Near Exogeneity and Weak Identification in Generalized Empirical Likelihood Estimators: Many Moment Asymptotics

Mehmet Caner
North Carolina State University *

May 5, 2014

Abstract

This paper investigates the Generalized Empirical Likelihood (GEL) Estimators when there are local violations of the exogeneity condition (near exogeneity) in the case of many weak moments. We also examine the tradeoff between the degree of violation of the exogeneity and the number of nearly exogenous instruments. In this respect, this paper extends many weak moment asymptotics of Newey and Windmeijer (2009a). The overidentifying restrictions test can detect both mild and large violations of exogeneity. In the case of minor violations, the Anderson-Rubin (1949) and Wald tests are not size distorted.

Keywords: Violation of Exogeneity, Anderson-Rubin test, Asymptotic Size

JEL CODES: C3, C13, C30.

*Department of Economics, 4168 Nelson Hall, Campus Box 8110, Raleigh, NC, 27695. email: mcaner@ncsu.edu.

We thank Co-Editor Cheng Hsiao, the associate editor, and two anonymous referees for the comments that improved the paper substantially. We also thank Lauren Jones for professional text editing for this article.
1 Introduction

The issue of violation of the exogeneity condition in instrumental variable estimation is one important problem that must be addressed. We define near exogeneity as the local to zero violation of the perfect exogeneity condition. Selection of a perfectly exogenous instrument is often difficult. Note that this violation of the exogeneity condition in Generalized Method of Moments (GMM) setting is discussed by Newey (1985), and Hall and Inoue (2003) for the case of a fixed number of strong instruments. Bound, Jaeger, and Baker (1995) demonstrate that minor violations of the exogeneity condition, coupled with weak instruments, increases the bias in the coefficient estimates. We think that many instruments setup is a natural place to find nearly exogenous instruments. A key issue is also the interaction of weak identification with near exogeneity in many instruments setup. In related literature, Newey and Windmeijer (2009a) develop many weak moment asymptotics for Generalized Empirical Likelihood (GEL) Estimators with perfectly exogenous instruments. The new limit has larger asymptotic variance than the standard GEL limit of Newey and Smith (2004). This approximation improves the finite sample results. Chao and Swanson (2005) derive the linear case, and Han and Phillips (2006) consider GMM. Several recent papers have explored the testing of exogeneity violations. Guggenberger (2012) analyzes the size distortion of various tests in fixed number of instruments with a linear setup and exogeneity violations. He concludes that the Anderson-Rubin (1949) test is less size distorted than the other tests. Berkowitz, Caner, and Fang (2012) provide a new resampling technique when there are local exogeneity violations. Also, Caner and Morrill (2010) tackle the inference problem of strong but invalid instruments. They show that a joint test of structural parameters and correlation parameters may be useful. Finally, in a working paper, Kolesar et al. (2011) analyze estimation when there are many invalid instruments in a linear context. Our paper goes in a different direction and analyzes the tests in the case of many weak and nearly exogenous instruments.

This article extends the previous literature in several ways. First, we allow many weak and nearly exogenous instruments. Their number may be equal to the total number of instruments. Second, we analyze various degrees of the violation of exogeneity, unlike the root $n$ case in the previous literature. We identify the tradeoff between the degrees of violation and the number of nearly exogenous instruments. Next, we show that the overidentifying restrictions test can detect both mild and large violations. Note that the Anderson-Rubin (1949) test is not affected by minor violations.

We provide assumptions and the limits of the GEL Estimators in Section 2. In Section 3, we discuss tests under the condition of many weak moments and near exogeneity. In Section 4, we conduct several simulations. Conclusion is in Section 5. The Appendix covers the proofs, including the Supplement Appendix which provides the details of the proofs in Section 2.
2 Many Weak Moment Asymptotics

The model is

\[ Eg(Z_i, \theta_0) = \frac{C_1}{n^\kappa}, \] (1)

where \(C_1\) is a \(q_n \times 1\) vector of constants, \(0 < \kappa < \infty\), and \(\theta_0\) is the true structural parameter vector of dimension \(p\). The data \(\{Z_i\}_{i=1}^n\) is iid. The elements of \(C_1\) are not necessarily zeros as in the standard model, and are in a compact set. This is specified below in Assumption M.1. The setup is a generalization of Hall and Inoue (2003), and Newey (1985), where they impose \(\kappa = 1/2\). \(E(\cdot)\) denotes the expectation taken with respect to \(Z_i\) for sample size \(n\), we suppress the subscript \(n\).

In section 3.2, we will generalize equation (1). For our analysis, we will use (1) as the basis for Assumption M.1.

The GEL estimator is defined as in Newey and Windmeijer (2009a): set \(g_i(\theta) = g(Z_i, \theta)\), for all \(i = 1, \cdots, n\),

\[ \hat{\theta} = \arg \min_{\theta \in \Theta} \sup_{\lambda \in \hat{\Lambda}_n(\theta)} \frac{1}{n} \sum_{i=1}^n \rho(\lambda' g_i(\theta))/n. \]

Let \(\rho(.)\) be a real valued function, \(V \to \mathbb{R}\), where \(V\) is an open interval of the real line that contains zero, and \(\hat{\Lambda}_n(\theta) = \{ \lambda : \lambda' g_i(\theta) \in V \text{ for } i = 1, \cdots, n \}\). Also, we set \(\theta \in \Theta\), where \(\Theta\) is a compact subset of \(R^p\), and define \(\rho_j(\nu) = \partial^j \rho(\nu)/\partial \nu^j\), with \(\rho_j = \rho_j(0)\) for nonnegative integers \(j\). We want to estimate the unknown \(\theta_0\), which is the true parameter vector. We normalize \(\rho(\nu)\), \(\nu \in V\) so that \(\rho(0) = 0, \partial \rho(0)/\partial \nu = -1, \partial^2 \rho(0)/\partial \nu^2 = -1\). The moment function \(g_i(\theta)\) is of the dimension \(q_n \times 1\), and \(q_n\) increases with \(n\). The relationship between \(q_n\) and \(n\) will be explained in Assumptions, but \(q_n/n \to 0 \text{ as } n \to \infty\). Therefore \(q_n\) will grow slower than \(n\). This is the approach taken by Newey and Windmeijer (2009a) to control the dimension of the variance term.

GEL consists of several interesting sub cases. There are specifically three estimators, that we use in econometrics. The first one is Empirical Likelihood estimator of Owen (2001), Qin and Lawless (1994), Imbens (1997). This is obtained from GEL when we set \(\rho(\nu) = ln(1 - \nu), V = (-\infty, 1)\). Next, we have the Exponential Tilting estimator of Kitamura and Stutzer (1997), where we set \(\rho(\nu) = -e^\nu + 1\) in the GEL estimator. The last one is Continuous Updating estimator, where we set \(\rho(\nu) = -\nu - \nu^2/2\), and the objective function has GMM like form

\[ \hat{Q}(\theta) = \frac{1}{2} (n^{-1} \sum_{i=1}^n g_i(\theta))' [n^{-1} \sum_{i=1}^n g_i(\theta) g_i(\theta)']^{-1} (n^{-1} \sum_{i=1}^n g_i(\theta)), \]

which is shown in Newey and Smith (2004).

We can rewrite the GEL estimator in the following way. Denote for each \(\theta \in \Theta\)

\[ \hat{Q}(\theta) = \sup_{\lambda \in \hat{\Lambda}_n(\theta)} \frac{1}{n} \sum_{i=1}^n \rho(\lambda' g_i(\theta))/n, \]
and

\[ \hat{\theta} = \arg \min_{\theta \in \Theta} \hat{Q}(\theta). \]

Newey and Windmeijer (2009a) present detailed explanations about why the GEL estimator is consistent under many weak asymptotics, whereas GMM cannot be consistent. The limit of the objective function in GMM consists of a “noise” term and a “signal” term. The noise term consists of a weight matrix multiplied by \( \Omega(\theta) = E g_i(\theta) g_i(\theta)' \). This noise does not disappear in large samples and contaminates the limit, which leads to inconsistency. This issue is shown by Han and Phillips (2006) and, subsequently, by Newey and Windmeijer (2009a). However, the noise term in the Continuous Updating Estimator (CUE) in GEL does not depend on \( \theta \) since the weight matrix is \( \Omega(\theta)^{-1} \). The CUE is consistent under many weak moment asymptotics. Since Newey and Windmeijer (2009a) have shown the GEL objective function is well approximated by CUE, any GEL estimator is also consistent.

There are several reasons that we use GEL rather than the two-step GMM. First, exogeneity violations in a two-step GMM are analyzed in Hall and Inoue (2003), where there are strong and valid instruments. Second, the two-step GMM is inconsistent in the many weak moments case. Note that GEL estimators are consistent even when there are many weak moments with near exogeneity. They are, therefore, more robust to the problems in data compared with GMM. But we also see that the asymptotic limit described by Newey and Windmeijer (2009a) may change when we have nearly exogenous instruments.

2.1 Assumptions

We start with the near exogeneity assumption.

**Assumption M.1.** (i)

\[ E g_i(\theta_0) = \frac{C_1}{n^\kappa}, \]

where \( 0 < \kappa < \infty \) and \( C_1 \) is a \( q_n \times 1 \) vector. \( C_1 = (0_{q_n-l_n}^\prime, C_{l_n}^\prime)^\prime \), \( C_{l_n} \) is an \( l_n \times 1 \) vector, where \( l_n \to \infty \) when \( n \to \infty \). For each \( j = 1, \ldots, l_n \), \(-\infty < C_a < C_{l_n,j} < C_b < \infty \), where \( C_a, C_b \) are scalars. \( C_{l_n,j} \) is in a compact set \( S \). \( 0_{q_n-l_n} \) represents a zero vector of the dimension \( q_n-l_n \).

We allow for two possibilities regarding the ratio of the number of imperfect moment conditions \( l_n \) to total number of moment conditions \( q_n \):

(ii) Let \( l_n/q_n \to f \), as \( n \to \infty \), where \( 0 < f \leq 1 \),

or

(iii) Let \( l_n/q_n \to 0 \), as \( n \to \infty \).

Assumption M.1(i) expresses a very general form of violation of exogeneity. \( \kappa = 1/2 \) is a mild violation of exogeneity, but we consider the case of \( 1/2 < \kappa < \infty \), a minor violation of exogeneity. We consider \( 0 < \kappa < 1/2 \) as the range of major violation of perfect exogeneity condition. This approach is more general than the setups by Newey (1985), Hall and Inoue (2003), and Berkowitz,
Caner, and Fang (2008, 2012), where there are fixed number of invalid instruments \( (l_n = l, \text{ and } l \text{ is constant.}) \) and \( \kappa = 1/2 \). Note that we cannot allow for \( \kappa = 0 \), because it would violate one of the conditions needed for consistency. Assumptions M.1(ii) and (iii) illustrate two distinct possibilities between the ratio of imperfect moments \( (l_n) \) to the total number of moments \( (q_n) \). The first possibility is that the number of imperfect moments can be a positive fraction of all orthogonality conditions. This potentiality may include all the invalid orthogonality restrictions \( (l_n = q_n) \). The next possibility is Assumption M.1(iii), where we allow \( l_n \rightarrow \infty \), but \( l_n/q_n \rightarrow 0 \). We separate these two cases, since consistency conditions are different in each. Note that \( C_a, C_b \) does not depend on \( n \).

The following assumption explains the nature of many weak moment asymptotics very well. This is Assumption 1 of Newey and Windmeijer (2009a). Many weak moment asymptotics provide improvements in overidentified models. In those models, the finite sample improvements are substantial when the variance of Jacobian of the moment functions is large relative to its average. The many weak moments approximation is better than the standard Gaussian approximation when there are many weak moments. In the many weak moments case, the asymptotic variance is larger than the usual one. Note that \( \text{diag}(M) \) represents a diagonal matrix \( M \).

**Assumption M.2.**

(i) There is a \( p \times p \) matrix \( S_n = \tilde{S}_n \text{diag}(\mu_1, \ldots, \mu_p) \) so that \( \tilde{S}_n \) is bounded, the smallest eigenvalue of \( \tilde{S}_n S_n' \) is bounded away from zero, for each \( j \) either \( \mu_{jn} = \sqrt{n} \) or \( \mu_{jn}/\sqrt{n} \rightarrow 0 \), where \( \mu_n = \min_{1 \leq j \leq p} \mu_{jn} \rightarrow \infty \), \( q_n \rightarrow \infty \) as \( n \rightarrow \infty \), and \( q_n/\mu_n^2 \) is bounded.

(ii)

\[
\begin{align*}
nS^{-1}_n G' \Omega^{-1} G S^{-1}_n' \rightarrow H,
\end{align*}
\]

where \( H \) is nonsingular and

\[
G(\theta) = E \left[ \frac{\partial g_i(\theta)}{\partial \theta'} \right],
\]

and \( G = G(\theta_0), \Omega(\theta) = Eg_i(\theta)g_i(\theta)', \Omega = \Omega(\theta_0) \).

Note that when \( \mu_{jn} = \sqrt{n} \), the reduced form coefficients are strongly identified. Weak identification is defined as \( \mu_{jn}/n^{1/2} \rightarrow 0 \). This assumption also tells that \( q_n \) grows with sample size. The limit variance in standard GEL converges to zero. But \( nG\Omega^{-1}G \) grows so \( \theta_0 \) is identified.

Next, set \( \delta(\theta) = S_n' (\theta - \theta_0)/\mu_n \). The dependence of \( \delta(\theta) \) on \( n \) is suppressed. We use the following notation. Set \( \hat{g}(\theta) = n^{-1} \sum_{i=1}^{n} g_i(\theta) \), and \( \tilde{g}(\theta) = Eg_i(\theta) \). Furthermore, set \( \tilde{\Omega}(\theta) = \frac{1}{n} (\sum_{i=1}^{n} g_i(\theta)g_i(\theta)') \), and \( \tilde{\Omega} = \tilde{\Omega}(\theta_0) \).

The following assumption, Assumption 2 in Newey and Windmeijer (2009a), is used for global identification, and it simplifies the consistency proof in Newey and Windmeijer (2009a).

**Assumption M.3.** (i) There is \( C > 0 \) with \( \| \delta(\theta) \| \leq C \sqrt{n} \| \hat{g}(\theta) \| / \mu_n \) for all \( \theta \in \Theta \). (ii) There is \( C > 0 \) and \( \dot{M} = O_p(1) \), with probability approaching one, and \( \| \delta(\theta) \| \leq C \sqrt{n} \| \hat{g}(\theta) \| / \mu_n + \dot{M} \), for all \( \theta \in \Theta \).
We use the following Assumptions 3-4 of Newey and Windmeijer (2009a), for the stochastic equicontinuity of the objective function.

**Assumption M.4.** \( g_i(\theta) \) is continuous in \( \theta \), and there is \( C > 0 \) so that (i) \( \sup_{\theta \in \Theta} E[g_i(\theta) g_i(\theta)]^2 / n \to 0 \), (ii) \( 1 / C \leq \text{Eigmin}(\Omega(\theta)) \leq \text{Eigmax}(\Omega(\theta)) \leq C \), for all \( \theta \in \Theta \). (iii) \( \sup_{\theta \in \Theta} \|(\Omega(\theta) - \Omega(\theta))\| \stackrel{p}{\to} 0 \), (iv) \( |a'(\Omega(\theta_2) - \Omega(\theta_1))b| \leq C\|a\|\|b\|\|\theta_2 - \theta_1\| \), for all \( a, b \in R^\mu, \theta_1, \theta_2 \in \Theta \). (v) For every \( \bar{C} > 0 \), there is \( C \) and \( M = O_p(1) \), so that for all \( \theta_1, \theta_2 \in \Theta \), \( \|\delta(\theta_2)\| \leq \bar{C}, \|\delta(\theta_1)\| \leq \bar{C} \), \( \sqrt{n}\|\bar{g}(\theta_2) - \bar{g}(\theta_1)\| / \mu_n \leq C\|\delta(\theta_2 - \theta_1)\| \), and \( \sqrt{n}\|ar{g}(\theta_2) - \bar{g}(\theta_1)\| / \mu_n \leq \bar{M}\|\delta(\theta_2 - \theta_1)\| \). (vi) \( \rho(\nu) \) is three times continuously differentiable and there is \( \gamma > 2 \) so that \( n^{1/\gamma} E \sup_{\theta \in \Theta} \|g_i(\theta)\|^{\gamma} \sqrt{\frac{q_n}{n}} \to 0 \).

First, \( \text{Eigmin}(M) \) and \( \text{Eigmax}(M) \) show the minimal and maximal eigenvalues of the matrix \( M \), respectively. Via Assumption M.4(i), the growth rate of \( E[g_i(\theta) g_i(\theta)]^2 \) is controlled. The central limit theorem in Newey and Windmeijer (2009a) is valid under that assumption. Assumptions M.4(ii)-(iii) are standard arguments for the nonsingularity and the finiteness of the variance matrix. This Assumption also includes uniform convergence of an estimator for the limit variance. Mainly, Assumptions M.4(iv)-(v) are primitives for the stochastic equicontinuity of the sample objective function. Assumption M.4(vi) is standard in GEL literature, and it is used to get consistency and the rate of convergence of the Lagrange Multiplier (see Newey and Smith (2004)). Assumption M.4 restricts the relative number of the moment restrictions compared to the number of observations. If the moment functions are bounded uniformly in parameters, then a sufficient condition for \( \|\tilde{\Omega}(\theta) - \Omega(\theta)\| \stackrel{p}{\to} 0 \) is \( q_n^2 / n \to 0 \). The other parts of the Assumption M.4 provide both invertible variance matrix and the Lipschitz continuity of functions.

The consistency of GEL estimators under Assumptions M.1, M.3, and M.4 follows from the Newey and Windmeijer (2009b) proof. Even though there is near exogeneity and weak identification in many moments setup, we still have the consistency. The detailed proof of consistency, in our case, is contained in the Supplement Appendix. For the limit result, we use Assumptions 6-9 of Newey and Windmeijer (2009a) which are described below.

**Assumption M.5.**

(i) \( g(z_i, \theta) \) is twice continuously differentiable in a neighborhood \( N \) of \( \theta_0 \), setting \( g_i = g_i(\theta_0) \), \( G_i = \frac{\partial g_i}{\partial \theta} \), \( E[\|g_i\|^4 + E[|G_i|^4]|g_i/n \to 0 \), and for a constant \( C \), \( j = 1, \ldots, p \), \( \text{Eigmax}(E[|G_i G'_i|]) \leq C \), \( \text{Eigmax}(E[\|\frac{\partial G_i}{\partial \theta} \frac{\partial G'_i}{\partial \theta}\|]) \leq C \), \( \sqrt{n}\|E[|\frac{\partial G_i}{\partial \theta} \frac{\partial G'_i}{\partial \theta}|S_n^{-1}']\| \leq C \). (ii) If \( \tilde{\theta} \to \theta_0 \), define \( \hat{G}(\theta) = \frac{\partial \hat{g}(\theta)}{\partial \theta} \) for each \( \theta \in \Theta \), then \( \|\sqrt{n}[\hat{G}(\theta) - \hat{G}(\theta_0)]S_n^{-1}'\| \to 0 \). (iii) \( |\frac{\partial G_i}{\partial \theta} - \frac{\partial G_i}{\partial \theta_k}|S_n^{-1}'\| \to 0 \), for \( k = 1, \ldots, p \).

Assumption M.5(i) imposes conditions on the growth rate of the moment conditions, and Assumption M.5(ii) restricts the behavior of derivatives. If \( g_{ij}(\theta) \) are uniformly bounded, a sufficient condition for Assumption M.5(ii) to hold, would be \( q_n^2 / n \to 0 \). So in this sense, this is more restrictive than the implication of Assumption M.4 of \( q_n^2 / n \to 0 \). The assumption on partial derivatives is
used for obtaining the limit of the score of the objective function. This shows that the derivatives are affected by weak instruments and collapse to zero unless multiplied by a scale.

The next assumption imposes uniform convergence and smoothness for the variance terms. Basically, Assumption M.6 further restricts the growth rate of $q_n$. For that purpose define for each $k = 1, \cdots, p$, $l = 1, \cdots, p$, $\hat{\Omega}^k(\theta) = \frac{1}{n} \sum_{i=1}^{n} g_i(\theta) \frac{\partial g_i(\theta)'}{\partial \theta_k}$, $\Omega^k(\theta) = E[\hat{\Omega}^k(\theta)]$, $\hat{\Omega}^{kl}(\theta) = \frac{1}{n} \sum_{i=1}^{n} g_i(\theta) \frac{\partial^2 g_i(\theta)'}{\partial \theta_k \partial \theta_l}$, $\Omega^{kl}(\theta) = E[\hat{\Omega}^{kl}(\theta)]$.

**Assumption M.6.** For all $\theta$ on a neighborhood $N$ of $\theta_0$, $A(\theta)$ equals $\Omega^k, \Omega^{kl}, \Omega^{kl}$, with estimators denoted by $\hat{A}(\theta)$. (i) $\sup_{\theta \in N} ||\hat{A}(\theta) - A(\theta)|| \to 0$. (ii) $|a'[A(\theta_2) - A(\theta_1)]b| \leq C ||a|| ||b|| ||\theta_2 - \theta_1||$, for $\theta_2, \theta_1 \in \Theta$, $a$ and $b$ are vectors. (iii) Either $\hat{\theta}$ is CUE, or there is $\gamma > 2$ so that $n^{1/\gamma} E[\sup_{\theta \in \Theta} ||g_i(\theta)||^\gamma]^{1/\gamma} (q_n + \mu_n)/\sqrt{n} \to 0$. (iv) $\mu_n^2 E[d_i^4]/n \to 0$, for $d_i = \max_{\theta \in \Theta} \max_j \{||g_i(\theta)||, ||\frac{\partial g_i(\theta)}{\partial \theta_j}||, ||\frac{\partial^2 g_i(\theta)}{\partial \theta_k \partial \theta_l}||\}$. (v) $\Omega^k$ is a positive definite matrix.

Assumptions M.6(i) and (ii) impose Lipschitz continuity of certain moments. Assumption M.6(iii) shows the possible stochastic order of the moments. Assumption M.6(iv) shows the relation between the number of moments and the number of observations. If the first and second derivatives are uniformly bounded, we have $E[d_i^4] = O(q_n^2)$. Then since $q_n/\mu_n^2$ is bounded, we see $\mu_n^2 E[d_i^4]/n \to 0$ when $q_n^2/n \to 0$. The number of moments still has to be substantially smaller than the number of observations. Simulations in Newey and Windmeijer (2009a) provide an idea about the number of instruments that offer good finite sample properties.

### 2.2 The Limit for GEL Estimators

In this subsection, we provide two important theorems. First, we obtain a consistency result, and we show that the interaction between $\kappa$ and $l_n$ is important. Next, we obtain the limit for the GEL estimators and show the difference from the Newey and Windmeijer (2009a) result. The following theorem shows two possibilities for getting consistency.

**Theorem 1.**

(i). Under Assumptions M.1(i)(ii), $M.2 - M.4$, for $1/2 < \kappa < \infty$

$$\hat{\theta} - \theta_0 \overset{p}{\to} 0.$$  

(ii). Under Assumptions M.1(i)(iii), $M.2 - M.4$, for $0 < \kappa < \infty$, with $n^{1 - 2\kappa} l_n/q_n \to 0$

$$\hat{\theta} - \theta_0 \overset{p}{\to} 0.$$  

Remarks. 1. Theorem 1(i) shows that even with a positive fraction of all moments violating exogeneity, it is still possible to get consistency as long as the violation is minor ($1/2 < \kappa < \infty$). With all the orthogonality instruments being non-exogenous $l_n = q_n$, we can still have the consistency of GEL as long as the violations are minor.
2. Theorem 1(ii) shows another possibility: even with a larger degree of violation of exogeneity compared to Theorem 1(i), \((0 < \kappa \leq 1/2)\) consistency of the GEL can be achieved. This time, however, we cannot have a large \(l_n\). The restriction is \(n^{1-2\kappa}l_n/q_n \to 0\) which is stronger than Assumption M.1(iii), and it also shows the tradeoff between the number of invalid instruments \(l_n\) and their degree of violation \((\kappa)\). So Theorem 1(i) allows more generality in the number of invalid instruments \(l_n\) compared to Theorem 1(ii), but Theorem 1(ii) also allows for large violation of exogeneity compared to Theorem 1(i), within a smaller number of instruments. Theorem 1(ii) also satisfies \(n^{1-2\kappa}l_n/q_n \to 0\) condition automatically with \(1/2 < \kappa < \infty\) with Assumption M.1(iii).

