7 Hypothesis Testing

In this chapter we consider the properties of versions of the Wald, Lagrange multiplier, and "likelihood ratio" statistics for testing the hypothesis

$$H_o: h(\theta_n^*) = h_n^o \quad \text{a.a. } n$$

versus

$$H_a: h(\theta_n^*) \neq h_n^o \quad \text{i.o.},$$

where $h: \Theta \to \mathbb{R}^q$, $q \in \mathbb{N}$. Each statistic we consider, say T_n, is decomposed into a sum of two random variables

$$T_n = X_n + a_n$$

where a_n converges in probability to zero and X_n has a known finite sample distribution. Such a decomposition permits the statement that for any $t \in \mathbb{R}$

$$\lim_{n \to \infty} [P(T_n > t) - P(X_n > t)] = 0.$$

Because we allow specification error and nonstationarity we will not necessarily have T_n converging in distribution to a random variable X. However, the utility of convergence in distribution in applications derives from the statement

$$\lim_{n \to \infty} [P(T_n > t) - P(X > t)] = 0,$$

because $P(X > t)$ is computable. Since our $P(X_n > t)$ is computable, our results carry the full force of classical results.

The starting point for our results is the following set of conditions, which we collect together as a single assumption. All of the results given in this chapter rely on these conditions.

Assumption HT (hypothesis testing)

Let (Ω, F, P) be a complete probability space and let $\Theta \subset \mathbb{R}^k$, $k \in \mathbb{N}$, be a compact set. Assume:

(a) $Q_n: \Omega \times \Theta \to \mathbb{R}$ is a random function continuously differentiable of order 2 on Θ, a.s., $n = 1, 2, \ldots$.

(b) There exist sequences of functions $\{\bar{Q}_n: \Theta \to \mathbb{R}\}$ and $\{A_n: \Theta \to \mathbb{R}^{k \times k}\}$ such that $\bar{Q}_n$ is differentiable on Θ and

$$Q_n(\theta) - \bar{Q}_n(\theta) \to 0 \quad \text{a.s. uniformly on } \Theta,$$
$$\nabla_\theta Q_n(\theta) - \nabla_\theta \bar{Q}_n(\theta) \to 0 \quad \text{a.s. uniformly on } \Theta,$$
$$\nabla_\theta^2 Q_n(\theta) - A_n(\theta) \to 0 \quad \text{a.s. uniformly on } \Theta,$$

where $\{\bar{Q}_n\}$, $\{\nabla_\theta \bar{Q}_n\}$, and $\{A_n\}$ are continuous on Θ uniformly in n.

(c) $\{\bar{Q}_n\}$ has identifiably unique minimizers $\{\theta_n^*\}$ on Θ, interior to Θ uniformly in n. Define

$$\Theta_n \equiv \{\theta \in \Theta : h(\theta) = h_n^o\}$$

where $h: \Theta \to \mathbb{R}^q$, $q \in \mathbb{N}$, is continuously differentiable on Θ, $h_n^* \equiv h(\theta_n^*)$, and $\{h_n^o\}$ is chosen so that

$$\sqrt{(n)}(h_n^* - h_n^o) = O(1).$$

Assume that $\{\bar{Q}_n\}$ has identifiably unique minimizers $\{\theta_n^o\}$ on $\{\Theta_n\}$, interior to Θ uniformly in n.

(d) $\sqrt{(n)}(\theta_n^* - \theta_n^o) = O(1)$.

(e) There exist sequences $\{B_n^o\}$ and $\{B_n^*\}$ of $O(1)$ uniformly positive definite symmetric $k \times k$ matrices such that

$$\sqrt{(n)} B_n^{o\,-1/2} (\nabla_\theta Q_n^o - \nabla_\theta \bar{Q}_n^o)' \overset{A}{\sim} N(0, I_k)$$
$$\sqrt{(n)} B_n^{*\,-1/2} \nabla_\theta Q_n^{*\prime} \overset{A}{\sim} N(0, I_k).$$

(f) There exist sequences $\{\bar{B}_n: \Omega \to \mathbb{R}^{k \times k}\}$ and $\{\hat{B}_n: \Omega \to \mathbb{R}^{k \times k}\}$ measurable-$F/B(\mathbb{R}^{k \times k})$ and $O(1)$ nonstochastic sequences $\{U_n^o\}$ and $\{U_n^*\}$ such that

$$\bar{B}_n - (B_n^o + U_n^o) \xrightarrow{P} 0$$
$$\hat{B}_n - (B_n^* + U_n^*) \xrightarrow{P} 0.$$

(g) There exists a closed sphere $S \subset \Theta$ of finite nonzero radius such that for some $\varepsilon > 0$

$$\cup_{n=1}^{\infty} \{\theta \in \Theta : |\theta - \theta_n^*| < \varepsilon\} \subset S$$

and $\{A_n(\theta)\}$ is $O(1)$ and uniformly positive definite uniformly on S.

(h) There exist sequences $\{\tilde{A}_n:\Omega \to \mathbb{R}^{k\times k}\}$ and $\{\hat{A}_n:\Omega \to \mathbb{R}^{k\times k}\}$ measurable-$F/B(\mathbb{R}^{k\times k})$ such that

$$\tilde{A}_n - A_n^o \to 0 \quad \text{a.s.}$$
$$\hat{A}_n - A_n^* \to 0 \quad \text{a.s.}$$

where $A_n^o \equiv A_n(\theta_n^o)$, $A_n^* \equiv A_n(\theta_n^*)$. □

Given assumption HT(a), theorem 2.2 ensures the existence of a measurable function $\hat{\theta}_n$ which solves the problem $\min_\Theta Q_n(\theta)$ a.s.; further, theorem 3.19 ensures that $\hat{\theta}_n - \theta_n^* \to 0$ a.s. given assumptions HT(a), (b), and (c). The same theorems ensure that there exists a measurable function $\tilde{\theta}_n$ which solves the constrained problem $\min_{\Theta_n} Q_n(\theta)$ a.s. for Θ_n as defined in assumption HT(c) and that $\tilde{\theta}_n - \theta_n^o \to 0$ a.s.

The requirement that h_n^o be chosen in such a way that

$$\sqrt{(n)}(h_n^* - h_n^o) = O(1)$$

is the way in which we impose a Pitman drift. We keep the data generating process (defined by P) fixed, and suppose that the investigator gradually adjusts his hypothesis in such a way as to approach the "truth" asymptotically. This is in contrast to the traditional approach in which the hypothesis is held fixed, and the data generating process (different for each sample size) drifts in such a way that the data generation process satisfies the hypothesis asymptotically. In some respects our approach is philosophically more palatable than the standard approach. It makes more sense to assume that an investigator slowly discovers the truth as more data become available than to assume that nature slowly accommodates to the investigator's pig-headedness. But withal, the drift is only a technical artifice to obtain approximations to the sampling distributions of test statistics that are reasonably accurate in applications, so that philosophical nitpicking of this sort is ultimately irrelevant.

