Generalised Empirical Likelihood Kernel Based Bootstrapping

Paulo M.D.C. Parente
Department of Economics,
University of Exeter
and
CReMic

Richard J. Smith
*c*emmap
U.C.L and I.F.S.
and
Faculty of Economics
University of Cambridge

March, 2012

Abstract

This article unveils how the kernel based bootstrap method for dependent data of Parente and Smith (2009) can be applied to make inferences on models defined through moment restrictions. We introduce bootstrap procedures that resort to generalised empirical likelihood implied probabilities to draw observations and prove that they are asymptotically valid. The advantage of resampling methods based such probabilities is that they were shown to be efficient by Brown and Newey (2002). We also consider bootstrap testing and establish the first-order asymptotic validity of bootstrapped test statistics for overidentifying moment restrictions, parametric restrictions and additional moment restrictions.

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 (2009). Simultaneously, we amend some 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 specially 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 those that have 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 (asymptotic) variance. From the practical standpoint it has the benefit of not requiring the application of complicated formulas 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 did not consider bootstrapped test statistics based on GMM. Hall and Horowitz (1996), Andrews (2002) and Inoue and Shintani (2006) propose the use 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 the 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 denominated 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 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 (2004). 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 the bootstrapped $\mathcal{J}$ statistic and bootstrapped Wald ($\mathcal{W}$) statistics tests for parametric restrictions. Finally, also 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 blocks of data and introduce an alternative to these resampling schemes based on the novel Kernel Based Bootstrap (KBB) method of Parente and Smith (2009). 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 these types of kernels. 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, 2004) to construct the bootstrap sample. We prove the first order validity of the bootstrapped distribution of the GMM estimator, the bootstrapped $\mathcal{J}$ statistic and the bootstrap test statistics 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 actually in error in that when applied to the inefficient GMM estimator the EL bootstrap distribution of the latter has to be centered at the efficient GMM estimator, not at an inefficient estimator as stated there. Hence the results presented in their Theorems 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)]. Despite the fact that we only prove this result for the new bootstrap methods introduced in this article, the demonstration for EL block bootstrapping is similar.

We generalise the work of Bravo and Crudu (2010) by considering bootstrapped statistics for parametric restrictions and additional moment conditions using GEL implied probabilities computed under the null. We note that the formula for the $\mathcal{W}$ statistic presented in Bravo and Crudu (2011) is only valid if the implied probabilities were computed under the maintained 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 test 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.

The essence of Bravo and Crudu (2011)' bootstrap method for parametric restrictions based on the implied probabilities computed under the null is similar to the idea of imposing the null hypothesis in the bootstrap data generating process. This approach was studied by numerous authors in the context of a regression model namely Li and Maddala (1996), Nankervis and Savin (1996), Flaichaire (1999), Park (2000), van Giersbergen and Kiviet (2001) and Swensen (2003). However, Paparoditis and Politis (2005) showed that imposing the null to generate the bootstrap samples usually leads a loss in power if a non-pivotal statistic is used. The Monte Carlo study of Bravo and Crudu (2011) reveals that the power of pivotal test statistics in their simulation study is higher imposing the null than not imposing.

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 the GMM and GEL estimators in the time-series context that are required to implement the bootstrap methods proposed in this article. The KBB method is briefly summarized in section 3. In section 4 we present the first order asymptotic theory on the KBB methods computed using the following different types of probabilities to draw bootstrap samples: 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. Finally section 5 concludes. A selection of the proofs of the results in the main text are sketched in the Appendix. Proofs of Theorems and Lemmata not given here are available upon request.

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 α-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 $\mathrm{E}[\bullet]$ 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 $g_t(\beta) \equiv g(z_t, \beta)$, $(t = 1, ..., T)$, $\hat{g}(\beta) \equiv \sum_{s=1}^{T} g_t(\beta)/T$, $G_t(\beta) \equiv \partial g_t(\beta)/\partial \beta'$, $(t = 1, ..., T)$, $G \equiv \mathrm{E}[G_t(\beta_0)]$ 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) proved that the GMM estimator is consistent for $\hat{\beta}$ and asymptotically normal. We present these results under assumptions different from those stated in the original article of Hansen (1982) to facilitate comparisons with the assumptions that we make in the next sections for GEL and KBB.

