# Improved Projection GMM-LM Tests for Linear Restrictions*

Saraswata Chaudhuri†

This version: August 15, 2017.

**Abstract**

This paper extends the two-step GMM-LM projection test for subvectors in Chaudhuri and Zivot (2011) to testing null hypotheses on linear restrictions on a parameter vector. This extension retains all the properties of the original two-step projection test, in particular, its asymptotic rejection probability of the true and the false null hypothesis, but under substantively more general conditions on the identification of the parameter vector. Since we allow for multiple rates (strength) of identification, the linear restriction being tested may be rate-entangled. This leads to novel issues when we consider the rejection of the locally-false null hypothesis to establish the two-step test's asymptotic efficiency. Our results address them. We present a simple example to illustrate the results. A simulation study shows that these asymptotic results are reliable approximations even in samples of small size.

*I thank M. Carrasco, J-M Dufour, A. Mikusheva, E. Renault, J. Robins, T. Richardson, V. Zinde-Walsh and E. Zivot for helpful discussions. I am particularly grateful to D. Andrews for helpful discussions about this paper and for pointing out and discussing with me his related but more extensive work Andrews (2017a,b). To keep this paper short I focus on issues related to local efficiency of a test for parameters with rate-entangled identification that, to my knowledge, have not been in addressed in Andrews (2017a,b) or elsewhere in the literature.

†Department of Economics, McGill University; and Cireq, Montreal, Canada. Email: saraswata.chaudhuri@mcgill.ca.

# 1 Introduction

Consider a parameter vector $\theta \in \Theta \subset \mathbb{R}^{d_\theta}$ whose unknown true value $\theta^0$ satisfies the moment restrictions:

$$E[g(Z_t; \theta^0)] = 0 \tag{1}$$

where $\{Z_t\}_{t=1}^T$ are $\mathbb{R}^{d_z}$-valued random vectors, $g(.;\theta) : \mathbb{R}^{d_z} \times \Theta \mapsto \mathbb{R}^{d_g}$ is a known (up to $\theta$) function, and $d_g \geq d_\theta$. Suppose that we are interested in testing the null hypothesis:

$$H_0 : R\theta^0 = r_0 \text{ against the alterative hypothesis } H : R\theta^0 \neq r_0 \tag{2}$$

where $R$ is a fixed, full row-rank, $d_R \times d_\theta$ known matrix, and $r_0$ is a $d_R \times 1$ known vector, and $d_R < d_\theta$.

This paper extends the improved two-step GMM-LM projection subvector (of $\theta$) test presented in Chaudhuri and Zivot (2011) and built on the original work of Robins (2004), to testing $H_0$ in (2). (Also see Chaudhuri (2008), Zivot and Chaudhuri (2009), Chaudhuri et al. (2010), Chaudhuri and Renault (2011).) For brevity, in the sequel we generally refer to this test and its extension as the *two-step test*.

Chaudhuri and Zivot (2011) established the following results in the context of testing for subvectors. First, allowing for identification failure of $\theta^0$ in (1) as in Stock and Wright (2000), this test's asymptotic rejection probability of the true null can be non-trivially bounded from above, a property that the subvector-plug-in tests cannot possess in general [see Guggenberger et al. (2012a)]. Second, this test is generally more powerful than the standard projection tests as in Dufour (1997), Dufour and Jasiak (2001), Dufour et al. (2006), Dufour and Taamouti (2005, 2007), etc. Indeed, under the classical conditions as in Newey and McFadden (1994)'s Theorem 9.2 (henceforth NM-9.2) and a global identification condition for the nuisance subvector, this test is locally efficient. This efficiency, which is also the focus of our paper, is what fundamentally distinguishes the two-step test from the standard projection tests noted above, or the Bonferroni tests in Dufour (1990), Berger and Boos (1994), Silvapulle (1996), etc.

In this paper, we establish that all these results of Chaudhuri and Zivot (2011) remain valid for testing the hypothesis on general linear restrictions in (2) under more general characterization of the identification failure of $\theta^0$. Importantly, the efficiency result is also strengthened to cover non-classical scenarios that do not satisfy the NM-9.2 conditions, and this is the main contribution of our paper.

This paper is organized as follows. Section 2 summarizes at the outset the key features of our paper, their novelty, consequences, and relation with the literature. Section 3 describes the two-step test and heuristically discusses the efficiency result under the classical NM-9.2 setup. The key to efficiency is the test statistic used in the second step, and for this we use the (GMM) LM version of Neyman (1959)'s C-alpha statistic. We show its numerical equivalence with the efficient score statistic used in Chaudhuri and Zivot (2011). Section 4 allows for identification failure of $\theta^0$ and treats the NM-9.2 setup as a special case. The two-step test's asymptotic rejection probability of the true null is discussed under the setup of Andrews and Guggenberger (2014). Efficiency for suitable local deviations of the null from the truth is discussed by ruling out weak or worse identification of $\theta^0$ but still allowing for multiple rates (strength) of identification in between weak and strong. For this, we impose more structure that nevertheless covers the setups of Stock and Wright (2000), Antoine and Renault (2012), etc. A linear instrumental variables model is used to assess the local efficiency with a simulation study. Technical materials are collected in Appendices A (for the explicit construction of local deviations), B (for Section 3) and C (for Section 4).

## 2 The key features of our paper, and the relation with the literature

To put the contributions of this paper into perspective, before proceeding further, we briefly summarize the key features (F1)-(F4) of our paper, their consequences, and their relation with the literature.

**(F1)** To facilitate imposing $H_0$ in (2) in the two-step test, we re-parameterize the system in (1) as follows. Consider a $(d_\theta - d_R) \times d_\theta$ matrix $S$ such that the $d_\theta \times d_\theta$ matrix $A_S = [R', S']'$, indexed by $S$, is nonsingular. $S$ exists since $R$ is full row-rank. Now, for this $S$, rotate the parameter $\theta$ to define:

$$(\beta', \gamma_S')' := A_S\theta. \tag{3}$$

(1) and (3) imply that $\beta^0 := R\theta^0$ and $\gamma_S^0 := S\theta^0$ are the true values for $\beta$ and $\gamma_S$. The parameter space for $(\beta', \gamma_S')'$ is $\mathcal{B} \times \Gamma_S$ where $\mathcal{B} := \{R\theta : \theta \in \Theta\} \subset \mathbb{R}^{d_R}$ and $\Gamma_S := \{S\theta : \theta \in \Theta\} \subset \mathbb{R}^{d_\theta - d_R}$.

This rotation is different from that considered in Sargan (1983), Phillips (1989), and later in Choi and Phillips (1992), Zivot et al. (2006), Antoine and Renault (2009, 2012), Andrews and Cheng (2012, 2014), Cheng (2015), etc. The rotation in (3) isolates the components/directions in $\theta$ that are identified by the null hypothesis in (2) regardless of the identification that is due to the model (1), whereas in the aforementioned papers, a rotation is employed to isolate the directions identified by the model itself.

One could then simply follow Chaudhuri and Zivot (2011)'s framework of a standard subvector (for $(\beta', \gamma_S')'$) test for $\beta = r_0$ by directly specifying the identification or lack thereof for $\beta^0$ and $\gamma_S^0$ [also see Andrews (2017a)]. However, in practice, moment restrictions such as (1) and, more generally, the GMM framework are typically conceived in terms of $\theta$, and not $\beta := R\theta$ or a nuisance parameter $\gamma_S := S\theta$.

**(F2)** Therefore, we use the re-parameterization in (3) only for computational convenience with the help of an explicitly defined nuisance parameter $\gamma_S$, while the description of the framework is provided entirely in terms of $\theta$, more precisely, $\theta^0$, by strictly adhering to (1).[1] When $H_0$ is false, taking this route

[1]Given this importance that we assign to $\theta^0$, we let the representation in (1) and (2) suffer from two drawbacks that

leads to novel issues since we allow for multiple rates of identification for the elements of $\theta^0$. The issues arise precisely because this allowance can make, *concurrently*, the identification of $\beta^0$ rate-entangled and the strength of identification of $\gamma_S^0$ dependent on the choice of $S$.[2] Our results address them.

To elaborate, suppose, that $d_\theta = 2$, $\theta = (\theta_1, \theta_2)'$ and $R = [1, 1]$, i.e., $\beta = \theta_1 + \theta_2$. Consider a $\gamma_S$ (e.g., $\gamma_S = \theta_1$ if $S = [1, 0]$, $\gamma_S = \theta_2$ if $S = [0, 1]$) for (3). The local efficiency property of the two-step test appeals to its asymptotic equivalence with an infeasible LM test that uses the unknown true $\gamma_S^0$ (hence, infeasible). However, the infeasible test is not invariant to $S$ unless $H_0$ is true. Indeed, it can be very different even locally for the different choices of $S$ if the strengths of identification for the corresponding $\gamma_S^0$'s are not the same. As a result, the interpretation of the local deviations of $H_0$ from the truth, for which the said asymptotic equivalences hold, needs particular care. For example, let $\theta_1^0$ be nearly strongly and $\theta_2^0$ strongly identified. Then, fixing $\gamma_S = \theta_1$ at $\theta_1^0$ leads the local deviation in $\beta = \theta_1 + \theta_2$ to be along the strong direction $\theta_2$, while fixing $\gamma_S = \theta_2$ at $\theta_2^0$ leads the deviation to be along the nearly strong directions $\theta_1$. We consider both (and others, e.g., $\gamma_S = \theta_1 - \theta_2$) under a unified framework maintained in terms of $\theta$ (as noted above). We present our results such that they automatically reflect that the asymptotic equivalence with the latter, which is the less powerful infeasible test ($\gamma_S = \theta_2$), holds in a larger region, i.e., less locally in terms of the local deviation, than that with the former ($\gamma_S = \theta_1$).

**(F3)** One might infer from (F2) and the non-invariance of the infeasible test that the two-step could have better power if it does not impose $H_0$ in the first step. This intuition better suits the plug-in tests which, then lose invariance to $S$ and resemble the corresponding $S$-dependent infeasible test. (However, as we show in Appendix D, this turns out to be bad for the size of the plug-in tests in cases where a standard, i.e., restricted-by-$H_0$, plug-in would have otherwise worked.) On the other hand, this intuition has no justification for the two-step test since it is still invariant to $S$. In fact, this leads to very poor power in small samples except in the NM-9.2 setup. We do not recommend this for the two-step test.

**(F4)** Finally, we note that the notion of optimality/efficiency in relation to the infeasible test is less ambitious than that considered in the literature on identification failure inspired by Moreira (2002, 2003). See Andrews et al. (2006), Moreira and Moreira (2013), Andrews (2016a), Andrews and Mikusheva (2016), Montiel-Olea (2016), etc. By contrast, our use of the term is similar to that in Section 9 of An-

hinder a satisfactory treatment of issues related to similarity and, hence, size of the two-step test. First, (1) and (2) do not adequately distinguish between the truth for $\theta$ (i.e., $\theta = \theta^0$), and what would be the truth for $\beta$ (i.e., $\beta = \beta^0$, equivalently, $R\theta = R\theta^0$) in a standard subvector test representation. The former is a point in $\Theta$, while the latter is a $(d_\theta - d_R)$ dimensional linear subspace of $\Theta$. Second, we consider the true $\theta$ (i.e, $\theta^0$) and hence, given the first drawback, the true $\beta$ (i.e., $\beta^0$) as fixed but let the hypothesized value $r_0$ vary (possibly with $T$), which is what determines if $H_0$ is true or false. Accordingly, our assumptions focus on the fixed true $\theta^0$. While both drawbacks could be bypassed by maintaining the setup in terms of $\beta$ and $\gamma_S$, given our focus on highlighting the issues related to local efficiency, we do not do so for brevity and instead refer the reader to Andrews (2017a) for a comprehensive treatment along this route for the asymptotic size of two-step tests.

[2]While related, these issues are not the same as in Antoine and Renault (2009). Also, they do not arise in the study of efficiency with a single rate of identification as in Chaudhuri and Zivot (2011), (I.)Andrews (2016b), Andrews (2017a), etc. They are also standard if assumptions on identification are instead maintained directly on $\beta$ and the specific $\gamma_S$ being used.

drews and Guggenberger (2015), or Comment (iii) following Theorem 4.1 of Andrews and Guggenberger (2014), or the oracle equivalence considered in Andrews (2017a). Indeed, the LM-principle generally does not lead to optimality other than in a local sense since it is only based on the slope of the moment vector. Furthermore, as originally noted by Kleibergen (2005), allowing for identification failure necessitates the use of an estimator for the Jacobian matrix that is not simply the sample mean of the derivative of the moment vector, but the sample mean of the residual of the regression of this derivative on the moment vector itself. In certain cases of identification failure, this affects the intended direction along which the LM-principle maximizes local power; see, e.g., Antoine and Renault (2009). Even otherwise, this may lead to a spurious decline in power away from the truth [see Kleibergen (2005)], which, however, is partially addressed by the two-step test by virtue of its first step [see Chaudhuri and Zivot (2011)].

Related literature: While we generalize the use of the LM and C-alpha principle in Chaudhuri (2008), Zivot and Chaudhuri (2009), Chaudhuri et al. (2010) and Chaudhuri and Zivot (2011); the LM and/or C-alpha tests were originally used in the context of identification failure by Wang and Zivot (1998), Dufour and Jasiak (2001), Kleibergen (2002), Moreira (2003), Kleibergen (2005), Guggenberger and Smith (2005), Antoine and Renault (2009), etc. It has also been considered more recently in Magnusson and Mavroeidis (2010), Guggenberger et al. (2012b), Qu (2014), Dufour et al. (2015), Andrews and Mikusheva (2015), Andrews and Guggenberger (2014), etc. Even more recently, McCloskey (2015) and Andrews (2016b) propose sophisticated related methods to improve the performance of such tests.

The papers closest to ours are Andrews (2017a,b), where this basic two-step test is substantially generalized, extended, refined and demonstrably improved. However, our results, and hence the above discussion, i.e., (F2)-(F4), apply to all these papers if one considers the local efficiency of the respective tests under our framework, whose importance and practical relevance were noted at the end of (F1).

# 3 Definition and an overview of the improved two-step projection test

This section maintains the classical NM-9.2 setup as default. Define $\bar{g}_T(\theta) := \frac{1}{T}\sum_{t=1}^{T} g(Z_t;\theta)$, $G(\theta) := \frac{\partial}{\partial\theta'}E\left[g(Z_t;\theta)\right]$ and $V(\theta) := Var\left(g(Z_t;\theta)\right)$. Then, the efficient GMM estimator of $R\theta^0$ has the asymptotically linear representation: $\sqrt{T}(\widehat{R\theta^0} - R\theta^0) = -\sqrt{T}l_T(\theta^0) + o_p(1)$ [see Appendix B.1] where

$$l_T(\theta) := R\left(G'(\theta)V^{-1}(\theta)G(\theta)\right)^{-1}G'(\theta)V^{-1}(\theta)\bar{g}_T(\theta),$$

if it exists. Therefore, for local optimality/efficiency, a test for $H_0$ in (2) can be based on a consistent estimator of $l_T(\theta)$:

$$\widehat{l}_T(\theta) := R\left(\widehat{G}'_T(\theta)\widehat{V}_T^{-1}(\theta)\widehat{G}_T(\theta)\right)^{-}\widehat{G}'_T(\theta)\widehat{V}_T^{-1}(\theta)\bar{g}_T(\theta)$$

where $\widehat{G}_T(\theta) \xrightarrow{P} G(\theta)$ and $\widehat{V}_T(\theta) \xrightarrow{P} V(\theta)$, as appropriate, and $(.)^-$ denotes a g-inverse.[3]

For this reason, we use the following quadratic form of $\widehat{l}_T(\theta)$ for the second step of the two-step test:

$$\begin{aligned} LM_T(\theta) &:= T \times \widehat{l}_T'(\theta) \left( R \left( \widehat{G}_T'(\theta) \widehat{V}_T^{-1}(\theta) \widehat{G}_T(\theta) \right)^- R' \right)^- \widehat{l}_T(\theta) \\ &= T \times \left( \widehat{V}_T^{-1/2}(\theta) \bar{g}_T(\theta) \right)' P \left( \widehat{V}_T^{-1/2}(\theta) \widehat{G}_T(\theta) \left( \widehat{G}_T'(\theta) \widehat{V}_T^{-1}(\theta) \widehat{G}_T(\theta) \right)^- R' \right) \left( \widehat{V}_T^{-1/2}(\theta) \bar{g}_T(\theta) \right) \end{aligned} \tag{4}$$

where we use the notation $P(D) := D(D'D)^- D'$ to define the projection matrix for any matrix $D$; and if $D$ is positive semidefinite then we define $D^{1/2}$ to be the upper triangular matrix such that $D = D^{1/2'} D^{1/2}$.

$LM_T(\theta)$ is Smith (1987)'s $LLM_T$, Dagenais and Dufour (1991)'s $PC$ or Newey and McFadden (1994)'s $LM_{2n}$ statistic for testing linear restrictions. It falls under the class of Neyman (1959)'s C-alpha statistic.

Note that, (3) and (4) give:

$$LM_T(\theta) \equiv LM_T \left( A_S^{-1} (\beta', \gamma_S')' \right). \tag{5}$$

**Definition:** For $\epsilon, \alpha > 0$ and $\epsilon + \alpha < 1$, the improved two-step GMM-LM projection test, or simply the two-step test, for $H_0$ in (2) is defined as:

$$\begin{aligned} &\text{Step 1: obtain a nominal } (1-\epsilon)\text{-level confidence set } CI_T(\gamma_S; \epsilon) \text{ for } \gamma_S^0; \\ &\text{Step 2: reject } H_0 \text{ if } CI_T(\gamma_S; \epsilon) \text{ is empty or if } \inf_{\gamma_0 \in CI_T(\gamma_S; \epsilon)} LM_T \left( A_S^{-1} (r_0', \gamma_0')' \right) > \chi^2_{d_R}(1-\alpha) \end{aligned} \tag{6}$$

where $\chi^2_{d_R}(1-\alpha)$ is the $(1-\alpha)$-th quantile of a central $\chi^2$ distribution with $d_R$ degrees of freedom.

$CI_T(\gamma_S; \epsilon)$ is what we refer to as the first-step confidence set, i.e., the preliminary non-point (set) estimator of the nuisance parameter $\gamma_S^0 := S\theta^0$ that is not specified by $H_0$. As noted below, it plays an important role in influencing the asymptotic properties and the computational ease of the two-step test.

**Remark 1:** Invariance of the two-step test to the choice of $S$ in (3) is preserved by the conventional confidence sets $CI_T(\gamma_S; \epsilon)$ regardless of the non-uniqueness of the infimum in the second step. If possible, however, choosing an $S$ with a better identified $\gamma_S^0$ might help with the computation in the second step.

**Remark 2:** To accommodate for identification failures of $\theta^0$, $CI_T(\gamma_S; \epsilon)$ can be obtained by inverting, e.g., the S-test, the K-test, modifications of Moreira (2003)'s CLR test (see Kleibergen (2005), Andrews and Guggenberger (2014, 2015)) for $\gamma_S$, while treating $\beta = r_0$ as known. In practice, the operations required in steps one and two can be simultaneously conducted since, to fail to reject $H_0$, it is sufficient to find a single point $\gamma_0$ that would belong in $CI_T(\gamma_S; \epsilon)$ and also satisfy the condition for step two.

**Remark 3:** It is sufficient to focus on the second-step test statistic to see the connection between (6) and the two-step projection subvector test in Chaudhuri and Zivot (2011). For a given $S$, let $R_S^1$ and $S_S^1$

[3] The non-classical setup in Section 4 lets $G(\theta)$ be column-rank deficient. But we always maintain nonsingularity of $V(\theta)$. Our use of the g-inverse in the definitions of $\widehat{l}_T(\theta)$ and $LM_T(\theta)$ reflects this asymmetric treatment of $G(\theta)$ and $V(\theta)$.

be the $d_\theta \times d_R$ and $d_\theta \times (d_\theta - d_R)$ matrices respectively such that $A_S^{-1} = [R_S^1, S_S^1]$. Then, in this context, the test statistic in Chaudhuri and Zivot (2011)'s second step would be [see Appendix B.2]:

$$LM_{T,S}^{\text{ES}}(\theta) := T \times \left(\widehat{V}_T^{-1/2}(\theta)\bar{g}_T(\theta)\right)' P\left(\left(I_{d_g} - P\left(\widehat{V}_T^{-1/2}(\theta)\widehat{G}_T(\theta)S_S^1\right)\right)\widehat{V}_T^{-1/2}(\theta)\widehat{G}_T(\theta)R_S^1\right)\left(\widehat{V}_T^{-1/2}(\theta)\bar{g}_T(\theta)\right),$$

which they refer to as the efficient score (ES) statistic for $\beta = R\theta$. We show in Appendix B.3 that:

**Lemma 1** *Let $\theta \in interior(\Theta)$ be such that $\widehat{G}_T'(\theta)\widehat{V}_T^{-1}(\theta)\widehat{G}_T(\theta)$ is positive definite almost surely for some $T$. Then $LM_{T,S^*}^{ES}(\theta) = LM_{T,S^\dagger}^{ES}(\theta)$ almost surely for any $S = S^*, S^\dagger$ for which $[R', S']'$ is nonsingular.*

**Lemma 2** *Let $\theta \in interior(\Theta)$ be such that $\widehat{G}_T'(\theta)\widehat{V}_T^{-1}(\theta)\widehat{G}_T(\theta)$ is positive definite almost surely for some $T$. Then $LM_T(\theta) = LM_{T,S}^{ES}(\theta)$ almost surely for any choice of $S$ for which $[R', S']'$ is nonsingular.*

Lemma 1 establishes invariance for $LM_{T,S}^{\text{ES}}(\theta)$ similar to that for $LM_T(\theta)$ in (5). Lemma 2 reconciles between $LM_T(\theta)$ and $LM_{T,S}^{\text{ES}}(\theta)$ in the context of testing (2) with a general $R$. While expected from the relationship between the efficient influence function and the efficient score function, on which $LM_T(\theta)$ and $LM_{T,S}^{\text{ES}}(\theta)$ are based respectively, this is, to our knowledge, a new result in the C-alpha literature.
**Remark 4:** It is useful to note here that the conventional projection test rejects $H_0$ at the level $\alpha$ if:

$$\inf_{\theta_0 \in \Theta: R\theta_0 = r_0} \widetilde{LM}_T(\theta_0) > \chi^2_{d_\theta}(1-\alpha) \quad \text{or equivalently,} \quad \inf_{\gamma_0 \in \Gamma_S} \widetilde{LM}_T\left(A_S^{-1}(r_0', \gamma_0')'\right) > \chi^2_{d_\theta}(1-\alpha) \qquad (7)$$

where [see the $LM_{3n}$ statistic in Newey and McFadden (1994) and the K statistic in Kleibergen (2005)]:

$$\widetilde{LM}_T(\theta) := T \times \left(\widehat{V}_T^{-1/2}(\theta)\bar{g}_T(\theta)\right)' P\left(\widehat{V}_T^{-1/2}(\theta)\widehat{G}_T(\theta)\right)\left(\widehat{V}_T^{-1/2}(\theta)\bar{g}_T(\theta)\right).$$

The second version in (7) explicitly imposes $H_0$ by using (3) and thus streamlines the computation for the test. Note that, $\widetilde{LM}_T(\widetilde{\theta}_T) = LM_T(\widetilde{\theta}_T)$ where $\widetilde{\theta}_T$ is the restricted-by-$H_0$ GMM estimator of $\theta_T$ [see Appendix B.4]. Then, NM-9.2 gives: $LM_T(\widetilde{\theta}_T) \xrightarrow{d} \chi^2_{d_R}$ distribution, which is central if $H_0$ is true, and non-central under local deviations of $H_0$. Hence, the conservativeness (and thereby, the inefficiency) of the conventional projection test, that the two-step test would address, is due to the use of a $\chi^2_{d_\theta}$ critical value in (7), while the test statistic is actually $\inf_{\theta_0 \in \Theta: R\theta_0 = r_0} \widetilde{LM}_T(\theta_0) \leq \widetilde{LM}_T(\widetilde{\theta}_T) = LM_T(\widetilde{\theta}_T) \xrightarrow{d} \chi^2_{d_R}$.
**Remark 5:** By contrast, the local efficiency of the two-step test, that one would expect from (4), results as follows. For the given $S$, let $\theta_0 := A_S^{-1}(r_0', \gamma_0')'$ where $r_0 := \beta^0 + \mu_\beta/\sqrt{T}$, $\gamma_0 := \gamma_S^0 + \mu_{\gamma S}/\sqrt{T}$, $\mu_\beta$ is a constant and $\mu_{\gamma_S} = O_p(1)$. Then, NM-9.2 gives: $\widehat{G}_T(\theta_0) \xrightarrow{P} G(\theta^0)$, $\widehat{V}_T(\theta_0) \xrightarrow{P} V(\theta^0)$ and, crucially,

$$\sqrt{T}\widehat{l}_T(\theta_0) \xrightarrow{P} \sqrt{T}l_T(\theta^0) + \mu_\beta$$

by using $RR_S^1 = I_{d_\theta}$ and $RS_S^1 = 0$ that follow from $A_S A_S^{-1} = I_{d_\theta}$. Therefore, NM-9.2 gives: $LM_T(\theta_0) \xrightarrow{d} \chi^2_{d_R}$ with non-centrality parameter $\mu'_\beta \left(R\left(G'(\theta^0)V^{-1}(\theta^0)G(\theta^0)\right)^{-1} R'\right)^{-1} \mu_\beta$, which does not depend on the $\sqrt{T}$-deviation of $\gamma_0$ from $\gamma_S^0$. On the other hand, under the same conditions and a global strong identification condition for $\gamma_S^0$ given $\beta^0$, it can be shown that [see Lemma 7 in Section 4 for a more general result]:

$$\sup_{\gamma_0 \in \in CI_T(\gamma_S;\epsilon)} \sqrt{T}\|\gamma_0 - \gamma_S^0\| = O_p(1)$$

for conventional confidence sets, provided they are non-empty (to fix ideas for now). Hence, by construction, $\gamma_{S,T}^\dagger := \arg\inf_{\gamma_0 \in CI_T(\gamma_S;\epsilon)} LM_T\left(A_S^{-1}(r_0', \gamma_0')'\right) = \gamma_S^0 + \mu_{\gamma_S,T}/\sqrt{T}$ for some $\mu_{\gamma_S,T} = O_p(1)$. Therefore, the two-step test in (6), with a non-empty first-step confidence set, is locally efficient since it is asymptotically equivalent to the locally optimal/efficient infeasible test – based on the infeasible efficient influence function and the infeasible unknown true $\gamma_S^0$ – that rejects $H_0$ at the level $\alpha$ if for $\theta_{0,S}^{infs} := A_S^{-1}(r_0', \gamma_S^{0\prime})'$:

$$LM_T^{infs}(\theta_{0,S}^{infs}) := T \times l_T'(\theta_{0,S}^{infs}) \left(R\left(G'(\theta_{0,S}^{infs})V^{-1}(\theta_{0,S}^{infs})G(\theta_{0,S}^{infs})\right)^{-1} R'\right)^{-1} l_T(\theta_{0,S}^{infs}) > \chi^2_{d_R}(1-\alpha). \quad (8)$$

This optimality discussion is only under the classical NM-9.2 setup. Section 4 presents a general treatment allowing for identification failure of $\theta^0$ and, hence, considering the NM-9.2 setup as a special case.