Note that we cannot allow \(\kappa = 0\), even in Theorem 1(ii). We then see that condition \(n^{1-2\kappa}l_n/q_n \to 0\) is violated because of \(q_n/n \to 0\). This is true even in the case of \(l_n = l\).

We provide the limit of estimators under many weak moments with near exogeneity. Instead of the zero mean normal distribution found in Newey and Windmeijer (2009a), we observe that the limit changes with the violation of exogeneity. This development is new in this literature. We use the following notation: let \(e_j\) denote the \(j^{th}\) unit vector, \(B_j = \Omega^{-1}E[g_i e_j' G_i']\), where \(g_i, G_i\) are defined in Assumption M.5(i), and \(j = 1, \ldots, p\). Then

\[
U^j_i = G_i e_j - Ge_j - B_j' g_i,
\]

\[
U_i = [U_{i1}, \ldots, U_{ip}].
\]

where \(U_i\) represents the residual from regressing the partial derivatives on the moment function.

In the following Theorem 2, the proof of Lemma A.1 shows that the drift depends on the behavior of

\[
nS_n^{-1}G_i'\Omega^{-1}E\tilde{g} = \sqrt{n}S_n^{-1}G_i' n^{1/2-\kappa}\Omega^{-1}(0'_{q_n-l_n}, C_i') C_1 C_1' \Omega^{-1/2} S_n^{-1} (U_i - EU_i).
\]

where \(\tau \neq 0\) or \(\tau = 0\), and \(E\tilde{g} = E[1/n \sum_{i=1}^n g_i(\theta_0)] = E g_i = C_1/n^\kappa\). The value of \(\tau\) depends largely on the behavior of the partial derivative \(G\). \(\tau\) is a \(p \times 1\) vector of constants.

An important element in the analysis of Theorem 2 is directly related to the violation of exogeneity. Because of the violation of exogeneity, there is a new normally distributed random variable. The new limit is normally distributed but with a larger variance term compared to the standard limit in Newey and Windmeijer (2009a). The extra variance term stems from local violation of exogeneity, and this extra variance term: \(\Delta\) is defined as

\[
\Delta = \lim_{n \to \infty} S_n^{-1} E[(U_i - EU_i)'\Omega^{-1/2}(n^{1-2\kappa}\Omega^{-1/2} C_1 C_1' \Omega^{-1/2}) \Omega^{-1/2} (U_i - EU_i)] S_n^{-1}'.
\]

The details of the proof are shown in Lemma SA.6.

The limit of the estimators is different from the one in Newey and Windmeijer (2009a) under certain violations of exogeneity. These limits are explained in detail after Theorem 2. Note that \(\nu_1, \nu_2, \nu_{2nd}\) are normal random variables explained below in Theorem 2.
Theorem 2. Under Assumptions M.1-M.6, and $S_n^{-1}E[(U_i - EU_i)'\Omega^{-1}(U_i - EU_i)]S_n^{-1} \to \Lambda$, (i). If $n^{1/2-\kappa} l_n^{1/2} = O(1)$ and $\tau \neq 0$, then

$$S_n' (\hat{\theta} - \theta_0) \overset{d}{\to} H^{-1}\nu_2,$$

where $\nu_2 \equiv N(\tau, H + \Delta + \Lambda)$.

(ii). If $n^{1/2-\kappa} l_n^{1/2} = O(1)$ and $\tau = 0$, then

$$S_n' (\hat{\theta} - \theta_0) \overset{d}{\to} H^{-1}\nu_2nd,$$

where $\nu_2nd \equiv N(0, H + \Delta + \Lambda)$.

(iii). If $n^{1/2-\kappa} l_n^{1/2} = o(1)$, then

$$S_n' (\hat{\theta} - \theta_0) \overset{d}{\to} H^{-1}\nu_1,$$

where $H^{-1}\nu_1 \equiv N(0, V), V = H^{-1}(H + \Lambda)H^{-1}$.

Remarks.

1. This theorem extends Theorem 3 in Newey and Windmeijer (2009a) to near exogeneity with the many weak moments case. Note that the limits in Theorem 2(i) – (ii) can be simplified as (i).

$$H^{-1}\nu_2 \equiv N(H^{-1}\tau, V + H^{-1}\Delta H^{-1}),$$

where $V = H^{-1}(H + \Lambda)H^{-1}$. The additional variance, due to exogeneity violation, is $H^{-1}\Delta H^{-1}$, and also $H^{-1}\tau$ drift terms stems from exogeneity violation.

(ii). When $\tau = 0$,

$$H^{-1}\nu_2nd \equiv N(0, V + H^{-1}\Delta H^{-1}).$$

This is the case of $\tau = 0$, which is weak identification with a violation of exogeneity.

(iii). Note that with $n^{1/2-\kappa} l_n^{1/2} = o(1)$, everything simplifies to standard many weak moment asymptotics of Newey and Windmeijer (2009a). In that case, $\Delta = 0, \tau = 0, \nu_2 \equiv \nu_1$. These details of the proofs can be seen in Lemma SA.6. When $1/2 < \kappa < \infty$, we can obtain this case. This is basically the case of minor violations with a moderate number of instruments.

2. The violation of exogeneity introduced a different normal distribution than the one in Newey and Windmeijer (2009a). In the limit there is an additional variance term compared with Theorem 3 in Newey and Windmeijer (2009a) and a drift term. These variables are seen in cases (i) and (ii) which come from the change in the limit of the score. In Newey and Windmeijer (2009a), since $Eg_i = 0$, this additional term is zero.

3. We should also pinpoint the importance of $n^{1/2-\kappa} l_n^{1/2}$, which determines the limits and shows the tradeoff between degree of violation of exogeneity and the number of imperfect instruments.
We require \(1/2 \leq \kappa < \infty\) for our Theorem 2. The case of \(\kappa = 1/2\) is compatible with \(l_n = l\) a fixed number of imperfect instruments only. When \(1/2 < \kappa < \infty\), it is possible to have all the instruments being imperfect as in \(l_n = q_n\). If the degree of violation is a mild one (\(\kappa\) is slightly larger than \(1/2\)), then we require \(l_n\) to grow to infinity at a certain rate so that \(n^{1/2-\kappa}l_n^{1/2} = O(1)\) is satisfied for cases (i) and (ii). This is basically \(l_n = O(n^{2\kappa-1})\). For case (iii), we require \(l_n = o(n^{2\kappa-1})\). When \(l_n \rightarrow \infty\), we cannot have \(0 < \kappa \leq 1/2\). In that scenario, we obtain \(n^{1/2-\kappa}l_n^{1/2} \rightarrow \infty\). So the limit of the score will diverge to infinity as well as the GEL estimators. This divergence can be seen from the proof of Theorem 2 by considering the limit of Hessian in Lemma A.2 under the score will diverge to infinity as well as the GEL estimators.

4. Here we outline the possibility of \(\tau \neq 0\) as well as \(\tau = 0\). For the nonzero \(\tau\) to happen, there has to be strong or nearly-weak identification such as

\[
\|n^{1/2} S_n^{-1} G'\| = O(1).
\]

This is shown in Example 1 below, and is described in linear models by Hahn and Kuersteiner (2002), and in nonlinear models by Bertille and Renault (2009), and Caner (2010). There are two possibilities to have a nonzero drift. The first possibility is \(1/2 < \kappa < \infty\), \(n^{1/2-\kappa}l_n^{1/2} = O(1)\). This allows for even \(l_n = q_n\). So all the instruments may be invalid, and with a minor violation of exogeneity in each one, we get a nonzero drift. We allow for both Assumptions M.1(ii) and (iii) which will be shown in a numerical example in Remark 5 below. The second possibility is \(\kappa = 1/2, l_n = l\). This possibility is well illustrated in Example 1. There are two cases, for the possibility of \(\tau = 0\). The first case is \(n^{1/2-\kappa}l_n^{1/2} = O(1)\). It is possible to get a zero drift with weak identification when \(1/2 < \kappa < \infty\). This is true with both Assumptions M.1(ii) and (iii). Also, with weak instruments only, the zero drift is possible by relaxing \(\kappa\) being equal to \(1/2\) with \(l_n = l\) which represents fixed number of invalid instruments. The second case is when \(n^{1/2-\kappa}l_n^{1/2} = o(1)\), which is used in Example 2 below.

5. We use a numerical example to illustrate the conditions in Theorem 2. For the cases of (i) and (ii), the assumption is \(n^{1/2-\kappa}l_n^{1/2} = O(1)\). Assume that \(\kappa = 2/3\) is at each instrument (\(l_n = q_n\), minor violation of exogeneity in instruments). So we have \(l_n = q_n = O(n^{1/6})\). To get a zero drift, we require \(n^{1/2-\kappa}l_n^{1/2} \rightarrow 0\), which means \(l_n = q_n = o(n^{1/6})\).

The next two examples represent cases when there is a new term in the limit due to exogeneity violation (Theorem 2(i)), and another one where we get the standard limit in Newey and Windmeijer (2009a) (Theorem 2(iii)). The first example is taken from p.690-698 of Newey and Windmeijer (2009a).

**Example 1.** In this example, we assume \(n^{1/2-\kappa}l_n^{1/2} = O(1)\). Suppose the model is

\[
y_i = x_i' \beta_0 + \epsilon_i,
\]

\[
x_i = \Psi_i + \eta_i,
\]
where $E[\eta_i|Z_i, \Psi_i] = 0$, $E[Z_i\epsilon_i] = C_1/n^\kappa$, and $C_1$ is described in Assumption M.1 with $1/2 < \kappa < \infty$. $Z_i$ represents a vector of control variables and the feasible instruments. We set $g_i = Z_i(y_i - x_i\beta_0)$. Suppose as in Newey and Windmeijer (2009a), the reduced form equation and the instruments are given by $x_i = (z_{1i}', x_{2i}')'$, $Z_i = (z_{1i}', Z_{2i}')'$,

$$x_{2i} = \pi_{2i}z_{1i} + (\mu_n/n^{1/2})z_{2i} + \eta_{2i},$$

where $z_{1i}$ is $p_1 \times 1$ vector of included exogenous variables, $z_{2i}$ is a $(p - p_1) \times 1$ vector of excluded exogenous variables, and $Z_{2i}$ is $(q_n - p) \times 1$ vector of instruments. If $z_{2i}$ is not observed, this is approximated by $Z_{2i}$, which can be a polynomial. Then

$$\tilde{S}_n = \begin{pmatrix} I_{p_1} & 0 \\ \pi_{2i} & I_{p-p_1} \end{pmatrix}.$$ 

Also, $\mu_{jn} = n^{1/2}$ for $j = 1, 2, \cdots p_1$, and $\mu_{jn} = \mu_n$ for $j = p_1, \cdots p$. Therefore, the first $p_1$ instruments are strong and the rest are nearly-weak. Then, for $z_i = (z_{1i}', z_{2i}')'$, the reduced form is written

$$\Psi_i = \begin{pmatrix} \frac{z_{1i}}{\pi_{2i}z_{1i} + (\mu_n/n^{1/2})z_{2i}} \\ z_{2i} \end{pmatrix},$$

and $G = -E[Z_i\Psi_i]' = -E[Z_i\tilde{z}_i]'S_n^{-1}/n^{1/2}$. Since $E\tilde{g} = C_1/n^\kappa$ in our case, by using Assumption M.1

$$nS_n^{-1}G'\Omega^{-1}E\tilde{g} = n^{1/2}S_n^{-1}G'\Omega^{-1}n^{1/2-\kappa}\Omega^{-1}(0_{q_n-l_n}, C_i)' = E[Z_i\tilde{z}_i]'n^{1/2-\kappa}\Omega^{-1}(0_{q_n-l_n}, C_i)' = \tau.$$  \hspace{1cm} (3)

If $n^{1/2-\kappa}l_{1n}/2 = O(1)$, then $\tau \neq 0$. For details about how that rate is obtained, see (SA.46). So with both strong and nearly-weak instruments, we still have the drift. The key factor will be how fast the partial derivative $G$ converges to zero. We should also note that Newey and Windmeijer (2009a) show that Assumption M2(ii) is also satisfied in this example. Of course, we can get a nonzero drift with $\kappa = 1/2$ and $l_n = l$. We can relax $1/2 < \kappa < \infty$, but then $l_n$ cannot increase with $n$. Analyzing the second element that contributes to the new limit, this is

$$nS_n^{-1}\tilde{U}'\Omega^{-1}E\tilde{g} = n^{1/2}S_n^{-1}\tilde{U}'n^{1/2-\kappa}\Omega^{-1}C_1,$$  \hspace{1cm} (4)

where $E\tilde{g} = E\tilde{g}_i = C_1/n^\kappa$ by data being iid and Assumption M.1. We can simplify $U_i^j = -Z_i\Psi_{ij} + E[Z_i\Psi_{ij}] + u_{ij}, u_{ij} = -Z_i\eta_{ij} + B'\eta Z_i\epsilon_i, B^j = -\Omega^{-1}E[Z_i\tilde{z}_i\eta_{ij}\epsilon_i], j = 1, \cdots, p, i = 1 \cdots, n$. Then, the extra term $\Delta$ in the variance comes from when we take the limit, of the variance of (4) when $n^{1/2-\kappa}l_{1n}/2 = O(1)$, and

$$\Delta = \lim_{n \to \infty} S_n^{-1}E[(u_i - E u_i)'(u_i - E u_i)]S_n^{-1},$$

$u_i = (u_{i1}, \cdots, u_{ip})'$. For these derivations see pages 703-704 of Newey and Windmeijer (2009b).

**Example 2.** In this example, we assume $n^{1/2-\kappa}l_{1n}/2 = o(1)$. Let

$$y_i = x_i'\theta_0 + \epsilon_i,$$
\[ x_i = \pi'_n z_i + v_i, \]

\[ \pi_n = \frac{d}{\sqrt{n}}, \epsilon_i \sim \text{iid}(0, \sigma_\epsilon^2), 0 < \sigma_\epsilon^2 < \infty, \quad z_i \text{ is iid, } v_i \sim \text{iid}(0, \sigma_v^2), 0 < \sigma_v^2 < \infty, \quad x_i : p \times 1, \]
\[ \pi_n : q_n \times p, \quad z_i : q_n \times 1. \]

Set \( E(\epsilon_i^2 / z_i) = \sigma_\epsilon^2, E(\epsilon_i \epsilon_j / z_i) = 0 \) for all \( i \neq j \). Then we can see that \( g_i(\theta) = z_i(y_i - x'_i \theta), \) and set \( d = (d'_1, d'_2)' \) where \( d_1 : (q_n - l_n) \times p, d_2 : l_n \times p. \) \( \epsilon \) and \( v \) are correlated. Assume \( \text{Eigmax}(dd') < \infty, \) \( E z_i \epsilon_i = C_1/n^\kappa, \) where \( 1/2 < \kappa < \infty. \) With the \( \pi_n \) definition, this is the weak instruments setup of Stock and Wright (2000). So
\[
G = -E[z_i x'_i] = -E[z_i z'_i] \frac{d}{\sqrt{n}}.
\]

Also we have \( \Omega = \sigma_\epsilon^2 E z_i z'_i \) which yields
\[
n S_n^{-1} G^\prime \Omega^{-1} E g_i = -\sigma_\epsilon^{-2} S_n^{-1} n^{1/2 - \kappa} d'(0_{q_n - l_n}, C'_l) \leq -\sigma_\epsilon^{-2} S_n^{-1} n^{1/2 - \kappa} \| d'_2 C_l \|
\]
\[
= O(n^{1/2 - \kappa} l_n^{1/2} / \mu_n) \rightarrow 0,
\]

where \( \| d'_2 C_l \| = O(l_n^{1/2}) \) by \( \| C_l \| = O(l_n^{1/2}) \), via Assumptions M.1 and M2(i). So it is possible to have \( \tau = 0 \) in the following Theorem 2. Since we assume \( n^{1/2 - \kappa} l_n^{1/2} = o(1) \), we have \( \Delta = 0. \) Therefore there is no extra variance term in the limit as in Example 1. The details are all in the proof of Lemma SA.6. Of course, we have to check whether this example satisfies Assumption M2(ii). To see that
\[
n S_n^{-1} G^\prime \Omega^{-1} G S_n^{-1}' = n S_n^{-1} (E(z_i z'_i) \frac{d}{n^{1/2}})' (\sigma_\epsilon^2 E(z_i z'_i))^{-1} (E(z_i z'_i) \frac{d}{n^{1/2}}) S_n^{-1}'
\]
\[
= \sigma_\epsilon^{-2} S_n^{-1} d'(E(z_i z'_i))' d S_n^{-1}'
\]
\[
\rightarrow H
\]

This derivation is true since \( E(z_i z'_i) \) is a positive definite matrix, and \( d \) is \( q_n \times p \) matrix and by Assumption M2(i), \( q_n / \mu_n^2 \) is bounded.

The differences in both examples can be understood from the behavior of \( G \) and \( n^{1/2 - \kappa} l_n^{1/2} \) conditions. If the instruments are nearly-weak as in Newey and Windmeijer (2009a), Example 1, with mild exogeneity violations, then we have an extra variance term in the limit with a drift compared with standard case of perfect exogeneity. If the instruments are as weak as described in Stock and Wright (2000), and the violations are many but minor, then Example 2 shows that the limit will be the same as in Newey and Windmeijer (2009a).

3 Tests

In this section, we discuss inference when there are many weak moments and near exogeneity (Assumptions M.1 and M2). Dufour (1997) argues that in linear models, the confidence intervals for structural parameters must be unbounded with positive probability if there are identification
problems. Newey and Windmeijer (2009a) find that the Kleibergen (2005), Anderson-Rubin (1949),
and Conditional Likelihood Ratio tests achieve correct asymptotic size. These results can be seen
as Theorems 3-5 in Newey and Windmeijer (2009a). We extend these cases to Assumption M.1
which is violation of exogeneity. For both tests, the null is \( H_0 : \theta = \theta_0 \). The Anderson-Rubin
(1949) type of test is

\[
AR(\theta_0) = 2n\hat{Q}(\theta_0) .
\]  

(5)

Theorem 5 of Newey and Windmeijer (2009a) analyzes the Kleibergen (2005) type of test statistic

\[
K(\theta_0) = n \left[ \frac{\partial \hat{Q}(\theta_0)}{\partial \theta'} \right]' \left[ \hat{D}(\theta_0)' \hat{\Omega}(\theta_0)^{-1} \hat{D}(\theta_0) \right]^{-1} \left[ \frac{\partial \hat{Q}(\theta_0)}{\partial \theta'} \right] ,
\]  

(6)

where

\[
\hat{D}(\theta) = \sum_{i=1}^{n} \hat{\rho}_1(\theta) \partial g_i(\theta)/\partial \theta' / \sum_{i=1}^{n} \hat{\rho}_1(\theta),
\]

and \( \rho_1(.) \) represents the partial derivative of \( \rho(.) \) with respect to its arguments: \( \hat{\rho}_1(\theta) = \rho_1(\hat{\lambda}(\theta)'g_i(\theta)) \).

This test has \( \chi^2_p \) limit under many weak moments in Newey and Windmeijer (2009a). Newey and
Windmeijer (2009a) also consider the Conditional Likelihood Ratio test. This test is computa-
tionally burdensome and it is very similar to the asymptotic analysis for the \( K(\theta_0) \) test. We will
not consider this test here. The Wald/t type of tests can be formed with the new standard errors
taking into account many weak moments. If there is \( r_n \) and \( c^* \neq 0 \) so that \( r_nS_n^{-1}c \rightarrow c^* \), we define the \( t-test \)

\[
t = \frac{c'(\hat{\theta} - \theta_0)}{\sqrt{c'\hat{V}c/n}},
\]

where \( \hat{V} = \hat{H}(\hat{\theta})^{-1}(\hat{D}(\hat{\theta})'\hat{\Omega}(\hat{\theta})^{-1}\hat{D}(\hat{\theta}))\hat{H}(\hat{\theta})^{-1} \), and \( \hat{H}(\hat{\theta}) = \partial^2 \hat{Q}(\hat{\theta})/\partial \theta \partial \theta' \). This test has the standard normal distribution as it is shown in Newey and Windmeijer (2009a), when \( C_1 = 0 \) (perfect exogeneity case). Newey and Windmeijer (2009a) also analyze the overidentifying restrictions test

\[
OI = 2n\hat{Q}(\hat{\theta}).
\]

This test is used to detect whether the overidentifying restrictions associated with moment condi-
tions are correct. Under perfect exogeneity and many weak moments, the limit has the asymptoti-
cally correct level.

Theorem 3 shows pointwise asymptotic null rejection probabilities of several tests, and it is one
of the main results of this paper. It shows that various tests differ in their limits when there is a
violation of perfect exogeneity.

**Theorem 3.**

(i). We want to test \( H_0 : c'\theta = c'\theta_0 \) for a \( p \times 1 \) vector. Under Assumptions M.1-M6, and
\( S_n^{-1}E[(U_i - EU_i)\Omega^{-1}(U_i - EU_i)]S_n^{-1}' \rightarrow \Lambda \), if there is \( r_n \) and \( c^* \neq 0 \) so that \( r_nS_n^{-1}c \rightarrow c^* \), then there are three distinct cases.
First, if \( n^{1/2-\kappa}l_n^{1/2} = o(1) \),
\[
\frac{c'(\hat{\theta} - \theta_0)}{\sqrt{c'\hat{\Omega}c/n}} \xrightarrow{d} N(0,1).
\]

Second, if \( n^{1/2-\kappa}l_n^{1/2} = O(1) \) and \( \tau = 0 \),
\[
\frac{c'(\hat{\theta} - \theta_0)}{\sqrt{c'\hat{\Omega}c/n}} \xrightarrow{d} N(0,1 + \frac{\alpha_0'}{\alpha_0}\Delta_{n}^{-1}c^*),
\]
where \( \Delta \) is described in Theorem 2.

Third, if \( n^{1/2-\kappa}l_n^{1/2} = O(1) \) and \( \tau \neq 0 \),
\[
\frac{c'(\hat{\theta} - \theta_0)}{\sqrt{c'\hat{\Omega}c/n}} \xrightarrow{d} N(\frac{\alpha_0'}{\alpha_0}\Delta_{n}^{-1}\tau, 1 + \frac{\alpha_0'}{\alpha_0}\Delta_{n}^{-1}c^*).
\]

(ii). We want to test \( H_0: \theta = \theta_0 \). If \( q_nE||g||^4/n \to 0 \), \( \text{Eigmin}(\Omega) \geq C \), there is \( \gamma > 2 \), so that \( n^{1/\gamma}E||g||^\tau/q_n/n^{1/2} \to 0 \), and if \( n^{1/2-\kappa}(l_n/q_n)^{1/2} = o(1) \)
\[
P(\text{AR}(\theta_0) \geq \sigma_{\alpha}^2) \to \alpha,
\]
where \( \sigma_{\alpha}^2 \) is the \( (1-\alpha)^{th} \) quantile of a \( \chi^2_{\alpha} \) distribution. Note that when centered and standardized, this quantile converges to the quantile of standard normal distribution.

(iii). We want to test \( H_0: \theta = \theta_0 \). If \( S_n^{-1}E[(U_i - E\Omega)^\tau(U_i - E\Omega)]S_n^{-1'} \to \Lambda \), and under Assumptions M.1-M6, there are three possibilities.
First, if \( n^{1/2-\kappa}l_n^{1/2} = o(1) \), then
\[
K(\theta_0) \xrightarrow{d} \chi^2_p,
\]
where \( \chi^2_p \) represents the standard \( \chi^2 \) distribution with \( p \) degrees of freedom.

Second, if \( n^{1/2-\kappa}l_n^{1/2} = O(1) \) and \( \tau = 0 \),
\[
K(\theta_0) \xrightarrow{d} \nu_{2nd}(H + \Lambda)^{-1}\nu_{2nd},
\]
where \( \nu_{2nd} \) is described in detail in Theorem 2.

Third, if \( n^{1/2-\kappa}l_n^{1/2} = O(1) \) and \( \tau \neq 0 \),
\[
K(\theta_0) \xrightarrow{d} \nu_2(H + \Lambda)^{-1}\nu_2,
\]
where \( \nu_2 \) is described in detail in Theorem 2.

