COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 1060R

HYPOTHESIS TESTING WITH A RESTRICTED PARAMETER SPACE

Donald W. K. Andrews

June 1994

HYPOTHESIS TESTING

WITH A RESTRICTED PARAMETER SPACE

Donald W. K. Andrews

Cowles Foundation for Research in Economics

Yale University

April 1993

Revised: June 1994

ABSTRACT

This paper considers hypothesis tests when the parameter space is restricted under the alternative hypothesis. Multivariate one-sided tests are a leading example. Optimal tests, called *directed* tests, are derived using a weighted average power criterion. The likelihood ratio test is shown to be admissible and to maximize power against alternatives that are arbitrarily distant from the null hypothesis. Exact results are established first for Gaussian linear regression models with known variance. Asymptotic analogues are then established for dynamic nonlinear models.

Simulation is used to compare the tests discussed in the paper. The D–W_∞ directed test is found to perform best in an overall sense for multivariate one-sided alternatives. The D–W_∞ and LR tests are found to perform likewise for mixed one- and two-sided alternatives.

INTRODUCTION

This paper considers tests of hypotheses in parametric models in which the parameter of interest, β, is restricted under the alternative hypothesis. More specifically, we consider tests of $\mathrm{H}_0 : \beta = 0$ versus $\mathrm{H}_1^* : \beta \in B$, where $\beta \in R^p$ is a subvector of an unknown parameter $\theta \in R^s$ and B is a subset of R^p that does not include a neighborhood of zero. Leading examples of hypotheses of this form include multivariate one-sided hypotheses, joint one- and two-sided hypotheses, and multivariate non-negativity hypotheses (of the form $\mathrm{H}_1^* : \beta \not\geq 0$). Econometric applications of hypotheses of these types have been noted in the literature, e.g., see Gourieroux, Holly, and Monfort (1982) and Gourieroux and Monfort (1989). Examples include (1) applications where the signs of regression parameters are known (such as the coefficients on own price and income in demand analysis), (2) tests of positivity of the variances of random coefficients, (3) tests of positivity of the variances of error components in panel data models with individual and time error components, (4) tests of positive definiteness of variance matrices, (5) tests of skewness and excess kurtosis, (6) tests of positive first-order serial correlation plus ARCH effects, and (7) tests of ordered alternatives (of the form $\mathrm{H}_1^* : \mu_1 < \mu_2 < \mu_3$, where $\beta = (\mu_2 - \mu_1,\ \mu_3 - \mu_2)'$).

Hypotheses of the above sort are non-standard when $p \geq 2$. In consequence, the likelihood ratio (LR) statistic does not have its usual chi-square asymptotic distribution nor does it possess its usual optimality properties of the type established by Wald (1942, 1943). Nevertheless, most of the literature on this subject (see references below) has considered the LR test or asymptotically equivalent tests and has focussed its attention on obtaining suitable critical values for such tests.

In contrast, the focus of this paper is on the choice and optimality properties of test statistics. We specify a weighted average power optimality criterion and derive tests that are optimal or asymptotically optimal (depending on the model) according to this criterion. This approach is similar to that used by Andrews and Ploberger (1994) for a different testing problem. The weight function employed is a truncated multivariate normal density truncated to be zero when $\beta \notin B$. The contours of this density are the ellipses that Wald (1942, 1943) considered in his analysis of the optimal (asymptotic) weighted average power of Wald and LR tests (for the case where

B contains a neighborhood of zero).[2] The optimal test statistic is found to be a *directed* Wald statistic (or asymptotically equivalent directed LR or Lagrange multiplier (LM) statistic).

The directed Wald statistic equals the standard Wald statistic for testing $H_0 : \beta = 0$ against the unrestricted alternative $H_1 : \beta \neq 0$ multiplied by a weighting factor that depends on the location of the unrestricted maximum likelihood (ML) estimator $\widehat{\beta}$ relative to the restricted parameter space B. The directed LR and LM statistics are defined analogously. None require the computation of the ML estimator for the restricted parameter space B, which can be difficult to compute and is required by the standard LR test and asymptotically equivalent versions of it. For the case of a univariate one-sided test, the directed Wald statistic reduces to the standard one-sided Wald test, which has known optimality properties. The asymptotic null distribution of the directed statistics is shown to be a function of a multivariate normal random variable (rv). Critical values and p–values can be obtained by simulation. An interactive GAUSS program, written by E. Fiuza and the author, that does this is available from the author.

The weighted average power optimality criterion that we consider depends on a scalar parameter c that indexes whether more or less weight is placed on alternatives that are close to the null hypothesis. For each value of c, one obtains a different optimal directed test statistic. Thus, it is important to know whether the directed tests have power that is sensitive to c and to have guidance regarding the choice of c. These issues are discussed below.

This paper also analyzes the power properties of the most widely used test statistics in the literature, viz., the LR test and various asymptotically equivalent Wald, Kuhn–Tucker (KT), and modified LM tests. It is shown that the LR test is admissible or asymptotically admissible (depending on the model) when B is a positively homogeneous set (i.e., $\beta \in B$ implies $\tau\beta \in B$ $\forall \tau > 0$). The latter condition is satisfied in most examples. Furthermore, it is shown that the LR test is the best test in a certain sense against alternatives in B that are arbitrarily far from the null hypothesis. This result is useful in terms of understanding the properties of the LR test, but is only a weak optimality result, because all good tests have high power against such alternatives.

We note that the admissibility of the LR test in the context considered here is not an obvious

result. The LR test is known to be inadmissible in several closely related testing problems concerning the means of normal rv's, e.g., see Berger and Sinclair (1984), Warrack and Robertson (1984), and Berger (1989). In addition, the inadmissibility of the ML estimator for simple ordered normal means (see Sackrowitz and Strawderman (1974)) has led Robertson, Wright, and Dykstra (1988, p. 109) to (incorrectly) anticipate that the LR test is not admissible in the present context.

For clarity of presentation of the results and their proofs, exact versions of the results outlined above are derived first for linear regression models with Gaussian errors and known error variances. Then, analogous asymptotic results are provided for parametric models that satisfy standard ML regularity conditions.

The theoretical results outlined above are supplemented in the paper by simulation results. The simulations compare the power of the LR test, several directed Wald tests, viz., D–W_c for $c = 0$, $1/3$, 1, 3, and ∞, the two-sided Wald test, and the power envelope. The model considered is the multivariate normal location model with known variance. For this model, the theoretical results hold exactly. Furthermore, the asymptotic local power of the above tests for nonlinear models equals their exact power for the model above. In consequence, the finite sample power comparisons given in the paper also provide asymptotic local power comparisons for a wide variety of models.

The simulation results can be summarized as follows: We find the D–W_∞ test to be the best overall test for multivariate one-sided alternatives. For mixed one- and two-sided alternatives, the D–W_∞ and LR tests are best with the D–W_∞ test doing better in the middle of the parameter space and the LR test doing better along the edges. The D–W_0 test does very poorly for mixed one- and two-sided alternatives, for reasons given below. The D–$W_{1/3}$, D–W_1, and D–W_3 tests have similar power and are almost as good as D–W_∞. All of the above tests usually have much higher power than the standard two-sided Wald test, which ignores the restrictions on the parameter space. The D–W_∞ test is often near the power envelope for alternatives in the middle of the parameter space, but not for alternatives on the edge of the parameter space. Finally, the relative performances of the tests do not vary greatly as the distance of the alternative from the null is

varied, at least across the range for which the LR test has power between .3 and .9.

The fact that the power of the directed Wald tests does not vary greatly with c, provided $c \neq 0$, has useful consequences. First, it implies that the choice of c is not crucial. The choice of $c = \infty$ does very well in all cases considered. Second, it implies that the power of a directed test for a given value of c is nearly optimal for a wide range of weight functions. In consequence, directed test statistics (with $c \neq 0$) have nearly the same optimality properties as classical Wald, LM, and LR tests for standard two-sided alternative hypotheses.

The optimality properties of the directed tests and the classical tests can be criticized for using weight functions whose contours are arbitrary. In fact, the contours are not arbitrary. They are chosen to deliver a computationally tractable test statistic. Given that it is hard to argue in favor of one particular shape of contour over another on a priori grounds for a general class of testing problems, and given that an applied researcher has to choose *some* test, it seems prudent to choose contours that ease the computational burden as much as possible. This is what has been done in this paper and in Wald (1942, 1943).

We now briefly review the literature concerning the testing problems considered here. The early literature focussed on one-sided testing problems in multivariate analysis. Much of it is concerned with the finite sample distribution of the LR statistic. See Perlman (1969) and Barlow *et al.* (1972) for references. More recent work along similar lines is referenced in Robertson, Wright, and Dykstra (1988) and includes papers by Hillier (1986), Shapiro (1988), and Goldberger (1992), among others. An exception to the focus of the early literature is the paper by Chernoff (1954), which considers the asymptotic distribution of the LR statistic for more general models.

The econometrics literature has focussed on deriving the distribution (asymptotic and finite sample) of the LR statistic for one-sided alternatives for regression and nonlinear models and on deriving asymptotically equivalent tests to the LR test. For linear regression models, references include Gourieroux, Holly, and Monfort (1982), Hillier (1986), Wolak (1987, 1989b), and Dufour (1989). For nonlinear models, references include Gourieroux, Holly, and Monfort (1980), Kodde and Palm (1986), Rogers (1986), and Wolak (1989a). Results for the LR test for mixed one-

and two-sided alternatives were initiated by Perlman (1969) for Gaussian location models and extended to more general models by Kodde and Palm (1986) and Wolak (1987, 1989a). The papers above all consider the LR test or asymptotically equivalent tests. There is also a number of papers that consider tests based on contrasts. These include Hillier (1986), King and Smith (1986), and King and Wu (1990). See Robertson *et al.* (1988) for further references. King and Wu (1990) established a locally mean most powerful property of their additive t–test.

The testing problems considered in this paper are ones in which the null hypothesis is defined by equality restrictions. We do not consider tests of "multivariate inequality constraints." A simple example of such a testing problem is $\mathrm{H}_0 : \beta \geq 0$ versus $\mathrm{H}_1 : \beta \ngeq 0$. For results concerning problems of this sort see Perlman (1969), Robertson and Wegman (1978), and Wolak (1987, 1989a, b). For further references, see Robertson *et al.* (1988).

The remainder of the paper is organized as follows. Section 2 derives the optimal directed tests for the Gaussian regression model. Section 3 establishes the admissibility of the LR test for this model. Section 4 establishes the asymptotic null distribution and asymptotic optimality properties of the directed statistics for dynamic nonlinear models under a set of high-level assumptions. Section 5 shows that the LR statistic is asymptotically admissible in such models using similar conditions. Section 6 provides primitive sufficient conditions for the high-level assumptions of Sections 4 and 5. Section 7 describes the experimental design and the results from a simulation experiment that compares several tests of one-sided and mixed one- and two-sided alternatives. Two Appendices contain proofs of the results stated in the text.

2. REGRESSION: OPTIMAL TESTS

This section derives optimal tests for Gaussian linear regression models with known variance using a weighted average power criterion. We assume:

ASSUMPTION 1: *The model is given by*

$$Y_t = X_t'\beta + G_t'\delta + U_t \quad \text{for} \quad t = 1, ..., T \ ,$$

where $U_t \sim$ iid $N(0, \sigma^2)$, $\sigma^2 > 0$ *is known,* $X_t \in R^p$, $G_t \in R^q$, $\beta \in B \subset R^p$, $\delta \in \Delta \subset R^q$, $\{(X_t, G_t) : t = 1, ..., T\}$ *are non-random, and* $[X \vdots G]$ *is full rank* $s = p+q$ $(\leq T)$ *for* $X = [X_1 \vdots X_2 \vdots \cdots \vdots X_T]'$ *and* $G = [G_1 \vdots G_2 \vdots \cdots \vdots G_T]'$.

The null and alternative hypotheses of interest are:

$$\text{(2.1)} \qquad \text{H}_0 : \beta = 0 \text{ and } \text{H}_1^* : \beta \in B/\{0\} ,$$

where $B/\{0\}$ denotes the set B minus the zero vector. The regression parameter vector is $\theta = (\beta', \delta)'$. The information matrix for θ is

$$\text{(2.2)} \qquad \mathcal{I} = \begin{bmatrix} \mathcal{I}_1 & \mathcal{I}_2 \\ \mathcal{I}_2' & \mathcal{I}_3 \end{bmatrix} = \begin{bmatrix} X'X & X'G \\ G'X & G'G \end{bmatrix} /\sigma^2 .$$

The parameter space Θ^* $(\subset R^s)$ of θ is required to satisfy:

ASSUMPTION 2: $\Theta^* = B \times \Delta$ *for* $B \subset R^p$ *and* $\Delta \subset R^q$ *and* B *has positive Lebesgue measure.*

ASSUMPTION 3: $d - \mathcal{I}_3^{-1}\mathcal{I}_2' b \in \Delta \ \ \forall b \in B, \ \forall d \in \Delta$.

The two main cases where Assumption 3 is satisfied are when (1) the nuisance parameter δ is unrestricted (i.e., $\Delta = R^q$) or (2) the regressors X and G are orthogonal (i.e., $\mathcal{I}_2 = X'G/\sigma^2 = 0$). In each of these cases, Assumption 3 places no restrictions on the shape of the parameter space B. Common shapes include (i) $\{\beta \in R^p : \beta_j \geq 0 \ \forall j = 1, ..., p\}$, (ii) $\{\beta \in R^p : \beta_j \in R \ \forall j \leq J; \beta_j \geq 0 \ \forall j = J{+}1, ..., p\}$, (iii) $\{\beta \in R^p : \beta_j > 0 \text{ for some } j \leq p\}$, and (iv) $\{\beta \in R^p : \beta_j - \beta_{j-1} \geq 0 \ \forall j = 1, ..., p, \text{ where } \beta_0 = 0\}$ (which corresponds to the non-negative and ordered alternative $\text{H}_1^* : 0 \leq \beta_1 \leq \beta_2 \ \cdots \leq \beta_p$). In addition to the two main cases listed above, Assumption 3 is satisfied in a variety of other special cases. For example, if B is contained in the positive orthant of R^p, Δ is the negative orthant of R^q, and $\mathcal{I}_3^{-1}\mathcal{I}_2'$ contains non-negative elements, then Assumption 3 holds.

Let θ_0 denote some parameter vector in the null hypothesis: $\theta_0 = (0', \delta_0')' \in R^s$ for some $\delta_0 \in R^q$. Any parameter vector $\theta \in R^s$ can be written as the sum of the null parameter vector θ_0 and some perturbation vector $h \in R^s$. That is, $\theta = \theta_0 + h$. Given θ_0, we specify a weight function

$Q_c(\cdot)$ over perturbation vectors h. The weight function we use is a singular multivariate normal distribution whose support lies in the orthogonal complement (with respect to a particular inner product) of the linear subspace of R^s defined by the null hypothesis.

More specifically, let V denote the linear subspace of R^s defined by

$$V = \{\theta \in R^s \,:\, \theta = (0',\ \delta')' \text{ for some } \delta \in R^q\}\ . \tag{2.3}$$

The null hypothesis can be expressed as $\mathrm{H}_0 \,:\, \theta \in V \cap \Theta$. Define the inner product $\langle h,\, \ell\rangle = h'\mathcal{I}\ell$, for $h,\, \ell \in R^s$. Denote the orthogonal complement of V under $\langle\cdot,\ \cdot\rangle$ by $V^\perp$. Since V is a q dimensional subspace of R^s, $V^\perp$ is a p dimensional subspace of R^s. Let $\{a_1,\, ...,\, a_p\}$ be some basis of $V^\perp$ and define $A = [a_1 \,\vdots\, a_2 \,\vdots\, \cdots \,\vdots\, a_p] \in R^{s\times p}$. For example, one can take $A = [I_p \,\vdots\, -\mathcal{I}_2'\mathcal{I}_3^{-1}]'$. Note that

$$A'\mathcal{I}A = \mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2' = X'M_GX/\sigma^2\ ,\quad \text{where}\quad M_G = I_T - G(G'G)^{-1}G'\ . \tag{2.4}$$

Next, let

$$\Sigma = A(A'\mathcal{I}A)^{-1}A' = \begin{bmatrix} \left(\mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2'\right)^{-1} & -\left(\mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2'\right)^{-1}\mathcal{I}_2\mathcal{I}_3^{-1} \\ -\mathcal{I}_3^{-1}\mathcal{I}_2'\left(\mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2'\right)^{-1} & \mathcal{I}_3^{-1}\mathcal{I}_2'\left(\mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2'\right)^{-1}\mathcal{I}_2\mathcal{I}_3^{-1} \end{bmatrix}\ . \tag{2.5}$$

Also, let $N(0,\ \Sigma)$ denote a multivariate normal distribution with mean 0 and covariance matrix Σ (possibly singular).

ASSUMPTION 4: $Q_c = N(0,\ c\Sigma)$ *for some positive constant* c.

Note that the support of Q_c is $V^\perp$.

The weight function Q_c gives equal weight to different alternatives $\theta = \theta_0 + h$ that are equally difficult to detect (as measured by the power of the best test of $\widetilde{\mathrm{H}}_0 \,:\, \theta = \theta_0$ versus $\widetilde{\mathrm{H}}_1 \,:\, \theta = \theta_0 + h$. Thus, the contours of Q_c are the same as those considered by Wald (1942, 1943).

The constant c, which scales the variance matrix of the weight function Q_c, determines the relative weight given to alternatives that are close to the null versus alternatives that are distant from the null. A small value of c corresponds to giving high weight to close alternatives. The

larger is c, the more weight is given to distant alternatives. As $c \to \infty$, the weight function gets closer and closer to giving equal weight to alternatives of different proximity to the null.

The weighted average power criterion that we consider is given by

$$\int 1(\theta_0 + h \in \Theta^*) P(\varphi \text{ rejects } H_0 | \theta_0 + h) dQ_c(h)/K \ , \tag{2.6}$$

where φ is some level α test and $K = \int 1(\theta_0 + h \in \Theta^*) dQ_c(h)$. K is positive, because B has positive Lebesgue measure. Note that the weight function $Q_c(h)$ is truncated so that it only gives non-zero weight to parameter values θ $(= \theta_0 + h)$ in Θ^*. (The constant K merely ensures that the truncated weight function integrates to one.) An optimal test of level α maximizes the above weighted average power criterion over all tests of level α.

We determine an optimal test as follows. Let $f(y, \theta)$ denote the density of the T vector of observations $Y = (Y_1, ..., Y_T)'$ evaluated at $y = (y_1, ..., y_T)'$. Let $\varphi = \varphi(y)$ denote a test of H_0. That is, $\varphi(Y)$ is a [0, 1]-valued function of Y that rejects H_0 with probability γ when $\varphi(Y) = \gamma$. (Of course, φ depends on the non-stochastic regressors as well as on Y.) The power of φ against $\theta = \theta_0 + h$ is given by $\int \varphi(y) f(y, \theta_0 + h) dy$. The weighted average power of φ equals

$$\begin{aligned} & \int 1(\theta_0 + h \in \Theta^*) P(\varphi \text{ rejects } H_0 | \theta_0 + h) dQ_c(h)/K \\ = \ & \int 1(\theta_0 + h \in \Theta^*) \int_{R^T} \varphi(y) f(y, \theta_0 + h) dy dQ_c(h)/K \\ = \ & \int_{R^T} \varphi(y) \left[\int 1(\theta_0 + h \in \Theta^*) f(y, \theta_0 + h) dQ_c(h)/K \right] dy \end{aligned} \tag{2.7}$$

by Fubini's Theorem.

Equation (2.7) shows that the weighted average power of φ equals the power of φ against the single alternative density specified by

$$g(y, \theta_0) = \int 1(\theta_0 + h \in \Theta^*) f(y, \theta_0 + h) dQ_c(h)/K \ . \tag{2.8}$$

Hence, a test that maximizes power against the simple alternative $g(\cdot, \theta_0)$ also maximizes weighted average power.

The Neyman–Pearson Lemma shows that the best test for testing the simple null $Y \sim f(\cdot, \theta_0)$ against the simple alternative $Y \sim g(\cdot, \theta_0)$ is based on the likelihood ratio statistic $LR(\theta_0)$:

$$LR(\theta_0) = g(Y, \theta_0)/f(Y, \theta_0) = \left[\int 1(\theta_0 + h \in \Theta^*) f(Y, \theta_0 + h) dQ_c(h)/K \right] / f(Y, \theta_0) \ . \tag{2.9}$$

We show below that this statistic does not depend on θ_0. In addition, we show below that it can be written in a simplified form that involves only a p dimensional multivariate normal probability rather than an s $(= p+q)$ dimensional integral. These simplifications are a consequence of our choice of the contours of the weight function Q_c. One could consider different contours, but this would leave one with a test statistic that involves a higher (often much higher) dimensional integral to compute and an integral which is less well understood from a computational perspective than the multivariate normal probabilities that arise (see below) with the given choice of Q_c.

For $B \subset R^p$, $\mu \in R^p$, and Ω a positive semi-definite $p \times p$ matrix, let

$$\Phi_p(B,\ \mu,\ \Omega) = P(Z \in B)\ , \quad \text{where} \quad Z \sim N(\mu,\ \Omega)\ . \tag{2.10}$$

For notational simplicity, we often suppress the subscript p.

The statistic $LR(\theta_0)$ is shown below to equal a constant times the *directed Wald statistic* D–W_c defined by

$$\begin{aligned} D\text{–}W_c &= (1+c)^{-p/2} \exp\left(\tfrac{1}{2}\ \tfrac{c}{1+c} W\right) \Phi\left(B,\ \tfrac{c}{1+c}\widehat{\beta},\ \tfrac{c}{1+c}(X'M_GX)^{-1}\sigma^2\right)\ , \quad \text{where} \\ W &= \widehat{\beta}'(X'M_GX/\sigma^2)\widehat{\beta}\ , \quad \text{and} \quad \widehat{\beta} = (X'M_GX)^{-1}X'M_GY\ . \end{aligned} \tag{2.11}$$

Note that $\widehat{\beta}$ is the least squares estimator of β from the (unrestricted) regression of Y_t on X_t and G_t. W is just the standard Wald test statistic for testing $\mathrm{H}_0 : \beta = 0$ against $\mathrm{H}_1 : \beta \neq 0$ (for the case where σ^2 is known). The calculation of D–W_c requires the computation of a multivariate normal integral. For $p = 2$ or 3, this can be done by numerical quadrature. For $p \geq 4$, it can be done by simulation methods. See Hajivassiliou, McFadden, and Ruud (1994) for a review and evaluation of methods of doing so. One rejects the null hypothesis for large values of D–W_c.

The directed Wald test differs from the standard Wald test by the appearance of the factor $\Phi(\cdot,\ \cdot,\ \cdot)$ in the test statistic. Due to the Φ factor, the directed Wald statistic differentially weights realizations of W depending on the specification of the alternative parameter space B and the length and direction of $\widehat{\beta}$. Consider the case where $p = 2$, B is the positive orthant, and $(X'M_GX)^{-1}\sigma^2$ is the identity matrix. Table I provides the values of $\Phi_2(R^2_+,\ \lambda,\ I_2)$ for several values of λ on the unit circle. For λ in the middle of R^2_+, i.e., $\lambda = (.707,\ .707)$, the factor

$\Phi_2(R_+^2,\, \lambda,\, I_2)$ is almost ten times as large as when λ is on the opposite side of the circle. Thus, the factor Φ has a substantial impact on the value of the directed Wald statistic.

