

2014s-25

**Bootstrapping the GMM overidentification test  
Under first-order underidentification**

*Prosper Dovonon, Sílvia Gonçalves*

---

**Série Scientifique**  
*Scientific Series*

---

**Montréal**  
**Avril 2014**

© 2014 *Prosper Dovonon, Sílvia Gonçalves*. Tous droits réservés. *All rights reserved*. Reproduction partielle permise avec citation du document source, incluant la notice ©.  
*Short sections may be quoted without explicit permission, if full credit, including © notice, is given to the source.*



Centre interuniversitaire de recherche en analyse des organisations

## **CIRANO**

Le CIRANO est un organisme sans but lucratif constitué en vertu de la Loi des compagnies du Québec. Le financement de son infrastructure et de ses activités de recherche provient des cotisations de ses organisations-membres, d'une subvention d'infrastructure du Ministère de l'Enseignement supérieur, de la Recherche, de la Science et de la Technologie, de même que des subventions et mandats obtenus par ses équipes de recherche.

*CIRANO is a private non-profit organization incorporated under the Québec Companies Act. Its infrastructure and research activities are funded through fees paid by member organizations, an infrastructure grant from the Ministère de l'Enseignement supérieur, de la Recherche, de la Science et de la Technologie, and grants and research mandates obtained by its research teams.*

### **Les partenaires du CIRANO**

#### **Partenaire majeur**

Ministère de l'Enseignement supérieur, de la Recherche, de la Science et de la Technologie

#### **Partenaires corporatifs**

Autorité des marchés financiers  
Banque de développement du Canada  
Banque du Canada  
Banque Laurentienne du Canada  
Banque Nationale du Canada  
Banque Scotia  
Bell Canada  
BMO Groupe financier  
Caisse de dépôt et placement du Québec  
Fédération des caisses Desjardins du Québec  
Financière Sun Life, Québec  
Gaz Métro  
Hydro-Québec  
Industrie Canada  
Intact  
Investissements PSP  
Ministère des Finances et de l'Économie  
Power Corporation du Canada  
Rio Tinto Alcan  
Transat A.T.  
Ville de Montréal

#### **Partenaires universitaires**

École Polytechnique de Montréal  
École de technologie supérieure (ÉTS)  
HEC Montréal  
Institut national de la recherche scientifique (INRS)  
McGill University  
Université Concordia  
Université de Montréal  
Université de Sherbrooke  
Université du Québec  
Université du Québec à Montréal  
Université Laval

Le CIRANO collabore avec de nombreux centres et chaires de recherche universitaires dont on peut consulter la liste sur son site web.

Les cahiers de la série scientifique (CS) visent à rendre accessibles des résultats de recherche effectuée au CIRANO afin de susciter échanges et commentaires. Ces cahiers sont écrits dans le style des publications scientifiques. Les idées et les opinions émises sont sous l'unique responsabilité des auteurs et ne représentent pas nécessairement les positions du CIRANO ou de ses partenaires.

*This paper presents research carried out at CIRANO and aims at encouraging discussion and comment. The observations and viewpoints expressed are the sole responsibility of the authors. They do not necessarily represent positions of CIRANO or its partners.*

ISSN 2292-0838 (en ligne)

Partenaire financier

Enseignement supérieur,  
Recherche, Science  
et Technologie  
Québec 

# Bootstrapping the GMM overidentification test Under first-order underidentification

*Prosper Dovonon*<sup>\*</sup>, *Sílvia Gonçalves*<sup>†</sup>

## Résumé/abstract

The main contribution of this paper is to study the applicability of the bootstrap to estimating the distribution of the standard test of overidentifying restrictions of Hansen (1982) when the model is globally identified but the rank condition fails to hold (lack of first order local identification). An important example for which these conditions are verified is the popular test of common conditionally heteroskedastic features proposed by Engle and Kozicki (1993). As Dovonon and Renault (2013) show, the Jacobian matrix for this model is identically zero at the true parameter value, resulting in a highly nonstandard limiting distribution that complicates the computation of critical values. We first show that the standard bootstrap method of Hall and Horowitz (1996) fails to consistently estimate the distribution of the overidentification restrictions test under lack of first order identification. We then propose a new bootstrap method that is asymptotically valid in this context. The modification consists of adding an additional term that recenters the bootstrap moment conditions in a way as to ensure that the bootstrap Jacobian matrix is zero when evaluated at the GMM estimate.

**Mots clés/Key words:** Bootstrapping, overidentification, overidentification

---

<sup>\*</sup> *Concordia University, [Prosper.Dovonon@concordia.ca](mailto:Prosper.Dovonon@concordia.ca).*

<sup>†</sup> *Université de Montréal.*

# 1 Introduction

GMM estimators are commonly used in economics to estimate parameters defined by moment conditions. Under standard regularity conditions, including the rank identification condition, the GMM estimator is  $\sqrt{T}$ -consistent and asymptotically normal, as shown by Hansen (1982) in his seminal paper. Nevertheless, the bootstrap is often used to estimate the distribution of the GMM estimator and related test statistics because the first order asymptotic distribution is a poor approximation to the GMM finite sample distribution for the sample sizes typically found in practice.

The main goal of this paper is to study the applicability of the bootstrap for GMM inference when the standard rank identification condition fails but the model is still globally identified. Global identification, which requires that a unique parameter value  $\theta_0$  solves the moment conditions, ensures that the GMM estimator is consistent for  $\theta_0$ . As is well known, global identification is equivalent to the rank identification condition for linear models. The rank identification condition, which requires that the expected value of the Jacobian matrix of the moment conditions with respect to  $\theta$  is of full column rank, is an example of a local identification condition that is important to derive the first order asymptotic distribution of the GMM estimator. For nonlinear models, and as first discussed by Sargan (1983), it is possible to maintain global identification of the model and at the same time have a rank deficient Jacobian matrix. Failure of the rank identification condition (or first order underidentification) implies that the first order asymptotic distribution of the GMM estimator and related test statistics are highly nonstandard, thus motivating the use of the bootstrap as an alternative method of inference.

In this paper, we focus on bootstrapping the test of overidentification restrictions proposed by Hansen (1982) when the model is globally identified but the expected Jacobian matrix is identically zero. Our motivation for considering this special case of rank deficiency is the test for common conditionally heteroskedastic factors studied by Dovonon and Renault (2013) (henceforth D&R (2013)), for which the expected Jacobian matrix is nil when evaluated at the true parameter value. In this case, local identification is ensured by imposing identification conditions on higher order derivatives of the moment conditions. Under these conditions, D&R (2013) show that this popular test (which amounts to a test of appropriately defined overidentifying restrictions) is no longer asymptotically distributed as a  $\chi^2_{H-p}$  random variable (where  $H$  denotes the number of moment conditions and  $p$  the number of parameter coefficients in  $\theta$ ). Instead, its asymptotic distribution is a fifty-fifty mixture of  $\chi^2_{H-1}$  and  $\chi^2_H$  when  $p = 1$ , leading to an oversized test under the null of no common conditionally heteroskedastic features when the standard  $\chi^2_{H-1}$  distribution is used, independently of the sample size. When  $p > 1$ , the correct asymptotic distribution is the minimum of a certain stochastic process whose distribution is highly nonstandard, implying that critical values are difficult to obtain. Our main goal here is to propose a bootstrap method that is able to estimate the correct distribution of the overidentification test in this context. The bootstrap will be especially useful when  $p > 1$  because it avoids the use

of conservative critical values, which were proposed by D&R (2013) based on their proof that the asymptotic distribution of the test lies between  $\chi_{H-p}^2$  and  $\chi_H^2$  when  $p > 1$ .

Our contributions are as follows. First, we show that the standard bootstrap overidentification test statistics (as proposed by Hall and Horowitz (1996)), where one resamples the moment conditions recentered around the bootstrap population mean evaluated at the GMM estimate  $\hat{\theta}_T$ , is not valid when the Jacobian matrix is degenerate. The reason for this failure is that the bootstrap Jacobian matrix is the sample Jacobian matrix evaluated at  $\hat{\theta}_T$  and this is typically not zero. Thus the standard GMM bootstrap does not mimic the degeneracy of the expected Jacobian matrix at  $\theta_0$ . To remedy this problem, we propose two alternative bootstrap methods where the bootstrap GMM estimator is defined as the minimum of a quadratic form of a set of *modified* recentered bootstrap moment conditions. The first modification that we propose consists on recentering the moment condition that the bootstrap resamples twice: first we subtract off the bootstrap mean of the bootstrap moment conditions evaluated at  $\hat{\theta}_T$ , as proposed by Hall and Horowitz (1996). Second, we subtract off a term that is equal to the sample Jacobian matrix evaluated at  $\hat{\theta}_T$  multiplied by the difference between  $\theta$  and  $\hat{\theta}_T$ . We label this the corrected bootstrap. The second modification differs from the first one by letting the Jacobian matrix be evaluated at the unknown parameter. Because of its continuous correction nature, we label this as the continuously corrected bootstrap. In both alternatives, the modified bootstrap moment conditions are equal to zero when evaluated at  $\hat{\theta}_T$  (as in Hall and Horowitz (1996)), but in addition are such that the first order derivative with respect to  $\theta$  is zero when evaluated at  $\hat{\theta}_T$ . Consequently, the bootstrap expected Jacobian matrix is degenerate and this restores the asymptotic validity of the bootstrap for estimating the overidentification test when the expected Jacobian matrix is zero under the true model.

The rest of this paper is organized as follows. In Section 2, we introduce our assumptions and provide the asymptotic distribution of the overidentification test when the expected Jacobian matrix is zero. These results generalize some of the results in D&R (2013) by allowing for general nonlinear moment conditions that are not necessarily quadratic in  $\theta$ . Section 3 establishes the invalidity of the standard bootstrap method for the test of overidentification conditions when first order underidentification holds. Section 4 introduces the new bootstrap methods based on the doubly recentered bootstrap moment conditions and proves their asymptotic validity. Section 5 contains Monte Carlo simulation results and Section 6 concludes. Two mathematical appendices contain the proofs.

## 2 Setup, assumptions and asymptotic theory

We consider a sample  $\{X_t : t = 1, \dots, T\}$  of random variables described by the moment conditions

$$E(\psi(X_t, \theta)) = 0, \tag{1}$$

where  $\psi(\cdot, \cdot) \in \mathbb{R}^H$  is a known function and  $\theta \in \Theta \subset \mathbb{R}^p$  is the parameter of interest. In the following, we will often write  $\psi_t(\theta) \equiv \psi(X_t, \theta)$  whenever convenient. Letting

$$Q_T(\theta) \equiv \bar{\psi}'_T(\theta) W_T \bar{\psi}_T(\theta),$$

where  $\bar{\psi}_T(\theta) = \frac{1}{T} \sum_{t=1}^T \psi(X_t, \theta)$  and  $W_T$  a symmetric positive definite random matrix, we can define the GMM estimator  $\hat{\theta}_T$  as

$$\hat{\theta}_T = \arg \min_{\theta \in \Theta} Q_T(\theta).$$

The statistic of interest is

$$\hat{J}_T = T \bar{\psi}'_T(\hat{\theta}_T) W_T \bar{\psi}_T(\hat{\theta}_T).$$

When  $W_T$  is such that  $W_T \xrightarrow{P} W = \Sigma^{-1}$ , where  $\Sigma = \text{Var}(\psi(X_1, \theta_0))$ , we obtain the standard Hansen's (1982) GMM overidentification test statistic. Our aim in this section is to derive the asymptotic distribution of  $\hat{J}_T$  whenever the Jacobian matrix of the moment conditions is identically zero. An important example that fits into this framework is the test for common conditionally heteroskedastic features recently analyzed by D&R (2013). Our results in this section extend the results of D&R (2013) by considering more general moment conditions (which in particular are not necessarily quadratic in  $\theta$ ).

## 2.1 Assumptions

Our assumptions are as follows.

**Assumption 1 (DGP)** Let  $(\Omega, \mathcal{F}, P)$  denote a complete probability space. We observe an i.i.d. sample given by  $\{X_t : \Omega \rightarrow \mathbb{R}^l, l \in \mathbb{N}, t = 1, \dots, T\}$ .

**Assumption 2 (global identification)**  $\theta_0$  is an interior point of  $\Theta$ , a compact subset of  $\mathbb{R}^p$ ,  $p \in \mathbb{N}$ , and it is the unique solution in  $\Theta$  to the equation  $E(\psi(X_t, \theta)) = 0$ .

**Assumption 3 (regularity conditions on  $\psi$ )**

- (i)  $\{\psi(x, \theta)\}$  is continuous on  $\Theta$  for all  $x$  in the support of  $X_1$ .
- (ii)  $E(\sup_{\theta \in \Theta} \|\psi(X_1, \theta)\|) < \infty$ .
- (iii)  $E(\|\psi(X_1, \theta_0)\|^2) < \infty$  and  $\Sigma \equiv \text{Var}(\psi(X_1, \theta_0))$  is positive definite.

**Assumption 4 (regularity conditions on derivatives on  $\psi$ )**

- (i)  $\{\psi(x, \theta)\}$  is twice continuously differentiable with respect to  $\theta$  in a neighborhood  $\mathcal{N}$  of  $\theta_0$  for all  $x$  in the support of  $X_1$  and there exists a function  $m(x)$  such that  $E(m(X_1)) < \infty$  and for any  $\theta_1, \theta_2 \in \mathcal{N}$ ,  $\left\| \frac{\partial^2 \psi_h}{\partial \theta \partial \theta'}(x, \theta_1) - \frac{\partial^2 \psi_h}{\partial \theta \partial \theta'}(x, \theta_2) \right\| \leq m(x) \|\theta_1 - \theta_2\|$ , for  $h = 1, \dots, H$ , for all  $x$  in the support of  $X_1$ .

(ii)  $E \left( \left\| \frac{\partial}{\partial \theta'} \psi(X_1, \theta_0) \right\|^2 \right) < \infty$  and  $E \left( \left\| \frac{\partial^2}{\partial \theta \partial \theta'} \psi_h(X_1, \theta_0) \right\| \right) < \infty$ , for  $h = 1, \dots, H$ .

**Assumption 5 (weighting matrix)**  $W_T \xrightarrow{P} W$ , a symmetric positive definite matrix.

**Assumption 6 (local identification)**

(i)  $E \left( \frac{\partial}{\partial \theta'} \psi(X_1, \theta_0) \right) = \frac{\partial}{\partial \theta'} \rho(\theta_0) = 0$ , where  $\rho(\theta) \equiv E(\psi(X_1, \theta))$ .

(ii)  $\forall \theta \in \Theta$ ,  $\left[ (\theta - \theta_0)' E \left( \frac{\partial^2}{\partial \theta \partial \theta'} \psi_h(X_1, \theta_0) \right) (\theta - \theta_0) \right]_{1 \leq h \leq H} = 0 \iff \theta = \theta_0$ .

Assumptions 1, 2, 3 and 5 are standard in the GMM literature but Assumptions 4 and 6 are not. Assumption 4(i) is useful to control the remainder of second-order Taylor expansions of the estimating function while the first part of Assumption 4(ii) ensures the applicability of some central limit theorem to the sample mean of the Jacobian matrix of the estimating function evaluated at the true value  $\theta_0$ . The second part of Assumption 4(ii) along with Assumption 4(i) ensures that the second order derivatives of the estimating function is uniformly dominated in a neighborhood of  $\theta_0$  allowing for the application of the uniform law of large numbers. Part (i) of Assumption 6 states that the expected value of the Jacobian matrix is zero. This is a violation of the standard rank identification condition, which requires the rank of  $\frac{\partial}{\partial \theta'} \rho(\theta_0)$  to be equal to  $p$ . Given the lack of first order identification, part (ii) of Assumption 6 ensures local identification of  $\theta_0$  through second-order derivatives. In particular, letting

$$G \equiv \left[ \text{vec} \left( \frac{\partial^2 \rho_1(\theta_0)}{\partial \theta \partial \theta'} \right) \quad \cdots \quad \text{vec} \left( \frac{\partial^2 \rho_H(\theta_0)}{\partial \theta \partial \theta'} \right) \right]'$$

denote an  $H \times p^2$  matrix, Assumption 6(ii) is equivalent to

$$G \text{vec}((\theta - \theta_0)(\theta - \theta_0)') \neq 0 \text{ for all } \theta \neq \theta_0. \quad (2)$$

When  $p = 1$ , (2) is equivalent to  $G \neq 0$ , i.e.  $\frac{\partial^2 \rho_h(\theta_0)}{\partial \theta^2} \neq 0$  for at least one  $h = 1, \dots, H$ . In the general case, the condition that  $G \neq 0$  is obviously necessary for (2) but it is not sufficient. It is hard to provide more primitive conditions in this case as this amounts to giving conditions for a unique solution of a set of  $H$  nonlinear equations in  $\theta$ . An important example that satisfies (2) are the moment conditions underlying the test for common conditionally heteroskedastic features. See D&R's (2013) Lemma 2.3.

Next, we derive the asymptotic distribution of  $\hat{J}_T$  under Assumptions 1-6.

## 2.2 Asymptotic results

**Proposition 2.1** *Under Assumptions 1-3 and 5,  $\hat{\theta}_T - \theta_0 = o_P(1)$ .*

Proposition 2.1 shows that the GMM estimator is consistent for  $\theta_0$  despite the lack of first order identification. This is an immediate consequence of global identification and the uniform convergence of  $Q_T(\theta)$  towards

$$Q(\theta) = E(\psi(X_t, \theta))' W E(\psi(X_t, \theta)) \equiv \rho(\theta)' W \rho(\theta),$$

where  $\rho(\theta) \equiv E(\psi(X_t, \theta))$ .

Our next result derives the rate of convergence of  $\hat{\theta}_T - \theta_0$  when local identification is achieved through second-order local conditions.

**Proposition 2.2** *Under Assumptions 1-6, (i)  $\|\hat{\theta}_T - \theta_0\| = O_P(T^{-1/4})$  and (ii)  $T^{1/4}(\hat{\theta}_T - \theta_0)$  has at least a subsequence that converges in distribution to some random variable  $V$  such that  $P(V \neq 0) > 0$ .*

Part (i) of Proposition 2.2 shows that the convergence rate of  $\hat{\theta}_T - \theta_0$  is at least  $T^{-1/4}$  whereas part (ii) shows that this rate is sharp, i.e. there is a subsequence of  $T^{1/4}(\hat{\theta}_T - \theta_0)$  which converges in distribution to a nondegenerate random variable  $V$ . This implies that  $T^{1/4}(\hat{\theta}_T - \theta_0)$  cannot be  $o_P(1)$  (if it were then any subsequence would converge in distribution to 0).

To understand why the lack of first order identification implies a slower than  $\sqrt{T}$  rate of convergence, note that we can write a second-order Taylor expansion of the moment conditions  $\sqrt{T}\bar{\psi}_T(\hat{\theta}_T)$  as

$$\sqrt{T}\bar{\psi}_T(\hat{\theta}_T) = \sqrt{T}\bar{\psi}_T(\theta_0) + \underbrace{\frac{\partial\bar{\psi}_T(\theta_0)}{\partial\theta'}}_{=o_P(1) \text{ under Assumption 6(i)}}\sqrt{T}(\hat{\theta}_T - \theta_0) + \frac{1}{2}R_T(\ddot{\theta}_T), \quad (3)$$

where the remainder is such that

$$R_T(\ddot{\theta}_T) = \underbrace{\bar{G}(\ddot{\theta}_T)}_{\substack{\downarrow P \\ G \\ \text{under A6(ii)}}} \text{vec}\left(T^{1/4}(\hat{\theta}_T - \theta_0)T^{1/4}(\hat{\theta}_T - \theta_0)'\right),$$

where for any  $\theta$ ,

$$\bar{G}(\theta) \equiv \left[ \text{vec}\left(\frac{\partial^2\bar{\psi}_{1,T}(\theta)}{\partial\theta\partial\theta'}\right) \quad \cdots \quad \text{vec}\left(\frac{\partial^2\bar{\psi}_{H,T}(\theta)}{\partial\theta\partial\theta'}\right) \right]' \quad (4)$$

is a random  $H \times p^2$  matrix. Because the Jacobian matrix is nil,  $\sqrt{T}\frac{\partial\bar{\psi}_T(\theta_0)}{\partial\theta'} = o_P(1)$  (by a CLT) and since  $\hat{\theta}_T - \theta_0 = o_P(1)$ , it follows that the second term in (3) is  $o_P(1)$ . Thus, identification is achieved at the second-order. In particular, Assumption 6(ii) guarantees that  $R_T(\ddot{\theta}_T)$  is an  $O_P(1)$  term that is *not*  $o_P(1)$ , implying that  $T^{1/4}(\hat{\theta}_T - \theta_0) = O_P(1)$ .

To describe the asymptotic distribution of  $\hat{J}_T$  we need to introduce some additional notation. Following D&R (2013), let

$$Z_T(\theta_0) = \left( \frac{\partial^2\rho'(\theta_0)}{\partial\theta_i\partial\theta_j} W \sqrt{T}\bar{\psi}_T(\theta_0) \right)_{1 \leq i, j \leq p}$$

denote a symmetric random  $p \times p$  matrix whose limit in distribution is given by the following random Gaussian matrix

$$Z(X) = \left( \frac{\partial^2\rho'(\theta_0)}{\partial\theta_i\partial\theta_j} W X \right)_{1 \leq i, j \leq p},$$



with  $X \sim N(0, \Sigma)$ . Similarly, define the  $\mathbb{R}^p$ -indexed stochastic process

$$J(v) \equiv X'WX + X'WGvec(vv') + \frac{1}{4}vec'(vv')G'WGvec(vv'),$$

where  $v \in \mathbb{R}^p$  and note that the sample paths of  $J$  are continuous functions of  $v$ .

**Theorem 2.1** *If Assumptions 1-6 hold, then as  $T \rightarrow \infty$ ,*

$$\hat{J}_T \xrightarrow{d} \underline{J} \equiv \min_{v \in \mathbb{R}^p} J(v).$$

and  $x \rightarrow P(\underline{J} \leq x)$  is continuous at  $x$ , for all  $x \in \mathbb{R}$ .

The first part of Theorem 2.1 extends Theorem 3.1 of D&R (2013) to the case of general moment conditions satisfying Assumptions 1-6. In particular, we do not restrict ourselves to quadratic functions of  $\theta$ . The second part of this theorem is new. It shows that the asymptotic distribution of  $J_T$  is continuous. This result is essential to prove that the bootstrap methods that we introduce in this paper (see Sections 4) are uniformly consistent. Regarding the first part of Theorem 2.1, the main insights of D&R's (2013) proof remain valid here. Specifically, we consider the following stochastic process indexed by  $v \in \mathbb{R}^p$ ,

$$J_T(v) = T\bar{\psi}_T \left( \theta_0 + T^{-1/4}v \right)' W_T \bar{\psi}_T \left( \theta_0 + T^{-1/4}v \right),$$

where  $v$  is implicitly defined as  $v = T^{1/4}(\theta - \theta_0)$ , with  $\theta \in \Theta$ . Let  $\ell^\infty(\mathbb{K})$  denote the space of bounded real-valued functions on a compact subset  $\mathbb{K} \subset \mathbb{R}^p$ , equipped with the supremum norm  $\sup_{v \in \mathbb{K}} |z(v)|$ .<sup>1</sup> We show in the Appendix that  $J_T \Rightarrow J$  in  $\ell^\infty(\mathbb{K})$  for every compact  $\mathbb{K} \subset \mathbb{R}^p$ , which is equivalent to  $Eh(J_T) \rightarrow E(h(J))$ , for any  $h : \ell^\infty(\mathbb{K}) \rightarrow \mathbb{R}$  bounded and continuous with respect to the sup norm. Let  $\hat{v}_T = T^{1/4}(\hat{\theta}_T - \theta_0)$ . Since  $\hat{J}_T = J_T(\hat{v}_T) = \min_{v \in \mathbb{H}_T} J_T(v)$ , where  $\mathbb{H}_T \equiv \{v \in \mathbb{R}^p : v = T^{1/4}(\theta - \theta_0), \theta \in \Theta\}$  is such that  $\cup_{T \geq 0} \mathbb{H}_T = \mathbb{R}^p$ , uniform tightness of  $\hat{v}_T \in \arg \min_{v \in \mathbb{H}_T} J_T(v)$  and of  $\hat{v} \in \arg \min_{v \in \mathbb{R}^p} J(v)$  suffice to show that the minimum of  $J_T(v)$  converges in distribution to the minimum of  $J(v)$  (see Lemma B.5 of D&R (2013)).

Theorem 2.1 provides the asymptotic distribution for the GMM overidentification test statistic  $\hat{J}_T$  under second-order identification (the standard case where  $W_T \rightarrow^P W = \Sigma^{-1}$  is included as a special case). As it makes clear, and as was already discussed by D&R (2013) in the context of the test for common conditionally heteroskedastic features, the lack of first order identification implies that this distribution is no longer the standard chi-squared distribution with  $H - p$  degrees of freedom. Instead, the correct distribution of  $\hat{J}_T$  is the distribution of  $\underline{J}$  which, as we show in this paper, is continuous. Except for the special case when  $p = 1$ , for which D&R (2013) showed that this distribution is a fifty-fifty mixture of  $\chi_H^2$  and  $\chi_{H-1}^2$ , critical values of  $\underline{J}$  are not available. This is the main motivation for proposing the bootstrap.

