

Now by Theorem A.1 we have

$$\frac{T^{1/2}}{S_T}\hat{\varphi}_r = -T^{1/2}P_r\hat{h}(\beta_0) + o_p(1).$$

Thus

$$\begin{aligned} T^{1/2}\tilde{w}_r &= T^{-1/2}\sum\nolimits_{t=1}^{T} w_{tT} + \frac{S_T}{T}\sum\nolimits_{t=1}^{T} w_{tT}\hat{g}'_{tT}[-T^{1/2}P_r\hat{h}(\beta_0)+o_p(1)](k+o_p(1)) \\ &\quad + \sum\nolimits_{t=1}^{T} w_{tT}O_p(\frac{S_T}{T})], \end{aligned}$$

where $\hat{h}_T(\beta) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{S_T}\sum_{s=t-T}^{t-1} k\left((s/S_T)\, h_{t-s}(\beta)\right.$, and $\hat{h}_{tT} = \hat{h}_{tT}(\hat{\beta})$. Now note that as in Lemma A.2 of Smith (2011) we have

$$\begin{aligned} T^{-1/2}\sum\nolimits_{t=1}^{T} w_{tT} &= T^{-1/2}\sum\nolimits_{t=1}^{T} w_t + O_p(T^{-1/2}), \\ T^{1/2}\hat{h}_T(\beta) &= T^{1/2}\sum\nolimits_{t=1}^{T} h_t(\beta_0) + O_p(T^{-1/2}), \end{aligned}$$

and

$$\begin{aligned} \sum\nolimits_{t=1}^{T} w_{tT}O_p(\frac{S_T}{T}) &= O_p(\frac{S_T}{T^{1/2}})[T^{-1/2}\sum\nolimits_{t=1}^{T} w_t + O_p(T^{-1/2})] \\ &= o_p(1). \end{aligned}$$

By a similar argument arguments of the proof of Lemma A3 of Smith (2011) we have

$$\frac{S_T}{T}k\sum\nolimits_{t=1}^{T} w_{tT}\hat{h}'_{tT} = J_0 + o_p(1),$$

where $J_0 = \sum_{s=-\infty}^{\infty} E[w_t h_{t-s}(\beta_0)']$. Thus

$$\sqrt{T}\tilde{w}_r = T^{-1/2}\sum\nolimits_{t=1}^{T} w_t - J_0P_rT^{1/2}\hat{h}(\beta_0) + o_p(1).$$

Now note that if $w_t = h(z_t,\beta_0)$ we have $T^{-1/2}\sum_{t=1}^{T} w_t = T^{1/2}\hat{h}(\beta_0)$ and $J_0 = \Xi$ and hence

$$T^{-1/2}\sum\nolimits_{t=1}^{T} w_t - J_0P_rT^{1/2}\hat{h}(\beta_0) = DS_1[\Lambda - \Lambda R'\left[R\Lambda R'\right]^{-1}R\Lambda]D'\Xi^{-1}]\sqrt{T}\hat{h}(\beta_0) + o_p(1).$$

■

A.3 Proofs of the results in sub-section 3 and auxiliary Lemmata on the weighted kernel block bootstrap method

In this Appendix we present bootstrap LLN, CLT and UWL that are required to prove the results.

Proof of Theorem 3.1: Let

$$\begin{aligned} q_{tT} &\equiv \tilde{Y} + (S_T/k_2)^{1/2}(Y_{tT} - \tilde{Y}), (t=1,...,T), \\ q^*_{tT} &\equiv \tilde{Y} + (S_T/k_2)^{1/2}(Y^*_{tT} - \tilde{Y})\ (t=1,...,m_T), \end{aligned}$$

and

$$\begin{aligned} \tilde{q} &\equiv \sum\nolimits_{t=1}^{T} q_{tT}p_{tT} = \sum\nolimits_{t=1}^{T}\tilde{Y}p_{tT} + (S_T/k_2)^{1/2}\sum\nolimits_{t=1}^{T}(w_{tT}-\tilde{w})p_{tT} \\ &= \tilde{Y}, \\ \tilde{q}^* &\equiv \sum\nolimits_{t=1}^{m_T} q^*_{tT}/m_T. \end{aligned}$$

Thus

$$\begin{aligned} \sqrt{m_T}(\tilde{q}^* - \tilde{q}) &= \sqrt{m_T}((S_T/k_2)^{1/2}(\bar{Y}^* - \tilde{Y})) \\ &= \sqrt{T/k_2}(\bar{Y}^* - \tilde{Y}). \end{aligned}$$

and consequently

$$\mathcal{P}^*\{\sqrt{T/k_2}(\bar{Y}^* - \tilde{Y}) \le x\} = \mathcal{P}^*\{\sqrt{m_T}(\tilde{q}^*_T - \tilde{q}) \le x\},$$

where $\bar{Y}^* = \frac{1}{m_T}\sum_{j=1}^{m_T} Y^*_{jT}$.The result is proven if we are able to show the following steps:. Step 1: $\bar{X} \xrightarrow{p} 0$. Step 2: $T^{1/2}\bar{X}/\sigma_\infty \xrightarrow{d} N(0,1)$. Step 3: $\sup_{x\in\mathbb{R}}\left|\mathcal{P}\{T^{1/2}\bar{X} \le x\} - \Phi(x/\sigma_\infty)\right| \to 0$, where $\Phi(\cdot)$ is the c.d.f. of the standard normal distribution. Step 4: $m_T\text{var}^*[\tilde{q}^{e*}_T] \xrightarrow{p} \sigma^2_\infty$. Step 5:

$$\lim_{T\to\infty}\mathcal{P}\left\{\sup_{x\in\mathbb{R}}\left|\mathcal{P}^*\{\frac{\sqrt{m_T}(\tilde{q}^* - \tilde{q}^e)}{\text{var}^*[\sqrt{m_T}\tilde{q}^*]^{1/2}} \le x\} - \Phi(x)\right| \ge \varepsilon\right\} = 0.$$

Step 1: Follows from the ergodic theorem (Theorem 3.34 of White, 1999).

Step 2: By White (1999, Theorem 5.20).

Step 3: From Step 2 and the Polya Theorem, Serfling (2002, p.18), as $\Phi(\cdot)$ is a continuous c.d.f.

Step 4: To prove this note that

$$\mathrm{E}^*(q_{tT}^*) = \mathrm{E}^*(\tilde{Y} + (S_T/k_2)^{1/2}(Y_{tT}^* - \tilde{Y})) = \tilde{Y}$$

and

$$\begin{aligned}
\mathrm{var}^*(q_{tT}^*) &= \mathrm{var}^*(\tilde{Y} + (S_T/k_2)^{1/2}(Y_{tT}^* - \tilde{Y})) \\
&= S_T/k_2 \mathrm{var}^*(Y_{tT}^*) \\
&= \frac{S_T}{k_2}\sum_{t=1}^{T}(Y_{tT} - \tilde{Y})^2 p_{tT} \\
&= \frac{S_T}{k_2}\sum_{t=1}^{T} Y_{tT}^2 p_{tT} - \frac{S_T}{k_2}\tilde{Y}^2 \\
&= \frac{S_T}{k_2}\frac{1}{T}\sum_{t=1}^{T} Y_{tT}^2 T p_{tT} + O_p(\frac{S_T}{T}) \\
&= \frac{S_T}{k_2}\frac{1}{T}\sum_{t=1}^{T} Y_{tT}^2 (1 + o_p(1)) + O_p(\frac{S_T}{T}) \\
&= \sigma_\infty^2 + o_p(1),
\end{aligned}$$

as $\max_{1\leq t\leq T} |Tp_{tT}| \xrightarrow{p} 1$ and the fact that $\tilde{Y} = O_p(\frac{1}{\sqrt{T}})$ and Lemma A.3 of Smith (2011).

Step 5: Since the bootstrap sample observations are independent, we can apply Berry-Esséen inequality. Thus

$$\begin{aligned}
\sup_{x\in\mathbb{R}}\left|\mathcal{P}^*\{\frac{\sqrt{m_T}(\tilde{q}^* - \tilde{q})}{\mathrm{var}^*[\sqrt{m_T}\tilde{q}^*]^{1/2}} \leq x\} - \Phi(x)\right| &\leq \frac{C}{m_T^{1/2}}\mathrm{E}^*[(\frac{\left|q_{tT}^* - \tilde{q}^e\right|}{\mathrm{var}^*[q_{tT}^*]^{1/2}})^3] \\
&= \frac{C}{m_T^{1/2}}\mathrm{var}^*[q_{tT}^*]^{-3/2}\mathrm{E}^*[|q_{tT}^* - \tilde{q}^e|^3].
\end{aligned}$$

Note that $\mathrm{var}^*[q_{tT}^*] = \sigma_\infty^2 + o_p(1)$ and that

$$\begin{aligned}
\mathrm{E}^*[|q_{tT}^* - \tilde{q}|^3] &= \sum_{t=1}^{T} |q_{tT} - \tilde{q}|^3 \, \hat{\pi}_t. \\
&\leq \max_t |q_{tT} - \tilde{q}| \sum_{t=1}^{T} |q_{tT} - \tilde{q}|^2 \, \hat{\pi}_t.
\end{aligned}$$

Now

$$\begin{aligned}
\max_t |q_{tT} - \tilde{q}| &= O(S_T^{1/2}) \max_t \left|Y_{tT} - \tilde{Y}\right| \\
&= O_p(S_T^{1/2} T^{1/\alpha})
\end{aligned}$$

by Lemma A.1 of Smith (2011) and M with $\alpha > \max(4v, 1/\eta)$.

Hence

$$\mathrm{E}^*[|q_{tT}^* - \tilde{q}|^3] = O_p(S_T^{1/2} T^{1/\alpha}).$$

Thus

$$\begin{aligned}
\frac{C}{m_T^{1/2}}\mathrm{E}^*[|q_{tT}^* - \tilde{q}^e|^3] &= S_T^{1/2} T^{-1/2} O_p(S_T^{1/2} T^{1/\alpha}) \\
&= O(S_T T^{1/\alpha - 1/2}) = \\
&= O(T^{1/\alpha - \eta}) o_p(1) \qquad \text{(A.7)}
\end{aligned}$$

since $S_T = O(T^{1/2-\eta})$. Now as $\alpha > \max(4v, 1/\eta) > 1/\eta$ we have $1/\alpha < \eta$ and the result follows as $\mathrm{var}^*[q_{tT}^*] = \sigma_\infty^2 + o_p(1)$. ■

Assumption A.2 **(a)** $\mathrm{E}[||X_t||^{4v}] < \infty$; **(b)** $\Sigma_\infty \equiv \lim_{T\to\infty} \mathrm{var}[T^{1/2}\bar{X}]$ *is finite and positive definite.*

Theorem A.2 *Let Assumptions 2.4, 2.5 and A.2 be satisfied. If* $\mathrm{E}[X_t] = 0$ $m_T = T/S_T$, *then*

$$\lim_{T\to\infty} \mathcal{P}\left\{\sup_{x\in\mathbb{R}^d}\left|\mathcal{P}^*\{T^{1/2}\left(\bar{Y}^* - \tilde{Y}\right) \leq x\} - \mathcal{P}\{T^{1/2}\bar{X} \leq x\}\right| \geq \varepsilon\right\} = 0.$$

Proof of Theorem A.2: Let q_{tT}, q^*_{tT}, $\tilde{q}$ and $\tilde{q}^*$ defined as in the proof of Theorem 3.1. The result is proven if we are able to show the following steps; cf. Politis and Romano (1992b, Proof of Theorem 2). Step 1: $\bar{X} \xrightarrow{p} 0$. Step 2: $T^{1/2}\Sigma_\infty^{-1/2}\bar{X} \xrightarrow{d} N(0,1)$. Step 3: $\sup_{x\in\mathbb{R}}\left|\mathcal{P}\{T^{1/2}\bar{X} \leq x\} - \Phi_d(\Sigma_\infty^{-1/2}x)\right| \to 0$, where $\Phi_d(\cdot)$ is the c.d.f. of the standard d-variate normal distribution. Step 4: $Tvar^*[\tilde{q}^*] \xrightarrow{p} \Sigma_\infty$. Step 5:

$$\lim_{T\to\infty} \mathcal{P}\left\{\sup_{x\in\mathbb{R}^d}\left|\mathcal{P}^*\{\text{var}^*[\tilde{q}^*]^{-1/2}(\tilde{q}^* - \tilde{q}) \leq x\} - \Phi_d(x)\right| \geq \varepsilon\right\} = 0.$$

The proofs of results 1-4 are analogous to the proofs of results 1-4 in Theorem 3.1 As pointed out by Cattaneo et al. (2010) to prove 5 we need to show that

$$\lim_{T\to\infty} \mathcal{P}\left\{\sup_{\lambda\in\Lambda_d}\sup_{x\in\mathbb{R}}\left|\mathcal{P}^*\{m_T^{1/2}\frac{\lambda'(\tilde{q}^* - \tilde{q})}{\text{var}^*[\lambda' q^*_{1T}]^{1/2}} \leq x\} - \Phi(x)\right| \geq \varepsilon\right\} = 0.$$

where $\Lambda_d = \{\lambda \in \mathbb{R}^d : \lambda'\lambda = 1\}$. Let $\bar{\Lambda}_d = \{\lambda \in \mathbb{R}^d : \lambda'\lambda \leq 1\}$ and note that $\Lambda_d \subset \bar{\Lambda}_d$.

Given the sample, the bootstrap observations are independent. Hence we can apply Berry-Esséen inequality. Thus

$$\begin{aligned}
\sup_{\lambda\in\Lambda_d}\sup_{x\in\mathbb{R}}\left|\mathcal{P}^*\{m_T^{1/2}\frac{\lambda'\tilde{q}^* - \lambda'\tilde{q}}{\text{var}^*[\lambda' q^*_{1T}]^{1/2}} \leq x] - \Phi(x)\right| &\leq \sup_{\lambda\in\Lambda_d}\frac{C}{m_T^{1/2}}\text{E}^*[(\frac{\lambda'\left|q^*_{1T} - \tilde{q}\right|}{var^*[\lambda' q^*_{1T}]^{1/2}})^3] \\
&= \sup_{\lambda\in\Lambda_d}\frac{C}{m_T^{1/2}}\text{var}^*[\lambda' q^*_{1T}]^{-3/2}\text{E}^*[\left|\lambda' q^*_{1T} - \lambda'\tilde{q}\right|^3] \\
&\leq \frac{C}{\inf_{\lambda\in\Lambda_d}\text{var}^*[\lambda' q^*_{1T}]^{3/2}}\sup_{\lambda\in\bar{\Lambda}_d}\frac{S_T^{1/2}}{T^{1/2}}\text{E}^*[\left|\lambda' q^*_{1T} - \lambda'\tilde{q}\right|^3].
\end{aligned}$$

Now for fixed λ we have

$$\begin{aligned}
\frac{S_T^{1/2}}{T^{1/2}}\text{E}^*[\left|\lambda' q^*_{1T} - \lambda'\tilde{q}\right|^3] &= \frac{S_T^{1/2}}{T^{1/2}}\frac{1}{T}\sum\nolimits_{t=1}^{T}\left|\lambda' q_{tT} - \lambda'\tilde{q}\right|^3 \\
&= o_p(1)
\end{aligned}$$

as in A.7. Since $\bar{\Lambda}_d$ is a is compact and convex and since $|.|^3$ is a convex function we can apply Pollard (1991,p.187) Convexity Lemma to strength pointwise convergence to uniform convergence and therefore $\sup_{\lambda\in\bar{\Lambda}_d} S_T^{1/2}T^{-1/2}\text{E}^*[\left|\lambda' q^*_{1T} - \lambda'\tilde{q}\right|^3] = o_p(1)$, using also the fact that $\text{E}[\sup_{\lambda\in\Lambda_d}|\lambda' X_t|^{4v}] \leq \text{E}[|||X_t|||^{4v}] < \Delta$ by CS.

Additionally, by Lemma A.3 of Smith (2011) we have

$$\begin{aligned}
\inf_{\lambda\in\Lambda_d}\frac{1}{T}\sum\nolimits_{t=1}^{T}\left|\lambda' q_{tT} - \lambda'\tilde{q}\right|^2 &= \inf_{\lambda\in\Lambda_d}\lambda'(\frac{1}{T}\sum\nolimits_{t=1}^{T}(q_{tT} - \tilde{q})(q_{tT} - \tilde{q})')\lambda \\
&= \inf_{\lambda\in\Lambda_d}\lambda'\Sigma_\infty\lambda + o_p(1) = \inf_{\lambda\in\Lambda_d}\lambda' QPQ'\lambda + o_p(1) \\
&= \inf_{\lambda\in\Lambda_d}\lambda' P\lambda + o_p(1) > p_{\min} + o_p(1)
\end{aligned}$$

where P is a diagonal matrix of eigenvectors of Σ_∞ and Q is the corresponding orthonormal matrix of eigenvectors and $p_{\min} > 0$ is the smallest eigenvalue of P. Hence the result follows. ■

Let $\bar{Y} \equiv \frac{1}{T}\sum_{t=1}^{T} Y_{tT}$ and $\tilde{Y} \equiv \sum_{t=1}^{T} Y_{tT}p_{tT}$

Assumption A.3 **(a)** *The finite dimensional stochastic process* $\{X_t\}_{t=1}^\infty$ *is stationary and ergodic;* **(b)** $\text{E}[|X_t|^\tau] < \infty$ *for some* $\tau \geq 1$*;* **(c)** $T^{1/\tau}/m_T = o(1)$.

Lemma A.5 *Let the both A.3, 3.2, 3.3* **(a)**, *Then*

$$\bar{Y}^* - \bar{Y} \to 0,\ prob\text{-}\mathcal{P}^*,\ prob\text{-}\mathcal{P}, \tag{A.8}$$

$$\bar{Y}^* - \tilde{Y} \to 0,\ prob\text{-}\mathcal{P}^*,\ prob\text{-}\mathcal{P} \tag{A.9}$$

Proof: If we prove $(A.9)$, $(A.8)$ follows from this result and the fact that

$$\begin{aligned}\left|\bar{Y}-\tilde{Y}\right| &= \left|\frac{1}{T}\sum\nolimits_{t=1}^{T}Y_{tT}-\sum\nolimits_{t=1}^{T}Y_{tT}p_{tT}\right| \\ &= \left|\frac{1}{T}\sum\nolimits_{t=1}^{T}Y_{tT}(1-Tp_{tT})\right| \\ &\leq \left|\frac{1}{T}\sum\nolimits_{t=1}^{T}Y_{tT}\right|\max_{t}|1-Tp_{tT}|\xrightarrow{p}0\end{aligned}$$

by the ergodic theorem and the fact that $\max_{1\leq t\leq T}|Tp_{tT}|=1+o_p(1)$. First note that

$$\begin{aligned}\mathrm{E}^*[\left|Y_{jT}^*\right|] &= \sum\nolimits_{t=1}^{T}p_{tT}\left|\frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1}k(\frac{s}{S_T})X_{t-s}\right| \\ &\leq (1+o_p(1))\frac{1}{T}\sum\nolimits_{t=1}^{T}\left|\frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1}k(\frac{s}{S_T})X_{t-s}\right| \\ &= (1+o_p(1))\frac{1}{T}\sum\nolimits_{t=1}^{T}\left|\frac{1}{S_T}\sum\nolimits_{-j=-T}^{-1}k(\frac{t-j}{S_T})X_j\right| \\ &= (1+o_p(1))\frac{1}{T}\sum\nolimits_{t=1}^{T}\left|\frac{1}{S_T}\sum\nolimits_{j=1}^{T}k(\frac{t-j}{S_T})X_j\right| \\ &\leq (1+o_p(1))\frac{1}{T}\sum\nolimits_{t=1}^{T}\frac{1}{S_T}\sum\nolimits_{j=1}^{T}\left|k(\frac{t-j}{S_T})\right||X_j| \\ &= (1+o_p(1))\frac{1}{T}\sum\nolimits_{j=1}^{T}|X_j|\frac{1}{S_T}\sum\nolimits_{s=1-j}^{T-j}\left|k(\frac{s}{S_T})\right|\end{aligned}$$

by Smith (2011, equation, (A.4) we have

$$\frac{1}{S_T}\sum\nolimits_{s=1-j}^{T-j}\left|k(\frac{s}{S_T})\right|=O(1)$$

uniformly in j. Also by the ergodic theorem (White, 1999, Theorem 2.34) $\sum\nolimits_{j=1}^{T}|X_j|/T=O_p(1)$. Thus $\mathrm{E}^*[\left|Y_{jT}^*\right|]=O_p(1)$. In addition by T

$$\begin{aligned}\left|\sum\nolimits_{t=1}^{T}|Y_{tT}|\hat{\pi}_t-\sum\nolimits_{t=1}^{T}|Y_{tT}|p_{tT}\mathrm{I}(|Y_{tT}|<\delta m_T)\right| &\leq (1+o_p(1))\frac{1}{T}\sum\nolimits_{t=1}^{T}|Y_{tT}|\,\mathrm{I}(|Y_{tT}|\geq\delta m_T) \\ &\leq (1+o_p(1))\frac{1}{T}\sum\nolimits_{t=1}^{T}|Y_{tT}|\max_{t}\mathrm{I}(|Y_{tT}|\geq\delta m_T)\end{aligned}$$

Now by M

$$\max_{t}|Y_{tT}|=O(1)\max_{t}|X_t|=O_p(T^{1/\tau}).$$

Since $T^{1/\tau}/m_T=o(1)$ it follows that $\max_t\mathrm{I}(|Y_{tT}|\geq\delta m_T)=o_p(1)$. Thus

$$\frac{1}{T}\sum\nolimits_{t=1}^{T}|Y_{tT}|\,\mathrm{I}(|Y_{tT}|\geq\delta m_T)=o_p(1).$$

The remaining part of the proof is similar to the proof of Khinchine's weak law of large numbers given in Rao (2002). Define a pair of new random variables for each T, $(t=1,...,m_T)$,

$$\begin{aligned}W_{tT} &= Y_{tT}^*, Z_{tT}=0 \text{ if } |Y_{tT}^*|<\delta m_T, \\ W_{tT} &= 0, Z_{tT}=Y_{tT}^* \text{ if } |Y_{tT}^*|\geq\delta m_T.\end{aligned}$$

Hence $Y_{tT}^*=W_{tT}+Z_{tT}$. Define

$$\begin{aligned}\mu_T &= \mathrm{E}^*[W_{tT}] \\ &= \sum\nolimits_{t=1}^{T}p_{tT}Y_{tT}\mathrm{I}[|Y_{tT}|<\delta m_T].\end{aligned}$$

Note that $\tilde{Y}=\mathrm{E}^*[\left|Y_{tT}^*\right|]$ and $\left|\tilde{Y}-\mu_T\right|<\varepsilon$ for any $\varepsilon>0$ and T large enough. The latter claim holds since by T

$$\begin{aligned}\left|\sum\nolimits_{t=1}^{T}p_{tT}Y_{tT}I[|w_{tT}|<\delta m_T]-\sum\nolimits_{t=1}^{T}p_{tT}Y_{tT}\right| &\leq (1+o_p(1))\frac{1}{T}\sum\nolimits_{t=1}^{T}|Y_{tT}|\,I(|Y_{tT}|\geq\delta m_T) \\ &= o_p(1).\end{aligned}$$

Now

$$\mathrm{var}^*[W_{tT}]=\mathrm{E}^*[W_{tT}^2]-\mu_T^2\leq\mathrm{E}^*[W_{tT}^2]\leq\delta m_T\mathrm{E}^*[|W_{tT}|].$$

Thus, writing $\bar{W}=\sum\nolimits_{t=1}^{m_T}W_{tT}/m_T$, using C,

$$\begin{aligned}\mathcal{P}^*\{\left|\bar{W}-\mu_T\right| &\geq \varepsilon\}\leq\frac{\mathrm{var}^*[W_{tT}]}{\varepsilon^2 m_T} \\ &\leq \frac{\delta\mathrm{E}^*[|W_{tT}|]}{\varepsilon^2}.\end{aligned}$$

Hence, since $\left|\tilde{Y}-\mu_T\right|<\varepsilon$ for any $\varepsilon>0$ and T large enough,

$$\mathcal{P}^*\{\left|\bar{W}-\tilde{Y}\right|\geq 2\varepsilon\}\leq \frac{\delta \mathrm{E}^*[|W_{tT}|]}{\varepsilon^2}. \tag{A.10}$$

Now by M it follows that

$$\begin{aligned}\mathcal{P}^*\{Z_{tT} &\neq 0\}=\mathcal{P}^*\{|Y^*_{tT}|\geq \delta m_T\}\\ &\leq \frac{1}{\delta m_T}\mathrm{E}^*[|Y^*_{tT}|\,\mathrm{I}[|Y^*_{tT}|\geq \delta m_T]]\leq \frac{\delta}{m_T}.\end{aligned}$$

w.p.a. To see this, as $\mathrm{E}^*[\left|Y^*_{tT}\right|]=O_p(1)$, it follows that $\mathrm{E}^*[\left|Y^*_{tT}\right| I[\left|Y^*_{tT}\right|\geq \delta m_T]]=o_p(1)$. Thus, we can always choose a constant δ^2 such that for T large enough $\mathrm{E}^*[\left|Y^*_{tT}\right| I[\left|Y^*_{tT}\right|\geq \delta m_T]]\leq \delta^2$ w.p.a.1. Write $\bar{Z}=\sum_{t=1}^{m_T} Z_{tT}/m_T$. Note that

$$\begin{aligned}\mathcal{P}^*\{\bar{Z} &\neq 0\}\leq \mathcal{P}^*\{\max_t Z_{tT}\neq 0\}\\ &\leq \sum_{t=1}^{m_T}\mathcal{P}^*\{Z_{tT}\neq 0\}\leq \delta.\end{aligned} \tag{A.11}$$

From eqs. (A.10) and (A.11)

$$\begin{aligned}\mathcal{P}^*\{\left|\bar{Y}^*-\tilde{Y}\right| &\geq 4\varepsilon\}=\mathcal{P}^*\{\left|\bar{W}-\tilde{Y}+\bar{Z}\right|\geq 4\varepsilon\}\\ &\leq \mathcal{P}^*\{\left|\bar{W}-\tilde{Y}\right|+\left|\bar{Z}\right|\geq 4\varepsilon\}\\ &\leq \mathcal{P}^*\{\left|\bar{W}-\tilde{Y}\right|\geq 2\varepsilon\}+\mathcal{P}^*\{\left|\bar{Z}\right|\geq 2\varepsilon\}\\ &\leq \frac{\delta \mathrm{E}^*[|W_{tT}|]}{\varepsilon^2}+\mathcal{P}^*\{\left|\bar{Z}\right|\neq 0\}=\frac{\delta \mathrm{E}^*[|W_{tT}|]}{\varepsilon^2}+\delta.\end{aligned}$$

Now choose δ small enough. As $\mathrm{E}^*[|W_{tT}|]\leq \mathrm{E}^*[\left|Y^*_{tT}\right|]=O_p(1)$, the result follows from M. ■

The following Theorem is due to Ranga Rao [see Wooldridge, 1994]

Theorem A.3 *Let $\Theta\subset R^p$, let $\{X_t\in\mathbb{X}:t=1,2,...,\}$ be a sequence of stationary and ergodic $m\times 1$ random vectors with and let $f_t:\mathbb{X}\times\Theta\to\mathbb{R}$ be a real valued function. Assume that (a) Θ is compact., (b) for each θ, $f(.,\theta)$ is measurable and for each $X_t\in\mathbb{X}$, $f(x_t,.)$ is continuous on Θ; (c) $\mathrm{E}\left[\sup_{\theta\in\Theta}|f(X_t,\theta)|\right]<\infty$ then*

$$\sup_{\theta\in\Theta}\left|\frac{1}{T}\sum_{t=1}^{T} f(X_t,\theta)-\mathrm{E}[f(X_t,\theta)]\right|=o_p(1).$$

The following Lemma corresponds to a weak uniform law of large numbers for kernel block bootstrapped sequences.

Lemma A.6 *Let $\{X_t\in\mathbb{X}:t=1,2,...,\}$ be a sequence of stationary and ergodic $m\times 1$ random vectors and let*

$$q_{tT}(\theta)=\frac{1}{S_T}\sum_{s=t-T}^{t-1} k(\frac{s}{S_T})g(X_{t-s},\theta), \tag{A.12}$$

and consider the sample $q_{tT}(\theta)$, $(t=1,...,T)$. Draw a random sample of size m_T with replacement from $q_{tT}(\theta)$, $(t=1,...,T)$, to obtain the bootstrap sample $q^*_{tT}(\theta)$, $(t=1,...,m_T)$ where $\mathcal{P}(q^*_{tT}(\theta)=q_{tT}(\theta))=p_{tT}$ for $s=1,...,m_T$ and $t=1,...,T$. Assume that 3.2, 3.3 **(a)** *hold and that :* **(a)** *Bootstrap Pointwise Weak Law of Large Numbers. for each fixed $\theta\in\Theta\subset\mathbb{R}^p$, Θ a compact set,*

$$\frac{1}{m_T}\sum_{t=1}^{m_T} q^*_{tT}(\theta)-\sum_{t=1}^{T} q_{tT}(X_t,\theta)p_{tT}\to 0,\ \textit{prob-}\mathcal{P}^*,\ \textit{prob-}\mathcal{P};$$

(b) *Uniform Convergence:*

$$\sup_{\theta\in\Theta}\left|\frac{1}{T}\sum_{t=1}^{T} q_{tT}(\theta)-\frac{1}{T}\sum_{t=1}^{T} g(X_t,\theta)\right|\xrightarrow{p} 0,$$
$$\mathrm{E}[\sup_{\theta\in\Theta}|g(X_t,\theta)|]\leq\Delta$$

(c) *or each θ, $g(.,\theta)$ is measurable and for each $x_t\in X$, $g(x_t,.)$ is continuous on Θ. Then, as $m_T\to\infty$ and $S_T=o_p(T^{1/2})$, for any $\delta>0$ and $\xi>0$*

$$\begin{aligned}\lim_{T\to\infty}\mathcal{P}\{\mathcal{P}^*\{\sup_{\theta\in\Theta}\left|\frac{1}{m_T}\sum_{t=1}^{m_T} q^*_{tT}(\theta)-\frac{1}{T}\sum_{t=1}^{T} g(X_t,\theta)\right| &> \delta\}>\xi\}=0,\\ \lim_{T\to\infty}\mathcal{P}\{\mathcal{P}^*\{\sup_{\theta\in\Theta}\left|\frac{1}{m_T}\sum_{t=1}^{m_T} q^*_{tT}(\theta)-\sum_{t=1}^{T} q_{tT}(\theta)p_{tT}\right| &> \delta\}>\xi\}=0.\end{aligned}$$

Proof: First write

$$A_T = \mathcal{P}^*\{\sup_{\theta\in\Theta}\left|\frac{1}{m_T}\sum\nolimits_{t=1}^{m_T} q^*_{tT}(\theta) - \frac{1}{T}\sum\nolimits_{t=1}^{T} g(X_t,\theta)\right| > \delta\}$$

and by M $\mathcal{P}\{A_T > \delta\} \leq \delta^{-1}\mathrm{E}[A_T]$. Note that the Lebesgue convergence theorem is valid for sequences that converge in probability by Proposition 20 of Royden (1988, p.96.). Therefore as $A_T \leq 1$ the result follows from this theorem if we show that $A_T \xrightarrow{p} 0$.