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* $z \equiv \{z_t : \Omega \to \mathbb{R}^n, n \in \mathbb{N}, t = 1, 2, ...\}$ *on the complete probability space* $(\Omega, \mathcal{F}, P)$ *where* $\Omega = \times_{t=1}^{\infty}\mathbb{R}^k$ *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 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)** $\mathrm{E}[\sup_{\beta\in B} \|g(z_t,\beta)\|] < \Delta < \infty$, **(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).[1]

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.[2]

Assumption 2.2 **(i)** $\{z_t, -\infty < t < \infty\}$ *is a strong mixing process with mixing coefficients of size* $-r/(r-2)$, $r > 2$, $\mathrm{E}[\|g(z_t,\beta_0)\|^r] < \Delta < \infty$, $r \geq 2$; **(ii)** $G_t(\beta)$ *exists and is continuous on* $\mathcal{B}$ *for each* $z_t \in \mathcal{Z}$ **(iii)** $rank(G) = p$ *(iii)* $\mathrm{E}[\sup_{\beta\in\mathcal{N}} \|G_t(\beta)\|] < \Delta < \infty$, *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 $\mathrm{avar}(\hat{\beta}) = (G'WG)^{-1}G'W\Omega WG(G'WG)^{-1}$.

Denote $\Sigma \equiv (G'\Omega^{-1}G)^{-1}$. Hansen (1982) proved also that the most efficient GMM estimator $\hat{\beta}^e$ is obtained when $W = \Omega^{-1}$ and in this case $\mathrm{avar}(\hat{\beta}^e) = \Sigma$.

To obtain an efficient estimator we need to estimate Ω. Numerous estimators for Ω have been proposed in the literature under different assumptions, most of them motivated by estimators proposed in the 50's and 60's is statistics for the spectral density of the

[1]For a version of the Theorem on *almost surely* convergence of the GMM estimator see Theorem 2.1 of Hansen (1982).

[2]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.

process at frequency zero [see White (1984), Newey and West (1987), Gallant (1981), Andrews (1991), Ng and Perron (1996)]. Let $\hat{\Omega} = \Omega + o_p(1)$, the efficient two-step GMM estimator is defined as

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

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

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

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 Theorem 2.4.

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)** $\mathrm{E}[\|q(z_t,\beta_0)\|^r] \leq \Delta < \infty$, $r \geq 2$ $Q_t(\beta)$ *exists and is continuous on* $\mathcal{B}$ *for each* $z_t \in \mathcal{Z}$; **(v)** *rank*$(Q) = s$; **(vi)** $\mathrm{E}[\sup_{\beta\in\mathcal{N}} \|Q_t(\beta)\|] < \Delta < \infty$; **(vii)** Ξ *is non-singular and* $\hat{\Xi} = \Xi + o_p(1)$.

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 (2004). Consider the smoothed moments

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

where the kernel function $k(\bullet)$ 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 $\rho(\bullet)$ be a function that is concave on its domain $\mathcal{V}$, an open interval containing zero. It is convenient to impose a normalization on $\rho(\bullet)$. Let $\rho_j(\bullet) = \partial^j \rho(\bullet)/\partial v^j$ 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 (2004) as

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

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 (2004) 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 (2004) 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 (2004).*

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 *$S_T = O(T^{1/2-\eta})$ for some $\eta \in (1/6, 1/2)$.*

Assumption 2.7 **(i)** *$\beta_0 \in \mathcal{B}$ is the unique solution of* $\mathrm{E}[g_t(\beta)] = 0$; **(ii)** *$\mathcal{B}$ is compact*; **(iii)** *$g_t(\beta)$ is continuous at each $\beta \in B$*; **(iv)** $\mathrm{E}[\sup_{\beta\in\mathcal{B}} \|g_t(\beta)\|^{\alpha}] < \infty$ *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)** *$\rho(\bullet)$ 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 (2004).

Theorem 2.5 *If Assumptions 2.4, 2.5 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 (2004) also required the following assumptions.

