# DEPARTMENT OF ECONOMICS

## Working Paper

EMPIRICALLY RELEVANT POWER COMPARISONS
FOR LIMITED DEPENDENT VARIABLE MODELS

N.E. Savin
Allan H. Würtz

Working Paper No. 1999-20
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

EMPIRICALLY RELEVANT POWER COMPARISONS
FOR LIMITED DEPENDENT VARIABLE MODELS

N.E. Savin
Allan H. Würtz

Working Paper No. 1999-20

## 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

**EMPIRICALLY RELEVANT POWER COMPARISONS FOR LIMITED DEPENDENT VARIABLE MODELS***

N.E. Savin

Department of Economics, 108 Pappajohn Bus. Bldg.
University of Iowa, Iowa City, IA 52242
(319)-335 0855

Allan H. Würtz

Department of Economics, University of Aarhus, Building 350,
DK-8000 Aarhus C, Denmark

August 18, 1999

Keywords: Bootstrap, censored regression, critical values, composite hypotheses, finite sample power, probit model.

JEL: C12, C15, C24, C25

*We gratefully acknowledge discussions with Joel Horowitz, Hidehiko Ichimura, and George R. Neumann and the comments of Jim Powell.

## 1. Introduction.

Most hypotheses in limited dependent variable (LDV) models are composite, meaning that the null hypothesis $H_0$ does not completely specify the data generating process (DGP). In this case, the null specifies only that the DGP belongs to a given set. As a consequence, the sampling distribution of a test statistic under $H_0$ is unknown except in special cases because it depends on the true DGP in the set specified by $H_0$. The problem is how to test the null hypothesis in this situation. In this paper, the bootstrap is used to solve the hypothesis testing problem.

The LDV models considered in this paper are the simple binary probit model and the simple censored normal linear regression model. For these models, a typical null hypothesis $H_0$ is that the slope coefficient is zero. This null is composite, the remaining parameters being nuisance parameters. This null is tested using the Lagrange Multiplier (LM), likelihood ratio (LR), and Wald test statistics. In our Monte Carlo experiments, we compare the powers of the competing tests when the tests use bootstrap-based critical values. We argue that the powers of the tests with bootstrap-based critical values are empirically relevant because these critical values can be calculated in applications.

There are two basic approaches to obtaining critical values for testing a composite null hypothesis. One approach employs the concept of the size of a test. The size is the supremum of the test's rejection probability over all DGP's contained in $H_0$. The $\alpha$-level *size-corrected critical value* is the critical value that makes the size equal to $\alpha$. In principle, the exact, size-corrected critical value can be calculated if it exists, but this is rarely done in applications, possibly because doing so typically entails very difficult computations.

In addition to the difficulty of computing size-corrected critical values, size-corrected tests (that is, tests based on size-corrected critical values) have two fundamental problems. First, the size-corrected critical value may be infinite, in which case a test based on the size concept has no power. Second, even if the size-corrected critical value is finite, the power of the test may be less than or equal to its size. Dufour (1997) gives several examples of the first problem. Bahadur and Savage (1956) demonstrate the second for the case of testing a hypothesis about a population mean. Savin and Würtz (1999a) consider testing the hypothesis that the slope parameter is zero in a binary logit model with one explanatory variable. In their example, the size-corrected critical value of the outer product LM test exists, but it is sensitive to the interval of admissible values for the intercept. If this interval is large, then the power of the test is zero for empirically relevant sample sizes.

A second approach to dealing with a composite $H_0$ is to base the test on an estimator of the *Type I critical value*. Horowitz and Savin (1998) define this critical value as the one that would be obtained if the exact finite-sample distribution of the test statistic under the true DGP were known. In general, the true Type I critical value is unknown because the exact finite-sample distribution of the test statistic depends on population parameters that are not specified by $H_0$. Thus, an approximation to the Type I critical value is required to implement the second approach.

An approximation to the Type I critical value often can be obtained by using the (first-order) asymptotic distribution of the test statistic to approximate its finite-sample distribution. This approximation is useful because most test statistics in econometrics are asymptotically pivotal: their asymptotic distributions do not depend on unknown population parameters when the hypothesis being tested is true. Thus, an approximate Type I critical value can be obtained from asymptotic distribution theory without knowledge of the true DGP. Critical values obtained from asymptotic distribution theory are widely used in applications. However, Monte Carlo experiments have shown that first-order asymptotic theory often gives a poor approximation to the exact distributions of test statistics with the sample sizes available in applications. As a result, the true and nominal probabilities that a test makes a Type I error can be very different when an asymptotic critical value is used.

Under certain conditions, the bootstrap provides an approximation to the Type I critical value that is more accurate than the approximation of first-order asymptotic theory. These conditions are satisfied for the test statistics considered in this paper. Given that the LDV models in this paper are fully parametric, the Type I critical can be estimated by the parametric bootstrap. Our Monte Carlo results show that the parametric bootstrap estimates of the Type I critical values provide good control over the probability of making a Type I error. Thus, for the examples we consider, the bootstrap provides a way to obtain empirically relevant critical values for the purpose of making power comparisons. As consequence, the powers of tests with bootstrap-based critical values are empirically relevant because these critical values can be calculated in actual applications.

