Accepted Manuscript

Bootstrapping the GMM overidentification test under first-order underidentification

Prosper Dovonon, Sílvi Gonçalves

Please cite this article as: Dovonon, P., Gonçalves, S., Bootstrapping the GMM overidentification test under first-order underidentification. *Journal of Econometrics* (2017), http://dx.doi.org/10.1016/j.jeconom.2017.06.021

Bootstrapping the GMM overidentification test under first-order underidentification[*]

Prosper Dovonon[†] and Sílvia Gonçalves[‡]

Concordia University and University of Western Ontario

July 31, 2017

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 (2013b) 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 GMM bootstrap 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.

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 θ_0 solves the moment conditions, ensures

---

[†]Economics Department, Concordia University, 1455 de Maisonneuve Blvd. West, H 1155, Montreal, Quebec, Canada, H3G 1M8. Email: prosper.dovonon@concordia.ca.

[‡]Department of Economics, Faculty of Social Science, University of Western Ontario, 1151 Richmond Street N., London, Ontario, Canada, N6A 5C2. Email: sgoncal9@uwo.ca.

that the GMM estimator is consistent for θ_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 θ 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 (2013b) (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 χ^2_{H-p} random variable (where H denotes the number of moment conditions and p the number of parameter coefficients in θ). Instead, its asymptotic distribution is a fifty-fifty mixture of χ^2_{H-1} and χ^2_H when $p = 1$, leading to an oversized test under the null of common conditionally heteroskedastic features when the standard χ^2_{H-1} distribution is used, independently of the sample size. When $p > 1$, the correct asymptotic distribution is the minimum over $\mathbb{R}^p$ of a certain stochastic process whose distribution is highly nonstandard and depends on nuisance parameters. 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 χ^2_{H-p} and χ^2_H when $p > 1$. An alternative to the bootstrap in this case is to simulate critical values from the nonstandard asymptotic distribution by replacing the nuisance parameters with consistent estimates and solving for the minimum of the simulated stochastic process. The bootstrap avoids solving this minimization problem explicitly, which can be computationally very challenging when p is large.

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

2

GMM bootstrap does not mimic the degeneracy of the expected Jacobian matrix at θ_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 of 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 θ 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 θ 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.

Another approach to conduct overidentification tests under first-order underidentification is to exploit the information contained in the zero Jacobian matrix. This idea was recently proposed by Lee and Liao (2016) and Sentana (2015) as a way to overcome the difficulties created by the lack of first-order local identification for GMM-based tests (see also Doz and Renault (2006) for an earlier implementation of this idea in the context of conditionally heteroskedastic common factors). In particular, by augmenting the set of original moment conditions with the moment conditions implied by the zero Jacobian matrix, the GMM estimator recovers the standard $\sqrt{T}$-rate of convergence and tests based on the augmented GMM estimator have the standard asymptotic distributions. Our simulations show that the overidentification test based on the augmented set of moment conditions for which the standard chi-squared critical values apply can be very oversized in finite samples. An alternative test also proposed by Lee and Liao (2016) is an overidentification test for the augmented set of moment conditions that relies on a GMM estimator obtained from a subset of these conditions (in particular, those induced by the zero Jacobian matrix). This test is no longer asymptotically distributed as a chi-squared distribution and also requires simulated critical values. Although it is much better behaved in finite samples than the efficient GMM test, it can lead to over-rejections under the null that are avoided by the continuously corrected bootstrap. In addition, the null hypothesis that it tests nests the original moment conditions, which can result in a loss of power in certain directions.

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 θ. Section 3 establishes the inconsistency of the standard bootstrap distribution for the test of overidentification conditions when first-order

underidentification occurs. 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, examples, assumptions, and asymptotic theory

We consider a sample $\{X_t : t = 1, \ldots, T\}$ 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. Throughout we maintain the assumption that the model is globally identified, i.e. there exists a unique parameter vector θ_0 that solves the moment conditions $E(\psi(X_t, \theta_0)) = 0$. 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 $\Sigma = \mathrm{Var}(\psi(X_1, \theta_0))$, we obtain Hansen's (1982) GMM overidentification test statistic. This statistic has the standard χ^2_{H-p} as $T \to \infty$ under the rank identification condition that $Rank\left(\frac{\partial}{\partial \theta'} \rho(\theta_0)\right) = p$, where $\rho(\theta) \equiv E(\psi(X_1, \theta))$.

Our aim in this section is to derive the asymptotic distribution of $\hat{J}_T$ whenever the rank identification condition fails. First, we provide two examples where this is the case.

2.1 Examples of first-order underidentification

Our first example is studied by D&R (2013) and is the main motivation for our paper. It illustrates a model that is globally identified at some parameter vector θ_0 for which the Jacobian matrix is nil whatever the value of θ_0.

Example 2.1 *(Testing for common conditionally heteroskedastic factors)*
Following D&R (2013), consider n assets and $(Y_{t+1})_t$ the vector process of their returns that is assumed stationary with each component conditionally heteroskedastic with respect to an increasing filtration $\mathfrak{F}_t$ that contains at least returns up to t. These n returns have common conditionally heteroskedastic

features if there exists a $n \times K$ matrix Λ of rank $K < n$, a symmetric positive definite $n \times n$ matrix Ω, and a diagonal $K \times K$ matrix D_t that is $\mathfrak{F}_t$-measurable such that

$$Var(Y_{t+1}|\mathfrak{F}_t) = \Lambda D_t \Lambda' + \Omega.$$

A natural consequence of the common conditionally heteroskedastic features is the existence of so-called co-feature vectors $\theta \neq 0$ in $\mathbb{R}^n$ that offset the conditionally heteroskedastic patterns from $(Y_{t+1})_t$:

$$Var(\theta' Y_{t+1}|\mathfrak{F}_t) = \theta' \Omega \theta,$$

a constant matrix. Assuming that $E(Y_{t+1}|\mathfrak{F}_t) = 0$, a test for common conditionally heteroskedastic factors can be based on the conditional moment restriction:

$$E((\theta' Y_{t+1})^2|\mathfrak{F}_t) = c,$$

for some constant c. If we let $c = c(\theta) = E(\theta' Y_{t+1})^2$, this moment restriction can be translated into an unconditional moment restriction:

$$E\left(\psi(z_t, Y_{t+1}, \theta)\right) \equiv E\left(z_t \left((\theta' Y_{t+1})^2 - c(\theta)\right)\right) = 0, \tag{2}$$

where z_t is a vector of instruments belonging to $\mathfrak{F}_t$. Testing for common conditionally heteroskedastic features amounts to testing the validity of the moment conditions in (2) and the standard overidentification test statistic is routinely used to do so. D&R (2013) show that the model (2) globally identifies the parameters of interest under suitable normalization of θ if $K = n - 1$. They also establish that regardless of the normalization chosen,

$$E\left(\frac{\partial \psi}{\partial \theta^{1\prime}}(z_t, Y_{t+1}, \theta_0)\right) = 0,$$

where θ^1 is the collection of free parameters after normalization.

Our next example considers a less extreme case of first-order underidentification in which the Jacobian matrix is not identically zero although it is rank deficient.

Example 2.2 *(Nonstationary panel AR(1) model with individual fixed effects)*
Consider the standard AR(1) panel data model with individual specific effects,

$$y_{it} = \rho y_{i,t-1} + \eta_i + \varepsilon_{it}, \quad i = 1, \ldots, N, \quad t = 0, 1, 2.$$

Assume that the vector $(y_{i0}, \eta_i, \varepsilon_{i1}, \ldots, \varepsilon_{iT})$ is i.i.d. across i with mean 0 and that $E(\varepsilon_{it}^2) = \sigma_\varepsilon^2$, $E(\varepsilon_{is}\varepsilon_{it}) = 0$ for $s \neq t$, $s, t = 0, 1, 2$, $E(\eta_i^2) = \sigma_\eta^2$, $E(y_{i0}^2) = \sigma_0^2$, $E(\varepsilon_{it}\eta_i) = 0$, $E(\varepsilon_{it}y_{i0}) = 0$, $t = 0, 1, 2$, and $E(y_{i0}\eta_i) = \sigma_{0\eta}$.

Let $\theta = (\rho, \sigma_0^2, \sigma_{0\eta}, \sigma_\eta^2, \sigma_\varepsilon^2)'$ *denote the vector of parameters in this model. Writing y_{i1} and y_{i2} in terms of y_{i0}, and then $Var((y_{i0}, y_{i1}, y_{i2})')$ in terms of θ, we can easily show the following moment*

restrictions hold:

$$E(\psi(y_i, \theta)) = 0, \tag{3}$$

where $y_i = (y_{i0}, y_{i1}, y_{i2})'$ and

$$\psi(y_i, \theta) = \begin{pmatrix} y_{i0}^2 - \sigma_0^2 \\ y_{i0}y_{i1} - \rho y_{i0}^2 - \sigma_{0\eta} \\ y_{i0}y_{i2} - \rho y_{i0}y_{i1} - \sigma_{0\eta} \\ y_{i1}^2 - y_{i0}y_{i2} - \sigma_\eta^2 + (1-\rho)\sigma_{0\eta} - \sigma_\varepsilon^2 \\ y_{i1}y_{i2} - \rho y_{i1}^2 - \sigma_\eta^2 - \rho\sigma_{0\eta} \\ (y_{i2} - \rho y_{i1})^2 - \sigma_\eta^2 - \sigma_\varepsilon^2 \end{pmatrix}.$$

These moment restrictions are equivalent, up to a linear one-to-one mapping, to those considered by Madsen (2009, Eq. (5)). We can show that this model is globally identified over a large set of parameter values, i.e. if the true parameter value θ^* belongs to that set, (3) is uniquely solved at θ^*.

However, as it turns out, the usual rank condition is not satisfied for all θ^* in that parameter space. In particular, it fails when the true value is $\theta^* = (1, \sigma_0^{*2}, 0, 0, \sigma_\varepsilon^{*2})$. This parameter configuration corresponds to a panel AR(1) model with a unit root and no individual fixed effects:

$$y_{it} = y_{i,t-1} + \varepsilon_{it}, \quad i = 1, \ldots, n, t = 0, 1, 2.$$

More specifically, we can show that the Jacobian matrix evaluated at $\theta^* = (1, \sigma_0^{*2}, 0, 0, \sigma_\varepsilon^{*2})$ is

$$E\left(\frac{\partial \psi}{\partial \theta'}(y_i, \theta^*)\right) = \begin{pmatrix} 0 & -1 & 0 & 0 & 0 \\ -\sigma_0^{*2} & 0 & -1 & 0 & 0 \\ -\sigma_0^{*2} & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & -1 \\ -\sigma_0^{*2} - \sigma_\varepsilon^{*2} & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & -1 & -1 \end{pmatrix},$$

which is of rank $4 < 5$. Hence, the usual rank identification condition fails, implying that the overidentification test statistic does not have the usual limiting chi square distribution. Similarly, a test for unit root will not have the standard limiting distribution.

The failure of the rank condition in Example 2.2 is not as extreme as in Example 2.1 since the expected Jacobian matrix is not nil. Although our assumptions in the next section impose that the expected Jacobian matrix is zero, we discuss in Section 4 how our bootstrap methods can be modified to cover situations like Example 2.2.

2.2 Assumptions

The previous examples show that for nonlinear moment condition models, local first-order identification cannot be taken for granted even when the model is globally identified. As D&R (2013) showed for Example 2.1, the lack of first-order identification results in nonstandard limiting distributions for the overidentification test statistic. Here we extend their results by considering general moment conditions and by studying bootstrap inference. 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 $\{X_t : \Omega \to \mathbb{R}^l, l \in \mathbb{N}, t = 1, \ldots, T\}$.

Assumption 2 (global identification) θ_0 is an interior point of Θ, a compact subset of $\mathbb{R}^p$, $p \in \mathbb{N}$, and it is the unique solution in Θ to the equation $E(\psi(X_t, \theta)) = 0$.

Assumption 3 (regularity conditions on ψ)

(i) $\{\psi(x, \theta)\}$ is continuous on Θ for all x in the support of X_1.

(ii) $E\left(\sup_{\theta \in \Theta} \|\psi(X_1, \theta)\|\right) < \infty.$

(iii) $E\left(\|\psi(X_1, \theta_0)\|^2\right) < \infty$ and $\Sigma \equiv \text{Var}(\psi(X_1, \theta_0))$ is positive definite.

Assumption 4 (regularity conditions on derivatives on ψ)

(i) $\{\psi(x, \theta)\}$ is twice continuously differentiable with respect to θ in a neighborhood $\mathcal{N}$ of θ_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 $h = 1, \ldots, H$, 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 $h = 1, \ldots, H$.

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 a central limit theorem to the sample mean of the Jacobian matrix of the estimating function evaluated at the true value θ_0. The second part of Assumption 4(ii) along with Assumption 4(i) ensures that the second-order derivatives of the estimating function are uniformly dominated in a neighborhood of θ_0, allowing for the application of a 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 $\frac{\partial \rho(\theta_0)}{\partial \theta'}$ to be equal to p. Given the lack of first-order identification, part (ii) of Assumption 6 ensures local identification of θ_0 through second-order derivatives. In particular, letting

$$G \equiv \left[\begin{array}{ccc} vec\left(\frac{\partial^2 \rho_1(\theta_0)}{\partial \theta \partial \theta'}\right) & \cdots & vec\left(\frac{\partial^2 \rho_H(\theta_0)}{\partial \theta \partial \theta'}\right) \end{array}\right]'$$

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

$$G \cdot vec\left((\theta - \theta_0)(\theta - \theta_0)'\right) \neq 0 \text{ for all } \theta \neq \theta_0. \tag{4}$$

When $p = 1$, (4) is equivalent to $G \neq 0$, i.e. $\frac{\partial^2 \rho_h(\theta_0)}{\partial \theta^2} \neq 0$ for at least one $h = 1, \ldots, H$. In the general case, the condition that $G \neq 0$ is obviously necessary for (4) 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 θ. An important example that satisfies (4) are the moment conditions underlying the test for common conditionally heteroskedastic features (Example 2.1). See D&R's (2013) Lemma 2.3.

Before introducing our asymptotic results, we further discuss the previous assumptions in connection with Examples 2.1 and 2.2. Even though Assumption 1 imposes i.i.d samples for simplicity of presentation, our main results continue to hold if we assume that X_t is a stationary process and $\psi(X_t, \theta_0)$ is a martingale difference sequence with respect to the past of X_t. In that respect, Example 2.1 is also covered by our results. The fact that the estimating functions in both examples are quadratic functions of parameters makes Assumption 4(i) hold in both settings. In Example 2.1, the existence of some moments imposed by Assumptions 3(ii, iii) and 4(ii) rules out some conditionally heteroskedastic processes with heavy tails. Nevertheless, a large class of processes fits in, including GARCH processes with innovations having finite eighth moment.[1]

Example 2.2 does not fit into the second-order local identification pattern imposed by Assumption 6 since the Jacobian of the moment function (3) at the true value is not nil. The kind of rank deficiency presented by this example is covered by the general formulation of second-order local identification in Assumption 6' in Section 4. Assumptions 1-5 are fulfilled in this example by a wide choice of joint distribution for $(y_{i0}, \eta_i, \varepsilon_{i1}, \varepsilon_{i2})$.

Next, we derive the asymptotic distribution of $\hat{J}_T$ under Assumptions 1-6. Although in practice $\hat{J}_T$ requires the choice of the weighting matrix W_T, we abstract from this choice in the paper. However, we note that our Assumption 5 is consistent with the usual choice of W_T that guarantees efficiency when the rank condition for identification is satisfied. In particular, under our assumptions, a consistent estimator of $W = \Sigma^{-1}$ 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 θ_0 (e.g. a GMM estimator based on $W_T = I_p$).

2.3 Asymptotic results

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

---

[1] See He and Teräsvirta (1999) for a discussion on the existence of moments of GARCH processes.

Proposition 2.1 shows that the GMM estimator is consistent for θ_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)* $\left\|\hat{\theta}_T - \theta_0\right\| = O_P\left(T^{-1/4}\right)$ *and (ii)* $T^{1/4}\left(\hat{\theta}_T - \theta_0\right)$ *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 $T^{1/4}\left(\hat{\theta}_T - \theta_0\right)$ which converges in distribution to a nondegenerate random variable V. This implies that $T^{1/4}\left(\hat{\theta}_T - \theta_0\right)$ 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 $\sqrt{T}\bar{\psi}_T\left(\hat{\theta}_T\right)$ as

$$\sqrt{T}\bar{\psi}_T\left(\hat{\theta}_T\right) = \sqrt{T}\bar{\psi}_T(\theta_0) + \underbrace{\frac{\partial\bar{\psi}_T(\theta_0)}{\partial\theta'}\sqrt{T}\left(\hat{\theta}_T - \theta_0\right)}_{=o_P(1) \text{ under Assumption 6(i)}} + \frac{1}{2}R_T\left(\ddot{\theta}_T\right), \tag{5}$$

where the remainder is such that

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

$$\text{under A6(ii)}$$

where for any θ,

$$\bar{G}(\theta) \equiv \left[\begin{array}{ccc} vec\left(\frac{\partial^2\bar{\psi}_{1,T}(\theta)}{\partial\theta\partial\theta'}\right) & \cdots & vec\left(\frac{\partial^2\bar{\psi}_{H,T}(\theta)}{\partial\theta\partial\theta'}\right) \end{array}\right]' \tag{6}$$

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 the CLT) and since $\hat{\theta}_T - \theta_0 = o_P(1)$, it follows that the second term in (5) is $o_P(1)$. Thus, identification is achieved at the second order. In particular, Assumption 6(ii) guarantees that $R_T\left(\ddot{\theta}_T\right)$ is an $O_P(1)$ term that is *not* $o_P(1)$, implying that $T^{1/4}\left(\hat{\theta}_T - \theta_0\right) = O_P(1)$.