---

<sup>1</sup>Note that for given  $\omega \in \Omega$ , the sample paths of  $v \mapsto J(\omega, v)$  and  $v \mapsto J_T(\omega, v)$ ,  $v \in \mathbb{K}$  are elements of  $\ell^\infty(\mathbb{K})$ .

### 3 Asymptotic invalidity of the standard bootstrap

Given Assumption 1, we let  $\{X_t^* : t = 1, \dots, T\}$  denote a (conditionally) i.i.d. bootstrap sample obtained by resampling with replacement the original sample  $\mathcal{X}_T$ .<sup>2</sup> A standard application of the bootstrap for GMM estimators involves recentering the bootstrap moment conditions by subtracting the term<sup>3</sup>  $E^* \left( \psi \left( X_t^*, \hat{\theta}_T \right) \right)$  from  $\frac{1}{T} \sum_{t=1}^T \psi_t \left( X_t^*, \theta \right)$  when defining the bootstrap criterion function, i.e.

$$Q_T^* (\theta) = \bar{\psi}_{c,T}^* (\theta)' W_T^* \bar{\psi}_{c,T}^* (\theta),$$

where  $W_T^*$  is a symmetric positive definite weighting matrix that may depend on the bootstrap sample and

$$\begin{aligned} \bar{\psi}_{c,T}^* (\theta) &= T^{-1} \sum_{t=1}^T \psi_c \left( X_t^*, \theta \right), \text{ with} \\ \psi_c \left( X_t^*, \theta \right) &= \psi \left( X_t^*, \theta \right) - E^* \left( \psi \left( X_t^*, \hat{\theta}_T \right) \right). \end{aligned}$$

Letting  $\psi_t^* (\theta) \equiv \psi \left( X_t^*, \theta \right)$  and  $\psi_{c,t}^* (\theta) \equiv \psi_c \left( X_t^*, \theta \right)$  for  $t = 1, \dots, T$  and  $\theta \in \Theta$ , it follows that

$$\psi_{c,t}^* (\theta) = \psi_t^* (\theta) - E^* \left( \psi_t^* \left( \hat{\theta}_T \right) \right).$$

Recentering ensures that the bootstrap moment conditions are equal to zero when evaluated at the “true parameter”  $\hat{\theta}_T$ , i.e. that we have  $E^* \left( \bar{\psi}_{c,T}^* \left( \hat{\theta}_T \right) \right) = 0$ . Instead, by the properties of the i.i.d. bootstrap, without recentering, we have that

$$E^* \left( \frac{1}{T} \sum_{t=1}^T \psi_t^* \left( \hat{\theta}_T \right) \right) = \frac{1}{T} \sum_{t=1}^T \psi_t \left( \hat{\theta}_T \right) = \frac{1}{T} \sum_{t=1}^T \psi \left( X_t, \hat{\theta}_T \right),$$

which is not necessarily zero when the model is overidentified.

As shown by Hahn (1996), recentering of the moment conditions in the bootstrap world is not necessary for the consistency of the bootstrap distribution of the bootstrap GMM estimator. Nevertheless, it is important to obtain asymptotic refinements for bootstrap tests and intervals based on Wald tests, as first shown by Hall and Horowitz (1996) and further studied by Andrews (2002) and Inoue and Shintani (2006), among others. For bootstrapping the distribution of the overidentification test, recentering is crucial even for first-order asymptotic validity of the bootstrap. This was discussed by Brown and Newey (2002), who proposed using a weighted bootstrap scheme, where the bootstrap probabilities are the implied empirical likelihood probabilities instead of the empirical probabilities given by  $1/T$ .

The goal of this section is to study the asymptotic properties of the standard GMM bootstrap

<sup>2</sup>Under weak dependence of  $\{X_t : t = 1, \dots, T\}$ , a block bootstrap would be appropriate. We do not consider this possibility here because our focus is on the impact of the first order local underidentification on bootstrap validity rather than on the impact of weak dependence.

<sup>3</sup>Here and throughout, we let  $E^*$ ,  $Var^*$  and  $P^*$  denote the bootstrap expectation, variance and probability measure induced by the resampling, conditional on the original sample. Appendix B gives more details on these bootstrap definitions.

when local identification is achieved at the second-order as described by Assumption 5. The bootstrap GMM estimator and the corresponding overidentification test statistic are defined as

$$\hat{\theta}_T^* = \arg \min_{\theta \in \Theta} Q_T^*(\theta) \quad \text{and} \quad \hat{J}_T^* = T \bar{\psi}_{c,T}^* \left( \hat{\theta}_T^* \right) W_T^* \bar{\psi}_{c,T}^* \left( \hat{\theta}_T^* \right).$$

We require  $W_T^*$  to converge to  $W$  under the bootstrap measure  $P^*$  with probability  $P$  approaching one, i.e.  $W_T^* \xrightarrow{P^*} W$  in prob- $P$  (see Appendix B for the formal definition of this mode of convergence). Remark 1 below provides more discussion on how to choose  $W_T^*$  in practice.

Our first result shows that  $\hat{\theta}_T^*$  is consistent for  $\theta_0$  under  $P^*$  with probability approaching one. Given Proposition 2.1, this result implies the usual result that  $\hat{\theta}_T^*$  is consistent for  $\hat{\theta}_T$  in probability.

**Proposition 3.1** *Under Assumptions 1-6 and if  $W_T^* \xrightarrow{P^*} W$  in prob- $P$ , then  $\hat{\theta}_T^* - \theta_0 = o_{P^*}(1)$  in prob- $P$ .*

The proof of this result is rather standard and requires showing that  $\sup_{\theta \in \Theta} |Q_T(\theta) - Q(\theta)| = o_P(1)$  and  $\sup_{\theta \in \Theta} |Q_T^*(\theta) - Q_T(\theta)| = o_{P^*}(1)$  in prob- $P$ . This together with the fact that  $\theta_0$  is the unique solution to the moment conditions  $\rho(\theta) = 0$  (and hence the unique minimizer of  $Q(\theta)$  over  $\Theta$ ) deliver the result.

Our next result shows that the standard GMM bootstrap method does not consistently estimate the distribution of  $\hat{J}_T$ . Specifically, we show that the unconditional limiting distribution of the bootstrap overidentification statistic  $\hat{J}_T^*$  does not coincide with the asymptotic distribution of  $\hat{J}_T$ . To simplify the arguments, we consider only the case where  $p = 1$ , but the asymptotic invalidity of the standard bootstrap method extends to the general case  $p > 1$ .

Because our proof of invalidity is based on the unconditional distribution of  $\hat{J}_T^*$ , we need to introduce the joint probability measure  $\mathbb{P} = P \times P^*$  that accounts for the two sources of randomness in  $\hat{J}_T^*$ : the randomness that comes from the original data (and which is described by  $P$ ) and the randomness that comes from the resampling, conditional on the original sample (described by  $P^*$ ). See Appendix B for more details on the properties of  $\mathbb{P}$  and its relation to  $P^*$  and  $P$ .

To characterize the bootstrap distribution of  $\hat{J}_T^*$ , we introduce the following stochastic process indexed by  $v \in \mathbb{R}$ ,

$$J_T^*(v) = T \bar{\psi}_{c,T}^* \left( \hat{\theta}_T + T^{-1/4} v \right)' W_T^* \bar{\psi}_{c,T}^* \left( \hat{\theta}_T + T^{-1/4} v \right),$$

where  $v$  is implicitly defined as  $v = T^{1/4} (\theta - \hat{\theta}_T)$ , with  $\theta \in \Theta$ . Note that

$$\hat{J}_T^* = J_T^*(\hat{v}_T^*) = \min_{v \in \mathbb{V}_T} J_T^*(v),$$

where  $\mathbb{V}_T = \{v \in \mathbb{R} : v = T^{1/4} (\theta - \hat{\theta}_T), \theta \in \Theta\}$ .

**Theorem 3.1** *Suppose that the assumptions of Proposition 3.1 hold with  $p = 1$ . It follows that:*

(i)  $\left\| \hat{\theta}_T^* - \theta_0 \right\| = O_{\mathbb{P}}(T^{-1/4}).$

(ii) *There exists at least one subsequence of*

$\left(\sqrt{T}\bar{\psi}_T(\theta_0), \sqrt{T}(\bar{\psi}_T^*(\theta_0) - \bar{\psi}_T(\theta_0)), T^{1/4}(\hat{\theta}_T - \theta_0), T^{1/4}(\hat{\theta}_T^* - \hat{\theta}_T)\right)$  *which converges in distribution under  $\mathbb{P}$  towards  $(X, X^*, V, U^*)$ , where  $X$  and  $X^*$  have the same distribution  $N(0, \Sigma)$ ,  $X^*$  is independent of  $(X, V)$ , and  $P(U^* \neq 0) > 0$ .*

(iii) *Along that same subsequence,  $\hat{J}_T^*$  converges in distribution under  $\mathbb{P}$  towards  $\underline{J}^* \equiv \min_{v \in \mathbb{R}} J^*(v)$ , where*

$$J^*(v) = \left(X^* - \frac{1}{2}GV^2\right)' W \left(X^* - \frac{1}{2}GV^2\right) + \left(X^* - \frac{1}{2}GV^2\right)' W G v^2 + \frac{1}{4}G' W G v^4.$$

(iv) *If, in addition,  $W = \Sigma^{-1}$ , then  $E(\underline{J}^*) = E(\underline{J}) - \frac{1}{2\pi}$ .*

Part (i) of Theorem 3.1 shows that the convergence rate of  $\hat{\theta}_T^* - \theta_0$  is  $T^{-1/4}$  whereas part (ii) shows that this rate is sharp. Thus, the standard bootstrap method replicates the convergence rate of the GMM estimator despite first order underidentification. Nevertheless, and as shown by part (iii), the limit process of  $J_T(v)$  does not look like the unconditional limit of the bootstrap process  $J_T^*(v)$  since the latter depends on  $V^2$ , the limit distribution of  $\sqrt{T}(\hat{\theta}_T - \theta_0)^2$ . This is a strong indication that the minima ( $\underline{J}$ , and  $\underline{J}^*$ ) of these limit processes may not have the same distribution; highlighting the invalidity of this standard bootstrap method. This is precisely the point made by part (iv) which, for  $W = \Sigma^{-1}$ , shows that the expected values of these minima differ by  $1/2\pi$ , which in particular implies that the standard bootstrap distribution is asymptotically biased.

The main reason for the failure of the standard bootstrap when applied to the overidentification test is that while the sample mean of the Jacobian matrix of the estimating function evaluated at the population value  $\theta_0$ ,  $\partial\bar{\psi}(\theta_0)/\partial\theta'$ , is of order  $O_P(T^{-1/2})$ , its bootstrap analogue is of order  $O_{\mathbb{P}}(T^{-1/4})$ . This rate is not fast enough to make the terms depending on this Jacobian matrix vanish from the expansion of the bootstrap test statistic, creating a discrepancy between the limiting distributions of the original statistic and its bootstrap analogue.

## 4 Modified bootstrap moment conditions

In this section, we propose two alternative modifications to the standard GMM bootstrap. Both alternatives involve a double recentering of the bootstrap moment conditions. This double recentering ensures that not only the bootstrap expected value of the bootstrap moment conditions at  $\hat{\theta}_T$  is zero, but that the bootstrap expected Jacobian matrix at  $\hat{\theta}_T$  is also zero.

### 4.1 The corrected GMM bootstrap

The first method considers the following bootstrap moment conditions,

$$\psi_t^{*(1)}(\theta) = \psi_t^*(\theta) - E^*\left(\psi_t^*(\hat{\theta}_T)\right) - E^*\left(\frac{\partial}{\partial\theta'}\psi_t^*(\hat{\theta}_T)\right)(\theta - \hat{\theta}_T),$$

where  $\psi_t^{*(1)}(\theta) \equiv \psi^{(1)}(X_t^*, \theta)$  for any  $t = 1, \dots, T$  and  $\theta \in \Theta$ . Thus, we recenter the original bootstrap moment function  $\psi_t^*(\theta) = \psi(X_t^*, \theta)$  twice: first by subtracting off its bootstrap expected value evaluated at  $\hat{\theta}_T$  (as in the standard GMM bootstrap), and second by subtracting off the product of the expected bootstrap Jacobian matrix evaluated at  $\hat{\theta}_T$  with the factor  $\theta - \hat{\theta}_T$ . We call this method the ‘‘corrected’’ GMM bootstrap. Similarly to the standard GMM bootstrap, we have that

$$E^* \left( \psi_t^{*(1)} \left( \hat{\theta}_T \right) \right) = 0,$$

which ensures  $E^* \left( \bar{\psi}_T^{*(1)}(\theta) \right) = 0$  when  $\theta = \hat{\theta}_T$ , where  $\bar{\psi}_T^{*(1)}(\theta) \equiv T^{-1} \sum_{t=1}^T \psi_t^{*(1)}(\theta)$ . Nevertheless, and contrary to the standard GMM bootstrap, the second recentering ensures that the expected value of the bootstrap Jacobian matrix

$$E^* \left( \frac{\partial}{\partial \theta'} \psi_t^{*(1)}(\theta) \right) = E^* \left( \frac{\partial}{\partial \theta'} \psi_t^*(\theta) \right) - E^* \left( \frac{\partial}{\partial \theta'} \psi_t^*(\hat{\theta}_T) \right)$$

is zero when  $\theta = \hat{\theta}_T$ . Thus, the corrected GMM bootstrap is able to mimic the lack of first-order identification at  $\theta_0$  that affects the original moment conditions.

Let

$$Q_T^{*(1)}(\theta) = \bar{\psi}_T^{*(1)'}(\theta) W_T^{*(1)} \bar{\psi}_T^{*(1)}(\theta),$$

where  $W_T^{*(1)}$  is a symmetric positive definite weighting matrix that may depend on the bootstrap sample. The modified bootstrap GMM estimator  $\hat{\theta}_T^{*(1)}$  is defined as the minimum of  $Q_T^{*(1)}(\theta)$  over  $\theta \in \Theta$ . The corresponding overidentification test statistic is given by

$$\hat{J}_T^{*(1)} = T \bar{\psi}_T^{*(1)'} \left( \hat{\theta}_T^{*(1)} \right) W_T^{*(1)} \bar{\psi}_T^{*(1)} \left( \hat{\theta}_T^{*(1)} \right).$$

We note that since  $E^* \left( \frac{\partial}{\partial \theta'} \psi_t^{*(1)} \left( \hat{\theta}_T \right) \right) = 0$ , the second recentering term ensures that  $\sqrt{T} \partial \bar{\psi}_T^{*(1)} \left( \hat{\theta}_T \right) / \partial \theta' = O_{P^*}(1)$  in prob- $P$ , thus mimicking the fact that  $\sqrt{T} \partial \bar{\psi}_T(\theta_0) / \partial \theta' = O_P(1)$  under first order underidentification.

We first show that  $\hat{\theta}_T^{*(1)}$  is consistent for  $\theta_0$  under the bootstrap probability measure  $P^*$  with probability approaching one. We impose the additional moment condition, which is a strengthening of Assumption 4 (ii).

**Assumption 7**  $E \left( \sup_{\theta \in \Theta} \left\| \frac{\partial \psi(X_t, \theta)}{\partial \theta'} \right\| \right) < \infty$ .

**Proposition 4.1** *Under Assumptions 1-7, if  $W_T^{*(1)} \rightarrow^{P^*} W$  in prob- $P$ , then  $\hat{\theta}_T^{*(1)} = \theta_0 + o_{P^*}(1)$  in prob- $P$ .*

Given Proposition 2.1, Proposition 4.1 implies that  $\hat{\theta}_T^{*(1)} = \hat{\theta}_T + o_{P^*}(1)$  in prob- $P$ . The next result shows that the convergence rate of the bootstrap GMM estimator  $\hat{\theta}_T^{*(1)}$  is  $T^{-1/4}$  and that this rate is sharp.

**Proposition 4.2** *Under the same assumptions as in Proposition 4.1, (i)  $\|\hat{\theta}_T^{*(1)} - \hat{\theta}_T\| = O_{P^*}(T^{-1/4})$  in prob- $P$ , and (ii)  $T^{1/4}(\hat{\theta}_T^{*(1)} - \hat{\theta}_T)$  has at least a subsequence that converges in distribution to some random variable  $V^*$  under  $P^*$ , a.s. - $P$ , such that for some  $\delta > 0$ ,  $P(\|V^*\| \neq 0) \geq \delta$ .*

Part (ii) shows that  $\hat{v}_T^{*(1)} \equiv T^{1/4}(\hat{\theta}_T^{*(1)} - \hat{\theta}_T)$  is not  $o_{P^*}(1)$ , a.s.- $P$ , which suffices to show that the rate of convergence derived in (i) is sharp. Next, we show that the modified bootstrap statistic  $\hat{J}_T^{*(1)}$  has the same asymptotic distribution as  $\hat{J}_T$ , the original test statistic, under first-order underidentification. To characterize the bootstrap distribution we introduce the following stochastic process indexed by  $v \in \mathbb{R}^p$ ,

$$J_T^{*(1)}(v) = T\bar{\psi}_T^{*(1)}(\hat{\theta}_T + T^{-1/4}v)' W_T^{*(1)} \bar{\psi}_T^{*(1)}(\hat{\theta}_T + T^{-1/4}v),$$

where  $v$  is implicitly defined as  $v = T^{1/4}(\theta - \hat{\theta}_T)$ , with  $\theta \in \Theta$ . As before, note that

$$\hat{J}_T^{*(1)} = J_T^{*(1)}(\hat{v}_T^{*(1)}) = \min_{v \in \mathbb{V}_T} J_T^{*(1)}(v),$$

where  $\mathbb{V}_T = \{v \in \mathbb{R}^p : v = T^{1/4}(\theta - \hat{\theta}_T), \theta \in \Theta\}$ . Our first result shows that conditionally on the original sample  $\mathcal{X}_n$ ,  $J_T^{*(1)}(v)$  converges weakly to  $J(v)$  in  $\ell^\infty(\mathbb{K})$  in probability for every compact  $\mathbb{K} \subset \mathbb{R}^p$ . This is denoted as  $J_T^{*(1)} \Rightarrow^{P^*} J$  in  $\ell^\infty(\mathbb{K})$ , in prob- $P$ , and it means that  $E^*(h(J_T^{*(1)})) \rightarrow E(h(J))$  in prob- $P$  for any  $h : \ell^\infty(\mathbb{K}) \rightarrow \mathbb{R}$  bounded and continuous with respect to the sup norm.

**Lemma 4.1** *Under Assumptions 1-7, if  $W_T^{*(1)} \rightarrow^{P^*} W$  in prob- $P$ , we have that  $J_T^{*(1)}(v) \Rightarrow^{P^*} J(v)$  in  $\ell^\infty(\mathbb{K})$ , in prob- $P$ .*

Lemma 4.1 is instrumental in deriving the following result.

**Theorem 4.1** *Under Assumptions 1-7, if  $W_T^{*(1)} \rightarrow^{P^*} W$  in prob- $P$ , we have that (i)  $\hat{J}_T^{*(1)} \xrightarrow{d^*} \min_{v \in \mathbb{R}^p} J(v) \equiv \underline{J}$ , in prob- $P$ , and (ii)  $\sup_{x \in \mathbb{R}} |P^*(\hat{J}_T^{*(1)} \leq x) - P(\hat{J}_T \leq x)| \rightarrow 0$ , in prob- $P$ .*

Theorem 4.1 shows that the bootstrap distribution of the corrected bootstrap overidentification test statistic  $\hat{J}_T^{*(1)}$  is consistent for the distribution of the original statistic  $\hat{J}_T$ . This result justifies using the corrected bootstrap method to compute the critical values of  $\hat{J}_T$  when testing for overidentifying restrictions. This is particularly useful when  $p > 1$ , for which there is no closed form expression for the asymptotic distribution of  $\hat{J}_T$  and the only available alternative is the use of conservative critical values, as proposed by D&R (2013).

## 4.2 The continuously-corrected GMM bootstrap

The second modification we consider is based on the following modified moment conditions:

$$\psi_t^{*(2)}(\theta) = \psi_t^*(\theta) - E^*(\psi_t^*(\hat{\theta}_T)) - E^*\left(\frac{\partial}{\partial \theta'} \psi_t^*(\theta)\right)(\theta - \hat{\theta}_T),$$

where the expected bootstrap Jacobian matrix is evaluated at  $\theta$  instead of  $\hat{\theta}_T$ . Because this bears a resemblance with the continuous-updated GMM, in which the weighting matrix is evaluated at  $\theta$  and not at  $\hat{\theta}_T$ , we call this method the “continuously-corrected” GMM bootstrap. As our simulations show, the finite sample null rejection rates of this method are closer to desired nominal level than those of the corrected bootstrap method, which is our main motivation for studying its theoretical properties here.

The Jacobian matrix of  $\psi_{h,t}^{*(2)}(\theta)$  ( $h = 1, \dots, H$ ) is given by

$$\frac{\partial}{\partial \theta} \psi_{h,t}^{*(2)}(\theta) = \frac{\partial}{\partial \theta} \psi_{h,t}^*(\theta) - E^* \left( \frac{\partial^2}{\partial \theta \partial \theta'} \psi_{h,t}^*(\theta) \right) (\theta - \hat{\theta}_T) - E^* \left( \frac{\partial}{\partial \theta} \psi_{h,t}^*(\theta) \right),$$

implying that its bootstrap expectation at  $\theta = \hat{\theta}_T$  is also equal to zero:

$$\begin{aligned} E^* \left( \frac{\partial}{\partial \theta} \psi_{h,t}^{*(2)}(\hat{\theta}_T) \right) &= E^* \left( \frac{\partial}{\partial \theta} \psi_{h,t}^*(\hat{\theta}_T) \right) - E^* \left( \frac{\partial^2}{\partial \theta \partial \theta'} \psi_{h,t}^*(\hat{\theta}_T) \right) (\hat{\theta}_T - \hat{\theta}_T) \\ &\quad - E^* \left( \frac{\partial}{\partial \theta} \psi_{h,t}^*(\hat{\theta}_T) \right) = 0. \end{aligned}$$

The set of moment conditions that the “continuously-corrected” GMM bootstrap implements,  $\bar{\psi}_T^{*(2)}(\theta)$ , converge uniformly towards a modified set of moment conditions given by

$$E(\Phi(X_1, \theta)) = E(\psi(X_1, \theta)) - E \left( \frac{\partial \psi(X_1, \theta)}{\partial \theta'} \right) (\theta - \theta_0),$$

where we let

$$\Phi(x, \theta) \equiv \psi(x, \theta) - \frac{\partial \psi(x, \theta)}{\partial \theta'} (\theta - \theta_0).$$

To ensure that these modified moment conditions identify  $\theta_0$ , we need to impose the following assumption.

**Assumption 8**  $\theta_0$  is the unique solution to the equation  $E(\Phi(X_1, \theta)) = 0$ .

Assumption 8 imposes the restriction that  $\theta_0$ , the parameter vector that uniquely identifies the original moment conditions given by Assumption 2, also uniquely identifies the modified moment conditions  $E(\Phi(X_1, \theta)) = 0$ . One leading case where this requirement is satisfied is when the original moment conditions  $\psi(x, \theta)$  are quadratic in  $\theta$  and the expected Jacobian matrix is nil at  $\theta_0$ . Actually, in this case,  $E(\Phi(x, \theta)) = -E(\psi(x, \theta))$ , for all  $\theta$ . Note that the test for common conditionally heteroskedastic features fits into this category. It is also worthwhile to mention that  $E(\Phi(x, \theta)) = 0$  has the same local identification properties at  $\theta_0$  as  $E(\psi(x, \theta)) = 0$ . In particular, we can see that

$$E \left( \frac{\partial \Phi(x, \theta_0)}{\partial \theta'} \right) = 0, \quad \text{and that} \quad \left[ (\theta - \theta_0)' E \left( \frac{\partial^2 \Phi_h(x, \theta_0)}{\partial \theta \partial \theta'} \right) (\theta - \theta_0) \right]_{1 \leq h \leq H} = 0 \Leftrightarrow (\theta = \theta_0).$$

Under Assumptions 1-8, we can show that  $\hat{\theta}_T^{*(2)}$ , the bootstrap GMM estimator that minimizes

$$Q_T^{*(2)}(\theta) = \bar{\psi}_T^{*(2)'}(\theta) W_T^{*(2)} \bar{\psi}_T^{*(2)}(\theta),$$

over  $\theta \in \Theta$ , is consistent towards  $\theta_0$ . Here and throughout,  $W_T^{*(2)}$  denotes a symmetric bootstrap random weighting matrix that converges to  $W$  in probability  $P^*$ , in prob- $P$ .

**Proposition 4.3** *Under Assumptions 1-8, if  $W_T^{*(2)} \xrightarrow{P^*} W$  in prob- $P$ , then  $\hat{\theta}_T^{*(2)} = \theta_0 + o_{P^*}(1)$  in prob- $P$ .*

Given  $\hat{\theta}_T^{*(2)}$ , we can form the bootstrap overidentification test statistic

$$\hat{J}_T^{*(2)} = T \bar{\psi}_T^{*(2)'} \left( \hat{\theta}_T^{*(2)} \right) W_T^{*(2)} \bar{\psi}_T^{*(2)} \left( \hat{\theta}_T^{*(2)} \right).$$

To show that the bootstrap distribution of  $\hat{J}_T^{*(2)}$  is consistent for the distribution of  $\hat{J}_T$ , we need to impose the following additional regularity condition, which strengthens Assumption 4(i).