By contrast, conventional Monte Carlo studies comparing the finite-sample powers of tests usually take a different approach to obtaining critical values. Most studies report powers based on critical values that are called "size-corrected," but are really Type I critical values for essentially arbitrary chosen simple null hypotheses. These Type I critical values ignore the uncertainty about the values of the nuisance parameters; they implicitly assume that the values of the nuisance parameters are known. Therefore, conventional Monte Carlo studies compare tests using powers that can be misleading in empirical research. In short, the Type I critical values used by conventional Monte Carlo studies are both misnamed and irrelevant in empirical research.

The organization of this paper is as follows. The Type I critical value is defined in Section 2. The bootstrap critical value when $H_0$ is true is developed in Section 3. Section 4 considers the Type I critical value when $H_0$ is false and its bootstrap estimate. Monte Carlo results on the numerical performance of the bootstrap are presented in Section 5 for a simple binary probit model and in Section 6 for a simple censored normal linear regression model. Section 7 concludes the paper.

## 2. Type I Critical Values

Let the data be a random sample of size $n$ from a probability distribution whose cumulative distribution function (CDF) is $F$. Denote the data by $\{X_i, i = 1, ..., n\}$. For the purpose of this paper, $F$ is assumed to belong to a family of CDF's that is indexed by the finite-dimensional parameter $\theta$ whose population value is $\theta*$. We write $F(x, \theta*)$ for $P(X \leq x)$ and $F(\cdot, \theta)$ for a general member of the parametric family. The unknown parameter $\theta$ is restricted to a parameter set $\Theta$. The null hypothesis $H_0$ restricts $\theta$ to a subset $\Theta_0$ of $\Theta$. If $H_0$ is composite, then $\Theta_0$ contains two or more points.

Let $T_n = T_n(X_1, \ldots, X_n)$ be a statistic for testing $H_0$. Let $G_n[\tau, F(\cdot, \theta)] \equiv P(T_n \leq \tau | \theta)$ be the exact finite sample CDF of $T_n$ when the CDF of the sampled distribution is $F(\cdot, \theta)$. Consider a symmetrical, two-tailed test of $H_0$. This is the kind of test typically used to test the parameters of LDV models. $H_0$ is rejected by such a test if $|T_n|$ exceeds a suitable critical value and accepted otherwise. For $\theta*$ in $\Theta_0$, the exact, $\alpha$-level Type I critical value of $|T_n|$, $z_{n\alpha}$, is defined as the solution to the equation $G_n[z_{n\alpha}, F(\cdot, \theta*)] - G_n[-z_{n\alpha}, F(\cdot, \theta*)] = 1 - \alpha$. A test based on this critical value rejects $H_0$ if $|T_n| > z_{n\alpha}$. Such a test makes a Type I error with probability $\alpha$. However, $z_{n\alpha}$ can be calculated in applications only in special cases. If $H_0$ is simple so that $\Theta_0$ contains only one point, then $\theta^*$ is specified by $H_0$, and $z_{n\alpha}$ can be calculated or estimated with arbitrary accuracy by Monte Carlo simulation.

If, as usually happens in econometrics, $H_0$ is composite, then $\theta^*$ is unknown and $z_{n\alpha}$ cannot be evaluated unless $G_n[\tau, F(\cdot, \theta)]$ does not depend on $\theta$ when $H_0$ is true. In this special case, $T_n$ is said to be *pivotal*. The Student $t$ statistic for testing a hypothesis about the mean of a normal population or a slope coefficient in a normal linear regression model is pivotal. However, pivotal test statistics are generally not available in most econometric applications, and, in particular, for LDV models. When $T_n$ is not pivotal, its Type I critical value $z_{n\alpha}$ can be very different at different points in $\Theta_0$. This is illustrated by the examples in Sections 5 and 6.

When $H_0$ is composite and $T_n$ is not pivotal, it is necessary to replace the true Type I critical value with an approximation or estimator. First-order asymptotic distribution theory provides one approximation. Most test statistics in econometrics are asymptotically pivotal. Indeed, the asymptotic distributions of most commonly used test statistics are standard normal or chi-square under $H_0$, regardless of the details of the DGP. If $n$ is sufficiently large and $T_n$ is asymptotically pivotal, then $G_n[\cdot, F(\cdot, \theta^*)]$ can be approximated accurately by the asymptotic distribution of $T_n$. The asymptotic distribution is the same for any $\theta$ in $H_0$ (including $\theta^*$) when $T_n$ is asymptotically pivotal, so approximate critical values for $T_n$ can be obtained from the asymptotic distribution without having to know $\theta^*$.

Critical values obtained from asymptotic distribution theory are widely used in applications of LDV models. However, Monte Carlo experiments have shown that first-order asymptotic theory often gives a poor approximation to the distributions of test statistics with the sample sizes available in applications. As a result, the true and nominal probabilities that a test makes a Type I error can be very different when an asymptotic critical value is used. Davidson and MacKinnon (1984) have documented such distortions for the case of logit and probit models.

## 3. Bootstrap Critical Values when the Null Hypothesis is True

The bootstrap provides a way to obtain approximations to the Type I critical value of a test that are more accurate than the approximations of first-order asymptotic distribution theory. The bootstrap does this by using the information in the sample to estimate $\theta$ and, thereby, $G_n[\cdot, F(\cdot, \theta^*)]$. The estimator of $G_n[\cdot, F(\cdot, \theta^*)]$ is $G_n(\cdot, F_n)$ where $F_n(x) = F(x, \theta_n)$ and $\theta_n$ is a $n^{1/2}$-consistent estimator of $\theta^*$ when $\theta^* \in \Theta_0$. The idea is that $\theta_n$ has a high probability of being close to $\theta^*$ when $H_0$ is true. Therefore, $F_n$ is close to $F(\cdot, \theta^*)$. The bootstrap estimator of the $\alpha$-level Type I critical value for $|T_n|$, $z_{n\alpha}^*$, solves $G_n(z_{n\alpha}^*, F_n) - G_n(-z_{n\alpha}^*, F_n) = 1-\alpha$. Thus, the bootstrap estimator of the Type I critical value is, in fact, the exact Type I critical value at $\theta_n$.