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

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

9

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

$$J(v) \equiv X'WX + X'WG\, vec\,(vv') + \frac{1}{4} vec'\,(vv')\, G'WG\, vec\,(vv'), \qquad (7)$$

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

Theorem 2.3 *If Assumptions 1-6 hold, then as $T \to \infty$,*

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

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

The first part of Theorem 2.3 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 θ. The second part of this theorem is new. It shows that the asymptotic distribution of $\hat{J}_T$ 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 2.3, 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 $\sup_{v \in \mathbb{K}} |z(v)|$.[2] 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 $E(h(J_T)) \to E(h(J))$, for any $h : \ell^\infty(\mathbb{K}) \to \mathbb{R}$ bounded and continuous with respect to the sup norm. Let $\hat{v}_T = T^{1/4} \left(\hat{\theta}_T - \theta_0 \right)$. Since $\hat{J}_T = J_T(\hat{v}_T) = \min_{v \in \mathbb{H}_T} J_T(v)$, where $\mathbb{H}_T \equiv \left\{ v \in \mathbb{R}^p : v = T^{1/4} (\theta - \theta_0), \theta \in \Theta \right\}$ 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 2.3 provides the asymptotic distribution for the GMM overidentification test statistic $\hat{J}_T$ under second-order local identification (the standard case where $W_T \xrightarrow{P} W = \Sigma^{-1}$ 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.

---

[2]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 $\ell^\infty(\mathbb{K})$.

Except for the special case when $p = 1$, for which D&R (2013) showed that this distribution is a fifty-fifty mixture of χ_H^2 and χ_{H-1}^2, critical values of $\underline{J}$ are not available. This is the main motivation for proposing the bootstrap.

3   Asymptotic inconsistency of the standard bootstrap distribution

Given Assumption 1, we let $\{X_t^* : t = 1, \ldots, T\}$ denote a (conditionally) i.i.d. bootstrap sample obtained by resampling with replacement the original sample $\mathcal{X}_T \equiv \{X_t : t = 1, \ldots, n\}$.[3] A standard application of the bootstrap for GMM estimators involves recentering the bootstrap moment conditions by subtracting the term[4] $E^* \left(\psi \left(X_t^*, \hat{\theta}_T \right) \right)$ from $\frac{1}{T} \sum_{t=1}^T \psi_t (X_t^*, \theta)$ 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

$$\bar{\psi}_{c,T}^* (\theta) = T^{-1} \sum_{t=1}^T \psi_c (X_t^*, \theta), \text{ with}$$

$$\psi_c (X_t^*, \theta) = \psi (X_t^*, \theta) - E^* \left(\psi \left(X_t^*, \hat{\theta}_T \right) \right).$$

Letting $\psi_t^* (\theta) \equiv \psi (X_t^*, \theta)$ and $\psi_{c,t}^* (\theta) \equiv \psi_c (X_t^*, \theta)$ for $t = 1, \ldots, T$ 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. (See also Lee (2014).) 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

---

[3]Under weak dependence of $\{X_t : t = 1, \ldots, T\}$, 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.

[4]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.

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 when local identification is achieved at the second order as described by Assumption 6. The bootstrap GMM estimator and the corresponding overidentification test statistic are defined as

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

We require W_T^* to converge to W under the bootstrap measure P^* with probability P approaching one, i.e. $W_T^* \overset{P^*}{\to} W$ in prob-P (see Appendix B for the formal definition of this mode of convergence). One possible choice for W_T^* which is covered by our assumptions is the inverse of

$$\frac{1}{T} \sum_{t=1}^{T} \psi_{c,t}^*(\tilde{\theta}_T^*) \psi_{c,t}^*(\tilde{\theta}_T^*)', \tag{8}$$

where $\tilde{\theta}_T^*$ is a first-step GMM estimator.

Our first result shows that $\hat{\theta}_T^*$ is consistent for θ_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 $W_T^* \overset{P^*}{\to} W$ 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 θ_0 is the unique solution to the moment conditions $\rho(\theta) = 0$ (and hence the unique minimizer of $Q(\theta)$ over Θ) 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 $v = T^{1/4}\left(\theta - \hat{\theta}_T\right)$, with $\theta \in \Theta$. Note that

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

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

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

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

(ii) *There exists at least one subsequence of*
$\left(\sqrt{T}\bar{\psi}_T(\theta_0), \sqrt{T}\left(\bar{\psi}_T^*(\theta_0) - \bar{\psi}_T(\theta_0)\right), T^{1/4}\left(\hat{\theta}_T - \theta_0\right), T^{1/4}\left(\hat{\theta}_T^* - \hat{\theta}_T\right)\right)$ *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*

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

(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. 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 inconsistency of the standard bootstrap distribution of the overidentification test is that while the sample mean of the Jacobian matrix of the estimating function evaluated at the population value θ_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,

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

where $\psi_t^{*(1)}(\theta) \equiv \psi^{(1)}(X_t^*, \theta)$ for any $t = 1, \ldots, T$ 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

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

is zero when $\theta = \hat{\theta}_T$. Thus, the corrected GMM bootstrap is able to mimic the lack of first-order identification at θ_0 that affects the original moment conditions.

Let

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

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

$$\hat{J}_T^{*(1)} = T\bar{\psi}_T^{*(1)\prime}\left(\hat{\theta}_T^{*(1)}\right)W_T^{*(1)}\bar{\psi}_T^{*(1)}\left(\hat{\theta}_T^{*(1)}\right).$$

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

14

This assumption is not particularly restrictive when ψ is a polynomial function of θ as in Examples 2.1 and 2.2. In this case, since Θ is compact, it is not needed if Assumption 4(ii) is maintained.

Proposition 4.1 *Under Assumptions 1-7, if* $W_T^{*(1)} \overset{P^*}{\to} W$ *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)* $\left\| \hat{\theta}_T^{*(1)} - \hat{\theta}_T \right\| = O_{P^*}\left(T^{-1/4}\right)$ *in prob-P, and (ii)* $T^{1/4}\left(\hat{\theta}_T^{*(1)} - \hat{\theta}_T\right)$ *has at least a subsequence that converges in distribution to some random variable* V^* *under* P^*, *a.s. -P, such that for some* $\delta > 0$, $P\left(\|V^*\| \neq 0\right) \geq \delta$.

Part (ii) shows that $\hat{v}_T^{*(1)} \equiv T^{1/4}\left(\hat{\theta}_T^{*(1)} - \hat{\theta}_T\right)$ 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 $v \in \mathbb{R}^p$,

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

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

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

where $\mathbb{V}_T = \left\{ v \in \mathbb{R}^p : v = T^{1/4}\left(\theta - \hat{\theta}_T\right), \theta \in \Theta \right\}$. Our first result shows that conditionally on the original sample, $J_T^{*(1)}(v)$ 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 $E^*\left(h\left(J_T^{*(1)}\right)\right) \to E(h(J))$ in prob-P for any $h : \ell^\infty(\mathbb{K}) \to \mathbb{R}$ bounded and continuous with respect to the sup norm.

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

15

restrictions, which is particularly useful when $p > 1$, for which there is no closed form expression for the asymptotic distribution of $\hat{J}_T$.

Letting $\hat{c}_T^*(1 - \alpha)$ denote the $100(1 - \alpha)\%$ quantile of the bootstrap distribution of $\hat{J}_T^{*(1)}$, Theorem 4.1 implies that the bootstrap test has correct pointwise asymptotic size in the sense that

$$\limsup_{T \to \infty} \Pr_{(P,\theta_0)} \left(\hat{J}_T > \hat{c}_T^*(1 - \alpha) \right) = \alpha,$$

where $\Pr_{(P,\theta_0)}$ denotes the probability of the indicated event for a given distribution of the data (as summarized by P) and a true parameter value θ_0. Note that under our set of assumptions, pointwise validity of the test implies uniform validity (in the sense that $\limsup_{T \to \infty} \sup_{(P,\theta_0)} \Pr_{(P,\theta_0)} \left(\hat{J}_T > \hat{c}_T^*(1 - \alpha) \right) = \alpha$). This is true because we assume that the Jacobian matrix is nil at θ_0 (whatever value of θ_0 is) and therefore we rule out the possibility that there is a discontinuity of the asymptotic distribution of the overidentification test. If we were to relax this assumption, then a discontinuity in the asymptotic distribution of $\hat{J}_T$ would arise and pointwise results would not necessarily imply uniform results. Next, we offer a brief discussion of what happens if we relax Assumption 6(i).

Consider the special case where $p = 1$, where relaxing Assumption 6 (i) amounts to allowing for the possibility that the Jacobian is either non-zero (and of rank 1) or zero (and of rank 0). In the first case (which is the regular case), $\hat{J}_T$ is asymptotically distributed as χ^2_{H-1} whereas in the second case it is asymptotically distributed as a fifty-fifty mixture of χ^2_H and χ^2_{H-1} (as shown by D&R 2013). Take a fixed value of θ_0 such that the expected Jacobian matrix is either zero or non-zero and consider a bootstrap test based on the modified moment conditions. Given that by construction $\sqrt{T} \partial \bar{\psi}_T^{*(1)} \left(\hat{\theta}_T \right) / \partial \theta' = O_{P^*}(1)$ in prob-P, we can verify that the conclusion of part (i) of Theorem 4.1 holds without Assumption 6(i) provided the bootstrap estimator converges to θ_0 in probability, i.e. $\hat{\theta}_T^{*(1)} = \theta_0 + o_{P^*}(1)$ in prob-P. Letting $D = E\left(\partial \psi(X_t, \theta_0) / \partial \theta' \right)$, we can show that this is the case if the equation $E\left(\psi(X_t, \theta) \right) - D(\theta - \theta_0) = 0$ is uniquely solved at $\theta = \theta_0$. Under these conditions, the bootstrap distribution is asymptotically equal to a fifty-fifty mixture of χ^2_H and χ^2_{H-1}, independently of whether the Jacobian is zero or not, and since the critical values from χ^2_{H-1} are smaller than those from the mixture, we can conclude that the bootstrap is pointwise valid in this case (in the sense that its asymptotic size is smaller or equal than the nominal level).

When $p > 1$, asymptotic (pointwise or uniform) validity of the bootstrap is even harder to explore when we relax Assumption 6 (i). The main reason is that in this case a failure of the rank condition of the expected Jacobian matrix is equivalent to the rank being any integer $0 \leq r \leq p - 1$. Although the bootstrap distribution of $\hat{J}_T^*$ is still consistent towards the distribution of $\underline{J}$, it is hard to compare the critical values of this distribution with the critical values of $\hat{J}_T$ when r takes any value from 0 to p. However, if we consider the case where $r = p$ (which amounts to the regular situation of full column rank), a sufficient condition for our corrected bootstrap method to be pointwise valid is that the critical values of $\underline{J}$ are larger or equal to those of a χ^2_{H-p} distribution. We can show that $\underline{J}$ stochastically dominates $\chi^2_{H-rank(G)}$, implying that the modified bootstrap method we propose is (pointwise) valid

16

when the Jacobian matrix is of full column-rank if $rank\,(G) \leq p$. D&R (2013) show that this condition is verified for Example 2.1 (see their Lemma 3.1).

Because our bootstrap methods are based on knowing the rank of the Jacobian matrix, they are not designed to be uniformly valid against departures of this assumption. Whether they remain valid in this more general context is a challenging question left for future research. A recent paper that proposes specification tests that are uniformly valid and robust to the rank deficiency of the Jacobian matrix in the context of moment (in)equality models is Bugni, Canay and Shi (2015).

To end this section, we now discuss the power properties of the proposed tests. We consider fixed alternatives under which the model is not correctly specified in the sense that no parameter value in the parameter space Θ solves the moment condition model, i.e., $\mu(\theta) \equiv E(\psi_t(\theta)) \neq 0, \quad \forall \theta \in \Theta$. In this case and as mentioned by D&R (2013), $\hat{J}_T$ diverges to infinity, which guarantees that the overidentification test is consistent under fixed alternatives. This together with the fact that we can show that the bootstrap statistic $\hat{J}_T^{*(1)}$ is $O_{P^*}(1)$, in prob-P, under the alternative, imply the consistency of the bootstrap test. To see why $\hat{J}_T^{*(1)}$ is $O_{P^*}(1)$, in prob-P, note that

$$
\begin{aligned}
\hat{J}_T^{*(1)} &\equiv T\bar{\psi}_T^{*(1)\prime}\left(\hat{\theta}_T^{*(1)}\right) W_T^{*(1)} \bar{\psi}_T^{*(1)}\left(\hat{\theta}_T^{*(1)}\right) \\
&\leq T\bar{\psi}_T^{*(1)\prime}\left(\hat{\theta}_T\right) W_T^{*(1)} \bar{\psi}_T^{*(1)}\left(\hat{\theta}_T\right) \\
&= T\left(\bar{\psi}_T^{*}(\hat{\theta}_T) - \bar{\psi}_T(\hat{\theta}_T)\right)' W_T^{*(1)}\left(\bar{\psi}_T^{*}(\hat{\theta}_T) - \bar{\psi}_T(\hat{\theta}_T)\right),
\end{aligned}
$$

with $\bar{\psi}_T^{*}(\theta) = \frac{1}{T}\sum_{t=1}^{T}\psi_t^{*}(\theta)$. Under our assumptions, $\sqrt{T}\left(\bar{\psi}_T^{*}(\hat{\theta}_T) - \bar{\psi}_T(\hat{\theta}_T)\right) = O_{P^*}(1)$, in prob-$P$, implying the result.

4.2 The continuously-corrected GMM bootstrap

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

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

where the expected bootstrap Jacobian matrix is evaluated at θ instead of $\hat{\theta}_T$. Because this bears a resemblance with the continuous-updated GMM, in which the weighting matrix is evaluated at θ 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 $\psi_{h,t}^{*(2)}(\theta)$ $(h = 1, \ldots, H)$ is given by

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

17

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

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

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

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

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 θ_0, we need to impose the following assumption.

Assumption 8 θ_0 is the unique solution to the equation $E \left(\Phi \left(X_1, \theta \right) \right) = 0$.

Assumption 8 imposes the restriction that θ_0, the parameter vector that uniquely identifies the original moment conditions given by Assumption 2, also uniquely identifies the modified moment conditions $E \left(\Phi \left(X_1, \theta \right) \right) = 0$. One leading case where this requirement is satisfied is when the original moment conditions $\psi(x, \theta)$ are quadratic in θ and and the expected Jacobian matrix is nil at θ_0. Actually, in this case, $E(\Phi(x, \theta)) = -E(\psi(x, \theta))$, for all θ. Note that the test for common conditionally heteroskedastic features (Example 2.1) fits into this category. It is also worthwhile to mention that $E(\Phi(x, \theta)) = 0$ has the same local identification properties at θ_0 as $E(\psi(x, \theta)) = 0$. In particular, we can see that

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

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

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

over $\theta \in \Theta$, is consistent towards θ_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

$$
\hat{J}_T^{*(2)} = T \bar{\psi}_T^{*(2)\prime} \left(\hat{\theta}_T^{*(2)} \right) W_T^{*(2)} \bar{\psi}_T^{*(2)} \left(\hat{\theta}_T^{*(2)} \right).
$$

18

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 θ in a neighborhood $\mathcal{N}$ of θ_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 $i, j = 1, \dots, p$.

Note that this assumption is trivially satisfied by Examples 2.1 and 2.2 since their respective estimating functions are quadratic in the parameters.

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 (and the same remarks apply here regarding the implications for pointwise/uniform validity and power properties of the test). 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.

4.3 Extension to rank deficient but non-vanishing Jacobian matrix

So far we have focused on the particular case where first-order local identification failure is expressed in the form of a null Jacobian matrix at the true parameter value. The goal of this section is to show how our bootstrap methods can be adapted to the case where the Jacobian matrix is rank deficient but different from zero, as in Example 2.2.

Suppose that

$$Rank\left(\frac{\partial \rho}{\partial \theta'}(\theta_0)\right) = r, \quad 0 < r < p.$$

Dovonon and Renault (2009) study the asymptotic properties of the overidentification test statistic in this setting. They show in particular that when $r = p - 1$, $\hat{J}_T$ is asymptotically distributed as a fifty-fifty mixture of χ^2_{H-r} and χ^2_{H-p}. When $r < p - 1$, the asymptotic distribution of $\hat{J}_T$ lies between χ^2_{H-r} and χ^2_{H-p} with non-tractable critical values. In this case, it is possible to perform conservative tests with the χ^2_{H-r} bound.

