POWER OF TESTS IN BINARY RESPONSE MODELS

By N. E. Savin and A. H. Würtz[1]

Department of Economics,

108 Pappajohn Bus. Adm. Bldg.

University of Iowa

Iowa City, IA 52242.

May 22, 1996.

Most hypotheses in binary response models are composite. The null hypothesis is usually that one or more slope coefficients are zero. Typically, the sequence of alternatives of interest is one in which the slope coefficients are increasing in absolute value. In this paper, we prove that the power goes to zero for this sequence of alternatives of interest in cases which often occur in practice. The practical implication is that for the sequence of alternatives of interest the power is nonmonotonic. This is true for any non-randomized test with size less than one and for a wide class of binary response models which includes the logit and probit models.

Keywords: Composite hypothesis, finite sample power, likelihood ratio test, logit model, nonmonotonic power, probit model.

---

[1]We gratefully acknowledge the comments of Lars Muus, Forrest Nelson, George Neumann and Richard Smith. We are especially indebted to Joel Horowitz and Narayana Kocherlakota. The comments of a co-editor and a referee led to considerable improvement of the paper.

1. INTRODUCTION

Most hypotheses in binary response models are composite, that is, the null hypothesis restricts only a subset of the parameters. The remaining parameters are referred to as nuisance parameters. The null hypothesis is usually that one or more of the slope coefficients are zero. Typically, the sequence of alternatives of interest is one in which the slope coefficients are increasing in absolute value. In this paper, we prove that the power goes to zero for this sequence of alternatives in cases which often occur in practice.

The alternatives can conveniently be expressed using a ray in the parameter space. The power along a ray corresponds to a slice of the power surface. We show that there always exist a ray for which the power function of any non-randomized test with size less than one goes to zero. The conventional alternatives of interest correspond to a ray where the nuisance parameters are being held fixed. Such a ray is called an orthogonal ray. The power function evaluated along an orthogonal ray goes to zero under a condition which is often satisfied in practice.

Testing the null hypothesis that the slope coefficient is zero, the condition is satisfied if the corresponding regressor is positive and the model includes an intercept. As an application we consider the LR test of a zero slope coefficient in a simple logit model. In this simple model, the intercept is the nuisance parameter. The empirical power function of the LR test is nonmonotonic and approaches zero for increasing values of the slope coefficient when the value of the intercept is held fixed. This result holds for any sample size even though the LR test is a consistent test.

In fields in which qualitative response models have traditionally been used, the regressors are positive. For example, in biological experiments a stimulus is applied to the subject at a stated dose, an intensity specified in units of concentration, weight, time, etc (Finney (1977)). In pesticide bioassays, the dose refers to the amount of a toxicant (Robertson and Preisler (1992)). In applied economics, regressors often take on only positive values. Many examples are cited in Amemiya (1981). In labor economics, female labor force participation is typically estimated using a logit or probit model with explanatory variables such as income of the spouse, years of education and age (Killingsworth and Heckman

(1986)). In travel demand, the explanatory variables include travel time, cost, income and age (Domencich and McFadden (1975)).

Regressors with only positive values can be demeaned. This transformation of the regressors changes the power function of the test. The power along the orthogonal ray in the transformed model can be monotonic. The reason is that an orthogonal ray in the transformed model is not an orthogonal ray in the model with the original regressors. The orthogonal ray in the transformed model translates into a ray in the original model where the values of the nuisance parameters are no longer fixed. Hence, it is not surprising that the power can be monotonic along the orthogonal ray in the transformed model. Demeaning the regressors is equivalent to choosing a different sequence of alternatives of interest. It is not the solution to the nonmonotonicity problem.

Nelson and Savin (1988,1990) reported nonmonotonic power of a sequence of alternatives for a Wald test of a zero slope coefficient in a logit model. In the case of the Wald test, nonmonotonic power occurs because the standard errors explode. This is a special feature of the Wald test. Our results apply to any non-randomized test with size less than one; for example, in a logit model this includes the uniformly most powerful test (Cox (1970)) and tests based on exact conditional inference (Cytel (1993)). It is important to emphasize that our results are derived for a finite sample and that the usual asymptotic results (Amemiya (1985)) hold. Hence, for finite samples, consistent tests can have nonmonotonic power for the sequence of alternatives of interest.

