DEPARTMENT OF ECONOMICS

Working Paper

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

N.E. Savin
Allan H. Würtz

Working Paper No. 1997-17
Centre for Non-linear Modelling in Economics

ISSN 1396-2426

UNIVERSITY OF AARHUS • DENMARK

CENTRE FOR NON-LINEAR MODELLING IN ECONOMICS
DEPARTMENT OF ECONOMICS - UNIVERSITY OF AARHUS - DK - 8000 AARHUS C - DENMARK
☎ +45 89 42 11 33 - TELEFAX +45 86 13 63 34

WORKING PAPER

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

N.E. Savin
Allan H. Würtz

Working Paper No. 1997-17

DEPARTMENT OF ECONOMICS
SCHOOL OF ECONOMICS AND MANAGEMENT - UNIVERSITY OF AARHUS - BUILDING 350
8000 AARHUS C - DENMARK ☎ +45 89 42 11 33 - TELEFAX +45 86 13 63 34

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

N. E. Savin and Allan H. Würtz[1]

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

and

Centre for Non-Linear Modelling in Economics
Department of Economics
Bygn. 350, Universitetsparken
University of Aarhus,
DK-8000 Aarhus C
Denmark

September 29, 1997

In econometrics, most null hypotheses are composite, dividing the parameters into parameters of interest and nuisance parameters. Typically, a composite hypothesis can be tested using two or more testing procedures. Competing testing procedures are commonly compared using size-corrected powers. What is often overlooked is that the size-corrected critical value of a test can be sensitive to the set of admissible values of the nuisance parameters, and hence its size-corrected power. As a result, different choices for the admissible set can produce different conclusions about which test is best. This fact complicates the interpretation of Monte Carlo power studies because in many cases there is no natural definition of the set of admissible values. We find this fact to be crucial when choosing a Lagrange Multiplier test in the case of a logit model. A theoretical explanation for this effect is developed using large parameter asymptotics.

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

JEL : C12, C15, C25.

---

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

1. INTRODUCTION

In logit models, the hypotheses tested usually only involve a subset of the parameters, in particular, one or more slope coefficients, the others being nuisance parameters. Typically, such a composite hypothesis can be tested using two or more testing procedures. Competing testing procedures are commonly compared using size-corrected powers. What is often overlooked is that the size-corrected critical value of a test can be sensitive to the set of admissible values of the nuisance parameters, and hence its size-corrected power. In Monte Carlo studies designed to compare size-corrected powers, this set is chosen by the experimenter. Different choices can produce different conclusions about which test is best. Thus, the choice for the set of admissible values of the nuisance parameters can play a key role in size-corrected power comparisons. This fact complicates the interpretation of experimental power studies because in many cases there is no natural definition of the set of admissible values.

This paper illustrates the influence of the admissible set using variants of the Lagrange Multiplier (LM) test. The variants of the LM test differ in the estimator of the information matrix. One variant is based on the expectation of the Hessian matrix, a second on the Hessian matrix and a third 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. Dagenais and Dufour (1991) show that the expected Hessian and OP LM tests enjoy certain invariance properties to reparameterization, while the observed Hessian does not. For this reason we only consider these two variants of the LM test. Although the tests are asymptotically equivalent (Amemiya (1985)), Monte Carlo evidence suggests that the performance of the tests can be very different in finite samples. The OP LM test has been reported to have poor size performance in logit and probit models, see Davidson and MacKinnon (1984). For other tests using the outer product of the score vectors, see, for example Chesher and Spady (1991), Davidson and MacKinnon (1983, 1985, 1992), Godfrey, McAleer and McKenzie (1988) and Orme (1990).

In this paper, large parameter asymptotics is used to show that the size of the OP LM test is strongly influenced by the nuisance parameters. The large parameter asymptotic distribution is derived from a sequence of test statistics indexed by the values of the parameters, assuming the sample size is fixed. The purpose is to obtain the limiting distribution of such a sequence as the values of the parameters increase in absolute value. The critical value obtained from this limiting distribution is a lower bound on the size-corrected critical value of the LM statistics. For the OP LM test these critical values are many orders of magnitude larger than that of the Hessian LM test and the asymptotic critical value from the chi-square distribution.

For the logit model, Davidson and MacKinnon (1984) concluded that despite the poor size performance of the OP LM test, the size-corrected power of the OP LM test is good compared to the expected Hessian LM test. We find, however, that this conclusion is a result of the treatment of the set of admissible values of the nuisance parameters; in particular, the set contains only one point (Savin and Würtz (1996)). This treatment of the set of admissible values of the nuisance parameters is not uncommon in Monte Carlo experiments with tests of composite hypotheses. In effect, the null hypothesis is more restricted than claimed, and hence the tests are actually size-corrected. A similar point is emphasized, along with other examples, in Campbell and Dufour (1995), Dufour (1994) and Dufour and Torres (1995). Our Monte Carlo simulations demonstrate that the OP LM test can have about the same power as the Hessian LM test if the set of admissible values of the nuisance parameters is small. If the set is enlarged, then the Hessian LM test has better size-corrected power. The poor power performance of the OP LM test is partly due to its poor size-performance, but not entirely. Hence, when choosing a LM test in logit models, the expected Hessian LM test is recommended.

The organization of the paper is the following. The model and the expected Hessian and OP LM tests are presented in Section 2. The size properties of the LM tests are investigated in Section 3. In Section 4 the large parameter asymptotic distributions are derived. The implications for the size-corrected powers are analyzed in Section 5 and Section 6 concludes the paper.

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

$$P(Y_i=1) = F(\beta' x_i) = \frac{1}{1+e^{-\beta' x_i}}$$

where $\{Y_i\}$ is a sequence of independent binary random variables taking the value 1 or 0, β is a k-vector of parameters and x_i is the i'th k-vector of regressors from a matrix of regressors X with n observations. The hypothesis is assumed to be composite, restricting only a subset of the parameters. Partition the parameter vector β into β_1 and β_2 such that β_1 is the k_1 vector of restricted parameters and the k_2-vector β_2 the remaining parameters. The null hypothesis is

(1) $H_0 = \{\beta=(\beta_1,\beta_2) \mid \beta_1 \in B_1^0 \subset \mathbb{R}^{k_1},\ \beta_2 \in B_N \subset \mathbb{R}^{k_2}\}$

and the alternative hypothesis

(2) $H_A = \{\beta=(\beta_1,\beta_2) \mid \beta_1 \in B_1^A \subset \mathbb{R}^{k_1},\ \beta_2 \in B_N \subset \mathbb{R}^{k_2}\}$,

where B_1^0 is a bounded subset disjoint with B_1^A. The null hypothesis does not restrict the parameters β_2 and, hence, β_2 are nuisance parameters.

The LM test statistics use the score vector and a covariance matrix. Let S be the score vector and V the covariance matrix. Then the LM test statistic is

$$LM = S(\beta)'\ V(\beta)\ S(\beta),$$

where both S and V are evaluated at parameter values satisfying the null hypothesis. Two versions of the LM test statistics are obtained by using different consistent estimators of the covariance matrix. One version, the expected Hessian LM test is calculated using the inverse of the expected information matrix

(3) $V_{EX}(\beta) = \left(\sum_{i=1}^{n} F(\beta' x_i)(1-F(\beta' x_i))x_i x_i'\right)^{-1}$

Note, in a logit model the Hessian matrix equals the expected Hessian matrix. The other version, the OP LM test, is calculated using the inverse of the outer product of the score vectors

(4) $V_{OP}(\beta) = \left(\sum_{i=1}^{n} (y_i - F(\beta' x_i))^2 x_i x_i' \right)^{-1}$

The score vector, used in both versions of the LM test, is given by

(5) $S(\beta) = \sum_{i=1}^{n} (y_i - F(\beta' x_i)) x_i$

The two versions are denoted LM_{EX} and LM_{OP} for the expected Hessian LM test and the OP LM test, respectively.

