HYPOTHESIS TESTING WITH ARESTRICTED PARAMETER SPACE

BY

DONALD W.K. ANDREWS

COWLES FOUNDATION PAPER NO. 960

COWLES FOUNDATION FOR RESEARCH IN ECONOMICSA T YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281
1998
Hypothesis testing with a restricted parameter space

Donald W.K. Andrews

Cowles Foundation for Research in Economics, Yale University, New Haven, Connecticut 06520-8281, USA

Received 1 May 1995; received in revised form 1 December 1996; accepted 11 March 1997

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. 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 $D-W_\infty$ directed test is found to perform best in an overall sense for multivariate one-sided alternatives with the likelihood ratio test being a close second. The $D-W_\infty$ and likelihood ratio tests are found to perform best for mixed one- and two-sided alternatives. © 1998 Elsevier Science S.A.

Keywords: Directed tests; Linear regression; Nonlinear models; One-sided alternative; Optimal test; Weighted average power

JEL classification: C12

1. Introduction

This paper considers tests of hypotheses in parametric models in which the parameter of interest, $\beta$, is restricted under the alternative hypothesis. More specifically, we consider tests of $H_0$: $\beta = 0$ versus $H_1^*$: $\beta \in B$, where $\beta \in R^p$ is a sub-vector 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 $H_1^*$: $\beta \not\leq 0$). Econometric applications of hypotheses of these types have been noted in the literature, e.g.,

The author gratefully acknowledges the research support of the National Science Foundation via grant numbers SES-9121914 and SBR-9410675.
see Gourieroux et al. (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 no error components against the alternative of positivity of the variances of error components in panel data models with individual and/or time error components, (3) tests of no skewness or excess kurtosis against the alternative of skewness and/or excess kurtosis, (4) tests of no serial correlation or ARCH effects against the alternative of positive first-order serial correlation and/or ARCH effects, and (5) tests of the equality of several parameters against the alternative of ordered alternatives (of the form \( H_1^\beta: \mu_1 < \mu_2 < \mu_3 \), 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 \( \beta \notin \mathcal{B} \). 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 \( \mathcal{B} \) contains a neighborhood of zero).\(^1\) 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 \( \hat{\beta} \) relative to the restricted parameter space \( \mathcal{B} \). The directed LR and LM statistics are defined analogously. None require the computation of the ML estimator for the restricted parameter space \( \mathcal{B} \), which sometimes is difficult to compute and is required by the standard LR test. 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 (r.v.). Critical values and \( p \)-values can be obtained by simulation. An

---

\(^{1}\)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.
interactive GAUSS program, written by 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.

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 $\infty$, 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 results for the multivariate normal location model with known variance are supplemented by results for a normal linear regression model with unknown variance. The latter show that the effect of estimation of the variance on the power of the tests is small for the sample sizes considered, viz., 25, 50, and 100. In consequence, the multivariate normal location model results provide accurate comparisons of the power of the tests for normal regression models with unknown variances even for small sample sizes.

The multivariate normal location model simulation results can be summarized as follows: We find the $D-W_{\infty}$ test to be the best overall test for multivariate one-sided alternatives. The $D-W_1$ and LR tests are a fairly close second. For mixed one- and two-sided alternatives, the $D-W_{\infty}$ and LR tests are best. In all cases, the $D-W_{\infty}$ test has comparative advantage in the middle of the parameter space, while the LR test has comparative advantage 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 $D-W_{3/2}$ tests have similar power and are almost as good as $D-W_{\infty}$. As expected, 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 $D-W_{\infty}$ 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 0.3 and 0.9.
The fact that the power of the directed Wald tests does not vary much 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 fact that the average power of the LR test (across different directions) does not differ greatly from that of the \( DH_c \) tests for \( c \neq 0 \) also has useful consequences. It implies that the LR test is close to being optimal for a wide range of weight functions. This supplements the results of Andrews (1996), which shows that the LR test is strictly optimal for certain weight functions that place all weight on distant alternatives from the null.

The optimality properties of the directed tests and the classical (two-sided) tests might 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 et al. (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 et al. (1982), Hillier (1986), Wolak (1987, 1989b), and Dufoü (1989). For nonlinear models, references include Gourieroux et al. (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 admissibility or asymptotic admissibility of the LR test (depending on the model) is established by Andrews (1996). Andrews (1996) also shows that the LR test directs power at alternatives that are arbitrarily distant from the null hypothesis.
The papers above all consider the L.R 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 (1996). See Robertson et al. (1988) for further references. King and Wu (1997) 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 $H_0$: $\beta \geq 0$ versus $H_1$: $\beta \not\geq 0$. For results concerning problems of this sort, see Perlman (1969), Robertson and Wegman (1978), Farebrother (1986), Kodde and Palm (1986), Wolak (1987, 1989a,b), and Dufour (1989). 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 asymptotic null distribution and asymptotic optimality properties of the directed statistics for dynamic nonlinear models under a set of high-level assumptions. Section 4 provides primitive sufficient conditions for the high-level assumptions of Section 3. Section 5 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. An appendix contains proofs of the results stated in the text.

2. Regression

2.1. 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

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

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

The null and alternative hypotheses of interest are

$$H_0: \beta = 0 \quad \text{and} \quad H_1: \beta \in \mathbb{B}/\{0\},$$

(1)
where $B\setminus\{0\}$ denotes the set $B$ minus the zero vector. The regression parameter vector is $\theta = (\beta', \delta')'$. The information matrix for $\theta$ is

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

(2)

The parameter space $\Theta^* (\subset R^d)$ of $\theta$ is required to satisfy:

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

**Assumption 3.** $d = \mathcal{I}_3^{-1}\mathcal{I}_2' b \in A \forall b \in B$, $\forall d \in \Delta$.

The two main cases where Assumption 3 is satisfied are when (i) the nuisance parameter $\delta$ is unrestricted (i.e., $A = R^q$) or (ii) 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) $\{\beta \in R^p: \beta_j \geq 0 \forall j = 1, \ldots, p\}$,
(ii) $\{\beta \in R^p: \beta_j \in R \forall j \leq J; \beta_j \geq 0 \forall j = J + 1, \ldots, p\}$,
(iii) $\{\beta \in R^p: \beta_j > 0 \text{ for some } j \leq p\}$,
(iv) $\{\beta \in R^p: \beta_j - \beta_{j-1} \geq 0 \forall j = 1, \ldots, p\}$ (which corresponds to the nonnegative and ordered alternative $H^*: 0 \leq \beta_1 \leq \beta_2 \leq \cdots \leq \beta_p$).

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 $R^p$, $A$ is the negative orthant of $R^q$, and $\mathcal{I}_3^{-1}\mathcal{I}_2'$ contains nonnegative elements, then Assumption 3 holds.

Let $\theta_0$ denote some parameter vector in the null hypothesis: $\theta_0 = (0', \delta_0')' \in R^d$ for some $\delta_0 \in R^q$. Any parameter vector $\theta \in R^d$ can be written as the sum of the null parameter vector $\theta_0$ and some perturbation vector $h \in R^p$. That is, $\theta = \theta_0 + h$.

Given $\theta_0$, we specify a weight function $Q_0(\cdot)$ 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 $R^p$ defined by the null hypothesis.

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

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

(3)

The null hypothesis can be expressed as $H_0: \theta \in V \cap \Theta^*$. Define the inner product $\langle h, f \rangle = h'Ff$, for $h, f \in R^q$. Denote the orthogonal complement of $V$ under $\langle \cdot, \cdot \rangle$ by $V^\perp$. Since $V$ is a $q$ dimensional subspace of $R^p$, $V^\perp$ is a $p$ dimensional subspace of $R^p$. Let $\{a_1, \ldots, a_p\}$ be some basis of $V^\perp$ and define $A = [a_1; a_2; \cdots; a_p] \in R^{q \times p}$. (The optimal test statistics developed below are invariant with respect to the choice of basis of $V^\perp$.) For example, one can take $A = [I_p; \mathcal{I}_2\mathcal{I}_3^{-1}]'$. 


Note that

\[ A' \mathcal{J} A = \mathcal{J}_1 - \mathcal{J}_2 \mathcal{J}_3^{-1} \mathcal{J}_2' = X' M_G X / \sigma^2, \]

where \( M_G = I_T - G (G' G)^{-1} G' \).

Next, let

\[ \Sigma = A (A' \mathcal{J} A)^{-1} A' \]

\[ = \begin{bmatrix}
(\mathcal{J}_1 - \mathcal{J}_2 \mathcal{J}_3^{-1} \mathcal{J}_2')^{-1} & -(\mathcal{J}_1 - \mathcal{J}_2 \mathcal{J}_3^{-1} \mathcal{J}_2')^{-1} \mathcal{J}_2 \mathcal{J}_3^{-1} \\
-\mathcal{J}_2^{-1} \mathcal{J}_2' (\mathcal{J}_1 - \mathcal{J}_2 \mathcal{J}_3^{-1} \mathcal{J}_2')^{-1} & \mathcal{J}_2^{-1} \mathcal{J}_2' (\mathcal{J}_1 - \mathcal{J}_2 \mathcal{J}_3^{-1} \mathcal{J}_2')^{-1} \mathcal{J}_2 \mathcal{J}_3^{-1}
\end{bmatrix}. \]

(5)

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

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

Note that the support of \( Q_c \) is \( \nu^\perp \).

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 \( H_0: \theta = \theta_0 \) versus \( H_1: \theta = \theta_0 + h \). 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

\[ \int 1(\theta_0 + h \in \Theta^*) P(\varphi \text{ rejects } H_0 \mid \theta_0 + h) dQ_c(h) / K, \]

where \( \varphi \) is some level \( \alpha \) 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 nonzero weight to parameter values \( \theta (= \theta_0 + h) \) in \( \Theta^* \). (The constant \( K \) merely ensures that the truncated weight function integrates to one.) An optimal test of level \( \alpha \) maximizes the above weighted average power criterion over all tests of level \( \alpha \).

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

\[
\int 1(\theta_0 + h \in \Theta^*) P(\varphi \text{ rejects } H_0 \mid \theta_0 + h) dQ_c(h)/K
\]

\[
= \int 1(\theta_0 + h \in \Theta^*) \int_{\mathcal{R}^p} \varphi(y) f(y, \theta_0 + h) dy dQ_c(h)/K
\]

\[
= \int_{\mathcal{R}^p} \varphi(y) \left[ \int 1(\theta_0 + h \in \Theta^*) f(y, \theta_0 + h) dQ_c(h)/K \right] dy \quad (7)
\]

by Fubini's Theorem.

Eq. (7) shows that the weighted average power of \( \varphi \) equals the power of \( \varphi \)

against the single alternative density specified by

\[
g(y, \theta_0) = \int 1(\theta_0 + h \in \Theta^*) f(y, \theta_0 + h) dQ_c(h)/K. \quad (8)
\]

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 \( LR(\theta_0) \):

\[
LR(\theta_0) = g(Y, \theta_0)/f(Y, \theta_0)
\]

\[
= \left[ \int 1(\theta_0 + h \in \Theta^*) f(Y, \theta_0 + h) dQ_c(h)/K \right] / f(Y, \theta_0). \quad (9)
\]

We show below that this statistic does not depend on \( \theta_0 \). 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 require that one place a weight function over the \( q \) dimensional nuisance parameter \( \delta \). It would also 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 \( B \subset \mathcal{R}^p, \mu \in \mathcal{R}^p, \) and \( \Omega \) a positive semi-definite \( p \times p \) matrix, let

\[
\Phi_p(B, \mu, \Omega) = P(Z \in B) \quad \text{where } Z \sim N(\mu, \Omega). \quad (10)
\]

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

$$D-W_c = (1 + c)^{-n/2} \exp\left(\frac{c}{2} \frac{1}{1 + c} W\right) \Phi\left(B, \frac{c}{1 + c} \hat{\beta}, \frac{c}{1 + c} (X'M_GX)^{-1} \sigma^2\right),$$

where $W = \hat{\beta}'(X'M_GX/\sigma^2)\hat{\beta}$ and $\hat{\beta} = (X'M_GX)^{-1}X'M_GY$.

Note that $\hat{\beta}$ is the least squares estimator of $\beta$ from the (unrestricted) regression of $Y_t$ on $X_t$ and $G_t$. $W$ is just the standard Wald test statistic for testing $H_0: \beta = 0$ against $H_1: \beta \neq 0$ (for the case where $\sigma^2$ is known). The calculation of $D-W_c$ requires the computation of a multivariate normal integral, see Section 3.5 below regarding computation. 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 $\Phi$ 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 $\hat{\beta}$. Consider the case where $p = 2$, $B$ is the positive orthant, and $(X'M_GX)^{-1}\sigma^2$ is the identity matrix. Table 1 provides the values of $\Phi_2(R^2, \lambda, I_2)$ for several values of $\lambda$ on the unit circle. For $\lambda$ in the middle of $R^2$, i.e., $\lambda = (0.707, 0.707)$, the factor $\Phi_2(R^2, \lambda, I_2)$ is almost ten times as large as when $\lambda$ is on the opposite side of the circle. Thus, the factor $\Phi$ has a substantial impact on the value of the directed Wald statistic.

Note that the directed test statistics depend on the weight function $Q_2(\cdot, \cdot)$ 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:

$$D-W_\infty = \lim_{c \to \infty} \frac{1}{2} \log[(1 + c)^{-n/2} D-W_c]\right]$$

$$= W + 2 \log[\Phi(B, \hat{\beta}, (X'M_GX)^{-1} \sigma^2)],$$

where log denotes the natural logarithm here and below.

