EQUIVALENCE OF THE HIGHER ORDER ASYMPTOTIC EFFICIENCY OF k-STEP AND EXTREMUM STATISTICS

BY

DONALD W.K. ANDREWS

COWLES FOUNDATION PAPER NO. 1044

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281
2002

Econometric Theory, **18**, 2002, 1040–1085. Printed in the United States of America.
DOI: 10.1017.S0266466602185021

EQUIVALENCE OF THE HIGHER ORDER ASYMPTOTIC EFFICIENCY OF *k*-STEP AND EXTREMUM STATISTICS

DONALD W.K. ANDREWS
Cowles Foundation for Research in Economics, Yale University

It is well known that a one-step scoring estimator that starts from any $N^{1/2}$-consistent estimator has the same first-order asymptotic efficiency as the maximum likelihood estimator. This paper extends this result to k-step estimators and test statistics for $k \geq 1$, higher order asymptotic efficiency, and general extremum estimators and test statistics.

The paper shows that a k-step estimator has the same higher order asymptotic efficiency, to any given order, as the extremum estimator toward which it is stepping, provided (i) k is sufficiently large, (ii) some smoothness and moment conditions hold, and (iii) a condition on the initial estimator holds.

For example, for the Newton–Raphson k-step estimator based on an initial estimator in a wide class, we obtain asymptotic equivalence to integer order s provided $2^k \geq s + 1$. Thus, for $k = 1$, 2, and 3, one obtains asymptotic equivalence to first, third, and seventh orders, respectively. This means that the maximum differences between the probabilities that the ($N^{1/2}$-normalized) k-step and extremum estimators lie in any convex set are $o(1)$, $o(N^{-3/2})$, and $o(N^{-3})$, respectively.

1. INTRODUCTION

In this paper, we consider the differences between statistics that are based on an extremum estimator $\hat{\theta}_N$ and corresponding statistics that are based on a k-step estimator $\hat{\theta}_{N,k}$ that starts from some initial estimator $\hat{\theta}_{N,0}$ and takes k steps toward $\hat{\theta}_N$. Robinson (1988, Theorem 2) shows that the stochastic difference between such estimators declines to zero as $N \to \infty$ and that the magnitude of the difference declines very quickly as a function of k. Here, we show that the convex variational distance (defined subsequently) between the distributions of such estimators declines to zero very quickly as $N \to \infty$ at a rate that increases very quickly as a function of k. This result establishes the equivalence of the

The author thanks Carol Copeland for proofreading the manuscript. The author gratefully acknowledges the research support of the National Science Foundation via grant number SBR-9730277. Address correspondence to: Donald W.K. Andrews, Cowles Foundation, P.O. Box 208281, New Haven, CT 06520-8281; e-mail: donald.andrews@yale.edu.

  0266-4666/02 $9.50

higher order asymptotic efficiency of the k-step and extremum estimators. The magnitude of the order of equivalence depends on k, on moment and smoothness conditions, and on the initial estimator.

We also establish analogous results that hold under the null hypothesis for the t, Wald, Lagrange multiplier (LM), quasi-likelihood ratio (QLR), and J test statistics based on the k-step and extremum estimators. The results hold for a variety of different types of k-step estimators, including Newton–Raphson (NR), default NR, line-search NR, and Gauss–Newton (GN) k-step estimators. The results hold for stationary asymptotically weakly dependent time series observations and also for independent and identically distributed (i.i.d.) observations. The results hold for a variety of different extremum estimators, including generalized method of moments (GMM), maximum likelihood (ML), and least squares (LS) estimators. The results cover GMM estimators with a fixed weight matrix, called FW-GMM estimators, and GMM estimators with an estimated weight matrix based on a preliminary FW-GMM estimator, called EW-GMM estimators.

Let $\mathcal{B}_L$ denote the class of all convex Borel measurable sets in R^L. The convex variational (CV) distance between the distributions of two L_Y-valued random vectors Y_1 and Y_2 is defined to be

$$d_{CV}(Y_1, Y_2) = \sup_{B \in \mathcal{B}_{L_Y}} |P(Y_1 \in B) - P(Y_2 \in B)|. \tag{1.1}$$

We say that two $N^{1/2}$-consistent estimators $\hat{\theta}_{1,N}$ and $\hat{\theta}_{2,N}$ of a parameter $\theta_0 \in \Theta \subset R^{L_\theta}$ have equal s-order asymptotic efficiency if

$$d_{CV}(N^{1/2}(\hat{\theta}_{1,N} - \theta_0), N^{1/2}(\hat{\theta}_{2,N} - \theta_0)) = o(N^{-a}) \quad \text{for } a = (s-1)/2. \tag{1.2}$$

Analogously, two test statistics $T_{1,N}$ and $T_{2,N}$ have equal s-order asymptotic efficiency if $d_{CV}(T_{1,N}, T_{2,N}) = o(N^{-a})$ for $a = (s-1)/2$.[1,2]

Higher order asymptotic efficiency is defined in terms of CV distances rather than stochastic differences, because the main use of asymptotic results is to provide approximations to the distributions of statistics. The magnitudes of the errors of these approximations are assessed directly by CV distances. Higher order asymptotic efficiency measures the rate at which these errors go to zero as $N \to \infty$.

We now summarize some of the results for the case where the estimator used to initiate the k-step estimator satisfies an $N^{1/2}$-consistency type of condition that is shown to hold for a broad class of estimators.[3] For the NR, default NR, and line-search NR k-step estimators, we show that the CV distance between the distributions of the ($N^{1/2}$-normalized) k-step estimator and the corresponding extremum estimator is of order $o(N^{-a})$ for any $a \geq 0$ with $2a$ an integer, provided $2^k \geq 2a + 2$. In terms of equality of s-order asymptotic efficiency, the requirement is $2^k \geq s + 1$. Hence, for $k = 2$, we have $a = 1$ and $s = 3$; for $k = 3$, we have $a = 3$ and $s = 7$; for $k = 4$, we have $a = 7$ and

$s = 15$; etc. Analogous results are shown to hold for t, Wald, and LM test statistics. For the QLR statistic based on an EW-GMM estimator and for the J-statistic for testing overidentifying restrictions, somewhat weaker conditions suffice: $2^k \geq 2a + 1$ or $2^k \geq s$. For the QLR statistic in likelihood contexts, even weaker conditions suffice: $2^{k+1} \geq 2a + 3$ or $2^{k+1} \geq s + 2$.

For GN k-step estimators, we show that the CV distance between the distributions of the ($N^{1/2}$-normalized) k-step estimator and the corresponding extremum estimator is of order $o(N^{-a})$ for any $a \geq 0$ with $2a$ an integer, provided $k \geq 2a + 1$. In terms of equality of s-order asymptotic efficiency, the requirement is $k \geq s$. Hence, in this scenario, for $k = 2$, we have $a = \frac{1}{2}$ and $s = 2$; for $k = 3$, we have $a = 1$ and $s = 3$; for $k = 4$, we have $a = \frac{3}{2}$ and $s = 4$; etc. Analogous results are shown to hold for t, Wald, and LM test statistics. For the QLR statistic based on an EW-GMM estimator and for the J-statistic, weaker conditions suffice: $k \geq 2a$ or $k \geq s - 1$. For the QLR statistic in likelihood contexts, even weaker conditions suffice: $2k \geq 2a + 1$ or $2k \geq s$.

The results of the paper can be useful in practice to obtain an estimator that has the same desirable higher order asymptotic efficiency properties as some extremum estimator without having to compute the extremum estimator. The results show that it suffices to compute any extremum estimator based on a well-behaved criterion function and to take a sufficiently large number of steps k from it and toward the extremum estimator of interest. The results also can be useful to obtain a $N^{1/2}$-consistent estimator, which may have desirable first- or higher order asymptotic efficiency properties, starting from an initial estimator that is only N^{-c}-consistent for some $c \in (0, \frac{1}{2})$.

On the other hand, one has to be careful in applying the theoretical results of the paper, because they rely on the initial estimator being in a neighborhood of the true value. If the initial estimator is far from the true value and the extremum estimator criterion function at hand has multiple local minima, then the asymptotic results will not be reflected closely in the finite sample behavior.

The results of the paper extend results of Pfanzagl (1974), Pfanzagl and Wefelmeyer (1978), Janssen, Jureckova, and Veraverbeke (1985), Robinson (1988), and others. One-step estimators were first considered by Fisher (1925) and LeCam (1956). Papers in the literature that consider higher order asymptotic efficiency of estimators include Pfanzagl (1974), Pfanzagl and Wefelmeyer (1978), Akahira and Takeuchi (1981), Rothenberg (1984), and Robinson (1988), among others. Papers that consider k-step bootstrap estimators include Davidson and MacKinnon (1999) and Andrews (2002). Davidson and MacKinnon (1999) point out that k-step likelihood ratio bootstrap statistics require fewer steps than other k-step bootstrap test procedures, which is analogous to what we find here.

Proofs in this paper rely heavily on methods used by Hall and Horowitz (1996), who consider higher order properties of bootstrap procedures for GMM estimators. In turn, the methods of Hall and Horowitz (1996) build on those of Bhattacharya and Ghosh (1978) and Götze and Hipp (1983, 1994). Parts of our proofs are similar to those of Robinson (1988). The methods of Robinson (1988) are related to those of Pfanzagl (1974) and to results in the numerical analysis literature on convergence of iterative optimization algorithms (e.g., see Dennis and Schnabel, 1983, Sect. 5.2).

Throughout the paper, $\ln^b(N)$ denotes $(\ln(N))^b$.

The remainder of the paper is organized as follows. Section 2 provides an outline of the results and their proof. Section 3 defines the extremum estimators and test statistics. Section 4 introduces the *k*-step estimators and test statistics. Section 5 presents the assumptions used. Section 6 states the higher order equivalence results. Section 7 contains proofs of the results.

2. OUTLINE OF THE RESULTS

In this section, we provide an outline of the methods and results established in detail in the sections that follow.

An extremum estimator $\hat{\theta}_N$ of a parameter $\theta \in \Theta$ is defined to minimize a criterion function $J_N(\theta)$ over Θ. For example, $J_N(\theta)$ could be a GMM or an ML criterion function. The true parameter value is θ_0.

Let $\hat{\theta}_{N,0}$ denote the estimator used to initiate the *k*-step estimator. The *k*-step estimator is defined recursively as follows:

$$\hat{\theta}_{N,j} = \hat{\theta}_{N,j-1} - (Q_{N,j-1})^{-1} \frac{\partial}{\partial\theta} J_N(\hat{\theta}_{N,j-1}) \quad \text{for } j = 1, \ldots, k, \tag{2.1}$$

where $Q_{N,j-1}$ is a matrix that depends on $\hat{\theta}_{N,j-1}$. For NR steps, $Q_{N,j-1} = (\partial^2/\partial\theta\partial\theta')\ J_N(\hat{\theta}_{N,j-1})$. In this case, the definition of $\hat{\theta}_{N,k}$ is motivated by the approximation of $(\partial/\partial\theta)J_N(\theta)$ at the $k-1$ step by the affine function

$$A_{N,k-1}(\theta) = (\partial/\partial\theta)J_N(\hat{\theta}_{N,k-1}) + (\partial^2/\partial\theta\partial\theta')J_N(\hat{\theta}_{N,k-1})(\theta - \hat{\theta}_{N,k-1}). \tag{2.2}$$

The value of θ that solves the approximate first-order conditions $A_{N,k-1}(\theta) = 0$ is easily seen to be $\hat{\theta}_{N,k}$. For brevity, in this section we only consider the NR choice of $Q_{N,j-1}$.

We want to show for some $a \geq 0$, where $2a$ is an integer, that

$$\sup_{B \in \mathcal{B}_{L_\theta}} |P(N^{1/2}(\hat{\theta}_{N,k} - \theta_0) \in B) - P(N^{1/2}(\hat{\theta}_N - \theta_0) \in B)| = o(N^{-a}). \tag{2.3}$$

This implies that $N^{1/2}(\hat{\theta}_{N,k} - \theta_0)$ and $N^{1/2}(\hat{\theta}_N - \theta_0)$ are asymptotically equivalent to integer order $s = 2a + 1$. First, we show that the distribution of $N^{1/2}(\hat{\theta}_N - \theta_0)$ possesses a well-behaved Edgeworth expansion with remainder of order $o(N^{-a})$. In consequence, a small change in z yields a small change in $P(N^{1/2}(\hat{\theta}_N - \theta_0) + z \in B)$. This is used to show that (2.3) holds if $N^{1/2}(\hat{\theta}_{N,k} - \theta_0)$ and $N^{1/2}(\hat{\theta}_N - \theta_0)$ are close in the sense that

$$P(\|N^{1/2}(\hat{\theta}_{N,k} - \theta_0) - N^{1/2}(\hat{\theta}_N - \theta_0)\| > \omega_N) = o(N^{-a}) \tag{2.4}$$

for some constants $\omega_N = o(N^{-a})$. Note that the left-hand side of (2.4) equals $P(\|\hat{\theta}_{N,k} - \hat{\theta}_N\| > N^{-1/2}\omega_N)$.

The initial estimator $\hat{\theta}_{N,0}$ is assumed to satisfy the following condition. For some finite constant C_1,

$$P(\|\hat{\theta}_{N,0} - \theta_0\| > C_1\gamma_N) = o(N^{-a}), \tag{2.5}$$

where $\{\gamma_N : N \geq 1\}$ is a sequence of constants that satisfies $\lim_{N\to\infty}\gamma_N = 0$ and $\gamma_N \geq (\ln(N)/N)^{1/2}$ for all $N \geq 1$. For example, we show that (2.5) holds with $\gamma_N = (\ln(N)/N)^{1/2}$ if $\hat{\theta}_{N,0}$ is an extremum estimator, provided the estimator criterion function is sufficiently smooth and terms that arise in its Taylor expansion about θ_0 have sufficiently many finite moments. (See Lemma 1, which follows.) But, it could be the case that the initial estimator converges more slowly and $\gamma_N = N^{-c}$ for some $c \in (0, \frac{1}{2})$. For example, this occurs if one minimizes an extremum estimator criterion function over a discrete grid of fixed points or over a set of randomly selected points (see Robinson, 1988, Theorem 8).

Given that the initial estimator satisfies (2.5) for some sequence $\{\gamma_N : N \geq 1\}$, we show that (2.4) holds with $\omega_N = C_3 N^{1/2}\gamma_N^{2^k}$ for some $C_3 < \infty$ for the NR choice of $Q_{N,j-1}$. The expression $\omega_N = C_3 N^{1/2}\gamma_N^{2^k}$ corresponds to quadratic convergence of $\hat{\theta}_{N,k}$ to $\hat{\theta}_N$ as the number of steps k increases, which is very fast.

In the leading case where $\gamma_N = (\ln(N)/N)^{1/2}$, we have $\omega_N = o(N^{-a})$ (as is required for (2.4)) provided $2^k \geq 2a + 2$. For $k = 2$, this holds for $a = 1$, which corresponds to asymptotic equivalence of $\hat{\theta}_{N,k}$ and $\hat{\theta}_N$ to order $s = 3$, because $s = 2a + 1$. For $k = 3$, this holds for $a = 3$, which corresponds to asymptotic equivalence to order $s = 7$.

For an initial estimator for which $\gamma_N = N^{-c}$ for $c \in (0, \frac{1}{2})$, $\omega_N = C_3 N^{1/2}\gamma_N^{2^k} = o(N^{-a})$ provided $2^k > (a + \frac{1}{2})/c$. For example, for $c = \frac{1}{4}$ and $k = 2$, this holds for $a = 0$, which corresponds to asymptotic equivalence of order $s = 1$. For $c = \frac{1}{4}$ and $k = 3$, this holds for $a = 1$, which corresponds to asymptotic equivalence of order $s = 3$. For $c = \frac{1}{4}$ and $k = 4$, this holds for $a = 3$ and $s = 7$. A larger number of steps k are needed to achieve a given order s of asymptotic equivalence when the initial estimator $\hat{\theta}_{N,0}$ has a slower rate of convergence.

For the GN choice of $Q_{N,j-1}$, the expression for ω_N is different from $C_3 N^{1/2}\gamma_N^{2^k}$, the rate convergence of $\hat{\theta}_{N,k}$ to $\hat{\theta}_N$ is slower, and k needs to be larger to obtain the same order of asymptotic equivalence of $\hat{\theta}_{N,k}$ and $\hat{\theta}_N$.

To establish (2.4), we show that (i) the difference between $\hat{\theta}_{N,k}$ and $\hat{\theta}_N$ depends on the difference between $(\partial/\partial\theta)J_N(\theta)$ and its affine approximation $A_{N,k-1}(\theta)$ both evaluated at $\theta = \hat{\theta}_N$ and (ii) the latter difference is a quadratic function of the difference between $\hat{\theta}_{N,k-1}$ and $\hat{\theta}_N$. Our proof parallels the standard proof in the numerical analysis literature of the quadratic convergence of the NR algorithm (e.g., see Dennis and Schnabel, 1983, Sect. 5.2). For notational simplicity, let $\nabla^2 J_{N,k-1}$ denote $(\partial^2/\partial\theta\partial\theta')J_N(\hat{\theta}_{N,k-1})$. By the definition of $\hat{\theta}_{N,k}$,

$$\begin{aligned}\hat{\theta}_{N,k} - \hat{\theta}_N &= \hat{\theta}_{N,k-1} - (\nabla^2 J_{N,k-1})^{-1} \frac{\partial}{\partial \theta} J_N(\hat{\theta}_{N,k-1}) - \hat{\theta}_N \\ &= (\nabla^2 J_{N,k-1})^{-1} \left(\frac{\partial}{\partial \theta} J_N(\hat{\theta}_N) - \frac{\partial}{\partial \theta} J_N(\hat{\theta}_{N,k-1}) \right. \\ &\qquad \left. - \nabla^2 J_{N,k-1}(\hat{\theta}_N - \hat{\theta}_{N,k-1}) \right) \\ &= (\nabla^2 J_{N,k-1})^{-1} \left(\frac{\partial}{\partial \theta} J_N(\hat{\theta}_N) - A_{N,k-1}(\hat{\theta}_N) \right), \end{aligned} \tag{2.6}$$

where the second equality holds because $(\partial/\partial\theta)J_N(\hat{\theta}_N) = 0$ with probability $1 - o(N^{-a})$ by the first-order conditions for $\hat{\theta}_N$. Element by element Taylor expansions of $(\partial/\partial\theta)J_N(\hat{\theta}_N)$ about $\hat{\theta}_{N,k-1}$ give

$$\begin{aligned}&\frac{\partial}{\partial \theta} J_N(\hat{\theta}_N) - A_{N,k-1}(\hat{\theta}_N) \\ &\quad = \left[(\hat{\theta}_N - \hat{\theta}_{N,k-1})' \frac{\partial^3}{\partial \theta_u \partial \theta \partial \theta'} J_N(\theta^+_{N,k-1,u})(\hat{\theta}_N - \hat{\theta}_{N,k-1})/2 \right]_{\text{vec}}, \end{aligned} \tag{2.7}$$

where $[b_u]_{\text{vec}}$ denotes a vector whose uth element is b_u and $\theta^+_{N,k-1,u}$ lies between $\hat{\theta}_N$ and $\hat{\theta}_{N,k-1}$.

Combining (2.6) and (2.7) gives

$$\begin{aligned}&\|\hat{\theta}_{N,k} - \hat{\theta}_N\| \le \zeta_N \|\hat{\theta}_{N,k-1} - \hat{\theta}_N\|^2, \quad \text{where} \\ &\qquad \zeta_N = \max_{j=1,\ldots,k} \|(\nabla^2 J_{N,j-1})^{-1}\| \cdot \sum_{u=1}^{L_\theta} \left\| \frac{\partial^3}{\partial \theta_u \partial \theta \partial \theta'} J_N(\theta^+_{N,j-1,u})/2 \right\|. \end{aligned} \tag{2.8}$$

We show that there exists a constant $K < \infty$ such that

$$P(\zeta_N > K) = o(N^{-a}). \tag{2.9}$$

Repeated substitution into the right-hand side of the inequality in (2.8) gives

$$\|\hat{\theta}_{N,k} - \hat{\theta}_N\| \le \zeta_N^{\phi} \|\hat{\theta}_{N,0} - \hat{\theta}_N\|^{2^k}, \tag{2.10}$$

where $\phi = \sum_{j=1}^{k} 2^{j-1}$.

We show that the extremum estimator $\hat{\theta}_N$ satisfies the following conditions. For some finite constant C_4,

$$P(\|\hat{\theta}_N - \theta_0\| > C_4(\ln(N)/N)^{1/2}) = o(N^{-a}). \tag{2.11}$$

Equations (2.5) and (2.11) and the triangle inequality combine to yield

$$P(\|\hat{\theta}_{N,0} - \hat{\theta}_N\| > C\gamma_N) = o(N^{-a}), \tag{2.12}$$

where $C = \max\{C_1, C_4\}/2$.

Combining (2.9), (2.10), and (2.12) gives the following result. For some finite constant C_3,

$$
\begin{aligned}
P(\|\hat{\theta}_{N,k} - \hat{\theta}_N\| > C_3 \gamma_N^{2^k}) \\
&\le P(\zeta_N^{\phi} \|\hat{\theta}_{N,0} - \hat{\theta}_N\|^{2^k} > C_3 \gamma_N^{2^k}) \\
&\le P(K^{\phi}(C\gamma_N)^{2^k} > C_3 \gamma_N^{2^k}) + o(N^{-a}) \\
&= o(N^{-a}), \qquad\qquad \textbf{(2.13)}
\end{aligned}
$$

where the equality holds for C_3 sufficiently large. This establishes (2.4) with $\omega_N = C_3 N^{1/2} \gamma_N^{2^k}$, as desired.

The proof of analogous results for the GN k-step estimator is similar, though more complicated, and requires k to be larger for a given value of a. The reason that k needs to be larger for the GN k-step estimator than the NR k-step estimator is that additional terms arise in (2.6) when $Q_{N,j-1}$ does not equal $\nabla^2 J_{N,k-1}$ and these terms increase the difference between $\hat{\theta}_{N,k}$ and $\hat{\theta}_N$.