**Assumption 9**  $\{\psi(x, \theta)\}$  is three times continuously differentiable with respect to  $\theta$  in a neighborhood  $\mathcal{N}$  of  $\theta_0$  for all  $x$  in the support of  $X_1$  and  $E \left( \sup_{\theta \in \mathcal{N}} \left\| \frac{\partial^3}{\partial \theta_i \partial \theta_j \partial \theta} \psi(X_1, \theta) \right\| \right) < \infty$  for all  $i, j = 1, \dots, p$ .

**Theorem 4.2** *Under Assumptions 1-9, the conclusions of Theorem 4.1 hold with  $\hat{J}_T^{*(1)}$  replaced with  $\hat{J}_T^{*(2)}$  provided  $W_T^{*(2)} \xrightarrow{P^*} W$  in prob- $P$ .*

Theorem 4.2 is the analogue of Theorem 4.1 for the continuously-corrected GMM bootstrap. As the proof in Appendix B shows, it follows by establishing the analogues of Proposition 4.2 and Lemma 4.1 for this bootstrap method.

**Remark 1 (the choice of  $W_T$  and  $W_T^*$ )** *In practice, implementing the overidentification test requires the choice of the weighting matrices in the original sample and in the bootstrap samples. The usual practice is to choose  $W_T$  such that  $W_T \xrightarrow{P} W = \Sigma^{-1}$ , which amounts to the efficient GMM when the rank condition for identification is satisfied. Under our assumptions, a consistent estimator of  $\Sigma^{-1}$  is given by the inverse of*

$$\frac{1}{T} \sum_{t=1}^T \psi_t(\tilde{\theta}_T) \psi_t(\tilde{\theta}_T)',$$

where  $\tilde{\theta}_T$  is a first-step GMM estimator that is consistent for  $\theta_0$  (e.g. a GMM estimator based on  $W_T = I_p$ ). Similarly, one possible choice for  $W_T^*$  is the inverse of

$$\frac{1}{T} \sum_{t=1}^T \psi_t^*(\tilde{\theta}_T^*) \psi_t^*(\tilde{\theta}_T^*)',$$

where  $\tilde{\theta}_T^*$  is a first-step GMM estimator based on  $I_p$ . Our theoretical results require only that  $W_T^*$  is consistent for  $W$ , which follows for this particular choice of  $W_T^*$  when  $W = \Sigma^{-1}$  under standard assumptions (since our focus is on the consequences of the rank identification failure for the bootstrap rather than on the implementation of the two-step GMM estimator, we do not provide the particular set of assumptions that justify this choice here).



## 5 Monte Carlo simulations

In this section, we illustrate the finite sample properties of the proposed tests in the context of testing for common conditionally heteroskedastic factors. Specifically, we consider an  $n \times 1$  return vector  $Y_{t+1}$  with the following conditionally heteroskedastic factor representation,

$$Y_{t+1} = \Lambda F_{t+1} + U_{t+1},$$

where  $\Lambda$  is the matrix of factor loadings,  $F_{t+1}$  is the vector of conditionally heteroskedastic and mutually independent factors, and  $U_{t+1}$  is the vector of idiosyncratic shocks. In particular, we let  $U_{t+1} \sim$  i.i.d.  $N(0, \frac{1}{2}I)$ , where  $I$  denotes the identity matrix. The generic component  $f_{t+1}$  of  $F_{t+1}$  follows a Gaussian-GARCH model,

$$f_{t+1} = \sigma_t \varepsilon_{t+1}, \quad \sigma_t^2 = \omega + \alpha f_t^2 + \beta \sigma_{t-1}^2; \quad \omega, \alpha, \beta > 0 \text{ and } \varepsilon_t \sim \text{i.i.d. } N(0, 1).$$

The null of interest is the null of common conditionally heteroskedastic features. As Engle and Kozicki (1993) explained, if there exists a parameter vector  $\delta \neq 0$  such that  $Var(\delta' Y_{t+1} | \mathcal{F}^t)$  is constant (where  $\mathcal{F}^t$  denotes the information set available at time  $t$ ), then the  $n$  return series in  $Y_{t+1}$  share common conditionally heteroskedastic features described by the vector  $\delta$ . This restriction is a testable restriction. In particular, if  $z_t \in \mathcal{F}^t$ , we can test for a constant conditional variance of  $\delta' Y_{t+1}$  by considering the following moment conditions:

$$H_0 : E \left( (z_t - E(z_t)) \left( (\delta' Y_{t+1})^2 - c(\delta) \right) \right) = 0,$$

where  $c(\delta) \equiv E \left( (\delta' Y_{t+1})^2 \right)$  denotes the unconditional variance of  $\delta' Y_{t+1}$ . By casting the problem of testing for common features as a GMM testing problem, a natural way to test  $H_0$  is to rely on Hansen's (1982) overidentification test statistic. Nevertheless, and as recently pointed out by D&R (2013), this is precisely an example where the expected Jacobian matrix is zero under  $H_0$ . Our goal in this section is to compare the performance of the bootstrap methods studied in this paper with the asymptotic tests studied in D&R (2013).

As explained by D&R (2013), in practice, we consider the feasible moment conditions,

$$H_0 : E(\psi_{t,T}(\delta)) = 0$$

where

$$\psi_{t,T}(\delta) \equiv (z_t - \bar{z}_T) \left( (\delta' Y_{t+1})^2 - \bar{c}_T(\delta) \right),$$

where we replace  $c(\delta)$  with its sample version  $\bar{c}(\delta) \equiv T^{-1} \sum_{t=1}^T (\delta' Y_{t+1})^2$ . In order to globally identify  $\delta$ , a normalization on  $\delta$  is imposed. We follow D&R (2013) and assume that  $\sum_{i=1}^n \delta_i = 1$  and set  $\delta_n = 1 - \sum_{i=1}^{n-1} \delta_i$  in our derivations.

Our simulation design is as in D&R (2013). In particular, we consider models with  $n = 2$  and 3 asset return series. Designs 1 and 2 generate bivariate return series with 1 and 2 heteroskedastic

factors, respectively, while Designs 3, 4 and 5 generate trivariate returns with 2, 1 and 3 factors, respectively. Designs 1, 3 and 4 generate returns under the null hypothesis. In particular, Designs 1 and 3 imply the existence of  $n - 1$  common factors whereas Design 4 implies the existence of  $n - 2$  common factors. While our theory does not provide an exact asymptotic distribution for the case of  $n$  asset returns with less than  $n - 1$  factors (unless the co-feature vectors are identified through some additional restrictions), this setup is nevertheless of obvious interest in practice. This is why we include Design 4 in our simulations. Designs 2 and 5 imply the existence of no common conditionally heteroskedastic factors and hence allow us to investigate power.

The simulation designs are summarized in the following table.

	Number of assets	Number of factors	GARCH Parameters	Factor loading
Design 1	2	1	$F_1 : (\omega, \alpha, \beta) = (0.2, 0.2, 0.6)$	$(1, 0.5)'$
Design 2	2	2	$F_1 : (\omega, \alpha, \beta) = (0.2, 0.2, 0.6)$ $F_2 : (\omega, \alpha, \beta) = (0.2, 0.4, 0.4)$	$I_2$
Design 3	3	2	$F_1 : (\omega, \alpha, \beta) = (0.2, 0.2, 0.6)$ $F_2 : (\omega, \alpha, \beta) = (0.2, 0.4, 0.4)$	$(\lambda_1   \lambda_2): \lambda_1 = (1, 1, 0.5)'$ $\lambda_2 = (0, 1, 0.5)'$
Design 4	3	1	$F_1 : (\omega, \alpha, \beta) = (0.2, 0.2, 0.6)$	$(1, 1, 0.5)'$
Design 5	3	3	$F_1 : (\omega, \alpha, \beta) = (0.2, 0.2, 0.6)$ $F_2 : (\omega, \alpha, \beta) = (0.2, 0.4, 0.4)$ $F_3 : (\omega, \alpha, \beta) = (0.1, 0.1, 0.8)$	$I_3$

The GMM estimations are carried out with  $H = 2$  and  $z_t = (Y_{1,t}^2, Y_{2,t}^2)'$  for the designs with 2 assets, and with  $H = 3$  and  $z_t = (Y_{1,t}^2, Y_{2,t}^2, Y_{3,t}^2)'$  for the designs with 3 assets.

The sample sizes are  $T = 1000, 2000, 5000, 10,000, 20,000, 40,000$  and  $50,000$  throughout. Because the convergence rate of the GMM estimator is  $T^{1/4}$ , it is important to allow for larger than usual sample sizes to evaluate the performance of the methods in this context. The simulation rejection rates are based on 10,000 Monte Carlo replications with 199 bootstrap replications for each Monte Carlo replication throughout.

The bootstrap samples consist of random draws with replacement of  $T$  realizations from  $\{(Y_2, z_1), (Y_3, z_2), \dots, (Y_T, z_{T-1})\}$ . We consider the standard bootstrap and the modified bootstrap methods proposed in Section 4. The choice of the weighting matrices is done as explained in Remark 1 and the nominal level is 5% throughout.

Design 1 corresponds to the case where the number  $p$  of parameters in  $\delta$  is equal to 1. Thus, the resulting overidentification test statistic is asymptotically distributed as a fifty-fifty mixture of chi-square distributions with 1 and 2 degrees of freedom. For Design 3,  $p = 2$  and therefore we do not know the critical values of the limiting distribution of  $\hat{J}_T$  in this case. However, as suggested by D&R (2013), this test can be carried out conservatively using the quantiles from a chi-square distribution with 3 degrees of freedom. The bootstrap methodologies developed in this paper are expected to provide asymptotically correct critical values to perform this test. Design 4 corresponds to a case of common conditionally heteroskedastic factor structure but, without some identifying restrictions on  $\delta$ , no exact asymptotic distribution is available in the literature for the test statistic and our theory

does not apply either. Following D&R (2013), we carry out the simulations without such restrictions to assess the performances of the tests in this case. Designs 2 and 5 correspond to the alternative where no co-feature vector exists and the purpose of these designs is to assess the power of the tests.

The results are displayed in Figures 1 through 5. Figure 1 gives the simulated null rejection rates for the bootstrap and asymptotic tests for Design 1 for which critical values for the asymptotic test are available. Results for the standard bootstrap are labeled “BootST”, whereas “Boot-(1)” refers to the corrected bootstrap and “Boot-(2)” refers to the continuously-corrected bootstrap. “Asym” indicates results based on the asymptotic distribution based on the fifty-fifty mixture of  $\chi_1^2$  and  $\chi_2^2$ . The results confirm the failure of the standard bootstrap reflected by a systematic over-rejection of the null hypothesis of factor structure with a rejection rate at around 8% in large samples. The corrected bootstrap also over-rejects for small sample sizes, but its rejection rate declines as the sample size grows, reaching around 6% for large samples. The continuously-corrected bootstrap tracks closely the asymptotic distribution and displays a correct size for almost all the sample sizes considered. Its rejection rate is of about 5% for large values of  $T$ . From these results, the continuously-corrected bootstrap is noticeably better than the corrected bootstrap. This means that correcting the standard bootstrap using the bootstrap expected Jacobian evaluated at the bootstrap estimator yields better small sample properties (for the overidentification test) than evaluated at the original GMM estimator. This feature is observed throughout the remaining designs satisfying the null hypothesis of conditionally heteroskedastic factor structure: the simulated size of the corrected bootstrap converges towards the nominal level more slowly than the continuously-corrected bootstrap.

Figure 2 shows the results for Design 3. Note that critical values for the asymptotic distribution are not available for this design. The figure shows that the corrected and continuously-corrected bootstrap methods yield null rejection rates that tend to the nominal level of 5%. The corrected bootstrap has a rejection rate of about 5.7% for large  $T$  whereas the the continuously-corrected bootstrap rejection rate lies between 4.0 and 5.0% across all sample sizes. As for Design 1, the standard bootstrap approximation over rejects systematically, yielding a rejection rate of about 8.5% in large samples, confirming its theoretical invalidity. This figure also shows that using critical values from the  $\chi_3^2$  yields rejection rates well below the desired nominal level, which implies an unnecessary loss of power in comparison to the corrected bootstrap methods (this is particularly true when comparing the conservative bound to the continuously-corrected bootstrap). We can also see that using critical values from the standard  $\chi_1^2$  (thereby ignoring the local identification failure) leads to large over-rejections under the null.

The results for Design 4 are shown in Figure 3. Here, the null of common conditionally heteroskedastic factor structure holds, but the co-feature vector is not identifiable. Though no asymptotic result is available, we expect the overidentification test to be conservative. This is the case for the two valid bootstrap methods as well as for the standard  $\chi_1^2$  approximation. Instead, the standard bootstrap method over-rejects even in large samples, which makes it undesirable in empirical works

since it can potentially rule out present and ‘easy’ to detect common conditionally heteroskedastic factor structures.

The results for Designs 2 and 5 are displayed in Figures 4 and 5, respectively. Figure 4 shows that the standard bootstrap and the corrected bootstrap have similar power and this is slightly larger than the power associated with the use of critical values based on the fifty-fifty mixture of  $\chi_1^2$  and  $\chi_2^2$ . The continuously-corrected bootstrap has smaller power than these three tests, which is as expected given that it yields lower rejection rates under the null. In this case, where the true asymptotic distribution is known, the trade-off between power and size favors the asymptotic test based on the mixture of chi-square distributions. Nevertheless, this is no longer the case for Design 5, where the exact asymptotic null distribution is not known and the choice is between using the bootstrap or the conservative bound based on the  $\chi_3^2$  distribution. Indeed, as Figure 5 shows, the latter approximation yields lower rejection rates under the alternative compared to the continuously-corrected bootstrap, which in turn is dominated by the standard bootstrap and the corrected bootstrap in terms of power. However, if we want to control the size of the test under the null and still achieve the largest possible power, the continuously-corrected bootstrap should be the preferred method.

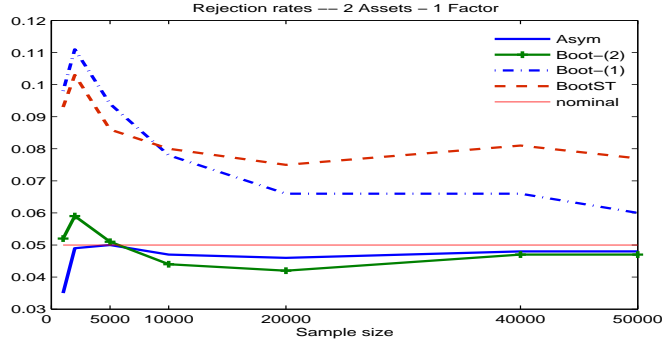


Figure 1: Rejection rates under the null for Design 1:  $\hat{J}_T \sim^a \frac{1}{2}\chi_1^2 + \frac{1}{2}\chi_2^2$ ;

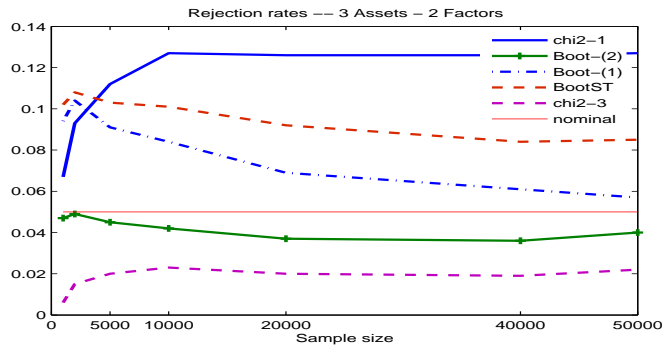


Figure 2: Rejection rates under the null for Design 3:  $\hat{J}_T \sim^a \min_v J(v)$

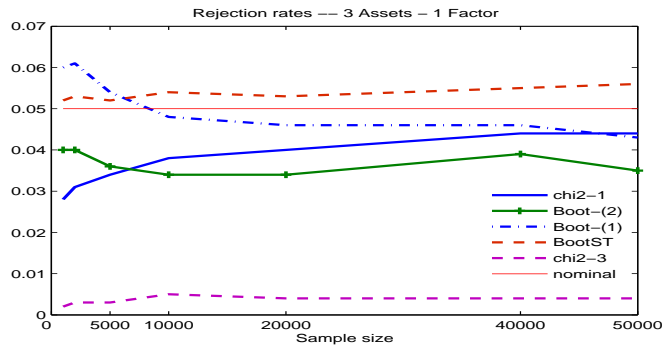


Figure 3: Rejection rates under the null for Design 4: asymptotic theory unavailable

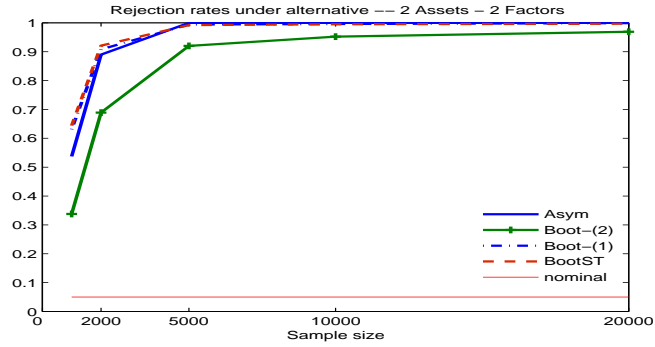


Figure 4: Rejection rates under the alternative for Design 2

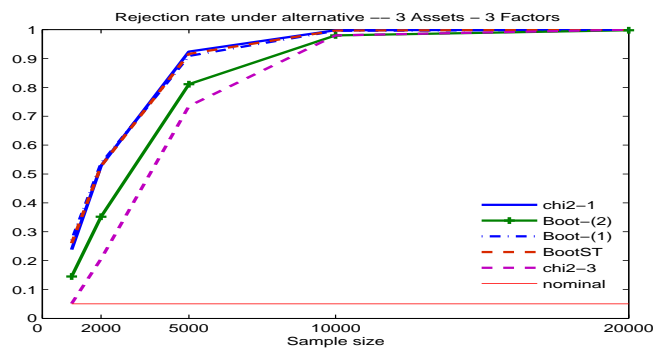


Figure 5: Rejection rates under the alternative for Design 5

## 6 Conclusions

The main contribution of this paper is to propose a new bootstrap method for GMM inference in the context of nonlinear overidentified models that are globally identified but not locally identified to first order. In particular, we focus on the special case of a degenerate rank identification condition, which was recently analyzed by Dovonon and Renault (2013) in the context of tests for common conditionally heteroskedastic factors.

We show that the standard method of bootstrapping the overidentification test statistic as proposed by Hall and Horowitz (1996) fails under this condition. The main reason for the failure of the bootstrap is that the bootstrap moment conditions do not replicate the singularity of the Jacobian matrix present in the population. We offer an easy modification of the standard bootstrap method that consists of further recentering the bootstrap moment function by subtracting off a term that is proportional to the sample Jacobian matrix multiplied by  $\theta - \hat{\theta}$ . This second recentering of the bootstrap moment condition ensures that the bootstrap Jacobian matrix is also degenerate in the bootstrap world, restoring the asymptotic validity of the bootstrap overidentification test.

## Appendix A: Proof of results in Section 2

**Proof of Proposition 2.1.** The proof that  $\hat{\theta}_T - \theta_0 = o_P(1)$  follows by Theorem 2.6 of Newey and McFadden (1994) under our Assumptions 1-3.

**Proof of Proposition 2.2.** For part (i), we follow the proof of Proposition 3.1 of D&R (2013). The only difference is the fact that the moment conditions  $\psi$  considered here are not necessarily quadratic. For each  $h = 1, \dots, H$ , by a Taylor expansion of  $\psi_{t,h}(\hat{\theta}_T) \equiv \psi_h(X_t, \hat{\theta}_T)$  around  $\theta_0$ , we have that

$$\psi_{t,h}(\hat{\theta}_T) = \psi_{t,h}(\theta_0) + \frac{\partial \psi_{t,h}(\theta_0)}{\partial \theta'} (\hat{\theta}_T - \theta_0) + \frac{1}{2} (\hat{\theta}_T - \theta_0)' \frac{\partial^2 \psi_{t,h}(\check{\theta}_T)}{\partial \theta \partial \theta'} (\hat{\theta}_T - \theta_0),$$

where  $\check{\theta}_T$  is a  $p \times 1$  vector on the segment connecting  $\hat{\theta}_T$  and  $\theta_0$  that may depend on  $h$ . The superscript  $(h)$  reflects the fact that the mean value may be different for each element  $\psi_{t,h}(\theta)$  of  $\psi_t(\theta)$ . Stacking this equation over  $h = 1, \dots, H$ , summing over  $t$  and multiplying the result by  $\sqrt{T}$  yields

$$\sqrt{T} \bar{\psi}_T(\hat{\theta}_T) = \sqrt{T} \bar{\psi}_T(\theta_0) + \sqrt{T} \frac{\partial \bar{\psi}_T(\theta_0)}{\partial \theta'} (\hat{\theta}_T - \theta_0) + \frac{1}{2} R_T(\check{\theta}_T), \quad (5)$$

where the remainder  $R_T(\check{\theta}_T)$  is a  $H \times 1$  vector defined as

$$R_T(\check{\theta}_T) = \sqrt{T} \left[ (\hat{\theta}_T - \theta_0)' \frac{\partial^2 \bar{\psi}_{h,T}(\check{\theta}_T^{(h)})}{\partial \theta \partial \theta'} (\hat{\theta}_T - \theta_0) \right]_{h=1, \dots, H}.$$

For  $h = 1, \dots, H$ , we can write

$$\left(\hat{\theta}_T - \theta_0\right)' \frac{\partial^2 \bar{\psi}_{h,T} \left(\ddot{\theta}_T^{(h)}\right)}{\partial \theta \partial \theta'} \left(\hat{\theta}_T - \theta_0\right) = \text{vec} \left( \frac{\partial^2 \bar{\psi}_{h,T} \left(\ddot{\theta}_T^{(h)}\right)}{\partial \theta \partial \theta'} \right)' \text{vec} \left( \left(\hat{\theta}_T - \theta_0\right) \left(\hat{\theta}_T - \theta_0\right)' \right),$$

since  $\text{tr}(AB) = \text{vec}(A)' \text{vec}(B)$ . It follows that

$$R_T \left(\ddot{\theta}_T\right) = \sqrt{T} \bar{G} \left(\ddot{\theta}_T\right) \text{vec} \left( \left(\hat{\theta}_T - \theta_0\right) \left(\hat{\theta}_T - \theta_0\right)' \right) \equiv \bar{G} \left(\ddot{\theta}_T\right) \text{vec} \left(\hat{v}_T \hat{v}_T'\right),$$

where  $\hat{v}_T \equiv T^{1/4} \left(\hat{\theta}_T - \theta_0\right)$  and  $\bar{G} \left(\ddot{\theta}_T\right)$  is the random  $H \times p^2$  matrix  $\bar{G}(\theta)$  defined in (4) evaluated at  $\ddot{\theta}_T$ . Given our assumptions and the fact that  $\ddot{\theta}_T \rightarrow^P \theta_0$ , we can show that  $\bar{G} \left(\ddot{\theta}_T\right) = G + o_P(1)$  (see e.g. Lemmas 2.4 and 2.9 of Newey and McFadden (1994)). Thus,

$$R_T \left(\ddot{\theta}_T^{(h)}\right) = G \text{vec} \left(\hat{v}_T \hat{v}_T'\right) + o_P \left(\|\hat{v}_T\|^2\right).$$

This implies that

$$\begin{aligned} \sqrt{T} \bar{\psi}_T \left(\hat{\theta}_T\right) &= \sqrt{T} \bar{\psi}_T \left(\theta_0\right) + \sqrt{T} \frac{\partial \bar{\psi}_T \left(\theta_0\right)}{\partial \theta'} \left(\hat{\theta}_T - \theta_0\right) + \frac{1}{2} G \text{vec} \left(\hat{v}_T \hat{v}_T'\right) + o_P \left(\|\hat{v}_T\|^2\right) \\ &= \sqrt{T} \bar{\psi}_T \left(\theta_0\right) + \frac{1}{2} G \text{vec} \left(\hat{v}_T \hat{v}_T'\right) + o_P(1) + o_P \left(\|\hat{v}_T\|^2\right), \end{aligned} \quad (6)$$

where the  $o_P(1)$  term is equal to  $\sqrt{T} \frac{\partial \bar{\psi}_T \left(\theta_0\right)}{\partial \theta'} \left(\hat{\theta}_T - \theta_0\right) = O_P(1) \times o_P(1)$  since under Assumption 6 (i),  $E \left(\frac{\partial \bar{\psi}_T \left(\theta_0\right)}{\partial \theta'}\right) = 0$  and thanks to Assumption 4(ii), therefore  $\sqrt{T} \frac{\partial \bar{\psi}_T \left(\theta_0\right)}{\partial \theta'} = O_P(1)$  by a CLT applied to  $\left\{\frac{\partial \psi(X_t, \theta_0)}{\partial \theta'}\right\}$ . It follows that

$$\begin{aligned} T \bar{\psi}_T' \left(\hat{\theta}_T\right) W_T \bar{\psi}_T \left(\hat{\theta}_T\right) &= T \bar{\psi}_T' \left(\theta_0\right) W_T \bar{\psi}_T \left(\theta_0\right) + \frac{1}{4} \text{vec}' \left(\hat{v}_T \hat{v}_T'\right) G' W_T G \text{vec} \left(\hat{v}_T \hat{v}_T'\right) \\ &\quad + \text{vec}' \left(\hat{v}_T \hat{v}_T'\right) G' W_T \sqrt{T} \bar{\psi}_T \left(\theta_0\right) + o_P(1) + o_P \left(\|\hat{v}_T\|^2\right) + o_P \left(\|\hat{v}_T\|^4\right). \end{aligned}$$

The expected result follows from the same arguments as those used by D&R (2013) to prove their Proposition 3.1. The additional  $o_P(\|\hat{v}_T\|^4)$  term that appears here does not alter the reasoning.