Assumption 2.9 **(i)** *$\beta_0 \in int(\mathcal{B})$*; **(ii)** *$g(\bullet, \beta)$ 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 (2004) proved the following theorem.

Theorem 2.6 *If Assumptions 2.4, 2.5, 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 Based 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 validly of the method was proven by Parente and Smith (2009).

In this article we modify the original method as now each transformed 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 $0 \leq p_{tT} \leq 1$ and $\sum_{t=1}^{T} p_{tT} = 1$. The standard KBB method of Parente and Smith (2009) 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 require the following assumption taken from Parente and Smith (2009).

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)** $\mathrm{E}[|X_t|^{\alpha}] < \Delta < \infty$, *for some* $\alpha > \max(4v, 1/\eta)$, **(iii)** $\sigma_{\infty}^2 \equiv \lim_{T\to\infty} \mathrm{var}[T^{1/2}\bar{X}]$ *is finite.*

Assumption 3.3 **(i)** $0 \leq p_{tT} \leq 1$, $\sum_{t=1}^{T} p_{tT} = 1$, $\max_{1\leq t\leq T} |Tp_{tT}| = o_p(1)$, **(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 the measures of magnitude of bootstrapped sequences as defined by Hanh (1996): Let $\xi_T^* = O_p^\nu(a_T)$ if ξ_T^*. when conditioned on ν is $O_p(a_T)$ and $\xi_T^* = o_p^\nu(a_T)$ if ξ_T^*. when conditioned on ν 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 $O_p^\nu(1)$. Similarly we write $\xi_T^* = o_B(1)$ if, for a given subsequence $\{T'\}$ there exists a further subsequence $\{T''\}$ such that $o_p^\nu(1)$.

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* $\mathrm{E}[X_t] = 0$

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

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

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

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

4 Kernel based 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 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 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)** $\mathrm{E}[\sup_{\beta \in \mathcal{B}} \|g(z_t, \beta)\|^{\alpha}] < \Delta < \infty$ *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 $W = \Omega^{-1}$. We now unveil how to obtain consistent estimator for Ω using the bootstrap. Let $\hat{\Omega}^*(\tilde{\beta}^*) \equiv \frac{S_T}{m_T k_2} \sum_{t=1}^{m_T} g_t^*(\tilde{\beta}^*) g_t^*(\tilde{\beta}^*)'$ 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,*

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

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 $\hat{h}_T^*(\beta) = \frac{1}{m_T}\sum_{s=1}^{m_T} h_{sT}^*(\beta)$,

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

and

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

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 define the following bootstrapped statistics

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

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)** $\mathrm{E}[\sup_{\beta\in\mathcal{N}}\|\partial q(z,\beta)/\partial\beta'\|^{a}] < \Delta < \infty$; **(vii)** *$\Xi$ exists and is positive definite and $\hat{\Xi} = \Xi + o_p(1)$.*

Theorem 4.4 reveals that under Assumptions 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 based 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 based 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, 2.8 and 2.9* $\tilde{\beta} \xrightarrow{p} \beta_0$.

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*

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

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

$$\hat{\beta}^{\star} = \arg\min_{\beta\in\mathcal{B}} \hat{g}_T^{\star}(\beta)' W_T^{\star} \hat{g}_T^{\star}(\beta),$$

where $W_T^{\star} = W_T + o_B(1)$.

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 ,*

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

We note that $\hat{\beta}^{\star}$ is centered at the efficient estimator $\tilde{\beta}$, not at 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 $W = \Omega^{-1}$.

Remark: Although in the proofs we assume that $\hat{\pi}_t$, $t = 1, ..., T$ was obtained using GEL, it can be obtained using the two step efficient GMM estimator. In this case $\hat{\pi}_t$ is replaced by the Back and Brown (1993) estimator (see also Guay and Pelgrin, 2008):

$$\begin{aligned} p_{\text{GMM},t} &= \frac{1}{T} - \frac{1}{T-p}\frac{S_T}{k_2}[g_{tT}(\hat{\beta}^e) - \hat{g}_T(\hat{\beta}^e)]'\hat{\Omega}^{-1}\hat{g}_T(\hat{\beta}^e), \\ t &= 1, ..., T. \end{aligned}$$

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

$$\hat{\Omega}^\star(\bar{\beta}^\star) \equiv \frac{S_T}{m_T k_2}\sum\nolimits_{t=1}^{m_T} g_t^\star(\bar{\beta}^\star)g_t^\star(\bar{\beta}^\star)'.$$

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 reveals that the bootstrap distribution of $\mathcal{J}^\star$ converges to the asymptotic distribution of $\mathcal{J}$.