Remarks. 1. These results clearly show that the \( \text{AR}(\theta_0), K(\theta_0), \) and t-tests are not affected by the near exogeneity problem in some cases. We still get the limit in Theorem 5 of Newey and Windmeijer (2009a). Even with \( l_n = q_n \), it is possible to get the standard limit with both \( \text{AR}(\theta_0), K(\theta_0) \) tests. However this is only possible with minor violations in exogeneity: \( 1/2 < \kappa < \infty \) as long as \( n^{1/2-\kappa}l_n^{1/2} = o(1) \).
2. Using Theorem 3(ii), we will show that, based on $AR(\theta_0)$ results, if the violation of exogeneity is such that $n^{1/2-\kappa l_n^{1/2}} = O(1)$, then the $AR(\theta_0)$ rejection probability of the null will not be affected. Therefore, we do not expect size distortions in this important case, unlike the $K(\theta_0)$ test. To see this, the null rejection probability of the $AR(\theta_0)$ test is $\alpha$ when $n^{1/2-\kappa (l_n/q_n)^{1/2}} = o(1)$. However, by analyzing (A.15) in the proof of Lemma A.4, we see that if $n^{1-2\kappa C_1^{1/2}} C_1^{1/2} C_1 q_n \rightarrow \Xi \neq 0$, then the $AR(\theta_0)$ test is size distorted. This distortion can be shown via the proof of Lemma A.4 and Theorem 3(ii). Also, from the proof of Lemma A.4, under $n^{1/2-\kappa l_n^{1/2}} = O(1)$, we get $\Xi \neq 0$.

3. Note that both the Kleibergen (2005) type (K test) of test and the $t/Wald$ type of tests are affected when $n^{1/2-\kappa l_n^{1/2}} = O(1)$. The K test depends on the score of the objective function. From Lemma A.1, we see that the limit of the score of the objective function is a new normal distribution compared with the one in Newey and Windmeijer (2009a). In the new limit, there is an extra variance term ($\Delta$) compared with standard case of Newey and Windmeijer (2009a). This extra variance term depends on $n^{1/2-\kappa l_n^{1/2}}$, and it is also described in Lemma A.1 and Theorem 2.

4. We form the Wald test for the null of $H_0 : \theta = \theta_0$. From Newey and Windmeijer (2009a), with $\hat{V}$ as the consistent estimate for $V = H^{-1}(H + \Lambda)H^{-1}$,

$$Wald = n(\hat{\theta} - \theta_0)/\hat{V}^{-1}(\hat{\theta} - \theta_0).$$

See Remark 1 of Theorem 2 and Theorem 3(i) to get the limit

$$Wald \overset{d}{\to} \nu_2^2 H^{-1}[H^{-1}(H + \Lambda)H^{-1}]^{-1} H^{-1} \nu_2 \equiv \nu_2^2 (H + \Lambda)^{-1} \nu_2,$$

under $n^{1/2-\kappa l_n^{-1/2}} = O(1)$, and $\tau \neq 0$. This limit is the same as in Theorem 3(iii), under the same conditions. Also, we see that if $n^{1/2-\kappa l_n^{1/2}} = O(1)$, and $\tau = 0$, then

$$Wald \overset{d}{\to} \nu_{2nd}^2 (H + \Lambda)^{-1} \nu_{2nd}.$$

With $n^{1/2-\kappa l_n^{1/2}} = o(1)$,

$$Wald \overset{d}{\to} \chi_2^2.$$

In short, asymptotically the Wald test behaves like the Kleibergen (2005) test even under exogeneity violations in many weak moment asymptotics.

5. After studying the theorem above, we want to know what matters in an application? What are the factors that affect $AR(\theta_0)$, $K(\theta_0)$, and the sizes of Wald tests? For the $AR(\theta_0)$ test from Remark 2, we see that critical terms are the ratio of the number of imperfect instruments to all instruments $(l_n/q_n)$ as well as the magnitude of the violation $C_1/n^\kappa$. As the researchers we can control $q_n$. But $l_n, C_1, \kappa$ are not observed first hand. So the researchers, by being careful, and by picking really relevant and valid instruments, can make $l_n, C_1/n^\kappa$ very small. Therefore, the size of the $AR(\theta_0)$ test will be good. The $K(\theta_0)$ test is affected by $l_n, C_1/n^\kappa$, which can be seen in Theorem 2. The test, therefore, is not affected by the total number of instruments, $q_n$. Increasing
$q_n$ decreases size distortion, keeping everything constant. We can understand this distinction, also by comparing the conditions on the $AR(\theta_0), K(\theta_0)$ tests. Condition $n^{1/2-\kappa l_n^{1/2}/q_n^{1/2}}$ is much weaker than $n^{1/2-\kappa l_n^{1/2}}$. The Wald test behaves in the same way as the Kleibergen (2005) test.

Berkowitz, Caner, and Fang (2008), and Doko and Dufour (2008) make the point that the $AR(\theta_0)$ test is invalid under near exogeneity when the number of instruments is fixed. When $\kappa = 1/2$ as in their case, the non-centrality parameter is $C_1^\top \Omega^{-1} C_1$, and our condition in Theorem 3(ii) becomes $n^{1/2-\kappa (l_n/q_n)^{1/2}} = (l/q)^{1/2} \neq 0$ since $l$ and $q$ are fixed in Berkowitz, Caner, and Fang (2008).

### 3.1 OI test and Practical Advice

In this part, we first note that in Newey and Windmeijer (2009a), with no violations of exogeneity, the OI test has the perfect size. We consider the power of the OI test under violations of exogeneity.

**Corollary 1.** If $S_n^{-1} E[(U_i - EU_i)\Omega^{-1}(U_i - EU_i)] S_n^{-1} \rightarrow \Lambda$, and under Assumption M.1-M.6

(i). If $n^{1-2\kappa l_n}/\sqrt{q_n} = o(1)$, then

$$\lim_{n \rightarrow \infty} P(OI \geq s_{\alpha}^{q_n-p}) = \alpha,$$

where $s_{\alpha}^{q_n-p}$ is the $(1-\alpha)$th quantile of a $\chi_{q_n-p}^2$ limit.

(ii). If $n^{1-2\kappa l_n}/\sqrt{q_n} = O(1)$, then

$$\lim_{n \rightarrow \infty} P(OI \geq s_{\alpha}^{q_n-p}) > \alpha.$$

(iii). If $n^{1-2\kappa l_n}/\sqrt{q_n} \rightarrow \infty$, and $n^{1-2\kappa l_n}/q_n = O(1)$,

$$\lim_{n \rightarrow \infty} P(OI \geq s_{\alpha}^{q_n-p}) = 1.$$

Remarks. 1. Corollaries 1(i)-(iii) show three important cases. First, with Corollary 1(i), we see that, with minor violations, the OI test has no power. If the violations are slightly larger and more numerous, then Corollary 1(ii) shows that the OI test has power. Then, in Corollary 1(iii), we show that under certain alternatives the test is consistent. We discuss these below. If we have mild violations of exogeneity, as in $0 < \kappa \leq 1/2$, then $n^{1-2\kappa l_n} \rightarrow \infty$, we can detect these types of violations as long as $l_n/q_n$ is not too small. In Corollary 1(iii), the OI test is consistent. For example when $l_n/q_n \rightarrow f$, $0 < f \leq 1$, with $\kappa = 1/2$, we have $n^{1-2\kappa l_n}/\sqrt{q_n} = n^{1-2\kappa} \frac{l_n}{q_n} \sqrt{q_n} \rightarrow \infty$ and $n^{1-2\kappa l_n}/q_n = O(1)$. The power of the OI test will be smaller with few minor violations compared with a large number of mild violations of exogeneity. This is when $l_n/q_n \rightarrow 0, 0 < f \leq 1/2 < \kappa < \infty$ in comparison to $l_n/q_n \rightarrow f$, $0 < f \leq 1$, and $\kappa = 1/2$. All these equations can be seen in cases (i) and (ii). Corollary 1(i) shows that the OI test cannot detect the cases where $n^{1-2\kappa l_n} = o(\sqrt{q_n})$. We expected non-detection of the violation of exogeneity to occur when $\kappa$ is very large, a very minor violation of exogeneity, or with a very small number of imperfect instruments. Also, the conditions in Corollary 1 may satisfy all the conditions in Theorem 1.
2. If \( n^{1-2\kappa}l_n/\sqrt{q_n} = O(1) \) or \( n^{1-2\kappa}l_n/\sqrt{q_n} \to \infty \), the OI test will detect violations of exogeneity. For example, with \( \kappa = 1/2 \), \( l_n/q_n = 1/2 \), \( n^{1-2\kappa}l_n/\sqrt{q_n} \to \infty \), the OI test will be consistent in this case. In that setup, the \( AR(\theta_0) \) will be size distorted since \( n^{1/2-\kappa}(l_n/q_n)^{1/2} = (1/2)^{1/2} \neq 0 \), from Theorem 3(ii) as well as the \( K(\theta_0) \) and Wald tests in Theorem 3(i) and (iii). In the case of Corollary 1(iii), if \( \kappa = 1 \) and \( l_n/q_n \to f, 0 < f \leq 1 \), then \( n^{1-2\kappa}l_n/\sqrt{q_n} = n^{-1}q_n^{1/2}(l_n/q_n) \to 0 \) since \( q_n/n \to 0 \). That type of violation cannot be detected with the OI test, but the asymptotic null rejection probability of the \( AR(\theta_0) \) test will not be affected since \( n^{1/2-\kappa}(l_n/q_n)^{1/2} = n^{-1/2}f^{1/2} \to 0 \) in Theorem 3(ii). The good news is that when the OI test cannot detect the violation with \( n^{1-2\kappa}l_n/\sqrt{q_n} \to 0 \), then we know from Theorem 3(iii) that the null rejection probability of the \( AR(\theta_0) \) test will not be affected since \( n^{1/2-\kappa}(l_n/q_n)^{1/2} \to 0 \) as well. When the OI test does not detect any violation of exogeneity, we prefer to use the Anderson-Rubin (1949) test since its conditions are weaker than the Wald or \( K(\theta_0) \) test. Of course, since some researchers may be interested in testing a single restriction, or a subset of restrictions on endogenous variables, or if the weak identification is not an issue, the Wald/t type of test may be useful compared to Anderson-Rubin (1949) or Kleibergen (2005) type of tests.

3.2 Generalization of Violation of Exogeneity

In this subsection, we generalize Assumption M.1 and try to understand how the conditions in Theorem 3 and Corollary 1 will change. We assume a general violation of exogeneity which means invalid instruments with two differing degrees of violation. Of course, this type of violation can be generalized to each individual instrument having its own degree of invalidity. We will comment on that after we show how the conditions of Theorem 3 and Corollary 1 are changing.

**Assumption M.7.** (i).

\[
Eg_i(\theta_0) = R_n^{-1}C_1,
\]

where

\[
R_n = \begin{bmatrix}
I_{q_n-l_n} & 0_{l_1n} & 0_{l_2n} \\
0_{q_n-l_n} & n^{\kappa_1}I_{l_1n} & 0_{l_2n} \\
0_{q_n-l_n} & 0_{l_1n} & n^{\kappa_2}I_{l_2n}
\end{bmatrix},
\]

and

\[
C_1 = \begin{bmatrix}
0_{q_n-l_n} \\
C_{1l_1n} \\
C_{1l_2n}
\end{bmatrix}.
\]

So \( R_n^{-1}C_1 = (0_{q_n-l_n}^\prime, C_{1l_1n}/n^{\kappa_1}, C_{1l_2n}/n^{\kappa_2})^\prime \). \( R_n \) is a \( q_n \times q_n \) matrix, which is invertible. \( C_{1l_1n} \), and \( C_{1l_2n} \) are \( l_{1n} \times 1, l_{2n} \times 1 \) vectors respectively. We also know that \( l_{1n} \to \infty, l_{2n} \to \infty \) when \( n \to \infty \), and \( l_{1n} + l_{2n} = l_n \). For each \( s = 1, 2, j = 1, 2, \cdots, l_{sn} < C_a < C_{sn,j} < C_b < \infty \) where \( C_a \), and \( C_b \) are scalars. \( C_{l_{sn,j}} \) is in a compact set. \( 0_{q_n-l_n} \) represents a zero vector of \( q_n - l_n \) dimension. Other
quantities in $R_n$ are self explanatory. We allow for two possibilities for the ratio of the number of imperfect moment conditions $l_{sn}$, $s = 1, 2$ to total number of moment conditions $q_n$

(ii) Let $l_{sn}/q_n \rightarrow f$, as $n \rightarrow \infty$, where $0 < f \leq 1$.

or

(iii) Let $l_{sn}/q_n \rightarrow 0$, as $n \rightarrow \infty$.

The results of Corollary 1 and Theorem 3 will remain the same, but the conditions will change slightly. The new condition for Corollary 1 will address on the behavior of $\max(n^{1-2\kappa_1}l_{1n}/q_n^{1/2}, n^{1-2\kappa_2}l_{2n}/q_n^{1/2})$. For Theorem 3(ii), the condition should become $\max(n^{1-2\kappa_1}l_{1n}/q_n^{1/2}, n^{1-2\kappa_2}l_{2n}^{1/2}/q_n^{1/2})$. Then the condition for Theorem 3(iii) will become $\max(n^{1/2-\kappa_1}l_{1n}^{1/2}, n^{1/2-\kappa_2}l_{2n}^{1/2})$. These conditions can be seen from the proofs easily. In the case of more than two groups of invalid instruments, the instrument group which has the largest degree of violation, is the one which determines the size of the tests.

4 Simulation

In this section, we attempt to answer several questions. If there is an exogeneity violation as described in Assumption M.1 in many weak moments case, can the OI test detect that? If not, are the AR, K, and Wald tests in that setup robust to these violations of exogeneity? In other words, we will check the power of the OI test. We will then calculate the null rejection probability of the AR, K and Wald tests, when the null is true. We will also discuss how some of the parameters can contribute to exogeneity violations and thus affect the finite sample properties of the AR, K, and Wald, OI tests. However, as we see in the finite samples, the null rejection probability of the Wald test is somewhat better than the K test, so we report the Wald test rather than the K test. When we start our setup, we consider the following simple linear model:

$$y_i = Y_i\theta_0 + \epsilon_i,$$

$$Y_i = \Pi'X_i + V_i,$$

where $\theta_0 = 0$. $\Pi$ can take two values, $\bar{e}/n^{1/3}$, and $\bar{e}/n^{1/12}$, where $\bar{e}$ is $q_n$ vector of ones. Since the results of these tests are similar, we report the ones with $\bar{e}/n^{1/3}$. That setup allows for nearly-weak instruments. The number of near exogenous instruments is either $l_n = 1$ or $l_n = 4$. $Y_i$ is scalar. $(X_{i1}, \epsilon_i, V_i)$ are jointly normally distributed with a variance covariance matrix $\Sigma$ (dimension: $(q_n + 2) \times (q_n + 2)$). The variance terms in $\Sigma$ are normalized at one. We have the following setup for $\text{cov}(X_{i1}, \epsilon_i) = C_{l_n}/n^\kappa$, where $X_{i1}$ represents the invalid instruments, and $C_{l_n}$ varies between 0, 1, 2, 3. All the values of $C_{l_n}$ are the same for simplification. The other parameter that is relevant in exogeneity violation is $\kappa$. We set $\kappa = 1/4, 1/2, 1$. So the small values of $\kappa$ display larger exogeneity violation.
We are interested in three specific setups. First, we consider the minor violations with a small $C_{ln} = 1$ and $\kappa = 1$. Then we analyze moderate violations with $C_{ln} = 2, 3$ and $\kappa = 1/2$. Next, the large violations become the center of interest with $\kappa = 1/4$. We also tried some simulations with $C_{ln} = 5$ which are similar in nature to $C_{ln} = 3$ results, so we do not report them here. The other instruments are uncorrelated with both the structural and reduced form errors in $\Sigma$. Furthermore, $X_{i1}$ is uncorrelated with the reduced form error term. Also, we set $cov(\epsilon_i, V_i) = 0.5$. The total number of instruments is $15, 25, 50, 60, 75, 80, 100$ for $n = 200, 500$. The sample sizes of 200, 500 are common in labor/education studies.

For both tests, we try both the heteroskedasticity corrected and uncorrected versions. The results are very similar, so we use a conditionally homoskedastic version. The limits are described in Theorem 3 and Corollary 1. We analyze the tests at the 5% nominal level. We report the rejection percentage out of 1000 iterations. We also conduct 10000 iterations in certain cases, but this increase does not make any significant difference, so we use 1000 iterations given the large sample sizes analyzed. The results are displayed in Tables 1-4. The case for $C_{ln} = 0$ shows the size under the standard perfect exogeneity assumption.

In this exercise, we test the power of the OI test specifically. Then when the null is true, we calculate the null rejection frequency of the AR and Wald tests with an exogeneity violation. For the AR and Wald tests, there is no power exercise. The theories tell us specifically that through (A.15), the null rejection frequency of the AR test will depend on $C_{ln}, \kappa, l_n/q_n$. The null rejection frequency of the Wald test depends on all the factors that impact the AR test except the total number of the instruments $q_n$. The power of the OI test is affected by all the factors that affect the AR test. Of course, the conditions for these tests are different, and the implications on the power of the OI test and the null rejection frequency of the AR and Wald tests are discussed in detail in Remark 2 after Corollary 1. Assumptions M.1, M.5, and M.6 make the point that $q_n$ cannot be too small or too large for the theories to hold. Newey and Windmeijer (2009a) also make the point that for uniformly bounded functions $q_n^{3/2}/n \to 0$ for the K test.

Tables 1a through 1c show the power of the OI test with $l_n = 1$ under the alternatives $\kappa = 1/4, 1/2, 1$, respectively. Tables 2a through 2f present the null rejection frequency for the AR and Wald tests with $l_n = 1$, when the null hypothesis is true. Tables 3 and 4 repeat Tables 1 and 2 with $l_n = 4$. We begin with Table 1a, where $\kappa = 1/4$, and we consider the power of the OI test. From Corollary 1 and the proof, the power of test depends on $C_{ln}$ and $n^{1-2\kappa} l_n^{1/4} q_n^{-1/2}$ in Table 1a. So we expect the test to be consistent and detect the violations of exogeneity in this setup. For example, with $l_n = 1, q_n = 50, C_{ln} = 2, n = 200$, the power is 98.6%. The power increases with large $n, C_{ln}$ and decreases with large $q_n$. In Table 1a with $n = 200, C_{ln} = 1, q_n = 100$, the power of the OI test is 3.9%. So the OI test cannot detect that type of violation. But when we look at the AR test, its null rejection frequency is 3.9% in the same setup ($n = 200, C_{ln} = 1, q_n = 100$) in Table 2c. The Wald test does not perform as well as the AR test. The Wald test rejects the true null 34.9%
in Table 2c. Even though the OI test cannot detect that violation, the null rejection frequency of AR test is not affected; hence, the researcher can use these tests in sequence. We try the same exercise with $\kappa = 1/2$, which represents a mild violation of exogeneity. We expect some local power as shown in Corollary 1(ii). With $C_l_n = 2, q_n = 15, n = 200$, in Table 1b, the power of the OI test is 15.2%. In Table 2b, the null rejection frequency of the AR and Wald tests are 16.7% and 5.9% at a nominal 5% level, respectively. Thus, the Wald test does well after the OI test cannot detect the violation. We also analyze the case of minor violations ($\kappa = 1$), where the OI test cannot detect the violations since $n^{1-2\kappa}l_n/\sqrt{q_n} = n^{-1}l_n/\sqrt{q_n} \rightarrow 0$ as in Corollary 1(i) in Table 1c. The power of the test is even smaller than the nominal size in Table 1c. However, as we see from Table 2a, the AR test null rejection frequency is between 1-5%, hence the size of the AR test is not affected when the null is true, but there is an exogeneity violation. This violation is due to the condition in Theorem 3(ii) being satisfied, namely $n^{1/2-\kappa}(l_n/q_n)^{1/2} = n^{-1/2}(l_n/q_n)^{1/2} \rightarrow 0$ either under Assumption M.1(ii) or M.1(iii). The null rejection frequency of the Wald test in this setup of $\kappa = 1$ is not very good at $n = 200$ but dramatically improves with $n = 500$. For the Wald test, the null rejection frequency depends on the condition in Theorem 3(iii), which is $n^{1/2-\kappa}l_n^{1/2} = n^{-1/2}l_n^{1/2}$, but it is not clear this will go to zero. The pointwise null rejection frequency of the Wald test is between 4-30% in Table 2a, but improves to 3-8% in Table 2e, with $n = 500$. As we show in the other tables with $l_n = 4$ or $n = 500$, we reach the same conclusions. The first diagnostic of the violation should be the OI test. If this test indicates violations, we can try to learn which instruments are invalid, by deleting the certain suspect instruments and running the OI test again. If the OI test does not detect violation, then the AR and Wald tests are generally robust to both minor and mild violations.
Table 1a: Power of the OI test, \( l_n = 1 \), \( \Pi = \bar{e}/n^{1/3} \), \( \kappa = 1/4 \)

<table>
<thead>
<tr>
<th>( q_n )</th>
<th>( C_{l_n} = 1 )</th>
<th>( C_{l_n} = 2 )</th>
<th>( C_{l_n} = 3 )</th>
<th>( C_{l_n} = 1 )</th>
<th>( C_{l_n} = 2 )</th>
<th>( C_{l_n} = 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>63.8</td>
<td>100.0</td>
<td>100.0</td>
<td>88.5</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>25</td>
<td>46.1</td>
<td>100.0</td>
<td>100.0</td>
<td>79.2</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>50</td>
<td>26.5</td>
<td>98.6</td>
<td>100.0</td>
<td>56.7</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>60</td>
<td>17.6</td>
<td>97.9</td>
<td>100.0</td>
<td>49.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>75</td>
<td>11.1</td>
<td>90.7</td>
<td>100.0</td>
<td>42.5</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>80</td>
<td>7.3</td>
<td>89.1</td>
<td>100.0</td>
<td>40.1</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>100</td>
<td>3.9</td>
<td>67.2</td>
<td>100.0</td>
<td>32.8</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Note: \( q_n \) represents the number of instruments. Critical values from \( \chi^2 \) distribution at 5% level for \( q_n - 1 = 14, 24, 49, 59, 79, 99 \) are 23.59, 36.34, 66.39, 77.38, 95.28, 100.74, 123.21, respectively. We report here the rejection percentages out of 1000 iterations for the OI test compared with \( \chi^2_{q_n-1} \) critical values for differing degrees of near exogeneity, \( C_{l_n} = 1, 2, 3 \). We have \( l_n = 1 \) invalid instrument, with one endogenous variable and \( C_{l_n}, \kappa \) show the degree of violation of exogeneity from Assumption M.1.

Table 1b: Power of the OI test, \( l_n = 1 \), \( \Pi = \bar{e}/n^{1/3} \), \( \kappa = 1/2 \)

<table>
<thead>
<tr>
<th>( q_n )</th>
<th>( C_{l_n} = 1 )</th>
<th>( C_{l_n} = 2 )</th>
<th>( C_{l_n} = 3 )</th>
<th>( C_{l_n} = 1 )</th>
<th>( C_{l_n} = 2 )</th>
<th>( C_{l_n} = 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>7.1</td>
<td>15.2</td>
<td>39.8</td>
<td>6.2</td>
<td>17.7</td>
<td>42.3</td>
</tr>
<tr>
<td>25</td>
<td>4.1</td>
<td>11.4</td>
<td>24.2</td>
<td>5.9</td>
<td>12.5</td>
<td>29.0</td>
</tr>
<tr>
<td>50</td>
<td>3.9</td>
<td>5.3</td>
<td>13.9</td>
<td>5.7</td>
<td>8.0</td>
<td>18.1</td>
</tr>
<tr>
<td>60</td>
<td>3.0</td>
<td>4.9</td>
<td>8.9</td>
<td>4.5</td>
<td>6.3</td>
<td>27.0</td>
</tr>
<tr>
<td>75</td>
<td>2.0</td>
<td>2.2</td>
<td>6.1</td>
<td>2.7</td>
<td>6.8</td>
<td>12.8</td>
</tr>
<tr>
<td>80</td>
<td>1.8</td>
<td>1.9</td>
<td>4.0</td>
<td>3.7</td>
<td>6.6</td>
<td>12.1</td>
</tr>
<tr>
<td>100</td>
<td>1.1</td>
<td>2.2</td>
<td>1.7</td>
<td>4.9</td>
<td>5.2</td>
<td>11.7</td>
</tr>
</tbody>
</table>

Note: See the note for Table 1a.
Table 1c: Power of the OI test, $l_n = 1$, $\Pi = \bar{e}/n^{1/3}$, $\kappa = 1$

<table>
<thead>
<tr>
<th>$q_n$</th>
<th>$C_{l_n} = 1$</th>
<th>$C_{l_n} = 2$</th>
<th>$C_{l_n} = 3$</th>
<th>$C_{l_n} = 1$</th>
<th>$C_{l_n} = 2$</th>
<th>$C_{l_n} = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>3.7</td>
<td>4.9</td>
<td>4.8</td>
<td>5.4</td>
<td>4.8</td>
<td>4.5</td>
</tr>
<tr>
<td>25</td>
<td>4.8</td>
<td>3.2</td>
<td>4.2</td>
<td>4.3</td>
<td>5.6</td>
<td>4.1</td>
</tr>
<tr>
<td>50</td>
<td>3.3</td>
<td>3.3</td>
<td>2.6</td>
<td>4.8</td>
<td>5.1</td>
<td>4.5</td>
</tr>
<tr>
<td>60</td>
<td>2.8</td>
<td>2.7</td>
<td>2.1</td>
<td>4.5</td>
<td>4.0</td>
<td>4.5</td>
</tr>
<tr>
<td>75</td>
<td>1.6</td>
<td>1.0</td>
<td>1.9</td>
<td>3.6</td>
<td>3.3</td>
<td>4.1</td>
</tr>
<tr>
<td>80</td>
<td>1.0</td>
<td>1.0</td>
<td>1.8</td>
<td>4.1</td>
<td>3.2</td>
<td>3.6</td>
</tr>
<tr>
<td>100</td>
<td>1.0</td>
<td>1.0</td>
<td>1.2</td>
<td>2.6</td>
<td>3.7</td>
<td>3.1</td>
</tr>
</tbody>
</table>

Note: See the note for Table 1a.