The proofs for results concerning t, Wald, LM, QLR, and J test statistics under the null hypothesis also are similar to the proof outlined earlier but more complicated. The conditions relating k and a required for the t, Wald, and LM statistics are the same as those for the normalized estimator $N^{1/2}(\hat{\theta}_{N,k} - \theta_0)$, because the differences between the k-step and extremum versions of these test statistics are approximately linear functions of $N^{1/2}(\hat{\theta}_{N,k} - \hat{\theta}_N)$. The conditions required for the QLR and J-statistics are weaker than for the other statistics. The reason is that the differences between the k-step and extremum versions of these statistics are approximately quadratic functions of $N^{1/2}(\hat{\theta}_{N,k} - \hat{\theta}_N)$ and $N^{1/2}(\bar{\theta}_{N,k} - \bar{\theta}_N)$, where $\bar{\theta}_{N,k}$ and $\bar{\theta}_N$ are restricted analogues of $\hat{\theta}_{N,k}$ and $\hat{\theta}_N$ that satisfy the null hypothesis.

3. EXTREMUM STATISTICS

In this section, we define the extremum estimators and corresponding test statistics that are considered. We consider extremum estimators that are either GMM estimators or estimators that minimize a sample average. We call the latter "minimum ρ estimators," because the sample average is taken to be $N^{-1}\sum_{i=1}^N \rho(X_i,\theta)$, where $X_i \in R^{L_x}$ is a random vector, $\theta \in \Theta \subset R^{L_\theta}$ is an unknown parameter, and $\rho(\cdot,\cdot)$ is a known real function. ML, LS, and regression M estimators are examples of minimum ρ estimators. GMM estimators are based on the moment conditions $Eg(X_i,\theta_0) = 0$, where $g(\cdot,\cdot)$ is a known L_g-valued function, X_i is as before, $\theta_0 \in \Theta \subset R^{L_\theta}$ is the true unknown parameter, and $L_g \ge L_\theta$.

Minimum ρ estimators can be written as GMM estimators with $g(X_i,\theta) = (\partial/\partial\theta)\,\rho(X_i,\theta)$. It is useful to consider minimum ρ estimators separately, however, for two reasons. First, the k-step estimator may differ depending on whether the extremum estimator is written in minimum ρ or GMM form. The traditional one-step scoring estimator is obtained by writing the ML estimator as a minimum ρ estimator, not as a GMM estimator. Second, the identification condition for consistency of a minimum ρ estimator requires that there is a unique minimum of $E\rho(X_i,\theta)$ over $\theta \in \Theta$, whereas the identification condition for consistency of the GMM estimator based on the first-order conditions of the minimum ρ estimator requires that there is a unique solution to the equations $E(\partial/\partial\theta)\rho(X_i,\theta) = 0$ over $\theta \in \Theta$. The latter may have multiple solutions even though the former has a unique minimum.

The observations are $\{X_i : i = 1,\ldots,n\}$. They are assumed to be from a (strictly) stationary and ergodic sequence of random vectors. We assume that the true moment functions $\{g(X_i,\theta_0) : i \geq 1\}$ (for a GMM or minimum ρ estimator) are uncorrelated beyond lags of length κ for some $0 \leq \kappa < \infty$. That is, $Eg(X_i,\theta_0)g(X_{i+j},\theta_0)' = 0$ for all $j > \kappa$. This assumption is satisfied with $\kappa = 0$ in many time series models in which the estimator moment functions form a martingale difference sequence as a result of optimizing behavior by economic agents, because of inheritance of this property from a regression error term, or because of the martingale difference property of the ML score function. It also holds with $0 < \kappa < \infty$ in many models with rational expectations and/or overlapping forecast errors, such as Hansen and Hodrick (1980), Brown and Maital (1981), and Hansen and Singleton (1982). For additional references, see Hansen and Singleton (1996).

A consequence of the assumption that $Eg(X_i,\theta_0)g(X_{i+j},\theta_0)' = 0$ for all $j > \kappa$ is that the covariance matrix estimator and the asymptotically optimal weight matrix for the GMM estimator only depend on terms of the form $g(X_i,\theta)g(X_{i+j},\theta)'$ for $0 \leq j \leq \kappa$. This means that the covariance matrix estimator and the weight matrix can be written as sample averages, which allows us to use the Edgeworth expansion results of Götze and Hipp (1983, 1994) for sample averages of stationary dependent random vectors, as in Hall and Horowitz (1996).

To this end, we let

$$\widetilde{X}_i = (X_i', X_{i+1}', \ldots, X_{i+\kappa}')' \quad \text{for } i = 1,\ldots,n-\kappa. \tag{3.1}$$

All of the statistics considered subsequently can be closely approximated by sample averages of functions of the random vectors $\widetilde{X}_i$ in the sample χ_N:

$$\chi_N = \{\widetilde{X}_i : i = 1,\ldots,N\}, \tag{3.2}$$

where $N = n - \kappa$.

We consider two types of GMM estimator. The first is a FW-GMM estimator that utilizes an $L_g \times L_g$ nonrandom positive-definite symmetric weight matrix Ω. In practice, Ω is often taken to be the identity matrix I_{L_g}. The second is an EW-GMM estimator that uses a weight matrix that depends on a preliminary FW-GMM estimator and is asymptotically optimal to first order. In the literature this estimator is sometimes called a *two-step GMM estimator*. We do not use this terminology, because we reserve the term *k-step GMM estimator* for the iterative estimator that is the main focus of this paper.

The FW-GMM estimator, $\hat{\theta}_N$, minimizes $J_N(\theta)$ over Θ, where

$$J_N(\theta) = \left(N^{-1}\sum_{i=1}^{N} g(X_i,\theta)\right)'\Omega\left(N^{-1}\sum_{i=1}^{N} g(X_i,\theta)\right). \tag{3.3}$$

The EW-GMM estimator, which, for economy of notation, we also denote by $\hat{\theta}_N$, minimizes $J_N(\theta,\tilde{\theta}_N)$ over Θ, where

$$\begin{aligned} J_N(\theta,\tilde{\theta}_N) &= \left(N^{-1}\sum_{i=1}^{N} g(X_i,\theta)\right)'\Omega_N(\tilde{\theta}_N)\left(N^{-1}\sum_{i=1}^{N} g(X_i,\theta)\right), \quad \text{where} \\ \Omega_N(\theta) &= \overline{W}_N^{-1}(\theta), \\ \overline{W}_N(\theta) &= N^{-1}\sum_{i=1}^{N}\left(g(X_i,\theta)g(X_i,\theta)' + \sum_{j=1}^{\kappa} H(X_i,X_{i+j},\theta)\right), \\ H(X_i,X_{i+j},\theta) &= g(X_i,\theta)g(X_{i+j},\theta)' + g(X_{i+j},\theta)g(X_i,\theta)', \end{aligned} \tag{3.4}$$

and $\tilde{\theta}_N$ minimizes (3.3).[4]

The minimum ρ estimator, which we also denote by $\hat{\theta}_N$, minimizes $\rho_N(\theta)$ over Θ, where

$$\rho_N(\theta) = N^{-1}\sum_{i=1}^{N}\rho(X_i,\theta). \tag{3.5}$$

For this estimator, we let $g(X_i,\theta)$ denote $(\partial/\partial\theta)\rho(X_i,\theta)$.

The asymptotic covariance matrix, σ, of the extremum estimator $\hat{\theta}_N$ is

$$\sigma = \begin{cases} (D'\Omega D)^{-1}D'\Omega\Omega_0^{-1}\Omega D(D'\Omega D)^{-1} & \text{for FW-GMM} \\ (D'\Omega_0 D)^{-1} & \text{for EW-GMM} \\ D^{-1}\Omega_0^{-1}D^{-1} & \text{for minimum } \rho, \quad \text{where} \end{cases}$$

$$\Omega_0 = (E\overline{W}_N(\theta_0))^{-1} \quad \text{and} \quad D = E\frac{\partial}{\partial\theta'}g(X_i,\theta_0). \tag{3.6}$$

A consistent estimator of σ is

$$\sigma_N = \sigma_N(\hat{\theta}_N), \quad \text{where}$$

$$\sigma_N(\theta) = \begin{cases} (D_N'(\theta)\Omega D_N(\theta))^{-1}D_N(\theta)'\Omega\Omega_N^{-1}(\theta)\Omega D_N(\theta) & \\ \quad\times(D_N(\theta)\Omega D_N(\theta))^{-1} & \text{for FW-GMM} \\ (D_N(\theta)'\Omega_N(\theta)D_N(\theta))^{-1} & \text{for EW-GMM} \\ D_N^{-1}(\theta)\Omega_N^{-1}(\theta)D_N^{-1}(\theta) & \text{for minimum } \rho, \end{cases}$$

and

$$D_N(\theta) = N^{-1}\sum_{i=1}^{N}\frac{\partial}{\partial\theta'}g(X_i,\theta). \tag{3.7}$$

Let θ_r, $\theta_{0,r}$, and $(\hat{\theta}_N)_r$ denote the rth elements of θ, θ_0, and $\hat{\theta}_N$, respectively.[5] Let $(\sigma_N)_{rr}$ denote the (r,r)th element of σ_N. The t-statistic for testing the null hypothesis $H_0\colon \theta_r = \theta_{0,r}$ is

$$T_N = N^{1/2}((\hat{\theta}_N)_r - \theta_{0,r})/(\sigma_N)_{rr}^{1/2}. \tag{3.8}$$

Under H_0 and the assumptions given subsequently, T_N has an asymptotic $N(0,1)$ distribution.

Let $\eta(\theta)$ be an R^{L_η}-valued function (for some integer $L_\eta \geq 1$) that is continuously differentiable at θ_0. The Wald statistic for testing $H_0\colon \eta(\theta_0) = 0$ versus $H_1\colon \eta(\theta_0) \neq 0$ is

$$\mathcal{W}_N = N\eta(\hat{\theta}_N)'\left(\frac{\partial}{\partial\theta'}\eta(\hat{\theta}_N)\sigma_N\left(\frac{\partial}{\partial\theta'}\eta(\hat{\theta}_N)\right)'\right)^{-1}\eta(\hat{\theta}_N). \tag{3.9}$$

Under H_0 and the assumptions given subsequently, $\mathcal{W}_N$ has an asymptotic chi-squared distribution with L_η degrees of freedom.

Next, we consider the LM statistic for testing $H_0\colon \beta = 0$ versus $H_1\colon \beta \neq 0$, where $\theta = (\tau', \beta')'$ and $\beta \in R^{L_\beta}$. By definition, the restricted FW-GMM estimator, denoted $\bar{\theta}_N = (\bar{\tau}_N',0')'$, minimizes $J_N(\theta)$ over $\Theta_0 = \{\theta \in \Theta : \theta = (\tau',0')'$ for some $\tau \in R^{L_\tau}\}$. The restricted EW-GMM and minimum ρ estimators, also denoted by $\bar{\theta}_N = (\bar{\tau}_N',0')'$, minimize $J_N(\theta,\theta_N^*)$ and $\rho_N(\theta)$, respectively, over Θ_0, where θ_N^* denotes the restricted FW-GMM estimator.

The LM statistic is

$$\begin{aligned}
LM_N &= U_N(\bar{\theta}_N)'U_N(\bar{\theta}_N), \quad \text{where}\\
U_N(\theta) &= U_{1,N}(\theta)U_{2,N}(\theta),\\
U_{1,N}(\theta) &= ([0\vdots I_{L_\beta}]\sigma_N(\theta)[0\vdots I_{L_\beta}]')^{-1/2}[0\vdots I_{L_\beta}], \quad \text{and}\\
U_{2,N}(\theta) &= \begin{cases} N^{1/2}(D_N(\theta)\Omega D_N(\theta))^{-1}\dfrac{\partial}{\partial\theta}J_N(\theta) & \text{for FW-GMM}\\[2ex] N^{1/2}(D_N(\theta)\Omega_N(\theta)D_N(\theta))^{-1}\dfrac{\partial}{\partial\theta}J_N(\theta,\theta_N^*) & \text{for EW-GMM}\\[2ex] N^{1/2}D_N^{-1}(\theta)\dfrac{\partial}{\partial\theta}\rho_N(\theta) & \text{for minimum } \rho. \end{cases}
\end{aligned} \tag{3.10}$$

Under H_0 and the assumptions given subsequently, LM_N has an asymptotic chi-squared distribution with L_β degrees of freedom.

The QLR statistic for testing $H_0: \beta = 0$ versus $H_1: \beta \neq 0$ is

$$QLR_N = \begin{cases} N(J_N(\bar{\theta}_N, \theta_N^*) - J_N(\hat{\theta}_N, \tilde{\theta}_N)) & \text{for EW-GMM} \\ 2N(\rho_N(\bar{\theta}_N) - \rho_N(\hat{\theta}_N)) & \text{for minimum } \rho. \end{cases} \quad \textbf{(3.11)}$$

Under H_0 and the assumptions given subsequently, QLR_N has an asymptotic chi-squared distribution with L_β degrees of freedom when QLR_N is based on the EW-GMM estimator. When QLR_N is based on the minimum ρ estimator, the asymptotic chi-squared result requires $D = \Omega_0^{-1}$. For example, the latter holds in an ML context by the information matrix equality, provided the model is correctly specified.

We do not consider a QLR_N-statistic that is based on the FW-GMM estimator, because such a statistic has an asymptotic chi-squared null distribution only if $\Omega = \Omega_0^{-1}$. The latter is rarely satisfied in practice, because one rarely knows Ω_0.

The J-statistic for testing overidentifying restrictions is

$$J_N = K_N(\hat{\theta}_N)' K_N(\hat{\theta}_N), \quad \text{where}$$

$$K_N(\theta) = \Omega_N^{1/2}(\tilde{\theta}_N) N^{-1/2} \sum_{i=1}^{N} g(X_i, \theta), \quad \textbf{(3.12)}$$

$\hat{\theta}_N$ is the EW-GMM estimator, and $\tilde{\theta}_N$ is the FW-GMM estimator. If $L_g > L_\theta$ and the overidentifying restrictions hold, then J_N has an asymptotic chi-squared distribution with $L_g - L_\theta$ degrees of freedom under the assumptions given subsequently. (This is not true if $\hat{\theta}_N$ is the FW-GMM estimator and $\Omega_N^{1/2}(\tilde{\theta}_N)$ is replaced by $\Omega^{1/2}$ in (3.12).)

4. *k*-STEP STATISTICS

Here, we define the k-step estimators and k-step t, Wald, LM, QLR, and J-statistics. The k-step estimator is denoted $\hat{\theta}_{N,k}$. The starting value for the k-step estimator is a consistent estimator $\hat{\theta}_{N,0}$. For the FW-GMM estimator, we define recursively

$$\hat{\theta}_{N,j} = \hat{\theta}_{N,j-1} - (Q_{N,j-1})^{-1} \frac{\partial}{\partial\theta} J_N(\hat{\theta}_{N,j-1}) \quad \text{for } 1 \leq j \leq k. \quad \textbf{(4.1)}$$

For EW-GMM and minimum ρ estimators, $\hat{\theta}_{N,k}$ is defined in the same way with $(\partial/\partial\theta) J_N(\hat{\theta}_{N,j-1})$ replaced by $(\partial/\partial\theta) J_N(\hat{\theta}_{N,j-1}, \tilde{\theta}_{N,k_1})$ and $N^{-1}\sum_{i=1}^N \times g(X_i, \hat{\theta}_{N,j-1})$, respectively, where the derivative is taken with respect to the first argument of $J_N(\cdot,\cdot)$ and $\tilde{\theta}_{N,k_1}$ denotes the k_1-step FW-GMM estimator, defined in (4.1), that starts from the same estimator $\hat{\theta}_{N,0}$ as the k-step EW-GMM estimator. We assume that $k_1 \geq k$.

The $L_\theta \times L_\theta$ random matrix $Q_{N,j-1}$ depends on $\hat{\theta}_{N,j-1}$. It determines whether the k-step estimator is an NR, a default NR, a line-search NR, a GN, or some other k-step estimator. The NR, default NR, and line-search NR choices of $Q_{N,j-1}$ yield k-step estimators that have the same higher order asymptotic efficiency. The results that follow show that they require fewer steps, k, to approximate the extremum estimator $\hat{\theta}_N$ to a specified accuracy than does the GN k-step estimator. The NR choice of $Q_{N,j-1}$ is

$$Q^{NR}_{N,j-1} = \begin{cases} \dfrac{\partial^2}{\partial\theta\partial\theta'} J_N(\hat{\theta}_{N,j-1}) & \text{for FW-GMM} \\ \dfrac{\partial^2}{\partial\theta\partial\theta'} J_N(\hat{\theta}_{N,j-1}, \tilde{\theta}_{N,k_1}) & \text{for EW-GMM} \\ D_N(\hat{\theta}_{N,j-1}) & \text{for minimum } \rho, \end{cases} \tag{4.2}$$

where the derivatives of $J_N(\cdot,\cdot)$ are with respect to its first argument and $\tilde{\theta}_{N,k_1}$ is defined as before. Note that the expression for $\hat{\theta}_{N,k}$ for a minimum ρ estimator with the NR matrix $Q^{NR}_{N,j-1}$ is just the usual one-step scoring estimator starting from $\hat{\theta}_{N,k-1}$ in the case of the ML estimator with score function $g(x,\theta)$ $(= (\partial/\partial\theta)\rho(x,\theta))$. It is possible for that NR steps may move one away from the target extremum estimator. For this reason, we also consider default and line-search NR matrices $Q_{N,j-1}$.

The default NR choice of $Q_{N,j-1}$, denoted $Q^{D}_{N,j-1}$, equals $Q^{NR}_{N,j-1}$ if $Q^{NR}_{N,j-1}$ leads to an estimator $\hat{\theta}_{N,j}$ via (4.1) for which $J_N(\hat{\theta}_{N,j}) \leq J_N(\hat{\theta}_{N,j-1})$ for the FW-GMM estimator, but it equals some other matrix otherwise. In practice, one wants this other matrix to be such that $J_N(\hat{\theta}_{N,j}) < J_N(\hat{\theta}_{N,j-1})$ (but the theoretical results do not require this). For example, one might use the matrix $(1/\varepsilon)I_{L_\theta}$ for some small $\varepsilon > 0$. (For a result that indicates that such a choice will decrease the criterion function, see Ortega and Rheinboldt, 1970, Theorem 8.2.1.) For the EW-GMM and minimum ρ estimators, $J_N(\cdot)$ is replaced by $J_N(\cdot, \tilde{\theta}_{N,k_1})$ and $\rho_N(\cdot)$, respectively.

The line-search NR choice of $Q_{N,j-1}$, denoted $Q^{LS}_{N,j-1}$, uses a scaled version of the NR matrix $Q^{NR}_{N,j-1}$ that optimizes the step length. Specifically, let A be a finite subset of (0,1] of step lengths that includes 1. One computes $\hat{\theta}_{N,j}$ via (4.1) for $Q_{N,j-1} = (1/\alpha)Q^{NR}_{N,j-1}$ for each $\alpha \in A$. One takes $Q^{LS}_{N,j-1}$ to be the matrix $(1/\alpha)Q^{NR}_{N,j-1}$ for the value of α that minimizes $J_N(\hat{\theta}_{N,j})$ over all $\alpha \in A$ for the FW-GMM estimator. (If the minimizing of value of α is not unique, one takes the largest minimizing value of α in A.) For the EW-GMM and minimum ρ estimators, one replaces $J_N(\hat{\theta}_{N,j})$ by $J_N(\hat{\theta}_{N,j}, \tilde{\theta}_{N,k_1})$ and $\rho_N(\hat{\theta}_{N,j})$, respectively.

The GN choice of $Q_{N,j-1}$, denoted $Q^{GN}_{N,j-1}$, uses a matrix that differs from, but is a close approximation to, the NR matrix $Q^{NR}_{N,j-1}$. In particular,

$$Q^{GN}_{N,j-1} = \begin{cases} 2D'_{N,j-1}\Omega D_{N,j-1} & \text{for FW-GMM} \\ 2D'_{N,j-1}\Omega_N(\tilde{\theta}_{N,k_1})D_{N,j-1} & \text{for EW-GMM} \\ D_{N,j-1} & \text{for minimum } \rho, \end{cases} \tag{4.3}$$

where $D_{N,j-1}$ is determined by some function $\Delta(\cdot,\cdot)$ as follows:

$$D_{N,j-1} = N^{-1} \sum_{i=1}^{N} \Delta(\widetilde{X}_i, \hat{\theta}_{N,j-1}) \in R^{L_g \times L_\theta} \qquad \text{and}$$

$$E\Delta(\widetilde{X}_i, \theta_0) = E \frac{\partial}{\partial \theta'} g(X_i, \theta_0). \tag{4.4}$$

The latter condition is responsible for $D_{N,j-1}$ being a close approximation to $D_N(\hat{\theta}_{N,j-1})$, which appears in $Q^{NR}_{N,j-1}$. Note that, for the FW-GMM and EW-GMM estimators, $Q^{NR}_{N,j-1}$ is the sum of two terms, one of which contains $N^{-1}\sum_{i=1}^{N}(\partial^2/\partial\theta\partial\theta')\ g(X_i, \hat{\theta}_{N,j-1})$. The latter term is omitted in $Q^{GN}_{N,j-1}$. It is close to zero, because it is multiplied by the factor $N^{-1}\sum_{i=1}^{N} g(X_i, \hat{\theta}_{N,j-1})$, which is close to zero.