The outline of the paper is the following. In Section 2 we present sufficient conditions for the power along a ray to approach zero and we prove that such a ray always exist. In Section 3 we illustrate the theoretical results for the power function of the LR test of a zero slope coefficient in a simple logit model. Section 4 analyzes the effect of transforming the regressors and Section 5 concludes the paper.

2. ZERO AND NONMONOTONIC POWER

In this section, we prove under conditions often satisfied in practice that the power goes to zero along a ray in the parameter space. Further, we prove that there always exists a ray along which the power goes to zero.

Consider the following class of binary response models. Let Y_i, i=1,..n denote n independent binary random variables where $Y_i \varepsilon \{0,1\}$. The response function is

$$P_i \equiv P(\ Y_i=1 \mid X_i=x_i\ ) = E(\ Y_i \mid X_i=x_i\ ) = F(x_i'\beta)$$

where X_i is a K-vector of regressors, F is a known distribution function and β is a K-vector of unknown parameters in the parameter space $\Re^K$. For the purpose of exposition, it is convenient to treat X_i as fixed in repeated samples. A sample is denoted by the n×(K+1) matrix [Y,X] whose i'th row is the i'th observation $(Y_i,X_i^{'})$. Since the range of the random variable Y_i contains two values there exists at most 2^n distinct samples, or sample points, in the sample space. The probability of a sample is

$$P(\ Y=y \mid X\ ) = \prod_{i=1}^{n} P_i^{y_i}(1-P_i)^{1-y_i}\ ,\quad P_i = F(\ x_i^{/}\beta\ ) \tag{1}$$

A composite hypothesis restricts parameters of interest. The other parameters are nuisance parameters. Partition the parameter vector into $\beta = (\beta_1,\beta_2)$ where $\beta_1\ \varepsilon\ \Re^k$ are the parameters of interest and $\beta_2\ \varepsilon\ \Re^{K-k}$ are the nuisance parameters and where $0 < k < K$. The composite null hypothesis is

$$H_0\!: \beta_1\ \varepsilon\ B,\ \beta_2\ \varepsilon\ \Re^{K-k}, \tag{2}$$

where B is a bounded subset of $\Re^k$. The composite alternative hypothesis is $H_A\!: \Re^K \backslash H_0$. Hence, β_2 is unrestricted both under the null and alternative hypothesis. The vector of regressors is partitioned conformably with the vector of parameters as $x_i=[x_{i1}^{'}\ x_{i2}^{'}]'$, where x_{i1} is a k vector and x_{i2} is a (K-k) vector.

It is convenient to express the alternative in the following linear form of a scaler j.

$$\beta^{A} = \begin{pmatrix} \beta_1^{A} \\ \beta_2^{A} \end{pmatrix} = \begin{pmatrix} \beta_1^{s} \\ \beta_2^{s} \end{pmatrix} + \begin{pmatrix} \beta_1^{d} \\ \beta_2^{d} \end{pmatrix} j \qquad \textbf{(3)}$$

where β_1^s, β_1^d ε $\Re^k$ and β_2^s, β_2^d ε $\Re^{K-k}$ are all fixed vectors. The alternative (3) as a function of j is a ray in $\Re^K$. If β_1^s ε B and $\beta_1^d \neq 0$, then using the Euclidean metric the alternative moves away from the null hypothesis as j increases in absolute value. The advantage of representing the alternatives by the ray (3) is that the power can be graphed in two-dimensions. Geometrically, this corresponds to a slice of the power surface. Hence, the power function is a function of j given by $\pi(\beta_1^s+\beta_1^d j,\beta_2^s+\beta_2^d j)$.

In order to predict the shape of the power function evaluated along a ray, we derive the limit behavior for $j \rightarrow \infty$. The following assumption insures that in the limit as $j \rightarrow \infty$ the sampling distributions under the null and alternative hypotheses converges to the same limit sample point. Let $\beta^0 = (\beta_1^0,\beta_2^0)$, β_1^0 ε B, β_2^0 ε $\Re^{K-k}$ and, thus, β^0 ε H_0.

ASSUMPTION 1 (Limit sample point): Let Sgn(•) be the sign function. For all i, $x_{i1}'\beta_1^d+x_{i2}'\beta_2^d \neq 0$, $x_{i2}'\beta_2^0 \neq 0$ and $Sgn(x_{i1}'\beta_1^d+x_{i2}'\beta_2^d) = Sgn(x_{i2}'\beta_2^0)$.