Table 1

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>$\Phi_2(R^2, \lambda, I_2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0.707, 0.707)</td>
<td>0.58</td>
</tr>
<tr>
<td>(0, 1)</td>
<td>0.42</td>
</tr>
<tr>
<td>(−0.707, 0.707)</td>
<td>0.18</td>
</tr>
<tr>
<td>(−1, 0)</td>
<td>0.08</td>
</tr>
<tr>
<td>(−0.707, −0.707)</td>
<td>0.06</td>
</tr>
</tbody>
</table>
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

$$D-W_0 = \lim_{c \to 0} J_c(D-W_c) = d' \hat{\beta} / (d' (X' M_G X/\sigma^2)^{-1} d)^{1/2},$$

(13)

where

$$d = \frac{\partial}{\partial \mu} \Phi(B, 0, (X' M_G X)^{-1}),$$

$$\frac{\partial}{\partial \mu} \Phi(B, 0, \Omega) = \frac{\partial}{\partial \mu} \Phi(B, \mu, \Omega) \bigg|_{\mu=0} = \Omega^{-1} E Z1(Z \leq B) \quad \text{for} \quad Z \sim N(0, \Omega),$$

and

$$J_c(x) = \frac{2}{\sqrt{c}} \left[ \log((1+c)^{p/2} x) - \log \Phi(B, 0, (X' M_G X)^{-1}) \right]$$

$$\times \frac{\Phi(B, 0, (X' M_G X)^{-1})}{(d' (X' G X/\sigma^2)^{-1} d)^{1/2}}.$$ 

Except for special cases (such as $B = R^p$), we have $d \neq 0$ and $D-W_0$ is well-defined. Since $D-W_0 \sim N(0, 1)$ 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 $D-W_0$.) Thus, to carry out a test based on $D-W_0$, the only potential complication is in computing

$$d = \frac{\partial}{\partial \mu} \Phi(B, 0, (X' M_G X)^{-1}).$$

For $p = 2$ or 3, $d$ can be computed by numerical quadrature. For $p \geq 4$, it can be computed by simulation methods. See Hajivassiliou et al. (1996) 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 $(X' M_G X)^{-1} \sigma^2$ has all nondiagonal 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 favour 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 5 below.
The null distributions of the directed Wald statistics equal those of the following r.v.'s: \( 0 < c < \infty \),

\[
\mathcal{L}_c = (1 + c)^{-p/2} \exp \left( \frac{1}{2} \frac{c}{1 + c} Z'Z \right) \\
\times \Phi \left( B, \frac{c}{1 + c} (X'MGX)^{-1/2} \sigma Z, \frac{c}{1 + c} (X'MGX)^{-1} \sigma^2 \right), \\
\mathcal{L}_\infty = Z'Z + \log[\Phi(B, (X'MGX)^{-1/2} \sigma Z, (X'MGX)^{-1} \sigma^2)], \\
\mathcal{L}_0 = Z_1, \tag{14}
\]

where

\[ Z \sim N(0, I_p) \text{ and } Z = (Z_1, \ldots, Z_p)'. \]

Let \( \xi_c \) denote a test of level \( \alpha \) 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) \( D-W_c \sim LR(\theta_0) \times K \) for all \( 0 < c < \infty \),

(b) \( D-W_c \sim \mathcal{L}_c \) under the null hypothesis \( H_0 \) for all \( 0 < c < \infty \) and

(c) for any level \( \alpha \) test \( \varphi \), the directed Wald test \( \xi_c \) satisfies

\[
\int \{ \varphi(\xi_c) \} \left[ \int \varphi(f(\theta_0 + h)) d\mu \right] dQ_c(h) \\
\leq \int \{ \varphi(\xi_c) \} \left[ \int \varphi(f(\theta_0 + h)) d\mu \right] dQ_c(h)
\]

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 \( \mu \) is Lebesgue measure on \( R^p \).

The proofs of Theorem 1 and other results stated below are given in the Appendix.

**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 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: \( Y^1 = X^1\theta^1 + U^1 \), \( U^1 \sim N(0, \sigma^2 \Omega) \), \( H_0^1 \); \( R^1\theta^1 = r \), and \( H_1^1 \); \( R^1\theta \in B^1/\{r\} \), where \( Y^1 \) is an observed r.v. in \( R^T \); \( X^1 \) is an observed nonstochastic \( T \times s \) regressor matrix; \( \theta^1 \) is an unknown parameter vector in \( R^p \); \( U^1 \) is an unobserved error vector in \( R^T \); \( \Omega \) is a known \( T \times T \) positive-definite covariance matrix; \( \sigma^2 \) is a known positive constant; \( R^1 \) is a known \( p \times s \) matrix.
of constants; \( r \) is a known \( p \)-vector of constants; and \( B^1 \) is a known subset of \( \mathbb{R}^p \). An example of the above testing problem is a test of the equality of a sub-vector of parameters \( H^1_0: \theta^1_1 = \theta^1_2 = \cdots = \theta^1_{p+1} \) against the ordered alternative \( H^1_1: \theta^1_1 \leq \theta^1_2 \leq \cdots \leq \theta^1_{p+1} \) 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 (1), but it can be transformed to the form of (1). (See King and Smith (1986, Section 2) for details of the transformation.)

(3) If \( \sigma^2 \) is unknown, then an exactly similar test can be constructed, provided \( B \) is positively homogeneous, by replacing \( \sigma^2 \) in the definition of \( D-W \) by the unrestricted estimator

\[
\hat{\sigma}^2 = \frac{1}{T - p} \sum_{i=1}^{T} (Y_i - X_i \hat{\beta} - G_i \delta)^2.
\]

The resultant test statistic is a directed \( F \) statistic. The null distribution of the directed \( F \) statistic is given by that of

\[
Q_c = (1 + c)^{-p/2} \exp \left( \frac{1}{2} \frac{c}{1 + c} Z' Z (T - p) / \chi^2 \right) \times \Phi \left( B, \frac{c}{1 + c} (X'M_cX)^{-1}Z, \frac{c}{1 + c} (X'M_cX)^{-1} \chi^2 / (T - p) \right),
\]

where \( Z \sim \mathcal{N}(0, I_p) \), \( \chi^2 \) has a chi-squared distribution with \( p \) degrees of freedom, and \( Z \) and \( \chi^2 \) are independent. Theorem 1 does not establish finite sample optimality properties of the directed \( F \) statistic, but Theorem 2 below provides asymptotic optimality properties for it. Furthermore, the simulation results of Section 5 show that the estimation of \( \sigma^2 \) has a small effect on power. Thus, the directed \( F \) test must be (at least) very close to being optimal in the weighted average power sense considered here for the case of unknown \( \sigma^2 \).

(4) If \( B \) is convex and positively homogeneous, then the directed Wald test is strictly unbiased when \( c \in (0, \infty] \) (i.e., the test has power greater than its significance level \( \alpha \in (0, 1) \) for all \( \beta \in B \) with \( \beta \neq 0 \)) and is unbiased for \( c = 0 \) (i.e., the test has power greater than or equal to its significance level \( \alpha \in (0, 1) \) for all \( \beta \in B \)). Examples of convex and positively homogeneous sets include convex (one-sided) cones, which include orthants, half-spaces, etc.

The proof of this result for \( c \in (0, \infty] \) follows from the fact that (i) \( W \) is stochastically larger when \( \beta \neq 0 \) than when \( \beta = 0 \) for any \( \beta \in \mathbb{R}^p \) with \( \beta \neq 0 \), because \( W \) has noncentral chi-square distribution in the former case and central chi-square distribution in the latter, and (ii) for all \( \beta \in B \) with \( \beta \neq 0 \),

\[
\Phi \left( B, \frac{c}{1 + c} \tilde{\beta}, \frac{c}{1 + c} (X'M_cX)^{-1} \sigma^2 \right) = \Phi \left( B, \frac{c}{1 + c} (\xi + \beta), \frac{c}{1 + c} (X'M_cX)^{-1} \sigma^2 \right)
\]
\[ \Phi \left( B, \frac{c}{1 + c} \xi, \frac{c}{1 + c} (X'MGX)^{-1} \sigma^2 \right) \]
\[ > \Phi \left( B, \frac{c}{1 + c} \xi, \frac{c}{1 + c} (X'MGX)^{-1} \sigma^2 \right), \]  

(15)

where \( \xi = (X'MGX)^{-1}X'MGU \) and the inequality holds because, for \( B \subset B + \beta = \{ \lambda + \beta : \lambda \in B \} \) by convexity and positive homogeneity of \( B \). For \( c = 0 \), the directed Wald test is unbiased by the previous result and the continuity of its power function as \( c \to 0 \). The \( c = 0 \) directed Wald test is not necessarily strictly unbiased for convex and positively homogeneous \( B \), because \( d' \beta \) could equal zero for some \( \beta \in B \) with \( \beta \neq 0 \). For example, this occurs for the half-space \( B = \{ (\beta_1, \beta_2)' \in R^2 : \beta_1 \geq 0 \} \) when \( \beta = (0, -1)' \), since \( d = (1, 0)' \) in this case.

For parameter spaces \( B \) that are not both convex and positively homogeneous, the unbiasedness of directed Wald tests is an open question. The term involving \( W \) always works to increase the rejection rate when \( \beta \) is changed from 0 to \( \beta \), but it is possible that such a change can reduce the term
\[ \Phi \left( B, \frac{c}{1 + c} (\xi + \beta), \frac{c}{1 + c} (X'MGX)^{-1} \sigma^2 \right). \]

In fact, one can show that for some parameter spaces \( B \) and for some \( c \) sufficiently close to zero, the directed Wald test is biased. For example, consider the parameter space \( B = \{ (\beta_1, \beta_2)' \in R^2 : \beta_1 \geq 0 \text{ or } \beta_2 \geq 0 \} \). The \( c = 0 \) directed Wald test has \( d = (1, 1)' \) and is biased for all \( \beta \) with \( \beta_1 + \beta_2 < 0 \). By continuity of the power function as \( c \to 0 \), the directed Wald test also is biased for sufficiently small \( c > 0 \). Whether the directed Wald test can be biased for some parameter spaces for large \( c \), such as \( c = \infty \), is an open question, but it seems unlikely.

3. 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.

3.1. Notation

Let \( Y_T \) denote the data matrix when the sample size is \( T \) for \( T = 1, 2, \ldots \). Consider a parametric family \( \{ f_T(y_T, \theta) : \theta \in \Theta^* \cup \Theta \} \) of densities of \( Y_T \) with respect to some \( \sigma \)-finite measure \( \mu_T \), where \( \Theta^* \subset R^r \) and \( \Theta \subset R^r \) are two parameter spaces defined below. The likelihood function of the data is given by \( f_T(\theta) = f_T(y_T, \theta) \). In many cases, the likelihood function \( f_T(\theta) \) can be written as a product of two terms, one that depends on \( \theta \) and another that does not.
Often the latter term is the product over \( t = 1, \ldots, T \) 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 \( \theta \) is taken to be of the form \( \theta = (\beta', \delta')' \), where \( \beta \in \mathbb{R}^p \), \( \delta \in \mathbb{R}^q \), and \( s = p + q \). The null hypothesis of interest is \( H_0: \beta = 0 \), as in (1). We let \( \theta_0 \) denote a parameter vector in the null hypothesis. That is, \( \theta_0 \) is of the form \( \theta_0 = (0', \delta')' \) for some \( \delta \in \mathbb{R}^q \).

For standard large-sample two-sided testing problems, the parameter space is taken to be a subset \( \Theta \) of \( \mathbb{R}^q \) that contains a neighborhood of \( \theta_0 \) for all \( \theta_0 \) in the null hypothesis. We refer to such a parameter space \( \Theta \) as the unrestricted alternative (UA) parameter space. The alternative hypothesis corresponding to the UA parameter space is \( H_1: \beta \neq 0 \). 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 \( \theta \). Let \( D^2\ell_T(\theta) \) denote the \( s \times s \) matrix of second partial derivatives of \( \ell_T(\theta) \) with respect to \( \theta \). 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 \( 1/\sqrt{T} \) and \( 1/T \) respectively. Let \( I(\theta) = \text{plim}_{T \to \infty} (1/T)D^2\ell_T(\theta) \) (\( I(\theta) \) is the limiting information matrix for \( \theta \)).

Let \( \hat{\theta} \) be the UA ML estimator of \( \theta \). By definition, \( \hat{\theta} \) maximizes the log likelihood function over \( \Theta \) at least with probability that goes to one as \( T \to \infty \) (i.e., \( wp \to 1 \)). That is, \( \hat{\theta} \) satisfies

\[
\ell_T(\hat{\theta}) = \max_{\theta \in \Theta} \ell_T(\theta) \quad wp \to 1 \text{ under } \theta_0.
\]  

Let \( \hat{\theta} \) be the restricted ML estimator of \( \theta \) (restricted by the null hypothesis \( H_0 \)). By definition, \( \theta_0 \) satisfies

\[
\hat{\theta} \in \hat{\Theta} = \{ \theta \in \Theta: \theta = (0', \delta')' \text{ for some } \delta \in \mathbb{R}^q \}.
\]

Wald, LM, and LR test statistics for testing \( H_0 \) against \( H_1 \) are given by

\[
W_T = T(H(\hat{\theta})' [H^2]^{-1}(\hat{\theta})H')^{-1}H \hat{\beta} = T\hat{\beta}[\hat{\beta}_1 - \hat{\beta}_2][\hat{\beta}_1 - \hat{\beta}_2]' \hat{\beta}.
\]

\[
LM_T = (D\ell_T(\hat{\theta})/\sqrt{T})' [I(\hat{\theta})^{-1}(\hat{\theta})D\ell_T(\hat{\theta})/\sqrt{T}]
\]
\[ L_{RT} = -2 (\ell_T(\hat{\theta}) - \ell_T(\bar{\theta})) \quad \text{where } H = [I_p : 0] \subset \mathbb{R}^{p \times s}, \]

\[ J_T(\bar{\theta}) = -\frac{1}{T} D^2 \ell_T(\bar{\theta}) = \begin{bmatrix} J_{1T}(\bar{\theta}), J_{2T}(\bar{\theta}) \\ J_{2T}(\bar{\theta})^T, J_{3T}(\bar{\theta}) \end{bmatrix}, \quad \tilde{J}_j = J_{jt}(\bar{\theta}), \]

\[ \tilde{J}_j = J_{jt}(\bar{\theta}) \quad \text{for } j = 1, 2, 3. \]

Alternatively, one can define \( J_T(\bar{\theta}) \) to be of outer product rather than Hessian form or equal \( E_{\bar{\theta}} J_T(\bar{\theta}) \).