3. SIZES OF THE TESTS

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 non-central chi-square distribution under sequences of local alternatives. As a result, the asymptotic critical values are the same for the two tests, and they have the same local power. In finite samples, however, the tests are not pivotal: their finite sample distributions depend on the true values of the parameters under the null hypothesis and, in particular, the values of the nuisance parameters. Before considering the power of the tests in section 5, we investigate the size properties.

In our Monte Carlo experiments we find, that the actual size of the OP LM test can differ substantially from the nominal size, which is consistent with the findings of other investigators. These results hold in both simple and more complicated logit and probit models. Therefore, to illustrate the effects of the nuisance parameters, it suffices to consider a simple binary logit model.

In a simple logit model with a slope β_1 and an intercept β_2, consider a test of the null hypothesis that the slope is zero. The values of the corresponding regressor x were generated using the perfect normal N(0, 1): $x_i = \Phi^{-1}(i/(n+1))$, i = 1,2,...,n, where Φ is the standard normal cdf. The Monte Carlo sample size is 5000. To calculate the Hessian and OP LM statistics, only the constrained ML estimate of β_2 subject to the constraint $\beta_1 = 0$ is needed. The ML estimate is, however, not finite for some samples. In other words, 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 sample points “bad” points.

For the constrained ML estimator 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. Hence, the estimate of the rejection probability at the parameter point (β_1, β_2) is R/G where R is the number of rejections in the G non-deleted samples.

Using the 0.95 quantile of the chi-square one distribution as the critical value, Figure 1 shows the empirical rejection probabilities of the tests when H_0 is true for values of the nuisance parameter, that is, the intercept β_2, which range from -6 to 6. For the Hessian LM test with n = 100, 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_2 is less than 2 and then increases as the absolute value of β_2 increases. For example, the empirical rejection probability is about 0.20 when the absolute value of β_2 is 3. For n = 200, the empirical rejection probability is roughly about 0.05 when the absolute value of β_2 is less than 3 and then increases as the absolute value of β_2 increases. Hence, the distribution of the OP LM test shifts to the right as the value of the nuisance parameter increases.

Recall that the type I error or the size of a test of a composite null hypothesis is defined as the supremum of the rejection probabilities under the null hypothesis, see Hogg and Craig (1978, p. 239) and Lehmann (1959, p. 61):

$$\alpha = \underset{\beta \in H_0}{Sup}\ \pi(\beta),$$

where $\pi(\beta)$ is the probability of the test rejecting at the true β, that is, the power function. To find a size-corrected critical value for, say, a 0.05 significant level, select the critical value such that the maximum rejection probability taken over all $\beta \in H_0$ is 0.05. In the simple logit model, $\beta_1 = 0$. Define the set of admissible values of the nuisance parameter, denoted by B_N, to be $\beta_2 \in B_N = [-\beta_{2,max}, \beta_{2,max}]$ where $\beta_{2,max}$ is a positive number. Hence, $H_0 = \{(\beta_1,\beta_2) \mid \beta_1 = 0, \beta_2 \in [-\beta_{2,max}, \beta_{2,max}]\}$; see (1).

The choice of admissible values of the nuisance parameter has a substantial influence on the size-corrected critical value. Figure 2 shows the 0.05 size-corrected critical values for both the Hessian and OP LM test for different choices of the set of admissible values of β_2. For the Hessian LM test, the asymptotic critical value is approximately the size-corrected critical value for all choices for $\beta_{2,max}$. In the case of the OP LM test, however, the size-corrected critical value increases sharply as $\beta_{2,max}$ increases. For example, with n=100 observations it approaches 17.9 as the absolute value of $\beta_{2,max}$ approaches 3 and 80 as $\beta_{2,max}$ approaches 6. For n=200, the size-corrected critical value is 158, which is much larger than in the n=100 case. This counterintuitive effect is explained below by large parameter asymptotics.

The size-corrected critical value of the OP LM test increases sharply as the set of admissible values of the nuisance parameters is enlarged. This implies that the distribution of the test statistic is very different for large values of the nuisance parameters compared to small values. Moreover, the distributions for large values of the nuisance parameters are dominating in the sense that the size-corrected critical values are picked from these distributions. This suggests that a general proof of the size-distortion of the OP LM test can be found by analysing a sequence of distributions indexed by nuisance parameters increasing in value. This technique is denoted large parameter asymptotics.

4. LARGE PARAMETER ASYMPTOTICS

We now use large parameter asymptotics to prove that the size-corrected critical values of the OP LM test are substantially different from the critical values obtained from the conventional asymptotics. In the conventional asymptotics the limit is taken with respect to the sample size whereas in large parameter asymptotics the limit is taken with respect to an index of the parameters. In other words, we find the limit of a sequence of test statistics indexed by the parameters. The index of the parameters is constructed by selecting a ray in the parameter space. Since we are interested in investigating the size properties, attention is restricted to rays satisfying the null hypothesis, namely,

(6) $\alpha(r) = (\alpha_1, \ r\alpha_2), \quad r \geq 0, \ \alpha_1 \in B_1^0 \subset \Re^{k_1}, \ \alpha_2 \in \Re^{k_2}$

With this specification, r can naturally be used as the scalar index on the sequence of test statistics. To insure that the LM statistics depends on the value of the nuisance parameters, make the following assumption

ASSUMPTION 1. $\alpha_2' x_{i2} \neq 0$ for all i.

For the tests in the logit model considered in Section 3, $\alpha_1 = 0$ and α_2 can be any non-zero value since the nuisance parameter space is one-dimensional.

The Monte Carlo experiments showed that the distributions of the LM test statistics for large values of the nuisance parameters determined the size-corrected critical values. Moreover, in the Monte Carlo experiments bad points were deleted implying that the distribution over the sample space is renormed. For large values of the nuisance parameters, the sampling distribution in the renormed sample space concentrates on n sample points; see the proof to the proposition. The large parameter limiting distribution of the LM test statistics in the renormed sample space is given in the following proposition.

PROPOSITION: *Make assumption 1. Assume that bad points are deleted from the sample space. In testing* $H_0 = \{(\beta_1,\beta_2) \mid \beta_1 = 0, \beta_2 \in \mathbb{R}\}$, *the large parameter limiting distributions of the LM statistics along the ray (6) are*

$$\textbf{(7)} \quad P(LM_{EX}=a) = \frac{1}{n}(\#J) \quad where \ \ J=\{\ j \mid a = \frac{n}{n-1}\frac{(x_j-\bar{x})^2}{s_x^2},\ j=1,2,...,n\} \ \ and$$

$$\textbf{(8)} \quad P(LM_{OP}=b) = \frac{1}{n}(\#J) \quad where \ \ J=\{\ j \mid b = n\ \frac{(x_j-\bar{x})^2}{s_x^2+\frac{n-1}{n}(x_j-\bar{x})^2},\ j=1,2,...,n\}$$

where #J is the cardinality of the set J, and $\bar{x}$ *and* s_x *are the sample mean and standard deviation of the regressor.*

Proof: See the appendix.