The requirement that $\sqrt{(n)}(\theta_n^* - \theta_n^o) = O(1)$ is essentially a requirement that θ_n^* approach θ_n^o at the same rate at which h_n^* approaches h_n^o. This can be ensured by placing further restrictions on h. Lemmas 7.1 and 7.2 below enable us to place restrictions on h ensuring that $\sqrt{(n)}(\theta_n^* - \theta_n^o) = O(1)$ as required.

Assumption HT(e) is the conclusion of theorem 5.4 and corollary 5.5, while assumption HT(f) posits the existence of consistent estimators for $B_n^o + U_n^o$ and for $B_n^* + U_n^*$. For this any appropriate result of chapter 6 will suffice.

Assumption HT(g) is a strengthening of assumption PD(ii); however, in cases in which $\theta_n^* = \theta^*$ for all n, S can be chosen as a sufficiently small neighborhood of θ^* without further conditions, as the continuity of the determinant function, the uniform continuity of A_n on Θ, and the uniform positive definiteness of $\{A_n^*\}$ ensure that $\{A_n(\theta)\}$ is $O(1)$ and uniformly positive definite uniformly on such a neighborhood S. We strengthen assumption PD.

Assumption PD′

Assumption PD(i) holds, and

(ii) There exists a closed sphere $S \subset \Theta$ of finite nonzero radius such that for some $\varepsilon > 0$

$$\cup_{n=1}^{\infty} \{\theta \in \Theta : |\theta - \theta_n^*| < \varepsilon\} \subset S$$

and $\{A_n(\theta)\}$ is $O(1)$ and uniformly positive definite uniformly on S. □

Finally, assumption HT(h) is the conclusion of theorem 6.1. This suggests that the results of the previous chapters provide sufficient conditions for assumption HT, once we ensure that $\sqrt{(n)}(\theta_n^* - \theta_n^o) = O(1)$. To do this, we first establish that $\theta_n^* - \theta_n^o \to 0$ as $h_n^* - h_n^o \to 0$. The following assumption plays a key role in ensuring this.

Assumption CN (constraint)

Suppose $h:\Theta \to \mathbb{R}^q$, $q \in \mathbb{N}$, is continuously differentiable of order 2 on Θ with Jacobian $H(\cdot) = \nabla_\theta h(\cdot)$ such that the eigenvalues of $H(\theta)H(\theta)'$ are bounded below on S by $\delta > 0$ and above by $\Delta < \infty$.

For $q = k$, let h be one-to-one with a continuous inverse on S. For $q < k$, suppose there exists $r:\Theta \to \mathbb{R}^{k-q}$ continuous on Θ such that the mapping

$$(\rho', \tau') = (r(\theta)', h(\theta)')$$

has a continuous inverse

$$\theta = \Psi(\rho, \tau)$$

defined over $M = \{(\rho, \tau): \rho = r(\theta), \tau = h(\theta), \theta \in S\}$. Moreover, $\Psi(\rho, \tau)$ has a continuous extension to the set

$$R \times T = \{\rho: \rho = r(\theta), \theta \in \Theta\} \times \{\tau: \tau = h(\theta), \theta \in \Theta\}. \qquad \square$$

Lemma 7.1

Given assumptions HT(b), HT(c), and CN, if $h_n^* - h_n^o \to 0$ as $n \to \infty$, then $\theta_n^* - \theta_n^o \to 0$ as $n \to \infty$. $\square$

We now formally impose the Pitman drift assumption.

Assumption DR (Pitman drift)

The sequence $\{h_n^o\}$ is chosen such that

$$\sqrt{(n)}(h_n^* - h_n^o) = O(1). \qquad \square$$

By making use of lemma 7.1 and the following result, it is straightforward to establish that $\sqrt{(n)}(\theta_n^* - \theta_n^o) = O(1)$.

Lemma 7.2

Let A be a symmetric $k \times k$ matrix and let H be a $q \times k$ matrix with $q < k$. Suppose that the eigenvalues of A are bounded below by $\delta > 0$ and above by $\Delta < \infty$ and that those of HH' are bounded below by δ^2 and above by Δ^2. Then there is a $k \times (k-q)$ matrix G with orthonormal columns such that $HG = 0$, the elements of

$$J = \begin{bmatrix} G'A \\ H \end{bmatrix}$$

are bounded above by $k\Delta$, and $|\det J| \geqslant \delta^{2k}$. $\square$

We can now state a result ensuring that $\sqrt{(n)}(\theta_n^* - \theta_n^o) = O(1)$, along with several other useful facts.

Lemma 7.3

Given assumptions DG, OP′, MX, SM, DM′, NE′, ID′, PD′, and CN:

(a) There exists $\Delta_o < \infty$ such that $|\theta_n^* - \theta_n| \leqslant \Delta_o |h_n^* - h_n^o|$.

If assumption DR also holds, then

(b) $\tilde\theta_n - \theta_n^* \to 0$ a.s.
(c) $\sqrt{(n)}(\theta_n^* - \theta_n^o) = O(1)$.
(d) $\sqrt{(n)}(\tilde\theta_n - \theta_n^o) = O_p(1)$. $\square$

The following result gives a formal statement of the fact that our previous results provide conditions which, when combined with those added here, suffice for assumption HT. Given this result we can proceed to consider the test statistics of interest.

Theorem 7.4

Suppose that any one of the following three sets of assumptions holds with $\Theta_n \equiv \{\theta \in \Theta: h(\theta) = h_n^o\}$, where h satisfies assumption CN and $\{h_n^o\}$ satisfies assumption DR:

(i) DG, OP′, MX′, SM, DM″, NE″, ID′, PD′, and $E(M_{nt}^o | F^{t-1}) = E(M_{nt}^* | F^{t-1}) = 0, n, t = 1, 2, \ldots$;
(ii) DG, OP′, MX′, SM DM″, NE″, ID′, PD′, and $E(M_{nt}^o M_{nt-\tau}^{o\prime}) = E(M_{nt}^* M_{nt-\tau}^{*\prime}) = 0$ for all $\tau > m \in \mathbb{N}, n, t = 1, 2, \ldots$;
(iii) DG, OP′, MX′, SM, DM″, NE‴, ID′, TL, WT, PD′, and $\{U_n^o\}$, $\{U_n^*\} O(1)$.