# 4 Asymptotic Properties: When identification failure of $\theta^0$ is allowed

The choice of $\widehat{G}_T(\theta)$ in the definition of $LM_T(\theta)$ in (4) is important. To account for a possible identification failure of $\theta^0$, it is imperative that the choice follows Kleibergen (2005). That is:

$$\widehat{G}_T(\theta) := \left[\widehat{G}_{1,T}(\theta), \ldots, \widehat{G}_{d_\theta,T}(\theta)\right] \quad \text{where} \quad \widehat{G}_{T,j}(\theta) := \frac{\partial}{\partial \theta_j}\bar{g}_T(\theta) - \widehat{V}_{j,g,T}(\theta)\widehat{V}_T^{-1}(\theta)\bar{g}_T(\theta),$$

$\widehat{V}_{j,g,T}(\theta)$ and $\widehat{V}_T(\theta)$ are respectively $d_\theta \times d_g$ and $d_g \times d_g$ matrices, and $\theta_j$ is the $j$-th element of $\theta$ for $j = 1, \ldots, d_\theta$. Unless otherwise noted [see Remark 9], we require $\widehat{V}_{j,g,T}(\theta)$ and $\widehat{V}_T(\theta)$ to be estimators of

$$\begin{aligned} V_{j,g}(\theta) &:= \lim_{T\to\infty} T \times E\left[\left(\frac{\partial}{\partial \theta_j}\bar{g}_T(\theta) - E\left[\frac{\partial}{\partial \theta_j}\bar{g}_T(\theta)\right]\right)\bar{g}_T(\theta)'\right] \text{ for } j = 1, \ldots, d_\theta \\ \text{and } V(\theta) &:= \lim_{T\to\infty} T \times E\left[\left(\bar{g}_T(\theta) - E[\bar{g}_T(\theta)]\right)\bar{g}_T(\theta)'\right], \end{aligned}$$

provided they exist. Also applicable are the choices of $\widehat{G}_T(\theta)$ considered in Guggenberger and Smith (2005, 2008) that only deviate from $\widehat{G}_T(\theta)$ defined above by an order of magnitude of $o_p(1/\sqrt{T})$.

We maintain high-level but standard assumptions on the joint distribution $F_T$ of the data $\{Z_t\}_{t=1}^T$. Allowing for a drifting data generating process (DGP) in what follows is important, and to emphasize it we index by $T$ the key parameters defined in terms of $F_T$; see, e.g., Stock and Wright (2000), Andrews and Guggenberger (2014). However, irrespective of the drifting DGP $\{F_T : T \geq 1\}$, we take the truth $\theta^0$ satisfying the moment restrictions in (1) as fixed. $H_0$ in (2) is true if the hypothesized value $r_0$ is equal to $R\theta^0$, it is false otherwise. Apart from characterizing the false $H_0$ by locally deviating (made precise later in (14) and (15)) $r_0$ from $R\theta^0$, no other assumptions involve $r_0$. For convenience, we maintain that:

**Assumption O:**

$\theta^0 \in \text{int}(\Theta)$ where $\Theta$ is compact in $\mathbb{R}^{d_\theta}$ and $\text{int}(\Theta) := \text{interior}(\Theta)$.

**Notation:** We suppress the triangular array $\{Z_{t,T} : t = 1, \ldots, T; T \geq 1\}$ notation, and instead denote $Z_{t,T}$ by $Z_t$. Let $\underline{c} > 0$ and $\bar{c} > 0$ denote generic constants. For any matrix $D$, define $\|D\| := \sqrt{\text{trace}(D'D)}$. For any $a \times b$ matrix $D = [D_1, \ldots, D_b]$ define $D_{(j:k)} := [D_j, \ldots, D_k]$ as the $a \times (k-j+1)$ matrix for $1 \leq j \leq k \leq b$. $D_{(k:j)}$ is an empty matrix for $0 \leq j < k \leq b+1$. For an $(ab) \times 1$ vector $D = (d_1, \ldots, d_{ab})'$, define $devec_b(D) := [(d_1, \ldots, d_b)', (d_{b+1}, \ldots, d_{2b})', \ldots, (d_{(a-1)b+1}, \ldots, d_{ab})']$ as a $b \times a$ matrix.

### 4.1 Rejection of the null hypothesis in (2) when it is true

**Assumption M:**

M1. $\frac{\partial}{\partial\theta'} g(z;\theta^0)$ exists for each $z \in \mathbb{R}^{d_z}$. Let $\bar{G}_T(\theta) := \frac{1}{T}\sum_{t=1}^{T} \frac{\partial}{\partial\theta'} g(Z_t;\theta)$ and $G_T := E_T[\bar{G}_T(\theta^0)]$. Then,

$$\left[\sqrt{T}\bar{g}_T'(\theta^0),\ \sqrt{T}vec(\bar{G}_T(\theta^0) - G_T)'\right] \xrightarrow{d} [\psi',\ \psi_G'] \sim N(0, \Sigma)$$

$$\text{and } \lim_{T\to\infty} Var_T \begin{pmatrix} \sqrt{T}\bar{g}_T(\theta^0) \\ \sqrt{T}vec(\bar{G}_T(\theta^0)) \end{pmatrix} \equiv \lim_{T\to\infty} \begin{bmatrix} V_T & V_{gG,T} \\ V_{Gg,T} & V_{GG,T} \end{bmatrix} = \Sigma := \begin{bmatrix} V & V_{gG} \\ V_{Gg} & V_{GG} \end{bmatrix}.$$

M2. $\max\{\|G_T\|, \|V_T\|, \|V_T^{-1}\|, \|V_{Gg,T}\|\} \leq \bar{c}$ for all $T$. $\|\widehat{V}_T - V_T\| = o_p(1)$ and $\|\widehat{V}_{Gg,T} - V_{Gg,T}\| = o_p(1)$.

To characterize any identification failure of $\theta^0$, consider the singular value decomposition of $V_T^{-1/2}G_T$:

$$V_T^{-1/2}G_T = C_T\bar{\Delta}_T B_T' \tag{9}$$

where $C_T$ and $B_T$ are $d_g \times d_g$ and $d_\theta \times d_\theta$ orthogonal matrices whose columns are respectively the eigenvectors of the matrices $V_T^{-1/2}G_TG_T'V^{-1/2'}$ and $G_T'V^{-1}G_T$. $\bar{\Delta}_T := [\Delta_T, 0]'$ is the $d_g \times d_\theta$ matrix where $\Delta_T := \text{diag}(\delta_{T,1}, \ldots, \delta_{T,d_\theta})$ is the $d_\theta \times d_\theta$ diagonal matrix with its diagonal elements $\delta_{T,1} \geq \delta_{T,2} \geq \ldots \geq \delta_{T,d_\theta}$ ($\geq 0$, without loss of generality) as the singular values of $V_T^{-1/2}G_T$.

**Assumption M** (continued)**:** (identification failure of $\theta^0$ following Andrews and Guggenberger (2014))

M3. For the singular value decomposition in (9), there exists a $p \in \{0, 1, \ldots, d_\theta\}$ such that:

(a) $\delta_{T,j} \to \delta_j$, a constant, and $\sqrt{T}\delta_{T,j} \to \infty$ for $j = 1, \ldots, p$ as $T \to \infty$ (M3(a) is void if $p = 0$);

(b) $\sqrt{T}\delta_{T,j} \to l_j$, a constant, for $j = p+1, \ldots, d_R$ as $T \to \infty$ (M3(b) is void if $p = d_\theta$);

(c) $C_T \to C$ and $B_T \to B$ as $T \to \infty$ where $B$ is a nonsingular matrix;

(d) $G^* := [C_{(1:p)}, C_{(p+1:d_\theta)}L + V^{-1/2}(\theta^0)devec_{d_g}(\psi_G - V_{Gg}V^{-1}\psi)B_{(p+1:d_\theta)}]$ is a $d_g \times d_\theta$ matrix with full column-rank $d_\theta$ almost surely, where $L := \text{diag}\,(l_{p+1}, \ldots, l_{d_\theta})$ is a $(d_\theta - p) \times (d_\theta - p)$ diagonal matrix with $l_{p+1}, \ldots, l_{d_\theta}$ as its diagonal elements if $p < d_\theta$, and $L$ is empty if $p = d_\theta$.

**Remark 6:** $p$ is the number of directions in $\theta$ that are better than weakly identified. The remaining $d_\theta - p$ directions in $\theta$ are at best weakly identified and necessitate the particular choice of $\widehat{G}_T(\theta)$. Assumption M3 and the representation involved in it are entirely based on the original work of Andrews and Guggenberger (2014). M3 is actually slightly stronger than what Andrews and Guggenberger (2014) require; it helps to avoid certain peripheral complications arising from the fact that $d_R < d_\theta$. The other assumptions, O, M1 and M2, are standard; see, e.g., Kleibergen (2005), Guggenberger and Smith (2005).

**Lemma 3** *Let assumptions O and M1-M3 hold. Then, for $LM_T(\theta^0)$ defined in* (4), $LM_T(\theta^0) \xrightarrow{d} \chi^2_{d_R}$.

**Proposition 4** *Let the null hypothesis $H_0$ in* (2) *be true, i.e., $r_0 = R\theta^0$ for $\theta^0$ defined in* (1). *Let the joint distribution $\{F_T : T \geq 1\}$ of $\{Z_t\}_{t=1}^T$ be constrained by the assumptions O and M1-M3. Let $\epsilon, \alpha > 0$ and $\epsilon + \alpha < 1$. Let $CI_T(\gamma_S; \epsilon)$ be a confidence set for $\gamma_S$ defined in* (3) *with asymptotic coverage* $(1 - \epsilon)$ *for $\gamma_S^0 := S\theta^0$. Then, the probability with which the improved two-step projection test in* (6) *rejects $H_0$ cannot exceed $(\epsilon + \alpha)$ asymptotically.*

**Remark 7:** The result follows by Bonferroni's inequality applied to Lemma 3 and the asymptotic coverage of $CI_T(\gamma_S; \epsilon)$. Importantly, the upper bound $(\epsilon + \alpha)$ is entirely under the control of the user.

**Remark 8:** An example of the first-step confidence set $CI_T(\gamma_S; \epsilon)$ that possesses the desired property is:

$$CI_T^{SW}(\gamma_S; r_0, \epsilon) := \left\{\gamma_0 \in \Gamma_S : T \times Q_T(A_S^{-1}(r_0', \gamma_0)') \leq \chi^2_{d_g}(1 - \epsilon)\right\}, \tag{10}$$

obtained by inverting the $S$-test of Stock and Wright (2000) (SW). $r_0$ is the hypothesized value from $H_0$ in (2). $CI_T^{SW}(\gamma_S; r_0, \epsilon)$ imposes $H_0$ and utilizes the re-parameterization in (3) for the computation of

$$Q_T(\theta) := \bar{g}_T'(\theta)\widehat{V}_T^{-1}(\theta)\bar{g}_T(\theta), \tag{11}$$

the standard continuously updated (CU) GMM criterion function. Theorem 2 of Stock and Wright (2000) establishes that the asymptotic coverage of $CI_T^{SW}(\gamma_S; r_0, \epsilon)$ for $\gamma_S^0 := S\theta^0$ is $(1 - \epsilon)$ when $H_0$ in (2) is true and when: (a) $\sqrt{T}\bar{g}_T(\theta^0) \xrightarrow{d} \psi$ and (b) $\widehat{V}_T \xrightarrow{P} V$. Since (a) and (b) are included in M1 and M2, the asymptotic coverage for $CI_T^{SW}(\gamma_S; r_0, \epsilon)$ holds under weaker conditions than what we maintain here. What also works here, but not recommended otherwise, is an (ur)unrestricted-by-$H_0$ version:

$$CI_T^{ur-SW}(\gamma_S; \epsilon) := \left\{\gamma = S\theta : \theta \in \Theta,\ T \times Q_T(\theta) \leq \chi^2_{d_g}(1 - \epsilon)\right\}. \tag{12}$$

However, we have found $CI_T^{SW}(\gamma_S; r_0, \epsilon)$ in (10) to be much more useful in practice for all purposes.[4,5]

### 4.2 Rejection of the null hypothesis in (2) when it locally deviates from the truth

We focus on constructing local deviations of the null from the truth that make the null locally false, and generate results resembling that in the classical NM-9.2 setup. While the construction is generic, as noted in Section 2 (F2), its interpretation for specific choices of $S$ in (3) helps to clarify the local efficiency properties of the test. This is a key aspect of this subsection. To appeal to contiguity arguments, we rule out weak or worse identification of $\theta^0$. In terms of assumption M3, it means $p = d_\theta$.

Given this, a key issue is the rate at which $CI_T(\gamma_S; \epsilon)$ shrinks to $\gamma_S^0$ [see Remark 5]. This necessitates characterizing $E_T[\bar{g}_T(\theta)]$ globally for $\theta \in \Theta$ (more specifically, $E_T[\bar{g}_T(A_S^{-1}(r_0', \gamma_S')')]$ for $\gamma_S \in \Gamma_S$).[6] Two well-known setups for this are Stock and Wright (2000) and Antoine and Renault (2012). They differ in terms of their consequences in that the former results in rate-disentangled local identification of $\theta^0$, while the latter generally leads to rate-entanglement. Both will lead to rate-entangled $R\theta^0$ in general.

Both setups (and also that in Section 4.1) require similar local (near $\theta^0$) analysis of $LM_T(\theta)$ since the local behavior of $E_T[\bar{g}_T(\theta)]$ under them can be characterized by the existence of a nonsingular matrix (depends on $T$ and the setup) which, when *pre-multiplied by* the (scaled) expected Jacobian, converges to a finite $d_g \times d_\theta$ matrix of full column-rank. However, the analyses of the rate at which $CI_T(\gamma_S; \epsilon)$ shrinks to $\gamma_S^0$ are different. A relatively simple but notation-heavy extension of Chaudhuri and Zivot (2011) gives this analysis under Stock and Wright (2000), but not under Antoine and Renault (2012).[7]

Therefore, we take the following strategy for the sake of brevity: We explicitly follow Antoine and Renault (2012) for the discussion in this subsection, and remark on the simplifications that are possible under Stock and Wright (2000). Also, since modeling $E_T[\bar{g}_T(\theta)]$ explicitly has implications on its derivative, there is no reason to continue with the general setup from Section 4.1. Hence, we abandon it, and refer to Andrews (2017a) for a rigorous treatment with the general setup [also see footnote 6].

[4]Amongst the well-known identification-robust confidence sets [see Remark 2], Chaudhuri and Zivot (2011) recommend the use of $CI_T^{SW}(\gamma_S; r_0, \epsilon)$ because of its: (i) validity under weak and general conditions, (ii) computational simplicity, and (iii) effectiveness in eliminating certain spurious declines in power of the GMM-LM test from the second step of the improved projection test. The $\epsilon$ in the upper bound in Proposition 4 is, in practice, primarily due to the fact that $CI_T^{SW}(\gamma_S; r_0, \epsilon)$ can be empty with nonzero probability (that increases with $\epsilon$). While possibly unsatisfying in theory [see Andrews (2017a) for ways to avoid it], this feature is actually useful for (iii) and also for (ii), and hence is accommodated in the definition of the improved two-step projection test in (6). Thus, the recommendation is in spite of the concern raised in Davidson and MacKinnon (2014) and Muller and Norets (2016) (page 2184) that $CI_T^{SW}(\gamma_S; r_0, \epsilon)$ does not properly reflect the parameter uncertainty, since this concern is at least partly addressed by the second step of the improved two-step projection test.

[5]This is a continuation of footnote 4. While it is clear that $CI_T(\gamma_S; \epsilon)$ based on Kleibergen (2005)'s GMM-LM principle cannot be helpful for (iii) in general, it should be noted that $CI_T(\gamma_S; \epsilon)$ based on Moreira (2003)'s conditional likelihood ratio principle may not also be helpful for (iii). Simulation evidence and discussion on this can be found in Section 7.2.1 of Andrews (2016b). Both such $CI_T(\gamma_S; \epsilon)$'s can also be less appealing in terms of (i) and (ii) [see Mikusheva (2010) for (ii)].

[6]See Remark 5. Andrews (2017a) ensures this in a similar setup by maintaining a global strong-identification condition SI2(i)/(13.2) for $\gamma_S^0$ given $\beta = r_0$, and a local strong-identification condition SI2(ii)/SS2$_{\mathrm{LM}}$(i) for $\theta^0$. By contrast, our local and global conditions are going to be inter-related, and they allow for worse than strong but better than weak identification.

[7]Lemma A.1 and Theorem 3.1 of Antoine and Renault (2012) work with point estimators, and do not help with this.

Accordingly, following Antoine and Renault (2012), for some $\rho : \Theta \mapsto \mathbb{R}^{d_g}$ and a sequence of diagonal matrices $\{\Lambda_T : T \geq 1\}$ , let

$$E_T\left[\bar{g}_T(\theta)\right] = \frac{\Lambda_T}{\sqrt{T}}\rho(\theta). \tag{13}$$

**Notation:** Let $1_c$ denote the $1 \times c$ vector with all elements equal to 1. For a set of $d_j$-dimensional vectors $\{a_j\}_{j=1}^q$, let $\text{diag}(a_1, \ldots, a_q)$ denote the $\sum_{j=1}^q d_j$-dimensional diagonal matrix with diagonal elements as the elements of $a_1, \ldots, a_q$ respectively. Let $\mathcal{N}(\theta^0) \subset \Theta$ denote a generic open neighborhood of $\theta^0$.

**Assumption N:** (following Antoine and Renault (2012))

N1. $\rho(\theta) = 0$ if and only if $\theta = \theta^0$.

N2. $\psi_T(\theta) := \sqrt{T}\left(\bar{g}_T(\theta) - E_T\left[\bar{g}_T(\theta)\right]\right) \Rightarrow \psi(\theta)$ where $\psi(\theta)$ is a Gaussian process on $\Theta$ with mean zero and covariance function $V(\theta_1, \theta_2)$. $V(\theta^0) = V$ (as in M1) where $V(\theta) := V(\theta, \theta)$.

N3. $\{\Lambda_T : T \geq 1\}$ is a deterministic sequence of $d_g \times d_g$ diagonal matrix with positive diagonal elements. $I^*$ is a $d_g \times d_g$ matrix whose rows are a suitable permutation of the rows of $I_{d_g}$ such that $I^* \Lambda_T I^{*'} = \text{diag}(\lambda_{T,1} 1_{k_1}, \ldots, \lambda_{T,l} 1_{k_l})$ where $k_j > 0$ for $j = 1, \ldots, l$ and $\sum_{j=1}^l k_j = d_g$. $\lambda_{T,j} = o(\lambda_{T,j+1})$ for $j = 1, \ldots, l-1$; $\lim_T \lambda_{T,1} = \infty$ but $\lim_T \lambda_{T,l}/\sqrt{T} < \infty$.[8]

N4. The $d_g \times d_\theta$ matrix $\rho_\theta(\theta) := \frac{\partial}{\partial \theta'}\rho(\theta)$ exists, has full column-rank $d_\theta$, and is continuous in $\theta \in \mathcal{N}(\theta^0)$.

N5. $g(z; \theta)$ is differentiable in $\theta \in \mathcal{N}(\theta^0)$ for each $z \in \mathbb{R}^{d_z}$.

N6. $\frac{\partial}{\partial \theta'}\psi_T(\theta^0) = \sqrt{T}\left[\bar{G}_T(\theta^0) - \frac{\Lambda_T}{\sqrt{T}}\rho_\theta(\theta^0)\right] = O_p(1)$.

N7. N7 (a) and (b) are grouped together to allow us to briefly discuss a tradeoff in Remark 9 below.

(a) $\rho(\theta)$ is twice continuously differentiable in $\theta \in \mathcal{N}(\theta^0)$. $g(z; \theta)$ is twice differentiable in $\theta \in \mathcal{N}(\theta^0)$ for each $z \in \mathbb{R}^{d_z}$. $\sup_{\theta \in \mathcal{N}(\theta^0)}\left\|\frac{\partial}{\partial \theta_i}\left[\bar{G}_T(\theta) - \frac{\Lambda_T}{\sqrt{T}}\rho_\theta(\theta)\right]\right\| = o_p(\lambda_{T,l}/\sqrt{T})$ for $i = 1, \ldots, d_\theta$.

(b) $\lambda_{T,1}^2/\lambda_{T,l} \to \infty$ as $T \to \infty$.

N8. $\sup_{\theta \in \Theta}\|\widehat{V}_T(\theta) - V(\theta)\| = o_p(1)$. $\sup_{\theta \in \mathcal{N}(\theta^0)}\|\widehat{V}_{Gg,T}(\theta) - V_{Gg}(\theta)\| = o_p(1)$. $V(\theta)$ and $V_{Gg}(\theta)$ are continuous in $\theta \in \mathcal{N}(\theta^0)$. $\sup_{\theta \in \Theta}\max[\text{eigen values}(V(\theta))] \leq \bar{c} < \infty$ and $\inf_{\theta \in \Theta}\min[\text{eigen values}(V(\theta))] \geq \underline{c} > 0$. Also, $V_{Gg}(\theta)$ is finite for $\theta \in \mathcal{N}(\theta^0)$.[9]

**Remark 9:** Assumption N describes the setup following Antoine and Renault (2012), who also provide discussion on each of them. Some assumptions are maintained locally since that suffices for us. Given the rest, N7(b) can be slightly weakened at the cost of messy notation; in particular, it suffices if $\lambda_{T,j_1}^2/\lambda_{T,l} \to \infty$ as $T \to \infty$ where $\lambda_{T,j_1}$ has a rather long definition stated in Appendix A.2.2. It

[8] $I^{*-1} = I^{*'}$. $I^*$ is not unique unless $k_1 = \ldots = k_l = 1$ and thus $l = d_g$. The multiplicity of the elements can be made dependent on $T$ and $\theta$ at the cost of significantly involved notation, but such generalizations may not be relevant in practice.

[9] We deviate from M2 in N8 by abstracting from the $\{F_T\}$-(i.e., $T$)-dependence of the second moments $V_T(\theta)$ and $V_{Gg,T}(\theta)$ to focus on the key issue under our scope.

should, however, be noted that there is a tradeoff between the smoothness assumption N7(a) and the rate assumption in N7(b). This explains why, e.g., in a linear instrumental variables regression where the higher order derivatives are necessarily zero, one does not, unlike N7(b), need to impose a lower bound to the rate at which $\lambda_{T,1} \to \infty$.[10] Finally, note that, while $V_{Gg}(\theta)$ in N8 should ideally be such that $V_{Gg}(\theta^0) = V_{Gg}$ defined in M1, this is not necessary here since we no longer allow for weak identification.

**Remark 10:** In this setup, Antoine and Renault (2012) characterize the orthogonal rotation of $\theta^0$ with rate-disentangled identification as $\Pi'_{\rho_\theta}\theta^0$ and with the rates given by $\sqrt{T}D^{-1}_{T,\rho_\theta}$. $\Pi_{\rho_\theta}$ and $D_{T,\rho_\theta}$ are, respectively, $d_\theta \times d_\theta$ orthogonal and diagonal matrices dependent on $\frac{\Lambda_T}{\sqrt{T}}\rho_\theta(\theta^0)$ and with rather involved definitions that, for the sake of readability, are presented in detail in Appendices A.2.1 and A.2.2.