We now establish (ii). Since  $Z_T(\theta_0) = O_P(1)$  and  $\hat{v}_T = O_P(1)$ , we have  $(\text{vec}'(Z_T(\theta_0)), \hat{v}_T')' = O_P(1)$ . Therefore, from Prohorov's theorem (see Theorem 2.4 of van der Vaart (1998)), the joint sequence has a subsequence that converges in distribution to  $(\text{vec}'(Z(X)), V)'$ , say, where

$$Z(X) = \left( \frac{\partial^2 \rho'(\theta_0)}{\partial \theta_i \partial \theta_j} W X \right)_{1 \leq i, j \leq p},$$

and  $X \sim N(0, \Sigma)$ . Let “ $Z(X) \geq 0$ ” denote the event “ $Z(X)$  is positive semi-definite and “ $\overline{Z(X) \geq 0}$ ” its complement. Under Assumption 6(ii), D&R (2013, Proposition 3.2) show that  $P(Z(X) \geq 0) \leq 1/2$ , which implies that  $P(\overline{Z(X) \geq 0}) \geq 1/2 > 0$ . Since

$$P(V \neq 0) \geq P(V \neq 0, \overline{Z(X) \geq 0}) = P(V \neq 0 | \overline{Z(X) \geq 0}) P(\overline{Z(X) \geq 0}),$$

it suffices to show that  $P(V = 0 | \overline{Z(X) \geq 0}) = 0$ . To this end, we follow the proof of Proposition



3.2 of D&R (2013). The second order necessary condition for an interior solution for a minimization problem implies that, for any unit vector  $e \in \mathbb{R}^p$ ,

$$e' \left( \frac{\partial^2}{\partial \theta \partial \theta'} \bar{\psi}_T(\theta)' W_T \bar{\psi}_T(\theta) \Big|_{\theta=\hat{\theta}_T} \right) e \geq 0,$$

which is equivalent to  $e'(\tilde{Z}_T + N_T)e \geq 0$ , with

$$\tilde{Z}_T = \left( \frac{\partial^2}{\partial \theta_i \partial \theta_j} \bar{\psi}'_T(\hat{\theta}_T) W_T \sqrt{T} \bar{\psi}_T(\hat{\theta}_T) \right)_{1 \leq i, j \leq p}, \quad \text{and} \quad N_T = \sqrt{T} \frac{\partial}{\partial \theta} \bar{\psi}'_T(\hat{\theta}_T) W_T \frac{\partial}{\partial \theta} \bar{\psi}_T(\hat{\theta}_T).$$

The result follows exactly along the lines of the proof of D&R (2013) once we establish that

$$\tilde{Z}_T = Z_T(\theta_0) + \frac{1}{2} \left( \frac{\partial^2 \rho'(\theta_0)}{\partial \theta_i \partial \theta_j} W \text{Gvec}(\hat{v}_T \hat{v}'_T) \right)_{1 \leq i, j \leq p} + o_P(1), \quad (7)$$

and

$$N_T = \left( \hat{v}'_T \frac{\partial^2 \rho'(\theta_0)}{\partial \theta_i \partial \theta} W \frac{\partial^2 \rho(\theta_0)}{\partial \theta_j \partial \theta'} \hat{v}_T \right)_{1 \leq i, j \leq p} + o_P(1). \quad (8)$$

To show (7), we observe that from part (i), (6) implies that

$$\sqrt{T} \bar{\psi}_T(\hat{\theta}_T) = \sqrt{T} \bar{\psi}_T(\theta_0) + \frac{1}{2} \text{Gvec}(\hat{v}_T \hat{v}'_T) + o_P(1). \quad (9)$$

Given the dominance condition in Assumption 4 (i) and the fact that  $\hat{\theta}_T \xrightarrow{P} \theta_0$ , it follows that

$$\frac{\partial^2 \bar{\psi}'_T(\hat{\theta}_T)}{\partial \theta_i \partial \theta_j} W_T = \frac{\partial^2 \rho'(\theta_0)}{\partial \theta_i \partial \theta_j} W + o_P(1),$$

which together with (9) implies (7). To obtain (8), by a mean-value expansion of  $\frac{\partial \bar{\psi}_T(\hat{\theta}_T)}{\partial \theta_i}$  around  $\theta_0$ , we get that for all  $i = 1, \dots, p$ ,

$$W_T^{1/2} T^{1/4} \frac{\partial \bar{\psi}_T(\hat{\theta}_T)}{\partial \theta_i} = W_T^{1/2} T^{1/4} \frac{\partial \bar{\psi}_T(\theta_0)}{\partial \theta_i} + W_T^{1/2} \frac{\partial^2 \bar{\psi}_T(\bar{\theta})}{\partial \theta_i \partial \theta'} T^{1/4} (\hat{\theta}_T - \theta_0) = W_T^{1/2} \frac{\partial^2 \rho(\theta_0)}{\partial \theta_i \partial \theta'} \hat{v}_T + o_P(1),$$

where  $\bar{\theta} \in (\theta_0, \hat{\theta}_T)$  and may differ from row to row. We obtain (8) by writing this equation for any  $i, j = 1, \dots, p$  and taking the inner product of each hand side.

**Proof of Theorem 2.1.** We follow the proof of Lemma B.6 of D&R (2013). By a second-order mean value expansion of  $v \mapsto \sqrt{T} \bar{\psi}_T(\theta_0 + T^{-1/4}v)$  at 0, we have that

$$\sqrt{T} \bar{\psi}_T(\theta_0 + T^{-1/4}v) = \sqrt{T} \bar{\psi}_T(\theta_0) + T^{1/4} \frac{\partial \bar{\psi}_T}{\partial \theta'}(\theta_0) v + \frac{1}{2} \bar{G}(\bar{\theta}(v)) \text{vec}(vv'),$$

where  $\bar{\theta}(v) \in (\theta_0, \theta_0 + T^{-1/4}v)$  and may differ from row to row. It follows that

$$\begin{aligned} \sqrt{T} \bar{\psi}_T(\theta_0 + T^{-1/4}v) &= \sqrt{T} \bar{\psi}_T(\theta_0) + \frac{1}{2} \text{Gvec}(vv') + \frac{1}{2} (\bar{G}(\theta_0) - G) \text{vec}(vv') \\ &\quad + \frac{1}{2} (\bar{G}(\bar{\theta}(v)) - \bar{G}(\theta_0)) \text{vec}(vv') + T^{1/4} \frac{\partial \bar{\psi}_T}{\partial \theta'}(\theta_0) v. \end{aligned}$$

Next, we show that the last three terms in the equation above are  $o_p(1)$  uniformly over any compact subset  $\mathbb{K}$  of  $\mathbb{R}^p$ . Starting with the first of these terms, note that  $\bar{G}(\theta_0) - G = o_P(1)$  and since

$v \in \mathbb{K} \subset \mathbb{R}^p$ , this term converges to zero uniformly over  $\mathbb{K}$ . For the term that follows, note that since  $\theta_0$  is an interior point of  $\Theta$  and  $\mathbb{K}$  is compact (hence bounded), then  $\ddot{\theta}(v) \in \mathcal{N}$ , a neighborhood of  $\theta_0$ , for all  $v \in \mathbb{K}$  and  $T$  sufficiently large. Hence, from Assumption 3(i), using the triangle inequality, the  $h$ -th row of  $\bar{G}(\ddot{\theta}(v)) - \bar{G}(\theta_0)$ , say  $\bar{G}_h(\ddot{\theta}(v)) - \bar{G}_h(\theta_0)$  satisfies

$$\|\bar{G}_h(\ddot{\theta}(v)) - \bar{G}_h(\theta_0)\| \leq \frac{1}{T} \sum_{t=1}^T m(X_t) \|\ddot{\theta}(v) - \theta_0\| = \frac{1}{T} \sum_{t=1}^T m(X_t) (1 - \alpha) T^{-1/4} \|v\|,$$

for some  $0 \leq \alpha \leq 1$ . Thus, this term is  $o_P(1)$  uniformly over  $\mathbb{K}$ . Finally, note that  $\frac{\partial \bar{\psi}_T}{\partial \theta'}(\theta_0) = O_P(T^{-1/2})$  by a CLT since  $E\left(\frac{\partial \psi_t}{\partial \theta'}(\theta_0)\right) = 0$ . This implies that the last term is also  $o_P(1)$  uniformly over  $\mathbb{K}$ . As a result,

$$\sqrt{T} \bar{\psi}(\theta_0 + T^{-1/4}v) = \sqrt{T} \bar{\psi}_T(\theta_0) + \frac{1}{2} G \text{vec}(vv') + o_P(1),$$

implying that

$$J_T(v) = \sqrt{T} \bar{\psi}'_T(\theta_0) W \sqrt{T} \bar{\psi}_T(\theta_0) + \text{vec}'(vv') G' W \sqrt{T} \bar{\psi}_T(\theta_0) + \frac{1}{4} \text{vec}'(vv') G' W G \text{vec}(vv') + o_P(1),$$

where the neglected terms are uniformly (asymptotically) negligible over any compact subset  $\mathbb{K}$ . Since this is the analogue of equation (B.4) of D&R (2013), the rest of the proof follows exactly the same lines as the proof of Lemma B.6 of D&R (2013).

To complete the proof of Theorem 2.1, we show that the random variable  $\underline{J} \equiv \min_{v \in \mathbb{R}^p} J(v)$  has a continuous distribution. We adopt the notation  $J(X, v)$  for  $J(v)$  to highlight its dependence on  $X$  which is its only source of randomness. We show that  $\forall c \in \mathbb{R}$ ,  $\text{Prob}(\underline{J} = c) = 0$ . It is easy to see that

$$J(X, v) = \left( X + \frac{1}{2} G \text{vec}(vv') \right)' W \left( X + \frac{1}{2} G \text{vec}(vv') \right).$$

Since  $W$  is symmetric positive definite,  $J(x, v) \geq 0$ ,  $\forall x \in \mathbb{R}^H$  and  $v \in \mathbb{R}^p$ . As a result,  $\text{Prob}(\underline{J} = c) = 0$ ,  $\forall c < 0$ .

We show next that  $\underline{J}$  cannot have an atom of probability at any  $c > 0$ . Let us assume by contradiction that there exists  $c^* > 0$  such that  $\text{Prob}(\underline{J} = c^*) > 0$ . Let

$$A = \left\{ x \in \mathbb{R}^H : \min_{v \in \mathbb{R}^p} J(x, v) = c^* \right\}.$$

By definition,  $\text{Prob}(X \in A) = \text{Prob}(\underline{J} = c^*) > 0$ . As a result, since the distribution of  $X$  is absolutely continuous with respect to the Lebesgue measure on  $\mathbb{R}^H$ ,  $A$  contains an open ball centered at a certain  $x_0 \in \mathbb{R}^H$ , say  $\mathcal{B}(x_0, \epsilon) = \{x \in \mathbb{R}^H : \|x - x_0\| < \epsilon\}$  for some  $\epsilon > 0$ . This implies that  $\forall d \in \mathbb{R}^H : \|d\| < \epsilon$ ,  $\min_{v \in \mathbb{R}^p} J(x_0 + d, v) = c^*$ . Let  $v_0 \in \arg \min_{v \in \mathbb{R}^p} J(x_0, v)$  (such an argmin exists for almost every  $x_0 \in \mathbb{R}^H$  by D&R's (2013) Lemma B6(ii)). Since  $x_0 \in A$ ,  $J(x_0, v_0) = c^*$  and  $\forall d \in \mathbb{R}^H : \|d\| < \epsilon$ , we have that

$$J(x_0, v_0) = \min_{v \in \mathbb{R}^p} J(x_0 + d, v) \leq J(x_0 + d, v_0) = J(x_0, v_0) + 2d'Wx_0 + d'Wd + d'W \text{vec}(v_0v_0').$$

(The second equality above is obtained by expanding the expression of  $J(x_0 + d, v_0)$ .) That is,

$$2d' \left( W(x_0 + \frac{1}{2}G\text{vec}(v_0v_0')) + \frac{1}{2}Wd \right) \geq 0, \quad \forall d \in \mathcal{B}(0, \epsilon). \quad (10)$$

Note that since  $J(x_0, v_0) = c^* = a'W^{-1}a > 0$ , we must have  $a \equiv W(x_0 + \frac{1}{2}G\text{vec}(v_0v_0')) \neq 0$ . Fix  $a_j \neq 0$ , the  $j^{\text{th}}$  element of  $a$  that is different from zero. Equation (10) implies that if we choose  $d$  with all entries equal to zero except  $d_j$  such that  $-\epsilon < d_j < \epsilon$ , then  $d_j(2a_j + W_{jj}d_j) \geq 0$  for all such  $d_j$ . Since  $W_{jj} > 0$ ,  $d_j(2a_j + W_{jj}d_j) = 0$  has two roots, 0 and  $-2a_j/W_{jj} \neq 0$ . It follows that the sign of the polynomial  $d_j(2a_j + W_{jj}d_j)$  must change in a neighborhood of zero, implying that it is impossible to have  $d_j(2a_j + W_{jj}d_j) \geq 0$  for all  $-\epsilon < d_j < \epsilon$ . Thus, it must be that  $\text{Prob}(\underline{J} = c^*) = 0$  for any  $c^* > 0$ . To complete the proof, we need to show that  $\underline{J}$  does not have an atom of probability at 0. To this end, we show that there exists a random variable  $L$  that is such that  $0 \leq L \leq \underline{J}$  and  $\text{Prob}(L = 0) = 0$ . Let  $L = \min_{u \in \mathbb{R}^{p^2}} (X + \frac{1}{2}Gu)'W(X + \frac{1}{2}Gu)$ . Clearly,  $L \leq \underline{J}$ . Assume that  $\text{Rank}(G) = r$ . Given the second-order identification condition,  $G \neq 0$  and  $r > 0$ . Consider a rank factorization of  $G$ :  $G = G_1G_2$ , where  $G_1$  is  $H \times r$  and  $G_2$  is  $r \times p^2$ , both of rank  $r$ . The first order condition associated with  $L$  solved at  $\hat{u}$  implies that  $G_2\hat{u} = -2(G_1'WG_1)^{-1}G_1'WX$ . Consequently, we can write

$$L = \left( X + \frac{1}{2}G_1G_2\hat{u} \right)' W \left( X + \frac{1}{2}G_1G_2\hat{u} \right) = X'W^{1/2}PW^{1/2}X, \quad (11)$$

where  $P = I_H - W^{1/2}G_1(G_1'WG_1)^{-1}G_1'W^{1/2}$  is the orthogonal projection matrix on the subspace of dimension  $H - r$  that is orthogonal to the space spanned by the columns of  $W^{1/2}G_1$ . Hence, there exists an  $H \times H$  matrix  $Q$  such that  $Q'Q = I_H$  and  $P = Q'DQ$ , where  $D = \text{diag}(I_{H-r}, 0)$ . Thus,  $L = Y'DY$ , with  $Y = QW^{1/2}X \sim N(0, S)$  and  $S = QW^{1/2}\Sigma W^{1/2}Q'$  positive definite. Hence,  $L = \sum_{i=1}^{H-r} Y_i^2$ . Clearly,  $\text{Prob}(L = 0) \leq \text{Prob}(Y_1 = 0) = 0$ . This concludes the proof.

## Appendix B: Proofs of results in Section 3 and 4

### .1 Preliminaries

A convenient way to formalize the bootstrap is as follows (see Gonçalves and White (2004) for a similar framework). Given the underlying probability space  $(\Omega, \mathcal{F}, P)$ , we observe a sample of size  $T$ ,  $\mathcal{X}_T \equiv \{X_1(\omega), X_2(\omega), \dots, X_T(\omega)\}$  from a given realization  $\omega \in \Omega$ . Suppose we obtain a bootstrap sample  $\mathcal{X}_T^* = \{X_1^*, X_2^*, \dots, X_T^*\}$  by resampling from  $\mathcal{X}_T$ . For each  $\omega \in \Omega$ , we view  $\{X_t^* : t = 1, \dots, T\}$  as the realization of a stochastic process defined on  $(\Lambda, \mathcal{G}, P^*)$ , another probability space, such that for each  $t = 1, \dots, T$ ,

$$X_t^*(\omega, \lambda) = X_{\tau_t(\lambda)}(\omega), \quad (12)$$

where  $\tau_t : \Lambda \rightarrow \{1, 2, \dots, N\}$  denotes the random index generated by the resampling scheme for each  $t = 1, 2, \dots, T$  (independently of  $\omega$ ). Thus,  $P^*$  describes the probability of bootstrap random variables, conditional on the observed data  $\mathcal{X}_T$ , i.e.  $P^*$  describes the probability induced by  $\lambda$ , conditional on  $\omega$ .

We write  $E^*$  and  $Var^*$  to denote the expected value and the variance with respect to  $P^*$ , respectively.

As (12) makes clear,  $X_t^*$  depends on two sources of randomness, one related to the observed data (and indexed by  $\omega$ ) and the other related to the resampling mechanism (as measured by  $\tau_t(\lambda)$ ). When the joint randomness is of interest, we can view the bootstrap statistic as being defined on the product probability space  $(\Omega, \mathcal{F}, P) \times (\Lambda, \mathcal{G}, P^*) = (\Omega \times \Lambda, \mathcal{F} \times \mathcal{G}, \mathbb{P})$ , where  $\mathbb{P} = P \times P^*$  denotes the unconditional (or joint) probability. We write  $\mathbb{E}$  and  $\mathbb{V}ar$  to denote expected value and variance with respect to  $\mathbb{P}$ , respectively.

Given any bootstrap statistic  $Z_T^*$ , we say that  $Z_T^* \xrightarrow{P^*} 0$  in prob- $P$  (or  $Z_T^* = o_{P^*}(1)$  in prob- $P$ ) if for any  $\varepsilon, \delta > 0$ ,  $P(P^*(|Z_T^*| > \varepsilon) > \delta) \rightarrow 0$  as  $T \rightarrow \infty$ . Similarly, we say that  $Z_T^* = O_{P^*}(1)$  in prob- $P$  if for any  $\delta > 0$ , there exists  $0 < M < \infty$  such that  $P(P^*(|Z_T^*| \geq M) > \delta) \rightarrow 0$  as  $T \rightarrow \infty$ . Lemma .1 in Appendix B describes the transition between the bootstrap and the non-bootstrap stochastic orders.

The following lemma is very useful when proving bootstrap results as it describes the transition between the bootstrap and the non-bootstrap stochastic orders. The proof is found in Cheng and Huang (2010, *Annals of Statistics*, Lemma 3).

**Lemma .1** *Suppose that*

$$Z_T^* = o_{P^*}(1) \text{ in prob-}P, \text{ and } W_T^* = O_{P^*}(1) \text{ in prob-}P.$$

*Then, we have that (a1) if  $A_T$  is defined only on  $(\Omega, \mathcal{F}, P)$  and it is  $o_P(1)$  [ $O_P(1)$ ], then it is also  $o_{P^*}(1)$  in prob- $P$  [ $O_{P^*}(1)$  in prob- $P$ ]; (a2) if  $A_T$  is defined only on  $(\Omega, \mathcal{F}, P)$  and it is  $o_P(1)$  [ $O_P(1)$ ], then it is also  $o_{\mathbb{P}}(1)$  [ $O_{\mathbb{P}}(1)$ ]; (a3)  $A_T^* = o_{P^*}(1)$  in prob- $P$  [ $O_{P^*}(1)$  in prob- $P$ ]  $\iff A_T^* = o_{\mathbb{P}}(1)$  [ $O_{\mathbb{P}}(1)$ ]; (a4)  $B_T^* = Z_T^* \times W_T^* = o_{P^*}(1) \times O_{P^*}(1) = o_{P^*}(1)$  in prob- $P$ ; (a5)  $C_T^* = Z_T^* \times O_P(1) = o_{P^*}(1) \times O_P(1) = o_{P^*}(1)$  in prob- $P$ , and (a6)  $A_T^* = Z_T^* \times o_P(1) = o_{P^*}(1) \times o_P(1) = o_{P^*}(1)$  in prob- $P$ .*

For a sequence of random variables (or vectors)  $Z_T^*$ , we also need the definition of convergence in distribution in prob- $P$ . In particular, we write  $Z_T^* \xrightarrow{d^*} Z$ , in prob- $P$ , if  $E^*f(Z_T^*) \rightarrow E(f(Z))$  in prob- $P$  for every continuous and bounded function  $f$ .

Finally, we define weak convergence of a random process  $Z_T^*(v)$  in  $\ell^\infty(\mathbb{K})$  in prob- $P$ . We write  $Z_T^* \Rightarrow^{d^*} Z$  in  $\ell^\infty(\mathbb{K})$  in prob- $P$  if  $\sup_{h \in BL_1(\ell^\infty(\mathbb{K}))} |E^*(h(Z_T^*)) - E(h(Z))| \xrightarrow{P} 0$  where  $BL_1(\ell^\infty(\mathbb{K}))$  is the space of functions  $h : \ell^\infty(\mathbb{K}) \rightarrow \mathbb{R}$  with Lipschitz norm bounded by 1, i.e. for any  $h \in BL_1(\ell^\infty(\mathbb{K}))$ ,  $\sup_{z \in \ell^\infty(\mathbb{K})} |h(z)| \leq 1$  and  $|h(z_1) - h(z_2)| \leq d(z_1, z_2)$  for all  $z_1, z_2$  in  $\ell^\infty(\mathbb{K})$ , where  $d(z_1, z_2) = \sup_{v \in \mathbb{K}} |z_1(v) - z_2(v)|$ . The bounded Lipschitz distance between distribution functions metrizes weak convergence in distribution and therefore this definition is equivalent to saying that  $E^*(h(Z_T^*)) \rightarrow E(h(Z))$  in prob- $P$  for any  $h : \ell^\infty(\mathbb{K}) \rightarrow \mathbb{R}$  continuous and bounded with respect to the sup norm.

The following result is the bootstrap version of Lemma B.6 of D&R (2003) and is useful to prove Lemma 4.1.

**Lemma .2** Suppose  $J_T^*(v) \Rightarrow^{P^*} J(v)$ , in  $\ell^\infty(\mathbb{K})$ , in prob- $P$ , for any compact subset  $\mathbb{K}$  of  $\mathbb{R}^p$ . If there exist  $\hat{v}_T^* \in \arg \min_{v \in \mathbb{R}^p} J_T^*(v)$  and  $\hat{v} \in \arg \min_{v \in \mathbb{R}^p} J(v)$  such that  $\hat{v}_T^* = O_{P^*}(1)$  in prob- $P$  and  $\hat{v} = O_P(1)$ , then  $\hat{J}_T^* \equiv J_T^*(\hat{v}_T^*) = \min_{v \in \mathbb{R}^p} J_T^*(v) \xrightarrow{d^*} \underline{J} \equiv \min_{v \in \mathbb{R}^p} J(v)$ , in prob- $P$ .