For an example of a GN matrix for FW-GMM or EW-GMM estimators, consider a nonlinear instrumental variables (IV) estimator for which

$$g(X_i, \theta) = U(X_i, \theta) L(Z_i, \theta) \quad \text{and} \quad E(U(X_i, \theta_0) | Z_i) = 0 \quad \text{a.s.}, \tag{4.5}$$

where $U(X_i, \theta) \in R$ is a residual, $L(Z_i, \theta) \in R^{L_g}$ is a function of some IVs Z_i, and Z_i is a subvector of X_i. In this case,

$$\frac{\partial}{\partial \theta'} g(X_i, \theta) = L(Z_i, \theta) \frac{\partial}{\partial \theta'} U(X_i, \theta) + U(X_i, \theta) \frac{\partial}{\partial \theta'} L(Z_i, \theta). \tag{4.6}$$

The GN choice of $Q_{N,j-1}$ omits the second summand of $(\partial/\partial\theta')g(X_i,\theta)$ in $D_{N,j-1}$ because $EU(X_i, \theta_0)(\partial/\partial\theta')L(Z_i, \theta_0) = 0$. That is, $Q^{GN}_{N,j-1}$ is as in (4.3) and (4.4) with

$$\Delta(\widetilde{X}_i, \theta) = L(Z_i, \theta) \frac{\partial}{\partial \theta'} U(X_i, \theta). \tag{4.7}$$

For an example of a GN matrix for a minimum ρ estimator, consider the LS estimator of a nonlinear regression model:

$$\begin{aligned}
Y_i &= q(Z_i, \theta_0) + U_i \quad \text{for } i = 1, \ldots, n, \\
\rho(X_i, \theta) &= (Y_i - q(Z_i, \theta))^2/2, \\
g(X_i, \theta) &= -(Y_i - q(Z_i, \theta)) \frac{\partial}{\partial \theta} q(Z_i, \theta), \quad \text{and} \\
\frac{\partial}{\partial \theta'} g(X_i, \theta) &= \frac{\partial}{\partial \theta} q(Z_i, \theta) \frac{\partial}{\partial \theta'} q(Z_i, \theta) + (Y_i - q(Z_i, \theta)) \frac{\partial^2}{\partial \theta \partial \theta'} q(Z_i, \theta),
\end{aligned} \tag{4.8}$$

where Y_i is a scalar dependent variable, Z_i is a vector of regressor variables, U_i is an unobserved scalar error with $E(U_i | Z_i) = 0$ a.s., and $q(\cdot,\cdot)$ is a known real function that is twice differentiable in its second argument. The GN matrix $Q^{GN}_{N,j-1}$ omits the second summand of $(\partial/\partial\theta')g(X_i,\theta)$, because $E(Y_i - q(Z_i,\theta_0))\,(\partial^2/\partial\theta\partial\theta')q(Z_i,\theta_0) = 0$. That is, $Q^{GN}_{N,j-1}$ is as in (4.3) (for minimum ρ estimators) and (4.4) with

$$\Delta(\widetilde{X}_i,\theta) = \frac{\partial}{\partial\theta}\,q(Z_i,\theta)\,\frac{\partial}{\partial\theta'}\,q(Z_i,\theta). \tag{4.9}$$

A second example of a GN matrix $Q^{GN}_{N,j-1}$ for a minimum ρ estimator is the sample outer-product estimator of the information matrix in a ML scenario. Suppose that $\rho_N(\theta)$ is a normalized negative log likelihood function and $g(X_i,\theta) = (\partial/\partial\theta)\rho(X_i,\theta)$ is the negative score (or conditional score) function for the X_ith observation. By the information matrix equality,

$$E\,\frac{\partial}{\partial\theta'}\,g(X_i,\theta_0) = Eg(X_i,\theta_0)g(X_i,\theta_0)' \tag{4.10}$$

when the model is correctly specified. In this case, the NR matrix $Q^{NR}_{N,j-1}$ is the sample analogue of the expectation on the left-hand side of (4.10): $Q^{NR}_{N,j-1} = N^{-1}\sum_{i=1}^{N}(\partial/\partial\theta')g(X_i,\hat{\theta}_{N,j-1})$. The GN matrix $Q^{GN}_{N,j-1}$ is the sample analogue of the expectation on the right-hand side of (4.10). Thus, $Q^{GN}_{N,j-1}$ is as in (4.3) (for minimum ρ estimators) and (4.4) with

$$\Delta(\widetilde{X}_i,\theta) = g(X_i,\theta)g(X_i,\theta)'. \tag{4.11}$$

The GN matrix does not require calculation of the second derivative of the log likelihood function.

Alternatively, in an ML scenario, one can use a GN matrix $Q_{N,j-1}$ based on the expected information matrix:

$$Q^{GN2}_{N,j-1} = E_\theta\,\frac{\partial}{\partial\theta'}\,g(\widetilde{X}_i,\theta)\bigg|_{\theta=\hat{\theta}_{N,j-1}}, \tag{4.12}$$

where E_θ denotes expectation when the true parameter is θ. In this case, the function $\Delta(\widetilde{X}_i,\theta)$ of (4.4) is $E_\theta(\partial/\partial\theta')g(\widetilde{X}_i,\theta)$, which is nonrandom and does not depend on $\widetilde{X}_i$. The expected information matrix is often used in the statistical literature on one-step and k-step estimators in likelihood scenarios (e.g., see Pfanzagl, 1974).

For GMM estimators that have the same number of moment conditions as the dimension of θ, such as ML estimators defined via the likelihood equations, $\hat{\theta}_{N,k}$ is the same whether defined using Ω or $\Omega_N(\tilde{\theta}_{N,k_1})$ (because the moment conditions $N^{-1}\sum_{i=1}^{N} g(X_i,\theta)$ have an exact zero with probability that goes to one at an appropriate rate as $N \to \infty$).

Next, we define the restricted k-step estimator $\bar{\tau}_{N,k}$ of τ that is used by the k-step LM and QLR statistics when the null hypothesis is $H_0\colon \beta = 0$. The restricted estimator $\bar{\tau}_{N,k}$ of τ yields the corresponding restricted estimator $\bar{\theta}_{N,k} = (\bar{\tau}'_{N,k},0')'$ of θ. The starting value for the restricted k-step estimator is an estimator $\bar{\theta}_{N,0} = (\bar{\tau}_{N,0}', 0')'$ that is consistent under $H_0: \beta = 0$. For the restricted FW-GMM estimator, we define $\bar{\tau}_{N,k}$ recursively via

$$\bar{\tau}_{N,j} = \bar{\tau}_{N,j-1} - (Q_{N,j-1}^{\tau})^{-1} \frac{\partial}{\partial \tau} J_N(\bar{\theta}_{N,j-1}) \quad \text{for } 1 \le j \le k, \tag{4.13}$$

where $(\partial/\partial\tau)J_N(\theta)$ denotes the vector of partial derivatives of $J_N(\theta)$ with respect to the first L_τ elements of θ and $Q_{N,j-1}^{\tau}$ is an $L_\tau \times L_\tau$ matrix that depends on $\bar{\tau}_{N,j-1}$. The matrix $Q_{N,j-1}^{\tau}$ determines whether the restricted k-step estimator is an NR, a default NR, a line-search NR, a GN, or some other k-step estimator. Often, $Q_{N,j-1}^{\tau}$ equals the upper $L_\tau \times L_\tau$ block of $Q_{N,j-1}$ defined with $\bar{\theta}_{N,j-1}$ in place of $\hat{\theta}_{N,j-1}$.

For EW-GMM and minimum ρ estimators, $\bar{\tau}_{N,k}$ is defined as in (4.13) with $(\partial/\partial\tau)J_N(\bar{\theta}_{N,j-1})$ replaced by $(\partial/\partial\tau)J_N(\bar{\theta}_{N,j-1}, \theta_{N,k_1}^*)$ and $(\partial/\partial\tau)\rho_N(\bar{\tau}_{N,j-1})$, respectively, where the derivative is taken with respect to the first L_τ elements of the first argument of $J_N(\cdot,\cdot)$ and θ_{N,k_1}^* denotes the restricted k_1-step FW-GMM estimator that starts at the same estimator $\bar{\theta}_{N,0}$ as the restricted k-step EW-GMM estimator. We assume that $k_1 \ge k$.

The restricted NR matrix, $Q_{N,j-1}^{\tau,NR}$, default NR matrix, $Q_{N,j-1}^{\tau,D}$, line-search NR matrix, $Q_{N,j-1}^{\tau,LS}$, and GN matrix, $Q_{N,j-1}^{\tau,GN}$, are defined as in (4.2)–(4.4) but with $\partial^2/\partial\theta\partial\theta'$, $\hat{\theta}_{N,j-1}$, $\tilde{\theta}_{N,j-1}$, $D_N(\theta)$, and $D_{N,j-1}$ replaced by $\partial^2/\partial\tau\partial\tau'$, $\bar{\theta}_{N,j-1}$, $\theta_{N,j-1}^*$, the first L_τ rows of $D_N(\theta)$, and the first L_τ rows of $D_{N,j-1}$, respectively.

We define the k-step t-statistic, $T_{N,k}$, Wald statistic, $\mathcal{W}_{N,k}$, LM statistic, $LM_{N,k}$, QLR statistic, $QLR_{N,k}$, and J-statistic, $J_{N,k}$, as in (3.8)–(3.12), but with $(\hat{\theta}_N)_r$, $\hat{\theta}_N$, $\tilde{\theta}_N$, $\bar{\theta}_N$, and θ_N^*, replaced by $\hat{\theta}_{N,k,r}$, $\hat{\theta}_{N,k}$, $\tilde{\theta}_{N,k_1}$, $\bar{\theta}_{N,k}$, and θ_{N,k_1}^*, respectively, in all parts of their definitions, where $\hat{\theta}_{N,k,r}$ denotes the rth element of $\hat{\theta}_{N,k}$, $\tilde{\theta}_{N,k_1}$ denotes the k_1-step FW-GMM estimator, and θ_{N,k_1}^* denotes the restricted k_1-step FW-GMM estimator.

5. ASSUMPTIONS

We now introduce the assumptions. They apply to the FW-GMM, EW-GMM, or minimum ρ estimator.

Let a be a non-negative constant such that $2a$ is an integer. The following assumptions depend on a and are used to show that the CV distances between the distributions of the k-step and the extremum statistics are $o(N^{-a})$. This corresponds to equality of $s = 2a + 1$–order asymptotic efficiency. The larger is a, the stronger are the assumptions.

Let $f(\widetilde{X}_i, \theta)$ denote the vector containing the unique components of X_i, $g(X_i, \theta)$, and $g(X_i, \theta)g(X_{i+j}, \theta)'$ for $j = 0, \dots, \kappa$, and their derivatives with respect to θ through order $d = \max\{2a + 2, 3\}$. Let $f(\widetilde{X}_i) = f(\widetilde{X}_i, \theta_0)$. Let $(\partial^j/\partial\theta^j)g(X_i, \theta)$ denote the vector of partial derivatives with respect to θ of order j of $g(X_i, \theta)$. For a matrix A, $\|A\|$ denotes $(\mathrm{tr}(A'A))^{1/2}$.

Assumption 1. There is a sequence of i.i.d. vectors $\{\varepsilon_i : i = -\infty,\ldots,\infty\}$ of dimension $L_\varepsilon \geq L_x$ and an $L_x \times 1$ function h such that $f(\widetilde{X}_i) = h(\varepsilon_i, \varepsilon_{i-1}, \varepsilon_{i-2}, \ldots)$. There are constants $K < \infty$ and $\xi > 0$ such that for all $m \geq 1$

$$E\|h(\varepsilon_i, \varepsilon_{i-1}, \ldots) - h(\varepsilon_i, \varepsilon_{i-1}, \ldots, \varepsilon_{i-m}, 0, 0, \ldots)\| \leq K \exp(-\xi m).$$

Assumption 2. (a) Θ is compact and θ_0 is an interior point of Θ. (b) Either (i) $\hat{\theta}_N$ minimizes $J_N(\theta)$ or $J_N(\theta, \tilde{\theta}_N)$ over $\theta \in \Theta$; θ_0 is the unique solution in Θ to $Eg(X_1, \theta) = 0$; for some function $C_g(x)$, $\|g(x,\theta_1) - g(x,\theta_2)\| \leq C_g(x)\|\theta_1 - \theta_2\|$ for all x in the support of X_1 and all $\theta_1, \theta_2 \in \Theta$; and $EC_g^{q_0}(X_1) < \infty$ and $E\|g(X_1,\theta)\|^{q_0} < \infty$ for all $\theta \in \Theta$ for $q_0 = \max\{2a+1, 2\}$ or (ii) $\hat{\theta}_N$ minimizes $\rho_N(\theta)$ over $\theta \in \Theta$; θ_0 is the unique minimum of $E\rho(X_1,\theta)$ over $\theta \in \Theta$; and $E|\rho(X_1,\theta)|^{q_0} < \infty$ for all $\theta \in \Theta$ and $E\sup_{\theta\in\Theta}\|g(X_1,\theta)\|^{q_0} < \infty$ for $q_0 = \max\{2a+1,\ 2\}$, where $g(x,\theta) = (\partial/\partial\theta)\rho(x,\theta)$.

Assumption 3. (a) $Eg(X_1,\theta_0)g(X_{1+j},\theta_0)' = 0$ for all $j > \kappa$ for $0 \leq \kappa < \infty$. (b) Ω and Ω_0 are positive definite and D is full rank L_θ. (c) $g(x,\theta)$ is $d = \max\{2a+2,\ 3\}$ times differentiable with respect to θ on N_0, some neighborhood of θ_0, for all x in the support of X_1. (d) $E\|f(\widetilde{X}_1)\|^{q_1} < \infty$ for $q_1 = 2a+3$. (e) There is a function $C_{\partial f}(\widetilde{X}_1)$ such that $\|f(\widetilde{X}_1,\theta) - f(\widetilde{X}_1,\theta_0)\| \leq C_{\partial f}(\widetilde{X}_1)\|\theta - \theta_0\|$ for all $\theta \in N_0$ and $EC_{\partial f}^{q_1}(\widetilde{X}_1) < \infty$ for $q_1 = 2a+3$. (f) If the Wald statistic is considered, the R^{L_η}-valued function $\eta(\cdot)$ is d times continuously differentiable at θ_0 and $(\partial/\partial\theta')\eta(\theta_0)$ is full rank $L_\eta \leq L_\theta$. If the LM or QLR statistic is considered, the true parameter $\theta_0 = (\tau_0', 0')'$ under H_0 is such that τ_0 is in the interior of $\{\tau : (\tau', 0')' \in \Theta_0\}$.

Assumption 4. There exist constants $K_1 < \infty$ and $\delta > 0$ such that for arbitrarily large $\zeta > 1$ and all integers $m \in (\delta^{-1}, N)$ and $t \in R^{\dim(f)}$ with $\delta < \|t\| < N^\zeta$,

$$E\left|E\left(\exp\left(\sqrt{-1}t'\sum_{s=1}^{2m+1} f(\widetilde{X}_s)\right)\middle|\{\varepsilon_j : |j-m| > K_1\}\right)\right| \leq \exp(-\delta).$$

Assumption 5. The initial estimator $\hat{\theta}_{N,0}$ satisfies the following conditions. For some finite constant C_1 and for some sequence of constants $\{\gamma_N : N \geq 1\}$ with $\lim_{N\to\infty}\gamma_N = 0$ and $\gamma_N \geq (\ln(N)/N)^{1/2}$ for all $N \geq 1$, we have

$$P(\|\hat{\theta}_{N,0} - \theta_0\| > C_1\gamma_N) = o(N^{-a}).$$

If the LM or QLR statistic is considered, the restricted initial estimator $\bar{\theta}_{N,0} = (\bar{\tau}_{N,0}', 0')'$ satisfies the same condition under H_0.

Assumption 6. The matrices $\{Q_{N,j-1} : j = 1,\ldots,k\}$ satisfy the following conditions. For some finite constant C_2 and for some sequences of constants $\{\psi_{N,j-1} : N \geq 1\}$ for $j = 1,\ldots,k$ that satisfy either (i) $\psi_{N,j-1} = 0$ for all $N \geq 1$ and $j = 1,\ldots,k$ or (ii) $\psi_{N,j-1} = \max\{\gamma_N^{2^{j-1}}, (\ln(N)/N)^{1/2}\}$, we have

$$P\left(\|Q_{N,j-1} - \frac{\partial^2}{\partial\theta\partial\theta'} J_N(\hat{\theta}_{N,j-1})\| > C_2\psi_{N,j-1}\right) = o(N^{-a}) \quad \text{for } j = 1,\ldots,k$$

for FW-GMM estimators. For EW-GMM and minimum ρ estimators, analogous conditions hold with $(\partial^2/\partial\theta\partial\theta')J_N(\hat{\theta}_{N,j-1})$ replaced by $(\partial^2/\partial\theta\partial\theta')J_N(\hat{\theta}_{N,j-1},\tilde{\theta}_{N,k_1})$ and $D_N(\hat{\theta}_{N,j-1})$, respectively. If the LM or QLR statistic is considered, $\{Q^{\tau}_{N,j-1} : j = 1,\ldots,k\}$ satisfy the same condition with the same constants $\{\psi_{N,j-1} : N \geq 1\}$ but with $(\partial^2/\partial\tau\partial\tau')J_N(\bar{\theta}_{N,j-1})$ in place of $(\partial^2/\partial\theta\partial\theta')J_N(\hat{\theta}_{N,j-1})$ for the restricted FW-GMM estimator and analogously for the restricted EW-GMM and minimum ρ estimators.

When considering the $QLR_{N,k}$-statistic, we use the following assumption.

Assumption 7. The QLR_N-statistic has an asymptotic χ^2 expansion with remainder $o(N^{-a})$. That is, there exist polynomials $\{\pi_i(z) : i = 1,\ldots,[a]\}$ in z whose coefficients are $O(1)$ such that

$$\sup_{B\in\mathcal{B}_1} |P(A_N \in B) - \int_B \left(1 + \sum_{i=1}^{[a]} N^{-i}\pi_i(z)\right) f_{\chi^2}(z)\,dz| = o(N^{-a}),$$

where $f_{\chi^2}(\cdot)$ denotes the density of some χ^2 random variable.

Assumption 1 is the same as condition (1) of Götze and Hipp (1994). It is an assumption of asymptotically weak temporal dependence of the sequence of random vectors $\{f(\widetilde{X}_i) : i \geq 1\}$. It implies that $\{f(\widetilde{X}_i) : i \geq 1\}$ are strong mixing. Assumption 1 holds automatically if $\{X_i : i \geq 1\}$ are i.i.d. Assumption 2 is a standard assumption used to obtain consistency of extremum estimators. Assumption 3 is similar to conditions in the literature used to obtain asymptotic normality of extremum estimators. But, when $a > 0$, it imposes stronger smoothness and moments restrictions than is typical. In addition, Assumption 3(a) is more restrictive than usual. See Section 3 for a discussion of Assumption 3(a). Assumption 4 is the same as condition (4) of Götze and Hipp (1994). It reduces to the standard Cramér condition if $\{X_i : i \geq 1\}$ are i.i.d.

The condition $\gamma_N \geq (\ln(N)/N)^{1/2}$ in Assumption 5 (concerning the initial estimator $\hat{\theta}_{N,0}$) is not restrictive because Assumption 5 typically does not hold for constants γ_N that are smaller than $(\ln(N)/N)^{1/2}$. For some estimators, Assumption 5 may hold only when $\gamma_N > (\ln(N)/N)^{1/2}$, such as $\gamma_N = N^{-1/4}$.

On the other hand, the following lemma shows that for initial estimators in a broad class of extremum estimators Assumption 5 holds with γ_N given by the minimal values

$$\gamma_N = (\ln(N)/N)^{1/2}. \tag{5.1}$$

LEMMA 1. *Suppose $\hat{\theta}_{N,0}$ is an extremum estimator that minimizes a criterion function $J_{N,0}(\theta)$, $J_{N,0}(\theta,\tilde{\theta}_{N,0})$, or $\rho_{N,0}(\theta)$ over Θ, where $J_{N,0}(\theta)$, $J_{N,0}(\theta,\tilde{\theta}_{N,0})$, and $\rho_{N,0}(\theta)$ are defined as in (3.3)–(3.5), respectively, with $g(X_i,\theta)$, Ω, $\Omega_N(\cdot)$, $\tilde{\theta}_N$, $\rho(X_i,\theta)$, and $f(\widetilde{X}_i,\theta)$ replaced by some quantities $g_0(X_i,\theta)$, Ω^0, $\Omega_{N,0}(\cdot)$,*

$\tilde{\theta}_{N,0}$, $\rho_0(X_i,\theta)$, and $f_0(\widetilde{X}_i,\theta)$, respectively. Suppose Assumptions 1–4 hold with the same changes. Suppose $\bar{\theta}_{N,0}$ is a restricted extremum estimator that minimizes one of the preceding criterion functions over Θ_0 but with $\tilde{\theta}_{N,0}$ replaced by $\theta^*_{N,0}$ in $J_{N,0}(\theta,\tilde{\theta}_{N,0})$, where $\theta^*_{N,0}$ minimizes $J_{N,0}(\theta)$ over Θ_0. Then, Assumption 5 holds with $\gamma_N = (\ln(N)/N)^{1/2}$.

Remarks.

(1) Given the result of Lemma 1, the leading case of interest for γ_N in Assumption 5 is $\gamma_N = (\ln(N)/N)^{1/2}$. Much of the discussion that follows focuses on this case. But, we provide results that allow for initial estimators that have a slower rate of convergence, such as those for which $\gamma_N = N^{-c}$ for $c \in (0,\frac{1}{2})$. This occurs with an initial estimator that is defined by minimizing a criterion function over a discrete grid of points or over a set of randomly selected points, rather than all points in the parameter space.

(2) In Lemma 1, the condition on d in Assumption 3 for the initial estimators $\hat{\theta}_{N,0}$ and $\bar{\theta}_{N,0}$ can be weakened to $d \geq 3$. (This holds because the proof of Lemma 1 only relies on Lemma 5 in Section 7 and not on Lemma 8 or 9. See the remark following Lemma 9 in Section 7.)

Next, we provide sufficient conditions for Assumption 6 for the NR, default NR, line-search NR, and GN choices of matrices $Q_{N,j-1}$. Other choices of matrices $Q_{N,j-1}$ are possible.

LEMMA 2. *Suppose Assumptions 1–5 hold for some $a \geq 0$ with $2a$ an integer. Then, Assumption 6 holds with $\psi_{N,j} = 0$ for all $N \geq 1$ and $j = 1,\dots,k$ for the* NR, *default* NR, *and line-search* NR *choices of $Q_{N,j-1}$. In addition, Assumption 6 holds with $\psi_{N,j-1} = \max\{\gamma_N^{2^{j-1}}, (\ln(N)/N)^{1/2}\}$ for the* GN *choice of $Q_{N,j-1}$ for $j = 1,\dots,k$, provided Assumptions 1 and 4 hold with the elements of $\Delta(\widetilde{X}_i,\theta_0)$ (defined in (4.4)) added to $f(\widetilde{X}_i)$, the function $\Delta(\cdot,\cdot)$ satisfies $E(\Delta(\widetilde{X}_i,\theta_0) - (\partial/\partial\theta')g(X_i,\theta_0)) = 0$, $\Delta(\widetilde{X}_i,\theta)$ is continuously differentiable in θ for $\theta \in N_0$, $E\|\Delta(\widetilde{X}_i,\theta_0) - (\partial/\partial\theta')g(X_i,\theta_0))\|^{2a+3} < \infty$, and $E\sup_{\theta\in N_0}\|(\partial/\partial\theta_r)\Delta(\widetilde{X}_i,\theta)\|^{q_2} < \infty$ for all $r = 1,\dots,L_\theta$ for $q_2 = \max\{2a+1, 2\}$.*

Remark. Suppose Assumption 5 holds with $\gamma_N = (\ln(N)/N)^{1/2}$, as Lemma 1 shows occurs for a broad class of extremum estimators. Then, for GN choices of $Q_{N,j-1}$, Assumption 6 holds by Lemma 2 with $\psi_{N,j-1} = (\ln(N)/N)^{1/2}$ for all $N \geq 1$ and $j = 1,\dots,k$.