**Definition:** For a fixed $d_R \times 1$ vector $\mu_\beta$, define the local deviation of the null $H_0$ from the truth as:

$$\sqrt{T}D_{T,R}\Pi'_R(r_0 - \beta^0) = \mu_\beta, \tag{14}$$

where $D_{T,R}$ and $\Pi_R$ are, respectively, $d_R \times d_R$ diagonal and orthogonal matrices dependent on $R$, $\Pi_{\rho_\theta}$ and $D_{T,\rho_\theta}$. The rather involved definitions of $\Pi_R$ and $D_{T,R}$ are presented in Appendices A.3.1 and A.3.2.[11]

However, as noted in Section 2 (F2), (14) is not sufficient for a study of local efficiency in a rate-entangled setup like ours. More is needed to complete the description of "local". Accordingly, consider an arbitrary and possibly random sequence $\{\gamma_{S,T} : T \geq 1\} \in \Gamma_S$ such that $\theta_T := A_S^{-1}(r'_0, \gamma'_{S,T})'$ satisfies:

$$\sqrt{T}D^{-1}_{T,\rho_\theta}\Pi'_{\rho_\theta}(\theta_T - \theta^0) \equiv \sqrt{T}D^{-1}_{T,\rho_\theta}\Pi'_{\rho_\theta}\left(R^1_S(r_0 - \beta^0) + S^1_S(\gamma_{S,T} - \gamma^0_S)\right) = \mu_{T,\theta} \text{ for some } \mu_{T,\theta} = O_p(1). \tag{15}$$

**Definition:** Define the full row-rank matrix $R^*$ [Appendix A.3.3 contains the details of its construction in (24)] as:

$$R^* := \lim_{T\to\infty} D_{T,R}\Pi'_R R \Pi_{\rho_\theta} D_{T,\rho_\theta} \quad \text{and note that,} \quad R^*\mu_{T,\theta} \xrightarrow{P} \mu_\beta.$$

**Definition:** Define the full column-rank matrix $G^*$ [Appendix A.2.3 contains the details of its construction in (21)] as:

$$G^* := \lim_{T\to\infty} \frac{\Lambda_T}{\sqrt{T}}\rho_\theta(\theta^0)\Pi_{\rho_\theta} D_{T,\rho_\theta}.$$

**Remark 11:** $R^*$ and $G^*$ correspond to $R$ and $G(\theta^0)$ respectively in the classical NM-9.2 setup [see Remark 5]. However, $R$ and $R^*$ can be very different otherwise [see Remark 16 for an example].

**Remark 12:** Finally, note the modifications to these definitions that would also make the results in the

[10] A sketch demonstrating this tradeoff is given in Appendix C.2 by referencing to Lemma 11. Also see footnote 13.

[11] The somewhat nonstandard nature of the deviations in (14) is not an artifact of $d_R < d_\theta$, but of rate-entanglement and since we want the results to resemble NM-9.2. For example, if instead one wishes to test $\theta^0 = \theta_0$, i.e., $d_R = d_\theta$, then under the setup of Section 4.1 the proper local deviation should be $\lim_T \text{diag}(1/\delta_{T,1}, \dots, 1/\delta_{T,p})B'_T(\theta_0 - \theta^0) = \mu_\theta$ for a fixed $\mu_\theta$; while under the setup here in Section 4.2, this should resemble (15). Both are still nonstandard.

sequel applicable to the setup of Stock and Wright (2000). To distinguish between the setups, use an "upper bar" for the corresponding quantities now. Resembling N3, define $\bar{\Lambda}_T := \text{diag}(\bar{\lambda}_{T,1}1_{\bar{k}_1}, \ldots, \bar{\lambda}_{T,l}1_{\bar{k}_l})$ where $\bar{k}_j > 0$ for $j = 1, \ldots, l$, $\sum_{j=1}^{l} \bar{k}_j = d_\theta$, $\bar{\lambda}_{T,j}/\bar{\lambda}_{T,j+1} = o(1)$ for $j = 1, \ldots, l-1$; $\lim_T \lambda_{T,1} = \infty$ and $\lim_T \bar{\lambda}_{T,l}/\sqrt{T} < \infty$. Suppose that: $E_T[\bar{G}_T(\theta)] = \bar{\rho}_{T,\theta}(\theta)\frac{\bar{\Lambda}_T}{\sqrt{T}}$ where $\bar{\rho}_{T,\theta}(\theta)$ is some $d_g \times d_\theta$ function with limit $\bar{\rho}_\theta(\theta)$ uniformly in $\theta \in \Theta$. The crucial difference here is the location of $\bar{\Lambda}_T$, as opposed to $\Lambda_T$ earlier, in the matrix-multiplication [c.f. (13)]. Let $\theta = (\theta_1, \ldots, \theta_{d_\theta})'$. If analogous assumptions in N are now maintained in terms of $\bar{\rho}_\theta(\theta)$ and $\bar{\Lambda}_T$, then $\bar{\lambda}_{T,j}$ would be the rate of identification of $(\theta^0_{\bar{k}_{j-1}+1}, \ldots, \theta^0_{\bar{k}_j})'$ for $j = 1, \ldots, l$.[12] Hence, the constructions in (14) and (15) can proceed with the corresponding $\Pi_{\rho_\theta}, D_{T,\rho_\theta}, \Pi_R, D_{T,R}$ and $R^*$, call them $\bar{\Pi}_{\bar{\rho}_\theta}, \bar{D}_{T,\bar{\rho}_\theta}, \bar{\Pi}_R, \bar{D}_{T,R}$ and $\bar{R}^*$ now, with $\bar{\Pi}_{\bar{\rho}_\theta} = I_{d_\theta}$ and $\bar{D}_{T,\bar{\rho}_\theta} = \sqrt{T}\bar{\Lambda}_T^{-1}$ giving $\bar{\Pi}_R$, $\bar{D}_{T,R}$ and $\bar{R}^*$ following Appendices A.3.1-A.3.3. What follows next holds under a similar or weaker versions of assumption N; e.g., N7(b) is not needed anymore as $\bar{\Pi}_{\bar{\rho}_\theta} = I_{d_\theta}$.

### 4.2.1 Asymptotic results:

**Lemma 5** *Let assumptions O and N hold. Let $r_0$ in (14) be such that $\theta^{infs}_{0,S} := A_S^{-1}(r_0', \gamma_S^{0\prime})' \equiv R_S^1 r_0 + S_S^1\gamma_S^0$ satisfies (15). Consider any sequence $\{\theta_T = A_S^{-1}(r_0', \gamma_{S,T}')' : T \geq 1\}$ where $\{\gamma_{S,T} : T \geq 1\}$ is such that (15) holds. Then, the following results hold for $LM_T(\theta_T)$ defined in (4), as $T \to \infty$:*

*(a) $LM_T(\theta_T) = LM_T^{infs}(\theta^{infs}_{0,S}) + o_p(1)$ where $LM_T^{infs}(\theta^{infs}_{0,S})$ is as defined in (8).*

*(b) $LM_T(\theta_T) \xrightarrow{d} \chi^2_{d_R}$ with non-centrality parameter $\mu_\beta' \left(R^*(G^{*\prime}V^{-1}G^*)^{-1}R^{*\prime}\right)^{-1}\mu_\beta$.*

**Lemma 6** *Let assumptions O and N hold. Let $r_0$ in (14) be such that $\theta^{infs}_{0,S} := A_S^{-1}(r_0', \gamma_S^{0\prime})' \equiv R_S^1 r_0 + S_S^1\gamma_S^0$ satisfies (15). For the given choice of S, and some $\epsilon, \alpha > 0$ such that $\epsilon + \alpha < 1$, let $CI_T(\gamma_S; \epsilon)$ be a confidence set for $\gamma_S^0$, the true value of $\gamma_S$, such that:*

$$\sup_{\gamma_0 \in CI_T(\gamma_S;\epsilon)} \sqrt{T}\left\|D^{-1}_{T,\rho_\theta}\Pi'_{\rho_\theta}\left((R_S^1(r_0 - \beta^0) + S_S^1(\gamma_0 - \gamma_S^0)\right)\right\| = O_p(1) \tag{16}$$

*where $\Pi_{\rho_\theta}$ and $D_{T,\rho_\theta}$ are defined in (18) and (19) in Appendices A.2.1 and A.2.2 respectively. Then,*

$$\inf_{\gamma_0 \in CI_T(\gamma_S;\epsilon)} LM_T\left(A_S^{-1}(r_0', \gamma_0')'\right) = LM_T^{infs}(\theta^{infs}_{0,S}) + o_p(1)$$

*where $LM_T(\theta)$ and $LM_T^{infs}(\theta^{infs}_{0,S})$ is as defined in (4) and (8) respectively.*

[12] A direct generalization of Chaudhuri and Zivot (2011) adhering to Stock and Wright (2000)'s setup would lead to the above structure giving rate-disentangled $\theta^0$. As opposed to (13), here one could, for example, model $E_T[\bar{g}_T(\theta)]$ as [c.f. (13)]:

$$E_T[\bar{g}_T(\theta)] = \sum_{j=1}^{l} \frac{\bar{\lambda}_{T,j}}{\sqrt{T}}\left(\bar{\rho}_T^{(j)}(\theta_{\bar{k}_{j-1}+1}, \ldots, \theta_{d_\theta}) - \bar{\rho}_T^{(j)}(\theta^0_{\bar{k}_{j-1}+1}, \ldots, \theta^0_{d_\theta})\right)$$

where $\theta = (\theta_1, \ldots, \theta_{d_\theta})'$, $\bar{k}_0 = 0$, $\bar{k}_j > 0$ for $j = 1, \ldots, l$, $\sum_{j=1}^{l} \bar{k}_j = d_\theta$, and, for $j = 1, \ldots, l$, the functions $\bar{\rho}_T^{(j)}(.)$ are $d_g \times 1$ deterministic functions satisfying the restrictions in Section 3.3 of Stock and Wright (2000), and other conditions as needed.

**Remark 13:** Lemmas 5 and 6 formalize the discussion of asymptotic equivalence from Remark 5. As we will see in Lemma 7 below, the asymptotic behavior of the two-step test *itself* does not require imposing the condition (15) as the first step based on $CI_T^{SW}(\gamma_S; r_0, \epsilon)$ in (10) automatically incorporates (15) by virtue of (16). This condition will, however, be needed to completely characterize the local deviation of $H_0$ for which the asymptotic equivalence with the corresponding ($S$-dependent) infeasible test holds. This is an important observation that fundamentally distinguishes our paper from the related literature on efficiency. Sections 4.2.2 and 4.2.3, respectively, analytically and visually illustrate this observation.

**Remark 14:** Lemma 5(a) establishes the key property of $LM_T(\theta)$ that its asymptotic behavior is invariant to certain local deviations of the nuisance parameter $\gamma_S$ from $\gamma_S^0$. This motivated Chaudhuri and Zivot (2011)'s use of LM C-alpha in a projection test. Lemma 5(b) describes the asymptotic distribution of $LM_T(\theta)$, which closely resembles that from the NM-9.2 classical setup; the latter is a special case. Lemma 6 utilizes Lemma 5(a) to establish that the second-step test statistic is asymptotically equivalent to the infeasible statistic under the condition (16) imposed on the first-step confidence set. Then, the asymptotic rejection rate of the improved two-step projection test in (6) follows from Lemma 5(b).

**Remark 15:** Condition (16) is important for benefitting from the use of (LM) C-alpha, and this is what we ideally expect the first-step confidence set to satisfy. Given the local deviation in (14), it characterizes the worst possible convergence of the first-step confidence set to $\gamma_S^0$ that still allows appealing to (15), which is necessary for discussing the local asymptotic properties of $LM_T(\theta)$. In the case where $\Lambda_T = \lambda_T I_{d_g}$ for some $\lambda_T \to \infty$ (but $\lim_T \lambda_T/\sqrt{T} < \infty$), i.e., all the rates are equal, and if our interest lies in testing subvectors of $\theta$ (e.g., $R = [I_{d_R}, 0]$), then, by virtue of (14), the condition in (16) boils down to $\sup_{\gamma_0 \in CI_T(\gamma_S;\epsilon)} \lambda_T \left\|\gamma_0 - \gamma_S^0\right\| = O_p(1)$, which becomes $\sup_{\gamma_0 \in CI_T(\gamma_S;\epsilon)} \sqrt{T} \left\|\gamma_0 - \gamma_S^0\right\| = O_p(1)$ if, additionally, we focus only on the strong identification classical setup as done in the related literature.

It is, however, clear that our recommended confidence set $CI_T^{SW}(\gamma_S; r_0, \epsilon)$ from Section 4.1 cannot satisfy (16) since $CI_T^{SW}(\gamma_S; r_0, \epsilon)$ can be empty with positive probability. Nevertheless, as noted earlier in footnote 4, we are of the opinion that as long as it satisfies the requirement of Proposition 4, the practical benefit of an empty $CI_T^{SW}(\gamma_S; r_0, \epsilon)$ (with a small $\epsilon$) in: (i) eliminating spurious declines in power and (ii) easing computation, may be worth the cost associated with it. With the caveat of emptiness, which we sidestep by redefining the supremum in (16), we now show that $CI_T^{SW}(\gamma_S; r_0, \epsilon)$ would satisfy (16).

**Lemma 7** *Let assumptions O and N hold. Let $r_0$ satisfy* (14). *Define the supremum in* (16) *as zero if $CI_T(\gamma_S, \epsilon)$ is empty for a given $\epsilon > 0$. Then, $CI_T(\gamma_S, \epsilon) = CI_T^{SW}(\gamma_S; r_0, \epsilon)$ satisfies* (16) *for $\epsilon > 0$, i.e.,*

$$\sup_{\gamma_0 \in CI_T^{SW}(\gamma_S; r_0, \epsilon)} \sqrt{T} \left\| D_{T,\rho_\theta}^{-1} \Pi'_{\rho_\theta} \left( (R_S^1(r_0 - \beta^0) + S_S^1(\gamma_0 - \gamma_S^0) \right) \right\| = O_p(1).$$

For completeness, we summarize Lemmas 6 and 7 to present our final result in Proposition 8.

**Proposition 8** *Let assumptions O and N hold. Let the sequence of hypothesized value $r_0$ in* (2) *locally deviates from the truth $\beta^0 = R\theta^0$ following* (14)*, and such that $\theta_{0,S}^{infs} := A_S^{-1}(r_0', \gamma_S^{0\prime})' \equiv R_S^1 r_0 + S_S^1 \gamma_S^0$ satisfies* (15)*. Then, for $\epsilon, \alpha > 0$ such that $\epsilon + \alpha < 1$, the asymptotic probability of rejection of $r_0$ by the improved two-step projection test in* (6)*, based on the choice $CI_T^{SW}(\gamma_S; r_0, \epsilon)$ in* (10)*, cannot be smaller than that by the infeasible test in* (8)*.*

#### 4.2.2 Discussion of the asymptotic results with an example (reference: Remark 13):

Consider the example from Section 2 (F2) where $d_\theta = 2$, $\theta = (\theta_1, \theta_2)'$ and $R = [1, 1]$. Consider the two choices of $S$: $S^* = [1, 0]$ and $S^\dagger = [0, 1]$, giving $\gamma_{S^*} = \theta_1$ and $\gamma_{S^\dagger} = \theta_2$ respectively. We focus on what our local results say about the asymptotic equivalence with the infeasible statistic, in this case for the two choices $S^*$ and $S^\dagger$ respectively, and how, through (15), that restricts the local deviation in (14).

Complying with both Antoine and Renault (2012) and Stock and Wright (2000), let, for example,

$$E_T[\bar{G}_T(\theta^0)] = \begin{bmatrix} \frac{\lambda_{T,1}}{\sqrt{T}}\rho_{11} & 0 \\ 0 & \frac{\lambda_{T,2}}{\sqrt{T}}\rho_{22} \end{bmatrix}, \text{ and define } \rho_1 := (\rho_{11}', 0')' \neq 0 \text{ and } \rho_2 := (0', \rho_{22}')' \neq 0. \qquad (17)$$

Therefore, $\lambda_{T,1} = \bar{\lambda}_{T,1}$, $\lambda_{T,2} = \bar{\lambda}_{T,2}$, $\Pi_{\rho_\theta} = \bar{\Pi}_{\bar{\rho}_\theta} = I_2$, $D_{T,\rho_\theta} = \bar{D}_{T,\bar{\rho}_\theta} = \sqrt{T}\text{diag}(\lambda_{T,1}^{-1}, \lambda_{T,2}^{-1})$, $\Pi_R = \bar{\Pi}_R = 1$ and $D_{T,R} = \bar{D}_{T,R} = \lambda_{T,1}/\sqrt{T}$. This is a setup where identification of $\theta^0$, but not $R\theta^0$, is rate-disentangled.
**Remark 16:** This setup gives $R^* = \bar{R}^* = \lim_T[1, \lambda_{T,1}/\lambda_{T,2}]$. Unless $\lambda_{T,1} = \lambda_{T,2}$, we have $R^* = \bar{R}^* = [1, 0] \neq R = [1, 1]$ [see Remark 11]. On the other hand, here $G^* = [\rho_1, \rho_2]$ (block-diagonal). Therefore, defining $m_{ij} := \rho_i' V^{-1} \rho_j$ for $j = 1, 2$, this implies that the non-centrality parameter in Lemma 5(b) is $\mu_\beta^2/(m_{11} - m_{12}^2/m_{22})$, where the denominator is what one should obtain in a subvector test for $\theta_1^0$, the element with less strong identification [c.f. Proposition 2.2, Antoine and Renault (2009)].
**Remark 17:** In this example, $\mu_\beta = \lambda_{T,1}(r_0 - \beta^0)$ by (14). However, (15) gives: $\lambda_{T,1}(\theta_1 - \theta_1^0) + \lambda_{T,2}(\theta_2 - \theta_2^0) = \mu_{T,\theta}$, which implies that $\lambda_{T,1}(r_0 - \beta^0) + (\lambda_{T,2} - \lambda_{T,1})(\theta_2 - \theta_2^0) = \mu_{T,\theta}$. Hence, while (14) captures deviations of order up to $\lambda_{T,1}^{-1}$ for $\beta := \theta_1 + \theta_2$, any deviation of order bigger than $\lambda_{T,2}^{-1}$ along the $\theta_2$-axis, that causes this, is beyond the scope of our results. (On the other hand, everything is standard along the $\theta_1$-axis.) This is a consequence of allowing for multiple rates that make $R\theta^0 := \theta_1^0 + \theta_2^0$ rate-entangled.
**Remark 18:** Remark 17 is not relevant for the asymptotic behavior of the two-step test itself, but is important for its asymptotic equivalence with the infeasible tests. Note that, for the two choices $S = S^*, S^\dagger$, we have $\theta_{0,S}^{infs}$ as $\theta_{0,S^*}^{infs} = (\gamma_{S^*}^0 := \theta_1^0, r_0 - \gamma_{S^*}^0)'$ and $\theta_{0,S^\dagger}^{infs} = (r_0 - \gamma_{S^\dagger}^0, \gamma_{S^\dagger}^0 := \theta_2^0)'$ respectively. Consider $r_0$ satisfying (14), i.e., $(r_0 - \beta^0) = \mu_\beta/\lambda_{T,1}$ for some constant $\mu_\beta \neq 0$, which, for $\theta_{0,S^*}^{infs}$ and $\theta_{0,S^\dagger}^{infs}$

respectively, means that $(\theta_{T,2} - \theta_2^0) = \mu_\beta/\lambda_{T,1}$ and $(\theta_{T,1} - \theta_1^0) = \mu_\beta/\lambda_{T,1}$. Hence, from Remark 17 we know that for a constant $\mu_\beta \neq 0$ in (14), $\theta_{0,S^\dagger}^{infs}$ falls under the scope of our results, while $\theta_{0,S^*}^{infs}$ does not. For the latter, we need $\mu_\beta = O(\lambda_{T,1}/\lambda_{T,2})$, a deviation more local than what is implied by (14) alone.

**Remark 19:** This is the role of (15). It refines the deviation in (14) to precisely tell where the asymptotic equivalence holds. Interestingly, since the two-step test (and the plug-in tests) are generally invariant to $S$, this means that irrespective of what it uses for $S$, the asymptotic equivalence holds in a larger region with the infeasible test with $S = S^\dagger$, i.e., the one using the better identified nuisance parameter $\gamma_S = \theta_2$. This nicely aligns with the practical implementation of the two-step projection test since it is generally computationally easiest to use the same, i.e., the most strongly identified, $\gamma_S$, if possible [see Remark 1].

#### 4.2.3 Simulation study with a linear instrumental variables model:

Let us now visually demonstrate the points made in Section 4.2.2 with the help of a linear instrumental variables model, the leading example of GMM. A linear (in $\theta$) $g(Z_t;\theta)$ gives the following simplifications.

**Remark 20:** The rate-restrictions $\lambda_{T,j}(\theta_{T,j} - \theta_j^0) = O_p(1)$ for $j = 1, 2$ in (15) can be weakened for the properly scaled terms inside the projection matrix $P(.)$ in $LM_T(\theta)$ in (4) to converge to the concerned limits. All we need are $(\theta_{T,1} - \theta_1^0) = o_p(1)$ and $\lambda_{T,2}(\theta_{T,2} - \theta_2^0) = o_p(\lambda_{T,1})$.[13] Hence, now, for deviations satisfying this condition instead of (15), our results give that *only* the infeasible test with $S = [1, 0]$, i.e., with $\gamma_S = \theta_1$, rejects $H_0$ with probability approaching one, except spuriously [c.f. Remarks 17 and 19].

**Simulation design:** Consider an i.i.d. sample $\{Z_t := (y_t, X_{1t}, X_{2t}, W_t')'\}_{t=1}^T$ where

$$\begin{aligned} \text{dependent variable: } y_t &= X_{1t}\theta_1^0 + X_2\theta_2^0 + u_t, \\ \text{endogenous regressors: } X_{jt} &= W_t'\pi_{jT} + v_{jt} \text{ for } j = 1, 2, \end{aligned}$$

while the instruments $W_t \sim N(0, I_4)$ are independent of the model errors $(u_t, v_{1t}, v_{2t})$; $u_t \sim N(0, 1)$, $v_{jt} \sim N(0, 1)$ with $Cov(u_t, v_{jt}) = .8$ and $Cov(v_{1t}, v_{2t}) = .3$ for $j = 1, 2$. The moment vector is $g(Z_t;\theta) = W_t(y_t - X_{1t}\theta_1 - X_2\theta_2)$ is $4 \times 1$ dimensional. Take $\pi_{jT}$, for $j = 1, 2$, such that (17) holds with:

- $\rho_1 = \sqrt{4/3}(1, 1, 0, 0)'$ if $\lambda_{T,1} \neq \sqrt{T}$, and $\rho_1 = 20(1, 1, 0, 0)'$ if $\lambda_{T,1} = \sqrt{T}$,
- $\rho_2 = \sqrt{4/3}(0, 0, 1, 1)'$ if $\lambda_{T,2} \neq \sqrt{T}$, and $\rho_2 = 20(0, 0, 1, 1)'$ if $\lambda_{T,2} = \sqrt{T}$.

This design is exactly the same as that in Chaudhuri and Zivot (2011) except that the structure of $[\pi_{1T}, \pi_{2T}]$ now conforms to the setups of both Stock and Wright (2000) and Antoine and Renault (2012).

[13] Let $\theta_T \xrightarrow{P} \theta^0$. Then, Lemma 11 (a), (b) and (c) hold [see Appendix C.1]. On the other hand, for $j = 1, 2$, use N2 to obtain that $\frac{\sqrt{T}}{\lambda_{T,j}}\bar{g}_T(\theta_T) = O_p\left(\frac{1}{\lambda_{T,j}}\right) + \frac{\lambda_{T,1}}{\lambda_{T,j}}\rho_1(\theta_{T,1} - \theta_1^0) + \frac{\lambda_{T,2}}{\lambda_{T,j}}\rho_2(\theta_{T,2} - \theta_2^0)$, which is $o_p(1)$ if, additionally, $\frac{\lambda_{T,2}}{\lambda_{T,j}}(\theta_{T,2} - \theta_2^0) = o_p(1)$, a less stringent rate restriction than (15). It is also worth noting here that since identification is better than weak, i.e., $\lambda_{T,1} \to \infty$, when taken in conjunction with Lemma 7, this discussion explains the part in Remark 9 that led to footnote 10.

We take $T = 100$ and consider six specifications for the identification strength of the elements of $\theta^0$: (i) $\lambda_{T,1} = \lambda_{T,2} = 1$, (ii) $\lambda_{T,1} = 1, \lambda_{T,2} = T^{1/6}$, (iii) $\lambda_{T,1} = 1, \lambda_{T,2} = \sqrt{T}$, (iv) $\lambda_{T,1} = T^{1/6}, \lambda_{T,2} = T^{1/6}$, (v) $\lambda_{T,1} = T^{1/6}, \lambda_{T,2} = \sqrt{T}$, and (vi) $\lambda_{T,1} = \sqrt{T}, \lambda_{T,2} = \sqrt{T}$. (i)-(iii) are out of the scope of Section 4.2. However, as noted in Remark 20, (iv)-(v) are covered, in spite of violating N7(b), especially since $g(Z_t; \theta)$ is linear in $\theta$. ((i)-(vi) are under the scope of Section 4.1 if interest lies is in the empirical size.)