Note that the directed test statistics depend on the weight function $Q_c(\cdot)$ only through the scalar constant c. As mentioned above, a small value of c corresponds to giving high weight to close alternatives. In contrast, as $c \to \infty$, the weight function gets closer to giving equal weight to alternatives of different proximity to the null.

The limiting value as $c \to \infty$ of the directed Wald statistic (after a suitable monotone transformation) is as follows:

$$\text{(2.12)} \qquad D\text{–}W_\infty = \lim_{c\to\infty} 2\log[(1+c)^{p/2} D\text{–}W_c] = W + 2\log[\Phi(B,\, \widehat{\beta},\, (X'M_GX)^{-1}\sigma^2)] \;,$$

where log denotes the natural logarithm here and below.

If B is positively homogeneous (i.e., $\beta \in B \Rightarrow \tau\beta \in B\ \forall \tau > 0$), which includes the four examples listed following Assumption 3, the limit as $c \to 0$ is

$$\begin{aligned} &D\text{–}W_0 = \lim_{c\to 0} J_c(D\text{–}W_c) = d'\widehat{\beta}/\left(d'(X'M_GX/\sigma^2)^{-1}d\right)^{1/2} \;, \\ &\qquad \text{where } d = \tfrac{\partial}{\partial\mu}\Phi(B,\, 0,\, (X'M_GX)^{-1}) \;, \\ &\tfrac{\partial}{\partial\mu}\Phi(B,\, 0,\, \Omega) = \tfrac{\partial}{\partial\mu}\Phi(B,\, \mu,\, \Omega)\Big|_{\mu=0} = \Omega^{-1}EZ1(Z \in B) \text{ for } Z \sim N(0,\, \Omega) \;, \text{ and} \\ &J_c(x) = \tfrac{2}{\sqrt{c}}\left[\log\left((1+c)^{p/2}x\right) - \log\Phi(B,\, 0,\, (X'M_GX)^{-1})\right] \frac{\Phi(B,\, 0,\, (X'M_GX)^{-1})}{(d'(X'GX/\sigma^2)^{-1}d)^{1/2}} \;. \end{aligned} \tag{2.13}$$

Except for special cases (such as $B = R^p$), we have $d \neq 0$ and $D\text{–}W_0$ is well-defined. Since $D\text{–}W_0 \sim N(0,1)$ under the null hypothesis, it is easy to obtain the desired one-sided critical values for this test statistic. (One rejects H_0 for large values of $D\text{–}W_0$.) Thus, to carry out a test based on $D\text{–}W_0$, the only potential complication is in computing $d = \frac{\partial}{\partial\mu}\Phi(B,\, 0,\, (X'M_GX)^{-1})$. For $p = 2$ or 3, d can be computed by numerical quadrature. For $p \geq 4$, it can be computed by simulation methods. As noted above, see Hajivassiliou, McFadden, and Ruud (1994) regarding the latter. In addition, d (or d up to a constant of proportionality) can be obtained by symmetry arguments in some cases. For example, suppose $p = 2$ and B is the positive orthant (or R^2 minus the negative orthant), then by symmetry d is proportional to (1,1)′. For $p > 2$, an analogous result holds if the correlation matrix that corresponds to the covariance matrix $(X'M_GX)^{-1}\sigma^2$

has all non-diagonal elements equal. Although a test based on D–W_0 is easier to carry out than one based on D–W_c for $0 < c \leq \infty$ (since one does not need to simulate critical values), we do not favor it on theoretical grounds (because it is designed only for very local alternatives to the null hypothesis) or based on the results of power simulations; see Section 7 below.

The null distributions of the directed Wald statistics equal those of the following rv's: $\forall 0 < c < \infty$,

$$
\begin{aligned}
&\mathcal{L}_c = (1+c)^{-p/2} \exp\left(\tfrac{1}{2}\ \tfrac{c}{1+c} Z'Z\right) \Phi\left(B,\ \tfrac{c}{1+c}(X'M_GX)^{-1/2}\sigma Z,\ \tfrac{c}{1+c}(X'M_GX)^{-1}\sigma^2\right)\ , \\
&\mathcal{L}_\infty = Z'Z + \log\left[\Phi(B,\ (X'M_GX)^{-1/2}\sigma Z,\ (X'M_GX)^{-1}\sigma^2)\right]\ , \quad \text{and} \\
&\mathcal{L}_0 = Z_1\ , \quad \text{where} \quad Z \sim N(0,\ I_p) \quad \text{and} \quad Z = (Z_1,\ ...,\ Z_p)'\ .
\end{aligned}
\tag{2.14}
$$

Let ξ_c denote a test of level α based on the directed Wald statistic D–W_c. Properties of this test are given in the following theorem.

THEOREM 1: *Suppose Assumptions* 1–4 *hold. Then,*

(a) D–$W_c = LR(\theta_0) \times K$ *for all* $0 < c < \infty$,

(b) D–$W_c \sim \mathcal{L}_c$ *under the null hypothesis* H_0 *for all* $0 \leq c \leq \infty$ *and*

(c) *for any level* α *test* φ*, the directed Wald test* ξ_c *satisfies*

$$\int 1(\theta_0 + h \in \Theta^*) \left[\int \varphi f(\theta_0 + h) d\mu\right] dQ_c(h) \leq \int 1(\theta_0 + h \in \Theta^*) \left[\int \xi_c f(\theta_0 + h) d\mu\right] dQ_c(h)$$

for all $0 < c < \infty$*, with strict inequality unless* $\varphi = \xi_c$ *(Lebesgue) almost everywhere, where* $f(\theta)$ *is the Gaussian density of the data and* μ *is Lebesgue measure on* R^T*.*

The proofs of Theorem 1 and Theorem 2 below are given in Appendix A.

COMMENTS: 1. By Theorem 1(b), the directed Wald test is an exactly similar test. Its null distribution depends on the regressors, however, so it is not possible to provide tables of exact critical values. Instead, critical values can be obtained on a case by case basis by simulation.

2. The result of Theorem 1 (and Theorem 2 below) applies to a more general class of regression testing problems than those that satisfy Assumption 1. In particular, the following regression model and hypotheses do not satisfy Assumption 1 but can be transformed to do so: $Y^\dagger$

$= X^\dagger\theta^\dagger + U^\dagger$, $U^\dagger \sim N(0,\, \sigma^2\Omega)$, $\mathrm{H}_0^\dagger : R^\dagger\theta^\dagger = r$, and $\mathrm{H}_1^\dagger : R^\dagger\theta^\dagger \in B^\dagger/\{r\}$, where $Y^\dagger$ is an observed rv in R^T; $X^\dagger$ is an observed non-stochastic $T \times s$ regressor matrix; $\theta^\dagger$ is an unknown parameter vector in R^s; $U^\dagger$ is an unobserved error vector in R^T; Ω is a known $T \times T$ positive definite covariance matrix; σ^2 is a known positive constant; $R^\dagger$ is a known $p \times s$ matrix of constants; r is a known p–vector of constants; and $B^\dagger$ is a known subset of R^p. An example of the above testing problem is a test of the equality of a sub-vector of parameters $\mathrm{H}_0^\dagger : \theta_1^\dagger = \theta_2^\dagger = \cdots = \theta_{p+1}^\dagger$ against the ordered alternative $\mathrm{H}_1^\dagger : \theta_1^\dagger \le \theta_2^\dagger \le \cdots \le \theta_{p+1}^\dagger$ with one or more of the inequalities being strict. This testing problem cannot be written directly as a special case of the testing problem of (2.1), but it can be transformed to the form of (2.1). (See King and Smith (1986, Sec. 2) for details of the transformation.)

3. If σ^2 is unknown, then an exactly similar test can be constructed, provided B is positively homogeneous, by replacing σ^2 in the definition of D–W_c by the unrestricted estimator $\widehat{\sigma}^2 = \frac{1}{T-p}\Sigma_{t=1}^T (Y_t - X_t'\widehat{\beta} - G_t'\widehat{\delta})^2$. The resultant test statistic is a *directed F statistic*. The null distribution of the directed F statistic is given by that of

$$\mathcal{L}_c = (1+c)^{-p/2} \exp\Big(\tfrac{1}{2}\, \tfrac{c}{1+c} Z'Z(T-p)/\chi^2\Big)\Phi\Big(B,\, \tfrac{c}{1+c}(X'M_GX)^{-1/2}Z,\, \tfrac{c}{1+c}(X'M_GX)^{-1}\chi^2/(T-p)\Big)\ ,$$

where $Z \sim N(0,\, I_p)$, χ^2 has a chi-squared distribution with p degrees of freedom, and Z and χ^2 are independent. Theorem 1 does not establish finite sample optimality properties of the directed F statistic, but Theorem 3 below provides asymptotic optimality properties for it.

3. REGRESSION: THE LIKELIHOOD RATIO TEST

3.1. Admissibility

In this section we consider the standard LR test for the hypotheses given in (2.1). We show that the LR test is admissible. Roughly speaking, we show that it maximizes weighted average power for weight functions that place weight only on distant alternatives. Analogous properties of the LR test, but for a different non-standard testing problem (viz., that of testing when a nuisance parameter is present only under the alternative) are established in Andrews and Ploberger (1993).

The standard LR test statistic equals minus two times the likelihood ratio:

$$\begin{aligned} &\mathcal{LR} = -2(\ell(\widetilde{\theta}) - \ell(\theta^*)) \ , \quad \text{where} \\ &\ell(\theta) = \log f(Y, \theta) \ , \quad \ell(\widetilde{\theta}) = \sup_{\theta \in V \cap \Theta^*} \ell(\theta) \ , \quad \text{and} \quad \ell(\theta^*) = \sup_{\theta \in \Theta^*} \ell(\theta) \ . \end{aligned} \tag{3.1}$$

Let $U(\cdot)$ denote the uniform distribution on the p–dimensional unit sphere. Let $U^*(\cdot)$ denote the distribution of $b^* = (A'\mathcal{I}A)^{1/2}\xi$, where $\xi \sim U$.

We consider the case where the parameter space Θ satisfies Assumptions 2 and 3 plus the following assumptions.

ASSUMPTION 5: *B is positively homogeneous.*

ASSUMPTION 6: $\forall \varepsilon > 0$, $\inf_{b \in B : b'A'\mathcal{I}Ab = 1} \int 1(\|b - b^*\| \leq \varepsilon, \, b^* \in B) dU^*(b^*) > 0$.

Assumption 6 does not allow B to contain isolated rays. It is satisfied if B is an orthant, a union of orthants, etc.

The weight functions we consider place all weight on ellipses of radius proportional to r for $r > 0$. As in the previous section, we write $\theta = \theta_0 + h$ and we let Q^r be a weight function over vectors h.

ASSUMPTION 7: *Q^r is the distribution of $rA(A'\mathcal{I}A)^{-1/2}\xi$, where $\xi \sim U$.*

The weighted average power criterion considered here is given by (2.6) with Q_c replaced by Q^r. By (2.7), the weighted average power of a test equals its power against the single alternative density specified in (2.8) with Q_c replaced by Q^r. By the Neyman–Pearson Lemma, then, the best test statistic in terms of weighted average power is given by the simple likelihood ratio:

$$LR(\theta_0, Q^r) = \left(\int 1(\theta_0 + h \in \Theta^*) f(Y, \theta_0 + h) dQ^r(h) / K \right) / f(Y, \theta_0) \ . \tag{3.2}$$

(Note that the constant $K = \int 1(\theta_0 + h \in \Theta^*) dQ^r(h)$ does not depend on r, since B is positively homogeneous.)

It turns out that $\mathcal{LR}$ equals the limit as $r \to \infty$ of $LR(\theta_0, Q^r)$ (after suitable normalization). In consequence, the level α likelihood ratio test $\Lambda = 1(\mathcal{LR} > k_\alpha)$ inherits a weighted average power optimality property for distant alternatives.

Let φ denote a test of H_0. Let E_0 denote expectation under H_0. We say that φ is *distinct* from Λ if $E_0(1-\varphi)1(\mathcal{LR} > k_\alpha) > 0$ (i.e., if φ accepts and Λ rejects with some positive probability under H_0).

THEOREM 2: *Suppose Assumptions* 1–3 *and* 5–7 *hold. Then,*

(a) $\mathcal{LR} = \beta^{*\prime}(X'M_GX/\sigma^2)\beta^* = \lim_{r\to\infty}\left(\max\left\{\frac{r}{2} + \frac{1}{r}\log LR(\theta_0, Q^r), 0\right\}\right)^2$, *where* $\theta^* = (\beta^{*\prime}, \delta^{*\prime})'$,

(b) $\mathcal{LR} \sim (\max\{\sup_{\xi\in\mathcal{S}_p(B)} \xi'Z, 0\})^2$ *under* θ_0, *where* $Z \sim N(0, I_p)$ *and* $\mathcal{S}_p(B) = \{\xi \in R^p : \|\xi\| = 1 \text{ and } (A'\mathcal{I}A)^{-1/2}\xi \in B\}$, *and*

(c) *for any test* φ *that is distinct from* Λ, *there exists a constant* $r_0 < \infty$ *such that for all* $r \geq r_0$

$$\int 1(\theta_0 + h \in \Theta^*)\left[\int \varphi f(\theta_0 + h)d\mu\right]dQ^r(h) < \int 1(\theta_0 + h \in \Theta^*)\left[\int \Lambda f(\theta_0 + h)d\mu\right]dQ^r(h) \ .$$

COMMENTS: 1. Theorem 2(c) provides only a weak optimality property for $\mathcal{LR}$, because power is directed only at very distant alternatives.

2. Critical values for $\mathcal{LR}$ can be obtained by simulation for arbitrary choices of parameter space B, provided the computation of β^* is not particularly onerous. If B is a closed, convex, positively homogeneous set, then the null distribution of $\mathcal{LR}$ is a chi-bar-squared distribution, which is a mixture of chi-squared distributions, see Shapiro (1988).

3. The proof of Theorem 2(c) actually establishes the stronger result that the ratio of the weighted average type II error of φ over that of Λ diverges to infinity as $r \to \infty$.

3.2. Computational Issues

We conclude this section by discussing some of the relative merits of the directed tests and the LR test from a computational perspective.

To obtain critical values for multivariate one-sided or mixed one- and two-sided hypotheses, both tests require the calculation of multivariate normal orthant probabilities. (The LR test needs them to determine the weights in the mixture of chi-squared distributions.) For either test, approximations could be used to circumvent the calculation of such probabilities. Simulation methods for calculating such probabilities, however, are now sufficiently easy and fast that there

seems to be little reason to rely on approximations. See Hajivassiliou, McFadden, and Ruud (1994) for GAUSS and FORTRAN programs that compute the requisite orthant probabilities.

An advantage of the LR test is that once one has calculated the orthant probabilities, one can obtain a p–value using just the distribution function of a chi-squared random variable. For a directed test, one has to simulate the p–value or critical values. On the other hand, simulating p–values and/or critical values is easy and fast and can be programmed simply to handle a wide variety of different alternative hypotheses. As noted above, an interactive GAUSS program that does this is available from the author. To write an analogous general program for the LR test is more complicated, because one has to determine the appropriate probability weights for any given restricted alternative.

A computational advantage of the directed tests is that they do not require computation of the ML estimator for the restricted alternative (RA) hypothesis $H_1^* : \beta \in B/\{0\}$. The estimator employed by the directed statistics is just the unrestricted LS estimator regardless of the specification of B. On the other hand, for common specifications of B, the LR statistic requires that one solve a quadratic programming problem that depends on B. For nonlinear models considered below, avoiding the computation of the ML estimator for the RA hypothesis can be particularly advantageous. Furthermore, the directed LM statistic (defined below), only requires calculation of the ML estimator under the null and not under the RA hypothesis or the unrestricted hypothesis $H_1 : \beta \in R^p$. As is well known from classical testing problems, this yields considerable computational simplicity in a variety of nonlinear models. In contrast, the LR test and asymptotically equivalent versions of it, such as the Kuhn–Tucker multiplier test, require computation of the ML estimator under the RA hypothesis.

4. NONLINEAR MODELS: OPTIMAL TESTS

In this section we extend the finite sample optimal test results of Section 2 to nonlinear dynamic models using asymptotics. We introduce tests called *directed* Wald, LM, and LR tests.

4.1. Notation

Let $\underset{\sim}{Y}_T$ denote the data matrix when the sample size is T for $T = 1, 2, ...$ Consider a parametric family $\{f_T(\underset{\sim}{y}_T, \theta) : \theta \in \Theta^* \cup \Theta\}$ of densities of $\underset{\sim}{Y}_T$ with respect to some σ–finite measure μ_T, where Θ^* ($\subset R^s$) and Θ ($\subset R^s$) are two parameter spaces defined below. The likelihood function of the data is given by $f_T(\theta) = f_T(\underset{\sim}{Y}_T, \theta)$. In many cases, the likelihood function $f_T(\theta)$ can be written as a product of two terms, one that depends on θ and another that does not. Often the latter term is the product over $t = 1, ..., T$ of the conditional distribution of some weakly exogenous variables at time t given all of the preceding variables (exogenous or not). In such cases, these conditional distributions of the weakly exogenous variables need not be known in order for one to construct the test statistics considered here. The optimality results stated below hold for any distribution for which the assumptions on $f_T(\theta)$ hold.

The parameter θ is taken to be of the form $\theta = (\beta', \delta')'$, where $\beta \in R^p$, $\delta \in R^q$, and $s = p+q$. The null hypothesis of interest is $\mathrm{H}_0 : \beta = 0$, as in (2.1). We let θ_0 denote a parameter vector in the null hypothesis. That is, θ_0 is of the form $\theta_0 = (0', \delta')'$ for some $\delta \in R^q$.

For standard large-sample two-sided testing problems, the parameter space is taken to be a subset Θ of R^s that contains a neighborhood of θ_0 for all θ_0 in the null hypothesis. We refer to such a parameter space Θ as the *unrestricted alternative* (UA) parameter space. The alternative hypothesis corresponding to the UA parameter space is $\mathrm{H}_1 : \beta \neq 0$. In this paper, our interest centers not on testing H_0 versus H_1, but on testing H_0 versus a restricted alternative. Nevertheless, we define the standard Wald, LM, and LR test statistics here, because it is necessary to establish notation that is used below when discussing the main problem of interest.

Let $\ell_T(\theta) = \log f_T(\theta)$. Let $D\ell_T(\theta)$ denote the s–vector of partial derivatives of $\ell_T(\theta)$ with respect to θ. Let $D^2\ell_T(\theta)$ denote the $s \times s$ matrix of second partial derivatives of $\ell_T(\theta)$ with respect to θ. We consider the standard case where the appropriate norming factors for $D\ell_T(\theta)$ and $D^2\ell_T(\theta)$ (so that each is $O_p(1)$ but not $o_p(1)$) are $\frac{1}{\sqrt{T}}$ and $\frac{1}{T}$ respectively. Let $\mathcal{I}(\theta) = \mathrm{plim}_{T\to\infty} -\frac{1}{T}D^2\ell_T(\theta)$. ($\mathcal{I}(\theta)$ is the limiting information matrix for θ.)

Let $\widehat{\theta}$ be the UA ML estimator of θ. By definition, $\widehat{\theta}$ maximizes the log likelihood function

over Θ at least with probability that goes to one as $T \to \infty$ (i.e., wp $\to 1$). That is, $\widehat{\theta}$ satisfies

$$\ell_T(\widehat{\theta}) = \max_{\theta \in \Theta} \ell_T(\theta) \quad \text{wp} \to 1 \quad \text{under} \quad \theta_0 \ . \tag{4.1}$$

Let $\widetilde{\theta}$ be the restricted ML estimator of θ (restricted by the null hypothesis H_0). By definition, $\widetilde{\theta}$ satisfies

$$\begin{aligned} &\widetilde{\theta} \in \widetilde{\Theta} = \{\theta \in \Theta \,:\, \theta = (0', \delta')' \ \text{ for some } \delta \in R^q\} \ \text{ and} \\ &\ell_T(\widetilde{\theta}) = \sup_{\theta \in \widetilde{\Theta}} \ell_T(\theta) \quad \text{wp} \to 1 \quad \text{under} \quad \theta_0 \ . \end{aligned} \tag{4.2}$$

The standard Wald, LM, and LR test statistics for testing H_0 against H_1 are given by

$$\begin{aligned} W_T &= T(H\widehat{\theta})' \left[H \mathcal{I}_T^{-1}(\widehat{\theta}) H' \right]^{-1} H\widehat{\theta} = T\widehat{\beta}'[\widehat{\mathcal{I}}_1 - \widehat{\mathcal{I}}_2 \widehat{\mathcal{I}}_3^{-1} \widehat{\mathcal{I}}_2'] \widehat{\beta} \ , \\ LM_T &= (D\ell_T(\widetilde{\theta})/\sqrt{T})' \mathcal{I}_T^{-1}(\widetilde{\theta}) D\ell_T(\widetilde{\theta})/\sqrt{T} \\ &= \left(\tfrac{\partial}{\partial \beta} \ell_T(\widetilde{\theta})/\sqrt{T} \right)' \left[\widetilde{\mathcal{I}}_1 - \widetilde{\mathcal{I}}_2 \widetilde{\mathcal{I}}_3^{-1} \widetilde{\mathcal{I}}_2' \right]^{-1} \tfrac{\partial}{\partial \beta} \ell_T(\widetilde{\theta})/\sqrt{T} \ , \\ LR_T &= -2(\ell_T(\widetilde{\theta}) - \ell_T(\widehat{\theta})) \ , \quad \text{where} \quad H = [I_p \,\vdots\, 0] \subset R^{p \times s} \ , \\ \mathcal{I}_T(\theta) &= -\tfrac{1}{T} D^2 \ell_T(\theta) = \begin{bmatrix} \mathcal{I}_{1T}(\theta) & \mathcal{I}_{2T}(\theta) \\ \mathcal{I}_{2T}(\theta)' & \mathcal{I}_{3T}(\theta) \end{bmatrix} \ , \quad \widehat{\mathcal{I}}_j = \mathcal{I}_{jT}(\widehat{\theta}) \ , \quad \text{and} \\ \widetilde{\mathcal{I}}_j &= \mathcal{I}_{jT}(\widetilde{\theta}) \quad \text{for} \quad j = 1, 2, 3 \ . \end{aligned} \tag{4.3}$$

Alternatively, one can define $\mathcal{I}_T(\theta)$ to be of outer product rather than Hessian form.

4.2. Optimal Tests for Restricted Alternatives

The alternative hypothesis that is of primary interest in this paper is $\mathrm{H}_1^* \,:\, \beta \in B/\{0\}$, as in (2.1). The corresponding restricted alternative (RA) parameter space is $\Theta^* = B \times \Delta$, where $B \subset R^p$ and $\Delta \subset R^q$. For standard asymptotic results, the parameter space B is required to be *positively homogeneous* (i.e., $\beta \in B \Rightarrow \tau\beta \in B \ \forall \tau > 0$) and to have positive Lebesgue measure; otherwise, its shape is arbitrary. For example, B could be an orthant, a half-space, a cone, or unions or intersections of such sets. By using non-standard asymptotics, the assumption of positive homogeneity can be circumvented; see the comments following Theorem 3 below.

