THE EFFECT OF NUISANCE PARAMETERS ON THE POWER OF LM TESTS IN LOGIT AND PROBIT MODELS

By N. E. Savin and Allan Würtz[1]

Department of Economics
108 Pappajohn Bus. Adm. Bldg.
University of Iowa
Iowa City, IA 52242.

May 26, 1996

In econometrics, most null hypotheses are composite, dividing the parameters into parameters of interest and nuisance parameters. The domain of the nuisance parameters can influence the size-corrected critical value and hence the power of a test. We show that the domain of the nuisance parameters determines which version of the LM test to use in logit and probit models. In these models there are two commonly used ways to construct the LM test: it can be based on the Hessian matrix or the outer product (OP) matrix of the score vectors. For the OP based LM test, the domain of the nuisance parameters strongly influences the size-corrected critical value whereas the same is not true for the Hessian based LM test. A theoretical explanation is developed using large nuisance parameter asymptotics. For empirically relevant domains, the experimental evidence shows that the Hessian based LM test has better finite sample power than the OP based LM test.

Keywords: Composite hypothesis, finite sample power, Hessian information matrix, Lagrange multiplier test, logit model, nuisance parameter, outer product information matrix

---

[1] We gratefully acknowledge the comments of Joel Horowitz, Narayana Kocherlakota, Forrest Nelson and George Neumann.

1. INTRODUCTION

In logit and probit models the hypotheses tested usually involve only a subset of the parameters, in particular, one or more slope coefficients, the others being nuisance parameters. When testing such composite hypotheses, the size-corrected critical value is potentially sensitive to the domain of the nuisance parameters and hence so is the power. As a consequence, when comparing the powers of competing tests, the domain of the nuisance parameters can play an important role in choosing a test. We will show that is the case for the LM test. As for other nonlinear models, the LM test is attractive since only the constrained maximum likelihood estimates need to be calculated. There are essentially two different versions of the LM test. Thus, the question is which version is best where by best we mean the test which has the most power.

The LM tests differ in the method used to estimate the information matrix. There are three commonly used estimators of the information matrix. The first is based on the expectation of the Hessian matrix; the second is based on the Hessian matrix and the third is based on the outer product (OP) matrix of the score vectors. The three estimators produce three variants of the LM test: the expected Hessian LM test, the observed Hessian LM test and the OP LM test. A special property of the logit model is that the expected and observed Hessian tests are the same. Although the three tests have different invariance properties (Dagenais and Dufour (1991)), they are asymptotically equivalent (Amemiya (1985)). In finite samples, however, the tests can have very different powers against the same alternative.

In this paper, we compare the powers of the Hessian and OP LM tests for testing the composite null hypothesis that an individual slope coefficient is zero, the other parameters being the nuisance parameters. The results for the probit and logit models are very similar and are most easily illustrated using a simple logit model, that is, a logit model with an intercept and a slope where the intercept is the nuisance parameter. The conclusions are not qualitatively changed by adding additional regressors.

The nuisance parameters can affect the power of the tests in two ways. The first is the actual values of the nuisance parameters can influence the distributions of the test statistics. Our experiments show that the power of the OP LM test is especially sensitive to the specific values of the nuisance parameters and, at best, has about the same power as the Hessian LM test.

The second is the domain of the nuisance parameters can influence the size-corrected critical values of the tests and hence the critical regions. Our experiments show that the size-corrected critical value of the OP LM test is sensitive to the domain of the nuisance parameters. For a given alternative, the larger the domain of the nuisance parameters, the smaller the power of the OP LM test. For a sufficiently small domain, the powers of the two tests can be about the same. For other empirically relevant domains, the Hessian LM test is superior.

When the domain of the nuisance parameter is large, the large sample asymptotic distribution is a poor approximation to the finite sample distribution in the case of the OP LM statistic. A better approximation can be obtained using large nuisance parameter asymptotics. In large nuisance parameter asymptotics, a sequence of test statistics is indexed by the values of the nuisance parameters, assuming that the null is true and the sample size is fixed. The purpose is to obtain the limiting distribution of such a sequence as the values of the nuisance parameters increase in absolute value. For the simple logit model, this approach is equivalent to finding the limiting distribution of a sequence indexed by the intercept. This limiting distribution provides a critical value which is close to the Monte Carlo size-corrected critical value when the domain of the intercept is large. The critical value of the OP LM test based on this limiting distribution is shown to be many orders of magnitude larger than that of the Hessian LM test.

The size-corrected critical value is calculated under the assumption that only the domain of the nuisance parameters is known. If the true values of the nuisance parameters are known, then the level-corrected critical value can be calculated. Treating the values of the nuisance parameters as known, the composite hypothesis is converted into a simple hypothesis. The level-corrected critical value is the critical value used to test the simple hypothesis. In the Monte Carlo experiment the true values of the nuisance parameters are known and hence the level-corrected critical value can, in fact, be computed. Of course, these critical values are not in general appropriate for testing a composite hypothesis.

The power of a test based on the level-corrected critical value can differ substantially from the power based on the size-corrected critical value and hence which test is best can depend on which critical value is used. We illustrate this point using Davidson and MacKinnon (1984). They compared the LM tests using level-corrected critical values and concluded that the Hessian and OP LM tests have about the same power. For a logit model with two regressors in addition to the intercept, we obtain results which are similar to those of Davidson and MacKinnon (1984) when the tests are based on level-corrected critical values calculated using their values for the nuisance parameters. For these same values of the nuisance parameters, we find that the Hessian LM test is better than the OP LM test when the power comparisons are based on size-corrected critical values. The power comparison based on the size-corrected critical values is the empirically relevant comparison when the hypothesis is composite.

The organization of the paper is the following. The model and the Hessian and OP LM test are presented in Section 2. The Monte Carlo design of the experiments is described in Section 3. The empirical results on the size and powers of the tests are reported in Section 4. The large nuisance parameter asymptotics is developed in Section 5. Section 6 reports experimental results on level- and size-corrected powers using a model motivated by Davidson and MacKinnon. The concluding comments are in Section 7.

2. HESSIAN AND OP LM TESTS

In this section we describe the Hessian and OP LM test statistics for a binary logit model. The binary logit model is defined by

$$P(Y_i\!=\!1) \;=\; \Lambda(\beta' x_i) \;=\; \frac{1}{1+EXP\{-\beta' x_i\}}$$

where $\{Y_i\}$ is a sequence of independent binary random variables taking the value 0 or 1, x_i is a K-vector of known constants, β_0 is the true K-vector of unknown parameters and Λ is the logistic distribution function.