Usually, $G_n(\cdot, F_n)$ and $z_{n\alpha}^*$ cannot be evaluated analytically. They can, however, be estimated with arbitrary accuracy by carrying out a Monte Carlo experiment in which random samples are drawn from $F_n$. Although the bootstrap is usually implemented by Monte Carlo simulation, its essential characteristic is the use of $F_n$ to approximate $F$ in $G_n[\cdot, F(\cdot, \theta^*)]$, not the method that is used to evaluate $G_n(\cdot, F_n)$. From this perspective, the bootstrap is an analog estimator in the sense of Manski (1988); it simply replaces the unknown $F$ with the sample analog $F_n$.

The bootstrap provides a good approximation to $G_n[\cdot, F(\cdot, \theta^*)]$ and $z_{n\alpha}$ if $n$ is sufficiently large. This is because under mild regularity conditions, $\sup_x|F_n(x) - F(x, \theta^*)|$ and $\sup_\tau|G_n(\tau, F_n) - G_n[(\tau, F(\cdot, \theta^*)]|$ converge to zero in probability or almost surely. Of course, first-order asymptotic distribution theory also provides a good approximation if $n$ is sufficiently large. It turns out, however, that if $T_n$ is asymptotically pivotal and certain technical conditions are satisfied, the bootstrap approximations are more accurate than those of first-order asymptotic theory. See Beran (1988) and Hall (1992) for the details.

In particular, the bootstrap is more accurate than first-order asymptotic theory for estimating the distribution of a "smooth" asymptotically pivotal statistic. It can be shown that

$$P(|T_n| > z_{n\alpha}^*) = \alpha + O(n^{-2})$$

when $H_0$ is true and regularity conditions hold. (The bootstrap does not achieve the same accuracy for one-tailed tests.) Thus, with the bootstrap critical value, the difference between the true and nominal probabilities that a symmetrical test makes a Type I error is $O(n^{-2})$ if the test statistic is asymptotically pivotal. In contrast, when a critical value based on first-order asymptotic theory is used, the difference is $O(n^{-1})$. This ability to improve upon first-order asymptotic approximations makes the bootstrap an attractive method for estimating Type I critical values. Horowitz (1997) presents results of Monte Carlo experiments showing that the use of bootstrap critical values can dramatically reduce the difference between the true and nominal probability that a test makes a Type I error.

## 4. Bootstrap Critical Values when the Null Hypothesis is False

The discussion of Type I critical values up to this point has assumed that $H_0$ is true so that there is a $\theta^* \in \Theta_0$ corresponding to the true DGP. The exact, $\alpha$-level Type I critical value of a symmetrical test based on the statistic $T_n$ is the $(1 - \alpha)$ quantile of the distribution of $|T_n|$ that is induced by the DGP corresponding to $\theta = \theta^*$. When $H_0$ is false, $\theta^*$ is not in $\Theta_0$, and so there is no Type I critical value corresponding to $\theta^*$. To obtain a Type I critical value, we follow Horowitz and Savin (1998) and propose using the Type I critical value corresponding to a specific $\theta$ under $H_0$ called the *pseudo-true value*. The bootstrap estimates the exact Type I critical value at the pseudo-true value of $\theta$. Therefore, when $H_0$ is false, the bootstrap provides an empirical analog of a test based on the exact Type I critical value evaluated at the pseudo-true value of $\theta$.

The problem of choosing a critical value when $H_0$ is false does not arise with size-corrected critical values or asymptotic critical values for asymptotically pivotal test statistics. This is because size-corrected critical values and asymptotic critical values for asymptotically pivotal statistics do not depend on $\theta$, that is, do not vary over the $\theta$ values in $H_0$. Thus, the problem discussed in this section arises only in connection with higher-order approximations to the Type I critical value such as that provided by the bootstrap.

When $H_0$ is false the true parameter value $\theta^*$ is in the complement of $\Theta_0$ in $\Theta$. There is no $\theta \in \Theta_0$ that corresponds to the true DGP. What value of $\theta$ should then be used to define the Type I critical value? If $H_0$ is simple, $\Theta_0$ consists of a single point, and this point can be used to define the Type I critical value. If $H_0$ is composite, however, there are many points in $\Theta_0$, and it is not clear which of them should be used to define the Type I critical value. We now review the solution to this problem proposed by Horowitz and Savin (1998).

As has already been explained, the bootstrap estimator of the Type I critical value is obtained from the distribution whose CDF is $G_n[\cdot, F(\cdot, \theta_n)]$, where $\theta_n$ is an estimator of $\theta$. Under regularity conditions (see, e.g., White (1982) and Amemiya (1985)) $\theta_n$ converges in probability or almost surely as $n \to \infty$ to a nonstochastic limit $\theta^0$ and $n^{1/2}(\theta_n - \theta^0) = O_p(1)$. The pseudo-true value is the limit $\theta^0$ of $\theta_n$ when $\theta_n$ is restricted to $\Theta_0$. If $H_0$ is true, then $\theta^0 = \theta*$. Regardless of whether $H_0$ is true, the bootstrap computes the distribution of $T_n$ at a point $\theta_n$ that is a consistent estimator of $\theta^0$. It can be shown that when $H_0$ is false, the bootstrap provides a higher-order approximation to the Type I critical value based on $\theta^0$ (Horowitz (1994)). For the LDV models considered in this paper, the bootstrap samples are generated imposing the constraint of $H_0$ and replacing the nuisance parameters by their constrained estimates.