The proof is similar to the proof of a standard UWL (eg. Amemya, 1985).First note that

$$\begin{aligned}\left|\frac{1}{m_T}\sum\nolimits_{t=1}^{m_T} q^*_{tT}(\theta) - \frac{1}{T}\sum\nolimits_{t=1}^{T} g(X_t,\theta)\right| &\leq \left|\frac{1}{m_T}\sum\nolimits_{t=1}^{m_T} q^*_{tT}(\theta) - \sum\nolimits_{t=1}^{T} q_{tT}(\theta)p_{tT}\right| \\ &+\left|\frac{1}{T}\sum\nolimits_{t=1}^{T} g_t(\theta) - \sum\nolimits_{t=1}^{T} q_{tT}(\theta)p_{tT}\right|\end{aligned}$$

and that

$$\begin{aligned}\sup_{\theta\in\Theta}\left|\frac{1}{T}\sum\nolimits_{t=1}^{T} g(X_t,\theta) - \sum\nolimits_{t=1}^{T} q_{tT}(\theta)p_{tT}\right| &\leq \sup_{\theta\in\Theta}\left|\frac{1}{T}\sum\nolimits_{t=1}^{T} g(X_t,\theta) - \frac{1}{T}\sum\nolimits_{t=1}^{T} q_{tT}(\theta)\right| \\ &+\sup_{\theta\in\Theta}\left|\frac{1}{T}\sum\nolimits_{t=1}^{T} q_{tT}(\theta)Tp_{tT} - \frac{1}{T}\sum\nolimits_{t=1}^{T} q_{tT}(\theta)\right|.\end{aligned}$$

By Smith (Lemma A.1, 2004) we have

$$\sup_{\theta\in\Theta}\left|\frac{1}{T}\sum\nolimits_{t=1}^{T} g(X_t,\theta) - \frac{1}{T}\sum\nolimits_{t=1}^{T} q_{tT}(\theta)\right| = o_p(1). \tag{A.13}$$

Also

$$\begin{aligned}\sup_{\theta\in\Theta}\left|\frac{1}{T}\sum\nolimits_{t=1}^{T} q_{tT}(\theta)Tp_{tT} - \frac{1}{T}\sum\nolimits_{t=1}^{T} q_{tT}(\theta)\right| &\leq \sup_{\theta\in\Theta}\left|\frac{1}{T}\sum\nolimits_{t=1}^{T} q_{tT}(\theta)\right| \max_{1\leq t\leq T}|Tp_{tT}-1| \\ &= o_p(1)\end{aligned}$$

since

$$\sup_{\theta\in\Theta}\left|\frac{1}{T}\sum\nolimits_{t=1}^{T} q_{tT}(\theta)\right| \leq O(1)\frac{1}{T}\sum\nolimits_{t=1}^{T}\sup_{\theta\in\Theta}|g(x_t,\theta)| = O_p(1)$$

by the ergodic theorem ergodic theorem (White, 1999, Theorem 2.34) and the fact that $\max_{1\leq t\leq T}|Tp_{tT}| = 1 + o_p(1)$.We prove now that

$$\lim_{T\to\infty}\mathcal{P}\{\mathcal{P}^*\{\sup_{\theta\in\Theta}\left|\frac{1}{m_T}\sum\nolimits_{t=1}^{m_T} q^*_{tT}(\theta) - \sum\nolimits_{t=1}^{T} q_{tT}(\theta)p_{tT}\right| > \delta\} > \xi\} = 0.$$

Since Θ is compact it follows that there is a finite number of $\theta's$ for instance $\theta_1, \theta_2, ..., \theta_{n_\delta}$ such that $\Theta \subset \bigcup_{i=1}^{n}\Gamma(\theta_i,\delta)$ where $\Gamma(\theta_i,\delta)$ is an open ball with center θ_i and radius δ. Thus

$$\begin{aligned}\mathcal{P}^*\{\sup_{\theta\in\Theta}\left|\frac{1}{m_T}\sum\nolimits_{t=1}^{m_T} q^*_{tT}(\theta) - \sum\nolimits_{t=1}^{T} q_{tT}(\theta)p_{tT}\right| &> \delta\} \leq \\ \mathcal{P}^*\{\bigcup\nolimits_{i=1}^{n_\delta}\sup_{\theta\in\Gamma(\theta_i,\delta)}\left|\frac{1}{m_T}\sum\nolimits_{t=1}^{m_T} q^*_{tT}(\theta) - \sum\nolimits_{t=1}^{T} q_{tT}(\theta)p_{tT}\right| &> \delta\} \leq \\ \sum\nolimits_{i=1}^{n_\delta}\mathcal{P}^*\{\sup_{\theta\in\Gamma(\theta_i,\delta)}\left|\frac{1}{m_T}\sum\nolimits_{t=1}^{m_T} q^*_{tT}(\theta) - \sum\nolimits_{t=1}^{T} q_{tT}(\theta)p_{tT}\right| &> \delta\}.\end{aligned}$$

Now

$$\begin{gathered}\mathcal{P}^*\{\sup_{\theta\in\Gamma(\theta_i,\delta)}\left|\frac{1}{m_T}\sum\nolimits_{t=1}^{m_T} q^*_{tT}(\theta) - \sum\nolimits_{t=1}^{T} q_{tT}(\theta)p_{tT}\right| > \delta\} \leq \\ \mathcal{P}^*\{\left|\frac{1}{m_T}\sum\nolimits_{t=1}^{m_T} q^*_{tT}(\theta_i) - \sum\nolimits_{t=1}^{T} q_{tT}(\theta_i)p_{tT}\right| > \frac{\delta}{3}\} \\ +\mathcal{P}^*\{\sup_{\theta\in\Gamma(\theta_i,\delta)}\left|\frac{1}{m_T}\sum\nolimits_{t=1}^{m_T} q^*_{tT}(\theta) - \frac{1}{m_T}\sum\nolimits_{t=1}^{m_T} q^*_{tT}(\theta_i)\right| > \frac{\delta}{3}\} \\ \mathcal{P}\{\sup_{\theta\in\Gamma(\theta_i,\delta)}\left|\sum\nolimits_{t=1}^{T} q_{tT}(\theta)p_{tT} - \sum\nolimits_{t=1}^{T} q_{tT}(\theta_i)p_{tT}\right| > \frac{\delta}{3}\} \\ B_{1,T} + B_{2,T} + B_{3,T.}\end{gathered}$$

Now

$$\left|\frac{1}{m_T}\sum\nolimits_{t=1}^{m_T} q^*_{tT}(\theta_i) - \sum\nolimits_{t=1}^{T} p_{tT}q_{tT}(\theta_i)\right| = o_B(1)$$

by the KBB Law of large numbers. Thus $B_{1,T} = o_p(1)$.

By M

$$\begin{aligned}
&\mathcal{P}^*\{\sup_{\theta\in\Gamma(\theta_i,\delta)}\left|\frac{1}{m_T}\sum\nolimits_{t=1}^{m_T} q^*_{tT}(\theta) - \frac{1}{m_T}\sum\nolimits_{t=1}^{m_T} q^*_{tT}(\theta_i)\right| > \frac{\delta}{3}\} \\
&\leq \frac{3}{\delta}\mathrm{E}^*[\sup_{\theta\in\Gamma(\theta_i,\delta)}\left|\frac{1}{m_T}\sum\nolimits_{t=1}^{m_T} q^*_{tT}(\theta) - \frac{1}{m_T}\sum\nolimits_{t=1}^{m_T} q^*_{tT}(\theta_i)\right|] \\
&\quad \leq \frac{3}{\delta}\frac{1}{m_T}\sum\nolimits_{t=1}^{m_T}\mathrm{E}^*[\sup_{\theta\in\Gamma(\theta_i,\delta)}|q^*_{tT}(\theta) - q^*_{tT}(\theta_i)|] \\
&\quad = \frac{1}{\delta}\sum\nolimits_{t=1}^{T} p_{tT}\sup_{\theta\in\Gamma(\theta_i,\delta)}|q_{tT}(\theta) - q_{tT}(\theta_i)| \\
&\quad = \frac{1}{\delta}\frac{1}{T}\sum\nolimits_{t=1}^{T} Tp_{tT}\sup_{\theta\in\Gamma(\theta_i,\delta)}|q_{tT}(\theta) - q_{tT}(\theta_i)| \\
&\quad = (1+o_p(1))\frac{1}{\delta}\frac{1}{T}\sum\nolimits_{t=1}^{T}\sup_{\theta\in\Gamma(\theta_i,\delta)}|q_{tT}(\theta) - q_{tT}(\theta_i)|\,,
\end{aligned}$$

where the second inequality follows from T. But by M

$$\mathcal{P}(\frac{1}{T}\sum\nolimits_{t=1}^{T}\sup_{\theta\in\Gamma(\theta_i,\delta)}|q_{tT}(\theta) - q_{tT}(\theta_i)| > \epsilon) \leq \frac{1}{\epsilon T}\sum\nolimits_{t=1}^{T}\mathrm{E}(\sup_{\theta\in\Gamma(\theta_i,\delta)}|q_{tT}(\theta) - q_{tT}(\theta_i)|)$$

also

$$\begin{aligned}
\sup_{\theta\in\Gamma(\theta_i,\delta)}|q_{tT}(\theta) - q_{tT}(\theta_i)| &= \sup_{\theta\in\Gamma(\theta_i,\delta)}\left|\frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1} k(\frac{s}{S_T})(g(X_{t-s},\theta) - g(X_{t-s},\theta_i)\right| \\
&\leq \frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1}\left|k(\frac{s}{S_T})\right|\sup_{\theta\in\Gamma(\theta_i,\delta)}|(g(X_{t-s},\theta) - g(X_{t-s},\theta_i)|
\end{aligned}$$

by T. Now taking expectations we have

$$\begin{aligned}
&\mathrm{E}[\frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1}\left|k(\frac{s}{S_T})\right|\sup_{\theta\in\Gamma(\theta_i,\delta)}|(g(x_{t-s},\theta) - g(x_{t-s},\theta_i)|] \\
= &\frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1}\left|k(\frac{s}{S_T})\right|\mathrm{E}[\sup_{\theta\in\Gamma(\theta_i,\delta)}|(g(x_{t-s},\theta) - g(x_{t-s},\theta_i)|]
\end{aligned}$$

and $\frac{1}{S_T}\sum_{s=t-T}^{t-1}\left|k(\frac{s}{S_T})\right| \leq C$, $\mathrm{E}[\sup_{\theta\in\Gamma(\theta_i,\delta)}|(g(x_{t-s},\theta) - g(x_{t-s},\theta_i)|] \to 0$ by as $g(X_{t-s},\theta)$ is continuous and dominated convergence as $\delta \to 0$. Consequently we have

$$\begin{aligned}
\frac{1}{\epsilon T}\sum\nolimits_{t=1}^{T}\mathrm{E}(\sup_{\theta\in\Gamma(\theta_i,\delta)}|q_{tT}(\theta) - q_{tT}(\theta_i)|) &= \frac{1}{\epsilon T}\sum\nolimits_{t=1}^{T}\frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1}\left|k(\frac{s}{S_T})\right|\xi \\
&= o(1).
\end{aligned}$$

Thus $B_{2,T} = o_p(1)$.

Finally

$$\begin{aligned}
B_{3,T} &= \sup_{\theta\in\Gamma(\theta_i,\delta)}\left|\sum\nolimits_{t=1}^{T} p_{tT}q_{tT}(\theta) - \sum\nolimits_{t=1}^{T} p_{tT}q_{tT}(\theta_i))\right| \leq \frac{1}{T}\sum\nolimits_{t=1}^{T} Tp_{tT}\sup_{\theta\in\Gamma(\theta_i,\delta)}|q_{tT}(\theta) - q_{tT}(\theta_i)| \\
&\leq (1+o_p(1))\frac{1}{T}\sum\nolimits_{t=1}^{T}\sup_{\theta\in\Gamma(\theta_i,\delta)}|q_{tT}(\theta) - q_{tT}(\theta_i)|
\end{aligned}$$

by T and the result follows as above.

The second result follows from the first and $(A.13)$ and $(A.14)$.

■

Lemma A.7 *Let $\{X_t \in \mathbb{X} : t = 1, 2, ...,\}$ be a sequence of stationary and ergodic $m \times 1$ random vectors and let*

$$q_{tT}(\theta) = \frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1} k(\frac{s}{S_T})g(X_{t-s},\theta), \tag{A.14}$$

and consider the sample $q_{tT}(\theta)$, $(t = 1, ..., T)$. Draw a random sample of size m_T with replacement from $q_{tT}(\theta)$, $(t = 1, ..., T)$, to obtain the bootstrap sample $q^*_{tT}(\theta)$, $(t = 1, ..., m_T)$ where $\mathcal{P}(q^*_{tT}(\theta) = q_{tT}(\theta)) = p_{tT}$ for $s = 1, ..., m_T$ and $t = 1, ..., T$. Assume that 3.2, 3.3 **(a)** *hold and that :* **(a)** *Bootstrap Pointwise Weak Law of Large Numbers. for ,*

$$\frac{1}{m_T}\sum\nolimits_{t=1}^{m_T} q^*_{tT}(\theta_0) - \sum\nolimits_{t=1}^{T} q_{tT}(\theta_0)p_{tT} \to 0,\ prob\text{-}\mathcal{P}^*,\ prob\text{-}\mathcal{P};$$

(b) $\mathrm{E}[\sup_{\theta\in\mathcal{N}}|g(X_t,\theta)|]\leq\Delta$ *where* $\mathcal{N}$ *is a neighbourhood of* θ_0. **(c)** *or each* θ, $g(.,\theta)$ *is measurable and for each* $x_t\in X$, $g(x_t,.)$ *is continuous on* Θ. *Then,*

$$\sup_{\theta\in\mathcal{N}}\left|\frac{1}{T}\sum_{t=1}^{T}g(X_t,\theta)-\mathrm{E}[g(X_t,\theta)]\right|\xrightarrow{p}0$$

and as $m_T\to\infty$ *and* $S_T=o_p(T^{1/2})$, *for any* $\delta>0$ *and* $\xi>0$

$$\lim_{T\to\infty}\mathcal{P}\{\mathcal{P}^*\{\sup_{\theta\in\mathcal{N}}\left|\frac{1}{m_T}\sum_{t=1}^{m_T}q_{tT}^*(\theta)-\sum_{t=1}^{T}q_{tT}(\theta)p_{tT}\right|>\delta\}>\xi\}=0,$$
$$\lim_{T\to\infty}\mathcal{P}\{\mathcal{P}^*\{\sup_{\theta\in\mathcal{N}}\left|\frac{1}{m_T}\sum_{t=1}^{m_T}q_{tT}^*(\theta)-\frac{1}{T}\sum_{t=1}^{T}g(X_t,\theta)\right|>\delta\}>\xi\}=0.$$

Proof: Let $\mathcal{N}=\Gamma(\theta_0,\delta)$, where $\Gamma(\theta_0,\delta)$ is an open ball with center θ_0 and radius δ. First note that

$$\begin{aligned}\sup_{\theta\in\Gamma(\theta_0,\delta)}\left|\frac{1}{T}\sum_{t=1}^{T}g(X_t,\theta)-\mathrm{E}[g(X_t,\theta)]\right| &\leq \sup_{\theta\in\Gamma(\theta_0,\delta)}\left|\frac{1}{T}\sum_{t=1}^{T}g(X_t,\theta)-\frac{1}{T}\sum_{t=1}^{T}g(X_t,\theta_0)\right|\\ &+\left|\frac{1}{T}\sum_{t=1}^{T}g(X_t,\theta_0)-\mathrm{E}[g(X_t,\theta_0)]\right|\\ &+\sup_{\theta\in\Gamma(\theta_0,\delta)}|\mathrm{E}[g(X_t,\theta_0)]-\mathrm{E}[g(X_t,\theta)]|\end{aligned}$$

$$\mathcal{P}\{\sup_{\theta\in\Gamma(\theta_0,\delta)}\left|\frac{1}{T}\sum_{t=1}^{T}g(X_t,\theta)-\frac{1}{T}\sum_{t=1}^{T}g(X_t,\theta_0)\right|\}>\varepsilon\leq\frac{1}{\varepsilon}\mathrm{E}[\sup_{\theta\in\Gamma(\theta_0,\delta)}|g(X_t,\theta)-g(X_t,\theta_0)|]$$

Now by T

$$\mathrm{E}[\sup_{\theta\in\Gamma(\theta_0,\delta)}|g(X_t,\theta)-g(X_t,\theta_0)|]\leq 2\mathrm{E}[\sup_{\theta\in\Gamma(\theta_0,\delta)}|g(X_t,\theta)|]$$

Thus by the Dominated convergence theorem and continuity of $g(X_t,\theta)$ thus as $\delta\to0$ we have

$$\lim_{\delta\to0}E[\sup_{\theta\in\Gamma(\theta_0,\delta)}|g(X_t,\theta)-g(X_t,\theta_0)|]=0.$$

Let us consider now

$$\sup_{\theta\in\Gamma(\theta_0,\delta)}\left|\frac{1}{T}\sum_{t=1}^{T}q_{tT}(\theta)-\frac{1}{T}\sum_{t=1}^{T}g(X_t,\theta)\right|\xrightarrow{p}0$$

Note that

$$\begin{aligned}\sup_{\theta\in\Gamma(\theta_0,\delta)}\left|\frac{1}{T}\sum_{t=1}^{T}q_{tT}(\theta)-\frac{1}{T}\sum_{t=1}^{T}g(X_t,\theta)\right| &\leq \sup_{\theta\in\Gamma(\theta_0,\delta)}\left|\frac{1}{T}\sum_{t=1}^{T}g(X_t,\theta)-E[g(X_t,\theta)]\right|+\\ &\quad\sup_{\theta\in\Gamma(\theta_0,\delta)}\left|\frac{1}{T}\sum_{t=1}^{T}q_{tT}(\theta)-E[g(X_t,\theta)]\right|\\ &= A_{1,T}+A_{2,T}.\end{aligned}$$

The proofs that $A_{1,T}\xrightarrow{p}0$ was proven before and the proof that $A_{2,T}\xrightarrow{p}0$ is identical to the proof of Lemma A.1 of Smith (2011) [and uses the fact that $A_{1,T}\xrightarrow{p}0$]

We prove now that

$$\lim_{T\to\infty}\mathcal{P}\{\mathcal{P}^*\{\sup_{\theta\in\mathcal{N}}\left|\frac{1}{m_T}\sum_{t=1}^{m_T}q_{tT}^*(\theta)-\sum_{t=1}^{T}q_{tT}(\theta)p_{tT}\right|>\delta\}>\xi\}=0.$$

Note that

$$\begin{aligned}&\mathcal{P}^*\{\sup_{\theta\in\Gamma(\theta_0,\delta)}\left|\frac{1}{m_T}\sum_{t=1}^{m_T}q_{tT}^*(\theta)-\sum_{t=1}^{T}q_{tT}(\theta)p_{tT}\right|>\delta\}\leq\\ &\mathcal{P}^*\{\left|\frac{1}{m_T}\sum_{t=1}^{m_T}q_{tT}^*(\theta_0)-\sum_{t=1}^{T}q_{tT}(\theta_0)p_{tT}\right|>\frac{\delta}{3}\}\\ &+\mathcal{P}^*\{\sup_{\theta\in\Gamma(\theta_0,\delta)}\left|\frac{1}{m_T}\sum_{t=1}^{m_T}q_{tT}^*(\theta)-\frac{1}{m_T}\sum_{t=1}^{m_T}q_{tT}^*(\theta_0)\right|>\frac{\delta}{3}\}\\ &\mathcal{P}\{\sup_{\theta\in\Gamma(\theta_0,\delta)}\left|\sum_{t=1}^{T}q_{tT}(\theta)p_{tT}-\sum_{t=1}^{T}q_{tT}(\theta_0)p_{tT})\right|>\frac{\delta}{3}\}\\ &C_{1,T}+C_{2,T}+C_{3,T}.\end{aligned}$$

Now

$$\left|\frac{1}{m_T}\sum_{t=1}^{m_T}q_{tT}^*(\theta_0)-\sum_{t=1}^{T}p_{tT}q_{tT}(\theta_0)\right|=o_B(1)$$

by the KBB Law of large numbers. Thus $C_{1,T} = o_p(1)$.

By M

$$\begin{aligned}
&\mathcal{P}^*\{\sup_{\theta\in\Gamma(\theta_0,\delta)}\left|\frac{1}{m_T}\sum\nolimits_{t=1}^{m_T} q_{tT}^*(\theta) - \frac{1}{m_T}\sum\nolimits_{t=1}^{m_T} q_{tT}^*(\theta_0)\right| > \frac{\delta}{3}\} \\
&\leq \frac{3}{\delta}\mathrm{E}^*[\sup_{\theta\in\Gamma(\theta_0,\delta)}\left|\frac{1}{m_T}\sum\nolimits_{t=1}^{m_T} q_{tT}^*(\theta) - \frac{1}{m_T}\sum\nolimits_{t=1}^{m_T} q_{tT}^*(\theta_0)\right|] \\
&\leq \frac{3}{\delta}\frac{1}{m_T}\sum\nolimits_{t=1}^{m_T}\mathrm{E}^*[\sup_{\theta\in\Gamma(\theta_0,\delta)}|q_{tT}^*(\theta) - q_{tT}^*(\theta_0)|] \\
&= \frac{3}{\delta}\sum\nolimits_{t=1}^{T} p_{tT}\sup_{\theta\in\Gamma(\theta_0,\delta)}|q_{tT}(\theta) - q_{tT}(\theta_0)| \\
&= \frac{3}{\delta}\frac{1}{T}\sum\nolimits_{t=1}^{T} Tp_{tT}\sup_{\theta\in\Gamma(\theta_0,\delta)}|q_{tT}(\theta) - q_{tT}(\theta_0)| \\
&= (1+o_p(1))\frac{1}{\delta}\frac{1}{T}\sum\nolimits_{t=1}^{T}\sup_{\theta\in\Gamma(\theta_0,\delta)}|q_{tT}(\theta) - q_{tT}(\theta_0)|,
\end{aligned}$$

where the second inequality follows from T. But

$$\mathcal{P}(\frac{1}{T}\sum\nolimits_{t=1}^{T}\sup_{\theta\in\Gamma(\theta_0,\delta)}|q_{tT}(\theta) - q_{tT}(\theta_0)| > \epsilon) \leq \frac{1}{\epsilon T}\sum\nolimits_{t=1}^{T}\mathrm{E}(\sup_{\theta\in\Gamma(\theta_0,\delta)}|q_{tT}(\theta) - q_{tT}(\theta_0)|)$$

also

$$\begin{aligned}
\sup_{\theta\in\Gamma(\theta_0,\delta)}|q_{tT}(\theta) - q_{tT}(\theta_i)| &= \sup_{\theta\in\Gamma(\theta_0,\delta)}\left|\frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1} k(\frac{s}{S_T})(g(X_{t-s},\theta) - g(X_{t-s},\theta_0)\right| \\
&\leq \frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1}\left|k(\frac{s}{S_T})\right|\sup_{\theta\in\Gamma(\theta_0,\delta)}|(g(X_{t-s},\theta) - g(X_{t-s},\theta_0)|
\end{aligned}$$

by T. Now taking expectations we have

$$\begin{aligned}
&\mathrm{E}[\frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1}\left|k(\frac{s}{S_T})\right|\sup_{\theta\in\Gamma(\theta_0,\delta)}|(g(x_{t-s},\theta) - g(x_{t-s},\theta_0)|] \\
&= \frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1}\left|k(\frac{s}{S_T})\right|\mathrm{E}[\sup_{\theta\in\Gamma(\theta_0,\delta)}|(g(x_{t-s},\theta) - g(x_{t-s},\theta_0)|]
\end{aligned}$$

and $\frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1}\left|k(\frac{s}{S_T})\right| \leq C$, $\mathrm{E}[\sup_{\theta\in\Gamma(\theta_0,\delta)}|(g(x_{t-s},\theta) - g(x_{t-s},\theta_0)|] \to 0$ as $g(X_{t-s},\theta)$ is continuous and dominated convergence as $\delta \to 0$. Consequently we have

$$\begin{aligned}
\frac{1}{\epsilon T}\sum\nolimits_{t=1}^{T}\mathrm{E}(\sup_{\theta\in\Gamma(\theta_0,\delta)}|q_{tT}(\theta) - q_{tT}(\theta_0)|) &= \frac{1}{\epsilon T}\sum\nolimits_{t=1}^{T}\frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1}\left|k(\frac{s}{S_T})\right|\xi \\
&= o(1).
\end{aligned}$$

Thus the result follows.

The second result follows from the fact that

$$\begin{aligned}
\sup_{\theta\in\Gamma(\theta_0,\delta)}\left|\frac{1}{m_T}\sum\nolimits_{t=1}^{m_T} q_{tT}^*(\theta) - \frac{1}{T}\sum\nolimits_{t=1}^{T} g(X_t,\theta)\right| &\leq \sup_{\theta\in\Gamma(\theta_0,\delta)}\left|\frac{1}{m_T}\sum\nolimits_{t=1}^{m_T} q_{tT}^*(\theta) - \sum\nolimits_{t=1}^{T} q_{tT}(\theta)p_{tT}\right| \\
&+ \sup_{\theta\in\Gamma(\theta_0,\delta)}\left|\frac{1}{T}\sum\nolimits_{t=1}^{T} g(X_t,\theta) - \sum\nolimits_{t=1}^{T} q_{tT}(\theta)p_{tT}\right|.
\end{aligned}$$

The first term of the RHS was shown to converge to zero

$$\begin{aligned}
\sup_{\theta\in\Gamma(\theta_0,\delta)}\left|\frac{1}{T}\sum\nolimits_{t=1}^{T} g(X_t,\theta) - \sum\nolimits_{t=1}^{T} q_{tT}(\theta)p_{tT}\right| &\leq \sup_{\theta\in\Gamma(\theta_0,\delta)}\left|\frac{1}{T}\sum\nolimits_{t=1}^{T} q_{tT}(\theta) - \sum\nolimits_{t=1}^{T} q_{tT}(\theta)p_{tT}\right| \\
&+ \sup_{\theta\in\Gamma(\theta_0,\delta)}\left|\frac{1}{T}\sum\nolimits_{t=1}^{T} g(X_t,\theta) - \frac{1}{T}\sum\nolimits_{t=1}^{T} q_{tT}(\theta)\right| \\
&\leq \sup_{\theta\in\Gamma(\theta_0,\delta)}\left|\frac{1}{T}\sum\nolimits_{t=1}^{T} q_{tT}(\theta)\right|\max_{1\leq t\leq T}|1 - Tp_{tT}| + o_p(1).
\end{aligned}$$

As $A_{1,T} + A_{2,T} \xrightarrow{p} 0$. Now $\sup_{\theta\in\Gamma(\theta_0,\delta)}\left|\frac{1}{T}\sum_{t=1}^{T} q_{tT}(\theta)\right| = O_p(1)$ and $\max_{1\leq t\leq T}|1 - Tp_{tT}| = o_p(1)$. Hence the result follows. ■

Lemma A.8 *If the finite dimensional stochastic process* $\{X_t\}_{t=1}^{\infty}$ *satisfy assumptions 3.1, 2.5 and 3.3 (a) hold and if* $m_T = T/S_T$*, and* $S_T = o(T^{1/2})$ *and if* $E[X_t] = 0$*,*

$$\lim_{n\to\infty} \mathcal{P}[\mathcal{P}^*(|\frac{S_T}{m_T k_2}\sum_{t=1}^{m_T} Y_{tT}^{*2} - \frac{S_T}{Tk_2}\sum_{t=1}^{T} Y_{tT}^2| \quad > \quad \varepsilon) > \delta] = 0,$$

$$\lim_{n\to\infty} \mathcal{P}[\mathcal{P}^*(|\frac{S_T}{m_T k_2}\sum_{t=1}^{m_T} Y_{tT}^{*2} - \frac{S_T}{k_2}\sum_{t=1}^{T} Y_{tT}^2 p_{tT}| \quad > \quad \varepsilon) > \delta] = 0.$$

Proof: The result is proved if we show that

$$\mathcal{P}^*(|\frac{S_T}{m_{Tk_2}}\sum_{t=1}^{m_T} Y_{tT}^{*2} - \frac{S_T}{Tk_2}\sum_{t=1}^{T} Y_{tT}^2| > \varepsilon) = o_p(1).$$

Note that

$$\left|\frac{S_T}{k_2}\sum_{t=1}^{T} Y_{tT}^2 p_{tT} - \frac{S_T}{Tk_2}\sum_{t=1}^{T} Y_{tT}^2\right| = \max_t |Tp_{tT} - 1| \left|\frac{S_T}{Tk_2}\sum_{t=1}^{T} Y_{tT}^2\right|.$$

Thus as $\left|\frac{S_T}{Tk_2}\sum_{t=1}^{T} Y_{tT}^2\right| = O_p(1)$ by Lemma A.3 of Smith (2004, pA.4) and $\max_{1\le t\le T}|Tp_{tT} - 1| = o_p(1)$ by assumption.