### 3.2. Optimal tests for restricted alternatives

The alternative hypothesis that is of primary interest in this paper is \( H^* : \beta \in B / \{0\} \), as in (1). The corresponding restricted alternative (RA) parameter space is \( \Theta^* = B \times \Delta \), where \( B \subset \mathbb{R}^p \) and \( \Delta \subset \mathbb{R}^d \). For standard asymptotic results, the parameter space \( B \) is required to be positively homogeneous (i.e., \( \beta \in B \Rightarrow t\beta \in B \forall t > 0 \)) 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 2 below.

To derive asymptotically optimal tests of \( H_0 \) versus \( H_1 \), we consider local alternatives to \( H_0 \) of the form \( f_1(\theta_0 + h/\sqrt{T}) \) for some \( h \in \mathbb{R}^d \). 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 (16), rather than than as in Section 2. We consider an asymptotic weighted average power criterion, which is the limit superior as \( T \to \infty \) of (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

\[
D_{\text{WT}} = (1 + c)^{-p/2} \exp \left( \frac{1 + c}{2} W_T \right) \times \Phi \left( \frac{c}{1 + c}, \frac{c}{1 + c} \right) \left( \tilde{J}_1 - \tilde{J}_2 \tilde{J}_3^{-1} \tilde{J}_2' \right)^{-1/2} \left( \frac{1}{T} \right).
\]  

(19)

for \( 0 < c < \infty \).

\[
D_{\text{WP}} = W_T + 2 \log \left[ \Phi \left( \tilde{J}_1 - \tilde{J}_2 \tilde{J}_3^{-1} \tilde{J}_2', \tilde{J}_1 - \tilde{J}_2 \tilde{J}_3^{-1} \tilde{J}_2', \frac{1}{T} \right) \right],
\]

\[
D_{\text{WT}} = \sqrt{T} \tilde{d}' \tilde{\beta} / \left( \tilde{d}' \left( \tilde{J}_1 - \tilde{J}_2 \tilde{J}_3^{-1} \tilde{J}_2' \right) \tilde{d} \right)^{1/2},
\]

where \( \tilde{d} = \frac{\partial}{\partial \mu} \Phi(B, 0, (\tilde{J}_1 - \tilde{J}_2 \tilde{J}_3^{-1} \tilde{J}_2')^{-1}/T). \)

One rejects \( H_0 \) if \( D_{\text{WT}} \) exceeds a critical value \( k_2 \) that is determined using the asymptotic null distribution of \( D_{\text{WT}} \).
The directed LM statistic, $D-\text{LM}_{c,T}$, is defined analogously to $D-\text{W}_{c,T}$ with $W_T$ replaced by $LM_T$, $\hat{\beta}$ replaced by
\[
[\hat{\mathcal{F}}_1 - \hat{\mathcal{F}}_2, \hat{\mathcal{F}}_3]^{-1} \frac{\partial}{\partial \hat{\beta}} f_T(\hat{\theta})/T,
\]
and
\[
(\hat{\mathcal{F}}_1 - \hat{\mathcal{F}}_2, \hat{\mathcal{F}}_3)^{-1}/T \text{ replaced by } (\hat{\mathcal{F}}_1 - \hat{\mathcal{F}}_2, \hat{\mathcal{F}}_3)^{-1}/T.
\]
Note that the $D-\text{LM}_{c,T}$ statistic is constructed using only the restricted ML estimator $\hat{\theta}$. The directed LR statistic, $D-\text{LR}_{c,T}$, is defined analogously to $D-\text{W}_{c,T}$ with $W_T$ replaced by $LR_T$. (One also could replace $\hat{\beta}$ and $(\hat{\mathcal{F}}_1 - \hat{\mathcal{F}}_2, \hat{\mathcal{F}}_3)^{-1}/T$ by the expressions above involving $\hat{\theta}$ without affecting the large sample properties of $D-\text{LR}_{c,T}$.) The test statistics $D-\text{W}_{c,T}$, $D-\text{LM}_{c,T}$, and $D-\text{LR}_{c,T}$ 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 $D-\text{LM}_{c,T}$ and $D-\text{LR}_{c,T}$, respectively, exceed $k_2$, where $k_2$ 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.

3.3. Assumptions

In this section, we state high-level assumptions under which the asymptotic results hold. Section 4 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 $\theta_0$ denote the true value of $\theta$ under the null $H_0$. We say that a statement holds 'under $\theta_0$' (i.e., under the null hypothesis) if it holds when the true density of $y_T$ is $f_T(\theta_0)$ for $T = 1, 2, \ldots$. We introduce a sequence of local alternatives to the null parameter vector $\theta_0$:

\[
\theta_T = \theta_0 + h_1/\sqrt{T} \quad \text{for } T \geq 1,
\]

where $h \in \mathbb{R}^l$. 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 $\Theta$ is some set that contains a neighborhood of $\theta_0$. The parametric model is assumed to be sufficiently regular that the following assumptions hold.

Assumption NLL. (a) $\theta_0$ is an interior point of $\Theta$.

(b) $f_T(\theta)$ is twice continuously partially differentiable in $\theta$ for all $\theta \in \Theta_0$ with probability one under $\theta_0$, where $\Theta_0 (\subset \Theta)$ is some neighborhood of $\theta_0$.

(c) $-T^{-1}D^2 f_T(\theta) \overset{P}{\to} \mathcal{J}(\theta)$ uniformly over $\theta \in \Theta_0$ under $\theta_0$ for some non-random $s \times s$ matrix function $\mathcal{J}(\theta)$.

(d) $\mathcal{J}(\theta)$ is uniformly continuous on $\Theta_0$. 
(e) $\mathcal{J} = \mathcal{J}(\theta_0)$ is positive definite.

**Assumption NL2.** $T^{-1/2}D_T(\theta_T)^{1/2} \overset{D}{\to} Z^* \sim N(0, \mathcal{J})$ under $\{\theta_T: T \geq 1\}$.

**Assumption NL3.** $\theta_T \overset{p}{\to} \theta_0$ under $\theta_0$.

**Assumption NL4.** $\theta_T \overset{p}{\to} \theta_0$ under $\theta_0$.

**Assumption NL5.** For each $d \in \Lambda$ and $b \in B$, $\exists \varepsilon_0 > 0$ such that $\forall \varepsilon \leq \varepsilon_0$ we have $d - \mathcal{J}_d^{-1}\mathcal{J}_b^2 \varepsilon \in \Lambda$, where

$$
\mathcal{J} = \begin{pmatrix}
\mathcal{J}_1 \\
\mathcal{J}_2 \\
\mathcal{J}_3
\end{pmatrix}
$$

and $\mathcal{J}_i \subset \mathbb{R}^{p \times p}$.

We comment now on Assumptions NL1–NL5. Assumptions NL1(a), (b), (d), and (e) are fairly common ML regularity conditions. Differentiability in $\theta$ 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 $-T^{-1}D_T(\theta)$ 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.

---

2 Assumption NL1 requires that the parametric family of densities is defined in a full neighbourhood of $\theta_0$. Given our interest in testing against a restricted alternative parameter space (which generally does not include a neighbourhood of $\theta_0$), this assumption can be restrictive. For example, if $\beta$ is a variance parameter, then $f_1(\tilde{\beta})$ 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

$$
Y_i = X_i(\delta_1 + \eta_i) + \varepsilon_i \quad \text{for } i = 1, \ldots, T,
$$

where $\{(X_i, \eta_i, \varepsilon_i): i \leq T\}$ are iid mutually independent non-degenerate scalar random variables, $\eta_i \sim N(0, \beta)$, $\varepsilon_i \sim N(0, \delta_2)$, $(Y_i, X_i)$ are observed, $(\eta_i, \varepsilon_i)$ are unobserved, and the unknown parameter is $\theta = (\beta, \delta_1, \delta_2)'$. The null hypothesis is $H_0: \beta = 0$ and the alternative is $H_1: \beta > 0$. The density $f_T(\theta)$ is given by

$$
f_T(\theta) = (2\pi)^{-T/2}(\delta_2 + X_i^2\beta)^{-T} \exp \left( -\frac{1}{2} \sum_{i=1}^{T} (Y_i - X_i\delta_1)^2 / (\delta_2 + X_i^2\beta) \right) \prod_{i=1}^{T} g(X_i),
$$

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

On the other hand, there are many applications in which the parameter $\theta$ can take on any value in a neighbourhood of $\theta_0$ 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 et al. (1980), Gourieroux and Monfort (1989, Ch. XXI), and Wold (1989a). Hence, we are not imposing more restrictive conditions than appear elsewhere in the literature.
(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 \( T^{-1/2} Df_T(\theta_T) \) 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 \( \mathcal{A} \) is open or (2) the information matrix is block diagonal between \( \beta \) and \( \delta \) (i.e., \( \mathcal{F}_2 = 0 \)). In addition, Assumption NL5 is satisfied in a variety of special cases.

### 3.4. Asymptotic Results

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

\[
\mathcal{L}_c(h) = \begin{cases} 
(1 + c)^{-p/2} \exp \left[ \frac{1}{2} \frac{-c}{1+c} Z'Z \right] & \times \Phi(B, c, 1+c) \left( \mathcal{F}_1 - \mathcal{F}_2 \mathcal{F}_5^{-1} \mathcal{F}_2' \right)^{-1/2} Z, \\
\left( c, \mathcal{F}_1 - \mathcal{F}_2 \mathcal{F}_5^{-1} \mathcal{F}_2' \right)^{-1} & \quad \text{for } 0 < c < \infty \end{cases}
\]

\[
Z'Z + 2 \log(\Phi(B, c)) + \left( \mathcal{F}_1 - \mathcal{F}_2 \mathcal{F}_5^{-1} \mathcal{F}_2' \right)^{-1/2} Z, \\
\left( \mathcal{F}_1 - \mathcal{F}_2 \mathcal{F}_5^{-1} \mathcal{F}_2' \right)^{-1} \right] & \quad \text{for } c = \infty, \\
d' \left( \mathcal{F}_1 - \mathcal{F}_2 \mathcal{F}_5^{-1} \mathcal{F}_2' \right)^{-1/2} Z & \quad \text{for } c = 0,
\]

where \( Z \sim N((\mathcal{F}_1 - \mathcal{F}_2 \mathcal{F}_5^{-1} \mathcal{F}_2')^{1/2} h_1, I_p) \), \( h = (h_1', h_2')' \) for \( h_1 \epsilon R^p \), and

\[
\mathcal{F} = \begin{pmatrix} \mathcal{F}_1 \\
\mathcal{F}_2 \\
\mathcal{F}_3 \end{pmatrix}
\]

for \( \mathcal{F}_1 \in R^{p \times p} \). 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 \( \varphi_T \) denote a test of \( H_0 \). The test \( \varphi_T \) is of asymptotic significance level \( \alpha \) if \( \int \varphi_T f_T(\theta_0) d\mu_T \rightarrow \alpha \) as \( T \rightarrow \infty \) for all \( \theta_0 \) 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 \( \varphi_T \) when \( \theta_0 \) is true. Similarly, the power of \( \varphi_T \) against the local alternative \( f_T(\theta_0 + g/\sqrt{T}) \) is denoted \( \int \varphi_T f_T(\theta_0 + g/\sqrt{T}) d\mu_T \).

Let \( \{ k_T : T \geq 1 \} \) 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 \( \alpha \). Let \( \xi_{cT} \) denote a directed test, i.e., \( \xi_{cT} = 1(D-W_{cT} > k_{Tc}) \), \( \xi_{cT} = 1(D-LM_{cT} > k_{Tc}) \), or \( \xi_{cT} = 1(D-LR_{cT} > k_{Tc}) \) for \( 0 \leq c \leq \infty \).

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

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

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

\[
\lim_{T \to \infty} \int f(\theta_0 + h/\sqrt{T} \in \Theta^* \Bigg| f_T \big(\theta_0 + h/\sqrt{T}\big) \mathrm{d}\mu_T \Bigg| \mathrm{d}Q_c(h) \leq \lim_{T \to \infty} \int f(\theta_0 + h/\sqrt{T} \in \Theta^* \Bigg| f_T \big(\theta_0 + h/\sqrt{T}\big) \mathrm{d}\mu_T \Bigg| \mathrm{d}Q_c(h)
\]

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

**Comments.** (1) One can extend the scope of the results by taking a model parameterized by \( \gamma \in \Gamma \subset \mathbb{R}^p \) with restrictions \( \bar{H}_0: h(\gamma) = 0 \), say, and transforming it into a model parameterized by \( \theta = (\beta', \delta')' \in \Theta^* \subset \mathbb{R}^q \) 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 \( B = \{ \beta \in \mathbb{R}^3: \beta_j \geq 0 \; \forall \; j \leq 3 \} \) yields a test against the ordered alternative \( \bar{H}_1^* \): \( \gamma_1 \leq \gamma_2 \leq \gamma_3 \leq \gamma_4 \) (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 is 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 \( \Theta^* \) to \( h/\sqrt{T^*/T} \) and \( \Theta^*_T = B/\sqrt{T^*/T} \times \Delta \), respectively, in the beginning of Section 3.2 and in (20) and \( B \) to \( B/\sqrt{T^*/T} \) in (19). 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 \( \beta \), viz., \( B/\sqrt{T^*/T} \times \Delta \), 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 \( B \subset \mathbb{R}^p \) and \( \Delta \subset \mathbb{R}^q \), where \( B \) has positive Lebesgue measure’. In (21), the set \( B \) remains as is – it is not changed to \( B/\sqrt{T^*/T} \). With the above changes, Theorem 2 holds with \( h/\sqrt{T} \) and \( \Theta^* \) changed to \( h/\sqrt{T^*/T} \) and \( \Theta^*_T \) respectively.
(3) Comment 4 following Theorem 1, concerning unbiasedness, is applicable to the directed tests considered here with unbiasedness replaced by local asymptotic unbiasedness.