With Assumption 1 the limit of $(x_{i1}'\beta_1^d+x_{i2}'\beta_2^d)j$ as $j \rightarrow \infty$ matches the limit of $x_{i2}'\beta_2^0 j$ as $j \rightarrow \infty$, namely, $-\infty$ or ∞. In the proof to the Theorem we show that the limit of the sampling distribution along the ray $(\beta_1^s+\beta_1^d j,\beta_2^s+\beta_2^d j)$ as $j \rightarrow \infty$ degenerates to a single sample point and that the same is true for $(\beta_1^0,\beta_2^0 j)$ ε H_0 and that the two limit sample points are identical.

We now state the result that any non-randomized test with size less than one has power which approaches zero along the ray $(\beta_1^s+\beta_1^d j,\beta_2^s+\beta_2^d j)$.

THEOREM (Zero Power): *Make Assumption 1. In testing the composite null hypothesis (2), then for any non-randomized test with size less than one, the power function evaluated along the ray* $(\beta_1^s+\beta_1^d j,\beta_2^s+\beta_2^d j)$ *goes to zero for any choice of* β_1^s *and* β_2^s:

$$Lim_{j\rightarrow\infty}\ \pi(\beta_1^s+\beta_1^d j,\beta_2^s+\beta_2^d j) = 0.$$

PROOF: First it is shown that the sampling distribution along the ray converges to a single sample point. For each observation i,

$$\text{Lim}_{j\rightarrow\infty} F(\ (x_{i1}'\beta_1^d+x_{i2}'\beta_2^d)j + (x_{i1}'\beta_1^s+x_{i2}'\beta_2^s)\ ) =$$
$$\text{Lim}_{j\rightarrow\infty} F(\ (x_{i1}'\beta_1^d+x_{i2}'\beta_2^d)j\ ) =$$
$$I((x_{i1}'\beta_1^d+x_{i2}'\beta_2^d) > 0)$$

where I is the indicator function. The first equality follows because $(x_{i1}'\beta_1^d+x_{i2}'\beta_2^d) \neq 0$ by assumption and the second equality follows since as $j \rightarrow \infty$, $(x_{i1}'\beta_1^d+x_{i2}'\beta_2^d)j$ converges to either $-\infty$ or $+\infty$ and F is a cdf. Since F is either 0 or 1 in the limit, one sample point has probability 1 of occurring. This sample point is $[y^*,X]$ where $y_i^* = I((x_{i1}'\beta_1^d+x_{i2}'\beta_2^d) > 0)$.

Under the null hypothesis, the sample point $[y^*,X]$ also has probability 1 of occurring. The parameter point $\beta^0 = (\beta_1^0,\beta_2^0 j)$ is in H_0 for all j, if $\beta_1^0 \varepsilon B$. Following the same line of arguments as above,

$$\text{Lim}_{j\rightarrow\infty} F(x_{i1}'\beta_1^0+x_{i2}'\beta_2^0 j) =$$
$$\text{Lim}_{j\rightarrow\infty} F(x_{i2}'\beta_2^0 j) =$$
$$I(x_{i2}'\beta_2^0 > 0)$$

Since $I(x_{i2}'\beta_2^0 > 0) = I((x_{i1}'\beta_1^d+x_{i2}'\beta_2^d) > 0)$ by Assumption 1, the sample point $[y^*,X]$ gets probability 1 of occurring as $j \rightarrow \infty$ for $\beta^0 = (\beta_1^0,\beta_2^0 j)$. Hence, the limit sample point $[y^*,X]$ along the ray representing the alternatives is also the limit sample point for a sequence of parameter values under the null hypothesis.

The limit sample point $[y^*,X]$ must be in the acceptance region of a test if that test is to have size less than 1. Recall, the size of a test is the supremum of the rejection probabilities under the null hypothesis (see e.g. Lehmann (1959)). Since the sample point $[y^*,X]$ has probability 1 of occurring under the null hypothesis, any non-randomized test with $[y^*,X]$ in the rejection region has size 1. Hence, $[y^*,X]$ must be in the acceptance region for the test to have size less than 1.

Along the ray, that is, under the alternative hypothesis, in the limit the probability is 1 of getting the sample point $[y^*,X]$. Since this sample point is in the acceptance region of the test, the test has zero probability of rejecting. Hence, $\text{Lim}_{j\to\infty}\ \pi(\beta_1^s+\beta_1^d j,\beta_2^s+\beta_2^d j) = 0$. Q.E.D.