There are two important considerations when selecting the parameter point used to calculate the Type I critical value. The first is that the $\theta$ value used to obtain the Type I critical value coincides with $\theta*$ when $H_0$ is true and has an empirical analog that can be implemented in applications regardless of whether $H_0$ is true. Second, the resulting test has good power in comparison to alternatives when $H_0$ is false. As already explained, $\theta^0$ satisfies the first of these conditions. Moreover, convergence of the Type I critical value based on $\theta^0$ to the asymptotic critical value insures that a test based on this Type I critical value inherits any asymptotic optimality properties of a test based on the asymptotic critical value.

**5. Binary Probit Model**

This section reports the numerical performance of the bootstrap for tests of the null hypothesis that the slope parameter is zero in a simple binary probit model.

The binary probit model is

$$P(Y = 1 | X = x) = \Phi(\beta' x)$$

where $X = (X_1, X_2)'$ is a vector of explanatory variables, $\beta = (\beta_1, \beta_2)'$ is a vector of parameters, and $\Phi$ is the standard normal CDF; see Amemiya (1985). Here $X$ consists of an intercept, $X_1$, and one covariate $X_2$. The composite null hypothesis is $H_0$: $\beta_2 = 0$, $\beta_1$ being the nuisance parameter, that is, the parameter not specified by $H_0$.

The null hypothesis is tested using the LM, LR and Wald test statistics. The LM and Wald test statistics require an estimator of the asymptotic covariance matrix of the maximum likelihood (ML) estimator of $\beta$. Three consistent estimators can be constructed from the expected Hessian (EX), the observed Hessian (HS) and the outer product (OP) matrix of the score vectors, respectively. Hence, there are seven test statistics in total.

The test statistics defined as t-statistics are asymptotically distributed as standard normal. In this and the following section, it is more convenient to consider the squares of the t-statistics, that is, $T_n^2$, all of which are distributed asymptotically as a chi-square with one degree of freedom when $H_0$ is true. Hence, the asymptotic approximation to the $\alpha = 0.05$ level Type I critical value is 3.84. Moreover, the tests have the same asymptotic non-central chi-square distribution under sequences of local alternatives. Although the test statistics are asymptotically pivotal, they are not pivotal: their finite sample distributions depend on the value of the intercept $\beta_1$ under $H_0$.

The test statistics are calculated using the ML estimator. The ML estimate is 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 sample points "bad" points. The LR statistic uses both a constrained and unconstrained ML estimator. The value of the LR statistic, however, can be calculated for all sample points because the value of the likelihood function is defined at the bad points; again see Albert and Anderson (1984).

The situation is different for the LM and Wald statistics. To calculate the LM statistic only the ML estimator of $\beta_1$ subject to the constraint $\beta_2 = 0$ is needed. For the constrained ML estimator 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 a positive probability of occurring. If a bad point occurs in the Monte Carlo experiments, it is deleted and not replaced. We note that the probability of a bad point goes to zero at an exponential rate in n when $H_0$ is true.

The Wald statistic employs the unconstrained ML estimator. For this estimator there are many bad points. The procedure for detecting bad points is fairly straightforward for ungrouped designs with one covariate plus intercept. A bad point can be detected by sorting the observations by the values of the covariate. A sample point is a bad point if the first i observations are all 0's and the remaining n-i are all 1's, or the first i observations are all 1's and the remaining n-i are all 0's. Under the alternative there are 2n bad points out of $2^n$ sample points. If a bad point occurs in the experiment, it is deleted and not replaced. It is important to note that each sample has to be checked *before* calculating the ML estimate. A misleading procedure is to delete points only when the ML estimation routine fails; standard estimation routines often produce finite estimates for bad points; see Hughes and Savin (1994).

The Monte Carlo experiment consists of repeating the following steps 1,000 times for each test statistic:

1. Generate an estimation dataset of size n = 50 by random sampling from the model with $\beta_2 = 0$ with X fixed in repeated samples. Estimate the parameters of the model by the relevant ML method(s) and compute the test statistic.
2. Generate a bootstrap sample of size n = 50 by random sampling from the model with $\beta_2 = 0$ using the constrained ML estimate $\beta_1$ instead of its true value. Using this bootstrap sample, compute the test statistic.
3. Repeat step 2 above 1000 times. Estimate the 0.05 Type I critical value by the 0.95 quantile of the empirical distribution of the each test statistic. Let $z_{n0.05}$* denote the estimated Type I critical value.
4. Reject $H_0$ at the nominal $\alpha = 0.05$ level if the value of the test statistic exceeds $z_{n0.05}$*.

The powers of the tests with bootstrap-based critical values are estimated by carrying out the same steps except that the value of $\beta_2$ is set equal to a nonzero number in step 1. Note that the estimate of the Type I critical value is R/G where R is the number of rejections of $H_0$ in G non-deleted samples.