The critical value obtained from the large parameter limiting distribution in the Proposition is a good approximation to the size-corrected critical value of the OP LM test when $\beta_{2,\max}$ is large. For our problem, a large value of $\beta_{2,\max}$ is 6. From Table 1 we see that for this value of the intercept the percentage of bad points is 77.9%. Figure 2 displays the size-corrected 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 0.95 quantile of the large parameter limiting distribution found by substitution of the regressor values directly into formula (8). These values are 80.0 for n =100 and 161 for n = 200. They are close to the empirical size-corrected critical values.

The large parameter limiting distributions can be further explored by assuming that $x_1,...,x_n$ is a random sample from a normal distribution. In this case, the large parameter limiting distribution of the Hessian LM test is given by

$$\textbf{(9)} \quad LM_{EX} \frac{n-1}{n} \sim F(1,n-1),$$

where F(1,n-1) is the F distribution with 1 and n-1 degrees of freedom. For n = 100, the 0.95 quantile of the large parameter limiting distribution is 3.98, which is close to the critical value in our Monte Carlo experiment, namely, 4.10. Notice that for n large, F(1,n-1) is approximately equal to the $\chi^2(1)$, which is the distribution of the Hessian LM test statistic obtained from conventional asymptotics.

In the case of the OP LM test, the large parameter limiting distribution does not mimic the $\chi^2(1)$ distribution. The large parameter limiting distribution of the OP LM test can be expressed as

$$\textbf{(10)} \quad n \frac{LM_{OP}}{n^2-(n-1)LM_{OP}} \sim F(1,n-1).$$

Using this distribution, the 0.95 quantile is 80.4 for n = 100 and 160 for n = 200. These are very close to 80 and 158, respectively, which are the critical values obtained in our Monte Carlo experiments. Conversely, the 0.95 quantile from the $\chi^2(1)$ is 3.84 which corresponds to the 0.16 quantile of the LM_{OP} distribution (10). Hence, using the asymptotic critical value implies a size distortion of the OP LM test equal to 84 percent. The effect of the sample size n can be seen both in (8) and (10). A larger n shifts the distribution of the OP LM test to the left, exactly the result found in the Monte Carlo simulations.

The Proposition is valid for a logit model with 2 parameters. To obtain a general result with k parameters, the large parameter asymptotics is derived for the full sample space, that is, by not renorming the sample space by deleting bad points. The large parameter asymptotic distribution for the full sample space is given in the following theorem.

THEOREM. *Make assumption 1. In testing* $H=\{(\beta_1,\beta_2)\,|\,\beta_1 \in B \subset \mathbb{R}^{k_1},\ \beta_2 \in \mathbb{R}^{k_2}\}$*, the large parameter asymptotic distributions of the LM statistics along the ray (6) are*

$$F_{EX}(t) = \begin{cases} 0 & if\ t < a \\ 1 & if\ t \geq a \end{cases} \quad \text{with } a = \iota' XW'(X'WX)^{-1}WX'\iota \quad and$$

$$F_{OP}(t) = \begin{cases} 0 & if\ t < b \\ 1 & if\ t \geq b \end{cases} \quad \text{with } b = \iota' XW'(X'WWX)^{-1}WX'\iota ,$$

where $\iota = (1,1,..,1,1)_{n\times 1}'$*, W is a (n×n) diagonal matrix with (i,i)'th element* $\dfrac{c_i}{r_i}$,

$$c_i = \begin{cases} e^{-\alpha_1' x_{i1}} & if\ \alpha_2' x_{i2} > 0 \\ e^{\alpha_1' x_{i1}} & if\ \alpha_2' x_{i2} < 0 \end{cases} \quad and \quad r_i = e^{r|\alpha_2' x_{i2}|} .$$

The large parameter limiting distributions are $LM_{EX} \to^{d} H_{EX}$ *and* $LM_{OP} \to^{d} H_{OP}$ *with cdfs* F_{EX} *and* F_{OP} *where a and b are replaced by* $\lim_{r\to\infty} a$ *and* $\lim_{r\to\infty} b$*, respectively.*

Proof. See the appendix.

The distributions of both LM tests are degenerate at a single point in the sample space. The values of the test statistics in the limit depend on the weight matrix W which is the only quantity depending on r. Each element i in the weight matrix W converges to 0 at a rate determined by the choice of α_2, that is, the ray in the parameter space.

The Theorem is useful to find a lower bound on the size-corrected critical values. The size-corrected critical values of the LM tests cannot be smaller than the limit values a^* and b^* for the Hessian and OP LM tests, respectively, if these tests are to have a size less than one. The reason is that the limit values are derived for parameter values which satisfy the null hypothesis by construction of the parameter ray. For example, since the probability is 1 for the LM_{EX} to be equal to a^* in the limit, the size of the LM_{EX} test is 1 if the size-corrected critical value is chosen below a^*. Hence, the Theorem provides lower bounds on the size-corrected critical values of the LM test statistics.

A number of interesting special cases can be derived from the Theorem. Typically, an intercept is included in the model where the intercept is a nuisance parameter. By selecting a ray in the parameter space in the direction of the intercept, it can be seen that all the terms in the weight matrix converge with the same speed, that is, $r_i = r_j$, i,j=1,2,..,n. Hence, by pulling the r terms outside, the LM_{EX} test equals a constant divided by r_i, which implies that the limit value a^* equals 0. For the LM_{OP} test, however, all the r_i terms cancel and what is left is a constant, whose value is straight forward to calculate. In particular, if the null hypothesis is $\beta_1 = 0$, then all the c_i's are the same and they cancel. The remaining part is

$$b^{*} = \iota'[X(X'X)^{-1}X'\iota].$$

The term in the bracket is an orthogonal projection of ι onto the plane spanned by X. Since the intercept is included, ι is in the plane spanned by X and, thus, $b^* = \iota'\iota = n$. The conclusion is that the size-corrected critical value for the LM_{EX} test is bounded below by 0, that is, no restriction, whereas the size-corrected critical value of the LM_{OP} test is bounded below by n, the sample size. Hence, a larger sample size increases the size-corrected critical value, which exactly mimics the result from the Monte Carlo experiments. It also demonstrates in general that the size-corrected critical value of the LM_{OP} test is substantially larger than the critical value provided by the chi-square distribution.

5. POWERS OF THE TESTS

To judge the performance of the two LM tests, the size-corrected powers are compared. In Section 3, the size properties of the tests were analysed and the OP LM test shown to have large size distortions using the asymptotic chi-square critical value. This, however, does not imply that the LM OP test is inferior to the Hessian LM test in terms of power. To make a fair comparison, the size-corrected powers are used. We will show that even with this adjustment the LM OP test has inferior power which partly is a result of its poor size performance.

For the same simple binary logit model as in section 3, we compare the powers of the 0.05 size-corrected LM tests for three values of $\beta_{2,max}$, that is, three different sets of admissible values of the intercept. Our results show that the power comparison is strongly influenced by the value of $\beta_{2,max}$. The first value is $\beta_{2,max} = 2$. For sample size 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 shows the empirical powers when $\beta_2 = 0$. The empirical powers are essentially the same for the Hessian and OP LM tests.

The second value is $\beta_{2,max} = 3$. 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 and 5 show the empirical powers when $\beta_2 = 0$, 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 5 the OP LM test has empirical power 0.05 when $\beta_1 = 0$. This is because the $\beta_2 = 3$ is on the boundary of the set of admissible values for β_2 and the rejection probability is increasing in β_2.