A practical implication of the Theorem is that if a test has power larger than its size at any alternative along the ray, then the power function is nonmonotonic along that ray. The power along the ray can be large for some alternatives. But eventually the power decreases to zero as the distance from the null hypothesis increases.

The only restrictions on the tests are that they are not randomized and have sizes less than 1. This implies that the Theorem applies to a large class of tests. The class includes the LR, Wald and LM tests. The Theorem does not restrict the distribution function F. Hence, the Theorem can be invoked in logit and probit models.

In general, there always exists rays for which the power function goes to zero. The result is stated in the following corollary.

COROLLARY (Existence of zero power): *In testing the composite hypothesis (2) assume that* $x_{i2} \neq 0_{(K-k)\times 1}$, $\forall i$. *Then for any non-randomized test with size less than one, there exist a ray (3) such that the power along this ray goes to zero.*

PROOF: We construct one possible ray. Let β_2^0 be a linear combination of x_{i2}, i=1,..,n such that $x_{i2}'\beta_2^0 \neq 0$ for all i. For this β_2^0 Assumption 1 is satisfied with $\beta_2^d = \beta_2^0$ and $\beta_1^d = 0$. We are interested in a ray which moves away from the null hypothesis as $j \to \infty$. Hence, instead of $\beta_1^d = 0$, select a β_1^d such that $|x_{i1}'\beta_1^d| < |x_{i2}'\beta_2^d|\ \forall i$. This is possible since $x_{i2}'\beta_2^d \neq 0$. Selecting β_1^d this way implies that $\text{Sgn}(x_{i1}'\beta_1^d+x_{i2}'\beta_2^d) = \text{Sgn}(x_{i1}'\beta_2^d)$. Therefore Assumption 1 is also satisfied with the new $\beta_1^d \neq 0$ and, thus, the Theorem applies for this ray. Q.E.D.

The consequence is that in testing composite hypotheses in binary response models, the problem of nonmonotonic power will always be present. Whether the slice of the power surface has a nonmonotonic shape depends on the alternatives of interest chosen by the researcher.

The alternatives of interest $\beta^A = (\beta_1^A,\beta_2^A)$ are typically different values of the parameters of interest β_1^A with the values of the nuisance parameters β_2^A being held fixed. The different alternatives of interest can be specified using (3) where the parameters β_1^s, β_2^s, β_1^d and β_2^d are chosen in the following way: $\beta_1^s \varepsilon B$, β_2^s plausible values of the nuisance parameters, $\beta_1^d \neq 0$ and β_2^d equal to the 0 vector. The choice of β_2^d equal to 0 is a consequence of the nuisance parameters being held fixed. When β_2^d is zero, the ray is orthogonal to the space spanned by the nuisance parameters. For this reason, we call it an orthogonal ray. Whether or not there exists an orthogonal ray for which the power goes to zero depends on the regressor matrix.

The hypothesis that one slope coefficient is zero is often tested in empirical studies. In this case, there is only one orthogonal ray, namely, different values of the slope coefficient. If all the values of the corresponding regressor have matching signs with the values of one of the other regressors, then the Theorem applies and the power function goes to zero along the orthogonal ray. One such case is when the regressor is positive and there is an intercept in the model. This important case is studied in detail in the next section.

3. TEST OF ONE SLOPE COEFFICIENT

In this section we consider the hypothesis that one slope coefficient is zero. The interesting alternatives are different values of the slope coefficient with the remaining parameters held fixed. To illustrate the cases when the Theorem applies and when it does not, it is sufficient to consider a model with one regressor and an intercept. We consider a simple logit model and the LR test. The power of the LR test is nonmonotonic in the slope coefficient when the Theorem applies and monotonic when it does not apply.

The simple binary logit model with an intercept and one regressor x_{i1} is

$$P(Y_i = 1) = \Lambda(x_{i1}\beta_1 + \beta_2) = \frac{1}{1+\exp(-(x_{i1}\beta_1+\beta_2))}$$

where Λ is the logistic distribution function, β_1 is the slope and β_2 is the intercept. Consider a 0.05 size LR test of the composite null hypothesis H_0: $\beta_1=0$ against the alternative H_A: $\beta_1\neq0$. The interesting alternatives are values of $\beta_1 \neq 0$ and the nuisance parameter β_2 being held fixed. Hence, the corresponding orthogonal ray is $(\beta_1^d j, \beta_2^s)$ where β_1^d can be set to 1.