The experiment can also be used to estimate the rejection probability of the test based on the asymptotic critical value. In this case, $H_0$ is rejected at the nominal $\alpha = 0.05$ level if the test statistic exceeds the 0.95 quantile of chi-square distribution with one degree of freedom, namely 3.84.

The Monte Carlo experiment is carried out for two designs.

*Design 1*. In the first design, $\beta_1$ ranges over the interval [-1.5, 0] and $\beta_2$ over the interval [0, 0.8]. The values of $X_2$ are generated using a perfect standard normal N (0,1): $x_i = \Phi^{-1}(i/(n+1))$, i = 1, 2,…,n. For this design, the probability of a bad point is negligible.

The second column of Table 1 reports the estimated probabilities (in paratheses) of making a Type I error when the tests are based on the 0.05 asymptotic critical value 3.84. Note that the estimated rejection probabilities tend to be sensitive to the value of the nuisance parameter $\beta_1$. The largest distortions in the rejection probabilities under $H_0$ occur for the asymptotic OP LM and OP Wald tests; for $\beta_1 = -1.5$, the OP LM test massively over-rejects, and the OP Wald substantially under-rejects. The asymptotic critical value works satisfactorily for the EX LM and LR tests. For the remaining tests, the asymptotic critical works satisfactorily except when $\beta_1 = -1.5$.

The remaining columns of Table 1 report the empirical rejection probabilities of the tests with bootstrap-based critical values, both under $H_0$ and the alternative hypotheses. The bootstrap eliminates the distortions in the rejection probabilities under $H_0$ except for OP LM test when $\beta_1 = -1.5$. Hence, the bootstrap provides good control over the probability of making a Type I error.

The empirical powers are very sensitive to the value of the nuisance parameter; the power tends to be largest when $\beta_1 = 0.0$ and smallest when $\beta_1 = -1.5$. The powers are similar for all the tests, except for the OP LM test and the OP Wald test. In particular, when $\beta_1 = -1.5$, the estimated powers of these two OP-based tests are lower than that of the LR test. This is due, at least in part, to the fact that the bootstrap does not fully correct the distortions in the rejection probabilities under $H_0$ of these OP-based tests. For example, when $\beta_1 = -1.5$, the estimated rejection probability under $H_0$ is 0.021 for the OP LM test and 0.055 for the LR test. The estimated power functions of the OP LM and LR tests are shown in Figure 1. The figure shows the lower power of the OP LM test compared to the LR test when $\beta_1 = -1.5$

Table 2 reports the average bootstrap critical value for the 1000 Monte Carlo samples for each pair of parameter values. The average bootstrap critical values for the OP LM and OP Wald tests are very sensitive to the true parameter values. Under $H_0$, the average critical value depends on the value of the intercept; the average critical value ranges from about 3.9 to 23 for the OP LM test and about 2.3 to 3.7 for the OP Wald test. When $H_0$ is false, the average critical value varies across the values of the slope, for a given value of the intercept. For example, for the OP LM test, the average critical value ranges from 12 to 22 when $\beta_1 = -1.5$. This is in sharp contrast to the behavior of the Type I critical value used in a conventional Monte Carlo study: for a given value of $\beta_1$, the Type I critical value is constant across alternative values of $\beta_2$.

In the introduction, we noted that the powers produced by the conventional Monte Carlo study could differ substantially from those produced by a study that uses bootstrap-based critical values. This is now illustrated for the case of the OP LM test. Assuming that the value of $\beta_1$ is known, the Type I critical value can be calculated. The Type I critical value is 25 for the OP LM test when $\beta_1 = -1.5$. Using this critical value, the power of the test increases from 0.050 to 0.066 as the alternative $\beta_2$ increases from 0.0 to 0.8. Using bootstrap-based critical values, the power of the test increases from 0.021 to 0.56 as $\beta_2$ increases from 0.0 to 0.8. For this example, the powers of the bootstrap-based OP LM test are dramatically higher than those produced by the conventional Monte Carlo study.

The behavior of the average bootstrap critical values for the LR test differs markedly from those of the OP LM test. This is shown graphically in Figure 2. For the LR test, the average bootstrap critical values are close to 4.0, both when $H_0$ is true and when it is false. This reflects the fact that for Design 1 the LR test statistic is almost pivotal. Hence, for the LR test, the powers produced by the conventional Monte Carlo study are very close to those obtained using bootstrap-based critical values.

*Design 2*. In the second design, $\beta_1$ ranges over the interval [-1.5, 0] and $\beta_2$ over the interval [0, 0.8]. The values of $X_2$ are now generated using a perfect uniform [2, 4] : $x_i = 2 + 2(i-1)/(n-1)$, i = 1,2, …,n. For this design, the power functions of the test statistics are nonmonotonic; in particular, the power first increases and then decreases as the value of $\beta_2$ increases. Savin and Würtz (1999b) show that the power function is nonmonotonic for a large class of tests when all values of $X_2$ are positive.

The second column of Table 3 reports the estimated probabilities of making a Type I error for the tests with asymptotic critical values. Again, the asymptotic critical value of 3.84 does not work satisfactorily for most of the tests when $\beta_1 = -1.5$. In particular, for Design 2, the LR test signficantly over-rejects when $\beta_1 = -1.5$. As in the case of Design 1, the largest distortions in the rejection probabilities under $H_0$ occur for the OP LM and OP Wald tests.