The log likelihood function for the sample is

$$L(\beta;y_1,\ldots,y_n) = \sum_{i=1}^{n} L_i = \sum_{i=1}^{n} y_i \log\Lambda_i + (1-y_i)\log(1-\Lambda_i)$$

and the score vector for the i-th observation is

$$S_i(\beta) = y_i(1-\Lambda_i)-(1-y_i)\Lambda_i x_i = (y_i-\Lambda_i)x_i.$$

The score vector for all observations is

$$S(\beta) = \frac{\partial L}{\partial \beta} = X^{\prime}(y-\Lambda),$$

where y = $(y_1,\ldots,y_n)$', X = $(x_1,\ldots x_n)$' and Λ = $(\Lambda_1,\ldots\Lambda_n)$'.

Consider testing the composite null hypothesis that the parameter vector β satisfies q < K restrictions of the form

H_0: h(β) = 0.

In this case, the dimension of the vector of nuisance parameters is K-q. The LM test, that is, the score test, of H_0 is based on the test statistic

$$LM = [S(\tilde{\beta})]^{\prime}\ [n\tilde{I}]^{-1}\ [S(\tilde{\beta})],$$

where S(β) is the score vector evaluated at the constrained maximum likelihood (ML) estimate of β, β, and $\tilde{I}$ is an estimator of the limiting information matrix evaluated at the constrained ML estimate β. The matrix $[n\tilde{I}]^{-1}$ is an estimator of the covariance matrix of β when H_0 is true.

Two versions of the LM statistic are obtained by using different consistent estimators of the limiting information matrix. The limiting information matrix evaluated at β_0 is given by

$$I(\beta_0) = lim_{n\to\infty}\left(-n^{-1}E[\frac{\partial^2 L(\beta)}{\partial\beta\partial\beta^{\prime}}]\right)\Bigg|_{\beta=\beta_0} = \lim_{n\to\infty}\left(n^{-1}E\sum_{i=1}^{n}[\frac{\partial L_i(\beta)}{\partial\beta}][\frac{\partial L_i(\beta)}{\partial\beta^{\prime}}]\right)\Bigg|_{\beta=\beta_0}$$

The Hessian based estimator of the limiting information matrix *I*(β_0) is

$$I_{OH}(\tilde{\beta}) = -n^{-1}\frac{\partial^2 L(\beta)}{\partial\beta\partial\beta'}\bigg|_{\beta=\tilde{\beta}} = n^{-1}\sum_{i=1}^{n}\Lambda_i(1-\Lambda_i)x_i x_i'\bigg|_{\beta=\tilde{\beta}} = n^{-1}X'[diag(\Lambda_i(1-\Lambda_i))]X\Big|_{\beta=\tilde{\beta}}$$

Since the Hessian matrix evaluated at β_0 is a matrix of constants, the Hessian with and without the expectation operator is the same, that is, the expected Hessian is equal to the observed Hessian. The asymptotic covariance matrix of β can be approximated by the Hessian based estimator:

$$E(\tilde{\beta}-\beta_0)(\tilde{\beta}-\beta_0)' \cong \left[\frac{\partial^2 L(\beta)}{\partial\beta\partial\beta'}\bigg|_{\beta=\tilde{\beta}}\right]^{-1} = [nI_{OH}(\tilde{\beta})]^{-1}$$

For details see Amemiya (1985).

The OP estimator of the information matrix is

$$I_{OP}(\tilde{\beta}) = n^{-1}\sum_{i=1}^{n}[\frac{\partial L_i(\beta)}{\partial\beta}][\frac{\partial L_i(\beta)}{\partial\beta'}]\bigg|_{\beta=\tilde{\beta}} = n^{-1}\sum_{i=1}^{n}(y_i-\Lambda_i)^2 x_i x_i'$$

In this case, the asymptotic covariance matrix of β is approximated by

$$E(\tilde{\beta}-\beta_0)(\tilde{\beta}-\beta_0)' = \left[\sum_{i=1}^{n}[\frac{\partial L_i(\beta)}{\partial\beta}][\frac{\partial L_i(\beta)}{\partial\beta'}]\bigg|_{\beta=\tilde{\beta}}\right]^{-1} = [nI_{OP}(\tilde{\beta})]^{-1}$$

The two versions of the LM test are asymptotically equivalent. Using first-order asymptotics, the two versions of the LM tests have the same limiting chi-square distribution when H_0 is true and the same limiting noncentral chi-square distribution under sequences of local alternatives. As a result, the asymptotic critical values are the same for the two tests and the tests have the same local power.

3. MONTE CARLO DESIGN

In this section, we describe the design of the Monte Carlo experiments used to investigate the size and power of the LM tests in a simple logit model.

The simple logit model is defined by

$$P(y_i=1) = \Lambda(\beta_1+\beta_2 x_i) = \frac{1}{1+e^{-\beta_1-\beta_2 x_i}}, \quad i = 1,..,n \tag{1}$$

Using this model, we test the composite null hypothesis H_0: $\beta_2 = 0$, the nuisance parameter being the intercept β_1. The alternative hypothesis is $\beta_2 \neq 0$. Only results for the 0.05 size tests are reported since this is the nominal size commonly used in practice. The results are qualitatively the same for other sizes.

In the Monte Carlo experiments, the values of the regressor x for a sample size n were generated using the perfect normal N(m, s^2):

$$x_i = m + s\Phi^{-1}(i/(n+1)), \; i = 1,2,...,n,$$

where Φ is the standard normal cdf. We report results for the perfect standard normal: m = 0,s = 1. We also conducted experiments using a perfect uniform distribution. The results for the perfect uniform were qualitatively the same as for the perfect standard normal. The experiments were conducted for various sample sizes. We report results for n = 100 and, in some cases, for n =200.

To calculate the Hessian and OP LM statistics for testing H_0, only the constrained ML estimate of β_1 is needed, namely, the estimate of β_1 subject to the constraint $\beta_2 = 0$. The ML estimate is, however, not finite for some samples, that is, the ML estimator is not defined for certain points in the sample space; see Albert and Anderson (1984), or, for a brief discussion , Amemiya (1985). We call these points “bad” points. For the constrained ML estimate of β_1 there are only two bad points; one is y = (0,0,...,0)’ and the other is y = (1,1,...,1)’. For finite n, these bad points have positive probability of occurring. If a bad point occurs, it is deleted and not replaced.