First, we examine the power function of the LR test for a sample size n = 50 when the regressor x_{i1} equals the perfect uniform [2,4], $x_i^p = 2+2(i-1)/(n-1)$, i = 1,..,n and hence is positive. Figure 1 shows the empirical power function obtained by evaluating the power at 147 different parameter points (β_1,β_2). The empirical power at each point is calculated using 5000 Monte Carlo repetitions. We used the asymptotic critical value given by the χ^2 distribution for the 0.05 significance level, namely 3.841, since this critical value is very close to the size-corrected critical value.

Figure 1 shows that for any choice of the nuisance parameter β_2, the power function $\pi(\beta_1,\beta_2)$ is nonmonotonic in β_1. Further, the power function is sensitive to the choice of β_2. For a high value of β_2, the power is low at any value of β_1 whereas for a low value of β_2, the power can be high for a range of values of β_1. Typically, the sequence of alternatives of interest is increasing values of the slope β_1 with the nuisance parameter β_2 being held fixed. Figure 2 shows that the power for such a sequence with β_2 = -1 is nonmonotonic. Even at the peak, the power is below 0.4.

The nonmonotonicity of the empirical power function $\pi(\beta_1,\beta_2)$ with β_2 held fixed is predictable using the Theorem. The alternatives of interest are $(\beta_1^d j, \beta_2^s)$ where $\beta_1^d = 1$. Assumption 1 is satisfied for $\beta_2^0 = 1$ because for all i, $\beta_1^d x_i^p > 0$ and $\beta_2^0 1 > 0$. The particular choice of β_2^0 is not important, merely that one can be found. Hence, the Theorem applies and the power along the orthogonal ray $(\beta_1^d j, \beta_2^s)$ goes to zero.

The application of the Theorem does not depend on the logistic distribution function nor does it depend on the LR test. In fact, it is enough that the regressor takes on only positive values and similarly for the constant. This implies that the result holds for other tests, for example, the LM and Wald tests and also for the uniformly most powerful test against one-sided alternatives in a logit model (Cox (1970)). The same is true for other distribution functions, for example, the probit model. Finally, the result also holds in the case where more regressors and nuisance parameters are added.

Second, we examine the power function of the LR test for a sample size n = 50 when the regressor takes on both positive and negative values. Assume the regressor x_{i1} equals the perfect uniform [-0.5,0.5], $x_i^b = -0.5 + (i-1)/(n-1)$, i = 1,..,n. Figure 3 shows the empirical power function obtained by evaluating the power at 147 different parameter points (β_1,β_2). As in the first example, the empirical power at each point is calculated using 5000 Monte Carlo repetitions. We used the critical value 3.841 since it also is very close to the size-corrected critical value in this example.

In this example, the empirical power function $\pi(\beta_1,\beta_2)$ with β_2 held fixed is monotonic in β_1. The value of the nuisance parameter β_2 again influences the power for any value of β_1. For a sufficiently large value of β_2, the power can be low even for a large value of β_1. The sequence of alternatives of interest usually corresponds to a slice in the power surface along an orthogonal ray. The power along the orthogonal ray with β_2 = -1 is shown in Figure 4. The power along the ray is monotonic and approaches 1 as β_1 increases.

The Theorem does not apply to any orthogonal ray. For instance, in the second example, setting $\beta_1^d = 1$ and $\beta_2^d = 0$, it is not possible to find a β_2^0 which satisfies Assumption 1. Since the regressor takes on both positive and negative values, $\beta_1^d x_i^b$ is negative for some i and positive for others. But $\beta_2^0 1$ is either negative or positive for all i. The Corollary guarantees the existence of a ray for which the power goes to zero. For this example, the ray with $\beta_1^d = 1$ and $\beta_2^d = 2$ satisfies Assumption 1 when β_2^0 is set to 1. In fact, there are infinitely many rays for which the power goes to zero, but none of them are orthogonal.

4. TRANSFORMATION OF THE REGRESSORS

The two examples in Section 3 suggest that demeaning the regressor is a solution to the problem of nonmonotonicity along an orthogonal ray. In this section we analyze the connection between the power function with a positive regressor and the power function when the regressor is demeaned. We show that the two power functions are transformations of each other. Then we explain the nature of the solution to the nonmonotonicity problem provided by demeaning.