We take $R = [1, 1]$, and for $S$ we take $S^* = [1, 0], S^\dagger = [0, 1]$ giving $\gamma_S = \theta_1$ and $\gamma_S = \theta_2$ respectively. For the simulation study we consider six different tests, three for each choice of $S$ ($\gamma_S$):

- The infeasible test in (8).
- The standard plug-in test that rejects $H_0$ if $LM_T(r_0, \widehat{\gamma_S}(r_0)) > \chi^2_{d_R}(1-\alpha)$ where $\widehat{\gamma_S}(r_0)$ is the restricted-by-$H_0$ CU-GMM estimator of $\gamma_S$, i.e., $\widehat{\gamma_S}(r_0) := \arg\min_{\gamma \in \Gamma_S} Q_T(A_S^{-1}(r_0', \gamma')')$ with $Q_T(.)$ as in (11). Here, $\widehat{\gamma_S}(r_0)$ is the restricted limited information maximum likelihood (LIML) estimator.
- The two-step projection test in (6) with $CI_T^{SW}(\gamma_S; r_0, \epsilon)$ in (10) as the first-step confidence set.

**Simulation results:** Figure 1 plots, under specifications (i)-(vi), the empirical rejection probability of these tests with $S = S^*, S^\dagger$. Since the plug-in test (in this case) and the two-step projection test are both invariant to the choice of $S$, only one plot for each test is reported.[14] (Note that the plug-in test is not invariant if $H_0$ is not imposed while estimating $\gamma_S$. This has serious adverse consequences on size, and is studied in Appendix D.) Results under different variations of this specification are similar.

It must be acknowledged that while considerable care — e.g., at least seven different starting values, far more stringent than default conditions for optimization in Matlab, cross-checking with both and other choices of $S$, etc. — is taken in the implementation of the two-step test, it is technically possible that the second step minimization is not accurate and hence the rejection probabilities are overestimated.[15]

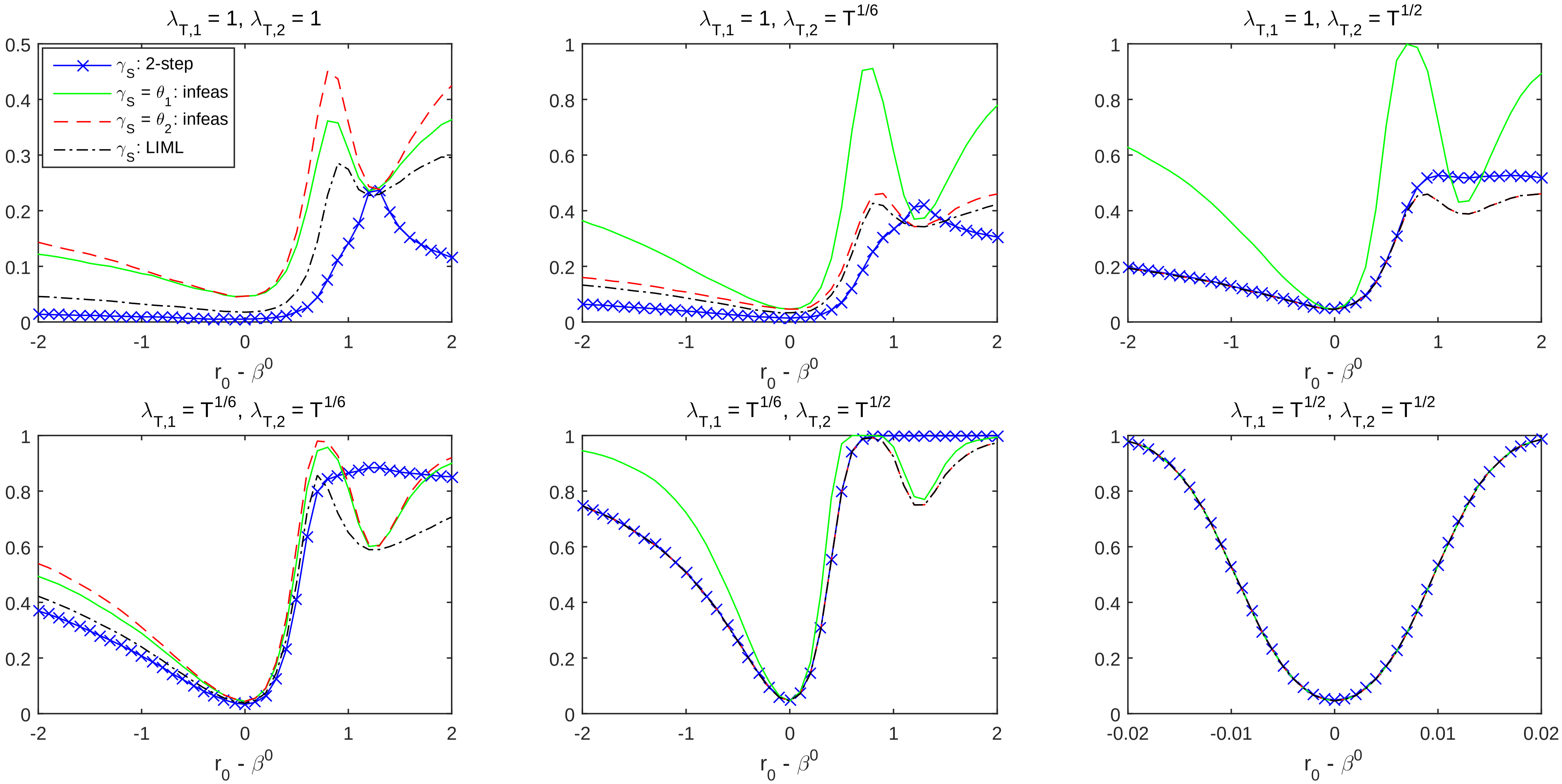
We find that the simulation results under (v) and (vi) and, to a lesser extent, under (iv) corroborate our local asymptotic results even for $T = 100$. They are also encouraging for the two-step test under (ii) and (iii) that are actually outside our scope. When $\lambda_{T,1} \neq \lambda_{T,2}$, the two infeasible tests behave differently as $H_0$ deviates a bit far from the truth, and we find that the two-step and the plug-in tests, which are invariant to $S$, resemble the infeasible test using $\gamma_S = \theta_2$, the better identified component of $\theta^0$.[16] This confirms a key aspect of the local efficiency result that we sought to highlight in our paper.

[14] The invariance to $S$ for the two-step in (6) follows once we note (a) and (b) below. (a) For any $S$, the optimization problem in (6) is: $\min_{\gamma_S} LM_T(A_S^{-1}(r_0', \gamma_s')')$ such that $h_T(A_S^{-1}(r_0', \gamma_s')') \leq c$ where the inequality constraint represents the first step of the two-step test, with $h_T(.)$ denoting the first-step test statistic and $c$ the critical value [see, e.g., (10)]. (b) The minimization problem in (a) is the same as: $\min_\theta LM_T(\theta)$ such that $h_T(\theta) \leq c$ and $R\theta = r_0$, where the last equality relates to (2) and (3). The problem in (b) does not depend on $S$.

[15] This is more likely for the results in Appendix D, but not problematic in that context since the overestimation actually reinforces the point about the poor power of the unrestricted-by-$H_0$ version of the test that we wish to make there.

[16] More extensive simulations (1 million trials and grid size .001) not reported here suggest that when $\lambda_{T,1} \neq \lambda_{T,2}$, the two infeasible tests behave similarly roughly in the interval $[-.1, .1]$ around $r_0 - \beta^0$. Their difference under (i) and (iv), where $\lambda_{T,1} = \lambda_{T,2}$, is due to small sample size; the difference vanishes if, for example, $T = 2000$ for (i) and $T = 250$ for (iv).

Figure 1: Empirical rejection probabilities of the two-step projection test (2-step) in (6) with $\epsilon = .005, \alpha = .045$, the infeasible test (infeas) in (8) with $\alpha = .045$, and the standard plug-in test (LIML) based on the restricted-by-$H_0$ CU-GMM (LIML) estimator for $\gamma_S$, with $\alpha = .045$. Two choices of $\gamma_S$, i.e., $\gamma_S = \theta_1$ and $\gamma_S = \theta_2$ are employed for the infeasible test. The other tests are invariant to $S$. Results are based on 10,000 Monte Carlo trials. Horizontal axis: deviation of $H_0$ in (2) from the truth [see (14)]. Title: Identification strength that corresponds to specifications (i)-(vi) respectively.



## References


Andrews, D. W. K. (2017a). Identification-Robust Subvector Inference. Yale University Working Paper.

Andrews, D. W. K. (2017b). Identification-Robust Subvector Inference for Likelihood-Based Models. Yale University Working Paper.

Andrews, D. W. K. and Cheng, X. (2012). Estimation and inference with weak, semi-strong, and strong identification. *Econometrica*, 80: 2153–2211.

Andrews, D. W. K. and Cheng, X. (2014). Gmm Estimation and Uniform Subvector Inference with Possible Identification Failure. *Econometric Theory*, 20: 287–333.

Andrews, D. W. K. and Guggenberger, P. (2014). Asymptotic Size of Kleibergen's LM and Conditional LR Tests for Moment Condition Models. Working Paper.

Andrews, D. W. K. and Guggenberger, P. (2015). Identification And Singularity Robust Inference For Moment Condition Models. Working Paper.

Andrews, D. W. K., Moreira, M., and Stock, J. H. (2006). Optimal Invariant Similar Tests for Instrumental Variables Regression. *Econometrica*, 74: 715–752.

Andrews, I. (2016a). Conditional Linear Combination Tests. Forthcoming: Econometrica.

Andrews, I. (2016b). Robust Two-Step Confidence Sets, and the Trouble with the First Stage F-Statistic. Working Paper.

Andrews, I. and Mikusheva, A. (2015). Maximum Likelihood Inference in Weakly Identified DSGE Models. *Quantitative Economics*, 6: 123–152.

Andrews, I. and Mikusheva, A. (2016). Conditional Inference with a Functional Nuisance Parameter. Forthcoming: Econometrica.

Antoine, B. and Renault, E. (2009). Efficient GMM with nearly-weak instruments. *Econometrics Journal*, 12: S135–S171.

Antoine, B. and Renault, E. (2012). Efficient minimum distance estimation with multiple rates of convergence. *Journal of Ecnometrics*, 170: 350–367.

Berger, R. L. and Boos, D. D. (1994). P-Values Maximized Over a Confidence Set for the Nuisance Parameter. *Journal of the American Statistical Association*, 89: 1012–1016.

Chaudhuri, S. (2008). *Projection-Type Score Tests for Subsets of Paramaters*. PhD thesis, University of Washington.

Chaudhuri, S. and Renault, E. (2011). Finite-sample improvements of score tests by the use of implied probabilities from Generalized Empirical Likelihood. Technical report, UNC, Chapel Hill.

Chaudhuri, S., Richardson, T., Robins, J., and Zivot, E. (2010). Split-Sample Score Tests in Linear Instrumental Variables Regression. *Econometric Theory*, 26: 1820–1837.

Chaudhuri, S. and Zivot, E. (2011). A new method of projection-based inference in GMM with weakly identified nuisance parameters. *Journal of Econometrics*, 164: 239–251.

Cheng, X. (2015). Robust inference in nonlinear models with mixed identification strength. *Journal of Ecnometrics*, 189: 207–228.

Choi, I. and Phillips, P. C. B. (1992). Asymptotic and Finite Sample Distribution Theory for IV Estimators and Tests in Partially Identified Structural Equations. *Journal of Econometrics*, 51: 113–150.

Dagenais, M. and Dufour, J. M. (1991). Invariance, Nonlinear Models and Asymptotic Tests. *Econometrica*, 59: 1601–1615.

Davidson, R. and MacKinnon, J. (2014). Confidence sets based on inverting anderson-rubin tests. *Econometrics Journal*, 17: S39–S58.

Dufour, J. M. (1990). Exact Tests and Confidence Sets in Linear Regressions with Autocorrelated Errors. *Econometrica*, 58: 475–494.

Dufour, J. M. (1997). Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models. *Econometrica*, 65: 1365–1388.

Dufour, J. M. and Jasiak, J. (2001). Finite Sample Limited Information Inference Methods for Structural Equations and Models with Generated Regressors. *International Economic Review*, 42: 815–843.

Dufour, J.-M., Khalaf, L., and Kichian, M. ( 2006). Inflation dynamics and the New Keynesian Phillips Curve: An identification robust econometric analysis. *Journal of Economic Dynamics and Control*, 30: 1707–1727.

Dufour, J. M. and Taamouti, M. (2005). Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments. *Econometrica*, 73: 1351–1365.

Dufour, J. M. and Taamouti, M. (2007). Further Results on Projection-Based Inference in IV Regressions with Weak, Collinear or Missing Instruments. *Journal of Econometrics*, 139: 133–153.

Dufour, J.-M., Trognon, A., and Tuvaandorj, P. (2015). Invariant tests based on M-estimators, estimating functions, and the generalized method of moments. Forthcoming in Econometric Reviews.

Guggenberger, P., Kleibergen, F., Mavroeidis, S., and Chen, L. (2012a). On the Asymptotic Sizes of Subset AndersonRubin and Lagrange Multiplier Tests in Linear Instrumental Variables Regression. *Economoetrica*, 80: 2649–2666.

Guggenberger, P., Ramalho, J., and Smith, R. (2012b). GEL Statistics under Weak Identification. *Journal of Econometrics*, 170: 331–349.

Guggenberger, P. and Smith, R. (2005). Generalized Empirical Likelihood Estimators and Tests under Partial, Weak and Strong Identification. *Econometric Theory*, 21: 667–709.

Guggenberger, P. and Smith, R. (2008). Generalized empirical likelihood tests in time series models with potential identification failure. *Journal of Econometrics*, 142: 134–161.

Kleibergen, F. (2002). Pivotal Statistics for Testing Structural Parameters in Instrumental Variables Regression. *Econometrica*, 70: 1781–1803.

Kleibergen, F. (2005). Testing Parameters In GMM Without Assuming That They Are Identified. *Econometrica*, 73: 1103–1123.

Magnusson, L. and Mavroeidis, S. ( 2010). Identification-Robust Minimum Distance Estimation of the New Keynesian Phillips Curve. *Journal of Money, Credit and Banking*, 42: 465–481.

McCloskey, A. (2015). Bonferroni-Based Size-Correction for Nonstandard Testing Problems. Brown University.

Mikusheva, A. (2010). Robust Confidence Sets in the Presence of Weak Instruments. *Journal of Econometrics*, 157: 236–247.

Montiel-Olea, J. L. (2016). Admissible, Similar Tests: A Characterization. New York University.

Moreira, H. and Moreira, M. (2013). Contributions to the Theory of Similar Tests. FGV/EPGE.

Moreira, M. J. (2002). *Tests with Correct Size in the Simultaneous Equations Model.* PhD thesis, UC Berkeley.

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

Muller, U. and Norets, A. (2016). Credibility of Confidence Sets in Nonstandard Econometric Problems. *Econometrica*, 84: 2183–2213.

Newey, W. K. and McFadden, D. L. (1994). Large Sample Estimation and Hypothesis Testing. In Engle, R. F. and McFadden, D., editors, *Handbook of Econometrics*, volume IV, chapter 36, pages 2212–2245. Elsevier Science Publisher.

Neyman, J. (1959). Optimal Asymptotic Test of Composite Statistical Hypothesis. In Grenander, U., editor, *Probability and Statistics, the Harald Cramer Volume*, pages 313–334. Almqvist and Wiksell, Uppsala.

Phillips, P. C. B. (1989). Partially Identified Econometric Models. *Econometric Theory*, 5: 181–240.

Qu, Z. (2014). Inference in DSGE Models with Possible Weak Identification. *Quantitative Economics*, 5: 457–494.

Robins, J. M. (2004). Optimal Structural Nested Models for Optimal Sequential Decisions. In Lin, D. Y. and Heagerty, P., editors, *Proceedings of the Second Seattle Symposium on Biostatistics*. NY: Springer.

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

Silvapulle, M. J. (1996). A Test in the Presence of Nuisance Parameters. *Journal of the American Statistical Association*, 91: 1690–1693.

Smith, R. J. (1987). Alternative Asymptotically Optimal Tests and Their Application to Dynamic Specification. *The Review of Economic Studies*, 54: 665–680.

Stock, J. H. and Wright, J. H. (2000). GMM with Weak Identification. *Econometrica*, 68: 1055–1096.

Wang, J. and Zivot, E. (1998). Inference on a Structural Parameter in Instrumental Variables Regression with Weak Instruments. *Econometrica*, 66: 1389–1404.

Zivot, E. and Chaudhuri, S. (2009). Comment: Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve by F. Kleibergen and S. Mavroeidis. *Journal of Business and Economic Statistics*, 27: 328–331.

Zivot, E., Startz, R., and Nelson, C. (2006). Inference in Weakly Identified Instrumental Variables Regression. In Corbae, D., Durlauf, S. N., and Hansen, B. E., editors, *Frontiers in Analysis and Applied Research: Essays in Honor of Peter C. B. Phillips*, pages 125–166. Cambridge U. Press.

# Appendix A: Important constructions and definitions for Section 4

## A.1 UBT and LBT Constructions:

We extensively use the following constructions that are adapted from the original work of Antoine and Renault (2012), Andrews and Cheng (2014), Cheng (2015), etc. Let $\{W_T = [W_{T,1}, \dots, W_{T,m_T}] : T \geq 1\}$ be a sequence of $r \times c$ (for some $r, c$) matrix of full row-rank $r(\leq c)$ where $W_{T,j}$ is $r \times c_{T,j}$ (and empty if $c_{T,j} = 0$) for $j = 1, \dots, m_T$ and such that $\sum_{j=1}^{m_T} c_{T,j} = c$ for each $T \geq 1$.

### A.1.1 UBT-Construction: An upper block-triangular (UBT) construction

We construct a sequence of $r \times r$ matrix $\{\Pi_T = [\Pi_{T,1}, \dots, \Pi_{T,m_T}] : T \geq 1\}$ such that the $c \times r$ matrix $W_T' \Pi_T$ has an UBT structure for each $T \geq 1$. For any given $T$, the following steps give such a $\Pi_T$.

- Let $\text{rank}(W_{T,m_T}) = c^*_{T,m_T} \leq \min(r, c_{m_T})$. Define $\Pi_{T,m_T}$ as the $r \times c^*_{T,m}$ matrix such that its columns form an orthogonal basis for the column space of $W'_{T,m_T}$. Stop if $m_T = 1$.
- Let $\text{rank}([W_{T,m_T-1}, W_{T,m_T}]) - \text{rank}(W_{T,m_T}) = c^*_{T,m_T-1} \leq \min(r, c_{m_T-1})$. Define $\Pi_{T,m_T-1}$ as the $r \times c^*_{T,m_T-1}$ matrix such that the columns of $[\Pi_{T,m_T-1}, \Pi_{T,m_T}]$ form an orthogonal basis for the column space of $[W_{T,m_T-1}, W_{T,m_T}]'$. Stop if $m_T = 2$.
- Continue step-by-step, as above, for $j = m_T - 2, \dots, 1$ and for each $j$, define $\Pi_{T,j}$ as the $r \times c^*_{T,j}$ matrix, where $c^*_{T,j} = \text{rank}([W_{T,j}, \dots, W_{T,m_T}]) - \text{rank}([W_{T,j+1}, \dots, W_{T,m_T}) \leq \min(r, c_{T,j})$, such that the columns of $[\Pi_{T,j}, \dots, \Pi_{T,m_T}]$ form an orthogonal basis for the column space of $[W_{T,j}, \dots, W_{T,m_T}]'$.

As a convention, $\Pi_{T,j}$ is an empty matrix if $c^*_{T,j} = 0$. $\Pi_T$ is an orthogonal matrix by construction and

(i) for some integer $q_T \in \{1, \dots, \min(r, m_T)\}$, the $q_T$ blocks $W'_{T,j_{k,T}} \Pi_{T,j_{k,T}}$ for $k = 1, \dots, q_T$, and where $1 \leq j_{1,T} < \dots < j_{q_T,T} \leq m_T$, each has full column-rank $c^*_{T,j_{k,T}} > 0$ satisfying $\sum_{k=1}^{q_T} c^*_{T,j_{k,T}} = r$;

(ii) $W'_{T,j} \Pi_{T,k} = 0$, a zero matrix of suitable (according to the above) dimension, for all $1 \leq k < j \leq m_T$.

### A.1.2 LBT-Construction: A lower block-triangular (LBT) construction

We construct a sequence of $r \times r$ matrix $\{\Pi_T = [\Pi_{T,1}, \dots, \Pi_{T,m_T}] : T \geq 1\}$ such that the $c \times r$ matrix $W_T' \Pi_T$ has a BLT structure for each $T \geq 1$. For any given $T$, the following steps give such a $\Pi_T$. (This is same as the UBT-Construction, but in reverse order. Hence to save new notation, we continue to use the same notation as in the UBT-Construction and hope that this is not confusing.)

- Let $\text{rank}(W_{T,1}) = c^*_{T,1} \leq \min(r, c_1)$. Define $\Pi_{T,1}$ as the $r \times c^*_{T,1}$ matrix such that its columns form an orthogonal basis for the column space of $W'_{T,1}$. Stop if $m_T = 1$.
- Let $\text{rank}([W_{T,1}, W_{T,2}]) - \text{rank}(W_{T,1}) = c^*_{T,2} \leq \min(r, c_2)$. Define $\Pi_{T,2}$ as the $c \times c^*_{T,2}$ matrix such that the columns of $[\Pi_{T,1}, \Pi_{T,2}]$ form an orthogonal basis for the column space of $[W_{T,1}, W_{T,2}]'$. Stop if $m_T = 2$.

- Continue step-by-step, as above, for $j = 3, \ldots, m_T$ and for each $j$, define $\Pi_{T,j}$ as the $r \times c^*_{T,j}$ matrix, where $c^*_{T,j} = \text{rank}([W_{T,1}, \ldots, W_{T,j}]) - \text{rank}([W_{T,1}, \ldots, W_{T,j-1}) \leq \min(r, c_{T,j})$, such that the columns of $[\Pi_{T,1}, \ldots, \Pi_{T,j}]$ form an orthogonal basis for the column space of $[W_{T,1}, \ldots, W_{T,j}]'$.

As a convention, $\Pi_{T,j}$ is an empty matrix if $c^*_{T,j} = 0$. $\Pi_T$ is an orthogonal matrix by construction and

(i) for some integer $q_T \in \{1, \ldots, \min(r, m_T)\}$, the $q_T$ blocks $W'_{T,j_{k,T}}\Pi_{T,j_{k,T}}$ for $k = 1, \ldots, q_T$, and where $1 \leq j_{1,T} < \ldots < j_{q_T,T} \leq m_T$, each has full column-rank $c^*_{T,j_{k,T}} > 0$ satisfying $\sum_{k=1}^{q_T} c^*_{T,j_{k,T}} = r$;

(ii) $W'_{T,j}\Pi_{T,k} = 0$, a zero matrix of suitable (according to the above) dimension, for all $1 \leq j < k \leq m_T$.

## A.2 Construction of $\Pi_{\rho_\theta}$, $D_{T,\rho_\theta}$ and $G^*$:

The efficient rate-disentangled directions of $\theta$ that are identified from (13) under our assumptions are given by $\Pi^{-1}_{\rho_\theta}\theta$ where $\Pi_{\rho_\theta}$ is a $d_\theta \times d_\theta$ orthogonal matrix, and the appropriate rates along these directions, in the given order, are given by the $d_\theta \times d_\theta$ diagonal matrix $\sqrt{T}D^{-1}_{T,\rho_\theta}$ [see Antoine and Renault (2012)].

### A.2.1 Construction of $\Pi_{\rho_\theta}$:

Let $\rho_\theta := \rho_\theta(\theta^0)$, i.e., $\partial\rho(\theta^0)/\partial\theta'$. Using N3 write $I^*\Lambda_T\rho_\theta \equiv I^*\Lambda_T I^{*'} I^*\rho_\theta = \left[\lambda_{T,1}\rho'_{\theta,1}, \ldots, \lambda_{T,l}\rho'_{\theta,l}\right]'$ where $\rho_{\theta,j}(\theta)$ is a $k_j \times d_\theta$ matrix for $j = 1, \ldots, l$. Take $W_T = \left[\rho'_{\theta,1}, \ldots, \rho'_{\theta,l}\right] = (I^*\rho_\theta)'$ (not depending on $T$) in the UBT-Construction in Appendix A.1.1. To emphasize the non-dependence on $T$, write $W_T$ as $W$, and accordingly write the rest of the notation from the UBT-Construction. Thus $r = d_\theta$, $c = d_g$ and $m = l$ in terms of the notation from the UBT-Construction. $W$ is full row-rank $r$ $(= d_\theta)$ by N4.