3.5. Computational issues

We now discuss some computational issues. There are two aspects to computation. One is programming time and the second is execution time. Although there have been significant technological advances on both fronts in recent years, by far the greatest improvements have been in the reduction of execution time. These improvements seem likely to continue in the future. Thus, programming time will become relatively more important as time passes.

To program the directed test statistics (for $c \neq 0$), one needs to compute the standard two-sided Wald, LM, or LR test statistic, the (completely) unrestricted ML estimator of $\beta$, and an estimator of the asymptotic covariance matrix of the unrestricted ML estimator of $\beta$. (For the directed LM statistic one substitutes the normalized score function evaluated at the restricted (by the null) estimator of $\theta$.) Programs and methods for carrying out these calculations are readily available.

An interactive GAUSS program that takes the above statistics as input and provides the desired directed test statistic and its asymptotic $p$-value and critical values is available from the author. Thus, no further programming time is required beyond that involved in calculating the two-sided test statistics if one uses the author's program.

If one wishes to program the directed test statistic and its asymptotic $p$-value and critical values oneself, then one needs a program that calculates a $p$-dimensional multivariate normal probability. For $p$ equal to two or three, numerical integration can be used. Gauss and Matlab have built-in programs for doing so and FORTRAN programs for this are available in the standard FORTRAN libraries. For $p$ greater than or equal to four, one can use simulation methods for computing the multivariate normal probabilities. Methods for doing so are described and compared by Hajivassiliou et al. (1996). The GIII algorithm they discuss seems to be the best for present purposes. GAUSS and FORTRAN programs for this (and other) algorithm(s) are available from Hajivassiliou by ftp. (His GAUSS program is used in the author's interactive program for the calculation of the directed test statistics.) Once one has programmed the computation of the directed test statistic, it is fairly easy to program the calculation of asymptotic critical values. One needs to draw a large number of iid realizations of the random variable in (14). This requires a program that draws multivariate normal random variables. These are built into Gauss and Matlab and are readily available for FORTRAN. By vectorizing the calculation of the random variable in (14) or by using a DO LOOP, one can generate a specified number of draws of the random variable in (14). The sample quantiles from these sorted draws give the asymptotic critical values of the directed test. The asymptotic $p$-value is given by the fraction of
draws that exceed the value of the directed test statistic that was computed using the actual data.

Next, we consider the execution time for the directed test statistics, their $p$-values, and their critical values. The execution time depends upon $p$, the number of quantile simulation repetitions, and the shape of $B$. If simulation is used to compute the multivariate normal probabilities, then it also depends on the number of repetitions used for this, which we call ‘GHK repetitions’ (because the GHK algorithm is used in the author’s interactive GAUSS program). The execution time is increasing in each of these quantities.

We now provide some illustrative execution times using the author’s GAUSS program and a Pentium 90 PC. For $p=2$ and $B=(R^+)^2$, it takes 5 seconds to compute the directed test statistic, its asymptotic $p$-value, and its asymptotic critical values with 5000 quantile repetitions. These results are very quick because numerical integration is used to compute the multivariate normal probabilities (with a reputed accuracy of six or more digits). The execution time is linear in the number of quantile repetitions.

For $p=5$, $B=(R^+)^5$, 800 GHK repetitions, and 1000 quantile repetitions, it takes 4 min and 28 s to compute the directed test statistic, $p$-value, and critical values. The above numbers of repetitions are enough to get rough and ready $p$-values and critical values. For precise $p$-value and critical values, one could use 1600 GHK repetitions and 5000 quantile repetitions. In this case, it takes 47 min to compute. (The execution time is linear in the product of the number of GHK and quantile repetitions.) When the parameter space is taken to be $B=(R^+)^2 \times R^3$, the computations are about 25% faster. Next, for $p=10$, $B=(R^+)^{10}$, 800 GHK repetitions, and 1000 quantile repetitions, it takes 10 min and 41 s to compute.

In sum, the execution time for a directed test statistic, $p$-value, and critical values ranges from a few seconds for $p=2$, to a few minutes for $p=5$ or 10 with a relatively low number of repetitions, to an hour or two with a high number of repetitions. For the quickest PC’s available or a work station, the execution times would be much quicker.

We conclude this section by discussing some of the relative merits of the directed tests and the LR test from a computational perspective. First, we define the LR test. The standard LR test statistic equals minus twice the likelihood ratio:

\[
L^R = -2(\ell(\hat{\theta}) - \ell(\theta^*)),
\]

where \(\ell(\theta) = \log f(Y, \theta), \quad \ell(\theta^*) = \sup_{\theta \in \Theta^*} \ell(\theta)\)

and \(\ell(\theta^*) = \sup_{\theta \in \Theta^*} \ell(\theta)\). \hspace{1cm} (22)

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.

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.

A computational advantage of the directed tests is that they do not require computation of the ML estimator for the restricted alternative (RA) hypothesis \( H_1^r: \beta \in \mathcal{B} \setminus \{0\} \). The estimator employed by the directed statistics is just the unrestricted LS estimator regardless of the specification of \( \mathcal{B} \). On the other hand, for common specifications of \( \mathcal{B} \), the LR statistic for linear models requires that one solve a quadratic programming problem that depends on \( \mathcal{B} \). For nonlinear models, avoiding the computation of the ML estimator for the RA hypothesis can be particularly advantageous. Furthermore, the directed LM statistic only requires calculation of the ML estimator under the null and not under the RA hypothesis or the unrestricted hypothesis \( H_1^u: \beta \in \mathbb{R}^p \). 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 the asymptotically equivalent Kuhn–Tucker multiplier test require computation of the ML estimator under the RA hypothesis. On the other hand, the Kodde and Palm (1986) test, which also is asymptotically equivalent to the LR test, does not require computation of the ML estimator under the RA hypothesis.

4. Nonlinear models: primitive sufficient conditions

In this section, we provide primitive sufficient conditions for Assumptions NL1–NL4 of Section 3 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 nonstationary nontrending 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

\[
\{g_t(0): \theta \in \Theta^* \cup \Theta\} = \{g_t(S_t, S_{t-1}, \ldots, X_t, \ldots, \lambda; \theta): \theta \in \Theta^* \cup \Theta\}
\]

(23)

denote a parametric family of conditional densities (with respect to some measure \( \lambda \)) of \( S_t \) given \( S_{t-1}, \ldots, X_t, \ldots \), evaluated at the random variables
\[ S_1, \ldots, S_t, X_1, \ldots, X_t, \text{ where } \Theta^* \subset R^p \text{ and } \Theta \subset R^p. \text{ Let} \]
\[ h_t = h_t(X_t | S_1, \ldots, S_{t-1}; X_1, \ldots, X_{t-1}) \]  
(24)

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

We consider the case where \( \{(S_t, X_t): t \geq 1\} \) is part of a doubly infinitely strictly stationary ergodic sequence \( \{(S_t, X_t): t = \ldots, 0, 1, \ldots\} \) and \( \{S_t: t = \ldots, 0, 1, \ldots\} \) is \( m \)th order Markov for some integer \( m \geq 0 \). In this case, the function \( \mathcal{J}(\theta) \) equals
\[ -E \frac{\partial^2}{\partial \theta \partial \theta'} \log g_i(\theta). \]

By definition, \( \{S_t: t = \ldots, 0, 1, \ldots\} \) is \( m \)th order Markov if the conditional distribution of \( S_t \) given \( \mathcal{F}_{t-1} = \sigma(S_{t-2}, \ldots, S_{t-1}, X_{t-1}, X_t) \) equals the conditional distribution of \( S_t \) given \( S_{t-m} = (S_{t-m}, \ldots, S_{t-1}) \) and \( X_{t-m} = (X_{t-m}, \ldots, X_t) \) for all \( t \). The Markov assumption yields the simplification that the summands \( \log g_i(\theta) \) 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:

**Assumption A.** (a) \( \Theta \) is compact and \( \theta_0 \) lies in the interior of \( \Theta \).
(b) \( \{(S_t, X_t): t = \ldots, 0, 1, \ldots\} \) is strictly stationary and ergodic and \( \{S_t: t = \ldots, 0, 1, \ldots\} \) is \( m \)th order Markov under \( \theta \) for each \( \theta \in \Theta \).
(c) \( g_i(\theta) \) is continuous in \( \theta \) on \( \Theta \) and twice continuously partially differentiable in \( \theta \) on \( \Theta_0 \) with probability one under \( \theta_0 \), where \( \Theta_0 \) is some compact set that contains a neighborhood of \( \theta_0 \).
(d) \( g_i(\theta) \neq g_i(\theta_0) \) with positive probability under \( \theta_0 \) \( \forall \theta \in \Theta \) with \( \theta \neq \theta_0 \).
(e) \( E \sup_{\theta \in \Theta_0} |\log g_i(\theta)| < \infty, E \sup_{\theta \in \Theta_0} \|\partial / \partial \theta \log g_i(\theta)\| < \infty, E \|\partial^2 / \partial \theta \partial \theta' \log g_i(\theta)\| < \infty, \]
and \( E \sup_{\theta \in \Theta_0} \|\partial^3 / \partial \theta^2 \partial \theta' \log g_i(\theta)\| < \infty. \)
(f) \( \mathcal{J} = -E \frac{\partial^2}{\partial \theta \partial \theta'} \log g_i(\theta_0) \) is positive definite.

(The expectations in parts (c) and (f) are taken under \( \theta_0 \).
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 \( \theta \) for each fixed \( \theta \) in \( \Theta \), but does not place such restrictions on sequences of local alternatives.

**Lemma 1.** Assumption A implies Assumptions NL1–NL4.
5. Monte Carlo power comparisons

In this section, we compare the power of several tests of hypotheses of the form (1) by Monte Carlo simulation. Sections 5.1 and 5.2 describe results for the multivariate normal model with known covariance matrix. These results provide asymptotic local power comparisons. In addition, they provide comparisons for the normal linear regression model with known variance. Section 5.3 discusses comparisons for the normal linear regression model with unknown variance for sample sizes \( T = 25, 50, \) and \( 100. \) The results are described only briefly because they differ very little from the results of Sections 5.1 and 5.2.

5.1. Experimental design

The model we consider is a \( p \) variate normal location model with unknown mean \( \beta \) and known covariance matrix \( \Omega. \)

\[
Y \sim \mathcal{N}(\beta, \Omega). \tag{25}
\]

The data consist of a single realization of \( Y \in \mathbb{R}^p. \) The null and alternative hypotheses of interest are as in (1). Results for model (25) 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 (25). Second, the asymptotic local power of tests in a wide variety of nonlinear models equals their exact power in model (25).

To illustrate the first point, consider the linear regression model: \( y = X\beta + \epsilon + U, \) \( U \sim \mathcal{N}(0, \sigma^2 I_p) \), where \( \sigma^2 \) is known. Premultiplication of the regression model by \( (X'\Sigma X)^{-1}X'M \), yields a model of the form (25) with \( Y = (X'\Sigma X)^{-1}X'M \) and \( \Omega = \sigma^2(X'\Sigma X)^{-1}. \)

To establish the second point, one sets \( \Omega = (\mathcal{J}_1 - \mathcal{J}_2 \mathcal{J}_3^{-1} \mathcal{J}_2')^{-1}/T \) and \( \beta = h_1/\sqrt{T}. \) Then, the exact power of the directed tests (and the LR test) for the multivariate normal location model equal their asymptotic local power given in Theorem 2 for nonlinear models.\(^3\)

We consider four different choices for the parameter space \( B: \)

\[
B_1 = \{ \mu \in \mathbb{R}^2: \mu_1 \geq 0, \mu_2 \geq 0 \} \quad \text{where} \quad \mu = (\mu_1, \mu_2)',
\]

\[
B_2 = \{ \mu \in \mathbb{R}^2: \mu_1 \geq 0, \mu_2 \in \mathbb{R} \} \quad \text{where} \quad \mu = (\mu_1, \mu_2)', \tag{26}
\]

\[
B_3 = \{ \mu \in \mathbb{R}^6: \mu_j \geq 0, \forall j \leq 6 \} \quad \text{where} \quad \mu = (\mu_1, \ldots, \mu_6)',
\]

\(^3\) Since \( \Phi(B, \mu, \Sigma) = \Phi(B, \mu/\sqrt{T}, \Sigma/T) \) by positive homogeneity of \( B, \) the distribution \( \mathcal{J}_1(b) \) (given in (3.6)) equals the distribution of \( D - k \) when \( Y \sim \mathcal{N}(\beta, \Omega). \)
\[ B_4 = \{ \mu \in \mathbb{R}^6: \mu_j \geq 0, \ \forall j \leq 3, \ \text{where} \ \mu = (\mu_1, \ldots, \mu_6)' \}. \]

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:

\[ \Omega_j = \begin{pmatrix} 1 & \rho_j \\ \rho_j & 1 \end{pmatrix} \quad \text{for} \ j = 1, 2, 3 \ \text{where} \ \rho_1 = 0, \ \rho_2 = 0.6 \ \text{and} \ \rho_3 = -0.6. \]

(27)

Note that the correlation \( \rho_j \) in (27) 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 \( \rho_j \) equal to zero, or to \( \rho_j \) positive, than to \( \rho_j \) negative.