Hence the result follows by T if we show that

$$\mathcal{P}^*\left(\left|\frac{S_T}{m_{Tk_2}}\sum_{t=1}^{m_T} Y_{tT}^{*2} - \frac{S_T}{k_2}\sum_{t=1}^{T} Y_{tT}^2 p_{tT}\right| > \varepsilon\right) = o_p(1).$$

The proof of this result is similar to that of Lemma B.2 of Gonçalves and White (2004).First note that $\mathrm{E}^*\left[Y_{tT}^{*2}\right] = \sum_{t=1}^{T} p_{tT} Y_{tT}^2$ Thus by M we have

$$\mathcal{P}^*\left(\left||\frac{S_T}{m_T k_2}\sum_{t=1}^{m_T} Y_{tT}^{*2} - -\frac{S_T}{k_2}\sum_{t=1}^{T} Y_{tT}^2 p_{tT}\right| > \varepsilon\right) \le \varepsilon^{-p}\mathrm{E}^*(|\frac{S_T}{m_T k_2}\sum_{t=1}^{m_T} Y_{tT}^{*2} - \frac{S_T}{k_2}\sum_{t=1}^{T} p_{tT} Y_{tT}^2|^p)$$

for some $p > 1$. Now

$$\begin{aligned}
\mathrm{E}^*(|\frac{S_T}{m_T k_2}\sum_{t=1}^{m_T} Y_{tT}^{*2} - \frac{S_T}{k_2}\sum_{t=1}^{T} p_{tT} Y_{tT}^2|^p) &= (\frac{S_T}{m_T k_2})^p \mathrm{E}^*(|\sum_{t=1}^{m_T}(Y_{tT}^{*2} - \mathrm{E}^*\left[Y_{tT}^{*2}\right])|^p) \\
&\le (\frac{S_T}{m_T k_2})^p C\mathrm{E}^*((\sum_{t=1}^{m_T}|(Y_{tT}^{*2} - \mathrm{E}^\star\left[Y_{tT}^{\star}\right])^2|)^{p/2})
\end{aligned}$$

for some $C < \infty$ by an extension of the Burkholder inequality due to White and Chen (1996, Lemma A.2) as $(Y_{tT}^{*2} - \mathrm{E}^*\left[Y_{tT}^{*2}\right])$ are i.i.d zero mean. But for $1 < p \le 2$ we have by the c_r inequality (Davidson, 1994, p140) with $r = p/2$

$$\begin{aligned}
(\frac{S_T}{m_T k_2})^p \mathrm{E}^*((\sum_{t=1}^{m_T}|(Y_{tT}^{*2} - \mathrm{E}^*\left[Y_{tT}^{*2}\right])^2|)^{p/2}) &\le (\frac{S_T}{m_T k_2})^p \sum_{t=1}^{m_T} \mathrm{E}^*(\left|(Y_{tT}^{*2} - \mathrm{E}^*\left[Y_{tT}^{*2}\right])^2\right|^{p/2}) \\
&= \frac{S_T^p}{m_T^{p-1}k_2^p}\mathrm{E}^*\left(\left|(Y_{tT}^{*2} - \mathrm{E}^*\left[Y_{tT}^{*2}\right])\right|^p\right) \\
&\le \frac{S_T^p}{m_T^{p-1}k_2^p} 2^p \mathrm{E}^*\left(|(Y_{tT}^{*}|^{2p}\right) \\
&= \frac{S_T^{3/2}}{m_T^{1/2}k_2^{3/2}} 2^p \mathrm{E}^*\left(|(Y_{tT}^{*}|^{3}\right) \\
&= \frac{S_T^2}{T^{1/2}k_2^{3/2}} 2^{3/2}\sum_{t=1}^{T}|Y_{tT}|^3 p_{tT} \\
&= \frac{S_T^2}{T^{1/2}k_2^{3/2}} 2^{3/2}\frac{1}{T}\sum_{t=1}^{T}|Y_{tT}|^3 Tp_{tT} \\
&= \frac{S_T^2}{T^{1/2}k_2^{3/2}} 2^{3/2}(1 + o_p(1))\frac{1}{T}\sum_{t=1}^{T}|Y_{tT}|^3
\end{aligned}$$

as $\max_{1\le t\le T}|Tp_{tT} - 1| \xrightarrow{p} 1$ and for $p = 3/2$. Now note that

$$\frac{S_T}{T}\sum_{t=1}^{T}|Y_{tT}|^3 = \frac{S_T}{T}\sum_{t=1}^{T}|Y_{tT}|^2 \max_t |Y_{tT}| = O_p(T^{1/\alpha})$$

by Lemma A.3 of Smith (2011) and by M. Thus

$$\frac{S_T}{T^{1/2}k_2^{3/2}} 2^{3/2}(1 + o_p(1))\frac{S_T}{T}\sum_{t=1}^{T}|Y_{tT}|^3 = O_p(T^{-\eta+1/\alpha}).$$

Now note $\alpha > \max(4v, 1/\eta) > 1/\eta$, thus $\eta > 1/\alpha$ and the result follows ■

Lemma A.9 *If the finite dimensional stochastic process* $\{(X_t, Z_t\}_{t=1}^{\infty}$ *is strictly stationary and ergodic and satisfy* $E(|X_t|^{dp}) < \Delta$ *and* $E(|Z_t|^{\frac{dp}{d-1}}) < \Delta$*, for some* $1 < p \leq 2$ *and* $d > 1$ *and if assumptions assumptions 2.5 and 3.3 hold and if* $m_T = T/S_T$ *and* $S_T = o(T^{1/2})$ *then*

$$\lim_{n\to\infty} \mathcal{P}[\mathcal{P}^*(|\frac{S_T}{m_T}\sum_{t=1}^{m_T} Y_{tT}^* Z_{tT}^*| > T^{1/2}\varepsilon) > \varepsilon] = 0,$$

where $Z_{tT} = \frac{1}{S_T}\sum_{s=t-T}^{t-1} k(\frac{s}{S_T})Z_{t-s}, (t = 1, ..., T)$*, and* $(Z_{1T}^{\star}, ..., Z_{m_T T}^{\star})$ *is a bootstrap sample drawn from* $(Z_{1T}, ..., Z_{TT})$*.*

Proof: The proof is similar to the proof of Lemma B.2 of Gonçalves and White (2004). First note that by M for some $1 < p \leq 2$ we have

$$\begin{aligned}\mathcal{P}^*(|\frac{S_T}{m_T}\sum_{t=1}^{m_T} Y_{tT}^* Z_{tT}^*| > T^{1/2}\varepsilon) &\leq \frac{1}{\varepsilon^p T^{p/2}}\mathrm{E}^*[|\frac{S_T}{m_T}\sum_{t=1}^{m_T} Y_{tT}^* Z_{tT}^*|^p] \\ &\leq C\frac{1}{\varepsilon^p T^{p/2}}\mathrm{E}^*[|\frac{S_T}{m_T}\sum_{t=1}^{m_T} Y_{tT}^* Z_{tT}^* - \mathrm{E}^*[Y_{tT}^* Z_{tT}^*]|^p] \\ &\quad + \frac{C}{\varepsilon^p T^{p/2}}\mathrm{E}^*[|\frac{S_T}{m_T}\sum_{t=1}^{m_T}\mathrm{E}^*[Y_{tT}^* Z_{tT}^*]|^p] = F_1 + F_2\end{aligned}$$

by c_r inequality. Now

$$\begin{aligned}F_1 &\equiv \frac{1}{\varepsilon^p T^{p/2}}\frac{S_T^p}{m_T^p}\mathrm{E}^*[|\sum_{t=1}^{m_T} Y_{tT}^* Z_{tT}^* - \mathrm{E}^*[Y_{tT}^* Z_{tT}^*]|^p] \\ &\leq C\frac{1}{\varepsilon^p T^{p/2}}\frac{S_T^p}{m_T^p}\mathrm{E}^*[(\sum_{t=1}^{m_T}(|Y_{tT}^* Z_{tT}^* - \mathrm{E}^*[Y_{tT}^* Z_{tT}^*]|)^2)^{p/2}] \\ &\leq C\frac{1}{\varepsilon^p T^{p/2}}\frac{S_T^p}{m_T^p}\mathrm{E}^*[\sum_{t=1}^{m_T}|(Y_{tT}^* Z_{tT}^* - \mathrm{E}^*[Y_{tT}^* Z_{tT}^*])^p|] \\ &\leq C\frac{1}{\varepsilon^p T^{p/2}}\frac{S_T^p}{m_T^{p-1}}\mathrm{E}^*[|Y_{tT}^* Z_{tT}^*|^p]\end{aligned}$$

by an extension of the Burkholder inequality due to White and Chen (1996, Lemma A.2) and c_r inequality with $r = p/2$.

Also

$$\begin{aligned}F_2 &= \frac{C}{\varepsilon^p T^{p/2}}\left|\frac{S_T}{m_T}\sum_{t=1}^{m_T}\mathrm{E}^*[Y_{tT}^* Z_{tT}^*]\right|^p \\ &\leq \frac{C}{\varepsilon^p T^{p/2}}\frac{S_T^p}{m_T^p}\left|\sum_{t=1}^{m_T}\mathrm{E}^*[Y_{tT}^* Z_{tT}^*]\right|^p \\ &= \frac{C}{\varepsilon^p T^{p/2}}S_T^p\,|\mathrm{E}^*[Y_{tT}^* Z_{tT}^*]|^p \\ &\leq \frac{C}{\varepsilon^p T^{p/2}}S_T^p\mathrm{E}^*[|Y_{tT}^* Z_{tT}^*|^p]\end{aligned}$$

by Jensen. Now

$$\begin{aligned}\frac{C}{\varepsilon^p T^{p/2}}S_T^p\mathrm{E}^*[|Y_{tT}^* Z_{tT}^*|^p] &= \frac{C}{\varepsilon^p T^{p/2}}S_T^p\sum_{t=1}^{T}|Y_{tT}|^p\,|Z_{tT}|^p\,p_{tT} \\ &= \frac{C}{\varepsilon^p T^{p/2}}S_T^p\frac{1}{T}\sum_{t=1}^{T}|Y_{tT}|^p\,|Z_{tT}|^p\,Tp_{tT} \\ &= \frac{C}{\varepsilon^p T^{p/2}}S_T^p(1 + o_p(1))\frac{1}{T}\sum_{t=1}^{T}|Y_{tT}|^p\,|Z_{tT}|^p\end{aligned}$$

as $\max_{1\leq t\leq T}|Tp_{tT} - 1| \xrightarrow{p} 1$.

But by M and Holder Inequality

$$\begin{aligned}\mathcal{P}[\frac{C}{\varepsilon^p T^{p/2}}S_T^p\frac{1}{T}\sum_{t=1}^{T}|Y_{tT}|^p\,|Z_{tT}|^p &> \delta] \leq \frac{C}{\delta\varepsilon^p T^{p/2}}S_T^p\mathrm{E}[\frac{1}{T}\sum_{t=1}^{T}|Y_{tT}|^p\,|Z_{tT}|^p] \\ &= \frac{C}{\delta\varepsilon^p T^{p/2}}S_T^p\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}[|Y_{tT}|^p\,|Z_{tT}|^p] \\ &\leq \frac{C}{\delta\varepsilon^p T^{p/2}}S_T^p\frac{1}{T}\sum_{t=1}^{T}(\mathrm{E}[|Y_{tT}|^{\alpha p}])^{1/\alpha}\left(\mathrm{E}[|Z_{tT}|^{\frac{\alpha p}{1-\alpha}}]\right)^{(1-\alpha)/\alpha}.\end{aligned}$$

Now by T and Jensen inequalities

$$
\begin{aligned}
\mathrm{E}[|Y_{tT}|^{\alpha p}] &= \mathrm{E}[|\frac{1}{S_T}\sum_{s=t-T}^{t-1} k(\frac{s}{S_T})X_{t-s}|^{\alpha p}] \\
&\leq \mathrm{E}[(\frac{1}{S_T}\sum_{s=t-T}^{t-1} |k(\frac{s}{S_T})||X_{t-s}|)^{\alpha p}] \\
&= \mathrm{E}[(\frac{1}{S_T}\sum_{s=t-T}^{t-1} |k(\frac{s}{S_T})|)^{\alpha p}(\frac{\frac{1}{S_T}\sum_{s=t-T}^{t-1} |k(\frac{s}{S_T})|\,|X_{t-s}|}{\frac{1}{S_T}\sum_{s=t-T}^{t-1} |k(\frac{s}{S_T})|})^{\alpha p}] \\
&\leq \mathrm{E}[(\frac{1}{S_T}\sum_{s=t-T}^{t-1} |k(\frac{s}{S_T})|)^{\alpha p-1}\frac{1}{S_T}\sum_{s=t-T}^{t-1} |k(\frac{s}{S_T})|\,|X_{t-s}|^{\alpha p}] \\
&= (\frac{1}{S_T}\sum_{s=t-T}^{t-1} |k(\frac{s}{S_T})|)^{\alpha p-1}\frac{1}{S_T}\sum_{s=t-T}^{t-1} |k(\frac{s}{S_T})|\mathrm{E}\,|X_{t-s}|^{\alpha p} \\
&= (\frac{1}{S_T}\sum_{s=t-T}^{t-1} |k(\frac{s}{S_T})|)^{\alpha p}\Delta = O(1)
\end{aligned}
$$

as $E(|X_t|^{\alpha p})$ is bounded . By the same reasoning $E[|Z_{tT}|^{\frac{\alpha p}{1-\alpha}}] = O(1)$. Thus the result follows since $S_T/T^{1/2} = o(1)$. ■

A.4 Proofs of the results in section 4.1

In this subsection of the appendix we take $p_{tT} = 1/T$ and consequently Assumption 3.3 (i) is automatically satisfied. Assumption 3.3 (ii) follow from Lemma A.2 of Smith (2011).

Proof of Theorem 4.1: The result is proven if we show that the conditions of Lemma A.2 of Gonçalves and White (2004) are satisfied. Conditions (a1), (a2) and (b1) and (b2) are satisfied by assumption 4.1 (i) and (iii) (see Jennrich, 1969, Lemma 2). Note that uniqueness of the minimum follows from Lemma 2.3 of Newey and MacFadden (1994). To prove (a3) define $Q_0(\beta) = \mathrm{E}[g(z_t,\beta)]'W\mathrm{E}[g(z_t,\beta)]$ and note that as in the proof of Theorem 2.6 of Newey and MacFadden (1994) using T and CS

$$
\begin{aligned}
|Q_T(\beta) - Q_0(\beta)| \leq\; & \|\hat{g}(\beta) - \mathrm{E}[g(z_t,\beta)]\|^2 \|W_T\| \\
& +2\|\mathrm{E}[g(z_t,\beta)]\|\,\|\hat{g}(\beta) - \mathrm{E}[g(z_t,\beta)]\|\,\|W_T\| \\
& +\|\mathrm{E}[g(z_t,\beta)]\|^2 \|W_T - W\|.
\end{aligned}
$$

By the the Lemma A.3 we have $\sup_{\beta\in B}\|\hat{g}(\beta) - \mathrm{E}[g(z_t,\beta)]\| = o_p(1)$ Also by assumption $\|\mathrm{E}[g(z_t,\beta)]\|$ is bounded and $\|W_T - W\| = o_p(1)$.

It remains to prove (b3). By T and CS

$$
\begin{aligned}
|Q_T^*(\beta) - Q_T(\beta)| \leq\; & \|\hat{g}_T^*(\beta) - \hat{g}(\beta)\|^2 \|W_T^*\| + 2\|\hat{g}(\beta)\|\,\|\hat{g}_T^*(\beta) - \hat{g}(\beta)]\|\,\|W_T^*\| \\
& +\|\hat{g}(\beta)\|^2 \|W_T^* - W_T\|.
\end{aligned}
$$

Now by Lemma A.6 we have $\sup_{\beta\in B}\left\|\hat{g}_T^*(\beta) - \hat{g}(\beta)\right\| = o_B(1)$ also

$$
\begin{aligned}
\sup_{\beta\in B}\|\hat{g}(\beta)\| &\leq \sup_{\beta\in B}\|\hat{g}(\beta) - \mathrm{E}[g(z_t,\beta)]\| + \sup_{\beta\in B}\|\mathrm{E}[g(z_t,\beta)]\| \\
&= o_p(1) + C.
\end{aligned}
$$

thus the result follows as $\left\|W_T^* - W_T\right\| = o_B(1)$. ■

Proof of Theorem 4.2: Let $\hat{G}_T^* \equiv \partial\hat{g}_T^*(\hat{\beta}^*)\partial\beta'$ To prove asymptotic Normality notice that by the first order conditions we have $\sqrt{T/k_2}\hat{G}_T^{*\prime}W_T^*\hat{g}_T^*(\hat{\beta}^*) = 0$. Hence a first order Taylor expansion around $\hat{\beta}$ yields

$$\sqrt{T/k_2}\hat{G}_T^{*\prime}W_T^*\hat{g}_T^*(\hat{\beta}) + \hat{G}_T^{*\prime}W_T^*\tilde{G}_T^*\sqrt{T/k_2}(\hat{\beta}^* - \hat{\beta}) = 0,$$

where $\tilde{G}_T^* \equiv \partial\hat{g}_T^*(\tilde{\beta}^*)\partial\beta'$ and $\tilde{\beta}^*$ is on a line joining $\hat{\beta}$ and $\hat{\beta}^*$. Solving for $\sqrt{T/k_2}(\hat{\beta}^* - \hat{\beta})$ we obtain

$$\sqrt{T/k_2}(\hat{\beta}^* - \hat{\beta}) = -[\hat{G}_T^{*\prime}W_T^*\tilde{G}_T^*]^{-1}\sqrt{T/k_2}\hat{G}_T^{*\prime}W_T^*\hat{g}_T^*(\hat{\beta}).$$

By a Taylor expansion we have

$$
\begin{aligned}
& \sqrt{T/k_2}\hat{G}_T^{*\prime}W_T^*\hat{g}_T^*(\hat{\beta}) = \sqrt{T/k_2}\hat{G}_T^{*\prime}W_T^*\hat{g}_T^*(\beta_0) + \sqrt{T/k_2}\hat{G}_T^{*\prime}W_T^*\check{G}_T^*(\hat{\beta} - \beta_0) \\
= \; & \sqrt{T/k_2}\hat{G}_T^{*\prime}W_T^*[\hat{g}_T^*(\beta_0) - \hat{g}_T(\beta_0)] + \sqrt{T/k_2}\hat{G}_T^{*\prime}W_T^*\hat{g}_T(\beta_0) + \sqrt{T/k_2}\hat{G}_T^{*\prime}W_T^*\tilde{G}_T^*(\hat{\beta} - \beta_0),
\end{aligned}
$$

where $\check{G}_T^* = \partial\hat{g}_T^*(\check{\beta})\partial\beta'$ and $\check{\beta}$ is on a line joining $\hat{\beta}$ and β_0.

We prove now that

$$\sqrt{T/k_2}\hat{G}_T^{*\prime}W_T^*\hat{g}_T(\beta_0) + \sqrt{T/k_2}\hat{G}_T^{*\prime}W_T^*\tilde{G}_T^*(\hat{\beta} - \beta_0) = o_B(1).$$

Note that by the first order conditions of the original GMM problem we have

$$\sqrt{T}(\hat{\beta} - \beta_0) = -[\hat{G}_T'W_T\ddot{G}_T]^{-1}\sqrt{T/k_2}\hat{G}_T'W_T\hat{g}_T(\beta_0),$$

where $\ddot{G}_T = \partial \hat{g}_T(\ddot{\beta})\partial\beta'$ and $\ddot{\beta}$ is on a line joining $\hat{\beta}$ and β_0. Thus

$$\begin{aligned} &\sqrt{T/k_2}\hat{G}_T^{*\prime}W_T^*\hat{g}_T(\beta_0) + \sqrt{T/k_2}\hat{G}_T^{*\prime}W_T^*\tilde{G}_T^*(\hat{\beta}-\beta_0) \\ = \quad & [\hat{G}_T^{*\prime}W_T^* - \hat{G}_T^{*\prime}W_T^*\tilde{G}_T^*\left[\hat{G}_T'W_T\ddot{G}_T\right]^{-1}\hat{G}_T'W_T]\sqrt{T/k_2}\hat{g}_T(\beta_0). \end{aligned}$$

Now by Assumption $W_T^* = W_T + o_B(1)$, $W_T = W + o_p(1)$, also by the bootstrap uniform convergence Lemma A.7 consistency of $\hat{\beta}^*$ and $\hat{\beta}$, $\hat{G}_T^* - G = o_B(1)$, $\tilde{G}_T^* - G = o_B(1)$, $\ddot{G}_T - G = o_p(1)$, $\hat{G}_T - G = o_p(1)$ and by the CLT of Wooldridge and White (Theorem 5.20 of White ,1999) $\sqrt{T/k_2}\hat{g}_T(\beta_0) = O(1)$.

Now $\sqrt{T/k_2}2\hat{G}_T^{*\prime}W_T^*[\hat{g}_T^*(\beta_0) - \hat{g}_T(\beta_0)]$ converges to $N(0, (G'WG)^{-1}G'W\Omega WG(G'WG)^{-1})$ by bootstrap CLT Theorem A.2 and the fact that $\hat{G}_T^* - G = o_B(1)$ and $W_T^* = W + o_B(1)$. The result follows as the $\sqrt{T/k_2}(\hat{\beta}^* - \hat{\beta})$ converges to the same asymptotic distribution of $T^{1/2}(\hat{\beta} - \beta_0)$ and by Polya Theorem, Serfling (2002, p.18), as $\Phi(\cdot)$ is a continuous c.d.f.. ■

Proof of Lemma 4.1: We use the same strategy of the proof of Theorem 4.1 of Gonçalves and White (2004). First consider the unfeasible estimator of Ω :

$$\hat{\Omega}^*(\beta_0) = \frac{S_T}{m_T k_2}\sum\nolimits_{t=1}^{m_T} g_t^*(\beta_0)g_t^*(\beta_0)'.$$

Fix any $\lambda \in \mathbb{R}^m$. Now

$$\begin{aligned} \lambda'\hat{\Omega}^*(\beta_0)\lambda \quad &= \quad \frac{S_T}{m_T k_2}\sum\nolimits_{t=1}^{m_T} \lambda' g_t^*(\beta_0)g_t^*(\beta_0)'\lambda \\ &= \quad \frac{S_T}{m_T k_2}\sum\nolimits_{t=1}^{m_T} (\lambda' g_t^*(\beta_0))^2. \end{aligned}$$

Now applying Lemma A.8 with $X_t = \lambda' g_t(\beta_0)$ and $p_{tT} = 1/T$, $t = 1, ..., T$ it follows that

$$\frac{S_T}{m_T k_2}\sum\nolimits_{t=1}^{m_T} (\lambda' g_t^*(\beta_0))^2 - \frac{S_T}{T k_2}\sum\nolimits_{t=1}^{T} (\lambda' g_{tT}(\beta_0))^2 = o_B(1)$$

and by Smith (2011) Lemma A.3

$$\frac{S_T}{T k_2}\sum\nolimits_{t=1}^{T} (\lambda' g_{tT}(\beta_0))^2 = \lambda'\Omega\lambda + o_p(1).$$

Thus it remains to prove that $\left|\lambda'\hat{\Omega}^*(\tilde{\beta}^*)\lambda - \lambda'\hat{\Omega}^*(\beta_0)\lambda\right| = o_B(1)$. Note that by first order Taylor expansion of $(\lambda' g_{tT}(\tilde{\beta}^*))^2$ around β_0 we have

$$(\lambda' g_{tT}(\tilde{\beta}^*))^2 = \left(\lambda' g_t^*(\beta_0)\right)^2 + 2(\lambda' g_t^*(\ddot{\beta}^*)\lambda' G_t^*(\ddot{\beta}^*)(\tilde{\beta}^* - \beta_0))$$

where $\ddot{\beta}^*$ is in a line joining $\tilde{\beta}^*$ and β_0. Thus

$$\begin{aligned} \lambda'\hat{\Omega}^*(\tilde{\beta}^*)\lambda \quad &= \quad \frac{S_T}{m_T k_2}\sum\nolimits_{t=1}^{m_T} (\lambda' g_t^*(\tilde{\beta}^*))^2 \\ &= \quad \frac{S_T}{m_T k_2}\sum\nolimits_{t=1}^{m_T} (\left(\lambda' g_t^*(\beta_0)\right)^2 + 2(\lambda' g_t^*(\ddot{\beta}^*)\lambda' G_t^*(\ddot{\beta}^*)(\tilde{\beta}^* - \beta_0))) \\ &= \quad \lambda'\hat{\Omega}^*(\beta_0)\lambda + \frac{S_T}{m_{Tk_2}}\sum\nolimits_{t=1}^{m_T} 2(\lambda' g_t^*(\ddot{\beta}^*)\lambda' G_t^*(\ddot{\beta}^*)(\tilde{\beta}^* - \beta_0)). \end{aligned}$$

Now denote $G_{t,j}^*(\ddot{\beta}^*)$ the column j of $G_t^*(\ddot{\beta}^*)$ thus

$$\begin{aligned} &|\frac{S_T}{m_T}\sum\nolimits_{t=1}^{m_T} 2(\lambda' g_t^*(\ddot{\beta}^*)\lambda' G_t^*(\ddot{\beta}^*)(\tilde{\beta}^* - \beta_0))| \\ &= |\frac{S_T}{m_T}\sum\nolimits_{t=1}^{m_T} 2(\lambda' g_t^*(\ddot{\beta}^*)\sum\nolimits_{j=1}^{p} \lambda' G_{t,j}^*(\ddot{\beta}^*)(\tilde{\beta}_j^* - \beta_{0,j}))| \\ &= |\sum\nolimits_{j=1}^{p}\frac{S_T}{m_T}\sum\nolimits_{t=1}^{m_T} 2(\lambda' g_t^*(\ddot{\beta}^*)\lambda' G_{t,j}^*(\ddot{\beta}^*)(\tilde{\beta}_j^* - \beta_{j,0}))| \\ &\leq \sum\nolimits_{j=1}^{p}|\frac{S_T}{m_T}\sum\nolimits_{t=1}^{m_T} 2(\lambda' g_t^*(\ddot{\beta}^*)\lambda' G_{t,j}^*(\ddot{\beta}^*)(\tilde{\beta}_j^* - \beta_{j,0}))| \\ &= \sum\nolimits_{j=1}^{p} O_B(\frac{1}{\sqrt{T}})\left|\frac{S_T}{m_T}\sum\nolimits_{t=1}^{m_T} 2(\lambda' g_t^*(\ddot{\beta}^*)\lambda' G_{tj}^*(\ddot{\beta}^*)\right| \end{aligned}$$

by T and the fact that $(\tilde{\beta}_j^* - \beta_{j,0}) = O_B(1/T^{1/2})$. Note that

$$\begin{aligned} \left|\frac{S_T}{m_T}\sum\nolimits_{t=1}^{m_T} 2(\lambda' g_t^*(\ddot{\beta}^*)\lambda' G_{tj}^*(\ddot{\beta}^*)\right| \quad &\leq \quad \frac{S_T}{m_T}\sum\nolimits_{t=1}^{m_T} 2\left|(\lambda' g_t^*(\ddot{\beta}^*)\lambda' G_{tj}^*(\ddot{\beta}^*)\right| \\ &\leq \quad \frac{S_T}{m_T}\sum\nolimits_{t=1}^{m_T} \sup_{\beta\in B} 2\left|(\lambda' g_t^*(\beta)\right| \sup_{\beta\in B}\left|\lambda' G_{tj}^*(\beta)\right|. \end{aligned}$$

Now define $|Y_{tT}| = 2\sup_{\beta\in B}|(\lambda' g_t^*(\beta)|$ and $|Z_{tT}| = \sup_{\beta\in\mathcal{N}}\left|\lambda' G_{tj}^*(\beta)\right|$ and apply Lemma A.9 above with $p = 2$, $d = \alpha/2$ and $p_{tT} = 1/T$, $t = 1, ..., T$ which shows that

$$\frac{S_T}{m_T}\sum_{t=1}^{m_T} 2\sup_{\beta\in B} 2\left|(\lambda' g_t^*(\beta)\right|\sup_{\beta\in B}\left|\lambda' G_{tj}^*(\beta)\right| = o_p(T^{-1/2})$$

and hence the result follows. ■

Proof of Theorem 4.3: Note that by a Taylor expansion

$$\sqrt{T/k_2}\hat{g}^*(\hat{\beta}^{e*}) = \sqrt{T/k_2}\hat{g}^*(\hat{\beta}^e) + \tilde{G}^*\sqrt{T/k_2}(\hat{\beta}^{e*} - \hat{\beta}^e),$$

where $\tilde{G}_T^* \equiv \partial\hat{g}_T^*(\tilde{\beta})/\partial\beta'$ and $\tilde{\beta}^*$ is in a line joining $\hat{\beta}^{e*}$ and $\hat{\beta}^e$.

Note that by Theorem 4.2 with $W_T^* = \tilde{\Omega}^{*-1}$

$$\sqrt{T/k_2}(\hat{\beta}^{e*} - \hat{\beta}^e) = -[\hat{G}_T^{*\prime}\tilde{\Omega}^{*-1}\tilde{G}_T^*]^{-1}\hat{G}_T^{*\prime}\tilde{\Omega}^{*-1}\sqrt{T/k_2}[\hat{g}_T^*(\beta_0) - \hat{g}_T(\beta_0)] + o_B(1).$$

Also by a Taylor expansion

$$\begin{aligned}\sqrt{T/k_2}(\hat{g}^*(\hat{\beta}^e) - \hat{g}^*(\beta_0) - \hat{g}(\hat{\beta}^e) + \hat{g}(\beta_0)) &= (\check{G}_T^* - \ddot{G}_T)\sqrt{T/k_2}\left(\hat{\beta}^e - \beta_0\right) \qquad (A.15)\\ &= o_B(1)O_p(1) = o_B(1),\end{aligned}$$

where $\check{G}_T^* \equiv \partial\hat{g}_T^*(\check{\beta})/\partial\beta'$ where $\check{\beta}$ is in a line joining $\hat{\beta}^e$ and β_0 and $\ddot{G}_T \equiv \partial\hat{g}_T(\ddot{\beta})/\partial\beta'$ where $\ddot{\beta}$ is in a line joining $\hat{\beta}^e$ and β_0.