$\Pi_{\rho_\theta} = [\Pi_{\rho_\theta,1}, \ldots, \Pi_{\rho_\theta,l}]$ is the $d_\theta \times d_\theta$ matrix $\Pi$ from the UBT-Construction with $W = \left(I^*\rho_\theta(\theta^0)\right)'$. (18)

### A.2.2 Construction of $D_{T,\rho_\theta}$:

The construction of $D_{T,\rho_\theta}$ depends on the matrix $I^*\Lambda_T I^{*'} I^*\rho_\theta(\theta^0)\Pi_{\rho_\theta}$. Let $c^*_{\rho_\theta,j} = c^*_j \geq 0$ denote the number of columns of $\Pi_{\rho_\theta,j}$ for $j = 1, \ldots, l$, and $q_{\rho_\theta} = q$ from (i) in the UBT-Construction (of $\Pi_{\rho_\theta}$). Let $(j_1, \ldots, j_{q_{\rho_\theta}})$ denote the indices such that the block $\rho_{\theta,j_i}\Pi_{\rho_\theta,j_i}$ of dimension $k_{j_i} \times c^*_{\rho_\theta,j_i}$ is full column-rank $c^*_{\rho_\theta,j_i} > 0$ for $i = 1, \ldots, q_{\rho_\theta}$ and $\sum_{i=1}^{q_{\rho_\theta}} c^*_{\rho_\theta,j_i} = d_\theta$. Thus, the corresponding block of $I^*\Lambda_T I^{*'} I^*\rho_\theta(\theta^0)\Pi_{\rho_\theta}$ is $\lambda_{T,j_i}\rho_{\theta,j_i}\Pi_{\rho_\theta,j_i}$. Accordingly, for $I^*\Lambda_T I^{*'} I^*\rho_\theta(\theta^0)\Pi_{\rho_\theta}$, the columns from $(d_\theta - \sum_{i'=i}^{q_{\rho_\theta}} c^*_{\rho_\theta,j_{i'}})$ to $(d_\theta - \sum_{i'=i}^{q_{\rho_\theta}} c^*_{\rho_\theta,j_{i'}} + c^*_{\rho_\theta,j_i})$ for $i = 1, \ldots, q_{\rho_\theta}$ are represented by the $d_g \times c^*_{\rho_\theta,j_i}$ matrix:

$$\left[\lambda_{T,1}(\rho_{\theta,1}\Pi_{\rho_\theta,1})', 0'\right]' \text{ if } j_i = 1,$$

$$\left[\lambda_{T,1}(\rho_{\theta,1}\Pi_{\rho_\theta,j_i})', \ldots, \lambda_{T,j_i}(\rho_{\theta,j_i}\Pi_{\rho_\theta,j_i})', 0'\right]' \text{ otherwise.}$$

In both cases: $j_i = 1$ and $j_1 > 1$, the 0's inside the big matrices denote sub-matrices of zeros with number of rows, which can be zero, such that the number of rows of the corresponding big matrix is $d_g$.

Now, conforming to this above structure, define the $d_\theta \times d_\theta$ matrix $D_{T,\rho_\theta}$ as:

$$D_{T,\rho_\theta} := \sqrt{T}\mathrm{diag}\left(\lambda_{T,j_1}^{-1} 1_{c^*_{\rho_\theta,j_1}}, \ldots, \lambda_{T,j_{q_{\rho_\theta}}}^{-1} 1_{c^*_{\rho_\theta,j_{q_{\rho_\theta}}}}\right). \tag{19}$$

**A.2.3 Construction of $G^*$:**

Define the $d_g \times d_\theta$ matrix $G^\dagger$ as the following limit:

$$G^\dagger := \lim_{T\to\infty} \frac{1}{\sqrt{T}} I^* \Lambda_T I^{*'} I^* \rho_\theta(\theta^0)\Pi_{\rho_\theta} D_{T,\rho_\theta}. \tag{20}$$

By construction, $G^\dagger$ is finite, and its columns from $(d_\theta - \sum_{i'=i}^{q_{\rho_\theta}} c^*_{\rho_\theta,j_{i'}})$ to $(d_\theta - \sum_{i'=i}^{q_{\rho_\theta}} c^*_{\rho_\theta,j_{i'}} + c^*_{\rho_\theta,j_i})$ for $i = 1, \ldots, q_{\rho_\theta}$ are represented by the $d_g \times c^*_{\rho_\theta,j_i}$ matrix:

$$\begin{aligned}
&\left[(\rho_{\theta,1}\Pi_{\rho_\theta,1})', 0'\right]' \;\; j_i = 1, \\
&\left[0', (\rho_{\theta,j_i}\Pi_{\rho_\theta,j_i})', 0'\right]' \;\; \text{otherwise.}
\end{aligned}$$

(As above, 0 denotes sub-matrices of zeros with number of rows, which can be zero, such that the number of rows of the corresponding matrix is $d_\theta$.) Naturally, under our assumptions $G^\dagger$ is full column-rank.

Now define the $d_g \times d_\theta$ finite matrix of full column-rank $G^*$ as:

$$G^* := I^{*'} G^\dagger. \tag{21}$$

## A.3 Construction of $\Pi_R$, $D_{T,R}$ and $R^*$:

$\Pi_R$ and $D_{T,R}$ are quantities used to characterize the appropriate local deviation of the null from the truth. Their construction depends on the constructions of $\Pi_{\rho_\theta}$ and $D_{T,\rho_\theta}$.

**A.3.1 Construction of $\Pi_R$:**

Take $W_T = R\Pi_{\rho_\theta} = [W_{T,1} = R\Pi_{\rho_\theta,j_1}, \ldots, W_{T,q_{\rho_\theta}} = R\Pi_{\rho_\theta,j_{\rho_\theta}}]$ (not depending on $T$) in the LBT-Construction in Appendix A.1.2. Note that the partition of $W_T$ was informed by the indices $j_1, \ldots, j_{q_{\rho_\theta}}$ defined immediately after constructing $\Pi_{\rho_\theta}$. These indices do not depend on $T$, and they also informed the construction of $D_{T,\rho_\theta}$. Once again, to emphasize the non-dependence on $T$, write $W_T$ as $W$, and accordingly write the rest of the notation from the LBT-Construction. Thus $r = d_R$, $c = d_\theta$ and $m = q_{\rho_\theta}$. $W$ is full row-rank by the definition of $R$, $\Pi_{\rho_\theta}$ and Lemma 10 (in Appendix C).

$$\Pi_R = [\Pi_{R,1}, \ldots, \Pi_{R,q_{\rho_\theta}}] \text{ is the } d_R \times d_R \text{ matrix } \Pi_T \text{ from the LBT-Construction with } W = R\Pi_{\rho_\theta}. \tag{22}$$

### A.3.2 Construction of $D_{T,R}$:

The construction of $D_{T,R}$ depends on the matrix $D_{T,\rho\theta}\Pi'_{\rho\theta}R'\Pi_R$. Let $c^*_{R,j} = c^*_j \geq 0$ denote the number of columns of $\Pi_{R,j}$ for $j = 1, \ldots, q_{\rho\theta}$, and $q_R = q$ from (i) in the LBT-Construction (of $\Pi_R$). Let $(j_{n_1}, \ldots, j_{n_{q_R}})$ denote the sub-indices of the indices $(j_1, \ldots, j_{q_{\rho\theta}})$ such that the block $\Pi'_{\rho\theta,j_{n_i}}R'\Pi_{R,n_i}$ of dimension $c^*_{\rho\theta,j_{n_i}} \times c^*_{R,n_i}$ is full column-rank $c^*_{R,n_i} > 0$ for $i = 1, \ldots, q_R$ and $\sum_{i=1}^{q_R} c^*_{R,n_i} = d_R$. Thus, the corresponding block of $D_{T,\rho\theta}\Pi'_{\rho\theta}R'\Pi_R$ is $\frac{\sqrt{T}}{\lambda_{T,j_{n_i}}}\Pi'_{\rho\theta,j_{n_i}}R'\Pi_{R,n_i}$. Accordingly, for $D_{T,\rho\theta}\Pi'_{\rho\theta}R'\Pi_R$, the columns from $(d_R - \sum_{i'=i}^{q_R} c^*_{R,n_{i'}})$ to $(d_R - \sum_{i'=i}^{q_R} c^*_{R,n_{i'}} + c^*_{R,n_i})$ for $i = 1, \ldots, q_R$ are represented by the $d_\theta \times c^*_{R,n_i}$ matrix:

$$\begin{aligned}
&\left[0', \frac{\sqrt{T}}{\lambda_{T,j_{q_{\rho\theta}}}}\left(\Pi'_{\rho\theta,j_{q_{\rho\theta}}}R'\Pi_{R,q_{\rho\theta}}\right)'\right]' \text{ if } n_i = q_{\rho\theta},\\
&\left[0', \frac{\sqrt{T}}{\lambda_{T,j_{n_i}}}\left(\Pi'_{\rho\theta,j_{n_i}}R'\Pi_{R,n_i}\right)', \ldots, \frac{\sqrt{T}}{\lambda_{T,j_{q_{\rho\theta}}}}\left(\Pi'_{\rho\theta,j_{q_{\rho\theta}}}R'\Pi_{R,n_i}\right)'\right]' \text{ otherwise.}
\end{aligned}$$

In both cases, 0 represents the sub-matrix of zeros with number of rows that make the number of rows of the corresponding matrix equal to $d_\theta$.

Now, conforming to this above structure, define the $d_R \times d_R$ matrix $D_{T,R}$ as:

$$D_{T,R} := T^{-1/2}\mathrm{diag}\left(\lambda_{T,j_{n_1}}1_{c^*_{R,n_1}}, \ldots, \lambda_{T,j_{n_{q_R}}}1_{c^*_{R,n_{q_R}}}\right). \tag{23}$$

### A.3.3 Construction of $R^*$:

Define the $d_R \times d_\theta$ matrix $R^*$ as the transpose of the following limit:

$$R^{*'} := \lim_{T\to\infty} D_{T,\rho\theta}\Pi'_{\rho\theta}R'\Pi_R D_{T,R}. \tag{24}$$

By construction, $R^{*'}$ is finite, and its columns from $(d_R - \sum_{i'=i}^{q_R} c^*_{R,n_{i'}})$ to $(d_R - \sum_{i'=i}^{q_R} c^*_{R,n_{i'}} + c^*_{R,n_i})$ for $i = 1, \ldots, q_R$ are represented by the $d_\theta \times c^*_{R,n_i}$ matrix:

$$\begin{aligned}
&\left[0', \left(\Pi'_{\rho\theta,j_{q_{\rho\theta}}}R'\Pi_{R,q_{\rho\theta}}\right)'\right]' \text{ if } n_i = q_{\rho\theta},\\
&\left[0', \left(\Pi'_{\rho\theta,j_{n_i}}R'\Pi_{R,n_i}\right)', 0'\right]' \text{ otherwise.}
\end{aligned}$$

(As above, 0 denotes sub-matrices of zeros with number of rows, which can be zero, such that the number of rows of the corresponding matrix is $d_\theta$). Naturally, under our assumptions $R^*$ is full row-rank.

## Technical Appendix B: For the references from Section 3

### Appendix B.1: Efficient influence function for $\beta^0 := R\theta^0$ under (1)

It is well-known that under the assumptions that (1) holds, $G(\theta^0)$ is full column-rank, and $V(\theta^0)$ is positive definite: the efficient estimator of $R\theta^0$ has an asymptotically linear representation $-\sqrt{T}l_T(\theta^0) + o_p(1)$. Unfortunately, we could not find a paper to cite the proof of it. So we provide a standard proof.

**Lemma 9** *Let $\{Z_t\}_{t=1}^T$ be i.i.d. copies of a random variable $Z$, and let* (1) *holds. If $G := \frac{\partial}{\partial\theta'}E[g(Z;\theta)]_{\theta=\theta^0}$ is a full column-rank $d_g \times d_\theta$ matrix and $V := E[g(Z;\theta^0)g'(Z;\theta^0)]$ is a $d_g \times d_g$ positive definite matrix, then the asymptotic variance lower bound for any regular estimator of the $d_R \times 1$ parameter vector $\beta^0 := R\theta^0$ where $d_R \leq d_\theta$ is $(R(G'V^{-1}G)R')^{-1}$. The regular estimator whose asymptotic variance attains this bound has the asymptotically linear representation $\sqrt{T}(\widehat{\beta^0} - \beta^0) = -\sqrt{T}l_T(\theta^0) + o_p(1)$.*

**Proof:** Consider a parametric path $\xi$ of the distribution of $Z$ such that for the unique value $\xi^0$ we have the joint density $f_{\xi^0}(z) = f(z)$, the true density. Let $s_\xi(Z)$ denote the score with respect to $\xi$. Without any other restrictions, the tangent space for the model is simply $\mathcal{T} = a(z)$ where $a(z)$ satisfies $E[a(Z)] = 0$, and $E[.]$ equivalently stands for $E_{\xi^0}[.]$. Since $d_g > d_R$, (1) equivalently requires that for any given $d_R \times d_g$ matrix $B$, the relation $BE[g(Z;\theta^0)] = 0$ holds. Take $B$ as full row-rank without loss of generality. Now, differentiating with respect to $\xi$ under the expectation we obtain $\frac{\partial\theta(\xi^0)}{\partial\xi} = -(BG)^{-1}E[Bg(Z;\theta^0)s_{\xi^0}(Z)]$ and thus $\frac{\partial\beta^0(\xi^0)}{\partial\xi} = -R(BG)^{-1}E[Bg(Z;\theta^0)s_{\xi^0}(Z)]$. Therefore, any regular estimator for $\beta^0$ will be asymptotically linear with the influence function $\varphi(B) := -R(BG)^{-1}Bg(Z;\theta^0)$. Given the structure of the tangent space $\mathcal{T}$, (1) implies that the projection of this influence function $\varphi(B)$ onto $\mathcal{T}$ is $\varphi(B)$ itself. For this given $B$, $Var(\varphi(B)) = \Sigma(B) := R(BG)^{-1}BVB'(BG)^{-1'}R'$. Thus the efficient influence function is obtained by choosing $B^* := \arg\min_B \Sigma(B) = G'V^{-1}$, giving $\Sigma(B^*) = R(G'V^{-1}G)^{-1}R'$ and $\varphi(B^*) = -R(G'V^{-1}G)^{-1}G'V^{-1}g(Z;\theta^0)$. This completes the proof. ■

### Appendix B.2: The second-step test statistic $LM_{T,S}^{\mathbf{ES}}(\theta)$ in Chaudhuri and Zivot (2011):

Given the choice of $S$ in (3), the scores for $\beta$ and $\gamma_s$, by which we mean here the population version of the optimal rotations, in the efficient GMM sense, of $\bar{g}_T(\theta^0)$ along the directions of $\beta$ and $\gamma_s$ are:

$$l_{\beta,S,T}(\theta) := R_S^{1'}G'(\theta)V^{-1}(\theta)\bar{g}_T(\theta) \quad \text{and} \quad l_{\gamma_S,S,T}(\theta) := S_S^{1'}G'(\theta)V^{-1}(\theta)\bar{g}_T(\theta)$$

respectively. It is important to note that while the definition of $\beta := R\theta$ does not depend on $S$, the score for $\beta$ in the re-parameterized model generally depends on $S$ through $R_S^1$ [see Remark 21].

Following Chaudhuri and Zivot (2011), the efficient score for $\beta$ would be the residual from a regression:

$$l_{\beta.\gamma_S,S,T}(\theta) := l_{\beta,S,T}(\theta) - \text{Cov}\left(\sqrt{T}l_{\beta,S,T}(\theta), \sqrt{T}l_{\gamma_S,S,T}(\theta)\right)\text{Var}^{-1}\left(\sqrt{T}l_{\gamma_S,S,T}(\theta)\right)l_{\gamma_S,S,T}(\theta).$$

Define $\Omega(\theta) := G'(\theta)V^{-1}(\theta)G(\theta)$. Then, NM-9.2 gives $\sqrt{T}\left(l_{\beta.\gamma_S,S,T}(\theta) - E\left[l_{\beta.\gamma_S,S,T}(\theta)\right]\right) \xrightarrow{d} N(0,\Xi(\theta))$

$$\begin{aligned}\text{where}\quad \Xi_S(\theta) &:= \left(R_S^{1\prime}\Omega(\theta)R_S^1\right) - \left(R_S^{1\prime}\Omega(\theta)S_S^1\right)\left(S_S^{1\prime}\Omega(\theta)S_S^1\right)^{-1}\left(S_S^{1\prime}\Omega(\theta)R_S^1\right)\\ &= R_S^{1\prime}G'(\theta)V^{-1/2\prime}(\theta)\left(I_{d_g} - P\left(V^{-1/2}(\theta)G(\theta)S_S^1\right)\right)V^{-1/2}(\theta)G(\theta)R_S^1.\end{aligned}$$

Using the definitions of $G(\theta)$ and $V(\theta)$, the feasible version for $l_{\beta.\gamma_S,S,T}(\theta)$ (as is $\widehat{l}_T(\theta)$ for $l_T(\theta)$) is:

$$\widehat{l}_{\beta.\gamma_S,S,T}(\theta) = R_S^{1\prime}\widehat{G}_T'(\theta)\widehat{V}_T^{-1/2\prime}(\theta)\left(I_{d_g} - P\left(\widehat{V}_T^{-1/2}(\theta)\widehat{G}_T(\theta)S_S^1\right)\right)\widehat{V}_T^{-1/2}(\theta)\bar{g}_T(\theta),$$

and, similarly, $\widehat{\Xi}_T(\theta)$ for $\Xi_T(\theta)$. Then, the statistic in Chaudhuri and Zivot (2011) would be defined as:

$$\begin{aligned}LM_{T,S}^{\text{ES}}(\theta) &:= T\times\widehat{l}'_{\beta.\gamma_S,T}(\theta)\;\widehat{\Xi}_{S,T}^{-1}(\theta)\;\widehat{l}_{\beta.\gamma_S,T}(\theta)\\ &= T\times\left(\widehat{V}_T^{-1/2}(\theta)\bar{g}_T(\theta)\right)'P\left(\left(I_{d_g} - P\left(\widehat{V}_T^{-1/2}(\theta)\widehat{G}_T(\theta)S_S^1\right)\right)\widehat{V}_T^{-1/2}(\theta)\widehat{G}_T(\theta)R_S^1\right)\left(\widehat{V}_T^{-1/2}(\theta)\bar{g}_T(\theta)\right).\end{aligned}$$

Chaudhuri and Zivot (2011) noted that Remark 5 from Section 3 is equally applicable to $LM_{T,S}^{\text{ES}}(\theta)$.

**Appendix B.3: Proofs of Lemmas 1 and 2 from Section 3:**

We will repeatedly use the following relations that follow since $A_S = [R', S']'$ and $A_S^{-1} = [R_S^1, S_S^1]$:

$$RR_S^1 = I_{d_R},\; RS_S^1 = 0,\; SR_S^1 = 0,\; SS_S^1 = I_{d_\theta - d_R}\;\text{ and }\; R_S^1R + S_S^1S = I_{d_\theta}. \tag{25}$$

We will suppress the dependence of the quantities on $\theta$ to avoid notational clutter. We will not consider the negligible set on which the maintained assumptions are allowed to not hold since we only require to show the intended results hold almost surely.

**Proof of Lemma 1:** Consider any $(d_\theta - d_R)\times d_\theta$ full row-rank matrix $S$ in (3) such that $[R', S']'$ is nonsingular. Let $\zeta$ be a $d_\theta\times(d_\theta - d_R)$ matrix whose columns form a basis for the null space of $R$. Therefore, since $RS_S^1 = 0$ by (25) while $S_S^1$ is full column-rank by definition, we have $S_S^1 = \zeta B_S$ for some $(d_\theta - d_R)\times(d_\theta - d_R)$ nonsingular matrix $B_S$. Therefore, for any two such choices of $S = S^*, S^\dagger$, we have the corresponding $S_{S^*}^1 = \zeta B_{S^*}$ and $S_{S^\dagger}^1 = \zeta B_{S^\dagger}$ for some $(d_\theta - d_R)\times(d_\theta - d_R)$ nonsingular matrices $B_{S^*}$ and $B_{S^\dagger}$. This implies that $S_{S^*}^1 = S_{S^\dagger}^1 B$ where $B = B_{S^\dagger}^{-1}B_{S^*}$ is $(d_\theta - d_R)\times(d_\theta - d_R)$ and nonsingular.

Now for any $d_\theta \times d_\theta$ nonsingular matrix $M = [M_1, M_2]$, where $M_1$ is $d_\theta \times d_R$, define:

$$\begin{aligned}
\Phi_T(M) &:= T \times \left(\widehat{V}_T^{-1/2}\bar{g}_T\right)' P\left(\widehat{V}_T^{-1/2}\widehat{G}_T M\right)\left(\widehat{V}_T^{-1/2}\bar{g}_T\right), && (26)\\
\Phi_{1.2,T}(M) &:= T \times \left(\widehat{V}_T^{-1/2}\bar{g}_T\right)' P\left(\left(I_{d_g} - P\left(\widehat{V}_T^{-1/2}\widehat{G}_T M_2\right)\right)\widehat{V}_T^{-1/2}\widehat{G}_T M_1\right)\left(\widehat{V}_T^{-1/2}\bar{g}_T\right), && (27)\\
\Phi_{2,T}(M_2) &:= T \times \left(\widehat{V}_T^{-1/2}\bar{g}_T\right)' P\left(\widehat{V}_T^{-1/2}\widehat{G}_T M_2\right)\left(\widehat{V}_T^{-1/2}\bar{g}_T\right), && (28)
\end{aligned}$$

and note that, by construction:

(i) $\Phi_T(M) = \Phi_T(I_{d_\theta})$ since $M$ is nonsingular,

(ii) $\Phi_T(M) = \Phi_{1.2,T}(M) + \Phi_{2,T}(M_2)$ since is $M$ is partitioned as $M = [M_1, M_2]$,

(iii) $\Phi_{2,T}(M_2) = \Phi_{2,T}(M_2 B)$ since $B$ is a $(d_\theta - d_R) \times (d_\theta - d_R)$ nonsingular matrix.

In the above, now take $M = M^*, M^\dagger$ where $M^* = [R^1_{S^*}, S^1_{S^*}]$ and $M^\dagger = [R^1_{S^\dagger}, S^1_{S^\dagger}]$ correspond to the two choices $S = S^*, S^\dagger$ respectively. Thus we obtain:

$$\begin{aligned}
\Phi_T(M^*) &= \Phi_T(M^\dagger) \quad \text{[by (i)]}\\
\Phi_{1.2,T}(M^*) + \Phi_{2,T}(M_2^*) &= \Phi_{1.2,T}(M^\dagger) + \Phi_{2,T}(M_2^\dagger) \quad \text{[by (ii)]}\\
\Phi_{1.2,T}(M^*) &= \Phi_{1.2,T}(M^\dagger) \quad \text{[by (iii), since } M_2^* := S^1_{S^*} = S^1_{S^\dagger} B =: M_2^\dagger B\text{]}.
\end{aligned}$$

Thus, by its definition, we obtain that $LM^{\text{ES}}_{T,S^*}(\theta) = \Phi_{1.2,T}(M^*) = \Phi_{1.2,T}(M^\dagger) = LM^{\text{ES}}_{T,S^\dagger}\theta)$. ■

**Proof of Lemma 2:** Define $\widehat{\Omega} := \widehat{G}_T'(\theta)\widehat{V}_T^{-1}(\theta)\widehat{G}_T(\theta)$. We are ignoring the negligible set outside which $\widehat{\Omega}$ is positive definite. Thus, the null space of $R\widehat{\Omega}^{-1}$ is of dimension $d_\theta - d_R$ since $R$ is full row rank. Now consider a $(d_\theta - d_R) \times d_\theta$ matrix $S$ whose rows form the basis for the null space of $R\widehat{\Omega}^{-1}$.

*Claim 1:* With this $S$, we have a nonsingular $A_S := [R', S']'$ in (3).

*Proof:* Suppose not. Then, the full row-rank of $R = [R_1', \ldots, R_{d_R}']'$ implies that there exists a $(d_\theta - d_R) \times 1$ vector $c \neq 0$ such that $R_1 = \sum_{j=2}^{d_R} a_j R_j + c'S$ for some scalar coefficients $a_2, \ldots, a_{d_R}$. Since $\widehat{\Omega}^{-1}$ is positive definite except in the negligible set that we are ignoring, it means that for this $c \neq 0$, we have $R_1\widehat{\Omega}^{-1} = \sum_{j=2}^{d_R} a_j R_j\widehat{\Omega}^{-1} + c'S\widehat{\Omega}^{-1}$. Post-multiply both sides by $S'$ and note that the rows of $S$ belong in the null space of $R\widehat{\Omega}^{-1}$, i.e. $R_j\widehat{\Omega}^{-1}S' = 0$ for $j = 1, \ldots, d_R$. Hence, it follows that $0 = c'S\widehat{\Omega}^{-1}S'$. Since $S\widehat{\Omega}^{-1}S'$ is positive definite (as $\widehat{\Omega}^{-1}$ is positive definite and as the rows of $S$ are linearly independent), this is only possible if $c = 0$, which contradicts our supposition. Therefore, Claim 1 is true. ■

*Claim 2:* $R\widehat{\Omega}^{-1}S' = 0$ if and only if $R^{1'}_S\widehat{\Omega}S^1_S = 0$.