First we obtain the power function of a test when the regressors are replaced with an affine transformation. Demeaning or standardization are typical affine transformations used in practice. Assume that one of the regressors is a constant. Accordingly, the parameter vector and regressor matrix are partitioned as $\beta = (\delta',\alpha)'$ and $x_i = (z_i',1)'$ where α is the intercept and δ and z_i are the remaining (K-1) parameters and (K-1) regressors. The affine transformation of the regressors is

$$z_i^* = a + Dz_i, \tag{4}$$

where z_i^* are the (K-1) transformed regressors, a is a (K-1) vector and D is a (K-1)×(K-1) nonsingular matrix.

The distribution function F with the original regressors relates to the distribution function with the transformed regressors in the following way.

$$F(z_i'\delta + \alpha) = F((D^{-1}(z_i^*-a))'\delta + \alpha) = F(z_i^{*\prime}[D^{-1\prime}\delta] + [-a'D^{-1\prime}\delta+\alpha])$$

Hence, there is a linear transformation of the parameters that neutralizes the affine transformation of the regressors, namely,

$$\beta^* = \begin{pmatrix} \delta^* \\ \alpha^* \end{pmatrix} = \begin{bmatrix} D^{-1\prime} & 0 \\ -a'D^{-1\prime} & 1 \end{bmatrix} \begin{pmatrix} \delta \\ \alpha \end{pmatrix} \tag{5}$$

This implies that the power of the test $\pi^*(\delta^*,\alpha^*)$ with the transformed regressors equals the power $\pi(\delta,\alpha)$ with the original regressors provided that δ^* and α^* are given by (5). Thus, $\pi^*(\delta^*,\alpha^*)$ is obtained from $\pi(\delta,\alpha)$ by the linear coordinate transformation (5).

One commonly used transformation of the regressors is demeaning. Consider a model with one slope coefficient δ and an intercept α. The composite null hypothesis is $\delta = 0$ against $\delta \neq 0$ and α unrestricted. The interesting alternatives are different values of δ with the nuisance parameter α held fixed; that is, the orthogonal ray $(\delta^d j,\alpha^s)$. Let the original regressor z_i take on only positive values. The demeaned regressor z_i^* is given by the affine transformation (4) where $a = -\bar{z}$ $(\neq 0)$, $\bar{z}$ is the mean of z_i and D = 1.

The power along the orthogonal ray with the original regressor is not the same as the power along an orthogonal ray with the demeaned regressor. Following (5), the power along the orthogonal ray $(\delta^d j,\alpha^s)$ with the original regressor corresponds to the power along the ray $(\delta^d j, \alpha^s + z\overline{\delta^d} j)$ with the demeaned regressor; that is,

$$\pi(\delta^d j, \alpha^s) = \pi^*(\delta^d j, \alpha^s + z\overline{\delta^d} j),$$

where π^* is the power function with the demeaned regressor. The ray $(\delta^d j, \alpha^s + z\overline{\delta^d} j)$ is not an orthogonal ray; the nuisance parameter is not being held fixed. Hence, $\pi(\delta^d j, \alpha^s)$ is not equal to $\pi^*(\delta^d j, \alpha^s)$ unless the test statistic is pivotal in the nuisance parameter, that is, its distribution does not depend on the nuisance parameter (see e.g. Hall (1992)).

The two empirical power functions of the LR test, shown in Figure 1 and 3, are transformations of each other. This is the case because the regressor used in the second example, x_i^b, is the regressor from the first example, x_i^p, demeaned and standardized to be in a unit interval; that is, the affine transformation (4) with a = $-\frac{1}{2}\overline{x^p}$ and D = ½, where $\overline{x^p}$ is the mean of x_i^p. In both examples, β_1 (=δ) is the slope and β_2 (=α) is the intercept. The composite null hypothesis is $\beta_1 = 0$ against $\beta_1 \neq 0$ with β_2 unrestricted. In the first example, the alternatives of interest are represented by the orthogonal ray $(\beta_1^d j, \beta_2^s)$. The power along this orthogonal ray differs in the two examples, as can be seen in Figure 2 and 4. This is not surprising since the alternatives are not comparable. Using (5), however, the alternatives can be made comparable. The orthogonal ray in the first example is equivalent to the non-orthogonal ray $(2\beta_1^d j, (\beta_2^s + 2\overline{x^p}\beta_1^d j))$ in the second example.