Table 2a: Null rejection probability of the tests at 5% level, $n = 200$, $l_n = 1$, $\Pi = \bar{e}/n^{1/3}$, $\kappa = 1$

<table>
<thead>
<tr>
<th>$q_n$</th>
<th>Anderson Rubin Test</th>
<th>Wald Test</th>
<th>Anderson Rubin Test</th>
<th>Wald Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$C_{l_n} = 0$</td>
<td>$C_{l_n} = 1$</td>
<td>$C_{l_n} = 2$</td>
<td>$C_{l_n} = 3$</td>
</tr>
<tr>
<td>15</td>
<td>4.9</td>
<td>4.2</td>
<td>4.0</td>
<td>5.0</td>
</tr>
<tr>
<td>25</td>
<td>3.4</td>
<td>3.4</td>
<td>3.4</td>
<td>3.4</td>
</tr>
<tr>
<td>50</td>
<td>3.2</td>
<td>3.4</td>
<td>2.6</td>
<td>3.2</td>
</tr>
<tr>
<td>60</td>
<td>1.7</td>
<td>2.3</td>
<td>1.3</td>
<td>1.9</td>
</tr>
<tr>
<td>75</td>
<td>1.9</td>
<td>1.7</td>
<td>2.4</td>
<td>1.9</td>
</tr>
<tr>
<td>80</td>
<td>1.5</td>
<td>1.6</td>
<td>1.3</td>
<td>1.6</td>
</tr>
<tr>
<td>100</td>
<td>0.5</td>
<td>0.9</td>
<td>1.4</td>
<td>1.3</td>
</tr>
</tbody>
</table>

Note: $q_n$ represents the number of instruments. Critical values from $\chi^2$ distribution at 5% level for $q_n = 15, 25, 50, 60, 75, 80, 100$ are 25.06, 37.70, 67.14, 79.10, 95.50, 101.77, 123.67, respectively. We report here the rejection percentages out of 1000 iterations for the Anderson-Rubin (CUE) test compared with $\chi^2_{q_n}$ critical values for differing degrees of near exogeneity, $C_{l_n} = 0, 1, 2, 3$. The Wald test is only compared with $\chi^2_1$ distribution. We have $l_n = 1$ invalid instrument, and $C_{l_n}$, $\kappa$ show the degree of violation of exogeneity from Assumption M.1. The null hypothesis is correct under this DGP.
Table 2b: Null rejection probability of the tests at 5% level, $n = 200$, $l_n = 1 \Pi = \bar{e}/n^{1/3}$, $\kappa = 1/2$

<table>
<thead>
<tr>
<th>$q_n$</th>
<th>$C_{l_n} = 0$</th>
<th>$C_{l_n} = 1$</th>
<th>$C_{l_n} = 2$</th>
<th>$C_{l_n} = 3$</th>
<th>$C_{l_n} = 0$</th>
<th>$C_{l_n} = 1$</th>
<th>$C_{l_n} = 2$</th>
<th>$C_{l_n} = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>4.1</td>
<td>5.8</td>
<td>16.7</td>
<td>34.6</td>
<td>5.0</td>
<td>4.2</td>
<td>5.9</td>
<td>6.9</td>
</tr>
<tr>
<td>25</td>
<td>3.5</td>
<td>5.3</td>
<td>11.7</td>
<td>26.1</td>
<td>4.1</td>
<td>5.2</td>
<td>6.1</td>
<td>4.6</td>
</tr>
<tr>
<td>50</td>
<td>2.0</td>
<td>3.3</td>
<td>4.8</td>
<td>12.0</td>
<td>7.6</td>
<td>8.8</td>
<td>7.3</td>
<td>7.7</td>
</tr>
<tr>
<td>60</td>
<td>2.1</td>
<td>2.8</td>
<td>4.6</td>
<td>10.6</td>
<td>11.8</td>
<td>8.8</td>
<td>9.7</td>
<td>9.8</td>
</tr>
<tr>
<td>75</td>
<td>2.3</td>
<td>2.5</td>
<td>2.3</td>
<td>7.1</td>
<td>16.2</td>
<td>15.3</td>
<td>16.0</td>
<td>18.5</td>
</tr>
<tr>
<td>80</td>
<td>1.4</td>
<td>2.6</td>
<td>1.7</td>
<td>5.9</td>
<td>21.8</td>
<td>18.2</td>
<td>19.8</td>
<td>18.5</td>
</tr>
<tr>
<td>100</td>
<td>0.9</td>
<td>1.4</td>
<td>1.8</td>
<td>3.5</td>
<td>32.1</td>
<td>31.1</td>
<td>28.2</td>
<td>29.9</td>
</tr>
</tbody>
</table>

Note: See the note for Table 2a.

Table 2c: Null rejection probability of the tests at 5% level, $n = 200$, $l_n = 1 \Pi = \bar{e}/n^{1/3}$, $\kappa = 1/4$

<table>
<thead>
<tr>
<th>$q_n$</th>
<th>$C_{l_n} = 0$</th>
<th>$C_{l_n} = 1$</th>
<th>$C_{l_n} = 2$</th>
<th>$C_{l_n} = 3$</th>
<th>$C_{l_n} = 0$</th>
<th>$C_{l_n} = 1$</th>
<th>$C_{l_n} = 2$</th>
<th>$C_{l_n} = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>3.9</td>
<td>63.7</td>
<td>100.0</td>
<td>100.0</td>
<td>3.8</td>
<td>7.7</td>
<td>6.5</td>
<td>4.3</td>
</tr>
<tr>
<td>25</td>
<td>4.6</td>
<td>44.2</td>
<td>100.0</td>
<td>100.0</td>
<td>2.5</td>
<td>5.3</td>
<td>7.1</td>
<td>7.4</td>
</tr>
<tr>
<td>50</td>
<td>3.5</td>
<td>23.9</td>
<td>99.2</td>
<td>100.0</td>
<td>7.2</td>
<td>7.3</td>
<td>13.9</td>
<td>14.5</td>
</tr>
<tr>
<td>60</td>
<td>1.7</td>
<td>16.9</td>
<td>98.3</td>
<td>100.0</td>
<td>10.3</td>
<td>14.9</td>
<td>15.0</td>
<td>18.1</td>
</tr>
<tr>
<td>75</td>
<td>2.1</td>
<td>11.7</td>
<td>93.5</td>
<td>100.0</td>
<td>14.1</td>
<td>18.8</td>
<td>21.4</td>
<td>23.6</td>
</tr>
<tr>
<td>80</td>
<td>2.0</td>
<td>9.4</td>
<td>88.4</td>
<td>100.0</td>
<td>18.2</td>
<td>18.7</td>
<td>25.2</td>
<td>27.4</td>
</tr>
<tr>
<td>100</td>
<td>0.8</td>
<td>3.9</td>
<td>70.9</td>
<td>100.0</td>
<td>27.3</td>
<td>34.9</td>
<td>31.0</td>
<td>36.6</td>
</tr>
</tbody>
</table>

Note: See the note for Table 2a.

Table 2d: Null rejection probability of the tests at 5% level, $n = 500$, $l_n = 1 \Pi = \bar{e}/n^{1/3}$, $\kappa = 1$

<table>
<thead>
<tr>
<th>$q_n$</th>
<th>$C_{l_n} = 0$</th>
<th>$C_{l_n} = 1$</th>
<th>$C_{l_n} = 2$</th>
<th>$C_{l_n} = 3$</th>
<th>$C_{l_n} = 0$</th>
<th>$C_{l_n} = 1$</th>
<th>$C_{l_n} = 2$</th>
<th>$C_{l_n} = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>3.9</td>
<td>5.4</td>
<td>3.5</td>
<td>5.4</td>
<td>4.2</td>
<td>4.4</td>
<td>3.7</td>
<td>4.3</td>
</tr>
<tr>
<td>25</td>
<td>3.9</td>
<td>4.1</td>
<td>3.6</td>
<td>5.8</td>
<td>3.7</td>
<td>4.2</td>
<td>4.6</td>
<td>3.2</td>
</tr>
<tr>
<td>50</td>
<td>3.7</td>
<td>3.6</td>
<td>4.8</td>
<td>2.5</td>
<td>2.5</td>
<td>3.6</td>
<td>2.3</td>
<td>2.8</td>
</tr>
<tr>
<td>60</td>
<td>4.1</td>
<td>3.9</td>
<td>4.9</td>
<td>3.4</td>
<td>3.3</td>
<td>4.7</td>
<td>3.0</td>
<td>3.3</td>
</tr>
<tr>
<td>75</td>
<td>3.4</td>
<td>4.2</td>
<td>4.0</td>
<td>5.1</td>
<td>2.6</td>
<td>2.4</td>
<td>3.3</td>
<td>3.3</td>
</tr>
<tr>
<td>80</td>
<td>4.2</td>
<td>4.6</td>
<td>3.5</td>
<td>3.7</td>
<td>2.8</td>
<td>2.8</td>
<td>3.4</td>
<td>4.1</td>
</tr>
<tr>
<td>100</td>
<td>2.5</td>
<td>4.0</td>
<td>3.1</td>
<td>3.2</td>
<td>2.7</td>
<td>3.8</td>
<td>2.7</td>
<td>3.4</td>
</tr>
</tbody>
</table>

Note: See the note for Table 2a.
Table 2e: Null rejection probability of the tests at 5% level, $n = 500$, $l_n = 1$ $\Pi = \bar{e}/n^{1/3}$, $\kappa = 1/2$

<table>
<thead>
<tr>
<th>$q_n$</th>
<th>$C_{l_n} = 0$</th>
<th>$C_{l_n} = 1$</th>
<th>$C_{l_n} = 2$</th>
<th>$C_{l_n} = 3$</th>
<th>$C_{l_n} = 0$</th>
<th>$C_{l_n} = 1$</th>
<th>$C_{l_n} = 2$</th>
<th>$C_{l_n} = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>5.0</td>
<td>7.1</td>
<td>15.3</td>
<td>40.0</td>
<td>4.5</td>
<td>6.1</td>
<td>6.2</td>
<td>8.2</td>
</tr>
<tr>
<td>25</td>
<td>3.8</td>
<td>5.6</td>
<td>11.5</td>
<td>30.3</td>
<td>2.9</td>
<td>4.4</td>
<td>5.3</td>
<td>4.3</td>
</tr>
<tr>
<td>50</td>
<td>4.3</td>
<td>6.6</td>
<td>10.3</td>
<td>19.9</td>
<td>3.3</td>
<td>3.9</td>
<td>2.4</td>
<td>3.5</td>
</tr>
<tr>
<td>60</td>
<td>3.4</td>
<td>4.6</td>
<td>8.7</td>
<td>14.4</td>
<td>3.0</td>
<td>2.2</td>
<td>3.9</td>
<td>3.6</td>
</tr>
<tr>
<td>75</td>
<td>3.8</td>
<td>5.4</td>
<td>7.8</td>
<td>16.6</td>
<td>3.3</td>
<td>3.3</td>
<td>2.1</td>
<td>2.5</td>
</tr>
<tr>
<td>80</td>
<td>3.3</td>
<td>3.8</td>
<td>7.5</td>
<td>16.6</td>
<td>2.7</td>
<td>3.0</td>
<td>3.7</td>
<td>3.7</td>
</tr>
<tr>
<td>100</td>
<td>4.5</td>
<td>4.2</td>
<td>6.5</td>
<td>10.2</td>
<td>3.4</td>
<td>2.8</td>
<td>4.0</td>
<td>3.2</td>
</tr>
</tbody>
</table>

Note: See the note for Table 2a.

Table 2f: Null rejection probability of the tests at 5% level, $n = 500$, $l_n = 1$ $\Pi = \bar{e}/n^{1/3}$, $\kappa = 1/4$

<table>
<thead>
<tr>
<th>$q_n$</th>
<th>$C_{l_n} = 0$</th>
<th>$C_{l_n} = 1$</th>
<th>$C_{l_n} = 2$</th>
<th>$C_{l_n} = 3$</th>
<th>$C_{l_n} = 0$</th>
<th>$C_{l_n} = 1$</th>
<th>$C_{l_n} = 2$</th>
<th>$C_{l_n} = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>4.8</td>
<td>88.9</td>
<td>100.0</td>
<td>100.0</td>
<td>4.2</td>
<td>5.4</td>
<td>5.7</td>
<td>4.1</td>
</tr>
<tr>
<td>25</td>
<td>3.2</td>
<td>79.9</td>
<td>100.0</td>
<td>100.0</td>
<td>3.8</td>
<td>3.6</td>
<td>8.4</td>
<td>11.1</td>
</tr>
<tr>
<td>50</td>
<td>4.3</td>
<td>57.4</td>
<td>100.0</td>
<td>100.0</td>
<td>3.9</td>
<td>3.5</td>
<td>4.8</td>
<td>16.3</td>
</tr>
<tr>
<td>60</td>
<td>4.6</td>
<td>47.8</td>
<td>100.0</td>
<td>100.0</td>
<td>3.3</td>
<td>2.4</td>
<td>5.4</td>
<td>15.3</td>
</tr>
<tr>
<td>75</td>
<td>5.1</td>
<td>45.0</td>
<td>100.0</td>
<td>100.0</td>
<td>3.0</td>
<td>2.1</td>
<td>3.2</td>
<td>15.0</td>
</tr>
<tr>
<td>80</td>
<td>4.2</td>
<td>37.9</td>
<td>100.0</td>
<td>100.0</td>
<td>2.6</td>
<td>3.0</td>
<td>4.1</td>
<td>13.3</td>
</tr>
<tr>
<td>100</td>
<td>4.4</td>
<td>33.1</td>
<td>100.0</td>
<td>100.0</td>
<td>2.7</td>
<td>3.6</td>
<td>4.7</td>
<td>13.5</td>
</tr>
</tbody>
</table>

Note: See the note for Table 2a.

Table 3a: Power of the OI test at 5% level, $l_n = 4$, $\Pi = \bar{e}/n^{1/3}$, $\kappa = 1/4$

<table>
<thead>
<tr>
<th>$q_n$</th>
<th>$C_{l_n} = 1$</th>
<th>$C_{l_n} = 2$</th>
<th>$C_{l_n} = 1$</th>
<th>$C_{l_n} = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>25</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>50</td>
<td>99.9</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>60</td>
<td>97.1</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>75</td>
<td>91.8</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>80</td>
<td>88.2</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>100</td>
<td>69.1</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Note: Compared to Table 1, the only difference is $l_n = 4$.  

23
Table 3b: Power of the OI test at 5% level, $l_n = 4$, $\Pi = \bar{e}/n^{1/3}, \kappa = 1/2$  

<table>
<thead>
<tr>
<th>$q_n$</th>
<th>$C_{l_n} = 1$</th>
<th>$C_{l_n} = 2$</th>
<th>$C_{l_n} = 1$</th>
<th>$C_{l_n} = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>12.7</td>
<td>64.3</td>
<td>14.1</td>
<td>61.5</td>
</tr>
<tr>
<td>25</td>
<td>12.9</td>
<td>52.7</td>
<td>11.2</td>
<td>50.8</td>
</tr>
<tr>
<td>50</td>
<td>5.9</td>
<td>28.7</td>
<td>8.3</td>
<td>36.4</td>
</tr>
<tr>
<td>60</td>
<td>5.3</td>
<td>22.5</td>
<td>9.2</td>
<td>30.7</td>
</tr>
<tr>
<td>75</td>
<td>3.8</td>
<td>13.8</td>
<td>5.2</td>
<td>24.6</td>
</tr>
<tr>
<td>80</td>
<td>2.4</td>
<td>11.5</td>
<td>7.0</td>
<td>14.7</td>
</tr>
<tr>
<td>100</td>
<td>1.0</td>
<td>4.7</td>
<td>5.6</td>
<td>21.0</td>
</tr>
</tbody>
</table>

Note: Compared to Table 1, the only difference is $l_n = 4$.

Table 3c: Power of the OI test at 5% level, $l_n = 4$, $\Pi = \bar{e}/n^{1/3}, \kappa = 1$  

<table>
<thead>
<tr>
<th>$n$ = 200</th>
<th>$n$ = 500</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q_n$</td>
<td>$C_{l_n} = 1$</td>
</tr>
<tr>
<td>15</td>
<td>5.1</td>
</tr>
<tr>
<td>25</td>
<td>3.4</td>
</tr>
<tr>
<td>50</td>
<td>2.4</td>
</tr>
<tr>
<td>60</td>
<td>2.5</td>
</tr>
<tr>
<td>75</td>
<td>1.8</td>
</tr>
<tr>
<td>80</td>
<td>1.7</td>
</tr>
<tr>
<td>100</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Note: Compared to Table 1, the only difference is $l_n = 4$.

Table 4a: Null rejection probability of the tests at 5% level, $l_n = 4$, $\Pi = \bar{e}/n^{1/3}, \kappa = 1$  

<table>
<thead>
<tr>
<th>$n$ = 200</th>
<th>$n$ = 500</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q_n$</td>
<td>Anderson Rubin Test</td>
</tr>
<tr>
<td>15</td>
<td>4.3</td>
</tr>
<tr>
<td>25</td>
<td>4.6</td>
</tr>
<tr>
<td>50</td>
<td>3.1</td>
</tr>
<tr>
<td>60</td>
<td>2.0</td>
</tr>
<tr>
<td>75</td>
<td>2.2</td>
</tr>
<tr>
<td>80</td>
<td>1.3</td>
</tr>
<tr>
<td>100</td>
<td>1.8</td>
</tr>
</tbody>
</table>

Note: Compared to Table 2, the only difference is $l_n = 4$.  

24
Table 4b: Null rejection probability of the tests at 5% level, $l_n = 4$, $\Pi = \bar{e}/n^{1/3}, \kappa = 1/2$

<table>
<thead>
<tr>
<th>$q_n$</th>
<th>$C_{l_n} = 1$</th>
<th>$C_{l_n} = 2$</th>
<th>$C_{l_n} = 1$</th>
<th>$C_{l_n} = 2$</th>
<th>$C_{l_n} = 1$</th>
<th>$C_{l_n} = 2$</th>
<th>$C_{l_n} = 1$</th>
<th>$C_{l_n} = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>13.5</td>
<td>69.6</td>
<td>13.5</td>
<td>19.3</td>
<td>17.5</td>
<td>71.8</td>
<td>16.3</td>
<td>25.1</td>
</tr>
<tr>
<td>25</td>
<td>11.3</td>
<td>54.2</td>
<td>8.1</td>
<td>12.2</td>
<td>17.2</td>
<td>56.4</td>
<td>8.7</td>
<td>13.6</td>
</tr>
<tr>
<td>50</td>
<td>5.4</td>
<td>26.7</td>
<td>7.6</td>
<td>13.0</td>
<td>11.1</td>
<td>35.9</td>
<td>5.2</td>
<td>6.3</td>
</tr>
<tr>
<td>60</td>
<td>4.2</td>
<td>22.4</td>
<td>12.1</td>
<td>15.9</td>
<td>8.9</td>
<td>33.3</td>
<td>4.4</td>
<td>6.1</td>
</tr>
<tr>
<td>75</td>
<td>3.6</td>
<td>15.8</td>
<td>15.2</td>
<td>21.7</td>
<td>6.5</td>
<td>28.4</td>
<td>2.9</td>
<td>5.5</td>
</tr>
<tr>
<td>80</td>
<td>2.7</td>
<td>9.8</td>
<td>16.3</td>
<td>22.4</td>
<td>10.0</td>
<td>25.6</td>
<td>3.9</td>
<td>4.4</td>
</tr>
<tr>
<td>100</td>
<td>1.3</td>
<td>5.8</td>
<td>28.4</td>
<td>32.4</td>
<td>5.6</td>
<td>21.0</td>
<td>4.5</td>
<td>4.6</td>
</tr>
</tbody>
</table>

Note: Compared to Table 2, the only difference is $l_n = 4$.

Table 4c: Null rejection probability of the tests at 5% level, $l_n = 4$, $\Pi = \bar{e}/n^{1/3}, \kappa = 1/4$

<table>
<thead>
<tr>
<th>$q_n$</th>
<th>$C_{l_n} = 1$</th>
<th>$C_{l_n} = 2$</th>
<th>$C_{l_n} = 1$</th>
<th>$C_{l_n} = 2$</th>
<th>$C_{l_n} = 1$</th>
<th>$C_{l_n} = 2$</th>
<th>$C_{l_n} = 1$</th>
<th>$C_{l_n} = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>100.0</td>
<td>100.0</td>
<td>19.6</td>
<td>13.2</td>
<td>100.0</td>
<td>100.0</td>
<td>18.7</td>
<td>26.8</td>
</tr>
<tr>
<td>25</td>
<td>100.0</td>
<td>100.0</td>
<td>13.4</td>
<td>8.8</td>
<td>100.0</td>
<td>100.0</td>
<td>7.6</td>
<td>3.6</td>
</tr>
<tr>
<td>50</td>
<td>99.4</td>
<td>100.0</td>
<td>16.4</td>
<td>14.2</td>
<td>100.0</td>
<td>100.0</td>
<td>4.0</td>
<td>5.3</td>
</tr>
<tr>
<td>60</td>
<td>97.4</td>
<td>100.0</td>
<td>18.2</td>
<td>17.2</td>
<td>100.0</td>
<td>100.0</td>
<td>4.1</td>
<td>7.5</td>
</tr>
<tr>
<td>75</td>
<td>92.3</td>
<td>100.0</td>
<td>23.1</td>
<td>21.6</td>
<td>100.0</td>
<td>100.0</td>
<td>4.6</td>
<td>8.1</td>
</tr>
<tr>
<td>80</td>
<td>89.4</td>
<td>100.0</td>
<td>22.6</td>
<td>22.4</td>
<td>100.0</td>
<td>100.0</td>
<td>3.8</td>
<td>9.9</td>
</tr>
<tr>
<td>100</td>
<td>66.8</td>
<td>100.0</td>
<td>33.2</td>
<td>34.2</td>
<td>100.0</td>
<td>100.0</td>
<td>5.0</td>
<td>11.0</td>
</tr>
</tbody>
</table>

Note: Compared to Table 2, the only difference is $l_n = 4$. 

25
5 Conclusion

In this paper, we extend the many weak moments asymptotics in GEL from Newey and Windmeijer (2009a) to the nearly exogenous case. As a result of our research, we can see the effects of a nearly exogenous instrument. The OI test can detect large violations of exogeneity. The OI test cannot detect minor violations, but the AR/Wald tests are robust to these violations. We think that an important extension for the future work may entail estimation of these models with structural change.