Thus

$$\sqrt{\frac{T}{k_2}}[\hat{g}^*(\hat{\beta}^{e*}) - \hat{g}(\hat{\beta}^e)] = [I_m - \tilde{G}^*[\hat{G}_T^{*\prime}\tilde{\Omega}^{*-1}\tilde{G}_T^*]^{-1}\hat{G}_T^{*\prime}\tilde{\Omega}^{*-1}]\sqrt{T/k_2}[\hat{g}_T^*(\beta_0) - \hat{g}_T(\beta_0)] + o_B(1).$$

Now since $\tilde{G}^* = G + o_B(1)$, $\hat{G}_T^* = G + o_B(1)$, $\tilde{\Omega}^{*-1} = \Omega^{-1} + o_B(1)$ and by the bootstrap CLT Theorem A.2 $\sqrt{T/k_2}[\hat{g}_T^*(\beta_0) - \hat{g}_T(\beta_0)]$ converges to $N(0,\Omega)$ It follows that

$$\sqrt{\frac{T}{k_2}}[\hat{g}^*(\hat{\beta}^{e*}) - \hat{g}(\hat{\beta}^e)] = [I_m - G\left[G'\Omega^{-1}G\right]^{-1}G'\Omega^{-1}]\sqrt{T/k_2}[\hat{g}_T^*(\beta_0) - \hat{g}_T(\beta_0)] + o_B(1).$$

Thus

$$\mathcal{J}^* = \frac{T}{k_2}[\hat{g}_T^*(\beta_0) - \hat{g}_T(\beta_0)]'[\Omega^{-1} - \Omega^{-1}G\left[G'\Omega^{-1}G\right]^{-1}G'\Omega^{-1}][\hat{g}_T^*(\beta_0) - \hat{g}_T(\beta_0)] + o_B(1).$$

As

$$\begin{aligned}&\Omega[\Omega^{-1} - \Omega^{-1}G\left[G'\Omega^{-1}G\right]^{-1}G'\Omega^{-1}]\Omega[\Omega^{-1} - \Omega^{-1}G\left[G'\Omega^{-1}G\right]^{-1}G'\Omega^{-1}]\Omega\\ =\ &\Omega[\Omega^{-1} - \Omega^{-1}G\left[G'\Omega^{-1}G\right]^{-1}G'\Omega^{-1}]\Omega\end{aligned}$$

and $tr([\Omega^{-1} - \Omega^{-1}G\left[G'\Omega^{-1}G\right]^{-1}G'\Omega^{-1}]\Omega) = m - p$, it follows from Rao and Mitra(1972), the fact that $\mathcal{J}^* \Rightarrow^{d_{P^*}} \chi^2(m-p)$. Since $\mathcal{J} \xrightarrow{d} \chi^2(m-p)$, the result stated in the Theorem is a consequence of Polya Theorem (Serfling, 2002, p.18), as the chi-squared distribution has a continuous c.d.f. ■

Proof of Theorem 4.4: We start by deriving the asymptotic distribution of $\mathcal{W}^*$. Define $h_t^{a,*}(\beta,\gamma) \equiv (g^*(z_t,\beta)', [q^*(z_t,\beta) - \gamma]')$, $\hat{h}^{a,*}(\beta,\gamma) = \sum_{t=1}^{m_T} h_t^{a,*}(\beta,\gamma)/m_T$ and $\tilde{Q}^*(\beta) = \hat{h}^{a,*}(\beta,\gamma)'\hat{\Xi}^{*-1}\hat{h}^{a,*}(\beta,\gamma)$. Note that the unrestricted bootstrapped GMM estimator solves

$$(\hat{\beta}^{e*\prime}, \hat{\gamma}^{*\prime}) = \underset{\beta\in B,\gamma\in\Gamma}{\arg\min}\,\tilde{Q}^*(\beta,\gamma),$$

where Γ is a compact parameter space. The solution is given by

$$\begin{aligned}\hat{\beta}^{e*} &= \underset{\beta\in B}{\arg\min}\,\hat{g}^*(\beta)'\hat{\Omega}^{*-1}\hat{g}^*(\beta),\\ \hat{\gamma}^* &= \hat{q}^*(\hat{\beta}^*) - \hat{\Xi}_{21}^*\hat{\Omega}^{*-1}\hat{g}^*(\hat{\beta}^*).\end{aligned}$$

We note that by Theorem 4.1 $\hat{\beta}^{e*} = \hat{\beta} + o_B(1)$ and by Lemma A.6 and $\hat{\Xi}^* = \hat{\Xi} + o_B(1)$ we have $\hat{\gamma}^* = \hat{\gamma} + o_B(1)$.Since these estimators satisfy the first order conditions we have $\hat{D}^*(\hat{\beta}^{e*})'\hat{\Xi}^{*-1}\hat{h}^{a,*}(\hat{\beta}^{e*},\hat{\gamma}^*) = 0$ with

$$\hat{D}^*(\beta) \equiv \begin{pmatrix}\sum_{i=1}^{m_T}\hat{G}_t^*(\beta)/m_T & 0\\ \sum_{i=1}^{m_T}\hat{Q}_t^*(\beta)/m_T & -I_s\end{pmatrix}.$$

Thus by a Taylor expansion around $(\hat{\beta}^{e\prime}, \hat{\gamma}')'$

$$\hat{D}^*(\hat{\beta}^{e*})'\hat{\Xi}^{*-1}\hat{h}^*(\hat{\beta}^e,\hat{\gamma}) + \hat{D}^*(\hat{\beta}^*)'\hat{\Xi}^{*-1}\hat{D}^*(\tilde{\beta}^*)\begin{pmatrix}\hat{\beta}^* - \hat{\beta}\\ \hat{\gamma}^* - \hat{\gamma}\end{pmatrix} = 0,$$

where $\tilde{\beta}^*$ is in a line joining $\hat{\beta}^{e*}$ and $\hat{\beta}^e$. Thus

$$\sqrt{T}\left(\begin{array}{c}\hat{\beta}^{e*}-\hat{\beta}^e\\ \hat{\gamma}^*-\hat{\gamma}\end{array}\right)=-[\hat{D}^{*\prime}\hat{\Xi}^{*-1}\tilde{D}^*]^{-1}\hat{D}^{*\prime}\hat{\Xi}^{*-1}\sqrt{T}\hat{h}^*(\hat{\beta},\hat{\gamma}).$$

Now notice that as in the proof of Theorem 4.2

$$\sqrt{\frac{T}{k_2}}\left(\begin{array}{c}\hat{\beta}^{e*}-\hat{\beta}^e\\ \hat{\gamma}^*-\hat{\gamma}\end{array}\right)=-[D'\Xi^{-1}D]^{-1}D'\Xi^{-1}\sqrt{\frac{T}{k_2}}[\hat{h}_T^{a,*}(\beta_0,0)-\hat{h}_T^a(\beta_0,0)]+o_B(1).$$

Thus by a Taylor expansion we have

$$\begin{aligned}\sqrt{\frac{T}{k_2}}\left(\begin{array}{c}a(\hat{\beta}^{e*})-a(\hat{\beta}^e)\\ \hat{\gamma}^*-\hat{\gamma}\end{array}\right) &= -R(\ddot{\beta}^*)[D'\Xi^{-1}D]^{-1}D'\Xi^{-1}\sqrt{\frac{T}{k_2}}[\hat{h}_T^{a,*}(\beta_0,0)-\hat{h}_T^a(\beta_0,0)]+o_B(1)\\ &= -R[D'\Xi^{-1}D]^{-1}D'\Xi^{-1}\sqrt{\frac{T}{k_2}}[\hat{h}_T^{a,*}(\beta_0,0)-\hat{h}_T^a(\beta_0,0)]+o_B(1),\end{aligned}$$

where $\ddot{\beta}^*$ is in a line joining $\hat{\beta}^{e*}$ and $\hat{\beta}^e$ as $R(\ddot{\beta}^*)=R+o_B(1)$.

Thus

$$\begin{aligned}\mathcal{W}^* &= (T/k_2)[\hat{r}^*-\hat{r}]'\left[\hat{R}^*(\hat{D}^{*\prime}\hat{\Xi}^{*-1}\hat{D}^*)^{-1}\hat{R}^{*\prime}\right]^{-1}[\hat{r}^*-\hat{r}]\\ &= \sqrt{\frac{T}{k_2}}[\hat{h}_T^{a,*}(\beta_0,0)-\hat{h}_T^a(\beta_0,0)]'\Xi^{-1}D[D'\Xi^{-1}D]^{-1}R'\left[R(D'\Xi^{-1}D)^{-1}R'\right]^{-1}\\ &\quad\times R(D'\Xi^{-1}D)^{-1}D'\Xi^{-1}\sqrt{\frac{T}{k_2}}[\hat{h}_T^{a,*}(\beta_0,0)-\hat{h}_T^a(\beta_0,0)]+o_B(1).\end{aligned}$$

Since as $\hat{D}^*=D+o_B(1)$ by Lemma A.7 and $\hat{\Xi}^*=\Xi+o_B(1)$ by Lemma A.6and the fact that $\sqrt{T/k_2}[\hat{h}_T^{a,*}(\beta_0,0)-\hat{h}_T^a(\beta_0,0)]=O_B(1)$ by the bootstrap CLT. Thus as in Theorem 2.4 above $\mathcal{W}^*$ converges to a chi-squared distribution with $s+r$ degrees of freedom.

We consider know the score statistic $\mathcal{S}^*$. We derive the distribution of the bootstrap restricted GMM estimator. Note that the Lagrangian of the restricted problem is

$$L^*=\tilde{Q}^*(\beta,\gamma)-a(\beta)'\lambda^*-\gamma'\mu.$$

Denote $\hat{\varphi}^*=(\hat{\lambda}',\hat{\mu}')'$ the vector of Lagrange multipliers evaluated at the optimum. Thus the first order conditions are

$$\hat{D}^*(\hat{\beta}_r^{e*})'\hat{\Xi}^{*-1}\hat{h}(\hat{\beta}_r^{e*},0)-R(\hat{\beta}_r^{e*})\hat{\varphi}^*=0.$$

Multiplying both sides by $R(\hat{\beta}_r^{e*})'(\hat{D}^{*\prime}\hat{\Xi}^{*-1}\hat{D}^*)^{-1}$ we have

$$R(\hat{\beta}_r^{e*})'(\hat{D}^*(\hat{\beta}_r^{e*})'\hat{\Xi}^{*-1}\hat{D}^*(\hat{\beta}_r^{e*}))^{-1}\hat{D}^*(\hat{\beta}_r^{e*})'\hat{\Xi}^{*-1}\hat{h}^{a,*}(\hat{\beta}_r^{e*},0)-R(\hat{\beta}_r^{e*})'(\hat{D}^*(\hat{\beta}_r^{e*})'\hat{\Xi}^{*-1}\hat{D}^*(\hat{\beta}_r^{e*}))^{-1}R(\hat{\beta}_r^{e*})\hat{\varphi}^*=0. \tag{A.16}$$

Thus

$$\hat{\varphi}^*=[R(\hat{\beta}_r^{e*})'(\hat{D}^*(\hat{\beta}_r^{e*})'\hat{\Xi}^{*-1}\hat{D}^*(\hat{\beta}_r^{e*}))^{-1}R(\hat{\beta}_r^{e*})]^{-1}R(\hat{\beta}_r^{e*})'(\hat{D}^*(\hat{\beta}_r^{e*})'\hat{\Xi}^{*-1}\hat{D}^*(\hat{\beta}_r^{e*}))^{-1}\hat{D}^*(\hat{\beta}_r^{e*})'\hat{\Xi}^{*-1}\hat{h}(\hat{\beta}_r^{e*},0). \tag{A.17}$$

Hence replacing $(A.17)$ in $(A.16)$ we have

$$\begin{aligned}&\hat{D}^*(\hat{\beta}_r^{e*})'\hat{\Xi}^{*-1}\hat{h}^{a,*}(\hat{\beta}_r^{e*},0)\\ &-R(\hat{\beta}_r^{e*})[R(\hat{\beta}_r^{e*})'(\hat{D}^*(\hat{\beta}_r^{e*})'\hat{\Xi}^{*-1}\hat{D}^*(\hat{\beta}_r^{e*}))^{-1}R(\hat{\beta}_r^{e*})]^{-1}R(\hat{\beta}_r^{e*})'(\hat{D}^*(\hat{\beta}_r^{e*})'\hat{\Xi}^{*-1}\hat{D}^*(\hat{\beta}_r^{e*}))^{-1}\hat{D}^*(\hat{\beta}_r^{e*})'\hat{\Xi}^{*-1}\hat{h}^{a,*}(\hat{\beta}_r^{e*},0)=0.\end{aligned}$$

But by a Taylor expansion $\hat{h}^{a,*}(\hat{\beta}_r^{e*},0)=\hat{h}^{a,*}(\hat{\beta}_r^e,0)+\tilde{D}^*(\bar{\beta}^*)S_1(\hat{\beta}_r^{e*}-\hat{\beta}_r^e)$ where $\bar{\beta}^*$ is in a line joining $\hat{\beta}_r^{e*}$ and $\hat{\beta}_r^e$ and S_1 is a selection matrix such that

$$\tilde{D}^*(\bar{\beta}^*)S_1=\left(\begin{array}{c}\sum_{i=1}^T\hat{G}_t^*(\bar{\beta}^*)/T\\ \sum_{i=1}^T\hat{Q}_t^*(\bar{\beta}^*)/T\end{array}\right).$$

Thus we have

$$\begin{aligned}&[I-R(\hat{\beta}_r^{e*})[R(\hat{\beta}_r^{e*})'(\hat{D}^*(\hat{\beta}_r^{e*})'\hat{\Xi}^{*-1}\hat{D}^*(\hat{\beta}_r^{e*}))^{-1}R(\hat{\beta}_r^{e*})]^{-1}R(\hat{\beta}_r^{e*})'(\hat{D}^*(\hat{\beta}_r^{e*})'\hat{\Xi}^{*-1}\hat{D}^*(\hat{\beta}_r^{e*}))^{-1}]\\ &[\hat{D}^*(\hat{\beta}_r^{e*})'\hat{\Xi}^{*-1}\sqrt{T}\hat{h}^{a,*}(\hat{\beta}_r^e,0)+\hat{D}^*(\hat{\beta}_r^{e*})'\hat{\Xi}^{*-1}\tilde{D}^*(\bar{\beta}^*)S_1\sqrt{T}(\hat{\beta}_r^{e*}-\hat{\beta}_r^e)]=0,\end{aligned}$$

and consequently

$$\begin{aligned}&S_1\sqrt{\frac{T}{k_2}}(\hat{\beta}_r^{e*}-\hat{\beta}_r^e)=-[\hat{D}^*(\hat{\beta}_r^{e*})'\hat{\Xi}^{*-1}\tilde{D}^*(\bar{\beta}^*)]^{-1}\\ &\times[I-R(\hat{\beta}_r^{e*})[R(\hat{\beta}_r^{e*})'(\hat{D}^*(\hat{\beta}_r^{e*})'\hat{\Xi}^{*-1}\hat{D}^*(\hat{\beta}_r^{e*}))^{-1}R(\hat{\beta}_r^{e*})]^{-1}R(\hat{\beta}_r^{e*})'(\hat{D}^*(\hat{\beta}_r^{e*})'\hat{\Xi}^{*-1}\hat{D}^*(\hat{\beta}_r^{e*}))^{-1}]\\ &\times\hat{D}^*(\hat{\beta}_r^{e*})'\hat{\Xi}^{*-1}\sqrt{\frac{T}{k_2}}\hat{h}^{a,*}(\hat{\beta}_r^e,0).\end{aligned}$$

Now as in $(A.15)$ above we have

$$\sqrt{\frac{T}{k_2}}(\hat{h}^{a,*}(\hat{\beta}^e_r,0)-\hat{h}^{a,*}(\beta_0,0)-\hat{h}^a(\hat{\beta}^e_r,0)+\hat{h}^a(\beta_0,0))=o_B(1). \tag{A.18}$$

Therefore we have

$$\begin{aligned}S_1\sqrt{\frac{T}{k_2}}(\hat{\beta}^{e*}_r-\hat{\beta}^e_r)&=-[\hat{D}^*(\hat{\beta}^{e*}_r)'\hat{\Xi}^{*-1}\tilde{D}^*(\bar{\beta}^*)]^{-1}\\ &\times[I-R(\hat{\beta}^{e*}_r)[R(\hat{\beta}^{e*}_r)'(\hat{D}^*(\hat{\beta}^{e*}_r)'\hat{\Xi}^{*-1}\hat{D}^*(\hat{\beta}^{e*}_r))^{-1}R(\hat{\beta}^{e*}_r)]^{-1}R(\hat{\beta}^{e*}_r)'(\hat{D}^*(\hat{\beta}^{e*}_r)'\hat{\Xi}^{*-1}\hat{D}^*(\hat{\beta}^{e*}_r))^{-1}]\\ &\times\hat{D}^*(\hat{\beta}^{e*}_r)'\hat{\Xi}^{*-1}\sqrt{T/k_2}(\hat{h}^{a,*}(\beta_0,0)-\hat{h}^a(\beta_0,0))+A^*_T,\end{aligned}$$

where

$$\begin{aligned}A^*_T\;=\;&-[\hat{D}^*(\hat{\beta}^{e*}_r)'\hat{\Xi}^{*-1}\tilde{D}^*(\bar{\beta}^*)]^{-1}\\ &\times[I_p-R(\hat{\beta}^{e*}_r)[R(\hat{\beta}^{e*}_r)'(\hat{D}^*(\hat{\beta}^{e*}_r)'\hat{\Xi}^{*-1}\hat{D}^*(\hat{\beta}^{e*}_r)')^{-1}R(\hat{\beta}^{e*}_r)]^{-1}R(\hat{\beta}^{e*}_r)'(\hat{D}^*(\hat{\beta}^{e*}_r)'\hat{\Xi}^{*-1}\hat{D}^*(\hat{\beta}^{e*}_r))^{-1}]\\ &\times\hat{D}^*(\hat{\beta}^{e*}_r)'\hat{\Xi}^{*-1}\sqrt{T/k_2}\hat{h}^a(\hat{\beta}^e_r,0).\end{aligned}$$

We show now that $A^*_T=o_B(1)$. But by the first order conditions of the original restricted problem we have

$$\hat{D}'\hat{\Xi}^{-1}\hat{h}(\hat{\beta}^e_r,0)-R(\hat{\beta}^e_r)[R(\hat{\beta}^e_r)'(\hat{D}'\hat{\Xi}^{-1}\hat{D})^{-1}R(\hat{\beta}^e_r)]^{-1}R(\hat{\beta}^e_r)'(\hat{D}'\hat{\Xi}^{-1}\hat{D})^{-1}\hat{D}'\hat{\Xi}^{-1}\hat{h}^a(\hat{\beta}^e_r,0)=0.$$

Hence

$$\begin{aligned}A^*_T\;=\;&-[\hat{D}^*(\hat{\beta}^{e*}_r)'\hat{\Xi}^{*-1}\tilde{D}^*(\bar{\beta}^*)]^{-1}\\ &\times[I_p-R(\hat{\beta}^{e*}_r)[R(\hat{\beta}^{e*}_r)'(\hat{D}^*(\hat{\beta}^{e*}_r)'\hat{\Xi}^{*-1}\hat{D}^*(\hat{\beta}^{e*}_r)')^{-1}R(\hat{\beta}^{e*}_r)]^{-1}R(\hat{\beta}^{e*}_r)'(\hat{D}^*(\hat{\beta}^{e*}_r)'\hat{\Xi}^{*-1}\hat{D}^*(\hat{\beta}^{e*}_r))^{-1}]\\ &\times\hat{D}^*(\hat{\beta}^{e*}_r)'\hat{\Xi}^{*-1}\sqrt{T/k_2}\hat{h}^a(\hat{\beta}^e_r,0)\\ &+[\hat{D}^*(\hat{\beta}^{e*}_r)'\hat{\Xi}^{*-1}\tilde{D}^*(\bar{\beta}^*)]^{-1}[I_p-R(\hat{\beta}^e_r)[R(\hat{\beta}^e_r)'(\hat{D}'\hat{\Xi}^{-1}\hat{D})^{-1}R(\hat{\beta}^e_r)]^{-1}R(\hat{\beta}^e_r)'(\hat{D}'\hat{\Xi}^{-1}\hat{D})^{-1}]\\ &\times\hat{D}'\hat{\Xi}^{-1}\hat{h}^a(\hat{\beta}^e_r,0)\\ =\;&o_B(1)\end{aligned}$$

by the bootstrap local UWL and $\sqrt{T/k_2}\hat{h}^a(\hat{\beta}^e_r,0)=O_p(1)$ which can be proven using Taylor expansion and the fact that $\sqrt{T}(\hat{\beta}^e_r-\beta_0)=O_p(1)$.

It follows that

$$\begin{aligned}S_1\sqrt{\frac{T}{k_2}}(\hat{\beta}^{e*}_r-\hat{\beta}^e_r)\;=\;&-[D'\Xi^{-1}D]^{-1}[I-R[R'(D'\Xi^{-1}D)^{-1}R]^{-1}R'(D'\Xi^{-1}D)^{-1}]\\ &\times D'\Xi^{-1}\sqrt{T/k_2}(\hat{h}^{a,*}(\beta_0,0)-\hat{h}^a(\beta_0,0))\\ &+o_B(1).\end{aligned}$$

Consider now the bootstrapped score statistic

$$\mathcal{S}^*=(\frac{T}{k_2})\left[\hat{h}^*(\hat{\beta}^{e*}_r)-\hat{h}(\hat{\beta}^e_r)\right]'\hat{\Xi}^{*-1}\hat{D}^*(\hat{D}^{*\prime}\hat{\Xi}^{*-1}\hat{D}^*)^{-1}\hat{D}^{*\prime}\hat{\Xi}^{*-1}\left[\hat{h}^*(\hat{\beta}^{e*}_r)-\hat{h}(\hat{\beta}^e_r)\right].$$

Note that by a Taylor expansion of $\hat{h}^*(\hat{\beta}^*_r)$ around $\hat{\beta}_r$ we have

$$\begin{aligned}\hat{D}^{*\prime}\hat{\Xi}^{*-1}\sqrt{\frac{T}{k_2}}(\hat{h}^*(\hat{\beta}^{e*}_r)-\hat{h}(\hat{\beta}^e_r))&=\sqrt{\frac{T}{k_2}}\hat{D}^{*\prime}\hat{\Xi}^{*-1}\hat{D}^*(\ddot{\beta}_r)S_1(\hat{\beta}^{e*}_r-\hat{\beta}^e_r)+\sqrt{\frac{T}{k_2}}\hat{D}^{*\prime}\hat{\Xi}^{*-1}(\hat{h}^{a,*}(\hat{\beta}^e_r)-\hat{h}^a(\hat{\beta}^e_r))\\ &=-\hat{D}^{*\prime}\hat{\Xi}^{*-1}\hat{D}^*(\ddot{\beta}_r)[D'\Xi^{-1}D]^{-1}[I-R[R'(D'\Xi^{-1}D)^{-1}R]^{-1}\\ &\quad\times R'(D'\Xi^{-1}D)^{-1}]D'\Xi^{-1}\sqrt{T/k_2}(\hat{h}^{a,*}(\beta_0,0)-\hat{h}^a(\beta_0,0))\\ &\quad+\hat{D}^{*\prime}\hat{\Xi}^{*-1}\sqrt{T/k_2}(\hat{h}^{a,*}(\beta_0,0)-\hat{h}^a(\beta_0,0))+o_B(1)\\ &=[R[R'(D'\Xi^{-1}D)^{-1}R]^{-1}R'(D'\Xi^{-1}D)^{-1}]D'\Xi^{-1}\sqrt{T/k_2}(\hat{h}^{a,*}(\beta_0,0)-\hat{h}(\beta_0,0))+o_B(1)\end{aligned}$$

by A.18, the local bootstrap UWL, and the bootstrap CLT.

Thus

$$\begin{aligned}\mathcal{S}^*\;=\;&(\frac{T}{k_2})\left[\hat{h}^*(\hat{\beta}^{e*}_r)-\hat{h}(\hat{\beta}^e_r)\right]'\hat{\Xi}^{*-1}\hat{D}^*(\hat{D}^{*\prime}\hat{\Xi}^{*-1}\hat{D}^*)^{-1}\hat{D}^{*\prime}\hat{\Xi}^{*-1}\left[\hat{h}^*(\hat{\beta}^{e*}_r)-\hat{h}(\hat{\beta}^e_r)\right]\\ =\;&\sqrt{T/k_2}(\hat{h}^{a,*}(\beta_0,0)-\hat{h}(\beta_0,0))'\Xi^{-1}D(D'\Xi^{-1}D)^{-1}\\ &\times[R'(D'\Xi^{-1}D)^{-1}R]^{-1}R'(\hat{D}^{*\prime}\hat{\Xi}^{*-1}\hat{D}^*)^{-1}[R[R'(D'\Xi^{-1}D)^{-1}R]^{-1}\\ &\times R'(D'\Xi^{-1}D)^{-1}]D'\Xi^{-1}\sqrt{T/k_2}(\hat{h}^{a,*}(\beta_0,0)-\hat{h}(\beta_0,0))+o_B(1)\\ =\;&\sqrt{T/k_2}(\hat{h}^{a,*}(\beta_0,0)-\hat{h}(\beta_0,0))'\Xi^{-1}D(D'\Xi^{-1}D)^{-1}[R'(D'\Xi^{-1}D)^{-1}R]^{-1}R'(D'\Xi^{-1}D)^{-1}]\\ &\times D'\Xi^{-1}\sqrt{T/k_2}(\hat{h}^{a,*}(\beta_0,0)-\hat{h}(\beta_0,0))+o_B(1)\\ =\;&\mathcal{W}^*+o_B(1),\end{aligned}$$

and the result follows.

Now we consider the distance statistic

$$\begin{aligned}\mathcal{D}^* &= (\frac{T}{k_2})[\left[\hat{h}^*(\hat{\beta}_r^{e*}) - \hat{h}(\hat{\beta}_r^e)\right]' \hat{\Xi}^{*-1}\left[\hat{h}^*(\hat{\beta}_r^{e*}) - \hat{h}(\hat{\beta}_r^e)\right] - [\hat{g}^*(\hat{\beta}^{e*}) - \hat{g}(\hat{\beta}^e)]'\tilde{\Omega}^{*-1}[\hat{g}^*(\hat{\beta}^{e*}) - \hat{g}(\hat{\beta}^e)] \\ &= (\frac{T}{k_2})[\left[\hat{h}^{a,*}(\hat{\beta}_r^{e*},0) - \hat{h}^a(\hat{\beta}_r^e,0)\right]' \hat{\Xi}^{*-1}\left[\hat{h}^{a,*}(\hat{\beta}_r^{e*},0) - \hat{h}^a(\hat{\beta}_r^e,0)\right] \\ &\quad - \left[\hat{h}^{a,*}(\hat{\beta}^{e*},\hat{\gamma}^*) - \hat{h}^a(\hat{\beta}^e,\hat{\gamma})\right]' \tilde{\Xi}^{*-1}\left[\hat{h}^{a,*}(\hat{\beta}^{e*},\hat{\gamma}^*) - \hat{h}^{a,*}(\hat{\beta}^e,\hat{\gamma})\right] + o_B(1),\end{aligned}$$

as

$$\hat{g}^*(\hat{\beta}^{e*})'\tilde{\Omega}^{*-1}\hat{g}^*(\hat{\beta}^{e*}) = \hat{h}^{a,*}(\hat{\beta}^{e*},\hat{\gamma}^*)'\tilde{\Xi}^{*-1}\hat{h}^{a,*}(\hat{\beta}^{e*},\hat{\gamma}^*),$$

and

$$\begin{aligned}T\hat{g}(\hat{\beta}^e)'\tilde{\Omega}^{-1}\hat{g}(\hat{\beta}^e) &= T\hat{h}^a(\hat{\beta}^e,\hat{\gamma})'\tilde{\Xi}^{-1}\hat{h}^a(\hat{\beta}^e,\hat{\gamma}) \\ &= T\hat{h}^a(\hat{\beta}^e,\hat{\gamma})'\tilde{\Xi}^{*-1}\hat{h}^a(\hat{\beta}^e,\hat{\gamma}) \\ &\quad + T\hat{h}^a(\hat{\beta}^e,\hat{\gamma})'[\tilde{\Xi}^{-1} - \tilde{\Xi}^{*-1}]\hat{h}^a(\hat{\beta}^e,\hat{\gamma}) \\ &= T\hat{h}^a(\hat{\beta}^e,\hat{\gamma})'\tilde{\Xi}^{*-1}\hat{h}^a(\hat{\beta}^e,\hat{\gamma}) + o_B(1),\end{aligned}$$

since $\sqrt{T}\hat{h}^a(\hat{\beta}^e,\hat{\gamma}) = O_p(1)$ and $\tilde{\Xi}^{-1} - \tilde{\Xi}^{*-1} = o_B(1)$.