Then assumption HT holds. $\square$

The first test statistic considered is the *Wald test statistic* (Wald 1943). Letting $\hat h_n \equiv h(\hat\theta_n)$ and $\hat H_n \equiv H(\hat\theta_n)$, this is defined as

$$W_n \equiv n(\hat h_n - h_n^o)'[\hat H_n \hat A_n^{-1} \hat B_n \hat A_n^{-1} \hat H_n']^{-1}(\hat h_n - h_n^o).$$

As shown below, one rejects the hypothesis $H_o: h_n^* = h_n^o$ at the α level when W_n exceeds the $1-\alpha$ quantile of the chi-square distribution with q degrees of freedom. The principal advantage of the Wald test is that it requires only one unconstrained optimization to compute. The principal disadvantages are that for nonlinear hypotheses it is not invariant to reparameterization (see for example Gregory and Veall 1985), and its sampling distribution is not as well approximated by our characterizations as are the "likelihood ratio" and Lagrange multiplier test statistics (Gallant 1987). The lack of invariance means that two researchers with the same model, the same data, and the same nonlinear null hypothesis can obtain conflicting results because they happened to parameterize differently.

We have the following result.

Theorem 7.5

Let assumption HT hold. Then

$$W_n \sim Y_n + o_p(1)$$

where, letting $H_n^* \equiv H(\theta_n^*)$,

$$Y_n = Z_n'[H_n^* A_n^{*-1}(B_n^* + U_n^*)A_n^{*-1}H_n^{*\prime}]^{-1}Z_n$$

and

$$Z_n \sim N(\sqrt{(n)}(h_n^* - h_n^o), H_n^* A_n^{*-1} B_n^* A_n^{*-1} H_n^{*\prime}).$$

If $U_n^* = o(1)$ then Y_n is to within $o_p(1)$ distributed as the noncentral chi-square distribution with q degrees of freedom and noncentrality parameter

$$\alpha = (n/2)(h_n^* - h_n^o)'[H_n^* A_n^{*-1} B_n^* A_n^{*-1} H_n^{*\prime}]^{-1}(h_n^* - h_n^o).$$

Under the null hypothesis, $\alpha = 0$. □

In order to characterize the distribution of the "likelihood ratio" and Lagrange multiplier test statistics we need the following characterization of the score vector when evaluated at the constrained value θ_n^o.

Theorem 7.6

Let assumption HT hold. Then

$$\sqrt{(n)}\nabla_\theta Q_n^{o\prime} \sim X_n + o_{as}(1)$$

where

$$X_n \sim N(\sqrt{(n)}\nabla_\theta \bar{Q}_n^{o\prime}, B_n^o). \quad \square$$

Both the "likelihood ratio" test statistic and the Lagrange multiplier test statistic are effectively functions of the score vector evaluated at $\tilde{\theta}_n$. The following result gives an essential representation.

Theorem 7.7

Let assumption HT hold. Then

$$\begin{aligned}\sqrt{(n)}\nabla_\theta \tilde{Q}_n' &= H_n^{o\prime}[H_n^o A_n^{o-1} H_n^{o\prime}]^{-1} H_n^o A_n^{o-1} \sqrt{(n)}\nabla_\theta Q_n^{o\prime} + o_p(1)\\ &= O_p(1). \quad \square\end{aligned}$$

The second test statistic considered is the "*likelihood ratio*" *test statistic*

$$L_n = 2n[Q_n(\tilde{\theta}_n) - Q_n(\hat{\theta}_n)].$$

We enclose "likelihood ratio" in quotes because Q need not be the likelihood function. It is instead (as usual) the optimand which defines the extremum estimator. As we show below, one rejects the hypothesis $H_o{:}h_n^* = h_n^o$ at the α level when L_n exceeds the $1-\alpha$ quantile of the chi-square distribution with q degrees of freedom. The principal disadvantages of the "likelihood ratio" test are that it takes two minimizations to compute, and it requires that $B_n^o = A_n^o + o(1)$ to achieve its null case asymptotic distribution. In cases where this condition holds, there is some Monte Carlo evidence to indicate that the asymptotic approximation is quite accurate if degrees of freedom corrections are applied (Gallant 1987).

The result for the "likelihood ratio" test is as follows.

Theorem 7.8

Let assumption HT hold. Then

$$L_n \sim Y_n + o_p(1)$$

where

$$Y_n = Z_n' A_n^{o-1} H_n^{o\prime}[H_n^o A_n^{o-1} H_n^{o\prime}]^{-1} H_n^o A_n^{o-1} Z_n$$

and

$$Z_n \sim N(\sqrt{(n)}\nabla_\theta \bar{Q}_n^{o\prime}, B_n^o).$$

If $B_n^o = A_n^o + o(1)$ then Y_n is to within $o_p(1)$ distributed as the noncentral chi-square distribution with q degrees of freedom and noncentrality parameter

$$\alpha = (n/2)\nabla_\theta \bar{Q}_n^o A_n^{o\,-1} H_n^{o\prime}[H_n^o A_n^{o\,-1} H_n^{o\prime}]^{-1} H_n^o A_n^{o\,-1} \nabla_\theta \bar{Q}_n^{o\prime}.$$

Under the null hypothesis, $\alpha = 0$. □

The last test statistic considered is the *Lagrange multiplier test statistic* (Rao 1947; Aitchison and Silvey 1958)

$$LM_n = n\nabla_\theta \tilde{Q}_n \tilde{A}_n^{-1} \tilde{H}_n' [\tilde{H}_n \tilde{A}_n^{-1} \tilde{B}_n \tilde{A}_n^{-1} \tilde{H}_n']^{-1} \tilde{H}_n \tilde{A}_n^{-1} \nabla_\theta \tilde{Q}_n',$$

where we write $\nabla_\theta \tilde{Q}_n \equiv \nabla_\theta Q_n(\tilde{\theta}_n)$ and $\tilde{H}_n \equiv H(\tilde{\theta}_n)$. As we show below, one rejects the hypothesis $H_o: h_n^* = h_n^o$ at the α level when LM_n exceeds the $1-\alpha$ quantile of the chi-square distribution with q degrees of freedom. Using the saddle point conditions

$$\nabla_\theta L_n(\tilde{\theta}_n, \tilde{\lambda}_n) = \nabla_\theta[Q_n(\tilde{\theta}_n) + \tilde{\lambda}_n'(h(\tilde{\theta}_n) - h_n^o)] = 0$$

for the problem minimize $Q_n(\theta)$ subject to $h(\theta) = h_n^o$ where L_n denotes the Lagrangean for this problem, one can replace $\nabla_\theta Q_n(\tilde{\theta}_n)$ by $\tilde{\lambda}_n' \nabla_\theta h(\tilde{\theta}_n) = \tilde{\lambda}_n' H(\tilde{\theta}_n)$ in the expression for LM_n, whence the term Lagrange multiplier test; it is also called the efficient score test. Its principal advantage is that it requires only one constrained optimization for its computation. If the constraint $h(\theta) = h_n^o$ completely specifies $\tilde{\theta}_n$ or results in a linear model, this can be an overwhelming advantage. The test can have rather bizarre structural characteristics Suppose $h(\theta) = h_n^o$ completely specifies $\tilde{\theta}_n$. Then the test will accept any h_n^o for which $\tilde{\theta}_n$ is a local minimum, maximum, or saddle point of $Q_n(\theta)$ regardless of how large is $h - h_n^o$. Monte Carlo simulations suggest that the asymptotic approximation can be reasonably accurate (Gallant 1987).