Assumption 7 is shown to hold under regularity conditions in i.i.d. likelihood contexts by Chandra and Ghosh (1979, Sect. 4). Furthermore, it should be possible to use the same line of argument in the non-i.i.d. likelihood case and in the EW-GMM case making use of the lemmas given in Section 7. However, the arguments for these cases would be quite long and involved. For brevity, we do not provide such results.

6. EQUIVALENCE OF THE HIGHER ORDER ASYMPTOTIC EFFICIENCY OF *k*-STEP AND EXTREMUM STATISTICS

The higher order asymptotic equivalence of the k-step and extremum statistics is established in parts (b)–(d) of Theorem 1, which follows. Part (b) gives conditions under which the CV distances between $(N^{1/2}(\hat{\theta}_{N,k} - \theta_0), T_{N,k}, \mathcal{W}_{N,k}, LM_{N,k})$ and $(N^{1/2}(\hat{\theta}_N - \theta_0), T_N, \mathcal{W}_N, LM_N)$, respectively, are $o(N^{-a})$ for some $a \geq 0$. Part (c) does likewise for $(QLR_{N,k}, J_{N,k})$ and (QLR_N, J_N) when QLR_N is based on the EW-GMM estimator. Part (d) does likewise for $QLR_{N,k}$ and QLR_N when QLR_N is based on the minimum ρ estimator. The conditions required for part (d) are weaker than those for part (c), which, in turn, are weaker than those for part (b).

In part (a) of the theorem, the difference between the k-step estimator and the corresponding extremum estimator is shown to be of greater magnitude than $\mu_{N,k}$ with probability $o(N^{-a})$, where

$$\mu_{N,k} = \begin{cases} \gamma_N^{2^k} & \text{when Assumption 6 holds with } \psi_{N,j-1} = 0 \\ \gamma_N \prod_{j=1}^{k} \psi_{N,j-1} & \text{when Assumption 6 holds with } \psi_{N,j-1} \\ & \quad = \max\{\gamma_N^{2^{j-1}}, (\ln(N)/N)^{1/2}\}. \end{cases} \tag{6.1}$$

Thus, $\mu_{N,k} = \gamma_N^{2^k}$ for NR, default NR, and line-search NR matrices and $\mu_{N,k} = \gamma_N \prod_{j=1}^{k} \psi_{N,j-1}$ for GN matrices.

If Assumption 5 holds with $\gamma_N = (\ln(N)/N)^{1/2}$, as it does for the extremum estimators of Lemma 1, then for GN matrices $\psi_{N,j-1} = (\ln(N)/N)^{1/2} = \gamma_N$ and $\mu_{N,k}$ simplifies to

$$\mu_{N,k} = \gamma_N^{k+1}. \tag{6.2}$$

We see that for the NR procedures the difference, $\mu_{N,k}$ $(= \gamma_N^{2^k})$, decreases very quickly as k increases, whereas for GN procedures the difference, $\mu_{N,k}$ $(= \gamma_N^{k+1})$, decreases more slowly as k increases. Simplified formulae for $\mu_{N,k}$ for GN matrices when $\gamma_N = N^{-c}$ for $c < \frac{1}{2}$ are given in Remark 3 following the theorem.

The key condition in part (b) of Theorem 1 is

$$\mu_{N,k} = o(N^{-(a+1/2)}), \tag{6.3}$$

where $2a$ is a non-negative integer. Given this condition, the CV distances between the k-step and extremum statistics are $o(N^{-a})$, and these statistics have equal asymptotic efficiency to order $s = 2a + 1$.

If Assumption 5 holds with $\gamma_N = (\ln(N)/N)^{1/2}$, as it does for initial estimators that are extremum estimators, and Assumption 6 holds with $\psi_{N,j} = 0$, as it does for NR, default NR, and line-search NR procedures, then (6.3) holds if

$$2^k \geq 2a + 2 \quad \text{or, equivalently,} \quad 2^k \geq s + 1, \tag{6.4}$$

where $2a$ and s are integers. Thus, for $k = 1$, we have $a = 0$ and $s = 1$; for $k = 2$, we have $a = 1$ and $s = 3$; for $k = 3$, we have $a = 3$ and $s = 7$; for $k = 4$, we have $a = 7$ and $s = 15$; etc.

If Assumption 5 holds with $\gamma_N = (\ln(N)/N)^{1/2}$ and Assumption 6 holds with $\psi_{N,j} = (\ln(N)/N)^{1/2}$, as it does for GN procedures under the conditions in Lemma 2, then (6.3) holds if

$$k \geq 2a + 1 \quad \text{or, equivalently,} \quad k \geq s, \tag{6.5}$$

where $2a$ and s are integers. Thus, for $k = 1$, we have $a = 0$ and $s = 1$; for $k = 2$, we have $a = \frac{1}{2}$ and $s = 2$; for $k = 3$, we have $a = 1$ and $s = 3$; for $k = 4$, we have $a = \frac{3}{2}$ and $s = 4$; etc.

Conditions under which (6.3) holds when Assumption 5 holds with $\gamma_N = N^{-c}$ for $c \in (0, \frac{1}{2})$ are given in Remarks 2 and 3 following the theorem.

The conditions used in parts (c) and (d) of the theorem for $QLR_{N,k}$ and $J_{N,k}$ are discussed in Remarks 4 and 5 following the theorem.

The main result of the paper is the following theorem. It holds when $\hat{\theta}_{N,k}$ is the FW-GMM, EW-GMM, or minimum ρ k-step estimator. As previously, $\mathcal{B}_L$ denotes the class of convex sets in R^L.

THEOREM 1. *Suppose Assumptions 1–6 hold for some $a \geq 0$ with $2a$ an integer in parts (a)–(d), which follows. When considering test statistics in parts (a)–(d), the null hypothesis is assumed to hold.*

(a) Then, for some finite constant C_3,

$$P(\|\hat{\theta}_{N,k} - \hat{\theta}_N\| > C_3\,\mu_{N,k}) = o(N^{-a}),$$

$$P(|T_{N,k} - T_N| > C_3 N^{1/2}\mu_{N,k}) = o(N^{-a}),$$

$$P(|\mathcal{W}_{N,k} - \mathcal{W}_N| > C_3 N^{1/2}\mu_{N,k}) = o(N^{-a}),$$

$$P(|LM_{N,k} - LM_N| > C_3 N^{1/2}\mu_{N,k}) = o(N^{-a}),$$

$$P(|QLR_{N,k} - QLR_N| > C_3\, q_N\, \mu_{N,k}) = o(N^{-a})$$

when $\hat{\theta}_N$ and $\bar{\theta}_N$ are EW-GMM *estimators,*

$$P(|QLR_{N,k} - QLR_N| > C_3 N\mu_{N,k}^2) = o(N^{-a})$$

when $\hat{\theta}_N$ and $\bar{\theta}_N$ are minimum ρ estimators, and

$$P(|J_{N,k} - J_N| > C_3\, q_N\, \mu_{N,k}) = o(N^{-a}),$$

where $q_N = \max\{\ln(N), N\mu_{N,k}\}$.

(b) *Suppose* $\mu_{N,k} = o(N^{-(a+1/2)})$. *Then,*

$$\sup_{B\in\mathcal{B}_{L_\theta}} |P(N^{1/2}(\hat{\theta}_{N,k} - \theta_0) \in B) - P(N^{1/2}(\hat{\theta}_N - \theta_0) \in B)| = o(N^{-a}),$$

$$\sup_{B\in\mathcal{B}_1} |P(T_{N,k} \in B) - P(T_N \in B)| = o(N^{-a}) \quad \textit{under } H_0,$$

$$\sup_{B\in\mathcal{B}_1} |P(\mathcal{W}_{N,k} \in B) - P(\mathcal{W}_N \in B)| = o(N^{-a}) \quad \textit{under } H_0,$$

and

$$\sup_{B\in\mathcal{B}_1} |P(LM_{N,k} \in B) - P(LM_N \in B)| = o(N^{-a}) \quad \textit{under } H_0.$$

(c) *Suppose* $\mu_{N,k} = o(N^{-a}q_N^{-1})$. *Suppose Assumption 7 holds when considering the* $QLR_{N,k}$*-statistic. Then, for the* $QLR_{N,k}$*-statistic based on the* EW-GMM *estimator and for the* $J_{N,k}$*-statistic,*

$$\sup_{B\in\mathcal{B}_1} |P(QLR_{N,k} \in B) - P(QLR_N \in B)| = o(N^{-a}) \quad \textit{under } H_0 \quad \textit{and}$$

$$\sup_{B\in\mathcal{B}_1} |P(J_{N,k} \in B) - P(J_N \in B)| = o(N^{-a}) \quad \textit{under } H_0.$$

(d) *Suppose* $\mu_{N,k} = o(N^{-(a+1)/2})$ *and Assumption 7 holds. Then, for the* $QLR_{N,k}$*-statistic based on the minimum* ρ *estimator,*

$$\sup_{B\in\mathcal{B}_1} |P(QLR_{N,k} \in B) - P(QLR_N \in B)| = o(N^{-a}) \quad \textit{under } H_0.$$

Remarks.

1. When $a = 0$, part (a) gives the stochastic differences between the statistics $\hat{\theta}_{N,k}$ and $\hat{\theta}_N$, etc., as in Robinson (1988) (although Robinson, 1988, does not consider test statistics). When $a = 0$ and $\gamma_N = N^{-c}$ for some $c \in (0, \frac{1}{2})$, the results of part (a) for GN procedures are stronger than those in Robinson (1988) because we exploit the fact that Assumption 6 holds with lower bounds $\psi_{N,j}$ that decrease in j in this case, rather than being independent of j. When $a > 0$, part (a) gives stronger results than stochastic difference results. It shows that the difference between k-step and extremum statistics is very small except on sets with very small probabilities. These stronger results are used to establish parts (b)–(d) of the theorem. Parts (b)–(d) show that the cv distances between the distributions of $N^{1/2}(\hat{\theta}_{N,k} - \theta_0)$ and $N^{1/2}(\hat{\theta}_N - \theta_0)$, etc., are $o(N^{-a})$. Parts (b)–(d) establish that the k-step and extremum estimators and test statistics have equal s-order asymptotic efficiency for $s = 2a + 1$.

2. Here we discuss the condition $\mu_{N,k} = o(N^{-(a+1/2)})$ in part (b) of the theorem when Assumption 5 holds with $\gamma_N = N^{-c}$ for some $c \in (0, \frac{1}{2})$ and NR, default NR, or line-search NR procedures are employed. In this case, $\mu_{N,k} = N^{-c2^k}$ and the condition holds if $2^k > (a + \frac{1}{2})/c$. For example, if $c = \frac{1}{4}$, it holds if $2^k > 4a + 2$. If $c = \frac{1}{4}$ and $k = 2$, the condition holds with $a = 0$ and $s = 1$. If $c = \frac{1}{4}$ and $k = 3$, the condition holds with $a = 1$ and $s = 3$.

3. Some calculations show that if Assumption 5 holds with $\gamma_N = N^{-c}$ for $c \in [2^{-b}, 2^{-b+1})$ for some $b \in \{2, 3, \ldots\}$, then for GN matrices $\mu_{N,k}$ satisfies

$$\begin{aligned} \mu_{N,k} &= \gamma_N^{2^k} && \text{for } k < b \\ \mu_{N,k} &= \gamma_N^{2^{b-1}} (\ln(N)/N)^{(k-b+1)/2} && \text{for } k \geq b. \end{aligned} \tag{6.6}$$

In this case, $\psi_{N,j-1} = \gamma_N^{2^{k-1}}$ for $j < b$ and $\psi_{N,j-1} = (\ln(N)/N)^{1/2}$ for $j \geq b$.

Suppose $\gamma_N = N^{-c}$ for $c \in [\frac{1}{4}, \frac{1}{2})$; then $b = 2$, $\mu_{N,k} = N^{-2c}$ for $k = 1$, and $\mu_{N,k} = N^{-2c}(\ln(N)/N)^{(k+1)/2}$ for $k \geq 2$. In this case, the condition $\mu_{N,k} = o(N^{-(a+1/2)})$ in part (b) of the theorem holds for $k = 1$ if $c > a/2 + \frac{1}{4}$, which holds for $a = 0$ and $s = 1$ provided $c > \frac{1}{4}$ and does not hold even for $a = 0$ when $c = \frac{1}{4}$. The condition $\mu_{N,k} = o(N^{-(a+1/2)})$ holds for $k \geq 2$ if $k > 2a - 4c$. For $k = 2$, this condition holds with $a = 1$ and $s = 3$ for $c = \frac{1}{4}$ and with $a = \frac{3}{2}$ and $s = 4$ for $c > \frac{1}{4}$. For $k = 3$, the condition holds with $a = \frac{3}{2}$ and $s = 4$ for $c = \frac{1}{4}$ and with $a = 2$ and $s = 5$ for $c > \frac{1}{4}$.

4. Here we discuss the condition $\mu_{N,k} = o(N^{-a} q_N^{-1})$ in part (c) of the theorem. When Assumption 5 holds with $\gamma_N = (\ln(N)/N)^{1/2}$, as it does for extremum estimators under the conditions of Lemma 1, and Assumption 6 holds with $\psi_{N,j} = 0$, as it does for the NR, default NR, and line-search NR procedures, then $q_N = \ln(N)$ and the condition $\mu_{N,k} = o(N^{-a} q_N^{-1})$ holds provided

$$2^k \geq 2a + 1 \quad \text{or, equivalently,} \quad 2^k \geq s. \tag{6.7}$$

Thus, for $k = 1$, we have $a = \frac{1}{2}$ and $s = 2$; for $k = 2$, we have $a = \frac{3}{2}$ and $s = 4$; for $k = 3$, we have $a = \frac{7}{2}$ and $s = 8$; for $k = 4$, we have $a = \frac{15}{2}$ and $s = 16$; etc.

If Assumption 5 holds with $\gamma_N = (\ln(N)/N)^{1/2}$ and Assumption 6 holds with $\psi_{N,j} = (\ln(N)/N)^{1/2}$, as it does for the GN procedure under the conditions in Lemma 2, then $q_N = \ln(N)$ and the condition $\mu_{N,k} = o(N^{-a} q_N^{-1})$ in part (c) holds if

$$k \geq 2a \quad \text{or, equivalently,} \quad k \geq s - 1, \tag{6.8}$$

where $2a$ and s are integers. Thus, for $k = 1$, we have $a = \frac{1}{2}$ and $s = 2$; for $k = 2$, we have $a = 1$ and $s = 3$; for $k = 3$, we have $a = \frac{3}{2}$ and $s = 4$; for $k = 4$, we have $a = 2$ and $s = 5$; etc.

When Assumption 5 holds with $\gamma_N = N^{-c}$ for $c \in (0, \frac{1}{2})$ and Assumption 6 holds with $\psi_{N,j} = 0$, then the condition $\mu_{N,k} = o(N^{-a} q_N^{-1})$ in part (c) holds provided $2^k > \max\{a/c, (a + 1)/(2c)\}$. For example, if $c = \frac{1}{4}$ and $k = 1$, the condition does not hold even with $a = 0$. If $c = \frac{1}{4}$ and $k = 2$, the condition holds with $a = \frac{1}{2}$ and $s = 2$. If $c = \frac{1}{4}$ and $k = 3$, the condition holds with $a = \frac{3}{2}$ and $s = 4$.

5. Next, we discuss the condition $\mu_{N,k} = o(N^{-(a+1)/2})$ in part (d). When Assumption 5 holds with $\gamma_N = (\ln(N)/N)^{1/2}$ and when Assumption 6 holds with $\psi_{N,j} = 0$, as it does for the NR, default NR, and line-search NR procedures, this condition holds provided

$$2^{k+1} \geq 2a + 3 \quad \text{or, equivalently,} \quad 2^{k+1} \geq s + 2. \tag{6.9}$$

Thus, for $k = 1$, we have $a = \frac{1}{2}$ and $s = 2$; for $k = 2$, we have $a = \frac{5}{2}$ and $s = 6$; for $k = 3$, we have $a = \frac{13}{2}$ and $s = 14$; for $k = 4$, we have $a = \frac{29}{2}$ and $s = 30$; etc.

If Assumption 5 holds with $\gamma_N = (\ln(N)/N)^{1/2}$ and Assumption 6 holds with $\psi_{N,j} = (\ln(N)/N)^{1/2}$ for all $j = 1, \ldots, k$, as it does for GN procedures under the conditions in Lemma 2, then the condition $\mu_{N,k} = o(N^{-(a+1)/2})$ in part (d) holds if

$$2k \geq 2a + 1 \quad \text{or, equivalently,} \quad 2k \geq s, \tag{6.10}$$

where $2a$ and s are integers. Thus, for $k = 1$, we have $a = \frac{1}{2}$ and $s = 2$; for $k = 2$, we have $a = \frac{3}{2}$ and $s = 4$; for $k = 3$, we have $a = \frac{5}{2}$ and $s = 6$; for $k = 4$, we have $a = \frac{7}{2}$ and $s = 8$; etc.

When Assumption 5 holds with $\gamma_N = N^{-c}$ for $c \in (0, \frac{1}{2})$ and Assumption 6 holds with $\psi_{N,j} = 0$, then the condition $\mu_{N,k} = o(N^{-(a+1)/2})$ in part (d) holds provided $2^k > (a+1)/(2c)$. For example, if $c = \frac{1}{4}$ and $k = 1$, the condition does not hold even with $a = 0$. If $c = \frac{1}{4}$ and $k = 2$, the condition holds with $a = \frac{1}{2}$ and $s = 2$. If $c = \frac{1}{4}$ and $k = 3$, the condition holds with $a = \frac{5}{2}$ and $s = 6$.

6. The condition on $\mu_{N,k}$ in part (d) of the theorem is weaker than those in parts (b) and (c). Also, the condition on $\mu_{N,k}$ in part (c) of the theorem is weaker than that in part (b). The reason this occurs is that part (a) of the theorem holds for the statistics considered in parts (c) and (d) with the lower bounds in the probability being $\ln(N)\mu_{N,k}$ and $N\mu_{N,k}^2$, respectively, rather than the larger quantity $N^{1/2}\mu_{N,k}$, which is the lower bound for the statistics considered in part (b).

The reason for these results in part (a) is as follows. Consider the $QLR_{N,k}$-statistic based on the minimum ρ estimator, for which part (a) holds with lower bound $N\mu_{N,k}^2$. We have

$$QLR_{N,k} - QLR_N = 2N(\rho_N(\hat{\theta}_{N,k}) - \rho_N(\hat{\theta}_N)) - 2N(\rho_N(\bar{\theta}_{N,k}) - \rho_N(\bar{\theta}_N)). \tag{6.11}$$

The first and second terms on the right-hand side are quadratic forms in $\hat{\theta}_{N,k} - \hat{\theta}_N$ and $\bar{\theta}_{N,k} - \bar{\theta}_N$, respectively. Hence, $|QLR_{N,k} - QLR_N|$ is of the same order as $N\|\hat{\theta}_{N,k} - \hat{\theta}_N\|^2$ and $N\|\bar{\theta}_{N,k} - \bar{\theta}_N\|^2$. The result of part (a) for $\|\hat{\theta}_{N,k} - \hat{\theta}_N\|$ and $\|\bar{\theta}_{N,k} - \bar{\theta}_N\|$ holds with lower bound $\mu_{N,k}$. Thus, the result of part (a) for $|QLR_{N,k} - QLR_N|$ holds with lower bound $N\mu_{N,k}^2$.

The reason that the first term on the right-hand side of (6.11) is a quadratic form in $\hat{\theta}_{N,k} - \hat{\theta}_N$ is that a two-term Taylor expansion of $\rho_N(\hat{\theta}_{N,k})$ about $\hat{\theta}_N$ gives

$$\rho_N(\hat{\theta}_{N,k}) - \rho_N(\hat{\theta}_N) = \frac{1}{2}(\hat{\theta}_{N,k} - \hat{\theta}_N)' \frac{\partial^2}{\partial\theta\partial\theta'} \rho_N(\theta_N^+)(\hat{\theta}_{N,k} - \hat{\theta}_N), \tag{6.12}$$

where the linear term in $\hat{\theta}_{N,k} - \hat{\theta}_N$ is zero with probability $1 - o(N^{-a})$, because $(\partial/\partial\theta)\rho_N(\hat{\theta}_N) = 0$ by the first-order conditions for $\hat{\theta}_N$. An analogous result holds for $\rho_N(\bar{\theta}_{N,k}) - \rho_N(\bar{\theta}_N)$.

For the $QLR_{N,k}$-statistic based on the EW-GMM estimator and the $J_{N,k}$-statistic, the preceding calculations need to be altered somewhat because of the difference between $(\partial/\partial\theta)J_N(\hat{\theta}_N, \tilde{\theta}_{N,k_1})$ and $(\partial/\partial\theta)J_N(\hat{\theta}_N, \tilde{\theta}_N)$. The former appears in the Taylor expansion (with respect to the first argument) of $J_N(\hat{\theta}_{N,k}, \tilde{\theta}_{N,k_1})$ about $\hat{\theta}_N$, which is analogous to the Taylor expansion of (6.12). But, it is the latter, $(\partial/\partial\theta)J_N(\hat{\theta}_N, \tilde{\theta}_N)$, that equals zero with probability $1 - o(N^{-a})$. Hence, the linear term in the Taylor expansion is not identically zero. In consequence, the lower bounds in part (a) for the $QLR_{N,k}$-statistic based on the EW-GMM estimator and the $J_{N,k}$-statistic are larger than for the $QLR_{N,k}$-statistic based on the minimum ρ estimator but smaller than for the other statistics considered. In turn, this implies that the condition needed in part (c) is stronger than that required in part (d) but weaker than that required in part (b).

7. Results analogous to those given previously for test statistics under the null hypothesis could be established under local alternatives. For brevity, we do not do so.

7. PROOFS

In Section 7.1, we state Lemmas 3–9, which are used in the proofs of Lemmas 1 and 2 and Theorem 1. In Section 7.2, we prove Theorem 1. In Section 7.3, we prove Lemmas 1–9.