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

\[ \Omega_4 = I_6. \]

(28)

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 (26) are invariant under multiplication by positive definite diagonal matrices. In consequence, there is no loss in generality in (27) and (28) by taking the diagonal elements of \( \Omega_j \) to be equal to unity – if \( \Omega \) has non-unit diagonal elements, premultiplication of (25) by \( \text{Diag}^{1/2}(\Omega) \) 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 (25). The Wald, LM, and LR versions of these statistics are numerically identical for model (25). In consequence, it suffices to consider the directed Wald statistics. By definition,

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

(29)

Note that for \( B_1, B_2, B_3 \), and \( B_4, d \) is proportional to \((1, 1)', (1, 0)', (1, 1, 1, 1, 1)', \) 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 $\alpha$ and true parameter value $\beta$, its power is given by $1 - \Phi_+(\Phi^{-1}_c(1 - \alpha) - d'd\beta/(d'd\sigma^2)^{1/2})$, where $\Phi_+(\cdot)$ and $\Phi^{-1}_c(\cdot)$ 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 (25):

$$L^R = Y'\Omega^{-1}Y - \inf_{\beta \in \mathcal{B}} (Y - \beta)'\Omega^{-1}(Y - \beta).$$  (30)

Algebraic manipulations yield the following expressions for $L^R$ for the parameter spaces $B_1 - B_4$:

$$L^R = Y'\Omega_1^{-1}Y1(Y_1 > 0, Y_2 > 0) - Y_2^21(Y_1 < 0, Y_2 > \rho_j Y_1)$$

$$- Y_2^21(Y_2 < 0, Y_1 > \rho_j Y_1) \quad \text{when } B = B_1 \text{ and } \Omega = \Omega_1,$$

$$L^R = Y'\Omega_2^{-1}Y - Y_1^21(Y_1 < 0) \quad \text{when } B = B_2 \text{ and } \Omega = \Omega_2,$$

$$L^R = \sum_{i=1}^6 Y_i^21(Y_i > 0) \quad \text{when } B = B_3 \text{ and } \Omega = \Omega_3,$$

$$L^R = \sum_{i=1}^3 Y_i^21(Y_i > 0) + \sum_{i=4}^6 Y_i^2 \quad \text{when } B = B_4 \text{ and } \Omega = \Omega_4.$$

(31)

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 $H_0$: $\beta = 0$ versus $H_1$: $\beta \neq 0$. By definition:

$$2S-W = Y'\Omega^{-1}Y.$$  (32)

Results for the $2S-W$ test are included to quantify the magnitude of the power gains that occur when the information that $\beta \in \mathcal{B}$ 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 $\alpha$ and true parameter $\beta$, 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 $\lambda$ 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 $\beta_1$, the Neyman–Pearson Lemma implies that
the envelope power is given by the power of the likelihood ratio test of $H_0$: $\beta = 0$ versus $H_1$: $\beta = \beta_1$. This test statistic, denoted $\text{ENV}(\beta_1)$, is defined by

$$\text{ENV}(\beta_1) = \beta_1' \Omega^{-1} Y' (\beta_1' \Omega^{-1} \beta_1)^{1/2}.$$  \hspace{1cm} (33)

Clearly, $\text{ENV}(\beta_1)$ depends only on the direction of $\beta_1$ 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 $D-W_r$, 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 $\alpha$ and true parameter value $\beta_1$, 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 $B_1$, $\beta$ is taken proportional to $(1,0)'$ and $(1,1)';$ for $B_2$, $\beta$ is taken proportional to $(0,1)', (1,1)',$ and $(1,0)';$ for $B_3$, $\beta$ is taken proportional to $(1,1,\ldots,1)', (1,1,0,0,0)',$ and $(1,0,\ldots,0)';$ and for $B_4$, $\beta$ is taken proportional to $(1,1,\ldots,1)', (1,1,1,0,0,0)', (1,0,0,\ldots,0)', (0,0,0,1,1,1)',$ $(0,0,\ldots,0,1)',$ and $(1,0,0,\ldots,0,1)'.$ For each direction of $\beta$, four distances from the origin are considered. These distances are chosen so that the LR test has powers $0.3$, $0.5$, $0.7$, and $0.9$. These choices ensure the chosen distances are reasonable and provide for easy comparisons between the $D-W_r$ 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 $\beta$. For $B_2$, power is invariant under changes in sign of the second element of $\beta$. For $B_4$, power is invariant under permutations of the first three elements of $\beta$, permutations of the last three elements of $\beta$, and changes in sign of the last three elements of $\beta$.

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

5.2. Simulation results

Table 2 gives the power of the $D-W_r$, LR, and $2S-W$ tests and the envelope power function for the parameter space $B_1 = R^2$ for all eight values of $\beta$ and all three values of $\rho$. 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 $D-W_0$, $D-W_1$, $D-W_\infty$, 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 more with the direction of departure from the null and the value of the correlation coefficient than with the distance from the null. Table 2 provides six different direction-$\rho$ combinations. The average power results for these six cases show that the $D$-$W_{\infty}$ test is best overall by a very small margin. It is best in two of the six cases, is within 0.02 of the best in four of the six cases, and is within 0.05 of the best in all cases.

**Table 2**

Power results for parameter space $R_1^2 = \mathbb{R}^2_+$

(a) $\rho = 0$

<table>
<thead>
<tr>
<th>Statistic/(\beta)</th>
<th>(0.88, 0.88)</th>
<th>(1.28, 1.28)</th>
<th>(1.57, 1.67)</th>
<th>(2.24, 2.24)</th>
<th>Average over $\beta \propto (1, 1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D$-$W_0$</td>
<td>0.34</td>
<td>0.57</td>
<td>0.76</td>
<td>0.94</td>
<td>0.65</td>
</tr>
<tr>
<td>$D$-$W_1$</td>
<td>0.34</td>
<td>0.55</td>
<td>0.75</td>
<td>0.93</td>
<td>0.64</td>
</tr>
<tr>
<td>$D$-$W_{\infty}$</td>
<td>0.33</td>
<td>0.54</td>
<td>0.74</td>
<td>0.92</td>
<td>0.64</td>
</tr>
<tr>
<td>$\rho$(\beta)</td>
<td>0.30</td>
<td>0.50</td>
<td>0.70</td>
<td>0.90</td>
<td>0.60</td>
</tr>
<tr>
<td>2S-$W$</td>
<td>0.18</td>
<td>0.35</td>
<td>0.55</td>
<td>0.82</td>
<td>0.48</td>
</tr>
<tr>
<td>Envelope</td>
<td>0.34</td>
<td>0.57</td>
<td>0.76</td>
<td>0.94</td>
<td>0.65</td>
</tr>
</tbody>
</table>

(b) $\rho = 0.6$

<table>
<thead>
<tr>
<th>Statistic/(\beta)</th>
<th>(1.37, 0)</th>
<th>(1.92, 0)</th>
<th>(2.48, 0)</th>
<th>(3.26, 0)</th>
<th>Average over $\beta \propto (1, 0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D$-$W_0$</td>
<td>0.25</td>
<td>0.39</td>
<td>0.54</td>
<td>0.75</td>
<td>0.48</td>
</tr>
<tr>
<td>$D$-$W_1$</td>
<td>0.28</td>
<td>0.46</td>
<td>0.65</td>
<td>0.86</td>
<td>0.56</td>
</tr>
<tr>
<td>$D$-$W_{\infty}$</td>
<td>0.29</td>
<td>0.47</td>
<td>0.67</td>
<td>0.87</td>
<td>0.58</td>
</tr>
<tr>
<td>$\rho$(\beta)</td>
<td>0.30</td>
<td>0.50</td>
<td>0.70</td>
<td>0.90</td>
<td>0.60</td>
</tr>
<tr>
<td>2S-$W$</td>
<td>0.21</td>
<td>0.39</td>
<td>0.60</td>
<td>0.84</td>
<td>0.51</td>
</tr>
<tr>
<td>Envelope</td>
<td>0.39</td>
<td>0.61</td>
<td>0.80</td>
<td>0.95</td>
<td>0.69</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Statistic/(\beta)</th>
<th>(1.08, 1.08)</th>
<th>(1.57, 1.57)</th>
<th>(2.07, 2.07)</th>
<th>(2.77, 2.77)</th>
<th>Average over $\beta \propto (1, 1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D$-$W_0$</td>
<td>0.33</td>
<td>0.54</td>
<td>0.75</td>
<td>0.93</td>
<td>0.64</td>
</tr>
<tr>
<td>$D$-$W_1$</td>
<td>0.31</td>
<td>0.51</td>
<td>0.72</td>
<td>0.91</td>
<td>0.61</td>
</tr>
<tr>
<td>$D$-$W_{\infty}$</td>
<td>0.30</td>
<td>0.50</td>
<td>0.70</td>
<td>0.90</td>
<td>0.60</td>
</tr>
<tr>
<td>$\rho$(\beta)</td>
<td>0.27</td>
<td>0.45</td>
<td>0.65</td>
<td>0.87</td>
<td>0.56</td>
</tr>
<tr>
<td>2S-$W$</td>
<td>0.17</td>
<td>0.33</td>
<td>0.53</td>
<td>0.80</td>
<td>0.46</td>
</tr>
<tr>
<td>Envelope</td>
<td>0.33</td>
<td>0.54</td>
<td>0.75</td>
<td>0.93</td>
<td>0.64</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Statistic/(\beta)</th>
<th>(1.38, 0)</th>
<th>(1.93, 0)</th>
<th>(2.48, 0)</th>
<th>(3.28, 0)</th>
<th>Average over $\beta \propto (1, 0)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D$-$W_0$</td>
<td>0.19</td>
<td>0.29</td>
<td>0.40</td>
<td>0.57</td>
<td>0.36</td>
</tr>
<tr>
<td>$D$-$W_1$</td>
<td>0.32</td>
<td>0.55</td>
<td>0.77</td>
<td>0.95</td>
<td>0.65</td>
</tr>
<tr>
<td>$D$-$W_{\infty}$</td>
<td>0.25</td>
<td>0.58</td>
<td>0.80</td>
<td>0.96</td>
<td>0.67</td>
</tr>
<tr>
<td>$\rho$(\beta)</td>
<td>0.39</td>
<td>0.64</td>
<td>0.85</td>
<td>0.98</td>
<td>0.72</td>
</tr>
<tr>
<td>2S-$W$</td>
<td>0.32</td>
<td>0.57</td>
<td>0.80</td>
<td>0.96</td>
<td>0.66</td>
</tr>
<tr>
<td>Envelope</td>
<td>0.53</td>
<td>0.78</td>
<td>0.93</td>
<td>0.99</td>
<td>0.81</td>
</tr>
</tbody>
</table>
Table 2 (Contd.)

\[\begin{array}{lcccc}
\text{Statistic/} & \beta = -0.6 & \beta = 0.61, 0.61 & 0.87, 0.87 & 1.12, 1.12 & 1.46, 1.46 & \text{Average over } \beta \propto (1,1) \\
\hline
\text{Stat} & \text{LR} & \text{LR} & \text{LR} & \text{LR} & \text{LR} & \text{LR} \\
(0.61,0.61) & 0.39 & 0.62 & 0.80 & 0.95 & 0.69 \\
(0.87,0.87) & 0.38 & 0.61 & 0.80 & 0.95 & 0.69 \\
(1.12,1.12) & 0.38 & 0.61 & 0.80 & 0.95 & 0.69 \\
(1.46,1.46) & 0.37 & 0.59 & 0.78 & 0.94 & 0.67 \\
\text{Stat} & \text{LR} & \text{LR} & \text{LR} & \text{LR} & \text{LR} & \text{LR} \\
\text{LR} & 0.21 & 0.40 & 0.60 & 0.84 & 0.51 \\
(2S-W) & 0.39 & 0.62 & 0.80 & 0.95 & 0.69 \\
(0.61,0.61) & 0.34 & 0.54 & 0.73 & 0.91 & 0.63 \\
(0.87,0.87) & 0.34 & 0.55 & 0.74 & 0.92 & 0.64 \\
(1.12,1.12) & 0.35 & 0.55 & 0.75 & 0.93 & 0.65 \\
(1.46,1.46) & 0.35 & 0.56 & 0.75 & 0.93 & 0.65 \\
\text{Stat} & \text{LR} & \text{LR} & \text{LR} & \text{LR} & \text{LR} & \text{LR} \\
\text{LR} & 0.22 & 0.40 & 0.60 & 0.86 & 0.52 \\
(2S-W) & 0.39 & 0.62 & 0.81 & 0.95 & 0.69 \\
\hline
\end{array}\]

The LR and $D-W_1$ tests are a close second in overall performance. The $DW_{\infty}$ test does better than the LR test in the middle of $B_1$ (i.e., $B \propto (1,1)'$), but the opposite is true at the boundary of $B_1$ (i.e., $\beta \propto (1,0)'$) (except when $\rho = -0.6$, where there are equal). The $D-W_0$ test suffers from poor relative performances in two cases. It does very well in the middle of $B_1$ in fact, it attains the envelope power there — but sacrifices considerable power at the boundaries of $B_1$, except when $\rho = -0.6$.

Table 3 presents power results for the parameter space $B_2 = R_1 \times R$. 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 0.3, 0.5, 0.7, and 0.9. Thus, the results of Table 3 are analogous to those of column five of Table 2. Analogous average power summary statistics are given in Tables 4 and 5 (discussed below) for the same reasons.