Theorem 7.9

Let assumption HT hold. Then

$$LM_n \sim Y_n + o_p(1),$$

where

$$Y_n = Z_n' A_n^{o\,-1} H_n^{o\prime}[H_n^o A_n^{o\,-1}(B_n^o + U_n^o) A_n^{o\,-1} H_n^{o\prime}]^{-1} H_n^o A_n^{o\,-1} Z_n$$

and

$$Z_n \sim N(\sqrt{(n)}\nabla_\theta \bar{Q}_n^{o\prime}, B_n^o).$$

If $U_n^o = o(1)$ then Y_n is to within $o_p(1)$ distributed as the noncentral chi-square distribution with q degrees of freedom and noncentrality parameter

$$\alpha = (n/2)\nabla_\theta \bar{Q}_n^o A_n^{o\,-1} H_n^{o\prime}[H_n^o A_n^{o\,-1} B_n^o A_n^{o\,-1} H_n^{o\prime}]^{-1} H_n^o A_n^{o\,-1} \nabla_\theta \bar{Q}_n^{o\prime}.$$

Under the null hypothesis, $\alpha = 0$. □

If $U_n^o = U_n^* = o(1)$, the noncentrality parameters of the Wald and Lagrange multiplier test statistics are, respectively,

$$\alpha_W = (n/2)(h_n^* - h_n^o)'[H_n^* A_n^{*\,-1} B_n^* A_n^{*\,-1} H_n^{*\prime}]^{-1}(h_n^* - h_n^o)$$
$$\alpha_{LM} = (n/2)\nabla_\theta \bar{Q}_n^o A_n^{o\,-1} H_n^{o\prime}[H_n^o A_n^{o\,-1} B_n^o A_n^{o\,-1} H_n^{o\prime}]^{-1} H_n^o A_n^{o\,-1} \nabla_\theta \bar{Q}_n^{o\prime}.$$

By equicontinuity and the fact that $\sqrt{(n)}(\theta_n^* - \theta_n^o) = O(1)$, the matrices in brackets differ by terms of order $o(1)$. Further, by Taylor's theorem and familiar arguments

$$\sqrt{(n)}(h_n^* - h_n^o) = H_n^* \sqrt{(n)}(\theta_n^* - \theta_n^o) + o(1)$$
$$\sqrt{(n)}\nabla_\theta \bar{Q}_n^{o\prime} = A_n^o \sqrt{(n)}(\theta_n^* - \theta_n^o) + o(1)$$

so we can write

$$\begin{aligned}\alpha_W &= (n/2)(\theta_n^* - \theta_n^o)' H_n^{*\prime}[H_n^* A_n^{*\,-1} B_n^* A_n^{*\,-1} H_n^{*\prime}]^{-1} \\ &\quad \times H_n^*(\theta_n^* - \theta_n^o) + o(1) \\ \alpha_{LM} &= (n/2)(\theta_n^* - \theta_n^o)' H_n^{o\prime}[H_n^o A_n^{o\,-1} B_n^* A_n^{o\,-1} H_n^{o\prime}]^{-1} \\ &\quad \times H_n^o(\theta_n^* - \theta_n^o) + o(1) = \alpha_W + o(1).\end{aligned}$$

As α_W is $O(1)$ and the noncentral chi-square distribution is continuous in the noncentrality parameter we will have

$$\lim_{n \to \infty} |P(LM_n > x) - P(W_n > x)| = 0.$$

Thus, under the Pitman drift assumption, the Wald and Lagrange multiplier test have the same power in large samples – the same "local power", to use the standard parlance. It is obvious that the same sort of arguments will yield this result even when U_n^o and U_n^* are $O(1)$ instead of $o(1)$. If $A_n^o = B_n^o + o(1)$, then the same applies to the "likelihood ratio" test, as α_L and α_{LM} are identical in that instance.

One should not interpret this result to mean that the tests are equivalent in sample sizes ordinarily encountered in practice. As remarked earlier, the tests do have different structural characteristics, and Monte Carlo evidence suggests that the asymptotic approximations to the distribution of the "likelihood ratio" and Lagrange multiplier tests are more accurate than the Wald.

MATHEMATICAL APPENDIX

Proof of lemma 7.1

When $q = k$ the result is immediate, as the one-to-one mapping $\tau = h(\theta)$ has a Jacobian whose inverse has bounded elements.

Suppose $q < k$. Let $\varepsilon > 0$ be given. Given assumption HT(c) there exists $N_o \in \mathbb{N}$ such that

$$\xi = \inf_{n > N_o} \inf_{|\theta - \theta_n^*| > \varepsilon} |\bar{Q}_n(\theta) - \bar{Q}_n^*| > 0.$$

Let $\Psi(\rho, \tau)$ be as defined in assumption CN. Now $h_n^o = h(\theta_n^o)$ by definition, and put $\rho_n^o = r(\theta_n^o)$, $h_n^* = h(\theta_n^*)$, and $\rho_n^* = r(\theta_t^*)$. The image of a compact set is compact and the Cartesian product of two compact sets is compact, so $R \times T$ is compact. A continuous function on a compact set is uniformly continuous, so $|h_n^* - h_n^o| \to 0$ implies

$$\sup_R |\Psi(\rho, h_n^*) - \Psi(\rho, h_n^o)| \to 0.$$

In particular, putting $\theta_n^\# = \Psi(\rho_n^*, h_n^o)$ we have

$$|\theta_n^* - \theta_n^\#| \to 0.$$

By definition of S, it follows that for all $n > N_1$, say, $\theta_n^\#$ is in S. Because $\{\bar{Q}_n(\theta)\}$ is continuous uniformly in n given assumption HT(b), there exists N_2 such that $|\theta - \theta_n^*| < \eta$ implies

$$|\bar{Q}_n(\theta) - \bar{Q}_n(\theta_n^*)| < \xi$$

for all $n > N_2$. Choose N_3 large enough that $|\theta_n^* - \theta_n^\#| < \eta$ for all $n > N_3$. Because $h(\theta_n^\#) = h_n^o$, we must have $\bar{Q}_n(\theta_n^o) \leqslant \bar{Q}_n(\theta_n^\#)$ for $n > N_1$, as θ_n^o minimizes $\bar{Q}_n$ on Θ_n. For $n > \max(N_o, N_1, N_2, N_3)$ we have

$$\bar{Q}_n(\theta_n^o) \leqslant \bar{Q}_n(\theta_n^\#) < \bar{Q}_n(\theta_n^*) + \xi.$$