In order to apply the bootstrap to this context, we need to modify the bootstrap estimating function. Consider first the simpler case where the Jacobian matrix is full rank in the first r directions and nil in the last $p - r$ directions. More specifically, suppose that we can write $\theta = (\theta_1', \theta_2')'$, where θ_1 is a $r \times 1$ vector and θ_2 an $(p - r) \times 1$ vector and

$$Rank\left(\frac{\partial \rho}{\partial \theta_1'}(\theta_0)\right) = r, \quad \text{and} \quad \frac{\partial \rho}{\partial \theta_2'}(\theta_0) = 0.$$

In this case the second recentering of the bootstrap estimating function shall be applied only in the direction of θ_2. In particular, for our first bootstrap method, the estimating function used in the

bootstrap shall be:

$$\psi_t^{*(1)}(\theta) = \psi_t^*(\theta) - E^* \left(\psi_t^*(\hat{\theta}_T) \right) - E^* \left(\frac{\partial}{\partial \theta_2'} \psi_t^*(\hat{\theta}_T) \right) \left(\theta_2 - \hat{\theta}_{2T} \right).$$

A similar correction works for the continuously-corrected bootstrap method based on $\psi_t^{*(2)}(\theta)$ with the only difference that the bootstrap Jacobian matrix with respect to θ_2 is left as a function of θ instead of being evaluated at $\hat{\theta}_T$.

In general, a clear rank partition of the Jacobian matrix might not exist. But, since the null space of $\frac{\partial \rho}{\partial \theta'}(\theta_0)$ has dimension $p - r$, there exists a $p \times (p - r)$ full-rank matrix R_2 such that

$$\frac{\partial \rho}{\partial \theta'}(\theta_0) R_2 = 0.$$

Let R_1 be any $p \times r$ matrix such that $R = (R_1 | R_2)$ is a $p \times p$ nonsingular matrix and consider the parameterization $\eta = R^{-1}\theta$. Given R, it is obvious that the GMM overidentification test statistic is invariant when computed using the moment condition

$$\bar{\rho}(\eta) \equiv E \left(\psi(X_t, R\eta) \right) = 0$$

and the GMM estimator of η is $\hat{\eta}_T = R^{-1}\hat{\theta}_T$. The true value of η is $\eta_0 = R^{-1}\theta_0$. Let $\eta = (\eta_1', \eta_2')' \in \mathbb{R}^r \times \mathbb{R}^{p-r}$. Clearly, $\frac{\partial \bar{\rho}(\eta_0)}{\partial \eta_1'} = \frac{\partial \rho(\theta_0)}{\partial \theta'} R_1$ is full rank while $\frac{\partial \bar{\rho}(\eta_0)}{\partial \eta_2'} = \frac{\partial \rho(\theta_0)}{\partial \theta'} R_2 = 0$. In this context, the bootstrap estimating function is given by

$$\psi_t^{*(1)}(\eta) = \psi(X_t^*, R\eta) - E^* \left(\psi_t^*(\hat{\theta}_T) \right) - E^* \left(\frac{\partial}{\partial \theta'} \psi_t^*(\hat{\theta}_T) \right) R_2(\eta_2 - \hat{\eta}_{2T}), \tag{9}$$

where $\hat{\eta}_{2T}$ is the last $(p - r)$ components of $\hat{\eta}_T$. To implement the continuously corrected bootstrap, we can replace $\hat{\theta}_T$ in the first-order derivative by $R\eta$.

In applications, R has to be estimated in general. Next we show that the modified bootstrap based on (9) is still asymptotically valid when we replace R with a consistent estimator $\hat{R}$ and local identification is ensured at the second order.

We replace Assumption 6 with the following more general local identification assumption.

Assumption 6$'$

(i) $Rank \left(\frac{\partial \rho}{\partial \theta'}(\theta_0) \right) = r$ for some $0 \leq r \leq p$. If $r = 0$, let $R = R_2 = I_p$ and $R_1 = 0$; if $r = p$, let $R = R_1 = I_p$ and $R_2 = 0$. Otherwise, let $R = (R_1 | R_2)$ be a $p \times p$ nonsingular matrix such that $Rank \left(\frac{\partial \rho}{\partial \theta'}(\theta_0) R_1 \right) = r$ and $\frac{\partial \rho}{\partial \theta'}(\theta_0) R_2 = 0$.

(ii) For all u in the range of $\frac{\partial \rho'}{\partial \theta}(\theta_0)$ and for all v in the null space of its transpose, we have that

$$\left(\frac{\partial \rho}{\partial \theta'}(\theta_0) u + \left(v' \frac{\partial^2 \rho_h}{\partial \theta \partial \theta'}(\theta_0) v \right)_{1 \leq h \leq H} = 0 \right) \Rightarrow (u = v = 0).$$

Assumption 6$'$ generalizes Assumption 6 by allowing for any rank configuration of the expected Jacobian matrix as long as this is known information. The focus of this section is on the intermediate

rank deficiency case where $0 < r < p$, but Assumption 6' includes also as special cases $r = 0$ (studied in the previous section) as well as $r = p$ (the standard full column rank case). Part (ii) is the second-order local identification condition of Dovonon and Renault (2009). If $r = 0$, this condition boils down to our previous Assumption 6(ii). Assumption 6'(ii) is fulfilled by Examples 2.1 and 2.2 because the moment conditions involved are globally identified with second-order polynomial (in parameters) as estimating functions.

Let G_r be the $H \times (p - r)^2$ matrix such that, for all $v \in \mathbb{R}^{p-r}$,

$$\left(v' R_2' \frac{\partial^2 \rho_h}{\partial \theta \partial \theta'}(\theta_0) R_2 v \right)_{1 \leq h \leq H} \equiv G_r vec(vv')$$

and

$$J_1(v) = \left(X + \frac{1}{2} G_r vec(vv') \right)' W^{1/2} M W^{1/2} \left(X + \frac{1}{2} G_r vec(vv') \right),$$

where $X \sim N(0, \Sigma)$ and $M = I_H - W^{1/2} D (D'WD)^{-1} D'W^{1/2}$, $D = \frac{\partial \rho}{\partial \theta'}(\theta_0) R_1$.

In general, when $0 < r < p$, R is unknown and is not unique while a consistent estimator is required to implement the bootstrap algorithm. For this purpose, it is useful to fix one candidate that can be consistently estimated. Since $\partial \rho(\theta_0)/\partial \theta'$ is of rank r, it has r rows that are linearly independent. Let us assume, up to rearrangement of its rows and columns, that

$$\frac{\partial \rho}{\partial \theta'}(\theta_0) = \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix},$$

where M_{11} is an $r \times r$ nonsingular matrix and M_{ij}'s have conformable dimensions. Let

$$R_1 = \begin{pmatrix} M_{11}' \\ M_{12}' \end{pmatrix}, \quad R_2 = \begin{pmatrix} -M_{11}^{-1} M_{12} \\ I_{p-r} \end{pmatrix}, \quad \text{and} \quad R = (R_1 \mid R_2).$$

It is not hard to see that this choice of R fulfills the requirements of Assumption 6'(i). Also, R is a continuous function of the Jacobian matrix $\partial \rho(\theta_0)/\partial \theta'$ and a consistent estimator of R can be based on any consistent estimator of $\partial \rho(\theta_0)/\partial \theta'$, e.g. its sample mean analogue evaluated at the GMM estimator of θ_0.

Remark 1 *It is possible that $\partial \rho(\theta_0)/\partial \theta'$ is known to be of rank r $(0 < r < p)$ but its analysis does not make easily transparent a candidate nonsingular $r \times r$ submatrix M_{11} on which one can base the estimation procedure previously described. In this case, one can rely on the tests for rank of matrices proposed by Cragg and Donald (1997) and Wright (2003) that can be sequentially applied to detect the relevant submatrix.*

We have the following result.

Theorem 4.3 *Assume that Assumptions 1-5, 6' and 7 hold and $W_T^{*(1)} \overset{P^*}{\to} W$ in prob-P. Let $\hat{R}$ be a consistent estimator of R and let $\hat{J}_{T,1}^{*(1)}$ denote the bootstrap test statistic based on (9) with $\hat{R}$ replacing R. Then: (i) $\hat{J}_{T,1}^{*(1)} \overset{d^*}{\to} \min_{v \in \mathbb{R}^{p-r}} J_1(v)$, in prob-P, and (ii) $\sup_{x \in \mathbb{R}} \left| P^*(\hat{J}_{T,1}^{*(1)} \leq x) - P(\hat{J}_T \leq x) \right| \to 0$, in prob-P.*

21

Theorem 4.3 generalizes Theorem 4.1 by allowing for the possibility that the expected Jacobian matrix has an intermediate rank deficiency where $0 < r < p$. In this case, the theorem shows that using the bootstrap estimating function (9) with $\hat{R}$ in place of R is asymptotically valid provided $\hat{R}$ is consistent for R. Note that when $r = 0$, $R = R_2 = I_p$ and $R_1 = 0$, which implies that (9) is equivalent to the bootstrap estimating function underlying the corrected bootstrap studied in the previous section. Similarly, when $r = p$, $R = R_1 = I_p$ and $R_2 = 0$, in which case (9) is equivalent to the standard bootstrap estimating function of Hall and Horowitz (1996). Thus, Theorem 4.3 shows that the bootstrap is asymptotically valid for any value of r provided we know it and can use this information to design the bootstrap estimating function.

Remark 2 *Even though the stochastic process $J_1(v)$ depends on a specific choice of parameterization matrix $R = (R_1 \mid R_2)$, the limit distribution of the GMM overidentification test statistic, $\min_{v \in \mathbb{R}^{p-r}} J_1(v)$, does not depend on R. This is due to the fact that any candidate choice of parameterization matrix, say $\tilde{R}$, can be written $\tilde{R} = (R_1 Q_1 \mid R_2 Q_2)$ where Q_1 and Q_2 are (r, r) and $(p - r, p - r)$ nonsingular matrices, respectively and it is not hard to see that replacing R by $\tilde{R}$ in the definition of $J_1(v)$ does not change its minimum. This does not come as a surprise since the test statistic itself, as already mentioned, is not sensitive to such parameterizations. Nevertheless, one has to rely on a given parameterization to obtain the bootstrap critical values even though, similarly to the bootstrap test statistic, they are insensitive to any specific parameterization.*

We can use similar arguments to show the asymptotic validity of the continuously corrected bootstrap under Assumption 1-5, 6' and 7-9, provided Assumption 8 is rewritten as: "$E(\Phi(x, \eta)) = 0$ is uniquely solved by $\eta_0 = R^{-1}\theta_0$, with $\Phi(x, \eta) = \psi(x, R\eta) - \frac{\partial \psi}{\partial \theta'}(x, R\eta) R_2(\eta_2 - \eta_{02})$". As previously mentioned, this condition is not restrictive if the estimating function is quadratic in the parameters as in our two examples.

Theorem 4.3 can be applied to Example 2.2 by setting:

$$R_1 = \begin{pmatrix} 0 & -\sigma_0^2 & 0 & -\sigma_0^2 - \sigma_\varepsilon^2 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & -1 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & -1 & 0 \end{pmatrix}, \quad \text{and} \quad R_2 = \begin{pmatrix} 1/\sigma_\varepsilon^2 \\ 0 \\ -\sigma_0^2/\sigma_\varepsilon^2 \\ -1 \\ 1 \end{pmatrix}.$$

Consistent estimators of R_1 and R_2 can be obtained using GMM estimators of σ_0^2 and σ_ε^2.

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,

$$Y_{t+1} = \Lambda F_{t+1} + U_{t+1}, \tag{10}$$

where Λ is the $n \times K$ matrix of factor loadings, F_{t+1} is the $K \times 1$ vector of conditionally heteroskedastic and mutually independent factors and U_{t+1}, the $n \times 1$ vector of idiosyncratic shocks. We let $U_{t+1} \sim$ i.i.d. $N(0, aI_n)$, where a is a constant that determines the signal-to-noise ratio and I_n denotes the $n \times n$ identity matrix. The generic component f_{t+1} of F_{t+1} follows a Gaussian-GARCH model,

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

In addition, ε_t and U_t are mutually independent and independent of $\{F_\tau, Y_\tau : \tau \le t\}$.

Let $\mathfrak{F}_t$ be the increasing filtration given by the sigma algebra generated by returns heteroskedastic factors up to date t. We have:

$$Var(Y_{t+1}|\mathfrak{F}_t) = \Lambda D_t \Lambda' + \Omega,$$

where $\Omega = aI_n$ is the conditional variance of U_{t+1} and D_t is the diagonal $K \times K$ matrix of σ_t^2's. If $K = n$, there are no common conditionally heteroskedastic features for the returns in Y_{t+1} whereas $K < n$ implies that the returns have common conditionally heteroskedastic features (see Example 2.1).

We will test for common conditionally heteroskedastic features in Y_{t+1} using the moment condition (2) or more precisely its equivalent form exploited by D&R (2013), namely: $Cov(z_t, (\theta'Y_{t+1})^2) = 0$. However, this moment restriction is not feasible since $c(\theta) = E((\theta'Y_{t+1})^2)$ and $E(z_t)$ are unknown in general. As suggested by D&R (2013), we consider the feasible moment restriction that replaces these population means by their sample counterparts:

$$H_0 : E\left(\psi_{t,T}\left(\theta\right)\right) = 0, \quad \psi_{t,T}\left(\theta\right) \equiv \left(z_t - \bar{z}_T\right)\left(\left(\theta'Y_{t+1}\right)^2 - \bar{c}_T\left(\theta\right)\right),$$

with $\bar{c}_T\left(\theta\right) = T^{-1} \sum_{t=1}^{T} \left(\theta'Y_{t+1}\right)^2$ and $\bar{z}_T = \frac{1}{T} \sum_{t=1}^{T} z_t$. In order to globally identify θ, a normalization on θ is imposed. We follow D&R (2013) and assume that $\sum_{i=1}^{n} \theta_i = 1$ and set $\theta_n = 1 - \sum_{i=1}^{n-1} \theta_i$ in our derivations. Therefore, $p = n - 1$ in our applications.

We consider models with $n = 2$ and 3 asset return series and vary K, the number of factors, between 1 and 2 for $n = 2$ and 2 and 3 for $n = 3$. Each factor follows an independent GARCH(1,1) model as specified above with $\omega_1 = 0.2$, $\alpha_1 = 0.2$ and $\beta_1 = 0.6$ for the first factor; $\omega_2 = 0.2$, $\alpha_2 = 0.4$ and $\beta_2 = 0.4$ for the second factor, and $\omega_3 = 0.1$, $\alpha_3 = 0.1$ and $\beta_3 = 0.8$ for the third factor. The conditional variance-covariance matrix of the idiosyncratic errors U_{t+1} is $\Omega = aI_n$, where $a = 1/2$ and $a = 1$, corresponding to relatively high and low signal-to-noise ratios, respectively. This yields a total of 8 simulation designs, which are summarized in the following table.

Table 1: simulation designs

	Number of assets: n	Number of factors: K	Factor loadings: Λ	Signal-to-noise ratio: a $\Omega = aI_n$
Design 1	2	1	$\begin{pmatrix} 1 \\ 0.5 \end{pmatrix}$	0.5
Design 2	3	2	$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0.5 & 0.5 \end{pmatrix}$	0.5
Design 3	2	1	$\begin{pmatrix} 1 \\ 0.5 \end{pmatrix}$	1
Design 4	3	2	$\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0.5 & 0.5 \end{pmatrix}$	1
Design 5	2	2	I_2	0.5
Design 6	3	3	I_3	0.5
Design 7	2	2	$\begin{pmatrix} 1 & 0 \\ 0.5 & 1 \end{pmatrix}$	1
Design 8	3	3	$\begin{pmatrix} 1 & 0 & 2 \\ 1 & 1 & 0 \\ 0.5 & 0.5 & 2 \end{pmatrix}$	1

The first four designs correspond to models that have either 2 or 3 assets and where the number of factors is one less than the number of assets so that the null hypothesis of a common heteroskedastic factor is true. In particular, the factor loading matrix Λ is of dimension $n \times K$ with $K = n-1$. Given the normalization of θ, the moment conditions tested by H_0 have only one solution for these designs so that global identification of the co-feature vector is ensured and our theory applies. Designs 1 and 2 were considered by D&R (2013) and correspond to a relatively high level of signal-to-noise ratio with $a = 0.5$. Designs 3 and 4 are new and correspond to a lower signal-to-noise ratio $a = 1$. Fiorentini, Sentana and Shephard (2004) considered both scenarios in their seminal paper on the estimation of conditionally heteroskedastic factor models. Designs 5 through 8 simulate models under the alternative hypothesis that there are no common factors among the n assets so that $K = n$. The factor loadings and the value of Ω is the same as in D&R (2013) for designs 5 and 6; designs 7 and 8 correspond to $a = 1$.

The GMM estimations are carried out with $H = 2$ and $z_t = \left(Y_{1,t}^2, Y_{2,t}^2\right)'$ for the designs with 2 assets, and with $H = 3$ and $z_t = \left(Y_{1,t}^2, Y_{2,t}^2, Y_{3,t}^2\right)'$ for the designs with 3 assets. The sample sizes start with $T = 50, 500, 1000$ and then increase by increments of 1000 up to $20,000$. Because the convergence rate of the GMM estimator is $T^{1/4}$, it is important to allow for sample sizes larger than usual to evaluate the performance of the methods in this context. The simulated rejection rates, computed at the nominal level of 5%, are based on 10,000 Monte Carlo replications with 399 bootstrap replications for each Monte Carlo replication throughout. Figures 1 through 8 contain the results.