The results of Table 3 cover three different directions of the alternative from the null and three different values of the correlation $\rho$. 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 3 are worthy of note. First, the results for $D-W_{\infty}$, LR, 2S-W, and the envelope power function (the last four rows of the table) are insensitive to the value of $\rho$. In fact, $D-W_0$ is the only test that shows significant sensitivity to $\rho$. 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 $p_1$. Third, the $D-W_0$ test almost dominates the $D-W_1$ test. Thus, the $D-W_\infty$ and LR tests are the best overall tests with the $D-W_\infty$ 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 4 presents results for the high-dimensional one-sided parameter space $B_3 = R_+^n$ with the single covariance matrix $\Omega_3 = I_n$. In this case, $D-W_1$ and $D-W_\infty$ 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 0.02 of) the power envelope in the middle of the parameter space.

Table 5 presents results for the high-dimensional mixed one- and two-sided parameter space $B_4 = R_+^n \times R_+^m$. As with parameter space $B_3$, 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 $D-W_\infty$ tests have similar power, although $D-W_\infty$ is somewhat better due to its performance for the fourth and fifth directions. The comparison between $D-W_\infty$ and LR is much the same as with the low-dimensional parameter spaces $B_1$ and $B_2$. That is, $D-W_\infty$ typically does better for alternatives that are in the middle of the parameter space, but worse for alternatives that are on the boundary.

<table>
<thead>
<tr>
<th>Table 3</th>
<th>Average power results for parameter space $B_2 = \mathbb{R}^+ \times \mathbb{R}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistic</td>
<td>$\rho$</td>
</tr>
<tr>
<td>$D-W_0$</td>
<td>0.63</td>
</tr>
<tr>
<td>$D-W_1$</td>
<td>0.58</td>
</tr>
<tr>
<td>$D-W_\infty$</td>
<td>0.60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4</th>
<th>Average power results for parameter space $B_3 = R_+^n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistic $/ \beta$</td>
<td>$0.05$</td>
</tr>
<tr>
<td>$D-W_0$</td>
<td>0.70</td>
</tr>
<tr>
<td>$D-W_1$</td>
<td>0.70</td>
</tr>
<tr>
<td>$D-W_\infty$</td>
<td>0.68</td>
</tr>
</tbody>
</table>
Table 5
Average power results for parameter space $B_4 = R_+^2 \times R^3$

<table>
<thead>
<tr>
<th>Statistic $/\beta \propto (1,1,1,0,0,0,0)$</th>
<th>$(1,1,1,0,0,0,0)$</th>
<th>$(1,0,0,0,1,1,1)$</th>
<th>$(0,0,0,0,0,1,1)$</th>
<th>$(1,0,0,0,0,0,1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D-W_0$</td>
<td>0.60</td>
<td>0.79</td>
<td>0.51</td>
<td>0.05</td>
</tr>
<tr>
<td>$D-W_1$</td>
<td>0.67</td>
<td>0.72</td>
<td>0.59</td>
<td>0.47</td>
</tr>
<tr>
<td>$D-W_2$</td>
<td>0.66</td>
<td>0.69</td>
<td>0.60</td>
<td>0.53</td>
</tr>
<tr>
<td>$D-W_3^\infty$</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
<td>0.60</td>
</tr>
<tr>
<td>$2S-W_1$</td>
<td>0.50</td>
<td>0.49</td>
<td>0.54</td>
<td>0.57</td>
</tr>
<tr>
<td>Envelope</td>
<td>0.80</td>
<td>0.79</td>
<td>0.83</td>
<td>0.86</td>
</tr>
</tbody>
</table>

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 2–5 report results for only three values of $c$, viz., 0, 1, and $\infty$. 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 $\infty$ 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 $D-W_1$ is very much closer to that of $D-W_0$ 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.

5.3. Normal linear regression with unknown variance

The model considered here is

$$y = \delta I + X\beta + U,$$

where $y \in R_T$, $I = (1, \ldots, 1)' \in R_T$, $X \in R_T^{T \times 2}$, $U \sim N(0, \sigma^2 I_T)$, $\delta \in R$, and $\beta \in R^2$. The regressors $X$ are iid bivariate normal with variances equal to 1 and correlation coefficient $\rho_X$ for $\ell = 1, \ldots, T$. The error variance $\sigma^2$ is assumed to be unknown. Without loss of generality, power is computed for the case of $\delta = 0$ and $\sigma^2 = 1$. Three sample sizes, $T = 25, 50$, and 100, are considered. Three
values of $\rho_X$ are considered: 0, −0.6, and 0.6. The values of $\beta$ that are considered are the same as those in Table 2 multiplied by $(1 - \rho_X^2)^{1/2}/T^{1/2}$. (The latter multiplicative is employed to guarantee that the tests have almost the same power as in Table 2 for each value of $T$ and $\rho_X$.) We note that $\rho_X = 0.6$ in model (34) corresponds to $\rho = -0.6$ in model (25) and vice versa, because $\rho$ corresponds to the correlation between the two least squares estimators of $\beta_1$ and $\beta_2$, which has the opposite sign from the correlation between the regressors $X_{it}$ and $X_{at}$.

The tests that we consider are the directed $F$-tests described in Comment 3 following Theorem 1 and the LR test for unknown variance. We refer to these tests as estimated variance tests. For comparative purposes, we also consider the directed Wald tests of Section 2 and the LR test for known variance. We refer to these tests as known variance tests. We compare the tests' size-adjusted power using 40,000 simulations for the critical values and for the power results.

There are two items that one might be interested in concerning the power of tests for model (34). The first is the absolute difference in power between the known variance and estimated variance tests. The second is the effect of estimation on the relative power of the directed tests and the LR test. In particular, do the comparative results of Tables 2–5 hold up when the variance is estimated?

The answers to these questions are apparent without doing any simulations, so we keep our summary of the simulation results brief. The answers are as follows. First, the effect of estimating the variance can be expected to be smaller than that for a standard (two-sided) $F$-test, because the directed tests and LR test for a restricted alternative parameter space $B$ use more a priori information than the $F$-test. The effect of estimating the variance on a standard $F$-test is given by the noncentral $F$ distribution with $T - 1$ and $\infty$ denominator degrees of freedom and can be obtained by looking up tables or charts of the power of the $F$-test, e.g., see Scheffé (1959). For significance level $\alpha = 0.05$, the differences are small even for $T$ as small as 30. Since these differences are small for two-sided $F$-tests, they are also small for directed $F$-tests and the LR test with restricted parameter space. Furthermore, the differences are sufficiently small that estimation of the variance cannot have a significant effect on the power of the directed Wald tests relative to the LR test for known variance, as given in Tables 2–5.

The above argument is substantiated by the simulation results. We report results that are averaged over the 24 cases considered in Table 2 (three values of $\rho = -\rho_X$ by eight values of $\beta$). For the $DW_\infty$ test, the average power difference between the known variance and the estimated variance tests are 0.006, 0.012, and 0.021 for $T = 100$, 50, and 25 respectively (with simulation standard errors 0.0004, 0.0005, and 0.0007 respectively). For the LR test, the analogous power differences are 0.006, 0.012, and 0.023. They represent 0.9%, 1.9%, and either 3.3% or 3.6% decreases in power due to the estimation of the variance. Needless to say, these are quite small. Also, they are almost the same for the $DW_\infty$ and LR tests (and for the $DW_\epsilon$ tests with $\epsilon = 1/3$, 1, and 3).
One can assess the impact of variance estimation on the relative performances of the LR and $DW_{\infty}$ tests by comparing the difference in power between the LR and $DW_{\infty}$ tests for known variance with their difference in power with estimated variance. The smaller is the difference in power differences, the more accurately do Tables 2–5 reflect the relative powers of the LR and directed tests for the case of estimated variance. The absolute value of the difference in power differences over 24 cases are 0.0023, 0.0032, and 0.0055 for $T = 100, 50,$ and 25 respectively. These numbers are sufficiently small that Tables 2–5 give an accurate comparison of the directed tests and the LR test for normal linear regression models with unknown variance and sample size 25 or larger.

Appendix

For notational simplicity, we suppress the subscript $c$ on $Q_c$ in the proofs below. The proof of Theorem 1(a) uses the following Lemmas. Let

$$LR(\theta_0) = \exp\left[\frac{1}{2} \tilde{\theta}^T \mathcal{A} \tilde{\theta} \right] \{1(\theta_0 + h \in \Theta^*) \times \exp[-\frac{1}{2} (\tilde{\theta} - \theta)^T \mathcal{A} (\tilde{\theta} - \theta)] \frac{dQ_c(h)}{K},$$

(A.1)

where $\tilde{\theta} = \hat{\theta} - \theta_0$ and $\hat{\theta}$ is the (unrestricted) LS estimator of $\theta$ (i.e., $\hat{\theta} = (\sigma^2 \mathcal{A})^{-1} [X:G]Y$).

Lemma A.1. Under Assumption 1, $LR(\theta_0) = \overline{LR}(\theta_0)$.

Lemma A.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 $A = [I_p: - \mathcal{F}_3, \mathcal{F}_3^{-1}]'$ and $H = [I_p: 0] \in R^{p \times s}$.

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

$$LR(\theta_0) = \int \left(2\pi\right)^{-p/2} \det^{1/2}(A' \mathcal{A} A/c)$$

$$\times \exp\left(\frac{1}{2} \tilde{\theta} (\mathcal{A} \tilde{\theta} - \mathcal{A} \lambda - \lambda A' \mathcal{A} A \lambda/c)\right) d\lambda/K,$$

(A.2)
since $\theta_0 + h \in \Theta^\ast$ iff

$$
\lambda \in \{b \in R^p: \theta_0 + Ah \in \Theta^\ast\} = \left\{ b \in R^p: \left( \delta_0 - \mathcal{F}_2^{-1} \mathcal{F}_2 h \right) \in B \times A \right\} = B,
$$

where the second equality holds under Assumption 3.

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

$$
\mathcal{F}^\perp \mathcal{F} \lambda \mathcal{F}^\perp \lambda - (\mathcal{F}^\perp \lambda \mathcal{F}^\perp \lambda - h' \mathcal{F} h/c)
= \frac{c}{1+c}(P^\perp \bar{\lambda}^\prime P^\perp \bar{\lambda} - \left( h - P^\perp \bar{\lambda} \frac{c}{1+c} \right) \mathcal{F} \frac{1+c}{c} \left( h - P^\perp \bar{\lambda} \frac{c}{1+c} \right)),
$$

(A.3)

using the fact that $(P^\perp \bar{\lambda}^\prime P^\perp \lambda - 0) \in V^\perp$.

By Lemma A.2, $P^\perp \bar{\lambda} = A \bar{\lambda} \theta = A \bar{\beta}$. In addition, $A \mathcal{F} A' = \mathcal{F}_1 - \mathcal{F}_2 \mathcal{F}_2^{-1} \mathcal{F}_2' = X'M_GX/\sigma^2$. Thus, the r.h.s. of (A.3) equals

$$
\frac{c}{1+c} \mathcal{F} \left( X'M_GX/\sigma^2 \right) \bar{\beta} - \left( \bar{\lambda} - \bar{\beta} \frac{c}{1+c} \right) \mathcal{F} \left( X'M_GX/\sigma^2 \right) \left( \bar{\lambda} - \bar{\beta} \frac{c}{1+c} \right).
$$

(A.4)

Combining (A.2)–(A.4) gives the desired result $LR(\theta_0) = D^\prime W / K$.

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

Part (c) holds by (7), the Neyman–Pearson Lemma, and Theorem 1(a). □

**Proof of Lemma A.1.** Let $\ell(\theta) = \log f(Y, \theta)$ and $D\ell(\theta) = \partial/\partial \theta \ell(\theta)$. Simple algebra yields $D\ell(\theta_0) = (U'X', U'G)'$ and $\bar{\beta} = \mathcal{F}^{-1} D\ell(\theta_0)$. We can write

$$
LR(\theta_0) = \int 1(\theta_0 + h \in \Theta^\ast) \exp \left( \ell(\theta_0 + h) - \ell(\theta_0) \right) dQ(h)/K.
$$

(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:

$$
\ell(\theta_0 + h) - \ell(\theta_0)
= -\frac{1}{2\sigma^2} \left[ \Sigma_1^\prime (Y_i - X_i' h_1 - G_i h_2)^2 - \Sigma_1^\prime (Y_i - G_i \delta_0)^2 \right]
= -\frac{1}{2\sigma^2} \left[ -2\Sigma_1^\prime (Y_i - G_i \delta_0)(X_i' h_1 + G_i h_2) + \Sigma_1^\prime (X_i' h_1 + G_i h_2)^2 \right]
- D\ell(\theta_0)^\prime h - \frac{1}{2} \mathcal{F} h \mathcal{F} \bar{\lambda} - \frac{1}{2} \bar{\lambda} \mathcal{F} h - \frac{1}{2} \bar{\lambda} \mathcal{F} \bar{\lambda} - \frac{1}{2} (\bar{\lambda} - h' \mathcal{F} \bar{\lambda} - h).
$$

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

Proof of Lemma A.2. It suffices to show that (1) $AHd = 0$ ∀ $d \in V$ and (2) $AHm = m$ ∀ $m \in V^\perp$. To show (1), note that $d \in V$ iff $d = (0', d_2')'$ for some $d_2 \in R^q$. Thus,

$$AHd = A[l_p : 0]
\begin{pmatrix}
0 \\
\vdots \\
0 \\
d_2 \\
\end{pmatrix} = 0.$$

To show (2), note that $m \in V^\perp$ iff $d' \mathcal{S} m = 0$ ∀ $d \in V$ iff $[0 : l_q] \mathcal{S} m = 0$ iff

$$[\mathcal{S}_2' : \mathcal{S}_3] = \begin{pmatrix} m_1 \\ m_2 \\ \end{pmatrix} = 0,$$

where $m = (m_1', m_2')'$, iff $m_2 = -\mathcal{S}_3^{-1} \mathcal{S}_2' m_1$ for $m_1 \in R^p$ and $m_2 \in R^q$. Thus, for $m \in V^\perp$,

$$AHm = \begin{bmatrix}
I_p & 0 \\
\mathcal{S}_3^{-1} \mathcal{S}_2' & 0
\end{bmatrix}
\begin{bmatrix}
m_1 \\ -\mathcal{S}_3^{-1} \mathcal{S}_2' m_1
\end{bmatrix} = \begin{bmatrix}
m_1 \\ -\mathcal{S}_3^{-1} \mathcal{S}_2' m_1
\end{bmatrix} = m,$$

as desired. □

The proof of Theorem 2 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, L.M., and L.R. test statistics are asymptotically equivalent, under the null and local alternatives, to the Neyman–Pearson likelihood ratio $\ell_{\mathcal{R}}(\theta_0)$, defined below. The proof uses the following definitions, lemmas, and theorems: Let

$$\ell_{\mathcal{R}}(\theta_0) = \int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) f_{\theta}(\theta_0 + h/\sqrt{T}) dQ(h)/K,$$

$$\overline{\ell}_{\mathcal{R}} = \int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) \exp[\frac{1}{2} \mathcal{S} \mathcal{S} \bar{\theta} - \frac{1}{2} (\bar{\theta} - h)' \mathcal{S} (\bar{\theta} - h)] dQ(h)/K,$$

where

$$K = \int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) dQ(h)$$

$$= \int 1(z \in B) dN(z), \quad \bar{\theta} = \mathcal{S}^{-1} D'_{\mathcal{T}}(\theta_0)/\sqrt{T},$$

and $N(z)$ denotes a $N(0, c(A', \mathcal{S}A)^{-1})$ distribution function evaluated at $z$.