Appendix: Mathematical Proofs

In this section, we provide proofs that do not repeat the material in Newey and Windmeijer (2009b). The detailed proofs can be found in the Supplement Appendix. As on p.705 of Newey and Windmeijer (2009a), we set up the following reparameterization and simplification:

\[ \delta = \delta(\theta) = S_n'(\theta - \theta_0)/\mu_n. \]

We denote the new reparameterized objective function as \( \hat{Q}(\delta) \), where \( \hat{Q}(\delta) \) is simply \( \hat{Q}(\theta_0 + \mu_nS_n^{-1}\delta) \). To simplify the notation we set \( \hat{Q}(\delta = 0) = \hat{Q}(0) \), and \( \tilde{g} = \sum_{i=1}^{n} g_i(\theta_0)/n \). Consistency of the estimator is shown in the Supplement Appendix. Next, following the same steps on p.16-18 in Newey and Windmeijer (2009b) by Lemma A.12, we obtain equation (A.7) in Newey and Windmeijer (2009b)

\[ n\mu_n^{-1}\partial\hat{Q}(0)/\partial\delta = nS_n^{-1}[G'\Omega^{-1}\tilde{g} + (n^{-1}\sum_{i=1}^{n} U_i')\Omega^{-1}\tilde{g}] + o_p(1), \quad (A.1) \]

where the score is evaluated at \( \delta = 0 \). First, take a \( p \times 1 \) vector \( \zeta \) with \( ||\zeta|| = 1 \). Note that in (SA.17), by denoting \( \tilde{U} = n^{-1}\sum_{i=1}^{n} U_i \),

\[ n\mu_n^{-1}\zeta'\partial\hat{Q}(0)/\partial\delta = n\zeta'S_n^{-1}[G'\Omega^{-1}(\tilde{g}_d + E \tilde{g}) + \tilde{U}'\Omega^{-1}(\tilde{g}_d + E \tilde{g})] + o_p(1), \quad (A.2) \]

where \( \tilde{g}_d = \tilde{g} - E \tilde{g}_i \). We rewrite (A.2) as

\[ n\mu_n^{-1}\zeta'\partial\hat{Q}(0)/\partial\delta = n\zeta'S_n^{-1}[G'\Omega^{-1}(\tilde{g}_d + E \tilde{g}_i) + \tilde{U}'\Omega^{-1}(\tilde{g}_d + E \tilde{g}_i)] + o_p(1) \]

\[ = n\zeta'S_n^{-1}[G'\Omega^{-1}\tilde{g}_d + (\tilde{U} - E \tilde{U})'\Omega^{-1}\tilde{g}_d] + o_p(1) \]

\[ + n\zeta'S_n^{-1}E\tilde{U}'\Omega^{-1}\tilde{g}_d \]

\[ + n\zeta'S_n^{-1}[G + \tilde{U}']\Omega^{-1}E \tilde{g}_i + o_p(1). \quad (A.3) \]

The terms (A.3)-(A.5) can be rewritten as

\[ n\mu_n^{-1}\zeta'\partial\hat{Q}(0)/\partial\delta = n\zeta'S_n^{-1}[G'\Omega^{-1}\tilde{g}_d + (\tilde{U} - E \tilde{U})'\Omega^{-1}\tilde{g}_d] + o_p(1) \]

\[ + n\zeta'S_n^{-1}E\tilde{U}'\Omega^{-1}\tilde{g}_d \]

\[ + n\zeta'S_n^{-1}[G + \tilde{U}']\Omega^{-1}E \tilde{g}_i + o_p(1). \quad (A.4) \]

\[ \]
we have three possibilities: compared with Newey and Windmeijer (2009a). So the violation of exogeneity may add up to more case in detail in the Remarks after Lemma A.1.

\[ \tau \]

Lemma A.1, and they are related to \( nS \), where \( \nu \) and \( \nu \) where. We use the following to get (A.8)-(A.10), and \( Egi = C_1/n^\kappa \):

\[
n\mu_n^{-1} \zeta_n^{-1} [G' \Omega^{-1}(\hat{g}_d + Egi + U\Omega^{-1}(\hat{g}_d + Egi)] + o_p(1),
\]

\[
= n\zeta_n^{-1} [G' \Omega^{-1}(\hat{g}_d) + (\hat{U} - E\hat{U})\Omega^{-1}\hat{g}_d] + n^{1/2}\zeta_n^{-1} E\hat{U}'\Omega^{-1}\hat{g}_d + n^{1/2}\zeta_n^{-1} G'\Omega^{-1/2}(n^{1/2-\kappa}\Omega^{-1/2}C_1) + n^{1/2}\zeta_n^{-1}(\hat{U} - E\hat{U})\Omega^{-1/2}(n^{1/2-\kappa}\Omega^{-1/2}C_1) + n^{1/2}\zeta_n^{-1} E\hat{U}'\Omega^{-1/2}(n^{1/2-\kappa}\Omega^{-1/2}C_1) + o_p(1)
\]

We define the notation before the following lemma. First, we start with \( \tau \). We let \( nS_n^{-1}G'(n^{1/2-\kappa}\Omega^{-1}C_1) \to \tau \), and \( \tau \) can be zero or nonzero constant. There are three possible cases in Lemma A.1, and they are related to \( \tau \), and to the behavior of \( n^{1/2-\kappa}n^{1/2} \). We explain each case in detail in the Remarks after Lemma A.1.

The result shows that violation of exogeneity can alter the standard limit for the score in Newey and Windmeijer (2009b). The new limit is normally distributed but with an extra variance term compared with Newey and Windmeijer (2009a). So the violation of exogeneity may add up to more than a simple constant drift term. The following lemma provides the limit of the score.

**Lemma A.1.** Under Assumptions M.1–M.6 and \( S_n^{-1}E[(U_i - EU_i)'\Omega^{-1}(U_i - EU_i)]S_n^{-1}' \to \Lambda \), we have three possibilities:

(i). If \( n^{1/2-\kappa}n^{1/2} = O(1) \) and \( ||n^{1/2}S_n^{-1}G'|| = O(1) \), then

\[
n\mu_n^{-1} \frac{\partial \hat{Q}_n(0)}{\partial \delta} \to \nu_2,
\]

where \( \nu_2 \equiv N(\tau, H + \Delta + \Lambda) \), where

\[
\Delta = \lim_{n \to \infty} S_n^{-1}E[(U_i - EU_i)'\Omega^{-1/2}(n^{1/2-\kappa}\Omega^{-1/2}C_1\Omega^{-1/2})(U_i - EU_i)]S_n^{-1}'.
\]

(ii). If \( n^{1/2-\kappa}n^{1/2} = O(1) \) and \( ||n^{1/2}S_n^{-1}G'|| = o(1) \), then

\[
n\mu_n^{-1} \frac{\partial \hat{Q}_n(0)}{\partial \delta} \to \nu_{2nd},
\]

where \( \nu_{2nd} \equiv N(0, H + \Delta + \Lambda) \).

(iii). If \( n^{1/2-\kappa}n^{1/2} = o(1) \), then

\[
n\mu_n^{-1} \frac{\partial \hat{Q}_n(0)}{\partial \delta} \to \nu_1,
\]

where \( \nu_1 \equiv N(0, H + \Lambda) \).

Remarks.
1. First, the standard result is Case (iii) which is the limit of the score in Lemma A.12 in Newey and Windmeijer (2009b). So their lemma extends to minor violations with many imperfect instruments. It is critical to see even \( l_n = q_n \) is possible in that case, but to satisfy the condition of \( n^{1/2-\kappa}l_n = o(1) \), where the violation should be really minor, \( \kappa \) should be large. Case (iii) restricts \( 1/2 < \kappa < \infty \).

2. Nonstandard Cases are (i) and (ii). Compared with Case (iii), we allow for \( \kappa = 1/2 \) in Cases (i)(ii). Note that by using the definition of \( C_1 \), we can get \( \|C_1\| = O(l_n^{1/2}) \). We have \( n^{1/2-\kappa}\|\Omega^{-1/2}C_1\| = O(n^{1/2-\kappa}l_n^{1/2}) = O(1) \). Also in case (i) we assume \( \|n^{1/2}S_n^{-1}G\| = O(1) \). Case (ii) is the same as in case (i) but with weak identification, so the drift in the second normal distribution is zero. To see this point about relation of \( \tau \) to identification issue, see Examples 1-2.

3. The key to the limit results is the condition of \( n^{1/2-\kappa}l_n^{1/2} \). Depending on whether it goes to zero or not, we get the standard limit as in Case (iii) or non-standard limits in Cases (i)-(ii).

4. Note that the results in Lemma A.1 hold regardless of \( l_n/q_n \to f, 0 < f \leq 1 \), or \( f = 0 \) (i.e. Assumptions M.1(ii) or M.1(iii), respectively).

5. There are more possibilities than the above results. We can have \( 1/2 < \kappa < \infty \) and \( n^{1/2-\kappa}l_n^{1/2} \to \infty \). In the case of strong/near weak identification, we see \( \tau \to \infty \), and in the case of weak identification, \( \tau \) can be zero, constant, or infinity. So, there is uncertainty in that case. Also, in the case of \( 0 < \kappa \leq 1/2 \) with Assumption M.1(iii), \( n^{1/2-\kappa}l_n^{1/2} \to \infty \), and the resulting drift will diverge to infinity when there is strong/near weak identification. This divergence is clear from the proof of Lemma SA.6. The case of \( 0 < \kappa \leq 1/2 \) with \( l_n/q_n \to f, 0 < f \leq 1 \) (Assumption M.1(ii)) is not analyzed since the estimator is not consistent via the proof of Theorem 1. With \( 0 < \kappa \leq 1/2 \), when there is weak identification, there is uncertainty about the drift behavior, which can be seen in the proof.

**Proof of Lemma A.1.** The detailed proof of this lemma is in the Supplement Appendix as Lemma SA.6. We derive the limit using (A.6) through (A.10), and Lemma A.10 of Newey and Windmeijer (2009b). We consider the sum of the terms (A.6) and (A.9) first.

All the conditions of Lemma A.10 in Newey and Windmeijer (2009b) are satisfied by Assumptions M.1-M.6. This verification can be seen by following the proof of Lemma A.12 in Newey and Windmeijer (2009b). So the limit of the sum of (A.6) and (A.9) in the score is

\[
ns_n^{-1}[G'\Omega^{-1}(\tilde{g} - E\tilde{g})] + ns_n^{-1}[n^{-1}\sum_{i=1}^{n}(U_i - EU_i)'\Omega^{-1}(\tilde{g} - E\tilde{g})]
+ n^{1/2}\zeta_s n^{-1}(\bar{U} - E\bar{U})'\Omega^{-1/2}(n^{1/2-\kappa}\Omega^{-1/2}C_1)
\xrightarrow{d} N(0, H + \Delta + \Lambda),
\]

(A.11)

by using the Cramer-Wold device. This is a key result and the extra \( \Delta \) term compared to standard limit in Newey and Windmeijer (2009a) in the limit is due to exogeneity violation.

We show then, in (SA.41) that (A.7) becomes

\[
n\zeta_s n^{-1}E\bar{U}'\Omega^{-1}\tilde{g}_d \xrightarrow{p} 0.
\]

However, in our case, one of the key terms in the score is (A.8)

\[
n^{1/2}\zeta_s n^{-1}G'\Omega^{-1/2}(n^{1/2-\kappa}\Omega^{-1/2}C_1).
\]
When there is no violation of exogeneity, \( C_1 = 0 \), we know that \( Eg_i = 0 \), hence this term is zero under the asymptotics of Newey and Windmeijer (2009b). In this paper, because of the exogeneity violation, this term plays an important role. The term above can be written via Assumption M.1

\[
n^{1/2} S_n^{-1} G' \Omega^{-1/2} (n^{1/2} / C_1) \rightarrow \tau,\]

where \( \tau \neq 0 \) with \( n^{1/2} / C_1 = O(1) \), and with the case of strong/near-weak identification, we have \( \|n^{1/2} S_n^{-1} G'\| = O(1) \). If there is weak identification, we have \( \|n^{1/2} S_n^{-1} G'\| = o(1) \), then \( \tau = 0 \).

These are shown by Examples 1-2. If \( n^{1/2} / C_1 = o(1) \), then again \( \tau = 0 \). The proofs are shown in Lemma SA.6. Next, the proof of Lemma SA.6 shows that, since \( \hat{U} \) is local to zero, by \( EU_i \) definition, the term (A.10)

\[
n^{1/2} \zeta' S_n^{-1} \hat{G}^i \Omega^{-1/2} (n^{1/2} / C_1) \rightarrow 0.\]

Combine the results in the score to form three cases for the limit. The first case is when \( n^{1/2} / C_1 = O(1) \) and \( \|n^{1/2} S_n^{-1} G'\| = O(1) \). The limit is \( \nu_2 \). The second case is when \( n^{1/2} / C_1 = O(1) \) and \( \|n^{1/2} S_n^{-1} G'\| = o(1) \). The limit is \( \nu_{2nd} \). The third case is when \( n^{1/2} / C_1 = o(1) \), and the limit is \( \nu_1 \) only.

There are also very detailed proofs in Lemma SA.6 about these issues. Note that \( \tau \) can also go to infinity if \( n^{1/2} / C_1 \rightarrow \infty \) even with \( 1/2 < \kappa < \infty \). This divergence is also discussed in the proof of Lemma SA.6.

Q.E.D.

We next provide Lemma A.2, which deals with the limit of second order partial derivative of the objective function. Lemma A.2 is an extension of Lemma A.13 in Newey and Windmeijer (2009b) to the case of near exogeneity. The proof simply follows from Lemma A.13 of Newey and Windmeijer (2009b). The detailed proof is in Lemma SA.8 in the Supplement Appendix. Note that the acronym “wpa1” stands for “with probability approaching one” through the article.

**Lemma A.2.** Under Assumptions M.1 – M.6, there is an open convex set \( N_n \) so that \( 0 \in N_n \), and wpa1 \( \delta \in N_n \), \( Q(\delta) \) is twice continuously differentiable on \( N_n \), and for any \( \delta \), that is an element of \( N_n \) wpa1, with \( n^{1/2} / \sqrt{n} \) \( Q(\delta) = o(1) \)

\[
\mu_n^{-2} n \frac{\partial^2 Q(\delta)}{\partial \delta \partial \delta'} \Rightarrow H.
\]

Remarks

1. It is clear that this lemma works in cases of minor violations, \( 1/2 < \kappa < \infty \), as long as the number of invalid instruments satisfy the \( n^{1/2} / C_1 = O(1) \) assumption. This lemma is also valid when there is a mild exogeneity violation with a fixed number of invalid instruments. The case of \( \kappa = 1/2 \), \( l_n = l \) is also covered by this Lemma.

2. In one interesting case, if \( 0 < \kappa < 1/2 \) (larger violations) with \( l_n / q_n \rightarrow 0 \) (Assumption M.1(iii)), then it is possible to satisfy in, certain cases, \( n^{1/2} / \sqrt{n} \) \( Q(\delta) = o(1) \). This case can happen when the number of imperfect instruments is not numerous.

**Proof of Theorem 2.** Given the first order condition \( \frac{\partial Q(\delta)}{\partial \theta} = 0 \), and the Taylor series expansion around \( \theta_0 \), following Lemmata A.1, A.2 provides the result. This proof technique can also be seen in detail in the proof of Theorem 3 in p.26 of Newey and Windmeijer (2009b). Q.E.D.
The following result is useful in deriving the variance estimator. The proof follows from the proof of Lemma A.14 in p.25-26 of Newey and Windmeijer (2009b). The detailed proof is in Lemma SA.9 in the Supplement.

**Lemma A.3.** Under Assumptions M.1-M6, with \( n^{1/2-\kappa} \sqrt{\ln n} / \sqrt{q_n} \rightarrow 0 \)

\[
nS_n^{1/2} \hat{D}(\hat{\theta})' \hat{\Omega}^{-1} \hat{D}(\hat{\theta}) S_n^{1/2} \rightarrow H + \Lambda.
\]

The next result is an extension of Lemma A.15 of Newey and Windmeijer (2009b). This Lemma extends their result from the perfect exogeneity to the near exogeneity. This Lemma is one of the main results of this article.

**Lemma A.4.** If \( E[(g_i' \Omega^{-1} g_i)^2]/q_n n \rightarrow 0 \), with \( n^{1/2-\kappa} (t_n/q_n)^{1/2} = o(1) \), then

\[
\frac{n \hat{g}' \Omega^{-1} \hat{g} - q_n}{\sqrt{2q_n}} \xrightarrow{d} N(0,1).
\]

Remark. This result is crucial in establishing the subsequent results. Even though we have non-zero mean of the moment functions via Assumption M.1, we achieve the same result as in Lemma A.15, from p.28 of Newey and Windmeijer (2009b).

**Proof of Lemma A.4.** See that

\[
\frac{n \hat{g}' \Omega^{-1} \hat{g} - q_n}{\sqrt{2q_n}} = \frac{\sum_{i=1}^{n} g_i' \Omega^{-1} g_i/n - q_n}{\sqrt{2q_n}} + \frac{\sum_{i \neq j} g_i' \Omega^{-1} g_j/n}{\sqrt{2q_n}}.
\]

We consider each term. First, we analyze \( E(g_i' \Omega^{-1} g_i) = q_n \); so as in the proof of Lemma A.15 of Newey and Windmeijer (2009b),

\[
\frac{\sum_{i=1}^{n} g_i' \Omega^{-1} g_i/n - q_n}{\sqrt{2q_n}} \xrightarrow{p} 0.
\]

This proof is true since \( E[(g_i' \Omega^{-1} g_i)^2]/nq_n \rightarrow 0 \). (A.12)

(A.12) holds under Assumption M.4(i), and by the result that the minimum eigenvalue of \( \Omega^{-1} \) is being finite. Next, for the second term above, by adding and subtracting we have

\[
\frac{\sum_{i \neq j} g_i' \Omega^{-1} g_j/n}{\sqrt{2q_n}} = \frac{\sum_{i \neq j} (g_i - Eg_i)' \Omega^{-1} (g_j - Eg_j)/n}{\sqrt{2q_n}} + \frac{\sum_{i \neq j} (Eg_i)' \Omega^{-1} (g_j - Eg_j)/n}{\sqrt{2q_n}}
\]

\[
\xrightarrow{p} \sum_{i \neq j} (Eg_i)' \Omega^{-1} (Eg_j)/n.
\]

The first term on the right side of (A.13) can be analyzed by Lemma A.10 of Newey and Windmeijer (2009b), or Lemma SA.4 here. Note that \( \Sigma_{YY} = EY_i' Y_i', \Sigma_{ZZ} = EZ_i' Z_i', \) and \( \Sigma_{YZ} = EY_i' Z_i'. \) See that

\[
EY_i' Y_i' = \frac{I_{q_n}}{n \sqrt{2q_n}} - \frac{\Omega^{-1/2} C_1 C_1' \Omega^{-1/2}}{n^{2\kappa} \sqrt{2q_n}}.
\]

30
Note that the second term is smaller in order than the first one since \( \|\Omega^{-1/2}C_1C_1'\Omega^{-1/2}\| = O(l_n) \) by \( C_1 \) definition, and \( l_n/n^{2\kappa} \leq q_n \). Following the proof of Lemma A.15 of p.28 in Newey and Windmeijer (2009b), set \( Z_i = Y_i = \Omega^{-1/2}(g_i - Eg_i)/\sqrt{n}(2q_n)^{1/4} \)

\[
n^2 tr(\Psi) = n^2 tr(\Sigma_{yy}\Sigma_{zz} + \Sigma_{yz}^2) \to 1.
\]

Note that we define \( E\tilde{g}z = Eigmax(\Sigma_{zz}) \), \( E\tilde{g}y = Eigmax(\Sigma_{yy}) \). First, \( E\tilde{g}z = E\tilde{g}y \leq n^{-1}(2q_n)^{-1/2} \) by Assumptions M.1 and M.4. From p.28 of Newey and Windmeijer (2009b) or substituting \( Y_i, Z_i \), we get

\[
E(g_i\Omega^{-1}g_i^2)/q_n n \to 0,
\]

\[
g_n n^4 E\tilde{g}z^2 E\tilde{g}y^2 \to 0,
\]

\[
n^3[E\tilde{g}z^2 E\|Y_i\|^4 + E\tilde{g}y^2 E\|Z_i\|^4] \to 0,
\]

\[
n^2 E\|Y_i\|^4 E\|Z_i\|^4 \to 0.
\]

Then Lemma A.10 of Newey and Windmeijer (2009b) applies (without \( X_i \) term), and we have

\[
\frac{\sum_{i \neq j}(g_i - Eg_i)'\Omega^{-1}(g_j - Eg_j)/n}{\sqrt{2q_n}} \xrightarrow{d} N(0, 1).
\]

Next, for the second term,

\[
\frac{\sum_{i \neq j}(Eg_i)'\Omega^{-1}(g_j - Eg_j)/n}{\sqrt{2q_n}} = \frac{(\sum_{i=1}^n Eg_i)'\Omega^{-1}[\sum_{j=1}^n (g_j - Eg_j)]/n}{\sqrt{2q_n}} - \frac{\sum_{i=1}^n Eg_i'\Omega^{-1}(g_i - Eg_i)/n}{\sqrt{2q_n}}.
\]

Note that by Assumption M.1, and the analysis in (A.12), and \( n^{1/2 - \kappa}(l_n/q_n)^{1/2} = o(1) \),

\[
\frac{\sum_{i=1}^n (Eg_i)'\Omega^{-1}(g_i - Eg_i)/n}{\sqrt{2q_n}} \to 0.
\]

Then by \( g_i \) being iid, and \( Eg_i = C_1/n^\kappa \),

\[
\frac{(\sum_{i=1}^n Eg_i)'\Omega^{-1}[\sum_{j=1}^n (g_j - Eg_j)]/n}{\sqrt{2q_n}} = \frac{(nEg_i)'\Omega^{-1}(\sum_{j=1}^n g_j - Eg_j)/n}{\sqrt{2q_n}} = \frac{(n^{1/2} Eg_i)'\Omega^{-1}(\sum_{j=1}^n g_j - Eg_j)/n^{1/2}}{\sqrt{2q_n}} = \frac{(n^{1/2 - \kappa}C_1)'\Omega^{-1}(\sum_{j=1}^n g_j - Eg_j)/n^{1/2}}{\sqrt{2q_n}} \quad (A.14)
\]

Note that the term on the right side in (A.14) has zero mean. Then

\[
\frac{n^{1-2\kappa}}{2q_n} E[C_1'\Omega^{-1}[\sum_{j=1}^n (g_j - Eg_j)(g_j - Eg_j)'/n]\Omega^{-1}C_1] = \frac{n^{1-2\kappa}}{2q_n} C_1'\Omega^{-1} E\left[\sum_{j=1}^n (g_j - Eg_j)(g_j - Eg_j)'/n\right]\Omega^{-1}C_1
\]

\[
= n^{1-2\kappa} C_1'\Omega^{-1}C_1/2q_n + o(1) \to 0, \quad (A.15)
\]

by Assumption M.1 with \( C_1 = (0'_{q_n-l}, C_{1n}'). \) To get (A.15), we use

\[
n^{1-2\kappa} C_1'\Omega^{-1}C_1/2q_n \leq n^{1-2\kappa} \|C_1\|^2/2q_n = n^{1-2\kappa} O(l_n)/2q_n.
\]

31
Next, by Lemma A.1. So we use the conditions of Lemma A.1 in the statement of Theorem 3(iii). Q.E.D.

Apply Lemmata A.1 and A.3 to get the desired results. Note that the condition \( n \to \infty \) and Windmeijer (2009b), depending on \( n^{1/2-\kappa}l_n^{1/2} \), and \( \tau \).

**Proof of Theorem 3(i).** This proof follows from the assumptions, Theorem 2 by the continuous mapping theorem, and Lemmata SA.8-SA.9. The details can be seen from the proof of Theorem 3, p.26-27 in Newey and Windmeijer (2009b). The limit here may be different compared with Newey and Windmeijer (2009b), depending on \( n^{1/2-\kappa}l_n^{1/2} \), and \( \tau \).

**Proof of Theorem 3(ii).** By Lemma A.4

\[
\frac{2n\hat{Q}(\theta_0) - q_n}{\sqrt{2q_n}} \xrightarrow{d} N(0,1). \tag{A.17}
\]

Next, as \( q_n \to \infty \), the \((1-\alpha)\)th quantile \( s_\alpha^{q_n} \) of \( \chi^2_{q_n} \) distribution has the property that \( (s_\alpha^{q_n} - q_n)/\sqrt{2q_n} \) converges to the \((1-\alpha)\)th quantile \( s_\alpha \) of \( N(0,1) \). So

\[
P(2n\hat{Q}_n(\theta_0) \geq s_\alpha^{q_n}) = P\left(\frac{2n\hat{Q}(\theta_0) - q_n}{\sqrt{2q_n}} \geq \frac{s_\alpha^{q_n} - q_n}{\sqrt{2q_n}}\right) \to \alpha. \tag{A.18}
\]

**Proof of Theorem 3(iii).** This proof follows from Lemmata A.1 and A.3 here. First, see that at \( \hat{\theta} = \theta_0 \), Lemma A.14 of Newey and Windmeijer (2009b) or Lemma A.3 provides

\[
nS_n^{-1} \hat{D}(\theta_0)'\hat{\Omega}(\theta_0)^{-1}\hat{D}(\theta_0)S_n^{-1} \xrightarrow{P} H + \Lambda. \tag{A.19}
\]

Next, by \( K_n(\theta_0) \) definition

\[
K_n(\theta_0) = n\frac{\partial^2 \hat{Q}(\theta_0)'}{\partial \theta^2} S_n^{-1} [nS_n^{-1} \hat{D}(\theta_0)'\hat{\Omega}(\theta_0)^{-1}\hat{D}(\theta_0)S_n^{-1}]^{-1}nS_n^{-1} \frac{\partial \hat{Q}(\theta_0)}{\partial \theta}. \tag{A.20}
\]

Apply Lemmata A.1 and A.3 to get the desired results. Note that the condition \( n^{1/2-\kappa}\sqrt{l_n/q_n} = o(1) \) of Lemma A.3 is weaker than the conditions \( n^{1/2-\kappa}\sqrt{l_n} = o(1) \) or \( n^{1/2-\kappa}\sqrt{l_n} = O(1) \) in Lemma A.1. So we use the conditions of Lemma A.1 in the statement of Theorem 3(iii). Q.E.D.