It is misleading to interpret the lower power of the OP LM test in Figure 4 as due to size distortion. The test has the correct size by construction. What Figure 4 shows is that the rejection probability is less than the size when $\beta_2 = 0$. By contrast, the test has rejection probability equal to its size in Figure 5. Even in this case, the power of the OP LM test is substantially less than the power of the Hessian LM test.

In the present setting, there is no natural definition of the admissible set for β_2, except possibly the entire real line; the latter is the admissible set used in ML estimation. If this definition is used, then the power comparison is even more dramatic. Figure 6 shows the size-corrected powers of the Hessian and OP LM tests when $\beta_{2,max}$ is 6, $\beta_2 = 0$ and n = 100. In this case, the power function of the OP LM test is a horizontal straight line at 0; that is, the OP LM test never rejects H_0.

Part of the reason for the inferiority of the OP LM test is its poor size performance. Under the null hypothesis, the distributions of the OP LM tests shift to the right for larger values of the nuisance parameters. Therefore, if the true value of the nuisance parameter is small, then the true distribution is far to the left of the distribution from which the critical value is selected, namely, a distribution for which the value of the nuisance parameter is large. Even though the true distribution shifts to the right as the parameter of interest increases in value, the critical value is so far out in the tail to begin with that the power only increases slightly, and certainly much less than for the Hessian LM test. Hence, it is not surprising that a test statistic which has very different distributions for different parameter values satisfying the null hypothesis also has poor power performance and, as in the extreme case of Figure 6, no power at all.

In Monte Carlo experiments, a useful criterion of empirical relevance is the probability that the ML estimate blows-up. In Table 1 we report the percentage of bad points for selected positive values of β_2 when $\beta_1 = 0$ and n = 100. The results show that the percentage of bad points is 0.007 for $\beta_2 = 3$ and increases as the absolute value of β_2 increases. Since bad points are uncommon in practice (the ML estimate typically does not blow-up in applications), the results in Table 1 suggest to us that $\beta_2 = 3$ is a plausible empirically relevant value when the null is true.

6. CONCLUSION

It is well known for nonlinear models that the power function depends on the value of the nuisance parameters. What is often overlooked is that the size-corrected critical value of a test may depend on the set of admissible values of the nuisance parameters. In this paper, we show that this dependence can affect the conclusions of Monte Carlo experiments designed to compare the size-corrected power of competing tests. In particular, for a simple logit model, where the nuisance parameter is the intercept, the expected Hessian and OP LM tests have about the same size-corrected power when the range of admissible values for the intercept is sufficiently narrow; as the range increases so does the superiority of the expected Hessian LM test.

The OP LM test suffers from serious size distortion when using the conventional asymptotic critical value. This has been observed in Monte Carlo experiments for many tests based on the outer product information matrix. Large parameter asymptotics permits us to obtain an analytical result on the size-distortion of the OP LM test. In fact, in a general logit model with an intercept, the size-corrected critical value of the OP LM test is closely related to the sample size, and it is many order of magnitude larger than the conventional critical value. The main difference between conventional asymptotics and large parameter asymptotics is that the latter is derived using the nuisance parameters and for a fixed sample size.

The arguments for and against a particular admissible set of nuisance parameters are typically based on an appeal to empirical relevance. These arguments often tend to rely implicitly on a prior distribution for the nuisance parameters. An example of this is the argument that $\beta_{2,max} = 3$ is not empirically relevant because $\beta_2 = 3$ is “very unlikely” and produces rare events. It suffices to say that a prior on the nuisance parameters is not compatible with the classical analysis in this paper. But even with a prior, the results would not change qualitatively if the prior has the whole set of admissible values as support.

It is worth stressing that the admissible set refers to the possible, not the likely, values of the parameters. In this regard, it is instructive to consider the Challenger Space Shuttle data (Presidential Commission on the Space Shuttle *Challenger* Accident (1986)). This data includes the number of occurrences of distress (O-ring failures) and the launch temperature of the 23 space shuttle flights before the explosion of the Challenger space shuttle. Applying the logit model to this data, the estimated intercept is about 5. When the null of no temperature effect is true, the probability of distress is essentially one. Since the Challenger data is arguably one of the most analysed data sets in the last decade, the argument that for this data set that $\beta_2 = 5$ is not empirically relevant is not credible. Of course, the alternative also matters. In the case of the space shuttle, the temperature effect may not be zero.

For many models, there is no natural definition for the admissible set of values for the nuisance parameters. Consequently, different experimenters may use different definitions of the admissible set. As our results illustrate, different definitions can lead to different conclusions about which test is best. Hence, the conclusions of experimental power studies may be more problematical than they appear. Our recommendation is to conduct a power study using different admissible sets. If different sets produce different results, then the reader can draw his or her own conclusions.

APPENDIX

A. Proof of Proposition

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

$$Q(y) = \frac{P(y)}{1-P(bad\ points)} = \frac{a}{(1+a)^n - (1+a^n)}\ , \quad a = e^{-r\alpha_2}.$$

For $\alpha_2 > 0$, $r \rightarrow \infty$ implies $a \rightarrow 0$. Using L'Hospital's rule, $Q(y) \rightarrow 1/n$. A similar result can be obtained when α_2 is negative.

Next we calculate 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 $\tilde{\beta}_2 = F^{-1}(\sum_{i=1}^{n} y_i/n)$ or $F(\tilde{\beta}_2) = (n-1)/n$. By substituting into (5), the score vector is

$$S(\{0,\tilde{\beta}_2\}) = \begin{pmatrix} \sum (y_i - F(\tilde{\beta}_2))x_i \\ 0 \end{pmatrix} = \begin{pmatrix} \bar{x} - x_1 \\ 0 \end{pmatrix} , \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

The expected Hessian covariance matrix (3) is

$$V_{EX}(\{0,\tilde{\beta}_2\}) = \frac{n}{n-1}\frac{1}{V(x)}\begin{bmatrix} 1 & -\bar{x} \\ -\bar{x} & \frac{1}{n}\sum_{i=1}^{n} x_i^2 \end{bmatrix}, \quad V(x) = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2$$

The OP covariance matrix (4) can be written as

$$V_{op}(\{0,\tilde{\beta}_2\}) = \frac{1}{D}\begin{bmatrix} 1 & -\frac{n-1}{n}x_1 - \frac{1}{n}\bar{x} \\ -\frac{n-1}{n}x_1 - \frac{1}{n}\bar{x} & \frac{1}{n}(V(x) + \bar{x}^2) + \frac{n-1}{n}x_1^2 \end{bmatrix} \text{ where}$$

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

Because the second element of the score vector is zero, only the (1,1) element of the covariance matrix is used to calculate the value of the LM test statistic. The other n-1 values of the large parameter limiting distribution of the LM test can be obtained by replacing x_1 with x_i for the corresponding sample $y^{(i)}$. This completes the proof.