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 $\hat{\Xi}^\star = \Xi + o_B(1)$, $\hat{h}^\star(\beta) = \frac{1}{m_T}\sum_{s=1}^{m_T} h^\star_{sT}(\beta)$, $\tilde{h}_T(\beta) = \sum_{t=1}^{T} h_{t,T}(\beta)\hat{\pi}_t$ 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

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

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

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

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*

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

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 based bootstrap versions of the tests for parametric restrictions and additional moment conditions that impose the null hypothesis through the generalised empirical likelihood implied probabilities which is similar to the method of Bravo and Crudu (2011) for MBB.

Before describing the new the method we 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. Let $\hat{\varphi}(\beta)=\arg\sup_{\varphi\in\Delta_T}\bar{P}_T(\beta,\varphi)$, $\hat{\varphi}_r\equiv\hat{\varphi}(\hat{\beta}_{r,\text{GEL}})$ and

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

Consider a bootstrap sample of size m_T, $\left\{h_{tT}^{\dagger}(\beta)\right\}_{t=1}^{m_T}$, drawn from $\{h_{tT}(\beta)\}_{t=1}^{T}$, where $\mathcal{P}(h_{jT}^{\dagger}(\beta)=h_{tT}(\beta))=\tilde{\pi}_t$, $t=1,...,T$ and $j=1,...,m_T$. We consider the case that the bootstrap weighting matrix is $W_T^{\dagger}=\hat{\Xi}^{\dagger-1}$, where $\hat{\Xi}^{\dagger}=\Xi+o_B(1)$. Define $\hat{h}_T^{\dagger}(\beta)\equiv\frac{1}{m_T}\sum_{s=1}^{m_T}h_{sT}^{\dagger}(\beta)$, $\bar{Q}_T^{\dagger}(\beta)=\hat{h}_T^{\dagger}(\beta)'\hat{\Xi}^{\dagger-1}\hat{h}_T^{\dagger}(\beta)$ and let

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

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

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

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 uniformly close to the asymptotic distributions of the statistics 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 $\Delta_T = \left\{\varphi : \|\varphi\| \le D(T/S_T^2)^{-\zeta}\right\}$, 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.*

Remark: As in section 4.2.2 $\tilde{\pi}_t$, $t = 1, ..., T$ can be obtained using the two step efficient GMM estimator. In this case $\tilde{\pi}_t$ is replaced by the Back and Brown (1993) estimator :

$$\begin{aligned}
\tilde{p}_{\text{GMM},t} &= \frac{1}{T} - \frac{1}{T-p}\frac{S_T}{k_2}[h_{tT}(\hat{\beta}_r^e) - \hat{h}_T(\hat{\beta}_r^e)]'\hat{\Xi}^{-1}h_T(\hat{\beta}_r^e), \\
t &= 1, ..., T.
\end{aligned}$$

5 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 Based Bootstrap method of Parente and Smith (2009). These methods approximate the asymptotic distributions of GMM 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 generalising and correcting the work of Allen et al. (2011) and Bravo and Crudu (2011).

6 References

Allen, J., Gregory, A. W. and K. Shimotsu, (2011): "Empirical likelihood block bootstrapping," *Journal of Econometrics*, vol. 161(2), 110-121.

Amemiya, T. (1985): *Advanced Econometrics.* Harvard University

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.

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.

Eichenbaum, M.S.; Hansen, L. P. and Singleton, K. J., (1988): "A Time Series Analysis of Representative Agent Models of Consumption and Leisure Choice under Uncertainty," *The Quarterly Journal of Economics*, vol. 103(1), pages 51-78.