In the experiments, N samples were generated for selected parameter points (β_1, β_2). The number of nondeleted samples G equals N minus the number of times a bad point occurs. The estimate of the rejection probability at the parameter point (β_1, β_2) is R/G where R is the number of times the test rejects in the G samples. In other words, R/G is an estimate of the power of the test at the point (β_1, β_2).

In the experiments we use N = 5000 replications. With N = 5000 the Monte Carlo error for the estimate of the power is small. From the binomial model, the standard error of the estimator R/N is 0.00308 if the rejection probability is 0.05 for each of the N samples. Hence, the rejection probabilities are estimated with high precision if the number of times a bad point occurs is small. If the number of times a bad point occurs is as large as 10% of 5000, then G = 4,500 and the standard error of R/G is 0.00316, which is still small.

All calculations were performed using version 3.01 of Gauss 386i (Aptech 1992).

4. EMPIRICAL SIZES AND POWERS

In this section we obtain the size-corrected critical values for the Hessian and OP LM tests and calculate the empirical size-corrected powers of the tests.

We first investigate the sizes of the tests using the asymptotic 0.05 critical value. When H_0 is true, the asymptotic distribution of the test statistics is chi-square with one degree of freedom. Hence, the 0.05 asymptotic critical value is 3.84. It is well known that OP LM tests can suffer from substantial size distortions (Davidson and MacKinnon (1993, p. 477)).

In the case of a composite hypothesis, the size of a test is the supremum of the rejection probabilities over the domain of the nuisance parameters (Hogg and Craig (1978, p. 239) and Lehmann (1959 p. 61)). Figure 1 shows the empirical rejection probabilities of the tests when H_0 is true for values of the nuisance parameter, the intercept β_1 , which range from -6 to 6. The Hessian LM test for n = 100 has essentially the correct size when it is based on the asymptotic critical value: the empirical rejection probability is about 0.05 for all values of the intercept. On the other hand, the empirical rejection probability of the OP LM test is sensitive to the value of the intercept. For n = 100, the empirical rejection probability is roughly 0.05 when the absolute value of β_1 is less than 2 and then increases as the absolute value of β_1 increases. For example, the empirical rejection probability is about 0.20 when the absolute value of β_1 is 3. For n = 200, the empirical rejection probability is roughly about

0.05 when the absolute value of β_1 is less than 3 and then increases as the absolute value of β_1 increases. Hence, the OP LM test based on the asymptotic critical value can suffer from very large size distortions.

The size-corrected critical value is a function of the domain of the nuisance parameters. We define the domain for β_1 by $\beta_1 \in [-\beta_{1,max}, \beta_{1,max}]$ where $\beta_{1,max}$ is a positive number. Figure 2 shows the 0.05 size-corrected critical values for both the Hessian and OP LM test for different domains of β_1. For all choices for $\beta_{1,max}$, the asymptotic critical value is approximately the size-corrected critical value for the Hessian LM test. For the OP LM test, however, the size-corrected critical value increases sharply as $\beta_{1,max}$ increases beyond 2. For example, it approaches 17.9 as the absolute value of $\beta_{1,max}$ approaches 3. Hence, the 0.05 size-corrected critical value for the OP LM test is very sensitive to $\beta_{1,max}$. From Figure 2 it is clear that we need to have some idea about the empirically relevant values of β_1 when calculating size-corrected critical values.

In practice, bad points are uncommon. This suggests that it is reasonable to conclude that values of β_1 which produces a noticeable percentage of bad points are not empirically relevant. Thus, we calculated the percent of bad points for each value of β_1. In Table 1 we report the percentage of good points for selected positive values of β_1 when $\beta_2 = 0$ and n = 100. The results show that the percentage of bad points is 0.007 for $\beta_1 = 3$ and increases as the absolute value of β_1 increases. For the purpose of our experiments, we conclude that β_1 values less than or equal to 3 in absolute value are empirically relevant.

We next investigate the powers of the 0.05 size-corrected LM tests. The first case we consider is where the domain of β_1 is $\beta_{1,max} = 2$. For this case with n = 100, the size-corrected critical value is 4.10 for the Hessian LM test and 5.17 for the OP LM Test. Figure 3(a) shows the empirical powers when $\beta_1 = 0$ and Figure 3(b) when $\beta_1 = 2$. In Figures 3(a) and 3(b), the empirical powers are essentially the same for the Hessian and OP LM tests. Notice that in Figure 3(b), the OP LM test has empirical power 0.05 when $\beta_2 = 0$. This is because $\beta_1 = 2$ is on the boundary of the domain for β_1 and the rejection probability is largest for $\beta_1 = 2$.

The second case is where the domain of β_1 is $\beta_{1,max} = 3$. For this case, the size-corrected critical value is 4.10 for the Hessian LM test and 17.9 for the OP LM test. Figures

4(a), 4(b) and 4(c) show the empirical powers when $\beta_1 = 0$, 2, and 3, respectively. In these figures, the empirical powers of the size-corrected OP LM tests are substantially smaller than those of the Hessian LM tests. Also notice that in Figure 4(c) the OP LM test has empirical power 0.05 when $\beta_2 = 3$. Again, this is because $\beta_1 = 3$ is on the boundary of the domain for β_1 and the rejection probability is increasing in β_1.

It is misleading to interpret the lower power of the OP LM test in Figures 4(a) and 4(b) as due to size distortion. The test has the correct size by construction. What Figures 4(a) and 4(b) show is that the level is less than the size when $\beta_1 = 0$ and 2 where the level is the rejection probability when the null is true for a specific value of the nuisance parameter. By contrast, the test has level equal to size in Figure 4(c). Even in this case, the power of the OP LM test is substantially less than the power of the Hessian LM test. In fact, the power of the OP LM test is less than that of the Hessian LM test in all the experiments even though the powers are close when $\beta_{1,max}$ is small.

5. LARGE NUISANCE PARAMETER ASYMPTOTICS

An interesting question is why are the size-corrected critical values for the OP LM test so large in the simple logit model. The answer is that when H_0 is true the distribution of the OP LM test statistic shifts to the right as the value of nuisance parameter, β_1, increases. As β_1 increases with n fixed, consider the sequence of distributions of the Hessian and OP LM test statistics. We call the limit of such a sequence the large nuisance parameter asymptotic distribution. In the simple logit model this is equivalent to a large β_1 asymptotic distribution. For a large domain of the intercept the large β_1 asymptotic distribution provides a better approximation to the size-corrected critical value than the large sample asymptotic distribution.