Throughout this section, a denotes a constant that satisfies $a \geq 0$ and $2a$ is an integer.

7.1. Lemmas

LEMMA 3. *Suppose Assumption 1 holds.*

(a) Let $m(\cdot)$ be a matrix-valued function that satisfies $Em(\widetilde{X}_i) = 0$ and $E\|m(\widetilde{X}_i)\|^p < \infty$ for $p > 2a$ and $p \geq 2$. Then, for all $\varepsilon > 0$,

$$P\left(\left\|N^{-1}\sum_{i=1}^N m(\widetilde{X}_i)\right\| > \varepsilon\right) = o(N^{-a}).$$

(b) Let $m(\cdot)$ be a matrix-valued function that satisfies $E\|m(\widetilde{X}_i)\|^p < \infty$ for $p > 2a$ and $p \geq 2$. Then, there exists $K < \infty$ such that

$$P\left(\left\|N^{-1}\sum_{i=1}^N m(\widetilde{X}_i)\right\| > K\right) = o(N^{-a}).$$

(c) *Suppose Assumptions 3(c), 3(d), and 4 also hold. Then, for all constants* $C_4 > (2a)^{1/2}$,

$$P\left(\left\|N^{-1/2}\sum_{i=1}^{N}(f(\widetilde{X}_i) - Ef(\widetilde{X}_i))\right\| > C_4 \ln^{1/2}(N)\right) = o(N^{-a}).$$

LEMMA 4. *Suppose Assumptions 1–3 hold. Let* $\bar\theta_{1,N}$ *and* $\bar\theta_{2,N}$ *denote any estimators that satisfy the following condition. For all* $\varepsilon > 0$, $P(\|\bar\theta_{j,N} - \theta_0\| > \varepsilon) = o(N^{-a})$ *for* $j = 1,2$. *Then, for all* $\varepsilon > 0$ *and some* $K < \infty$,

$$P(\|D_N(\bar\theta_{1,N}) - D\| > \varepsilon) = o(N^{-a}),$$

$$P(\|\Omega_N(\bar\theta_{1,N}) - \Omega_0\| > \varepsilon) = o(N^{-a}),$$

$$P\left(\left\|\frac{\partial^2}{\partial\theta\partial\theta'} J_N(\bar\theta_{1,N},\bar\theta_{2,N}) - 2D'\Omega_0 D\right\| > \varepsilon\right) = o(N^{-a}),$$

$$P\left(\left\|\frac{\partial^3}{\partial\theta^3} J_N(\bar\theta_{1,N},\bar\theta_{2,N})\right\| > K\right) = o(N^{-a}),$$

$$P\left(\left\|N^{-1}\sum_{i=1}^{N} g(X_i,\bar\theta_{1,N})\right\| > \varepsilon\right) = o(N^{-a}),$$

and analogous results hold for $(\partial^2/\partial\theta\partial\theta')J_N(\bar\theta_{1,N}) - 2D'\Omega D$ *and* $(\partial^3/\partial\theta^3)J_N(\bar\theta_{1,N})$, *where* $(\partial^3/\partial\theta^3)J_N(\bar\theta_{1,N},\bar\theta_{2,N})$ *denotes a vector containing all of the partial derivatives of order three of* $J_N(\bar\theta_{1,N},\bar\theta_{2,N})$ *with respect to its first argument and likewise for* $(\partial^3/\partial\theta^3)J_N(\bar\theta_{1,N})$.

LEMMA 5. *Suppose Assumptions 1–4 hold. Let* $\hat\theta_N$ *denote the* FW-GMM, EW-GMM, *or minimum* ρ *estimator. Then, for some finite constant* C_5,

$$P(\|\hat\theta_N - \theta_0\| > C_5(\ln(N)/N)^{1/2}) = o(N^{-a}).$$

LEMMA 6. *Let* $\{A_N : N \ge 1\}$ *be a sequence of* $L_A \times 1$ *random vectors with an Edgeworth expansion or asymptotic* χ^2 *expansion with coefficients of order* $O(1)$ *and remainder of order* $o(N^{-a})$. *(That is, in the case of an Edgeworth expansion, there exist polynomials* $\{\pi_i(z) : i = 1,\ldots,2a\}$ *in* z *whose coefficients are* $O(1)$ *such that* $\sup_{B\in\mathcal{B}_{L_A}}|P(A_N \in B) - \int_B(1 + \sum_{i=1}^{2a} N^{-i/2}\pi_i(z))\phi_{\Sigma_N}(z)\,dz| = o(N^{-a})$, *where* $\phi_{\Sigma_N}(z)$ *is the density function of an* $N(0,\Sigma_N)$ *random variable,* Σ_N *has eigenvalues that are bounded away from zero and infinity as* $N \to \infty$, *and* $\mathcal{B}_{L_A}$ *denotes the class of all convex sets in* R^{L_A}. *In the case of an asymptotic* χ^2 *expansion,* $L_A = 1$ *and there exist polynomials* $\{\pi_i(z) : i = 1,\ldots,a\}$ *in* z *whose coefficients are* $O(1)$ *such that* $\sup_{B\in\mathcal{B}_1}|P(A_N \in B) - \int_B(1 + \sum_{i=1}^{[a]} N^{-i}\pi_i(z))f_{\chi^2}(z)\,dz| = o(N^{-a})$, *where* $f_{\chi^2}(z)$ *is the density function of a* χ^2 *random variable.) Let* $\{\xi_{j,N} : N \ge 1\}$ *be a sequence of random vectors with* $P(\|\xi_{j,N}\| > \omega_N) = o(N^{-a})$ *for some constants* $\omega_N = o(N^{-a})$ *for* $j = 1,2$, *where* $\xi_{1,N} \in R^{L_A}$ *and* $\xi_{2,N} \in R$. *Then,*

$$\sup_{B \in \mathcal{B}_{L_A}} |P(A_N + \xi_{1,N} \in B) - P(A_N \in B)| = o(N^{-a}) \quad \textit{and}$$

$$\sup_{B \in \mathcal{B}_1} |P(A_N' A_N + \xi_{2,N} \in B) - P(A_N' A_N \in B)| = o(N^{-a}).$$

For any function $m(\widetilde{X}_i, \theta)$, let $m_N(\theta) = N^{-1} \sum_{i=1}^N m(\widetilde{X}_i, \theta)$.

LEMMA 7. *Suppose Assumption 1 holds, $m(\widetilde{X}_i, \theta)$ is differentiable with respect to θ, and $E \sup_{\theta \in N_0} \|(\partial/\partial\theta) m(\widetilde{X}_1, \theta)\|^p < \infty$ for $p = \max\{2a+1, 2\}$. Let $\bar{\theta}_{1,N}$ and $\bar{\theta}_{2,N}$ be any estimators that satisfy $P(\|\bar{\theta}_{1,N} - \theta_0\| > \varepsilon) = o(N^{-a})$ for all $\varepsilon > 0$ and $P(\|\bar{\theta}_{2,N} - \bar{\theta}_{1,N}\| > \omega_N) = o(N^{-a})$ for some sequence of constants $\{\omega_N : N \geq 1\}$ for which $\omega_N \to 0$. Then, for some finite constant C_6,*

$$P(\|m_N(\bar{\theta}_{2,N}) - m_N(\bar{\theta}_{1,N})\| > C_6 \omega_N) = o(N^{-a}).$$

Let $S_N = N^{-1} \sum_{i=1}^N f(\widetilde{X}_i, \theta_0)$ and $S = ES_N$.
Let $H_N(\theta) = ((\partial/\partial\theta')\eta(\theta)\sigma_N(\theta)((\partial/\partial\theta')\eta(\theta))')^{-1/2} N^{1/2}\eta(\theta)$.

LEMMA 8. *Suppose Assumptions 1–4 hold. Let Δ_N denote $N^{1/2}(\hat{\theta}_N - \theta_0)$, T_N, $H_N(\hat{\theta}_N)$, $U_N(\hat{\theta}_N)$, or $K_N(\hat{\theta}_N)$ (where the statistics may be defined using* FW-GMM, EW-GMM, *or minimum ρ estimators in each case except $K_N(\hat{\theta}_N)$, in which case $\hat{\theta}_N$ is the* EW-GMM *estimator). Let L denote the dimension of Δ_N. For each definition of Δ_N, there is an infinitely differentiable function $G(\cdot)$ with $G(S) = 0$ such that*

$$\sup_{B \in \mathcal{B}_L} |P(\Delta_N \in B) - P(N^{1/2} G(S_N) \in B)| = o(N^{-a}).$$

We now establish Edgeworth expansions for the random vectors $\sigma^{-1/2} \times N^{1/2}(\hat{\theta}_N - \theta_0)$, T_N, $H_N(\hat{\theta}_N)$, $U_N(\hat{\theta}_N)$, and $K_N(\hat{\theta}_N)$. Let $\phi(\cdot)$ denote the density function of a vector of independent standard normal random variables. Let $\mathcal{B}_L$ denote the class of convex Borel measurable sets in R^L.

LEMMA 9. *Suppose Assumptions 1–4 hold. Then, there exist (vector-valued) polynomials $\pi_{\theta i}(\delta)$, $\pi_{Ti}(\delta)$, $\pi_{Hi}(\delta)$, $\pi_{Ui}(\delta)$, and $\pi_{Ki}(\delta)$ in $\delta = \partial/\partial z$ for $i = 1, \ldots, 2a$ such that*

$$\sup_{B \in \mathcal{B}_{L_\theta}} \left| P(\sigma^{-1/2} N^{1/2}(\hat{\theta}_N - \theta_0) \in B) - \int_B \left[1 + \sum_{i=1}^{2a} N^{-i/2} \pi_{\theta i}(\delta) \right] \phi(z)\, dz \right| = o(N^{-a}),$$

$$\sup_{B \in \mathcal{B}_1} \left| P(T_N \in B) - \int_B \left[1 + \sum_{i=1}^{2a} N^{-i/2} \pi_{Ti}(\delta) \right] \phi(z)\, dz \right| = o(N^{-a}),$$

$$\sup_{B \in \mathcal{B}_{L_\eta}} \left| P(H_N(\hat{\theta}_N) \in B) - \int_B \left[1 + \sum_{i=1}^{2a} N^{-i/2} \pi_{Hi}(\delta) \right] \phi(z)\, dz \right| = o(N^{-a}),$$

$$\sup_{B \in \mathcal{B}_{L_\theta}} \left| P(U_N(\hat{\theta}_N) \in B) - \int_B \left[1 + \sum_{i=1}^{2a} N^{-i/2} \pi_{Ui}(\delta) \right] \phi(z)\, dz \right| = o(N^{-a}),$$

and

$$\sup_{B\in\mathcal{B}_{L_g}}\left|P(K_N(\hat{\theta}_N)\in B)-\int_B\left[1+\sum_{i=1}^{2a}N^{-i/2}\pi_{Ki}(\delta)\right]\phi(z)\,dz\right|=o(N^{-a}).$$

Remark. The conditions on d in Assumption 3 are not needed in all of the preceding lemmas. In particular, Lemmas 4 and 5 only use $d=3$.

7.2. Proof of Theorem 1

We establish the first result of part (a) first. To start, suppose $\hat{\theta}_N$ is the FW-GMM estimator. A Taylor expansion about $\hat{\theta}_{N,k-1}$ gives

$$\begin{aligned}
0 &= \frac{\partial}{\partial\theta}J_N(\hat{\theta}_N)\\
&= \frac{\partial}{\partial\theta}J_N(\hat{\theta}_{N,k-1})+\frac{\partial^2}{\partial\theta\partial\theta'}J_N(\hat{\theta}_{N,k-1})(\hat{\theta}_N-\hat{\theta}_{N,k-1})+R_{N,k}\\
&= \frac{\partial}{\partial\theta}J_N(\hat{\theta}_{N,k-1})+Q_{N,k-1}(\hat{\theta}_{N,k}-\hat{\theta}_{N,k-1})+Q_{N,k-1}(\hat{\theta}_N-\hat{\theta}_{N,k})\\
&\quad+\left(\frac{\partial^2}{\partial\theta\partial\theta'}J_N(\hat{\theta}_{N,k-1})-Q_{N,k-1}\right)(\hat{\theta}_N-\hat{\theta}_{N,k-1})+R_{N,k}\\
&= Q_{N,k-1}(\hat{\theta}_N-\hat{\theta}_{N,k})+\left(\frac{\partial^2}{\partial\theta\partial\theta'}J_N(\hat{\theta}_{N,k-1})-Q_{N,k-1}\right)(\hat{\theta}_N-\hat{\theta}_{N,k-1})\\
&\quad+R_{N,k},\quad\text{where}
\end{aligned}$$

$$R_{N,k}=\left[(\hat{\theta}_N-\hat{\theta}_{N,k-1})'\frac{\partial^3}{\partial\theta_u\partial\theta\partial\theta'}J_N(\theta^+_{N,k-1,u})(\hat{\theta}_N-\hat{\theta}_{N,k-1})/2\right]_{L_\theta},\qquad\textbf{(7.1)}$$

$[\xi_u]_{L_\theta}$ denotes an L_θ vector whose uth element is ξ_u, $\theta^+_{N,k-1,u}$ lies between $\hat{\theta}_N$ and $\hat{\theta}_{N,k-1}$, the first equality holds with probability $1-o(N^{-a})$ by Lemma 5, and the fourth equality holds because $(\partial/\partial\theta)J_N(\hat{\theta}_{N,k-1})+Q_{N,k-1}\times(\hat{\theta}_{N,k}-\hat{\theta}_{N,k-1})=0$ by the definition of $\hat{\theta}_{N,k}$. Rearranging (7.1) yields

$$\begin{aligned}
\|\hat{\theta}_{N,k}-\hat{\theta}_N\| &\le \|(Q_{N,k-1})^{-1}R_{N,k}\|\\
&\quad+\left\|(Q_{N,k-1})^{-1}\left(\frac{\partial^2}{\partial\theta\partial\theta'}J_N(\hat{\theta}_{N,k-1})-Q_{N,k-1}\right)(\hat{\theta}_{N,k-1}-\hat{\theta}_N)\right\|\\
&\le \zeta_N(\|\hat{\theta}_{N,k-1}-\hat{\theta}_N\|^2+\psi_{N,j-1}\|\hat{\theta}_{N,k-1}-\hat{\theta}_N\|),\quad\text{where}\\
\zeta_N &= \max_{j=1,\dots,k}\left\{\|(Q_{N,j-1})^{-1}\|\cdot\sum_{u=1}^{L_\theta}\left\|\frac{\partial^3}{\partial\theta_u\partial\theta\partial\theta'}J_N(\theta^+_{N,j-1,u})/2\right\|\right.\\
&\quad\left.+\|(Q_{N,j-1})^{-1}\|\cdot\tilde{\psi}_{N,j-1}\left\|\frac{\partial^2}{\partial\theta\partial\theta'}J_N(\hat{\theta}_{N,j-1})-Q_{N,j-1}\right\|+1\right\},
\end{aligned}\qquad\textbf{(7.2)}$$

where $\tilde{\psi}_{N,j-1}=\psi^{-1}_{N,j-1}$ if $\psi_{N,j-1}>0$ and $\tilde{\psi}_{N,j-1}=0$ if $\psi_{N,j-1}=0$.

For the case where $\psi_{N,j-1} = 0$ for all j, repeated substitution into the right-hand side of the inequality in (7.2) gives the upper bound

$$\zeta_N^{\phi} \|\hat{\theta}_{N,0} - \hat{\theta}_N\|^{2^k} \leq C\zeta_N^{\phi} \|\hat{\theta}_{N,0} - \theta_0\|^{2^k} + C\zeta_N^{\phi} \|\hat{\theta}_N - \theta_0\|^{2^k} \tag{7.3}$$

for some $C < \infty$, where $\phi = \sum_{j=1}^{k} 2^{j-1}$. By Lemma 4 and Assumptions 3(b) and 6, there exists a finite constant K such that $P(\zeta_N > K) = o(N^{-a})$. Combining these results gives

$$\begin{aligned} P(\|\hat{\theta}_{N,k} - \hat{\theta}_N\| > C_3\,\mu_{N,k}) &\leq P(C\zeta_N^{\phi} \|\hat{\theta}_{N,k} - \theta_0\|^{2^k} > C_3\gamma_N^{2^k}/2) \\ &\quad + P(C\zeta_N^{\phi} \|\hat{\theta}_N - \theta_0\|^{2^k} > C_3(\ln(N)/N)^{2^{K-1}}/2) \\ &= o(N^{-a}), \end{aligned} \tag{7.4}$$

where the inequality uses $\mu_{N,k} = \gamma_N^{2^k}$ and $\gamma_N \geq (\ln(N)/N)^{1/2}$ and the equality uses Assumption 5 and Lemma 5.

For the case where Assumption 6 holds with $\psi_{N,j-1} = \max\{\gamma_N^{2^{j-1}}, (\ln(N)/N)^{1/2}\}$, we argue as follows. Let $x_k = \|\hat{\theta}_{N,k} - \hat{\theta}_N\|$. By Assumption 5, Lemma 5, and the triangle inequality, $x_0 \leq C_1\gamma_N + C_5(\ln(N)/N)^{1/2} \leq K_1\gamma_N$ with probability $1 - o(N^{-a})$ for some constant $K_1 < \infty$. As before, $P(\zeta_N \leq K) = 1 - o(N^{-a})$. Hence, using (7.2), $x_k \leq K(x_{k-1}^2 + \psi_{N,k-1}x_{k-1})$ with probability $1 - o(N^{-a})$. Note that $\psi_{N,0} = \gamma_N$ and $\psi_{N,j} \geq \gamma_N\psi_{N,j-1}$. Combining these results, with probability $1 - o(N^{-a})$, we have

$$\begin{aligned} x_1 &\leq Kx_0(x_0 + \psi_{N,0}) \leq KK_1\gamma_N(K_1\gamma_N + \psi_{N,0}) \\ &\leq KK_1(K_1+1)\gamma_N\psi_{N,0} = K'\gamma_N\psi_{N,0} \quad \text{and} \\ x_2 &\leq Kx_1(x_1 + \psi_{N,1}) \leq K(K'\gamma_N\psi_{N,0})(K'\gamma_N\psi_{N,0} + \psi_{N,1}) \\ &\leq KK'(K'+1)\gamma_N\psi_{N,0}\psi_{N,1}. \end{aligned} \tag{7.5}$$

Proceeding recursively, we obtain $x_k \leq C_3\gamma_N \prod_{j=1}^{k} \psi_{N,j-1} = C_3\mu_{N,k}$ for some constant $C_3 < \infty$ with probability $1 - o(N^{-a})$, which is the desired result. Hence, the first result of part (a) of the theorem holds for the FW-GMM estimator.

The proof of the first result of part (a) for the minimum ρ estimator is the same as for the FW-GMM estimator with $J_N(\theta)$ replaced by $\rho_N(\theta)$ throughout.

The proof for the EW-GMM estimator is similar to that given previously for the FW-GMM estimator with $J_N(\theta)$ replaced by $J_N(\theta, \tilde{\theta}_N)$ or $J_N(\theta, \tilde{\theta}_{N,k_1})$ in the appropriate places. However, two additional terms arise on the right-hand side of (7.1) because $J_N(\theta, \tilde{\theta}_{N,k_1}) \neq J_N(\theta, \tilde{\theta}_N)$. These terms are

$$\begin{aligned} M_{1,N} &= \left(\frac{\partial^2}{\partial\theta\partial\theta'} J_N(\hat{\theta}_{N,k-1}, \tilde{\theta}_N) - \frac{\partial^2}{\partial\theta\partial\theta'} J_N(\hat{\theta}_{N,k-1}, \tilde{\theta}_{N,k_1})\right)(\hat{\theta}_N - \hat{\theta}_{N,k-1}) \quad \text{and} \\ M_{2,N} &= \frac{\partial}{\partial\theta} J_N(\hat{\theta}_{N,k-1}, \tilde{\theta}_N) - \frac{\partial}{\partial\theta} J_N(\hat{\theta}_{N,k-1}, \tilde{\theta}_{N,k_1}). \end{aligned} \tag{7.6}$$

These terms can be shown to satisfy the following condition. For some finite constant C,

$$P(\|M_{j,N}\| > C\mu_{N,k}) = o(N^{-a}) \quad \text{for } j = 1,2. \tag{7.7}$$

In consequence, the result of part (a) of the theorem holds for the EW-GMM estimator.

To prove (7.7), we first show that, for some finite constant C_6,

$$P(\|\Omega_N^{-1}(\tilde{\theta}_{N,k_1}) - \Omega_N^{-1}(\tilde{\theta}_N)\| > C_6\,\mu_{N,k}) = o(N^{-a}) \tag{7.8}$$

using Lemma 7 with $m_N(\theta) = \Omega_N^{-1}(\theta)$, $\bar{\theta}_{1,N} = \tilde{\theta}_N$, $\bar{\theta}_{2,N} = \tilde{\theta}_{N,k_1}$, and $\omega_N = C_3\,\mu_{N,k}$. The conditions of Lemma 7 are verified using the result of part (a) of the theorem for the FW-GMM estimator, the assumption that $k_1 \geq k$, and Lemma 5. The proof of (7.7) also uses the first, second, and fifth results of Lemma 4 with $\bar{\theta}_{1,N} = \hat{\theta}_{N,k-1}$, where the condition on $\bar{\theta}_{1,N}$ holds by applying the proof of part (a) of the theorem for the EW-GMM estimator recursively for $k = 1,2,\ldots$. The proof of (7.7) also uses $P(\|\hat{\theta}_N - \hat{\theta}_{N,k-1}\| > K) = o(N^{-a})$ for some $1 \leq K < \infty$, which holds by applying the current proof recursively because $K \geq \mu_{N,k-1}$.