**Proof of Lemma .2.** We show that  $E^*(h(J_T^*(\hat{v}_T^*))) - E(h(J(\hat{v}))) \rightarrow^P 0$  for any continuous and bounded function  $h : \mathbb{R} \rightarrow \mathbb{R}$ . This is equivalent to showing that for any  $\epsilon, \delta > 0$ ,

$$P(|E^*(h(J_T^*(\hat{v}_T^*))) - E(h(J(\hat{v})))| > \epsilon) < \delta$$

for all  $T$  sufficiently large. Let  $\epsilon, \delta > 0$  and let  $M$  denote the upper bound on the absolute value of  $h$ . Since  $\hat{v}_T^* = O_{P^*}(1)$  in prob- $P$  and  $\hat{v} = O_P(1)$ , there exists  $A > 0$  such that

$$P\left(P^*(\|\hat{v}_T^*\| > A) > \frac{\epsilon}{6M}\right) < \frac{\delta}{3}, \quad \text{and} \quad P(\|\hat{v}\| > A) < \frac{\delta}{3}. \quad (13)$$

Since for all compact subset  $\mathbb{K}$ ,  $J_T^*(v)$  converges weakly towards  $J(v)$  in  $\ell^\infty(\mathbb{K})$ , in prob- $P$ , by the continuous mapping theorem we have that  $\min_{\|v\| \leq A} J_T^*(v) \xrightarrow{d^*} \min_{\|v\| \leq A} J(v)$ , in prob- $P$ . Hence, with  $\hat{v}_1^*$  and  $\hat{v}_2$  denoting the argument of the minimum over  $\mathbb{K} \equiv \{\|v\| \leq A\}$  of  $J_T^*(v)$  and  $J(v)$ , respectively, we have that  $E^*(h(J_T^*(\hat{v}_1^*))) - E(h(J(\hat{v}_2)))$  converges in probability to 0. As a result, for all  $T$  sufficiently large,

$$P\left(|E^*(h(J_T^*(\hat{v}_1^*))) - E(h(J(\hat{v}_2)))| > \frac{\epsilon}{3}\right) < \frac{\delta}{3}. \quad (14)$$

By the triangle inequality, we can write

$$\begin{aligned} |E^*(h(J_T^*(\hat{v}_T^*))) - E(h(J(\hat{v})))| &\leq |E^*(h(J_T^*(\hat{v}_T^*))) - E^*(h(J_T^*(\hat{v}_1^*)))| \\ &\quad + |E^*(h(J_T^*(\hat{v}_1^*))) - E(h(J(\hat{v}_2)))| \\ &\quad + |E(h(J(\hat{v}_2))) - E(h(J(\hat{v})))|. \end{aligned} \quad (15)$$

It follows that

$$\begin{aligned} P(|E^*(h(J_T^*(\hat{v}_T^*))) - E(h(J(\hat{v})))| > \epsilon) &\leq P\left(|E^*(h(J_T^*(\hat{v}_T^*))) - E^*(h(J_T^*(\hat{v}_1^*)))| > \frac{\epsilon}{3}\right) \\ &\quad + P\left(|E^*(h(J_T^*(\hat{v}_1^*))) - E(h(J(\hat{v}_2)))| > \frac{\epsilon}{3}\right) \\ &\quad + P\left(|E(h(J(\hat{v}_2))) - E(h(J(\hat{v})))| > \frac{\epsilon}{3}\right) \equiv I_1 + I_2 + I_3. \end{aligned} \quad (16)$$

To bound  $I_1$ , note that

$$\begin{aligned} E^*(h(J_T^*(\hat{v}_T^*))) - E^*(h(J_T^*(\hat{v}_1^*))) &= E^*(h(J_T^*(\hat{v}_T^*)) - h(J_T^*(\hat{v}_1^*))) | \hat{v}_T^* \in \mathbb{K} P^*(\hat{v}_T^* \in \mathbb{K}) \\ &\quad + E^*(h(J_T^*(\hat{v}_T^*)) - h(J_T^*(\hat{v}_1^*))) | \hat{v}_T^* \notin \mathbb{K} P^*(\hat{v}_T^* \notin \mathbb{K}), \end{aligned}$$

where the first term is zero since if  $\hat{v}_T^* \in \mathbb{K}$ ,  $J_T^*(\hat{v}_1^*)$  is necessarily equal to  $J_T^*(\hat{v}_T^*)$ . It follows that

$$|E^*(h(J_T^*(\hat{v}_T^*))) - E^*(h(J_T^*(\hat{v}_1^*)))| \leq 2MP^*(\hat{v}_T^* \notin \mathbb{K}),$$

where we have bounded the function  $h$  by its upper bound  $M$ . Thus, by (13),

$$I_1 \leq P\left(2MP^*(\hat{v}_T^* \notin \mathbb{K}) > \frac{\epsilon}{3}\right) = P\left(P^*(\hat{v}_T^* \notin \mathbb{K}) > \frac{\epsilon}{6M}\right) < \frac{\delta}{3}$$

for all  $T$  sufficiently large. Similarly,  $I_2 < \delta/3$  by (14). Finally, to bound  $I_3$ , note that by the same reasoning used above to bound  $I_1$ ,

$$\begin{aligned} I_3 &= P\left(|E(h(J(\hat{v}_2))) - E(h(J(\hat{v})))| > \frac{\epsilon}{3} | \hat{v} \in \mathbb{K}\right) P(\hat{v} \in \mathbb{K}) \\ &\quad + P\left(|E(h(J(\hat{v}_2))) - E(h(J(\hat{v})))| > \frac{\epsilon}{3} | \hat{v} \notin \mathbb{K}\right) P(\hat{v} \notin \mathbb{K}) \leq P(\hat{v} \notin \mathbb{K}) < \delta/3, \end{aligned}$$

given (13). The result now follows from (16).

## .2 Proofs of bootstrap results

**Proof of Proposition 3.1.** We would like to show that for any  $\epsilon > 0$ ,  $\lim_{T \rightarrow \infty} P\left(P^*\left(\|\hat{\theta}_T^* - \theta_0\| > \epsilon\right) > \epsilon\right) = 0$ . Because  $\theta_0$  is the unique minimizer of  $Q(\theta)$ , for any  $\epsilon > 0$  such that  $\|\theta - \theta_0\| > \epsilon$ , there is  $\delta > 0$  such that  $Q(\theta) - Q(\theta_0) \geq \delta > 0$ . It follows that

$$\begin{aligned} &P^*\left(\|\hat{\theta}_T^* - \theta_0\| > \epsilon\right) \leq P^*\left(Q(\hat{\theta}_T^*) - Q(\theta_0) > \delta\right) \\ &\leq P^*\left(Q(\hat{\theta}_T^*) - Q_T(\hat{\theta}_T^*) + Q_T(\hat{\theta}_T^*) - Q_T^*(\hat{\theta}_T^*) + Q_T^*(\hat{\theta}_T^*) - Q_T(\hat{\theta}_T) + Q_T(\hat{\theta}_T) - Q(\theta_0) > \delta\right) \\ &\leq P^*\left(Q(\hat{\theta}_T^*) - Q_T(\hat{\theta}_T^*) + Q_T(\hat{\theta}_T^*) - Q_T^*(\hat{\theta}_T^*) + Q_T^*(\hat{\theta}_T) - Q_T(\hat{\theta}_T) + Q_T(\theta_0) - Q(\theta_0) > \delta\right) \\ &\leq P^*\left(2 \sup_{\theta \in \Theta} |Q_T(\theta) - Q(\theta)| > \delta/2\right) + P^*\left(2 \sup_{\theta \in \Theta} |Q_T^*(\theta) - Q_T(\theta)| > \delta/2\right) \end{aligned}$$

where the third inequality uses the fact that  $Q_T^*(\hat{\theta}_T^*) \leq Q_T^*(\hat{\theta}_T)$  and  $Q_T(\hat{\theta}_T) \leq Q_T(\theta_0)$  by definition of  $\hat{\theta}_T^*$  and  $\hat{\theta}_T$ , respectively. Thus, for any  $\epsilon > 0$ ,

$$\begin{aligned} &P\left(P^*\left(\|\hat{\theta}_T^* - \theta_0\| > \epsilon\right) > \epsilon\right) \\ &\leq P\left(P^*\left(2 \sup_{\theta \in \Theta} |Q_T(\theta) - Q(\theta)| > \delta/2\right) > \epsilon/2\right) + P\left(P^*\left(2 \sup_{\theta \in \Theta} |Q_T^*(\theta) - Q_T(\theta)| > \delta/2\right) > \epsilon/2\right) \\ &\leq \frac{1}{\epsilon/2} P\left(2 \sup_{\theta \in \Theta} |Q_T(\theta) - Q(\theta)| > \delta/2\right) + P\left(P^*\left(2 \sup_{\theta \in \Theta} |Q_T^*(\theta) - Q_T(\theta)| > \delta/2\right) > \epsilon/2\right), \quad (17) \end{aligned}$$

where the first term in the second inequality is obtained by first noting that  $P^*(2 \sup_{\theta \in \Theta} |Q_T(\theta) - Q(\theta)| > \delta/2) = 1\{\cdot\}$  (where  $1\{\cdot\}$  denotes the indicator function containing the expression inside  $P^*(\cdot)$ ) and then applying Markov's inequality. From (17), it is clear that it suffices to show that (A)  $\sup_{\theta \in \Theta} |Q_T(\theta) - Q(\theta)| = o_P(1)$  and (B)  $\sup_{\theta \in \Theta} |Q_T^*(\theta) - Q_T(\theta)| = o_{P^*}(1)$  in prob- $P$ . (A) by standard arguments under our assumptions (see e.g. Theorem 2.6 of Newey and

McFadden (1994)). Thus, we only need to show (B). By definition,

$$Q_T^*(\theta) = \bar{\psi}_{c,T}^{*'}(\theta) W_T^* \bar{\psi}_{c,T}^*(\theta),$$

where  $\bar{\psi}_{c,T}^*(\theta) = \bar{\psi}_T^*(\theta) - E^* \left( \bar{\psi}_T^* \left( \hat{\theta}_T \right) \right) = \bar{\psi}_T^*(\theta) - \bar{\psi}_T \left( \hat{\theta}_T \right)$ , given the properties of the i.i.d. bootstrap. It follows that

$$\begin{aligned} Q_T^*(\theta) &= \left( \bar{\psi}_T^*(\theta) - \bar{\psi}_T \left( \hat{\theta}_T \right) \right)' W_T^* \left( \bar{\psi}_T^*(\theta) - \bar{\psi}_T \left( \hat{\theta}_T \right) \right) \\ &= \bar{\psi}_T^{*'}(\theta) W_T^* \bar{\psi}_T^*(\theta) - 2\bar{\psi}_T^{*'}(\theta) W_T^* \bar{\psi}_T \left( \hat{\theta}_T \right) + \bar{\psi}_T' \left( \hat{\theta}_T \right) W_T^* \bar{\psi}_T \left( \hat{\theta}_T \right) \\ &\equiv \bar{\psi}_T^{*'}(\theta) W_T^* \bar{\psi}_T^*(\theta) + r_{1T}^*(\theta) + r_{2T}^*, \end{aligned}$$

which implies that

$$\sup_{\theta \in \Theta} |Q_T^*(\theta) - Q_T(\theta)| \leq \sup_{\theta \in \Theta} \left| \bar{\psi}_T^{*'}(\theta) W_T^* \bar{\psi}_T^*(\theta) - \bar{\psi}_T'(\theta) W_T \bar{\psi}_T(\theta) \right| + \sup_{\theta \in \Theta} |r_{1T}^*(\theta)| + r_{2T}^*.$$

Now,  $r_{2T}^* \leq \left\| \bar{\psi}_T \left( \hat{\theta}_T \right) \right\|^2 W_T^* = o_P(1) \times O_{P^*}(1) = o_{P^*}(1)$  in prob- $P$ , given Lemma B.1 and given that  $\bar{\psi}_T \left( \hat{\theta}_T \right) = o_P(1)$  and  $W_T^* = O_{P^*}(1)$  in prob- $P$ . Similarly, we can show that  $\sup_{\theta \in \Theta} |r_{1T}^*(\theta)| = o_{P^*}(1)$  in prob- $P$  using in particular the fact that  $\sup_{\theta} \left\| \bar{\psi}_T^*(\theta) \right\| = O_{P^*}(1)$  in prob- $P$ . For the first term, note that we can write

$$\begin{aligned} &\bar{\psi}_T^{*'}(\theta) W_T^* \bar{\psi}_T^*(\theta) - \bar{\psi}_T'(\theta) W_T \bar{\psi}_T(\theta) \\ &= \left( \bar{\psi}_T^*(\theta) - \bar{\psi}_T(\theta) \right)' W_T^* \bar{\psi}_T^*(\theta) + \bar{\psi}_T(\theta)' W_T^* \left( \bar{\psi}_T^*(\theta) - \bar{\psi}_T(\theta) \right) + \bar{\psi}_T(\theta)' \left( W_T^* - W_T \right) \bar{\psi}_T(\theta) \\ &\leq \left\| \bar{\psi}_T^*(\theta) - \bar{\psi}_T(\theta) \right\| \left\| W_T^* \right\| \left( \left\| \bar{\psi}_T^*(\theta) \right\| + \left\| \bar{\psi}_T(\theta) \right\| \right) + \left\| \bar{\psi}_T(\theta) \right\|^2 \left\| W_T^* - W_T \right\|. \end{aligned}$$

Since  $\sup_{\theta \in \Theta} \left\| \bar{\psi}_T(\theta) \right\| = O_P(1)$ ;  $W_T^* - W_T = o_{P^*}(1)$ ;  $\left\| W_T^* \right\| = O_{P^*}(1)$ ; and  $\sup_{\theta \in \Theta} \left\| \bar{\psi}_T^*(\theta) \right\| = O_{P^*}(1)$  in prob- $P$ , it suffices to show that  $\sup_{\theta \in \Theta} \left\| \bar{\psi}_T^*(\theta) - \bar{\psi}_T(\theta) \right\| = o_{P^*}(1)$  in prob- $P$ . This follows by the bootstrap law of large numbers of Giné and Zinn (1990, Theorem 3.5), concluding the proof.

**Proof of Theorem 3.1.** (i) Given Proposition 3.1,  $\hat{\theta}_T \xrightarrow{P^*} \theta_0$ , prob- $P$ . Let  $\hat{u}_T^* = T^{1/4}(\hat{\theta}_T^* - \theta_0)$ . To show that  $\hat{u}_T^* = O_{\mathbb{P}}(1)$ , it suffices to show that for some  $\gamma_1 > 0$  and  $\gamma_2 > 0$ ,

$$\left\| \hat{u}_T^* \right\|^2 (\gamma_1 + o_{\mathbb{P}}(1)) \leq \gamma_2 O_{\mathbb{P}}(1) + \frac{o_{\mathbb{P}}(1)}{\left\| \hat{u}_T^* \right\|^2} + o_{\mathbb{P}}(1).$$

A second-order mean value expansion of  $\bar{\psi}_{c,T}^*(\hat{\theta}_T^*)$  around  $\theta_0$  gives:

$$\sqrt{T} \bar{\psi}_{c,T}^*(\hat{\theta}_T^*) = \sqrt{T} \bar{\psi}_{c,T}^*(\theta_0) + \sqrt{T} \frac{\partial \bar{\psi}_{c,T}^*}{\partial \theta'}(\theta_0) (\hat{\theta}_T^* - \theta_0) + \sqrt{T} \frac{1}{2} \bar{G}^*(\bar{\theta}_T^*) \text{vec}((\hat{\theta}_T^* - \theta_0)(\hat{\theta}_T^* - \theta_0)'),$$

where  $\bar{\theta}_T^* \in (\hat{\theta}_T^*, \theta_0)$  and may differ from row to row, and  $\bar{G}^*(\theta)$  is defined the same way as  $\bar{G}(\theta)$  but with bootstrap data. By a bootstrap uniform law of large numbers (cf. Giné and Zinn, 1990, Theorem 3.5),  $\sup_{\theta \in \mathcal{N}} \left\| \bar{G}^*(\theta) - \bar{G}(\theta) \right\| = o_{\mathbb{P}}(1)$ . This together with  $\sup_{\theta \in \mathcal{N}} \left\| \bar{G}(\theta) - G(\theta) \right\| = o_P(1)$  and the fact

that  $\hat{\theta}_T^* - \theta_0 = o_{\mathbb{P}}(1)$  imply that  $\bar{G}^*(\hat{\theta}_T^*) = G + o_{\mathbb{P}}(1)$ . Also,  $\sqrt{T} \frac{\partial \bar{\psi}_{c,T}^*}{\partial \theta'}(\theta_0) = O_{\mathbb{P}}(1)$ . To see this, write

$$\sqrt{T} \frac{\partial \bar{\psi}_{c,T}^*}{\partial \theta'}(\theta_0) = \sqrt{T} \left( \frac{\partial \bar{\psi}_{c,T}^*}{\partial \theta'}(\theta_0) - \frac{\partial \bar{\psi}_T}{\partial \theta'}(\theta_0) \right) + \sqrt{T} \frac{\partial \bar{\psi}_T}{\partial \theta'}(\theta_0)$$

and note that the first term is  $O_{\mathbb{P}}(1)$  by a bootstrap CLT applied to  $\frac{\partial \bar{\psi}_{c,T}^*}{\partial \theta'}(\theta_0) - \frac{\partial \bar{\psi}_T}{\partial \theta'}(\theta_0)$  whereas the second term is  $O_P(1)$  (hence  $O_{\mathbb{P}}(1)$ ) by a CLT for  $\frac{\partial \bar{\psi}_T}{\partial \theta'}(\theta_0)$  given that the expected value of the Jacobian is zero by the local identification condition. As a result,

$$\sqrt{T} \bar{\psi}_{c,T}^*(\hat{\theta}_T^*) = \sqrt{T} \bar{\psi}_{c,T}^*(\theta_0) + \frac{1}{2} G \text{vec} \left( \hat{u}_T^* \hat{u}_T^{*'} \right) + o_{\mathbb{P}}(1) + o_{\mathbb{P}} \left( \|\hat{u}_T^*\|^2 \right), \quad (18)$$

where  $\bar{\psi}_{c,T}^*(\theta_0) = (\bar{\psi}_T^*(\theta_0) - \bar{\psi}_T(\theta_0)) + \bar{\psi}_T(\theta_0) - \bar{\psi}_T(\hat{\theta}_T) = O_{\mathbb{P}}(T^{-1/2}) + O_P(T^{-1/2}) + O_P(T^{-1/2}) = O_{\mathbb{P}}(T^{-1/2})$ , and where the order of magnitude of  $\bar{\psi}_T(\hat{\theta}_T)$  is obtained through a second-order Taylor expansion of  $\bar{\psi}_T(\hat{\theta}_T)$  around  $\theta_0$ :

$$\sqrt{T} \bar{\psi}_T(\hat{\theta}_T) = \sqrt{T} \bar{\psi}_T(\theta_0) + \frac{1}{2} G \text{vec}(T^{1/4}(\hat{\theta}_T - \theta_0)T^{1/4}(\hat{\theta}_T - \theta_0)') + o_P(1). \quad (19)$$

Thus,

$$\begin{aligned} T \bar{\psi}_{c,T}^{*'}(\hat{\theta}_T^*) W_T^* \bar{\psi}_{c,T}^*(\hat{\theta}_T^*) &= \bar{\psi}_{c,T}^{*'}(\theta_0) W \bar{\psi}_{c,T}^*(\theta_0) + \sqrt{T} \bar{\psi}_{c,T}^{*'}(\theta_0) W G \text{vec}(\hat{u}_T^* \hat{u}_T^{*'}) \\ &\quad + \frac{1}{4} \text{vec}'(\hat{u}_T^* \hat{u}_T^{*'}) G' W G \text{vec}(\hat{u}_T^* \hat{u}_T^{*'}) + o_{\mathbb{P}}(1) + o_{\mathbb{P}}(\|\hat{u}_T^*\|^2) + o_{\mathbb{P}}(\|\hat{u}_T^*\|^4). \end{aligned}$$

By definition,

$$T \bar{\psi}_{c,T}^{*'}(\hat{\theta}_T^*) W_T^* \bar{\psi}_{c,T}^*(\hat{\theta}_T^*) \leq T \bar{\psi}_{c,T}^{*'}(\theta_0) W \bar{\psi}_{c,T}^*(\theta_0) + o_{\mathbb{P}}(1),$$

which implies that

$$\frac{1}{4} \text{vec}'(\hat{u}_T^* \hat{u}_T^{*'}) G' W G \text{vec}(\hat{u}_T^* \hat{u}_T^{*'}) \leq -\sqrt{T} \bar{\psi}_{c,T}^{*'}(\theta_0) W G \text{vec}(\hat{u}_T^* \hat{u}_T^{*'}) + o_{\mathbb{P}}(1) + o_{\mathbb{P}}(\|\hat{u}_T^*\|^2) + o_{\mathbb{P}}(\|\hat{u}_T^*\|^4).$$

From Lemma A.1 of D&R (2013) and given Assumption 6 (ii), we have that  $\frac{1}{4} \text{vec}'(\hat{u}_T^* \hat{u}_T^{*'}) G' W G \text{vec}(\hat{u}_T^* \hat{u}_T^{*'}) \geq \gamma_1 \|\hat{u}_T^*\|^4$  for some  $\gamma_1 > 0$ . Thus, the previous inequality together with Cauchy-Schwarz inequality imply that

$$\gamma_1 \|\hat{u}_T^*\|^4 \leq \gamma_2 \|\sqrt{T} \bar{\psi}_{c,T}^{*'}(\theta_0)\| \|\hat{u}_T^*\|^2 + o_{\mathbb{P}}(1) + o_{\mathbb{P}}(\|\hat{u}_T^*\|^2) + o_{\mathbb{P}}(\|\hat{u}_T^*\|^4),$$

where  $\gamma_2 = \|W\| \|G\|$ , which completes the proof of part (i).

Next, we prove part (ii). We can show that the random vector

$\left( \sqrt{T} \bar{\psi}'_T(\theta_0), \sqrt{T} (\bar{\psi}'_T(\theta_0) - \bar{\psi}'_T(\theta_0))', T^{1/4} (\hat{\theta}_T - \theta_0)', T^{1/4} (\hat{\theta}_T^* - \hat{\theta}_T)' \right)' = O_{\mathbb{P}}(1)$ . In particular,  $\sqrt{T} \bar{\psi}_T(\theta_0) \rightarrow^d X \sim N(0, \Sigma)$ , implying that the first term is  $O_P(1)$  (hence  $O_{\mathbb{P}}(1)$ ), whereas by an application of a bootstrap CLT,  $\sqrt{T} (\bar{\psi}'_T(\theta_0) - \bar{\psi}'_T(\theta_0)) \rightarrow^{d^*} X^* \sim N(0, \Sigma)$ , implying that the second term is  $O_{P^*}(1)$  in prob- $P$  (hence, it is  $O_{\mathbb{P}}(1)$  by Lemma .1.a3)). Finally, Propostion 2.1 and part (i) of this theorem justify the  $O_{\mathbb{P}}(1)$  for the last two terms. By Prohorov's theorem (cf. van



der Vaart and Welner (1990)), it follows that this random vector has at least one subsequence<sup>4</sup> which converges in distribution under  $\mathbb{P}$  towards  $(X, X^*, V, U^*)$ , say, where  $X^*$  is independent of  $(X, V)$  (to obtain this last result, we apply Lemma 3.1 of Sen, Banerjee and Woodroffe (2010)). Next, we will show that  $P(|U^*| > 0) > 0$ . By the second-order condition for the minimization problem underlying  $\hat{\theta}_T^*$ ,

$$\left. \frac{\partial^2}{\partial \theta^2} \bar{\psi}_{c,T}^*(\theta) W_T^* \bar{\psi}_{c,T}^*(\theta) \right|_{\theta=\hat{\theta}_T^*} \geq 0,$$

which can be written as

$$\tilde{Z}_T^* + N_T^* \geq 0, \tag{20}$$

where

$$\tilde{Z}_T^* = \frac{\partial^2 \bar{\psi}_{c,T}^*}{\partial \theta^2}(\hat{\theta}_T^*) W_T^* \sqrt{T} \bar{\psi}_{c,T}^*(\hat{\theta}_T^*) \quad \text{and} \quad N_T^* = \sqrt{T} \frac{\partial \bar{\psi}_{c,T}^*}{\partial \theta}(\hat{\theta}_T^*) W_T^* \frac{\partial \bar{\psi}_{c,T}^*}{\partial \theta}(\hat{\theta}_T^*).$$

From part (i) of this theorem, and equations (18) and (19), we can show that

$$\tilde{Z}_T^* = Z_T^* - \frac{1}{2} G' W G \hat{v}_T^2 + \frac{1}{2} G' W G \hat{u}_T^{*2} + o_{\mathbb{P}}(1),$$

where  $Z_T^* = G' W \sqrt{T} (\bar{\psi}_T^*(\theta_0) - \bar{\psi}_T(\theta_0))$  and  $N_T^* = G' W G \hat{u}_T^{*2} + o_{\mathbb{P}}(1)$ . From (20), we have that

$$-Z_T^* + A \hat{v}_T^2 - 3A \hat{u}_T^{*2} \leq o_{\mathbb{P}}(1),$$

with  $A = G' W G / 2 > 0$  since  $G \neq 0$  given the second-order identification assumption. By the continuous mapping theorem,  $-Z_T^* + A \hat{v}_T^2 - 3A \hat{u}_T^{*2} \rightarrow^d -Z^* + AV^2 - 3AU^{*2}$  under  $\mathbb{P}$ . Using the same arguments as in Lemma B.2 of D&R (2013b), we can claim that  $P(-Z^* + AV^2 - 3AU^{*2} \leq 0) = 1$ , which implies that  $P(U^{*2} \neq 0 | Z^* < 0, V = 0) = 1$ . Hence,  $P(U^* \neq 0) \geq P(Z^* < 0, V = 0)$ . Since  $Z^* = G' W X^*$  is independent of  $V$ , we have that  $P(Z^* < 0, V = 0) = P(Z^* < 0)P(V = 0)$ , where  $P(Z^* < 0) = 1/2$  (since  $Z^*$  is a non-degenerate Gaussian random variable) and we can show (similarly to the proof of Proposition 3.1 of D&R (2013)) that  $P(V = 0) = 1/2$  when  $p = 1$ . As a result,  $P(U^* \neq 0) \geq 1/4 > 0$ .