B. Proof of Theorem

First the convergence in distribution is established. Let the discrete random variables LM_{EX} and LM_{OP} have probability mass functions P_{EX} and P_{OP}, respectively. Then

$$P_{EX}(LM_{EX} = h) = P(y \in A),$$

where y is a sample point, A = { y : LM_{EX}(y) = h } and P the probability mass function over the sample space. The probability mass function over the sample space is given by

$$P(y) = \prod_{i=1}^{n} F(\beta' x_i)^{y_i}(1-F(\beta' x_i))^{1-y_i}$$

The limit distribution of P along the ray

$$\alpha(r) = (\alpha_1,\ r\alpha_2), \quad r \geq 0,\ \alpha_1 \in B \subset \Re^{k_1},\ \alpha_2 \in \Re^{k_2}$$

degenerates to a single sample point y^* given by

$$y^* = (y_1^*,..,y_n^*) \text{ where } y_i^* = I(\beta_2' x_{i2} > 0) \text{ and I is the indicator function,}$$

since

$$\lim_{r\to\infty} F(\alpha(r)' x_i) = \lim_{r\to\infty} F(\alpha_1' x_{i1} + r\alpha_2' x_{i2}) = \lim_{r\to\infty} F(r\alpha_2' x_{i2}) = I(\alpha_2' x_{i2} > 0)$$

by assumption 1. Then the limit cdf of the Hessian LM test, $F_{H_{EX}}$, and OP LM test, $F_{H_{OP}}$, are:

$$F_{H_{EX}}(a) = \begin{cases} 0 & if\ a < a^* \\ 1 & if\ a \geq a^* \end{cases} \quad \text{with } a^* = \lim_{r\to\infty} LM_{EX}(y^*)$$

and

$$F_{H_{OP}}(b) = \begin{cases} 0 & if\ b < b^* \\ 1 & if\ b \geq b^* \end{cases} \quad \text{with } b^* = \lim_{r\to\infty} LM_{OP}(y^*)$$

The cdf has to converge to the limiting cdf at every continuity point of the limiting cdf (Amemiya 1985). The limiting cdf only has one discontinuity point namely at a^*. Using the limit of the probability mass function over the sample space, it is seen that for all sample points except y^* the probability converges to zero as r goes to infinity.

Next, the limit values a^* and b^* are derived. For that purpose, the LM statistics are approximated. The LM statistics depends on the parameter values β only through F, the logistic distribution function. Therefore, the convergence properties of the expected information matrix will depend crucially on the behaviour of

$$F(\alpha' x_i)(1-F(\alpha' x_i)),$$

and the convergence properties of the score vector and OP information matrix will depend on

$$(y_i^* - F(\alpha' x_i)).$$

The convergence rate of these terms as r → ∞ is defined as a real positive function of r, g(r), which bounds a term away from 0 and ∞ when multiplied to the term. Hence, g(r) must satisfy the following condition

$$0 < \left|\lim_{r\to\infty} g(r)\left[F(\alpha(r)'x_i)(1-F(\alpha(r)'x_i))\right]\right| < M, \quad M \in \Re_+.$$

The convergence rate

$$g(r) = e^{r|\beta_2' x_{i2}|}$$

is next shown to satisfy the condition.

First, consider $F(\alpha(r)'x_i)(1-F(\alpha(r)'x_i))$. The limit of F(α(r)'$x_i$) is

$$\lim_{r\to\infty} F(\alpha(r)'x_i) = \lim_{r\to\infty} F(r\alpha_2' x_{i2}) = I(\alpha_2' x_{i2} > 0)$$

where I is the indicator function. Therefore, two cases arise depending on the sign of $\alpha_2'x_{i2}$.

Consider the case where $\alpha_2'x_{i2} > 0$. Then $\lim_{r\to\infty} F(\alpha(r)'x_i) = 1$ implying that

$$\lim_{r\to\infty} F(\alpha(r)'x_i)(1-F(\alpha(r)'x_i))g(r) = \lim_{r\to\infty} (1-F(\alpha(r)'x_i))g(r).$$

Now, verify that the expression for g(r) given in the Theorem indeed satisfies the conditions for convergence. Rewrite (1-F(α(r)'x_i))g(r) as

$$\left(1-\frac{1}{1+e^{-(\alpha_1' x_{i1}+r\alpha_2' x_{i2})}}\right)e^{r|\alpha_2' x_{i2}|} = \left(\frac{1}{1+e^{\alpha_1' x_{i1}+r\alpha_2' x_{i2}}}\right)e^{r|\alpha_2' x_{i2}|} = \frac{1}{e^{-r|\alpha_2' x_{i2}|}+e^{\alpha_1' x_{i1}+r(\alpha_2' x_{i2}-|\alpha_2' x_{i2}|)}}$$

The first term in the denominator goes to zero as r goes to infinity. Since $\alpha_2'x_{i2} > 0$, the r part of the second term in the denominator cancels. Hence,

$$\lim_{r\to\infty} \frac{1}{e^{-r|\alpha_2' x_{i2}|}+e^{\alpha_1' x_{i1}+r(\alpha_2' x_{i2}-|\alpha_2' x_{i2}|)}} = e^{-\alpha_1' x_{i1}},$$

which is a non-zero constant.

Consider the second case where $\alpha_2'x_{i2} < 0$. Then $\lim_{r\to\infty} (1-F(\alpha(r)'x_i)) = 1$

implying that

$$\lim_{r\to\infty} F(\alpha(r)'x_i)(1-F(\alpha(r)'x_i))g(r) = \lim_{r\to\infty} F(\alpha(r)'x_i)g(r).$$

As above, verify that the expression for g(r) given in the Theorem indeed satisfies the conditions for convergence. Rewrite $F(\alpha(r)'x_i)g(r)$ as

$$\left(\frac{1}{1+e^{-(\alpha_1'x_{i1}+r\alpha_2'x_{i2})}}\right)e^{r|\alpha_2'x_{i2}|} = \frac{1}{e^{-r|\alpha_2'x_{i2}|}+e^{-\alpha_1'x_{i1}-r(\alpha_2'x_{i2}+|\alpha_2'x_{i2}|)}}$$

The first term in the denominator goes to zero as r goes to infinity. Since $\alpha_2'x_{i2} < 0$, the r part of the second term in the denominator cancels. Hence,

$$\lim_{r\to\infty} \frac{1}{e^{-r|\alpha_2'x_{i2}|}+e^{-\alpha_1'x_{i1}-r(\alpha_2'x_{i2}+|\alpha_2'x_{i2}|)}} = e^{\alpha_1'x_{i1}},$$

which is a non-zero constant.

The convergence of $(y_i^*-F(\alpha(r)'x_i))$ can be found in a similar manner as for $F(\alpha(r)'x_i)(1-F(\alpha(r)'x_i))$. Again, consider two cases. First, suppose $\alpha_2'x_{i2} > 0$. Then $y_i^* = 1$. From above,

$$\lim_{r\to\infty} F(\alpha(r)'x_i)(1-F(\alpha(r)'x_i))g(r) = \lim_{r\to\infty} (1-F(\alpha(r)'x_i))g(r),$$

and, hence, $(y_i^*-F(\alpha(r)'x_i))$ with $y_i^* = 1$ has the same limit as $F(\alpha(r)'x_i)(1-F(\alpha(r)'x_i))$ for $\alpha_2'x_{i2} > 0$. In the second case where $\alpha_2'x_{i2} < 0$, the limits are also the same apart from the sign. When $\alpha_2'x_{i2} < 0$, $y_i^* = 0$. From above,

$$\lim_{r\to\infty} F(\alpha(r)'x_i)(1-F(\alpha(r)'x_i))g(r) = \lim_{r\to\infty} F(\alpha(r)'x_i)g(r) = -\left[\lim_{r\to\infty} -F(\alpha(r)'x_i)g(r)\right],$$

and, hence, the last expression is the negative of the limit of $(y_i^*-F(\alpha(r)'x_i))$ with $y_i^* = 0$ for $\alpha_2'x_{i2} < 0$.