The bootstrap samples consist of random draws with replacement[5] of $T - 1$ realizations (Y_{t+1}^*, z_t^*)

---

[5]This i.i.d. bootstrap scheme is justified by the fact that the estimating function $\psi_{t,T}(\theta)$ has its sample mean

from $\{(Y_2, z_1), (Y_3, z_2), \cdots, (Y_T, z_{T-1})\}$. The bootstrap estimating function is given by

$$\psi^*_{t,T}(\theta) = (z^*_t - \bar{z}_T)(\theta' Y^*_{t+1} - \bar{c}_T(\theta)).$$

The standard bootstrap (labeled as "BootST" in the figures) and the modified bootstrap methods proposed in Section 4 are considered. The choice of the weighting matrices is done according to (8) with $\psi^*_{c,t}(\cdot)$ replaced by $\psi^{*(1)}_t(\cdot)$ and $\psi^{*(2)}_t(\cdot)$ for the corrected ("Boot-(1)") and continuously-corrected bootstraps ("Boot-(2)") , respectively.

For comparison purposes, we also include asymptotic-based tests. For the designs with 2 assets (cf. Figures 1, 3, 5 and 7), the number p of parameters in θ is equal to 1. Thus, the resulting overidentification test statistic is asymptotically distributed as a fifty-fifty mixture of χ^2_1 and χ^2_2 when the null is true (cf. Designs 1 and 3). Rejection rates obtained with this asymptotic distribution are labeled as "Asymp". For the designs with 3 assets (cf. Figures 2, 4, 6 and 8), $p = 2$ and therefore we do not know the critical values of the null 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 χ^2_3. We label these results as "Chi2-3". We also include the rejection rates associated with the standard χ^2_1 critical values ($H = 3$ and $p = 2$ implies $H - p = 1$). Alternatively, we can simulate critical values from the limiting distribution $\underline{J} \equiv \min_{v \in \mathbb{R}^p} J(v)$, where the stochastic process $J(v)$ is defined in (7) and depends on unknown quantities (such as Σ and G) that need to be replaced by consistent estimates. When p is relatively small (as in our designs with 3 assets, where $p = 2$), this approach is a feasible alternative to the bootstrap, but it might become computationally very demanding when p is large due to the fact that we need to minimize $J(v)$ over $v \in \mathbb{R}^p$. Results obtained with this approach are labeled as "Sim-Asymp" in Figures 2, 4, 6 and 8.

As mentioned already in the Introduction, another approach to conduct overidentification tests under first-order underidentification is to exploit the information contained in the zero Jacobian matrix and it is interesting to include this approach in our simulations. We follow Lee and Liao (2016) and consider two alternative implementations. Both are based on an overidentification test based the moment conditions $E(m_t(\theta)) = 0$, with $m_t(\theta) = \begin{pmatrix} \psi_{t,T}(\theta) & g_{t,T}(\theta) \end{pmatrix}'$, where $\psi_{t,T}(\theta)$ are the original moment conditions and where $g_{t,T}(\theta) = vec\left(\left(\frac{\partial \psi_{t,T}(\theta)}{\partial \theta'}\right)'\right)$ is of size $Hp \times 1$, but differ in the way they estimate θ. One method estimates θ with the efficient GMM estimator based on the full set of moment conditions $E(m_t(\theta)) = 0$. The resulting overidentification test is asymptotically distributed as χ^2_{H+Hp-p}. The results obtained with this method are labeled "Eff-GMM". A second method estimates θ using only the moment conditions implied by the zero Jacobian matrix $E(g_{t,T}(\theta)) = 0$ (these are sufficient to ensure the global and local first-order identification of θ_0, as shown by Lemma 4.3 of Lee and Liao (2016)) but then constructs an overidentification test for the full set of moment

---

that is equal to the sample mean of a martingale difference sequence up to an $o_P(T^{-1/2})$ term. Specifically, $\bar{\psi}_T(\theta) = \bar{v}(\theta) + \bar{\psi}_{0T}(\theta) + o_P(T^{-1/2})$, where $\bar{v}(\theta) = \frac{1}{T}\sum_{t=1}^T \mu_z(c(\theta) - (\theta' Y_{t+1})^2)$, $\bar{\psi}_{0T}(\theta) = \frac{1}{T}\sum_{t=1}^T z_t((\theta' Y_{t+1})^2 - c(\theta))$; $\mu_z = E(z_t)$ (see Equation (9) of D&R (2013)). Note that $(\theta' Y_{t+1})^2 - c(\theta)$ is a martingale difference sequence with respect to $\mathfrak{F}_t$ under the null of constant conditional variance of $\theta' Y_{t+1}$.

conditions $m_t(\theta)$. Note that the GMM estimator $\hat{\theta}_g$ is very easy to compute since the estimating equations $g_{t,T}(\theta)$ are linear in θ in our application. Following Lee and Liao (2016), we use the identity matrix to obtain $\hat{\theta}_g$ and to compute the corresponding overidentification test. Its asymptotic null distribution is non-pivotal but can be simulated (cf. Theorem 2.2 of Lee and Liao (2016)). We label the results of this test as "Sim-GMM".

Figure 1 gives the simulated null rejection rates for Design 1. This design was considered by D&R (2013) and corresponds to a design where the null hypothesis is true and the signal-to-noise ratio is relatively high. Since $n = 2$, $p = 1$ and therefore critical values for the standard overidentification asymptotic test under the non-standard asymptotics are available ("Asymp"). The results confirm the failure of the standard bootstrap ("BootST") reflected by a systematic over-rejection of the null hypothesis with a rejection rate above 8% in large samples. The corrected bootstrap ("Boot-(1)") also over-rejects for small sample sizes, but its rejection rate declines as the sample size grows. The continuously-corrected bootstrap ("Boot-(2)") tracks closely the asymptotic distribution, with rejection rates close to 5% for almost all samples sizes. From these results, the continuously-corrected bootstrap is noticeably better than the corrected bootstrap, a feature that we observe throughout the remaining designs satisfying the null hypothesis. The results from the overidentification test for the augmented set of moment conditions that include those implied by the zero Jacobian matrix are interesting. The efficient GMM approach ("Eff-GMM") over-rejects the null by a large amount, especially for the smaller sample sizes, where it is even worse than the standard bootstrap. However, its rejection rates converge to the nominal level as we increase the sample size. Instead, the simulated approach of Lee and Liao (2016) ("Sim-GMM") tends to under-reject under the null, even at large sample sizes. Overall, the best approaches for this DGP are "Boot-(2)" and "Asymp", closely followed by "Sim-GMM".

Figure 2 shows the results for Design 2 where the number of assets is 3 and the number of factors is 2 (so that the null is still true) and the signal-to-noise ratio is as in Figure 1. Note that critical values for the asymptotic distribution are not available for this design, but we can simulate them using consistent estimates of the nuisance parameters that enter $J(v)$. As observed 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 χ_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 this conservative test with the continuously-corrected bootstrap). We can also see that using critical values from the standard χ_1^2 (thereby ignoring the local identification failure) leads to large over-rejections under the null. Both "Boot-(1)" and "Boot-(2)" yield null rejection rates that tend to the nominal level of 5%, with the difference that "Boot-(1)" is oversized whereas "Boot-(2)" is undersized. The extent of size distortions for "Boot-(2)" are nevertheless very small, with rejection rates between 3.8 and 5% across all sample sizes. It is interesting to note that "Boot-(2)" tracks closely the performance

of "Sim-Asymp" for the largest sample sizes but dominates it for values of $T \leq 8000$, avoiding the larger under-rejections that characterize "Sim-Asymp". For these smaller sample sizes, "Boot-(2)" also outperforms "Sim-GMM", which tends to over-reject. As in Figure 1, "Sim-GMM" is much better behaved than its efficient GMM version which is grossly over-sized.

The results for Designs 3 and 4 are shown in Figures 3 and 4. These figures are the analogues of Figures 1 and 2, with the difference that we increase the idiosyncratic errors variances from 0.5 to 1. The effect of this lower signal-to-noise ratio is to increase the actual null rejection rates of all methods. For bivariate returns, a comparison between Designs 3 and 1 shows that the rejection curves for all methods shift up when a increases from 0.5 to 1. This translates into over-rejections for all methods for all sample sizes, except the smallest ones. The best bootstrap method is still "Boot-(2)", but this method now over-rejects (whereas it underrejected slightly in Design 1), except for $T \leq 1000$. The degree of over-rejection is much smaller than that of "Boot-(1)", "BootST" and "Eff-GMM" (which is the worst procedure for this DGP), but is larger than that for "Sim-GMM" and "Asymp", which show smaller size distortions. In particular, the results favor "Sim-GMM" for the largest sample sizes. Figure 4 considers the same low signal-to-noise ratio but for trivariate returns. The comparison with Figure 2 shows that increasing a to 1 when $n = 3$ also shifts all rejection curves upwards. This implies an over-rejection for all methods for all but the smallest values of T. An exception is the approach based on the χ_3^2 distribution, which is undersized for all values of T, although much less so than when $a = 0.5$. The ranking between "Boot-(2)" and "Sim-GMM" favors the bootstrap, whose rates are closer to the nominal 5% level and converge to those of "Sim-Asymp". For this DGP, "Sim-Asymp" is the best performing method when $T \geq 4000$ (excluding the conservative approach based on χ_3^2).

The main conclusion from Figures 1 through 4 is that the continuously-corrected bootstrap "Boot-(2)" is the best performing bootstrap method in terms of size control. It performs similarly to the approach based on either the true asymptotic distribution (for $n = 2$) or its simulated version (for $n = 3$) when T is large, but a clear ranking cannot be established for the smaller sample sizes as it depends on the DGP. The simulated GMM approach of Lee and Liao (2016) is also able to deliver good size control when T is large, but it is clearly dominated by the continuously-corrected bootstrap for all designs but Design 3 (with $n = 2$, $K = 1$ and $a = 1$).

We now turn our attention to the analysis of power, based on Figures 5 through 8. Starting with Figure 5, where $n = 2$, $K = 2$, and $a = 0.5$, we see that "Boot-(2)" has the lowest power among the methods considered, including "Sim-GMM" and "Asymp", the closest procedures in terms of size control. So, for this alternative, it is clear that the cost of the good size properties of "Boot-(2)" is a loss of power. The ranking between "Boot-(2)" and "Sim-GMM" is reversed in Figure 7, where $n = 2$, $K = 2$ but $a = 1$, implying that a clear ranking between the two methods cannot be given under the alternative. Figures 6 and 8 are the analogues of Figures 5 and 7, but for trivariate returns. The patterns are largely the same. For $a = 0.5$, Figure 6 shows that "Boot-(2)" and "Sim-Asymp" perform very similarly in terms of power, and both are less powerful than the remaining approaches, except

27

for the conservative approach based on χ_3^2. For $a = 1$, the main feature of notice in Figure 8 is that "Sim-GMM" is now the least powerful approach, being even less powerful than the "Chi2-3" method. This occurs despite the fact that "Sim-GMM" over-rejects under the null when $a = 1$ (cf. Figure 4). Overall, Figures 5-8 suggest that there is no clear ranking in terms of power among "Boot-(2)" and "Sim-GMM". When $p = 1$ and the asymptotic distribution is fully known, the trade-off between power and size favors the asymptotic test based on the mixture of chi-square distributions, but this is not necessarily true when $p > 1$. In this case, we need to simulate an estimated version of the asymptotic distribution of the test, making this approach comparable to the bootstrap in terms of power.
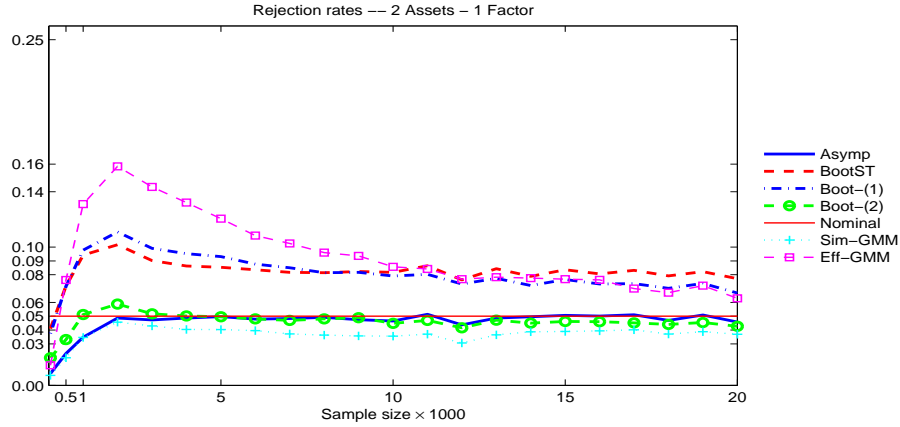

Figure 1: Size properties for Design 1 ($n = 2$, $K = 1$, $a = 0.5$)

Figure 2: Size properties for Design 2 ($n = 3$, $K = 2$, $a = 0.5$)

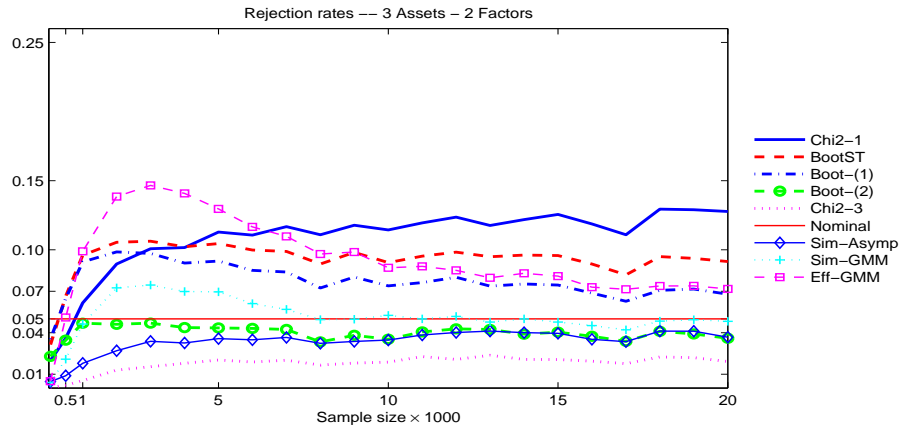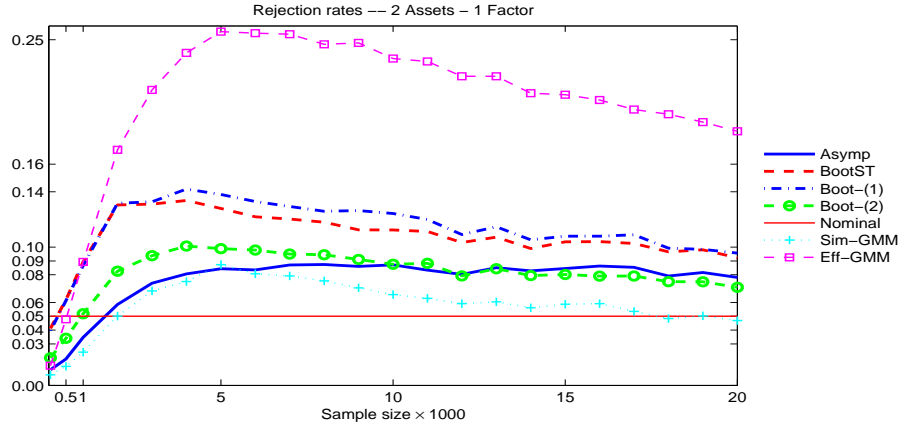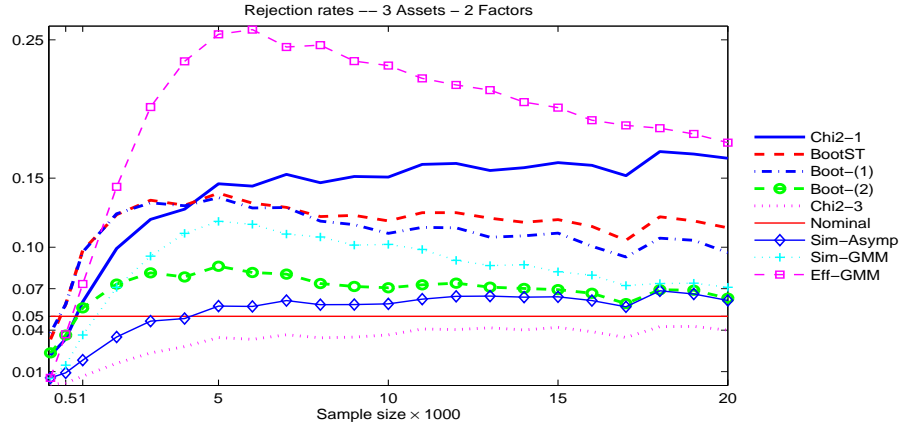Figure 3: Size properties for Design 3 ($n = 2$, $K = 1$, $a = 1$)