The estimated rejection probabilities of the bootstrap tests are also given in Table 3. The bootstrap eliminates the distortions in the rejection probabilities under $H_0$ except for OP LM and OP Wald tests when $\beta_1 = -1.5$. In contrast to Design 1, the power tends to be largest when $\beta_1 = -1.5$ and smallest when $\beta_1 = 0$. In general, the powers tend to be much lower for Design 2 than Design 1. This is partly due to the nonmonotonicity of the power functions. The estimated power functions of the OP LM and LR tests are shown in Figure 3. The nonmonotonicity is most clearly seen in the figure when $\beta_1 = 0$. Comparing the powers across the tests, the OP-based tests have lower powers than the other tests at $\beta_2 = 0.60$ and 0.80 when $\beta_1 = 0$. The EX and HS Wald tests also have nonmonotonic powers at $\beta_2 = 0.80$ when $\beta_1 = 0$.

Table 4 presents the average bootstrap critical values. As the case of the Design 1, there is substantial variation in the average critical value for the OP LM test. But in contrast to Design 1, the average bootstrap critical values do not always decrease as $\beta_2$ increases from 0.0 to 0.8. For example, the average bootstrap critical for the OP LM test increases from 3.9 to 35 as $\beta_2$ increases from 0.0 to 0.8 when $\beta_1 = 0$. This dramatic increase is shown in Figure 4. This because the distribution of the OP LM test statistic evaluated at the estimate of the pseudo-true value is shifting rapidly to the right as $\beta_2$ increases; see Savin and Würtz (1999a). By comparison, the average bootstrap critical value for the LR tests increases only from 4.0 to 4.6 as $\beta_2$ increases from 0.0 to 0.8 when $\beta_1 = 0$. Again see Figure 4

From the behavior average bootstrap critical values for the OP LM test, it is clear that the conventional Monte Carlo study will produce substantially different powers from a study using bootstrap-based critical values. Note further that in the case of Design 2, there is also substantial variation in the average bootstrap critical value for all the Wald tests, even though the variation is not as pronounced as in the case of the OP LM test.

## 6. Censored Normal Linear Regression Model

This section reports the numerical performance of the bootstrap for tests of the null hypothesis that the slope parameter is zero in a simple censored normal linear regression model.

The censored regression model is

$$Y = \max(0, X'\beta + U)$$

where $X = (X_1, X_2,)'$ is a vector of explanatory variables, $\beta = (\beta_1, \beta_2, )'$ is a vector of parameters, and $U$ is $N(0, \sigma^2)$; this is the standard Tobit model in Amemiya (1985). Here $X$ consists of an intercept, $X_1$, and one covariate $X_2$. The composite null hypothesis is $H_0$: $\beta_2 = 0$, $\beta_1$ and $\sigma^2$ being the nuisance parameters.

*Design 3.* In this design, the intercept varies, but not the variance: $\beta_1$ ranges over the interval [-0.75, 0.75], $\beta_2$ over [0, 0.8] and $\sigma^2 = 1$. The values of $X_2$ are generated using a perfect standard normal N (0,1): $x_i = \Phi^{-1}(i/(n+1))$, i = 1,2,…,n.

The experiment with Design 3 was carried out for six of the seven tests. Under the alternative it often happens that the Hessian matrix evaluated at the constrained estimate is not positive definite. Bera and McKenzie (1986) also encountered this problem. For this reason, we do not consider the HS LM test further in the censored normal regression model.

The estimated rejection probabilities in the second column of Table 5 show that the asymptotic critical value does not work satisfactorily for the OP-based tests when $\beta_1$ = -0.75 and –0.50 and the EX Wald test when $\beta_1$ = -0.75. The largest distortions in the rejection probabilities under $H_0$ occur for the OP LM and OP Wald tests. Recall that this was also true for Design 2.

The results in Table 5 also show that the bootstrap eliminates the distortions in the rejection probabilities under $H_0$ except in the case of the OP Wald test. The powers of the tests are sensitive to the value of $\beta_1$. The power against the alternative $\beta_2$ = 0.40 varies from about 0.32 to 0.70as $\beta_1$ varies from –0.75 to 0.75. For Design 3, the powers of the EX Wald and OP Wald tests are lower than the powers of the other tests when $\beta_1$ = -0.75. Aside from this case, the power functions of the tests are similar.

Table 6 presents the average bootstrap critical values. For each of the test, there is relatively little variation in the average critical value, except for the OP LM and OP Wald tests. Figure 5 shows the average critical values for the OP LM test. The same figure also shows that there is relatively little variation in the case of the LR test.

For Design 3, the powers produced by the conventional Monte Carlo study are not very different from those using bootstrap-based critical values, even for the OP LM and OP Wald tests. For example, the Type I critical value is 5.1 for the OP LM test when $\beta_1$ = -0.75. Using this critical value, the powers of the OP LM test are 0.05, 0.18, 0.52, 0.82 and 0.96 as $\beta_2$ increases from 0.0 to 0.8. These are somewhat higher than the powers reported in Table 5 for this case.

*Design 4.* The intercept is fixed and the variance varies: $\beta_1$ =-0.25, $\beta_2$ ranges over the interval [0, 0.8] and $\sigma^2$ over the interval [0.25, 2]. The values of $X_2$ are generated using a perfect normal N (0.5 , 1): $x_i = \Phi^{-1}(i/(n+1)) + 0.5$ , i = 1,2,…,n, to avoid too high a degree of censoring. The experiment with Design 4 was carried out for six of the seven tests. Again, the

HS LM test was not considered for the reason discussed under Design 3.