To derive asymptotically optimal tests of H_0 versus H_1, we consider local alternatives to H_0 of the form $f_T(\theta_0 + h/\sqrt{T})$ for some $h \in R^s$. We consider the same weight function $Q_c(h)$ over values of h as in Section 2, but with the information matrix $\mathcal{I}$ defined as $\mathcal{I}(\theta_0)$, where $\mathcal{I}(\theta)$ is

defined above (4.1), rather than as in Section 2. We consider an asymptotic weighted average power criterion, which is the limit superior as $T \to \infty$ of (2.6) with $\theta_0 + h$ replaced by $\theta_0 + h/\sqrt{T}$. For this criterion, directed Wald, LM, and LR tests are shown to be best.

The directed Wald statistic is defined as

$$
\begin{aligned}
&D\text{–}W_{cT} = (1+c)^{-p/2} \exp\left(\tfrac{1}{2}\tfrac{c}{1+c} W_T\right) \Phi\left(B, \tfrac{c}{1+c}\widehat{\beta}, \tfrac{c}{1+c}\left(\widehat{\mathcal{I}}_1 - \widehat{\mathcal{I}}_2\widehat{\mathcal{I}}_3^{-1}\widehat{\mathcal{I}}_2'\right)^{-1}/T\right) \text{ for } 0<c<\infty, \\
&D\text{–}W_{\infty T} = W_T + 2\log\left[\Phi\left(B, \widehat{\beta}, \left(\widehat{\mathcal{I}}_1 - \widehat{\mathcal{I}}_2\widehat{\mathcal{I}}_3^{-1}\widehat{\mathcal{I}}_2'\right)^{-1}/T\right)\right], \text{ and} \\
&D\text{–}W_{0T} = \sqrt{T}\widehat{d}'\widehat{\beta}/\left(\widehat{d}'\left(\widehat{\mathcal{I}}_1 - \widehat{\mathcal{I}}_2\widehat{\mathcal{I}}_3^{-1}\widehat{\mathcal{I}}_2'\right)\widehat{d}\right)^{1/2}, \\
&\qquad \text{where } \widehat{d} = \tfrac{\partial}{\partial\mu}\Phi\left(B, 0, \left(\widehat{\mathcal{I}}_1 - \widehat{\mathcal{I}}_2\widehat{\mathcal{I}}_3^{-1}\widehat{\mathcal{I}}_2'\right)^{-1}/T\right).
\end{aligned}
\tag{4.4}
$$

One rejects H_0 if $D\text{–}W_{cT}$ exceeds a critical value k_α that is determined using the asymptotic null distribution of $D\text{–}W_{cT}$.

The directed LM statistic, $D\text{–}LM_{cT}$, is defined analogously to $D\text{–}W_{cT}$ with W_T replaced by LM_T, $\widehat{\beta}$ replaced by $\left[\widetilde{\mathcal{I}}_1 - \widetilde{\mathcal{I}}_2\widetilde{\mathcal{I}}_3^{-1}\widetilde{\mathcal{I}}_2'\right]^{-1}\frac{\partial}{\partial\beta}\ell_T(\widetilde{\theta})/T$, and $\left(\widehat{\mathcal{I}}_1 - \widehat{\mathcal{I}}_2\widehat{\mathcal{I}}_3^{-1}\widehat{\mathcal{I}}_2'\right)^{-1}/T$ replaced by $\left(\widetilde{\mathcal{I}}_1 - \widetilde{\mathcal{I}}_2\widetilde{\mathcal{I}}_3^{-1}\widetilde{\mathcal{I}}_2'\right)^{-1}/T$. Note that the $D\text{–}LM_{cT}$ statistic is constructed using only the restricted *ML* estimator $\widetilde{\theta}$. The directed *LR* statistic, $D\text{–}LR_{cT}$, is defined analogously to $D\text{–}W_{cT}$ with W_T replaced by LR_T. (One also could replace $\widehat{\beta}$ and $\left(\widehat{\mathcal{I}}_1 - \widehat{\mathcal{I}}_2\widehat{\mathcal{I}}_3^{-1}\widehat{\mathcal{I}}_2'\right)^{-1}/T$ by the expressions above involving $\widetilde{\theta}$ without affecting the large sample properties of $D\text{–}LR_{cT}$.) The test statistics $D\text{–}W_{cT}$, $D\text{–}LM_{cT}$, and $D\text{–}LR_{cT}$ have the same asymptotic distributions under the null hypothesis and under local alternatives. In consequence, the directed LM and directed LR tests reject H_0 if $D\text{–}LM_{cT}$ and $D\text{–}LR_{cT}$, respectively, exceed k_α, where k_α is the same critical value as for the directed Wald test.

A GAUSS computer program is available from the author that calculates each of the above test statistics plus asymptotic p–values and critical values.

4.3. Assumptions

In this section, we state high-level assumptions under which the asymptotic results hold. Section 6 below gives one set of sufficient conditions for these high-level assumptions. All limits below are taken "as $T \to \infty$" unless stated otherwise. Let θ_0 denote the true value of θ under

the null H_0. We say that a statement holds "under θ_0" (i.e., under the null hypothesis) if it holds when the true density of y_T is $f_T(\theta_0)$ for $T = 1, 2, ...$ We introduce a sequence of local alternatives to the null parameter vector θ_0:

$$\theta_T = \theta_0 + h/\sqrt{T} \text{ for } T \geq 1 , \tag{4.5}$$

where $h \in R^s$. Of greatest interest are cases where h is such that $\theta_T \in \Theta^*$, but the asymptotic results given below do not require this. As stated in Section 2, the domain of the density functions $f_T(\theta)$ is (at least) $\Theta^* \cup \Theta$, where Θ is some set that contains a neighborhood of θ_0. The parametric model is assumed to be sufficiently regular that the following assumptions hold.

ASSUMPTION NL1: (a) *θ_0 is an interior point of Θ.*

(b) *$f_T(\theta)$ is twice continuously partially differentiable in θ for all $\theta \in \Theta_0$ with probability one under θ_0, where Θ_0 ($\subset \Theta$) is some neighborhood of θ_0.*

(c) *$-T^{-1}D^2\ell_T(\theta) \xrightarrow{p} \mathcal{I}(\theta)$ uniformly over $\theta \in \Theta_0$ under θ_0 for some non-random $s \times s$ matrix function $\mathcal{I}(\theta)$.*

(d) *$\mathcal{I}(\theta)$ is uniformly continuous on Θ_0.*

(e) *$\mathcal{I} = \mathcal{I}(\theta_0)$ is positive definite.*

ASSUMPTION NL2: *$T^{-1/2}D\ell_T(\theta_T) \xrightarrow{d} Z^* \sim N(0, \mathcal{I})$ under $\{\theta_T : T \geq 1\}$.*

ASSUMPTION NL3: *$\widehat{\theta} \xrightarrow{p} \theta_0$ under θ_0.*

ASSUMPTION NL4: *$\widetilde{\theta} \xrightarrow{p} \theta_0$ under θ_0.*

ASSUMPTION NL5: *For each $d \in \Delta$ and $b \in B$, $\exists \varepsilon_0 > 0$ such that $\forall \varepsilon \leq \varepsilon_0$ we have $d - \mathcal{I}_3^{-1}\mathcal{I}_2' b\varepsilon \in \Delta$, where $\mathcal{I} = \begin{bmatrix} \mathcal{I}_1 & \mathcal{I}_2' \\ \mathcal{I}_2 & \mathcal{I}_3 \end{bmatrix}$ and $\mathcal{I}_1 \subset R^{p \times p}$.*

We comment now on Assumptions NL1–NL5. Assumptions NL1(a), (b), (d), and (e) are fairly common ML regularity conditions.[3] Differentiability in θ is assumed for simplicity at the expense of some generality. As is well known, it is not needed for standard ML estimation results and undoubtedly could be relaxed here with some increase in complexity.

Assumption NL1(c) is a high-level assumption that requires a uniform weak law of large numbers (WLLN) to hold (since $-T^{-1}D^2\ell_T(\theta_0)$ can be written as a normalized sum of random variables by factoring the likelihood function using conditional distributions). The "uniformity" in Assumption NL1(c) can be established, e.g., by using the generic uniform convergence results of Andrews (1992). As stated, Assumption NL1(c) allows one to be relatively agnostic regarding the temporal dependence and heterogeneity of the data. To verify NL1(c), one needs to be more specific regarding these properties.

Assumption NL2 requires that the normalized score function satisfies a central limit theorem (CLT) (since $T^{-1/2}D\ell_T(\theta_T)$ can be written as a normalized sum of random variables that are mean zero under weak additional conditions). Assumptions NL3 and NL4 are not very restrictive. Given primitive sufficient conditions for Assumption 1, one typically needs few additional conditions to verify Assumptions NL3 and NL4.

Assumption NL5 is automatically satisfied if (1) the nuisance parameter space Δ is open or (2) the information matrix is block diagonal between β and δ (i.e., $\mathcal{I}_2 = 0$). In addition, Assumption NL5 is satisfied in a variety of special cases.

4.4. Asymptotic Results

The asymptotic distributions of the directed test statistics under the local alternatives $\{\theta_T : T \geq 1\}$ are given by

$$(4.6)\ \mathcal{L}_c(h) = \begin{cases} (1+c)^{-p/2}\exp\left[\frac{1}{2}\frac{c}{1+c}Z'Z\right]\Phi\Big(B, \frac{c}{1+c}\left(\mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2'\right)^{-1/2}Z, \\ \qquad \frac{c}{1+c}\left(\mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2'\right)^{-1}\Big) & \text{for } 0 < c < \infty \\ Z'Z + 2\log\left[\Phi\left(B,\ \left(\mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2'\right)^{-1/2}Z,\ \left(\mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2'\right)^{-1}\right)\right] & \text{for } c = \infty, \\ d'\left(\mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2'\right)^{-1/2}Z \Big/ \left[d'\left(\mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2'\right)^{-1}d\right]^{1/2} & \text{for } c = 0, \end{cases}$$

where $Z \sim N\left(\left(\mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2'\right)^{1/2}h_1,\ I_p\right)$, $h = (h_1',\ h_2')'$ for $h_1 \in R^p$, and $\mathcal{I} = \left(\begin{smallmatrix}\mathcal{I}_1 & \mathcal{I}_2 \\ \mathcal{I}_2' & \mathcal{I}_3\end{smallmatrix}\right)$ for $\mathcal{I}_1 \in R^{p\times p}$. Of course, the asymptotic distributions under the null are obtained by taking $h_1 = 0$. Note that the asymptotic null distribution of the $c = 0$ directed statistic, $\mathcal{L}_0(0)$, simplifies to the $N(0,1)$ distribution.

Next, to state the optimality properties of the directed tests, we introduce some additional notation. Let φ_T denote a test of H_0. The test φ_T is of asymptotic significance level α if $\int \varphi_T f_T(\theta_0) d\mu_T \to \alpha$ as $T \to \infty$ for all θ_0 in the null hypothesis H_0, where $\int \varphi_T f_T(\theta_0) d\mu_T$ denotes the probability of rejecting H_0 using φ_T when θ_0 is true. Similarly, the power of φ_T against the local alternative $f_T(\theta_0 + h/\sqrt{T})$ is denoted $\int \varphi_T f_T(\theta_0 + h/\sqrt{T}) d\mu_T$.

Let $\{k_{T\alpha} : T \geq 1\}$ be a sequence of critical values (possibly random, but with non-random probability limit) such that the directed Wald, LM, or LR test has asymptotic significance level α. Let ξ_{cT} denote a directed test, i.e., $\xi_{cT} = 1(D\text{–}W_{cT} > k_{T\alpha})$, $\xi_{cT} = 1(D\text{–}LM_{cT} > k_{T\alpha})$, or $\xi_{cT} = 1(D\text{–}LR_{cT} > k_{T\alpha})$ for $0 \leq c \leq \infty$.

The primary asymptotic properties of the directed tests are given in the following theorem:

THEOREM 3: (a) *Suppose Assumptions* NL1–NL4, 2, *and* 5 hold. Then, under the local alternatives $\{\theta_T : T \geq 1\}$, $D\text{–}W_{cT} \xrightarrow{d} \mathcal{L}_c(h)$, $D\text{–}LM_{cT} \xrightarrow{d} \mathcal{L}_c(h)$, and $D\text{–}LR_{cT} \xrightarrow{d} \mathcal{L}_c(h)$ for all $0 \leq c \leq \infty$.

(b) *Suppose Assumptions* NL1–NL5, 2, 4, *and* 5 *hold. Then, for any sequence of asymptotically level* α *tests* $\{\varphi_T : T \geq 1\}$*, a sequence of asymptotically level* α *directed Wald (LM or LR) tests* $\{\xi_{cT} : T \geq 1\}$ *satisfies*

$$\begin{aligned} &\overline{\lim_{T\to\infty}} \int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) \left[\int \varphi_T f_T(\theta_0 + h/\sqrt{T}) d\mu_T\right] dQ_c(h) \\ &\qquad \leq \underline{\lim_{T\to\infty}} \int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) \left[\int \xi_{cT} f_T(\theta_0 + h/\sqrt{T}) d\mu_T\right] dQ_c(h) \end{aligned}$$

for all $0 < c < \infty$*. (In addition, the* $\underline{\lim}_{T\to\infty}$ *on the right-hand side in part (b) equals* $\lim_{T\to\infty}$*.)*

The proofs of Theorem 3 and Theorem 4 below are extensions of the proofs of Theorems 1 and 2 for the exact Gaussian regression case to the asymptotic nonlinear models case. The proofs are given in Appendix B.

COMMENTS: 1. One can extend the scope of the results by taking a model parameterized by $\gamma \in \Gamma \in R^s$ with restrictions $H_0 : h(\gamma) = 0$, say, and transforming it into a model parameterized by $\theta = (\beta', \delta')' \in \Theta^* \subset R^s$ with $\beta = h(\gamma)$. For example, if $h(\gamma) = (\gamma_2 - \gamma_1, \gamma_3 - \gamma_2, \gamma_4 - \gamma_3)'$, then letting $\beta = h(\gamma)$ and $B = \{\beta \in R^3 : \beta_j \geq 0 \ \forall j \leq 3\}$ yields a test against the ordered alternative

H_1^* : $\gamma_1 \leq \gamma_2 \leq \gamma_3 \leq \gamma_4$ (without a nonnegativity constraint).

2. The assumption that B is positively homogeneous (Assumption 5) can be restrictive in some contexts. This assumption is not needed for the corresponding finite sample Gaussian linear regression results given above. It can be circumvented in the case of nonlinear models if one adopts a slightly different asymptotic framework than the usual one. In particular, suppose the sample size of interest is T^*. We embed the testing problem for sample size T^* in a sequence of testing problems indexed by $T \geq 1$ as follows: One changes $h/\sqrt{T}$ and Θ^* to $h/\sqrt{T^*/T}$ and $\Theta_T^* = B\sqrt{T^*/T} \times \Delta$, respectively, in the beginning of Section 4.2 and in (4.5) and B to $B\sqrt{T^*/T}$ in (4.4). For the sample size of interest T equals T^* and this has no effect on the definition of the test statistics. What the changes do is create an asymptotic framework in which the restricted alternative parameter space for β, viz., $B\sqrt{T^*/T}$, shrinks to zero at a suitable rate, yet equals the parameter space of interest B when $T = T^*$. Assumption 5 can now be dropped and Assumption 2 can be changed to: "$\Theta^* = \Theta_T^* = B\sqrt{T^*/T} \times \Delta$ for $B \subset R^p$ and $\Delta \subset R^q$, where B has positive Lebesgue measure." In (4.6), the set B remains as is—it is not changed to $B\sqrt{T^*/T}$. With the above changes, Theorem 3 holds with $h/\sqrt{T}$ and Θ^* changed to $h\sqrt{T^*/T}$ and Θ_T^* respectively.

5. NONLINEAR MODELS: THE LIKELIHOOD RATIO TEST

In this section we consider the LR test for the hypotheses given in (2.1) for dynamic nonlinear models. We show that the LR test is asymptotically admissible. In fact, we show that the LR test is asymptotically best, in a weighted average power sense, against alternatives that are sufficiently distant from the null hypothesis.

The LR test statistic $\mathcal{LR}_T$ is defined as in (3.1) with $\ell(\theta)$ replaced by $\ell_T(\theta)$ (defined in Section 4.1) and with the RA ML estimator θ^* ($\in \Theta^*$) defined such that $\ell_T(\theta^*) = \sup_{\theta \in \Theta^*} \ell_T(\theta)$ wp $\to 1$. We assume θ^* satisfies:

ASSUMPTION NL6: $\theta^* \xrightarrow{p} \theta_0$ *under* θ_0.

Given the previous assumptions, this assumption is not restrictive. It can be verified in the same

way as are Assumptions NL3 and NL4.

As in Section 3, we consider the case where Θ^* satisfies Assumptions 2, 5 (positive homogeneity), and 6 (no isolated rays in B) and the weight function is Q^r.

Let $\{k_T^\alpha : T \geq 1\}$ be a sequence of non-negative critical values (possibly random, but with non-random probability limit) such that the LR test, $\Lambda_T = 1(\mathcal{LR}_T > k_T^\alpha)$, has asymptotic significance level α. By definition, a sequence of tests $\{\varphi_T : T \geq 1\}$ is said to be *asymptotically distinct* from the sequence of asymptotically level α LR tests $\{\Lambda_T : T \geq 1\}$ if

$$\varlimsup_{T\to\infty} E_0(1-\varphi_T)\Lambda_T > 0 \ , \tag{5.1}$$

where E_0 denotes expectation under H_0.

Asymptotic admissibility of $\{\Lambda_T : T \geq 1\}$ is established in the following theorem.

THEOREM 4: (a) *Suppose Assumptions* NL1–NL4, NL6, 2, 5, *and* 6 *hold. Then, under the local alternatives* $\{\theta_T : T \geq 1\}$, $\mathcal{LR}_T \xrightarrow{d} (\max\{\sup_{\xi\in\mathcal{S}_p(B)} \xi' Z,\ 0\})^2$, *where* $Z \sim N\left(\left(\mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2'\right)^{1/2} h_1,\ I_p\right)$, *and* $h = (h_1', h_2')'$, *and* $\mathcal{S}_p(B) = \{\xi \in R^p : \|\xi\| = 1$ and $(A'\mathcal{I}A)^{-1/2}\xi \in B\}$.

(b) *Suppose Assumptions* NL1–NL6, 2, *and* 5–7 *hold. Then, for any sequence of tests* $\{\varphi_T : T \geq 1\}$ *that is asymptotically distinct from the sequence of asymptotically level* α *LR tests* $\{\Lambda_T : T \geq 1\}$, *there exists a constant* $r_0 < \infty$ *such that for all* $r \geq r_0$

$$\begin{aligned} &\varlimsup_{T\to\infty} \int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) \left[\int \varphi_T f_T(\theta_0 + h/\sqrt{T}) d\mu_T\right] dQ^r(h) \\ < &\varliminf_{T\to\infty} \int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) \left[\int \Lambda_T f_T(\theta_0 + h/\sqrt{T}) d\mu_T\right] dQ^r(h) \ . \end{aligned}$$

COMMENT: The comments following Theorem 2 apply here too.

6. NONLINEAR MODELS: PRIMITIVE SUFFICIENT CONDITIONS

In this section, we provide primitive sufficient conditions for Assumptions NL1–NL4 of Section 4 and Assumption NL6 of Section 5 for nonlinear dynamic models. For simplicity, we consider

strictly stationary m–th order Markov models. With some additional complexity in the assumptions, the results could be extended to allow for non-Markov models with non-stationary non-trending random variables.

The sample of observations is given by $\{(S_t, X_t) : t \leq T\}$, where $\{S_t : t \leq T\}$ are endogenous variables and $\{X_t : t \leq T\}$ are weakly exogenous variables. Let

$$\{g_t(\theta) : \theta \in \Theta^* \cup \Theta\} = \{g_t(S_t|S_1, ..., S_{t-1}; X_1, ..., X_t; \theta) : \theta \in \Theta^* \cup \Theta\} \tag{6.1}$$

denote a parametric family of conditional densities (with respect to some measure λ) of S_t given $S_1, ..., S_{t-1}, X_1, ..., X_t$ evaluated at the random variables $S_1, ..., S_t, X_1, ..., X_t$, where $\Theta^* \subset R^s$ and $\Theta \subset R^s$. Let

$$h_t = h_t(X_t|S_1, ..., S_{t-1}; X_1, ..., X_{t-1}) \tag{6.2}$$

denote the conditional density (with respect to some measure) of X_t given $S_1, ..., S_{t-1}, X_1, ..., X_{t-1}$ evaluated at the random variables $S_1, ..., S_{t-1}, X_1, ..., X_t$. By the assumption of weak exogeneity, h_t does not depend on θ. The log likelihood function $\ell_T(\theta)$ is given by $\ell_T(\theta) = \sum_{t=1}^T \log g_t(\theta)$ $+ \sum_{t=1}^T \log h_t$.

We consider the case where $\{(S_t, X_t) : t \geq 1\}$ is part of a doubly infinite strictly stationary ergodic sequence $\{(S_t, X_t) : t = ..., 0, 1, ...\}$ and $\{S_t : t = ..., 0, 1, ...\}$ is m–th order Markov for some integer $m \geq 0$. In this case, the function $\mathcal{I}(\theta)$ equals $-E\frac{\partial^2}{\partial\theta\partial\theta'} \log g_t(\theta)$. By definition, $\{S_t : t = ..., 0, 1, ...\}$ is m–th *order Markov* if the conditional distribution of S_t given $\mathcal{F}_{t-1} = \sigma(..., S_{t-2}, S_{t-1}; ..., X_{t-1}, X_t\}$ equals the conditional distribution of S_t given $S_{t,m} = (S_{t-m}, ..., S_{t-1})$ and $X_{t,m} = (X_{t-m}, ..., X_t)$ for all t. The Markov assumption yields the simplification that the summands $\log g_t(\theta)$ in the log-likelihood function are strictly stationary and ergodic for $t > m$. Without the Markov assumption this would not be the case, because the number of relevant observed variables in the conditioning set would vary with t.

The following assumption is sufficient for Assumptions NL1–NL4 and NL6:

ASSUMPTION A: (a) Θ *is compact and* θ_0 *lies in the interior of* Θ.

(b) $\{(S_t, X_t) : t = ..., 0, 1, ...\}$ *is strictly stationary and ergodic and* $\{S_t : t = ..., 0, 1, ...\}$ *is*

m–th order Markov under θ for each $\theta \in \Theta$.

(c) *$g_t(\theta)$ is continuous in θ on Θ and twice continuously partially differentiable in θ on Θ_0 with probability one under θ_0, where Θ_0 is some compact set that contains a neighborhood of θ_0.*

(d) *$g_t(\theta) \neq g_t(\theta_0)$ with positive probability under θ_0 $\forall \theta \in \Theta$ with $\theta \neq \theta_0$.*

(e) *$E \sup_{\theta\in\Theta} |\log g_t(\theta)| < \infty$, $E \sup_{\theta\in\Theta_0} \left\| \frac{\partial}{\partial\theta} \log g_t(\theta) \right\| < \infty$, $E \left\| \frac{\partial}{\partial\theta} \log g_t(\theta_0) \right\|^2 < \infty$, and $E \sup_{\theta\in\Theta_0} \left\| \frac{\partial^2}{\partial\theta\partial\theta'} \log g_t(\theta) \right\| < \infty$.*

(f) *$\mathcal{I} = -E \frac{\partial^2}{\partial\theta\partial\theta'} \log g_t(\theta_0)$ is positive definite.*

(The expectations in parts (e) and (f) are taken under θ_0.)