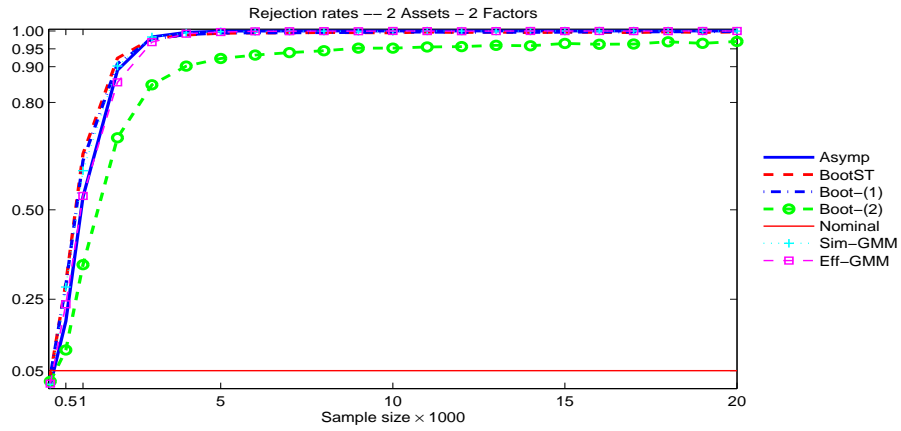Figure 4: Size properties for Design 4 ($n = 3$, $K = 2$, $a = 1$)

Figure 5: Power properties for Design 5 ($n = 2$, $K = 2$, $a = 0.5$)
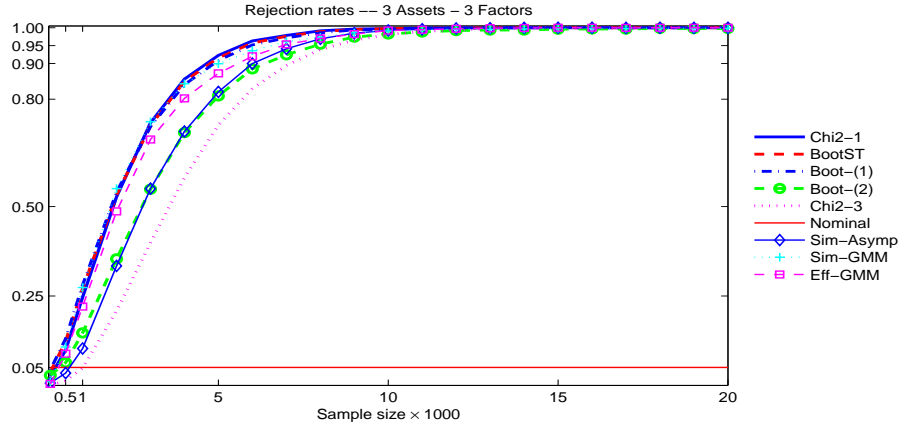
29

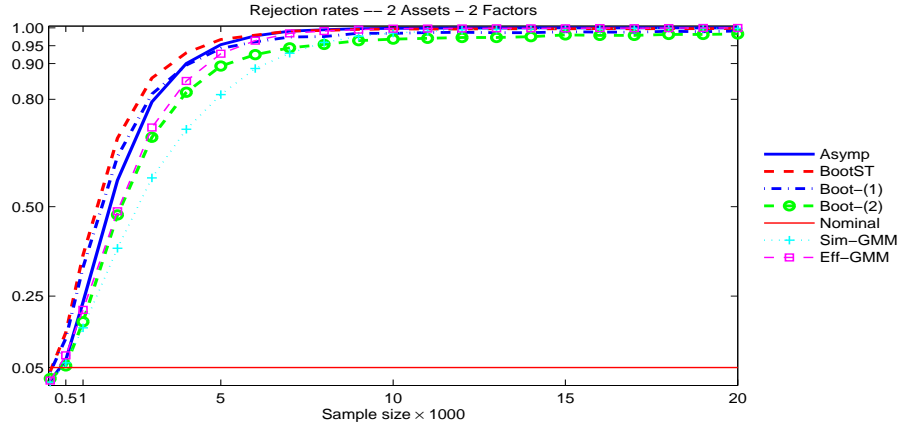Figure 6: Power properties for Design 6 ($n = 3$, $K = 3$, $a = 0.5$)

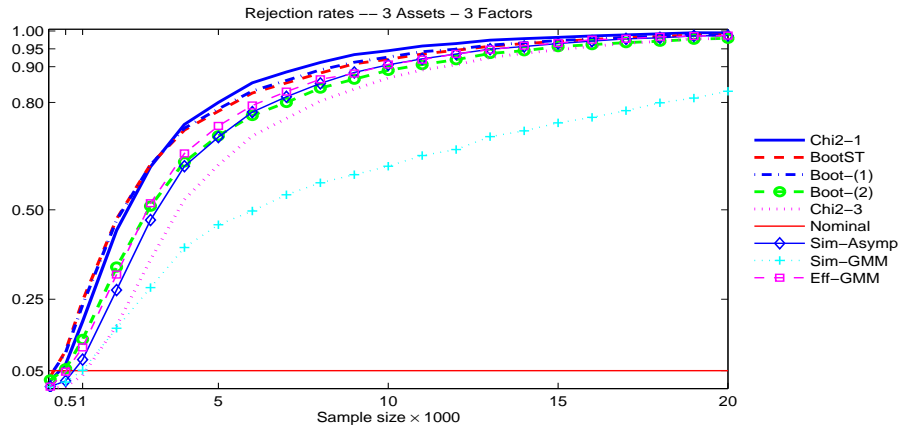Figure 7: Power properties for Design 7 ($n = 2$, $K = 2$, $a = 1$)

Figure 8: Power properties for Design 8 ($n = 3$, $K = 3$, $a = 1$)

30

6 Conclusion and some possible extensions

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 at the first order. In particular, we focus on the special case of a degenerate rank identification condition, which was recently analyzed by Dovonon and Renault (2013b) 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 $\theta - \hat{\theta}_T$. 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. Several extensions of our work are worth considering in future work. From the simulations, the continuously-corrected bootstrap test has better properties under the null than the corrected bootstrap test. It would be interesting to investigate this analytically through higher order expansions of the bootstrap statistics. Such expansions are made complicated by the non-standard nature of the problem being studied. Another interesting avenue for future research is the study of the asymptotic uniform validity of our bootstrap tests when we relax the assumption of a zero Jacobian matrix. A recent contribution along this line is Andrews and Guggenberger (2015). Finally, bootstrapping the distribution of the GMM estimators $\hat{\theta}_T$ themselves when the Jacobian matrix is rank deficient is an interesting unresolved question. This is a difficult extension because the asymptotic distribution of $\hat{\theta}_T$ has not yet been studied under general form of rank condition failure. In recent work, Dovonon and Hall (2015) have derived the asymptotic distribution of the GMM estimator in the special case where the Jacobian matrix is of rank $p - 1$. Their results highlight the non standard nature of this asymptotic distribution which also involves nuisance parameters. This echoes the results of Sargan (1983) who has derived the asymptotic distribution for nonlinear IV regressions when the rank of the Jacobian matrix is equal to $p - 1$ and showed that the asymptotic distribution of the IV estimators is a mixture of two conditional distributions. Given that these distributions are difficult to estimate in practice, developing a bootstrap method that replaces analytical approximations would be an important contribution to this literature, in particular with the focus on general form of rank condition failure.

Appendix A: Proof of results in Section 2

This Appendix contains the proofs of results Section 2. The following lemma will be useful to establish the continuity of the distribution of $\underline{J}$ as stated in Theorem 2.3. This result relies on some notions of real algebraic geometry and requires that we introduce the notion of real semialgebraic set. We refer to Bochnak, Coste and Roy (1998) for more details.

Definition A.1 *A real semialgebraic subset of $\mathbb{R}^n$ is a subset of $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ satisfying a boolean combination of polynomial equations and inequalities with real coefficients.*

Lemma A.1 *Let $(x, y) \mapsto f(x, y)$, with $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$ be a polynomial function of the components of x and y with coefficients in $\mathbb{R}$. Let C be a finite subset of $\mathbb{R}$ and define:*

$$A = \left\{ x \in \mathbb{R}^n : \min_{y \in \mathbb{R}^m} f(x, y) \text{ is reached and belongs to } C \right\}.$$

Then, A is a semialgebraic subset of $\mathbb{R}^n$ and, as such, if the Lebesgue measure of A is positive, then A has a nonempty interior.

Proof of Lemma A.1. We first show that A is a semialgebraic set. Since finite unions of semialgebraic sets are also semialgebraic sets, we assume without loss of generality that C is a singleton containing c^*. The sets

$$A^1 = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m : f(x, y) = c^*\},$$

$$A^2 = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m : f(x, y) < c^*\}$$

are semialgebraic subsets of $\mathbb{R}^n \times \mathbb{R}^m$. Therefore, by Theorem 2.2.1 of Bochnak, Coste and Roy (1998), $\pi(A^1)$ and $\pi(A^2)$ are also semialgebraic subsets of $\mathbb{R}^n$, where $\pi(\cdot)$ is the projection of $\mathbb{R}^n \times \mathbb{R}^m$ on the space of the first n coordinates. It is not hard to see that :

$$A = \pi(A^1) \setminus \pi(A^2)$$

showing that A is a semialgebraic subset of $\mathbb{R}^n$.

We now show the second part of the lemma which is a rather general property of semialgebraic sets with positive Lebesgue measure. By Proposition 2.9.10 of Bochnak, Coste and Roy (1998), as a semialgebraic set, A is a disjoint union of a finite number of Nash manifolds M_i of $\mathbb{R}^n$ that are each diffeomorphic to an open hypercube $]0, 1[^{dim(M_i)}$ (with $]0, 1[^0$ being a point). Those M_i's with dimension smaller than n have Lebesgue measure null in $\mathbb{R}^n$. Since A has positive Lebesgue measure, at least one of these manifolds has positive Lebesgue measure, hence is diffeomorphic to $]0, 1[^n$ and therefore is a (nonempty) open subset of $\mathbb{R}^n$. This shows that A has a nonempty interior.

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 ψ considered here are not necessarily quadratic. For each $h = 1, \ldots, H$, by a Taylor expansion of $\psi_{t,h}\left(\hat{\theta}_T\right) \equiv \psi_h\left(X_t, \hat{\theta}_T\right)$ around θ_0, we have that

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

where $\ddot{\theta}_T$ is a $p \times 1$ vector on the segment connecting $\hat{\theta}_T$ and θ_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 $h = 1, \ldots, H$, summing over t and multiplying the result by $\sqrt{T}$ yields

$$\sqrt{T}\bar{\psi}_T\left(\hat{\theta}_T\right) = \sqrt{T}\bar{\psi}_T(\theta_0) + \sqrt{T}\frac{\partial \bar{\psi}_T(\theta_0)}{\partial \theta'}\left(\hat{\theta}_T - \theta_0\right) + \frac{1}{2}R_T\left(\ddot{\theta}_T\right), \tag{A.1}$$

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

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

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

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

since $tr\left(AB\right) = vec\left(A'\right)' vec\left(B\right)$. It follows that

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

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 (6) evaluated at $\ddot{\theta}_T$. Given our assumptions and the fact that $\ddot{\theta}_T \xrightarrow{P} \theta_0$, 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,

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

This implies that

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

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

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

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 $(vec'\left(Z_T(\theta_0)\right), \hat{v}_T')' = O_P(1)$. 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 $(vec'(Z(X)), V')'$, say, where

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

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

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

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 $e \in \mathbb{R}^p$,

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

33

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

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

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

$$\tilde{Z}_T = Z_T(\theta_0) + \frac{1}{2}\left(\frac{\partial^2\rho'(\theta_0)}{\partial\theta_i\partial\theta_j}WG\,vec\,(\hat{v}_T\hat{v}'_T)\right)_{1\leq i,j\leq p} + o_P(1), \tag{A.3}$$

and

$$N_T = \left(\hat{v}'_T \frac{\partial^2\rho'(\theta_0)}{\partial\theta_i\partial\theta}W\frac{\partial^2\rho(\theta_0)}{\partial\theta_j\partial\theta'}\hat{v}_T\right)_{1\leq i,j\leq p} + o_P(1). \tag{A.4}$$

To show (A.3), we observe that from part (i), (A.2) implies that

$$\sqrt{T}\bar{\psi}_T\left(\hat{\theta}_T\right) = \sqrt{T}\bar{\psi}_T\left(\theta_0\right) + \frac{1}{2}G\,vec\,(\hat{v}_T\hat{v}'_T) + o_P(1). \tag{A.5}$$

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

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

which together with (A.5) implies (A.3). To obtain (A.4), by a mean-value expansion of $\frac{\partial\bar{\psi}_T(\hat{\theta}_T)}{\partial\theta_i}$ around θ_0, we get that for all $i = 1, \ldots, p$,

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

where $\bar{\theta} \in (\theta_0, \hat{\theta}_T)$ and may differ from row to row. We obtain (A.4) by writing this equation for any $i, j = 1, \ldots, p$ and taking the inner product of each hand side.

Proof of Theorem 2.3. We follow the proof of Lemma B.6 of Dovonon and Renault (2013a) (henceforth D&R (2013a)). 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

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

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

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

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 θ_0 is an interior point of Θ and $\mathbb{K}$ is compact (hence bounded), then $\ddot{\theta}(v) \in \mathcal{N}$, a neighborhood of θ_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 $\bar{G}\left(\ddot{\theta}(v)\right) - \bar{G}(\theta_0)$, say $\bar{G}_h\left(\ddot{\theta}(v)\right) - \bar{G}_h(\theta_0)$ satisfies

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

34

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

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

implying that

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

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 (2013a), the rest of the proof follows exactly the same lines as the proof of D&R's (2013a) Lemma B.6. To complete the proof of Theorem 2.3, 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 $\forall c \in \mathbb{R}, \quad \text{Prob}(\underline{J} = c) = 0$. It is easy to see that

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

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

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

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 has a positive Lebesgue measure. Also, $J(x,v)$ is a polynomial function of the components of x and v and for each $x \in \mathbb{R}^H$, $\min_{v \in \mathbb{R}^p} J(x,v)$ is reached. (see Lemma B.6(ii) of D&R (2013a), for a proof). This is essentially due to the fact that, for each $x \in \mathbb{R}^H$, $J(x,v) \to \infty$ as $\|v\| \to \infty$.) From Lemma A.1, we deduce that A has a nonempty interior and therefore, contains an open ball centered at a certain $x_0 \in \mathbb{R}^H$, say $\mathcal{B}(x_0, \epsilon) = \{x \in \mathbb{R}^H : \quad \|x - x_0\| < \epsilon\}$ for some $\epsilon > 0$. This implies that $\forall d \in \mathbb{R}^H : \quad \|d\| < \epsilon, \quad \min_{v \in \mathbb{R}^p} J(x_0 + d, v) = c^*$. Let $v_0 \in \arg\min_{v \in \mathbb{R}^p} J(x_0, v)$. Since $x_0 \in A$, $J(x_0, v_0) = c^*$ and $\forall d \in \mathbb{R}^H : \quad \|d\| < \epsilon$, we have that

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

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

$$2d'\left(W(x_0 + \frac{1}{2}G\,vec(v_0v_0')) + \frac{1}{2}Wd\right) \geq 0, \quad \forall d \in \mathcal{B}(0, \epsilon). \tag{A.6}$$

Note that since $J(x_0, v_0) = c^* = a'W^{-1}a > 0$, we must have $a \equiv W(x_0 + \frac{1}{2}GVec(v_0v_0')) \neq 0$. Fix $a_j \neq 0$, the jth element of a that is different from zero. Equation (A.6) 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 $L = \min_{u \in \mathbb{R}^{p^2}} \left(X + \frac{1}{2}Gu\right)' W \left(X + \frac{1}{2}Gu\right)$. Clearly, $L \leq \underline{J}$. Assume that $\text{Rank}(G) = r$. Given the second-order local 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

$$L = \left(X + \frac{1}{2}G_1G_2\hat{u}\right)' W \left(X + \frac{1}{2}G_1G_2\hat{u}\right) = X'W^{1/2}PW^{1/2}X, \tag{A.7}$$

35

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 $D = diag(I_{H-r}, 0)$. 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 Sections 3 and 4

B.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:
$\mathcal{X}_T \equiv \{X_1(\omega), X_2(\omega), \ldots, X_T(\omega)\}$ from a given realization $\omega \in \Omega$. Suppose we obtain a bootstrap sample $\mathcal{X}_T^* = \{X_1^*, X_2^*, \ldots, X_T^*\}$ by resampling from $\mathcal{X}_T$. For each $\omega \in \Omega$, we view $\{X_t^* : t = 1, \ldots, T\}$ as the realization of a stochastic process defined on $(\Lambda, \mathcal{G}, P^*)$, another probability space, such that for each $t = 1, \ldots, T$,

$$X_t^*(\omega, \lambda) = X_{\tau_t(\lambda)}(\omega), \tag{B.1}$$

where $\tau_t : \Lambda \to \{1, 2, \ldots, N\}$ denotes the random index generated by the resampling scheme for each $t = 1, 2, \ldots, T$ (independently of ω). 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 λ, conditional on ω. We write E^* and Var^* to denote the expected value and the variance with respect to P^*, respectively. As (B.1) makes clear, X_t^* depends on two sources of randomness, one related to the observed data (and indexed by ω) 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 $\varepsilon, \delta > 0$, $P(P^*(|Z_T^*| > \varepsilon) > \delta) \to 0$ as $T \to \infty$. Similarly, we say that $Z_T^* = O_{P^*}(1)$ in prob-P if for any $\delta > 0$, there exists $0 < M < \infty$ such that $P(P^*(|Z_T^*| \geq M) > \delta) \to 0$ as $T \to \infty$. Lemma B.2 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 B.2 *Suppose that*

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

*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 $Z_T^* \to^{d^*} Z$, in prob-P, if $E^*f(Z_T^*) \to E(f(Z))$ 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 $h : \ell^\infty(\mathbb{K}) \to \mathbb{R}$ 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 $E^*(h(Z_T^*)) \to E(h(Z))$ in prob-P for any $h : \ell^\infty(\mathbb{K}) \to \mathbb{R}$ continuous and bounded with respect to the sup norm. The following result is the bootstrap version of Lemma B.6 of D&R (2013a) and is useful to prove Lemma 4.1.