Note that $\ell_{\mathcal{R}}(\theta_0)$ is the Neyman–Pearson likelihood ratio statistic for testing the simple null hypothesis that $Y_T \sim f_{\theta}(\theta_0)$ against the simple alternative that $Y_T \sim \int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) f_{\theta}(\theta_0 + h/\sqrt{T}) dQ(h)/K$.

Lemma A.3. Under the null hypothesis and Assumptions NL1–NL3, $\sqrt{T}(\bar{\theta} - \theta_0) \rightarrow N(0, 0)$.\]
Lemma A.4. Under the null hypothesis and Assumptions NL1–NL3, \( \ell_r(\theta_0) - \ell_r^c \sim N(0, \mathcal{F}^{-1}) \) for all \( 0 \leq c \leq \infty \).

Lemma A.5. 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 \( h \in \mathbb{R}^p \).

The proofs of Lemmas A.3 and A.4 are analogous to those of Lemmas A.1 and A.2 of Andrews and Ploberger (1994). For brevity, they are not given here. The proof of Lemma A.5 is given below.

For \( 0 < c < \infty \), define

\[
\begin{align*}
\bar{W}_T &= (H \bar{\theta}')(H^{-1} H')^{-1} H \bar{\theta}, \\
D - \bar{W}^{cT} &= (1 + c)^{-\frac{p}{2}} \exp \left( \frac{c}{2 \sqrt{1 + c}} \bar{W}_T \right) \\
&\times \Phi \left( B \frac{c}{1 + c} \frac{\bar{\beta}}{\sqrt{T}}, \frac{c}{1 + c} (\mathcal{F}_1 - \mathcal{F}_2 \mathcal{F}_3^{-1} \mathcal{F}_2')^{-1} / T \right), \\
D - \bar{W}^{\infty T} &= \bar{W}_T + 2 \log \Phi(B \bar{\beta} / \sqrt{T}, (\mathcal{F}_1 - \mathcal{F}_2 \mathcal{F}_3^{-1} \mathcal{F}_2')^{-1} / T), \\
D - \bar{W}^{0T} &= d' \bar{\beta} / \sqrt{d'} (\mathcal{F}_2 \mathcal{F}_3^{-1} \mathcal{F}_2')^{-1} d)^{1/2},
\end{align*}
\]

where

\[
d = \frac{d}{C} \Phi(B, 0, (\mathcal{F}_1 - \mathcal{F}_2 \mathcal{F}_3^{-1} \mathcal{F}_2')^{-1} / T).
\]

Theorem A.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) \( \ell_r(\theta_0) - \ell_r^c \overset{p}{\to} 0 \), (b) \( \ell_r^c \times K = D - \bar{W}^{cT} \) provided Assumption NL5 also holds,

(c) \( D - \bar{W}^{cT} = D - \bar{W}^{cT} \overset{p}{\to} 0 \), (d) \( D - \bar{W}^{cT} = D - \bar{W}^{cT} \overset{p}{\to} 0 \), and (e) \( D - \bar{W}^{cT} = D - \bar{W}^{cT} \overset{p}{\to} 0 \).

In addition, parts (c), (d), and (e) hold for \( c = 0 \) and \( c = \infty \).

The proof of Theorem A.1 is given below.

Proof of Theorem 2(a). By the positive homogeneity of \( B \) (Assumption 5),

\[
\Phi \left( B \frac{c}{1 + c} \frac{\bar{\beta}}{\sqrt{T}}, \frac{c}{1 + c} (\mathcal{F}_1 - \mathcal{F}_2 \mathcal{F}_3^{-1} \mathcal{F}_2')^{-1} / T \right)
\]

\[
= \Phi \left( B \frac{c}{1 + c} \bar{\beta}, \frac{c}{1 + c} (\mathcal{F}_1 - \mathcal{F}_2 \mathcal{F}_3^{-1} \mathcal{F}_2')^{-1} \right).
\]

Theorem 2(a) now follows for all \( 0 \leq c \leq \infty \) from the combination of (i) Theorem A.1 parts (c)–(e), (ii) \( \theta_0 \overset{d}{\to} N(0, \mathcal{F}^{-1}) \) under \( \{ \theta_T : T \geq 1 \} \), (iii) the continuity of \( D - \bar{W}^{cT} \) as a function of \( \bar{\theta} \) and \( \mathcal{F} \), and (iv) the continuous mapping theorem.
Condition (ii) holds because

\[ \bar{\theta} = \mathcal{S}^{-1} D \ell_T(\theta_0) / \sqrt{T} \]
\[ = \mathcal{S}^{-1} D \ell_T(\tilde{\theta}_T) / \sqrt{T} + (\mathcal{S}^{-1} D^2 \ell_T(\tilde{\theta}_T) / \sqrt{T}) (-h / \sqrt{T}) \]
\[ \overset{d}{\rightarrow} N(h, \mathcal{S}^{-1}) \text{ under } \{ \theta_T : T \geq 1 \}. \tag{A.10} \]

where the second equality holds by element-by-element mean value expansions, \( \tilde{\theta}_T \) lies between \( \theta_0 \) and \( \theta_T \), \( -D^2 \ell_T(\tilde{\theta}_T) / T \overset{p}{\rightarrow} \mathcal{S} \) under \( \theta_0 \) and under \( \{ \theta_T : T \geq 1 \} \) using Assumption NL1 and contiguity, and the convergence in distribution uses Assumption NL2.

Proof of Lemma A.5. We make use of the following result, which follows, e.g., from Theorems 16.8 and 18.1 of Strasser (1985): If (i) \( f_T(\theta_T)/f_T(\theta_0) \overset{d}{\rightarrow} X \) under \( \theta_0 \) for some random variable \( X \) and (ii) \( EX = 1 \), then the densities \( \{ f_T(\theta_T) : T \geq 1 \} \) 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 \( \theta_0 \):

\[ f_T(\tilde{\theta}_T) / f_T(\theta_0) = \exp(\ell_T(\tilde{\theta}_T) - \ell_T(\theta_0)) \]
\[ = \exp(T^{-1/2} D \ell_T(\theta_0)'h + \frac{1}{2} h' D^2 \ell_T(\tilde{\theta}_T)h) \]
\[ \overset{d}{\rightarrow} \exp(Z' h - \frac{1}{2} h' \mathcal{S} h) \text{ under } \theta_0. \tag{A.11} \]

where \( \tilde{\theta}_T \) lies between \( \theta_T \) and \( \theta_0 \), \( Z \sim N(0, \mathcal{S}) \), and the convergence in distribution uses Assumptions NL1 and NL2.

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

Proof of Theorem A.1. By contiguity (Lemma A.5), it suffices to show that the Theorem holds under the null. In consequence, all probabilistic statements in this proof are made 'under \( \theta_0 \)': Part (a) holds by Lemma A.4. Next, consider part (b). Let \( \lambda \sim N(0, c(A' \mathcal{S} A)^{-1}) \) and \( h = A\lambda \). Then, \( h \sim Q = N(0, cA(A' \mathcal{S} A)^{-1} A' \mathcal{S} A) \) 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

\[ \bar{\ell}_{\mathcal{E}} = (2\pi)^{-p/2} \det^{1/2}(A' \mathcal{S} A / c) \]
\[ \times \int 1(\lambda/\sqrt{T} \in B) \exp(\frac{1}{2} \lambda' \mathcal{S} \bar{\lambda} - \frac{1}{2} A(\lambda - \theta)' \mathcal{S}(A\lambda - \bar{\theta})) \]
\[ - (A\lambda)' J_{\lambda}(\mathcal{E}) \lambda / c \) d\lambda / K, \tag{A.12} \]

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 (A.12), with $A\lambda^*$ replaced by $h$ for simplicity, can be shown to simplify as follows:

$$
\bar{\theta}' A \bar{\theta} - (h - \bar{\theta})' A (h - \bar{\theta}) - h'h/c
= \frac{c}{1+c} (P^\perp \bar{\theta})' A P^\perp \bar{\theta} - \left[ h - P^\perp \bar{\theta} - \frac{c}{1+c} \right]' \frac{1+c}{c} \left[ h - P^\perp \bar{\theta} - \frac{c}{1+c} \right],
$$

(A.13)

using the fact that $h'h/\ell = 0 \forall h \in V^\perp, \ell \in V$.

Combining (A.12) and (A.13) yields

$$
\bar{\ell} \bar{c} \times K = (1+c)^{-p/2} \exp \left[ \frac{c}{2} P^\perp \bar{\theta} A \bar{\theta} \right] \cdot \int_{1/\sqrt{T} \in B}(2\pi)^{-p/2} \det^{1/2} A \frac{1+c}{c} \exp \left[ -\frac{1}{2} \left( A \lambda^* - P^\perp \bar{\theta} - \frac{c}{1+c} \right)' A \frac{1+c}{c} \right] A \lambda^* - P^\perp \bar{\theta} - \frac{c}{1+c} d\lambda^*
$$

$$
= (1+c)^{-p/2} \exp \left[ \frac{c}{2} P^\perp \bar{\theta} A \bar{\theta} \right] \times \phi \left( B, \frac{c}{1+c} H \bar{\theta}/\sqrt{T}, \frac{c}{1+c} A' A \bar{\theta} A \bar{\theta} \right)
$$

$$
= (1+c)^{-p/2} \exp \left[ \frac{c}{2} P^\perp \bar{\theta} A \bar{\theta} \right] \times \phi \left( B, \frac{c}{1+c} H \bar{\theta}/\sqrt{T}, \frac{c}{1+c} \left( A' A \bar{\theta} A \bar{\theta} \right)^{-1}/T \right)
$$

where the second equality uses $P^\perp = AH$ (Lemma A.2) and a change of variables $(\lambda^* = \lambda^*/\sqrt{T})$ and the third equality uses $(A' A)^{-1} = H \bar{\theta}^{-1} H = (A_1 - A_2 A_3^{-1} A_2^T)^{-1}$. This completes the proof of part (b).

Part (c) is established as follows. Since $H\bar{\theta}_0 = 0$,

$$
\sqrt{T} H \bar{\theta} - H \bar{\theta} = \sqrt{T} \bar{\theta} - \bar{\theta} \sim P, 0
$$

(A.15)
by Lemma A.3. In addition,

$$\| \mathcal{F}_T(\hat{\theta}) - \mathcal{F}_T \| \leq \sup_{\theta \in \Theta_T^{0}} \| \mathcal{F}_T(\hat{\theta}) - \mathcal{F}(\theta) \| + \| \mathcal{F}(\hat{\theta}) - \mathcal{F} \| = o_p(1). \tag{A.16}$$

where the inequality holds wp → 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 < c ≤ ∞.

For parts (d) and (e) with 0 < c ≤ ∞, it suffices to show that

$$W_T - LM_T \overset{p}{\to} 0 \text{ and } LM_T - LR_T \overset{p}{\to} 0, \tag{A.17a}$$

$$\sqrt{T} \hat{\beta} - H\mathfrak{F}_T(\hat{\theta})/\sqrt{T} \overset{p}{\to} 0, \tag{A.17b}$$

$$(\hat{\mathcal{F}}_1 - \mathcal{F}_2)^{-1} (\hat{\mathcal{F}}_2 - \mathcal{F}_3)^{-1} (\hat{\mathcal{F}}_3 - \mathcal{F}_4)^{-1} \overset{p}{\to} 0 \tag{A.17c}$$

Condition (A.17c) holds by (A.16) and Assumption NL1(e). Condition (A.17a) is a standard result and its proof is similar to proofs in the literature. For brevity, its proof is omitted. Condition (A.17b) typically is established as part of the proof of condition (A.17a). Again, for brevity, its proof is omitted.

We now turn to the proof of Theorem 2(b). First, we establish another contiguity result.

**Lemma A.6.** Under Assumptions NL1–NL3, NL5, 2, 4, and 5, the densities \{f_T(\theta_0 + h/\sqrt{T} \in \Theta^*) \} are contiguous to the densities \{f_T(\theta_0); \theta \geq 1\} for all 0 < c < ∞.