Assumption A constitutes a fairly standard set of ML regularity conditions for stationary and ergodic situations. Note that Assumption A imposes stationarity on $\{(S_t, X_t) : t \geq 1\}$ under θ for each fixed θ in Θ, but does not place such restrictions on sequences of local alternatives.

LEMMA A: *Assumption* A *implies Assumptions* NL1–NL4 *and* NL6.

The proof of Lemma A is given in Appendix B.

7. MONTE CARLO POWER COMPARISONS

In this section, we compare the power of several tests of hypotheses of the form (2.1) by Monte Carlo simulation.

7.1. Experimental Design

The model we consider is a p–variate normal location model with unknown mean β and known covariance matrix Ω:

$$Y \sim N(\beta,\ \Omega)\ . \tag{7.1}$$

The data consist of a single realization of $Y \in R^p$. The null and alternative hypotheses of interest are as in (2.1). Results for model (7.1) are more general than they might appear at first glance. First, a normal linear regression model (with known variance) can be written in the form of (7.1).

Second, the asymptotic local power of tests in a wide variety of nonlinear models equals their exact power in model (7.1).

To illustrate the first point, consider the linear regression model: $y = X\beta + G\delta + U$, $U \sim N(0, \sigma^2 I_n)$, where σ^2 is known. Premultiplication of the regression model by $(X'M_GX)^{-1}X'M_G$, yields a model of the form (7.1) with $Y = (X'M_GX)^{-1}X'M_Gy$ and $\Omega = \sigma^2(X'M_GX)^{-1}$.

To establish the second point, one sets $\Omega = \left(\mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2'\right)^{-1}/T$ and $\beta = h_1/\sqrt{T}$. Then, the exact power of the directed tests and LR test for the multivariate normal location model equal their asymptotic local power given in Theorems 3 and 4 for nonlinear models.[4]

We consider four different choices for the parameter space B:

$$
\begin{aligned}
B_1 &= \{\mu \in R^2 : \mu_1 \geq 0,\ \mu_2 \geq 0\}\ , && \text{where } \mu = (\mu_1, \mu_2)'\ , \\
B_2 &= \{\mu \in R^2 : \mu_1 \geq 0,\ \mu_2 \in R\}\ , && \text{where } \mu = (\mu_1, \mu_2)'\ , \\
B_3 &= \{\mu \in R^6 : \mu_j \geq 0,\ \forall j \leq 6\}\ , && \text{where } \mu = (\mu_1, ..., \mu_6)'\ , \text{ and} \\
B_4 &= \{\mu \in R^6 : \mu_j \geq 0,\ \forall j \leq 3,\ \mu_j \in R,\ \forall 4 \leq j \leq 6\}\ , && \text{where } \mu = (\mu_1, ..., \mu_6)'\ .
\end{aligned} \tag{7.2}
$$

Parameter spaces B_1 and B_3 correspond to multivariate one-sided hypotheses. Parameter spaces B_2 and B_4 correspond to mixed one- and two-sided hypotheses. For parameter spaces B_1 and B_2, we consider three different covariance matrices:

$$
\Omega_j = \begin{pmatrix} 1 & \rho_j \\ \rho_j & 1 \end{pmatrix} \quad \text{for } j = 1, 2, 3\ , \quad \text{where } \rho_1 = 0\ , \quad \rho_2 = .6\ , \quad \text{and } \rho_3 = -.6\ . \tag{7.3}
$$

Note that the correlation ρ_j in (7.3) corresponds to the correlation between the least squares estimators of different regression coefficients in the linear regression model. Clearly, in applications this correlation often ranges from -1 to 1. Thus, there is no a priori reason to give greater weight to the results corresponding to ρ_j equal to zero, or to ρ_j positive, than to ρ_j negative.

For parameter spaces B_3 and B_4, we consider the single covariance matrix

$$
\Omega_4 = I_6\ . \tag{7.4}
$$

Note that for the parameter space B_1, the models and hypotheses above are the same as those considered by Goldberger (1992).

The parameter spaces $B_1 - B_4$ of (7.2) are invariant under multiplication by positive definite diagonal matrices. In consequence, there is no loss in generality in (7.3) and (7.4) by taking the diagonal elements of Ω_j to be equal to unity—if Ω has non-unit diagonal elements, premultiplication of (7.1) by $\mathrm{Diag}^{1/2}(\Omega)$ yields a data vector with unit variances and leaves the hypotheses unchanged.

We now introduce the test statistics that will be considered: First, we define the directed test statistics in the context of model (7.1). The Wald, LM, and LR versions of these statistics are numerically identical for model (7.1). In consequence, it suffices to consider the directed Wald statistics. By definition,

$$D\text{–}W_c = \begin{cases} d'Y/(d'\Omega d)^{1/2} & \text{for } c = 0 \\ (1+c)^{-p/2}\exp\left(\frac{1}{2}\frac{c}{1+c}Y'\Omega^{-1}Y\right)\Phi\left(B, \frac{c}{1+c}Y, \frac{c}{1+c}\Omega\right) & \text{for } 0 < c < \infty \\ Y'\Omega^{-1}Y + 2\log[\Phi(B, Y, \Omega)] & \text{for } c = \infty\ . \end{cases} \tag{7.5}$$

Note that for B_1, B_2, B_3, and B_4, d is proportional to $(1,1)'$, $(1,0)'$, $(1,1,1,1,1,1)'$, and $(1,1,1,0,0,0)'$ respectively (and its length is irrelevant).

Below we report results for $c = 0,\ 1,\ \infty$. Results for $c = 1/3$ and $c = 3$ also are discussed, but are not tabulated for brevity.

The power of the test based on D–W_0 can be calculated exactly and, hence, need not be simulated. For a test of significance level α and true parameter value β, its power is given by $1 - \Phi_*(\Phi_*^{-1}(1-\alpha) - d'\beta/(d'\Omega d)^{1/2})$, where $\Phi_*(\cdot)$ and $\Phi_*^{-1}(\cdot)$ are the normal and inverse normal distribution functions respectively. For $c > 0$, critical values and power of the D–W_c test are calculated by simulation. The multivariate normal probability $\Phi(B, \cdot, \cdot)$ that appears in the expression for D–W_c is calculated by numerical integration since it reduces to univariate or bivariate normal probabilities in the cases considered here.

Second, we define the LR statistic for the model (7.1):

$$\mathcal{LR} = Y'\Omega^{-1}Y - \inf_{\beta \in B}(Y-\beta)'\Omega^{-1}(Y-\beta)\ . \tag{7.6}$$

Algebraic manipulations yield the following expressions for $\mathcal{LR}$ for the parameter spaces $B_1 - B_4$:

$$
\begin{aligned}
\mathcal{LR} &= Y'\Omega_j^{-1}Y1(Y_1>0,\ Y_2>0) - Y_1^2 1(Y_1<0,\ Y_2>\rho_j Y_1) \\
&\quad - Y_2^2 1(Y_2<0,\ Y_1>\rho_j Y_1) \quad \text{when} \quad B=B_1 \quad \text{and} \quad \Omega=\Omega_j\ , \\
\mathcal{LR} &= Y'\Omega_j^{-1}Y - Y_1^2 1(Y_1<0) \qquad \text{when} \quad B=B_2 \quad \text{and} \quad \Omega=\Omega_j\ , \\
\mathcal{LR} &= \sum_{i=1}^{6} Y_i^2 1(Y_i>0) \qquad \text{when} \quad B=B_3 \quad \text{and} \quad \Omega=\Omega_4\ , \quad \text{and} \\
\mathcal{LR} &= \sum_{i=1}^{3} Y_i^2 1(Y_i>0) + \sum_{i=4}^{6} Y_i^2 \quad \text{when} \quad B=B_4 \quad \text{and} \quad \Omega=\Omega_4\ .
\end{aligned} \tag{7.7}
$$

For convenience, critical values and power of the LR test are computed by simulation.

Third, we define the "two-sided Wald" ($2S$–W) test. This is the Wald (Lagrange multiplier, and likelihood ratio) test of $\mathrm{H}_0 : \beta = 0$ versus $\mathrm{H}_1 : \beta \neq 0$. By definition:

$$
2S\text{–}W = Y'\Omega^{-1}Y\ . \tag{7.8}
$$

Results for the $2S$–W test are included to quantify the magnitude of the power gains that occur when the information that $\beta \in B$ is exploited in the definition of the test. The D–W_c and LR tests exploit this information, whereas the $2S$–W test does not.

The power of the $2S$–W test can be calculated exactly. For significance level α and true parameter β, it equals $1 - \chi_p(\chi_p^{-1}(1-\alpha),\ \beta'\Omega^{-1}\beta)$, where $\chi_p(y,\, \lambda)$ is the noncentral chi-squared distribution function with p degrees of freedom and noncentrality parameter λ evaluated at y and $\chi_p^{-1}(\cdot)$ is the inverse of the central chi-squared distribution function with p degrees of freedom.

Fourth, we define test statistics that map out the envelope power function. For a given alternative parameter vector β_1, the Neyman–Pearson Lemma implies that the envelope power is given by the power of the likelihood ratio test of $\mathrm{H}_0 : \beta = 0$ versus $\mathrm{H}_1 : \beta = \beta_1$. This test statistic, denoted $ENV(\beta_1)$, is defined by

$$
ENV(\beta_1) = \beta_1'\Omega^{-1}Y/(\beta_1'\Omega^{-1}\beta_1)^{1/2}\ . \tag{7.9}
$$

Clearly, $ENV(\beta_1)$ depends only on the direction of β_1 from the origin and not on its distance from the origin. Results for the envelope power function are included to quantify the magnitude

of the power losses that occur for the D–W_c, LR, and $2S$–W tests due to the lack of knowledge of the direction of the true parameter value from the origin.

The power envelope can be calculated exactly. For a test of significance level α and true parameter value β_1, it equals $1 - \Phi_*(\Phi_*^{-1}(1-\alpha) - (\beta_1'\Omega^{-1}\beta_1)^{1/2})$.

The alternative parameter values for which the power of the above tests are computed are as follows: For B_1, β is taken proportional to (1,0)′ and (1,1)′; for B_2, β is taken proportional to (0,1)′, (1,1)′, and (1,0)′; for B_3, β is taken proportional to (1,1, ..., 1)′, (1,1,1,0,0,0)′, and (1,0, ..., 0)′; and for B_4, β is taken proportional to (1,1, ..., 1)′, (1,1,1,0,0,0)′, (1,0,0, ..., 0)′, (0,0,0,1,1,1)′, (0,0, ..., 0,1)′, and (1,0,0, ..., 0,1)′. For each direction of β, four distances from the origin are considered. These distances are chosen so that the LR test has powers .3, .5, .7, and .9. These choices ensure the chosen distances are reasonable and provide for easy comparisons between the D–W_c and LR tests.

We note that the power functions of the tests considered here display certain symmetries, which increase the generality of the results provided below. For parameter spaces B_1 and B_3, power is invariant under permutations of the elements of β. For B_2, power is invariant under changes in sign of the second element of β. For B_4, power is invariant under permutations of the first three elements of β, permutations of the last three elements of β, and changes in sign of the last three elements of β.

Thirty thousand repetitions are used in the simulation of the critical values and power of the D–W_c and LR tests for $0 < c \leq \infty$. All calculations were carried out using the GAUSS computer program on a 486–66MHz PC.

7.2. Simulation Results

Table II gives the power of the D–W_c, LR, and $2S$–W tests and the envelope power function for the parameter space $B_1 = R_+^2$ for all eight values of β and all three values of ρ_j. The first feature of the table to notice is that the relative powers of the tests are not sensitive to the distance of the alternative from the null (at least within the range considered). In consequence, the average

power of each test over the four distances considered summarizes the results well. These averages are given in the fifth column of numbers in the table. This insensitivity is quite interesting given the theoretical derivation of the D–W_0, D–W_1, D–W_∞, and LR tests as tests that are better for alternatives that are progressively more distant from the null.

The relative powers of the D–W_c and LR tests vary much more with the direction of departure from the null and the value of the correlation coefficient than with the distance from the null. Table II provides six different direction/ρ_j combinations. The average power results for these six cases show that the D–W_∞ test is best overall. It is best in three of the six cases, is within .02 of the best in five of the six cases, and is within .04 of the best in all cases. The test D–W_1 is a close second in overall performance. The LR and D–W_0 tests both suffer from poor relative performances in two or more cases. This is especially true of the D–W_0 test, which does very well in the middle of B_1 (i.e., $\beta \propto (1,1)'$)—in fact, it attains the envelope power there—but sacrifices considerable power at the boundaries of B_1.

Table III presents power results for the parameter space $B_2 = R_+ \times R$. In this case too, the relative powers of the tests are insensitive to the distance of the alternative from the null. In consequence, for brevity, we do not report the results for all distances, but rather, just give averages of the powers over the four distances that yield the LR test to have powers .3, .5, .7, and .9. Thus, the results of Table III are analogous to those of column five of Table II. Analogous average power summary statistics are given in Tables IV and V (discussed below) for the same reasons.

The results of Table III cover three different directions of the alternative from the null and three different values of the correlation ρ_j. Note that the direction (1,0) is in the middle of B_2, direction (0,1) is on the boundary of B_2, and direction (1,1) is between the two. Several features of Table III are worthy of note. First, the results for D–W_∞, LR, $2S$–W, and the envelope power function (the last four rows of the table) are insensitive to the value of ρ_j. In fact, D–W_0 is the only test that shows significant sensitivity to ρ_j. This makes the comparison of the tests much simpler. Second, the D–W_0 test has disastrous power along the edge of B_2—its power equals

its size for direction (0,1) for all distances from the null and for all ρ_j. Third, the D–W_∞ test almost dominates the D–W_1 test. Thus, the D–W_∞ and LR tests are the best overall tests with the D–W_∞ test being preferable unless one places great weight on performances at or near the boundary of B_2. The D–W_0 test is not to be recommended as an omnibus test.

Table IV presents results for the high dimensional one-sided parameter space $B_3 = R^6_+$ with the single covariance matrix $\Omega_4 = I_6$. In this case, D–W_1 and D–W_∞ are the best tests overall. They both dominate LR and almost dominate D–W_0. They perform very much better than $2S$–W and are at or near (within .02 of) the power envelope in the middle of the parameter space.

Table V presents results for the high dimensional mixed one- and two-sided parameter space $B_4 = R^3_+ \times R^3$. As with parameter space B_2, the test D–W_0 has very poor overall power properties, since its power equals its size for the fourth and fifth directions, which lie on the boundary of B_4. The D–W_1 and D–W_∞ tests have similar power, although D–W_∞ is somewhat better due to its performance for the fourth and fifth directions. The comparison between D–W_∞ and LR is much the same as with the low dimensional mixed one- and two-sided parameter space B_2. That is, D–W_∞ typically does better for alternatives that are in the middle of the parameter space, but worse for alternatives that are on the boundary.

The poor performance of D–W_0 for alternatives on the boundary of the parameter spaces B_2 and B_4 may seem puzzling. This test is the limiting test of a sequence of tests that maximize weighted average power where the weight functions place increasingly great weight on alternatives close to the null. The explanation of the puzzle seems to be that for alternatives very close to the null, all tests have power almost equal to size in all directions, so the drawback of the D–W_0 test in certain directions is a relatively minor one that can be compensated for by high power in other directions. As soon as one considers alternatives that are not very close to the null, the deficiency of the D–W_0 test in certain directions is glaring and cannot be compensated for by high power in other directions.

For brevity, Tables II–V report results for only three values of c, viz., 0, 1, and ∞. Power calculations for c equal 1/3 and 3 also were carried out. The results for these tests lie between

those of the $c = 0$, 1, and ∞ tests. In particular, the monotonicity of power as a function of c, which is evident in the tables, also holds when the results for $c = 1/3$ and $c = 3$ are added. For those cases where D–W_0 has power equal to size, the power of D–$W_{1/3}$ is very much closer to that of D–W_1 than D–W_0. Overall, for those cases where D–W_0 does not have power equal to size, the power of the D–W_c tests is not very sensitive to the choice of c. For those cases where D–W_0 has size equal to power, there is a substantial difference between D–W_0 and D–W_c for $c > 0$ and larger values of c are preferable.

For a summary of the simulation results, see the Introduction.

APPENDIX A

This Appendix provides proofs of Theorems 1 and 2.

The proof of Theorem 1(a) uses the following Lemmas. Let

$$\text{(A.1)} \qquad \overline{LR}(\theta_0) = \exp\left[\tfrac{1}{2}\overline{\theta}'\mathcal{I}\overline{\theta}\right] \int 1(\theta_0 + h \in \Theta^*) \exp\left[-\tfrac{1}{2}(\overline{\theta} - h)'\mathcal{I}(\overline{\theta} - h)\right] dQ_c(h)/K \ ,$$

where $\overline{\theta} = \widehat{\theta} - \theta_0$ and $\widehat{\theta}$ is the (unrestricted) LS estimator of θ (i.e., $\widehat{\theta} = (\sigma^2\mathcal{I})^{-1}[X \vdots Z]'Y$).

LEMMA 1: *Under Assumption* 1, $LR(\theta_0) = \overline{LR}(\theta_0)$.

LEMMA 2: *The projection matrix $P^{\perp}$ onto the orthogonal complement $V^{\perp}$ of V with respect to $< \cdot, \cdot >$ is given by $P^{\perp} = AH$, where $A = [I_p \vdots -\mathcal{I}_2\mathcal{I}_3^{-1}]'$ and $H = [I_p \vdots 0] \in R^{p \times s}$.*

PROOF OF THEOREM 1: First we establish part (a). By Lemma 1, it suffices to show that $\overline{LR}(\theta_0) = D - W_c/K$. To do so, let $\lambda \sim N(0,\ c(A'\mathcal{I}A)^{-1})$ and $h = A\lambda$. Then, $h \sim Q_c = N(0,\ cA(A'\mathcal{I}A)^{-1}A')$ as desired. The statistic $\overline{LR}(\theta_0)$ can be written as

$$\text{(A.2)} \qquad \begin{aligned} \overline{LR}(\theta_0) \ &= \ \textstyle\int_B (2\pi)^{-p/2} \det{}^{1/2}(A'\mathcal{I}A/c) \\ &\qquad \times \exp\left(\tfrac{1}{2}[\overline{\theta}'\mathcal{I}\overline{\theta} - (\overline{\theta} - A\lambda)'\mathcal{I}(\overline{\theta} - A\lambda) - \lambda'A'\mathcal{I}A\lambda/c]\right) d\lambda/K \ , \end{aligned}$$

since $\theta_0 + h \in \Theta^*$ iff $\lambda \in \{b \in R^p : \theta_0 + Ab \in \Theta^*\} = \{b \in R^p : \left(\begin{smallmatrix} b \\ \delta_0 - \mathcal{I}_3^{-1}\mathcal{I}_2'b \end{smallmatrix}\right) \in B \times \Delta\} = B$, where the second equality holds under Assumption 3.

Let P and $P^{\perp}$ denote the projection matrices with respect to $\langle \cdot, \cdot \rangle$ onto V and $V^{\perp}$ respectively. The term in square brackets in the exponent on the rhs of (A.2), with $A\lambda$ replaced by h for simplicity, now simplifies as follows:

$$\text{(A.3)} \qquad \begin{aligned} &\overline{\theta}'\mathcal{I}\overline{\theta} - (\overline{\theta} - h)'\mathcal{I}(\overline{\theta} - h) - h'\mathcal{I}h/c \\ = \ &\tfrac{c}{1+c}(P^{\perp}\overline{\theta})'\mathcal{I}P^{\perp}\overline{\theta} - \left(h - P^{\perp}\overline{\theta}\tfrac{c}{1+c}\right)'\mathcal{I}\tfrac{1+c}{c}\left(h - P^{\perp}\overline{\theta}\tfrac{c}{1+c}\right) \ , \end{aligned}$$

using the fact that $(P\overline{\theta})'\mathcal{I}h = 0 \ \forall h \in V^{\perp}$.

By Lemma 2, $P^{\perp}\overline{\theta} = AH\overline{\theta} = A\widehat{\beta}$. In addition, $A\mathcal{I}A' = \mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2' = X'M_GX/\sigma^2$. Thus, the rhs of (A.3) equals

$$\text{(A.4)} \qquad \tfrac{c}{1+c}\widehat{\beta}'(X'M_GX/\sigma^2)\widehat{\beta} - \left(\lambda - \widehat{\beta}\tfrac{c}{1+c}\right)'\tfrac{1+c}{c}(X'M_GX/\sigma^2)\left(\lambda - \widehat{\beta}\tfrac{c}{1+c}\right) \ .$$

Combining (A.2), (A.3), and (A.4) gives the desired result $\overline{LR}(\theta_0) = D\text{–}W_c/K$.

Part (b) holds because $\widehat{\beta} = (X'M_GX)^{-1}X'M_GU \sim N(0, (X'M_GX)^{-1}\sigma^2)$ under H_0, where $U = (U_1, ..., U_T)'$.

Part (c) holds by (2.7), the Neyman–Pearson Lemma, and Theorem 1(a). □

PROOF OF LEMMA 1: Let $\ell(\theta) = \log f(Y, \theta)$ and $D\ell(\theta) = \frac{\partial}{\partial\theta}\ell(\theta)$. Simple algebra yields $D\ell(\theta_0) = (U'X, U'Z)'$ and $\overline{\theta} = \mathcal{I}^{-1}D\ell(\theta_0)$. We can write

$$LR(\theta_0) = \int 1(\theta_0 + h \in \Theta^*)\exp(\ell(\theta_0 + h) - \ell(\theta_0))dQ_c(h)/K \ . \tag{A.5}$$

Let $h = (h_1', h_2')'$ for $h_1 \in R^p$ and $h_2 \in R^q$. The integrand of (A.5) simplifies as follows:

$$\begin{aligned} & \ell(\theta_0 + h) - \ell(\theta_0) \\ = & -\tfrac{1}{2\sigma^2}\left[\Sigma_1^T(Y_t - X_t'h_1 - G_t'(\delta_0 + h_2))^2 - \Sigma_1^T(Y_t - G_t'\delta_0)^2\right] \\ = & -\tfrac{1}{2\sigma^2}\left[-2\Sigma_1^T(Y_t - G_t'\delta_0)(X_t'h_1 + G_t'h_2) + \Sigma_1^T(X_t'h_1 + G_t'h_2)^2\right] \\ = & \ D\ell(\theta_0)'h - \tfrac{1}{2}h'\mathcal{I}h = \overline{\theta}\mathcal{I}h - \tfrac{1}{2}h'\mathcal{I}h = \tfrac{1}{2}\overline{\theta}'\mathcal{I}\overline{\theta} - \tfrac{1}{2}(\overline{\theta} - h)'\mathcal{I}(\overline{\theta} - h) \ . \end{aligned} \tag{A.6}$$

Combining (A.5) and (A.6) gives the desired result $LR(\theta_0) = \overline{LR}(\theta_0)$. □

PROOF OF LEMMA 2: It suffices to show that (1) $AHd = 0 \ \forall d \in V$ and (2) $AHm = m \ \forall m \in V^\perp$. To show (1), note that $d \in V$ iff $d = (0', d_2')'$ for some $d_2 \in R^q$. Thus, $AHd = A[I_p \vdots 0]\begin{pmatrix} 0 \\ d_2 \end{pmatrix} = 0$. To show (2), note that $m \in V^\perp$ iff $d'\mathcal{I}m = 0 \ \forall d \in V$ iff $[0 \vdots I_q]\mathcal{I}m = 0$ iff $[\mathcal{I}_2' \vdots \mathcal{I}_3] = \begin{pmatrix} m_1 \\ m_2 \end{pmatrix} = 0$, where $m = (m_1', m_2')'$, iff $m_2 = -\mathcal{I}_3^{-1}\mathcal{I}_2'm_1$ for $m_1 \in R^p$ and $m_2 \in R^q$. Thus, for $m \in V^\perp$, $AHm = \begin{bmatrix} I_p & 0 \\ -\mathcal{I}_3^{-1}\mathcal{I}_2' & 0 \end{bmatrix}\begin{bmatrix} m_1 \\ -\mathcal{I}_3^{-1}\mathcal{I}_2'm_1 \end{bmatrix} = \begin{bmatrix} m_1 \\ -\mathcal{I}_3^{-1}\mathcal{I}_2'm_1 \end{bmatrix} = m$, as desired. □

The proof of Theorem 2 parts (a) and (b) follows from the following Lemmas 3–6. The proof of Theorem 2 part (c) is given below. It utilizes Lemmas 3–7.