Using the above result, the score and covariance matrices can be approximated. For convenience, define

$$c_i = \begin{cases} e^{-\alpha_1'x_{i1}} & if\ \alpha_2'x_{i2} > 0 \\ e^{\alpha_1'x_{i1}} & if\ \alpha_2'x_{i2} < 0 \end{cases}, \quad d_i = \begin{cases} e^{-\alpha_1'x_{i1}} & if\ \alpha_2'x_{i2} > 0 \\ -e^{\alpha_1'x_{i1}} & if\ \alpha_2'x_{i2} < 0 \end{cases} \quad \text{and} \quad r_i = e^{r|\alpha_2'x_{i2}|}$$

and form two diagonal matrices W and W_d with (i,i)'th element $\frac{c_i}{r_i}$ and $\frac{d_i}{r_i}$, respectively. Then the score vector and covariance matrices can be approximated by

$$S(\alpha(r)) \;=\; \sum_{i=1}^{n} \frac{d_i}{r_i} x_i \;=\; W_d X^{/} \iota \,,$$

$$V_{EX}(\alpha(r)) \;=\; \left[\sum_{i=1}^{n} \frac{c_i}{r_i} x_i x_i^{/} \right]^{-1} \;=\; (X^{/} W X)^{-1} \text{ and}$$

$$V_{OP}(\alpha(r)) \;=\; \left[\sum_{i=1}^{n} \frac{d_i^2}{r_i^2} x_i x_i^{/} \right]^{-1} \;=\; (X^{/} W_d W_d X)^{-1} \,,$$

where $\iota' = (1,..,1)_{n \times n}$. Since $W_d W_d = WW$, W_d can be replaced with W when writing the LM statistics.

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.

Campbell, B. and J.M. Dufour (1995): “Exact Nonparameter Tests of Orthogonality and Random Walk in the Presence of a Drift Parameter,” Forthcoming in *The International Economic Review.*

Chesher, A. and R. Spady (1991): “Asymptotic Expansions of the Information Matrix Test,“ *Econometrica*, 59, 787-815.

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

Davidson R., and J. G. MacKinnon.(1983): “Small Sample Properties of Alternative Forms of the Lagrange Multiplier Test,” *Economics letters* 12, 269-75.

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.(1985): “Testing Linear and Loglinear Regressions against Box-Cox Alternatives,” *The Canadian Journal of Economics* 18, 499-517.

Davidson R., and J. G. MacKinnon.(1992): “A New Form of the Information Matrix Test,” *Econometrica* 60, 145-57.

Dufour, J.M. (1994): “Some Impossibility Theorem in Econometrics, with Applications to Instrumental Variables, Dynamic Models and Cointegration,” Technical Report, C.R.D.E., University of Montreal.

Dufour, J.M. and O. Torres (1995): “Two-sided Autoregressive and Exact Inference for Stationary and Nonstationary Processes,” Technical Report, C.R.D.E., University of Montreal.

Godfrey, L. G., M. McAleer, and C. R. McKenzie (1988): “Variable Addition and Lagrange Multiplier Tests for Linear and Logarithmic Regression Models, “ *Review of Economics and Statistics,* 70, 492-503.

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.

Orme, C. (1990): “The Small-Sample Performance of the Information matrix Test,“ *Journal of Econometrics*, 46, 309-331.

Presidential Commission on the Space Shuttle *Challenger* Accident (1986)): *Report of the Presidential Commission on the Space Shuttle Challenger Accident* (Vol. 1& 2). Author, Washington, DC.

Savin and Würtz (1996): “The effect of Nuisance Parameters on the Power of LM Tests in Logit and Probit Models,” *University of Iowa Working Paper Series 96-05*.

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

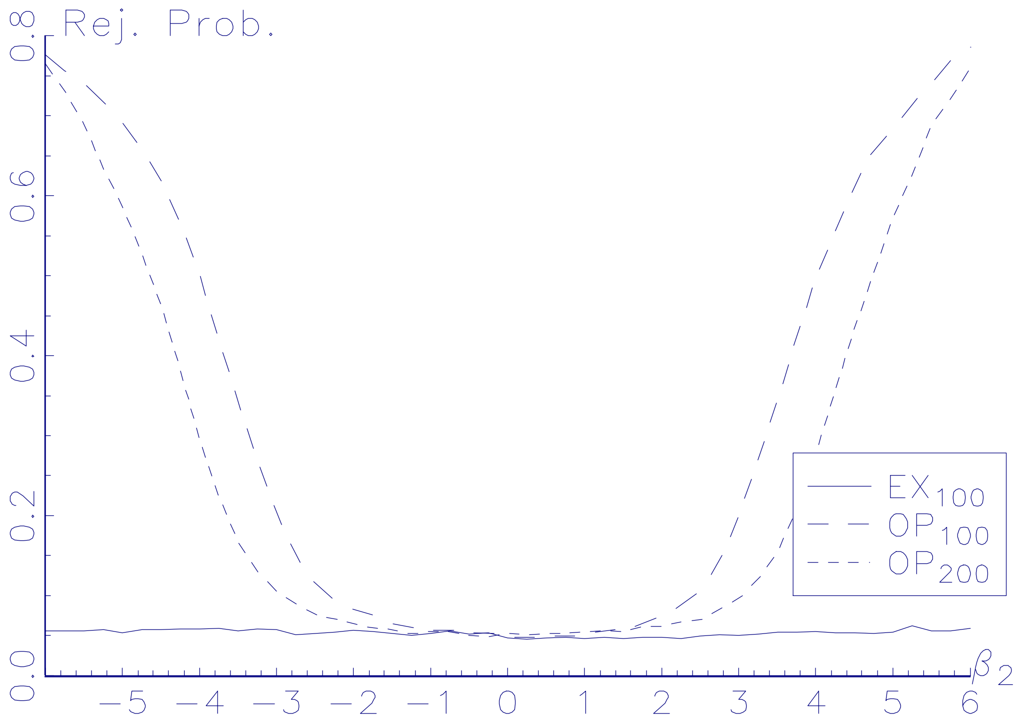

FIGURE 1.) Rejection probabilities for two-sided symmetric LM test with the 0.05 asymptotic critical value and β_2 equal to 0.