Before giving the formulas for the large β_1 asymptotic distributions of the test statistics, we consider the treatment of the two bad points, the points at which the ML estimates and hence the statistics are not defined. The probabilities of the two bad points are

$$Pr(y=(0,0,..,0)) = \left(\frac{e^{-\beta_1}}{1+e^{-\beta_1}}\right)^n \text{ and } Pr(y=(1,1,..,1)') = \left(\frac{1}{1+e^{-\beta_1}}\right)^n$$

where n is the sample size. Since the estimate of β_1 is not finite at the two bad points, we delete these points from the sample space and adjust the probabilities of the remaining points so that the probabilities in the reduced sample space sum to one.

In the reduced and renormed sample space, the probability concentrates on n points as β_1 increases with n fixed. Let $y^{(i)}$ be an n-vector where the i-th element is 0 and the rest are 1: $y^{(i)} = (1,1,..., 0,...1)'$. The n points of concentration are $y^{(i)}$, i = 1,...,n, and in the limit the probability of each point is 1/n. This is stated formally in the following

LEMMA: *Assume model* (1) *with* $\beta_2 = 0$. *Then, for fixed n, as* $\beta_1 \to \infty$, $\Pr(y = y^{(i)}) \to 1/n$.

The proof is given in the Appendix.

The large β_1 asymptotic distributions of the test statistics are constructed using the n points $y^{(i)}$. The large β_1 asymptotic distributions are given by the following.

PROPOSITION: *Assume model* (1) *and that the null hypothesis* H_0: $\beta_2 = 0$ *is true. Let*

$$LM_{OH,j} = \frac{n}{n-1}\frac{(x_j-\bar{x})^2}{s_x^2}, \quad j = 1,2,...,n$$

$$LM_{OP,j} = \frac{(x_j-\bar{x})^2}{\frac{1}{n}s_x^2+\frac{n-1}{n^2}(\bar{x}-x_j)^2}, \quad j = 1,2,...,n.$$

The sample mean and standard deviation of the regressor are denoted by $\bar{x}$ *and* s_x. *Then, for fixed* n, *as* $\beta_1 \to \infty$, $\Pr(LM_{OH} = LM_{OH,j}) \to 1/n$ *for all* j *and similarly* $\Pr(LM_{OP} = LM_{OP,j}) \to 1/n$.

The proof of the Proposition is given in the Appendix.

We now show that the large β_1 asymptotic critical value is a good approximation to the size-corrected critical value of the OP LM test when $\beta_{1,max}$ is large. For our problem, a large value of $\beta_{1,max}$ is 6. From Table 1 we see that for this value of the intercept the percentage of bad points is 77.9%. Figure 5 displays the critical values of the 0.05 size OP LM test when n = 100 and n = 200 for values of $\beta_{1,max}$ from 0 to 6. The horizontal lines represent the large β_1 asymptotic critical values. The large β_1 asymptotic critical value is 80.0 for n =100 and 161 for n = 200 The figure shows that the size-corrected critical values approach the large β_1 asymptotic critical value as $\beta_{1,max}$ approaches 6. In Monte Carlo experiments the size-corrected critical value is larger the larger the sample size. This is predicted by large β_1 asymptotics.

If, for a moment, we assume that $x_1,...,x_n$ is a random sample from a normal distribution, then the large β_1 asymptotic distributions of the Hessian and OP LM test statistics are functions of the F statistic. In the case of the Hessian LM test, the large β_1 asymptotic distribution is $[n/(n-1)]F(1,n-1)$ where $F(1,n-1)$ is an F distribution with 1 and n-1 degrees of freedom. For n = 100, the large β_1 asymptotic critical value is 3.98, which is close to the value in our Monte Carlo experiment, namely, 4.10. Notice that as $n \rightarrow \infty$, $[n/(n-1)]F(1,n-1) \rightarrow \chi^2(1)$, which is the large sample asymptotic distribution of the Hessian LM test statistic when the null hypothesis is true. In the case of the OP LM test, the derivation of the large β_1 asymptotic critical value is a bit more complicated. The large β_1 asymptotic critical value was found to be 80.4 for n = 100 and 160 for n = 200. These are very close to 80 and 158, respectively, which are the values obtained in our Monte Carlo experiments for a large $\beta_{1,max}$. Hence, for this case the large β_1 asymptotic distribution provides a good approximation to the size-corrected critical value compared to the large sample asymptotic distribution.

In practice, the domain of the nuisance parameters is implicitly assumed to be a very large set. This implies that the large β_1 asymptotic critical value is close to the size-corrected critical value. Figure 6 shows the size-corrected powers of the Hessian and OP LM tests when $\beta_{1,max}$ is 6, $\beta_1 = 0$ and n = 100. In this case, the power function of the OP LM test is a horizontal line at 0; that is, the OP LM test never rejects H_0.

6. LEVEL- VERSUS SIZE-CORRECTED POWER

In Monte Carlo experiments the critical values are often level-corrected instead of size-corrected. In this section we report the level- and size-corrected powers for testing that a single slope is zero in a logit model with three regressors. This design is motivated by Davidson and MacKinnon (1984). They compare the level-corrected powers of the Hessian and OP LM tests and conclude that the tests have about the same powers.

The logit model with three regressors is defined by

$$P(y_i=1) = \Lambda(\beta_1+\beta_2 x_{2i}+\beta_3 x_{3i}) = \frac{1}{1+e^{-\beta_1-\beta_2 x_{2i}-\beta_3 x_{3i}}}, \quad i = 1,..,n$$

Using this model, we conducted 0.05 size tests of the composite null hypothesis H_0: $\beta_2 = 0$, the nuisance parameters being the intercept β_1 and the slope β_3. The alternative hypothesis is $\beta_2 \neq 0$.

In the Monte Carlo experiments, the values of the regressors x_2 and x_3 for a sample size n are given by $x_{2i} = x_i$ and $x_{3i} = (-1)^i x_i$ for $i \leq 50$ and $x_{3i} = (-1)^{i+1} x_i$ for $i > 50$ where x_i is generated using the perfect normal $N(m, s^2)$:

$$x_i = m + s\Phi^{-1}(i/(n+1)), \; i = 1,2,...,n,$$

and where Φ is the standard normal cdf. We report results for the perfect standard normal: $m = 0$, $s = 1$ for $n = 100$. This implies that x_2 and x_3 are perfect standard normals with a correlation equal to -0.07. The use of the essentially independent standard normal regressors is motivated by the experimental design in Davidson and MacKinnon (1984).