Now note that by two first order Taylor expansions we have

$$\begin{aligned}\hat{h}^{a,*}(\hat{\beta}_r^{e*},0) - \hat{h}^a(\hat{\beta}_r^e,0) &= \hat{h}^{a,*}(\hat{\beta}^{e*},\hat{\gamma}^*) - \hat{h}^a(\hat{\beta}^e,\hat{\gamma}) \\ &\quad + D^*(\bar{\beta}^*)\begin{pmatrix}\hat{\beta}_r^{e*} - \hat{\beta}^{e*} \\ \hat{\gamma}^*\end{pmatrix} - D(\bar{\beta})\begin{pmatrix}\hat{\beta}_r^e - \hat{\beta}^e \\ \hat{\gamma}\end{pmatrix},\end{aligned}$$

where $\bar{\beta}^*$ is in a line joining $\hat{\beta}_r^{e*}$ and $\hat{\beta}^{e*}$ and $\bar{\beta}$ is in a line joining $\hat{\beta}_r^e$ and $\hat{\beta}^e$. Thus

$$\begin{aligned}&(\frac{T}{k_2})[\hat{h}^{a,*}(\hat{\beta}_r^{e*},0) - \hat{h}^a(\hat{\beta}_r^e,0)]'\hat{\Xi}^{*-1}[\hat{h}^{a,*}(\hat{\beta}_r^{e*},0) - \hat{h}^a(\hat{\beta}_r^e,0)] \\ =\ &(\frac{T}{k_2})[\hat{h}^{a,*}(\hat{\beta}^{e*},\hat{\gamma}^*) - \hat{h}^a(\hat{\beta}^e,\hat{\gamma})]'\hat{\Xi}^{*-1}[\hat{h}^{a,*}(\hat{\beta}^{e*},\hat{\gamma}^*) - \hat{h}^a(\hat{\beta}^e,\hat{\gamma})] \\ &+(\frac{T}{k_2})2[D^*(\bar{\beta}^*)\begin{pmatrix}\hat{\beta}_r^{e*} - \hat{\beta}^{e*} \\ \hat{\gamma}^*\end{pmatrix} - D(\bar{\beta})\begin{pmatrix}\hat{\beta}_r^e - \hat{\beta}^e \\ \hat{\gamma}\end{pmatrix}]'\hat{\Xi}^{*-1}[\hat{h}^{a,*}(\hat{\beta}^{e*},\hat{\gamma}^*) - \hat{h}^a(\hat{\beta}^e,\hat{\gamma})] \\ &+[D^*(\bar{\beta}^*)\begin{pmatrix}\hat{\beta}_r^{e*} - \hat{\beta}^{e*} \\ \hat{\gamma}^*\end{pmatrix} - D(\bar{\beta})\begin{pmatrix}\hat{\beta}_r^e - \hat{\beta}^e \\ \hat{\gamma}\end{pmatrix}]'\hat{\Xi}^{*-1}[D^*(\bar{\beta}^*)\begin{pmatrix}\hat{\beta}_r^{e*} - \hat{\beta}^{e*} \\ \hat{\gamma}^*\end{pmatrix} - D(\bar{\beta})\begin{pmatrix}\hat{\beta}_r^e - \hat{\beta}^e \\ \hat{\gamma}\end{pmatrix}].\end{aligned}$$

Note that $D^*(\hat{\beta}^{e*})'\hat{\Xi}^{*-1}\hat{h}^{a,*}(\hat{\beta}^{e*},\hat{\gamma}^*) = 0$ and $\hat{D}\hat{\Xi}\hat{h}^a(\hat{\beta}^e,\hat{\gamma}) = 0$. Thus

$$\begin{aligned}\sqrt{\frac{T}{k_2}}D^*(\bar{\beta}^*)'\hat{\Xi}^{*-1}\hat{h}^{a,*}(\hat{\beta}^{e*},\hat{\gamma}^*) &= \sqrt{\frac{T}{k_2}}(D^*(\bar{\beta}^*)'\hat{\Xi}^{*-1} - D^*(\hat{\beta}^{e*})'\hat{\Xi}^{*-1})\sqrt{\frac{T}{k_2}}\hat{h}^{a,*}(\hat{\beta}^{e*},\hat{\gamma}^*) = o_B(1), \\ \sqrt{\frac{T}{k_2}}D^*(\bar{\beta}^*)'\hat{\Xi}^{*-1}\hat{h}^a(\hat{\beta}^e,\hat{\gamma}) &= \sqrt{\frac{T}{k_2}}(D^*(\bar{\beta}^*)'\hat{\Xi}^{*-1} - \hat{D}\hat{\Xi})\hat{h}^a(\hat{\beta}^e,\hat{\gamma}) = o_B(1).\end{aligned}$$

by the bootstrap UWL, the standard UWL, $\sqrt{\frac{T}{k_2}}\hat{h}^{a,*}(\hat{\beta}^{e*},\hat{\gamma}^*) = O_B(1)$ and $\sqrt{T}\hat{h}^a(\hat{\beta}^e,\hat{\gamma}) = O_p(1)$. Thus

$$\sqrt{\frac{T}{k_2}}D^*(\bar{\beta}^*)'\hat{\Xi}^{*-1}\left[\hat{h}^{a,*}(\hat{\beta}^*,\hat{\gamma}^*) - \hat{h}^a(\hat{\beta},\hat{\gamma})\right] = o_B(1),$$

and similarly $\sqrt{\frac{T}{k_2}}D(\bar{\beta})'\hat{\Xi}^{*-1}\left[\hat{h}^{a,*}(\hat{\beta}^*,\hat{\gamma}^*) - \hat{h}^a(\hat{\beta},\hat{\gamma})\right] = o_B(1)$.

Now as $\bar{D}^{*\prime}\hat{\Xi}^{*-1}\left[\hat{h}^{a,*}(\hat{\beta}^*,\hat{\gamma}^*) - \hat{h}^a(\hat{\beta},\hat{\gamma})\right] = o_B(1/\sqrt{T})$ and $\bar{D}'\hat{\Xi}^{*-1}\left[\hat{h}^{a,*}(\hat{\beta}^*,\hat{\gamma}^*) - \hat{h}^a(\hat{\beta},\hat{\gamma})\right] = o_B(1/\sqrt{T})$ by the first order conditions and the bootstrap UWL and the standard UWL.

Also

$$\sqrt{\frac{T}{k_2}}[D^*(\bar{\beta}^*)\begin{pmatrix}\hat{\beta}_r^{e*} - \hat{\beta}^{e*} \\ \hat{\gamma}^*\end{pmatrix} - D(\bar{\beta})\begin{pmatrix}\hat{\beta}_r^e - \hat{\beta}^e \\ \hat{\gamma}\end{pmatrix}] = O_B(1),$$

as $\sqrt{T}\left(\hat{\beta}_r^{e*} - \hat{\beta}_r^e\right) = O_B(1)$, $\sqrt{T}\left(\hat{\beta}^{e*} - \hat{\beta}^e\right) = O_B(1)$, $\sqrt{T}\left(\hat{\beta}^e - \beta_0\right) = O_p(1)$, $\sqrt{T}\left(\hat{\beta}_r^e - \beta_0\right) = O_p(1)$. Thus

$$(\frac{T}{k_2})2[D^*(\bar{\beta}^*)\begin{pmatrix}\hat{\beta}_r^{e*} - \hat{\beta}^{e*} \\ \hat{\gamma}^*\end{pmatrix} - D(\bar{\beta})\begin{pmatrix}\hat{\beta}_r^e - \hat{\beta}^e \\ \hat{\gamma}\end{pmatrix}]'\hat{\Xi}^{*-1}[\hat{h}^{a,*}(\hat{\beta}^{e*},\hat{\gamma}^*) - \hat{h}^a(\hat{\beta}^e,\hat{\gamma})] = o_B(1).$$

Now notice that $D^*(\bar{\beta}^*) = D + o_B(1)$ and $D(\bar{\beta}) = D + o_p(1)$, thus

$$\sqrt{\frac{T}{k_2}}(D^*(\bar{\beta}^*)\begin{pmatrix}\hat{\beta}_r^{e*} - \hat{\beta}^{e*} \\ \hat{\gamma}^*\end{pmatrix} - D(\bar{\beta})\begin{pmatrix}\hat{\beta}_r^e - \hat{\beta}^e \\ \hat{\gamma}\end{pmatrix}) = \sqrt{\frac{T}{k_2}}D[\begin{pmatrix}\hat{\beta}_r^{e*} - \hat{\beta}_r^e \\ 0\end{pmatrix} - \begin{pmatrix}\hat{\beta}^{e*} - \hat{\beta}^e \\ \hat{\gamma}^* - \hat{\gamma}\end{pmatrix}] + o_B(1),$$

and consequently

$$\sqrt{\frac{T}{k_2}}D[\left(\begin{array}{c}\hat{\beta}_r^{e*}-\hat{\beta}_r^e\\0\end{array}\right)-\left(\begin{array}{c}\hat{\beta}^{e*}-\hat{\beta}^e\\\hat{\gamma}^*-\hat{\gamma}\end{array}\right)] = \sqrt{\frac{T}{k_2}}D[D'\Xi^{-1}D]^{-1}[R[R'(D'\Xi^{-1}D)^{-1}R]^{-1}R'(D'\Xi^{-1}D)^{-1}]$$
$$\times D'\Xi^{-1}\sqrt{T/k_2}(\hat{h}^{a,*}(\beta_0,0)-\hat{h}^a(\beta_0,0))+o_B(1).$$

Thus

$$\begin{aligned}
&\frac{T}{k_2}[D^*(\bar{\beta}^*)\left(\begin{array}{c}\hat{\beta}_r^{e*}-\hat{\beta}^{e*}\\\hat{\gamma}^*\end{array}\right)-D(\bar{\beta})\left(\begin{array}{c}\hat{\beta}_r^{e}-\hat{\beta}^{e}\\\hat{\gamma}\end{array}\right)]'\hat{\Xi}^{*-1}[D^*(\bar{\beta}^*)\left(\begin{array}{c}\hat{\beta}_r^{e*}-\hat{\beta}^{e*}\\\hat{\gamma}^*\end{array}\right)-D(\bar{\beta})\left(\begin{array}{c}\hat{\beta}_r^{e}-\hat{\beta}^{e}\\\hat{\gamma}\end{array}\right)]\\
= &\sqrt{T/k_2}(\hat{h}^{a,*}(\beta_0,0)-\hat{h}^a(\beta_0,0))'\Xi^{-1}D(D'\Xi^{-1}D)^{-1}[R'(D'\Xi^{-1}D)^{-1}R]^{-1}R'(D'\Xi^{-1}D)^{-1}]\\
&\times D'\Xi^{-1}\sqrt{T/k_2}(\hat{h}^{a,*}(\beta_0,0)-\hat{h}^a(\beta_0,0))+o_B(1)\\
= &\mathcal{W}^*+o_B(1).
\end{aligned}$$

■

A.5 Proofs of the results in section 4.2

Proof of 4.7: We only need to show that the regularity conditions of the lemma A.2 of Gonçalves and White (2004) . Condition (a1) is satisfied as $g(.,\beta)$ is measurable and continuous functions of measurable functions are measurable. Since $g(z_t,\beta)$ is continuous on B the objective function is continuous $\hat{g}(\beta)'W_T\hat{g}(\beta)$ is continuous. Also Note that by T.

$$\begin{aligned}
&\sup_{\beta\in B}\left|\hat{g}_T^\star(\beta)'W_T^\star\hat{g}_T^\star(\beta)-\mathrm{E}^\star[\hat{g}_T^\star(\beta)]'W_T\mathrm{E}^\star[\hat{g}_T^\star(\beta)]\right|\\
\leq\ &\sup_{\beta\in B}\left|\hat{g}_T^\star(\beta)'W_T^\star\hat{g}_T^\star(\beta)-\mathrm{E}^\star[\hat{g}_T^\star(\beta)]'W_T^\star\mathrm{E}^\star[\hat{g}_T^\star(\beta)]\right|\\
&+\sup_{\beta\in B}\left|\mathrm{E}^\star[\hat{g}_T^\star(\beta)]'W_T^\star\mathrm{E}^\star[\hat{g}_T^\star(\beta)]'-\mathrm{E}^\star[\hat{g}_T^\star(\beta)]'W_T\mathrm{E}^\star[\hat{g}_T^\star(\beta)]\right|.
\end{aligned}$$

Now by T

$$\begin{aligned}
&\sup_{\beta\in B}\left|\hat{g}_T^\star(\beta)'W_T^\star\hat{g}_T^\star(\beta)-\mathrm{E}[\hat{g}_T^\star(\beta)]'W_T^\star\mathrm{E}[\hat{g}_T^\star(\beta)]\right|=\\
&\leq\sup_{\beta\in B}\|\hat{g}_T^\star(\beta)-\mathrm{E}^\star[\hat{g}_T^\star(\beta)]\|^2\,\|W_T^\star\|\\
&+2\sup_{\beta\in B}\|\hat{g}_T^\star(\beta)-\mathrm{E}^\star[\hat{g}_T^\star(\beta)]\|\,\|W_T^\star\|\sup_{\beta\in B}\|\mathrm{E}^\star[\hat{g}_T^\star(\beta)]\|.
\end{aligned}$$

Now for $p_{tT}=\hat{\pi}_t$, note that by Lemma A.1 Assumption 3.3 (a) is satisfied. Hence the bootstrap UWL and the local UWL given by Lemmata A.6 and A.7 can be applied and therefore $\sup_{\beta\in B}\left\|\hat{g}_T^\star(\beta)-\mathrm{E}^\star[\hat{g}_T^\star(\beta)]\right\|=o_B(1)$. $\left\|W_T^\star-W_T\right\|=o_B(1)$ and $W_T=O_p(1)$.

$$\begin{aligned}
\mathrm{E}^\star[\hat{g}_T^\star(\beta)] &= \sum_{i=1}^T\hat{g}_{tT}(\beta)\hat{\pi}_t\\
&= (1+o_p(1))\frac{1}{T}\sum_{i=1}^T g_{tT}(\beta)\\
&= O_p(1),
\end{aligned}$$

by Lemma A1 of Smith (2011). Thus

$$\sup_{\beta\in B}\left|\hat{g}_T^\star(\beta)'W_T^\star\hat{g}_T^\star(\beta)-\mathrm{E}[\hat{g}_T^\star(\beta)]'W_T^\star\mathrm{E}[\hat{g}_T^\star(\beta)]\right|=o_B(1).$$

Now

$$\begin{aligned}
&\sup_{\beta\in B}\left|\mathrm{E}^\star[\hat{g}_T^\star(\beta)]'W_T^\star\mathrm{E}^\star[\hat{g}_T^\star(\beta)]'-\mathrm{E}^\star[\hat{g}_T^\star(\beta)]'W_T\mathrm{E}^\star[\hat{g}_T^\star(\beta)]'\right|\\
=\ &\sup_{\beta\in B}\left|\mathrm{E}^\star[\hat{g}_T^\star(\beta)]'[W_T^\star-W_T]\mathrm{E}^\star[\hat{g}_T^\star(\beta)]'\right|\\
\leq\ &\sup_{\beta\in B}\|\mathrm{E}^\star[\hat{g}_T^\star(\beta)]\|\,\|[W_T^\star-W_T]\|\\
=\ &O_p(1)o_p(1).
\end{aligned}$$

Uniqueness was proven in Lemma 2.3 of Newey and MacFadden (1994). ■

Proof of 4.8: Note that by that by Hansen (1982) we have

$$\sqrt{T}(\hat{\beta}-\beta)\xrightarrow{d}N(0,(G'WG)^{-1}G'W\Omega WG(G'WG)^{-1}),$$

as since the normal is continuous we have for $\Gamma = (G'WG)^{-1}G'W\Omega WG(G'WG)^{-1}$

$$\sup_{x\in\mathbb{R}^p}\left|\mathcal{P}\{\Gamma^{-1/2}T^{1/2}(\hat{\beta}-\beta_0)\leq x\} - \Phi(x)\right|$$

by Polya's Theorem.

We prove now that

$$\lim_{T\to\infty}\mathcal{P}\left\{\sup_{x\in\mathbb{R}^p}\left|\mathcal{P}^\star\{\Gamma^{-1/2}\sqrt{\frac{T}{k_2}}(\hat{\beta}^\star-\tilde{\beta})\leq x\} - \Phi(x)\right|\geq\varepsilon\right\} = 0.$$

Let $\hat{G}^\star_T \equiv \partial\hat{g}^\star_T(\hat{\beta}^*)/\partial\beta'$, To prove asymptotic normality notice that by the FOC $\sqrt{T/k_2}\hat{G}^{\star\prime}_T W^\star_T \hat{g}^\star_T(\hat{\beta}^\star) = 0$. Hence a first order Taylor expansion around $\hat{\beta}$ yields

$$\sqrt{T/k_2}\hat{G}^{\star\prime}_T W^\star_T \hat{g}^\star_T(\tilde{\beta}) + \hat{G}^{\star\prime}_T W^\star_T \tilde{G}^\star_T\sqrt{T/k_2}(\hat{\beta}^\star-\tilde{\beta}) = 0,$$

where $\bar{G}^*_T = \partial\hat{g}^\star_T(\bar{\beta}^\star)/\partial\beta'$ and $\bar{\beta}^*$ is on a line joining $\tilde{\beta}$ and $\hat{\beta}^\star$. Solving for $\sqrt{T/k_2}(\hat{\beta}^\star-\tilde{\beta})$ we obtain

$$\sqrt{T/k_2}(\hat{\beta}^\star-\tilde{\beta}) = -[\hat{G}^{\star\prime}_T W^\star_T \bar{G}^\star_T]^{-1}\sqrt{T/k_2}\hat{G}^{\star\prime}_T W^\star_T \hat{g}^\star_T(\tilde{\beta}).$$

Now notice that by a Taylor expansion

$$\begin{aligned}
&\sqrt{T/k_2}\hat{G}^{\star\prime}_T W^\star_T \hat{g}^\star_T(\tilde{\beta}) = \sqrt{T/k_2}\hat{G}^{\star\prime}_T W^\star_T \hat{g}^\star_T(\beta_0) + \sqrt{T/k_2}\hat{G}^{\star\prime}_T W^\star_T \check{G}^\star_T(\tilde{\beta}-\beta_0)\\
=\ &\sqrt{T/k_2}\hat{G}^{\star\prime}_T W^\star_T[\hat{g}^\star_T(\beta_0)-\tilde{g}_T(\beta_0)] + \sqrt{T/k_2}\hat{G}^{\star\prime}_T W^\star_T\tilde{g}_T(\beta_0) + \sqrt{T/k_2}\hat{G}^{\star\prime}_T W^\star_T\check{G}^\star_T(\tilde{\beta}-\beta_0),
\end{aligned}$$

where $\check{G}^\star_T \equiv \partial\hat{g}^\star_T(\check{\beta})/\partial\beta'$ and $\tilde{g}_T(\beta_0) = \sum_{t=1}^T g_{t,T}(\beta_0)\tilde{\pi}_t$ and $\check{\beta}$ is on a line joining $\hat{\beta}$ and β_0.

Now note that by $(A.2)$ we have

$$\sqrt{T}(\tilde{\beta}-\beta_0) = -(\hat{G}'_T W_T \ddot{G}_T)^{-1}\sqrt{T}\hat{G}'_T W_T\tilde{g}_T(\beta_0).$$

Thus

$$\begin{gathered}
\sqrt{T/k_2}\hat{G}^{\star\prime}_T W^\star_T\tilde{g}_T(\beta_0) + \sqrt{T/k_2}\hat{G}^{\star\prime}_T W^\star_T\tilde{G}^\star_T(\tilde{\beta}-\beta_0) = \\
\sqrt{T/k_2}\hat{G}^{\star\prime}_T W^\star_T\tilde{g}_T(\beta_0) - \sqrt{T/k_2}\hat{G}^{\star\prime}_T W^\star_T\tilde{G}^*_T(\hat{G}'_T W_T G^*_T)^{-1}\hat{G}'_T W_T\tilde{g}_T(\beta_0) = o_B(1)
\end{gathered}$$

since $\sqrt{T}\tilde{g}_T(\beta_0) = O_p(1)$, $W_T = W + o_p(1)$, $\hat{G}_T = G + o_p(1)$, $\hat{G}^\star_T = G + o_B(1)$, $W^\star_T = W_T + o_B(1)$.

Now $\sqrt{T/k_2}\hat{G}^{\star\prime}_T W^\star_T[\hat{g}^*_T(\beta_0)-\tilde{g}_T(\beta_0)]$ converges to $N(0,(G'WG)^{-1}G'W\Omega WG(G'WG)^{-1})$ by bootstrap CLT Theorem A.2 and the fact that $\hat{G}^*_T - G = o_B(1)$ and $W^*_T = W + o_B(1)$. The result follows as the $\sqrt{T/k_2}(\hat{\beta}^*-\tilde{\beta})$ converges uniformly to the same asymptotic distribution of $T^{1/2}(\hat{\beta}-\beta_0)$. We note that $\tilde{\beta}$ can be replaced by $\hat{\beta}_e$ because $\sqrt{T}(\tilde{\beta}-\beta_0) - \sqrt{T}(\hat{\beta}_e-\beta_0) = o_p(1)$. ■

Proof of Lemma 4.2: The proof of this Lemma is identical to the proof of Lemma 4.1 with $p_{tT} = \tilde{\pi}_t$ and uses the fact that $T\tilde{\pi}_t = 1 + o_p(1)$ by Lemma A.1. ■

Proof of Theorem 4.9: Note that by a Taylor expansion

$$\sqrt{T/k_2}\hat{g}^\star(\hat{\beta}^{e\star}) = \sqrt{T/k_2}\hat{g}^\star(\tilde{\beta}) + \tilde{G}^*\sqrt{T/k_2}(\hat{\beta}^{e*}-\tilde{\beta}),$$

where $\tilde{G}^*_T \equiv \partial\hat{g}^\star_T(\ddot{\beta})/\partial\beta'$ where $\ddot{\beta}$ is in a line joining $\hat{\beta}^{e\star}$ and $\tilde{\beta}$.

Note that by Theorem 4.8 with $W^\star_T = \tilde{\Omega}^{\star-1}$

$$\sqrt{T/k_2}(\hat{\beta}^{e\star}-\tilde{\beta}) = -[\hat{G}^{\star\prime}_T\tilde{\Omega}^{\star-1}\tilde{G}^\star_T]^{-1}\hat{G}^{\star\prime}_T\tilde{\Omega}^{\star-1}\sqrt{T/k_2}[\hat{g}^\star_T(\beta_0)-\tilde{g}_T(\beta_0)] + o_B(1).$$

Also by a Taylor expansion

$$\sqrt{T/k_2}(\hat{g}^\star(\tilde{\beta}) - \hat{g}^\star(\beta_0) - \tilde{g}_T(\tilde{\beta}) + \tilde{g}_T(\beta_0)) = (\check{G}^\star_T - \ddot{G}_T)\sqrt{T/k_2}\left(\tilde{\beta}-\beta_0\right),$$

where $\check{G}^\star_T = \partial\hat{g}^\star_T(\check{\beta})/\partial\beta'$ where $\check{\beta}$ is in a line joining $\tilde{\beta}$ and β_0 and $\ddot{G}_T = \partial\tilde{g}_T(\ddot{\beta})/\partial\beta'$ where $\ddot{\beta}$ is in a line joining $\tilde{\beta}$ and β_0.

Now $\check{G}_T = G + o_B(1)$ by Lemma A.7, $\ddot{G}_T = G + o_p(1)$ by Lemma A.1 of Smith (2011) and the fact that $T\tilde{\pi}_t = 1 + o_p(1)$ by Lemma A.1. Also by Theorem 4.8 $\sqrt{T}\left(\tilde{\beta}-\beta_0\right) = O_p(1)$.

Now we show that $\sqrt{T/k_2}\tilde{g}(\tilde{\beta}) = o_p(1)$.Note that by a Taylor expansion

$$\sqrt{T}\tilde{g}\left(\tilde{\beta}\right) = \sqrt{T}\tilde{g}\left(\beta_0\right) + \breve{G}_T\sqrt{T}(\tilde{\beta}-\beta_0),$$

where $\breve{G}_T = \partial\tilde{g}_T(\breve{\beta})/\partial\beta'$ where $\breve{\beta}$ is in a line joining $\tilde{\beta}$ and β_0. $\ddot{G}_T = G + o_p(1)$ by Lemma A.1 of Smith (2011)- and the fact that $T\tilde{\pi}_t = 1 + o_p(1)$ by Lemma A.1. Thus by Theorem 4.8 we have

$$\breve{G}_T\sqrt{T}(\tilde{\beta}-\beta_0) = G\Sigma G'\Omega^{-1}T^{1/2}\hat{g}(\beta_0) + o_p(1).$$

Now by Lemma A.2 we have

$$\sqrt{T}\tilde{g}(\beta_0) = [G\Sigma G'\Omega^{-1}]T^{1/2}\hat{g}(\beta_0) + o_p(1)$$

Hence$\sqrt{T}\tilde{g}(\tilde{\beta}) = o_p(1)$.Thus

$$\begin{aligned}\sqrt{\frac{T}{k_2}}\hat{g}^\star(\hat{\beta}^{e\star}) &= \sqrt{T/k_2}(\hat{g}^\star(\beta_0) - \tilde{g}(\beta_0)) - [\hat{G}_T^{\star\prime}\tilde{\Omega}^{\star-1}\tilde{G}_T^\star]^{-1}\hat{G}_T^{\star\prime}\tilde{\Omega}^{\star-1}\sqrt{T/k_2}[\hat{g}_T^\star(\beta_0) - \tilde{g}_T(\beta_0)] + o_B(1) \\ &= [I_m - \tilde{G}^\star[\hat{G}_T^{\star\prime}\tilde{\Omega}^{\star-1}\tilde{G}_T^\star]^{-1}\hat{G}_T^{\star\prime}\tilde{\Omega}^{\star-1}]\sqrt{T/k_2}[\hat{g}_T^\star(\beta_0) - \hat{g}_T(\beta_0)] + o_B(1).\end{aligned}$$

Now since $\tilde{G}^* = G + o_B(1)$, $\hat{G}_T^* = G + o_B(1)$ $\tilde{\Omega}^{*-1} = \Omega^{-1} + o_B(1)$ and by the bootstrap CLT Theorem A.2 $\sqrt{T/k_2}[\hat{g}_T^*(\beta_0) - \hat{g}_T(\beta_0)]$ converges to $N(0,\Omega)$ It follows as in Theorem 4.3 that $\mathcal{J}^\star = T\hat{g}^\star(\hat{\beta}^{e\star})'\tilde{\Omega}^{\star-1}\hat{g}^\star(\hat{\beta}^{e\star})/k_2$ converges to $\chi^2(m-p)$. Since $\mathcal{J} \xrightarrow{d} \chi^2(m-p)$ the result follows by Polya Theorem Serfling (2002, p.18), as the chi-squared distribution has a continuous c.d.f. ■

Proof of 4.10: We start by deriving the asymptotic distribution of $\mathcal{W}^\star$. Define $h_t^{a,\star}(\beta,\gamma) \equiv (g^\star(z_t,\beta)', [q^\star(z_t,\beta) - \gamma]')$, $\hat{h}^{a,\star}(\beta,\gamma) \equiv \sum_{t=1}^{m_T} h_t^{a,\star}(\beta,\gamma)/m_T$ and $\tilde{Q}^\star(\beta,\gamma) = \hat{h}^{a,\star}(\beta,\gamma)'\hat{\Xi}^{\star-1}\hat{h}^{a,\star}(\beta,\gamma)$. Note that the unrestricted GMM estimator solves

$$(\hat{\beta}^{e\star\prime}, \hat{\gamma}^{\star\prime})' = \underset{\beta\in B,\gamma\in\mathbb{R}^m}{\arg\min}\ \tilde{Q}^\star(\beta,\gamma).$$

As before the solution is given by

$$\begin{aligned}\hat{\beta}^{e\star} &= \underset{\beta\in B}{\arg\min}\hat{g}^\star(\beta)'\hat{\Omega}^{\star-1}\hat{g}^\star(\beta), \\ \hat{\gamma}^\star &= \hat{q}^\star(\hat{\beta}^{e\star}) - \hat{\Xi}_{21}^\star\hat{\Omega}^{\star-1}g^\star(\hat{\beta}^{e\star}).\end{aligned}$$

Consistency of $\hat{\beta}^{e\star}$ follows from Theorem 4.7. We note that by Lemma A.6 and $\hat{\Xi}^* = \hat{\Xi} + o_B(1)$ and $\hat{\gamma}^\star = \hat{\gamma} + o_B(1)$.

We derive now the asymptotic distribution of $(\hat{\beta}^{e\star\prime}, \hat{\gamma}^{\star\prime})'$.Since these estimators satisfy the first order conditions we have $\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{h}^{a,\star}(\hat{\beta}^{e\star},\hat{\gamma}^\star) = 0$. Thus by a Taylor expansion around $\left(\hat{\beta}^{e\prime},\hat{\gamma}'\right)'$

$$\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{h}^{a,\star}(\hat{\beta}^e,\hat{\gamma}) + \hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\tilde{D}^\star\begin{pmatrix}\hat{\beta}^{e\star} - \hat{\beta}^e \\ \hat{\gamma}^{e\star} - \hat{\gamma}\end{pmatrix} = 0,$$

where $\ddot{D}^\star \equiv D^\star(\ddot{\beta}^\star)$

$$D^\star(\beta) = \begin{pmatrix}\sum_{i=1}^{m_T}\hat{G}_t^\star(\beta)/m_T & 0 \\ \sum_{i=1}^{m_T}\hat{Q}_t^\star(\beta)/m_T & -I_s\end{pmatrix},$$

and $\ddot{\beta}^*$ is in a line joining $\hat{\beta}^{e\star}$ and $\hat{\beta}^e$. Thus

$$\sqrt{T}\begin{pmatrix}\hat{\beta}^{e\star} - \hat{\beta}^e \\ \hat{\gamma}^{e\star} - \hat{\gamma}\end{pmatrix} = -[\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\tilde{D}^\star]^{-1}\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\sqrt{T}\hat{h}^{a,\star}(\hat{\beta}^e,\hat{\gamma}).$$

Now by a Taylor expansion

$$\sqrt{T}\hat{h}^{a,\star}(\hat{\beta}^e,\hat{\gamma}) = T^{1/2}\hat{h}^{a,\star}(\beta_0,0) + \bar{D}^\star\begin{pmatrix}\hat{\beta}^e - \beta_0 \\ \hat{\gamma}\end{pmatrix},$$

where $\bar{D}^\star = D^\star(\bar{\beta})$,where $\bar{\beta}$ is in a line joining $\hat{\beta}$ and β_0.

We show now that

$$[\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\ddot{D}^\star]^{-1}\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}[T^{1/2}\tilde{h}_T^a(\beta_0,0) - T^{1/2}\bar{D}^\star\begin{pmatrix}\hat{\beta}^e - \beta_0 \\ \hat{\gamma}\end{pmatrix}] = o_B(1).$$

First notice that by Lemma A.2 above we have

$$\begin{aligned}T^{1/2}\tilde{h}_T^a(\beta_0,0) &= T^{-1/2}\hat{h}_T^a(\beta_0,0) - \begin{pmatrix}\Xi_{11} \\ \Xi_{21}\end{pmatrix}PT^{1/2}\hat{g}(\beta_0) + o_p(1) \qquad \text{(A.19)} \\ &= T^{-1/2}\sum_{t=1}^T\hat{h}_T^a(\beta_0,0) + \Xi S_1 PT^{1/2}\hat{g}(\beta_0) + o_p(1).\end{aligned}$$

Thus as $\hat{\Xi}^{\star-1} = \Xi^{-1} + o_B(1)$ we have

$$\begin{aligned}&[\hat{D}^{\star\prime}\hat{\Xi}\ddot{D}^\star]^{-1}\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}T^{1/2}\tilde{h}_T^a(\beta_0,0) \\ =\ &[\hat{D}^{\star\prime}\hat{\Xi}\ddot{D}^\star]^{-1}\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}T^{-1/2}\sum_{t=1}^T\hat{h}_T^a(\beta_0,0) \\ &-[\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\ddot{D}^\star]^{-1}\hat{D}^{\star\prime}S_1PT^{1/2}\hat{g}(\beta_0) + o_p(1) \\ =\ &[\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\ddot{D}^\star]^{-1}\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}T^{-1/2}\sum_{t=1}^T\hat{h}_T^a(\beta_0,0) \\ &-[\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\ddot{D}^\star]^{-1}\begin{bmatrix}\hat{G}^{\star\prime} & \hat{Q}^{\star\prime}\end{bmatrix}\begin{bmatrix}PT^{1/2}\hat{g}(\beta_0) \\ 0\end{bmatrix} \\ =\ &[\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\ddot{D}^\star]^{-1}\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}T^{-1/2}\sum_{t=1}^T\hat{h}_T^a(\beta_0,\gamma) + o_B(1),\end{aligned}$$

as $\hat{G}^{\star} = G + o_B(1)$ by Lemma A.7 and $G'P = 0$.