Next, we establish the second result of part (a) of the theorem. Let σ_r denote $(\sigma_N)_{rr}$. Let $\sigma_{k,r}$ denote σ_r with $\hat{\theta}_N$ replaced by $\hat{\theta}_{N,k}$ in all parts of its definition in (3.7). We use the following expression:

$$\begin{aligned} |T_{N,k} - T_N| \leq{}& N^{1/2}\|\hat{\theta}_{N,k} - \hat{\theta}_N\|/\sigma_{k,r}^{1/2} \\ &+ N^{1/2}\|\hat{\theta}_N - \theta_0\|\cdot|\sigma_{k,r}^{1/2} - \sigma_r^{1/2}|/(\sigma_{k,r}\sigma_r)^{1/2}. \end{aligned} \tag{7.9}$$

By (7.9), the second result of part (a) is implied by the first result plus the following condition. There exist constants $C < \infty$, $K < \infty$, and $\delta > 0$ such that

$$P(|\sigma_{k,r}^{1/2} - \sigma_r^{1/2}| > C\mu_{N,k}) = o(N^{-a}), \tag{7.10}$$

$$P(\|\hat{\theta}_N - \theta_0\| > K) = o(N^{-a}), \tag{7.11}$$

$$P(\sigma_{k,r} < \delta) = o(N^{-a}), \quad \text{and} \tag{7.12}$$

$$P(\sigma_r < \delta) = o(N^{-a}). \tag{7.13}$$

Equation (7.11) holds by Lemma 5. Equations (7.12) and (7.13) hold by Lemma 5, the first result of part (a) of the theorem, and the first and second results of Lemma 4.

By a mean value expansion, (7.10) is implied by (7.12), (7.13), and

$$P(|\sigma_{k,r} - \sigma_r| > \tilde{C}\mu_{N,k}) = o(N^{-a}) \tag{7.14}$$

for some finite constant $\tilde{C}$. Equation (7.14) is implied by

$$P(\|D_N(\hat{\theta}_{N,k}) - D_N(\hat{\theta}_N)\| > C'\mu_{N,k}) = o(N^{-a}) \quad \text{and}$$
$$P(\|\Omega_N^{-1}(\hat{\theta}_{N,k}) - \Omega_N^{-1}(\hat{\theta}_N)\| > C'\mu_{N,k}) = o(N^{-a}) \qquad \textbf{(7.15)}$$

for some finite constant C'. These results hold by Lemma 7 with $\omega_N = C_3\mu_{N,k}$, $\bar{\theta}_{1,N} = \hat{\theta}_N$, and $\bar{\theta}_{2,N} = \hat{\theta}_{N,k}$, using Lemma 5, the first result of part (a) of the theorem, and Assumption 3.

We now prove the third result of part (a). Let $H_N = H_N(\hat{\theta}_N)$ and $H_{N,k} = H_N(\hat{\theta}_{N,k})$. We have

$$\begin{aligned}|\mathcal{W}_{N,k} - \mathcal{W}_N| &= |(H_{N,k} - H_N)'H_{N,k} + H_N'(H_{N,k} - H_N)| \\ &\leq \|H_{N,k} - H_N\|(\|H_{N,k}\| + \|H_N\|). \qquad \textbf{(7.16)}\end{aligned}$$

Hence, it suffices to show that, for some finite constant C,

$$P(\|H_{N,k} - H_N\| > CN^{1/2}\mu_{N,k}) = o(N^{-a}) \quad \text{and}$$
$$P(\|H_N\| > M) = o(N^{-a}) \quad \text{for some } M < \infty. \qquad \textbf{(7.17)}$$

The second result of (7.17) holds by Lemma 9 by appropriate choice of the set B. The first result of (7.17) is implied by the matrix version of (7.14), mean value expansions of $\eta(\hat{\theta}_{N,k})$ and $(\partial/\partial\theta)\eta(\hat{\theta}_{N,k})$ about $\hat{\theta}_N$, and the first result of part (a) of the theorem.

The proof of the fourth result of part (a) is analogous to that of the third result with $H_N(\theta)$ replaced by $U_N(\theta)$.

To prove the sixth result of part (a), a Taylor expansion of $\rho_N(\hat{\theta}_{N,k})$ about $\hat{\theta}_N$ yields

$$N(\rho_N(\hat{\theta}_{N,k}) - \rho_N(\hat{\theta}_N)) = N(\hat{\theta}_{N,k} - \hat{\theta}_N)' \frac{\partial^2}{\partial\theta\partial\theta'}\rho_N(\theta_N^+)(\hat{\theta}_{N,k} - \hat{\theta}_N)/2 \qquad \textbf{(7.18)}$$

with probability $1 - o(N^{-a})$, where θ_N^+ lies between $\hat{\theta}_{N,k}$ and $\hat{\theta}_N$. The linear term in $\hat{\theta}_{N,k} - \hat{\theta}_N$ in the Taylor expansion is zero because $(\partial/\partial\theta)\rho_N(\hat{\theta}_N) = 0$ with probability $1 - o(N^{-a})$ by the first-order conditions for minimization of $\rho_N(\theta)$ over Θ using Lemma 5 and Assumption 2(a). By (7.18), part (a) of the theorem for $\|\hat{\theta}_{N,k} - \hat{\theta}_N\|$, and the first result of Lemma 4, we obtain

$$\begin{aligned}P(N|\rho_N(\hat{\theta}_{N,k}) - \rho_N(\hat{\theta}_N)| > CN\mu_{N,k}^2) &\leq P(\|\hat{\theta}_{N,k} - \hat{\theta}_N\|^2 > C'\mu_{N,k}^2) + o(N^{-a}) \\ &= o(N^{-a}) \qquad \textbf{(7.19)}\end{aligned}$$

for some finite constants C and C'.

By an analogous argument, (7.19) also holds with $\hat{\theta}_{N,k}$ and $\hat{\theta}_N$ replaced by $\bar{\theta}_{N,k}$ and $\bar{\theta}_N$, respectively, using the first-order conditions for $\bar{\tau}_N$. Because $QLR_{N,k} - QLR_N = 2N(\rho_N(\bar{\theta}_{N,k}) - \rho_N(\bar{\theta}_N)) - 2N(\rho_N(\hat{\theta}_{N,k}) - \rho_N(\hat{\theta}_N))$, this result and (7.19) imply the sixth result of part (a).

Next, we prove the seventh result of part (a). By the triangle inequality, we have

$$|J_{N,k} - J_N| \le N|J_N(\hat{\theta}_{N,k}, \tilde{\theta}_{N,k_1}) - J_N(\hat{\theta}_{N,k}, \tilde{\theta}_N)| \\ + N|J_N(\hat{\theta}_{N,k}, \tilde{\theta}_N) - J_N(\hat{\theta}_N, \tilde{\theta}_N)|. \tag{7.20}$$

To bound the second summand on the right-hand side of (7.20), a Taylor expansion of $J_N(\hat{\theta}_{N,k}, \tilde{\theta}_N)$ about $\hat{\theta}_N$ yields

$$J_N(\hat{\theta}_{N,k}, \tilde{\theta}_N) - J_N(\hat{\theta}_N, \tilde{\theta}_N) \\ = N(\hat{\theta}_{N,k} - \hat{\theta}_N)' \frac{\partial^2}{\partial\theta\partial\theta'} J_N(\theta_N^+, \tilde{\theta}_N)(\hat{\theta}_{N,k} - \hat{\theta}_N)/2 \tag{7.21}$$

with probability $1 - o(N^{-a})$, where θ_N^+ lies between $\hat{\theta}_{N,k}$ and $\hat{\theta}_N$ and the derivatives here and in the subsequent discussion are taken with respect to the first argument of $J_N(\cdot,\cdot)$. The linear term in $\hat{\theta}_{N,k} - \hat{\theta}_N$ in the Taylor expansion is zero because $(\partial/\partial\theta)J_N(\hat{\theta}_N, \tilde{\theta}_N) = 0$ with probability $1 - o(N^{-a})$ by the first-order conditions for minimization of $J_N(\theta, \tilde{\theta}_N)$ over Θ using Lemma 5 and Assumption 2(a).

By (7.21), part (a) of the theorem for $\|\hat{\theta}_{N,k} - \hat{\theta}_N\|$, and the third result of Lemma 4, we obtain

$$P(N|J_N(\hat{\theta}_{N,k}, \tilde{\theta}_N) - J_N(\hat{\theta}_N, \tilde{\theta}_N)| > CN\mu_{N,k}^2) \\ \le P(\|\hat{\theta}_{N,k} - \hat{\theta}_N\|^2 > C'\mu_{N,k}^2) + o(N^{-a}) \\ = o(N^{-a}) \tag{7.22}$$

for some finite constants C and C'.

The first summand on the right-hand side of (7.20) is

$$B_{N,k} = \left| N^{-1/2} \sum_{i=1}^N g(X_i, \hat{\theta}_{N,k})' [\Omega_N(\tilde{\theta}_{N,k_1}) - \Omega_N(\tilde{\theta}_N)] N^{-1/2} \sum_{i=1}^N g(X_i, \hat{\theta}_{N,k}) \right|. \tag{7.23}$$

The term in square brackets satisfies

$$P(\|\Omega_N(\tilde{\theta}_{N,k_1}) - \Omega_N(\tilde{\theta}_N)\| > C\mu_{N,k}) = o(N^{-a}) \tag{7.24}$$

for some finite constant C, by (7.8), the second result of Lemma 4, and the nonsingularity of Ω_0. By a mean value expansion about θ_0,

$$N^{-1/2} \sum_{i=1}^N g(X_i, \hat{\theta}_{N,k}) = N^{-1/2} \sum_{i=1}^N g(X_i, \theta_0) + N^{-1} \sum_{i=1}^N \frac{\partial}{\partial\theta'} g(X_i, \theta_N^+) \\ \times [N^{1/2}(\hat{\theta}_{N,k} - \hat{\theta}_N) + N^{1/2}(\hat{\theta}_N - \theta_0)], \tag{7.25}$$

where θ_N^+ lies between $\hat{\theta}_{N,k}$ and θ_0. The terms on the right-hand side of (7.25) satisfy

$$P\left(\left\|N^{-1/2}\sum_{i=1}^{N} g(X_i,\theta_0)\right\| > C_4 \ln^{1/2}(N)\right) = o(N^{-a}), \tag{7.26}$$

$$P(N^{1/2}\|\hat{\theta}_{N,k}-\hat{\theta}_N\| > C_3 N^{1/2}\mu_{N,k}) = o(N^{-a}), \tag{7.27}$$

$$P(N^{1/2}\|\hat{\theta}_N-\theta_0\| > C_5 \ln^{1/2}(N)) = o(N^{-a}), \quad \text{and} \tag{7.28}$$

$$P\left(\left\|N^{-1}\sum_{i=1}^{N}\frac{\partial}{\partial\theta'} g(X_i,\theta_N^+)\right\| > K\right) = o(N^{-a}) \quad \text{for some } K<\infty, \tag{7.29}$$

where (7.26) holds by Lemma 3(c), (7.27) holds by part (a) of the theorem, (7.28) holds by Lemma 5, and (7.29) holds by the first result of Lemma 4.

Combining (7.25)–(7.29) gives

$$P\left(\left\|N^{-1/2}\sum_{i=1}^{N} g(X_i,\hat{\theta}_{N,k})\right\| > C\max\{\ln^{1/2}(N), N^{1/2}\mu_{N,k}\}\right) = o(N^{-a}) \tag{7.30}$$

for some finite constant C. Combining (7.23), (7.24), and (7.30) gives

$$P(B_{N,k} > C\max\{\ln(N), N\mu_{N,k}^2\}\mu_{N,k}) = o(N^{-a}) \tag{7.31}$$

for some finite constant C. Combining (7.20), (7.22), and (7.31) and noting that $\max\{N\mu_{N,k}^2, \ln(N)\mu_{N,k}, N\mu_{N,k}^3\} = q_N\mu_{N,k}$ gives the seventh result of part (a) of the theorem.

To establish the fifth result of part (a), we write $QLR_{N,k}$ and QLR_N as $J_{N,k} - \bar{J}_{N,k}$ and $J_N - \bar{J}_N$, respectively, where $\bar{J}_{N,k}$ and $\bar{J}_N$ denote the k-step J-statistic and the J-statistic both based on $\bar{\tau}_N$, rather than $\hat{\theta}_N$. The seventh result of part (a) (for the $J_{N,k}$-statistic) and an analogous result for the $\bar{J}_{N,k}$-statistic (which holds by applying the seventh result of part (a) to the criterion function $J_N((\tau',0')'))$ gives the fifth result of part (a).

To establish part (b) of the theorem, we apply Lemma 6 four times with $\omega_N = C_3 N^{1/2}\mu_{N,k}$ and with $(A_N, \xi_{j,N})$ equal to $(N^{1/2}(\hat{\theta}_N - \theta_0)$, $N^{1/2}(\hat{\theta}_{N,k} - \hat{\theta}_N))$, $(T_N,\ T_{N,k} - T_N)$, $(H_N(\hat{\theta}_N),\ \mathcal{W}_{N,k} - \mathcal{W}_N)$, and $(U_N(\hat{\theta}_N),\ LM_{N,k} - LM_N)$. In the first two cases, we use the first result of Lemma 6. In the third and fourth cases, we use the second result of Lemma 6. By the assumption that $\mu_{N,k} = o(N^{-(a+1/2)})$, we have $\omega_N = o(N^{-a})$, as required by Lemma 6. The condition of Lemma 6 on $\xi_{j,N}$ holds by part (a) of the theorem. As required by Lemma 6, the random vectors $\sigma^{-1/2}N^{-1/2}(\hat{\theta}_N - \theta_0)$, T_N, $H_N(\hat{\theta}_N)$, and $U_N(\hat{\theta}_N)$ have Edgeworth expansions with remainder $o(N^{-a})$ by Lemma 9.

To establish part (c) of the theorem for the $J_{N,k}$-statistic, we apply Lemma 6 with $\omega_N = C_3 q_N\mu_{N,k}$ and with $(A_N, \xi_{2,N})$ equal to $(K_N(\hat{\theta}_N), J_{N,k} - J_N)$. By the assumption that $\mu_{N,k} = o(N^{-a}q_N^{-1})$, we have $\omega_N = o(N^{-a})$, as required by Lemma 6. The condition of Lemma 6 on $\xi_{2,N}$ holds by part (a) of the theorem.

The random vector $K_N(\hat{\theta}_N)$ has an Edgeworth expansion with remainder $o(N^{-a})$ by Lemma 9.

To establish part (c) of the theorem for the $QLR_{N,k}$-statistic based on the EW-GMM estimator, we apply Lemma 6 with $\omega_N = C_3 q_N \mu_{N,k}$ and with $(A_N, \xi_{1,N})$ equal to $(QLR_N,\ QLR_{N,k} - QLR_N)$. By the assumption that $\mu_{N,k} = o(N^{-a} q_N^{-1})$, we have $\omega_N = o(N^{-a})$. The condition of Lemma 6 on $\xi_{1,N}$ holds by part (a) of the theorem. The random vector QLR_N has an asymptotic χ^2 expansion with remainder $o(N^{-a})$ by Assumption 7.

To establish part (d) of the theorem, we apply Lemma 6 with $\omega_N = C_3 N \mu_{N,k}^2$ and with $(A_N, \xi_{1,N})$ equal to $(QLR_N,\ QLR_{N,k} - QLR_N)$. By the assumption that $\mu_{N,k} = o(N^{-(a+1)/2})$, we have $\omega_N = o(N^{-a})$. The condition of Lemma 6 on $\xi_{1,N}$ holds by part (a) of the theorem. The random variable QLR_N has an asymptotic χ^2 expansion with remainder $o(N^{-a})$ by Assumption 7. ■

7.3. Proofs of Lemmas

7.3.1. Proof of Lemma 1 The result holds for $\hat{\theta}_N$ by Lemma 5 with $g(X_i, \theta)$, etc., changed as stated in Lemma 1. The result holds for $\bar{\tau}_N$ by Lemma 5 with θ replaced by τ, with the same changes to $g(X_i, \theta)$, etc., as before and with Θ replaced by $\{\tau : \theta = (\tau', 0')' \in \Theta\}$ using Assumption 3(f) to ensure that the true value τ_0 lies in the interior of the latter set. The result of the lemma for $\bar{\tau}_N$ implies that the result holds for $\bar{\theta}_N$. ■

7.3.2. Proof of Lemma 2 The NR result of the lemma holds by definition of $Q_{N,j-1}^{NR}$. We now establish the default NR result of the lemma. Let $\hat{\theta}_{N,j}$ denote the NR j-step FW-GMM estimator for $j = 1, \ldots, k$. For the FW-GMM estimator, it suffices to show that

$$P(J_N(\hat{\theta}_{N,j}) - J_N(\hat{\theta}_{N,j-1}) > 0) = o(N^{-a}), \tag{7.32}$$

for all $j = 1, \ldots, k$, because this implies that $P(Q_{N,j-1}^{D} \neq Q_{N,j-1}^{NR}$ for some $j = 1, \ldots, k) = o(N^{-a})$. When $\hat{\theta}_{N,j} \neq \hat{\theta}_{N,j-1}$, a Taylor expansion of $J_N(\hat{\theta}_{N,j})$ about $\hat{\theta}_{N,j-1}$ gives

$$\begin{aligned}
&J_N(\hat{\theta}_{N,j}) - J_N(\hat{\theta}_{N,j-1}) \\
&\quad = \frac{\partial}{\partial \theta'} J_N(\hat{\theta}_{N,j-1}) \zeta_{N,j} \phi_{N,j} \\
&\qquad + \frac{1}{2} \zeta_{N,j}' \frac{\partial^2}{\partial \theta \partial \theta'} J_N(\hat{\theta}_{N,j-1}) \zeta_{N,j} \phi_{N,j}^2 + \Gamma_{N,j} \phi_{N,j}^3 \\
&\quad = -\frac{1}{2} \zeta_{N,j}' \frac{\partial^2}{\partial \theta \partial \theta'} J_N(\hat{\theta}_{N,j-1}) \zeta_{N,j} \phi_{N,j}^2 + \Gamma_{N,j} \phi_{N,j}^3, \quad \text{where}
\end{aligned}$$

$$\begin{aligned}
\Gamma_{N,j} &= \frac{1}{6} \sum_{r=1}^{L_\theta} \zeta_{N,j,r} \zeta_{N,j}' \frac{\partial^3}{\partial \theta_r \partial \theta \partial \theta'} J_N(\theta_{N,j-1}^+) \zeta_{N,j}, \\
\zeta_{N,j} &= (\hat{\theta}_{N,j} - \hat{\theta}_{N,j-1}) / \|\hat{\theta}_{N,j} - \hat{\theta}_{N,j-1}\|, \qquad \phi_{N,j} = \|\hat{\theta}_{N,j} - \hat{\theta}_{N,j-1}\|,
\end{aligned} \tag{7.33}$$

where $\zeta_{N,j,r}$ denotes the rth element of $\zeta_{N,j}$ and $\theta^+_{N,j-1}$ lies between $\hat{\theta}_{N,j}$ and $\hat{\theta}_{N,j-1}$. The second equality holds by the definition of $\hat{\theta}_{N,j}$. Using (7.33), the left-hand side of (7.32) is less than or equal to

$$P\left(-\lambda_{\min}\left(\frac{\partial^2}{\partial\theta\partial\theta'} J_N(\hat{\theta}_{N,j-1})\right)\Big/2 + \Gamma_{N,j}\phi_{N,j} > 0\right), \tag{7.34}$$

where $\lambda_{\min}(A)$ denotes the minimum eigenvalue of the matrix A. The latter is $o(N^{-a})$, because for $\delta = \lambda_{\min}(D'\Omega D)/2 > 0$,

$$P\left(\lambda_{\min}\left(\frac{\partial^2}{\partial\theta\partial\theta'} J_N(\hat{\theta}_{N,j-1})\right) < \delta\right) = o(N^{-a}),$$

$$P(|\Gamma_{N,j}| > K) = o(N^{-a}) \quad \text{for some } K < \infty, \quad \text{and}$$

$$P(\phi_{N,j} > \varepsilon) = o(N^{-a}), \tag{7.35}$$

where the first result holds by the third result of Lemma 4 for the FW-GMM estimator with $\bar{\theta}_{1,N} = \hat{\theta}_{N,j-1}$ and Assumption 3(b), the second holds by the fourth result of Lemma 4, and the third holds by two applications of part (a) of Theorem 1 for the NR FW-GMM estimator—one with $k = j-1$ and one with $k = j$. This completes the proof for the FW-GMM estimator. The proofs for the EW-GMM and minimum ρ estimators are analogous.

We now establish the line-search NR result of the lemma. We consider the FW-GMM estimator first. Let $\hat{\theta}_{N,j}$ be the NR j-step estimator:

$$\hat{\theta}_{N,j} = \hat{\theta}_{N,j-1} - \varphi_{N,j-1}\pi_{N,j-1}, \quad \text{where}$$

$$\varphi_{N,j-1} = \left\|(Q^{NR}_{N,j-1})^{-1}\frac{\partial}{\partial\theta} J_N(\hat{\theta}_{N,j-1})\right\| \quad \text{and}$$

$$\pi_{N,j-1} = (Q^{NR}_{N,j-1})^{-1}\frac{\partial}{\partial\theta} J_N(\hat{\theta}_{N,j-1})/\varphi_{N,j-1}. \tag{7.36}$$

Let

$$\hat{\theta}^{\alpha}_{N,j} = \hat{\theta}_{N,j-1} - \alpha(Q^{NR}_{N,j-1})^{-1}\frac{\partial}{\partial\theta} J_N(\hat{\theta}_{N,j-1}) = \hat{\theta}_{N,j} + (1-\alpha)\varphi_{N,j-1}\pi_{N,j-1}. \tag{7.37}$$

It suffices to show that

$$P(\inf_{\alpha\in A,\alpha\neq 1} J_N(\hat{\theta}^{\alpha}_{N,j}) - J_N(\hat{\theta}_{N,j}) < 0) = o(N^{-a}) \tag{7.38}$$

for all $j = 1,\ldots,k$, because this implies that $P(Q^{LS}_{N,j-1} \neq Q^{NR}_{N,j-1}$ for some $j = 1,\ldots,k) = o(N^{-a})$.

A Taylor expansion of $J_N(\hat{\theta}^{\alpha}_{N,j})$ about $\hat{\theta}_{N,j}$ gives

$$\begin{aligned}
&J_N(\hat{\theta}^{\alpha}_{N,j}) - J_N(\hat{\theta}_{N,j}) \\
&\quad = (1-\alpha)\varphi_{N,j-1}\pi'_{N,j-1}\frac{\partial}{\partial\theta}J_N(\hat{\theta}_{N,j}) \\
&\qquad + \frac{1}{2}(1-\alpha)^2\varphi^2_{N,j-1}\pi'_{N,j-1}\frac{\partial^2}{\partial\theta\partial\theta'}J_N(\hat{\theta}_{N,j})\pi_{N,j-1} \\
&\qquad + \frac{1}{6}(1-\alpha)^3\varphi^3_{N,j-1}\sum_{u=1}^{L_\theta}\pi_{N,j-1,r}\pi'_{N,j-1}\frac{\partial^3}{\partial\theta_u\partial\theta\partial\theta'}J_N(\theta^{+}_{N,j})\pi_{N,j-1},
\end{aligned} \tag{7.39}$$

where $\theta^{+}_{N,j}$ lies between $\hat{\theta}^{\alpha}_{N,j}$ and $\hat{\theta}_{N,j}$ and $\pi_{N,j-1,u}$ denotes the uth element of $\pi_{N,j-1}$.