Davidson and MacKinnon (1984) calculated level-corrected critical values assuming that the intercept and all slopes were zero except for one slope, which was set equal to 3 or 6. For the parameter values used in our experiments the probability of the bad points is negligible. For the constrained ML estimators of β_1 and β_3 there are 101 bad points and the probabilities of these points is less 0.01.

We investigated the sizes of the tests using the asymptotic 0.05 critical value.

When H_0 is true, the asymptotic distribution of the test statistics is chi-square with one degree of freedom. We calculated the rejection probabilities of the tests when the intercept β_1 takes on the value 0 or 3 and β_3 ranges from 0 to 6. Figure 7 shows the results for n = 100. The Hessian LM test has essentially the correct size when it is based on the asymptotic critical value: the empirical rejection probability is about 0.05 for all values of β_1 and β_3. On the other hand, the empirical rejection probability of the OP LM test is sensitive to the values of β_1 and β_3. As β_3 increases from 0 to 6, the empirical rejection probability increases monotonically from 0.05 to about 0.25 when $\beta_1 = 0$ and is a nonlinear function of β_3 when $\beta_1 = 3$. Again we find that the OP LM test based on asymptotic critical values can suffer from very large size distortions.

The finite sample size-corrected critical value of the OP LM test is very sensitive to the domain of the nuisance parameters. We define the domain of the nuisance parameters as $\beta_1 \in [0, \beta_{1,max}]$ and $\beta_3 \in [0, \beta_{3,max}]$. Figure 8 shows the 0.05 size-corrected critical values for both the Hessian and OP LM tests for different domains of β_1 and β_3, namely, $\beta_{1,max} = 1$ and 3 and $\beta_{3,max} \in [0, 6]$. The critical value for the Hessian LM test is close to 3.84 for the different choices for $\beta_{1,max}$ and $\beta_{3,max}$. In the case of the OP LM test, as $\beta_{3,max}$ goes from 0 to 6, the critical value increases from about 3.84 to about 15 when $\beta_{1,max} = 1$ and is roughly constant at about 12 when $\beta_{1,max}$ is less than 3. Note that including $\beta_1 \in [-\beta_{1,max}, 0]$ and $\beta_3 \in [-\beta_{3,max}, 0]$ in the domain does not significantly change the size-corrected critical values.

The differences between level- and size-corrected powers are shown in Figure 9. Figure 9 (a) shows that the level corrected powers of the Hessian and OP LM tests are about the same when $\beta_1 = 0$ and $\beta_3 = 3$, which are values used by Davidson and MacKinnon. By contrast, Figure 9(b) shows that the size-corrected power of the Hessian LM test is superior when the domain of the nuisance parameters is $\beta_{1,max} = 3$ and $\beta_{3,max} = 6$, again assuming that $\beta_1 = 0$ and $\beta_3 = 3$.

Figure 10 illustrates that the power of the OP LM test is sensitive to the domain of the nuisance parameters. In Figures 10 (a) and 10(b) the values of the nuisance parameters are the same, $\beta_1 = 0$ and $\beta_3 = 0$, but the domains of the nuisance parameters are different. In Figure 10(a) the domain is given by $\beta_{1,max} = 1$ and $\beta_{3,max} = 1$ whereas in Figure

10 (b) the domain is larger with $\beta_{1,\max} = 1$ and $\beta_{3,\max} = 6$. With the small domain the powers of the two tests are about the same whereas with the large domain the Hessian LM test has substantially better power.

The inferiority of the OP LM test is not simply due to the use of size-corrected critical values. This is illustrated by Figure 11 where the tests are based on level-corrected critical values assuming $\beta_1 = 3$ and $\beta_3 = 0$. In this case the power of the OP LM test is inferior, even though the influence of the domain of the nuisance parameters has been eliminated.

The results found in this section are consistent with the results obtained with the simple logit model; the Hessian LM test is well behaved and has superior power.

7. CONCLUDING COMMENTS

It is well known for nonlinear models that the power function depends on the value of the nuisance parameters. What is generally overlooked is that the size-corrected critical value of a test may also depend on the domain of the nuisance parameters. This dependence is crucial for the finite sample power performance of the OP LM test. The larger domain of the nuisance parameters, the larger the size-corrected critical value for the OP LM test. This has a negative impact on the power of the OP LM test.

The powers of the tests can be compared using level-corrected or size-corrected critical values. When the comparison is based on level-corrected critical values, the OP and Hessian LM tests can have about the same power for certain values of the nuisance parameters. But for other values the Hessian LM test is better. Turning to comparisons based on size-corrected critical values, we find that the tests have about the same power for certain small domains of the nuisance parameters. For other domains, however, the power of the Hessian LM test is superior.

APPENDIX. PROOF OF LEMMA AND PROPOSITION

The proof of the Lemma is the following. Without loss of generality consider y = $y^{(1)}$ = (0,1,1..,1)'. In the reduced and renormed sample space the probability of y is denoted Q(y)

$$Q(y) = \frac{P(y)}{1 - P(bad\ points)} = \frac{\dfrac{e^{-\beta_0}}{1+e^{-\beta_0}}\left(\dfrac{1}{1+e^{-\beta_0}}\right)^{n-1}}{1 - \left(\dfrac{1}{1+e^{-\beta_0}}\right)^{n} - \left(\dfrac{e^{-\beta_0}}{1+e^{-\beta_0}}\right)^{n}}$$

Rearranging the terms gives

$$Q(y) = \frac{a}{(1+a)^n - (1+a^n)} \;, \quad a = e^{-\beta_0}$$

Expand the denominator to get

$$Q(y) = \frac{a}{\sum_{i=0}^{n}\binom{n}{i}1^i a^{n-i} - (1+a^n)} = \frac{1}{n + \sum_{i=1}^{n-2} a^{n-i-1}}$$

When β_1 is large, a is close to zero. In other words, as $\beta_1 \to \infty$, Q(y) $\to$ 1/n. A similar result can be obtained when the negative of β_1 is large. In that case, all the probability over the reduced and renormed sample space concentrates on the n sample points having one element equal to one and the remaining equal to zero. Q.E.D.

The proof of the Proposition is established by calculating the value of the LM statistics at the sample points $y^{(i)}$, i=1,2,...,n.. Without loss of generality again suppose y = $y^{(1)}$ = (0,1,1..,1)'.