*Proof:* We use (25) repeatedly in this proof. Post-multiply $R^{1'}_S\widehat{\Omega}S^1_S = 0$ by $S$ to get $R^{1'}_S\widehat{\Omega}S^1_S S = 0$, i.e.,

$R_S^{1\prime}\widehat{\Omega}(I_{d_\theta} - R_S^1 R) = 0$ by (25). Hence,

$$R = (R_S^{1\prime}\widehat{\Omega}R_S^1)^{-1}R_S^{1\prime}\widehat{\Omega}. \tag{29}$$

Similarly obtain $S = (S_S^{1\prime}\widehat{\Omega}S_S^1)^{-1}S_S^{1\prime}\widehat{\Omega}$. Thus, $R\widehat{\Omega}^{-1}S' = (R_S^{1\prime}\widehat{\Omega}R_S^1)^{-1}(R_S^{1\prime}\widehat{\Omega}S_S^1)(S_S^{1\prime}\widehat{\Omega}S_S^1)^{-1}$. Hence, $R\widehat{\Omega}^{-1}S' = 0$ if and only if $R_S^{1\prime}\widehat{\Omega}S_S^1 = 0$, once again by using the positive definiteness of $\widehat{\Omega}$. ∎

Thus, using this specific choice of $S$ for which $R\widehat{\Omega}^{-1}S' = 0$ and hence $R_S^{1\prime}\widehat{\Omega}S_S^1 = 0$, we obtain from the definition of $LM_{T,S}^{\mathrm{ES}}(\theta)$ that $LM_{T,S}^{\mathrm{ES}}(\theta) = T \times \left(\widehat{V}_T^{-1/2}\bar{g}_T\right)' P\left(\widehat{V}_T^{-1/2}\widehat{G}_T R_S^1\right)\left(\widehat{V}_T^{-1/2}\bar{g}_T\right)$. On the other hand, (4) gives:

$$\begin{aligned} LM_T(\theta) &= T \times \left(\widehat{V}_T^{-1/2}\bar{g}_T\right)' P\left(\widehat{V}_T^{-1/2}\widehat{G}_T\widehat{\Omega}^{-1}R'\right)\left(\widehat{V}_T^{-1/2}\bar{g}_T\right) \\ &= T \times \left(\widehat{V}_T^{-1/2}\bar{g}_T\right)' P\left(\widehat{V}_T^{-1/2}\widehat{G}_T R_S^1(R_S^{1\prime}\widehat{\Omega}R_S^1)^{-1}\right)\left(\widehat{V}_T^{-1/2}\bar{g}_T\right) \end{aligned}$$

by using (29). However, since $(R_S^{1\prime}\widehat{\Omega}R_S^1)^{-1}$ is nonsingular, we have, by the construction of the projection matrix $P(.)$, that $P\left(\widehat{V}_T^{-1/2}\widehat{G}_T R_S^1(R_S^{1\prime}\widehat{\Omega}R_S^1)^{-1}\right) = P\left(\widehat{V}_T^{-1/2}\widehat{G}_T R_S^1\right)$. Therefore, $LM_T(\theta) = T \times \left(\widehat{V}_T^{-1/2}\bar{g}_T\right)' P\left(\widehat{V}_T^{-1/2}\widehat{G}_T R_S^1\right)\left(\widehat{V}_T^{-1/2}\bar{g}_T\right) = LM_{T,S}^{\mathrm{ES}}(\theta)$ (see above for the last equality). The desired result now follows from Lemma 1 for any general choice of $S$ in (3) such that $[R', S']'$ is nonsingular. ∎

**Remark 21:** The particular choice of $S$ employed to facilitate the proof of Lemma 2 has an interesting interpretation. To see it, consider the analogous population version of $S$, i.e., $S$ such that $R\Omega^{-1}S' = 0$. Similar to the proof of *Claim 1* above, it can be shown that $[R', S']'$ is nonsingular. Similar to the proof of *Claim 2* above, it can be shown that $R\Omega^{-1}S' = 0$ if and only if $R_S^{1\prime}\Omega S_S^1 = 0$, where the $R_S^1$ and $S_S^1$ correspond to this particular choice of $S$. Now, note from the discussion in Appendix B.2 that with this particular choice of $S$, the score for $\beta$, i.e., $l_{\beta,S,T}(\theta^0)$ is identical to the efficient score for $\beta$, i.e., $l_{\beta.\gamma_S,S,T}(\theta^0)$. In other words, this particular choice of $S$ in the re-parameterization (3) directly makes the scores for $\beta$ and $\gamma_S$ uncorrelated and, by asymptotic normality, asymptotically independent. A followup along this line in the case of nonlinear null hypotheses is the topic of our ongoing research.

### B.4 $\widetilde{LM}_T(\widetilde{\theta}_T) = LM_T(\widetilde{\theta}_T)$

From (26)-(28) and the definition in (3) it follows that $\widetilde{LM}_T(\theta) = LM_T(\theta) + \Phi_{2,T}(S_S^1, \theta)$ for all $\theta$ where the underlying quantities are defined. (Note that, by $\Phi_{2,T}(S_S^1, \theta)$ we mean $\Phi_{2,T}(S_S^1)$ with $\bar{g}_T$, $\widehat{G}_T$ and $\widehat{V}_T$ evaluated at $\theta$.) Now, by the definition of the $\widetilde{\theta}_T$, i.e., $(R_S^1 r_0 + S_S^1\widetilde{\gamma}_T)$ where $\widetilde{\gamma}_T$ is the GMM estimator of $\gamma$ under the restriction that $\beta = r_0$, it follows from the first order condition of the GMM optimization problem that $\Phi_{2,T}(S_S^1, \widetilde{\theta}_T) = 0$. This is because $\Phi_{2,T}(S_S^1, \theta)$ is simply a quadratic form of the first derivative of the GMM objective function with respect to $\gamma_S$, which is zero when evaluated at $\widetilde{\theta}_T$. Thus, $\widetilde{LM}_T(\widetilde{\theta}_T) = LM_T(\widetilde{\theta}_T)$. ■

# Technical Appendix C: Proofs and clarifications for Section 4

## C.1 Two useful lemmas

Since we use (have used) the following result in Lemma 10 often, let us state it here for reference.

**Lemma 10** *Let $X$ be an $a \times b$ matrix, and $P$ and $Q$ be $a \times a$ and $b \times b$ nonsingular matrices. Then, $rank(X) = rank(PX) = rank(XQ)$.*

**Proof:** $\text{rank}(X) \geq \text{rank}(PX) \geq \text{rank}(P^{-1}PX) = \text{rank}(X) \geq \text{rank}(XQ) \geq \text{rank}(XQQ^{-1}) = \text{rank}(X)$. ■

Lemma 11 lists a set of intermediate results useful for proving the results in the main text.

**Lemma 11** *Let assumptions O and N hold. Consider a sequence $\{\theta_T = R_S^1 r_0 + S_S^1 \gamma_{S,T} : T \geq 1\}$ where $r_0$ satisfies* (14) *and $\{\gamma_{S,T} : T \geq 1\}$ is such that $\theta_T$ satisfies* (15)*. Then, the following results hold as $T \to \infty$:*

*(a)* $\widehat{V}_T(\theta_T) \xrightarrow{P} V(\theta^0) \equiv V$.

*(b)* $\widehat{V}_{Gg,T}(\theta_T) \xrightarrow{P} V_{Gg}(\theta^0) \equiv V_{Gg}$.

*(c)* $\bar{G}_T(\theta_T)\Pi_{\rho_\theta} D_{T,\rho_\theta} \xrightarrow{P} G^*$ *where* $\Pi_{\rho_\theta}$, $D_{T,\rho_\theta}$ *and* $G^*$ *are as defined in* (18), (19) *and* (21) *respectively.*

*(d)* $\sqrt{T}\bar{g}_T(\theta_T) = \sqrt{T}\bar{g}_T(\theta^0) + G^* \mu_{T,\theta} + o_p(1)$ *where* $G^*$ *and* $\mu_{T,\theta}$ *are as defined in* (21) *and* (15) *respectively.*

*(e)* $\left[\widehat{V}_{1,g,T}(\theta_T)\widehat{V}_T^{-1}(\theta_T)\bar{g}_T(\theta_T), \ldots, \widehat{V}_{d_\theta,g,T}(\theta_T)\widehat{V}_T^{-1}(\theta_T)\bar{g}_T(\theta_T)\right] \Pi_{\rho_\theta} D_{T,\rho_\theta} = o_p(1)$ *(a $d_g \times d_\theta$ matrix).*

*(f)* $\widehat{G}_T(\theta_T)\Pi_{\rho_\theta} D_{T,\rho_\theta} \xrightarrow{P} G^*$ *where* $\Pi_{\rho_\theta}$, $D_{T,\rho_\theta}$ *and* $G^*$ *are as defined in* (18), (19) *and* (21) *respectively.*

**Proof:** (a) and (b) follow by assumption N8 since $\theta_T = \theta^0 + o_p(1)$.

(c) We prove it working term-by-term in the following decomposition:

$$\begin{aligned} & \bar{G}_T(\theta_T)\Pi_{\rho_\theta} D_{T,\rho_\theta} \\ = \; & \left[\bar{G}_T(\theta_T) - \bar{G}_T(\theta^0)\right] \Pi_{\rho_\theta} D_{T,\rho_\theta} + \sqrt{T}\left[\bar{G}_T(\theta^0) - \frac{\Lambda_T}{\sqrt{T}}\rho_\theta(\theta^0)\right] \frac{\Pi_{\rho_\theta} D_{T,\rho_\theta}}{\sqrt{T}} + \frac{\Lambda_T}{\sqrt{T}}\rho_\theta(\theta^0)\Pi_{\rho_\theta} D_{T,\rho_\theta}. \end{aligned} \quad (30)$$

From the definitions in (18) and (19) it follows that $\Pi_{\rho_\theta} D_{T,\rho_\theta} = o(\sqrt{T})$ by N3, and hence using N6 it follows that the second term on the right hand side (RHS) of (30) is $o_p(1)$. On the other hand, (20) and (21) imply that the third term on the RHS of (30) converges to $G^*$ by construction.

To complete the proof, now we show that the first term on the RHS of (30) is $o_p(1)$. We deviate from Antoine and Renault (2012) in the treatment of this term, and the result thus obtained has implications in terms of the allowable weakness of identification [see the part of Remark 9 that led to footnote 10]. Let $\bar{G}_{T,i}(\theta) := \frac{\partial}{\partial \theta_i}\bar{g}_T(\theta)$ denote the $i$-th column of $\bar{G}_T(\theta)$ for $i = 1, \ldots, d_\theta$ (recall that $\theta = (\theta_1, \ldots, \theta_{d_\theta})'$). Therefore, with a bad but common abuse of notation in denoting the mean values element by element, we obtain by a mean value expansion of $\bar{G}_{T,i}(\theta_T)$ around $\bar{G}_{T,i}(\theta^0)$ for $i = 1, \ldots, d_\theta$ that:

$$
\begin{aligned}
& \left[\bar{G}_T(\theta_T) - \bar{G}_T(\theta^0)\right] \Pi_{\rho_\theta} D_{T,\rho_\theta} \\
= \; & \left[\left\{\frac{\partial}{\partial \theta'}\bar{G}_{T,1}(\theta_T(\theta_1))\right\}(\theta_T - \theta^0), \ldots, \left\{\frac{\partial}{\partial \theta'}\bar{G}_{T,d_\theta}(\theta_T(\theta_{d_\theta}))\right\}(\theta_T - \theta^0)\right] \Pi_{\rho_\theta} D_{T,\rho_\theta} \\
= \; & \left[\left\{\frac{\partial}{\partial \theta_1}\bar{G}_T(\theta_T(\theta_1))\right\}(\theta_T - \theta^0), \ldots, \left\{\frac{\partial}{\partial \theta_{d_\theta}}\bar{G}_T(\theta_T(\theta_{d_\theta}))\right\}(\theta_T - \theta^0)\right] \Pi_{\rho_\theta} D_{T,\rho_\theta} \qquad (31)
\end{aligned}
$$

by twice interchanging the order in which the derivatives are taken in each of the $d_\theta$ columns. Note that, for $i = 1, \ldots, d_\theta$, we used $\theta_T(\theta_i)$ (such that $\|\theta_T(\theta_i) - \theta^0\| \leq \|\theta_T - \theta^0\|$) to denote the mean value in the first equality of the above equation. Recalling that $\mu_{T,\theta} = \sqrt{T} D^{-1}_{T,\rho_\theta}\Pi'_{\rho_\theta}(\theta_T - \theta^0)$ by (15), define $U_{T,i}$ for $i = 1, \ldots, d_\theta$ as the $d_g \times d_\theta$ matrix with

$$
\left\{\frac{\partial}{\partial \theta_i}\bar{G}_T(\theta_T(\theta_i))\right\}(\theta_T - \theta^0) = \left\{\frac{\partial}{\partial \theta_i}\bar{G}_T(\theta_T(\theta_i))\frac{\sqrt{T}}{\lambda_{T,l}}\right\}\frac{\Pi_{\rho_\theta} D_{T,\rho_\theta}\lambda_{T,j_1}}{\sqrt{T}}\mu_{T,\theta}\frac{\lambda_{T,l}}{\sqrt{T}\lambda_{T,j_1}}
$$

in the $i$-th column and zero everywhere else. [See Remark 9 for $\lambda_{T,j_1}$.] Therefore, (31) implies that:

$$
\left[\bar{G}_T(\theta_T) - \bar{G}_T(\theta^0)\right] \Pi_{\rho_\theta} D_{T,\rho_\theta} = \sum_{i=1}^{d_\theta} U_{T,i}\Pi_{\rho_\theta} D_{T,\rho_\theta},
$$

and thus,

$$
\begin{aligned}
& \left\|\left[\bar{G}_T(\theta_T) - \bar{G}_T(\theta^0)\right] \Pi_{\rho_\theta} D_{T,\rho_\theta}\right\| \\
\leq \; & \sum_{i=1}^{d_\theta} \|U_{T,i}\| \times \|\Pi_{\rho_\theta} D_{T,\rho_\theta}\| \\
\leq \; & \sum_{i=1}^{d_\theta} \left\|\frac{\partial}{\partial \theta_i}\bar{G}_T(\theta_T(\theta_i))\frac{\sqrt{T}}{\lambda_{T,l}}\right\| \times \left\|\frac{\Pi_{\rho_\theta} D_{T,\rho_\theta}\lambda_{T,j_1}}{\sqrt{T}}\right\| \times \|\mu_{T,\theta}\| \times \left\|\frac{\Pi_{\rho_\theta} D_{T,\rho_\theta}\lambda_{T,j_1}}{\sqrt{T}}\right\| \frac{\sqrt{T}\lambda_{T,l}}{\lambda^2_{T,j_1}\sqrt{T}} \\
\leq \; & \sum_{i=1}^{d_\theta} \sup_\theta \left\{\left\|\frac{\sqrt{T}}{\lambda_{T,l}}\frac{\partial}{\partial \theta_i}\frac{\Lambda_T}{\sqrt{T}}\rho_\theta(\theta)\right\| + \left\|\frac{\sqrt{T}}{\lambda_{T,l}}\frac{\partial}{\partial \theta_i}\left[\bar{G}_T(\theta) - \frac{\Lambda_T}{\sqrt{T}}\rho_\theta(\theta)\right]\right\|\right\} \times \|\mu_{T,\theta}\| \times \left\|\frac{\Pi_{\rho_\theta} D_{T,\rho_\theta}\lambda_{T,j_1}}{\sqrt{T}}\right\|^2 \frac{\lambda_{T,l}}{\lambda^2_{T,j_1}} \\
= \; & o_p(1)
\end{aligned}
$$

since, on the third line from above, the order of magnitude of the terms (from left to right) inside the sum is respectively: (i) $\sup_\theta \left\|\frac{\sqrt{T}}{\lambda_{T,l}}\frac{\partial}{\partial \theta_i}\frac{\Lambda_T}{\sqrt{T}}\rho_\theta(\theta)\right\| = O(1)$ by N3 and N4; (ii) $\sup_\theta \left\|\frac{\sqrt{T}}{\lambda_{T,l}}\frac{\partial}{\partial \theta_i}\left[\bar{G}_T(\theta) - \frac{\Lambda_T}{\sqrt{T}}\rho_\theta(\theta)\right]\right\| = o_p(1)$ by N3 and N7(a), (iii) $\|\mu_{T,\theta}\| = O_p(1)$ by (15); (iv) $\left\|\frac{\Pi_{\rho_\theta} D_{T,\rho_\theta}\lambda_{T,j_1}}{\sqrt{T}}\right\| = O(1)$ by N3, (18) and (19);

and (v) $\frac{\lambda_{T,l}}{\lambda_{T,j_1}^2} = o(1)$ by N7(b) [also see Remark 9].

(d) A mean value expansion (with similar abuse of notation as above to denote the mean value $\bar{\theta}_T$) gives $\sqrt{T}\bar{g}_T(\theta_T) = \sqrt{T}\bar{g}_T(\theta^0) + \bar{G}_T(\bar{\theta}_T)\sqrt{T}(\theta_T - \theta^0) = \sqrt{T}\bar{g}_T(\theta^0) + \bar{G}_T(\bar{\theta}_T)\Pi_{\rho_\theta} D_{T,\rho_\theta}\mu_{T,\theta} = \sqrt{T}\bar{g}_T(\theta^0) + G^*\mu_{T,\theta} + o_p(1)$ where the second equality uses (15) and the last one uses the result from Lemma 11(c).

(e) The result follows by Lemma 11 (a), (b), (d) since $\Pi_{\rho_\theta} D_{T,\rho_\theta} = o(\sqrt{T})$ by N3 and $\sqrt{T}\bar{g}_T(\theta^0) = O_p(1)$.

(f) The result follows by Lemma 11 (c) and (e). ■

## C.2 Clarification and details regarding footnote 10:

We briefly illustrate the said tradeoff by showing that if one strengthens the smoothness assumption by extending it to the second derivative, then this allows to weaken the rate assumption in N7(b). Since, at this point the definition of $\lambda_{T,j_1}$ is already stated, and since Lemma 11 works with $\lambda_{T,j_1}$ instead of $\lambda_{T,1}$, the discussion below uses $\lambda_{T,j_1}$. All we do hold if assumptions are maintained in terms of $\lambda_{T,1}$.

From Lemma 11, it is clear that the discussion here is pertinent mainly to part Lemma 11 (c) [see the proof of Lemma 11 (d)-(f)]. Indeed, the only part of (c) that needs attention is where we show that $\left[\bar{G}_T(\theta_T) - \bar{G}_T(\theta^0)\right]\Pi_{\rho_\theta} D_{T,\rho_\theta}$, i.e., the first term on the RHS of (30), is $o_p(1)$. So, let us focus on this.

**Remark 22:** If $g(Z_t;\theta)$ is linear in $\theta$, as in linear instrumental variables models, then this is trivially true since $\left[\bar{G}_T(\theta) - \bar{G}_T(\theta^0)\right] \equiv 0$ for all $\theta$. So, let us focus on a $g(Z_t;\theta)$ that is nonlinear in $\theta$.

Now, to accommodate for more smoothness we extend assumption N6 as N6' to the second derivative, and replace N7 by N7' as follows. (Assumptions N1-N5 and N8 remain the same.)

**Assumption N6':** (a one-time assumption for this clarification only)

(a) $\frac{\partial}{\partial\theta'}\psi_T(\theta^0) = \sqrt{T}\left[\bar{G}_T(\theta^0) - \frac{\Lambda_T}{\sqrt{T}}\rho_\theta(\theta^0)\right] = O_p(1)$. (This was the original N6.)

(b) For $i = 1,\ldots,d_\theta$: $\frac{\partial}{\partial\theta_i}\frac{\partial}{\partial\theta'}\psi_T(\theta^0) = \sqrt{T}\frac{\partial}{\partial\theta_i}\left[\bar{G}_T(\theta^0) - \frac{\Lambda_T}{\sqrt{T}}\rho_\theta(\theta^0)\right] = O_p(1)$. (This is the extension.)

**Assumption N7':** (a one-time assumption for this clarification only)

(a) $\rho(\theta)$ is thrice continuously differentiable in $\theta \in \mathcal{N}(\theta^0)$. $g(z;\theta)$ is thrice differentiable in $\theta \in \mathcal{N}(\theta^0)$ for each $z \in \mathbb{R}^{d_z}$ and $\sup_{\theta\in\mathcal{N}(\theta^0)}\left\|\frac{\partial}{\partial\theta_i}\frac{\partial}{\partial\theta_k}\left[\bar{G}_T(\theta) - \frac{\Lambda_T}{\sqrt{T}}\rho_\theta(\theta)\right]\right\| = o_p(\lambda_{T,l}/\sqrt{T})$ for $i,k = 1,\ldots,d_\theta$.

(b) $\lambda_{T,j_1}$ from (19) satisfies $\lambda_{T,j_1}^3/\lambda_{T,l} \to \infty$ as $T \to \infty$.

Comparing assumption N7 with N7' reveals the tradeoff in terms of parts (a) and (b) of these assumptions. We note that similar tradeoffs can be generated by working with higher order derivatives.

For clarity, specify further structure but without loss of generality. First, for $i = 1,\ldots,d_\theta$, define $\Pi_{\rho_\theta,i}$ and $D_{T,\rho_\theta,i}$ by the UBT-Construction like that in (18) and (19), but this time, by taking

$$W_T = \left[\frac{\partial}{\partial\theta_i}\rho'_{\theta,1}(\theta^0),\ldots,\frac{\partial}{\partial\theta_i}\rho'_{\theta,l}(\theta^0)\right] = \left(I^*\frac{\partial}{\partial\theta_i}\rho_\theta(\theta^0)\right)'$$

(instead of $W_T = \left[\rho'_{\theta,1}(\theta^0), \dots, \rho'_{\theta,l}(\theta^0)\right] = \left(I^*\rho_\theta(\theta^0)\right)'$) not depending on $T$ in the UBT-Construction. The corresponding quantities with full column-rank, and thus also the elements of $D_{T,\rho_\theta,i}$ will change. Indeed, no full-rank conditions are required, and instead, for the purpose of this proof, the only properties we will require are: For $i = 1, \dots d_\theta$,

$$\begin{aligned} \left(I^{*'}\left\{\frac{\partial}{\partial\theta_i} I^* \frac{\Lambda_T}{\sqrt{T}} I^{*'} I^* \rho_\theta(\theta^0)\right\} \Pi_{\rho_\theta,i} D_{T,\rho_\theta,i}\right) &= O(1), \qquad (32) \\ \Pi_{\rho_\theta,i} D_{T,\rho_\theta,i} &= o(\sqrt{T}) \qquad (33) \end{aligned}$$

and these will not change since (32) holds by the construction of $D_{T,\rho_\theta,i}$, while (33) follows from N3.

Start from (31). All we do below is to tease out further structure in the non-zero (i.e.,the $i$-th) column of $U_{T,i}$ (defined below (31)) so that assumption N7' could be effectively used to show that $\left[\bar{G}_T(\theta_T) - \bar{G}_T(\theta^0)\right] \Pi_{\rho_\theta} D_{T,\rho_\theta}$, i.e., the first term on the RHS of (30) is $o_p(1)$. With this purpose in mind, for each $i = 1, \dots, d_\theta$, consider a further mean value expansion (with similar abuse of notation, and this time using $\theta_T(\theta_i^k)$ to denote the mean value such that $\|\theta_T(\theta_i^k) - \theta^0\| \leq \|\theta_T(\theta_i) - \theta^0\| \leq \|\theta_T - \theta^0\|$ for $k = 1, \dots, d_\theta$):

$$\begin{aligned} & \left\{\frac{\partial}{\partial\theta_i} \bar{G}_T(\theta_T(\theta_i))\right\}(\theta_T - \theta^0) \\ = & \left\{\frac{\partial}{\partial\theta_i} \bar{G}_T(\theta^0)\right\}(\theta_T - \theta^0) \\ & + \left[\left\{\frac{\partial}{\partial\theta_i}\frac{\partial}{\partial\theta_1} \bar{G}_T(\theta_T(\theta_i^1))\right\}(\theta_T(\theta_i) - \theta^0), \dots, \left\{\frac{\partial}{\partial\theta_i}\frac{\partial}{\partial\theta_{d_\theta}} \bar{G}_T(\theta_T(\theta_i^{d_\theta}))\right\}(\theta_T(\theta_i) - \theta^0)\right](\theta_T - \theta^0) \end{aligned}$$

by similar (to above) interchange in the order of the derivatives. Since $\mu_{T,\theta} = \sqrt{T} D^{-1}_{T,\rho_\theta} \Pi'_{\rho_\theta}(\theta_T - \theta^0)$ by (15), it follows that:

$$\begin{aligned} \left\{\frac{\partial}{\partial\theta_i} \bar{G}_T(\theta^0)\right\}(\theta_T - \theta^0) = & \underbrace{\left(I^{*'}\left\{\frac{\partial}{\partial\theta_i} I^* \frac{\Lambda_T}{\sqrt{T}} I^{*'} I^* \rho_\theta(\theta^0)\right\} \Pi_{\rho_\theta,i} D_{T,\rho_\theta,i}\right) \mu_{T,\theta} \frac{1}{\sqrt{T}}}_{= u_{a,T,i} \text{ (say)}} \\ & + \underbrace{\left(\frac{\partial}{\partial\theta_i}\sqrt{T}\left(\bar{G}_T(\theta^0) - \frac{\Lambda_T}{\sqrt{T}}\rho_\theta(\theta^0)\right)\right)\left(\frac{\Pi_{\rho_\theta,i} D_{T,\rho_\theta,i}}{\sqrt{T}}\right) \mu_{T,\theta} \frac{1}{\sqrt{T}}}_{= u_{b,T,i} \text{ (say)}} \end{aligned}$$

for $i = 1, \dots, d_\theta$. Define the $d_g \times d_\theta$ matrices $U_{a,T,i}$ and $U_{b,T,i}$ such that all their columns are zeros, except for the $i$-th column, which for them is $u_{a,T,i}$ and $u_{b,T,i}$ respectively. Do this for all $i = 1, \dots, d_\theta$.