Lemma A.6 and Theorem A.1 combine to give:

**Theorem A.2.** Under the local alternative densities \{\int_0 \int_{\Theta^*} f_T(\theta_0 + h/\sqrt{T}) dQ(h)/K; \theta \geq 1\} and Assumptions NL1–NL5, 2, 4, and 5, for all 0 < c < ∞, we have (a) \(\ell_T(\theta_0) - \ell_T \overset{p}{\to} 0\), (b) \(\ell_T \times K = D - \hat{W}_{cT} \), (c) \(D - \hat{W}_{cT} \overset{p}{\to} 0\), (d) \(D - W_{cT} - D - LM_{cT} \overset{p}{\to} 0\), and (e) \(D - LM_{cT} - D - LR_{cT} \overset{p}{\to} 0\).

**Proof of Theorem 2(b).** Let \(\alpha_T\) be the rejection probability of \(\varphi_T\) under \(\theta_0\). Let \(k^*_c > 0\) and \(\lambda_{cT} \in [0, 1]\) be constants such that the likelihood ratio test \(\gamma_T = 1(\ell_T(\theta_0) > k^*_c) + \lambda_{cT} 1(\ell_T(\theta_0) = k^*_c)\) has rejection probability \(\alpha_T\) under \(\theta_0\). Then, by the Neyman–Pearson Lemma (e.g., see Lehmann (1959, Theorem 3.1, p. 65), for
all $T \geq 1$,
\[ \int \varphi_T \left[ \int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) f_T(\theta_0 + h/\sqrt{T}) \, dQ(h)/K \right] \, d\mu_T \]
\[ \leq \lim_{T \to \infty} \int \varphi_T \left[ \int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) f_T(\theta_0 + h/\sqrt{T}) \, dQ(h)/K \right] \, d\mu_T. \]  
(A.18)

By Corollary 15.11 of Strasser (1985), the $\lim_{T \to \infty}$ on the r.h.s of the inequality in the statement of Theorem 2(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 A.6 below.

This result, inequality (A.18), and Fubini’s Theorem yield
\[ \lim_{T \to \infty} \int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) \left[ \int \varphi_T f_T(\theta_0 + h/\sqrt{T}) \, d\mu_T \right] \, dQ(h)/K \]
\[ \leq \lim_{T \to \infty} \int \varphi_T \left[ \int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) f_T(\theta_0 + h/\sqrt{T}) \, dQ(h)/K \right] \, d\mu_T \]
\[ = \lim_{T \to \infty} \int 1(\mathcal{L}_T \geq k_{2T}/K) \left[ \int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) \right. \]
\[ \times f_T(\theta_0 + h/\sqrt{T}) \, dQ(h)/K \right] \, d\mu_T \]
\[ = \lim_{T \to \infty} \int 1(\mathcal{L}_T \geq k_{2T}) \left[ \int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) \right. \]
\[ \times f_T(\theta_0 + h/\sqrt{T}) \, dQ(h)/K \right] \, d\mu_T \]
\[ = \lim_{T \to \infty} \int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) \left[ \int \varphi_T f_T(\theta_0 + h/\sqrt{T}) \, d\mu_T \right] \, dQ(h)/K. \]  
(A.19)

where the first equality holds because $k_{2T}^* - k_{2T}/K \to 0$ and $\mathcal{L}_T(\theta_0)$ has an absolutely continuous asymptotic distribution under $\int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) \, dQ(h)/K$, the second equality holds because $\mathcal{L}_T(\theta_0) = D_2(\mathcal{L}_T)/K \to 0$ under $\int 1(\theta_0 + h/\sqrt{T} \in \Theta^*) f_T(\theta_0 + h/\sqrt{T}) \, dQ(h)/K$ by Theorem A.2, and the third inequality holds by Fubini’s Theorem. The proof is analogous for $D_2(LM_T)$ and $D_2(LR_T)$. □

**Proof of Lemma A.6.** As in the proof of Lemma A.5, it suffices to show that (i) $\mathcal{L}_T(\theta_0) \Rightarrow \mathcal{L}_T(\theta_0) \Rightarrow \mathcal{L}_T(\theta_0)$ under $\theta_0$ and (ii) $\mathcal{L}_T(\theta_0) = 1$. Condition (i) holds by Theorem A.1(a)–(c) and Theorem 2(a). Condition (ii) is obtained as follows: Let $\mathcal{E} = (\mathcal{F}_1 - \mathcal{F}_2)^{-1} \mathcal{F}_1^2 \mathcal{F}_2$. Then, we have
\[ E \mathcal{E} = (1 + c)^{-c/2} \int_{\mathcal{E}} \exp \left( \frac{1}{2} \frac{c}{1 + c} Z'Z \right) \int_{\mathcal{E}} (2\pi)^{-c/2} \det^{-1/2}(\mathcal{E}). \]
\[ \times \exp \left[ -\frac{1}{2} \left( \lambda - \frac{c}{1+c} \right)^{1/2} Z \right] \]

\[ \times y^{-1} \left( \frac{c}{1+c} \right)^{1/2} \left( \frac{c}{1+c} \right) \] \[ \times (2\pi)^{-p/2} \exp \left( -\frac{1}{2} z^2 \right) \] \[ \times (2\pi)^{-p/2} \exp \left( -\frac{1}{2} z^2 \right) \] \[ = (1+c)^{-p/2} \int_B \mathcal{L}^p \det^{-1/2} \left( \mathbf{y}^\prime \right) \] \[ \times \exp \left( -\frac{1}{2} \left[ \frac{1}{1+c} Z'Z + \left( \mathbf{Z} - \left( \frac{1+c}{c} \right)^{1/2} \mathbf{y}^{-1/2} \right)^{1/2} \right] \right) \] \[ \times \left( \mathbf{Z} - \left( \frac{1+c}{c} \right)^{1/2} \mathbf{y}^{-1/2} \mathbf{y}^{1/2} \left( \frac{c}{1+c} \right) \right) d\lambda /K, \] \[ (A.20) \]

where the second equality uses Fubini’s theorem. Via some algebra, the expression in square brackets in the exponent of the r.h.s of (A.20) equals

\[ \left( \mathbf{Z} - \left( \frac{c}{1+c} \right)^{1/2} \mathbf{y}^{-1/2} \right)^{1/2} \left( \mathbf{Z} - \left( \frac{c}{1+c} \right)^{1/2} \mathbf{y}^{-1/2} \right) + \frac{1}{1+c} \lambda^\prime \mathbf{y}^{-1/2} \lambda. \]

(A.21)

Substituting this result in (A.20) gives

\[ \mathcal{E} \mathcal{L}^p /K = (1+c)^{-p/2} \int_B \mathcal{L}^p \det^{-1/2} \left( \mathbf{y}^\prime \right) \exp \left( -\frac{1}{2} \left[ \frac{1}{1+c} \lambda^\prime \mathbf{y}^{-1/2} \right] \right) d\lambda /K \]

\[ = \Phi_p(B, 0, (1+c)\mathbf{y}^\prime)/K = 1, \] \[ (A.22) \]

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

\[ \square \]

Proof of Lemma 1. Assumption NL1(a) holds by Assumption A(a). NL1(b) holds by Assumption A(c). NL1(c) holds with

\[ F(\theta) = -E \frac{\theta^2}{\partial \theta \partial \theta} \log g_\theta(\theta) \]
provided a uniform WLLN can be established. The Markov property (Assumption 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 Assumptions A(b) and (e), the ergodic theorem implies that \(-T^{-1}D^2 \tau_T(\theta) \overset{a}{\to} \mathcal{H}(\theta)\) \(\forall \theta \in \Theta_0\). 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 \(\Theta_0\) using Assumptions A(b), (e), and (c).

Assumption NL1(d) holds, because \(\mathcal{H}(\theta)\) is continuous on \(\Theta_0\) by the dominated convergence theorem using Assumptions A(c) and (e) and \(\Theta_0\) is compact. Assumption NL1(e) holds by Assumption A(f).

To verify Assumption NL2, we use a martingale difference triangular array (MDTA) central limit theorem of Hall and Heyde (1980, Corollary 3.1, p. 58). By assumption, \(T^{-1/2}D^2 \tau_T(\theta_T) = T^{-1/2} \sum_{t \geq 1} \partial / \partial \theta \log g_t(\theta_T)\). Let \(E_T\) denote the expectation operator under \(\theta_T\).

\[
\left\{ \left( \frac{\partial}{\partial \theta} \log g_t(\theta_T), \mathcal{F}_{t-1} \right) : m < t \leq T, \ T \geq 1 \right\}
\]

is a MDTA under \(\{\theta_T; T \geq 1\}\), because

\[
E_T \left( \frac{\partial}{\partial \theta} \log g_t(\theta_T) | \mathcal{F}_{t-1} \right) = E_T \left( \frac{\partial}{\partial \theta} \log g_t(\theta_T) | S_{t,m}, X_{t,m} \right) = \int \frac{\partial}{\partial \theta} g_t(\theta_T) d\lambda(s_t) = \frac{\partial}{\partial \theta} \int g_t(\theta_T) d\lambda(s_t) = 0,
\]

(A.23)

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

\[
X_m = \frac{\partial}{\partial \theta} \log g_t(\theta_T) / \sqrt{T}.
\]

By the Cramer–Wold device, it suffices to consider the case where \(\theta\) is a scalar. Hall and Heyde’s condition (3.21) holds because \(\mathcal{F}_t\) does not depend on \(T\). The other two conditions of Corollary 3.1 are:

\[
\forall \varepsilon > 0, \Sigma_t E (X_m^2 1(\{X_m \geq \varepsilon\}) | \mathcal{F}_{t-1} ) \overset{P}{\to} 0, \quad (A.24)
\]

\[
\Sigma_t E (X_m^2 | \mathcal{F}_{t-1} ) \overset{P}{\to} \eta \text{ for some constant } \eta > 0. \quad (A.25)
\]
We will show that (A.24) and (A.25) hold under $\theta_0$ with

$$\eta = \mathbb{E} \frac{\partial}{\partial \theta} \log g_t(\theta) \bigg| \frac{\partial}{\partial \theta'} \log g_t(\theta) \bigg| = \mathcal{F}.$$ 

By contiguity (Lemma A.5), then, (A.24) and (A.25) also hold under $\{\theta_T: T \geq 1\}$, which is required for the application of Corollary 3.1.

Eq. (A.24) can be established under $\theta_0$ by establishing $L^1$-convergence of the left-hand side to 0. By stationarity, the $L^1$-norm of the left-hand side of (A.24) equals

$$\mathbb{E} \left( \frac{\partial}{\partial \theta} \log g_t(\theta_T) \right)^2 \mathbb{E} \left( \left| \frac{\partial}{\partial \theta} \log g_t(\theta_T) \right| > \sqrt{T} \right)$$

$$\leq \mathbb{E} \sup_{\theta \in \Theta} \left( \frac{\partial}{\partial \theta} \log g_t(\theta) \right)^2 \mathbb{E} \left( \left| \frac{\partial}{\partial \theta} \log g_t(\theta) \right| > \sqrt{T} \right)$$

$$\to 0 \quad (A.26)$$

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

Eq. (A.25) can be established by taking a mean value expansion of its left-hand side about $\theta_0$:

$$\frac{1}{T} \sum_{i=1}^{T} \mathbb{E} \left( \left( \frac{\partial}{\partial \theta} \log g_t(\theta_T) \right)^2 \bigg| \mathcal{F}_{t-1} \right)$$

$$= \frac{1}{T} \sum_{i=1}^{T} \mathbb{E} \left( \left( \frac{\partial}{\partial \theta} \log g_t(\theta) \right)^2 \bigg| \mathcal{F}_{t-1} \right)$$

$$+ \frac{2}{T} \sum_{i=1}^{T} \mathbb{E} \left( \frac{\partial}{\partial \theta} \log g_t(\theta_T) \frac{\partial^2}{\partial \theta \partial \theta'} \log g_t(\theta_T) \bigg| \mathcal{F}_{t-1} \right). \quad (A.27)$$

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

$$\frac{2\|h\|}{\varepsilon \sqrt{T}} \mathbb{E} \sup_{\theta \in \Theta} \left\| \frac{\partial}{\partial \theta} \log g_t(\theta) \right\| \cdot \left\| \frac{\partial^2}{\partial \theta \partial \theta'} \log g_t(\theta) \right\|, \quad (A.28)$$

which is finite for all $T$ by Assumption A(e) and, hence, converges to 0. Thus, the second term of (A.27) converges in probability to 0 and the verification of Assumption NL 2 is complete.

Sufficient conditions for Assumption NL 3 are: (i) $\Theta$ is compact, (ii) $\log g_t(\theta)$ is continuous in $\theta$ on $\Theta$ with probability one under $\theta_0$, (iii) $\sup_{\theta \in \Theta} \left( (1/T) \sum_{i=1}^{T} (\log$
\( g_t(\theta) - E \log g_t(\theta) \mid \theta \sim 0 \) under \( \theta_0 \), and (iv) \( E \log g_t(\theta) \) is uniquely maximized over \( \Theta \) at \( \theta_0 \) (e.g., see Amemiya (1985, Theorem 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 \( \theta \neq \theta_0 \),

\[
E \log g_t(\theta) - E \log g_t(\theta_0) = E \log [g_t(\theta)/g_t(\theta_0)] < \log E g_t(\theta)/g_t(\theta_0) = 0, 
\]

(A.29)

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

Assumption NL4 holds by the same argument as for Assumption NL3 with \( \Theta \) in place of \( \Theta \). ☐

References

Andrews, D.W.K., 1996. Admissibility of the likelihood ratio test when the parameter space is restricted under the alternative. Econometrica 64, 705–718.
Wald, A., 1943. Tests of statistical hypotheses concerning several parameters when the number of observations is large. Transactions of the American Mathematical Society 54, 426–482.