5. CONCLUSION

When testing a composite hypothesis in a binary response model, any non-randomized test with size less than 1 has a power function which goes to zero along a ray in the parameter space. This typically implies that the power function of a test is nonmonotonic along that ray; that is, the power along the ray is larger than the size for some alternatives but eventually the power decreases to zero along the ray.

In many applications in economics, the regressors take on only positive values. In a model with an intercept, a test of a zero slope coefficient has power which goes to zero for the alternatives of interest if the corresponding regressor is positive.

REFERENCES

Amemiya, T. (1981): "Qualitative Response Models: A Survey," *Journal of Economic Literature*, Vol XIX, 1483-1536.

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

Cox, D.R. (1970): *The Analysis of Binary Data*. Chapman and Hall, London.

Cytel (1993): *LogXact User Manual*. Cytel Software Corporation, Cambridge, MA.

Domencich, T. A. and D. McFadden (1975): *Urban Travel Demand*. North-Holland, Amsterdam.

Finney, D. J. (1977): *Probit Analysis*. Cambridge University Press, London.

Hall, P. (1992): *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, New York.

Killingsworth, M. and J. Heckman (1986): "Female Labor Supply: A Survey," in *Handbook of Labor Economics, Volume I,* ed. by O. Ashenfelter and R. Layard. North-Holland, Amsterdam.

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

Nelson, F.D. and Savin, N.E. (1988): "The Nonmonotonicity of the Power Function of the Wald Test in Nonlinear Models," *Working Paper Series No. 88-7, Department of Economics, University of Iowa*.

Nelson, F.D. and Savin, N.E. (1990): "The Danger of Extrapolating Asymptotic Local Power," *Econometrica*, 58, 977-981.

Robertson, J. L. and H. K. Preisler (1992): *Pesticide Bioassays with Arthropods.* CRC Press, Inc., Boca Raton.

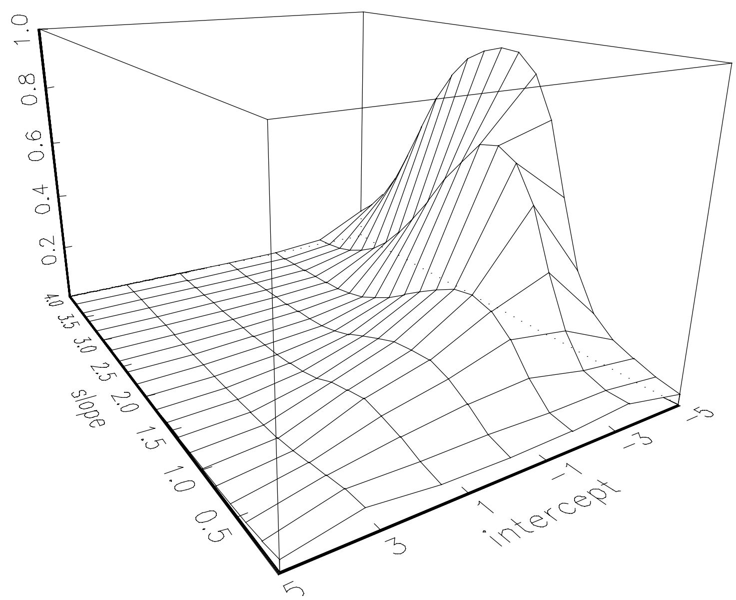

Figure 1. The power function of the LR test in a logit model with a perfect uniform regressor [2,4], n = 50.