36

Lemma B.3 *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 B.3. We show that $E^*(h(J_T^*(\hat{v}_T^*))) - E(h(J(\hat{v}))) \xrightarrow{P} 0$ for any continuous and bounded function $h : \mathbb{R} \to \mathbb{R}$. This is equivalent to showing that for any $\epsilon, \delta > 0$,

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

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

$$P\left(P^*(\|\hat{v}_T^*\| > A) > \frac{\epsilon}{6M}\right) < \frac{\delta}{3}, \quad \text{and} \quad P\left(\|\hat{v}\| > A\right) < \frac{\delta}{3}. \tag{B.2}$$

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 $\min_{\|v\| \le A} J_T^*(v) \xrightarrow{d^*} \min_{\|v\| \le A} J(v)$, in prob-P. Hence, with $\hat{v}_1^*$ and $\hat{v}_2$ denoting the argument of the minimum over $\mathbb{K} \equiv \{\|v\| \le A\}$ 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,

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

By the triangle inequality, we can write

$$
\begin{aligned}
|E^*(h(J_T^*(\hat{v}_T^*))) - E(h(J(\hat{v})))| \le {} & |E^*(h(J_T^*(\hat{v}_T^*))) - E^*(h(J_T^*(\hat{v}_1^*)))| \\
& + |E^*(h(J_T^*(\hat{v}_1^*))) - E(h(J(\hat{v}_2)))| \\
& + |E(h(J(\hat{v}_2))) - E(h(J(\hat{v})))| .
\end{aligned} \tag{B.4}
$$

It follows that

$$
\begin{aligned}
P\left(|E^*(h(J_T^*(\hat{v}_T^*))) - E(h(J(\hat{v})))| > \epsilon\right) \le {} & P\left(|E^*(h(J_T^*(\hat{v}_T^*))) - E^*(h(J_T^*(\hat{v}_1^*)))| > \frac{\epsilon}{3}\right) \\
& + P\left(|E^*(h(J_T^*(\hat{v}_1^*))) - E(h(J(\hat{v}_2)))| > \frac{\epsilon}{3}\right) \\
& + P\left(|E(h(J(\hat{v}_2))) - E(h(J(\hat{v})))| > \frac{\epsilon}{3}\right) \equiv I_1 + I_2 + I_3.
\end{aligned} \tag{B.5}
$$

To bound I_1, note that

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

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

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

where we have bounded the function h by its upper bound M. Thus, by (B.2),

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

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

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

given (B.2). The result now follows from (B.5).

37

B.2 Proofs of bootstrap results

Proof of Proposition 3.1. We would like to show that for any $\varepsilon > 0$, $\lim_{T \to \infty} P\left(P^*\left(\left\|\hat{\theta}_T^* - \theta_0\right\| > \varepsilon\right) > \varepsilon\right) = 0$. Because θ_0 is the unique minimizer of $Q(\theta)$, for any $\varepsilon > 0$ such that $\|\theta - \theta_0\| > \epsilon$, there is $\delta > 0$ such that $Q(\theta) - Q(\theta_0) \geq \delta > 0$. It follows that

$$P^*\left(\left\|\hat{\theta}_T^* - \theta_0\right\| > \varepsilon\right) \leq P^*\left(Q\left(\hat{\theta}_T^*\right) - Q(\theta_0) > \delta\right)$$

$$\leq P^*\left(Q\left(\hat{\theta}_T^*\right) - Q_T\left(\hat{\theta}_T^*\right) + Q_T\left(\hat{\theta}_T^*\right) - Q_T^*\left(\hat{\theta}_T^*\right) + Q_T^*\left(\hat{\theta}_T^*\right) - Q_T\left(\hat{\theta}_T\right) + Q_T\left(\hat{\theta}_T\right) - Q(\theta_0) > \delta\right)$$

$$\leq P^*\left(Q\left(\hat{\theta}_T^*\right) - Q_T\left(\hat{\theta}_T^*\right) + Q_T\left(\hat{\theta}_T^*\right) - Q_T^*\left(\hat{\theta}_T^*\right) + Q_T^*\left(\hat{\theta}_T\right) - Q_T\left(\hat{\theta}_T\right) + Q_T(\theta_0) - Q(\theta_0) > \delta\right)$$

$$\leq P^*\left(2\sup_{\theta \in \Theta}|Q_T(\theta) - Q(\theta)| > \delta/2\right) + P^*\left(2\sup_{\theta \in \Theta}|Q_T^*(\theta) - Q_T(\theta)| > \delta/2\right)$$

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

$$P\left(P^*\left(\left\|\hat{\theta}_T^* - \theta_0\right\| > \varepsilon\right) > \varepsilon\right)$$

$$\leq P\left(P^*\left(2\sup_{\theta \in \Theta}|Q_T(\theta) - Q(\theta)| > \delta/2\right) > \varepsilon/2\right) + P\left(P^*\left(2\sup_{\theta \in \Theta}|Q_T^*(\theta) - Q_T(\theta)| > \delta/2\right) > \varepsilon/2\right)$$

$$\leq \frac{1}{\varepsilon/2}P\left(2\sup_{\theta \in \Theta}|Q_T(\theta) - Q(\theta)| > \delta/2\right) + P\left(P^*\left(2\sup_{\theta \in \Theta}|Q_T^*(\theta) - Q_T(\theta)| > \delta/2\right) > \varepsilon/2\right), \qquad \text{(B.6)}$$

where the first term in the second inequality is obtained by first noting that $P^*\left(2\sup_{\theta \in \Theta}|Q_T(\theta) - Q(\theta)| > \delta/2\right) = 1\{\cdot\}$ (where $1\{\cdot\}$ denotes the indicator function containing the expression inside $P^*(\cdot)$) and then applying Markov's inequality. From (B.6), 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,

$$Q_T^*(\theta) = \bar{\psi}_{c,T}^{*\prime}(\theta)W_T^*\bar{\psi}_{c,T}^*(\theta),$$

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

$$Q_T^*(\theta) = \left(\bar{\psi}_T^*(\theta) - \bar{\psi}_T\left(\hat{\theta}_T\right)\right)'W_T^*\left(\bar{\psi}_T^*(\theta) - \bar{\psi}_T\left(\hat{\theta}_T\right)\right)$$

$$= \bar{\psi}_T^{*\prime}(\theta)W_T^*\bar{\psi}_T^*(\theta) - 2\bar{\psi}_T^{*\prime}(\theta)W_T^*\bar{\psi}_T\left(\hat{\theta}_T\right) + \bar{\psi}_T'\left(\hat{\theta}_T\right)W_T^*\bar{\psi}_T\left(\hat{\theta}_T\right)$$

$$\equiv \bar{\psi}_T^{*\prime}(\theta)W_T^*\bar{\psi}_T^*(\theta) + r_{1T}^*(\theta) + r_{2T}^*,$$

which implies that

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

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

$$\bar{\psi}_T^{*\prime}(\theta)W_T^*\bar{\psi}_T^*(\theta) - \bar{\psi}_T'(\theta)W_T\bar{\psi}_T(\theta)$$

$$= \left(\bar{\psi}_T^*(\theta) - \bar{\psi}_T(\theta)\right)'W_T^*\bar{\psi}_T^*(\theta) + \bar{\psi}_T(\theta)'W_T^*\left(\bar{\psi}_T^*(\theta) - \bar{\psi}_T(\theta)\right) + \bar{\psi}_T(\theta)'(W_T^* - W_T)\bar{\psi}_T(\theta)$$

$$\leq \left\|\bar{\psi}_T^*(\theta) - \bar{\psi}_T(\theta)\right\|\|W_T^*\|\left(\left\|\bar{\psi}_T^*(\theta)\right\| + \left\|\bar{\psi}_T(\theta)\right\|\right) + \left\|\bar{\psi}_T(\theta)\right\|^2\|W_T^* - W_T\|.$$

38

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 $\gamma_2 > 0$,

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

A second-order mean value expansion of $\bar{\psi}_{c,T}^*(\hat{\theta}_T^*)$ around θ_0 gives:

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

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), $\sup_{\theta \in \mathcal{N}} \|\bar{G}^*(\theta) - \bar{G}(\theta)\| = o_{\mathbb{P}}(1)$. This together with $\sup_{\theta \in \mathcal{N}} \|\bar{G}(\theta) - G(\theta)\| = o_P(1)$ and the fact that $\hat{\theta}_T^* - \theta_0 = o_{\mathbb{P}}(1)$ imply that $\bar{G}^*(\bar{\theta}_T^*) = G + o_{\mathbb{P}}(1)$. 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 the 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,

$$\sqrt{T}\bar{\psi}_{c,T}^*(\hat{\theta}_T^*) = \sqrt{T}\bar{\psi}_{c,T}^*(\theta_0) + \frac{1}{2}G\,vec\left(\hat{u}_T^*\hat{u}_T^{*'}\right) + o_{\mathbb{P}}(1) + o_{\mathbb{P}}\left(\|\hat{u}_T^*\|^2\right), \tag{B.7}$$

where $\bar{\psi}_{c,T}^*(\theta_0) = \left(\bar{\psi}_T^*(\theta_0) - \bar{\psi}_T(\theta_0)\right) + \bar{\psi}_T(\theta_0) - \bar{\psi}_T(\hat{\theta}_T) = O_{\mathbb{P}}(T^{-1/2}) + O_P(T^{-1/2}) + O_P(T^{-1/2}) = O_{\mathbb{P}}(T^{-1/2})$, and where the order of magnitude of $\bar{\psi}_T(\hat{\theta}_T)$ is obtained through a second-order Taylor expansion of $\bar{\psi}_T\left(\hat{\theta}_T\right)$ around θ_0:

$$\sqrt{T}\bar{\psi}_T(\hat{\theta}_T) = \sqrt{T}\bar{\psi}_T(\theta_0) + \frac{1}{2}G\,vec(T^{1/4}(\hat{\theta}_T - \theta_0)T^{1/4}(\hat{\theta}_T - \theta_0)') + o_P(1). \tag{B.8}$$

Thus,

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

By definition,

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

which implies that

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

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

$$\gamma_1\|\hat{u}_T^*\|^4 \le \gamma_2\left\|\sqrt{T}\bar{\psi}_{c,T}^*(\theta_0)\right\|\|\hat{u}_T^*\|^2 + o_{\mathbb{P}}(1) + o_{\mathbb{P}}(\|\hat{u}_T^*\|^2) + o_{\mathbb{P}}(\|\hat{u}_T^*\|^4),$$

where $\gamma_2 = \|W\|\|G\|$, which completes the proof of part (i). Next, we prove part (ii). We can show that the random vector

$$\left(\sqrt{T}\bar{\psi}_T'(\theta_0), \sqrt{T}\left(\bar{\psi}_T^*(\theta_0) - \bar{\psi}_T(\theta_0)\right)', T^{1/4}\left(\hat{\theta}_T - \theta_0\right)', T^{1/4}\left(\hat{\theta}_T^* - \hat{\theta}_T\right)'\right)' = O_{\mathbb{P}}(1).$$

39

In particular, $\sqrt{T}\bar{\psi}_T(\theta_0) \to^d X \sim N(0, \Sigma)$, implying that the first term is $O_P(1)$ (hence $O_{\mathbb{P}}(1)$), whereas by an application of the bootstrap CLT, $\sqrt{T}\left(\bar{\psi}_T^*(\theta_0) - \bar{\psi}_T(\theta_0)\right) \to^{d^*} X^* \sim N(0, \Sigma)$, implying that the second term is $O_{P^*}(1)$ in prob-P (hence, it is $O_{\mathbb{P}}(1)$ by Lemma B.2.a3)). Finally, Proposition 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 Wellner (1990)), it follows that this random vector has at least one subsequence[6] 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 Woodroofe (2010)). Next, we will show that $P(|U^*| > 0) > 0$. By the second-order condition for the minimization problem underlying $\hat{\theta}_T^*$,

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

which can be written as

$$\tilde{Z}_T^* + N_T^* \geq 0, \tag{B.9}$$

where

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

From part (i) of this theorem, and equations (B.7) and (B.8), we can show that

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

where $Z_T^* = G'W\sqrt{T}\left(\bar{\psi}_T^*(\theta_0) - \bar{\psi}_T(\theta_0)\right)$ and $N_T^* = G'WG\hat{u}_T^{*2} + o_{\mathbb{P}}(1)$. From (B.9), we have that

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

with $A = G'WG/2 > 0$ since $G \neq 0$ given the second-order local identification assumption. By the continuous mapping theorem, $-Z_T^* + A\hat{v}_T^2 - 3A\hat{u}_T^{*2} \to^d -Z^* + AV^2 - 3AU^{*2}$ under $\mathbb{P}$. Using the same arguments as in Lemma B.2 of D&R (2013a), 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 $Z^* = G'WX^*$ 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 (2013a). 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 θ_0 is interior point in Θ, $\bigcup_{T \geq 0} \mathbb{H}_T = \mathbb{R}$. 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 Lemma B.6(ii) of D&R (2013a). To show (i), it suffices to show that $Y_T^*(u) \equiv \bar{\psi}_{c,T}^*\left(\theta_0 + T^{-1/4}u\right)$ converges in distribution under $\mathbb{P}$ towards $Y^* \equiv X^* - \frac{1}{2}GV^2 + \frac{1}{2}Gu^2$, 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 $(Y_T^*(u_1), \ldots, Y_T^*(u_k))$ converge in distribution to $(Y^*(u_1), \ldots, Y^*(u_k))$ with respect to $\mathbb{P}$ for any $u_1, \ldots, u_k \in \mathbb{K}$, and (a2) The stochastic process $Y_T^*(u)$ is asymptotically equicontinuous, i.e. $\forall \epsilon > 0$,

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

(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 θ_0. This implies that

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

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

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

---

[6] To simplify the notation, we keep the same index T to denote this subsequence throughout this proof.

40

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

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

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

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

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

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

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

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

It follows that

$$E(\underline{J}^*) \quad = \quad E(\underline{J}_1^* | 2X^{*\prime}WG - G'WGV^2 \geq 0)q^* + E(\underline{J}_2^* | 2X^{*\prime}WG - G'WGV^2 < 0)(1 - q^*), \qquad (B.10)$$

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

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

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

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

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

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

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

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

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 $V^2 = -2\frac{G'WX}{G'WG}\mathbf{1}(G'WX < 0)$, which implies that

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

Hence,

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

41

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 $E(\underline{J}_1^* | 2X^{*\prime}WG - G'WGV^2 \geq 0)$. We have that

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

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

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

where the first equality follows because $V = 0$, the second uses the fact that $X^{*\prime}WX^*$ 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 $X^{*\prime}WX^* \sim \chi^2(H)$. Moreover,

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

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

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

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

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

where

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

Since X^* and X are independent, $(b1) = E(X^{*\prime}WX^* | G'WX + G'WX^* \geq 0)$. Let Q be the rotation matrix that transforms the canonical basis of $\mathbb{R}^H$ into the orthonormal basis $(a_1, a_2, \ldots, a_H)$ 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 $X^{*\prime}WX^* = Y^{*\prime}Y^* = Y_1^{*2} + \cdots + Y_{H-1}^{*2} + Y_H^{*2}$, with Y_i^*'s independent and identically distributed $N(0,1)$. Thus,

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

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,

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

42

by again relying on the properties of standard normal random variables. A similar argument shows that $(b_3) = 1$, implying that

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

and therefore $(b) = \frac{1}{8q^*}\left(H - \frac{4}{\pi} + 1\right)$. 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 (B.10) 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

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

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 Θ. Since $\frac{\partial}{\partial\theta'}\bar{\psi}_T\left(\hat{\theta}_T\right) \xrightarrow{P} 0$, $S_{2T}^*(\theta)$ is $o_{P^*}(1)$ in prob-P uniformly over $\theta \in \Theta$ given in particular the fact that $\left\|\theta - \hat{\theta}_T\right\|$ is bounded with probability P converging to 1 (given the compactness of Θ and the fact that $\hat{\theta}_T \xrightarrow{P} \theta_0 \in \Theta$). We also have:

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

That is $\sup_\theta |S_{1T}^*(\theta)| = o_{P^*}(1)$ in prob-P.