Column two of Table 7 reports the estimated probabilities of making a Type I error with asymptotic critical values. The effect of the nuisance parameter $\sigma^2$ is different for different tests. The asymptotic critical value does not work satisfactorily for the OP LM test and the Wald tests when $\sigma^2 = 0.25$ and for the LM tests and the LR test when $\sigma^2 = 2.0$. Note that for Design 4, the LR test over-rejects when $\sigma^2 = 1.0$ and 2.0. The largest distortions in the rejection probabilities under $H_0$ occur for the OP LM and OP Wald tests. Recall that this also true for Design 2.

The estimated rejection probabilities of the bootstrap tests are also given in Table 7. The bootstrap eliminated the distortions in the rejection probabilities under $H_0$ except for $\sigma^2 = 2.0$. The powers are very sensitive to the value of $\sigma^2$. The power against the alternative $\beta_2 = 0.20$ varies from about 1 to 0.1 as $\sigma^2$ varies from 0.25 to 2.0. For Design 4, all the tests have very similar power functions.

Table 8 presents the average bootstrap critical values. There is substantial variation in the average critical value for the OP LM test and the HS and OP Wald tests when $H_0$ is true. Figure 6 shows the average bootstrap critical values for the OP LM and LR tests. Note that for both tests there is relatively little variation in the average critical value when $H_0$ is false; the latter is true for the other tests as well.

We also note that for the OP LM test, the true Type I error is 7.4 when $\sigma^2 = 0.25$. This implies that the conventional Monte Carlo study will produce somewhat lower powers than a study using bootstrap-based critical values for the alternatives considered in the table.

## 7. Discussion

For the simple limited dependent variable models considered in this paper, the bootstrap provides better estimates of the Type I critical values of LM, LR and Wald tests than first-order asymptotic theory. The bootstrap substantially reduces or essentially eliminates the distortions in the probability of making a Type I error. The bootstrap estimates of the Type I critical values are empirically relevant since they can be calculated in empirical applications. The same is not true for the so-called “size-corrected” critical values calculated in most Monte Carlo studies, which are in fact Type I critical values for simple null hypotheses. The Type I critical value for a composite hypothesis depends on the value of the parameter vector $\theta$, both when the null is true (except if the test statistic is pivotal) and when the alternative is true. This is mirrored by the behavior of the bootstrap estimate of the Type I critical value: the estimate also depends on the value of $\theta$ under the null and the alternative. This point is illustrated by showing that the average bootstrap critical value depends on the value of $\theta$ in our Monte Carlo results. This behavior is in sharp contrast to that of size-corrrected critical values, which are invariant to the value of $\theta$ for a given value of the nuisance parameter.

We noted in the introduction that size-corrected tests have two fundamental problems. Both problems are the consequence of having a null-hypothesis set $\Theta_0$ that is too large. Dufour (1997) has shown that this happens in a wide variety of settings that are important in applied econometrics. Thus, even if computation of a size-corrected critical value is not an issue, size-corrected tests are available only when $\Theta_0$ is a sufficiently small set. Usually, this is accomplished by restricting $F(\cdot, \theta)$ to a suitably small, finite-dimensional family of functions. In applications, however, there is usually little justification for assuming such restrictions. This is reflected in the recent emphasis on semiparametric and nonparametric models and methods in econometrics, especially in the case of LDV models; for example, see Horowitz (1998a) and Powell (1994). Hence, we believe it is unlikely that size-corrected tests will play an important role in the future in testing of LDV models.

As also noted in the introduction, the parametric model LDV models considered in this paper satisfy the standard conditions under which the bootstrap provides improved estimates of the critical values. In particular, the test statistics can be approximated as smooth functions of sample moments. This is generally not the case for semiparametric and nonparametric models. The estimators and test statistics for these models typically depend on a bandwidth that decreases to zero as the sample size increases. Developing bootstrap methods for statistics that are not smooth functions of sample moments, even approximately, is a current area of research (Horowitz (1998b)). This research promises to provide improved estimates of critical values and hence empirically relevant power comparisons for semiparametric and nonparametric models.

**References**


Albert A., Anderson, J. A. 1994. On the existence of maximum likelihood estimates in logistic regression models, Biometrika 71, 1-10.

Amemiya, T., 1985. Advanced Econometrics. Harvard University Press, Cambridge.

Bahadur, R.R., Savage, L.J.,1956. The nonexistence of certain statistical procedures in nonparametric problems. Annals of Mathematical Statistics 27, 1115-1122.

Bera, A. K., McKenzie, C., 1986. Alternative forms and properties of the score test. Journal of Applied Statistics, 13, 13-25.

Beran, R., 1988. Prepivoting test statistics: a bootstrap view of asymptotic refinements. Journal of the American Statistical Association 85, 687-697.

Davidson, R., MacKinnon, J.G., 1984. Convenient specification tests for logit and probit models. Journal of Econometrics 25, 241-262.

Dufour, J.-M., 1997. Some impossibility theorems in econometrics with applications to structural and dynamic models. Econometrica 65, 1365-1387.

Hall, P., 1992. The Bootstrap and Edgeworth Expansion. Springer-Verlag, New York

Horowitz, J.L., 1994. Bootstrap-based critical values for the information matrix test. Journal of Econometrics 61, 395-411.

Horowitz, J.L., 1997. Bootstrap methods in econometrics: theory and numerical performance. In: Kreps, D.M., Wallis, K.F. (Eds.), Advances in Economics and Econometrics: Theory and Applications, vol. 3, pp. 188-222. Cambridge University Press, New York.