This implies $|\bar{Q}_n(\theta_n^o) - \bar{Q}_n(\theta_n^*)| < \xi$, so that $|\theta_n^o - \theta_n^*| < \varepsilon$. Because ε is arbitrary, $\theta_n^o - \theta_n^* \to 0$. □

Proof of lemma 7.2

Let

$$H = USV_1'$$

be the singular value decomposition (Lawson and Hanson 1974, chapter 4) of H, where S is a diagonal matrix of order q with positive entries on the diagonal, V_1' is of order q by k, and $U'U = UU' = V_1'V_1 = I_q$. From $HH' = US^2U'$ we see that $\delta \leqslant s_{ii}^2 \leqslant \Delta$. Choose V_2' of order $k-q$ by k such that

$$V = \begin{bmatrix} V_1' \\ V_2' \end{bmatrix}$$

satisfies

$$I_k = V'V = V_1V_1' + V_2V_2' = \begin{bmatrix} V_1'V_1 & V_1'V_2 \\ V_2'V_1 & V_2'V_2 \end{bmatrix} = \begin{bmatrix} I_q & 0 \\ 0 & I_{k-q} \end{bmatrix}.$$

Put $G' = V_2'$, note that $HG = 0$ and consider

$$\begin{aligned} JJ' &= \begin{bmatrix} V_2'A \\ USV_1' \end{bmatrix} [AV_2 \quad V_1SU'] \\ &= \begin{bmatrix} V_2'AAV_2 & V_2'AV_1SU' \\ USV_1'AV_2 & US^2U' \end{bmatrix} \\ &= \begin{bmatrix} V_2' & 0 \\ 0 & U \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & S \end{bmatrix} \begin{bmatrix} V_1 & V_2 \\ I_q & 0 \end{bmatrix} \begin{bmatrix} V_1' & I_q \\ V_2' & 0 \end{bmatrix} \begin{bmatrix} A & 0 \\ 0 & S \end{bmatrix} \begin{bmatrix} V_2 & 0 \\ 0 & U' \end{bmatrix} \\ &= BCDD'C'B'. \end{aligned}$$

The elements of B and D are bounded by one so we must have that each element of BCD is bounded above by $k\Delta$. Then each element of JJ' is bounded above by $k^2\Delta^2$. Since a diagonal element of JJ' has the form $\Sigma_i J_{ij}^2$ we must have $|J_{ij}| \leqslant k\Delta_1$. Now (Mood and Graybill 1963, p. 206)

$$\begin{aligned} \det JJ' &= \det(US^2U') \det |V_2'AAV_2 - V_2'AV_1SU'(US^2U')^{-1}USV_1'AV_2| \\ &= \det S^2 \det(V_2'AAV_2 - V_2'AV_1V_1'AV_2) \\ &= \det S^2 \det(V_2'AV_2V_2'AV_2) \\ &\geqslant \delta^{2k} \det^2(V_2AV_2). \end{aligned}$$

But

$$\delta x'x = \delta x'V_2'V_2x \leqslant x'V_2'AV_2x$$

whence $\delta^k \leqslant \det V_2'AV_2$ and

$$\delta^{4k} \leqslant \det J'J = \det^2 J. \qquad \square$$

Proof of lemma 7.3(*a*)

The first order conditions for the problem minimize $\bar{Q}_n(\theta)$ subject to $h(\theta) = h_n^o$ are

$$\nabla_\theta \bar{Q}_n^o + \lambda_n' H_n^o = 0$$
$$h(\theta_n^o) = h_n^o.$$

By Taylor's theorem we have

$$\nabla_\theta \bar{Q}_n^{*\prime} = \nabla_\theta \bar{Q}_n^{o\prime} + [\nabla_\theta^2 \bar{Q}_n^o + o(1)](\theta_n^* - \theta_n^o)$$
$$h_n^* - h_n^o = h(\theta_n^o) - h_n^o + [H_n^o + o(1)](\theta_n^* - \theta_n^o).$$

Using $\nabla_\theta \bar{Q}_n(\theta_n^*) = 0$, $h(\theta_n^o) = h_n^o$, we have upon substitution into the first order conditions

$$[\nabla_\theta^2 \bar{Q}_n^o + o(1)](\theta_n^* - \theta_n^o) = -H_n^{o\prime}\lambda_n^o$$
$$[H_n^o + o(1)](\theta_n^* - \theta_n^o) = h_n^* - h_n^o.$$

Let G_n^* be the matrix given by lemma 7.2 with orthonormal columns, $H_n^*G_n^* = 0$, $0 < \delta^{2k} \leqslant \det J_n^*$, and $\max_{ij}|J_{ijn}^*| \leqslant k\Delta < \infty$ where

$$J_n^* = \begin{bmatrix} G_n^{*\prime} A_n^* \\ H_n^* \end{bmatrix}.$$

Let J_{ij} denote the elements of a matrix J and consider the region

$$\{J_{ij}: 0 < \delta^{2k} - \varepsilon \leqslant \det J, |J_{ij}| \leqslant k\Delta + \varepsilon\}.$$

On this region we must have $|J^{ij}| \leqslant \Delta_o < \infty$ where J^{ij} denotes an element of J^{-1}. For large n the matrix J_n^* is in this region by lemma 7.2, as is the matrix

$$J_n = \begin{bmatrix} G_n^{*\prime}[A_n^* + o(1)] \\ H_n^* + o(1) \end{bmatrix},$$

because the elements of G_n^* are bounded by one. In consequence we

have

$$(\theta_n^* - \theta_n^o) = J_n^{-1}\begin{bmatrix} 0 \\ h_n^* - h_n^o \end{bmatrix},$$

where the elements of J_n^{-1} are bounded above by Δ_o for all n larger than some $N \in \mathbb{N}$. Thus we have $|\theta_n^* - \theta_n^o| \leqslant \Delta_o|h_n^* - h_n^o|$ for large n.

Proof of lemma 7.3(*b*)

By the triangle inequality

$$|\tilde{\theta}_n - \theta_n^*| \leqslant |\tilde{\theta}_n - \theta_n^o| + |\theta_n^o - \theta_n^*|$$
$$\leqslant |\tilde{\theta}_n - \theta_n^o| + \Delta_o|h_n^* - h_n^o|.$$

The first term converges to zero a.s. given assumptions DE, OP′, MX, SM(i), DM′, NE′, and ID′, and the second term converges to zero given assumptions DG, OP′, MX, SM(i), DM′, NE′, ID′, PD′, CN, and DR by lemma 7.3(a).

Proof of lemma 7.3(*c*)

Because $\sqrt{(n)}|\theta_n^* - \theta_n^o| \leqslant \Delta_o\sqrt{(n)}|h_n^* - h_n^o|$ by lemma 7.3(a) given assumptions DG, OP′, MX SM(i), DM′, NE′, ID′, PD′, CN, and DR, and because $\sqrt{(n)}|h_n^* - h_n^o|$ is $O(1)$ by assumption DR, it follows immediately that $\sqrt{(n)}(\theta_n^* - \theta_n^o) = O(1)$.