On the other hand, for the notation-abused quantity $\theta_T(\theta_i)$, define $\mu_{T,\theta(i)} := \sqrt{T} D^{-1}_{T,\rho_\theta} \Pi'_{\rho_\theta}(\theta_T(\theta_i) - \theta^0)$ where $\|\mu_{T,\theta(i)}\| \leq \|\mu_{T,\theta}\|$ by construction for $i = 1, \dots, d_\theta$ (recall that $\mu_{T,\theta} = \sqrt{T} D^{-1}_{T,\rho_\theta} \Pi'_{\rho_\theta}(\theta_T - \theta^0)$ by (15)). Now for each $i = 1, \dots, d_\theta$, define the $d_g \times d_\theta$ matrices $U_{c,T,i,k}$ for $k = 1, \dots, d_\theta$ such that all the columns of $U_{c,T,i,k}$ are zeros, except for the $k$-th column which is:

$$\left\{\frac{\partial}{\partial\theta_i}\frac{\partial}{\partial\theta_k}\bar{G}_T(\theta_T(\theta_i^k))\right\}(\theta_T(\theta_i)-\theta^0)=\left\{\frac{\partial}{\partial\theta_i}\frac{\partial}{\partial\theta_k}\bar{G}_T(\theta_T(\theta_i^k))\frac{\sqrt{T}}{\lambda_{T,l}}\right\}\frac{\Pi_{\rho_\theta}D_{T,\rho_\theta}\lambda_{T,j_1}}{\sqrt{T}}\mu_{T,\theta(i)}\frac{\lambda_{T,l}}{\sqrt{T}\lambda_{T,j_1}}.$$

Therefore, it follows that $U_{T,i}$ (defined below (31)) can be written as:

$$U_{T,i}=U_{a,T,i}+U_{b,T,i}+\left(\sum_{k=1}^{d_\theta}U_{c,T,i,k}\right)(\theta_T-\theta^0)=U_{a,T,i}+U_{b,T,i}+\left(\sum_{k=1}^{d_\theta}U_{c,T,i,k}\right)\frac{\Pi_{\rho_\theta}D_{T,\rho_\theta}\lambda_{T,j_1}}{\sqrt{T}}\mu_{T,\theta}\frac{1}{\lambda_{T,j_1}}.$$

And, therefore,

$$\begin{aligned}
&\left\|\left[\bar{G}_T(\theta_T)-\bar{G}_T(\theta^0)\right]\Pi_{\rho_\theta}D_{T,\rho_\theta}\right\|\\
\leq\ &\sum_{i=1}^{d_\theta}\|U_{T,i}\|\times\|\Pi_{\rho_\theta}D_{T,\rho_\theta}\|\\
\leq\ &\sum_{i=1}^{d_\theta}\|U_{a,T,i}\|\times\|\Pi_{\rho_\theta}D_{T,\rho_\theta}\|+\sum_{i=1}^{d_\theta}\|U_{b,T,i}\|\times\|\Pi_{\rho_\theta}D_{T,\rho_\theta}\|\\
&+\sum_{i=1}^{d_\theta}\sum_{k=1}^{d_\theta}\|U_{c,T,i,k}\|\times\left\|\frac{\Pi_{\rho_\theta}D_{T,\rho_\theta}\lambda_{T,j_1}}{\sqrt{T}}\right\|\times\frac{\|\mu_{T,\theta}\|}{\lambda_{T,j_1}}\times\|\Pi_{\rho_\theta}D_{T,\rho_\theta}\|.
\end{aligned}$$

Since $\|u_{a,T,i}\|=O(1/\sqrt{T})$ by its definition and using (32), it follows that $\sum_{i=1}^{d_\theta}\|U_{a,T,i}\|\times\|\Pi_{\rho_\theta}D_{T,\rho_\theta}\|=o_p(1)$ by using (33). Since $\|u_{b,T,i}\|=O(1/\sqrt{T})$ by its definition and using N6' and (33), it follows that $\sum_{i=1}^{d_\theta}\|U_{b,T,i}\|\times\|\Pi_{\rho_\theta}D_{T,\rho_\theta}\|=o_p(1)$ by using (33). Finally, note that $\sum_{i=1}^{d_\theta}\sum_{k=1}^{d_\theta}\|U_{c,T,i,k}\|\times\left\|\frac{\Pi_{\rho_\theta}D_{T,\rho_\theta}\lambda_{T,j_1}}{\sqrt{T}}\right\|\times\frac{\|\mu_{T,\theta}\|}{\lambda_{T,j_1}}\times\|\Pi_{\rho_\theta}D_{T,\rho_\theta}\|=o_p(1)$ since, collecting similar terms together,

$$\begin{aligned}
&\|U_{c,T,i,k}\|\times\left\|\frac{\Pi_{\rho_\theta}D_{T,\rho_\theta}\lambda_{T,j_1}}{\sqrt{T}}\right\|\times\frac{\|\mu_{T,\theta}\|}{\lambda_{T,j_1}}\times\|\Pi_{\rho_\theta}D_{T,\rho_\theta}\|\\
\leq\ &\sup_\theta\left\{\left\|\frac{\Lambda_T}{\sqrt{T}}\frac{\partial}{\partial\theta_i}\frac{\partial}{\partial\theta_k}\rho_\theta(\theta)\times\frac{\sqrt{T}}{\lambda_{T,l}}\right\|+\left\|\frac{\partial}{\partial\theta_i}\frac{\partial}{\partial\theta_k}\left[\bar{G}_T(\theta)-\frac{\Lambda_T}{\sqrt{T}}\rho_\theta(\theta)\right]\frac{\sqrt{T}}{\lambda_{T,l}}\right\|\right\}\\
&\times\left\|\frac{\Pi_{\rho_\theta}D_{T,\rho_\theta}\lambda_{T,j_1}}{\sqrt{T}}\right\|^3\times\|\mu_{T,\theta(i)}\|\times\|\mu_{T,\theta}\|\times\frac{\lambda_{T,l}}{\lambda_{T,j_1}^3}\\
=\ &O_p(1)\times O(1)\times O_p(1)\times O_p(1)\times o(1)
\end{aligned}$$

term by term: (i) $\sup_\theta\left\|\frac{\Lambda_T}{\sqrt{T}}\frac{\partial}{\partial\theta_i}\frac{\partial}{\partial\theta_k}\rho_\theta(\theta)\times\frac{\sqrt{T}}{\lambda_{T,l}}\right\|+\left\|\frac{\partial}{\partial\theta_i}\frac{\partial}{\partial\theta_k}\left[\bar{G}_T(\theta)-\frac{\Lambda_T}{\sqrt{T}}\rho_\theta(\theta)\right]\frac{\sqrt{T}}{\lambda_{T,l}}\right\|=O(1)+o_p(1)$ by using N3 and N7'(a); (ii) $\left\|\frac{\Pi_{\rho_\theta}D_{T,\rho_\theta}\lambda_{T,j_1}}{\sqrt{T}}\right\|^3=O(1)$ by using N3, (18) and (19); (iii) $\|\mu_{T,\theta(i)}\|=O_p(1)$ by using (15) and the definition of $\mu_{T,\theta(i)}$; (iv) $\|\mu_{T,\theta}\|=O_p(1)$ by using (15); and (v) $\frac{\lambda_{T,l}}{\lambda_{T,j_1}^3}=o(1)$ by using N7'(b). Thus $\left[\bar{G}_T(\theta_T)-\bar{G}_T(\theta^0)\right]\Pi_{\rho_\theta}D_{T,\rho_\theta}=o_p(1)$.

This completes the announced demonstration in the clarification for footnote 10. ■

### C.3 Proof of the results from Section 4

**Proof of Lemma 3:** The proof is based on the original work of Antoine and Renault (2012), Andrews and Guggenberger (2014), Andrews and Cheng (2014) and Cheng (2015), with suitable adjustments required by our setup. Let $\widehat{G}_T := \widehat{G}_T(\theta^0)$, $\widehat{V}_T := \widehat{V}_T(\theta^0)$. By M1 and M2, $\widehat{V}_T$ is positive definite with probability approaching one as $T \to \infty$. Thus, if defined, let $\widehat{V}_T^{-1/2}$ be such that $\widehat{V}_T^{-1/2'}\widehat{V}_T^{-1/2} = \widehat{V}_T^{-1}$ and let $\widehat{g}_T := \widehat{V}_T^{-1/2}\bar{g}_T(\theta^0)$ and $H_T := \widehat{V}_T^{-1/2}\widehat{G}_T$. Then, for $T$ sufficiently large, (4) gives:

$$\begin{aligned} LM_T(\theta^0) &= T\widehat{g}_T' P\left(H_T\{H_T'H_T\}^- R'\right)\widehat{g}_T \\ &= T\widehat{g}_T' P\Big(H_T B_T \Upsilon_T\left\{(H_T B_T \Upsilon_T)'(H_T B_T \Upsilon_T)\right\}^- \Upsilon_T B_T' R' \Pi_T^* D_T^*\Big)\widehat{g}_T \end{aligned}$$

where $\Upsilon_T := \mathrm{diag}(1/\delta_{T,1}, \ldots, 1/\delta_{T,p}, \sqrt{T}1_{d_\theta - p})$, a $d_\theta \times d_\theta$ diagonal matrix, nonsingular for any given $T$. ($1_c$ is the $1 \times c$ vector $(1, \ldots, 1)$.) $\Upsilon_T$ is $\mathrm{diag}(1/\delta_{T,1}, \ldots, 1/\delta_{T,p})$ if $d_\theta = p$ and is $\mathrm{diag}(\sqrt{T}1_{d_\theta - p})$ if $p = 0$. For a given $T$, $\Pi_T^*$ and $D_T^*$ are $d_R \times d_R$ nonsingular matrices defined as follows.

Step 1: Definition of $\Pi_T^*$ and $D_T^*$, and the asymptotic behavior of $\Upsilon_T B_T' R' \Pi_T^* D_T^*$

Under assumption M3(a) we can, without loss of generality, partition the set of elements $\delta_{T,1}, \ldots, \delta_{T,p}$ into $m-1$ groups containing $p_1, p_2, \ldots, p_{m-1}$ elements respectively as $(\delta_{T,1}, \ldots, \delta_{T,p_1})$, $(\delta_{T,\bar{p}_1+1}, \ldots, \delta_{T,\bar{p}_2})$,..., $(\delta_{T,\bar{p}_{m-2}+1}, \ldots, \delta_{T,\bar{p}_{m-1}})$ where $p_j \geq 0$ and $\bar{p}_j := \sum_{k=1}^{j} p_k$ for $j = 1, \ldots, m-1$ and $m \in \{1, \ldots, p+1\}$ (let $p_m := d_\theta - p$; and when $p = 0$ let $m = 1$; and also, by construction, $\bar{p}_{m-1} = p$ and $\bar{p}_m = d_\theta$), such that:

$$\delta_{T,\bar{p}_j} \neq o(\delta_{T,\bar{p}_j - p_j + 1}) \text{ for } j = 1, \ldots, m-1, \text{ and } \delta_{T,\bar{p}_j+1} = o(\delta_{T,\bar{p}_j}) \text{ for } j = 1, \ldots, m-2. \tag{34}$$

Taking $W_T := RB_T = [W_{T,1}, \ldots, W_{T,m}]$ where $W_{T,j} := RB_{T,(\bar{p}_j - p_j + 1:\bar{p}_j)}$ for $j = 1, \ldots, m$, define $\Pi_T^* = [\Pi_{T,1}^*, \ldots, \Pi_{T,m}^*]$ as the $\Pi_T$ matrix from the UBT-Construction in Appendix A.1.1. $B_T$ is orthogonal for each $T$ and also $B_T \to B$, which is nonsingular by M3(c). Therefore, by Lemma 10, quantities such as $q_T$ and $c_{T,j_i,T}^*$ in the UBT-Construction have well defined limits as $T \to \infty$. Denote these limits as $q$ and $c_{j_i}^*$ respectively, i.e., by dropping the subscript $T$, and note that $\sum_{i=1}^{q} c_{j_i}^* = d_R$.

Define $D_T^* = \mathrm{diag}(\delta_{T,\bar{p}_{j_1}} 1_{c_{j_1}^*}, \ldots, \delta_{T,\bar{p}_{j_q}} 1_{c_{j_q}^*})$ where we use the notation $\delta_{T,\bar{p}_{m-1}+1} = \ldots = \delta_{T,\bar{p}_m} = T^{-1/2}$ to allow for the possibility that $j_q = m$. $D_T^*$ is a $d_R \times d_R$ nonsingular diagonal matrix for each $T$.

Therefore, as $T \to \infty$, it follows by M3(a) and (34), and then again using Lemma 10, that

$$W^{*'} = \lim_{T \to \infty} \Upsilon_T B_T' R' \Pi_T^* D_T^* \tag{35}$$

is a finite, non-random, $d_\theta \times d_R$ matrix with full column-rank $d_R$.[17]

[17] To see its full column-rank, use arguments similar to those below (19) along with M3(a) to obtain that for $W^{*'}$, its

The rest of the proof is completely based on Andrews and Guggenberger (2014).

Step 2: Asymptotic behavior of $H_T B_T \Upsilon_T$

Under (9), $\|\Delta_T\| \leq c \times \bar{c}$ for some $c > 0$ by M2. Then, it follows that:

$$\begin{aligned} V_T^{-1/2}\widehat{G}_T B_T \Upsilon_T &= V_T^{-1/2}\widehat{G}_T \left[B_{T,(1:p)}\Delta_{T,(1:p)}^{-1}, \sqrt{T}B_{T,(p+1:d_\theta)}\right] \\ &= V_T^{-1/2}G_T \left[B_{T,(1:p)}\Delta_{T,(1:p)}^{-1}, \sqrt{T}B_{T,(p+1:d_\theta)}\right] \\ &\quad + V_T^{-1/2}\sqrt{T}\left(\widehat{G}_T - G_T\right)\left[B_{T,(1:p)}(\sqrt{T}\Delta_{T,(1:p)})^{-1}, B_{T,(p+1:d_\theta)}\right]. \end{aligned}$$

By the orthogonality of $B_T$, it follows from the relation $V_T^{-1/2}G_T = C_{T,(1:d_R)}\Delta_T B_T'$ (obtained from (9)) and M3, that the first term on the right hand side of the above equation converges to $[C_{(1:p)}, C_{(p+1:d_\theta)}L]$. On the other hand, M1 and M2 give $\sqrt{T}\left(\widehat{G}_T - G_T\right) \xrightarrow{d} devec_{d_g}(\psi_G - V_{Gg}V^{-1}\psi) = O_p(1)$ which, crucially, is independent of $\psi$. Also M3 implies that $[B_{T,(1:p)}(\sqrt{T}\Delta_{T,(1:p)})^{-1}, B_{T,(p+1:d_\theta)}] \to [0, B_{(p+1:d_\theta)}]$ as $T \to \infty$. Thus, by M1, for the second term on the right hand side of the above equation, we now have that $V_T^{-1/2}\sqrt{T}\left(\widehat{G}_T - G_T\right)\left[B_{T,(1:p)}(\sqrt{T}\Delta_{T,(1:p)})^{-1}, B_{T,(p+1:d_\theta)}\right] \xrightarrow{d} [0, V^{-1/2}devec_{d_g}(\psi_G - V_{Gg}V^{-1}\psi)B_{(p+1:d_\theta)}]$. Since M2 implies that $\widehat{V}_T^{-1/2}V_T^{1/2} \xrightarrow{p} I_{d_g}$, it now follows, by combining the two terms, that

$$H_T B_T \Upsilon_T = \widehat{V}_T^{-1/2}\widehat{G}_T B_T \Upsilon_T = \left(\widehat{V}_T^{-1/2}V_T^{1/2}\right)V_T^{-1/2}\widehat{G}_T B_T \Upsilon_T \xrightarrow{d} G^* \tag{36}$$

where $G^* := [C_{(1:p)}, C_{(p+1:d_\theta)}L + V^{-1/2}devec_{d_g}(\psi_G - V_{Gg}V^{-1}\psi)B_{(p+1:d_\theta)}]$ , as defined in M3(d).

Step 3: Asymptotic behavior of $LM_T(\theta^0)$

Therefore, $P(H_T B_T \Upsilon_T \{(H_T B_T \Upsilon_T)'(H_T B_T \Upsilon_T)\}^{-}\Upsilon_T B_T' R' \Pi_T^* D_T^*) \xrightarrow{d} P(G^*(G^{*'}G^*)^{-1}W^{*'})$, a finite matrix with full column-rank $d_R$ almost surely by (35), (36) and Lemma 10. Now, since M1 and M2 imply that $\sqrt{T}\widehat{g}_T \xrightarrow{d} V^{-1/2}\psi \sim N(0, I_{d_g})$, and since we have already noted the independence between $\psi$ and $G^*$, it follows that $LM_T(\theta^0) \xrightarrow{d} \chi^2_{d_R}$. ■

**Proof of Proposition 4:** Let $\{\phi_{\gamma_S,T} : T \geq 1\}$ denote the sequence of indicator variables where $\phi_{\gamma_S,T} = 0$ if $CI_T(\gamma_S;\epsilon)$ contains $\gamma_S^0$, and $\phi_{\gamma_S,T} = 1$ otherwise. It is given that $CI_T(\gamma_S;\epsilon)$ has asymptotic coverage

columns from $(d_R - \sum_{i'=i}^q c_{j_{i'}}^*)$ to $(d_R - \sum_{i'=i}^q c_{j_{i'}}^* + c_{j_i}^*)$ for $i = 1, \ldots, q$ are represented by the $d_g \times c_{j_i}^*$ matrix:

$$\begin{aligned} &\left[(\delta_{\bar{p}_1}\mathrm{diag}(\delta_1^{-1}, \ldots, \delta_{\bar{p}_1}^{-1})B'_{(1:p_1)}R'\bar{\Pi}_1)', 0'\right]' \text{ if } j_i = 1, \\ &\left[0', (\delta_{\bar{p}_{j_i}}\mathrm{diag}(\delta_{\bar{p}_{j_i}-p_{j_i}+1}^{-1}, \ldots, \delta_{\bar{p}_{j_i}}^{-1})B'_{(\bar{p}_{j_i}-p_{j_i}+1:\bar{p}_{j_i})}R'\bar{\Pi}_{j_i})', 0'\right]' \text{ otherwise} \end{aligned}$$

(where the 0 denotes sub-matrices of zeros with number of rows, which can be zero, such that the number of rows of the corresponding big matrix is $d_\theta$). Thus the non-zero blocks in such sets of columns (one block per set of columns) are: (i) at mutually non-overlapping positions (sets of rows); (ii) are finite by M1, M3(a); (iii) of full column-rank by Lemma 10, which tells that pre-multiplication by the nonsingular matrix $\delta_{\bar{p}_{j_i}}\mathrm{diag}(\delta_{\bar{p}_{j_i}-p_{j_i}+1}^{-1}, \ldots, \delta_{\bar{p}_{j_i}}^{-1})$ does not change the rank of $B'_{(\bar{p}_{j_i}-p_{j_i}+1:\bar{p}_{j_i})}R'\bar{\Pi}_{j_i}$. The latter has full column-rank $c_{j_i}^*$ for $i = 1, \ldots, q$ by (i) in the UBT-Construction. Therefore, full column-rank $d_R$ of $W^{*'}$ follows by noting that $\sum_{i=1}^q c_{j_1}^* = d_R$. Note that the additional structure in M3(a), that Andrews and Guggenberger (2014) did not require, was imposed here precisely for this step [see sentence 4 in Remark 6].

$(1-\epsilon)$ when $H_0$ is true. Hence, $\lim_{T\to\infty} Pr_T(\phi_{\gamma_S,T}=0) \geq (1-\epsilon)$ where $Pr_T(.)$ denotes the probability of an event under $F_T$ constrained by assumptions O and M1-M3 and when $\beta^0 = r_0$, equivalently, when $R\theta^0 = r_0$ [see Remark 23 and footnote 1 in that order]. Therefore, by construction:

$$\lim_{T\to\infty} Pr_T\left(\inf_{\gamma_0\in CI_T(\gamma_S;\epsilon)} LM_T\left(A_S^{-1}(r_0',\gamma_0')'\right) \leq LM_T(\theta^0)\right) \geq \lim_{T\to\infty} Pr_T(\phi_{\gamma_S,T}=0) \geq 1-\epsilon, \tag{37}$$

since for any $T\geq 1$, the event $\{\phi_{\gamma_S,T}=0\} \subseteq$ the event $\{\inf_{\gamma_0\in CI_T(\gamma_S;\epsilon)} LM_T\left(A_S^{-1}(r_0',\gamma_0')'\right) \leq LM_T(\theta^0)\}$.

Recall that the definition in (6) allows for the convention that $\inf_{\gamma_0\in CI_T(\gamma_S;\epsilon)} LM_T\left(A_S^{-1}(r_0',\gamma_0')'\right) = \infty$ if $CI_T(\gamma_S;\epsilon)$ is empty. Now, let $\{\phi_{\beta,T}: T\geq 1\}$ denote the sequence of indicator variables where $\phi_{\beta,T}=1$ if $\inf_{\gamma_0\in CI_T(\gamma_S;\epsilon)} LM_T\left(A_S^{-1}(r_0',\gamma_0')'\right) > \chi^2_{d_R}(1-\alpha)$, and $\phi_{\beta,T}=0$ otherwise. Therefore,

$$\begin{aligned}
& Pr_T\left(\phi_{\beta,T}=0\right) \\
= \; & Pr_T\left(\inf_{\gamma_0\in CI_T(\gamma_S;\epsilon)} LM_T\left(A_S^{-1}(r_0',\gamma_0')'\right) \leq \chi^2_{d_R}(1-\alpha)\right) \\
\geq \; & Pr_T\left(\left\{\inf_{\gamma_0\in CI_T(\gamma_S;\epsilon)} LM_T\left(A_S^{-1}(r_0',\gamma_0')'\right) \leq LM_T(\theta^0)\right\} \bigcap \left\{LM_T(\theta^0) \leq \chi^2_{d_R}(1-\alpha)\right\}\right) \\
= \; & 1 - Pr_T\left(\left\{\inf_{\gamma_0\in CI_T(\gamma_S;\epsilon)} LM_T\left(A_S^{-1}(r_0',\gamma_0')'\right) > LM_T(\theta^0)\right\} \bigcup \left\{LM_T(\theta^0) > \chi^2_{d_R}(1-\alpha)\right\}\right) \\
\geq \; & 1 - \left(Pr_T\left(\inf_{\gamma_0\in CI_T(\gamma_S;\epsilon)} LM_T\left(A_S^{-1}(r_0',\gamma_0')'\right) > LM_T(\theta^0)\right) + Pr_T\left(LM_T(\theta^0) > \chi^2_{d_R}(1-\alpha)\right)\right),
\end{aligned}$$

where the second line follows by the definition of $\phi_{\beta,T}$, the third line by the construction of the two-step projection test in (6), the fourth line by De Morgan's law, and the fifth line by Bonferroni's inequality. Taking limits on both sides gives:

$$\begin{aligned}
\lim_{T\to\infty} Pr_T\left(\phi_{\beta,T}=0\right) \;\geq\; & 1 - \lim_{T\to\infty} Pr_T\left(\inf_{\gamma_0\in CI_T(\gamma_S;\epsilon)} LM_T\left(A_S^{-1}(r_0',\gamma_0')'\right) > LM_T(\theta^0)\right) \\
& - \lim_{T\to\infty} Pr_T\left(LM_T(\theta^0) > \chi^2_{d_R}(1-\alpha)\right) \\
\geq\; & 1 - (\epsilon+\alpha),
\end{aligned}$$

where the last line follows by (37) and Lemma 3. ■

**Remark 23:** Since the way it is stated in the statement of the proposition, the coverage probability of $CI_T(\gamma_S;\epsilon)$ is $(1-\epsilon)$, possibly under a larger class of distributions than $F_T$ constrained by the assumptions O and M1-M3. This is the reason behind the inequality $\lim_{T\to\infty} Pr_T(\phi_{\gamma_S,T}=0) \geq (1-\epsilon)$. However, the confidence sets $CI_T(\gamma_S;\epsilon)$, e.g., $CI_T^{SW}(\gamma_S;r_0,\epsilon)$ defined in (10), that we actually mention [see Remark 2] are asymptotically similar and hence, for them, the above inequality will hold as an equality. ■

**Proof of Lemma 5:** (a) Utilizing the nonsingular matrices $\Pi_{\rho_\theta}$, $D_{T,\rho_\theta}$, $\Pi_R$ and $D_{T,R}$ in (18),(19), (22) and (23) respectively, recall from (4) and (8) that $LM_T(\theta)$ and $LM_T^{infs}(\theta)$ can be written as:

$$\begin{aligned} LM_T(\theta) &= T \times \left(\widehat{V}_T^{-1/2}(\theta)\bar{g}_T(\theta)\right)' P\left(\widehat{H}_T(\theta)(\widehat{H}_T'(\theta)\widehat{H}_T(\theta))^{-}D_{T,\rho\theta}\Pi_{\rho\theta}' R'\Pi_R D_{T,R}\right)\left(\widehat{V}_T^{-1/2}(\theta)\bar{g}_T(\theta)\right), \\ LM_T^{infs}(\theta) &= T \times \left(V_T^{-1/2}(\theta)\bar{g}_T(\theta)\right)' P\left(H_T(\theta)(H_T'(\theta)H_T(\theta))^{-}D_{T,\rho\theta}\Pi_{\rho\theta}' R'\Pi_R D_{T,R}\right)\left(V_T^{-1/2}(\theta)\bar{g}_T(\theta)\right), \end{aligned}$$

where $\widehat{H}_T(\theta) := \widehat{V}_T^{-1/2}(\theta)\widehat{G}_T(\theta)\Pi_{\rho\theta}D_{T,\rho\theta}$ and $H_T(\theta) := V_T^{-1/2}(\theta)E_T[\bar{G}_T(\theta)]\Pi_{\rho\theta}D_{T,\rho\theta}$ respectively. ($\widehat{H}_T(\theta^0)$ or $H_T(\theta^0)$ are not the same as $H_T$ in the proof of Lemma 3.) Essentially $LM_T^{infs}(\theta)$ is $LM_T^{infs}(\theta_{.,S}^{infs})$, but without plugging in $\theta_{.,S}^{infs}$ in place of the general $\theta$. Now recall that for $\theta_T$ defined in (15), we have:

(i) $\widehat{V}_T^{-1/2}(\theta_T) \xrightarrow{P} V^{-1/2}$ by N8 and $V_T^{-1/2}(\theta_T) \to V^{-1/2}$ by definition [also see N8];

(ii) $\widehat{V}_T^{-1/2}(\theta_T)\sqrt{T}\bar{g}_T(\theta_T) = V^{-1/2}[\sqrt{T}\bar{g}_T(\theta^0) + G^*\mu_{T,\theta}] + o_p(1)$ by (i) and Lemma 11(c); and this is $O_p(1)$ by N2 and N8;

(iii) $D_{T,\rho\theta}\Pi_{\rho\theta}' R'\Pi_R D_{T,R} \to R^{*'}$ by (24).