**Proof of Corollary 1(i).** The stochastic order of the score is

\[
n\mu_n^{-1} \frac{\partial \hat{Q}(0)}{\partial \delta} = O_p(1) + o_p(1) + O_p(n^{1/2-\kappa}\sqrt{l_n}), \tag{A.21}
\]

which are the rates of the terms in the score. The rates of the terms in can be understood in (SA.40), (SA.41), (SA.46), respectively. If \( n^{1/2-\kappa}\sqrt{l_n} \to \infty \), then

\[
n\mu_n^{-1} \frac{\partial \hat{Q}(0)}{\partial \delta} = O_p(n^{1/2-\kappa}\sqrt{l_n}). \tag{A.22}
\]
Next, we want to find out that if we can detect the violation of exogeneity. To see that, we use the assumption M.2(i), we see that \( n \to 0 \) as can be seen from (SA.56). Given \( \frac{n^{-2\kappa}}{\sqrt{q_n}} = o(1) \) in Corollary 1(i), and \( q_n/\mu_n^2 \) being bounded in Assumption M.2(i), we see that \( \frac{n^{1-2\kappa}}{\sqrt{n}} \to 0 \). Hence,
\[
\frac{n\mu_n^{-2} \partial^2 \hat{Q}(\delta)}{\partial \delta^2} = H + O_p(\frac{n^{1-2\kappa}}{\mu_n} \sqrt{\frac{n}{q_n}}),
\]
(A.23)
as can be seen from (SA.56). Given \( \frac{n^{-2\kappa}}{\sqrt{q_n}} = o(1) \) in Corollary 1(i), and \( q_n/\mu_n^2 \) being bounded in Assumption M.2(i), we see that \( \frac{n^{1-2\kappa}}{\sqrt{n}} \to 0 \). Hence,
\[
\frac{n\mu_n^{-2} \partial^2 \hat{Q}(\delta)}{\partial \delta^2} = H + o_p(1).
\]
(A.24)
From (A.22) and (A.24) it is clear that
\[
S_n'(\hat{\theta} - \theta_0) = O_p(n^{1-2\kappa} \sqrt{\frac{n}{q_n}}).
\]
(A.25)
By (A.22)(A.24) and p.30 of Newey and Windmeijer (2009b),
\[
2n[\hat{Q}(\theta_0) - \hat{Q}(\bar{\theta})] = n(\hat{\theta} - \theta_0)'[\partial^2 \hat{Q}(\bar{\theta})/\partial \theta \partial \theta'](\hat{\theta} - \theta_0) = (\hat{\theta} - \theta_0)' S_n(nS_n^{-1} \partial^2 \hat{Q}(\bar{\theta})/\partial \theta \partial \theta' S_n^{-1})S_n'(\hat{\theta} - \theta_0) = O_p(n^{-2\kappa} l_n). \quad (A.26)
\]
Next, we want to find out that if we can detect the violation of exogeneity. To see that, we use the quantiles from the null distribution \( s_{\alpha}^{q_n - p} \), and check to see if the OI test rejects more often than \( \alpha \). In that respect, by (A.26),
\[
\frac{2n\hat{Q}(\hat{\theta}) - (q_n - p)}{\sqrt{2(q_n - p)}} - \frac{2n\hat{Q}(\theta_0) - (q_n - p)}{\sqrt{2(q_n - p)}} = O_p(n^{-2\kappa} l_n) = o_p(1), \quad (A.27)
\]
since \( n^{-2\kappa} l_n/\sqrt{q_n} = o(1) \). From p.29 of Newey and Windmeijer (2009b), we know that
\[
\hat{Q}(\theta_0) = \frac{\bar{g}' \Omega^{-1} \bar{g}}{2} + o_p(\frac{\sqrt{q_n}}{n}).
\]
(A.28)
Since \( n^{-2\kappa} l_n/\sqrt{q_n} = o(1) \), (A.16) holds, using Lemma A.4 via (A.18) and (A.28); therefore,
\[
\frac{2n\hat{Q}(\theta_0) - (q_n - p)}{\sqrt{2(q_n - p)}} = \frac{\sqrt{2q_n}}{\sqrt{2(q_n - p)}} \left[ \frac{2n\hat{Q}(\theta_0) - q_n}{\sqrt{2q_n}} \right] + \frac{p}{\sqrt{2(q_n - p)}} + o_p(1) \overset{d}{\to} N(0,1). \quad (A.29)
\]
As displayed in Muirhead (2005), or by \( \chi^2 \) distribution with increasing degrees of freedom \( [s_{\alpha}^{q_n - p} - (q_n - p)]/\sqrt{2(q_n - p)} \to N(0,1) \). So, by (A.27)(A.29) we obtain
\[
P \left( 2n\hat{Q}(\hat{\theta}) \geq s_{\alpha}^{q_n - p} \right) = P \left( \frac{2n\hat{Q}(\hat{\theta}) - (q_n - p)}{\sqrt{2(q_n - p)}} \geq \frac{s_{\alpha}^{q_n - p} - (q_n - p)}{\sqrt{2(q_n - p)}} \right) \to \alpha. \quad (A.30)
\]

**Proof of Corollary 1(ii).** Note that (A.21)-(A.23) still apply here. Since we assume \( n^{-2\kappa} l_n/\sqrt{q_n} = O(1) \), we have
\[
O_p \left( \frac{n^{1-2\kappa} l_n}{\mu_n^2} \right) = O_p \left( \frac{n^{1-2\kappa} l_n}{q_n} \frac{q_n}{\mu_n^2} \right) = o_p(1),
\]

33
by Assumption M.2(i), where \( q_n/\mu_n^2 \) is bounded. So

\[
O_p(\frac{n^{1/2-\kappa}l_n^{1/2}}{\mu_n}) = o_p(1).
\]

Next, (A.24) is the same as in Corollary 1(i). Hence, we obtain (A.26).

Next, we want to find out if we can detect the violation of exogeneity. We use the quantiles from the null distribution \( s_n^{q_n-p} \) and check to see if the OI test rejects more often than \( \alpha \). In that respect, by (A.26),

\[
\frac{2n\hat{Q}(\hat{\theta}) - (q_n - p)}{\sqrt{2(q_n - p)}} - \frac{2n\hat{Q}(\theta_0) - (q_n - p)}{\sqrt{2(q_n - p)}} = O_p(\frac{n^{1-2\kappa}l_n}{\sqrt{q_n}}) = O_p(1) \equiv \nu_3,
\]

since \( n^{1-2\kappa}l_n/\sqrt{q_n} = O(1) \). So \( \nu_3 \) is stochastically bounded, and this will play an important role. By p.29 of Newey and Windmeijer (2009b), we know

\[
\hat{Q}(\theta_0) = \frac{\hat{g}'\Omega^{-1}\hat{g}}{2} + o_p(\frac{\sqrt{q_n}}{n}).
\]

By (A.16), since \( n^{1-2\kappa}l_n/\sqrt{q_n} = O(1) \), using Lemma A.4 via (A.32)

\[
\frac{2n\hat{Q}(\theta_0) - (q_n - p)}{\sqrt{2(q_n - p)}} = \frac{\sqrt{2q_n}}{\sqrt{2(q_n - p)}} \left[ \frac{2n\hat{Q}(\theta_0) - q_n}{\sqrt{2q_n}} \right] + \frac{p}{\sqrt{2(q_n - p)}} + o_p(1) \overset{d}{\rightarrow} N(0, 1). \tag{A.33}
\]

See that by Muirhead (2005), or by \( \chi^2 \) distribution with increasing degrees of freedom \( [s_n^{q_n-p} - (q_n - p)]/\sqrt{2(q_n - p)} \rightarrow N(0, 1) \). So, by (A.31)(A.30)

\[
P \left( 2n\hat{Q}(\hat{\theta}) \geq s_n^{q_n-p} \right) = P \left( \frac{2n\hat{Q}(\hat{\theta}) - (q_n - p)}{\sqrt{2(q_n - p)}} \geq \frac{s_n^{q_n-p} - (q_n - p)}{\sqrt{2(q_n - p)}} \right)
\]

\[
= P \left( \frac{2n\hat{Q}(\theta_0) - (q_n - p)}{\sqrt{2(q_n - p)}} + \nu_3 \geq \frac{s_n^{q_n-p} - (q_n - p)}{\sqrt{2(q_n - p)}} \right) > \alpha, \tag{A.34}
\]

since \( \nu_3 \) is positive, wpal with \( H \) being positive definite and with (A.26). Even though \( \frac{2n\hat{Q}(\theta_0) - (q_n - p)}{\sqrt{2(q_n - p)}} \)

approaches standard normal limit, \( \nu_3 \) is positive wpal, so test statistics exceed the critical values more than Corollary 1(i). Hence, the probability of rejecting the false null exceeds \( \alpha \).

**Proof of Corollary 1(iii).** From the proof of Corollary 1(ii), we also see that, via (A.31), if \( n^{1-2\kappa}l_n/\sqrt{q_n} \rightarrow \infty \), then \( \nu_3 \rightarrow \infty \). Thus, we have a consistent test. There are two details in the proof that need to be explained. First, given that \( n^{1-2\kappa}l_n/\mu_n^2 = O(1) \), the Hessian limit will be different but still uniformly bounded. Note that the limit of (A.30) will be a normal distribution, with the same variance but with a positive constant drift. This result is obtained from (A.15)(A.16). These two details do not affect the consistency of the test. Q.E.D.

**Supplement Appendix**

The Supplement Appendix provides the proofs that lead to Theorems 1-2 in Section 2. Here we first define some notation, explanations, and Lemmata before we derive the consistency result. As in Newey and Windmeijer (2009b), we use the following reparameterization and simplification:
\[ \delta = \delta(\theta) = S_n'(\theta - \theta_0)/\mu_n. \] We denote the new reparameterized objective function as \( \hat{Q}(\delta) \), where \( \hat{Q}(\delta) \) is simply \( \hat{Q}(\theta_0 + \mu_n S_n^{-1}\delta) \). We let

\[ \hat{Q}^*(\delta) = \hat{g}(\delta) \hat{\Omega}(\delta)^{-1} \hat{g}(\delta)/2, \]

where \( \hat{g}(\delta) = \frac{1}{n} \sum_{i=1}^n g_i(\theta_0 + \mu_n S_n^{-1}\delta) \), \( \hat{\Omega}(\delta) = \frac{1}{n} \sum_{i=1}^n g_i(\theta_0 + \mu_n S_n^{-1}\delta) g_i(\theta_0 + \mu_n S_n^{-1}\delta)' \).

\[ Q(\delta) = \hat{g}(\delta) \hat{\Omega}(\delta)^{-1} \hat{g}(\delta)/2 + q_n/2n, \quad (SA.1) \]

and \( \hat{g}(\delta) = E g_i(\theta_0 + \mu_n S_n^{-1}\delta) \), \( \hat{\Omega}(\delta) = E g_i(\theta_0 + \mu_n S_n^{-1}\delta) g_i(\theta_0 + \mu_n S_n^{-1}\delta)' \). The following results are Lemmata A.2, A.3, and A.4 in Newey and Windmeijer (2009b). These Lemmata do not use Assumption M.1, which is defined before Assumption M.3.

**Lemma SA.1.** If Assumptions M.4(i)-(v) are satisfied, then for any \( C > 0 \),

\[ \sup_{\theta \in \Theta, ||\delta|| \leq C} \mu_n^{-2} n |\hat{Q}^*(\delta) - Q(\delta)| \xrightarrow{p} 0. \]

We use the original parameters from Lemma A.3 of Newey and Windmeijer (2009b).

**Lemma SA.2** Under Assumption M.4, \( \hat{\theta} = \arg \min_{\theta} \hat{Q}(\theta) \), \( \hat{\lambda} = \arg \max_{\lambda \in \Lambda_n(\hat{\theta})} \sum_{i=1}^n \rho(\lambda' g_i(\hat{\theta}))/n \), and \( \hat{\lambda} = \arg \max_{\lambda \in \Lambda_n(\theta_0)} \sum_{i=1}^n \rho(\lambda' g_i(\theta_0))/n \) exist, \( ||\hat{\lambda}|| = O_p(\sqrt{\frac{2n}{n}}) \), \( ||\hat{\lambda}|| = O_p(\sqrt{\frac{2n}{n}}, ||\hat{g}(\hat{\theta})||) = O_p(\sqrt{\frac{2n}{n}}) \), and \( \hat{Q}^*(\hat{\theta}) \leq \hat{Q}^*(\theta_0) + o_p(q_n/n) \), where \( \hat{Q}^*(\hat{\theta}) = \hat{g}(\hat{\theta})' \hat{\Omega}(\hat{\theta})^{-1} \hat{g}(\hat{\theta})/2 \), and \( \hat{g}(\theta), \hat{\Omega}(\theta) \) are defined before Assumption M.3.

**Lemma SA.3** If Assumptions M.2(i), M.3, and M.4 are satisfied, then for \( \hat{\delta} = S_n'(\hat{\theta} - \theta_0)/\mu_n \),

\[ ||\hat{\delta}|| = O_p(1). \]

**Proof of Theorem 1.** We begin with an important result that is due to Newey and Windmeijer (2009b).

\[ \hat{Q}^*(\hat{\theta}) = \hat{Q}(\hat{\theta}) + o_p(q_n/n) \leq \hat{Q}(\theta_0) + o_p(q_n/n) = \hat{Q}^*(\theta_0) + o_p(q_n/n). \quad (SA.2) \]

This is the equation immediately above Lemma A.4 in Newey and Windmeijer (2009b). This result is independent from Assumption M.1 which here only affects the identifiability condition, not the uniform convergence condition in the consistency proof in Newey and Windmeijer (2009b). In (SA.2), see that by setting \( \delta = S_n'(\theta - \theta_0)/\mu_n, \delta_0 = 0 \), and \( q_n/\mu_n^2 \leq C \),

\[ \mu_n^{-2} n \hat{Q}^*(\hat{\delta}) \leq \mu_n^{-2} n \hat{Q}^*(0) + o_p(1). \quad (SA.3) \]

By Lemma SA.3, there is a \( C \), such that

\[ P\{ ||\hat{\delta}|| \leq C \} \geq 1 - \epsilon/3, \quad (SA.4) \]

for any \( \epsilon > 0 \). Then by Lemma SA.1, for any \( \gamma > 0 \),

\[ P\{ \sup_{||\delta|| \leq C} \mu_n^{-2} n |\hat{Q}^*(\delta) - Q(\delta)| < \gamma/3 \} \geq 1 - \epsilon/3. \quad (SA.5) \]
By (SA.3),

$$P\{\mu_n^{-2} n \hat{Q}^*(\delta) \leq \mu_n^{-2} n \hat{Q}^*(0) + \gamma/3\} \geq 1 - \epsilon/3.$$  \hspace{1cm} (SA.6)

Next, as in p.8 of Newey and Windmeijer (2009b), we denote each event in (SA.4)(SA.5)(SA.6) as $E_1, E_2,$ and $E_3$, respectively, where the probability of intersection of these events is larger than equal to $1 - \epsilon$, $\epsilon > 0$. On the intersection of these events,

$$\mu_n^{-2} n \hat{Q}(\delta) \leq \mu_n^{-2} n \hat{Q}^*(\delta) + \gamma/3 \leq \mu_n^{-2} n \hat{Q}^*(0) + 2\gamma/3 \leq \mu_n^{-2} n Q(0) + \gamma.$$  \hspace{1cm} (SA.7)

In our case, the analysis of $Q(0)$ differs from the one in Newey and Windmeijer (2009b). Using (SA.1) at $\delta = 0$ we have $\theta = \theta_0$, with Assumption M.1,

$$\frac{n}{\mu_n^2} \hat{Q}(\theta_0) = \frac{n}{\mu_n^2} Eg_i(\theta_0) (\Omega^{-1}) Eg_i(\theta_0) + \frac{q_n}{2\mu_n^2} = \frac{\left(n^{1/2}C_1\right)^{\Omega^{-1}}(\Omega^{-1}) \left(n^{1/2}C_1\right)}{2\mu_n^2} + \frac{q_n}{2\mu_n^2}. \hspace{1cm} (SA.8)$$

Note that with $C_1 = (0'_{q_n-l_n}, C'_l, \gamma)$ and Assumption M.4, we can see that

$$C'_l \Omega^{-1} C_1 \leq Eigmax(\Omega^{-1})\|C_1\|^2 = O(l_n),$$

$C_1$ is a $q_n \times 1$ vector, and where $0_{q_n-l_n}$ represents the $(q_n-l_n) \times 1$ vector of zeroes. In (SA.8), the first term on the right side can be expressed as

$$\frac{\left(n^{1/2}C_1\right)^{\Omega^{-1}}(\Omega^{-1}) \left(n^{1/2}C_1\right)}{2\mu_n^2} = \frac{n}{n^{2\kappa}} O\left(\frac{l_n}{q_n \mu_n^2}\right). \hspace{1cm} (SA.9)$$

In (SA.9) we first analyze the term on the right side under Assumption M.1(ii). There, with $1/2 < \kappa < \infty$, and since $q_n/\mu_n^2$ is bounded by Assumption M.2(i), we see that $n^{1-2\kappa} \to 0$. If $l_n/q_n \to f$, $0 < f \leq 1$, then

$$\frac{\left(n^{1/2}C_1\right)^{\Omega^{-1}}(\Omega^{-1}) \left(n^{1/2}C_1\right)}{2\mu_n^2} = \frac{n}{n^{2\kappa}} O\left(\frac{l_n}{q_n \mu_n^2}\right) = n^{1-2\kappa} O(1) = o(1). \hspace{1cm} (SA.10)$$

Under Assumption M.1(iii), we can get a better result for the term on the right side in (SA.9) regarding $\kappa$. Since $l_n/q_n \to 0$, and $q_n/\mu_n^2$ is bounded by Assumption M.2(i), with $n^{1-2\kappa}l_n/q_n \to 0$, we have

$$\frac{\left(n^{1/2}C_1\right)^{\Omega^{-1}}(\Omega^{-1}) \left(n^{1/2}C_1\right)}{2\mu_n^2} = \frac{n}{n^{2\kappa}} O\left(\frac{l_n}{q_n \mu_n^2}\right) = n^{1-2\kappa} \frac{l_n}{q_n} O(1) = o(1). \hspace{1cm} (SA.11)$$

In Newey and Windmeijer (2009b), with perfect exogeneity, the asymptotically negligible term in (SA.10)(SA.11) is exactly zero. Next, we use (SA.7),(SA.10), (SA.11), and (SA.1) to show

$$\frac{n}{\mu_n^2} \hat{g}(\delta)'(\Omega^{-1} \hat{g}(\delta)) + \frac{q_n}{2\mu_n^2} \leq o(1) + \frac{q_n}{2\mu_n^2} + \gamma. \hspace{1cm} (SA.12)$$

We subtract $q_n/2\mu_n^2$ from both sides, the result is

$$\mu_n^{-2} n \hat{g}(\delta)'(\Omega^{-1} \hat{g}(\delta)) \overset{P}{\to} 0.$$  \hspace{1cm} (SA.13)

Next, from Assumption M.3(i) and Assumption M.4(ii), the result is

$$\mu_n^{-2} n \hat{g}(\delta)'(\Omega^{-1} \hat{g}(\delta)) \geq C \mu_n^{-2} n \hat{g}(\delta)'(\Omega^{-1} \hat{g}(\delta)) \geq C \|\delta\|^2.$$  \hspace{1cm} (SA.14)
Therefore, from (SA.13) and (SA.14),

$$
\|\hat{\delta}\|^2 \overset{p}{\rightarrow} 0.
$$

**Q.E.D.**

Even though the notation is described before Theorem 1, we think it is necessary here to repeat it in more detail. We let $e_j$ denote the $j$th unit vector, $j = 1, \cdots, p$ and $u_i = g_i(\theta_0)$, $(q_n \times 1), G_i(\theta_0) = \frac{\partial u_i(\theta_0)}{\partial \theta}$. $B^j = \Omega^{-1}E[\partial_{\theta} G_i^j]$ $(q_n \times q_n), \ B^j_i = G_i e_j - Ge_j - B^j g_i$. (SA.15)

$$
U_i = [U_1, \cdots, U_p].
$$

(Note that $U_i$ is a $q_n \times p$ matrix. Basically, $U_i^j$ represents the residual from projecting the derivatives on the moment functions. Note also that $\hat{\theta} = \frac{1}{n} \sum_{i=1}^n g_i$. Next,

$$
E[\|\hat{\theta}\|^2] = tr(\Omega)/n \leq C q_n/n,
$$

from Eigmax $(\Omega) \leq C$, so $\|\hat{\theta}\| = O_p(\sqrt{q_n/n})$ by Markov’s inequality.

We use Lemma A.10 from Newey and Windmeijer (2009b). Let $X_i$ denote a scalar random variable, and dependence on $n$ is suppressed. Let $Z_i, Y_i$ be general $q_n \times 1$ random vectors. Set $\Psi = \Sigma_{ZZ} + \Sigma_{ZY}$, where $\Sigma_{ZZ} = \lim_{n \rightarrow \infty} EZ_i Z_i', \Sigma_{ZY} = \lim_{n \rightarrow \infty} EZ_i Y_i'$, and $\Sigma_{YY} = \lim_{n \rightarrow \infty} EY_i Y_i'$.

$Eigmax(\Sigma_{ZZ}) = Eigz$, and $Eigmax(\Sigma_{YY}) = Eigy$.

**Lemma SA.4** If $(X_i, Y_i, Z_i) (i = 1, \cdots, n)$ are iid, $EX_i = EZ_i = EY_i = 0$, and $\Sigma_{ZZ}, \Sigma_{YY}$ exist, $n E \|Z_i\|^2 \rightarrow A, n^2 tr \Psi \rightarrow \Lambda, n E \|X_i\|^4 \rightarrow 0, q_n n^2 Eigz^2 Eigy^2 \rightarrow 0, n^3 (Eigz^2 E\|Z_i\|^4 + Eigy^2 E\|Y_i\|^4) \rightarrow 0$, and $n^2 E[\|Y_i\|^4] E[\|Z_i\|^4] \rightarrow 0$, then

$$
\sum_{i=1}^n X_i + \sum_{j \neq i} Z_i Y_j \overset{d}{\rightarrow} N(0, A + \Lambda).
$$

Note that this Lemma is slightly different than the result in Newey and Windmeijer (2009b). Namely, they have $\Sigma_{ZZ} = EZ_i Z_i'$, and similar notation holds for the other moment matrices. Our definition does not change the proof of Lemma A.10 in Newey and Windmeijer (2009b).

We then provide Lemma A.11 in Newey and Windmeijer (2009b), which is useful in deriving a convenient expression for the score of the objective function. The following notation is used in Newey and Windmeijer (2009a), and is helpful in deriving the proofs. So we reparameterize $\delta = S_n'(\theta - \theta_0)/\mu_n$ and $\theta = \theta_0 + \mu_n S_n^{-1} \delta$, we see that at $\theta_0, \delta = 0$. We have the following notation: $g_i \delta_k = \frac{\partial u_i(\theta_0)}{\partial \theta_k} = G_i S_n^{-1} \partial_k \mu_n$, $\hat{\Omega} = \hat{\Omega}(\theta_0), \quad \hat{\Omega}^k = \sum_{i=1}^n g_i \hat{\theta}_k / n, \quad \Omega^k = E\hat{\Theta}^k, \quad \tilde{B}^k = \tilde{\Theta}^{-1} \tilde{\Omega}^k, \quad B^k = \Theta^{-1} \Omega^k$.