Now notice that

$$\sqrt{T}\begin{pmatrix} \hat{\beta}^e - \beta_0 \\ \hat{\gamma} \end{pmatrix} = \sqrt{T}\begin{pmatrix} \hat{\beta}^e - \beta_0 \\ \tilde{\gamma} \end{pmatrix} - \sqrt{T}\begin{pmatrix} 0 \\ \tilde{\gamma} - \hat{\gamma} \end{pmatrix} \quad (A.20)$$

and the usual asymptotic representation of the efficient GMM estimator yields

$$\sqrt{T}\begin{pmatrix} \hat{\beta}^e - \beta_0 \\ \hat{\gamma} \end{pmatrix} = -[\hat{D}'\hat{\Xi}^{-1}\ddot{D}]^{-1}\hat{D}'\hat{\Xi}^{-1}T^{-1/2}\sum\nolimits_{t=1}^{T}\hat{h}_t^a(\beta_0, 0) + o_p(1), \quad (A.21)$$

where $\ddot{D} = D(\ddot{\beta})$ and $\ddot{\beta}$ is in a line joining $\hat{\beta}^e$ and β_0. Hence by $(A.20)$ and $(A.21)$ we have

$$\begin{aligned} [\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\ddot{D}^{\star}]^{-1}\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\bar{D}^{\star}\sqrt{T}\begin{pmatrix} \hat{\beta}^e - \beta_0 \\ \hat{\gamma} \end{pmatrix} &= [\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\ddot{D}^{\star}]^{-1}\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\bar{D}^{\star}\sqrt{T}\begin{pmatrix} \hat{\beta}^e - \beta_0 \\ \tilde{\gamma} \end{pmatrix} \\ &\quad -O_B(1)\sqrt{T}\begin{pmatrix} 0 \\ \tilde{\gamma} - \hat{\gamma} \end{pmatrix} \end{aligned}$$

as $\hat{D}^{\star}$, $\ddot{D}^{\star}$ $\bar{D}^{\star}$ converge to D by the Lemma A.7 , and the fact that $\hat{\Xi}^{\star-1} = \Xi^{-1} + o_B(1)$. It remains to prove that $\sqrt{T}\,(\tilde{\gamma} - \hat{\gamma}) = o_p(1)$.

First note that

$$\sqrt{T}\,(\tilde{\gamma} - \hat{\gamma}) = \sqrt{T}\tilde{q}(\hat{\beta}^e) - \sqrt{T}\hat{q}(\hat{\beta}^e) + \hat{\Xi}_{21}\hat{\Xi}_{11}^{-1}\sqrt{T}\hat{g}(\hat{\beta}^e).$$

Now Lemma A.1 above yields

$$\begin{aligned} \sqrt{T}\tilde{q}(\hat{\beta}^e) &= \sqrt{T}\hat{q}(\hat{\beta}^e) + \sum\nolimits_{t=1}^{T}\sqrt{T}(\tilde{\pi}_i - 1)q_{tT}(\hat{\beta}^e) \\ &= \hat{q}(\hat{\beta}^e) + \frac{S_T}{T}\sum\nolimits_{t=1}^{T}q_{tT}(\hat{\beta}^e)g_{tT}(\hat{\beta}^e)'\frac{T^{1/2}}{S_T}\hat{\lambda}(1/k_2 + o_p(1)) + o_p(1). \end{aligned}$$

Also by the FOC of the GEL problem with respect to λ

$$\frac{1}{T}\sum\nolimits_{t=1}^{T}\rho_1(k\hat{\lambda}'g_{tT}(\hat{\beta}_{\text{GEL}}))g_{tT}(\hat{\beta}_{\text{GEL}}) = 0.$$

Thus by a Taylor expansion around 0 we have

$$-\frac{1}{T}\sum\nolimits_{t=1}^{T}g_{tT}(\hat{\beta}_{\text{GEL}}) + \frac{1}{T}\sum\nolimits_{t=1}^{T}\rho_2(k\tilde{\lambda}'g_{tT}(\hat{\beta}_{\text{GEL}}))g_{tT}(\hat{\beta}_{\text{GEL}})g_{tT}(\hat{\beta}_{\text{GEL}})'\hat{\lambda}/k_2 = 0.$$

Thus

$$\frac{T^{1/2}}{S_T}\hat{\lambda}/k_2 = [\frac{S_T}{T}\sum_{t=1}^{T}\rho_2(k\tilde{\lambda}'g_{tT}(\hat{\beta}_{\text{GEL}}))g_{tT}(\hat{\beta}_{\text{GEL}})g_{tT}(\hat{\beta}_{\text{GEL}})']^{-1}\sqrt{T}\hat{g}_T(\hat{\beta}_{\text{GEL}}).$$

Now

$$[\frac{S_T}{T}\sum_{t=1}^{T}\rho_2(k\tilde{\lambda}'g_{tT}(\hat{\beta}_{\text{GEL}}))g_{tT}(\hat{\beta}_{\text{GEL}})g_{tT}(\hat{\beta}_{\text{GEL}})']^{-1} = \Xi_{11}^{-1} + o_p(1)$$

by Theorem 2.5 of Smith (2011)-. Also by a Taylor expansion

$$\sqrt{T}\hat{g}_T(\hat{\beta}_{\text{GEL}}) = \sqrt{T}\hat{g}_T(\beta_0) + \breve{G}_T\sqrt{T}(\hat{\beta}_{\text{GEL}} - \beta_0)$$

and $\breve{G}_T \equiv \partial\hat{g}_T(\breve{\beta})/\partial\beta'$ and $\breve{\beta}$ is in a line joining $\hat{\beta}_{\text{GEL}}$ and β_0. now by Lemma A.2 of Smith (2011)- $\sqrt{T}\hat{g}_T(\beta_0) = \sqrt{T}\hat{g}(\beta_0) + O_p(T^{-1/2})$.

Since

$$\sqrt{T}\hat{g}(\hat{\beta}^e) = \sqrt{T}\hat{g}(\beta_0) + \bar{G}\sqrt{T}(\hat{\beta}^e - \beta_0),$$

where $\bar{G} \equiv \partial\hat{g}(\bar{\beta})/\partial\beta'$. It follows that $\sqrt{T}\hat{g}_T(\hat{\beta}_{\text{GEL}}) = \sqrt{T}\hat{g}(\hat{\beta}^e) + o_p(1)$ as $\breve{G}_T$ and $\bar{G}$ converge to G and $\sqrt{T}(\hat{\beta}_{\text{GEL}} - \hat{\beta}^e) = o_p(1)$. Consequently $\sqrt{T}(\tilde{\gamma} - \hat{\gamma}) = o_p(1)$ as $T^{1/2}\hat{g}(\hat{\beta}^e) = O_p(1)$.

Hence

$$\sqrt{T/k_2}\begin{pmatrix} \hat{\beta}^{e\star} - \hat{\beta}^e \\ \hat{\gamma}^{e\star} - \tilde{\gamma} \end{pmatrix} = -[\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\tilde{D}^{\star}]^{-1}\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\sqrt{T/k_2}[\hat{h}_T^{a,\star}(\beta_0, 0) - \tilde{h}_T^a(\beta_0, 0)] + o_B(1).$$

Thus by a Taylor expansion we have

$$\sqrt{T/k_2}\begin{pmatrix} a(\hat{\beta}^{e\star}) - a(\hat{\beta}^e) \\ \hat{\gamma}^{e\star} - \tilde{\gamma} \end{pmatrix} = -R(\ddot{\beta}^{\star})[\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\tilde{D}^{\star}]^{-1}\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\sqrt{T/k_2}[\hat{h}_T^{a,\star}(\beta_0, 0) - \tilde{h}_T^a(\beta_0, 0)] + o_B(1).$$

where $\ddot{\beta}^{\star}$ is in a line joining $\hat{\beta}^{e\star}$ and $\hat{\beta}^e$.

Thus

$$\begin{aligned} \mathcal{W}^{\star} &= (T/k_2)[\hat{r}^{\star} - \tilde{r}]'\left[\hat{R}^{\star}(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^{\star})^{-1}\hat{R}^{\star\prime}\right]^{-1}[\hat{r}^{\star} - \tilde{r}] \\ &= \sqrt{T/k_2}[\hat{h}_T^{a,\star}(\beta_0, 0) - \tilde{h}_T^a(\beta_0, 0)]'\left[\hat{R}^{\star}(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^{\star})^{-1}\hat{R}^{\star\prime}\right]^{-1}R(\ddot{\beta}^{\star})[\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\tilde{D}^{\star}]^{-1} \\ &\quad \times\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\sqrt{T/k_2}[\hat{h}_T^{a,\star}(\beta_0, 0) - \tilde{h}_T^a(\beta_0, 0)]. \end{aligned}$$

Thus as in Theorem 4.4 above $\mathcal{W}^\star$ converges to a chi-squared distribution with $s+r$ degrees of freedom as $\hat{D}^\star = D + o_B(1)$ by the bootstrap UWL Lemma A.7 and $\hat{\Xi}^\star = \Xi + o_B(1)$ and the fact that by the bootstrap CLT we have $\sqrt{T/k_2}[\hat{h}^{a,\star}_T(\beta_0,0) - \tilde{h}^a_T(\beta_0,0)]$ converging to $N(0,\Xi)$.

We consider now the $\mathcal{S}^\star$ statistic. First we derive the distribution of the bootstrapped restricted estimator. We first note that this estimator is consistent by Theorem 4.7 adapted to the moment restrictions $h(z_t,\beta)$ and considering the the compact parameter space $\{\beta \in \mathcal{B} : a(\beta) = 0\}$.

The Lagrangian of the restricted problem is

$$L^\star = \tilde{Q}^\star(\beta,\gamma) - a(\beta)'\lambda - \gamma'\mu.$$

Denote the value of the Lagrange Multiplier at the saddle point as $\hat{\varphi}^\star = (\hat{\lambda}^{\star\prime}, \hat{\mu}^{\star\prime})'$ thus the first order conditions are

$$\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{h}(\hat{\beta}^{e\star}_r,0) - R(\hat{\beta}^{e\star}_r)\hat{\varphi}^\star = 0.$$

Multiplying both sides by $R(\hat{\beta}^{e\star}_r)'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}$ we have

$$R(\hat{\beta}^{e\star}_r)'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{h}^{a,\star}(\hat{\beta}^{e\star}_r,0) - R(\hat{\beta}^{e\star}_r)'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}R(\hat{\beta}^{e\star}_r)\hat{\varphi}^\star = 0.$$

Thus

$$\hat{\varphi}^\star = [R(\hat{\beta}^{e\star}_r)'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}R(\hat{\beta}^{e\star}_r)]^{-1}R(\hat{\beta}^{e\star}_r)'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{h}(\hat{\beta}^{e\star}_r,0).$$

Hence

$$\begin{gathered}\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{h}^{a,\star}(\hat{\beta}^{e\star}_r,0)\\ -R(\hat{\beta}^{e\star}_r)[R(\hat{\beta}^{e\star}_r)'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}R(\hat{\beta}^{e\star}_r)]^{-1}R(\hat{\beta}^{e\star}_r)'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{h}^{a,\star}(\hat{\beta}^{e\star}_r,0) = 0.\end{gathered}$$

But by a Taylor expansion around $\hat{\beta}^e_r$ we have $\hat{h}^{a,\star}(\hat{\beta}^\star_r,0) = \hat{h}^{a,\star}(\hat{\beta}^e_r,0) + \check{D}^\star S_1(\hat{\beta}^{e\star}_r - \hat{\beta}^e_r)$ where $\check{D}^\star = D^\star(\check{\beta}^\star)$, $\check{\beta}^\star$ is in a line joining $\hat{\beta}^{e\star}_r$ and $\hat{\beta}^e_r$ and S_1 is a selection matrix such that

$$\check{D}^\star S_1 \equiv \begin{pmatrix} \sum_{i=1}^T \hat{G}^\star_t(\check{\beta}^\star)/T \\ \sum_{i=1}^T \hat{Q}^\star_t(\check{\beta}^\star)/T \end{pmatrix}.$$

Thus we have

$$\begin{gathered}[I - R(\hat{\beta}^{e\star}_r)[R(\hat{\beta}^{e\star}_r)'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}R(\hat{\beta}^{e\star}_r)]^{-1}R(\hat{\beta}^{e\star}_r)'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}]\\ [\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\sqrt{T}\hat{h}^{a,\star}(\hat{\beta}^e_r,0) + \hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\check{D}^\star S_1\sqrt{T}(\hat{\beta}^{e\star}_r - \hat{\beta}^e_r)] = 0.\end{gathered}$$

Hence

$$\begin{aligned} S_1\sqrt{T}(\hat{\beta}^{e\star}_r - \hat{\beta}^e_r) &= -[\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\check{D}^\star]^{-1}[I - R(\hat{\beta}^{e\star}_r)[R(\hat{\beta}^{e\star}_r)'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}R(\hat{\beta}^{e\star}_r)]^{-1}R(\hat{\beta}^\star_r)'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}] \\ &\quad \times\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\sqrt{T}\hat{h}^{a,\star}(\hat{\beta}^e_r,0).\end{aligned}$$

Now note that by a Taylor expansion around β_0 we have

$$\begin{aligned} \sqrt{T/k_2}(\hat{h}^{a,\star}(\hat{\beta}^e_r,0) - \hat{h}^{a,\star}(\beta_0,0) - \hat{h}^a(\hat{\beta}^e_r,0) + \hat{h}^a(\beta_0,0)) &= (\ddot{D}^\star - \bar{D})S_1\sqrt{T/k_2}(\hat{\beta}^e_r - \beta_0) \\ &= o_B(1), \end{aligned} \tag{A.22}$$

where $\ddot{D}^\star = D^\star(\ddot{\beta})$ and $\ddot{\beta}$ is in a line joining $\hat{\beta}^e_r$ and β_0 and $\bar{D} = D(\bar{\beta})$ and $\bar{\beta}$ is in a line joining $\hat{\beta}^e_r$ and β_0. The second line follows from a UWL an bootstrap UWL and the fact that $\sqrt{T/k_2}(\hat{\beta}^e_r - \beta_0) = O_p(1)$.

We show now that

$$\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\sqrt{T/k_2}\tilde{h}^a_T(\beta_0,0) + \hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\sqrt{T/k_2}\hat{h}^a(\beta_0,0) = o_B(1). \tag{A.23}$$

Note that by Lemma A.2 we have

$$\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\sqrt{T/k_2}\tilde{h}^a_T(\beta_0,0) = \hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\sqrt{T/k_2}\hat{h}^a(\beta_0,0) + \hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\Xi S_1 P\sqrt{T/k_2}\hat{g}(\beta_0) + o_p(1),$$

and $\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\Xi S_1 P = DS_1P + o_B(1) = o_B(1)$ as $\hat{D}^\star = D + o_B(1)$ by Lemma A.7, $\hat{\Xi}^{\star-1} = \Xi^{-1} + o_B(1)$, and the fact that $G'P = 0$ the result follows.

Thus we have

$$\begin{aligned} S_1\sqrt{T/k_2}(\hat{\beta}^{e\star}_r - \hat{\beta}^e_r) &= -[\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\check{D}^\star]^{-1} \\ &\quad \times[I - R(\hat{\beta}^{e\star}_r)[R(\hat{\beta}^{e\star}_r)'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}R(\hat{\beta}^{e\star}_r)]^{-1}R(\hat{\beta}^{e\star}_r)'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}] \\ &\quad \times\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\sqrt{T/k_2}(\hat{h}^{a,\star}(\beta_0,0) - \tilde{h}^a_T(\beta_0,0)) \\ &\quad +A^\star_T + o_B(1). \end{aligned} \tag{A.24}$$

where

$$\begin{aligned} A^\star_T &= -[\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\tilde{D}^\star]^{-1}[I - R(\hat{\beta}^{e\star}_r)[R(\hat{\beta}^{e\star}_r)'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}R(\hat{\beta}^{e\star}_r)]^{-1}R(\hat{\beta}^{e\star}_r)'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}] \\ &\quad \times\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\sqrt{T/k_2}\hat{h}(\hat{\beta}^e_r,0). \end{aligned}$$

But by the FOC of the original restricted problem we have

$$\hat{D}'\hat{\Xi}^{-1}\hat{h}(\hat{\beta}_r^e,0)-R(\hat{\beta}_r^e)[R(\hat{\beta}_r^e)'(\hat{D}'\hat{\Xi}^{-1}\hat{D})^{-1}R(\hat{\beta}_r^e)]^{-1}R(\hat{\beta}_r^e)'(\hat{D}'\hat{\Xi}^{-1}\hat{D})^{-1}\hat{D}'\hat{\Xi}^{-1}\hat{h}(\hat{\beta}_r^e,0)=0.$$

Thus

$$\begin{aligned}A_T^\star &= -[\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\tilde{D}^\star]^{-1}([I-R(\hat{\beta}_r^\star)[R(\hat{\beta}_r^\star)'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}R(\hat{\beta}_r^\star)]^{-1}R(\hat{\beta}_r^\star)'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}]\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\\ &\quad -[I-R(\hat{\beta}_r^e)[R(\hat{\beta}_r^e)'(\hat{D}'\hat{\Xi}^{-1}\hat{D})^{-1}R(\hat{\beta}_r^e)]^{-1}R(\hat{\beta}_r^e)'(\hat{D}'\hat{\Xi}^{-1}\hat{D})^{-1}]\hat{D}'\hat{\Xi}^{-1})\sqrt{T/k_2}\hat{h}(\hat{\beta}_r^e,0)\\ &= o_B(1).\end{aligned}$$

by the bootstrap UWL and $\sqrt{T/k_2}\hat{h}(\hat{\beta}_r^e,0)=O_p(1)$.

Now

$$\mathcal{S}^\star=(\frac{T}{k_2})\left[\hat{h}^\star(\hat{\beta}_r^{e\star})-\tilde{h}_T(\hat{\beta}_r^e)\right]'\hat{\Xi}^{\star-1}\hat{D}^\star(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\left[\hat{h}^\star(\hat{\beta}_r^{e\star})-\tilde{h}_T(\hat{\beta}_r^e)\right].$$

Notice that by two Taylor expansions

$$\begin{aligned}\sqrt{T/k_2}\left[\hat{h}^\star(\hat{\beta}_r^e)-\tilde{h}_T(\hat{\beta}_r^e)\right] &= \sqrt{T/k_2}(\hat{h}^\star(\beta_0)-\tilde{h}_T(\beta_0))+\sqrt{T/k_2}(\check{D}^\star-\bar{D})(\hat{\beta}_r-\beta_0)\\ &= \sqrt{T/k_2}(\hat{h}^\star(\beta_0)-\tilde{h}_T(\beta_0))+o_B(1),\end{aligned}$$

where $\check{D}^\star\equiv\partial\hat{h}^\star(\check{\beta}_r)/\partial\beta'$ where $\check{\beta}_r$ lies in a line between $\hat{\beta}_r$ and β_0 and $\bar{D}\equiv\partial\hat{h}_T(\bar{\beta}_r)/\partial\beta'$ where $\bar{\beta}_r$ lies in a line between $\hat{\beta}_r$ and β_0. The second line is due to a UWL a bootstrap UWL and the fact that $\sqrt{T/k_2}(\hat{\beta}_r-\beta_0)=O_p(1)$.

Thus by a Taylor expansion

$$\begin{aligned}&\hat{D}'\hat{\Xi}^{\star-1}\sqrt{T/k_2}(\hat{h}^\star(\hat{\beta}_r^{e\star})-\tilde{h}_T(\hat{\beta}_r))=\sqrt{T/k_2}\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\ddot{D}^\star(\hat{\beta}_r^{e\star}-\hat{\beta}_r)\\ &\quad+\sqrt{T/k_2}\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}(\hat{h}^\star(\hat{\beta}_r^e)-\tilde{h}_T(\hat{\beta}_r^e))\\ &=-\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\tilde{D}^\star[\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\tilde{D}^\star]^{-1}[I-R(\hat{\beta}_r^\star)[R(\hat{\beta}_r^\star)'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}R(\hat{\beta}_r^\star)]^{-1}R(\hat{\beta}_r^\star)'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}]\\ &\quad\times\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\sqrt{T/k_2}(\hat{h}^{a,\star}(\beta_0,0)-\tilde{h}_T^a(\beta_0,0))\\ &\quad+\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\sqrt{T/k_2}(\hat{h}^{a,\star}(\beta_0,0)-\tilde{h}_T^a(\beta_0,0))\\ &=[R(\hat{\beta}_r^\star)[R(\hat{\beta}_r^\star)'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}R(\hat{\beta}_r^\star)]^{-1}R(\hat{\beta}_r^\star)'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}]\\ &\quad\times\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\sqrt{T/k_2}(\hat{h}^{a,\star}(\beta_0,0)-\tilde{h}_T(\beta_0,0))+o_B(1),\end{aligned}$$

where $\ddot{D}^\star\equiv\partial\hat{h}^\star(\ddot{\beta}_r)/\partial\beta'$ and $\ddot{\beta}_r$ lies in a line between $\hat{\beta}_r^e$ and β_0 and using $(A.23)$.

Thus

$$\begin{aligned}\mathcal{S}^\star &= (\frac{T}{k_2})\left[\hat{h}^\star(\hat{\beta}_r^{e\star})-\tilde{h}_T(\hat{\beta}_r^e)\right]'\hat{\Xi}^{\star-1}\hat{D}^\star(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\left[\hat{h}^\star(\hat{\beta}_r^{e\star})-\tilde{h}_T(\hat{\beta}_r^e)\right]\\ &= \sqrt{T/k_2}(\hat{h}^{a,\star}(\beta_0,0)-\tilde{h}_T^a(\beta_0,0))'\hat{\Xi}^{\star-1}\hat{D}^\star(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}\\ &\quad\times[R(\hat{\beta}_r^{e\star})'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}R(\hat{\beta}_r^{e\star})]^{-1}R(\hat{\beta}_r^{e\star})'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}\\ &\quad\times[R(\hat{\beta}_r^{e\star})[R(\hat{\beta}_r^{e\star})'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}R(\hat{\beta}_r^{e\star})]^{-1}R(\hat{\beta}_r^\star)'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}]\\ &\quad\times\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\sqrt{T/k_2}(\hat{h}^{a,\star}(\beta_0,0)-\tilde{h}_T^a(\beta_0,0))+o_B(1)\\ &= \sqrt{T/k_2}(\hat{h}^{a,\star}(\beta_0,0)-\tilde{h}_T^a(\beta_0,0))'\hat{\Xi}^{\star-1}\hat{D}^\star(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}\\ &\quad\times[R(\hat{\beta}_r^{e\star})'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}R(\hat{\beta}_r^{e\star})]^{-1}R(\hat{\beta}_r^{e\star})'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}]\\ &\quad\times\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\sqrt{T/k_2}(\hat{h}^{a,\star}(\beta_0,0)-\tilde{h}_T^a(\beta_0,0))+o_B(1),\end{aligned}$$

and the result follows.

Now we consider the distance statistic

$$\begin{aligned}\mathcal{D}^\star &= (\frac{T}{k_2})[[\hat{h}^\star(\hat{\beta}_r^{e\star})-\tilde{h}(\hat{\beta}_r)]'\hat{\Xi}^{\star-1}[\hat{h}^\star(\hat{\beta}_r^{e\star})-\tilde{h}(\hat{\beta}_r)]-\hat{g}^\star(\hat{\beta}^{e\star})'\tilde{\Omega}^{\star-1}\hat{g}^\star(\hat{\beta}^{e\star})]\\ &= (\frac{T}{k_2})[[\hat{h}^{a,\star}(\hat{\beta}_r^{e\star},0)-\tilde{h}_T^a(\hat{\beta}_r,0)]'\hat{\Xi}^{\star-1}[\hat{h}^{a,\star}(\hat{\beta}_r^{e\star},0)-\tilde{h}_T^a(\hat{\beta}_r,0)]\\ &\quad-\hat{h}^{a,\star}(\hat{\beta}^{e\star},\hat{\gamma}^\star)'\tilde{\Xi}^{\star-1}\hat{h}^{a,\star}(\hat{\beta}^{e\star},\hat{\gamma}^\star)],\end{aligned}$$

as

$$\hat{g}^\star(\hat{\beta}^{e\star})'\tilde{\Omega}^{\star-1}\hat{g}^\star(\hat{\beta}^{e\star})=\hat{h}^{a,\star}(\hat{\beta}^{e\star},\hat{\gamma}^\star)'\tilde{\Xi}^{\star-1}\hat{h}^{a,\star}(\hat{\beta}^{e\star},\hat{\gamma}^\star).$$

Note now that by two Taylor expansions

$$\begin{aligned}\hat{h}^{a,\star}(\hat{\beta}_r^{e\star},0)-\tilde{h}_T^a(\hat{\beta}_r^e,0) &= \hat{h}^{a,\star}(\hat{\beta}^{e\star},\hat{\gamma}^\star)-\tilde{h}_T^a(\hat{\beta}^e,\hat{\gamma})\\ &\quad+\bar{D}^\star\begin{pmatrix}\hat{\beta}_r^{e\star}-\hat{\beta}^{e\star}\\ -\hat{\gamma}^\star\end{pmatrix}-\ddot{D}\begin{pmatrix}\hat{\beta}_r^e-\hat{\beta}^e\\ -\hat{\gamma}\end{pmatrix},\end{aligned}$$

where $\bar{D}^\star \equiv \partial\hat{h}^{a,\star}(\bar{\beta}^\star,\bar{\gamma}^\star)/\partial\beta'$ and $(\bar{\beta}^{\star\prime},\bar{\gamma}^{\star\prime})'$ is in a line joining $(\hat{\beta}_r^{e\star\prime},\hat{\gamma}^{\star\prime})'$ and $(\hat{\beta}^{e\star\prime},0')'$ and $\ddot{D} \equiv \partial\tilde{h}_T^a(\ddot{\beta},\ddot{\gamma})/\partial\beta'$ and $(\ddot{\beta}',\ddot{\gamma}')'$ is in a line joining $(\hat{\beta}_r^{e\prime},0')'$ and $(\hat{\beta}^{e\prime},\hat{\gamma}')'$. Thus

$$\begin{aligned}
&\frac{T}{k_2}[\hat{h}^{a,\star}(\hat{\beta}_r^{e\star},0)-\tilde{h}^a(\hat{\beta}_r^e,0)]'\hat{\Xi}^{\star-1}[\hat{h}^{a,\star}(\hat{\beta}_r^{e\star},0)-\tilde{h}^a(\hat{\beta}_r^e,0)] \\
=\ &\frac{T}{k_2}\{[\hat{h}^{a,\star}(\hat{\beta}^{e\star},\hat{\gamma}^\star)-\tilde{h}^a(\hat{\beta}^e,\hat{\gamma})+\bar{D}^\star\begin{pmatrix}\hat{\beta}_r^{e\star}-\hat{\beta}^{e\star}\\-\hat{\gamma}^\star\end{pmatrix}-\ddot{D}\begin{pmatrix}\hat{\beta}_r^e-\hat{\beta}^e\\-\hat{\gamma}\end{pmatrix}]'\hat{\Xi}^{\star-1} \\
&\times[\hat{h}^{a,\star}(\hat{\beta}^{e\star},\hat{\gamma}^\star)-\tilde{h}^a(\hat{\beta}^e,\hat{\gamma})+\bar{D}^\star\begin{pmatrix}\hat{\beta}_r^{e\star}-\hat{\beta}^{e\star}\\-\hat{\gamma}^\star\end{pmatrix}-\ddot{D}\begin{pmatrix}\hat{\beta}_r^e-\hat{\beta}^e\\-\hat{\gamma}\end{pmatrix}]\} \\
=\ &\frac{T}{k_2}[\hat{h}^{a,\star}(\hat{\beta}^{e\star},\hat{\gamma}^\star)-\tilde{h}^a(\hat{\beta}^e,\hat{\gamma})]'\hat{\Xi}^{\star-1}[\hat{h}^{a,\star}(\hat{\beta}^{e\star},\hat{\gamma}^\star)-\tilde{h}^a(\hat{\beta}^e,\hat{\gamma})] \\
&+\frac{T}{k_2}2[\bar{D}^\star\begin{pmatrix}\hat{\beta}_r^{e\star}-\hat{\beta}^{e\star}\\-\hat{\gamma}^\star\end{pmatrix}-\ddot{D}\begin{pmatrix}\hat{\beta}_r^e-\hat{\beta}^e\\-\hat{\gamma}\end{pmatrix}]'\hat{\Xi}^{\star-1}[\hat{h}^{a,\star}(\hat{\beta}^{e\star},\hat{\gamma}^\star)-\tilde{h}^a(\hat{\beta}^e,\hat{\gamma})] \\
&+\frac{T}{k_2}[\bar{D}^\star\begin{pmatrix}\hat{\beta}_r^{e\star}-\hat{\beta}^{e\star}\\-\hat{\gamma}^\star\end{pmatrix}-\ddot{D}\begin{pmatrix}\hat{\beta}_r^e-\hat{\beta}^e\\-\hat{\gamma}\end{pmatrix}]'\hat{\Xi}^{\star-1} \\
&\times[\bar{D}^\star\begin{pmatrix}\hat{\beta}_r^{e\star}-\hat{\beta}^{e\star}\\-\hat{\gamma}^\star\end{pmatrix}-\ddot{D}\begin{pmatrix}\hat{\beta}_r^e-\hat{\beta}^e\\-\hat{\gamma}\end{pmatrix}].
\end{aligned}$$