The constrained ML estimate is $\beta_1 = \Lambda^{-1}(\Sigma^{n}_{i=1}/n)$ or $\Lambda(\beta_1) = (n-1)/n$. The score vector is

$$S(\{\tilde{\beta}_0,0\}) = \begin{pmatrix} 0 \\ \sum (y_i - \Lambda(\tilde{\beta}_0))x_i \end{pmatrix} = \begin{pmatrix} 0 \\ \bar{x} - x_1 \end{pmatrix}, \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

The Hessian information matrix is

$$I_{OH}(\{\tilde{\beta}_0,0\}) = \sum_{i=1}^{n} \Lambda(\tilde{\beta}_0)(1-\Lambda(\tilde{\beta}_0)) \begin{bmatrix} 1 & x_i \\ x_i & x_i^2 \end{bmatrix} = \frac{n-1}{n} \begin{bmatrix} 1 & \bar{x} \\ \bar{x} & \frac{1}{n}\sum_{i=1}^{n} x_i^2 \end{bmatrix}$$

The covariance matrix is the inverse of the information matrix:

$$V_{OH}(\{\tilde{\beta}_0,0\}) = \frac{n}{n-1}\frac{1}{V(x)} \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} x_i^2 & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix}, \quad V(x) = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$$

The OP information matrix has the following form:

$$I_{op}(\{\tilde{\beta}_0,0\}) = \sum_{i=1}^{n} (y_i - \Lambda(\tilde{\beta}_0))^2 \begin{bmatrix} 1 & x_i \\ x_i & x_i^2 \end{bmatrix} = \frac{n-1}{n} \begin{bmatrix} 1 & x_1 \\ x_1 & x_1^2 \end{bmatrix} + \frac{1}{n} \begin{bmatrix} 1 & \bar{x} \\ \bar{x} & \frac{1}{n}\sum_{i=1}^{n} x_i^2 \end{bmatrix} =$$

$$\begin{bmatrix} 1 & \frac{n-1}{n}x_1 + \frac{1}{n}\bar{x} \\ \frac{n-1}{n}x_1 + \frac{1}{n}\bar{x} & \frac{1}{n}(V(x)+\bar{x}^2) + \frac{n-1}{n}x_1^2 \end{bmatrix}$$

This implies that the OP covariance matrix is

$$V_{op}(\{\tilde{\beta}_0,0\}) = \frac{1}{D} \begin{bmatrix} \frac{1}{n}(V(x)+\bar{x}^2) + \frac{n-1}{n}x_1^2 & -\frac{n-1}{n}x_1 - \frac{1}{n}\bar{x} \\ -\frac{n-1}{n}x_1 - \frac{1}{n}\bar{x} & 1 \end{bmatrix}$$

$$D = \frac{1}{n}s_x^2 + \frac{n-1}{n^2}\bar{x}^2 + \frac{n-1}{n^2}x_1^2 - \frac{n-1}{n^2}2x_1\bar{x} = \frac{1}{n}s_x^2 + \frac{n-1}{n^2}(\bar{x}-x_1)^2$$

Because the first element of the score vector is zero, only the (2,2) element of the covariance matrix is used to calculate the value of the LM test statistic.

The other n-1 values of the large β_1 asymptotic limiting distribution of the LM test can be obtained by replacing x_1 with x_i for the corresponding sample $y^{(i)}$. Q.E.D.

Table 1. Percent of Bad Points

Percent of Bad Points	
Intercept β_1	Percent
0.0	0.000
2.0	0.000
3.0	0.007
3.5	0.053
6.0	0.779

REFERENCES

Albert, A and J.Anderson (1984): “On the Existence of Maximum Likelihood Estimates in Logistic Regression Models, “ *Biometrika*, 71, 1-10.

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

Dagenais, M. and J.M. Dufour (1991): “Invariance, Nonlinear Models, and Asymptotic Tests, “ *Econometrica*, 50, 1601-1615.

Davidson R., and J. G. MacKinnon.(1984): “Convenient Specification for Logit and Probit Models,” *The Journal of Econometrics* 25, 241-262.

Davidson R., and J. G. MacKinnon.(1993)*: Estimation and Inference in Econometrics*. Oxford University Press, New York.

Hogg, R.V. and A.T. Craig.(1978): *Introduction to Mathematical Statistics*. Macmillian Publishing Co., New York.

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

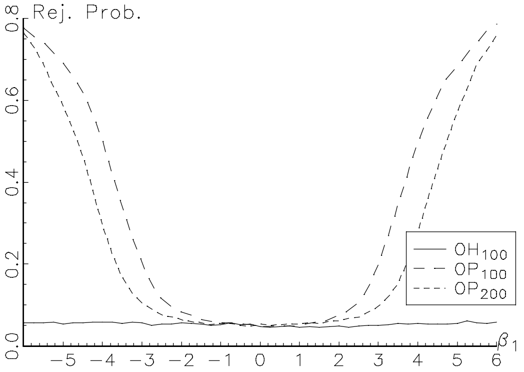


FIGURE 1. — Rejection probabilities for two-sided symmetric LM test with the 0.05 asymptotic critical value with β_2 equal to 0.

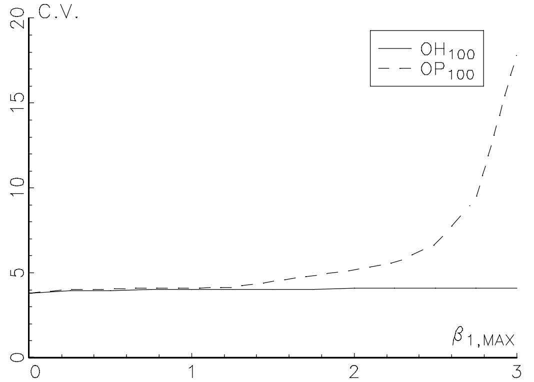
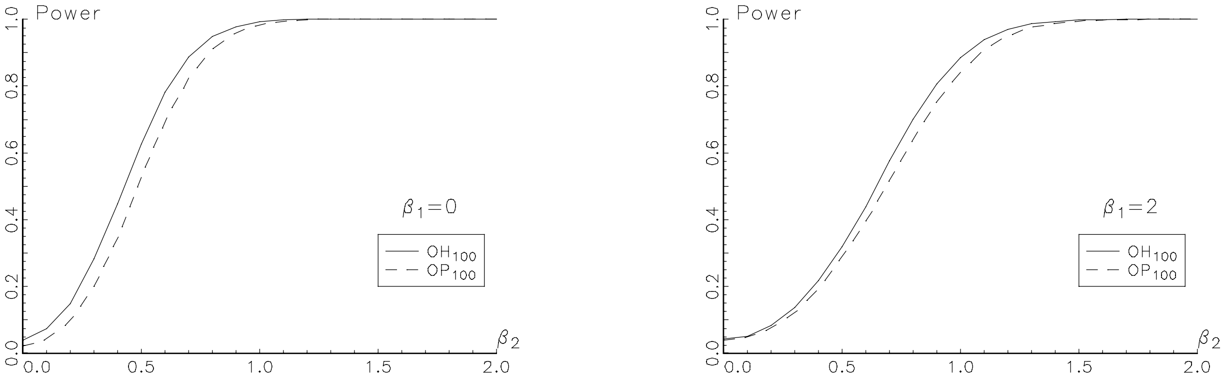