Proof of lemma 7.3(*d*)

By argument identical to that in part (a), we obtain

$$|\tilde{\theta}_n - \hat{\theta}_n| \leqslant \Delta_o|h(\hat{\theta}_n) - h_n^o| \quad \text{a.a. } n \quad \text{a.s.,}$$

where $Q_n(\theta)$ replaces $\bar{Q}_n(\theta)$, $\tilde{\theta}_n$ replaces θ_n^o, and $\hat{\theta}_n$ replaces θ_n^*; $A_n^* + o_{as}(1)$ replaces $A_n^* + o(1)$, $H_n^* + o_{as}(1)$ replaces $H_n^* + o(1)$, and the elements of J_n^{-1} are bounded above by Δ_o almost surely. By Taylor's theorem

$$h(\hat{\theta}_n) = h(\theta_n^*) + [H_n^* + o_{as}(1)](\hat{\theta}_n - \theta_n^*).$$

Therefore

$$\sqrt{(n)}(h(\hat{\theta}_n) - h_n^o) = \sqrt{(n)}(h_n^* - h_n^o) + [H_n^* + o_{as}(1)]\sqrt{(n)}(\hat{\theta}_n - \theta_n^*)$$
$$= O_p(1).$$

The first term is $O(1)$ given assumption DR, while the second term is $O_p(1)$ given that H_n^* is $O(1)$ by assumption CN and $\sqrt{(n)}(\hat{\theta}_n-\theta_n^*)$ is $O_p(1)$. That $\sqrt{(n)}(\hat{\theta}_n-\theta_n^*)$ is $O_p(1)$ follows because

$$\sqrt{(n)}(\hat{\theta}_n-\theta_n^*) = -A_n^{*-1}\sqrt{(n)}\nabla_\theta Q_n^{*'}+o_p(1)$$

under the conditions given, where $\{A_n^{*-1}\}$ is $O(1)$ and $\sqrt{(n)}\nabla_\theta Q_n^{*'} = O_p(1)$ as a consequence of McLeish's inequality, theorem 3.11 (see the proof of theorem 5.4). Thus

$$\sqrt{(n)}(\tilde{\theta}_n-\hat{\theta}_n) = O_p(1).$$

It follows that

$$\sqrt{(n)}(\tilde{\theta}_n-\theta_n^o) = \sqrt{(n)}(\tilde{\theta}_n-\hat{\theta}_n)+\sqrt{(n)}(\hat{\theta}_n-\theta_n^*)+\sqrt{(n)}(\theta_n^*-\theta_n^o)$$
$$= O_p(1)$$

because $\sqrt{(n)}(\tilde{\theta}_n-\hat{\theta}_n)$ is $O_p(1)$ as just established, $\sqrt{(n)}(\hat{\theta}_n-\theta_n^*)$ is $O_p(1)$ given assumptions DG, OP′, MX, SM, DM′, NE′, ID′, and PD as previously argued, and $\sqrt{(n)}(\theta_n^*-\theta_n^o)$ is $O(1)$ as established in lemma 7.3(c). □

Proof of theorem 7.4

We give the argument for conditions (iii) only. The result for conditions (i) follows using theorem 6.4 instead of theorem 6.9 and the result for conditions (ii) follows using theorem 6.5 instead of theorem 6.9.

Assumption HT(a) holds given assumptions DG and OP′ as established in the proof of theorem 2.2. Assumption HT(b) holds given assumptions DG, OP′, MX′, SM, DM″, NE‴, and MX′, since these conditions imply those of theorem 5.6. Assumption HT(c) is satisfied given assumption ID′. Similarly, assumption HT(d) holds as a consequence of lemma 7.3(c). Assumption HT(e) holds by the conclusion in theorem 5.4 and corollary 5.5. Assumption HT(f) holds by the conclusion of theorem 6.9. Assumption HT(g) holds given assumption PD′. Finally, assumption HT(h) holds by the conclusion of theorem 6.1. □

Proof of theorem 7.5

We may assume without loss of generality that $\hat{\theta}_n$ and θ_n^* are in S. Conditions HT(a), (b), and (c) ensure that $\nabla_\theta Q_n(\hat{\theta}_n) = o_p(n^{-1/2})$ and $\nabla_\theta \bar{Q}_n^* = o(n^{-1/2})$. By Taylor's theorem

$$\sqrt{(n)}[h_i(\hat{\theta}_n)-h_i(\theta_n^*)] = \nabla_\theta h_i(\bar{\theta}_n^i)\sqrt{(n)}(\hat{\theta}_n-\theta_n^*)$$

for $i = 1,2,\ldots,q$ and $\bar{\theta}_n^i$ on the line segment joining $\hat{\theta}_n$ to θ_n^*. By the almost sure convergence of $\theta_n^*-\hat{\theta}_n$ to zero, $\bar{\theta}_n^i-\theta_n^*\to 0$ almost surely, whence $\nabla_\theta h_i(\bar{\theta}_n^i)-\nabla_\theta h_i(\theta_n^*)\to 0$ almost surely. Thus we may write

$$\sqrt{(n)}(\hat{h}_n-h_n^*) = [H_n^*+o(1)]\sqrt{(n)}(\hat{\theta}_n-\theta_n^*).$$

By lemma 3 of Jennrich (1969) we have

$$\sqrt{(n)}B_n^{*-1/2}\nabla_\theta Q_n^{*'} = \sqrt{(n)}B_n^{*-1/2}\nabla_\theta \hat{Q}_n'$$
$$+B_n^{*-1/2}[A_n^*+o_{as}(1)]\sqrt{(n)}(\hat{\theta}_n-\theta_n^*)$$
$$=o_p(1)+B_n^{*-1/2}[A_n^*+o_{as}(1)]\sqrt{(n)}(\hat{\theta}_n-\theta_n^*).$$

Now $B_n^{*-1/2} = O(1)$ and $A_n^{*-1} = O(1)$, whence $\sqrt{(n)}(\hat{\theta}_n-\theta_n^*)$ is bounded in probability and

$$\sqrt{(n)}\nabla_\theta Q_n^{*'} = A_n^*\sqrt{(n)}(\hat{\theta}_n-\theta_n^*)+o_p(1).$$

Combining these two equations we have

$$\sqrt{(n)}(\hat{h}_n-h_n^o) = \sqrt{(n)}(h_n^*-h_n^o)$$
$$+H_n^*\sqrt{(n)}(\hat{\theta}_n-\theta_n^*)+o_p(1)$$
$$= \sqrt{(n)}(h_n^*-h_n^o)+H_n^*A_n^{*-1}B_n^{*1/2}B_n^{*-1/2}$$
$$\times\sqrt{(n)}\nabla_\theta Q_n^{*'}+o_p(1).$$