Note that

$$\begin{aligned}
\sqrt{T}\bar{D}^{\star\prime}\hat{\Xi}^{\star-1}\tilde{h}_T^a(\hat{\beta}^e,\hat{\gamma}) &= \bar{D}^{\star\prime}\hat{\Xi}^{\star-1}\sqrt{T}\hat{h}_T^a(\hat{\beta}^e,\hat{\gamma})+\bar{D}^{\star\prime}\hat{\Xi}^{\star-1}\sqrt{T}\sum\nolimits_{t=1}^T(n\hat{\pi}_t-1)h_{tT}(\hat{\beta}^e,\hat{\gamma}) \\
&= \bar{D}^{\star\prime}\hat{\Xi}^{\star-1}\sqrt{T}\hat{h}_T^a(\hat{\beta}^e,\hat{\gamma})+\bar{D}^{\star\prime}\hat{\Xi}^{\star-1}\frac{S_T}{T}\sum\nolimits_{t=1}^T h_{tT}(\hat{\beta}^e,\hat{\gamma})\hat{g}_{tT}'\frac{T^{1/2}}{S_T}\hat{\lambda}+o_p(1) \\
&= \bar{D}^{\star\prime}\hat{\Xi}^{\star-1}\sqrt{T}\hat{h}_T^a(\hat{\beta}^e,\hat{\gamma})-\bar{D}^{\star\prime}\hat{\Xi}^{\star-1}(\Xi S_1+o_p(1))(T^{1/2}P\hat{g}_T(\beta_0)+o_p(1))
\end{aligned}$$

using Lemma A.1 and the fact that $(T^{1/2}/S_T)\hat{\lambda}=-T^{1/2}P\hat{g}_T(\beta_0)+o_p(1)$ by Smith (2011) Proof of Theorem 2.3 (see expression B.2, p A.11). Now as $GP=0$ and $\bar{D}^*=D+o_p(1)$ by Lemma A.7 and $\hat{\Xi}^{\star-1}=\Xi^{-1}+o_p(1)$ and the fact $T^{1/2}\hat{g}_T(\beta_0)=O_p(1)$ that we have

$$\sqrt{T}\bar{D}^{\star\prime}\hat{\Xi}^{\star-1}\tilde{h}^a(\hat{\beta},\hat{\gamma})=\sqrt{T}\bar{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{h}^a(\hat{\beta},\hat{\gamma})+o_p(1), \tag{A.25}$$

Now by three Taylor expansions we have

$$\begin{aligned}
\hat{h}^{a,\star}(\hat{\beta}^{e\star},\hat{\gamma}^\star)-\tilde{h}_T^a(\hat{\beta}^e,\hat{\gamma}) &= \hat{h}^{a,\star}(\hat{\beta}^e,\hat{\gamma})-\tilde{h}_T^a(\hat{\beta}^e,\hat{\gamma})+\check{D}^\star\begin{pmatrix}\hat{\beta}^{e\star}-\hat{\beta}^e\\\hat{\gamma}^\star-\hat{\gamma}\end{pmatrix} \\
&= \hat{h}^{a,\star}(\beta_0,\gamma_0)-\tilde{h}_T^a(\beta_0,\gamma_0)+(\check{D}^\star-\dddot{D})\begin{pmatrix}\hat{\beta}^e-\beta_0\\\hat{\gamma}\end{pmatrix}+\ddot{D}^\star\begin{pmatrix}\hat{\beta}^{e\star}-\hat{\beta}^e\\\hat{\gamma}^\star-\hat{\gamma}\end{pmatrix} \\
&= \hat{h}^{a,\star}(\beta_0,\gamma_0)-\tilde{h}_T^a(\beta_0,\gamma_0)+\tilde{h}^a(\beta_0,\gamma_0)-\hat{h}^a(\beta_0,\gamma_0) \\
&\quad+(\check{D}^*-\dddot{D})\begin{pmatrix}\hat{\beta}^e-\beta_0\\\hat{\gamma}\end{pmatrix}+\ddot{D}^\star\begin{pmatrix}\hat{\beta}^{e\star}-\hat{\beta}^e\\\hat{\gamma}^\star-\hat{\gamma}\end{pmatrix} \\
&= O_B(1/\sqrt{T}),
\end{aligned}$$

where $\check{D}^\star\equiv\partial\hat{h}^{a,\star}(\check{\beta}^\star,\check{\gamma}^\star)/\partial\beta'$ and $(\check{\beta}^{\star\prime},\check{\gamma}^{\star\prime})'$ is in a line joining $(\hat{\beta}^{e\star\prime},\hat{\gamma}^{\star\prime})'$ and $(\hat{\beta}^{e\prime},\hat{\gamma}')'$ and $\dddot{D}\equiv\partial\tilde{h}_T^a(\dddot{\beta},\dddot{\gamma})/\partial\beta'$ and $(\dddot{\beta}',\dddot{\gamma}')'$ is in a line joining $(\hat{\beta}^{e\prime},\hat{\gamma}')'$ and $(\beta_0',0')'$ and $\ddot{D}^\star=\partial\hat{h}^{a,\star}(\ddot{\beta}^\star,\ddot{\gamma}^\star)/\partial\beta'$ and $(\ddot{\beta}',\ddot{\gamma}')'$ is in a line joining $(\hat{\beta}^{e\prime},\hat{\gamma}')'$ and $(\beta_0',0')'$. The result follows from the bootstrap CLT, the standard CLT, Lemma A.2 and asymptotic normality of $((\hat{\beta}^{e\star}-\hat{\beta}^e)',(\hat{\gamma}^\star-\hat{\gamma})')'$ and $((\hat{\beta}^e-\beta_0)',\hat{\gamma}')'$.

Thus by A.25

$$\begin{aligned}
\bar{D}^{\star\prime}\hat{\Xi}^{\star-1}\sqrt{T}[\hat{h}^{a,\star}(\hat{\beta}^{e\star},\hat{\gamma}^\star)-\tilde{h}_T^a(\hat{\beta}^e,\hat{\gamma})] &= \bar{D}^{\star\prime}\hat{\Xi}^{\star-1}\sqrt{T}[\hat{h}^{a,\star}(\hat{\beta}^{e\star},\hat{\gamma}^\star)-\hat{h}_T^a(\hat{\beta}^e,\hat{\gamma})]+o_p(1) \\
&= [\bar{D}^{\star\prime}\hat{\Xi}^{\star-1}-\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}]\sqrt{T}[\hat{h}^{a,\star}(\hat{\beta}^{e\star},\hat{\gamma}^\star)-\hat{h}_T^a(\hat{\beta}^e,\hat{\gamma})]+o_p(1),
\end{aligned}$$

as $\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{h}^{a,\star}(\hat{\beta}^{e\star},\hat{\gamma}^\star)=0$ and $\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\sqrt{T}\hat{h}^a(\hat{\beta}^e,\hat{\gamma})=[\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}-\hat{D}'\hat{\Xi}^{-1}]\sqrt{T}\hat{h}^a(\hat{\beta}^e,\hat{\gamma})=o_p(1)$ since $\hat{D}'\hat{\Xi}^{-1}\hat{h}^a(\hat{\beta}^e,\hat{\gamma})=0$ and consequently

$$\begin{aligned}
\hat{D}'\hat{\Xi}^{-1}\sqrt{T}\hat{h}_T^a(\hat{\beta}^e,\hat{\gamma}) &= \hat{D}'\hat{\Xi}^{-1}\sqrt{T}(\hat{h}_T^a(\hat{\beta}^e,\hat{\gamma})-\hat{h}^a(\hat{\beta}^e,\hat{\gamma}))+ \\
&= \hat{D}'\hat{\Xi}^{-1}\sqrt{T}(\hat{h}_T^a(\beta_0,0)-\hat{h}^a(\beta_0,0))+\hat{D}'\hat{\Xi}^{-1}(\ddot{D}_T-\ddot{D})\sqrt{T}\begin{pmatrix}\hat{\beta}^e-\beta_0\\\hat{\gamma}\end{pmatrix} \\
&= o_B(1)
\end{aligned}$$

with $\ddot{D}_T=\partial\tilde{h}_T^a(\ddot{\beta},\ddot{\gamma})/\partial\beta'$ and $\ddot{D}=\partial\tilde{h}^a(\ddot{\beta},\ddot{\gamma})/\partial\beta'$ and $(\ddot{\beta}',\ddot{\gamma}')'$ is in a line joining $(\hat{\beta}^{e\prime},\hat{\gamma}')'$ and $(\beta_0',0')'$. The result follows from equation (A11) page [A.3] and Lemma A.1 of Smith (2011).

Since $\sqrt{T}[\hat{h}^{a,\star}(\hat{\beta}^{e\star},\hat{\gamma}^\star)-\hat{h}^a(\hat{\beta}^e,\hat{\gamma})]=O_p(1)$ and $\bar{D}^\star=D+o_B(1)$, $\hat{D}^\star=D+o_B(1)$ by Lemma A.7 and $\hat{\Xi}^{\star-1}=\Xi^{-1}+o_p(1)$ and the fact that $\sqrt{T}[\hat{h}^{a,\star}(\hat{\beta}^{e\star},\hat{\gamma}^\star)-\hat{h}^a(\hat{\beta}^e,\hat{\gamma})]=O_p(1)$ it follows that $\bar{D}^{\star\prime}\hat{\Xi}^{\star-1}\sqrt{T}[\hat{h}^{a,\star}(\hat{\beta}^{e\star},\hat{\gamma}^\star)-\tilde{h}^a(\hat{\beta}^e,\hat{\gamma})]=o_B(1)$. Similarly $\ddot{D}\hat{\Xi}^{\star-1}\sqrt{T}[\hat{h}^{a,\star}(\hat{\beta}^{e\star},\hat{\gamma}^\star)-\tilde{h}^a(\hat{\beta}^e,\hat{\gamma})]=o_B(1)$.

Note also that $\hat{\beta}_r^{e\star}-\hat{\beta}^{e\star}=(\hat{\beta}_r^{e\star}-\hat{\beta}_r^e)-(\hat{\beta}^{e\star}-\hat{\beta}^e)+(\hat{\beta}_r^e-\beta_0)-(\hat{\beta}^e-\beta_0)=O_B(1/\sqrt{T})$, $\hat{\beta}_r^e-\hat{\beta}^e=(\hat{\beta}_r^e-\beta_0)-(\hat{\beta}^e-\beta_0)=O_p(1/\sqrt{T})$ and $\hat{\gamma}^\star-\hat{\gamma}+\hat{\gamma}=O_B(1/\sqrt{T})$ and $\hat{\gamma}=O_p(1/\sqrt{T})$.

Hence

$$\frac{T}{k_2}2[\bar{D}^\star\begin{pmatrix}\hat{\beta}_r^{e\star}-\hat{\beta}^{e\star}\\-\hat{\gamma}^\star\end{pmatrix}-\ddot{D}\begin{pmatrix}\hat{\beta}_r^e-\hat{\beta}^e\\-\hat{\gamma}\end{pmatrix}]'\hat{\Xi}^{\star-1}[\hat{h}^{a,\star}(\hat{\beta}^{e\star},\hat{\gamma}^\star)-\tilde{h}_T^a(\hat{\beta}^e,\hat{\gamma})]=o_B(1).$$

Now note that

$$\sqrt{\frac{T}{k_2}}[\bar{D}^\star\begin{pmatrix}\hat{\beta}_r^{e\star}-\hat{\beta}^{e\star}\\-\hat{\gamma}^\star\end{pmatrix}-\ddot{D}\begin{pmatrix}\hat{\beta}_r^e-\hat{\beta}^e\\-\hat{\gamma}\end{pmatrix}]=-\sqrt{\frac{T}{k_2}}D(\begin{pmatrix}\hat{\beta}^{e\star}-\hat{\beta}^e\\\hat{\gamma}^\star-\hat{\gamma}\end{pmatrix}-\begin{pmatrix}\hat{\beta}_r^{e\star}-\hat{\beta}_r^e\\0\end{pmatrix})+o_B(1)$$

as $\bar{D}^\star=D+o_B(1)$, $\ddot{D}=D+o_p(1)$ and $\hat{\beta}_r-\hat{\beta}=O_p(1/\sqrt{T})$ and $\hat{\gamma}^*=O_B(1/\sqrt{T})$ and $\hat{\gamma}=O_p(1/\sqrt{T})$.

Thus

$$-\sqrt{\frac{T}{k_2}}D(\begin{pmatrix}\hat{\beta}^{e\star}-\hat{\beta}^e\\\hat{\gamma}^\star-\hat{\gamma}\end{pmatrix}-\begin{pmatrix}\hat{\beta}_r^{e\star}-\hat{\beta}_r^e\\0\end{pmatrix})=\sqrt{\frac{T}{k_2}}D[\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\tilde{D}^\star]^{-1}[R(\hat{\beta}_r^\star)$$
$$\times[R(\hat{\beta}_r^\star)'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}R(\hat{\beta}_r^\star)]^{-1}R(\hat{\beta}_r^\star)'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}]\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\sqrt{T/k_2}(\hat{h}^{a,\star}(\beta_0,0)-\tilde{h}_T^a(\beta_0,0)),$$

using the asymptotic representations of $\sqrt{T/k_2}((\hat{\beta}^{e\star}-\hat{\beta}^e)',(\hat{\gamma}^\star-\hat{\gamma})')'$ given in Theorem 4.8 and of $S_1\sqrt{T/k_2}(\hat{\beta}_r^{e\star}-\hat{\beta}_r^e)$ given in (A.24).

Thus

$$\begin{aligned}
&\frac{T}{k_2}[\bar{D}^\star\begin{pmatrix}\hat{\beta}_r^{e\star}-\hat{\beta}^{e\star}\\-\hat{\gamma}^\star\end{pmatrix}-\ddot{D}\begin{pmatrix}\hat{\beta}_r^e-\hat{\beta}^e\\-\hat{\gamma}\end{pmatrix}]'\hat{\Xi}^{\star-1}[\bar{D}^\star\begin{pmatrix}\hat{\beta}_r^{e\star}-\hat{\beta}^{e\star}\\-\hat{\gamma}^\star\end{pmatrix}-\ddot{D}\begin{pmatrix}\hat{\beta}_r^e-\hat{\beta}^e\\-\hat{\gamma}\end{pmatrix}]\\
=\;&[\sqrt{T/k_2}(\hat{h}^{a,\star}(\beta_0,0)-\tilde{h}_T^a(\beta_0,0))'\hat{\Xi}^{\star-1}\hat{D}^\star(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}R(\hat{\beta}_r^{e\star})[R(\hat{\beta}_r^{e\star})'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}R(\hat{\beta}_r^{e\star})]^{-1}\\
&\times R(\hat{\beta}_r^{e\star})'[\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\tilde{D}^\star]^{-1}D'\hat{\Xi}^{\star-1}D[\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\tilde{D}^\star]^{-1}[R(\hat{\beta}_r^{e\star})[R(\hat{\beta}_r^{e\star})'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}R(\hat{\beta}_r^{e\star})]^{-1}\\
&\times R(\hat{\beta}_r^{e\star})'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}]\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\sqrt{T/k_2}(\hat{h}^{a,\star}(\beta_0,0)-\tilde{h}_T^a(\beta_0,0)))+o_B(1)\\
=\;&[\sqrt{T/k_2}(\hat{h}^{a,\star}(\beta_0,0)-\tilde{h}_T^a(\beta_0,0))'\hat{\Xi}^{\star-1}\hat{D}^\star(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}R(\hat{\beta}_r^{e\star})[R(\hat{\beta}_r^{e\star})'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}R(\hat{\beta}_r^{e\star})]^{-1}\\
&\times R(\hat{\beta}_r^{e\star})'(\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\hat{D}^\star)^{-1}]\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\sqrt{T/k_2}(\hat{h}^{a,\star}(\beta_0,0)-\tilde{h}_T^a(\beta_0,0))+o_B(1),
\end{aligned}$$

and the result follows as in the proof of Theorem 4.4. ■

Proof of Theorem 4.11: We start by deriving the asymptotic distribution of $\mathcal{W}^\dagger$. Define $h_t^{a,\dagger}(\beta,\gamma)\equiv(g^\dagger(z_t,\beta)',[q^\dagger(z_t,\beta)-\gamma]')$, $\hat{h}^{a,\dagger}(\beta,\gamma)=\sum_{t=1}^{m_T}h_t^{a,\dagger}(\beta,\gamma)/m_T$ and $\tilde{Q}^\dagger(\beta)=\hat{h}^{a,\dagger}(\beta,\gamma)'\hat{\Xi}^{\dagger-1}\hat{h}^{a,\dagger}(\beta,\gamma)$. Note that the unrestricted GMM estimator solves

$$(\hat{\beta}^{e\dagger\prime},\hat{\gamma}^{\dagger\prime})=\underset{\beta\in B,\gamma\in\mathbb{R}^m}{\arg\min}\,\tilde{Q}^\dagger(\beta,\gamma).$$

The solution is given by

$$\begin{aligned}
\hat{\beta}^{e\dagger}&=\underset{\beta\in B}{\arg\min}\,\hat{g}^\dagger(\beta)'\hat{\Omega}^{\dagger-1}\hat{g}^\dagger(\beta),\\
\hat{\gamma}^\dagger&=\hat{q}^\dagger(\hat{\beta}^{e\dagger})-\hat{\Xi}_{21}^\dagger\hat{\Omega}^{\dagger-1}g^\dagger(\hat{\beta}^{e\dagger}).
\end{aligned}$$

Consistency of $\hat{\beta}^{e\dagger}$ follows from Theorem 4.7 hence $\hat{\beta}^{e\dagger}=\hat{\beta}^e+o_B(1)$ and since $\hat{\beta}^e=\hat{\beta}_r^e+o_p(1)$ as $\hat{\beta}$ and $\hat{\beta}_r$ are both consistent we have $\hat{\beta}^{e\dagger}=\hat{\beta}_r^e+o_B(1)$. We note that by Lemma A.6 and $\hat{\Xi}^\dagger=\hat{\Xi}+o_B(1)$ we have $\hat{\gamma}^{e\dagger}=\hat{\gamma}+o_B(1)=o_B(1)$.

Since these estimators satisfy the first order conditions we have $\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{h}^\dagger(\hat{\beta}^{e\dagger},\hat{\gamma}^\dagger)=0$. Thus by a Taylor expansion around $(\hat{\beta}_r^{e\prime},0')'$ we have

$$\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{h}^\dagger(\hat{\beta}_r,0)+\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\breve{D}^\dagger\begin{pmatrix}\hat{\beta}^{e\dagger}-\hat{\beta}_r^e\\\hat{\gamma}^\dagger\end{pmatrix}=0,$$

where $\breve{D}^\dagger=D^\dagger(\breve{\beta}^\dagger)$,

$$D^\dagger(\beta)=\begin{pmatrix}\sum_{i=1}^{m_T}\hat{G}_t^\dagger(\beta)/m_T & 0\\\sum_{i=1}^{m_T}\hat{Q}_t^\dagger(\beta)/m_T & -I_s\end{pmatrix},$$

and $\breve{\beta}^\dagger$ is in a line joining $\hat{\beta}^{e\dagger}$ and $\hat{\beta}_r^e$. Thus

$$\sqrt{T}\begin{pmatrix}\hat{\beta}^{e\dagger}-\hat{\beta}_r^e\\\hat{\gamma}^\dagger\end{pmatrix}=-[\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\breve{D}^\dagger]^{-1}\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\sqrt{T}\hat{h}^{a,\dagger}(\hat{\beta}_r^e,0).$$

Now notice that expanding $\sqrt{T}\hat{h}^{a,\dagger}(\hat{\beta}_r^e,0)$ around β_0 yields

$$\begin{aligned}
\sqrt{T}\hat{h}^{a,\dagger}(\hat{\beta}_r^e,0)&=\sqrt{T}\hat{h}^{a,\dagger}(\beta_0,0)+\check{D}^\dagger S_1(\hat{\beta}_r^e-\beta_0)\\
&=\sqrt{T}\hat{h}^{a,\dagger}(\beta_0,0)-\sqrt{T}\tilde{h}_T^a(\beta_0,0)\\
&\quad+\sqrt{T}\tilde{h}_T^a(\beta_0,0)+\tilde{D}^\dagger S_1(\hat{\beta}_r^e-\beta_0),
\end{aligned}$$

where $\tilde{D}^{\dagger} = D^{\dagger}(\tilde{\beta}^{\dagger})$ and $\tilde{\beta}^{\dagger}$ is in a line joining $\hat{\beta}_r^e$ and β_0.

By the asymptotic representation of $\hat{\beta}_r^e$ we have

$$\check{D}^{\dagger} S_1 \sqrt{T}(\hat{\beta}_r^e - \beta_0) = -DS_1[\Lambda - \Lambda R' \left[R\Lambda R'\right]^{-1} R\Lambda]D'\Xi^{-1}\sqrt{T}\hat{h}_T^a(\beta_0, 0) + o_p(1).$$

as $\check{D}^{\dagger} = D + o_B(1)$ by Lemma A.7.

Also by Lemma A.4 we have

$$\sqrt{T}\tilde{h}_T^a(\beta_0, 0) = DS_1[\Lambda - \Lambda R' \left[R\Lambda R'\right]^{-1} R\Lambda]D'\Xi^{-1}]\sqrt{T}\hat{h}_T^a(\beta_0, 0) + o_p(1), \tag{A.26}$$

Consequently

$$\sqrt{T}\tilde{h}^a(\beta_0, 0) + \tilde{D}^{\dagger} S_1 \sqrt{T}(\hat{\beta}_r - \beta_0) = o_p(1). \tag{A.27}$$

It follows that

$$\sqrt{T/k_2}\begin{pmatrix} \hat{\beta}^{e\dagger} - \hat{\beta}_r^e \\ \hat{\gamma}^{\dagger} \end{pmatrix} = -[\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\check{D}^{\dagger}]^{-1}\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}[\sqrt{\frac{T}{k_2}}\hat{h}^{a,\dagger}(\beta_0, 0) - \sqrt{T}\tilde{h}_T^a(\beta_0, 0)].$$

Thus by a Taylor expansion we have

$$\begin{aligned}
\sqrt{T/k_2}\begin{pmatrix} a(\hat{\beta}^{e\dagger}) \\ \hat{\gamma}^{\dagger} \end{pmatrix} &= \sqrt{T/k_2}\begin{pmatrix} A(\ddot{\beta}^{\dagger})(\hat{\beta}^{e\dagger} - \hat{\beta}_r^e) \\ \hat{\gamma}^{\dagger} \end{pmatrix} = \begin{pmatrix} A(\ddot{\beta}^{\dagger}) & 0 \\ 0 & I \end{pmatrix}\sqrt{T}\begin{pmatrix} \hat{\beta}^{e\dagger} - \hat{\beta}_r^e \\ \hat{\gamma}^{\dagger} \end{pmatrix} \\
&= -R(\ddot{\beta}^{\dagger})[\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\tilde{D}^{\dagger}]^{-1}\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\sqrt{T/k_2}[\hat{h}_T^{a,\star}(\beta_0, 0) - \tilde{h}_T^a(\beta_0, 0)] + o_B(1),
\end{aligned}$$

where $\ddot{\beta}^{\dagger}$ is in a line joining $\hat{\beta}^{e\dagger}$ and $\hat{\beta}_r^e$.

Thus

$$\begin{aligned}
\mathcal{W}^{\dagger} &= (T/k_2)\hat{r}^{\dagger\prime}[\hat{R}^{\dagger}(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^{\dagger})^{-1}\hat{R}^{\dagger\prime}]^{-1}\hat{r}^{\dagger} \\
&= \sqrt{T/k_2}[\hat{h}_T^{a,\dagger}(\beta_0, 0) - \tilde{h}_T^a(\beta_0, \gamma)]'\hat{\Xi}^{\dagger-1}\hat{D}^{\dagger}[\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\tilde{D}^{\dagger}]^{-1}R(\ddot{\beta}^{\dagger})'[\hat{R}^{\dagger}(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^{\dagger})^{-1}\hat{R}^{\dagger\prime}]^{-1} \\
&\quad \times R(\ddot{\beta}^{\dagger})[\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\tilde{D}^{\dagger}]^{-1}\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\sqrt{T/k_2}[\hat{h}_T^{a,\star}(\beta_0, 0) - \tilde{h}_T^a(\beta_0, 0)].
\end{aligned}$$

Thus as in the proof of Theorem 2.4 above $\mathcal{W}^{\dagger}$ converges to a chi-squared distribution with $s + r$ degrees of freedom as $\hat{D}^{\dagger} = D + o_B(1)$ by the Lemma A.7 and $\hat{\Xi}^{\dagger} = \Xi + o_B(1)$ and the fact that by the bootstrap CLT we have $\sqrt{T/k_2}[\hat{h}_T^{a,\dagger}(\beta_0, \gamma) - \tilde{h}_T^a(\beta_0, \gamma)]$ converging to $N(0, \Xi)$.

We consider know the $\mathcal{S}^{\dagger}$ statistic. First we derive the distribution of the bootstrap restricted estimator. We note that $\hat{\beta}_r^{\dagger}$ is consistent by Theorem 4.7 applied to the moment indicators $h(z_t, \beta)$ and restricted parameter space $\mathcal{B}_r$. Note that the Lagrangian of the restricted problem is

$$L^{\dagger} = \tilde{Q}^{\dagger}(\beta, \gamma) - a(\beta)'\lambda^{\dagger} - \gamma'\mu.$$

Denote $\hat{\varphi}^{\dagger} = (\hat{\lambda}^{\dagger\prime}, \hat{\mu}^{\dagger\prime})'$ $\hat{\lambda}$ and $\hat{\mu}$ are the Lagrange multipliers evaluated at the optimum. Thus the first order conditions yield

$$\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{h}(\hat{\beta}_r^{e\dagger}, 0) - R(\hat{\beta}_r^{e\dagger})\hat{\varphi}^{\dagger} = 0.$$

Multiplying both sides by $R(\hat{\beta}_r^{e\dagger})'(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^{\dagger})^{-1}$ we have

$$R(\hat{\beta}_r^{e\dagger})'(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^{\dagger})^{-1}\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{h}^{a,\dagger}(\hat{\beta}_r^e, 0) - R(\hat{\beta}_r^{e\dagger})'(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^{\dagger})^{-1}R(\hat{\beta}_r^{e\dagger})\hat{\varphi}^{\dagger} = 0.$$

Thus

$$\hat{\varphi}^{\dagger} = [R(\hat{\beta}_r^{e\dagger})'(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^{\dagger})^{-1}R(\hat{\beta}_r^{e\dagger})]^{-1}R(\hat{\beta}_r^{e\dagger})'(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^{\dagger})^{-1}\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{h}(\hat{\beta}_r^{e\dagger}, 0).$$

Hence

$$\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{h}^{a,\dagger}(\hat{\beta}_r^{e\dagger}, 0) - R(\hat{\beta}_r^{e\dagger})[R(\hat{\beta}_r^{e\dagger})'(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^{\dagger})^{-1}R(\hat{\beta}_r^{e\dagger})]^{-1}R(\hat{\beta}_r^{e\dagger})'(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^{\dagger})^{-1}\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{h}^{a,\dagger}(\hat{\beta}_r^{e\dagger}, 0) = 0.$$

But by a Taylor expansion $\hat{h}^{a,\dagger}(\hat{\beta}_r^{e\dagger}, 0) = \hat{h}^{a,\dagger}(\hat{\beta}_r^e, 0) + \tilde{D}^{\dagger} S_1(\hat{\beta}_r^{e\dagger} - \hat{\beta}_r^e)$ where where S_1 is a selection matrix such that

$$\tilde{D}^{\dagger} S_1 = \begin{pmatrix} \sum_{i=1}^T \hat{G}_t^{\dagger}\left(\tilde{\beta}^{\dagger}\right)/T \\ \sum_{i=1}^T \hat{Q}_t^{\dagger}\left(\tilde{\beta}^{\dagger}\right)/T \end{pmatrix}$$

$\tilde{D}^{\dagger} = D^{\dagger}(\tilde{\beta}^{\dagger})$ where $\tilde{\beta}^{\dagger}$ is in a line joining $\hat{\beta}_r^{e\dagger}$ and $\hat{\beta}_r^e$, thus we have

$$\begin{gathered}
[I - R(\hat{\beta}_r^{e\dagger})[R(\hat{\beta}_r^{e\dagger})'(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^{\dagger})^{-1}R(\hat{\beta}_r^{e\dagger})]^{-1}R(\hat{\beta}_r^{e\dagger})'(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^{\dagger})^{-1}] \\
\times\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\sqrt{T}\hat{h}^{a,\dagger}(\hat{\beta}_r^e, 0) + \hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\tilde{D}^{\dagger} S_1\sqrt{T}(\hat{\beta}_r^{e\dagger} - \hat{\beta}_r^e) = 0.
\end{gathered}$$

Hence

$$\begin{aligned}
S_1\sqrt{T}(\hat{\beta}_r^{e\dagger} - \hat{\beta}_r^e) &= -[\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\tilde{D}^{\dagger}]^{-1}[I - R(\hat{\beta}_r^{e\dagger})[R(\hat{\beta}_r^{e\dagger})'(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^{\dagger})^{-1}R(\hat{\beta}_r^{e\dagger})]^{-1}R(\hat{\beta}_r^{e\dagger})'(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^{\dagger})^{-1}] \\
&\quad \times\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\sqrt{T}\hat{h}^{a,\dagger}(\hat{\beta}_r^e, 0).
\end{aligned}$$

Now note that by a Taylor expansion around β_0 we have

$$\begin{aligned}\sqrt{T/k_2}(\hat{h}^{a,\dagger}(\hat{\beta}_r^e,0)-\hat{h}^{a,\dagger}(\beta_0,0)-\tilde{h}_T^a(\hat{\beta}_r^e,0)+\tilde{h}_T^a(\beta_0,0)) &= (\ddot{D}^\dagger-\bar{D})S_1\sqrt{T/k_2}(\hat{\beta}_r^e-\beta_0) \\ &= o_B(1),\end{aligned} \quad (A.28)$$

where $\ddot{D}^\dagger = D^\dagger(\ddot{\beta})$ and $\ddot{\beta}$ lies in a line joining $\hat{\beta}_r^e$ and β_0 and $\bar{D} = D(\bar{\beta}_r)$ where

$$\tilde{D}(\beta) = \left(\begin{array}{cc} \sum_{i=1}^T G_t(\beta)\pi_{t,r} & 0 \\ \sum_{i=1}^T Q_t(\beta)\pi_{t,r} & -I_s \end{array}\right),$$

and $\bar{\beta}_r$ is in a line joining $\hat{\beta}_r^e$ and β_0.