Flaichaire, E. (1999): "A better way to bootstrap pairs," *Economic Letters*, vol. 64, 257-262.

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.

Guay, A. and F. Pelgrin (2008): "Using implied probabilities to improve estimation in unconditional moment models," unpublished working paper.

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. 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.

Kunsch, H. (1989): "The jacknife and the bootstrap for general stationary observations," *Annals of Statistics,* 17, 1217–1241.

Li, H., and G. S. Maddala (1996). "Bootstrapping Time Series Models," *Econometric Reviews*, 15, 297-318.

Liu, R., Singh, K., 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.

Nankervis, J. C. and N. E. Savin (1996). "The level and power of the bootstrap t test in the AR(1) model with trend," *Journal of Business and Economic Statistics*, 14, 161-168.

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. Politis (2005): "Bootstrap hypothesis testing in regression models," *Statistics and Probability Letters* , vol. 74, 356-365.

Paparoditis, E. and D. N. Politis (2001): "Tapered Block Bootstrap," *Biometrika*, vol. 88, No. 4, 1105-1119.

Parente, P.M.D.C. and R.J. Smith (2009): "A Kernel Based Bootstrap Method for dependent processes," *unpublished working paper.*

Park, J.Y., (2003): "Bootstrap Unit Root Tests," *Econometrica*, vol. 71(6), 1845-1895.

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.

Pollard, D. (1991): "Asymptotics for Least Absolute Deviation Regression Estimators," *Econometric Theory*, Vol 7, 186-199.

Ramalho, J.J.S., and R.J. Smith (2004): "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 (2004): "GEL Criteria for Moment Condition Models," *CeMMAP working papers*, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Swensen, A. R. (2003), "Bootstrapping unit root tests for integrated processes," *Journal of Time Series Analysis*, 24, 99–126.

Tauchen G. (1985): "Diagnostic Testing and Evaluation of Maximum Likelihood Models," *Journal of Econometrics*, 30, 415-443.

van Giersbergen, N. P. A. & Kiviet, J. F., 2002. "How to implement the bootstrap in static or stable dynamic regression models: test statistic versus confidence region approach," *Journal of Econometrics*, vol. 108(1), 133-156.

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.

Appendix: Proofs and Auxiliary Results

Throughout the Appendix, C and Δ will denote a 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 $\mathrm{E}(h_t^a(\beta_0, 0)) = 0$ and $a(\beta_0) = 0$. Define $\theta = (\beta', \gamma')'$ and $\hat{h}^a(\theta) = \sum_{t=1}^T h_t^a(\beta,\gamma)/T$.

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

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

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

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

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 $\mathcal{W}\overset{d}{\to}\chi^2(r+s)$.

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

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

Note that the restricted GMM estimator solves

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

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.

By the usual arguments based on a Lagrangean 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

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

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

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

It follows from replacing $\sqrt{T}\hat{h}^a(\hat{\theta}_r^e)$ 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 (2004) 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 $t=1,...,T$.

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 (2004).

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

$$\sqrt{T}\tilde{w}=T^{-1/2}\sum_{t=1}^T w_t-B_0PT^{1/2}\hat{g}(\beta_0)+o_p(1),$$

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

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

Also by a UWL

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

Now

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

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 (2004) 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

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

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

Let

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

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

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

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

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

and S_1 is a matrix such that $S_1\hat{\beta}_r = (\hat{\beta}_r', 0_{s\times 1}')'$.

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

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

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*

$$\sqrt{T}\tilde{w}_r = T^{-1/2}\sum_{t=1}^{T} w_t - J_0 P_r T^{1/2}\hat{h}(\beta_0) + o_p(1),$$

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

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

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

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

Proof of Theorem 3.1: The proof is similar to that of proof of Theorem 2.1 of Parente and Smith (2009). 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

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

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: $\sup_{x\in\mathbb{R}} \left|\mathcal{P}\{T^{1/2}\bar{X} \leq x\} - \Phi(x/\sigma_\infty)\right| \to 0$, where $\Phi(\cdot)$ is the c.d.f. of the standard normal distribution. Step 4: $m_T \text{var}^*[\tilde{q}_T^{e*}] \xrightarrow{p} \sigma_\infty^2$. Step 5:

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

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

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

and

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

as $\max\limits_{1\leq t\leq T} |Tp_{tT}| \xrightarrow{p} 1$ and the fact that $\tilde{Y} = O_p(\frac{1}{\sqrt{T}})$ and Lemma A.3 of Smith (2004).