This equation implies that the left hand side is bounded in probability, and because

$$\hat{H}_n\hat{A}_n^{-1}\hat{B}_n\hat{A}_n^{-1}\hat{H}_n' = H_n^*A_n^{*-1}(B_n^*+U_n^*)A_n^{*-1}H_n^{*'}+o_{as}(1)$$

we have that

$$W_n = n(\hat{h}_n-h_n^o)'[H_n^*A_n^{*-1}(B_n^*+U_n^*)A_n^{*-1}H_n^{*'}]^{-1}(\hat{h}_n-h_n^o)+o_p(1).$$

By the Skorohod representation theorem (Serfling 1980, section 1.6), there are random variables X_n with the same distribution as $B_n^{*-1/2}\sqrt{(n)}\nabla_\theta Q_n^{*'}$ such that $X_n = X+o_{as}(1)$ where $X\sim N(0,I_k)$. Then

$$\sqrt{(n)}(\hat{h}_n-h_n^o)\sim\sqrt{(n)}(h_n^*-h_n^o)+H_n^*A_n^{*-1}B_n^{*-1/2}X_n$$
$$=\sqrt{(n)}(h_n^*-h_n^o)+H_n^*A_n^{*-1}B_n^{*-1/2}X+o_{as}(1).$$

Let

$$Z_n = \sqrt{(n)}(h_n^* - h_n^o) + H_n^* A_n^{*-1} B_n^{*-1/2} X,$$

and the result follows. □

Proof of theorem 7.6

Given assumption HT(e) we have

$$\sqrt{(n)} B_n^{o\,-1/2} (\nabla_\theta Q_n^o - \nabla_\theta \bar{Q}_n^o)' \xrightarrow{L} N(0, I_k).$$

Now $\sqrt{(n)}\nabla_\theta \bar{Q}_n^o = O(1)$ because by Taylor's theorem

$$\begin{aligned}\sqrt{(n)}\nabla_\theta \bar{Q}_n^{o\prime} &= \sqrt{(n)}\nabla_\theta \bar{Q}_n^{*\prime} + [A_n^* + o(1)]\sqrt{(n)}(\theta_n^o - \theta_n^*)\\ &= A_n^* \sqrt{(n)}(\theta_n^o - \theta_n^*) + o(1)\end{aligned}$$

and $A_n^* = O(1)$ by assumption HT(g) and $\sqrt{(n)}(\theta_n^o - \theta_n^*) = O(1)$ by assumption HT(d). By the Skorohod representation theorem (Serfling 1980, section 1.6) there are random variables Y_n with the same distribution as $\sqrt{(n)} B_n^{o\,-1/2} \nabla_\theta Q_n^{o\prime}$ such that $Y_n - \sqrt{(n)} B_n^{o\,-1/2} \nabla_\theta \bar{Q}_n^{o\prime} = Y + o_{as}(1)$ where $Y \sim N(0, I_k)$. Let

$$X_n = B_n^{o\,1/2} Y + \sqrt{(n)}\nabla_\theta \bar{Q}_n^{o\prime}$$

whence

$$X_n \sim N[\sqrt{(n)}\nabla_\theta \bar{Q}_n^{o\prime}, B_n^o].$$

As $B_n^{o\,1/2}$ is bounded, $B_n^{o\,1/2} o_{as}(1) = o_{as}(1)$ and the result follows. □

Proof of theorem 7.7

By Taylor's theorem and the continuity of $\{\nabla_\theta^2 \bar{Q}_n(\theta)\}$ ard $H(\theta)$ uniformly in n on the compact set S we have

$$\begin{aligned}\sqrt{(n)}\nabla_\theta \tilde{Q}_n' &= \sqrt{(n)}\nabla_\theta Q_n^{o\prime} + [A_n^o + o_{as}(1)]\sqrt{(n)}(\tilde{\theta}_n - \theta_n^o)\\ \sqrt{(n)}\tilde{h}_n &= \sqrt{(n)} h(\theta_n^o) + [H_n^o + o_{as}(1)]\sqrt{(n)}(\tilde{\theta}_n - \theta_n^o).\end{aligned}$$

Recalling that $\sqrt{(n)}\tilde{h}_n - \sqrt{(n)} h(\theta_n^o) \equiv 0$ and the elements of A_n^o are bounded we have for large n that

$$\begin{aligned}\sqrt{(n)}(\tilde{\theta}_n - \theta_n^o) &= [A_n^o + o_{as}(1)]^{-1}\sqrt{(n)}\nabla_\theta \tilde{Q}_n'\\ &\quad - [A_n^o + o_{as}(1)]^{-1}\sqrt{(n)}\nabla_\theta \bar{Q}_n^{o\prime} + c_{as}(1)\end{aligned}$$

and

$$[H_n^o + o_{as}(1)]\sqrt{(n)}(\tilde{\theta}_n - \theta_n^o) = o_{as}(1)$$

whence

$$\begin{aligned}&[H_n^o + o_{as}(1)][A_n^o + o_{as}(1)]^{-1}\sqrt{(n)}\nabla_\theta \tilde{Q}_n'\\ &\quad = [H_n^o + o_{as}(1)][A_n^o + o_{as}(1)]^{-1}\sqrt{(n)}\nabla_\theta \bar{Q}_n^{o\prime} + o_{as}(1).\end{aligned}$$

Now $\sqrt{(n)} B_n^{o\,-1/2}[\nabla_\theta Q_n^o - \nabla_\theta \bar{Q}_n^o]'$ converges in distribution and by Taylor's theorem (Jernrich 1969, lemma 3)

$$\begin{aligned}\sqrt{(n)} B_n^{o\,-1/2}\nabla_\theta \bar{Q}_n^{o\prime} &= \sqrt{(n)} B_n^{o\,-1/2}\nabla_\theta \bar{Q}_n^{*\prime}\\ &\quad + \sqrt{(n)} B_n^{o\,-1/2}[A_n^* + o_{as}(1)](\theta_n^o - \theta_n^*)\\ &= o_{as}(1) + O(1).\end{aligned}$$

This implies that $\sqrt{(n)} B_n^{o\,-1/2}\nabla_\theta \bar{Q}_n^{o\prime}$ is $O(1)$. Because $B_n^{o\,-1/2}$ is $O(1)$, $\sqrt{(n)}\nabla_\theta Q_n^{o\prime}$ is bounded in probability given assumption HT(e).