**Lemma SA.5.** If Assumptions M.2-M.6 are satisfied, then

$$
\sqrt{n}||\hat{\Omega} - \Omega|| \overset{p}{\rightarrow} 0,
$$

$$
\sqrt{n}||\hat{\Omega}^k - \Omega^k|| \overset{p}{\rightarrow} 0,
$$

$$
\sqrt{n}||\tilde{B}^k - B^k|| \overset{p}{\rightarrow} 0.
$$

Note that there is a small typographical error in Newey and Windmeijer (2009b). The rate on $\hat{\Omega}_k$ is $\mu_n \sqrt{n}$, but in the proof, it is clear that the true rate is $\sqrt{n}$. Assumption M.1 is not used in the
derivation of the expansion of the score in the equation below. Following the same steps on p.16-18 in \cite{Newey2009b}, we obtain the equation

\[ n\mu_n^{-1} \frac{\partial \hat{Q}(0)}{\partial \delta} = nS_n^{-1}G'\Omega^{-1}\hat{g} + n^{-1} \sum_{i=1}^n U_i'\Omega^{-1}\hat{g} + o_p(1), \]  

(17)

which is (A.7) in \cite{Newey2009b}. In (SA.17), the score is evaluated at \( \delta = 0 \). Note that by (SA.15)(SA.16) \( EU_i^n = B^jE_{gi} \) for \( j = 1, \ldots, p \). So, from Assumption M.1, we obtain

\[ n^{1/2}EU_i^n = n^{1/2}\kappa B_{C_1}, \ldots, B_{C_1}. \]  

(18)

We define some notation before the following Lemma. First, we define \( \tau \) in the following way. Let \( nS_n^{-1}G'(n^{1/2}\kappa\Omega^{-1}C_1) \to \tau \), and \( \tau \) can be a zero or nonzero constant vector of dimension \( p \). There are three possible cases in Lemma SA.6, and they are related to \( \tau \), and the behavior of \( n^{1/2}\kappa l_n^{1/2} \). We explain each case in detail in the Remarks after Lemma SA.6. The result shows that violation of exogeneity can alter the standard limit for the score in \cite{Newey2009b}. The new limit is normally distributed with a large variance compared with standard limit due to exogeneity violation. Therefore, violation of exogeneity may add up more than a simple constant drift term.

**Lemma SA.6.** Under Assumptions M.1- M.6 and \( S_n^{-1}E[(U_i - EU_i)^r\Omega^{-1}(U_i - EU_i)]S_n^{-1} \to \Lambda \), we have three possibilities:

(i). If \( n^{1/2}\kappa l_n^{1/2} = O(1) \) and \( n^{1/2}S_n^{-1}G' = O(1) \), then

\[ n\mu_n^{-1} \frac{\partial \hat{Q}_n(0)}{\partial \delta} \xrightarrow{d} \nu_2, \]

where \( \nu_2 \equiv N(\tau, H + \Lambda + \Delta) \), where

\[ \Delta = \lim_{n \to \infty} S_n^{-1}E[(U_i - EU_i)^r\Omega^{-1/2}(n^{1-2\kappa}\Omega^{-1/2}C_1^r\Omega^{-1/2})\Omega^{-1/2}(U_i - EU_i)]S_n^{-1} \cdot \]

(ii). If \( n^{1/2}\kappa l_n^{1/2} = O(1) \) and \( n^{1/2}S_n^{-1}G' = o(1) \), then

\[ n\mu_n^{-1} \frac{\partial \hat{Q}_n(0)}{\partial \delta} \xrightarrow{d} \nu_{2nd}, \]

where \( \nu_{2nd} \equiv N(0, H + \Lambda + \Delta) \).

(iii). If \( n^{1/2}\kappa l_n^{1/2} = o(1) \), then

\[ n\mu_n^{-1} \frac{\partial \hat{Q}_n(0)}{\partial \delta} \xrightarrow{d} \nu_1, \]

where \( \nu_1 \equiv N(0, H + \Lambda) \).

**Remarks.**

1. Note that the standard result is Case (iii). This is the limit of the score in Lemma A.12 in \cite{Newey2009b}. So their lemma extends to minor violations with many imperfect instruments. It is critical to see even \( l_n = q_n \) is possible in that case, but to satisfy the condition of \( n^{1/2}\kappa l_n = o(1) \), \( \kappa \) (violation should be really minor) should be large. Case (iii) restricts \( 1/2 < \kappa < \infty \).

2. The nonstandard Cases are (i) and (ii). Compared with Case (iii), we allow for \( \kappa = 1/2 \) in Cases (i) and (ii). Note that by using the definition of \( C_1 \), we can get \( \|C_1\| = O(l_n^{1/2}) \). Then we
have \( n^{1/2-\kappa}\|\Omega^{-1/2}C_1\| = O(n^{1/2-\kappa}l_n^{1/2}) = O(1) \). In Case (i), we also assume \( \|n^{1/2}S_n^{-1}G'\| = O(1) \). Case (ii) is the same as Case (i), but with weak identification; therefore, the drift is zero. To see this point about relation of \( \tau \) to identification issues, see Examples 1-2.

3. We also see that the key for deriving limits is the behavior of \( n^{1/2-\kappa}l_n^{1/2} \). Depending on whether it goes to zero or not, we get the standard limit as in Case (iii), or non standard limits as in Cases (i) and (ii).

4. Note that the results in Lemma SA.6 hold regardless of \( l_n/q_n \to f \), \( 0 < f \leq 1 \) or \( f = 0 \) (i.e. Assumptions M.1(ii) or M.1(iii), respectively).

5. There are also other possibilities apart from the results above. We can have \( 1/2 < \kappa < \infty \) and \( n^{1/2-\kappa}l_n^{1/2} \to \infty \). In the case of strong/nearly-weak identification, we see that \( \tau \to \infty \). But in the case of weak identification, \( \tau \) can be zero, constant, or infinity. Thus, there is uncertainty in that case. Also, in the case of \( 0 < \kappa \leq 1/2 \) with Assumption M.1(iii), we will have \( n^{1/2-\kappa}l_n^{1/2} \to \infty \), and the resulting drift may diverge to infinity when there is strong/near weak identification. This divergence is clear from the proof of Lemma SA.6. We do not analyze the case of \( 0 < \kappa \leq 1/2 \) with \( l_n/q_n \to f \), \( 0 < f \leq 1 \) (Assumption M.1(ii)) since the estimator is not consistent via proof of Theorem 1.

Proof of Lemma SA.6.

The limit of the score depends crucially on whether \( n^{1/2-\kappa}l_n^{1/2} = O(1) \) or \( n^{1/2-\kappa}l_n^{1/2} = o(1) \). Note that by the same analysis in (SA.21) if \( n^{1/2-\kappa}l_n^{1/2} = O(1) \), then we have

\[
\|n^{1/2-\kappa}\Omega^{-1/2}C_1\| = O(1),
\]

(SA.19)

whereas if \( n^{1/2-\kappa}l_n^{1/2} = o(1) \), then we have

\[
\|n^{1/2-\kappa}\Omega^{-1/2}C_1\| = o(1).
\]

(SA.20)

In part (i) of the proof as well as (ii), we assume \( n^{1/2-\kappa}l_n^{1/2} = O(1) \). We need the following result for subsequent proofs:

\[
n^{1/2-\kappa}\Omega^{-1}C_1 \leq n^{1/2-\kappa}\|\Omega^{-1}C_1\| \leq Cn^{1/2-\kappa}\|C_1\| = O(n^{1/2-\kappa}l_n^{1/2}),
\]

(SA.21)

where we use \( C_1 = (0'_{q_n-l_n}, C'_n)' \), \( \|C_1\| = O(l_n^{1/2}) \).

(i). Here we assume \( n^{1/2-\kappa}l_n^{1/2} = O(1) \).

First, we take a \( p \times 1 \) vector \( \zeta \) with \( \|\zeta\| = 1 \). Next, we denote \( \bar{U} = n^{-1}\sum_{i=1}^n U_i \) and use this equation in (SA.17),

\[
n\mu_n^{-1}\zeta'\frac{d\hat{Q}(0)}{d\delta} = n\zeta'S_n^{-1}[G'\Omega^{-1}(\bar{g}_d + E_g) + \bar{U}'\Omega^{-1}(\bar{g}_d + E_g)] + o_p(1),
\]

(SA.22)

where \( \bar{g}_d = \bar{g} - E_g \). We rewrite (SA.22) as

\[
n\mu_n^{-1}\zeta'\frac{d\hat{Q}(0)}{d\delta} = n\zeta'S_n^{-1}[G'\Omega^{-1}(\bar{g}_d + E_g) + \bar{U}'\Omega^{-1}(\bar{g}_d + E_g)] + o_p(1),
\]

(SA.23)

\[
= n\zeta'S_n^{-1}[G'\Omega^{-1}\bar{g}_d + (\bar{U} - E\bar{U})'\Omega^{-1}\bar{g}_d]
\]

(SA.24)

\[
+ n\zeta'S_n^{-1}E\bar{U}'\Omega^{-1}\bar{g}_d 
\]

(SA.25)

\[
+ n\zeta'S_n^{-1}[G + \bar{U}]'\Omega^{-1}E_g + o_p(1).
\]
The terms (SA.23)-(SA.25) can be rewritten as

\[ n\mu_n^{-1}\zeta'\frac{\partial \hat{Q}(0)}{\partial \delta} = n\zeta'S_n^{-1}[G'\Omega^{-1}(\hat{g}_d + Egi) + U'\Omega^{-1}(\hat{g}_d + Egi)] + o_p(1), \]

\[ = n\zeta'S_n^{-1}[G'\Omega^{-1}\hat{g}_d + (U - E\bar{U})'\Omega^{-1}\hat{g}_d] \quad \text{(SA.26)} \]

\[ + n\zeta'S_n^{-1}E\bar{U}'\Omega^{-1}\hat{g}_d \quad \text{(SA.27)} \]

\[ + n^{1/2}\zeta'S_n^{-1}G'(n^{1/2}\kappa\Omega^{-1}C_1) \quad \text{(SA.28)} \]

\[ + n^{1/2}\zeta'S_n^{-1}(U - E\bar{U})'\Omega^{-1/2}(n^{1/2}\kappa\Omega^{-1/2}C_1) \quad \text{(SA.29)} \]

\[ + n^{1/2}\zeta'S_n^{-1}E\bar{U}'\Omega^{-1/2}(n^{1/2}\kappa\Omega^{-1/2}C_1) + o_p(1) \quad \text{(SA.30)} \]

We use the following to get (SA.28)-(SA.30), and \( Egi = C_1/n^\kappa \):

\[ n\zeta'S_n^{-1}[G + U]'\Omega^{-1}Egi = n^{1/2}\zeta'S_n^{-1}G'(n^{1/2}\kappa\Omega^{-1}C_1) \]

\[ + n^{1/2}\zeta'S_n^{-1}(U - E\bar{U})'\Omega^{-1/2}(n^{1/2}\kappa\Omega^{-1/2}C_1) \]

\[ + n^{1/2}\zeta'S_n^{-1}E\bar{U}'\Omega^{-1/2}(n^{1/2}\kappa\Omega^{-1/2}C_1). \]

We consider each of the terms (SA.26)-(SA.30). We apply Lemma SA.4 to the sum of terms (SA.26) and (SA.29). Then, we set

\[ X_i = \zeta'S_n^{-1}G'(g_i - Egi) + \zeta'S_n^{-1}\frac{(U_i - EU_i)'}{n^{1/2}}\Omega^{-1/2}(n^{1/2}\kappa\Omega^{-1/2}C_1), \quad \text{(SA.31)} \]

\[ Y_i = \Omega^{-1}(g_i - Egi), \quad \text{(SA.32)} \]

\[ Z_i = (U_i - EU_i)S_n^{-1}'/n. \quad \text{(SA.33)} \]

We then rewrite (SA.23) as

\[ \sum_{i=1}^{n} X_i + \sum_{i,j=1}^{n} Y_i'Z_i. \quad \text{(SA.34)} \]

Note that \( EY_i = EZ_i = 0 \). By definition of \( U_i, g_i \) in (SA.15) we derive \( EZ_i'Y_i \to 0 \). Then, since \( \|S_n^{-1}\| \leq C/\mu_n \),

\[ nE|Y_i'Z_i|^2 \leq CE\|(g_i - Egi)'\Omega^{-1}(U_i - EU_i)\|^2/n\mu_n^2 \]

\[ \leq CE\|g_i'\Omega^{-1}U_i\|^2/n\mu_n^2 + CE\|(Egi)'\Omega^{-1}U_i\|^2/n\mu_n^2 \]

\[ + CE\|g_i'\Omega^{-1}EU_i\|^2/n\mu_n^2 + CE\|(Egi)'\Omega^{-1}(EU_i)\|^2/n\mu_n^2. \quad \text{(SA.35)} \]

We analyze the first term on the right side of (SA.35), as in p.19 of Newey and Windmeijer (2009b)

\[ E\|g_i'\Omega^{-1}U_i\|^2/n\mu_n^2 \leq M[E\|g_i\|^4 + E\|G_i\|^4]/n\mu_n^2 \to 0, \]

by Assumption M.5(i), and \( M \) is a constant. The second, third, and fourth terms on the right side of (SA.35), are smaller than, or of the same stochastic order as the first term from Assumption M.1. So all of these converge in probability to zero as well. Then, by Markov’s inequality,

\[ \sum_{i=1}^{n} Z_i'Y_i \overset{p}{\to} 0. \quad \text{(SA.36)} \]

40
Therefore, from the last result, (SA.31)-(SA.33) and sum of (SA.26) with (SA.29) can be written as

\[
\begin{align*}
&n\zeta' S_n^{-1}[G'\Omega^{-1} (\bar{g} - E\bar{g})] + n^{1/2}\zeta' S_n^{-1}(\bar{U} - E\bar{U})'\Omega^{-1/2}(n^{1/2-\kappa}\Omega^{-1/2}C_1) \\
&+ \zeta' S_n^{-1}[n^{-1}\sum_{i=1}^{n}(U_i - EU_i)'\Omega^{-1}(\bar{g} - E\bar{g})] \\
&= \sum_{i=1}^{n} X_i + \sum_{i \neq j} Z_i Y_j + o_p(1).
\end{align*}
\]

We use Lemma SA.4, and we are, therefore, able to verify its conditions. First,

\[
\lim_{n \to \infty} \Omega^{-1}E(g_i - Eg_i)(g_i - Eg_i)'\Omega^{-1} = \Omega^{-1} = \Sigma_{YY},
\]

by Assumption M.1. We already show that \(\Sigma_{ZY} = 0\). We can then see that \(\Sigma_{Zn} = \lim_{n \to \infty} EZ_i Z_i' = \lim_{n \to \infty}[n^{-2}E(U_i - EU_i)S_n^{-1}\zeta' S_n^{-1}(U_i - EU_i)']\), which yields

\[
\Psi = \Sigma_{Zn}\Sigma_{YY} = \lim_{n \to \infty} [n^{-2}E[(U_i - EU_i)S_n^{-1}\zeta' S_n^{-1}(U_i - EU_i)']\Omega^{-1}].
\]

To analyze the limit behavior of the score, decompose

\[
X_i = X_{i1} + X_{i2},
\]

where

\[
X_{i1} = \zeta' S_n^{-1}G'\Omega^{-1}(g_i - Eg_i),
\]

and

\[
X_{i2} = \zeta' S_n^{-1}(U_i - EU_i)'\Omega^{-1/2}(n^{1/2-\kappa}\Omega^{-1/2}C_1).
\]

Note that \(X_{i1}, X_{i2}\) are asymptotically uncorrelated due to \(U_i, g_i\) definitions. We can see that \(E(U_i - EU_i)'(g_i - Eg_i) \to 0\). So we have

\[
nEX_i^2 = nEX_{i1}^2 + nEX_{i2}^2.
\]

Next, we consider from Assumption M.1 that

\[
nEX_{i1}^2 = n\zeta' S_n^{-1}G'\Omega^{-1}[E(g_i - Eg_i)(g_i - Eg_i)']\Omega^{-1}GS_n^{-1}\zeta = n\zeta' S_n^{-1}G'\Omega^{-1}GS_n^{-1}\zeta + o(1).
\]

From Assumption M.2, \(nEX_{i1}^2 \to \zeta' H\zeta\).

Next,

\[
nEX_{i2}^2 = \zeta' S_n^{-1}E[(U_i - EU_i)'\Omega^{-1/2}(n^{1-2\kappa}\Omega^{-1/2}C_1 C_1'\Omega^{-1/2})\Omega^{-1/2}(U_i - EU_i)]S_n^{-1}\zeta \to \zeta' \Delta\zeta,
\]

which also defines \(\Delta\). The same expression can also be rewritten as

\[
nEX_{i2}^2 = tr\{(n^{1-2\kappa}\Omega^{-1/2}C_1 C_1'\Omega^{-1/2})E[\Omega^{-1/2}(U_i - EU_i)]S_n^{-1}\zeta' S_n^{-1}(U_i - EU_i)'\Omega^{-1/2}]\}.
\]
Also, since $n^{1-2\kappa_1}\Omega^{-1/2}C_1\Omega^{-1/2}$ is symmetric and $E[\Omega^{-1/2}(U_i - EU_i)S_n^{-1'}\zeta' S_n^{-1}(U_i - EU_i)'\Omega^{-1/2}]$ is positive semidefinite, via (A.13) of Hansen (2012) these equations can be rewritten as

$$
tr\{n^{1-2\kappa_1}\Omega^{-1/2}C_1\Omega^{-1/2}\} \times E[\Omega^{-1/2}(U_i - EU_i)S_n^{-1'}\zeta' S_n^{-1}(U_i - EU_i)'\Omega^{-1/2}] \\
\leq Eigmax(n^{1-2\kappa_1}\Omega^{-1/2}C_1\Omega^{-1/2}) \\
\times tr\{E[\Omega^{-1/2}(U_i - EU_i)S_n^{-1'}\zeta' S_n^{-1}(U_i - EU_i)'\Omega^{-1/2}]\} \\
= Eigmax(n^{1-2\kappa_1}\Omega^{-1/2}C_1\Omega^{-1/2}) \\
\times \zeta' S_n^{-1}E[(U_i - EU_i)'\Omega^{-1}(U_i - EU_i)]S_n^{-1'}\zeta \\
\leq tr\{n^{1-2\kappa_1}\Omega^{-1/2}C_1\Omega^{-1/2}\} \\
\times \zeta' S_n^{-1}E[(U_i - EU_i)'\Omega^{-1}(U_i - EU_i)]S_n^{-1'}\zeta \\
= O(1), \quad (SA.37)
$$

with (SA.38)(SA.19), $n^{1/2-\kappa_1/2} = O(1)$, and $C_1 = (0_{q_n - l_n}, C_{l_n}')$.

So combining the results

$$
nEX^2_i = nEX^2_{i1} + nEX^2_{i2} \rightarrow \zeta'(H + \Delta)\zeta,
$$

where $\Delta$ is due to exogeneity violation.

By the statement in Lemma SA.6,

$$
n^2tr\Psi = \zeta' S_n^{-1}E[(U_i - EU_i)'\Omega^{-1}(U_i - EU_i)]S_n^{-1'}\zeta \rightarrow \zeta' \Lambda \zeta. \quad (SA.38)
$$

Next, we consider the following equation which benefited from Minkowski’s inequality:

$$
nEX^4_i = nE||\zeta' S_n^{-1}G'\Omega^{-1}(g_i - EG_i) + \zeta' S_n^{-1}(U_i - EU_i)'\Omega^{-1/2}(n^{1/2-\kappa_1/2}C_1)||^4 \\
\leq nE||\zeta' S_n^{-1}G'\Omega^{-1}g_i||^4 + nE||\zeta' S_n^{-1}G'\Omega^{-1}Eg_i||^4 \\
+ nE||\zeta' S_n^{-1}(U_i)'\Omega^{-1/2}(n^{1/2-\kappa_1/2}C_1)||^4 \\
+ nE||\zeta' S_n^{-1}(EU_i)'\Omega^{-1/2}(n^{1/2-\kappa_1/2}C_1)||^4. \quad (SA.39)
$$

So we can analyze the four terms on the right side of (SA.39). First,

$$
nE||\zeta' S_n^{-1}G'\Omega^{-1}g_i||^4 = nE||\zeta' \sqrt{n}S_n^{-1}G'\Omega^{-1}g_i||^4/n^2 \\
\leq CE||g_i||^4/n \rightarrow 0,
$$

since $||\sqrt{n}S_n^{-1}G'\Omega^{-1}|| \leq C$ from Assumptions M.2, M.4, and M.5(i). In the same way, from Assumption M.1, the second term on the right side of (SA.39) converges to zero. Note that also the third and fourth terms on the right side of (SA.39) converge to zero by Assumption M.5(i) and p.19 of Newey and Windmeijer (2009b). So, $nEX^4_i \rightarrow 0$.

Then, as in p.19 of Newey and Windmeijer (2009b), $Eigmax(S_n^{-1'}\zeta' S_n^{-1}) \leq C/\mu_n^2$, and $E\tilde{ig}z \leq C/\mu_n^2n^2$, $E\tilde{ig}y \leq C$

$$
q_n n^4 E\tilde{ig}y^2 \tilde{ig}y^2 \leq Cq_n n^4/\mu_n^2 n^2 \leq Cq_n/\mu_n^4 \rightarrow 0, \\
n^3(E\tilde{ig}z^2 EY_i||^4 + E\tilde{ig}y^2 E||Z_i||^4) \leq n^3C(E||g_i||^4 + E||G_i||^4)/\mu_n^4 n^4 \rightarrow 0,
$$

42
\[ n^2 E\|Y_i\|^4 E\|Z_i\|^4 \to 0, \]

where the last equation is obtained by \( Y_i, Z_i \) definition, and by Assumption M.5(i). So, all the conditions of Lemma SA.4 are satisfied. Then the sum of terms (SA.26)(SA.29) converges to

\[
nS_n^{-1}[G'\Omega^{-1}(\bar{g} - E\bar{g})] + nS_n^{-1}[n^{-1}\sum_{i=1}^{n}(U_i - EU_i)'\Omega^{-1}(\bar{g} - E\bar{g})] + n^{1/2}\zeta_s^{-1}(\bar{U} - EU\bar{U})'\Omega^{-1/2}(n^{1/2-\kappa}\Omega^{-1/2}C_1) \xrightarrow{d} N(0, H + \Delta + \Lambda), \tag{SA.40} \]

when we use the Cramer-Wold device, and we see that \( E\bar{g} = Eg_i \), by \( g_i \) being iid.

Note that the sum of (SA.26) with (SA.29) is stochastically bounded as shown in (SA.40). The term in (SA.27) goes through a similar analysis but is of a smaller stochastic order. This stochastic order can be observed since \( E\bar{U} \) involves \( Eg_i = C_1/n^\kappa \). It can also be seen that

\[
n\zeta_s^{-1}E\bar{U}'\Omega^{-1}\bar{g}_d = o_p(1), \tag{SA.41} \]

via the same type of analysis as in (SA.31)-(SA.39).

Next, we consider (SA.28), in which

\[
n^{1/2}S_n^{-1}G'(n^{1/2-\kappa}\Omega^{-1}C_1) \to \tau, \tag{SA.42} \]

where \( \tau \) can be a \( p \times 1 \) vector of zeros or \( \tau \neq 0 \). See Examples 1 and 2 for these possibilities. In this drift term, there are two factors that makes a difference. The first one is related to the strength and weakness of instruments. In a structure of weak instruments by Stock and Wright (2000),

\[
\|\sqrt{n}S_n^{-1}G'\| = o(1). \tag{SA.43} \]

This equation can be seen in Example 2 in section 2.3.

Conversely, the other possibility may be deemed as nearly-weak/strong instruments (Hahn and Kuersteiner (2002), Caner (2010), Bertille and Renault (2009)), and

\[
\|n^{1/2}S_n^{-1}G'\| = O(1). \tag{SA.44} \]

This can be seen in Example 1 in section 2.3. The second factor analysis can be seen from (SA.21).

The key result is on p.22 of Newey and Windmeijer (2009b) (line 2 from the bottom there), which states that

\[
n^{1/2}\|GS_n^{-1}'\| \leq C, \tag{SA.45} \]

given Assumption 1 and 3 in Newey and Windmeijer (2009a). These are Assumptions M.2(i)(ii), and M.4(i)-(iv) here. Note that (SA.45) is true regardless of weak or nearly-weak identification.