Element by element Taylor expansions of $(\partial/\partial\theta)J_N(\hat{\theta}_{N,j})$ about $\hat{\theta}_{N,j-1}$ give

$$\begin{aligned}
\frac{\partial}{\partial\theta}J_N(\hat{\theta}_{N,j}) &= \frac{\partial}{\partial\theta}J_N(\hat{\theta}_{N,j-1}) + \frac{\partial^2}{\partial\theta\partial\theta'}J_N(\hat{\theta}_{N,j-1})(\hat{\theta}_{N,j}-\hat{\theta}_{N,j-1}) \\
&\qquad + \frac{1}{2}\left[(\hat{\theta}_{N,j}-\hat{\theta}_{N,j-1})'\frac{\partial^3}{\partial\theta_r\partial\theta\partial\theta'}J_N(\theta^{++}_{N,j-1,r})(\hat{\theta}_{N,j}-\hat{\theta}_{N,j-1})\right]_{L_\theta} \\
&= 0 + \frac{1}{2}\varphi^2_{N,j-1}\left[\pi'_{N,j-1}\frac{\partial^3}{\partial\theta_r\partial\theta\partial\theta'}J_N(\theta^{++}_{N,j-1,r})\pi_{N,j-1}\right]_{L_\theta},
\end{aligned} \tag{7.40}$$

where $\theta^{++}_{N,j-1,r}$ lies between $\hat{\theta}_{N,j}$ and $\hat{\theta}_{N,j-1}$ and the second equality holds using the definition of $\hat{\theta}_{N,j}$.

The following properties hold. For $\delta = \lambda_{\min}(D'\Omega D) > 0$ and all $\varepsilon > 0$,

$$\begin{aligned}
&P\left(\lambda_{\min}\left(\frac{\partial^2}{\partial\theta\partial\theta'}J_N(\hat{\theta}_{N,j-1})\right) < \delta\right) = o(N^{-a}), \\
&\qquad P\left(\left\|\frac{\partial^3}{\partial\theta^3}J_N(\theta^{++}_{N,j-1})\right\| > K\right) = o(N^{-a}) \quad \text{for some } K < \infty, \quad \text{and} \\
&\qquad\qquad P(\varphi_{N,j} > \varepsilon) = o(N^{-a})
\end{aligned} \tag{7.41}$$

for $j = 1,\dots,k$, where the first result of (7.41) holds by the third result of Lemma 4 with $\bar{\theta}_{1,N} = \hat{\theta}_{N,j-1}$ and Assumption 3(b), the second holds by the fourth result of Lemma 4 with $\bar{\theta}_{1,N} = \theta^{++}_{N,j-1}$, and the third holds by the third result of Lemma 4 with $\bar{\theta}_{1,N} = \hat{\theta}_{N,j-1}$ and Assumption 3(b) to ensure that

$(Q_{N,j}^{NR})^{-1}$ is well behaved and by a mean value expansion of $(\partial/\partial\theta)J_N(\hat{\theta}_{N,j-1})$ about $\hat{\theta}_N$, application of part (a) of Theorem 1 with $k = j - 1$, and the first result of Lemma 4. The second result of (7.41) also holds with $\theta_{N,j-1}^{++}$ replaced by $\theta_{N,j-1}^{+}$.

Substituting (7.40) into the right-hand side of (7.39), dividing (7.39) by $\varphi_{N,j-1}^2$ (when $\varphi_{N,j-1} > 0$), and applying (7.41) yields the resultant first and third terms on the right-hand side of (7.39) to have norm greater than $\varepsilon > 0$ with probability $o(N^{-a})$ and the second term to be strictly positive with probability $1 - o(N^{-a})$ (uniformly over $\alpha \in A$ with $\alpha \neq 1$), which gives (7.38). This completes the proof for the FW-GMM estimator. The proofs for the EW-GMM and minimum ρ estimators are analogous.

Last, we establish the GN result of the lemma. Let $\alpha_N = (\ln(N)/N)^{1/2}$ and $\bar{\psi}_{N,j-1} = \mu_{N,j-1} \vee \alpha_N$ for $j = 1, \ldots, k$, where $b \vee c = \max\{b, c\}$. We have $\bar{\psi}_{N,j-1} = \psi_{N,j-1}$ for $j = 1, \ldots, k$. This holds because

$$\mu_{N,j} = \gamma_N \prod_{\ell=1}^{j} (\gamma_N^{2^{\ell-1}} \vee \alpha_n) \quad \text{for } j = 1, \ldots, k \quad \text{and}$$

$$\begin{aligned}\bar{\psi}_{N,j-1} &= \left(\gamma_N \prod_{\ell=1}^{j-1} (\gamma_N^{2^{\ell-1}} \vee \alpha_n)\right) \vee \alpha_N \\ &= \left(\gamma_N \prod_{\ell=1}^{j-1} \gamma_N^{2^{\ell-1}}\right) \vee \alpha_N \\ &= \gamma_N^{2^{j-1}} \vee \alpha_N = \psi_{N,j-1},\end{aligned} \tag{7.42}$$

where the second equality for $\bar{\psi}_{N,j-1}$ uses $\gamma_N^{j_1} \alpha_N^{j_2} \leq \alpha_N$ for all $j_1 \geq 0$ and $j_2 \geq 1$ because $\gamma_N \leq 1$ and $\alpha_N \leq 1$.

Given that $\bar{\psi}_{N,j-1} = \psi_{N,j-1}$, for the minimum ρ estimator, it suffices to show that

$$P\left(\left\|N^{-1} \sum_{i=1}^{N} \left(\Delta(\widetilde{X}_i, \hat{\theta}_{N,j-1}) - \frac{\partial}{\partial\theta'} g(X_i, \hat{\theta}_{N,j-1})\right)\right\| > C_2 \bar{\psi}_{N,j-1}\right) = o(N^{-a}). \tag{7.43}$$

For the FW-GMM estimator, we also need to show that

$$\begin{aligned}&P\left(\left\|N^{-1} \sum_{i=1}^{N} g(X_i, \hat{\theta}_{N,j-1})' \Omega N^{-1} \sum_{i=1}^{N} \frac{\partial^2}{\partial\theta_u \partial\theta'} g(X_i, \hat{\theta}_{N,j-1})\right\| > C\bar{\psi}_{N,j-1}\right) \\ &\quad = o(N^{-a})\end{aligned} \tag{7.44}$$

for $u = 1,\ldots,L_g$, for some finite constant C. For the EW-GMM estimator, (7.44) must hold with Ω replaced by $\Omega_N(\tilde{\theta}_{N,k_1})$.

First, we establish (7.43). By mean value expansions about θ_0 and the triangle inequality,

$$\begin{aligned}
&\left\| N^{-1}\sum_{i=1}^{N}\left(\Delta(\widetilde{X}_i,\hat{\theta}_{N,j-1}) - \frac{\partial}{\partial\theta'}\,g(X_i,\hat{\theta}_{N,j-1})\right)\right\| \\
&\quad\le \left\| N^{-1}\sum_{i=1}^{N}\left(\Delta(\widetilde{X}_i,\theta_0) - \frac{\partial}{\partial\theta'}\,g(X_i,\theta_0)\right)\right\| \\
&\qquad + N^{-1}\sum_{i=1}^{N}\sup_{\theta\in N_0,\,u\le L_\theta}\left\|\frac{\partial}{\partial\theta_u}\,\Delta(\widetilde{X}_i,\theta) - \frac{\partial^2}{\partial\theta_u\partial\theta'}\,g(X_i,\theta)\right\|\cdot\|\hat{\theta}_{N,j-1}-\theta_0\|.
\end{aligned} \tag{7.45}$$

In addition, $\|\hat{\theta}_{N,j-1}-\theta_0\| \le \|\hat{\theta}_{N,j-1}-\hat{\theta}_N\| + \|\hat{\theta}_N-\theta_0\|$. Hence, it suffices to show that

$$\begin{aligned}
&\text{(i)}\ \ P\left(\left\| N^{-1}\sum_{i=1}^{N}\left(\Delta(\widetilde{X}_i,\theta_0) - \frac{\partial}{\partial\theta'}\,g(X_i,\theta_0)\right)\right\| > C_4(\ln(N)/N)^{1/2}\right) = o(N^{-a}),\\
&\text{(ii)}\ \ P\left(N^{-1}\sum_{i=1}^{N}\sup_{\theta\in N_0,\,u\le L_\theta}\left\|\frac{\partial}{\partial\theta_u}\,\Delta(\widetilde{X}_i,\theta) - \frac{\partial^2}{\partial\theta_u\partial\theta'}\,g(X_i,\theta)\right\| > K\right) = o(N^{-a}),\\
&\text{(iii)}\ \ P(\|\hat{\theta}_{N,j-1}-\hat{\theta}_N\| > C_3\,\mu_{N,j-1}) = o(N^{-a}), \quad \text{and}\\
&\text{(iv)}\ \ P(\|\hat{\theta}_N-\theta_0\| > C_5(\ln(N)/N)^{1/2}) = o(N^{-a})
\end{aligned} \tag{7.46}$$

for all $j = 1,\ldots,k$ and some $C_j, K < \infty$. Condition (i) holds by Lemma 3(c), (ii) holds by Lemma 3(b) with $p = q_2$, (iv) holds by Lemma 5, (iii) holds for $j = 1$ by the assumption on the initial estimator $\hat{\theta}_{N,0}$ and Lemma 5, and (iii) holds for $j = 2,\ldots,k$ by recursively applying part (a) of Theorem 1 with $k = j-1$, which holds without assuming Assumption 6 by the present proof that the result of Assumption 6 holds for $Q_{N,i}$ for $i \le j-1$ under the assumptions.

Next, we establish (7.44). Element by element mean value expansions give

$$\begin{aligned}
N^{-1}\sum_{i=1}^{N} g(X_i,\hat{\theta}_{N,j-1}) &= N^{-1}\sum_{i=1}^{N} g(X_i,\theta_0)\\
&\quad + N^{-1}\sum_{i=1}^{N}\frac{\partial}{\partial\theta'}\,g(X_i,\theta^+_{N,j-1})(\hat{\theta}_{N,j-1}-\theta_0),
\end{aligned} \tag{7.47}$$

where $\theta^+_{N,j-1}$ lies between $\hat{\theta}_{N,j-1}$ and θ_0. By Lemma 3(c), $P(\|N^{-1}\sum_{i=1}^{N} \times g(X_i,\theta_0)\| > C_4(\ln(N)/N)^{1/2}) = o(N^{-a})$. Combining this with results (iii) and (iv) of (7.46), the first result of Lemma 4, and (7.47) gives

$$P\left(\left\|N^{-1}\sum_{i=1}^{N} g(X_i,\hat{\theta}_{N,j-1})\right\| > \tilde{C}\bar{\psi}_{N,j-1}\right) = o(N^{-a}) \tag{7.48}$$

for some $\tilde{C} < \infty$.

By mean value expansions about θ_0,

$$P\left(\left\|N^{-1}\sum_{i=1}^{N} \frac{\partial^2}{\partial\theta_u\partial\theta'} g(X_i,\hat{\theta}_{N,j-1})\right\| > K\right) = o(N^{-a}) \tag{7.49}$$

for some $K < \infty$, using Lemma 3(b) applied with $m(\widetilde{X}_i) = (\partial^2/\partial\theta_u\partial\theta')g(X_i,\theta_0)$ and results (iii) and (iv) of (7.46). Equations (7.48) and (7.49) combine to yield (7.44). Equation (7.44) holds with Ω replaced by $\Omega_N(\tilde{\theta}_{N,k_1})$ by the second result of Lemma 4 with $\bar{\theta}_{1,N} = \tilde{\theta}_{N,k_1}$ and the preceding proof of (7.44).

The results of the lemma for the restricted matrices $Q_{N,i}^{\tau,NR}$, $Q_{N,i}^{\tau,D}$, $Q_{N,i}^{\tau,LS}$, and $Q_{N,i}^{\tau,GN}$ are proved by the same arguments as for the unrestricted matrices by replacing θ by τ in the appropriate places. ∎

7.3.3. Proof of Lemma 3 A strong-mixing moment inequality of Yokoyama (1980) and Doukhan (1995, Theorem 2 and Remark 2, pp. 25–30) gives $E\|\sum_{i=1}^{N} m(\widetilde{X}_i)\|^p < CN^{p/2}$ provided $p \geq 2$. Application of Markov's inequality and the Yokoyama–Doukhan inequality yields the left-hand side in part (a) of the lemma to be less than or equal to

$$\varepsilon^{-p}N^{-p}E\left\|\sum_{i=1}^{N} m(\widetilde{X}_i)\right\|^p \leq \varepsilon^{-p}CN^{-p/2} = o(N^{-a}). \tag{7.50}$$

Part (b) follows from part (a) applied to $m(\widetilde{X}_i) - Em(\widetilde{X}_1)$ and the triangle inequality.

To establish part (c), we use the Edgeworth expansion given in Theorem 1.1 of Götze and Hipp (1994) (with our $f(\widetilde{X}_i)$ equal to their Z_j and their function $h(Z_j,\ldots,Z_{j+p-1})$ equal to Z_j, which makes their X_j equal to their Z_j). This theorem is a special case of Corollary 2.9 of Götze and Hipp (1983). Conditions (2)–(4) of Götze and Hipp (1994) hold by Assumptions 1, 3(c), 3(d), and 4. Because the result of the lemma can be proved element by element, we consider an arbitrary element $f_v(\cdot)$ of $f(\cdot)$. Let σ_ν^2 denote the variance of $f_\nu(\widetilde{X}_i)$. We assume $\sigma_\nu^2 > 0$; otherwise, the desired result holds trivially. Let $\Phi(\cdot)$ denote the standard normal distribution function. By the Edgeworth expansion, there are homogeneous polynomials $\pi_i(\delta)$ in $\delta = \partial/\partial z$ for $i = 1,\ldots,2a$ such that

$$\sup_{z\in R}\left|P\left(\sigma_\nu^{-1}N^{-1/2}\sum_{i=1}^{N} f_v(\widetilde{X}_i) - Ef_v(\widetilde{X}_i) \leq z\right) - \left(1 + \sum_{i=1}^{2a} N^{-i/2}\pi_i(\delta)\right)\Phi(z)\right|$$
$$= o(N^{-a}). \tag{7.51}$$

This implies that for any constant z_N

$$
\begin{aligned}
P&\left(\left|\sigma_\nu^{-1}N^{-1/2}\sum_{i=1}^{N} f_v(\widetilde{X}_i)-Ef_v(\widetilde{X}_i)\right|>z_N\right)\\
&=1-\left(1+\sum_{i=1}^{2a}N^{-i/2}\pi_i(\delta)\right)(\Phi(z_N)-\Phi(-z_N))+o(N^{-a})\\
&=2\Phi(-z_N)-\left(\sum_{i=1}^{2a}N^{-i/2}\pi_i(\delta)\right)(\Phi(z_N)-\Phi(-z_N))+o(N^{-a}). \qquad \textbf{(7.52)}
\end{aligned}
$$

Let $z_N = C_4 \ln^{1/2}(N)$ for $C_4 > (2a)^{1/2}$. The latter inequality implies that, for some $\varepsilon > 0$, $C_4^2/2 = a + \varepsilon$. Using this and $\Phi(-z) \leq C\exp(-z^2/2)$ for some constant C and all $z > 1$, we have

$$
\begin{aligned}
\Phi(-z_N) &\leq C\exp(-C_4^2\ln(N)/2) = C\exp(-(a+\varepsilon)\ln(N))\\
&= CN^{-(a+\varepsilon)} = o(N^{-a}). \qquad \textbf{(7.53)}
\end{aligned}
$$

The expression $\pi_i(\delta)\Phi(z_N)$ is a finite sum of terms of the form $bz_N^j\phi(z_N)$ for some integer j and real number b, where $\phi(\cdot)$ denotes the standard normal density. By an analogous calculation to that in (7.53), $z_N^j\phi(z_N) = C_4^j\ln^{j/2}(N)(2\pi)^{-1/2}\exp(-C_4^2\ln(N)/2) = o(N^{-a})$. This completes the proof. ■

7.3.4. Proof of Lemma 4 The second result of the lemma follows from

$$P(\|\Omega_N^{-1}(\bar\theta_{1,N})-\Omega_N^{-1}(\theta_0)\|>\varepsilon)=o(N^{-a}), \qquad \textbf{(7.54)}$$

$$P(\|\Omega_N^{-1}(\theta_0)-E\Omega_N^{-1}(\theta_0)\|>\varepsilon)=o(N^{-a}), \quad \text{and} \qquad \textbf{(7.55)}$$

$$E\Omega_N^{-1}(\theta_0)=\Omega_0^{-1}. \qquad \textbf{(7.56)}$$

To establish (7.54), we take mean value expansions about θ_0, apply Lemma 3(b) with $m(\widetilde{X}_i) = \sup_{\theta\in N_0}\|g(X_i,\theta)\|\cdot\|(\partial/\partial\theta')g(X_{i+j},\theta)\|$ for $j=-\kappa,\ldots,\kappa$ and $p = q_1$, and use the assumption on $\bar\theta_{1,N}$. To establish (7.55), we use Lemma 3(a) with $m(\widetilde{X}_i) = g(X_i,\theta_0)g(X_{i+j},\theta_0)' - Eg(X_1,\theta_0)g(X_{1+j},\theta_0)'$ for $j = -\kappa,\ldots,\kappa$ and $p = q_1$. Equation (7.56) holds by definition of Ω_0.

The third, fourth, and fifth results of the lemma follow from the first two results of the lemma and the following conditions. For some $K<\infty$ and all $\varepsilon>0$,

$$P\left(\left\|N^{-1}\sum_{i=1}^{N}\frac{\partial^j}{\partial\theta^j}g(X_i,\bar\theta_{1,N})\right\|>K\right)=o(N^{-a}) \quad \text{for } j=1,2,3, \quad \text{and} \qquad \textbf{(7.57)}$$

$$P\left(\left\|N^{-1}\sum_{i=1}^{N}g(X_i,\bar\theta_{1,N})\right\|>\varepsilon\right)=o(N^{-a}). \qquad \textbf{(7.58)}$$

The first result of the lemma, (7.57), and (7.58) hold by mean value expansions about θ_0, multiple applications of Lemma 3(b) with $m(\widetilde{X}_i) = (\partial^j/\partial\theta^j)g(X_i,\theta_0)$ for $j = 0,\ldots,3$ or $m(\widetilde{X}_i) = C_g(X_i)$, multiple applications of Lemma 3(a) with $m(\widetilde{X}_i) = (\partial^j/\partial\theta^j)\, g(X_i,\theta_0) - E(\partial^j/\partial\theta^j)\, g(X_i,\theta_0)$ for $j = 0,1$ and $p = q_1$, and the assumption on $\bar{\theta}_{1,N}$. ∎

7.3.5. Proof of Lemma 5 First, we show that for all $\varepsilon > 0$,

$$P\left(\sup_{\theta\in\Theta}\left\|N^{-1}\sum_{i=1}^{N} G_u(X_i,\theta)\right\| > \varepsilon\right) = o(N^{-a}) \quad \text{for } u = 1,2, \quad \text{where}$$
$$G_1(X_i,\theta) = g(X_i,\theta) - Eg(X_1,\theta) \quad \text{and} \quad G_2(X_i,\theta) = \rho(X_i,\theta) - E\rho(X_1,\theta). \tag{7.59}$$

Let $B(\theta,\varepsilon)$ denote the ball centered at θ with radius ε. By Assumption 2(a), Θ is compact. Hence, for any $\eta > 0$, there exist points $\{\theta_j \in \Theta : j \le J\}$ such that $\cup_{j=1}^{J} B(\theta_j,\eta)$ contains Θ. For $u = 1$, the left-hand side of (7.59) is less than or equal to

$$\begin{aligned}
&P\left(\max_{j\le J}\sup_{\theta\in B(\theta_j,\eta)}\left(\left\|N^{-1}\sum_{i=1}^{N} G_1(X_i,\theta) - G_1(X_i,\theta_j)\right\|\right.\right.\\
&\qquad\qquad\left.\left. + \left\|N^{-1}\sum_{i=1}^{N} G_1(X_i,\theta_j)\right\|\right) > \varepsilon\right)\\
&\quad\le P\left(\max_{j\le J}\sup_{\theta\in B(\theta_j,\eta)} N^{-1}\sum_{i=1}^{N}(C_g(X_i) + EC_g(X_i))\|\theta-\theta_j\| > \frac{\varepsilon}{2}\right)\\
&\qquad + P\left(\max_{j\le J}\left\|N^{-1}\sum_{i=1}^{N} G_1(X_i,\theta_j)\right\| > \frac{\varepsilon}{2}\right)\\
&\quad\le P\left(N^{-1}\sum_{i=1}^{N}(C_g(X_i) + EC_g(X_i))\eta > \frac{\varepsilon}{2}\right)\\
&\qquad + \sum_{j=1}^{J} P\left(\left\|N^{-1}\sum_{i=1}^{N} G_1(X_i,\theta_j)\right\| > \frac{\varepsilon}{2}\right)\\
&\quad = o(N^{-a}),
\end{aligned} \tag{7.60}$$

where the first inequality uses Assumption 2(b) and the equality holds by Lemma 3(b) with $p = q_0$ by taking η sufficiently small and Lemma 3(a) with $p = q_0$. The proof for $u = 2$ is the same except that $C_g(\cdot)$ is replaced by $\sup_{\theta\in\Theta}\|(\partial/\partial\theta)\rho(\cdot,\theta)\|$.