LEMMA 3: *Under Assumptions* 1–3 *and* 5, $\mathcal{LR} = \beta^{*\prime}A'\mathcal{I}A\beta^*$.

LEMMA 4: *Under Assumptions* 1, 2, 5, *and* 7, $LR(\theta_0, Q^r) = \int_{\mathcal{S}_p(B)} \exp[-r^2/2 + r\xi'(X'M_GX/\sigma^2)^{1/2}\widehat{\beta}]$ $dU(\xi)/K$.

LEMMA 5: *Under Assumptions* 2, 5, *and* 6,

$$\lim_{r\to\infty} \sup_{\lambda\in R^p:\|\lambda\|=1} \left| \tfrac{1}{r} \log \int_{\mathcal{S}_p(B)} \exp(r\xi'\lambda) dU(\xi)/K - \sup_{\xi\in\mathcal{S}_p(B)} \xi'\lambda \right| = 0 \ .$$

LEMMA 6: *Under Assumptions* 1, 2, *and* 5, $\max\{\sup_{\xi\in\mathcal{S}_p(B)} \xi'(A'\mathcal{I}A)^{1/2}\widehat{\beta}, 0\} = (\beta^{*\prime}A'\mathcal{I}A\beta^*)^{1/2}$.

LEMMA 7: *Let* φ *be a test of* H_0. *Under Assumptions* 1–3 *and* 5, *if* $E_0(1-\varphi)1(\mathcal{LR} > k_\alpha) > 0$, *then* $\exists\gamma > 0$, $\varepsilon > 0$, *and* $\delta > 0$ *such that*

$$E_0(1-\varphi)1(\mathcal{LR} > k_\alpha + 2\gamma,\ \varepsilon \le \|(X'M_GX/\sigma^2)^{1/2}\widehat{\beta}\| \le \varepsilon^{-1}) \ge \delta \ ,$$

where E_0 *denotes expectation under the null density* $f(y, \theta_0)$.

PROOF OF LEMMA 3: The likelihood function $\ell(\theta)$ equals

$$\begin{aligned} \ell(\theta) &= C - (\widehat{\theta}-\theta)'\mathcal{I}(\widehat{\theta}-\theta) \\ &= C - (P^\perp\widehat{\theta} - P^\perp\theta)'\mathcal{I}(P^\perp\widehat{\theta} - P^\perp\theta) - (P\widehat{\theta} - P\theta)'\mathcal{I}(P\widehat{\theta} - P\theta) \\ &= C - (\widehat{\beta}-\beta)'A'\mathcal{I}A(\widehat{\beta}-\beta) - (P\widehat{\theta} - P\theta)'\mathcal{I}(P\widehat{\theta} - P\theta) \ , \end{aligned} \tag{A.7}$$

where C is a constant, $P^\perp$ and P are as in the proof of Theorem 1, and the third equality uses Lemma 2. By Assumption 3, the imposition of the null hypothesis does not restrict the parameter space of $P\theta = \theta - AH\theta = (0', \delta' - \beta'\mathcal{I}_2\mathcal{I}_3^{-1})'$. Thus, $P\theta^* = P\widetilde{\theta}$. In addition, $P^\perp\widetilde{\theta} = \widetilde{\beta} = 0$ by definition. Hence,

$$\mathcal{LR} = \widehat{\beta}'A'\mathcal{I}A\widehat{\beta} - (\widehat{\beta}-\beta^*)'A'\mathcal{I}A(\widehat{\beta}-\beta^*) = \beta^{*\prime}A'\mathcal{I}A\beta^* \ , \tag{A.8}$$

where the second equality uses the orthogonality condition $(\widehat{\beta}^*-\beta)'A'\mathcal{I}A\beta^* = 0$, which holds because β^* minimizes $(\widehat{\beta}–\beta)'A'\mathcal{I}A(\widehat{\beta}–\beta)$ over the positively homogeneous set B (e.g., see Perlman (1969, Lemma 4.1(i))). □

PROOF OF LEMMA 4: By Lemma 1, $LR(\theta_0, Q^r)$ equals $\overline{LR}(\theta_0)$ defined in (A.1) with Q_c replaced by Q^r. The log of the integrand of (A.1) (including the term $\exp\left(\frac{1}{2}\overline{\theta}'\mathcal{I}\overline{\theta}\right)$) can be rewritten as

$$\tfrac{1}{2}\overline{\theta}'\mathcal{I}\overline{\theta} - \tfrac{1}{2}(h-\overline{\theta})'\mathcal{I}(h-\overline{\theta}) = \tfrac{1}{2}(P^\perp\overline{\theta})'\mathcal{I}P^\perp\overline{\theta} - \tfrac{1}{2}(h-P^\perp\overline{\theta})'\mathcal{I}(h-P^\perp\overline{\theta}) \ , \tag{A.9}$$

using the fact that $h'\mathcal{I}P\overline{\theta} = 0\ \forall h \in V^{\perp}$. Also, $P^{\perp}\overline{\theta} = AH\overline{\theta} = A\widehat{\beta}$ using Lemma 2. Let $\lambda = (A'\mathcal{I}A)^{1/2}\widehat{\beta}$. Let $h = rA(A'\mathcal{I}A)^{-1/2}\xi$, where ξ is a random p–vector with uniform distribution on the unit sphere. Then, $h \sim Q^r$ and

$$\begin{aligned} LR(\theta_0, Q^r) &= \int 1(\theta_0 + h \in \Theta^*) \exp\left[\tfrac{1}{2}\widehat{\beta}'A'\mathcal{I}A\widehat{\beta} - \tfrac{1}{2}(h - A\widehat{\beta})'\mathcal{I}(h - A\widehat{\beta})\right] dQ^r(h)/K \\ &= \int_{\mathcal{S}_p(B)} \exp\left[\tfrac{1}{2}\lambda'\lambda - \tfrac{1}{2}(r\xi - \lambda)'(r\xi - \lambda)\right] dU(\xi)/K \qquad \text{(A.10)} \\ &= \int_{\mathcal{S}_p(B)} \exp\left[-\tfrac{1}{2}r^2 + r\xi'\lambda\right] dU(\xi)/K\ . \quad \square \end{aligned}$$

Proof of Lemma 5: The supremum over λ in Lemma 5 is attained for each r. Hence, it suffices to show: For some λ_r with $\|\lambda_r\| = 1$,

$$\lim_{r\to\infty}\left[\tfrac{1}{r}\log\int_{\mathcal{S}_p(B)} \exp(r\xi'\lambda_r)dU(\xi)/K - \sup_{\xi_*\in\mathcal{S}_p(\lambda)} \xi_*'\lambda_r\right] = 0\ . \qquad \text{(A.11)}$$

By replacing $\xi'\lambda_r$ with $\sup_{\xi_*\in\mathcal{S}_p(B)}\xi_*'\lambda_r$, it is easy to see that the ℓhs of (A.11) is less than or equal to zero. To show the reverse inequality, let $\varepsilon > 0$ be given and choose a subset of $\mathcal{S}_p(B)$ for each r, say Ξ_r, for which $\xi'\lambda_r \geq \sup_{\xi_*\in\mathcal{S}_p(B)}\xi_*'\lambda_r - \varepsilon\ \forall\xi \in \Xi_r$ and $\int_{\Xi_r} dU(\xi) \geq \delta$ for some $\delta > 0$. Such sets exist by Assumption 6, since the latter is equivalent to: $\forall\varepsilon > 0$, $\inf_{\xi\in\mathcal{S}_p(B)}\int 1(\|\xi - \xi_*\| \leq \varepsilon,$ $\xi_* \in \mathcal{S}_p(B))dU(\xi_*) > 0$. The ℓhs of (A.11) is greater than or equal to

$$\underline{\lim}_{r\to\infty}\left[\tfrac{1}{r}\log\int_{\Xi_r}\exp(r[\sup_{\xi_*\in\mathcal{S}_p(B)}\xi_*'\lambda_r - \varepsilon])dU(\xi) - \sup_{\xi\in\mathcal{S}_p(B)}\xi'\lambda_r\right] = -\varepsilon\ . \qquad \text{(A.12)}$$

Since $\varepsilon > 0$ is arbitrary, the ℓhs of (A.12) is ≥ 0. $\square$

Proof of Lemma 6: Let $\overline{S}$ denote the closure of a set S. Any $\beta \in \overline{B}$ can be written as $\beta = c(A'\mathcal{I}A)^{-1/2}\xi$ for some $\xi \in \overline{\mathcal{S}}_p(B)$ and $c \geq 0$. In particular, there exists $c^* \geq 0$ and $\xi^* \in \overline{\mathcal{S}}_p(B)$ such that $\beta^* = c^*(A'\mathcal{I}A)^{-1/2}\xi^*$. Note that $\beta^{*\prime}A'\mathcal{I}A\beta^* = (c^*)^2$. By (A.7), β^* minimizes $(\widehat{\beta} - \beta)'A'\mathcal{I}A(\widehat{\beta} - \beta)$ over $\beta \in \overline{B}$. The latter holds if and only if (c^*, ξ^*) minimizes

$$\widehat{\beta}'A'\mathcal{I}A\widehat{\beta} - 2c\xi'(A'\mathcal{I}A)^{1/2}\widehat{\beta} + c^2 \ \text{ over } \ (c, \xi) \in [0, \infty) \times \overline{\mathcal{S}}_p(B)\ . \qquad \text{(A.13)}$$

To solve (A.13), we first maximize $\xi'(A'\mathcal{I}A)^{1/2}\widehat{\beta}$ over $\xi \in \overline{\mathcal{S}}_p(B)$ and then minimize

$$-2c\sup_{\xi\in\mathcal{S}_p(B)}\xi'(A'\mathcal{I}A)^{1/2}\widehat{\beta} + c^2 \ \text{ over } \ c \geq 0\ . \qquad \text{(A.14)}$$

If $\sup_{\xi \in \mathcal{S}_p(B)} \xi'(A'\mathcal{I}A)^{1/2}\widehat{\beta} > 0$, then we obtain $c^* = \sup_{\xi \in \mathcal{S}_p(B)} \xi'(A'\mathcal{I}A)^{1/2}\widehat{\beta}$. Combining this with the above expression for $(c^*)^2$ gives $\sup_{\xi \in \mathcal{S}_p(B)} \xi'(A'\mathcal{I}A)^{1/2}\widehat{\beta} = (\beta^{*\prime}A'\mathcal{I}A\beta^*)^{1/2}$. If $\sup_{\xi \in \mathcal{S}_p(B)} \xi'(A'\mathcal{I}A)^{1/2}\widehat{\beta} \leq 0$, we obtain $c^* = 0$, $\beta^* = 0$, and $(\beta^{*\prime}A'\mathcal{I}A\beta^*)^{1/2} = 0$. □

PROOF OF THEOREM 2(c): Let E_r and P_r denote expectations and probabilities, respectively, under the density $\int 1(\theta_0 + h \in \Theta^*) f(y, \theta_0 + h) dQ^r(h)/K$ for $r \geq 0$ (where for $r = 0$ the density is $f(y, \theta_0)$). It suffices to show that

$$E_r(1-\varphi)/P_r(\mathcal{LR} \leq k_\alpha) \to \infty \text{ as } r \to \infty \,. \tag{A.15}$$

Below we show that

$$P_r(\mathcal{LR} \leq k_\alpha) \leq \exp(-r^2/2 + r k_\alpha^{1/2}) \ \ \forall r > 0 \,. \tag{A.16}$$

We also show that for $\gamma > 0$ and $\delta > 0$ as in Lemma 7

$$E_r(1-\varphi) \geq \delta \, \exp(-r^2/2 + r(k_\alpha + \gamma)^{1/2}) \tag{A.17}$$

for r sufficiently large. Equation (A.15) follows immediately from (A.16) and (A.17).

We now establish (A.16). Let $\widehat{\widehat{\beta}} = \left(X'M_G X/\sigma^2\right)^{1/2} \widehat{\beta}$. We have

$$\begin{aligned}
P_r(\mathcal{LR} \leq k_\alpha) &= E_0 LR(\theta_0, Q^r) 1(\mathcal{LR} \leq k_\alpha) \\
&= \exp(-r^2/2) E_0 \int_{\mathcal{S}_p(B)} \exp(r\xi'\widehat{\widehat{\beta}}) dU(\xi) 1(\mathcal{LR} \leq k_\alpha)/K \\
&\leq \exp(-r^2/2) E_0 \exp(r \sup_{\xi \in \mathcal{S}_p(B)} \xi'\widehat{\widehat{\beta}}) 1(\mathcal{LR} \leq k_\alpha) \\
&\leq \exp(-r^2/2) E_0 \exp\left(r(\mathcal{LR})^{1/2}\right) 1(\mathcal{LR} \leq k_\alpha) \\
&\leq \exp(-r^2/2 + r k_\alpha^{1/2}) \,,
\end{aligned} \tag{A.18}$$

where the second equality holds by Lemma 4 and the second inequality holds by Lemmas 3 and 6.

Next, we establish (A.17). Let γ, ε, and δ be as in Lemma 7. Let $\widehat{\widehat{\beta}}_1 = \widehat{\widehat{\beta}}/\|\widehat{\widehat{\beta}}\|$. We have

$$\begin{aligned}
&E_r(1-\varphi) \geq E_0(1-\varphi) LR(\theta_0, Q^r) 1(\mathcal{LR} > k_\alpha + 2\gamma, \, \varepsilon \leq \|\widehat{\widehat{\beta}}\| \leq \varepsilon^{-1}) \\
&= \exp(-r^2/2) E_0(1-\varphi)[\int_{\mathcal{S}_p(B)} \exp(r\|\widehat{\widehat{\beta}}\| \xi'\widehat{\widehat{\beta}}_1) dU(\xi) 1(\mathcal{LR} > k_\alpha + 2\gamma, \, \varepsilon \leq \|\widehat{\widehat{\beta}}\| \leq \varepsilon^{-1})/K],
\end{aligned} \tag{A.19}$$

where the equality uses Lemma 4. Now, we use Lemma 5 to establish a lower bound on the term in square brackets on the right-hand side of (A.19). For λ such that $\|\lambda\| = 1$ and $\sup_{\xi \in \mathcal{S}_p(B)} \xi'\lambda > 0$, define $c(r, \lambda)$ to satisfy

$$\int_{\mathcal{S}_p(B)} \exp(r\xi'\lambda) dU(\xi)/K = \exp(rc(r, \lambda) \sup_{\xi \in \mathcal{S}_p(B)} \xi'\lambda) \ . \tag{A.20}$$

By Lemma 5, $\forall \zeta > 0$,

$$\lim_{r \to \infty} \sup\{|c(r, \lambda) - 1| \ : \ \|\lambda\| = 1, \ \sup_{\xi \in \mathcal{S}_p(B)} \xi'\lambda \geq \zeta\} = 0 \ . \tag{A.21}$$

In consequence, $\exists r_\gamma < \infty$ such that $\forall r \geq r_\gamma$ and $\forall \lambda$ with $\|\lambda\| = 1$ and $\sup_{\xi \in \mathcal{S}_p(B)} \xi'\lambda \geq (k_\alpha + 2\gamma)^{1/2}\varepsilon$, we have $c(r, \lambda) \geq ((k_\alpha + \gamma)/(k_\alpha + 2\gamma))^{1/2}$. Then, since $\sup_{\xi \in \mathcal{S}_p(B)} \xi'\widehat{\widehat{\beta}}_1 = (\mathcal{LR})^{1/2}/\|\widehat{\widehat{\beta}}\|$, we obtain: $\forall r \geq r_\gamma/\varepsilon$,

$$\begin{aligned} & c(r\|\widehat{\widehat{\beta}}\|, \ \widehat{\widehat{\beta}}_1) 1(\mathcal{LR} > k_\alpha + 2\gamma, \ \varepsilon \leq \|\widehat{\widehat{\beta}}\| \leq \varepsilon^{-1}) \\ & \quad \geq ((k_\alpha + \gamma)/(k_\alpha + 2\gamma))^{1/2} 1(\mathcal{LR} > k_\alpha + 2\gamma, \ \varepsilon \leq \|\widehat{\widehat{\beta}}\| \leq \varepsilon^{-1}) \ . \end{aligned} \tag{A.22}$$

By (A.20) and the substitution of (A.22) into (A.19), we get: $\forall r \geq r_\gamma/\varepsilon$,

$$\begin{aligned} E_r(1-\varphi) \ &\geq \ \exp(-r^2/2) E_0(1-\varphi) \exp(r\|\widehat{\widehat{\beta}}\| c(r\|\widehat{\widehat{\beta}}\|, \widehat{\widehat{\beta}}_1) \sup_{\xi \in \mathcal{S}_p(B)} \xi'\widehat{\widehat{\beta}}_1) \\ & \qquad \times 1(\mathcal{LR} > k_\alpha + 2\gamma, \ \varepsilon \leq \|\widehat{\widehat{\beta}}\| \leq \varepsilon^{-1}) \\ &\geq \ \exp(-r^2/2) E_0(1-\varphi) \exp\left(r((k_\alpha + \gamma)/(k_\alpha + 2\gamma))^{1/2} (\mathcal{LR})^{1/2}\right) \\ & \qquad \times 1(\mathcal{LR} > k_\alpha + 2\gamma, \varepsilon \leq \|\widehat{\widehat{\beta}}\| \leq \varepsilon^{-1}) \\ &\geq \ \delta \exp(-r^2/2 + r(k_\alpha + \gamma)^{1/2}) \ , \end{aligned} \tag{A.23}$$

where the last inequality uses Lemma 7. □

Proof of Lemma 7: The lemma follows from absolute continuity with respect to Lebesgue measure of $\|(X'M_G X/\sigma^2)^{1/2}\widehat{\beta}\|$ and of $\mathcal{LR}$ on $(0, \infty)$, where the latter follows from the expression $\mathcal{LR}^{1/2} = \max\{\sup_{\xi \in \mathcal{S}_p(B)} \xi'(A'\mathcal{I}A)^{1/2}\widehat{\beta}, 0\}$. □

APPENDIX B

This Appendix provides proofs of Theorems 3 and 4 and Lemma A.

For notational simplicity, we suppress the subscript c on Q_c in the proofs below.

The proof of Theorem 3 is similar to that of Theorems 1 and 2 of Andrews and Ploberger (1994). The idea of the proof is to show that the directed Wald, LM, and LR test statistics are asymptotically equivalent, under the null and local alternatives, to the Neyman–Pearson likelihood ratio $\ell r_c(\theta_0)$, defined below. The proof uses the following definitions, lemmas, and theorems: Let

$$
\begin{aligned}
&\ell r_c(\theta_0) = \left[\int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) f_T(\theta_0 + h/\sqrt{T}) dQ(h)/K\right] / f_T(\theta_0) \text{and} \\
&\overline{\ell r_c} = \int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) \exp\left[\tfrac{1}{2}\overline{\theta}\mathcal{I}\overline{\theta} - \tfrac{1}{2}(\overline{\theta} - h)'\mathcal{I}(\overline{\theta} - h)\right] dQ(h)/K \ , \quad \text{where} \\
&K = \int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) dQ(h) = \int 1(\lambda \in B) dN(\lambda), \, \overline{\theta} = \mathcal{I}^{-1} D\ell_T(\theta_0)/\sqrt{T} \ ,
\end{aligned}
\tag{B.1}
$$

and $N(\cdot)$ denotes a $N(0, c(A'\mathcal{I}A)^{-1})$ distribution function. Note that $\ell r_c(\theta_0)$ is the Neyman–Pearson likelihood ratio statistic for testing the simple null hypothesis that $\underset{\sim}{Y}_T \sim f_T(\theta_0)$ against the simple alternative that $\underset{\sim}{Y}_T \sim \int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) f_T(\theta_0 + h/\sqrt{T}) dQ(h)/K$.

LEMMA B–1: *Under the null hypothesis and Assumptions* NL1–NL3, $\sqrt{T}(\widehat{\theta} - \theta_0) - \overline{\theta} \xrightarrow{p} 0$.

LEMMA B–2: *Under the null hypothesis and Assumptions* NL1–NL3, $\ell r_c(\theta_0) - \overline{\ell r_c} \xrightarrow{p} 0$ $\forall 0 < c < \infty$.

LEMMA B–3: *The projection matrix $P^{\perp}$ onto the orthogonal complement $V^{\perp}$ of V with respect to $\langle \cdot, \cdot \rangle_{\mathcal{I}}$ is given by $P^{\perp} = AH$, where $A = [I_p \vdots -\mathcal{I}_2\mathcal{I}_3^{-1}]'$ and $H = [I_p \vdots 0] \in R^{p \times s}$.*

LEMMA B–4: *Under Assumptions* NL1 *and* NL2, *the densities $\{f_T(\theta_0 + h/\sqrt{T}) : T \geq 1\}$ are contiguous to the densities $\{f_T(\theta_0) : T \geq 1\}$ for all $h \in R^s$.*

The proofs of Lemmas B–1 to B–3 are analogous to those of Lemmas A–1 to A–3 of Andrews and Ploberger (1994). For brevity, they are not given here. The proof of Lemma B–4 is given below.