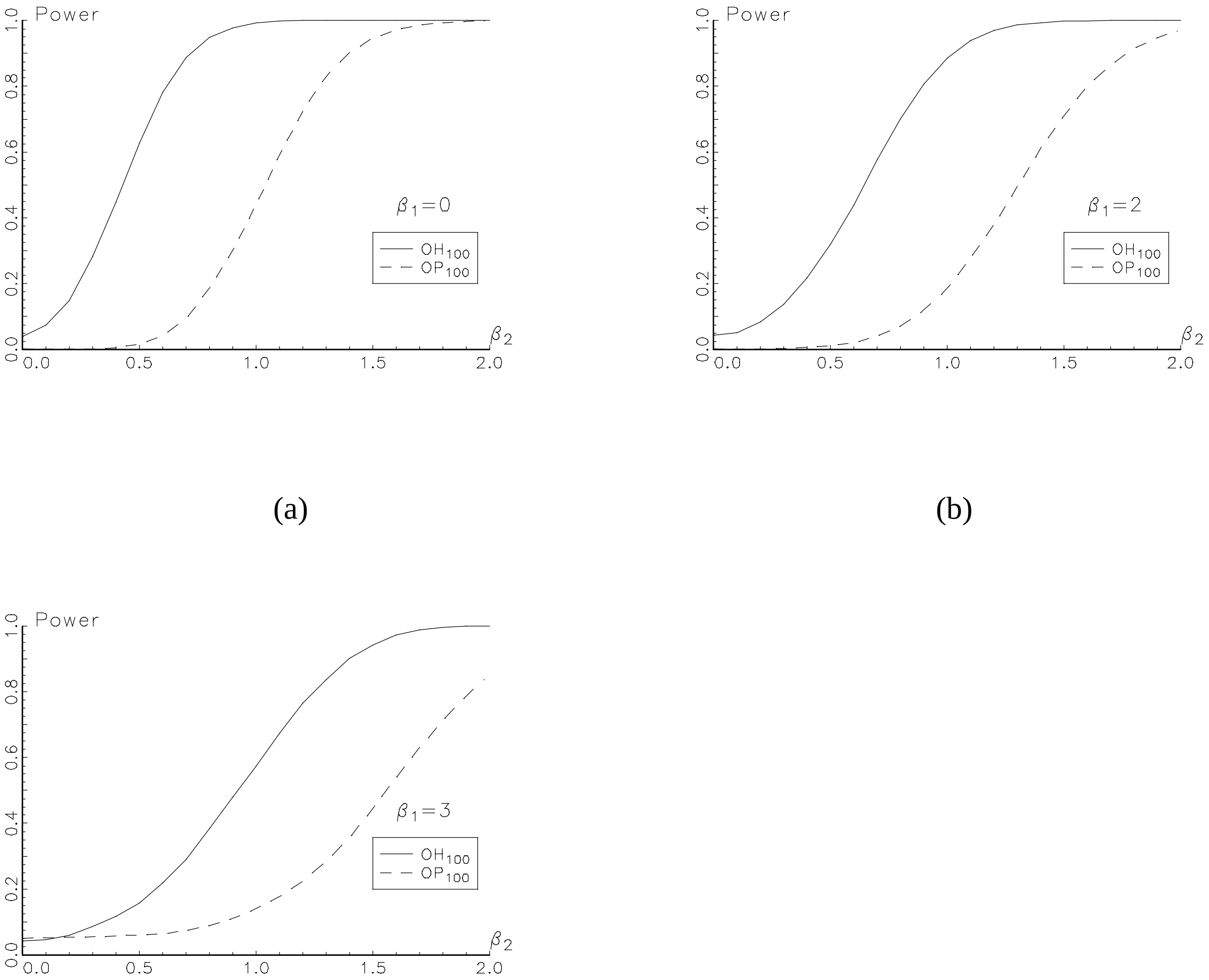
FIGURE 2. — Size-corrected critical values for 0.05 two-sided symmetric LM tests with β_2 equal to 0.



(a)

(b)

FIGURE 3. — Size-corrected powers of 0.05 two-sided symmetric LM tests with $\beta_{1,MAX}$ equal to 2.



(a)

(b)

(c)

FIGURE 4. — Size-corrected powers of 0.05 two-sided symmetric LM tests with $\beta_{1,MAX}$ equal to 3.

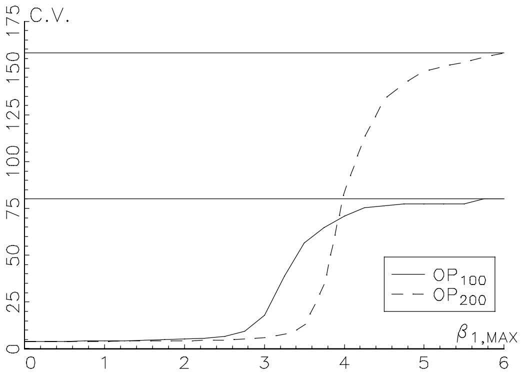


FIGURE 5. — Size-corrected critical values for 0.05 two-sided symmetric LM tests with β_2 equal to 0.