Next, to prove part (iii), we apply Lemma B.5 of D&R (2013b). First, note that  $\hat{J}_T^*(u)$  and  $J^*(u)$  have continuous sample paths (in particular,  $J^*(u)$  is a polynomial function in  $u$  for each value of  $V$  and  $X^*$ ). Also,  $\mathbb{H}_T$  is a non-decreasing sequence of sets and, since  $\theta_0$  is interior point in  $\Theta$ ,  $\bigcup_{T \geq 0} \mathbb{H}_T = \mathbb{R}$ . It remains to check conditions (i)-(iii) of that lemma. (ii) follows from part (i) of this Theorem by choosing  $u_T^* \in \arg \min_{u \in \mathbb{H}_T} \hat{J}_T^*(u)$  equal to  $\hat{u}_T^*$ . Similarly, we can show that  $u^* \in \arg \min_{u \in \mathbb{R}} J^*(u)$  is tight by relying on D&R (2013b)'s Lemma B.6 (ii). To show (i), it suffices to show that  $Y_T^*(u) \equiv \bar{\psi}_{c,T}^*(\theta_0 + T^{-1/4}u)$  converges in distribution under  $\mathbb{P}$  towards  $Y^* \equiv X^* - \frac{1}{2} G V^2 + \frac{1}{2} G u^2$ , in  $\ell^\infty(\mathbb{K})$ . For this, since  $\mathbb{K}$  equipped with the usual metric on  $\mathbb{R}$  is totally bounded as any compact subset of  $\mathbb{R}$ , we show that (a1) The marginals  $(Y_T^*(u_1), \dots, Y_T^*(u_k))$  converge in distribution to  $(Y^*(u_1), \dots, Y^*(u_k))$  with respect to  $\mathbb{P}$  for any  $u_1, \dots, u_k \in \mathbb{K}$ , and (a2) The stochastic process  $Y_T^*(u)$  is asymptotically

<sup>4</sup>To simplify the notation, we keep the same index  $T$  to denote this subsequence throughout this proof.

equicontinuous, i.e.  $\forall \epsilon > 0$ ,

$$\lim_{\delta \rightarrow 0} \limsup_{T \rightarrow \infty} \mathbb{P} \left( \sup_{u_1, u_2 \in \mathbb{K}: |u_1 - u_2| < \delta} \|Y_T^*(u_1) - Y_T^*(u_2)\| > \epsilon \right) = 0.$$

(a1) follows by two second-order mean-value expansions: one for  $\sqrt{T}Y_T^*(u)$  around 0, and another for  $\sqrt{T}\bar{\psi}_T(\hat{\theta}_T)$  around  $\theta_0$ . This implies that

$$\sqrt{T}Y_T^*(u) = \sqrt{T}(\bar{\psi}_T^*(\theta_0) - \bar{\psi}_T(\theta_0)) - \frac{1}{2}G\hat{v}_T^2 + \frac{1}{2}Gu^2 + o_{\mathbb{P}}(1),$$

where  $(\sqrt{T}(\bar{\psi}_T^*(\theta_0) - \bar{\psi}_T(\theta_0)), \hat{v}_T)$  converges in distribution towards  $(X^*, V)$  with respect to  $\mathbb{P}$ . We can then apply the continuous mapping theorem and conclude that  $(Y_T^*(u_1), \dots, Y_T^*(u_k))$  converge in distribution to  $(Y^*(u_1), \dots, Y^*(u_k))$  with respect to  $\mathbb{P}$ . To prove (a2), observe that

$$\|Y_T^*(u_1) - Y_T^*(u_2)\| \leq C\|G\|\|u_1 - u_2\| + o_{\mathbb{P}}(1),$$

for some  $C > 0$  and where the neglected terms are uniformly negligible over  $\mathbb{K}$ . This implies (a2) and ends the proof of part (iii). Finally, we prove part (iv). The first-order condition for the problem  $\min_{u \in \mathbb{R}} J^*(u)$  is

$$\left( (2X^*W'G - G'WGV^2) + G'WGu^{*2} \right) u^* = 0.$$

Hence, the possible minimizers are  $u^* = 0$  and  $u^*$  such that  $G'WGu^{*2} = G'WGV^2 - 2X^*W'G$ . The second-order necessary condition for a minimum imposes that

$$2X^*W'G - G'WGV^2 + 3G'WGu^{*2} \geq 0.$$

Thus, if  $2X^*W'G - G'WGV^2 \geq 0$ , then  $J^*(u)$  is minimized at  $u^* = 0$  and

$$\underline{J}^* = \underline{J}_1^* \equiv X^*W'X^* - X^*W'GV^2 + \frac{1}{4}G'WGV^4 = \left( X^* - \frac{1}{2}GV^2 \right)' W \left( X^* - \frac{1}{2}GV^2 \right);$$

otherwise, if  $2X^*W'G - G'WGV^2 < 0$ , then  $J^*(u)$  is minimized at  $u^*$  satisfying  $G'WGu^{*2} = G'WGV^2 - 2X^*W'G$ , in which case we can show that

$$\underline{J}^* = \underline{J}_2^* \equiv X^*W^{1/2} \left( I_H - W^{1/2}G(G'WG)^{-1}G'W^{1/2} \right) W^{1/2}X^*.$$

It follows that

$$E(\underline{J}^*) = E(\underline{J}_1^* | 2X^*W'G - G'WGV^2 \geq 0)q^* + E(\underline{J}_2^* | 2X^*W'G - G'WGV^2 < 0)(1 - q^*), \quad (21)$$

with  $q^* = P(2X^*W'G - G'WGV^2 \geq 0)$ . We claim that  $\underline{J}_2^*$  is independent of  $V$  and  $G'WX^*$ . To see this, note that

$$\underline{J}_2^* = X^*W^{1/2}M_{W^{1/2}G}W^{1/2}X^*,$$

where  $M_{W^{1/2}G} = I_H - W^{1/2}G(G'WG)^{-1}G'W^{1/2}$ . It follows that

$$E \left( G'WX^* \left( X^*W^{1/2}M_{W^{1/2}G} \right) \right) = G'W\Sigma W^{1/2}M_{W^{1/2}G} = G'W^{1/2}M_{W^{1/2}G} = 0,$$

where we used the fact that  $W = \Sigma^{-1}$ . Since  $X^*$  is Gaussian, a null covariance amounts to independence, implying that  $\underline{J}_2^*$  is independent of  $G'WX^*$ . Moreover,  $\underline{J}_2^*$  is also independent of  $V$  because it depends only on  $X^*$ , which is independent of  $V$ . As a result, we have that

$$E(\underline{J}_2^* | 2X^{*'}WG - G'WGV^2 < 0) = E(\underline{J}_2^*) = H - 1,$$

since  $\underline{J}_2^* \sim \chi^2(H - 1)$ . Next, we show that  $q^* = 3/8$ . Using the definition of  $q^*$ , we can show that

$$\begin{aligned} q^* &= P(G'WX^* \geq G'WGV^2/2 | V = 0)P(V = 0) + P(G'WX^* \geq G'WGV^2/2 | V \neq 0)P(V \neq 0) \\ &= P(G'WX^* \geq 0 | V = 0)P(V = 0) + P(G'WX^* \geq G'WGV^2/2 | V \neq 0)P(V \neq 0) \\ &= \frac{1}{4} + \frac{1}{2}P(G'WX^* \geq G'WGV^2/2 | V \neq 0), \end{aligned}$$

where the last equality uses the fact that  $P(G'WX^* \geq 0 | V = 0) = P(G'WX^* \geq 0) = 1/2$  (since  $X^*$  is independent of  $V$  and  $G'WX^*$  is a non-degenerate mean zero Gaussian variable) as well as the fact that  $P(V = 0) = P(V \neq 0) = 1/2$  when  $p = 1$ . In addition, we can show that  $V^2 = -2\frac{G'WX}{G'WG}1(G'WX < 0)$ , which implies that

$$\begin{aligned} P(G'WX^* \geq G'WGV^2/2 | V \neq 0) &= P(G'WX^* + G'WX \geq 0 | G'WX < 0) \\ &= P\left(\frac{G'WX^*}{\sqrt{G'WG}} + \frac{G'WX}{\sqrt{G'WG}} \geq 0 \mid \frac{G'WX}{\sqrt{G'WG}} < 0\right). \end{aligned}$$

Hence,

$$P(G'WX^* \geq G'WGV^2/2 | V \neq 0) = P(X_1 + X_2 \geq 0 | X_2 \leq 0) = \frac{P(X_1 + X_2 \geq 0, X_2 \leq 0)}{P(X_2 \leq 0)} = \frac{1/8}{1/2} = \frac{1}{4},$$

where  $P(X_2 \leq 0) = 1/2$  (since  $X_2$  is zero-mean non degenerate Gaussian on  $\mathbb{R}$ ) and  $P(X_1 + X_2 \geq 0, X_2 \leq 0) = 1/8$  (since it equals the probability that a zero-mean non-degenerate  $\mathbb{R}^2$ -valued random vector lies in half of a quadrant). Thus,  $q^* = \frac{3}{8}$ . To finish the proof, we evaluate  $E(\underline{J}_1^* | 2X^{*'}WG - G'WGV^2 \geq 0)$ . We have that

$$E(\underline{J}_1^* | 2X^{*'}WG - G'WGV^2 \geq 0) = (a) + (b),$$

with

$$(a) = E(\underline{J}_1^* | V = 0, 2G'WX^* - G'WGV^2 \geq 0)P(V = 0 | 2G'WX^* - G'WGV^2 \geq 0)$$

and

$$(b) = E(\underline{J}_1^* | V \neq 0, 2G'WX^* - G'WGV^2 \geq 0)P(V \neq 0 | 2G'WX^* - G'WGV^2 \geq 0).$$

We start with (a). We have that

$$E(\underline{J}_1^* | V = 0, 2G'WX^* - G'WGV^2 \geq 0) = E(X^{*'}WX^* | G'WX^* \geq 0) = E(X^{*'}WX^*) = H,$$

where the first equality follows because  $V = 0$ , the second uses the fact that  $X^{*'}WX^*$  is independent of the event  $(G'WX^* \geq 0)$  (see the proof of Corollary 3.2 of D&R (2013) for details on how to obtain this result) and the third follows because  $X^{*'}WX^* \sim \chi^2(H)$ . Moreover,

$$\begin{aligned} P(V = 0 | 2G'WX^* - G'WGV^2 \geq 0) &= \frac{P(V=0, G'WX^* \geq 0)}{q^*} = \frac{P(G'WX \geq 0, G'WX^* \geq 0)}{q^*} \\ &= \frac{P(G'WX \geq 0)P(G'WX^* \geq 0)}{q^*} = \frac{1}{4q^*}. \end{aligned}$$

Thus, (a) =  $\frac{H}{4q^*}$ . Let us now derive (b). We can show that

$$\begin{aligned} P(V \neq 0 | 2G'WX^* - G'WGV^2 \geq 0) &= \frac{P(G'WX \leq 0, G'WX^* + G'WX \geq 0)}{P(X_1 \leq 0, X_1 + X_2 \geq 0)} = \frac{q^*}{8q^*}, \\ &= \frac{1}{8q^*}, \end{aligned}$$

using the fact that  $X_1$  and  $X_2$  are two independent standard normal random variables. In addition,

$$\begin{aligned} &E(\underline{J}_1^* | V \neq 0, 2G'WX^* - G'WGV^2 \geq 0) \\ &= E\left(\left(X^* + G\frac{G'WX}{G'WG}\right)' W \left(X^* + G\frac{G'WX}{G'WG}\right) \middle| G'WX \leq 0, G'WX + G'WX^* \geq 0\right) \\ &= (b1) + 2(b2) + (b3), \end{aligned}$$

where

$$\begin{aligned} (b1) &= E(X^*WX^* | G'WX \leq 0, G'WX + G'WX^* \geq 0), \\ (b2) &= E\left(\frac{G'WX^*}{\sqrt{G'WG}} \frac{G'WX^*}{\sqrt{G'WG}} \middle| G'WX \leq 0, G'WX + G'WX^* \geq 0\right), \\ (b3) &= E\left(\frac{(G'WX^*)^2}{G'WG} \middle| G'WX \leq 0, G'WX + G'WX^* \geq 0\right). \end{aligned}$$

Since  $X^*$  and  $X$  are independent, (b1) =  $E(X^*WX^* | G'WX + G'WX^* \geq 0)$ . Let  $Q$  be the rotation matrix that transforms the canonical basis of  $\mathbb{R}^H$  into the orthonormal basis  $(a_1, a_2, \dots, a_H)$  with  $a_H = \frac{W^{1/2}G}{\sqrt{G'WG}}$ . Let  $Y^*$  be the coordinate of  $W^{1/2}X^*$  in this new basis. We have  $Y^* = Q'W^{1/2}X^*$  and since  $Q'Q = I_H$ ,  $Y^* \sim N(0, I_H)$ , where the last component of  $Y^*$  is  $Y_H^* = \frac{G'WX^*}{\sqrt{G'WG}}$  and  $X^*WX^* = Y_1^{*2} + \dots + Y_{H-1}^{*2} + Y_H^{*2}$ , with  $Y_i^*$ 's independent and identically distributed  $N(0, 1)$ . Thus,

$$\begin{aligned} (b1) &= E(Y_1^{*2} + \dots + Y_{H-1}^{*2} + Y_H^{*2} | G'WX + G'WX^* \geq 0) \\ &= E(Y_1^{*2} + \dots + Y_{H-1}^{*2}) + E(Y_H^{*2} | G'WX + G'WX^* \geq 0) \\ &= H - 1 + E(Y_H^{*2} | G'WX + G'WX^* \geq 0). \end{aligned}$$

But  $E(Y_H^{*2} | G'WX + G'WX^* \geq 0) = E(X_1^2 | X_1 + X_2 \geq 0) = 1$ , with  $X_1$  and  $X_2$  independent standard normal random variables (we can prove this last equality by relying on the properties of standard random normal variables). Thus, (b1) =  $H$ . Similarly,

$$\begin{aligned} (b2) &= E\left(\frac{G'WX^*}{\sqrt{G'WG}} \frac{G'WX^*}{\sqrt{G'WG}} \middle| \frac{G'WX}{\sqrt{G'WG}} \leq 0, \frac{G'WX}{\sqrt{G'WG}} + \frac{G'WX^*}{\sqrt{G'WG}} \geq 0\right) \\ &= E(X_1X_2 | X_1 \leq 0, X_1 + X_2 \geq 0) = -\frac{2}{\pi}, \end{aligned}$$

by again relying on the properties of standard normal random variables. A similar argument shows that (b3) = 1, implying that

$$E(\underline{J}_1^* | V \neq 0, 2G'WX^* - G'WGV^2 \geq 0) = H - \frac{4}{\pi} + 1,$$

and therefore (b) =  $\frac{1}{8q^*} (H - \frac{4}{\pi} + 1)$ . Hence,

$$E(\underline{J}_1^* | 2X^*WG - G'WGV^2 \geq 0) = (a) + (b) = \frac{H}{4q^*} + \frac{1}{8q^*} \left(H - \frac{4}{\pi} + 1\right) = \frac{1}{8q^*} \left(3H - \frac{4}{\pi} + 1\right).$$

Plugging these results in (21) gives the expected result.

**Proof of Proposition 4.1.** The proof follows closely that of Proposition 3.1, so we only highlight

the differences. To prove part (B) in that proof, using the definition of  $Q_T^{*(1)}(\theta)$ , we can show that

$$\begin{aligned} Q_T^{*(1)}(\theta) &= Q_T^*(\theta) - 2 \left( \bar{\psi}_T^*(\theta) - \bar{\psi}_T(\hat{\theta}_T) \right)' W_T^{*(1)} \frac{\partial}{\partial \theta'} \bar{\psi}_T(\hat{\theta}_T) (\theta - \hat{\theta}_T) \\ &\quad + (\theta - \hat{\theta}_T)' \left( \frac{\partial}{\partial \theta'} \bar{\psi}_T(\hat{\theta}_T) \right)' W_T^{*(1)} \frac{\partial}{\partial \theta'} \bar{\psi}_T(\hat{\theta}_T) (\theta - \hat{\theta}_T) \\ &\equiv Q_T^*(\theta) + S_{1T}^*(\theta) + S_{2T}^*(\theta), \end{aligned}$$

where  $Q_T^*(\theta)$  is as defined in Proposition 3.1 but using the weighting matrix  $W_T^{*(1)}$ . We proved already that  $\sup_{\theta} |Q_T^*(\theta) - Q_T(\theta)| = o_{P^*}(1)$ , in prob- $P$ , thus it suffices to show that the two last terms above are  $o_{P^*}(1)$ , in prob- $P$ , uniformly over  $\Theta$ . Since  $\frac{\partial}{\partial \theta'} \bar{\psi}_T(\hat{\theta}_T) \rightarrow^P 0$ ,  $S_{2T}^*(\theta)$  is  $o_{P^*}(1)$  in prob- $P$  uniformly over  $\theta \in \Theta$  given in particular the fact that  $\|\theta - \hat{\theta}_T\|$  is bounded with probability  $P$  converging to 1 (given the compactness of  $\Theta$  and the fact that  $\hat{\theta}_T \rightarrow^P \theta_0 \in \Theta$ ). Similarly, we can show that the second term can be bounded by a factor of  $\sup \|\bar{\psi}_T^*(\theta) - \bar{\psi}_T(\theta)\| = o_{P^*}(1)$ , in prob- $P$ , where the factor is  $O_{P^*}(1)$ , in prob- $P$ . In particular, we have that

$$\begin{aligned} |S_{1T}^*(\theta)| &\leq \left\| \bar{\psi}_T^*(\theta) - \bar{\psi}_T(\theta) + \bar{\psi}_T(\theta) - \bar{\psi}_T(\hat{\theta}_T) \right\| \left\| W_T^{*(1)} \right\| \left\| \frac{\partial}{\partial \theta'} \bar{\psi}_T(\hat{\theta}_T) \right\| \|\theta - \hat{\theta}_T\| \\ &\leq \left( \sup_{\theta} \|\bar{\psi}_T^*(\theta) - \bar{\psi}_T(\theta)\| + 2 \sup_{\theta} \|\bar{\psi}_T(\theta)\| \right) \left\| W_T^{*(1)} \right\| \left\| \frac{\partial}{\partial \theta'} \bar{\psi}_T(\hat{\theta}_T) \right\| \|\theta - \hat{\theta}_T\|, \end{aligned}$$

which implies that  $\sup_{\theta} |S_{1T}^*(\theta)| = o_{P^*}(1)$  in prob- $P$  provided  $\sup \|\bar{\psi}_T^*(\theta) - \bar{\psi}_T(\theta)\| = o_{P^*}(1)$ , and  $\left\| W_T^{*(1)} \right\| = O_{P^*}(1)$ , in prob- $P$ , given that  $\sup_{\theta} \|\bar{\psi}_T(\theta)\| = O_P(1)$ ,  $\frac{\partial}{\partial \theta'} \bar{\psi}_T(\hat{\theta}_T) \rightarrow^P 0$ , and  $\sup_{\theta} \|\theta - \hat{\theta}_T\| = O_P(1)$ .

**Proof of Proposition 4.2.** We follow the proof of Proposition 2.2. (i) A second-order mean value expansion of  $\bar{\psi}^{*(1)}(\hat{\theta}_T^{*(1)})$  around  $\hat{\theta}_T$  gives

$$\sqrt{T} \bar{\psi}_T^{*(1)}(\hat{\theta}_T^{*(1)}) = \sqrt{T} \bar{\psi}_T^{*(1)}(\hat{\theta}_T) + \sqrt{T} \frac{\partial \bar{\psi}_T^{*(1)}}{\partial \theta'}(\hat{\theta}_T) (\hat{\theta}_T^{*(1)} - \hat{\theta}_T) + \frac{1}{2} R_T^{*(1)}(\bar{\theta}^*), \quad (22)$$

where

$$R_T^{*(1)}(\bar{\theta}^*) = \sqrt{T} \bar{G}^{*(1)}(\bar{\theta}^*) \text{vec} \left( (\hat{\theta}_T^{*(1)} - \hat{\theta}_T) (\hat{\theta}_T^{*(1)} - \hat{\theta}_T)' \right) \equiv \bar{G}^{*(1)}(\bar{\theta}^*) \text{vec} \left( \hat{v}_T^{*(1)} \hat{v}_T^{*(1)'} \right)$$

with  $\bar{\theta}^* \in (\hat{\theta}_T^{*(1)}, \hat{\theta}_T)$  and may differ from row to row and  $\bar{G}^{*(1)}(\theta)$  is defined the same way as  $\bar{G}(\theta)$  but with  $\bar{\psi}_T(\theta)$  replaced with  $\bar{\psi}_T^{*(1)}(\theta)$ . By Proposition 4.1,  $\hat{\theta}_T^{*(1)} = \theta_0 + o_{P^*}(1)$  in prob- $P$ , which implies that  $\bar{G}^{*(1)}(\bar{\theta}^*) = G + o_{P^*}(1)$  in prob- $P$  (using the fact that  $\sup_{\theta \in \Theta} \|\bar{G}^{*(1)}(\theta) - \bar{G}(\theta)\| = o_{P^*}(1)$  in prob- $P$  and  $\sup_{\theta \in \Theta} \|\bar{G}(\theta) - G(\theta)\| = o_P(1)$ ). Thus,

$$R_T^{*(1)}(\bar{\theta}^*) = G \text{vec} \left( \hat{v}_T^{*(1)} \hat{v}_T^{*(1)'} \right) + o_{P^*} \left( \left\| \hat{v}_T^{*(1)} \right\|^2 \right),$$

in prob- $P$ . By a bootstrap CLT and under our assumptions, we have that  $\sqrt{T} \bar{\psi}_T^{*(1)}(\hat{\theta}_T) = \sqrt{T} (\bar{\psi}_T^*(\hat{\theta}_T) - \bar{\psi}_T(\hat{\theta}_T)) = O_{P^*}(1)$ , in prob- $P$ , and

$\sqrt{T} \frac{\partial \bar{\psi}_T^{*(1)}}{\partial \theta'} (\hat{\theta}_T) = \sqrt{T} \left( \frac{\partial \bar{\psi}_T^*}{\partial \theta'} (\hat{\theta}_T) - \frac{\partial \bar{\psi}_T}{\partial \theta'} (\hat{\theta}) \right) = O_{P^*}(1)$  in prob- $P$ . As a result,

$$\sqrt{T} \bar{\psi}_T^{*(1)} (\hat{\theta}_T^{*(1)}) = \sqrt{T} \bar{\psi}_T^{*(1)} (\hat{\theta}_T) + \frac{1}{2} \text{Gvec} (\hat{v}_T^{*(1)} \hat{v}_T^{*(1)'}) + o_{P^*}(1) + o_{P^*} (\|\hat{v}_T^{*(1)}\|^2), \quad (23)$$

in prob- $P$ . Thus,

$$\begin{aligned} T \bar{\psi}_T^{*(1)'} (\hat{\theta}_T^{*(1)}) W_T^{*(1)} \bar{\psi}_T^{*(1)} (\hat{\theta}_T^{*(1)}) &= T \bar{\psi}_T^{*(1)'} (\hat{\theta}_T) W \bar{\psi}_T^{*(1)} (\hat{\theta}_T) + \frac{1}{4} \text{vec}' (\hat{v}_T^{*(1)} \hat{v}_T^{*(1)'}) G' W \text{Gvec} (\hat{v}_T^{*(1)} \hat{v}_T^{*(1)'}) \\ &\quad + \text{vec}' (\hat{v}_T^{*(1)} \hat{v}_T^{*(1)'}) G' W \sqrt{T} \bar{\psi}_T^{*(1)} (\hat{\theta}_T) + o_{P^*}(1) \\ &\quad + o_{P^*} (\|\hat{v}_T^{*(1)}\|^2) + o_{P^*} (\|\hat{v}_T^{*(1)}\|^4), \end{aligned}$$

given in particular the fact that  $W_T^{*(1)} = W + o_{P^*}(1)$ , in prob- $P$ . By definition,

$$T \bar{\psi}_T^{*(1)'} (\hat{\theta}_T^{*(1)}) W_T^{*(1)} \bar{\psi}_T^{*(1)} (\hat{\theta}_T^{*(1)}) \leq T \bar{\psi}_T^{*(1)'} (\hat{\theta}_T) W_T^{*(1)} \bar{\psi}_T^{*(1)} (\hat{\theta}_T) = T \bar{\psi}_T^{*(1)'} (\hat{\theta}_T) W \bar{\psi}_T^{*(1)} (\hat{\theta}_T) + o_{P^*}(1),$$

in prob- $P$ . Thus,

$$\begin{aligned} \frac{1}{4} \text{vec}' (\hat{v}_T^{*(1)} \hat{v}_T^{*(1)'}) G' W \text{Gvec} (\hat{v}_T^{*(1)} \hat{v}_T^{*(1)'}) + \text{vec}' (\hat{v}_T^{*(1)} \hat{v}_T^{*(1)'}) G' W \sqrt{T} \bar{\psi}_T^{*(1)} (\hat{\theta}_T) \\ + o_{P^*}(1) + o_{P^*} (\|\hat{v}_T^{*(1)}\|^2) + o_{P^*} (\|\hat{v}_T^{*(1)}\|^4) \leq o_{P^*}(1), \end{aligned}$$

in prob- $P$ , implying that

$$\begin{aligned} \frac{1}{4} \text{vec}' (\hat{v}_T^{*(1)} \hat{v}_T^{*(1)'}) G' W \text{Gvec} (\hat{v}_T^{*(1)} \hat{v}_T^{*(1)'}) \leq \text{vec}' (\hat{v}_T^{*(1)} \hat{v}_T^{*(1)'}) G' W \sqrt{T} \bar{\psi}_T^{*(1)} (\hat{\theta}_T) \\ + o_{P^*}(1) + o_{P^*} (\|\hat{v}_T^{*(1)}\|^2) + o_{P^*} (\|\hat{v}_T^{*(1)}\|^4). \end{aligned}$$