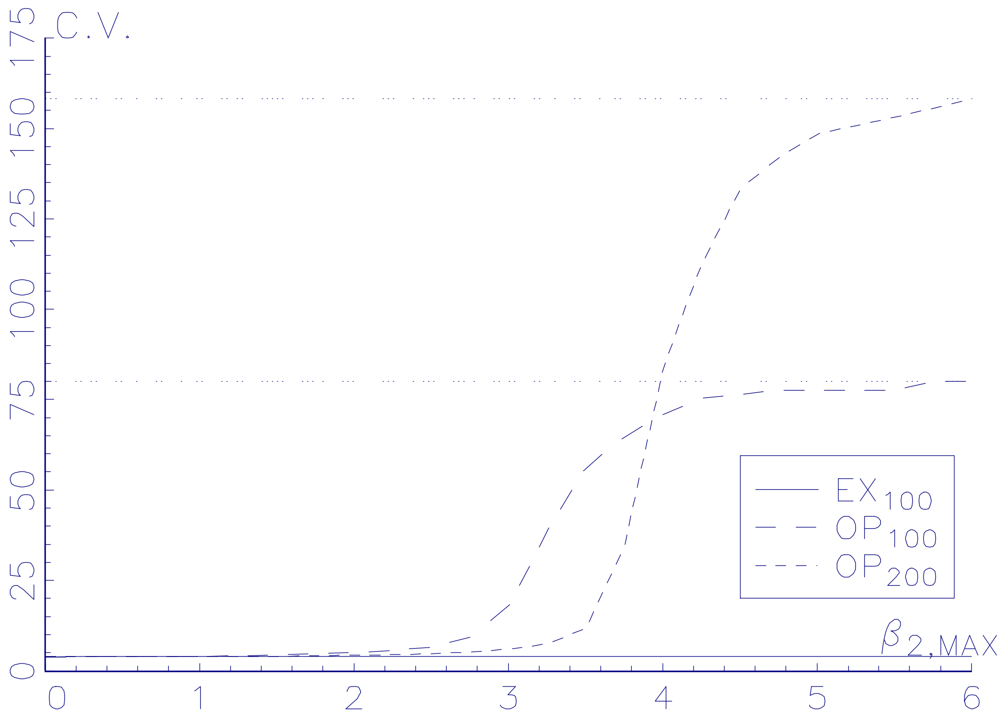

FIGURE 2.) Size-corrected critical values for 0.05 two-sided symmetric Hessian and OP LM tests with β_2 equal to 0.

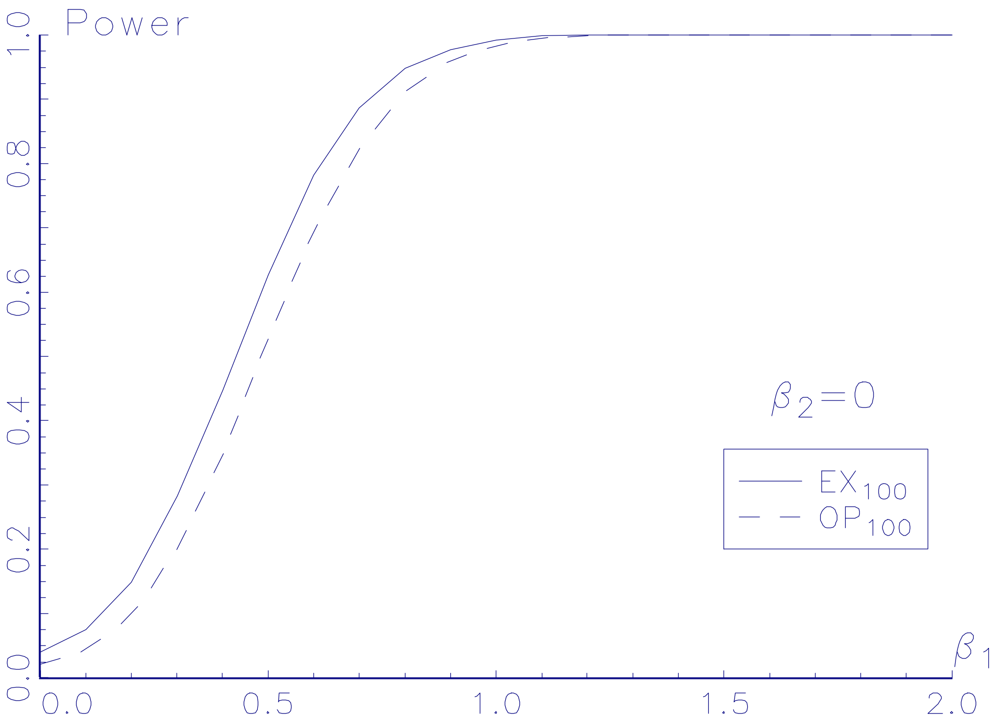

FIGURE 3.) Size-corrected powers of 0.05 two-sided symmetric LM tests with $\beta_{2,MAX}$ equal to 2.

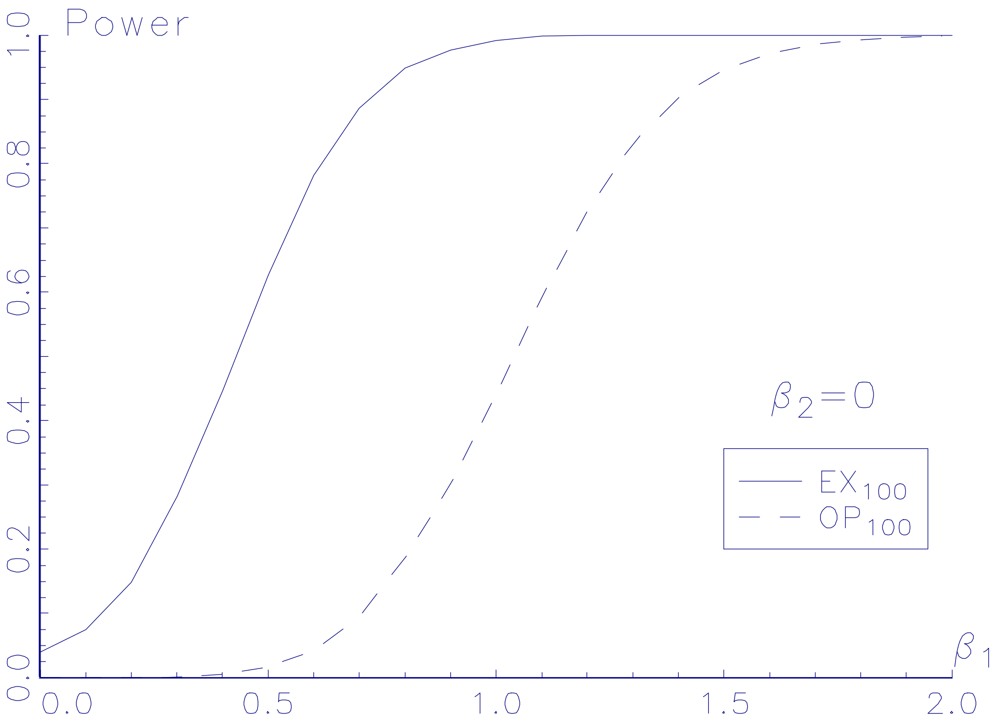

FIGURE 4.) Size-corrected powers of 0.05 two-sided symmetric LM tests with $\beta_{2,MAX}$ equal to 3.