For $0 < c < \infty$, define

$$\begin{aligned}
&\overline{W}_T = (H\overline{\theta})' \left(H\mathcal{I}^{-1}H'\right)^{-1} H\overline{\theta} \ , \\
&D\text{–}\overline{W}_{cT} = (1+c)^{-p/2} \exp\left(\tfrac{1}{2}\tfrac{c}{1+c}\overline{W}_T\right)\Phi\left(B,\ \tfrac{c}{1+c}\overline{\beta}/\sqrt{T},\ \tfrac{c}{1+c}\left(\mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2'\right)^{-1}/T\right) \ , \\
&D\text{–}\overline{W}_{\infty T} = \overline{W}_T + 2\log\left[\Phi\left(B, \overline{\beta}/\sqrt{T},\ \left(\mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2'\right)^{-1}/T\right)\right] \ , \text{ and} \\
&D\text{–}\overline{W}_{0T} = d'\overline{\beta}/\left(d'\left(\mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2'\right)^{-1} d\right)^{1/2} \ , \text{ where } d = \tfrac{\partial}{\partial\mu}\Phi\left(B,\ 0,\ \left(\mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2'\right)^{-1}/T\right) .
\end{aligned} \tag{B.2}$$

THEOREM B–1: *Under the local alternatives* $\{\theta_T : T \geq 1\}$ *and Assumptions* NL1–NL4, 2, *and* 5, *for all* $0 < c < \infty$, *we have*
(a) $\ell r_c(\theta_0) - \overline{\ell r_c} \xrightarrow{p} 0$, (b) $\overline{\ell r_c} \times K = D\text{–}\overline{W}_{cT}$ *provided Assumption* NL5 *also holds,* (c) $D\text{–}\overline{W}_{cT} - D\text{–}W_{cT} \xrightarrow{p} 0$, (d) $D\text{–}W_{cT} - D\text{–}LM_{cT} \xrightarrow{p} 0$, *and* (e) $D\text{–}LM_{cT} - D\text{–}LR_{cT} \xrightarrow{p} 0$. *In addition, parts* (c), (d), *and* (e) *hold for* $c = 0$ *and* $c = \infty$.

The proof of Theorem N–1 is given below.

PROOF OF THEOREM 3(a): By the positive homogeneity of B (Assumption 5),

$$\Phi\left(B,\ \tfrac{c}{1+c}\overline{\beta}/\sqrt{T},\ \tfrac{c}{1+c}\left(\mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2'\right)^{-1}/T\right) = \Phi\left(B,\ \tfrac{c}{1+c}\overline{\beta},\ \tfrac{c}{1+c}\left(\mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2'\right)^{-1}\right) . \tag{B.3}$$

Theorem 3(a) now follows for all $0 \leq c \leq \infty$ from the combination of (i) Theorem B–1 parts (c)–(e), (ii) $\overline{\theta} \xrightarrow{d} N(h, \mathcal{I}^{-1})$ under $\{\theta_T : T \geq 1\}$, (iii) the continuity of $D\text{–}\overline{W}_{cT}$ as a function of $\overline{\theta}$ and $\mathcal{I}$, and (iv) the continuous mapping theorem. Condition (ii) holds because

$$\begin{aligned}
\overline{\theta} &= \mathcal{I}^{-1} D\ell_T(\theta_0)/\sqrt{T} \\
&= \mathcal{I}^{-1} D\ell_T(\theta_T)/\sqrt{T} + (\mathcal{I}^{-1} D^2\ell_T(\ddot{\theta}_T)/\sqrt{T})(-h/\sqrt{T}) \\
&\xrightarrow{\text{d}} N(h,\ \mathcal{I}^{-1}) \text{ under } \{\theta_T : T \geq 1\},
\end{aligned} \tag{B.4}$$

where the second equality holds by element-by-element mean value expansions, $\ddot{\theta}_T$ lies between θ_0 and θ_T, $-D^2\ell_T(\ddot{\theta}_T)/T \xrightarrow{\text{p}} \mathcal{I}$ under θ_0 and under $\{\theta_T : T \geq 1\}$ using Assumption NL1 and contiguity, and the convergence in distribution uses Assumption NL2. □

PROOF OF LEMMA B–4: We make use of the following result, which follows, e.g., from Thms. 16.8

and 18.1 of Strasser (1985): If (i) $f_T(\theta_T)/f_T(\theta_0) \xrightarrow{\mathrm{d}} X$ under θ_0 for some random variable X and (ii) $EX = 1$, then the densities $\{f_T(\theta_0) : T \geq 1\}$ are contiguous to the densities $\{f_T(\theta_0) : T \geq 1\}$. To obtain condition (i), we do a two-term Taylor expansion of $\ell_T(\theta_T)$ about θ_0:

$$\begin{aligned} f_T(\theta_T)/f_T(\theta_0) &= \exp\left(\ell_T(\theta_T) - \ell_T(\theta_0)\right) \\ &= \exp\left(T^{-1/2} D\ell_T(\theta_0)'h + \tfrac{1}{2} h' D^2 \ell_T(\ddot{\theta}_T) h\right) \\ &\xrightarrow{\mathrm{d}} \exp\left(Z^{*\prime} h - \tfrac{1}{2} h' \mathcal{I} h\right) \text{ under } \theta_0, \end{aligned} \tag{B.5}$$

where $\ddot{\theta}_T$ lies between θ_T and θ_0, $Z^* \sim N(0, \mathcal{I})$, and the convergence in distribution uses Assumptions NL1 and NL2.

Condition (ii) holds because $E \exp(Z^{*\prime} h) = \exp(\frac{1}{2} h' \mathcal{I} h)$ using the formula for the moment generating function of a normal random variable. □

Proof of Theorem B–1: By contiguity (Lemma B–4), it suffices to show that the Theorem holds under the null. In consequence, all probabilistic statements in this proof are made "under θ_0." Part (a) holds by Lemma B–2. Next, consider part (b). Let $\lambda \sim N\left(0, c(A'\mathcal{I}A)^{-1}\right)$ and $h = A\lambda$. Then, $h \sim Q = N\left(0, cA(A'\mathcal{I}A)^{-1}A'\right)$ as desired. Note that $1(\theta_0 + A\lambda/\sqrt{T} \in \Theta^*)$ $= 1(\lambda/\sqrt{T} \in B)$ by Assumption NL5. Thus, we have

$$\begin{aligned} \overline{\ell r_c} &= (2\pi)^{-p/2} \det{}^{1/2}(A'\mathcal{I}A/c) \\ &\quad \times \int 1(\lambda/\sqrt{T} \in B) \exp\Big(\tfrac{1}{2}[\overline{\theta}'\mathcal{I}\overline{\theta} - (A\lambda - \overline{\theta})'\mathcal{I}(A\lambda - \overline{\theta}) - (A\lambda)'\mathcal{I}A\lambda/c]\Big)\, d\lambda/K \ , \end{aligned} \tag{B.6}$$

where $\det(\cdot)$ denotes the determinant operator.

Let P and $P^\perp$ denote the projection matrices with respect to $\langle \cdot, \cdot \rangle$ onto V and $V^\perp$ respectively. Using some algebra, the term in square brackets in (B.6), with $A\lambda$ replaced by h for simplicity, can be shown to simplify as follows:

$$\begin{aligned} &\overline{\theta}'\mathcal{I}\overline{\theta} - (h - \overline{\theta})'\mathcal{I}(h - \overline{\theta}) - h'\mathcal{I}h/c \\ &= \tfrac{c}{1+c}(P^\perp\overline{\theta})'\mathcal{I}P^\perp\overline{\theta} - \left[h - P^\perp\overline{\theta}\tfrac{c}{1+c}\right]' \mathcal{I}\tfrac{1+c}{c}\left[h - P^\perp\overline{\theta}\tfrac{c}{1+c}\right] \ , \end{aligned} \tag{B.7}$$

using the fact that $h'\mathcal{I}\ell = 0 \ \forall h \in V^\perp$, $\ell \in V$.

Combining (B.6) and (B.7) yields

$$\begin{aligned}\overline{\ell r_c} \times K &= (1+c)^{-p/2} \exp\left[\tfrac{1}{2}\ \tfrac{c}{1+c}(P^{\perp}\overline{\theta})'\mathcal{I}P^{\perp}\overline{\theta}\right] \cdot \int 1(\lambda/\sqrt{T} \in B)(2\pi)^{-p/2} \\ &\quad \times\ \det{}^{1/2}\left[A'\mathcal{I}A\tfrac{1+c}{c}\right] \exp\left[-\tfrac{1}{2}\left[A\lambda - P^{\perp}\overline{\theta}\tfrac{c}{1+c}\right]'\mathcal{I}\tfrac{1+c}{c}\left[A\lambda - P^{\perp}\overline{\theta}\tfrac{c}{1+c}\right]\right] d\lambda \\ &= (1+c)^{-p/2} \exp\left[\tfrac{1}{2}\ \tfrac{c}{1+c}(H\overline{\theta})'A'\mathcal{I}AH\overline{\theta}\right]\Phi\Big(B,\ \tfrac{c}{1+c}H\overline{\theta}/\sqrt{T},\ \tfrac{c}{1+c}(A'\mathcal{I}A)^{-1}/T\Big) \\ &= (1+c)^{-p/2} \exp\left[\tfrac{1}{2}\ \tfrac{c}{1+c}\overline{W}_T\right]\Phi\Big(B,\ \tfrac{c}{1+c}\overline{\beta}/\sqrt{T},\ \left(\mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2\right)^{-1}/T\Big)\ ,\end{aligned} \tag{B.8}$$

where the second equality uses $P^{\perp} = AH$ (Lemma B–3) and a change of variables ($\lambda^* = \lambda/\sqrt{T}$) and the third equality uses $(A'\mathcal{I}A)^{-1} = H\mathcal{I}^{-1}H' = \left(\mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2\right)^{-1}$. This completes the proof of part (b).

Part (c) is established as follows. Since $H\theta_0 = 0$,

$$\sqrt{T}H\widehat{\theta} - H\overline{\theta} = \sqrt{T}\widehat{\beta} - \overline{\beta} \xrightarrow{p} 0 \tag{B.9}$$

by Lemma B–1. In addition,

$$\|\mathcal{I}_T(\widehat{\theta}) - \mathcal{I}\| \leq \sup_{\theta\in\Theta_0}\|\mathcal{I}_T(\theta) - \mathcal{I}(\theta)\| + \|\mathcal{I}(\widehat{\theta}) - \mathcal{I}\| = o_p(1)\ , \tag{B.10}$$

where the inequality holds wp $\to 1$ using Assumption NL3 and the equality holds by Assumptions NL1(c), NL1(d), and NL3. Given Assumption NL1(e), this establishes part (c) for all $0 \leq c \leq \infty$.

For parts (d) and (e) with $0 \leq c \leq \infty$, it suffices to show that

$$\begin{aligned}&\text{(i)} \quad W_T - LM_T \xrightarrow{p} 0 \ \text{ and } \ \mathrm{LM_T} - \mathrm{LR_T} \xrightarrow{\mathrm{p}} 0\ , \\ &\text{(ii)} \quad \sqrt{T}\widehat{\beta} - H\widetilde{\mathcal{I}}^{-1}D\ell_T(\widetilde{\theta})/\sqrt{T} \xrightarrow{p} 0\ , \\ &\text{(iii)} \quad \left(\widehat{\mathcal{I}}_1 - \widehat{\mathcal{I}}_2\widehat{\mathcal{I}}_3^{-1}\widehat{\mathcal{I}}_2'\right)^{-1} - \left(\mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2'\right)^{-1} \xrightarrow{p} 0 \ \text{ and} \\ &\qquad\quad \left(\widehat{\mathcal{I}}_1 - \widehat{\mathcal{I}}_2\widehat{\mathcal{I}}_3^{-1}\widehat{\mathcal{I}}_2'\right)^{-1} - \left(\widetilde{\mathcal{I}}_1 - \widetilde{\mathcal{I}}_2\widetilde{\mathcal{I}}_3^{-1}\widetilde{\mathcal{I}}_2'\right)^{-1} \xrightarrow{p} 0\ .\end{aligned} \tag{B.11}$$

Condition (iii) holds by (B.10) and Assumption NL1(e). Condition (i) is a standard result and its proof is similar to proofs in the literature. For brevity, its proof is omitted. Condition (ii) typically is established as part of the proof of condition (i). Again, for brevity, its proof is omitted. □

We now turn to the proof of Theorem 3(b). First, we establish another *contiguity* result.

LEMMA B–5: *Under Assumptions* NL1–NL3, NL5, 2, 4, *and* 5, *the densities* $\{\int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) f_T(\theta_0 + h/\sqrt{T}) dQ(h)/K : T \geq 1\}$ *are contiguous to the densities* $\{f_T(\theta_0) : T \geq 1\}$ *for all* $0 < c < \infty$.

Lemma B–5 and Theorem B–1 combine to give:

THEOREM B–2: *Under the local alternative densities* $\{\int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) f_T(\theta_0 + h/\sqrt{T}) dQ(h)/K : T \geq 1\}$ *and Assumptions* NL1–NL5, 2, 4, *and* 5, *for all* $0 < c < \infty$, *we have* (a) $\ell r_c(\theta_0) - \overline{\ell r_c} \xrightarrow{p} 0$, (b) $\overline{\ell r_c} \times K = D\text{–}\overline{W}_{cT}$, (c) $D\text{–}\overline{W}_{cT} - D\text{–}W_{cT} \xrightarrow{p} 0$, (d) $D\text{–}W_{cT} - D\text{–}LM_{cT} \xrightarrow{p} 0$, *and* (e) $D\text{–}LM_{cT} - D\text{–}LR_{cT} \xrightarrow{p} 0$.

PROOF OF THEOREM 3(b): Let α_T be the rejection probability of φ_T under θ_0. Let $k^*_{\alpha_T} > 0$ and $\lambda_T \in [0,1]$ be constants such that the likelihood ratio test $\gamma_T = 1(\ell r_c(\theta_0) > k^*_{\alpha_T}) + \lambda_T 1(\ell r_c(\theta_0) = k^*_{\alpha_T})$ has rejection probability α_T under θ_0. Then, by the Neyman–Pearson Lemma (e.g., see Lehmann (1959, Thm. 3.1, p. 65), for all $T \geq 1$,

$$\begin{aligned} &\int \varphi_T \left[\int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) f_T(\theta_0 + h/\sqrt{T}) dQ(h)/K \right] d\mu_T \\ \leq\ &\int \gamma_T \left[\int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) f_T(\theta_0 + h/\sqrt{T}) dQ(h)/K \right] d\mu_T \ . \end{aligned} \tag{B.12}$$

By Corollary 15.11 of Strasser (1985), the $\underline{\lim}_{T\to\infty}$ on the rhs of the inequality in the statement of Theorem 3(b) is actually $\lim_{T\to\infty}$, because $\Delta_2(E_T, E) \to 0$ as $T \to \infty$ (in his notation) by the proof of Lemma B–5 below.

This result, inequality (B.12), and Fubini's Theorem yield

$$\begin{aligned} &\overline{\lim_{T\to\infty}} \int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) \left[\int \varphi_T f_T(\theta_0 + h/\sqrt{T}) d\mu_T \right] dQ(h)/K \\ \leq\ &\lim_{T\to\infty} \int \gamma_T \left[\int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) f_T(\theta_0 + h/\sqrt{T}) dQ(h)/K \right] d\mu_T \\ =\ &\lim_{T\to\infty} \int 1(\ell r_c(\theta_0) > k_{T\alpha}/K) \left[1(\theta_0 + h/\sqrt{T} \in \Theta^*) f_T(\theta_0 + h/\sqrt{T}) dQ(h)/K \right] d\mu_T \\ =\ &\lim_{T\to\infty} \int 1(D\text{–}W_{cT} > k_{T\alpha}) \left[1(\theta_0 + h/\sqrt{T} \in \Theta^*) f_T(\theta_0 + h/\sqrt{T}) dQ(h)/K \right] d\mu_T \\ =\ &\lim_{T\to\infty} \int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) \left[\int \xi_{cT} f_T(\theta_0 + h/\sqrt{T}) d\mu_T \right] dQ(h)/K \ , \end{aligned} \tag{B.13}$$

where the first equality holds because $k^*_{\alpha_T} - k_{T\alpha}/K \xrightarrow{p} 0$ and $\ell r_c(\theta_0)$ has an absolutely continuous asymptotic distribution under $\int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) dQ(h)/K$, the second equality holds because

$\ell r_c(\theta_0) = D\text{–}W_{cT}/K \xrightarrow{p} 0$ under $\int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) f_T(\theta_0 + h/\sqrt{T}) dQ(h)/K$ by Theorem B–2, and the third inequality holds by Fubini's Theorem. The proof is analogous for $D\text{–}LM_{cT}$ and $D\text{–}LR_{cT}$. □

PROOF OF LEMMA B–5: As in the proof of Lemma B–4, it suffices to show that (i) $\ell r_c(\theta_0) \xrightarrow{d} \mathcal{L}_c/K$ under θ_0 and (ii) $E\mathcal{L}_c/K = 1$. Condition (i) holds by Theorem B–1(a)–(c) and Theorem 3(a). Condition (ii) is obtained as follows: Let $\mathcal{V} = \left(\mathcal{I}_1 - \mathcal{I}_2\mathcal{I}_3^{-1}\mathcal{I}_2'\right)^{-1}$. Then, we have

$$
\begin{aligned}
E\mathcal{L}_c/K &= (1+c)^{-p/2} \int_{R^p} \exp\left(\tfrac{1}{2}\tfrac{c}{1+c} Z'Z\right) \int_B (2\pi)^{-p/2} \det^{-1/2}(\mathcal{V}) \\
&\quad \times \exp\left[-\tfrac{1}{2}\left(\lambda - \left(\tfrac{c}{1+c}\mathcal{V}\right)^{1/2} Z\right)' \mathcal{V}^{-1} \left(\lambda - \left(\tfrac{c}{1+c}\mathcal{V}\right)^{1/2} Z\right)\right] d\lambda \\
&\quad \times (2\pi)^{-p/2} \exp\left(-\tfrac{1}{2} Z'Z\right) dZ/K \\
&= (1+c)^{-p/2} \int_B (2\pi)^{-p/2} \det^{-1/2}(\mathcal{V}) \int_{R^p} (2\pi)^{-p/2} \\
&\quad \times \exp\left(-\tfrac{1}{2}\left[\tfrac{1}{1+c} Z'Z + \left(Z - \left(\tfrac{1+c}{c}\right)^{1/2} \mathcal{V}^{-1/2}\lambda\right)'\right.\right. \\
&\quad \times \left.\left.\left(Z - \left(\tfrac{1+c}{c}\right)^{1/2} \mathcal{V}^{-1/2}\lambda\right) \tfrac{c}{1+c}\right]\right) dz d\lambda/K \ ,
\end{aligned}
\tag{B.14}
$$

where the second equality uses Fubini's theorem. Via some algebra, the expression in square brackets in the exponent of the rhs of (B.14) equals

$$
\left(Z - \left(\tfrac{c}{1+c}\right)^{1/2} \mathcal{V}^{-1/2}\lambda\right)' \left(Z - \left(\tfrac{c}{1+c}\right)^{1/2} \mathcal{V}^{-1/2}\lambda\right) + \tfrac{1}{1+c}\lambda'\mathcal{V}^{-1}\lambda \ . \tag{B.15}
$$

Substituting this result in (B.14) gives

$$
\begin{aligned}
E\mathcal{L}_c/K &= (1+c)^{-p/2} \int_B (2\pi)^{-p/2} \det^{-1/2}(\mathcal{V}) \exp\left(-\tfrac{1}{2}\ \tfrac{1}{1+c}\lambda'\mathcal{V}^{-1}\lambda\right) d\lambda/K \\
&= \Phi_p(B,\, 0,\, (1+c)\mathcal{V})/K = 1 \ ,
\end{aligned}
\tag{B.16}
$$

where the first equality holds because the integral of a normal density equals one and the third equality holds because $K = \int 1(\lambda \in B) dN(\lambda) = \Phi_p(B,\, 0,\, c\mathcal{V}) = \Phi_p(B,\, 0,\, (1+c)\mathcal{V})$. □

The proof of Theorem 4 uses the following lemmas:

LEMMA B–6: *Under Assumptions* NL1–NL6, 2, *and* 5,

$$
\mathcal{LR}_T - T\beta_q^{*\prime} A'\mathcal{I}A\beta_q^* \xrightarrow{p} 0 \ \text{ under } \ \theta_0 \ ,
$$

where β_q^* *is such that* $(\widehat{\beta} - \beta_q^*)' A'\mathcal{I}A(\widehat{\beta} - \beta_q^*) = \inf_{\beta \in B}(\widehat{\beta} - \beta)' A'\mathcal{I}A(\widehat{\beta} - \beta)$.

LEMMA B–7: *Under Assumptions* NL1–NL3, NL5, 2, 5, *and* 7,

$$\mathcal{LR}_T(\theta_0, Q^r) - \int_{\mathcal{S}_p(B)} \exp(-r^2/2 + r\xi'(A'\mathcal{I}A)^{1/2}\sqrt{T}\widehat{\beta})dU(\xi)/K \xrightarrow{p} 0 \text{ under } \theta_0 .$$

LEMMA B–8: *Under Assumptions* 2, 5, *and* 6,

$$\lim_{r\to\infty} \sup_{\lambda\in R^p:\|\lambda\|=1} \left| \tfrac{1}{r} \log \int_{\mathcal{S}_p(B)} \exp(r\xi'\lambda)dU(\xi)/K - \sup_{\xi\in\mathcal{S}_p(B)} \xi'\lambda \right| = 0 .$$

LEMMA B–9: *Under Assumptions* NL1, 2, *and* 5,

$$\max\{\sup_{\xi\in\mathcal{S}_p(B)} \xi'(A'\mathcal{I}A)^{1/2}\widehat{\beta}, 0\} = (\beta_q^{*\prime}A'\mathcal{I}A\beta_q^*)^{1/2} \quad \forall T \geq 1 .$$

LEMMA B–10: *Under Assumptions* NL1–NL6, 2, *and* 5,

$$\mathcal{LR}_T \xrightarrow{d} (\max\{\sup_{\xi\in\mathcal{S}_p(B)} \xi'Z, 0\})^2 \text{ under } \theta_0 , \text{ where } Z \sim N(0, I_p) \text{ and}$$

$\sup_{\xi\in\mathcal{S}_p(B)} \xi'Z$ *has absolutely continuous distribution.*

LEMMA B–11: *Under Assumptions* NL1–NL3, NL5, 2, 5, *and* 7, *the densities* $\{\int 1(\theta_0 + h/\sqrt{T} \in \Theta^*)f_T(\theta_0 + h/\sqrt{T})dQ^r(h)/K : T \geq 1\}$ *are contiguous to the densities* $\{f_T(\theta_0) : T \geq 1\}$ *for all* $0 < r < \infty$.

PROOF OF THEOREM 4: Part (a) follows from Lemma B–10.

Next, we prove part (b). For simplicity we consider the case where k_T^α equals a constant $k_\alpha \geq 0$ $\forall T \geq 1$. For the case of random k_T^α, we must have $k_T^\alpha \xrightarrow{p} k_\alpha$ for some constant $k_\alpha \geq 0$ by Lemma B–10 and the corresponding adjustments to the proof are minor. Throughout, let $\widehat{\widehat{\beta}} = \sqrt{T}(A'\mathcal{I}A)^{1/2}\widehat{\beta}$ and $\mathcal{LR}_T^q = T\beta_q^{*\prime}A'\mathcal{I}A\beta_q^*$.

To prove Theorem 4(b), it suffices to show that

$$\lim_{T\to\infty} E_r(1 - \varphi_T) / \overline{\lim_{T\to\infty}} P_r(\mathcal{LR}_T \leq k_\alpha) \to \infty \text{ as } r \to \infty . \tag{B.17}$$

Below we show that

$$\overline{\lim_{T\to\infty}} P_r(\mathcal{LR}_T \leq k_\alpha) \leq 2\exp(-r^2/2 + rk_\alpha^{1/2}) \quad \forall r > 0 . \tag{B.18}$$

We also show that for some $\gamma > 0$ and $\delta > 0$,

$$\underline{\lim}_{T\to\infty} E_r(1-\varphi_T) \geq \tfrac{\delta}{4}\exp(-r^2/2 + r(k_\alpha+\gamma)^{1/2}) \tag{B.19}$$

for r sufficiently large. Equation (B.17) follows immediately from (B.18) and (B.19).