Therefore, to show that $LM_T(\theta_T) = LM_T^{infs}(\theta_T) + o_p(1)$, it suffices to show that $\widehat{H}_T(\theta_T) - H_T(\theta_T) = o_p(1)$. Thus, by virtue of (i) and Lemma 11(f), it suffices to show that $E_T[\bar{G}_T(\theta_T)]\Pi_{\rho\theta}D_{T,\rho\theta} \to G^*$. Hence, by virtue of (21), it is now sufficient to show that $E_T[\bar{G}_T(\theta_T) - \bar{G}_T(\theta^0)]\Pi_{\rho\theta}D_{T,\rho\theta} = o(1)$. This follows exactly by proceeding from (31) onward in the proof of Lemma 11 simply by replacing $\bar{G}_T(.)$ in that proof with $E_T[\bar{G}_T(.)]$. Thus, $LM_T(\theta_T) = LM_T^{infs}(\theta_T) + o_p(1)$.

Now, using (ii), (iii), N4, N8 and the fact that we just established $H_T(\theta_T) \to V^{-1/2}G^*$, note that:

$$\begin{aligned} & \left(H_T(\theta_T)(H_T'(\theta_T)H_T(\theta_T))^{-}D_{T,\rho\theta}\Pi_{\rho\theta}' R'\Pi_R D_{T,R}\right)'\left(V_T^{-1/2}(\theta_T)\sqrt{T}\bar{g}_T(\theta_T)\right) \\ =\ & R^*(G^{*'}V^{-1}G^*)^{-1}G^{*'}V^{-1}[\sqrt{T}\bar{g}_T(\theta^0) + G^*\mu_{T,\theta}] + o_p(1) \\ =\ & R^*(G^{*'}V^{-1}G^*)^{-1}G^{*'}V^{-1}\sqrt{T}\bar{g}_T(\theta^0) + R^*\mu_{T,\theta} + o_p(1) \\ =\ & R^*(G^{*'}V^{-1}G^*)^{-1}G^{*'}V^{-1}\sqrt{T}\bar{g}_T(\theta^0) + \mu_\beta + o_p(1), \end{aligned}$$

by using the relation $R^*\mu_{T,\theta} \xrightarrow{P} \mu_\beta$ (see below (15)). Therefore, it follows that the RHS on the last line does not depend on $\gamma_{S,T}$ at all as long as $\gamma_{S,T}$ is such that (15) holds.[18] Note that $\gamma_S^0$, a constant for all $T$, is trivially such a choice of the sequence $\{\gamma_{S,T} : T \geq 1\}$. Thus,

$$LM_T^{infs}(\theta_{0,S}^{infs}) = LM_T^{infs}(\theta_T) + o_p(1) = LM_T(\theta_T) + o_p(1).$$

(b) From (a), now it directly follows that $LM_T(\theta_T) \xrightarrow{d} \chi^2_{d_R}$ with non-centrality parameter given by $\mu_\beta'\left(R^*(G^{*'}V^{-1}G^*)^{-1}R^{*'}\right)^{-1}\mu_\beta$. ■

[18](15) is an important qualifier for this statement and the next. See Remarks 17-20.

**Proof of Lemma 6:** Define the sequence $\{\gamma_T^\dagger : T \geq 1\}$ such that:

$$\gamma_T^\dagger := \arg \inf_{\gamma_0 \in CI_T(\gamma_S;\epsilon)} LM_T\left(A_S^{-1}(r_0', \gamma_0')'\right).$$

By condition (16) on $CI_T(\gamma_S;\epsilon)$, it then follows that $\gamma_T^\dagger$ gives $\theta_T^\dagger = R_S^1 r_0 + S_S^1 \gamma_T^\dagger$, for which $\sqrt{T} D_{T,\rho_\theta}^{-1} \Pi_{\rho_\theta}'(\theta_T^\dagger - \theta^0) = O_p(1)$, i.e., (15) holds since (14) also holds. Therefore, the final result follows by Lemma 5. ■

**Proof of Lemma 7:** The lemma defines the supremum in case of an empty $CI_T^{SW}(\gamma_S; r_0, \epsilon)$ in a way that allows us to ignore those cases in the sequel, with the caveat from the discussion above the lemma.

Now, the proof follows in three steps. In Step 1 we show that $CI_T^{SW}(\gamma_S; r_0, \epsilon)$ shrinks to $\gamma_S^0$ in probability; more precisely, that the Hausdorff distance between $\gamma_S^0$ and $CI_T^{SW}(\gamma_S; r_0, \epsilon)$ converges in probability to zero. Using this, in Step 2 we obtain that this rate cannot be slower than $\lambda_{T,1}$. Using this, in Step 3 we obtain the final result. Details, once stated, are not repeated in the subsequent steps.

Step 1: Take any constant $\varpi > 0$. Consider the outside of the open ball of radius $\varpi$ around $\gamma_S^0$. Then, for some $\lambda_T^* \in \{\lambda_{T,1}, \ldots, \lambda_{T,l}\}$, it follows by N1 (well separated $\theta^0$) and N3 (elements of $\Lambda_T$) that:

$$\liminf_T \inf_{\beta \in \mathcal{B}, \gamma \in \Gamma_S : \|\gamma - \gamma_S^0\| \geq \varpi} \left\| \frac{\sqrt{T}}{\lambda_T^*} E_T[\bar{g}_T(R_S^1 \beta + S_S^1 \gamma)] \right\| \equiv \liminf_T \inf_{\beta \in \mathcal{B}, \gamma \in \Gamma_S : \|\gamma - \gamma_S^0\| \geq \varpi} \left\| \frac{\Lambda_T}{\lambda_T^*} \rho(R_S^1 \beta + S_S^1 \gamma)] \right\| > 0.$$

Taken together with N8 ($\sup_{\theta \in \Theta} \min[\text{eigen values}(V^{-1}(\theta))] > 0$), this gives:

$$\liminf_T \inf_{\beta \in \mathcal{B}, \gamma \in \Gamma_S : \|\gamma - \gamma_S^0\| \geq \varpi} \left\| V^{-1/2}(R_S^1 \beta + S_S^1 \gamma) \frac{\sqrt{T}}{\lambda_T^*} E_T[\bar{g}_T(R_S^1 \beta + S_S^1 \gamma)] \right\| > 0.$$

Now, note that:

$$V^{-1/2}(\theta) \frac{\sqrt{T}}{\lambda_T^*} \bar{g}_T(\theta) = V^{-1/2}(\theta) \frac{\sqrt{T}}{\lambda_T^*} \left(\bar{g}_T(\theta) - E_T[\bar{g}_T(\theta)]\right) + V^{-1/2}(\theta) \frac{\sqrt{T}}{\lambda_T^*} E_T[\bar{g}_T(\theta)].$$

By N2, N3 and N8, the first term on the RHS is $o_p(1)$ uniformly in $\theta \in \Theta$. Therefore, by using the uniform consistency of $\widehat{V}_T^{-1}(\theta)$ for $V^{-1}(\theta)$ from N8, and the definition of $Q_T(\theta)$ from (11), it follows that

$$\lim_T Pr_T\left(\inf_{\beta \in \mathcal{B}, \gamma \in \Gamma_S : \|\gamma - \gamma_S^0\| \geq \varpi} (\lambda_T^*)^{-2} \times T \times Q_T(R_S^1 \beta + S_S^1 \gamma) > c \text{ for some } c > 0\right) = 1,$$

$$\text{and hence,} \quad \lim_T Pr_T\left(\inf_{\beta \in \mathcal{B}, \gamma \in \Gamma_S : \|\gamma - \gamma_S^0\| \geq \varpi} T \times Q_T(R_S^1 \beta + S_S^1 \gamma) > \bar{c} \text{ for all } \bar{c} < \infty\right) = 1,$$

$$\text{and hence,} \quad \lim_T Pr_T\left(\inf_{\gamma \in \Gamma_S : \|\gamma - \gamma_S^0\| \geq \varpi} T \times Q_T(R_S^1 r_0 + S_S^1 \gamma) > \bar{c} \text{ for all } \bar{c} < \infty\right) = 1. \tag{38}$$

The second line follows since $\liminf_T \lambda_T^* = \infty$ by N3. The third line follows since $r_0 \in \mathcal{B}$ for large $T$.

Since $\varpi > 0$ is arbitrary, by the definition in (10) where the critical value is a fixed, finite positive number for a given $\epsilon < 1 - \alpha$, it follows from (38) that:[19]

$$\sup_{\gamma_0 \in CI_T^{SW}(\gamma_S; r_0, \epsilon)} \|\gamma_0 - \gamma_S^0\| = o_p(1).$$

Step 2: Take any constant $\varpi > 0$. Define $\{\Gamma_T(\varpi) : T \geq 1\}$, shrinking at rate slower than $\lambda_{T,1}$, as:

$$\begin{aligned}\Gamma_T(\varpi) \;&:=\; \{\gamma \in \Gamma_S : a_T\|\gamma - \gamma_S^0\| \leq \varpi \text{ and } \lambda_{T,1}\|\gamma - \gamma_S^0\| \geq b_T \text{ for some positive sequences} \\ &\qquad \{a_T : T \geq 1\} \text{ and } \{b_T : T \geq 1\} \text{ with } a_T \to \infty, b_T \to \infty \text{ as } T \to \infty\}.\end{aligned}$$

Consider a sequence $\{\gamma_T : T \geq 1\}$ such that $\gamma_T = \arg\inf_{\gamma \in \Gamma_T(\varpi)} T \times Q_T(R_S^1 r_0 + S_S^1 \gamma)$ for each $T \geq 1$. (The sequence need not be unique.) Hence, $\|\gamma_T - \gamma_S^0\| = o(1)$ (although $\lim_T \lambda_{T,1}\|\gamma_T - \gamma_S^0\| = \infty$). Also, $\lim_T \lambda_{T,1}\|r_0 - \beta^0\| < \infty$ by N3 and (14). Therefore, $\theta_T := R_S^1 r_0 + S_S^1 \gamma_T \in \mathcal{N}(\theta^0)$ (as in N4) for large $T$. This gives, by a mean value expansion of $\rho(\theta_T)$ around $\rho(\theta^0)$ with mean value $\bar{\theta}_T$ (element by element), that:

$$\lambda_{T,1}\Lambda_T^{-1}\sqrt{T}E_T[\bar{g}_T(\theta_T)] = 0 + \left[\rho_\theta(\bar{\theta}_T)R_S^1\right]\lambda_{T,1}(r_0 - \beta^0) + \left[\rho_\theta(\bar{\theta}_T)S_S^1\right]\lambda_{T,1}(\gamma_T - \gamma_S^0)$$

by using N1. N4 and Lemma 10 imply that the terms inside the squared brackets on the RHS are full column-rank. Hence, the second term on the RHS is $O(1)$ whereas the third term diverges to $\pm\infty$. Since $\liminf_T \Lambda_T/\lambda_{T,1} > 0$ by N3, it follows that $\lim_T \sqrt{T}\|E_T[\bar{g}_T(\theta_T)]\| = \infty$. Hence, using the definitions of $\gamma_T$ and $\Gamma_T(\varpi)$ in conjunction with the result of Step 1, the same arguments as in Step 1 now give:

$$\sup_{\gamma_0 \in CI_T^{SW}(\gamma_S; r_0, \epsilon)} \lambda_{T,1}\|\gamma_0 - \gamma_S^0\| = O_p(1).$$

Step 3: Equipped with the result from Step 2, now we further refine this rate as follows. Define

$$\begin{aligned}\Gamma_T \;&:=\; \{\gamma \in \Gamma_S : \lambda_{T,1}\|\gamma - \gamma_S^0\| < b_T \text{ for any positive sequence } \{b_T : T \geq 1\} \text{ with } b_T \to \infty, \\ &\qquad \|\sqrt{T}D_{T,\rho_\theta}^{-1}\Pi_{\rho_\theta}'(R_S^1(r_0 - \beta^0) + S_S^1(\gamma - \gamma_S^0))\| \geq a_T \text{ for some positive sequence} \\ &\qquad \{a_T : T \geq 1\} \text{ with } a_T \to \infty, \text{ and where } r_0 \text{ is as defined in (14)}\}.\end{aligned}$$

Consider a sequence $\{\gamma_T : T \geq 1\}$ such that $\gamma_T = \arg\inf_{\gamma \in \Gamma_T} T \times Q_T(R_S^1 r_0 + S_S^1 \gamma)$ for each $T \geq 1$. Note that, $\lim_T \lambda_{T,1}\|\gamma_T - \gamma_S^0\| < \infty$. Also, $\lim_T \lambda_{T,1}\|r_0 - \beta^0\| < \infty$ by N3 and (14). Therefore, for

[19] This argument is used at the end of all the three steps in the proof of this lemma. A rigorous version of this argument is presented (for the first time, to our knowledge) in the proof of Lemma 13.2 in Andrews (2017a). The proof of Lemma 7, on the other hand, focuses primarily on establishing an appropriate (for us) version of what Andrews (2017a) assumes as the global strong-identification condition for the nuisance parameter. See footnote 6.

$\theta_T := R_S^1 r_0 + S_S^1 \gamma_T$, it follows that $\lim \lambda_{T,1}\|\theta_T - \theta^0\| < \infty$, giving $\theta_T \in \mathcal{N}(\theta^0)$ (as in N4) for large $T$. Furthermore, since $\mu_\beta \neq 0$, $D_{T,\rho_\theta}^{-1}$ and $\Pi'_{\rho_\theta}$ are nonsingular, and $R_S^1$ and $S_S^1$ are full column-rank, it follows that $\eta_T := \sqrt{T} D_{T,\rho_\theta}^{-1} \Pi'_{\rho_\theta}(\theta - \theta^0) \neq 0$ and, by the definition of $\Gamma_T$, that $\lim_T \|\eta_T\| = \infty$.

To proceed, first we use (13) and N1, and by a mean value expansion of $\rho(\theta_T)$ around $\rho(\theta^0)$ with mean value $\bar{\theta}_T$ (element by element), we obtain that:

$$\begin{aligned}\sqrt{T}\bar{g}_T(\theta_T) &= \sqrt{T}\left(\bar{g}_T(\theta_T) - E_T[\bar{g}_T(\theta_T)]\right) + \Lambda_T\left\{\rho(\theta^0) + \rho_\theta(\theta^0)(\theta_T - \theta^0) + [\rho_\theta(\bar{\theta}_T) - \rho_\theta(\theta^0)](\theta_T - \theta^0)\right\} \\ &= \psi_T(\theta_T) + \left(\frac{\Lambda_T}{\sqrt{T}}\rho_\theta(\theta^0)\Pi_{\rho_\theta} D_{T,\rho_\theta}\right)\eta_T + \phi_T(\theta_T)\end{aligned}$$

using $\rho(\theta^0) = 0$ [see N1], and where $\psi_T(\theta_T)$, as defined in N2, is $O_p(1)$ using N8, while

$$\phi_T(\theta_T) := \Lambda_T[\rho_\theta(\bar{\theta}_T) - \rho_\theta(\theta^0)](\theta_T - \theta^0) = O_p\left(\frac{\lambda_{T,l}}{\lambda_{T,1}^2}\right) = o_p(1).$$

The first equality for $\phi_T(\theta)$ follows since $\rho(\theta)$ is twice continuous differentiable in $\mathcal{N}(\theta^0)$ [see N7(a)], and using N3 and the fact that $\lim \lambda_{T,1}\|\theta_T - \theta^0\| < \infty$ (as noted above). The second one uses N7(b).[20] Hence:

$$\frac{\sqrt{T}}{\|\eta_T\|}\bar{g}_T(\theta_T) = \frac{1}{\|\eta_T\|}\psi_T(\theta_T) + G^* \frac{\eta_T}{\|\eta_T\|} + o_p\left(\frac{1}{\|\eta_T\|}\right).$$

Hence, as in Step 1 and 2, by using N8, N1, the finiteness and full column-rank of $G^*$ [see (21)], and that $\eta_T \neq 0$ while $\lim_T \eta_T = \infty$, it follows that $T \times Q_T(\theta_T)$ diverges to $\infty$ in probability. Therefore, using the definitions of $\gamma_T$, $\theta_T$ and $\Gamma_T$ in conjunction with the result of Step 2, the same arguments as in Step 1 now give the final result of the lemma:

$$\sup_{\gamma_0 \in CI_T^{SW}(\gamma_S; r_0, \epsilon)} \sqrt{T}\left\| D_{T,\rho_\theta}^{-1}\Pi'_{\rho_\theta}\left((R_S^1(r_0 - \beta^0) + S_S^1(\gamma_0 - \gamma_S^0)\right)\right\| = O_p(1). \blacksquare$$

**Proof of Proposition 8:** The proof is omitted since it is exactly same as that of Theorem 3.2(ii). ∎

[20]Recall from Remark 12 that under the Stock and Wright (2000) setup, $\bar{\Pi}_{\bar{\rho}_\theta} = I_{d_\theta}$ and $\bar{D}_{T,\bar{\rho}_\theta} = \sqrt{T}\bar{\Lambda}_T^{-1}$. Hence, $\eta_T = \bar{\Lambda}_T(\theta_T - \theta^0)$. Hence, similar to above, an expansion of $\sqrt{T}\bar{g}_T(\theta_T)$ under Stock and Wright (2000)'s setup gives:

$$\sqrt{T}\bar{g}_T(\theta_T) = \psi_T(\theta_T) + \left(\bar{\rho}_\theta(\theta^0)\frac{\bar{\Lambda}_T}{\sqrt{T}}\bar{\Pi}_{\rho_\theta}\bar{D}_{T,\rho_\theta}\right)\eta_T + \phi_T(\theta_T) = \psi_T(\theta_T) + \bar{\rho}_\theta(\theta^0)\eta_T + \phi_T(\theta_T)$$

where, using the same arguments as in the main text but *without using N7(b)*, it follows that:

$$\phi_T(\theta_T) := [\bar{\rho}_\theta(\bar{\theta}_T) - \bar{\rho}_\theta(\theta^0)]\bar{\Lambda}_T(\theta_T - \theta^0) = O_p\left(\frac{1}{\bar{\lambda}_{T,1}}\right)\eta_T, \quad \text{i.e.,} \quad \|\phi_T(\theta_T)\| = O_p\left(\frac{\|\eta_T\|}{\bar{\lambda}_{T,1}}\right).$$

Therefore, $\frac{\sqrt{T}}{\|\eta_T\|}\bar{g}_T(\theta_T) = \frac{1}{\|\eta_T\|}\psi_T(\theta_T) + G^*\frac{\eta_T}{\|\eta_T\|} + O_p\left(\frac{1}{\bar{\lambda}_{T,1}}\right)$. Since weak identification is not allowed, i.e., $\bar{\lambda}_{T,1} \to \infty$, and since $G^*$ is full column-rank, this means that the rest of the steps in the main text of the proof can follow without change.

## Appendix D: Unrestricted-by-$H_0$ plug-in is not advisable

In reference to the key feature (F3) in Section 2, we conduct the same simulation study as in Section 4.2.3. However, instead of the standard plug-in tests, now we consider the unrestricted version of the plug-in tests that replace $\gamma_S$ in $LM_T(r_0, \gamma_S)$ by its unrestricted-by-$H_0$ CU-GMM estimator $\widetilde{\gamma}_S := S\widetilde{\theta}$, where $\widetilde{\theta} := \arg\min_{\theta\in\Theta} Q_T(\theta)$ and $Q_T(.)$ is as defined in (11). Here, $\widetilde{\gamma}_S$ is the unrestricted LIML estimator.

Plots similar to Figure 1 are now reported in Figure 2.

As evident from Figure 2, the intuition in (F3) is confirmed by the simulation results for the unrestricted plug-in tests. These tests are no longer invariant to $S$, and now behave in a way that resembles the behavior of the corresponding $S$-dependent infeasible tests when $H_0$ is false. This means that the unrestricted plug-in test has better power when the $\gamma_S$ with worse identification strength is used for the nuisance parameter. However, this comes at the cost of severe over-rejection of the true $H_0$ under specifications (ii) and (iii), where the standard plug-in test actually did not display over-rejection in Section 4.2.3 [see Figure 1]. To be clear, neither specification falls under the scope of our results, although Theorem 6 of Guggenberger and Smith (2005) indicates that the standard plug-in test (as in Section 4.2.3) would have had correct asymptotic size under specification (iii), i.e., when $\lambda_{T,1} = 1$ and $\lambda_{T,2} = \sqrt{T}$ and hence the nuisance parameters are strongly identified.

To avoid clutter, we do not report in Figure 2 the results for the unrestricted version of the two-step projection test. (They are available from the author.) But it should be mentioned that we find its empirical power to be extremely poor except under specification (vi) and, to a small extent, under specification (v). The problem of poor power can be roughly attributed to the all too frequent occurrences of very large and even unbounded[21] first-step confidence set when the sample size $T$ is small. It happens because this confidence set does not take advantage of the false $H_0$ by imposing it. The problem of poor power does not go away even in simulations with very large $T$, under the three specifications (i)-(iii) that involve at least one weakly identified component (and hence, outside the scope of Section 4.2).

However, we do find that this problem of poor power essentially disappears if, e.g., $T = 2000$, for all the three specifications (iv)-(vi) that are under the scope of our results in Section 4.2. In other words, under these three specifications (iv)-(vi), as in Figure 1, the two-step test behaves similar to the less powerful infeasible and plug-in tests that use the better identified $\gamma_S$ as the nuisance parameter. Notwithstanding, the intuition from (F3) is not applicable to the two-step test since it remains invariant to $S$, and, therefore, resembling the corresponding $S$-dependent infeasible test is not a possibility. In summary, as noted in (F3), we do not recommend this strategy for the two-step projection test.

[21]Since this a linear model, following convention, we did not impose compactness of $\Theta$ (or, $\mathcal{B}$ and $\Gamma_S$) in our computation.

Figure 2: Empirical rejection probabilities of the infeasible test (infeas) in (8), and the unrestricted version of the plug-in test (LIML (unres)) based on the unrestricted CU-GMM (in this case, unrestricted LIML) estimator for $\gamma_S$. Two choices of $\gamma_S$, i.e., $\gamma_S = \theta_1$ and $\gamma_S = \theta_2$ are employed for the infeasible test and the unrestricted version of the plug-in test. For all tests, we take $\alpha = .045$. Results are based on 10,000 Monte Carlo trials. Horizontal axis: deviation of $H_0$ in (2) from truth [also see (14)]. Title: Identification strength that corresponds to specifications (i)-(vi) respectively.