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

$$\sqrt{T}\bar{\psi}_T^{*(1)}\left(\hat{\theta}_T^{*(1)}\right) = \sqrt{T}\bar{\psi}_T^{*(1)}\left(\hat{\theta}_T\right) + \sqrt{T}\frac{\partial\bar{\psi}_T^{*(1)}}{\partial\theta'}\left(\hat{\theta}_T\right)\left(\hat{\theta}_T^{*(1)} - \hat{\theta}_T\right) + \frac{1}{2}R_T^{*(1)}\left(\bar{\theta}_T^*\right), \tag{B.11}$$

where

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

with $\bar{\theta}_T^* \in \left(\hat{\theta}_T^{*(1)}, \hat{\theta}_T\right)$ 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 $\bar{G}^{*(1)}\left(\bar{\theta}_T^*\right) = G + o_{P^*}(1)$ 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,

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

43

in prob-P. By the bootstrap CLT and under our assumptions, we have that

$$\sqrt{T}\bar{\psi}_T^{*(1)}\left(\hat{\theta}_T\right) = \sqrt{T}\left(\bar{\psi}_T^*\left(\hat{\theta}_T\right) - \bar{\psi}_T\left(\hat{\theta}_T\right)\right) = O_{P^*}(1), \text{ in prob-}P, \quad \text{and}$$

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

As a result,

$$\sqrt{T}\bar{\psi}_T^{*(1)}\left(\hat{\theta}_T^{*(1)}\right) = \sqrt{T}\bar{\psi}_T^{*(1)}\left(\hat{\theta}_T\right) + \frac{1}{2}G\,vec\left(\hat{v}_T^{*(1)}\hat{v}_T^{*(1)\prime}\right) + o_{P^*}(1) + o_{P^*}\left(\|\hat{v}_T^{*(1)}\|^2\right), \tag{B.12}$$

in prob-P. Thus,

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

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

$$T\bar{\psi}_T^{*(1)\prime}\left(\hat{\theta}_T^{*(1)}\right)W_T^{*(1)}\bar{\psi}_T^{*(1)}\left(\hat{\theta}_T^{*(1)}\right) \le T\bar{\psi}_T^{*(1)\prime}\left(\hat{\theta}_T\right)W_T^{*(1)}\bar{\psi}_T^{*(1)}\left(\hat{\theta}_T\right) = T\bar{\psi}_T^{*(1)\prime}\left(\hat{\theta}_T\right)W\bar{\psi}_T^{*(1)}\left(\hat{\theta}_T\right) + o_{P^*}(1),$$

in prob-P. Thus,

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

in prob-P, implying that

$$\begin{aligned}
\frac{1}{4}vec'\left(\hat{v}_T^{*(1)}\hat{v}_T^{*(1)\prime}\right)G'WG\,vec\left(\hat{v}_T^{*(1)}\hat{v}_T^{*(1)\prime}\right) &\le vec'(\hat{v}_T^{*(1)}\hat{v}_T^{*(1)\prime})G'W\sqrt{T}\bar{\psi}_T^{*(1)}\left(\hat{\theta}_T\right) \\
&\quad + o_{P^*}(1) + o_{P^*}(\|\hat{v}_T^{*(1)}\|^2) + o_{P^*}(\|\hat{v}_T^{*(1)}\|^4).
\end{aligned}$$

Given Lemma B.1 of D&R (2013a) and Assumption 6(ii), we have that

$$\frac{1}{4}vec'\left(\hat{v}_T^{*(1)}\hat{v}_T^{*(1)\prime}\right)G'WG\,vec\left(\hat{v}_T^{*(1)}\hat{v}_T^{*(1)\prime}\right) \ge \gamma_1\|\hat{v}_T^{*(1)}\|^4$$

for some $\gamma_1 > 0$. Hence,

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

implying that

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

This shows that $\|\hat{v}_T^{*(1)}\|^2$ is at most of the same order as $\left\|\sqrt{T}\bar{\psi}_T^{*(1)}\left(\hat{\theta}_T\right)\right\|$, 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 $e \in \mathbb{R}^p$,

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

This can be written as

$$e'\left(\tilde{Z}_T^* + N_T^*\right)e \ge 0, \tag{B.13}$$

44

where

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

From part (i), we can show that

$$\sqrt{T} \bar{\psi}_T^{*(1)}(\hat{\theta}_T^{*(1)}) = \sqrt{T} \bar{\psi}_T^{*(1)}(\hat{\theta}_T) + \frac{1}{2} G\, vec(\hat{v}_T^{*(1)} \hat{v}_T^{*(1)\prime}) + o_{P^*}(1), \quad \text{and}$$

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

in prob-P, for $i = 1, \ldots, p$. 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 convergence of $\bar{G}^*(\theta)$ and $\bar{G}(\theta)$ in a neighbourhood of θ_0 then allows to replace the sample means in the second-order 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,

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

$$N_T^* = \left(\hat{v}_T^{*(1)\prime} \frac{\partial^2 \rho'}{\partial \theta_i \partial \theta}(\theta_0) W \frac{\partial^2 \rho}{\partial \theta_i \partial \theta'}(\theta_0) \hat{v}_T^{*(1)} \right)_{1 \leq i,j \leq p} + o_{P^*}(1),$$

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 (B.13) 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 $\left(\sqrt{T} \bar{\psi}_T^{*(1)\prime}(\hat{\theta}_T), \hat{v}_T^{*(1)\prime} \right)' = O_{P^*}(1)$, in prob-$P$, it follows that $\left(Z_T^*, \hat{v}_T^{*(1)} \right) = O_{\mathbb{P}}(1)$ (by Lemma B.2). 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, $d\left(\mathcal{L}\left(Y_{T'}^* \right), \mathcal{L}\left(Y^* \right) \right) \xrightarrow{\mathbb{P}} 0$, where $\mathcal{L}(\cdot)$ denotes the law of the random variable in question. A second application of Lemma B.2 implies that $d\left(\mathcal{L}(Y_{T'}^*), \mathcal{L}(Y^*) \right) \xrightarrow{P^*} 0$ 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 $d\left(\mathcal{L}(Y_{T''}^*), \mathcal{L}(Y^*) \right) \xrightarrow{P^*} 0$ a.s.-P. Fix ω 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 $P\left(\|V^*\| > 0 \right) \geq 1 - P(Z^* \geq 0) \geq 1/2 \equiv \delta$ for all ω 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 $v = T^{1/4}\left(\theta - \hat{\theta}_T \right)$, 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 is suffices to show that

$$\left(W_T^{*(1)}, Y_T^*(v) \right) \Rightarrow^{P^*} (W, Y^*(v)), \text{ in } \ell^\infty(\mathbb{K}), \text{ in prob-}P, \tag{B.14}$$

where $Y^*(v) \equiv X^* + \frac{1}{2} G\, vec(vv')$ with $X^* \sim N(0, \Sigma)$, implying that $Y^*(v) =^d Y(v) \equiv X + \frac{1}{2} G\, vec(vv')$. 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 (Kosorok (2008, Theorem 7.5)) ensures that (B.14) 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

45

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

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

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

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

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

where $\ddot{\theta}^*\left(v\right) \in \left(\hat{\theta}_T, \hat{\theta}_T + T^{-1/4}v\right)$ and may differ from row to row; the matrix $\bar{G}^{*(1)}\left(\theta\right)$ is defined as in (6) with $\bar{\psi}_T^{*(1)}\left(\theta\right)$ in place of $\bar{\psi}_T\left(\theta\right)$. We can write

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

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

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

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

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

where we can show that $\bar{G}^{*(1)}\left(\hat{\theta}_T\right) - \bar{G}\left(\hat{\theta}_T\right) = o_{P^*}\left(1\right)$ in prob-P and $\bar{G}\left(\hat{\theta}_T\right) - G = o_P\left(1\right)$ (for the first result, we rely on the fact that $\sup_{\theta\in\Theta}\|\bar{G}^{*(1)}\left(\theta\right) - \bar{G}\left(\theta\right)\| = o_{P^*}\left(1\right)$ in prob-P whereas the second result follows from the uniform convergence of $\bar{G}\left(\theta\right)$ towards $G\left(\theta\right)$, the fact that $\hat{\theta}_T \xrightarrow{P} \theta_0$, and the definition $G \equiv G\left(\theta_0\right)$). 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 $\bar{G}^{*(1)}\left(\theta\right)$ towards $\bar{G}\left(\theta\right)$ and the fact that $\hat{\theta}_T^{*(1)} - \ddot{\theta}_T^* \xrightarrow{P^*} 0$, in prob-$P$ (see e.g. Lemma A.6 of Gonçalves and White (2004)). Finally, $r_{3T}^*\left(v\right) = o_{P^*}\left(1\right)$ in prob-P uniformly over $\mathbb{K}$ because $T^{1/2}\frac{\partial\bar{\psi}_T^{*(1)}}{\partial\theta'}\left(\hat{\theta}_T\right) = O_{P^*}\left(1\right)$ by the bootstrap CLT applied to $T^{1/2}\frac{\partial\bar{\psi}_T^{*(1)}}{\partial\theta'}\left(\hat{\theta}_T\right)$ (note that its mean is zero given the double recentering). Since $\sqrt{T}\bar{\psi}_T^{*(1)}\left(\hat{\theta}_T\right) \xrightarrow{d^*} X^* \sim N(0,\Sigma)$, in prob-$P$, we conclude that $Y_T^*\left(v\right) \to^{d^*} X^* + \frac{1}{2}G\,vec\left(vv'\right) \equiv Y^*\left(v\right)$, in prob-$P$ for any fixed $v \in \mathbb{K}$. Similarly, for any fixed $\left(v_1, v_2, \ldots, v_k\right) \in \mathbb{K}^k$, consider $c = \left(c_1,\ldots,c_k\right) \in \mathbb{R}^k$ such that $c'c = 1$. It follows that

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

where the last term is $o_{P^*}\left(1\right)$ in prob-P uniformly over $\mathbb{K}$ by the same arguments as those used above. It follows that $\sum_{j=1}^k c_j Y_T^*\left(v_j\right) \to^{d^*} c'X^* + \frac{1}{2}G\sum_{j=1}^k c_j vec\left(v_j v_j'\right)$ in prob-P and the result follows by the Cramer-Wold device. Next, we establish (b). From the second-order expansion of $Y_T^*\left(v\right)$ above, we can write

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

46

Hence,

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

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

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

implying that

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

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

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

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

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

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

(b3) $E^* \| \bar{G}^{*(1)}\left(\ddot{\theta}^*(v_2)\right) - \bar{G}^{*(1)}\left(\ddot{\theta}^*(v_1)\right) \| = o_P(1)$, 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,

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

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

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

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

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

47

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

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

where the last inequality uses the fact that $\ddot{\theta}^* (v_1) \in \left(\hat{\theta}_T, \hat{\theta}_T + T^{-1/4} v \right)$, which is included in $\mathcal{N}$, a neighborhood of θ_0, for all T sufficiently large with probability P approaching one (given that $\hat{\theta}_T \xrightarrow{P} \theta_0$). 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 Lipshitz 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

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

where the last inequality uses the fact that $\sup_{v_1, v_2 \in \mathbb{K}} \left\| \ddot{\theta}^* (v_1) - \ddot{\theta}^* (v_2) \right\| \leq C T^{-1/4}$ 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 $\{X_1, \ldots, X_n\}$, $E^* \left(\frac{1}{T} \sum_{t=1}^{T} m(X_t^*) \right) = \frac{1}{T} \sum_{t=1}^{T} m(X_t) = O_P(1)$ given Assumption 4(i) and we conclude that

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

uniformly in v_1 and v_2.

Proof of Theorem 4.1. Part (i) follows by Lemmas 4.1 and B.3. 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 (2013a). 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 $\hat{\theta}_T \xrightarrow{P} \theta_0$, an interior point of Θ, 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 2.3.

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 8), θ_0 is the unique $\arg \min_\theta Q^{(2)}(\theta)$, where

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

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

$$\sup_{\theta \in \Theta} \left| Q_T^{*(2)}(\theta) - Q^{(2)}(\theta) \right| \xrightarrow{P^*} 0. \tag{B.15}$$

Letting $Q_T^{(2)}(\theta) = \bar{\Phi}_T'(\theta) W_T \bar{\Phi}_T(\theta)$, where $\bar{\Phi}_T(\theta) = \bar{\psi}_T(\theta) - \frac{\partial}{\partial \theta'} \bar{\psi}_T(\theta) (\theta - \theta_0)$, (B.15) follows from: (A) $\sup_{\theta \in \Theta} \left| Q_T^{*(2)}(\theta) - Q_T^{(2)}(\theta) \right| \xrightarrow{P^*} 0$, in prob-$P$; and (B) $\sup_{\theta \in \Theta} \left| Q_T^{(2)}(\theta) - Q^{(2)}(\theta) \right| \xrightarrow{P} 0$. We can show that

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

48

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

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

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

$$
\sup_{\theta \in \Theta} |Q_T^* (\theta) - Q_T (\theta)| = o_{P^*} (1), \text{ in prob-}P; \tag{B.16}
$$

$$
\sup_{\theta \in \Theta} |R_{1T}^* (\theta) - R_{1T} (\theta)| = o_{P^*} (1), \text{ in prob-}P; \text{ and} \tag{B.17}
$$

$$
\sup_{\theta \in \Theta} |R_{2T}^* (\theta) - R_{2T} (\theta)| = o_{P^*} (1), \text{ in prob-}P. \tag{B.18}
$$

Proposition 3.1 implies (B.16) whereas we can show that (B.17) and (B.18) follow by relying on the fact that $\hat{\theta}_T \overset{P}{\to} \theta_0$; 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 Θ (see proof of Proposition 3.1). Moreover,

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

We can show that $[2] = o_P (1)$ uniformly on Θ 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 Θ 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 Θ.

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

Lemma B.4 *Under Assumptions 1-8, if $W_T^{*(2)} \overset{P^*}{\to} W$ in prob-P, (i) $\left\| \hat{\theta}_T^{*(2)} - \hat{\theta}_T \right\| = O_{P^*} \left(T^{-1/4} \right)$ 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 $\delta > 0$, $P \left(\|V^*\| \neq 0 \right) \geq \delta$.*

49

Lemma B.5 *Under Assumptions 1-9, if $W_T^{*(2)} \xrightarrow{P^*} W$ 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 B.4. We follow the proof of Proposition 4.2. First, note that for all $i, j = 1, \ldots, p$,

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

and

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

Hence,

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

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

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

where $R_T^{*(2)} \left(\bar{\theta}_T^* \right) = \bar{G}^{*(2)} \left(\bar{\theta}_T^* \right) vec \left(\hat{v}_T^{*(2)} \hat{v}_T^{*(2)\prime} \right)$ and $\hat{v}_T^{*(2)} \equiv T^{1/4} \left(\hat{\theta}_T^{*(2)} - \hat{\theta}_T \right)$ 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 $\hat{\theta}_T \xrightarrow{P} \theta_0$, 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, $\kappa_{i,j} \left(\bar{\theta}_T^* \right) = -\frac{\partial^2}{\partial \theta_i \partial \theta_j} \rho(\theta) + o_{P^*}(1)$, in prob-P, which implies that

$$\sqrt{T} \bar{\psi}_T^{*(2)} \left(\hat{\theta}_T^{*(2)} \right) = \sqrt{T} \bar{\psi}_T^{*(1)} \left(\hat{\theta}_T \right) - \frac{1}{2} G \, vec \left(\hat{v}_T^{*(2)} \hat{v}_T^{*(2)\prime} \right) + o_{P^*}(1) + o_{P^*} \left(\left\| \hat{v}_T^{*(2)} \right\| \right). \tag{B.19}$$

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

$$\sqrt{T} \bar{\psi}_T^{*(2)}(\hat{\theta}_T^{*(2)}) = \sqrt{T} \bar{\psi}_T^{*(1)}(\hat{\theta}_T) - \frac{1}{2} G \, vec(\hat{v}_T^{*(2)} \hat{v}_T^{*(2)\prime}) + o_{P^*}(1), \quad \text{and}$$

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

in prob-P, for $i = 1, \ldots, p$. 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

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

$$N_T^* = \left(\hat{v}_T^{*(2)\prime} \frac{\partial^2 \rho'}{\partial \theta_i \partial \theta}(\theta_0) W \frac{\partial^2 \rho}{\partial \theta_i \partial \theta'}(\theta_0) \hat{v}_T^{*(2)} \right)_{1 \leq i, j \leq p} + o_{P^*}(1),$$

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 $e \in \mathbb{R}^p$,

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

50

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

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

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 $P\left(\|V^*\| > 0 \right) > \delta$.

Proof of Lemma B.5. The proof follows exactly that of Lemma 4.1, given Lemma B.4. The main difference is that the stochastic process $Y_T^*(v) \equiv \sqrt{T} \bar{\psi}_T^{*(2)} \left(\hat{\theta}_T + T^{-1/4} v \right)$ now converges weakly in prob-P towards $\ddot{Y}(v) = X - \frac{1}{2} G \, vec(vv')$. Since $X \sim N(0, \Sigma)$, then $X =^d -X$, where $=^d$ denotes equality in distribution, which implies that $\ddot{Y}(v) =^d -Y(v) = -\left(X + \frac{1}{2} G \, vec(vv') \right)$, 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 $W_T^{*(2)} \overset{P^*}{\rightarrow} W$, in prob-$P$, concluding the proof.

Proof of Theorem 4.2. Part (i) follows as the proof of Theorem 4.1, given Lemmas B.5 and B.3 whereas part (ii) follows from Polya's theorem.