We now establish (B.18). Define the event D^*_{Tr} by

$$D^*_{Tr} = \left\{ \exp(-r^2/2)\left[\int_{\mathcal{S}_p(B)} \exp(r\xi'\widehat{\widehat{\beta}})dU(\xi)/K\right]/LR_T(\theta_0, Q^r) \in [1/2, 2]\right\} . \tag{B.20}$$

By Lemmas B–7 and B–11, $\lim_{T\to\infty} P_r(D^*_{Tr}) = 1\ \forall r \geq 0$. In addition, $\mathcal{LR}_T - \mathcal{LR}^q_T \xrightarrow{p} 0$ under P_r by Lemmas B–6 and B–11.

Using these results, we obtain: $\forall r > 0$,

$$\begin{aligned}
&\overline{\lim_{T\to\infty}}\ P_r(\mathcal{LR}_T \leq k_\alpha) \\
=\ &\overline{\lim_{T\to\infty}}\ P_r(\mathcal{LR}^q_T \leq k_\alpha, D^*_{Tr}) \\
=\ &\overline{\lim_{T\to\infty}}\ E_0 LR_T(\theta_0, Q^r)\cdot 1(\mathcal{LR}^q_T \leq k_\alpha, D^*_{Tr}) \\
\leq\ &2\exp(-r^2/2)\,\overline{\lim_{T\to\infty}}\ E_0\int_{\mathcal{S}_p(B)} \exp(r\xi'\widehat{\widehat{\beta}})dU(\xi)1(\mathcal{LR}^q_T \leq k_\alpha)/K \\
\leq\ &2\exp(-r^2/2)\,\overline{\lim_{T\to\infty}}\ E_0\exp(r\sup_{\xi\in\mathcal{S}_p(B)} \xi'\widehat{\widehat{\beta}})1(\mathcal{LR}^q_T \leq k_\alpha) \\
\leq\ &2\exp(-r^2/2)\,\overline{\lim_{T\to\infty}}\ E_0\exp\left(r(\mathcal{LR}^q_T)^{1/2}\right)1(\mathcal{LR}^q_T \leq k_\alpha) \\
\leq\ &2\exp(-r^2/2 + rk_\alpha^{1/2}) ,
\end{aligned} \tag{B.21}$$

where the third inequality holds by Lemma B–9. Note that the first equality of (B.21) actually relies on the results above plus Lemmas B–10 and B–11.

Next, we establish (B.19). The fact that φ_T and Λ_T are asymptotically distinct implies that $\exists\gamma > 0$ and $\varepsilon > 0$ such that

$$\underline{\lim}_{T\to\infty} E_0(1-\varphi_T)1(\mathcal{LR}_T > k_\alpha + 2\gamma,\ \varepsilon \leq \|\widehat{\widehat{\beta}}\| \leq \varepsilon^{-1}) \geq \delta/2 , \tag{B.22}$$

where $\delta = \underline{\lim}_{T\to\infty} E_0(1-\varphi_T)1(\mathcal{LR}_T > k_\alpha)$. This follows because the left-hand sides of (B.22) and (5.1) differ by less than

$$\overline{\lim_{T\to\infty}}\ P_0(\mathcal{LR}_T \in (k_\alpha,\ k_\alpha + 2\gamma]) + \overline{\lim_{T\to\infty}}\ P_0(\|\widehat{\widehat{\beta}}\| \in [0,\varepsilon)\cup(\varepsilon^{-1},\ \infty)) \leq \delta/2 , \tag{B.23}$$

where the inequality holds for some small $\gamma > 0$ and $\varepsilon > 0$ by Lemma B–10 and the fact that $\sqrt{T}\widehat{\beta} = H\sqrt{T}(\widehat{\theta} - \theta) = H\overline{\theta} \xrightarrow{d} N(0,\, H\mathcal{I}^{-1}H')$ by Lemma B–1 and Assumption NL2.

Let

$$D_{Tr} = \{\mathcal{LR}_T^q > k_\alpha + 2\gamma,\, \varepsilon \leq \|\widehat{\widehat{\beta}}\| \leq \varepsilon^{-1}\} \cap D_{Tr}^* \ . \tag{B.24}$$

By Lemmas B–6, B–7, and B–10 and (B.22), we obtain

$$\lim_{T\to\infty} E_0(1-\varphi_T)1(D_{Tr}) \geq \delta/2 \ . \tag{B.25}$$

Let $\widehat{\widehat{\beta}}_1 = \widehat{\widehat{\beta}}/\|\widehat{\widehat{\beta}}\|$. We have

$$\begin{aligned} \lim_{T\to\infty} E_r(1-\varphi_T) &\geq \lim_{T\to\infty} E_0(1-\varphi_T) LR_T(\theta_0,\, Q^r) 1(D_{Tr}) \\ &\geq \tfrac{1}{2}\exp(-r^2/2) \lim_{T\to\infty} E_0(1-\varphi_T) \left[\int_{\mathcal{S}_p(B)} \exp(r\|\widehat{\widehat{\beta}}\|\xi'\widehat{\widehat{\beta}}_1) dU(\xi)/K\right] 1(D_{Tr}) \ . \end{aligned} \tag{B.26}$$

Now, we use Lemma B–8 to establish a lower bound on the term in square brackets on the right-hand side of (B.26). For λ such that $\|\lambda\| = 1$ and $\sup_{\xi\in\mathcal{S}_p(B)} \xi'\lambda > 0$, define $c(\lambda,\, r)$ to satisfy

$$\int_{\mathcal{S}_p(B)} \exp(r\xi'\lambda) dU(\xi)/K = \exp(rc(r,\, \lambda) \sup_{\xi\in\mathcal{S}_p(B)} \xi'\lambda) \ . \tag{B.27}$$

By Lemma B–8, $\forall \zeta > 0$,

$$\lim_{r\to\infty} \sup\{|c(r,\, \lambda) - 1| \,:\, \|\lambda\| = 1,\ \sup_{\xi\in\mathcal{S}_p(B)} \xi'\lambda \geq \zeta\} = 0 \ . \tag{B.28}$$

In consequence, $\exists r_\gamma < \infty$ such that $\forall r \geq r_\gamma$ and $\forall \lambda$ with $\|\lambda\| = 1$ and $\sup_{\xi\in\mathcal{S}_p(B)} \xi'\lambda \geq (k_\alpha + 2\gamma)^{1/2}\varepsilon$, we have $c(r,\, \lambda) \geq ((k_\alpha + \gamma)/(k_\alpha + 2\gamma))^{1/2}$. Then, since $\sup_{\xi\in\mathcal{S}_p(B)} \xi'\widehat{\widehat{\beta}}_1 = (\mathcal{LR}_T^q)^{1/2}/\|\widehat{\widehat{\beta}}\|$ whenever $\mathcal{LR}_T^q > 0$ by Lemma B–9, we obtain: $\forall r \geq r_\gamma/\varepsilon$,

$$c(r\|\widehat{\widehat{\beta}}\|, \widehat{\widehat{\beta}}_1) 1(D_{Tr}) \geq ((k_\alpha + \gamma)/(k_\alpha + 2\gamma))^{1/2} 1(D_{Tr}) \ . \tag{B.29}$$

By the substitution of (B.27) and (B.29) into (B.26), we get: $\forall r \geq r_\gamma/\varepsilon$,

$$\begin{aligned} &\lim_{T\to\infty} E_r(1-\varphi_T) \\ &\geq \tfrac{1}{2}\exp\left(-\tfrac{r^2}{2}\right) \lim_{T\to\infty} E_0(1-\varphi_T) \exp(r\|\widehat{\widehat{\beta}}\| c(r\|\widehat{\widehat{\beta}}\|,\, \widehat{\widehat{\beta}}_1) \sup_{\xi\in\mathcal{S}_p(B)} \xi'\widehat{\widehat{\beta}}_1) 1(D_{Tr}) \\ &\geq \tfrac{1}{2}\exp\left(-\tfrac{r^2}{2}\right) \lim_{T\to\infty} E_0(1-\varphi_T) \exp\left(r((k_\alpha + \gamma)/(k_\alpha + 2\gamma))^{1/2} (\mathcal{LR}_T^q)^{1/2}\right) 1(D_{Tr}) \\ &\geq \tfrac{1}{2}\exp\left(-\tfrac{r^2}{2}\right) \lim_{T\to\infty} E_0(1-\varphi_T) \exp(r(k_\alpha + \gamma)^{1/2}) 1(D_{Tr}) \\ &\geq \tfrac{\delta}{4}\exp\left(-\tfrac{r^2}{2} + r(k_\alpha + \gamma)^{1/2}\right) \ , \end{aligned} \tag{B.30}$$

where the third inequality uses Lemma B–6 and the fourth inequality uses (B.25). □

PROOF OF LEMMA B–6: A two-term Taylor expansion of $\ell_T(\theta)$ about $\widehat{\theta}$ yields

$$\ell_T(\theta) = \ell_T(\widehat{\theta}) - \tfrac{T}{2}(\widehat{\theta}-\theta)'\left[-\tfrac{1}{T}D^2\ell_T(\widehat{\theta}(\theta))\right](\widehat{\theta}-\theta)\ , \tag{B.31}$$

where $\widehat{\theta}(\theta)$ lies on the line segment joining θ and $\widehat{\theta}$. (B.31) holds using the fact that $D\ell_T(\widehat{\theta})(\widehat{\theta}-\theta) = 0$ wp $\to 1$ by the first-order conditions for $\widehat{\theta}$. For any consistent estimator $\widehat{\widetilde{\theta}}$ of θ_0, $-\frac{1}{T}D^2\ell_T(\widehat{\theta}(\widehat{\widetilde{\theta}})) = \mathcal{I} + o_p(1)$ by Assumption NL1.

Let $P^{\perp}$ be as in Lemma B–3 and define $P = I_s - P^{\perp}$. By Assumption NL5, the imposition of the null hypothesis does not restrict the parameter space of $P\theta = \theta - AH\theta = (0',\ \delta' - \beta'\mathcal{I}_2\mathcal{I}_3^{-1})'$. In consequence, $P\widetilde{\theta} = P\widehat{\theta}$. In addition, $P^{\perp}\widehat{\theta} = AH\widehat{\theta} = A\widehat{\beta}$ and $P^{\perp}\widetilde{\theta} = 0$. These results yield $\sqrt{T}(\widehat{\theta}-\widetilde{\theta}) = \sqrt{T}(P^{\perp}\widehat{\theta} - P^{\perp}\widetilde{\theta} + P\widehat{\theta} - P\widetilde{\theta}) = \sqrt{T}A\widehat{\beta} = O_p(1)$. The latter result and (B.31) give

$$\begin{aligned}
\ell_T(\widetilde{\theta}) &= \ell_T(\widehat{\theta}) - \tfrac{T}{2}(\widehat{\theta}-\widetilde{\theta})'\mathcal{I}(\widehat{\theta}-\widetilde{\theta}) + o_p(1) \\
&= \ell_T(\widehat{\theta}) - \tfrac{T}{2}(P^{\perp}\widehat{\theta} - P^{\perp}\widetilde{\theta})'\mathcal{I}(P^{\perp}\widehat{\theta} - P^{\perp}\widetilde{\theta}) - \tfrac{T}{2}(P\widehat{\theta} - P\widetilde{\theta})'\mathcal{I}(P\widehat{\theta} - P\widetilde{\theta}) + o_p(1) \qquad \text{(B.32)} \\
&= \ell_T(\widehat{\theta}) - \tfrac{T}{2}\widehat{\beta}'A'\mathcal{I}A\widehat{\beta} + o_p(1)\ .
\end{aligned}$$

To obtain an analogous expression for $\ell_T(\theta^*)$, we need to show that $\sqrt{T}(\widehat{\theta}-\theta^*) = O_p(1)$. The likelihood ratio test statistic for testing $\beta = 0$ versus $\beta \neq 0$ is asymptotically chi-squared under θ_0 under the assumptions. In consequence,

$$O_p(1) = \ell_T(\widetilde{\theta}) - \ell_T(\widehat{\theta}) \leq \ell_T(\theta^*) - \ell_T(\widehat{\theta}) = -\tfrac{T}{2}(\widehat{\theta}\text{–}\theta^*)'[\mathcal{I}+o_p(1)](\widehat{\theta}\text{–}\theta^*) + o_p(1) \leq o_p(1) \tag{B.33}$$

using Assumption NL6. This yields $\sqrt{T}(\widehat{\theta}-\theta^*) = O_p(1)$. The analogue of (B.32) is

$$\ell_T(\theta^*) = \ell_T(\widehat{\theta}) - \tfrac{T}{2}(\widehat{\beta}-\beta^*)'A'\mathcal{I}A(\widehat{\beta}-\beta^*) + o_p(1)\ , \tag{B.34}$$

since $P\theta^* = P\widehat{\theta}$ for the same reason that $P\widetilde{\theta} = P\widehat{\theta}$.

Now, if we can show

$$\zeta_T = T(\widehat{\beta}-\beta^*)'A'\mathcal{I}A(\widehat{\beta}-\beta^*) - T(\widehat{\beta}-\beta_q^*)'A'\mathcal{I}A(\widehat{\beta}-\beta_q^*) \xrightarrow{p} 0 \ \text{ under } \ \theta_0\ , \tag{B.35}$$

then combining (B.32), (B.34), and (B.35) gives

$$\begin{aligned} \mathcal{LR}_T &= T\widehat{\beta}'A'\mathcal{I}A\widehat{\beta} - T(\widehat{\beta}-\beta_q^*)'A'\mathcal{I}A(\widehat{\beta}-\beta_q^*) + o_p(1) \\ &= 2T\beta_q^{*\prime}A'\mathcal{I}A\widehat{\beta} - T\beta_q^{*\prime}A'\mathcal{I}A\beta_q^* + o_p(1) = T\beta_q^{*\prime}A'\mathcal{I}A\beta_q^* + o_p(1) \end{aligned} \tag{B.36}$$

under θ_0, as desired, where the last equality uses the orthogonality condition $\beta_q^{*\prime}A'\mathcal{I}A(\widehat{\beta}-\beta_q^*) = 0$, which holds because β_q^* minimizes $(\widehat{\beta}-\beta)'A'\mathcal{I}A(\widehat{\beta}-\beta)$ over the positively homogeneous set B (e.g., see Perlman (1969, Lemma 4.1(i))).

To show (B.35), note that $\zeta_T \geq 0\ \forall T$ by definition of β_q^*. Let $\theta_q^* = A\beta_q^* + P\widehat{\theta} \in \Theta^*$. Then, $\ell_T(\theta_q^*) - \ell_T(\theta^*) \leq 0\ \forall T$ by definition of θ^*. In addition, $0 \leq T(\widehat{\beta}-\beta_q^*)'A'\mathcal{I}A(\widehat{\beta}-\beta_q^*) \leq T\widehat{\beta}'A'\mathcal{I}A\widehat{\beta} = O_p(1)$, so $\sqrt{T}(\widehat{\theta}-\theta_q^*) = A\sqrt{T}(\widehat{\beta}-\beta_q^*) = O_p(1)$. In consequence, (B.31) and simplifications as in (B.32) yield $\zeta_T = \ell_T(\theta_q^*) - \ell_T(\theta^*) + o_p(1) \leq o_p(1)$. This establishes (B.35). □

Proof of Lemma B–7: Lemma B–2 holds with Q_c replaced by Q^r. In consequence, $LR_T(\theta_0, Q^r)$ equals $\overline{\ell r_c} + o_p(1)$, where $\overline{\ell r_c}$ is as defined in (B.1) but with Q replaced by Q^r. The term in square brackets in the definition of $\overline{\ell r_c}$ is

$$\tfrac{1}{2}\overline{\theta}'\mathcal{I}\overline{\theta} - \tfrac{1}{2}(h-\overline{\theta})'\mathcal{I}(h-\overline{\theta}) = \tfrac{1}{2}(P^{\perp}\overline{\theta})'\mathcal{I}P^{\perp}\overline{\theta} - \tfrac{1}{2}(h-P^{\perp}\overline{\theta})'\mathcal{I}(h-P^{\perp}\overline{\theta})\ , \tag{B.37}$$

using the fact that $h'\mathcal{I}P\overline{\theta} = 0\ \forall h \in V^{\perp}$. Also, $P^{\perp}\overline{\theta} = AH\overline{\theta} = \sqrt{T}A\widehat{\beta} + o_p(1)$ using Lemmas B–1 and B–3. Let $\lambda = (A'\mathcal{I}A)^{1/2}\sqrt{T}\widehat{\beta}$. Let $h = rA(A'\mathcal{I}A)^{-1/2}\xi$, where $\xi \sim U$. Then, $h \sim Q^r$ and

$$\begin{aligned} LR_T(\theta_0, Q^r) &= \int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) \exp\left[\tfrac{T}{2}\widehat{\beta}'A'\mathcal{I}A\widehat{\beta} - \tfrac{1}{2}(h-A\sqrt{T}\widehat{\beta})'\mathcal{I}(h-A\sqrt{T}\widehat{\beta})\right] \\ &\quad \times dQ^r(h)/K + o_p(1) \\ &= \int_{\mathcal{S}_p(B)} \exp\left[\tfrac{1}{2}\lambda'\lambda - \tfrac{1}{2}(r\xi-\lambda)'(r\xi-\lambda)\right] dU(\xi)/K + o_p(1) \\ &= \int_{\mathcal{S}_p(B)} \exp\left[-\ \tfrac{1}{2}r^2 + r\xi'\lambda\right] dU(\xi)/K + o_p(1)\ . \quad \square \end{aligned} \tag{B.38}$$

Proof of Lemma B–8: The supremum over λ in Lemma B–8 is attained for each r. Hence, it suffices to show: For any λ_r with $\|\lambda_r\| = 1$,

$$\lim_{r\to\infty}\left[\tfrac{1}{r}\log\int_{\mathcal{S}_p(B)} \exp(r\xi'\lambda_r) dU(\xi)/K - \sup_{\xi_*\in\mathcal{S}_p(B)} \xi_*'\lambda_r\right] = 0\ . \tag{B.39}$$

By replacing $\xi'\lambda_r$ with $\sup_{\xi_*\in\mathcal{S}_p(B)}\xi_*'\lambda_r$, it is easy to see that the left-hand side of (B.39) is less than or equal to zero. To show the reverse inequality, let $\varepsilon > 0$ be given and choose a subset

of $\mathcal{S}_p(B)$ for each r, say Ξ_r, for which $\xi'\lambda_r \geq \sup_{\xi_* \in \mathcal{S}_p(B)} \xi_*'\lambda_r - \varepsilon \ \forall \xi \in \Xi_r$ and $\int_{\Xi_r} dU(\xi) \geq \delta$ for some $\delta > 0$. Such sets exist by Assumption 6, since the latter is equivalent to: $\forall \varepsilon > 0$, $\inf_{\xi \in \mathcal{S}_p(B)} \int 1(\|\xi - \xi_*\| \leq \varepsilon,\, \xi_* \in \mathcal{S}_p(B)) dU(\xi_*) > 0$. The left-hand side of (B.39) is greater than or equal to

$$\lim_{T \to \infty} \left[\tfrac{1}{r} \log \int_{\Xi_r} \exp(r[\sup_{\xi_* \in \mathcal{S}_p(B)} \xi_*'\lambda_r - \varepsilon]) dU(\xi) - \sup_{\xi_* \in \mathcal{S}_p(B)} \xi_*'\lambda_r \right] = -\varepsilon \ . \tag{B.40}$$

Since $\varepsilon > 0$ is arbitrary, the left-hand side of (B.40) is ≥ 0. □

Proof of Lemma B–9: Let $\overline{S}$ denote the closure of a set S. Any $\beta \in \overline{B}$ can be written as $\beta = c(A'\mathcal{I}A)^{-1/2}\xi$ for some $\xi \in \overline{\mathcal{S}}_p(B)$ and $c \geq 0$. In particular, there exists $c^* \geq 0$ and $\xi_q^* \in \overline{\mathcal{S}}_p(B)$ such that $\beta_q^* = c^*(A'\mathcal{I}A)^{-1/2}\xi_q^*$. Note that $\beta_q^{*\prime} A'\mathcal{I}A\beta_q^* = (c^*)^2$. Now, β_q^* minimizes $(\widehat{\beta} - \beta)'A'\mathcal{I}A(\widehat{\beta} - \beta)$ over $\beta \in \overline{B}$ if and only if $(c^*,\, \xi_q^*)$ minimizes

$$\widehat{\beta}' A'\mathcal{I}A\widehat{\beta} - 2c\xi'(A'\mathcal{I}A)^{1/2}\widehat{\beta} + c^2 \ \text{ over } \ (c, \xi) \in [0,\, \infty) \times \overline{\mathcal{S}}_p(B) \ . \tag{B.41}$$

To solve the latter problem, we first maximize $\xi'(A'\mathcal{I}A)^{1/2}\widehat{\beta}$ over $\xi \in \overline{\mathcal{S}}_p(B)$ and then minimize

$$-2c \sup_{\xi \in \mathcal{S}_p(B)} \xi'(A'\mathcal{I}A)^{1/2}\widehat{\beta} + c^2 \ \text{ over } \ c \geq 0 \ . \tag{B.42}$$

If $\sup_{\xi \in \mathcal{S}_p(B)} \xi'(A'\mathcal{I}A)^{1/2}\widehat{\beta} > 0$, we obtain $c^* = \sup_{\xi \in \mathcal{S}_p(B)} \xi'(A'\mathcal{I}A)^{1/2}\widehat{\beta}$. Combining this with the above expression for $(c^*)^2$ gives $\sup_{\xi \in \mathcal{S}_p(B)} \xi'(A'\mathcal{I}A)^{1/2}\widehat{\beta} = (\beta_q^{*\prime} A'\mathcal{I}A\beta_q^*)^{1/2}$. If $\sup_{\xi \in \mathcal{S}_p(B)} \xi'(A'\mathcal{I}A)^{1/2}\widehat{\beta} \leq 0$, we obtain $c^* = 0$, $\beta_q^* = 0$, and $(\beta_q^{*\prime} A'\mathcal{I}A\beta_q^*)^{1/2} = 0$. □

Proof of Lemma B–10: By Lemmas B–6 and B–9,

$$\mathcal{L}\mathcal{R}_T - (\max\{\sup_{\xi \in \mathcal{S}_p(B)} \xi'(A'\mathcal{I}A)^{1/2}\sqrt{T}\widehat{\beta},\, 0\})^2 \xrightarrow{p} 0 \ . \tag{B.43}$$

Since $(A'\mathcal{I}A)^{1/2}\sqrt{T}\widehat{\beta} \xrightarrow{d} Z \sim N(0,\, I_p)$ by Lemma B–1 and Assumption NL2, the result follows by the continuous mapping theorem. Absolute continuity follows by a result of Lifshits (1982). □

Proof of Lemma B–11: The proof is analogous to that of Lemma B–5 with $\ell r_c(\theta_0)$ replaced by $LR_T(\theta_0,\, Q^r)$ and $\mathcal{L}_c$ replaced by $\mathcal{L}^r = \int_{\mathcal{S}_p(B)} \exp(-r^2/2 + r\xi'Z) dU(\xi)$. We have $LR_T(\theta_0,\, Q^r)$

$\xrightarrow{d} \mathcal{L}^r/K$ by Lemmas B–1 and B–7 and Assumption NL2. It remains to show that $E\mathcal{L}^r/K = 1$. This follows by Fubini's Theorem and the standard formula for the normal moment generating function. □

PROOF OF LEMMA A: Assumption NL1(a) holds by Assumption A(a). NL1(b) holds by A(c). NL1(c) holds with $\mathcal{I}(\theta) = -E\frac{\partial^2}{\partial\theta\partial\theta'}\log g_t(\theta)$ provided a uniform WLLN can be established. The Markov property (A(b)) ensures that $\left\{\frac{\partial^2}{\partial\theta\partial\theta'}\log g_t(\theta) : t > m\right\}$ is part of a doubly infinite stationary and ergodic sequence. Thus, using A(b) and (e), the ergodic theorem implies that $-T^{-1}D^2\ell_T(\theta) \xrightarrow{p} \mathcal{I}(\theta)\ \forall\theta \in \Theta_0$. A generic uniform WLLN (e.g., Assumptions TSE–1D, BD, DM, and P–WLLN and Theorem 4 of Andrews (1992)) strengthens this result to uniform convergence over Θ_0 using A(b), (c), and (e).