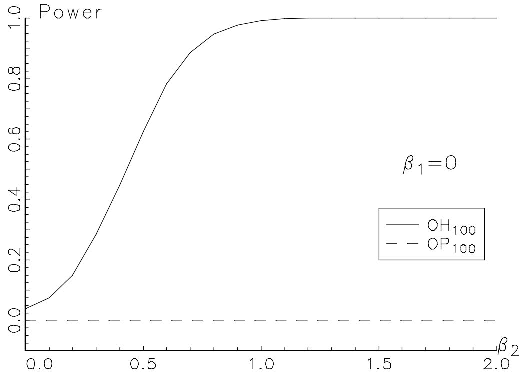


FIGURE 6. — Size-corrected powers of 0.05 two-sided symmetric LM tests with $\beta_{1,\mathrm{MAX}}$ equal to 6.

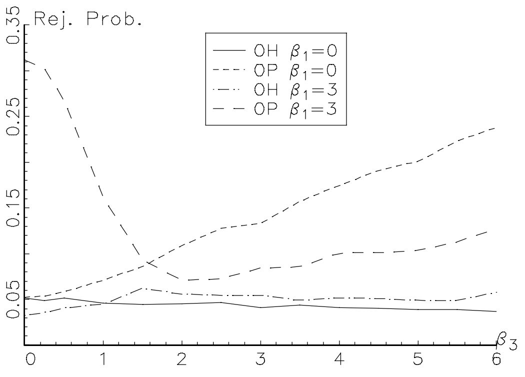


FIGURE 7. — Rejection probabilities for two-sided symmetric LM test with the 0.05 asymptotic critical value in the three regressor logit model.

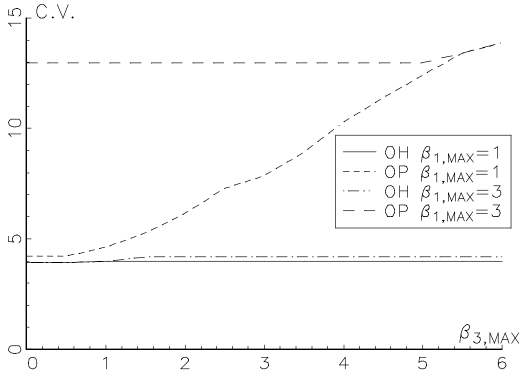
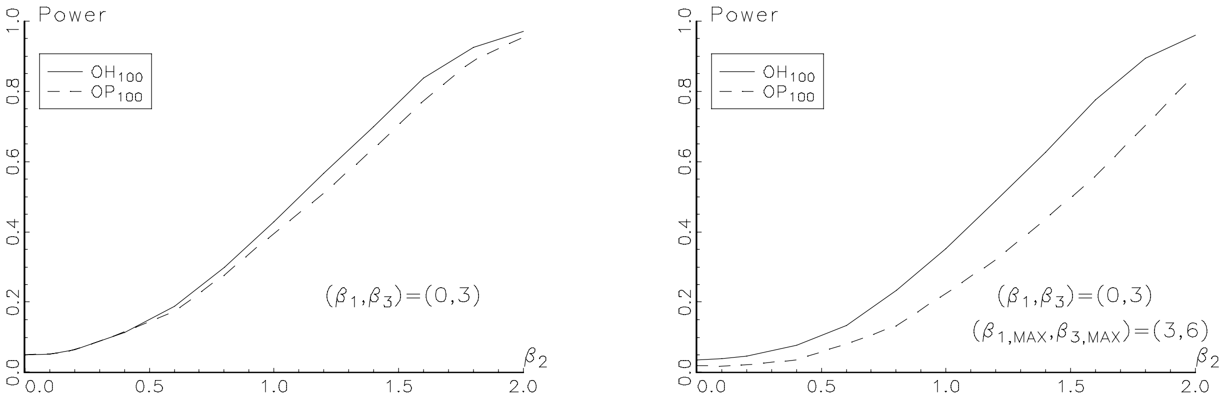
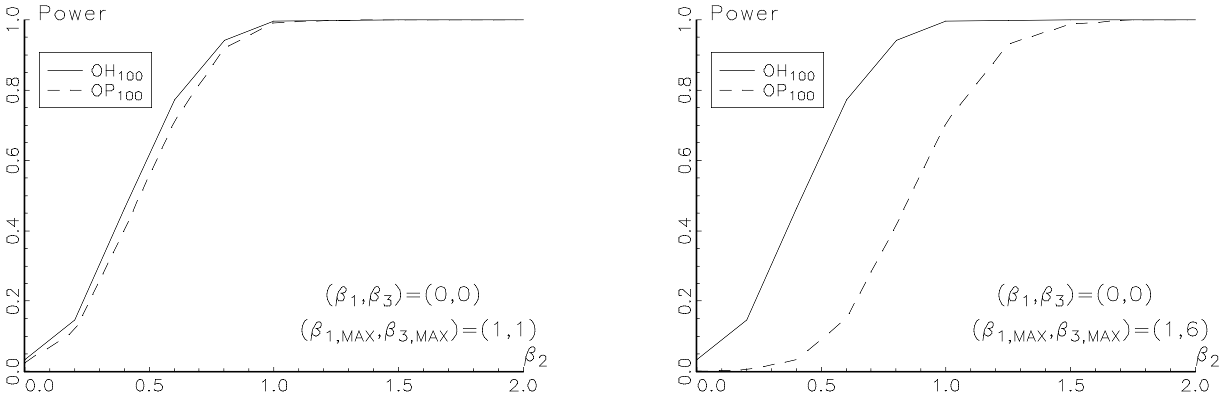


FIGURE 8. — Size-corrected critical values for 0.05 two-sided symmetric LM tests in the three regressor logit model.



(a) (b)

FIGURE 9. — Powers of 0.05 two-sided symmetric LM tests with critical value obtained by (a) level-correction and (b) size-correction.



(a) (b)

FIGURE 10. — Size-corrected powers of 0.05 two-sided symmetric LM tests.

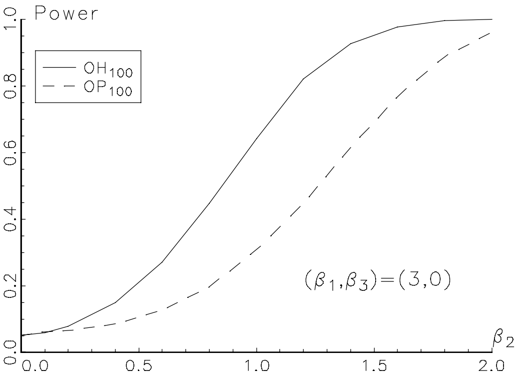


FIGURE 11. — Level-corrected powers of 0.05 two-sided symmetric LM tests.