Proof of Theorem 4.3. We provide a sketch of proof for this result. The bootstrap estimating function is:

$$\psi_t^{*(1)}(\eta) = \psi(X_t^*, \hat{R}\eta) - E^*\left(\psi_t^*(\hat{\theta}_T) \right) - E^*\left(\frac{\partial}{\partial \theta'} \psi_t^*(\hat{\theta}_T) \right) \hat{R}_2(\eta_2 - \hat{\eta}_{2T}),$$

with $\hat{R}\hat{\eta}_T = \hat{\theta}_T$ and $\hat{R} = (\hat{R}_1 | \hat{R}_2) \overset{P}{\rightarrow} R$. We use the partition $\eta = (\eta_1', \eta_2')' \in \mathbb{R}^r \times \mathbb{R}^{p-r}$.
1. Consistency of bootstrap GMM estimator: The proof follows by the same arguments as the proof of Proposition 3.1. Because $E^*\left(\frac{\partial}{\partial \theta'} \psi_t^*(\hat{\theta}_T) \right) \hat{R}_2 \overset{P}{\rightarrow} \frac{\partial \rho}{\partial \theta'}(\theta_0) R_2 = 0$ and $E^*\left(\psi_t^*(\hat{\theta}_T) \right) = \frac{1}{T} \sum_{t=1}^T \psi(X_t, \hat{\theta}_T) \overset{P}{\rightarrow} 0$, and using the bootstrap uniform law of large numbers, we have

$$\frac{1}{T} \sum_{t=1}^T \psi_t^{*(1)}(\eta) \overset{P^*}{\rightarrow} E(\psi(X_t, R\eta),$$

in prob-P uniformly over the compact parameter set $\{ R^{-1}\theta, \theta \in \Theta \}$. Then, the global identification condition ensures that the bootstrap GMM estimator converges in probability to the GMM estimator: $\hat{\eta}_T^* \overset{P^*}{\rightarrow} \hat{\eta}_T$ in prob-P.
2. Local identification property of bootstrap estimating function: We have:

$$E^*\left(\frac{\partial \psi_t^{*(1)}(\hat{\eta}_T)}{\partial \eta_2'} \right) = E^*\left(\frac{\partial \psi}{\partial \theta'}(X_t^*, \hat{\theta}_T) \right) \hat{R}_2 - E^*\left(\frac{\partial}{\partial \theta'} \psi_t^*(\hat{\theta}_T) \right) \hat{R}_2 = 0,$$

$$E^*\left(\frac{\partial \psi_t^{*(1)}(\hat{\eta}_T)}{\partial \eta_1'} \right) = E^*\left(\frac{\partial \psi}{\partial \theta'}(X_t^*, \hat{\theta}_T) \right) \hat{R}_1 \overset{P}{\rightarrow} \frac{\partial \rho}{\partial \theta'}(\theta_0) R_1,$$

which is full rank. The bootstrap central limit theorem guarantees that $\frac{1}{T} \sum_{t=1}^T \frac{\partial \psi}{\partial \theta'}(X_t^*, \hat{\theta}_T) - E^*\left(\frac{\partial}{\partial \theta'} \psi_t^*(\hat{\theta}_T) \right) = O_{P^*}(T^{-1/2})$ in prob-P and therefore,

$$\frac{1}{T} \sum_{t=1}^T \frac{\partial \psi_t^{*(1)}(\hat{\eta}_T)}{\partial \eta_2'} = O_{P^*}(T^{-1/2})$$

in prob-P.
3. Rate of convergence: Similar expansions to those in the proofs of Proposition 3.2 of Dovonon and Renault

51

(2009) and Theorem 1(a) of Dovonon and Hall (2015) also apply readily to this bootstrap setting with 'local identification' patterns as in 2. above and an analogue of Proposition 4.2 can be derived yielding:

$$\hat{\eta}_{1T}^* - \hat{\eta}_{1T} = O_{P^*}(T^{-1/2}), \quad \text{and} \quad \hat{\eta}_{2T}^* - \hat{\eta}_{2T} = O_{P^*}(T^{-1/4}),$$

in prob-P and these rates are sharp.

4. The asymptotic distribution of the bootstrap overidentification test statistic: The bootstrap test statistic is given by:

$$\hat{J}_{T,1}^{*(1)} = T\bar{\psi}^{*(1)}(\hat{\eta}_T^*)'W_T^{*(1)}\bar{\psi}^{*(1)}(\hat{\eta}_T^*).$$

Consider the local parameterization: $v_1 = \sqrt{T}(\eta_1 - \hat{\eta}_{1T})$ and $v_2 = T^{1/4}(\eta_2 - \hat{\eta}_{2T})$; $v \equiv (v_1', v_2')'$. Let:

$$J_{T,1}^{*(1)}(v) = T\bar{\psi}^{*(1)}\left(\hat{\eta}_{1T} + v_1/\sqrt{T}, \hat{\eta}_{2T} + v_2/T^{1/4}\right)' W_T^{*(1)} \bar{\psi}^{*(1)}\left(\hat{\eta}_{1T} + v_1/\sqrt{T}, \hat{\eta}_{2T} + v_2/T^{1/4}\right).$$

Letting $\hat{v}_{1T}^* = \sqrt{T}(\hat{\eta}_{1T}^* - \hat{\eta}_{1T})$ and $\hat{v}_{2T}^* = T^{1/4}(\hat{\eta}_{2T}^* - \hat{\eta}_{2T})$, we can see that:

$$\hat{J}_{T,1}^{*(1)} = J_{T,1}^{*(1)}(\hat{v}_T^*) = \min_{v \in \mathbb{R}^p} J_{T,1}^{*(1)}(v).$$

The derivation of the asymptotic distribution follows the same strategy as in Lemma 4.1 and Theorem 4.1(i): We show that $J_{T,1}^{*(1)}(v)$ converges in distribution to $J_0(v)$ uniformly over any compact set and then conclude that $J_{T,1}^{*(1)}(\hat{v}_T)$ converges in distribution to $\min_{v \in \mathbb{R}^p} J_0(v)$ using Lemma B.2, where

$$J_0(v) = \left(X + Dv_1 + \frac{1}{2}G_r vec(v_2 v_2')\right)' W \left(X + Dv_1 + \frac{1}{2}G_r vec(v_2 v_2')\right)$$

for all $v = (v_1', v_2')' \in \mathbb{R}^r \times \mathbb{R}^{p-r}$. The proof of uniform convergence follows similar lines as in the proof of Lemma 4.1: a second-order mean value expansion of $v \mapsto \sqrt{T}\bar{\psi}^{*(1)}\left(\hat{\eta}_{1T} + v_1/\sqrt{T}, \hat{\eta}_{2T} + v_2/T^{1/4}\right)$ around $v = 0$ yields:

$$\sqrt{T}\bar{\psi}^{*(1)}\left(\hat{\eta}_{1T} + v_1/\sqrt{T}, \hat{\eta}_{2T} + v_2/T^{1/4}\right)$$

$$= \sqrt{T}\bar{\psi}^{*(1)}(\hat{\eta}_T) + \sqrt{T}\frac{\partial\bar{\psi}^{*(1)}}{\partial\eta_1'}(\hat{\eta}_T)v_1/\sqrt{T} + \sqrt{T}\frac{\partial\bar{\psi}^{*(1)}}{\partial\eta_2'}(\hat{\eta}_T)v_2/T^{1/4} + \frac{1}{2\sqrt{T}}\left(v_1'\frac{\partial^2\bar{\psi}_h^{*(1)}}{\partial\eta_1\partial\eta_1'}(\dot{v})v_1\right)_{1 \le h \le H}$$

$$+ \frac{1}{T^{1/4}}\left(v_1'\frac{\partial^2\bar{\psi}_h^{*(1)}}{\partial\eta_1\partial\eta_2'}(\dot{v})v_2\right)_{1 \le h \le H} + \frac{1}{2}\left(v_2'\frac{\partial^2\bar{\psi}_h^{*(1)}}{\partial\eta_2\partial\eta_2'}(\dot{v})v_2\right)_{1 \le h \le H},$$

with $\dot{v} \in (0, v)$ and may differ from row to row. Under our assumptions, we have:

$$\sqrt{T}\bar{\psi}^{*(1)}\left(\hat{\eta}_{1T} + v_1/\sqrt{T}, \hat{\eta}_{2T} + v_2/T^{1/4}\right) = \sqrt{T}\bar{\psi}^{*(1)}(\hat{\eta}_T) + \frac{\partial\bar{\psi}^{*(1)}}{\partial\eta_1'}(\hat{\eta}_T)v_1 + \frac{1}{2}\left(v_2'\frac{\partial^2\bar{\psi}_h^{*(1)}}{\partial\eta_2\partial\eta_2'}(\dot{v})v_2\right)_{1 \le h \le H} + o_{P^*}(1),$$

with the $o_{P^*}(1)$ terms uniformly negligible over any compact set. Note that:

$$\frac{\partial\bar{\psi}^{*(1)}}{\partial\eta_1'}(\hat{\eta}_T) = \frac{1}{T}\sum_{t=1}^T \frac{\partial\psi}{\partial\theta'}\left(X_t^*, \hat{R}\hat{\eta}_T\right)\hat{R}_1, \quad \text{and} \quad \frac{\partial^2\bar{\psi}_h^{*(1)}}{\partial\eta_2\partial\eta_2'}(\dot{v}) = \hat{R}_2'\frac{1}{T}\sum_{t=1}^T \frac{\partial^2\psi_h}{\partial\theta\partial\theta'}\left(X_t^*, \hat{R}\dot{\eta}\right)\hat{R}_2,$$

with $\dot{\eta} \in (\hat{\eta}_T, \hat{\eta}_T^*)$. By the bootstrap uniform law of large numbers we have:

$$\frac{1}{T}\sum_{t=1}^T \frac{\partial\psi}{\partial\theta'}\left(X_t^*, \hat{R}\hat{\eta}_T\right)\hat{R}_1 \xrightarrow{P^*} \frac{\partial\rho}{\partial\theta'}(\theta_0)R_1 = D$$

in prob-P and

$$\left(v_2'\hat{R}_2'\frac{1}{T}\sum_{t=1}^T \frac{\partial^2\psi_h}{\partial\theta\partial\theta'}\left(X_t^*, \hat{R}\dot{\eta}\right)\hat{R}_2 v_2\right)_{1 \le h \le H} \xrightarrow{P^*} G_r vec(v_2 v_2')$$

52

in prob-P and in both cases, the convergence holds uniformly over any compact set of v's. We can then claim that:

$$\sqrt{T}\bar{\psi}^{*^{(1)}}\left(\hat{\eta}_{1T} + v_1/\sqrt{T}, \hat{\eta}_{2T} + v_2/T^{1/4}\right) = \sqrt{T}\bar{\psi}^{*^{(1)}}(\hat{\eta}_T) + Dv_1 + \frac{1}{2}G_r vec(v_2 v_2') + o_{P^*}(1)$$

with uniformly negligible $o_{P^*}(1)$ terms over compact sets. Note that, by the bootstrap central limit theorem, $\sqrt{T}\bar{\psi}^{*^{(1)}}(\hat{\eta}_T)$ converges in distribution to X. Thus the marginals of the stochastic process:
$\sqrt{T}\bar{\psi}^{*^{(1)}}\left(\hat{\eta}_{1T} + v_1/\sqrt{T}, \hat{\eta}_{2T} + v_2/T^{1/4}\right)$ converge in distribution to $X + Dv_1 + \frac{1}{2}G_r vec(v_2 v_2')$. The stochastic equicontinuity of this random process can be established along the same lines as in the proof of Lemma 4.1. We then conclude that

$$\sqrt{T}\bar{\psi}^{*^{(1)}}\left(\hat{\eta}_{1T} + v_1/\sqrt{T}, \hat{\eta}_{2T} + v_2/T^{1/4}\right) \xrightarrow{d^*} X + Dv_1 + \frac{1}{2}G_r vec(v_2 v_2'),$$

in prob-P uniformly over any compact set. The continuous mapping theorem allows to write that

$$J_{T,1}^{*^{(1)}}(v) \xrightarrow{d^*} J_0(v),$$

in prob-P, uniformly over any compact set. We can now apply Lemma B.2 to conclude that

$$\hat{J}_{T,1}^{*^{(1)}} \xrightarrow{d^*} \min_{v \in \mathbb{R}^p} J_0(v)$$

in prob-P.

5. To obtain (i), it suffices to show that $\min_{v \in \mathbb{R}^p} J_0(v) = \min_{v \in \mathbb{R}^{p-r}} J_1(v)$. This is straightforward once it is seen that $J_1(v_2) = \min_{v_1 \in \mathbb{R}^r} J_0(v_1, v_2)$. 6. (ii) follows from the fact that $\min_{v \in \mathbb{R}^{p-r}} J_1(v)$ is the asymptotic distribution of $\hat{J}_T$ under the conditions of the theorem.

References

Andrews, D. W. K. (2002). 'Higher-order improvements of a computationally attractive k-step bootstrap for extremum estimators', *Econometrica*, 70: 119–262.

Andrews, D. W. K., and Guggenberger, P. (2015). 'Identification-and singularity-robust inference for moment condition', Discussion paper, Cowles Foundation Discussion Paper No. 1978.

Bochnak, J., Coste, M., and Roy, M.-F. (1998). *Real algebraic geometry*. Springer Verlag, Berlin Heidelberg.

Brown, B., and Newey, K. W. (2002). 'Generalized method of moments, efficient bootstrapping, and improved inference', *Journal of Business and Economic Statistics*, 20: 507–510.

Bugni, F. A., Canay, I. A., and Shi, S. (2015). 'Specification tests for partially identified models defined by moment inequalities', *Journal of Econometrics*, 185: 259–282.

Cheng, G., and Huang, J. Z. (2010). 'Bootstrap consistency for general semiparameteric M-estimation', *Annals of Statistics*, 38: 2884–2915.

Cragg, J. G., and Donald, S. G. (1997). 'Inferring the rank of a matrix', *Journal of Econometrics*, 76: 223–250.

Dovonon, P., and Hall, A. R. (2015). 'The asymptotic properties of GMM and indirect inference under second-order identification', Discussion paper, Concordia University and University of Manchester.

Dovonon, P., and Renault, E. (2009). 'GMM overidentification test with first order underidentification', Discussion paper, Department of Economics, Concordia University, Montreal, Canada.

——— (2013a). 'Supplement to 'Testing for common conditionally heteroskedastic factors': Extensions and Proofs', *Econometrica Supplemental Material*, 81, http://www. econometricsociety.org/ecta/supmat/10082_proofs.pdf.

——— (2013b). 'Testing for common conditionally heteroskedastic factors', *Econometrica*, 81: 2561–2586.

Doz, C., and Renault, E. (2006). 'Factor volatility in mean models: a GMM approach', *Econometric Reviews*, 25: 275–309.

Engle, R. F., and Kozicki, S. (1993). 'Testing for common features', *Journal of Business and Economic Statistics*, 11: 369–395.

Fiorentini, G., Sentana, E., and Shephard, N. (2004). 'Likelihood-based estimation of latent generalized ARCH structures', *Econometrica*, 72: 1481–1517.

Giné, E., and Zinn, J. (1990). 'Bootstrapping general empirical measures', *Annals of Probability*, 18: 851–869.

Gonçalves, S., and White, H. (2004). 'Maximum likelihood and the bootstrap for nonlinear dynamic models', *Journal of Econometrics*, 119: 199–220.

Hahn, J. (1996). 'A note on bootstrapping generalized method of moments estimators', *Econometric Theory*, 12: 187–197.

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

Hansen, L. P. (1982). 'Large sample properties of generalized method of moments estimators', *Econometrica*, 50: 1029–1054.

He, C., and Teräsvirta, T. (1999). 'Properties of moments of a family of GARCH processes', *Journal of Econometrics*, 92: 173–192.

Inoue, A., and Shintani, M. (2006). 'Bootstrapping GMM estimators for time series', *Journal of Econometrics*, 133: 531–555.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer, New York.

Lee, J. H., and Liao, Z. (2016). 'On standard inference for GMM with local identication failure of known forms', Discussion paper, University of Illinois and UC Los Angeles.

Lee, S. (2014). 'Asymptotic refinements of a misspecification-robust bootstrap for generalized method of moments estimators', *Journal of Econometrics*, 178: 398–413.

Madsen, E. (2009). 'GMM-based inference in the AR(1) panel data model for parameter values where local identification fails', Discussion paper, Centre for Applied Microeconometrics, Department of Economics, University of Copenhagen, Copenhagen, Denmark.

Newey, W. K., and McFadden, D. L. (1994). 'Large sample estimation and hypothesis testing', in R. Engle and D. L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2113–2247. Elsevier Science Publishers, Amsterdam, The Netherlands.

Pollard, D. (1984). *Convergence of stochastic processes*. Springer, New York.

Rotnitzky, A., Cox, D. R., Bottai, M., and Robins, J. (2000). 'Likelihood-based inference with singular information matrix', *Bernoulli*, 6: 243–284.

Sargan, J. D. (1983). 'Identification and lack of identification', *Econometrica*, 51: 1605–1633.

Sen, B., Banerjee, M., and Woodroofe, M. (2010). 'Inconsistency of the bootstrap: The Grenander estimator', *Annals of Statistics*, 38: 1953–1977.

Sentana, E. (2009). 'Finite underidentification', Discussion paper, CEMFI, Working Paper No. 1508, Madrid, Spain.

van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge University Press, Cambridge.

van der Vaart, A. W., and Wellner, J. A. (1996). *Weak convergence and empirical processes*. Springer Verlag, New-York.

Wright, J. H. (2003). 'Detecting lack of identification in GMM', *Econometric Theory*, 19: 322–330.

54