Note that by a Taylor expansion

$$\sqrt{T/k_2}\tilde{h}_T^a(\hat{\beta}_r^e,0) = \sqrt{T/k_2}\tilde{h}_T^a(\beta_0,0) + \dddot{D}S_1(\hat{\beta}_r^e-\beta_0), \quad (A.29)$$

where $\dddot{D} = D(\dddot{\beta})$ and $\dddot{\beta}$ is on a line joining $\hat{\beta}_r^e$ and β_0. Note that similarly to $(A.27)$ the rhs of $(A.29)$ is $o_p(1)$. Thus we have

$$\begin{aligned}S_1\sqrt{T}(\hat{\beta}_r^{e\dagger}-\hat{\beta}_r^e) &= -[\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\tilde{D}^\dagger]^{-1}[I-R(\hat{\beta}_r^{e\dagger})[R(\hat{\beta}_r^{e\dagger})'(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^\dagger)^{-1}R(\hat{\beta}_r^{e\dagger})]^{-1}R(\hat{\beta}_r^{e\dagger})'(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^\dagger)^{-1}] \\ &\quad\times\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\sqrt{T}[\hat{h}^{a,\dagger}(\beta_0,0)-\sqrt{T}\tilde{h}_T^a(\beta_0,0)]+o_B(1).\end{aligned}$$

Now let us consider the score statistic:

$$\mathcal{S}^\dagger = (\frac{T}{k_2})\hat{h}^\dagger(\hat{\beta}_r^{e\dagger})'\hat{\Xi}^{\dagger-1}\hat{D}^\dagger(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^\dagger)^{-1}\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{h}^\dagger(\hat{\beta}_r^{e\dagger}).$$

We proved above that

$$\sqrt{T/k_2}(\hat{h}^{a,\dagger}(\hat{\beta}_r^e,0)-\hat{h}^{a,\dagger}(\beta_0,0)+\tilde{h}_T^a(\beta_0,0)) = o_B(1).$$

Notice also by a Taylor expansion

$$\sqrt{T/k_2}\hat{h}^{a\dagger}(\hat{\beta}_r^{e\dagger},0) = \sqrt{T/k_2}\hat{h}^{a,\dagger}(\hat{\beta}_r^e,0)+\sqrt{T/k_2}\check{D}^\dagger S_1(\hat{\beta}_r^{e\dagger}-\hat{\beta}_r^e),$$

where $\check{D}^\dagger = D^\dagger(\check{\beta})$ and $\check{\beta}$ is in a line joining $\hat{\beta}_r^e$ and β_0.Thus

$$\begin{aligned}\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\sqrt{T/k_2}\hat{h}^\dagger(\hat{\beta}_r^{e\dagger}) &= \hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}[\hat{h}^{a,\dagger}(\beta_0,0)-\tilde{h}(\beta_0,0)] \\ &\quad -\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}(\check{D}^\dagger)[\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\tilde{D}^\dagger]^{-1}\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\sqrt{T}[\hat{h}^{a,\dagger}(\beta_0,0)-\sqrt{T}\tilde{h}_T^a(\beta_0,0)] \\ &\quad +[\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\tilde{D}^\dagger]^{-1}R(\hat{\beta}_r^{e\dagger})[R(\hat{\beta}_r^{e\dagger})'(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^\dagger)^{-1}R(\hat{\beta}_r^{e\dagger})]^{-1} \\ &\quad \times R(\hat{\beta}_r^{e\dagger})'(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^\dagger)^{-1}\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\sqrt{T}[\hat{h}^{a,\dagger}(\beta_0,0)-\sqrt{T}\tilde{h}_T^a(\beta_0,0)]+o_p(1) \\ &= [\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\tilde{D}^\dagger]^{-1}R(\hat{\beta}_r^{e\dagger})[R(\hat{\beta}_r^{e\dagger})'(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^\dagger)^{-1}R(\hat{\beta}_r^{e\dagger})]^{-1} \\ &\quad \times R(\hat{\beta}_r^{e\dagger})'(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^\dagger)^{-1}\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\sqrt{T}[\hat{h}^{a,\dagger}(\beta_0,0)-\sqrt{T}\tilde{h}_T^a(\beta_0,0)]+o_p(1).\end{aligned}$$

Hence

$$\begin{aligned}\mathcal{S}^* &= (\frac{T}{k_2})\hat{h}^\dagger(\hat{\beta}_r^{e\dagger})'\hat{\Xi}^{\dagger-1}\hat{D}^\dagger(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^\dagger)^{-1}\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{h}^\dagger(\hat{\beta}_r^{e\dagger}) \\ &= \sqrt{T/k_2}(\hat{h}^{a,\dagger}(\beta_0,0)-\tilde{h}_T^a(\beta_0,0))'\hat{\Xi}^{\dagger-1}\hat{D}^\dagger(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^\dagger)^{-1} \\ &\quad\times[R(\hat{\beta}_r^{e\dagger})'(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^\dagger)^{-1}R(\hat{\beta}_r^{e\dagger})]^{-1}R(\hat{\beta}_r^{e\dagger})'(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^\dagger)^{-1}[R(\hat{\beta}_r^{e\dagger})[R(\hat{\beta}_r^{e\dagger})'(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^\dagger)^{-1}R(\hat{\beta}_r^{e\dagger})]^{-1} \\ &\quad\times R(\hat{\beta}_r^{e\dagger})'(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^\dagger)^{-1}]\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\sqrt{T/k_2}(\hat{h}^{a,\star}(\beta_0,0)-\tilde{h}_T^a(\beta_0,0))+o_B(1) \\ &= \sqrt{T/k_2}(\hat{h}^{a,\dagger}(\beta_0,0)-\tilde{h}_T^a(\beta_0,0))'\hat{\Xi}^{\dagger-1}\hat{D}^\dagger(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^\dagger)^{-1}[R(\hat{\beta}_r^{e\dagger})'(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^\dagger)^{-1}R(\hat{\beta}_r^{e\dagger})]^{-1} \\ &\quad\times R(\hat{\beta}_r^{e\dagger})'(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^\dagger)^{-1}]\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\sqrt{T/k_2}(\hat{h}^{a,\dagger}(\beta_0,0)-\tilde{h}_T^a(\beta_0,0))+o_B(1)\end{aligned}$$

as the result follows as in the proof of Theorem 2.4.

Now we consider the $\mathcal{D}^\dagger$ statistic:

$$\begin{aligned}\mathcal{D}^\dagger &= (\frac{T}{k_2})[\hat{h}^\dagger(\hat{\beta}_r^{e\dagger})'\hat{\Xi}^{\dagger-1}\hat{h}^\dagger(\hat{\beta}_r^{e\dagger})-\hat{g}^\dagger(\hat{\beta}^{e\dagger})'\tilde{\Omega}^{\dagger-1}\hat{g}^\dagger(\hat{\beta}^{e\dagger}) \\ &= (\frac{T}{k_2})[\hat{h}^{a,\dagger}(\hat{\beta}_r^{e\dagger},0)'\hat{\Xi}^{\dagger-1}\hat{h}^{a,\dagger}(\hat{\beta}_r^{e\dagger},0)-\hat{h}^{a,\dagger}(\hat{\beta}^{e\dagger},\hat{\gamma}^\dagger)'\hat{\Xi}^{\dagger-1}\hat{h}^{a,\dagger}(\hat{\beta}^{e\dagger},\hat{\gamma}^\dagger)].\end{aligned}$$

Expanding $\hat{h}^{a,\dagger}(\hat{\beta}_r^{e\dagger},0)$ around $(\hat{\beta}^{e\dagger},\hat{\gamma}^\dagger)$ yields

$$\begin{aligned}\hat{h}^{a,\dagger}(\hat{\beta}_r^{e\dagger},0) &= \hat{h}^{a,\dagger}(\hat{\beta}^{e\dagger},\hat{\gamma}^\dagger) \\ &\quad +\bar{D}^\dagger\left(\begin{array}{c}\hat{\beta}_r^{e\dagger}-\hat{\beta}^{e\dagger} \\ -\hat{\gamma}^\dagger\end{array}\right) \\ &= \hat{h}^{a,\dagger}(\hat{\beta}^{e\dagger},\hat{\gamma}^\dagger)+\bar{D}^\dagger[\left(\begin{array}{c}\hat{\beta}_r^{e\dagger}-\hat{\beta}_r^e \\ 0\end{array}\right)-\left(\begin{array}{c}\hat{\beta}^{e\dagger}-\hat{\beta}_r^e \\ \hat{\gamma}^\dagger\end{array}\right)],\end{aligned}$$

where $\bar{D}^{\dagger}=D(\bar{\beta})$ and $\bar{\beta}$ lies in a line joining $\hat{\beta}_r^{e\dagger}$ and $\hat{\beta}^{e\dagger}$. Thus

$$
\begin{aligned}
& (\frac{T}{k_2})\hat{h}^{a,\dagger}(\hat{\beta}_r^{e\dagger},0)'\hat{\Xi}^{\dagger-1}\hat{h}^{a,\dagger}(\hat{\beta}_r^{e\dagger},0) \\
= & (\frac{T}{k_2})[\hat{h}^{a,\dagger}(\hat{\beta}^{e\dagger},\hat{\gamma}^{\dagger})+\bar{D}^{\dagger}[\begin{pmatrix} \hat{\beta}_r^{e\dagger}-\hat{\beta}_r^{e} \\ 0 \end{pmatrix}-\begin{pmatrix} \hat{\beta}^{e\dagger}-\hat{\beta}_r^{e} \\ \hat{\gamma}^{\dagger} \end{pmatrix}]]'\hat{\Xi}^{\dagger-1} \\
& \times[\hat{h}^{a,\dagger}(\hat{\beta}^{e\dagger},\hat{\gamma}^{\dagger})+\bar{D}^{\dagger}[\begin{pmatrix} \hat{\beta}_r^{e\dagger}-\hat{\beta}_r^{e} \\ 0 \end{pmatrix}-\begin{pmatrix} \hat{\beta}^{e\dagger}-\hat{\beta}_r^{e} \\ \hat{\gamma}^{\dagger} \end{pmatrix}]] \\
= & (\frac{T}{k_2})\hat{h}^{a,\dagger}(\hat{\beta}^{e\dagger},\hat{\gamma}^{\dagger})'\hat{\Xi}^{\dagger-1}\hat{h}^{a,\dagger}(\hat{\beta}^{\dagger},\hat{\gamma}^{\dagger}) \\
& +(\frac{T}{k_2})2\{\bar{D}^{\dagger}[\begin{pmatrix} \hat{\beta}_r^{e\dagger}-\hat{\beta}_r^{e} \\ 0 \end{pmatrix}-\begin{pmatrix} \hat{\beta}^{e\dagger}-\hat{\beta}_r^{e} \\ \hat{\gamma}^{\dagger} \end{pmatrix}]\}'\hat{\Xi}^{\dagger-1}\hat{h}^{a,\dagger}(\hat{\beta}^{\dagger},\hat{\gamma}^{\dagger}) \\
& +(\frac{T}{k_2})\{\bar{D}^{\dagger}[\begin{pmatrix} \hat{\beta}_r^{e\dagger}-\hat{\beta}_r^{e} \\ 0 \end{pmatrix}-\begin{pmatrix} \hat{\beta}^{e\dagger}-\hat{\beta}_r^{e} \\ \hat{\gamma}^{\dagger} \end{pmatrix}]\}'\hat{\Xi}^{\dagger-1}\{\bar{D}^{\dagger}[\begin{pmatrix} \hat{\beta}_r^{e\dagger}-\hat{\beta}_r^{e} \\ 0 \end{pmatrix}-\begin{pmatrix} \hat{\beta}^{e\dagger}-\hat{\beta}_r^{e} \\ \hat{\gamma}^{\dagger} \end{pmatrix}\}.
\end{aligned}
$$

Now notice that by the first order conditions of the bootstrapped GMM problem we have

$$
\begin{aligned}
\sqrt{T}\bar{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{h}^{a,\dagger}(\hat{\beta}^{e\dagger},\hat{\gamma}^{\dagger}) &= \sqrt{T}(\bar{D}^{\dagger}-\hat{D}^{\dagger})'\hat{\Xi}^{\dagger-1}\hat{h}^{a,\dagger}(\hat{\beta}^{e\dagger},\hat{\gamma}^{\dagger}) \\
&= o_B(1)[\sqrt{T}\hat{h}^{a,\dagger}(\hat{\beta}^{\dagger},\hat{\gamma}^{\dagger}].
\end{aligned}
$$

Now note that by two Taylor expansions

$$
\begin{aligned}
\hat{h}^{a,\dagger}(\hat{\beta}^{e\dagger},\hat{\gamma}^{\dagger}) &= \hat{h}^{a,\dagger}(\hat{\beta}^{e},\hat{\gamma})+\ddot{D}^{\dagger}\begin{pmatrix} \hat{\beta}^{e\dagger}-\hat{\beta}^{e} \\ \hat{\gamma}^{\dagger}-\hat{\gamma} \end{pmatrix} \\
&= \hat{h}^{a,\dagger}(\beta_0,\gamma_0)+\breve{D}^{\dagger}\begin{pmatrix} \hat{\beta}^{e}-\beta_0 \\ \hat{\gamma} \end{pmatrix}+\ddot{D}^{\dagger}\begin{pmatrix} \hat{\beta}^{e\dagger}-\hat{\beta}^{e} \\ \hat{\gamma}^{\dagger}-\hat{\gamma} \end{pmatrix} \\
&= \hat{h}^{a,\dagger}(\beta_0,\gamma_0)-\tilde{h}^{a}(\beta_0,\gamma_0)+\tilde{h}^{a}(\beta_0,\gamma_0) \\
&\quad +\tilde{D}^{\dagger}\begin{pmatrix} \hat{\beta}^{e}-\beta_0 \\ \hat{\gamma} \end{pmatrix}+\ddot{D}^{\dagger}\begin{pmatrix} \hat{\beta}^{e\dagger}-\hat{\beta}^{e} \\ \hat{\gamma}^{\dagger}-\hat{\gamma} \end{pmatrix} \\
&= O_B(1/\sqrt{T}),
\end{aligned}
$$

where $\breve{D}^{\dagger}=D^{\dagger}(\breve{\beta})$ and $\breve{\beta}$ is in a line joining $\hat{\beta}^{e}$ and β_0 and $\ddot{D}^{\dagger}=D^{\dagger}(\ddot{\beta})$ and $\ddot{\beta}$ is in a line joining $\hat{\beta}^{e\dagger}$ and $\hat{\beta}^{e}$. Thus $\sqrt{T}\bar{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{h}^{a,\dagger}(\hat{\beta}^{\dagger},\hat{\gamma}^{\dagger})=o_B(1)$ and

$$
(\frac{T}{k_2})\{\bar{D}^{\dagger}[\begin{pmatrix} \hat{\beta}_r^{e\dagger}-\hat{\beta}_r^{e} \\ 0 \end{pmatrix}-\begin{pmatrix} \hat{\beta}^{e\dagger}-\hat{\beta}_r^{e} \\ \hat{\gamma}^{\dagger} \end{pmatrix}]\}'\hat{\Xi}^{\dagger-1}\hat{h}^{a,\dagger}(\hat{\beta}^{e\dagger},\hat{\gamma}^{\dagger})=o_B(1)
$$

as $\hat{\beta}_r^{e\dagger}-\hat{\beta}_r^{e}=O_B(1/\sqrt{T})$ and $\hat{\beta}^{e\dagger}-\hat{\beta}_r^{e}=O_B(1/\sqrt{T})$.

Additionally notice that

$$
\begin{aligned}
& (\frac{T}{k_2})\{\hat{h}^{a,\dagger}(\hat{\beta}^{e\dagger},\hat{\gamma}^{\dagger})'\hat{\Xi}^{\dagger-1}\hat{h}^{a,\dagger}(\hat{\beta}^{e\dagger},\hat{\gamma}^{\dagger})-\hat{h}^{a,\dagger}(\hat{\beta}^{e\dagger},\hat{\gamma}^{\dagger})'\tilde{\Xi}^{\dagger-1}\hat{h}^{a,\dagger}(\hat{\beta}^{e\dagger},\hat{\gamma}^{\dagger})\} \\
= & (\frac{T}{k_2})\{\hat{h}^{a,\dagger}(\hat{\beta}^{e\dagger},\hat{\gamma}^{\dagger})'(\hat{\Xi}^{\dagger-1}-\tilde{\Xi}^{\dagger-1})\hat{h}^{a,\dagger}(\hat{\beta}^{e\dagger},\hat{\gamma}^{\dagger})\}=o_B(1),
\end{aligned}
$$

as $\hat{\Xi}^{\dagger-1}-\tilde{\Xi}^{\dagger-1}=o_B(1)$ the result follows as we proved that $\hat{h}^{a,\dagger}(\hat{\beta}^{e\dagger},\hat{\gamma}^{\dagger})=O_B(1/\sqrt{T})$.

Also

$$
\sqrt{\frac{T}{k_2}}\bar{D}^{\dagger}\{\begin{pmatrix} \hat{\beta}_r^{e\dagger}-\hat{\beta}_r^{e} \\ 0 \end{pmatrix}-\begin{pmatrix} \hat{\beta}^{e\dagger}-\hat{\beta}_r^{e} \\ \hat{\gamma}^{\dagger} \end{pmatrix}\}=-\sqrt{\frac{T}{k_2}}D(\begin{pmatrix} \hat{\beta}^{e\dagger}-\hat{\beta}^{e} \\ \hat{\gamma}^{\dagger} \end{pmatrix}-\begin{pmatrix} \hat{\beta}_r^{e\dagger}-\hat{\beta}_r^{e} \\ 0 \end{pmatrix})+o_B(1).
$$

Thus

$$
\begin{aligned}
-\sqrt{\frac{T}{k_2}}D(\begin{pmatrix} \hat{\beta}^{e\dagger}-\hat{\beta}_r^{e} \\ \hat{\gamma}^{\dagger} \end{pmatrix}-\begin{pmatrix} \hat{\beta}_r^{e\dagger}-\hat{\beta}_r \\ 0 \end{pmatrix}) &= -\sqrt{\frac{T}{k_2}}D[\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\tilde{D}^{\dagger}]^{-1}[R(\hat{\beta}_r^{e\dagger}) \\
&\quad \times[R(\hat{\beta}_r^{e\dagger})'(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^{\dagger})^{-1}R(\hat{\beta}_r^{e\dagger})]^{-1} \\
&\quad \times R(\hat{\beta}_r^{e\dagger})'(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\hat{D}^{\dagger})^{-1}]\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\sqrt{T/k_2}(\hat{h}^{a,\dagger}(\beta_0,0)-\tilde{h}_T^{a}(\beta_0,0)) \\
&\quad +o_B(1),
\end{aligned}
$$

as

$$
\sqrt{T/k_2}\begin{pmatrix} \hat{\beta}^{e\dagger}-\hat{\beta}_r^{e} \\ \hat{\gamma}^{\dagger} \end{pmatrix}=-[\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}\breve{D}^{\dagger}]^{-1}\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger-1}[\sqrt{\frac{T}{k_2}}\hat{h}^{a,\dagger}(\beta_0,0)-\sqrt{T}\tilde{h}_T^{a}(\beta_0,0)],
$$

$$S_1\sqrt{T}(\hat{\beta}_r^{e\dagger}-\hat{\beta}_r^e) = -[\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger -1}\tilde{D}^{\dagger}]^{-1}[I-R(\hat{\beta}_r^{e\dagger})[R(\hat{\beta}_r^{e\dagger})^{\prime}(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger -1}\hat{D}^{\dagger})^{-1}R(\hat{\beta}_r^{e\dagger})]^{-1}R(\hat{\beta}_r^{e\dagger})^{\prime}(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger -1}\hat{D}^{\dagger})^{-1}]$$
$$\times\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger -1}\sqrt{T}[\hat{h}^{a,\dagger}(\beta_0,0)-\sqrt{T}\tilde{h}_T^a(\beta_0,0)]+o_p(1).$$

and the facts that $\hat{D}^{\dagger}=D+o_B(1)$ by Lemma A.7 also $R(\hat{\beta}_r^{e\dagger})=R+o_B(1)$, by continuity of $R(\beta)$ and $\hat{\Xi}^{\dagger -1}=\Xi^{-1}+o_p(1)$. Thus

$$\frac{T}{k_2}\{\bar{D}^{\dagger}[\begin{pmatrix}\hat{\beta}_r^{e\dagger}-\hat{\beta}_r^e\\ 0\end{pmatrix}-\begin{pmatrix}\hat{\beta}^{e\dagger}-\hat{\beta}_r^e\\ \hat{\gamma}^{\dagger}\end{pmatrix}]\}^{\prime}\hat{\Xi}^{\dagger -1}\{\bar{D}^{\dagger}[\begin{pmatrix}\hat{\beta}_r^{e\dagger}-\hat{\beta}_r^e\\ 0\end{pmatrix}-\begin{pmatrix}\hat{\beta}^{e\dagger}-\hat{\beta}_r^e\\ \hat{\gamma}^{\dagger}\end{pmatrix}]\}$$
$$=\frac{T}{k_2}[\sqrt{T/k_2}(\hat{h}^{a,\dagger}(\beta_0,0)-\tilde{h}_T^a(\beta_0,0))^{\prime}\hat{\Xi}^{\dagger -1}\hat{D}^{\dagger}(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger -1}\hat{D}^{\dagger})^{-1}R(\hat{\beta}_r^{e\dagger})[R(\hat{\beta}_r^{e\dagger})^{\prime}(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger -1}\hat{D}^{\dagger})^{-1}R(\hat{\beta}_r^{e\dagger})]^{-1}$$
$$\times R(\hat{\beta}_r^{e\dagger})^{\prime}[\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger -1}\tilde{D}^{\dagger}]^{-1}D^{\prime}\hat{\Xi}^{\dagger -1}D[\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger -1}\tilde{D}^{\dagger}]^{-1}$$
$$\times[R(\hat{\beta}_r^{e\dagger})[R(\hat{\beta}_r^{e\dagger})^{\prime}(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger -1}\hat{D}^{\dagger})^{-1}R(\hat{\beta}_r^{e\dagger})]^{-1}R(\hat{\beta}_r^{e\dagger})^{\prime}(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger -1}\hat{D}^{\dagger})^{-1}]$$
$$\times\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger -1}\sqrt{T/k_2}(\hat{h}^{a,\dagger}(\beta_0,0)-\hat{h}_T^a(\beta_0,0))$$
$$=\frac{T}{k_2}[\sqrt{T/k_2}(\hat{h}^{a,\dagger}(\beta_0,0)-\tilde{h}_T^a(\beta_0,0))^{\prime}\hat{\Xi}^{\dagger -1}\hat{D}^{\dagger}(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger -1}\hat{D}^{\dagger})^{-1}R(\hat{\beta}_r^{e\dagger})[R(\hat{\beta}_r^{e\dagger})^{\prime}(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger -1}\hat{D}^{\dagger})^{-1}R(\hat{\beta}_r^{e\dagger})]^{-1}$$
$$\times R(\hat{\beta}_r^{e\dagger})^{\prime}(\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger -1}\hat{D}^{\dagger})^{-1}]\hat{D}^{\dagger\prime}\hat{\Xi}^{\dagger -1}\sqrt{T/k_2}(\hat{h}^{a,\dagger}(\beta_0,0)-\tilde{h}_T^a(\beta_0,0))+o_B(1),$$

and the result follows as in the proof of Theorem 2.4. ■

A References

Allen, J., Gregory, A. W. and K. Shimotsu, (2011): "Empirical likelihood block bootstrapping," *Journal of Econometrics*, vol. 161(2), 110-121.

Andrews, D.W.K (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, vol. 59(3), 817-58.

Andrews, D.W.K., (2002): "Equivalence of the Higher Order Asymptotic Efficiency of k-step and Extremum Statistics," *Econometric Theory*, vol. 18(05), 1040-1085.

Andrews, D.W.K. and Monahan, J.C. (1992):"An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator," Econometrica, 60, 1992, 953—966.

Bravo F. and F. Crudu (2011): "Efficient Bootstrap with Weakly Dependent Processes," forthcoming in *Computational Statistics and Data Analysis*.

Brown, B.W. and W.K. Newey, (2002): "Generalized Method of Moments, Efficient Bootstrapping, and Improved Inference," *Journal of Business and Economic Statistics*, vol. 20(4), 507-17.

Burnside, C. and M.S. Eichenbaum, (1996): "Small-Sample Properties of GMM-Based Wald Tests," *Journal of Business and Economic Statistics*, vol. 14(3), 294-308.

Carlstein, E. (1986): "The use of subseries values for estimating the variance of a general statistic from a stationary sequence," *Annals of Statistics*, 14, 1171-1179.

Cattaneo,M.D.; Crump, R. K. and Jansson, M. (2010): "Bootstrapping Density-Weighted Average Derivatives," *CREATES Research Papers* 2010-23, School of Economics and Management, University of Aarhus.

Christiano, L. J. and den Haan, W.J., (1996):. "Small-Sample Properties of GMM for Business-Cycle Analysis," Journal of Business & Economic Statistics, American Statistical Association, vol. 14(3), pages 309-27, July

D'Antona, G. and A. Ferrero (2005): *Digital Signal Processing for Measurement Systems: Theory and Applications*, Springer, New York, U.S.A..

Davidson, J.E.H. (1994): *Stochastic Limit Theory.* Oxford: Oxford University Press.

Davidson, R. and J. G. MacKinnon, (1999): "Bootstrap Testing in Nonlinear Models," *International Economic Review*,, vol. 40(2), 487-508.

Efron B. (1979). "Bootstrap Methods: Another Look at the Jackknife," *Annals of Statistics,* 7 (1), 1–26.

Gallant, R. (1987): *Nonlinear Statistical Models*, John Wiley and Sons, New York,

Gonçalves S. and H. White (2004): "Maximum likelihood and the bootstrap for nonlinear dynamic models," *Journal of Econometrics*, vol. 119(1), 199-219.

Hahn, J. (1996): "A Note on Bootstrapping Generalized Method of Moments Estimators," *Econometric Theory*, vol. 12(01), 187-197.

Hall, A. (2005): *Generalized Method of Moments*, New York (NY): Oxford University Press.

Hall, P. (1992). The Bootstrap and Edgeworth Expansion, Springer-Verlag New York Inc.

Hall, P. and J. L. Horowitz (1996): "Bootstrap Critical Values for Tests Based on Generalized Method of Moments Estimators," *Econometrica*, 891-916.

Hansen, L. P., (1982): “Large Sample Properties of Generalized Method of Moments Estimators,” *Econometrica*, vol. 50(4), 1029-1054.

Inoue, A. and M. Shintani (2006): “Bootstrapping GMM estimators for time series,” *Journal of Econometrics*, vol. 133(2), 531-555.

Künsch, H. (1989): “The jacknife and the bootstrap for general stationary observations,” *Annals of Statistics,* 17, 1217–1241.

Liu, R. and K. Singh, (1992): “Moving blocks jacknife and bootstrap capture weak dependence,” in *Exploring the Limits of Bootstrap*. Wiley, New York, eds. Lepage, R. and Billard, L., 225–248.

Newey, W.K. and D. McFadden (1994): “Large Sample Estimation and Hypothesis Testing,” *Handbook of Econometrics*, Volume 4 R.F. Engle and D.L. McFadden (eds.), 2111-2245.

Newey, W.K. and K.D.West, (1987): “A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix,” *Econometrica*, 55, 703-08.

Newey, W. K, and K.D.West, (1994): “Automatic Lag Selection in Covariance Matrix Estimation,” *Review of Economic Studies*, vol. 61(4), 631-653.

Ng, S. and Perron, P. (1996): “The Exact Error in Estimating the Spectral Density at the Origin,” *Journal of Time Series Analysis* 17, 379–408.

Paparoditis, E. and D. N. Politis (2001): “Tapered Block Bootstrap,” *Biometrika*, vol. 88, No. 4, 1105-1119.

Parente, P.M.D.C., and Smith, R.J. (2018a): “Kernel Block Bootstrap,” CWP 48/18, Centre for Microdata Methods and Practice, U.C.L and I.F.S.

Parente, P.M.D.C. and R.J. Smith (2018b): “Quasi-Maximum Likelihood and The Kernel Block Bootstrap for Nonlinear Dynamic Models,” *working paper*.

Politis, D. and Romano, J. (1992b): “A general resampling scheme for triangular arrays of α-mixing random variables with application to the problem of spectral density estimation,” *Annals of Statistics*, 20, 1985-2007.

Politis, D. N. and Romano, J. P. (1995): “Bias-corrected nonparametric spectral estimation”. *Journal of Time Series Analysis* 16, 67–103.

Pollard, D. (1991): “Asymptotics for Least Absolute Deviation Regression Estimators,” *Econometric Theory*, Vol 7, 186-199.

Ramalho, J.J.S., and R.J. Smith (2011): “Goodness of Fit Tests for Moment Conditions Models,” *working paper*, Universidade de Evora.

Rao, C. (2002): *Linear Statistical Inference and its Applications*, Wiley.

Rao, C. R. and S. K. Mitra, (1971). *Generalized inverse of matrices and its applications.* New York: Wiley:

Royden H.L. (1988): *Real analysis*, 3ed., Macmillan.

Ruud, P. (2000): *An Introduction to Classical Econometric Theory*, Oxford University Press.

Serfling, R. (2002): *Approximation Theorems of Mathematical Statistics*, New York: Wiley.

Smith, R.J., (1997): “Alternative Semi-parametric Likelihood Approaches to Generalised Method of Moments Estimation,” *Economic Journal*, vol. 107(441), 503-19.

Smith. R.J. (2011): “GEL Criteria for Moment Condition Models,” *Econometric Theory*, 27, 1192-1235.

Tauchen G. (1985): “Diagnostic Testing and Evaluation of Maximum Likelihood Models,” *Journal of Econometrics*, 30, 415-443.

White, H., (1984): *Asymptotic theory for econometricians.* Academic Press.

White, H., (1999): *Asymptotic theory for econometricians.* Academic Press, 2nd ed.

White H. and X. Chen (1996): “Laws of Large Numbers for Hilbert Space-Valued Mixingales With Applications,” *Econometric Theory*, 12, 284-304 (1996).

Wooldridge,J. (1994): “Estimation and Inference for Dependent Processes,” in *Handbook of Econometrics*, Volume 4 R.F. Engle and D.L. McFadden (eds.), 2639-2738.