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

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

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

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

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 (2004) and M with $\alpha > \max(4v, 1/\eta)$.

Hence

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

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)** *$\mathrm{E}[\|\|X_t\|\|^{4v}] < \Delta < \infty$;* **(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*

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

Let also $\bar{Y} \equiv \frac{1}{T}\sum_{t=1}^T Y_{tT}$ 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)** *$\mathrm{E}[|X_t|^\tau] < \Delta < \infty$ 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*

$$\bar{Y}^* - \bar{Y} \to 0, \text{ prob-}\mathcal{P}^*, \text{ prob-}\mathcal{P}, \quad \text{(A.4)}$$

$$\bar{Y}^* - \tilde{Y} \to 0, \text{ prob-}\mathcal{P}^*, \text{ prob-}\mathcal{P} \quad \text{(A.5)}$$

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 based 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*

$$q_{tT}(\theta) = \frac{1}{S_T}\sum_{s=t-T}^{t-1} k(\frac{s}{S_T})g(X_{t-s},\theta), \quad \text{(A.6)}$$

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,*

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

(b) *Uniform Convergence:*

$$\sup_{\theta\in\Theta}\left|\frac{1}{T}\sum\nolimits_{t=1}^{T} q_{tT}(\theta) - \frac{1}{T}\sum\nolimits_{t=1}^{T} g(X_t,\theta)\right| \xrightarrow{p} 0,$$

$$\mathrm{E}[\sup_{\theta\in\Theta}|g(X_t,\theta)|] \le \Delta$$

(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 $\xi > 0$*

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

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

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*

$$q_{tT}(\theta) = \frac{1}{S_T}\sum\nolimits_{s=t-T}^{t-1} k(\frac{s}{S_T})g(X_{t-s},\theta), \tag{A.7}$$

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 **(i)** *hold and that :* **(a)** *Bootstrap Pointwise Weak Law of Large Numbers. for ,*

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

(b) $\mathrm{E}[\sup_{\theta\in\mathcal{N}}|g(X_t,\theta)|] \le \Delta$ *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,*

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

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

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

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

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

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

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

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 $1 < p \le 2$ 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*

$$\lim_{n\to\infty} \mathcal{P}[\mathcal{P}^*(|\frac{S_T}{m_T}\sum\nolimits_{t=1}^{m_T} Y^*_{tT}Z^*_{tT}| > T^{1/2}\varepsilon) > \varepsilon] = 0$$

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

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 (2004).

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

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

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

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

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

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

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

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

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

where $\tilde{G}_T^* \equiv \partial \hat{g}_T^\star(\tilde{\beta}^*) \partial \beta'$ 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

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

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

Using the first order conditions of the original GMM problem it can be shown that

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

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

$$\lambda'\hat{\Omega}^*(\beta_0)\lambda = \frac{S_T}{m_T k_2}\sum_{t=1}^{m_T}(\lambda' g_t^*(\beta_0))^2.$$

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

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

and by Smith (2004) Lemma A.3

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

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

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

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

$$\lambda'\hat{\Omega}^*(\tilde{\beta}^*)\lambda = \lambda'\hat{\Omega}^*(\beta_0)\lambda + \frac{S_T}{m_{Tk_2}}\sum_{t=1}^{m_T}2(\lambda' g_t^*(\ddot{\beta}^*)\lambda' G_t^*(\ddot{\beta}^*)(\tilde{\beta}^* - \beta_0)).$$

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

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

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

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

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

$$\sqrt{T/k_2}(\hat{g}^*(\hat{\beta}^e) - \hat{g}^*(\beta_0) - \hat{g}(\hat{\beta}^e) + \hat{g}(\beta_0)) = o_B(1). \tag{A.8}$$