Given Lemma B.1 of Dovonon and Renault (2013) and Assumption 5 (ii), we have that  $\frac{1}{4} \text{vec}' (\hat{v}_T^{*(1)} \hat{v}_T^{*(1)'}) G' W \text{Gvec} (\hat{v}_T^{*(1)} \hat{v}_T^{*(1)'}) \geq \gamma_1 \|\hat{v}_T^{*(1)}\|^4$  for some  $\gamma_1 > 0$ . Hence,

$$\gamma_1 \|\hat{v}_T^{*(1)}\|^4 \leq \|W\| \|G\| \|\sqrt{T} \bar{\psi}_T^{*(1)} (\hat{\theta}_T)\| \|\hat{v}_T^{*(1)}\|^2 + o_{P^*}(1) + o_{P^*} (\|\hat{v}_T^{*(1)}\|^2) + o_{P^*} (\|\hat{v}_T^{*(1)}\|^4),$$

implying that

$$\|\hat{v}_T^{*(1)}\|^2 (\gamma_1 + o_{P^*}(1)) \leq \|W\| \|G\| \|\sqrt{T} \bar{\psi}_T^{*(1)} (\hat{\theta}_T)\| + \frac{o_{P^*}(1)}{\|\hat{v}_T^{*(1)}\|^2} + o_{P^*}(1).$$

This shows that  $\|\hat{v}_T^{*(1)}\|^2$  is at most of the same order as  $\|\sqrt{T} \bar{\psi}_T^{*(1)} (\hat{\theta}_T)\|$ , which is  $O_{P^*}(1)$  in prob- $P$ , thus concluding the proof of part (i). To prove part (ii), note the second order optimality condition for an interior solution of a minimization problem implies that, for any vector  $e \in \mathbb{R}^p$ ,

$$e' \frac{\partial^2}{\partial \theta \partial \theta'} \left[ \bar{\psi}_T^{*(1)'} (\theta) W_T^{*(1)} \bar{\psi}_T^{*(1)} (\theta) \right] \Big|_{\theta = \hat{\theta}_T^{*(1)}} e \geq 0.$$

This can be written as

$$e' \left( \tilde{Z}_T^* + N_T^* \right) e \geq 0, \quad (24)$$

where

$$\tilde{Z}_T^* = \left( \frac{\partial^2 \bar{\psi}_T^{*(1)}}{\partial \theta_i \partial \theta_j} (\hat{\theta}_T^{*(1)}) W_T^{*(1)} \sqrt{T} \bar{\psi}_T^{*(1)} (\hat{\theta}_T^{*(1)}) \right)_{1 \leq i, j \leq p} \quad \text{and} \quad N_T^* = \sqrt{T} \frac{\partial \bar{\psi}_T^{*(1)'}}{\partial \theta} (\hat{\theta}_T^{*(1)}) W_T^{*(1)} \frac{\partial \bar{\psi}_T^{*(1)}}{\partial \theta'} (\hat{\theta}_T^{*(1)}).$$

From part (i), we can show that

$$\begin{aligned} \sqrt{T} \bar{\psi}_T^{*(1)} (\hat{\theta}_T^{*(1)}) &= \sqrt{T} \bar{\psi}_T^{*(1)} (\hat{\theta}_T) + \frac{1}{2} G \text{vec} (\hat{v}_T^{*(1)} \hat{v}_T^{*(1)'}) + o_{P^*}(1), \quad \text{and} \\ T^{1/4} \frac{\partial \bar{\psi}_T^{*(1)}}{\partial \theta_i} (\hat{\theta}_T^{*(1)}) &= \frac{\partial^2 \rho}{\partial \theta_i \partial \theta'} (\theta_0) T^{1/4} (\hat{\theta}_T^{*(1)} - \hat{\theta}_T) + o_{P^*}(1), \end{aligned}$$

in prob- $P$ , for  $i = 1, \dots, p$ . The last equality follows from the mean-value expansion of  $\partial \bar{\psi}_T^{*(1)} (\hat{\theta}_T^{*(1)}) / \partial \theta_i$  around  $\hat{\theta}_T$ ; the uniform convergences of  $\bar{G}^*(\theta)$  and  $\bar{G}(\theta)$  in a neighbourhood of  $\theta_0$  then allows to replace the sample means in the second derivatives by the population mean  $\rho(\cdot)$ . The uniform convergence argument also implies that  $\partial^2 \bar{\psi}_T^{*(1)} (\hat{\theta}_T^{*(1)}) / \partial \theta_i \partial \theta_j = \partial^2 \rho(\theta_0) / \partial \theta_i \partial \theta_j + o_{P^*}(1)$ , in prob- $P$ . Hence,

$$\begin{aligned} \tilde{Z}_T^* &= Z_T^* + \frac{1}{2} \left( \frac{\partial^2 \rho'}{\partial \theta_i \partial \theta_j} (\theta_0) W \text{vec} (\hat{v}_T^{*(1)} \hat{v}_T^{*(1)'}) \right)_{1 \leq i, j \leq p} + o_{P^*}(1) \quad \text{and} \\ N_T^* &= \left( \hat{v}_T^{*(1)' } \frac{\partial^2 \rho'}{\partial \theta_i \partial \theta} (\theta_0) W \frac{\partial^2 \rho}{\partial \theta_i \partial \theta'} (\theta_0) \hat{v}_T^{*(1)} \right)_{1 \leq i, j \leq p} + o_{P^*}(1), \end{aligned}$$

where  $Z_T^* = \left( \frac{\partial^2 \rho'}{\partial \theta_i \partial \theta_j} (\theta_0) W \sqrt{T} \bar{\psi}_T^{*(1)} (\hat{\theta}_T) \right)_{1 \leq i, j \leq p}$ . From (24) and some successive applications of the Cauchy-Schwarz inequality, we can claim that there exists  $A > 0$  such that for any (unit) vector  $e \in \mathbb{R}^p$ ,

$$-e' Z_T^* e - A \left\| \hat{v}_T^{*(1)} \right\|^2 \leq o_{P^*}(1).$$

Since  $\left( \sqrt{T} \bar{\psi}_T^{*(1)' } (\hat{\theta}_T), \hat{v}_T^{*(1)' } \right)' = O_{P^*}(1)$ , in prob- $P$ , it follows that  $\left( Z_T^*, \hat{v}_T^{*(1)} \right) = O_{\mathbb{P}}(1)$  (by Lemma .1). Thus, by Prohorov's Theorem (cf. Theorem 2.4 of van der Vaart (1998)), we can find a subsequence of  $\left( Z_T^*, \hat{v}_T^{*(1)} \right)$ ,  $\left( Z_{T'}^*, \hat{v}_{T'}^{*(1)} \right)$  say, which converges in distribution to  $(Z^*, V^*)$  under  $\mathbb{P}$ . Consequently, by the continuous mapping theorem,

$$Y_{T'}^* \equiv -e' Z_{T'}^* e - A \left\| \hat{v}_{T'}^{*(1)} \right\|^2 \xrightarrow{d_{\mathbb{P}}} Y^* \equiv -e' Z^* e - A \|V^*\|^2,$$

along this subsequence. This means that for any metric  $d$  metrizing weak convergence,  $d(\mathcal{L}(Y_{T'}^*), \mathcal{L}(Y^*)) \xrightarrow{\mathbb{P}} 0$ , where  $\mathcal{L}(\cdot)$  denotes the law of the random variable in question. A second application of Lemma .1 implies that  $d(\mathcal{L}(Y_{T'}^*), \mathcal{L}(Y^*)) \xrightarrow{P^*} 0$  in prob- $P$  along the subsequence indexed by  $T'$ . But this is equivalent to saying that there is a further subsequence  $Y_{T''}^*$  of  $Y_{T'}^*$  for which  $d(\mathcal{L}(Y_{T''}^*), \mathcal{L}(Y^*)) \xrightarrow{P^*} 0$  a.s.- $P$ . Fix  $\omega$  in the probability set on which this event occurs (whose probability  $P$  is one). By the same argument as used by D&R (2013) in their proof of Proposition 3.2, we can conclude that  $v_{T''}^*$  converges in distribution to  $V^*$  with  $P(\|V^*\| > 0) \text{geq} 1 - P(Z^* \geq 0) \geq 1/2 \equiv \delta$  for all  $\omega$  in a set of probability one. This completes the proof of part (ii).

**Proof of Lemma 4.1.** Letting  $Y_T^*(v) \equiv \sqrt{T} \bar{\psi}_T^{*(1)} (\hat{\theta}_T + T^{-1/4} v)$ , with  $v = T^{1/4} (\theta - \hat{\theta}_T)$ , and

noting that  $J_T^{*(1)}(v) = Y_T^*(v)' W_T^{*(1)} Y_T^*(v)$ , by the Continuous Mapping Theorem (see Pollard (1984, p. 70)) it suffices to show that

$$\left( W_T^{*(1)}, Y_T^*(v) \right) \Rightarrow^{P^*} (W, Y^*(v)), \text{ in } \ell^\infty(\mathbb{K}), \text{ in prob-}P, \quad (25)$$

where  $Y^*(v) \equiv X^* + \frac{1}{2} \text{Gvec}(vv')$  with  $X^* \sim N(0, \Sigma)$ , implying that  $Y^*(v) =^d Y(v) \equiv X + \frac{1}{2} \text{Gvec}(vv')$ . Since  $W_T^{*(1)} = W + o_{P^*}(1)$  in prob- $P$ ,  $W_T^{*(1)} \rightarrow^{d^*} W$ , in prob- $P$  and since it does not depend on  $v$ , it also converges weakly towards  $W$  in  $\ell^\infty(\mathbb{K})$ , in prob- $P$ . Given that  $W$  is constant, Slutsky's theorem (Korosok (2008, Theorem 7.5)) ensures that (25) holds once we show that  $Y_T^*(v) \Rightarrow^{P^*} Y^*(v)$ , in  $\ell^\infty(\mathbb{K})$ , in prob- $P$ . To establish this, observe that as a compact subset of  $\mathbb{R}^p$ ,  $\mathbb{K}$  equipped with the usual metric is totally bounded. Following the proof of Giné and Zinn (1990, Theorem 3.1), it remains to show that

(a)  $(Y_T^*(v_1), \dots, Y_T^*(v_k)) \rightarrow^{d^*} (Y^*(v_1), \dots, Y^*(v_k))$ , in prob- $P$ , for any  $v_1, \dots, v_k \in \mathbb{K}$ .

(b) For any  $\epsilon > 0$ , there exists  $\delta > 0$  sufficiently small such that

$$\limsup_{T \rightarrow \infty} P \left( E^* \left( \sup_{\|v_1 - v_2\| < \delta, v_1, v_2 \in \mathbb{K}} \|Y_T^*(v_1) - Y_T^*(v_2)\| \right) > \epsilon \right) = 0.$$

Starting with (a), by a second-order mean value expansion of  $Y_T^*(v)$  around 0, we have that

$$Y_T^*(v) = \sqrt{T} \bar{\psi}_T^{*(1)}(\hat{\theta}_T) + T^{1/4} \frac{\partial \bar{\psi}_T^{*(1)}}{\partial \theta'}(\hat{\theta}_T) v + \frac{1}{2} \bar{G}^{*(1)}(\ddot{\theta}^*(v)) \text{vec}(vv'),$$

where  $\ddot{\theta}^*(v) \in (\hat{\theta}_T, \hat{\theta}_T + T^{-1/4}v)$  and may differ from row to row; the matrix  $\bar{G}^{*(1)}(\theta)$  is defined as in (4) with  $\bar{\psi}_T^{*(1)}(\theta)$  in place of  $\bar{\psi}_T(\theta)$ . We can write

$$Y_T^*(v) = \sqrt{T} \bar{\psi}_T^{*(1)}(\hat{\theta}_T) + \frac{1}{2} \text{Gvec}(vv') + r_T^*(v),$$

where  $r_T^*(v) = r_{1T}^*(v) + r_{2T}^*(v) + r_{3T}^*(v)$  with

$$\begin{aligned} r_{1T}^*(v) &= \frac{1}{2} \left( \bar{G}^{*(1)}(\hat{\theta}_T) - G \right) \text{vec}(vv'), \\ r_{2T}^*(v) &= \frac{1}{2} \left( \bar{G}(\ddot{\theta}^*(v)) - \bar{G}^{*(1)}(\hat{\theta}_T) \right) \text{vec}(vv') \text{ and } r_{3T}^*(v) = T^{1/4} \frac{\partial \bar{\psi}_T^{*(1)}}{\partial \theta'}(\hat{\theta}_T) v. \end{aligned}$$

We can show that for  $j = 1, 2, 3$ ,  $r_{jT}^*(v)$  is  $o_{P^*}(1)$  in prob- $P$  uniformly over  $\mathbb{K}$ . For  $j = 1$ , note that

$$\bar{G}^{*(1)}(\hat{\theta}_T) - G = \bar{G}^{*(1)}(\hat{\theta}_T) - \bar{G}(\hat{\theta}_T) + \bar{G}(\hat{\theta}_T) - G,$$

where we can show that  $\bar{G}^{*(1)}(\hat{\theta}_T) - \bar{G}(\hat{\theta}_T) = o_{P^*}(1)$  in prob- $P$  and  $\bar{G}(\hat{\theta}_T) - G = o_P(1)$  (for the first result, we rely on the fact that  $\sup_{\theta \in \Theta} \|\bar{G}^{*(1)}(\theta) - \bar{G}(\theta)\| = o_{P^*}(1)$  in prob- $P$  whereas the second result follows from the uniform convergence of  $\bar{G}(\theta)$  towards  $G(\theta)$ , the fact that  $\hat{\theta}_T \rightarrow^P \theta_0$ , and the definition  $G \equiv G(\theta_0)$ ). Since  $v \in \mathbb{K}$ , a compact subset of  $\mathbb{R}^p$ , this proves the result for  $j = 1$ . For  $j = 2$ , the result follows from the uniform convergence of  $\bar{G}^{*(1)}(\theta)$  towards  $\bar{G}(\theta)$  and the fact



that  $\hat{\theta}_T^{*(1)} - \ddot{\theta}_T^* \rightarrow^{P^*} 0$ , in prob- $P$  (see e.g. Lemma A.6 of Gonçalves and White (2004)). Finally,  $r_{3T}^*(v) = o_{P^*}(1)$  in prob- $P$  uniformly over  $\mathbb{K}$  because  $T^{1/2} \frac{\partial \bar{\psi}_T^{*(1)}}{\partial \theta'}(\hat{\theta}_T) = O_{P^*}(1)$  by a bootstrap CLT applied to  $T^{1/2} \frac{\partial \bar{\psi}_T^{*(1)}}{\partial \theta'}(\hat{\theta}_T)$  (note that its mean is zero given the double recentering). Since  $\sqrt{T} \bar{\psi}_T^{*(1)}(\hat{\theta}_T) \xrightarrow{d^*} X^* \sim N(0, \Sigma)$ , in prob- $P$ , we conclude that  $Y_T^*(v) \rightarrow^{d^*} X^* + \frac{1}{2} G \text{vec}(vv')$   $\equiv Y^*(v)$ , in prob- $P$  for any fixed  $v \in \mathbb{K}$ . Similarly, for any fixed  $(v_1, v_2, \dots, v_k) \in \mathbb{K}^k$ , consider  $c = (c_1, \dots, c_k) \in \mathbb{R}^k$  such that  $c'c = 1$ . It follows that

$$\sum_{j=1}^k c_j Y_T^*(v_j) = c' \sqrt{T} \bar{\psi}_T^{*(1)}(\hat{\theta}_T) + \frac{1}{2} G \sum_{j=1}^k c_j \text{vec}(v_j v_j') + \sum_{j=1}^k c_j r_{3T}^*(v_j),$$

where the last term is  $o_{P^*}(1)$  in prob- $P$  uniformly over  $\mathbb{K}$  by the same arguments as those used above. It follows that  $\sum_{j=1}^k c_j Y_T^*(v_j) \rightarrow^{d^*} c' X^* + \frac{1}{2} G \sum_{j=1}^k c_j \text{vec}(v_j v_j')$  in prob- $P$  and the result follows by the Cramer-Wold device.

Next, we establish (b). From the second-order expansion of  $Y_T^*(v)$  above, we can write

$$\begin{aligned} Y_T^*(v_1) - Y_T^*(v_2) &= T^{1/4} \frac{\partial \bar{\psi}_T^{*(1)}}{\partial \theta'}(\hat{\theta}_T) (v_1 - v_2) + \frac{1}{2} \bar{G}^{*(1)}(\ddot{\theta}^*(v_1)) \text{vec}(v_1 v_1') + \frac{1}{2} \bar{G}^{*(1)}(\ddot{\theta}^*(v_2)) \text{vec}(v_2 v_2') \\ &= T^{1/4} \frac{\partial \bar{\psi}_T^{*(1)}}{\partial \theta'}(\hat{\theta}_T) (v_1 - v_2) + \frac{1}{2} \bar{G}^{*(1)}(\ddot{\theta}^*(v_1)) [\text{vec}(((v_1 - v_2) v_1' - v_2 (v_1 - v_2)'))] \\ &\quad + \frac{1}{2} [\bar{G}^{*(1)}(\ddot{\theta}^*(v_2)) - \bar{G}^{*(1)}(\ddot{\theta}^*(v_1))] \text{vec}(v_2 v_2'). \end{aligned}$$

Hence,

$$\begin{aligned} \|Y_T^*(v_1) - Y_T^*(v_2)\| &\leq \left\| T^{1/4} \frac{\partial \bar{\psi}_T^{*(1)}}{\partial \theta'}(\hat{\theta}_T) \right\| \|v_1 - v_2\| + \frac{1}{2} \left\| \bar{G}^{*(1)}(\ddot{\theta}^*(v_1)) \right\| \|v_1 - v_2\| (\|v_1\| + \|v_2\|) \\ &\quad + \frac{1}{2} \left\| \bar{G}^{*(1)}(\ddot{\theta}^*(v_2)) - \bar{G}^{*(1)}(\ddot{\theta}^*(v_1)) \right\| \|v_2\|^2. \end{aligned}$$

Since  $\mathbb{K}$  is compact, there exists  $M > 0$  such that

$$\begin{aligned} \|Y_T^*(v_1) - Y_T^*(v_2)\| &\leq \left\| T^{1/4} \frac{\partial \bar{\psi}_T^{*(1)}}{\partial \theta'}(\hat{\theta}_T) \right\| \|v_1 - v_2\| + M \left\| \bar{G}^{*(1)}(\ddot{\theta}^*(v_1)) \right\| \|v_1 - v_2\| \\ &\quad + M \left\| \bar{G}^{*(1)}(\ddot{\theta}^*(v_2)) - \bar{G}^{*(1)}(\ddot{\theta}^*(v_1)) \right\|, \end{aligned}$$

implying that

$$\begin{aligned} E^* \left( \sup_{\|v_1 - v_2\| < \delta, v_1, v_2 \in \mathbb{K}} \|Y_T^*(v_1) - Y_T^*(v_2)\| \right) &\leq E^* \left\| T^{1/4} \frac{\partial \bar{\psi}_T^{*(1)}}{\partial \theta'}(\hat{\theta}_T) \right\| \delta + M E^* \left\| \bar{G}^{*(1)}(\ddot{\theta}^*(v_1)) \right\| \delta \\ &\quad + M E^* \left\| \bar{G}^{*(1)}(\ddot{\theta}^*(v_2)) - \bar{G}^{*(1)}(\ddot{\theta}^*(v_1)) \right\|. \end{aligned}$$

Thus, for any  $\epsilon > 0$ , there exists  $\delta > 0$  such that

$$\begin{aligned} & P \left[ E^* \left( \sup_{\|v_1 - v_2\| < \delta, v_1, v_2 \in \mathbb{K}} \|Y_T^*(v_1) - Y_T^*(v_2)\| \right) > \epsilon \right] \\ & \leq P \left( E^* \left\| T^{1/4} \frac{\partial \bar{\psi}_T^{*(1)}}{\partial \theta'} (\hat{\theta}_T) \right\| > \frac{\epsilon}{3\delta} \right) + P \left( E^* \left\| \bar{G}^{*(1)} (\ddot{\theta}^*(v_1)) \right\| > \frac{\epsilon}{3M\delta} \right) \\ & \quad + P \left( E^* \left\| \bar{G}^{*(1)} (\ddot{\theta}^*(v_2)) - \bar{G}^{*(1)} (\ddot{\theta}^*(v_1)) \right\| > \frac{\epsilon}{3M} \right) \end{aligned}$$

can be made arbitrarily small as  $T \rightarrow \infty$ . In particular, it is sufficient to show that

$$(b1) \ E^* \left( \left\| T^{1/4} \frac{\partial \bar{\psi}_T^{*(1)}}{\partial \theta'} (\hat{\theta}_T) \right\| \right) = O_P(1),$$

$$(b2) \ E^* \left\| \bar{G}^{*(1)} (\ddot{\theta}^*(v_1)) \right\| = O_P(1), \text{ and}$$

(b3)  $E^* \left\| \bar{G}^{*(1)} (\ddot{\theta}^*(v_2)) - \bar{G}^{*(1)} (\ddot{\theta}^*(v_1)) \right\| = o_P(1)$ , uniformly on  $\mathbb{K}$ . Starting with (b1), by Jensen's inequality,

$$E^* \left( \left\| T^{1/4} \frac{\partial \bar{\psi}_T^{*(1)}}{\partial \theta'} (\hat{\theta}_T) \right\| \right) \leq \left( E^* \left( \left\| T^{1/4} \frac{\partial \bar{\psi}_T^{*(1)}}{\partial \theta'} (\hat{\theta}_T) \right\|^2 \right) \right)^{1/2}.$$

But,

$$E^* \left( \left\| T^{1/4} \frac{\partial \bar{\psi}_T^{*(1)}}{\partial \theta'} (\hat{\theta}_T) \right\|^2 \right) = T^{1/2} E^* \left( \sum_{\substack{h=1, \dots, H \\ j=1, \dots, p}} \left( \frac{\partial \bar{\psi}_{h,T}^{*(1)}}{\partial \theta_j} (\hat{\theta}_T) \right)^2 \right) = T^{1/2} \sum_{\substack{h=1, \dots, H \\ j=1, \dots, p}} E^* \left( \frac{\partial \bar{\psi}_{h,T}^{*(1)}}{\partial \theta_j} (\hat{\theta}_T) \right)^2,$$

where  $\bar{\psi}_{h,T}^{*(1)}$  is the  $h$ -th component of  $\bar{\psi}_T^{*(1)}$ . Since

$$\frac{\partial \bar{\psi}_{h,T}^{*(1)}}{\partial \theta_j} (\hat{\theta}_T) = \frac{1}{T} \sum_{t=1}^T \left( \frac{\partial \psi_h}{\partial \theta_j} (X_t^*, \hat{\theta}_T) - \frac{\partial \bar{\psi}_{h,T}}{\partial \theta_j} (\hat{\theta}_T) \right),$$

where  $X_t^*$  is i.i.d. from  $\{X_1, \dots, X_n\}$ , we have that

$$E^* \left( \frac{\partial \bar{\psi}_{h,T}^{*(1)}}{\partial \theta_j} (\hat{\theta}_T) \right)^2 = \frac{1}{T^2} \sum_{t=1}^T \left( \frac{\partial \psi_h}{\partial \theta_j} (X_t, \hat{\theta}_T) - \frac{\partial \bar{\psi}_{h,T}}{\partial \theta_j} (\hat{\theta}_T) \right)^2 = O_P(T^{-1}).$$

Thus,  $E^* \left( \left\| T^{1/4} \frac{\partial \bar{\psi}_T^{*(1)}}{\partial \theta'} (\hat{\theta}_T) \right\|^2 \right) = O_P(T^{-1/2}) = o_P(1)$  (uniformly on  $\mathbb{K}$  since it does not depend on  $v \in \mathbb{K}$ ), showing (b1). For (b2), for some constant  $C < \infty$ ,

$$\left\| \bar{G}^{*(1)} (\ddot{\theta}^*(v_1)) \right\| \leq C \sum_{h=1}^H \left\| \frac{\partial^2 \bar{\psi}_h^{*(1)} (\ddot{\theta}^*(v_1))}{\partial \theta \partial \theta'} \right\| \leq C \sum_{h=1}^H \frac{1}{T} \sum_{t=1}^T \sup_{\theta \in \mathcal{N}} \left\| \frac{\partial^2 \psi_h (X_t^*, \theta)}{\partial \theta \partial \theta'} \right\|^2,$$

where the last inequality uses the fact that  $\ddot{\theta}^*(v_1) \in (\hat{\theta}_T, \hat{\theta}_T + T^{-1/4}v)$ , which is included in  $\mathcal{N}$ , a neighborhood of  $\theta_0$ , for all  $T$  sufficiently large with probability  $P$  approaching one (given that

$\hat{\theta}_T \rightarrow^P \theta_0$ ). Hence,

$$E^* \left\| \bar{G}^{*(1)} \left( \ddot{\theta}^*(v_1) \right) \right\| \leq \sum_{h=1}^H \frac{1}{T} \sum_{t=1}^T E^* \left( \sup_{\theta \in \mathcal{N}} \left\| \frac{\partial^2 \psi_h(X_t^*, \theta)}{\partial \theta \partial \theta'} \right\| \right) = \sum_{h=1}^H \frac{1}{T} \sum_{t=1}^T \sup_{\theta \in \mathcal{N}} \left\| \frac{\partial^2 \psi_h(X_t, \theta)}{\partial \theta \partial \theta'} \right\|,$$

which is  $O_P(1)$  uniformly on  $\mathbb{K}$  given that  $\frac{\partial^2 \psi_h(X_t, \theta)}{\partial \theta \partial \theta'}$  is Lipschitz continuous on  $\mathcal{N}$  and  $E \left\| \frac{\partial^2 \psi_h(X_t, \theta_0)}{\partial \theta \partial \theta'} \right\| < \infty$ . Similarly, we have that

$$\begin{aligned} \left\| \bar{G}^{*(1)}(\ddot{\theta}^*(v_1)) - \bar{G}^{*(1)}(\ddot{\theta}^*(v_2)) \right\| &\leq C \sum_{h=1}^H \frac{1}{T} \sum_{t=1}^T \left\| \frac{\partial^2 \psi_h^{(1)}(X_t^*, \ddot{\theta}^*(v_1))}{\partial \theta \partial \theta'} - \frac{\partial^2 \psi_h^{(1)}(X_t^*, \ddot{\theta}^*(v_2))}{\partial \theta \partial \theta'} \right\| \\ &\leq C \sum_{h=1}^H \frac{1}{T} \sum_{t=1}^T m(X_t^*) \left\| \ddot{\theta}^*(v_1) - \ddot{\theta}^*(v_2) \right\| \leq \sum_{h=1}^H \left( \frac{1}{T} \sum_{t=1}^T m(X_t^*) \right) T^{-1/4} C, \end{aligned}$$

where the last inequality uses the fact that  $\sup_{v_1, v_2 \in \mathbb{K}} \left\| \ddot{\theta}^*(v_1) - \ddot{\theta}^*(v_2) \right\| \leq CT^{-1/4}$  for some constant  $C < \infty$  given the definitions of  $\ddot{\theta}^*(v_1)$  and  $\ddot{\theta}^*(v_2)$ . Because  $X_t^*$  is i.i.d. on  $\{X_1, \dots, X_n\}$ ,  $E^* \left( \frac{1}{T} \sum_{t=1}^T m(X_t^*) \right) = \frac{1}{T} \sum_{t=1}^T m(X_t) = O_P(1)$  given Assumption 4 (i) and we conclude that

$$E^* \left\| \bar{G}^{*(1)}(\ddot{\theta}^*(v_1)) - \bar{G}^{*(1)}(\ddot{\theta}^*(v_2)) \right\| \leq \sum_{h=1}^H \left( \frac{1}{T} \sum_{t=1}^T E^*(m(X_t^*)) \right) T^{-1/4} C = O_P(T^{-1/4}) = o_P(1),$$

uniformly in  $v_1$  and  $v_2$ .