Assumption NL1(d) holds, because $\mathcal{I}(\theta)$ is continuous on Θ_0 by the dominated convergence theorem using A(c) and (e) and Θ_0 is compact. NL1(e) holds by A(f).

To verify Assumption NL2, we use a martingale difference triangular array (MDTA) central limit theorem of Hall and Hyde (1980, Cor. 3.1, p. 58). By assumption, $T^{-1/2}D\ell_T(\theta_T) = T^{-1/2}\Sigma_1^T\frac{\partial}{\partial\theta}\log g_t(\theta_T)$. Let E_T denote the expectation operator under θ_T. $\{\left(\frac{\partial}{\partial\theta}\log g_t(\theta_T),\ \mathcal{F}_{t-1}\right) : m < t \leq T,\ T \geq 1\}$ is a MDTA under $\{\theta_T : T \geq 1\}$, because

$$\begin{aligned} &E_T\left(\tfrac{\partial}{\partial\theta}\log g_t(\theta_T)|\mathcal{F}_{t-1}\right) = E_T\left(\tfrac{\partial}{\partial\theta}\log g_t(\theta_T)|S_{t,m},\ X_{t,m}\right) \\ &= \int \tfrac{\partial}{\partial\theta}g_t(\theta_T)d\lambda(s_t) = \tfrac{\partial}{\partial\theta}\int g_t(\theta_T)d\lambda(s_t) = 0, \end{aligned} \tag{B.44}$$

where the third equality holds by the dominated convergence theorem using Assumptions A(c) and (e) and the last equality holds because $\int g_t(\theta)d\lambda(s_t) = 0\ \forall\theta \in \Theta_0$.

Now, we apply Corollary 3.1 under $\{\theta_T : T \geq 1\}$ with $X_{ni} = \frac{\partial}{\partial\theta}\log g_t(\theta_T)/\sqrt{T}$. By the Cramer–Wold device, it suffices to consider the case where θ is a scalar. Hall and Hyde's condition (3.21) holds because $\mathcal{F}_t$ does not depend on T. The other two conditions of Corollary 3.1 are:

$$\forall\varepsilon > 0,\ \Sigma_i E(X_{ni}^2 1(|X_{ni}| > \varepsilon)|\mathcal{F}_{i-1}) \xrightarrow{\text{p}} 0 \ \text{ and} \tag{B.45}$$

$$\Sigma_i E(X_{ni}^2|\mathcal{F}_{i-1}) \xrightarrow{\text{p}} \eta \ \text{ for some constant } \eta > 0. \tag{B.46}$$

We will show that (B.45) and (B.46) hold under θ_0 with $\eta = E\frac{\partial}{\partial\theta}\log g_t(\theta_0)\frac{\partial}{\partial\theta'}\log g_t(\theta_0) = \mathcal{I}$. By contiguity (Lemma B–4), then, (B.45) and (B.46) also hold under $\{\theta_T : T \geq 1\}$, which is required for the application of Corollary 3.1.

Equation (B.45) can be established under θ_0 by establishing L^1-convergence of the left-hand side to 0. By stationarity, the L^1-norm of the left-hand side of (B.45) equals

$$\begin{aligned} & E\left(\tfrac{\partial}{\partial\theta}\log g_t(\theta_T)\right)^2 1\left(\left|\tfrac{\partial}{\partial\theta}\log g_t(\theta_T)\right| > \sqrt{T}\varepsilon\right) \\ \leq\ & E\sup_{\theta\in\Theta_0}\left(\tfrac{\partial}{\partial\theta}\log g_t(\theta)\right)^2 1\left(\left|\tfrac{\partial}{\partial\theta}\log g_t(\theta)\right| > \sqrt{T}\varepsilon\right) \\ \to\ & 0 \end{aligned} \tag{B.47}$$

by the dominated convergence theorem, since $E\sup_{\theta\in\Theta_0}\|\frac{\partial}{\partial\theta}\log g_t(\theta)\|^2 < \infty$ by Assumption A(e).

Equation (B.46) can be established by taking a mean value expansion of its left-hand side about θ_0:

$$\begin{aligned} & \tfrac{1}{T}\Sigma_1^T E\left(\left(\tfrac{\partial}{\partial\theta}\log g_t(\theta_T)\right)^2|\mathcal{F}_{t-1}\right) \\ =\ & \tfrac{1}{T}\Sigma_1^T E\left(\left(\tfrac{\partial}{\partial\theta}\log g_t(\theta_0)\right)^2|\mathcal{F}_{t-1}\right) + \tfrac{2}{T}\Sigma_1^T E\left(\tfrac{\partial}{\partial\theta}\log g_t(\ddot{\theta}_T)\tfrac{\partial^2}{\partial\theta\partial\theta'}\log g_t(\ddot{\theta}_T)(-h/\sqrt{T})|\mathcal{F}_{t-1}\right), \end{aligned} \tag{B.48}$$

where $\ddot{\theta}_T$ lies between θ_T and θ_0. The first term is the average of stationary, ergodic, L^2–random variables under θ_0 and, hence, converges in probability to its expectation $\eta = E\left(\frac{\partial}{\partial\theta}\log g_t(\theta_0)\right)^2$. By Markov's inequality, the probability that the second term exceeds ε is less than or equal to

$$\tfrac{2\|h\|}{\varepsilon\sqrt{T}}E\sup_{\theta\in\Theta_0}\left\|\tfrac{\partial}{\partial\theta}\log g_t(\theta)\right\|\cdot\left\|\tfrac{\partial^2}{\partial\theta\partial\theta'}\log g_t(\theta)\right\|, \tag{B.49}$$

which is finite for all T by Assumption A(e) and, hence, converges to 0. Thus, the second term of (B.48) converges in probability to 0 and the verification of Assumption NL2 is complete.

Sufficient conditions for Assumption NL3 are: (i) Θ is compact, (ii) $\log g_t(\theta)$ is continuous in θ on Θ with probability one under θ_0, (iii) $\sup_{\theta\in\Theta}\left|\frac{1}{T}\Sigma_1^T(\log g_t(\theta) - E\log g_t(\theta))\right| \xrightarrow{p} 0$ under θ_0, and (iv) $E\log g_t(\theta)$ is uniquely maximized over Θ at θ_0 (e.g., see Amemiya (1985, Thm. 4.1.1, pp. 106–107). Parts (i) and (ii) hold by A(a) and (c) respectively. Part (iii) holds by the same argument as for NL1(c) above. To obtain part (iv), note that for $\theta \neq \theta_0$,

$$E\log g_t(\theta) - E\log g_t(\theta_0) = E\log[g_t(\theta)/g_t(\theta_0)] < \log Eg_t(\theta)/g_t(\theta_0) = 0\ , \tag{B.50}$$

where the inequality is an application of Jensen's inequality and is strict by A(d).

Assumptions NL4 and NL6 hold by the same argument as for Assumption NL3 with $\widetilde{\Theta}$ and Θ^*, respectively, in place of Θ. $\square$

TABLE I

λ	$\Phi_2(R_+^2, \lambda, I_2)$
(.707, .707)	.58
(0, 1)	.42
(–.707, .707)	.18
(–1, 0)	.08
(–.707, –.707)	.06

TABLE II. Power Results for Parameter Space $B_1 = R^2_+$

(a) $\rho = 0$

Statistic / $\beta =$	(.88,.88)	(1.28,1.28)	(1.67,1.67)	(2.24,2.24)	Average over $\beta \propto (1,1)$
D–W_0	.34	.57	.76	.94	.65
D–W_1	.34	.55	.75	.93	.64
D–W_∞	.33	.54	.74	.92	.64
$\mathcal{LR}$	.30	.50	.70	.90	.60
$2S$–W	.18	.35	.55	.82	.48
Envelope	.34	.57	.76	.94	.65

Statistic / $\beta =$	(1.37,0)	(1.92,0)	(2.48,0)	(3.26,0)	Average over $\beta \propto (1,0)$
D–W_0	.25	.39	.54	.75	.48
D–W_1	.28	.46	.65	.86	.56
D–W_∞	.29	.47	.67	.87	.58
$\mathcal{LR}$	.30	.50	.70	.90	.60
$2S$–W	.21	.39	.60	.84	.51
Envelope	.39	.61	.80	.95	.69

(b) $\rho = .6$

Statistic / $\beta =$	(1.08,1.08)	(1.57,1.57)	(2.07,2.07)	(2.77,2.77)	Average over $\beta \propto (1,1)$
D–W_0	.33	.54	.75	.93	.64
D–W_1	.31	.51	.72	.91	.61
D–W_∞	.30	.50	.70	.90	.60
$\mathcal{LR}$	.30	.50	.70	.90	.60
$2S$–W	.17	.33	.53	.80	.46
Envelope	.33	.54	.75	.93	.64

Statistic/$\beta =$	(1.38,0)	(1.93,0)	(2.48,0)	(3.28,0)	Average over $\beta \propto (1,0)$
D–W_0	.19	.29	.40	.57	.36
D–W_1	.32	.55	.77	.95	.65
D–W_∞	.35	.58	.80	.96	.67
$\mathcal{LR}$	.30	.50	.70	.90	.60
$2S$–W	.32	.57	.80	.96	.66
Envelope	.53	.78	.93	.99	.81

(c) $\rho = -.6$

Statistic / $\beta =$	(.61,.61)	(.87,.87)	(1.12,1.12)	(1.46,1.46)	Average over $\beta \propto (1,1)$
D–W_0	.39	.62	.80	.95	.69
D–W_1	.38	.61	.80	.95	.69
D–W_∞	.38	.61	.80	.95	.69
$\mathcal{LR}$	.30	.50	.70	.90	.60
$2S$–W	.21	.40	.60	.84	.51
Envelope	.39	.62	.80	.95	.69

Statistic/$\beta =$	(1.10,0)	(1.56,0)	(2.02,0)	(2.67,0)	Average over $\beta \propto (1,0)$
D–W_0	.34	.54	.73	.91	.63
D–W_1	.34	.55	.74	.92	.64
D–W_∞	.35	.55	.75	.93	.65
$\mathcal{LR}$	.30	.50	.70	.90	.60
$2S$–W	.22	.40	.61	.86	.52
Envelope	.39	.62	.81	.95	.69

TABLE III. Average Power Results for Parameter Space $B_2 = R_+ \times R$

	$\rho =$	0	0	0	.6	.6	.6	–.6	–.6	–.6
Statistic	$\beta \propto$	(1,1)	(1,0)	(0,1)	(1,1)	(1,0)	(0,1)	(1,1)	(1,0)	(0,1)
D–W_0		.51	.72	.05	.65	.58	.05	.29	.58	.05
D–W_1		.63	.66	.50	.66	.64	.50	.59	.64	.50
D–W_∞		.63	.76	.55	.64	.64	.55	.61	.64	.55
$\mathcal{LR}$		.60	.60	.60	.60	.60	.60	.60	.60	.60
$2S$–W		.54	.54	.58	.54	.54	.58	.55	.54	.58
Envelope		.72	.72	.75	.72	.72	.75	.72	.72	.75

TABLE IV. Average Power Results for Parameter Space $B_3 = R^6_+$

Statistic / $\beta \propto$	$(1,1,...,1)'$	$(1,1,1,0,0,0)'$	$(1,0,0,...,0)'$
D–W_0	.70	.56	.52
D–W_1	.70	.61	.60
D–W_∞	.68	.62	.61
$\mathcal{LR}$	.60	.60	.60
$2S$–W	.40	.46	.46
Envelope	.70	.77	.76

TABLE V. Average Power Results for Parameter Space $B_4 = R^3_+ \times R^3$

Statistic / $\beta \propto$	(1,1...1)′	1,1,1,0,0,0)′	(1,0,0...0)′	(0,0,0,1,1,1)′	(0,0...0,1)′	(1,0,0...0,1)′
D–W_0	.60	.79	.51	.05	.05	.49
D–W_1	.67	.72	.59	.47	.47	.87
D–W_∞	.66	.69	.60	.53	.53	.89
$\mathcal{LR}$	.60	.60	.60	.60	.60	.90
$2S$–W	.50	.49	.54	.57	.57	.86
Envelope	.80	.79	.83	.86	.86	.89

FOOTNOTES

[1]The author gratefully acknowledges the research support of the National Science Foundation via grant number SES–9121914.

[2]A difference between Wald's results and ours is that Wald's hold for arbitrary weightings of the contours referred to above, whereas ours require the specification of such a weighting. This is a consequence of the nature of the testing problems under consideration.

[3]Assumption NL1 requires that the parametric family of densities is defined in a full neighborhood of θ_0. Given our interest in testing against a restricted alternative parameter space (which generally does not include a neighborhood of θ_0), this assumption can be restrictive. For example, if β is a variance parameter, then $f_T(\theta)$ must be well-defined even when this variance parameter takes on some negative values. To illustrate the implications, consider a test of randomness of the coefficient in a simple random coefficients regression model. The model is

$$Y_t = X_t(\delta_1 + \eta_t) + \varepsilon_t \quad \text{for} \quad t = 1, ..., T \ ,$$

where $\{(X_t, \eta_t, \varepsilon_t) : t \leq T\}$ are iid mutually independent non-degenerate scalar random variables, $\eta_t \sim N(0, \beta)$, $\varepsilon_t \sim N(0, \delta_2)$, (Y_t, X_t) are observed, (η_t, ε_t) are unobserved, and the unknown parameter is $\theta = (\beta, \delta_1, \delta_2)'$. The null hypothesis is $\mathrm{H}_0 : \beta = 0$ and the alternative is $\mathrm{H}_1^* : \beta > 0$. The density $f_T(\theta)$ is given by

$$f_T(\theta) = (2\pi)^{-T/2} \left(\delta_2 + X_t^2\beta\right)^{-T} \exp\left(-\tfrac{1}{2}\textstyle\sum_{t=1}^T (Y_t - X_t\delta_1)^2/(\delta_2 + X_t^2\beta)\right) \displaystyle\prod_{t=1}^T g(X_t) \ ,$$

where $g(x)$ is the density of X_t with respect to some measure. In order for this density to be well-defined (i.e., to have $\delta_2 + X_t^2\beta > 0$) for θ in a neighborhood of $\theta_0 = (0, \delta_1, \delta_2)'$, it is necessary to assume that X_t is bounded. This restriction may be undesirable.

On the other hand, there are many applications in which the parameter θ can take on any value in a neighborhood of θ_0 without causing any problem with the definition of $f_T(\theta)$. For example, this is true of a test of positivity of the variances in an error components model. In addition, Assumption NL1 is a common assumption in the literature on one-sided testing, e.g., see Chernoff (1954), Gourieroux, Holly, and Monfort (1980), Gourieroux and Monfort (1989, Ch. XXI), and Wolak (1989a). Hence, we are not imposing more restrictive conditions than appear elsewhere in the literature.

[4]Since $\Phi(B, \mu, \Sigma) = \Phi(B, \mu/\sqrt{T}, \Sigma/T)$ by positive homogeneity of B, the distribution $\mathcal{L}_c(h)$ (given in (4.6)) equals the distribution of D–W_c when $Y \sim N(\beta, \Omega)$. To show that the asymptotic local distribution of $\mathcal{LR}_T$ given in Theorem 4(a) equals that of $\mathcal{LR}$ when $Y \sim N(\beta, \Omega)$, we rewrite $\mathcal{LR}$ as follows:

$$\mathcal{LR} = \sup_{\beta \in B}(2Y'\Omega^{-1}\beta - \beta'\Omega^{-1}\beta) = \sup_{c \geq 0} \sup_{\xi \in \mathcal{S}_p^*(B)} (2Z'\xi c - c^2) = (\max\{\sup_{\xi \in \mathcal{S}_p^*(B)} Z'\xi, \ 0\})^2,$$

where the first equality holds by simple algebra, the second equality holds by the change of variables $Z = \Omega^{-1/2}Y \sim N(\Omega^{-1/2}\beta, I_p)$, $\xi = \Omega^{-1/2}\beta/\|\Omega^{-1/2}\beta\| \in \mathcal{S}_p^*(B) = \{\xi \in R^p : \|\xi\| = 1$ and $\Omega^{-1/2}\xi \in B\}$, and $c = \|\Omega^{-1/2}\beta\|$, and the third equality holds by solving the maximization problem over $c \geq 0$.

REFERENCES

Amemiya, T. (1985): *Advanced Econometrics.* New York: Harvard University Press.

Andrews, D. W. K. (1992): "Generic Uniform Convergence," *Econometric Theory*, 8, 241–257.

——— (1994): "The Large Sample Correspondence Between Classical Hypothesis Tests and Bayesian Posterior Odds Tests," *Econometrica*, 62, forthcoming.

Andrews, D. W. K. and W. Ploberger (1993): "Admissibility of the Likelihood Ratio Test When a Nuisance Parameter Is Present Only Under the Alternative," Cowles Foundation Discussion Paper No. 1058, Yale University.

——— (1994): "Optimal Tests When a Nuisance Parameter Is Present Only under the Alternative," *Econometrica*, 62, forthcoming.

Barlow, R. E., D. J. Bartholomew, J. M. Bremner, and H. D. Brunk (1972): *Statistical Inference under Order Restrictions.* New York: John Wiley and Sons.

Berger, R. L. (1989): "Uniformly More Powerful Tests for Hypotheses Concerning Linear Inequalities and Normal Means," *Journal of the American Statistical Association*, 84, 192–199.

Berger, R. L. and D. F. Sinclair (1984): "Testing Hypotheses Concerning Unions of Linear Subspaces," *Journal of the American Statistical Association*, 79, 158–163.

Chernoff, H. (1954): "On the Distribution of the Likelihood Ratio," *Annals of Mathematical Statistics*, 54, 573–578.

Dufour, J.–M. (1989): "Nonlinear Hypotheses, Inequality Restrictions, and Non-nested Hypotheses: Exact Simultaneous Tests in Linear Regressions," *Econometrica*, 57, 335–355.

Goldberger, A. S. (1992): "One-sided and Inequality Tests for a Pair of Means," in *Contributions to Consumer Demand and Econometrics*, ed. by R. Bewley and T. Van Hoa. London: MacMillan Academic and Professional Press.

Gourieroux, C. and A. Monfort (1989): *Statistique et Modeles Econometriques*, Vol. 2. Paris: Economica.

Gourieroux, C., A. Holly, and A. Monfort (1980): "Kuhn–Tucker, Likelihood Ratio and Wald Tests for Nonlinear Models with Inequality Constraints on the Parameters," Harvard Institute of Economic Research Discussion Paper No. 770, Harvard University.

——— (1982): "Likelihood Ratio Test, Wald Test, and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters," *Econometrica*, 50, 63–80.

Hajivassiliou, V., D. McFadden, and P. Ruud (1994). "Simulation of Multivariate Normal Orthant Probabilities: Theoretical and Computational Results," *Journal of Econometrics*, forthcoming.

Hall, P. and C. C. Heyde (1980): *Martingale Limit Theory and its Application.* New York: Academic Press.

Hillier, G. H. (1986): "Joint Tests for Zero Restrictions on Non-negative Regression Coefficients," *Biometrika*, 73, 657–669.

King, M. L. and M. D. Smith (1986): "Joint One-sided Tests of Linear Regression Coefficients," *Journal of Econometrics*, 32, 367–383.

King, M. L. and P. X. Wu (1990): "Locally Optimal One-sided Tests for Multiparameter Hypotheses," Department of Econometrics Working Paper No. 2/90, Monash University.

Kodde, D. A. and F. C. Palm (1986): "Wald Criteria for Jointly Testing Equality and Inequality Restrictions," *Econometrica*, 54, 1243–1248.

Lehmann, E. L. (1959): *Testing Statistical Hypotheses.* New York: John Wiley and Sons.

Lifshits, M. A. (1982): "On the Absolute Continuity of Distributions of Functionals of Random Processes," *Theory of Probability and Its Applications*, 27, 600–607.

Perlman, M. D. (1969). "One-sided Problems in Multivariate Analysis," *Annals of Mathematical Statistics*, 40, 549–567. Corrections in *Annals of Mathematical Statistics*, 42, 1777.

Robertson, T. and E. J. Wegman (1978): "Likelihood Ratio Tests for Order Restrictions in Exponential Families," *Annals of Statistics*, 6, 485–505.

Robertson, T., F. T. Wright, and R. L. Dykstra (1988): *Order Restricted Statistical Inference.* New York: Wiley.

Rogers, A. J. (1986): "Modified Lagrange Multiplier Tests for Problems with One-sided Alternatives," *Journal of Econometrics*, 31, 341–361.

Sackrowitz, H. B. and W. E. Strawderman (1974): "On the Admissibility of the MLE for Ordered Binomial Parameters," *Annals of Statistics*, 2, 822–828.

Shapiro, A. (1988): "Towards a Unified Theory of Inequality Constrained Testing in Multivariate Analysis," *International Statistical Review*, 56, 49–62.

Strasser, H. (1985): *Mathematical Theory of Statistics: Statistical Experiments and Asymptotic Decision Theory.* New York: de Gruyter.

Wald, A. (1942): "On the Power Function of the Analysis of Variance Test," *Annals of Mathematical Statistics*, 13, 434–439.

——— (1943): "Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations Is Large," *Transactions of the American Mathematical Society*, 54, 426–482.

Warrack, G. and T. Robertson (1984): "A Likelihood Ratio Test Regarding Two Nested But Oblique Order-restricted Hypotheses," *Journal of the American Statistical Association*, 79, 881–886.

Wolak, F. A. (1987): An Exact Test for Multiple Inequality and Equality Constraints in the Linear Regression Model," *Journal of the American Statistical Association*, 82, 782–793.

——— (1989a): "Local and Global Testing of Linear and Nonlinear Inequality Constraints in Nonlinear Econometric Models," *Econometric Theory*, 5, 1–35.

——— (1989b): "Testing Inequality Constraints in Linear Econometric Models," *Journal of Econometrics*, 41, 205–235.