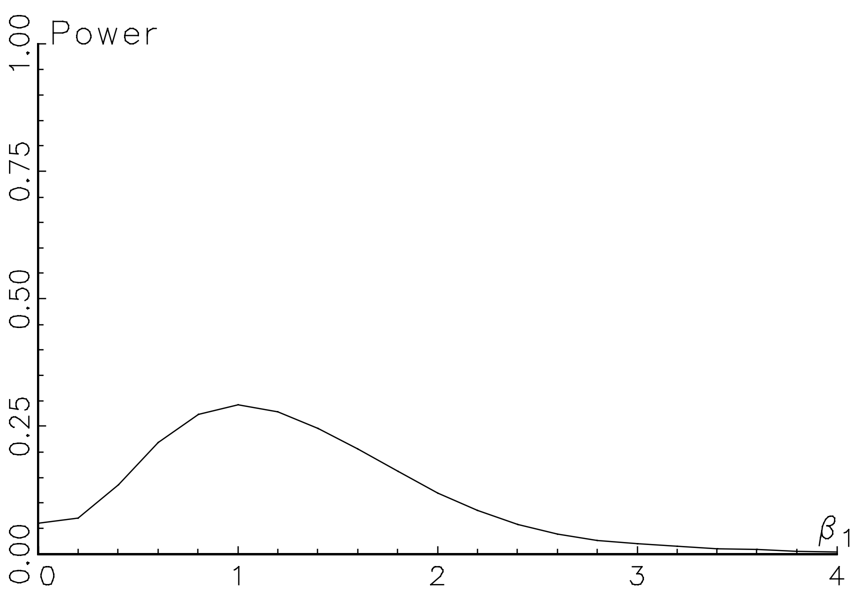

Figure 2. The power function of the LR test with the nuisance parameter β_2 = -1 in a logit model with a perfect uniform regressor [2,4], n = 50.

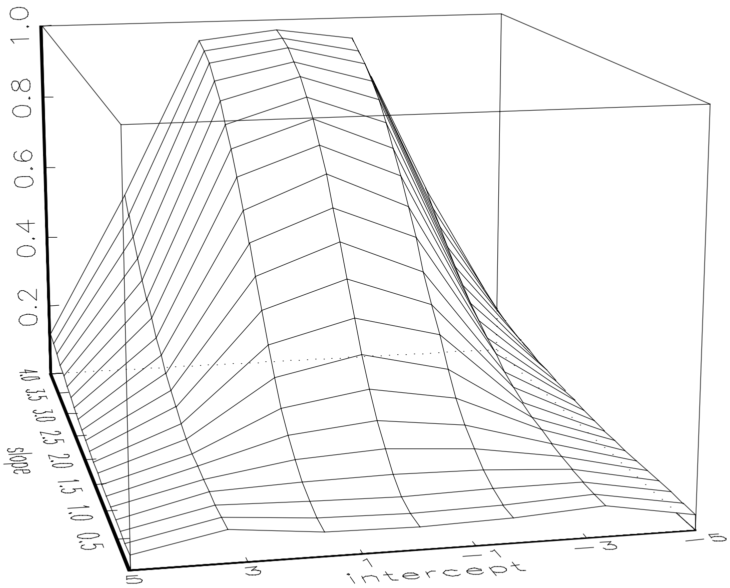

Figure 3. The power function of the LR test in a logit model with a perfect uniform regressor [-0.5,0.5], n = 50.

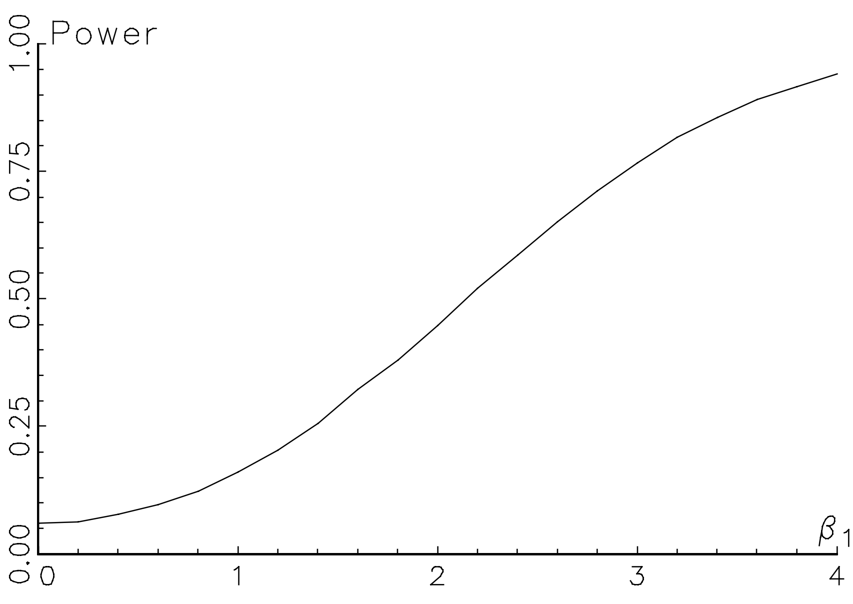

Figure 4. The power function of the LR test with the nuisance parameter β_2 = -1 in a logit model with a perfect uniform regressor [-0.5,0.5], n = 50.