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

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

Let $P = \Omega^{-1} - \Omega^{-1}G\left[G'\Omega^{-1}G\right]^{-1}G'\Omega^{-1}$ As $\Omega P\Omega P\Omega = \Omega P\Omega$ and $tr(P\Omega) = m - p$, 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

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

Now notice that as in the proof of Theorem 4.2

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

Thus by a Taylor expansion we have

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

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 $\mathcal{S}^*$. Using the usual arguments based on a Lagrangean we have

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

Consider now the bootstrapped score statistic

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

Now as in $(A.8)$ above we have

$$\sqrt{\frac{T}{k_2}}(\hat h^{a,*}(\hat\beta_r^e,0) - \hat h^{a,*}(\beta_0,0) - \hat h^a(\hat\beta_r^e,0) + \hat h^a(\beta_0,0)) = o_B(1). \tag{A.9}$$

Note that by a Taylor expansion of $\hat h^*(\hat\beta_r^*)$ around $\hat\beta_r$ we have

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

by A.9, the local bootstrap UWL, and the bootstrap CLT.

Thus

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

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

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

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

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

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

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

Note that $D^*(\hat{\beta}^{e*})'\hat{\Xi}^{*-1}\hat{h}^{a,*}(\hat{\beta}^{e*},\hat{\gamma}^*) = 0$ and $\hat{D}\hat{\Xi}\hat{h}^a(\hat{\beta}^e,\hat{\gamma}) = 0$. which can be used to show that

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

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

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

Thus

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

■

A.5 Proofs of the results in section 4.2

Proof of Theorem 4.7: This is similar to the proof of Theorem 4.1 ■

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

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

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

Now notice that by a Taylor expansion

$$\sqrt{T/k_2}\hat{G}_T^{\star\prime}W_T^\star\hat{g}_T^\star(\tilde{\beta})=\sqrt{T/k_2}\hat{G}_T^{\star\prime}W_T^\star[\hat{g}_T^\star(\beta_0)-\tilde{g}_T(\beta_0)]+\sqrt{T/k_2}\hat{G}_T^{\star\prime}W_T^\star\tilde{g}_T(\beta_0)+\sqrt{T/k_2}\hat{G}_T^{\star\prime}W_T^\star\check{G}_T^\star(\tilde{\beta}-\beta_0),$$

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

Now note that we have

$$\sqrt{T}(\tilde{\beta}-\beta_0)=-(G'WG)^{-1}\sqrt{T}G'W\tilde{g}_T(\beta_0)+o_p(1).$$

Thus

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

Now $\sqrt{T/k_2}\hat{G}_T^{*\prime}W_T^*[\hat{g}_T^*(\beta_0)-\tilde{g}_T(\beta_0)]$ 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 $p_{tT}=\hat{\pi}_t$ and uses the fact that $T\hat{\pi}_t=1+o_p(1)$ 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 $\tilde{G}_T^* \equiv \partial\hat{g}_T^\star(\ddot{\beta})/\partial\beta'$ where $\ddot{\beta}$ is in a line joining $\hat{\beta}^{e\star}$ and $\tilde{\beta}$.

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

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

Also by a Taylor expansion

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

where $\breve{G}_T^\star = \partial\hat{g}_T^\star(\breve{\beta})/\partial\beta'$ where $\breve{\beta}$ 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 $\breve{G}_T = G + o_B(1)$ by Lemma A.7, $\ddot{G}_T = G + o_p(1)$ by Lemma A.1 of Smith (2004)- and the fact that $T\hat{\pi}_t = 1 + o_p(1)$ 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 (2004)- and the fact that $T\hat{\pi}_t = 1 + o_p(1)$ 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

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

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

Thus

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

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 $\mathcal{J}^\star = T\hat{g}^\star(\hat{\beta}^{e\star})'\tilde{\Omega}^{*-1}\hat{g}^\star(\hat{\beta}^{e\star})/k_2$ 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 Theorem 4.10: We only derive the asymptotic distribution of $\mathcal{W}^\star$ here. The asymptotic equivalence $\mathcal{S}^\star, \mathcal{D}^\star$ and $.\mathcal{W}^\star$ can be shown as in the proof of Theorem 4.4. 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 $\left(\hat{\beta}^{e\prime}, \hat{\gamma}'\right)'$. Thus

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

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 $\hat{\beta}^{e}$.