**Proof of Theorem 4.1.** Part (i) follows by Lemmas 4.1 and .2. In particular,  $\hat{v} \in \arg \min_{v \in \mathbb{R}^p} J(v) = O_P(1)$  by Lemma B.6 (ii) of D&R (2013). Similarly,  $\hat{v}_T^{*(1)} = T^{1/4} \left( \hat{\theta}_T^{*(1)} - \hat{\theta}_T \right) \in \arg \min_{v \in \mathbb{V}_T} J_T^{*(1)}(v)$  is  $O_{P^*}(1)$  in prob- $P$  by Proposition 4.2 (i). Note that we can show that  $\hat{v}_T^{*(1)}$  is also the minimizer of  $J_T^{*(1)}(v)$  over  $\mathbb{R}^p$  since  $\hat{\theta}_T \rightarrow^P \theta_0$ , an interior point of  $\Theta$ , which implies that with probability  $P$  approaching one, the union of  $\mathbb{V}_T$  over all  $T \geq 1$  covers  $\mathbb{R}^p$ . Part (ii) follows from Polya's Theorem given that  $J(v)$  is continuous from Theorem 2.1.

**Proof of Proposition 4.3.** The proof follows by the same arguments as the proof of Proposition 3.1. In particular, under the new global identification (Assumption 7),  $\theta_0$  is the unique  $\arg \min_{\theta} Q^{(2)}(\theta)$ , where

$$Q^{(2)}(\theta) = E \left( \Phi'(X_1, \theta) \right) W E \left( \Phi(X_1, \theta) \right).$$

Thus, the result follows if we show that with probability  $P$  converging to 1,

$$\sup_{\theta \in \Theta} \left| Q_T^{*(2)}(\theta) - Q^{(2)}(\theta) \right| \rightarrow^{P^*} 0. \quad (26)$$

Letting  $Q_T^{(2)}(\theta) = \bar{\Phi}'_T(\theta) W_T \bar{\Phi}_T(\theta)$ , where  $\bar{\Phi}_T(\theta) = \bar{\psi}_T(\theta) - \frac{\partial}{\partial \theta'} \bar{\psi}_T(\theta) (\theta - \theta_0)$ , (26) follows from: (A)  $\sup_{\theta \in \Theta} \left| Q_T^{*(2)}(\theta) - Q_T^{(2)}(\theta) \right| \rightarrow^{P^*} 0$ , in prob- $P$ ; and (B)  $\sup_{\theta \in \Theta} \left| Q_T^{(2)}(\theta) - Q^{(2)}(\theta) \right| \rightarrow^P 0$ . We can

show that

$$\begin{aligned}\bar{\psi}_T^{*(2)}(\theta) &= \bar{\psi}_T^*(\theta) - E^* \left( \bar{\psi}_T^* \left( \hat{\theta}_T \right) \right) - E^* \left( \frac{\partial}{\partial \theta'} \bar{\psi}_T^*(\theta) \right) (\theta - \hat{\theta}_T) \\ &= \bar{\psi}_{c,T}^*(\theta) - \left( \frac{\partial}{\partial \theta'} \bar{\psi}_T^*(\theta) \right) (\theta - \hat{\theta}_T),\end{aligned}$$

where  $\bar{\psi}_{c,T}^*(\theta) \equiv \bar{\psi}_T^*(\theta) - E^* \left( \bar{\psi}_T^* \left( \hat{\theta}_T \right) \right) = \bar{\psi}_T^*(\theta) - \bar{\psi}_T^* \left( \hat{\theta}_T \right)$ . This implies that

$$\begin{aligned}Q_T^{*(2)}(\theta) &= \bar{\psi}_T^{*(2)'}(\theta) W_T^{*(2)} \bar{\psi}_T^{*(2)}(\theta) \\ &= Q_T^*(\theta) - 2\bar{\psi}_{c,T}^{*(2)'}(\theta) W_T^{*(2)} \frac{\partial}{\partial \theta'} \bar{\psi}_T^*(\theta) (\theta - \hat{\theta}_T) \\ &\quad + (\theta - \hat{\theta}_T)' \left( \frac{\partial}{\partial \theta'} \bar{\psi}_T^*(\theta) \right)' W_T^{*(2)} \frac{\partial}{\partial \theta'} \bar{\psi}_T^*(\theta) (\theta - \hat{\theta}_T) \\ &\equiv Q_T^*(\theta) + R_{1T}^*(\theta) + R_{2T}^*(\theta),\end{aligned}$$

where  $Q_T^*(\theta)$  is as defined in Proposition 3.1 but using the weighting matrix  $W_T^{*(2)}$ . Similarly, we can write

$$\begin{aligned}Q_T^{(2)}(\theta) &= \bar{\Phi}'_T(\theta) W_T \bar{\Phi}_T(\theta) \\ &= \bar{\psi}_T(\theta)' W_T \bar{\psi}_T(\theta) - 2\bar{\psi}_T(\theta)' W_T \frac{\partial}{\partial \theta'} \bar{\psi}_T(\theta) (\theta - \theta_0) \\ &\quad + (\theta - \theta_0)' \left( \frac{\partial}{\partial \theta'} \bar{\psi}_T(\theta) \right)' W_T \frac{\partial}{\partial \theta'} \bar{\psi}_T(\theta) (\theta - \theta_0) \\ &\equiv Q_T(\theta) + R_{1T}(\theta) + R_{2T}(\theta).\end{aligned}$$

Thus, to show (A), it suffices to show that

$$\sup_{\theta \in \Theta} |Q_T^*(\theta) - Q_T(\theta)| = o_{P^*}(1), \text{ in prob-}P; \quad (27)$$

$$\sup_{\theta \in \Theta} |R_{1T}^*(\theta) - R_{1T}(\theta)| = o_{P^*}(1), \text{ in prob-}P; \text{ and} \quad (28)$$

$$\sup_{\theta \in \Theta} |R_{2T}^*(\theta) - R_{2T}(\theta)| = o_{P^*}(1), \text{ in prob-}P. \quad (29)$$

Proposition 3.1 implies (27) whereas we can show that (28) and (29) follow by relying on the fact that  $\hat{\theta}_T \rightarrow^P \theta_0$ ; and that  $\sup_{\theta \in \Theta} \left| \bar{\psi}_{c,T}^*(\theta) - \bar{\psi}_T^*(\theta) \right| = o_{P^*}(1)$  and  $W_T^{*(2)} - W_T = o_{P^*}(1)$ , in prob- $P$ .

To prove (B), let  $\mu(\theta) = E(\psi(X_1, \theta))$  and  $d(\theta) = E \left( \frac{\partial \psi(X_1, \theta)}{\partial \theta'} \right)$ . We have that

$$\begin{aligned}Q_T^{(2)}(\theta) - Q^{(2)}(\theta) &= \left[ \bar{\psi}'_T(\theta) W_T \bar{\psi}_T(\theta) - \mu'(\theta) W \mu(\theta) \right] - 2 \left[ \bar{\psi}'_T(\theta) W_T \frac{\partial \bar{\psi}_T(\theta)}{\partial \theta'} - \mu'(\theta) W d(\theta) \right] (\theta - \theta_0) \\ &\quad + (\theta - \theta_0)' \left[ \frac{\partial \bar{\psi}'_T(\theta)}{\partial \theta} W_T \frac{\partial \bar{\psi}_T(\theta)}{\partial \theta'} - d'(\theta) W d(\theta) \right] (\theta - \theta_0) \\ &\equiv [1] + [2] + [3].\end{aligned}$$

By definition,  $[1] = Q_T(\theta) - Q(\theta) = o_P(1)$  uniformly over  $\Theta$  (see proof of Proposition 3.1). Moreover,

$$\begin{aligned} \sup_{\theta \in \Theta} \left| \frac{1}{2}[2] \right| &\leq (\sup_{\theta \in \Theta} \|\theta\| + \|\theta_0\|) \times \left( \sup_{\theta \in \Theta} \|\bar{\psi}_T(\theta) - \mu(\theta)\| \|W_T\| \frac{1}{T} \sum_{t=1}^T \sup_{\theta \in \Theta} \left\| \frac{\partial \bar{\psi}_T(\theta)}{\partial \theta'} \right\| \right. \\ &\quad + \sup_{\theta \in \Theta} \|\mu(\theta)\| \|W_T - W\| \frac{1}{T} \sum_{t=1}^T \sup_{\theta \in \Theta} \left\| \frac{\partial \bar{\psi}_T(\theta)}{\partial \theta'} \right\| \\ &\quad \left. + \sup_{\theta \in \Theta} \|\mu(\theta)\| \|W\| \sup_{\theta \in \Theta} \left\| \frac{\partial \bar{\psi}_T(\theta)}{\partial \theta'} - d(\theta) \right\| \right). \end{aligned}$$

We can show that  $[2] = o_P(1)$  uniformly on  $\Theta$  by relying in particular on the fact that  $\sup_{\theta \in \Theta} \|\bar{\psi}_T(\theta) - \mu(\theta)\| = o_P(1)$ ,  $\|W_T - W\| = o_P(1)$ , and  $\sup_{\theta \in \Theta} \left\| \frac{\partial \bar{\psi}_T(\theta)}{\partial \theta'} - d(\theta) \right\| = o_P(1)$  under our assumptions. The proof that  $[3] = o_P(1)$  uniformly on  $\Theta$  follows by similar arguments once we show that  $\frac{\partial \bar{\psi}'_T(\theta)}{\partial \theta} W_T \frac{\partial \bar{\psi}_T(\theta)}{\partial \theta'} - d'(\theta) W d(\theta) = o_P(1)$  uniformly on  $\Theta$ .

To prove Theorem 4.2, we first prove the two following auxiliary results, which are the continuously-corrected bootstrap analogues of Proposition 4.2 and Lemma 4.1, respectively.

**Lemma .3** *Under Assumptions 1-8, if  $W_T^{*(2)} \rightarrow^{P^*} W$  in prob- $P$ , (i)  $\left\| \hat{\theta}_T^{*(2)} - \hat{\theta}_T \right\| = O_{P^*}(T^{-1/4})$  in prob- $P$ , and (ii)  $T^{1/4} \left( \hat{\theta}_T^{*(2)} - \hat{\theta}_T \right)$  has at least a subsequence that converges in distribution to some random variable  $V^*$  under  $P^*$ , a.s. - $P$ , such that for some  $\delta > 0$ ,  $P(\|V^*\| \neq 0) \geq \delta$ .*

**Lemma .4** *Under Assumptions 1-8, if  $W_T^{*(2)} \rightarrow^{P^*} W$  in prob- $P$ , we have that  $J_T^{*(2)}(v) \Rightarrow^{P^*} J(v)$  in  $\ell^\infty(\mathbb{K})$ , in prob- $P$ , where  $J_T^{*(2)}(v)$  is defined as  $J_T^{*(1)}(v)$  but with  $\bar{\psi}_T^{*(1)}$  and  $W_T^{*(1)}$  replaced with  $\bar{\psi}_T^{*(2)}$  and  $W_T^{*(2)}$ , respectively.*

**Proof of Proposition .3.** We follow the proof of Proposition 4.2. First, note that for all  $i, j = 1, \dots, p$ ,

$$\frac{\partial}{\partial \theta_i} \bar{\psi}_T^{*(2)}(\theta) = \frac{\partial}{\partial \theta_i} \bar{\psi}_T^*(\theta) - \frac{\partial}{\partial \theta_i} \left( \frac{\partial}{\partial \theta'} \bar{\psi}_T(\theta) \right) (\theta - \hat{\theta}_T) - \frac{\partial}{\partial \theta_i} \bar{\psi}_T(\theta)$$

and

$$\frac{\partial^2}{\partial \theta_i \partial \theta_j} \bar{\psi}_T^{*(2)}(\theta) = \frac{\partial^2}{\partial \theta_i \partial \theta_j} \bar{\psi}_T^*(\theta) - \frac{\partial^2}{\partial \theta_i \partial \theta_j} \left( \frac{\partial}{\partial \theta'} \bar{\psi}_T(\theta) \right) (\theta - \hat{\theta}_T) - 2 \frac{\partial^2}{\partial \theta_i \partial \theta_j} \bar{\psi}_T(\theta).$$

Hence,

$$\bar{\psi}_T^{*(2)}(\hat{\theta}_T) = \bar{\psi}_T^{*(1)}(\hat{\theta}_T) \quad \text{and} \quad \frac{\partial}{\partial \theta_i} \bar{\psi}_T^{*(2)}(\hat{\theta}_T) = \frac{\partial}{\partial \theta_i} \bar{\psi}_T^{*(1)}(\hat{\theta}_T) \quad \text{for } i = 1, \dots, p.$$

By a second-order mean value expansion of  $\bar{\psi}_T^{*(2)}(\hat{\theta}_T^{*(2)})$  around  $\hat{\theta}_T$ , we can then write

$$\sqrt{T} \bar{\psi}_T^{*(2)}(\hat{\theta}_T^{*(2)}) = \sqrt{T} \bar{\psi}_T^{*(1)}(\hat{\theta}_T) + \sqrt{T} \frac{\partial \bar{\psi}_T^{*(1)}}{\partial \theta'}(\hat{\theta}_T) (\hat{\theta}_T^{*(2)} - \hat{\theta}_T) + \frac{1}{2} R_T^{*(2)}(\bar{\theta}_T^*),$$

where  $R_T^{*(2)}(\bar{\theta}_T^*) = \bar{G}^{*(2)}(\bar{\theta}_T^*) \text{vec}(\hat{v}_T^{*(2)} \hat{v}_T^{*(2)'})$  and  $\hat{v}_T^{*(2)} \equiv T^{1/4}(\hat{\theta}_T^{*(2)} - \hat{\theta}_T)$  and  $\bar{\theta}_T^*$  lies between  $\hat{\theta}_T^{*(2)}$  and  $\hat{\theta}_T$ . Here,  $\bar{G}^{*(2)}(\theta)$  is as  $\bar{G}^{*(1)}(\theta)$  but with  $\bar{\psi}_T^{*(1)}(\theta)$  replaced with  $\bar{\psi}_T^{*(2)}(\theta)$ . By a standard bootstrap uniform law of large numbers (see e.g. Giné and Zinn (1990)) and the fact that  $\hat{\theta}_T \rightarrow^P \theta_0$ , we can

show that  $\sup_{\theta} \left| \frac{\partial^2}{\partial \theta_i \partial \theta_j} \bar{\psi}_T^{*(2)}(\theta) - \kappa_{i,j}(\theta) \right| = o_{P^*}(1)$ , in prob- $P$ , where

$$\kappa_{i,j}(\theta) \equiv -\frac{\partial^2}{\partial \theta_i \partial \theta_j} \left( \frac{\partial}{\partial \theta'} \rho(\theta) \right) (\theta - \theta_0) - \frac{\partial^2}{\partial \theta_i \partial \theta_j} \rho(\theta).$$

By the continuous mapping theorem,  $\kappa_{i,j}(\hat{\theta}_T^*) = -\frac{\partial^2}{\partial \theta_i \partial \theta_j} \rho(\theta) + o_{P^*}(1)$ , in prob- $P$ , which implies that

$$\sqrt{T} \bar{\psi}_T^{*(2)}(\hat{\theta}_T^{*(2)}) = \sqrt{T} \bar{\psi}_T^{*(1)}(\hat{\theta}_T) - \frac{1}{2} \text{Gvec}(\hat{v}_T^{*(2)} \hat{v}_T^{*(2)'}) + o_{P^*}(1) + o_{P^*}(\|\hat{v}_T^{*(2)}\|). \quad (30)$$

We can now follow exactly the same arguments as in the proof of Proposition 4.2.(i). To prove part (ii), we also follow the proof of Proposition 4.2.(ii). We can show that

$$\begin{aligned} \sqrt{T} \bar{\psi}_T^{*(2)}(\hat{\theta}_T^{*(2)}) &= \sqrt{T} \bar{\psi}_T^{*(1)}(\hat{\theta}_T) - \frac{1}{2} \text{Gvec}(\hat{v}_T^{*(2)} \hat{v}_T^{*(2)'}) + o_{P^*}(1), \quad \text{and} \\ T^{1/4} \frac{\partial \bar{\psi}_T^{*(2)}}{\partial \theta_i}(\hat{\theta}_T^{*(2)}) &= -\frac{\partial^2 \rho}{\partial \theta_i \partial \theta'}(\theta_0) \hat{v}_T^{*(2)} + o_{P^*}(1), \end{aligned}$$

in prob- $P$ , for  $i = 1, \dots, p$ . Hence, the second order optimality condition for an interior solution of the minimization problem that  $\hat{\theta}_T^{*(2)}$  solves implies that for any vector  $e \in \mathbb{R}^p$ ,

$$e' \left( \tilde{Z}_T^* + N_T^* \right) e \geq 0,$$

where now

$$\begin{aligned} \tilde{Z}_T^* &= -Z_T^* - \frac{1}{2} \left( \frac{\partial^2 \rho'}{\partial \theta_i \partial \theta_j}(\theta_0) W \text{Gvec}(\hat{v}_T^{*(2)} \hat{v}_T^{*(2)'}) \right)_{1 \leq i, j \leq p} + o_{P^*}(1) \quad \text{and} \\ N_T^* &= \left( \hat{v}_T^{*(2)'} \frac{\partial^2 \rho'}{\partial \theta_i \partial \theta}(\theta_0) W \frac{\partial^2 \rho}{\partial \theta_i \partial \theta'}(\theta_0) \hat{v}_T^{*(2)} \right)_{1 \leq i, j \leq p} + o_{P^*}(1), \end{aligned}$$

where  $Z_T^* = \left( \frac{\partial^2 \rho'}{\partial \theta_i \partial \theta_j}(\theta_0) W \sqrt{T} \bar{\psi}_T^{*(1)}(\hat{\theta}_T) \right)_{1 \leq i, j \leq p}$ . Thus, there exists  $A > 0$  such that for any (unit) vector  $e \in \mathbb{R}^p$ ,

$$e' Z_T^* e - A \|\hat{v}_T^{*(2)}\|^2 \leq o_{P^*}(1).$$

Following the same argument as in the proof of Proposition 4.2, we can show that there is a subsequence of  $(Z_T^*, \hat{v}_T^{*(2)})$  which converges in distribution to  $(Z^*, V^*)$  under  $\mathbb{P}$ , where  $Z^*$  is a deterministic linear function of a Gaussian vector. Thus,

$$Y_{T'}^* \equiv e' Z_{T'}^* e - A \|\hat{v}_{T'}^{*(2)}\|^2 \xrightarrow{d_{\mathbb{P}}} Y^* \equiv e' Z^* e - A \|V^*\|^2,$$

along this subsequence. Because  $Z^*$  is a linear function of a Gaussian vector,  $Y^*$  is equal in distribution to  $-e' Z^* e - A \|V^*\|^2$  and the argument of D&R (2013) can now be applied to this limiting process to show that there exists a further subsequence of  $\hat{v}_T^{*(2)}$  that converges in distribution to  $V^*$  such that  $P(\|V^*\| > 0) > \delta$ .

**Proof of Lemma .4.** The proof follows exactly that of Lemma 4.1, given Lemma .3. The main difference is that the stochastic process  $Y_T^*(v) \equiv \sqrt{T} \bar{\psi}_T^{*(2)}(\hat{\theta}_T + T^{-1/4}v)$  now converges weakly in prob- $P$  towards  $\check{Y}(v) = X - \frac{1}{2} \text{Gvec}(vv')$ . Since  $X \sim N(0, \Sigma)$ , then  $X =^d -X$ , where  $=^d$  denotes equality in

distribution, which implies that  $\ddot{Y}(v) \stackrel{d}{=} -Y(v) = -(X + \frac{1}{2}Gvec(vv'))$ , where  $Y(v)$  was defined in Lemma 4.1. Thus,  $J_T^{*(2)}(v) = Y_T^*(v)' W_T^{*(2)} Y_T^*(v) \Rightarrow^{P^*} -Y(v)' W(-Y(v)) = Y(v)' W Y(v) \equiv J(v)$ , given that  $W_T^{*(2)} \rightarrow^{P^*} W$ , in prob- $P$ , concluding the proof.

**Proof of Theorem 4.2.** Part (i) follows as the proof of Theorem 4.1, given Lemmas .4 and .2 whereas part (ii) follows from Polya's Theorem.

## References

- [1] Ahn, S. C., and G. M. Thomas, 2006. "Likelihood Based Inference for Dynamic Panel Data Models," *Working Paper*, Arizona State University.
- [2] Altonji, J., and L. Segal, 1996. "Small-sample Bias in GMM Estimation of Covariance Structures," *Journal of Business & Economic Statistics*, 14, 353-366.
- [3] Andrews, D., 2002. "Higher-order Improvements of a Computationally Attractive  $k$ -step Bootstrap for Extremum Estimators," *Econometrica*, 70, 119-262.
- [4] Brown, B., and K. W. Newey, 2002. "Generalized Method of Moments, Efficient Bootstrapping, and Improved Inference," *Journal of Business & Economic Statistics*, 20, 507-517.
- [5] Chen, X., V. Chernozhukov, S. Lee and W. K. Newey, 2011. "Local Identification of Nonparametric and Semiparametric Models," *Working Paper*, MIT, <http://economics.mit.edu/files/6720>
- [6] Davidson, J., 1994. "Stochastic Limit Theory," *Oxford University Press*.
- [7] Dovonon, P. and E. Renault, 2013. "Testing for Common GARCH Factors," forthcoming in *Econometrica*.
- [8] Engle, R. F. and S. Kozicki, 1993. "Testing For Common Features," *Journal of Business & Economic Statistics*, 11(4), 369-395.
- [9] Giné, E. and J. Zinn, 1990. "Bootstrapping General Empirical Measures," *Annals of Probability* 18, 851-869.
- [10] Hahn, J., 1996. "A Note on Bootstrapping Generalized Method of Moments Estimators," *Econometric Theory*, 12, 187-197.
- [11] Hall, P., and J. Horowitz, 1996. "Bootstrap Critical Values for Tests Based on Generalized Method-of-Moments Estimators," *Econometrica*, 64, 891-916.
- [12] Hansen, L. P., 1982. "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029-1054.
- [13] Inoue, A., and M. Shintani 2006. "Bootstrapping GMM Estimators for Time Series," *Journal of Econometrics*, 133, 531-555.

- [14] Lee, L. F. and A. Chesher, 1986. "Specification Testing when Score Test Statistics are Identically Zero," *Journal of Econometrics*, 31, 121-149.
- [15] Madsen, E., 2009. "GMM-based inference in the AR(1) panel data model for parameter values where local identification fails," *Working Paper*, University of Copenhagen.
- [16] Melino, A., 1982. "Testing for Sample Selection Bias," *Review of Economic Studies*, 49, 151-153.
- [17] Newey, K. W. and D. McFadden, 1994. "Large Sample Estimation and Hypothesis Testing," *Handbook of Econometrics*, IV, Edited by R.F. Engle and D. L. McFadden, 2112-2245.
- [18] Pollard, D., 1984. "Convergence of Stochastic Processes," *Springer*, New York.
- [19] Rotnitzky, A., D. R. Cox, M. Bottai and J. Robins, 2000. "Likelihood-based Inference with Singular Information Matrix," *Bernoulli*, 6(2), 243-284.
- [20] Sargan, J. D., 1983. "Identification and lack of Identification," *Econometrica*, 51, 1605-1633.
- [21] Sen, B., M. Banerjee and M. Woodroffe, 2010. "Inconsistency of the Bootstrap: The Grenander Estimator," *Annals of Statistics*, 38, 1953-1977.