It is also possible to relax these conditions under two specific cases where we can get zero drift. If we have weak identification (i.e. (SA.43)) with \( 1/2 < \kappa < \infty \), then even with a more general \( n^{1/2-\kappa}l_n^{1/2} = O(1) \), we can get a zero drift. So this allows larger number of \( l_n \) with respect to first general scenario but only under weak identification. The second possibility is where \( \kappa = 1/2, l_n = l \), where \( l \) is fixed and uniformly bounded. There we see that \( n^{1/2-\kappa}l_n^{1/2} = l^{1/2} = O(1) \), and under weak identification (SA.43), we have a zero drift. In the second case, we allow for a mild exogeneity
violation with $\kappa = 1/2$, we still get a zero drift under only weak identification and a fixed number of imperfect instruments.

Now we provide the conditions for a nonzero drift. Consider (SA.42), and only under nearly-weak identification (SA.44), with $1/2 < \kappa < \infty$ and $n^{1/2-\kappa}l_n^{1/2} = O(1)$, we have

$$nS_n^{-1}[G'\Omega^{-1}E\tilde{y}] = \sqrt{n}S_n^{-1}G'\Omega^{-1}n^{1/2-\kappa}C_1 \to \tau \neq 0.$$  \hspace{1cm} (SA.46)

Thus, nonzero constant drift is possible under nearly-weak identification. We can generalize this slightly to $\kappa = 1/2$ when we have $l_n = l$. We talk about the zero versus nonzero drift issues in detail in the main text.

We should also point out that with nearly-weak/strong identification case with $0 < \kappa < 1/2$, we have $\tau \to \infty$, since $n^{1/2-\kappa}l_n^{1/2} \to \infty$, and (SA.44) is used in that case. Also with weak identification, the case of $0 < \kappa < 1/2$ brings uncertainty to $\tau$, since $n^{1/2}S_n^{-1}G' = o(1)$, and $n^{1/2-\kappa}l_n^{1/2} \to \infty$, it is not clear which one will dominate in (SA.46). From now on, we will not consider the case of $0 < \kappa < 1/2$, since either the drift is going to infinity, or as described earlier the drift term has an uncertain case.

After handling the drift term, we consider (SA.30):

$$n^{1/2}\zeta'S_n^{-1}EU'O^{-1/2}(n^{1/2-\kappa}\Omega^{-1/2}C_1)$$ \hspace{1cm} (SA.47)

In the above term, we consider the following expression, by using $n^{1/2-\kappa}l_n^{1/2} = O(1),

$$n^{1/2}\zeta'S_n^{-1}EU'O^{-1/2} = O(n^{1/2-\kappa}l_n^{1/2}/\mu_n) = o(1),$$

where we use $EU = n^{-1}\sum_{i=1}^{n}EU_i$ definition in (SA.18), $\|C_1\| = O(l_n^{1/2})$, and $S_n^{-1}$ definitions. Next, (SA.21) and the result immediately above show that

$$n^{1/2}\zeta'S_n^{-1}EU'O^{-1/2}(n^{1/2-\kappa}\Omega^{-1/2}C_1) \to 0.$$ \hspace{1cm} (SA.48)

Next, we combine (SA.40), (SA.41), (SA.46), with $n^{-2\kappa}l_n^{1/2} = O(1)$, and strong/near weak identification setup to have $\|n^{1/2}S_n^{-1}G'\| = O(1)$. The limit of the score is:

$$n\mu_n^{-1}\zeta'\frac{\partial Q(0)}{\partial \delta} \overset{d}{\to} \nu_2,$$  \hspace{1cm} (SA.49)

where $\nu_2 \equiv N(\tau, H + \Lambda + \Delta)$.

(ii). The second case is similar to case one in which $n^{1/2-\kappa}l_n^{1/2} = O(1)$, but with the possibility of weak identification in the sense of Example 2, which is $\|n^{1/2}S_n^{-1}G'\| = o(1)$.

$$nS_n^{-1}[G'\Omega^{-1}E\tilde{y}] = \sqrt{n}S_n^{-1}G'\Omega^{-1}n^{1/2-\kappa}C_1 \to 0.$$ \hspace{1cm} (SA.50)

So this last equation shows that, regardless of weak or nearly-weak/strong identification, and regardless of Assumption M.1(ii) or Assumption M.1(iii), we can get a zero drift with an increasing number of imperfect instruments ($l_n$). The violations should be minor as in $1/2 < \kappa < \infty$. , by (SA.50) $\tau = 0$, but the limit of the score is

$$n\mu_n^{-1}\zeta'\frac{\partial Q(0)}{\partial \delta} \overset{d}{\to} \nu_{2nd},$$

44
where \( \nu_{2nd} \equiv N(0, H + \Lambda + \Delta) \). So this is the normal distribution without the drift seen in case one.

(iii). Now we consider the case of \( n^{1/2-\kappa} l_n^{1/2} = o(1) \). This is basically either a high \( \kappa \) (minor violation of exogeneity) or a small \( l_n \) (number of imperfect instruments) compared to the previous case. There is no change in the analysis of (SA.23)(SA.24); however, the limit of (SA.25) will be different. We see that if \( n^{1/2-\kappa} l_n^{1/2} = o(1) \), then using the same analysis in (SA.21),

\[
n^{1/2-\kappa} \Omega^{-1/2} C_1 \to 0.
\]

Next, in (SA.37), the limit will converge to zero. Also, as discussed in (SA.50), \( \tau = 0 \). This means that one of the variance term \( \Delta \) of \( \nu_2 \) above is zero. There is no drift (\( \tau = 0 \)) and the variance term \( \Delta \) is zero. Therefore, the limit of the score when \( n^{1/2-\kappa} l_n^{1/2} = o(1) \), which is

\[
n\mu_n^{-1} \frac{\partial \hat{Q}(0)}{\partial \delta} \to \nu_1.
\]

\( \nu_1 \equiv N(0, H + \Lambda) \).

Q.E.D.

We use the following lemma from Newey and Windmeijer (Lemma A.1, 2009b). First, we define general \( q_n \times 1 \) random variables. They are iid and may depend on \( n \), but the additional subscript is avoided in this lemma for notational convenience. \( A \) is a \( q_n \times q_n \) matrix. \( F_i, W_i, A, d_n \) will be specified in the subsequent lemmata. We also set

\[
\hat{F} = \frac{\sum_{i=1}^n F_i}{n}, \quad \mu_F = EF_i, \quad \Sigma_{FF} = EF_iF_i', \quad \Sigma_{FW} = EF_iW_i', \quad \Sigma_{WW} = EW_iW_i'.
\]

**Lemma SA.7.** If \( \text{Eigmax}(AA') \leq C, \text{Eigmax}(A'A) \leq C, \text{Eigmax}(\Sigma_{FF}) \leq C, \text{Eigmax}(\Sigma_{WW}) \leq C, \text{Eigmax}(\Sigma_{WW}) \leq C, E(F_iF_i)^2/(nd_n^2) \to 0, E(W_iW_i)^2/(nd_n^2) \to 0, q_n/d_n^2 \to 0, d_n/n \leq C, n\mu_F^2/\mu_F^2/d_n^2 \to 0, n\mu_W^2/\mu_W^2/d_n^2 \to 0, \) then

\[
nF'AW/d_n = tr(\Sigma_{FW}')/d_n + n\mu_F^2\mu_W/d_n + o_p(1).
\]

Next, Lemma SA.8 deals with the limit of second order partial derivative of the objective function. This Lemma is an extension of Lemma A.13 in Newey and Windmeijer (2009b) to the case of near exogeneity. The proof simply follows from Newey and Windmeijer (2009b).

**Lemma SA.8.** Under Assumptions M.1-M.6, there is an open convex set \( N_n \) so that \( 0 \in N_n \), and wpa1 \( \delta \in N_n \), \( \hat{Q}(\delta) \) is twice continuously differentiable on \( N_n \), and for any \( \delta \), that is an element of \( N_n \) wpa1, with \( n^{1/2-\kappa} l_n^{1/2} = O(1) \) we obtain

\[
\mu_n^{-2n} \frac{\partial^2 \hat{Q}(\delta)}{\partial \delta \partial \delta'} \xrightarrow{p} H.
\]

Remark. It is also clear that the results of the lemma hold with \( n^{1/2-\kappa} l_n^{1/2} \to 0 \) condition. These results show that if \( \kappa = 1/2 \) then with \( l_n = l \) the result holds. If \( 1/2 < \kappa < \infty \), then \( n^{1/2-\kappa} l_n^{1/2} = O(1) \) may be satisfied with appropriate choice of \( l_n \).
Proof of Lemma SA.8. First, we denote $Q_{k,l}(0)$ as the second order partial derivative of the objective function $Q(\delta)$ at $\delta = 0$. Since we replace sample covariances with its limit terms, we use $\hat{Q}$ instead of $\hat{Q}$. An important fact to note is that our Assumption M.1 is not used in the proof of the limit of the Hessian in Lemma A.13 of Newey and Windmeijer (2009b) until p.24 in the proof. Assumption M.1 is needed when we evaluate the objective function at $\theta_0, (\delta = 0)$. Just as in the proof of Lemma A.13 of Newey and Windmeijer (2009b), we have the following notation: $\delta = S_i^{*}(\theta - \theta_0) / \mu_n, \tilde{g}_k = \delta \tilde{g}_k = \frac{\mu_n \tilde{g}_k}{\mu_n}$, $\bar{g} = \frac{1}{n} \sum_{i=1}^{n} g_i, \bar{g}_{k,\delta_i} = \frac{\partial^2 g_i(\delta = 0)}{\partial \delta_k \partial \delta_l}$, $\Omega^k = E g_i \tilde{g}_k^* \tilde{g}_i^*$ where $g_i = g_i(\delta = 0), \tilde{g}_{k,\delta_i} = \partial g_i(\delta = 0) / \partial \delta_k$. $\Omega^{kl} = E \left[ g_i g_{k,\delta_i} g_{l,\delta_i} \right], g_{k,\delta_i} (\delta = 0) = \partial^2 g_i(\delta = 0) / \partial \delta_k \partial \delta_l$, and $\Omega^{kl} = E \left[ g_{\delta_k} g_{l,\delta_l} \right] (for k = 1, \ldots, p, l = 1 \cdots p)$.

The second order partial derivative of the objective function evaluated at $\delta = 0$ in p.24 of Newey and Windmeijer (2009b) is

$$Q_{k,l}(0) = \tilde{g}_k^* \Omega^{-1} \tilde{g}_l + \bar{g}^* \Omega^{-1} \tilde{g}_{k,\delta_l} - \bar{g}^* \Omega^{-1} (\Omega^k + \Omega^{kl}) \Omega^{-1} \tilde{g}_{k,\delta_l} + \bar{g}^* \Omega^{-1} (\Omega^{kl} + \Omega^{kl}) \Omega^{-1} \tilde{g}_l.$$  \hspace{1cm} (SA.51)

We first consider the third term in (SA.51) and apply Lemma SA.7. We set $F_i = g_i, W_i = G_i S_n^{-1} \mu_n e_k, d_n = \mu_n^2, A = -\Omega^{-1} (\Omega^k + \Omega^{kl}) \Omega^{-1}$. Then by Assumption M.5, $Eigmax(AA') \leq C, Eigmax(A'A) \leq C$. By Assumption M.5, $Eigmax(\Sigma_{FF}) \leq C, Eigmax(\Sigma_{WW}) \leq C$. Then by Assumption M.2, $\mu_n^2 / n \leq C, \mu_n^4 / n \rightarrow 0$. $E(F_i'F_i^2) / nd_n^2 \rightarrow 0$ by Assumption M.4, and $E(W_i'W_i^2) / nd_n^2 \rightarrow 0$ by Assumption M.5. Next, by Assumption M.1,

$$n \mu_F^2 \mu_F^2 / d_n^2 = n \mu_n^4 (E g_i)'(E g_i) = n \mu_n^4 (C_1')^4 C_1 = n \mu_n^4 O(l_n).$$  \hspace{1cm} (SA.52)

by $C_1' C_1 = O(l_n)$ from the $C_1$ definition. Next, we use the assumption in Lemma SA.6, $1/2 < \kappa < \infty$ and $n^{1/2 - \kappa} l_n^{1/2} = O(1)$ (Note with $\kappa > 1/2, n^{1/2 - \kappa} l_n^{1/2} \rightarrow 0$ we obtain the same results below). We apply this assumption to have

$$n \mu_n^4 (C_1')^4 C_1 = n \mu_n^4 O(l_n) \rightarrow 0.$$  \hspace{1cm} (SA.53)

Another possibility in Lemma SA.6, is $\kappa = 1/2$, with $l_n = l$, then

$$n \mu_n^4 (C_1')^4 C_1 = n \mu_n^4 O(l_n) = l \mu_n^4 \rightarrow 0.$$  \hspace{1cm} (SA.54)

We apply (SA.53) or (SA.54) to get

$$\frac{n \mu_F^2 \mu_F^2}{d_n^2} \rightarrow 0.$$  \hspace{1cm} (SA.55)

Next, by Lemma SA.7,

$$\frac{n \mu_F^2 \mu_F^2}{d_n^2} = \frac{n \mu_n^4 \mu_n S_n^{-1} (E g_i)'(E g_i) S_n^{-1} \mu_n e_l}{d_n^2} \rightarrow 0,$$

by $\| \sqrt{n} S_n^{-1} E g_i \| \leq C$. Thus, by Lemma SA.7,

$$\frac{n \mu_n^4 \mu_n^2}{d_n^2} \rightarrow 0.$$  \hspace{1cm} (SA.55)
So we evaluate the right side term in (SA.55), with $1/2 < \kappa < \infty$, $n^{1/2 - \kappa} l_n^{1/2} = O(1)$

\[
\frac{n \mu_n^{1/2} \mu W}{\mu_n^2} = -\frac{n}{\mu_n^2} (E g_i)' [\Omega^{-1}(\Omega^k + \Omega^{k'})\Omega^{-1}] E g_i S_n^{-1} \mu_n e_i
\]

\[
= -\frac{n^{1/2 - \kappa} C_1}{\mu_n} [\Omega^{-1}(\Omega^k + \Omega^{k'})\Omega^{-1}] \sqrt{n} E g_i S_n^{-1} e_i
\]

\[
\leq C \frac{n^{1/2 - \kappa}}{\mu_n} \|C_1\| \|\sqrt{n} E g_i S_n^{-1}\|
\]

\[
= C \frac{n^{1/2 - \kappa}}{\mu_n} O(l_n^{1/2})
\]

\[
\rightarrow 0, \quad \text{(SA.56)}
\]

by Assumption M.1 and $\|\sqrt{n} E g_i S_n^{-1}\| \leq C$, $\text{Eigenmax}(A) \leq C$, $\|C_1\| = O(l_n^{1/2})$. Note also that (SA.56) holds with $1/2 < \kappa < \infty$, $n^{1/2 - \kappa} l_n^{1/2} = o(1)$ case in Lemma SA.6, and $\kappa = 1/2$, $l_n = l$ case in Lemma SA.6, as well. Next,

\[
\Sigma_{FW} = EF_{i} W_{i}^{1} = E g_i e_i S_n^{-1} G_i = \Omega^l. \quad \text{(SA.57)}
\]

So, by (SA.55)-(SA.57),

\[
\mu_n^{-2} n \tilde{g}' \Omega^{-1}(\Omega^k + \Omega^{k'})\Omega^{-1} \tilde{g}_b = \frac{-tr(\Omega^{-1}(\Omega^k + \Omega^{k'})\Omega^{-1} \Omega^l')}{\mu_n^2} + o_p(1). \quad \text{(SA.58)}
\]

In the same way, the fourth term is

\[
\mu_n^{-2} n \tilde{g}' \Omega^{-1}(\Omega^l + \Omega^{l'})\Omega^{-1} \tilde{g}_b = -tr(\Omega^{-1}(\Omega^l + \Omega^{l'})\Omega^{-1} \Omega^k')/\mu_n^2 + o_p(1). \quad \text{(SA.59)}
\]

We consider the fifth term on the right side of (SA.51):

\[
\mu_n^{-2} n \tilde{g}' \Omega^{-1}(\Omega^l + \Omega^{l'})\Omega^{-1}(\Omega^k + \Omega^{k'})\Omega^{-1} \tilde{g}.
\]

So, we apply Lemma SA.7, and set $A = \Omega^{-1}(\Omega^l + \Omega^{l'})\Omega^{-1}(\Omega^k + \Omega^{k'})\Omega^{-1}$, $F_i = g_i, W_i = g_i, d_i = \mu_n^2$. The conditions of Lemma SA.7 are satisfied as in the third term, and with $1/2 < \kappa < \infty$, $n^{1/2 - \kappa} l_n^{1/2} = O(1)$, we get

\[
\mu_n^{-2} n \tilde{g}' A \mu W = \frac{n}{\mu_n^2} (E g_i)' A (E g_i)
\]

\[
= \frac{n^{1 - 2\kappa} C_1}{\mu_n^2} A C_1 \leq C \frac{n^{1 - 2\kappa}}{\mu_n^2} \|C_1\|^2 = n^{1 - 2\kappa} O\left(\frac{l_n}{\mu_n^2}\right) = o(1),
\]

by Assumption M.1, the $C_1$ definition, $\|C_1\|^2 = O(l_n)$ and $\text{Eigenmax}(A) \leq C$ by Assumption M.5. The convergence to zero result above also holds when $1/2 < \kappa < \infty$, $n^{1/2 - \kappa} l_n^{1/2} = o(1)$, and when $\kappa = 1/2$, $l_n = l$. Then, by $\Sigma_{FW} = \Omega$, with the application of Lemma SA.7,

\[
n \mu_n^{-2} \tilde{g}' \Omega^{-1}(\Omega^l + \Omega^{l'})\Omega^{-1}(\Omega^k + \Omega^{k'})\Omega^{-1} \tilde{g} = tr(\Omega^{-1}(\Omega^l + \Omega^{l'})\Omega^{-1}(\Omega^k + \Omega^{k'})) / \mu_n^2 + o_p(1). \quad \text{(SA.60)}
\]

In the same way,

\[
n \mu_n^{-2} \tilde{g}' \Omega^{-1}(\Omega^{kl} + \Omega^{kl'})\Omega^{-1} \tilde{g} = tr(\Omega^{-1}(\Omega^{kl} + \Omega^{kl'})) / \mu_n^2 + o_p(1). \quad \text{(SA.61)}
\]
Next, we consider the second term via the analysis for the third term,

\[
\mu_n^{-2} \tilde{g}'\Omega^{-1}\tilde{g}_b = -tr(\Omega^{-1}\Omega^{k'})/\mu_n^2 + o_p(1). \tag{SA.62}
\]

Next, we consider the first term via Lemma SA.7. We set \( F_i = G_i S_n^{-1'} \mu_n e_k, W_i = G_i S_n^{-1'} \mu_n e_l, A = \Omega^{-1}, d_n = \mu_n^2 \). All the conditions of Lemma A.7 are satisfied easily. Then, we note that \( \Sigma_{FW} = \Omega^{k,l} \), and we define

\[
\frac{n\mu'_F A \mu_W}{d_n} = n S_n^{-1}(EG_i)'\Omega^{-1}(EG_i)S_n^{-1} = H_{n,kl}.
\]

Thus,

\[
n\mu_n^{-2}\tilde{g}_b\Omega^{-1}\tilde{g}_i = tr(\Omega^{-1}\Omega^{k,l})/\mu_n^2 + H_{n,k,l} + o_p(1). \tag{SA.63}
\]

Then, substituting (SA.58)-(SA.63) in (SA.51), following p.24 of Newey and Windmeijer (2009b) by simple matrix algebra,

\[
\mu_n^{-2}n\hat{Q}_{k,l}(0) = H_{n,k,l} + o_p(1).
\]

The conclusion follows by Assumption M.2. Q.E.D.

**Lemma SA.9.** Under the Assumptions M.1-M.6, and \( n^{1/2-\kappa_1/2}/q_n^{1/2} = o(1) \)

\[
n S_n^{-1} \hat{D}(\theta)'\hat{\Omega}^{-1}\hat{D}(\theta) S_n^{-1} \overset{p}{\rightarrow} H + \Lambda.
\]

**Proof of Lemma SA.9.** Note that

\[
\hat{D}(\theta) = \hat{\rho}_1(\theta) \frac{\partial g_i(\theta)}{\partial \theta} / \sum_{i=1}^n \hat{\rho}_i(\theta).
\]

Following the proof of Lemma A.14 on p.25 of Newey and Windmeijer (2009b), since this expansion does not use Assumption M.1, we have

\[
\frac{1}{n} \sum \rho_1(\hat{\lambda}_i) \hat{g}_i S_n^{-1} \hat{D}(\theta)'\hat{\Omega}^{-1}\hat{D}(\theta) S_n^{-1} e_i = \mu_n^{-2} n (\hat{g}_b - \Omega^{k'}\Omega^{-1}\hat{g})\Omega^{-1}(\hat{g}_i - \Omega^{k'}\Omega^{-1}\hat{g}) + o_p(1), \tag{SA.64}
\]

where \( \hat{\Omega} = \hat{\Omega}(\hat{\theta}), \hat{g}_i = g_i(\hat{\theta}) \). Next, we rewrite the right side of (SA.64)

\[
\mu_n^{-2} n (\hat{g}_b - \Omega^{k'}\Omega^{-1}\hat{g})\Omega^{-1}(\hat{g}_i - \Omega^{k'}\Omega^{-1}\hat{g}) = ne_k S_n^{-1}[G + \hat{U}]\Omega^{-1}[G + \hat{U}] S_n^{-1} e_l, \tag{SA.65}
\]

where \( \hat{U} = \sum_{i=1}^n U_i/n, U_i \) is defined in (SA.15). By Assumption M.2,

\[
H_n = n S_n^{-1} G'\Omega^{-1} GS_n^{-1} \rightarrow H. \tag{SA.66}
\]

By Assumption M.5, we have \( \text{Eigmax} \left( EU_i S_n^{-1'} e_l U_i' S_n^{-1} \right) \leq C/\mu_n^2 \). This inequality provides

\[
E[ne_k S_n^{-1} G'\Omega^{-1} \hat{U} S_n^{-1'} e_l e_l] \leq C ne_k S_n^{-1} G'\Omega^{-1} \hat{U} S_n^{-1'} e_l \leq Cne_k S_n^{-1} G'\Omega^{-2} GS_n^{-1'} e_k \leq Chn_{kk}/\mu_n^2 \rightarrow 0,
\]

where \( H_{n,kk} \) represents the matrix \( H_n \)'s (k,k) th element in (SA.66). We see that

\[
ne_k S_n^{-1}(\hat{U}'\Omega^{-1}\hat{U}) S_n^{-1'} e_l - ne_k S_n^{-1}[(\hat{U} - E\hat{U})'\Omega^{-1} (\hat{U} - E\hat{U})] S_n^{-1'} e_l \overset{p}{\rightarrow} 0,
\]

48
and 

$$n e_k' S_n^{-1} (\bar{U} - E\bar{U})' \Omega^{-1} (\bar{U} - E\bar{U}) S_n^{-1} e_l \xrightarrow{p} \Lambda_{k,l}.$$ 

The first result follows the definition of (SA.18) with $n^{1/2-\kappa_l/2}/\mu_n \to 0$, and Assumption M.2(ii). The result is 

$$n e_k' S_n^{-1} E\bar{U}' \Omega^{-1} E\bar{U} S_n^{-1} e_l \to 0.$$ 

With the same reasoning, we show that the cross product term also converges in probability to zero. The second result follows Lemma SA.7 and the assumption on the statement of Theorem 2. Namely, we apply Lemma SA.7 to $n e_k' S_n^{-1} (\bar{U} - E\bar{U})' \Omega^{-1} (\bar{U} - E\bar{U}) S_n^{-1} e_l$. We set $A = \Omega^{-1}$, $F_i = (U_i - E\bar{U})' S_n^{-1} e_k \mu_n$, $W_i = (U_i - E\bar{U})' S_n^{-1} e_l \mu_n$, $d_n = \mu_n^2$. The rest of the proof follows from p.26 of Newey and Windmeijer (2009b). So combining these in (SA.65) (for (k,l) th element of the matrices, $H$, $\Lambda$), we get 

$$e_k' S_n^{-1} \hat{D}(\hat{\theta})' \hat{\Omega}^{-1} \hat{D}(\hat{\theta}) S_n^{-1} e_l \xrightarrow{p} H_{kl} + \Lambda_{kl}.$$ 

Q.E.D.

References


Kleibergen, F., 2005. Testing in parameters in GMM without assuming that they are identified. Econometrica 73, 1103-1123.