There is a sequence of Lagrange multipliers $\tilde{\lambda}_n$ such that

$$\sqrt{(n)}\nabla_\theta \tilde{Q}_n' + \tilde{H}_n' \tilde{\lambda}_n = o_{as}(1).$$

By continuity of $H(\theta)$ and the fact that $\tilde{\theta}_n - \theta_n^o \to 0$ a.s. we have $\tilde{H}_n = H_n^o + o_{as}(1)$. Recalling that $\sqrt{(n)}\nabla_\theta \tilde{Q}_n'$ is $O_p(1)$ we have for large n that

$$\begin{aligned}&H_n^{o\prime}[H_n^o A_n^{o\,-1} H_n^{o\prime}]^{-1} H_n^o A_n^o \sqrt{(n)}\nabla_\theta Q_n^{o\prime}\\ &\quad = \tilde{H}_n'[(H_n^o + o_{as}(1))[A_n^o + o_{as}(1)]^{-1}\tilde{H}_n']^{-1}\\ &\qquad \times [H_i' + o_{as}(1)][A_n^o + o_{as}(1)]^{-1}\sqrt{(n)}\nabla_\theta Q_n^{o\prime} + o_p(1)\\ &\quad = \tilde{H}_n'[(H_n^o + o_{as}(1))[A_n^o + o_{as}(1)]^{-1}\tilde{H}_n']^{-1}\\ &\qquad \times [H_i' + o_{as}(1)][A_n^o + o_{as}(1)]^{-1}\sqrt{(n)}\nabla_\theta \tilde{Q}_n' + o_p(1)\\ &\quad = -\tilde{H}_n'[(H_n^o + o_{as}(1))[A_n^o + o_{as}(1)]^{-1}\tilde{H}_n']^{-1}\\ &\qquad \times [H_i' + o_{as}(1)][A_n^o + o_{as}(1)]^{-1}\tilde{H}_n'\tilde{\lambda}_n + o_p(1)\\ &\quad = -\tilde{H}_n'\tilde{\lambda}_n + o_p(1)\\ &\quad = \sqrt{(n)}\nabla_\theta \tilde{Q}_n' + o_p(1). \qquad \square\end{aligned}$$

Proof of theorem 7.8

By the mean value theorem (Jennrich 1969, lemma 3)

$$\begin{aligned}2n[Q_n(\tilde{\theta}_n) - Q_n(\hat{\theta}_n)] &= 2n\nabla_\theta Q(\hat{\theta}_n)(\tilde{\theta}_n - \hat{\theta}_n)\\ &\quad + n(\tilde{\theta}_n - \hat{\theta}_n)'\nabla_\theta^2 Q_n(\bar{\theta}_n)(\tilde{\theta}_n - \hat{\theta}_n)\end{aligned}$$

where $\bar{\theta}_n$ lies on the segment connecting $\hat{\theta}_n$ and $\tilde{\theta}_n$. Because $\{\nabla^2_\theta \bar{Q}_n(\theta)\}$ is continuous on Θ uniformly in n, $\bar{\theta}_n - \theta^\circ_n = o_{as}(1)$, $\hat{\theta}_n - \theta^*_n = o_{as}(1)$, and $\theta^*_n - \theta^\circ_n = o(1)$, we have

$$\nabla^2_\theta Q_n(\bar{\theta}_n) = \nabla^2_\theta \bar{Q}_n(\theta^\circ_n) + o_{as}(1).$$

Also, $n\nabla_\theta Q(\hat{\theta}_n) = o_{as}(1)$ so that

$$2n[\tilde{Q}_n - \hat{Q}_n] = n(\tilde{\theta}_n - \hat{\theta}_n)'[A^\circ_n + o_s(1)](\tilde{\theta}_n - \hat{\theta}_n) + o_{as}(1).$$

Again by the mean value theorem

$$[A^\circ_n + o_{as}(1)]\sqrt{(n)}(\tilde{\theta}_n - \hat{\theta}_n) = \sqrt{(n)}\nabla_\theta \tilde{Q}'_n$$

whence using the same type of argument as in theorem 7.7,

$$\begin{aligned}\sqrt{(n)}(\tilde{\theta}_n - \hat{\theta}_n) &= [A^\circ_n + o_{as}(1)]^{-1}[A^\circ_n + o_{as}(1)]\sqrt{(n)}(\tilde{\theta}_n - \hat{\theta}_n) + o_{as}(1)\\ &= [A^\circ_n + o_{as}(1)]^{-1}\sqrt{(n)}\nabla_\theta \tilde{Q}'_n + o_{as}(1),\end{aligned}$$

which is $O_p(1)$ by the argument of theorem 7.7. Thus

$$2n[\tilde{Q}_n - \hat{Q}_n] = n(\tilde{\theta}_n - \hat{\theta}_n)' A^\circ_n(\tilde{\theta}_n - \hat{\theta}_n) + o_p(1)$$

$$\sqrt{(n)}(\tilde{\theta}_n - \hat{\theta}_n) = A^{\circ -1}_n \sqrt{(n)}\nabla_\theta \tilde{Q}'_n + o_p(1)$$

whence

$$2n[\tilde{Q}_n - \hat{Q}_n] = n\nabla_\theta \tilde{Q}_n A^{\circ -1}_n \nabla_\theta \tilde{Q}'_n + o_p(1)$$

and the distributional results follow at once from theorems 7.6 and 7.7. □

Proof of theorem 7.9

Using arguments that by now are routine we have

$$\begin{aligned}LM_n = n\nabla_\theta Q^\circ_n A^{\circ -1}_n H^{\circ\prime}_n [H^\circ_n A^{\circ -1}_n (B^\circ_n + U^\circ_n) A^{\circ -1}_n H^{\circ\prime}_n]^{-1}\\ \times H^\circ_n A^{\circ -1}_n \nabla_\theta Q^{\circ\prime}_n + o_p(1).\end{aligned}$$

The result follows from theorems 7.6 and 7.7. □

REFERENCES

Aitchison, J. and S. D. Silvey 1958: Maximum likelihood estimation of parameters subject to restraints, *Annals of Mathematical Statistics* 29, 813–28.

Gallant, A. R 1987: *Nonlinear Statistical Models*. New York: John Wiley and Sons.

Gregory, A. W. and M. R. Veal 1985: Formulating Wald tests of nonlinear restrictions, *Econometrica* 53, 1465–8.

Jennrich, R. I. 1969: Asymptotic properties of non-linear least squares estimators, *Annals of Mathematical Statistics* 40, 633–43.

Lawson, C. and R. Hanson 1974: *Solving Least Squares Problems*. Englewood Cliffs: Prentice-Hall.

Mood, A. and F. Graybill 1963: *Introduction to the Theory of Statistics*, 2nd edn. New York: McGraw-Hill.

Rao, C. R. 1947: Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation, *Proceedings of the Cambridge Philosophical Society* 44, 50–7.

Serfling, R. 1980: *Approximation Theorems of Mathematical Statistics*. New York: John Wiley and Sons.

Wald, A. 1943: Tests of statistical hypotheses concerning several parameters when the number of observations is large, *Transactions of the American Mathematical Society* 54, 426–82.