Now, we prove that $P(\|\hat{\theta}_N - \theta_0\| > \varepsilon) = o(N^{-a})$ for $\varepsilon > 0$ for the minimum ρ estimator under Assumption 2(b)(ii). Let $\rho(\theta) = E\rho(X_1,\theta)$. Given $\varepsilon > 0$, there exists a $\delta > 0$ such that $\|\theta - \theta_0\| > \varepsilon$ implies that $\rho(\theta) - \rho(\theta_0) \ge \delta > 0$. Thus,

$$\begin{aligned}P(\|\hat{\theta}_N - \theta_0\| > \varepsilon) &\leq P(\rho(\hat{\theta}_N) - \rho_N(\hat{\theta}_N) + \rho_N(\hat{\theta}_N) - \rho(\theta_0) > \delta)\\ &\leq P(\rho(\hat{\theta}_N) - \rho_N(\hat{\theta}_N) + \rho_N(\theta_0) - \rho(\theta_0) > \delta)\\ &\leq P\left(2 \sup_{\theta \in \Theta} |\rho_N(\theta) - \rho(\theta)| > \delta\right)\\ &= o(N^{-a}) \end{aligned} \tag{7.61}$$

using (7.59) with $u = 2$. The corresponding proof for the FW-GMM estimator under Assumption 2(b)(i) is analogous with $\rho(\theta)$ and $\rho_N(\theta)$ replaced by $J(\theta) = Eg(X_1, \theta)'\Omega Eg(X_1, \theta)$ and $J_N(\theta)$, respectively.

For the minimum ρ estimator, the result that $P(\|\hat{\theta}_N - \theta_0\| > \varepsilon) = o(N^{-a})$ and the assumption that θ_0 is in the interior of Θ imply that with probability $1 - o(N^{-a})$ $\hat{\theta}_N$ is in the interior of Θ, $N^{-1}\sum_{i=1}^N g(X_i, \hat{\theta}_N) = 0$, and $\hat{\theta}_N$ minimizes not only $\rho_N(\theta)$ but $J_N(\theta)$ (defined with an arbitrary positive definite weight matrix Ω) over $\theta \in \Theta$. In consequence, in the remainder of this proof, we can treat the minimum ρ estimator as a FW-GMM estimator.

Next, we prove the result of the lemma for the FW-GMM estimator. We have the following conditions: $\hat{\theta}_N$ is in the interior of Θ and $(\partial/\partial\theta)J_N(\hat{\theta}_N) = 0$ with probability $1 - o(N^{-a})$. Hence, element by element mean value expansions of $(\partial/\partial\theta)J_N(\hat{\theta}_N)$ about θ_0 and rearrangement give

$$\hat{\theta}_N - \theta_0 = -\left(\frac{\partial^2}{\partial\theta\partial\theta'} J_N(\theta_N^+)\right)^{-1} \frac{\partial}{\partial\theta} J_N(\theta_0) \tag{7.62}$$

with probability $1 - o(N^{-a})$, where θ_N^+ lies between $\hat{\theta}_N$ and θ_0 and may differ across rows. In consequence, the result of the lemma follows from the third result of Lemma 4 for the FW-GMM estimator with $\bar{\theta}_{1,N} = \theta_N^+$, the first result of Lemma 4 with $\bar{\theta}_{1,N} = \theta_0$, and $P(\|N^{-1/2}\sum_{i=1}^N g(X_i, \theta_0)\| > C_4 \ln^{1/2}(N)) = o(N^{-a})$, which holds by Lemma 3(c) with $m(\widetilde{X}_i) = g(X_i, \theta_0)$ using the assumption that $q_1 \geq 2a + 3$.

Given the second result of Lemma 4, the proof of the lemma for the EW-GMM estimator is analogous to that for the FW-GMM estimator. ■

7.3.6. Proof of Lemma 6 For any convex set $B \subset R^{L_A}$ and any $\tau > 0$, let $B_\tau^+ = \{x \in R^{L_A} : \|x - y\| \leq \tau \text{ for some } y \in B\}$. We have

$$\begin{aligned}&P(A_N + \xi_{1,N} \in B) - P(A_N \in B)\\ &\quad = P(A_N + \xi_{1,N} \in B, \|\xi_{1,N}\| \leq \omega_N) - P(A_N \in B)\\ &\qquad + P(A_N + \xi_{1,N} \in B, \|\xi_{1,N}\| > \omega_N)\\ &\quad \leq (P(A_N \in B_{\omega_N}^+) - P(A_N \in B)) + P(\|\xi_{1,N}\| > \omega_N). \end{aligned} \tag{7.63}$$

The second term on the right-hand side is $o(N^{-a})$ by assumption. When A_N has an Edgeworth expansion with remainder $o(N^{-a})$, the first term on the last line of (7.63) is less than or equal to

$$\int_{B^+_{\omega_N}} \left(1 + \sum_{i=1}^{[2a]} N^{-i/2}\pi_i(z)\right)\phi_{\Sigma_N}(z)\,dz - \int_B \left(1 + \sum_{i=1}^{[2a]} N^{-i/2}\pi_i(z)\right)\phi_{\Sigma_N}(z)\,dz$$
$$+\, o(N^{-a}) \tag{7.64}$$

uniformly over convex sets B. The difference between the integrals is $O(\omega_N) = o(N^{-a})$ uniformly over convex sets $B \subset R^{L_A}$, because $\phi_{\Sigma_N}(z)$ and its derivatives of all orders are bounded over $z \in R^{L_A}$ given the assumptions on Σ_N. Hence, $P(A_N + \xi_{1,N} \in B) - P(A_N \in B) \leq o(N^{-a})$ uniformly over convex sets B.

Let $B^-_\tau = \{x \in B : \|x - y\| \geq \tau$ for all $y \in B^c\}$, where B^c denotes the complement of B. We have $P(A_N + \xi_{1,N} \in B, \|\xi_{1,N}\| \leq \omega_N) \geq P(A_N \in B^-_{\omega_N})$. Using this, an analogous argument shows that $P(A_N \in B) - P(A_N + \xi_{1,N} \in B) \leq o(N^{-a})$ uniformly over convex sets B, which completes the proof of the first result of the lemma.

The proof of the second result is analogous with $B \subset R^{L_A}$ and A_N replaced by $B \subset R$ and $A'_N A_N$, respectively, in (7.63) and B_{ω_N}, B, and $\phi_{\Sigma_N}(\cdot)$ replaced by $\{x \in R^{L_A} : x'x \in B_{\omega_N}\}$, $\{x \in R^{L_A} : x'x \in B\}$, and $f_{\chi^2}(\cdot)$, respectively, in (7.64). Again, the difference between the integrals is $O(\omega_N) = o(N^{-a})$ uniformly over convex sets B. ∎

7.3.7. Proof of Lemma 7 By a mean value expansion and the triangle inequality,

$$\|m_N(\bar\theta_{2,N}) - m_N(\bar\theta_{1,N})\|$$
$$\leq \left(N^{-1}\sum_{i=1}^{N} \sup_{\theta\in N_0} \|(\partial/\partial\theta) m(\widetilde{X}_i,\theta)\|\right)\cdot\|\bar\theta_{2,N} - \bar\theta_{1,N}\|. \tag{7.65}$$

Hence, the lemma holds by the assumption on $\|\bar\theta_{2,N} - \bar\theta_{1,N}\|$ and Lemma 3(b) with $m(\widetilde{X}_i) = \sup_{\theta\in N_0}\|(\partial/\partial\theta)m(\widetilde{X}_i,\theta)\|$. ∎

7.3.8. Proof of Lemma 8 First, we establish the result of the lemma with $\Delta_N = N^{1/2}(\hat\theta_N - \theta_0)$, where $\hat\theta_N$ is the FW-GMM estimator. By Lemma 5 and Assumption 2(a), $\hat\theta_N$ is in the interior of Θ and $(\partial/\partial\theta)J_N(\hat\theta_N) = 0$ with probability $1 - o(N^{-a})$. Element by element Taylor expansions of $(\partial/\partial\theta)J_N(\hat\theta_N)$ about θ_0 of order $d-1$ give

$$0 = \frac{\partial}{\partial\theta}J_N(\hat\theta_N) = \frac{\partial}{\partial\theta}J_N(\theta_0) + \sum_{j=1}^{d-1}\frac{1}{j!}D^j\frac{\partial}{\partial\theta}J_N(\theta_0)(\hat\theta_N - \theta_0,\ldots,\hat\theta_N - \theta_0)$$
$$+\,\zeta_N, \quad \text{where}$$
$$\zeta_N = \frac{1}{j!}\left(D^{d-1}\frac{\partial}{\partial\theta}J_N(\theta_N^+) - D^{d-1}\frac{\partial}{\partial\theta}J_N(\theta_0)\right)(\hat\theta_N - \theta_0,\ldots,\hat\theta_N - \theta_0), \tag{7.66}$$

θ_N^+ lies between $\hat{\theta}_N$ and θ_0, and $D^j(\partial/\partial\theta)J_N(\theta_0)(\hat{\theta}_N - \theta_0,\dots,\hat{\theta}_N - \theta_0)$ denotes $D^j(\partial/\partial\theta)J_N(\theta_0)$ as a j-linear map, whose coefficients are partial derivatives of $(\partial/\partial\theta)J_N(\theta_0)$ of order j, applied to the j-tuple $(\hat{\theta}_N - \theta_0,\dots,\hat{\theta}_N - \theta_0)$. Let R_N denote the column vector whose elements are the unique components of $(\partial/\partial\theta)J_N(\theta_0)$, $D^1(\partial/\partial\theta)J_N(\theta_0),\dots,D^{d-1}(\partial/\partial\theta)J_N(\theta_0)$. Each element of R_N is an infinitely differentiable function of S_N. Let R denote the probability limit of R_N. Let $e_N = (\zeta_N',0,\dots,0)'$ be conformable to R_N. The first equation in (7.66) can be written as $\nu(R_N + e_N, \hat{\theta}_N - \theta_0) = 0$, where $\nu(\cdot,\cdot)$ is an infinitely differentiable function, $\nu(R,0) = 0$, and $(\partial/\partial x)\nu(R,x)|_{x=0} = \mathrm{plim}_{N\to\infty}(\partial^2/\partial\theta\partial\theta')J_N(\theta_0) = 2\Omega$ is positive definite by Assumption 3(b). Hence, the implicit function theorem can be applied to $\nu(\cdot,\cdot)$ at the point $(R,0)$ to obtain

$$\hat{\theta}_N - \theta_0 = \Lambda(R_N + e_N) \tag{7.67}$$

with probability $1 - o(N^{-a})$, where Λ is a function that does not depend on N, is infinitely differentiable in a neighborhood of R, and satisfies $\Lambda(R) = 0$.

We apply Lemma 6 with $A_N = N^{1/2}\Lambda(R_N)$ and $\xi_N = N^{1/2}(\Lambda(R_N + e_N) - \Lambda(R_N))$ to obtain

$$\lim_{N\to\infty} \sup_{B\in\mathcal{B}_{L_\theta}} N^a |P(N^{1/2}\Lambda(R_N + e_N) \in B) - P(N^{1/2}\Lambda(R_N) \in B)| = 0. \tag{7.68}$$

Lemma 6 applies because (i) $P(\|\xi_N\| > \omega_N) \le P(CN^{1/2}\|e_N\| > \omega_N)$ by a mean value expansion, (ii) $\|e_N\| = \|\zeta_N\|$, (iii) ζ_N satisfies $\|\zeta_N\| \le C\|\hat{\theta}_N - \theta_0\|^d$ with probability $1 - o(N^{-a})$, (iv) ω_N, which is defined to equal $N^{1/2-d/2}\ln^d(N)$, is $o(N^{-a})$ because $d \ge 2a + 2$ by Assumption 3(c), (v) $P(N^{1/2}\|e_N\| > \omega_N) \le P(CN^{1/2}\|\hat{\theta}_N - \theta_0\|^d > \omega_N) + o(N^{-a}) = o(N^{-a})$ by Lemma 5, (vi) $\Lambda(R_N)$ can be written as $G(S_N)$, where $G(\cdot)$ is infinitely differentiable and $G(S) = 0$, and (vii) $A_N = N^{1/2}\Lambda(R_N) = N^{1/2}G(S_N)$ has an Edgeworth expansion by the proof of Lemma 9, which follows.

Equations (7.67) and (7.68) and $\Lambda(R_N) = G(S_N)$ yield the result of the lemma.

The proof for the minimum ρ estimator is identical because the latter satisfies $(\partial/\partial\theta)J_N(\hat{\theta}_N) = 0$ with probability $1 - o(N^{-a})$ by Lemma 5.

Next, suppose $\hat{\theta}_N$ is the EW-GMM estimator. We take a Taylor expansion of order $d - 1$ of $(\partial/\partial\theta)J_N(\hat{\theta}_N,\tilde{\theta}_N)$ about $(\hat{\theta}_N,\tilde{\theta}_N) = (\theta_0,\theta_0)$. Applying the implicit function theorem as before, there exists an infinitely differentiable function $\Lambda^*(\cdot,\cdot)$ such that

$$\hat{\theta}_N - \theta_0 = \Lambda^*(R_N^* + e_N^*, \tilde{\theta}_N - \theta_0) \tag{7.69}$$

with probability $1 - o(N^{-a})$, where $\Lambda^*(R^*,0) = 0$ and R_N^*, R^*, $e_N^* = (\zeta_N^{*\prime},0,\dots,0)'$, and ζ_N^* are defined analogously to R_N, R, e_N, and ζ_N. Substituting (7.67) with $\hat{\theta}_N$ replaced by $\tilde{\theta}_N$ into (7.69) and applying Lemma 6 as before gives a result analogous to (7.68) with $\Lambda(R_N + e_N)$ and $\Lambda(R_N)$ replaced by

$\Lambda^*(R_N^* + e_N^*, \Lambda(R_N + e_N))$ and $\Lambda^*(R_N^*, \Lambda(R_N))$, respectively. We can write $\Lambda^*(R_N^*, \Lambda(R_N))$ as $G(S_N)$, where $G(\cdot)$ is an infinitely differentiable function and $G(S) = \Lambda^*(R^*, \Lambda(R)) = \Lambda^*(R^*,0) = 0$. Combining this, the analogue of (7.68), and (7.69) gives the result of the lemma for the EW-GMM estimator.

Each of the remaining forms of Δ_N (namely, T_N, $H_N(\hat{\theta}_N)$, $U_N(\hat{\theta}_N)$, and $K_N(\hat{\theta}_N)$) is a function of $\hat{\theta}_N$ and, possibly, $\tilde{\theta}_N$. We take a Taylor expansion of $\Delta_N/N^{1/2}$ about $(\hat{\theta}_N, \tilde{\theta}_N) = (\theta_0, \theta_0)$ to order $d-1$ to obtain

$$\Delta_N = N^{1/2}(\Lambda^{**}(S_N, \hat{\theta}_N - \theta_0, \tilde{\theta}_N - \theta_0) + \zeta_N^{**}), \qquad \textbf{(7.70)}$$

where Λ^{**} is an infinitely differentiable function, $\Lambda^{**}(S,0,0) = 0$, ζ_N^{**} is the remainder term in the Taylor expansion, and $\|\zeta_N^{**}\| = O(\|\hat{\theta}_N - \theta_0\|^d) + O(\|\tilde{\theta}_N - \theta_0\|^d)$. Substituting (7.67) and/or (7.69) into (7.70) gives $\Delta_N = N^{1/2}(\Lambda^{**}(S_N,\ \Lambda^*(R_N^* + e_N^*,\ \Lambda(R_N + e_N)),\ \Lambda(R_N + e_N)) + \zeta_N^{**})$. We apply Lemma 6 again, using the preceding result for $\|\zeta_N^{**}\|$, to obtain an analogue of (7.68) with $A_N = N^{1/2}\Lambda^{**}(S_N, \Lambda^*(R_N^*, \Lambda(R_N)), \Lambda(R_N))$. We can write $G(S_N) = \Lambda^{**}(S_N,\ \Lambda^*(R_N^*, \Lambda(R_N)),\ \Lambda(R_N))$, where $G(\cdot)$ is infinitely differentiable and $G(S) = \Lambda^{**}(S,\ \Lambda^*(R^*, \Lambda(R)),\ \Lambda(R)) = \Lambda^{**}(S,0,0) = 0$. Combining this, the analogue of (7.68), and (7.70) gives the result of the lemma for Δ_N equal to T_N, $H_N(\hat{\theta}_N)$, $U_N(\hat{\theta}_N)$, or $K_N(\hat{\theta}_N)$. ■

7.3.9. Proof of Lemma 9 Given Lemma 8 and the triangle inequality, it suffices to show that the random vectors $N^{1/2}G(S_N)$ of Lemma 8 possess Edgeworth expansions with remainder $o(N^{-a})$. First, we obtain an Edgeworth expansion for $N^{1/2}(S_N - S)$ via Theorem 1.1 of Götze and Hipp (1994), as in the proof of Lemma 3(c). The Edgeworth expansion for $N^{1/2}G(S_N)$ is now obtained from that of $N^{1/2}(S_N - S)$ by the argument in Bhattacharya (1985, proof of Theorem 1) or Bhattacharya and Ghosh (1978, proof of Theorem 2) using the smoothness of $G(\cdot)$, $G(S) = 0$, and Assumption 3(b). ■

NOTES

1. Note that some authors, e.g., Rothenberg (1984), say that two statistics have equal s-order asymptotic efficiency if their distributions are of CV distance $o(N^{-(s-1)})$ apart, rather than $o(N^{-(s-1)/2})$.

2. As stated, these definitions of equivalence of higher order asymptotic efficiency apply for a single data generating process (DGP). They could be altered to cover multiple DGPs. For an estimator, one could require that the CV distance is $o(N^{-(s-1)/2})$ for all DGPs that correspond to a true parameter $\theta_0 \in \Theta$. For a test statistic, one could require that the CV distance is $o(N^{-(s-1)/2})$ for all distributions in the null hypothesis. The results of the paper cover definitions of this sort. One just needs the assumptions stated in Section 5 to hold for all DGPs of interest and then the results given apply to all such DGPs.

3. Specifically, the results stated in the second and third paragraphs following equation (1.2) hold when the initial estimator satisfies $P(\|\hat{\theta}_{N,0} - \theta_0\| > C_1\gamma_N) = o(N^{-a})$ with $\gamma_N = (\ln(N)/N)^{1/2}$ for some finite constant C_1. A wide class of extremum estimators satisfies this condition; see Lemma 1.

4. The matrix $\overline{W}_N(\theta)$ is positive definite with probability that goes to one at a rate that is sufficiently fast for the results of the paper to hold. In finite samples, however, $\overline{W}_N(\theta)$ is not necessarily positive definite. If $\overline{W}_N(\theta)$ is not positive definite, $\Omega_N(\theta)$ can be defined in an arbitrary fashion, and the results of the paper hold. For example, one could compute $\overline{W}_N(\theta)$ with κ replaced by a smaller value for which $\overline{W}_N(\theta)$ is positive definite.

5. The rth element of $\hat{\theta}_N$ is denoted $(\hat{\theta}_N)_r$, rather than $\hat{\theta}_{N,r}$, to distinguish it from the k-step estimator, $\hat{\theta}_{N,k}$, defined in Section 4.

REFERENCES

Akahira, M. & K. Takeuchi (1981) *Asymptotic Efficiency of Statistical Estimators: Concepts and Higher-Order Efficiency.* Lecture Notes in Statistics, vol. 7. New York: Springer-Verlag.

Andrews, D.W.K. (2002) Higher-order improvements of a computationally attractive k-step bootstrap. *Econometrica* 70, 119–162.

Bhattacharya, R.N. (1985) Some recent results on Cramér-Edgeworth expansions with applications. In P.K. Krishnaiah (ed.), *Multivariate Analysis*, vol. VI, pp. 57–75. New York: North-Holland.

Bhattacharya, R.N. & J.K. Ghosh (1978) On the validity of the formal Edgeworth expansion. *Annals of Statistics* 6, 434–451.

Brown, B.W. & S. Maital (1981) What do economists know? An empirical study of experts' expectations. *Econometrica* 49, 491–504.

Chandra, T.K. & J.K. Ghosh (1979) Valid asymptotic expansions for the likelihood ratio statistic and other perturbed chi-square variables. *Sankhya* 41, Series A, 22–47.

Davidson, R. & J.G. MacKinnon (1999) Bootstrap testing in nonlinear models. *International Economic Review* 40, 487–508.

Dennis, J.E. & R.B. Schnabel (1983) *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Englewood Cliffs, New Jersey: Prentice-Hall.

Doukhan, P. (1995) *Mixing: Properties and Examples*. New York: Springer-Verlag.

Fisher, R.A. (1925) Theory of statistical estimation. *Proceedings of the Cambridge Philosophical Society* 22, 700–725.

Götze, F. & C. Hipp (1983) Asymptotic expansions for sums of weakly dependent random vectors. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 64, 211–239.

Götze, F. & C. Hipp (1994) Asymptotic distribution of statistics in time series. *Annals of Statistics* 22, 2062–2088.

Hall, P. & J.L. Horowitz (1996) Bootstrap critical values for tests based on generalized-method-of-moment estimators. *Econometrica* 64, 891–916.

Hansen, L.P. & R.J. Hodrick (1980) Forward exchange rates as optimal predictors of future spot rates: An econometric analysis. *Journal of Political Economy* 88, 829–863.

Hansen, L.P. & K.J. Singleton (1982) Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica* 50, 1269–1286.

Hansen, L.P. & K.J. Singleton (1996) Efficient estimation of linear asset-pricing models with moving average errors. *Journal of Business and Economic Statistics* 14, 53–68.

Janssen, P., J. Jureckova, & N. Veraverbeke (1985) Rate of convergence of one- and two-step M-estimators with applications to maximum likelihood and Pitman estimators. *Annals of Statistics* 13, 1222–1229.

LeCam, L. (1956) On the asymptotic theory of estimation and testing hypotheses. In J. Neyman (ed.), *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 129–156. Berkeley: University of California Press.

Ortega, J.M. & W.C. Rheinboldt (1970) *Iterative Solution of Nonlinear Equations in Several Variables.* New York: Academic Press.

Pfanzagl, J. (1974) Asymptotic optimum estimation and test procedures. In J. Hájek (ed.), *Proceedings of the Prague Symposium on Asymptotic Statistics 3–6 September 1973*, vol. I. Prague: Charles University.

Pfanzagl, J. & W. Wefelmeyer (1978) A third-order optimum property of the maximum likelihood estimator. *Journal of Multivariate Analysis* 8, 1–29. Addendum: 9, 179–182.

Robinson, P.M. (1988) The stochastic difference between econometric statistics. *Econometrica* 56, 531–548.

Rothenberg, T.J. (1984) Approximating the distributions of econometric estimators and test statistics. In Z. Griliches & M.D. Intriligator (eds.), *Handbook of Econometrics*, pp. 881–935. Amsterdam: North-Holland.

Yokoyama, R. (1980) Moment bounds for stationary mixing sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 52, 45–57.