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

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

First notice that by Lemma A.2 above we have

$$T^{1/2}\tilde{h}^{a}_{T}(\beta_0,0) = T^{-1/2}\sum\nolimits_{t=1}^{T}\hat{h}^{a}_{T}(\beta_0,0) + \Xi S_1 P T^{1/2}\hat{g}(\beta_0) + o_p(1)$$

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

$$[\hat{D}^{\star\prime}\hat{\Xi}\ddot{D}^{\star}]^{-1}\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}T^{1/2}\tilde{h}^{a}_{T}(\beta_0,0) = [\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}\ddot{D}^{\star}]^{-1}\hat{D}^{\star\prime}\hat{\Xi}^{\star-1}T^{-1/2}\sum\nolimits_{t=1}^{T}\hat{h}^{a}_{T}(\beta_0,\gamma) + o_B(1).$$

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

Now notice that

$$\sqrt{T}\begin{pmatrix} \hat{\beta}^{e}-\beta_0 \\ \hat{\gamma} \end{pmatrix} = \sqrt{T}\begin{pmatrix} \hat{\beta}^{e}-\beta_0 \\ \tilde{\gamma} \end{pmatrix} - \sqrt{T}\begin{pmatrix} 0 \\ \tilde{\gamma}-\hat{\gamma} \end{pmatrix} \tag{A.10}$$

and the usual asymptotic representation of the efficient GMM estimator yields

$$\sqrt{T}\begin{pmatrix} \hat{\beta}^{e}-\beta_0 \\ \hat{\gamma} \end{pmatrix} = -[\hat{D}'\hat{\Xi}^{-1}\ddot{D}]^{-1}\hat{D}'\hat{\Xi}^{-1}T^{-1/2}\sum\nolimits_{t=1}^{T}\hat{h}^{a}_{t}(\beta_0,0) + o_p(1). \tag{A.11}$$

where $\ddot{D} = D(\ddot{\beta})$ and $\ddot{\beta}$ is in a line joining $\hat{\beta}^{e}$ and β_0. Hence by $(A.10)$ and $(A.11)$ 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

$$\sqrt{T}\tilde{q}(\hat{\beta}^{e}) = \hat{q}(\hat{\beta}^{e}) + \frac{S_T}{T}\sum\nolimits_{t=1}^{T} q_{tT}(\hat{\beta}^{e})g_{tT}(\hat{\beta}^{e})'\frac{T^{1/2}}{S_T}\hat{\lambda}(1/k_2 + o_p(1)) + o_p(1)$$

Also by the FOC of the GEL problem with respect to λ and by a Taylor expansion around 0 we have

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

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 (2004). 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 (2004) $\sqrt{T}\hat{g}_T(\beta_0) = \sqrt{T}\hat{g}(\beta_0) + O_p(T^{-1/2})$.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 $\sqrt{T/k_2}[\hat{h}_T^{a,\star}(\beta_0,0) - \tilde{h}_T^{a}(\beta_0,0)]$ converging to $N(0,\Xi)$. ■

Proof of Theorem 4.11: We only derive the asymptotic distribution of $\mathcal{W}^{\dagger}$ here. The asymptotic equivalence $\mathcal{S}^{\dagger}, \mathcal{D}^{\dagger}$ and .$\mathcal{W}^{\dagger}$ can be shown as in the proof of Theorem 4.4. 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

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

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

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

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

$$\sqrt{T}\tilde{h}_T^a(\beta_0, 0) = DS_1[\Lambda - \Lambda R' \left[R\Lambda R'\right]^{-1} R\Lambda]D'\Xi^{-1}]\sqrt{T}\hat{h}_T^a(\beta_0, 0) + o_p(1), \tag{A.12}$$

Consequently

$$\sqrt{T}\tilde{h}^a(\beta_0, 0) + \tilde{D}^{\dagger} S_1 \sqrt{T}(\hat{\beta}_r - \beta_0) = o_p(1). \tag{A.13}$$

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

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

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 $N(0, \Xi)$. ■