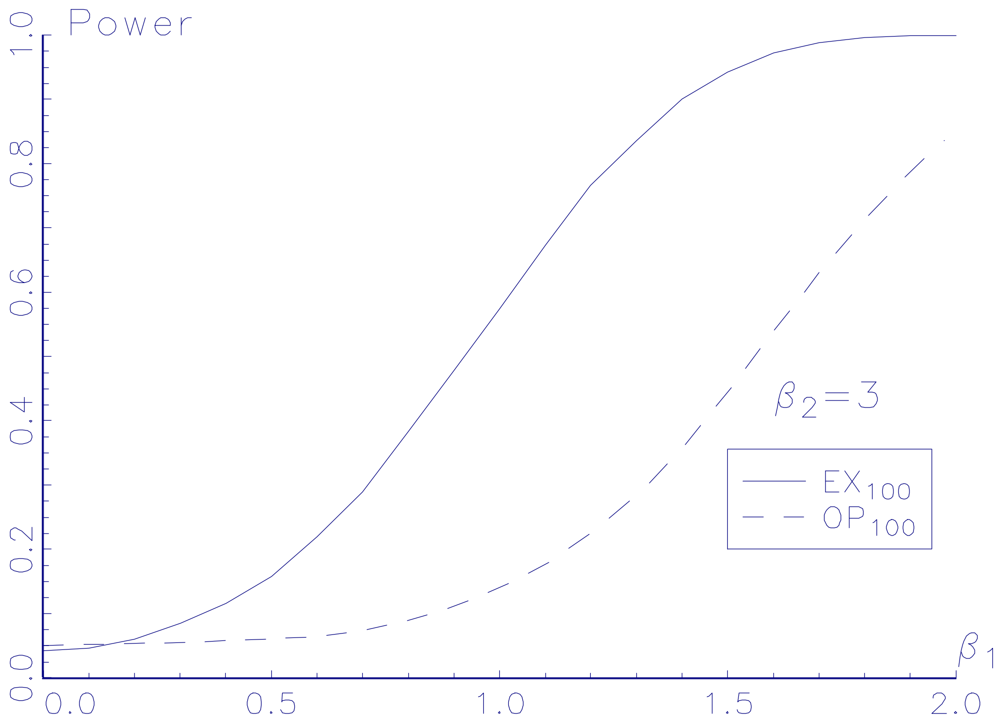

FIGURE 5.) Size-corrected powers of 0.05 two-sided symmetric LM tests with $\beta_{2,MAX}$ equal to 3.

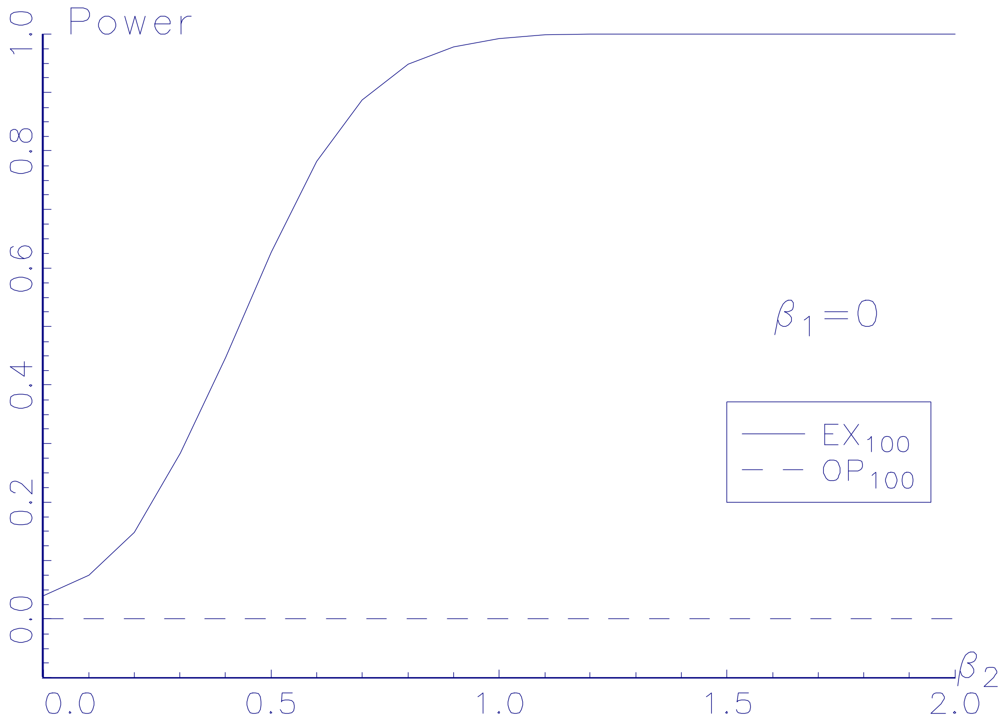

FIGURE 6.) Size-corrected powers of 0.05 two-sided symmetric LM tests with $\beta_{2,MAX}$ equal to 6.

Working Paper

1997-4	Alvaro Forteza: Multiple Equilibria in the Welfare State with Costly Policies.
1997-5	Torben M. Andersen and Morten Stampe Christensen: Contract Renewal under Uncertainty.
1997-6	Jan Rose Sørensen: Do Trade Unions Actually Worsen Economic Performance?
1997-7	Luca Fanelli: Estimating Multi-Equational LQAC Models with I(1) Variables: a VAR Approach.
1997-8	Bo Sandemann Rasmussen: Long Run Effects of Employment and Payroll Taxes in an Efficiency Wage Model.
1997-9	Bo Sandemann Rasmussen: International Tax Competition, Tax Co-operation and Capital Controls.
1997-10	Toke S. Aidt: Political Internalization of Economic Externalities. The Case of Environmental Policy in a Politico-Economic Model with Lobby Groups.
1997-11	Torben M. Andersen and Bo Sandemann Rasmussen: Effort, Taxation and Unemployment.
1997-12	Niels Haldrup: A Review of the Econometric Analysis of I(2) Variables.
1997-13	Martin Paldam: The Micro Efficiency of Danish Development Aid.
1997-14	Viggo Høst: Better Confidence Intervals for the Population Mean by Using Trimmed Means and the Iterated Bootstrap?
1997-15	Gunnar Thorlund Jepsen and Peter Skott: On the Effects of Drug Policy.
1997-16	Peter Skott: Growth and Stagnation in a Two-Sector Model: Kaldor's Mattioli Lectures.
1997-17	N.E. Savin and Allan H. Würtz: The Effect of Nuisance Parameters on Size and Power; LM Tests in Logit Models.

CENTRE FOR NON-LINEAR MODELLING IN ECONOMICS

DEPARTMENT OF ECONOMICS - UNIVERSITY OF AARHUS - DK - 8000 AARHUS C - DENMARK

☎ +45 89 42 11 33 - TELEFAX +45 86 13 63 34

Working papers, issued by the Centre for Non-linear Modelling in Economics:

1995-6 David Easley, Nicholas M. Kiefer, Maureen O'Hara and Joseph B. Paperman: Liquidity, Information, and Infrequently Traded Stocks.

1995-13 Tom Engsted, Jesus Gonzalo and Niels Haldrup: Multicointegration and Present Value Relations.

1996-1 Tom Engsted and Niels Haldrup: Estimating the LQAC Model with I(2) Variables.

1996-2 Peter Boswijk, Philip Hans Franses and Niels Haldrup: Multiple Unit Roots in Periodic Autoregression.

1996-3 Clive W.J. Granger and Niels Haldrup: Separation in Cointegrated Systems, Long Memory Components and Common Stochastic Trends.

1996-4 Morten O. Ravn and Martin Sola: A Reconsideration of the Empirical Evidence on the Asymmetric Effects of Money-Supply shocks: Positive vs. Negative or Big vs. Small?

1996-13 Robert F. Engle and Svend Hylleberg: Common Seasonal Features: Global Unemployment.

1996-14 Svend Hylleberg and Adrian R. Pagan: Seasonal Integration and the Evolving Seasonals Model.

1997-1 Tom Engsted, Jesus Gonzalo and Niels Halrup: Testing for Multicointegration.

1997-7 Luca Fanelli: Estimating Multi-Equational LQAC Models with I(1) Variables: a VAR Approach.

1997-12 Niels Haldrup: A Review of the Econometric Analysis of I(2) Variables.

1997-14 Viggo Høst: Better Confidence Intervals for the Population Mean by Using Trimmed Means and the Iterated Bootstrap?

1997-17 N.E. Savin and Allan H. Würtz: The Effect of Nuisance Parameters on Size and Power; LM Tests in Logit Models.