Horowitz, J.L., 1998a. The bootstrap. In: Heckman, J.J., Leamer,E.E.(Eds.), Handbook of Econometrics, vol 5. Elsevier Science, Amsterdam. Forthcoming.

Horowitz, J.L., 1998b.Bootstrap methods for median regression models. Econometrica, forthcoming

Horowitz, J.L., Savin, N.E., 1998. Empirically relevant critical values for hypothesis tests: a bootstrap approach. Journal of Econometrics. Forthcoming.

Hughes, G.A., Savin, N.E. 1994. Is the minimum chi-square estimator the winner in logit regression? Journal of Econometrics 61, 345-366.

Manski, C.F., 1988. Analog Estimation Methods in Econometrics. Chapman and Hall, New York.

Powell, J.L., 1994. Estimation of Semiparatmetric Models. In: Engle, R.F., McFadden, D. L., (Eds.), Handbook of Econometrics, vol 4. Elsevier Science, Amsterdam.

Savin, N. E., Würtz, A. H., 1999a. The power of LM tests in logit models: size-corrected versus Type I critical values. Department of Economics, University of Iowa, Iowa City, Iowa.

Savin, N. E., Würtz, A. H., 1999b. Power of tests in binary response models. Econometrica 67, 413- 421.

White, H., 1982. Maximum likelihood estimation of misspecified models. Econometrica 50, 1-26.

**Table 1**. Empirical rejection probabilities (percent): probit with normal regressor $X_2$ (Design 1)

| $\beta_1$ | $\beta_2$ | | | | | |
|---|---|---|---|---|---|---|
| | .00 | | .20 | .40 | .60 | .80 |
| | EX LM | | | | | |
| -1.5 | ( 4.4) | 4.6 | 8.8 | 24 | 50 | 77 |
| -1.0 | ( 5.8) | 5.5 | 14 | 38 | 70 | 91 |
| -.50 | ( 5.7) | 5.8 | 20 | 50 | 81 | 96 |
| .00 | ( 5.0) | 4.7 | 20 | 55 | 87 | 98 |
| | HS LM | | | | | |
| -1.5 | ( 3.1*) | 4.6 | 9.4 | 25 | 50 | 77 |
| -1.0 | ( 5.3) | 5.8 | 14 | 38 | 70 | 91 |
| -.50 | ( 5.7) | 5.8 | 19 | 51 | 82 | 96 |
| .00 | ( 5.1) | 4.8 | 20 | 55 | 86 | 98 |
| | OP LM | | | | | |
| -1.5 | (27*) | 2.1* | 3.1 | 12 | 30 | 56 |
| -1.0 | (9.8*) | 4.2 | 12 | 34 | 64 | 87 |
| -.50 | (6.5) | 5.2 | 19 | 49 | 82 | 96 |
| .00 | (5.2) | 4.5 | 19 | 54 | 86 | 97 |
| | EX WALD | | | | | |
| -1.5 | ( 1.9*) | 5.3 | 10 | 27 | 52 | 78 |
| -1.0 | ( 4.5) | 6.0 | 15 | 39 | 70 | 91 |
| -.50 | ( 5.5) | 5.7 | 19 | 51 | 82 | 97 |
| .00 | ( 4.7) | 4.7 | 20 | 55 | 87 | 98 |
| | HS WALD | | | | | |
| -1.5 | ( 1.9*) | 4.6 | 10 | 26 | 53 | 78 |
| -1.0 | ( 4.4) | 5.9 | 14 | 39 | 70 | 91 |
| -.50 | ( 5.4) | 5.7 | 19 | 51 | 82 | 96 |
| .00 | ( 4.8) | 4.8 | 20 | 55 | 87 | 98 |
| | OP WALD | | | | | |
| -1.5 | ( .93*) | 4.1 | 8.5 | 20 | 42 | 66 |
| -1.0 | ( 3.2*) | 6.0 | 13 | 36 | 66 | 88 |
| -.50 | ( 4.9) | 5.6 | 19 | 49 | 81 | 95 |
| .00 | ( 4.8) | 4.9 | 20 | 53 | 81 | 97 |
| | LR | | | | | |
| -1.5 | ( 6.3) | 5.5 | 9.0 | 24 | 50 | 76 |
| -1.0 | ( 6.3) | 5.6 | 14 | 38 | 70 | 91 |
| -.50 | ( 6.0) | 5.7 | 19 | 51 | 82 | 96 |
| .00 | ( 5.3) | 4.7 | 20 | 55 | 86 | 98 |

*Notes*: nominal .05 symmetric tests of $H_0$: $\beta_2 = 0$ using bootstrap-based critical values; n = 50. The number of Monte Carlo replications is 1,000, and the number of bootstrap replications is 1,000. The numbers in parentheses are the empirical rejection probabilities for a nominal .05 test using the asymptotic critical value. The asterisk denotes rejection of the null that the nominal rejection probability is .05 using a .05 symmetric asymptotic test.

**Table 2**. Average bootstrap critical values: probit with normal regressor $X_2$ (Design 1)

| $\beta_1$ | $\beta_2$ | | | | |
|---|---|---|---|---|---|
| | 0.0 | .20 | .40 | .60 | .80 |
| | EX LM | | | | |
| -1.5 | 3.8 | 3.8 | 3.8 | 3.8 | 3.8 |
| -1.0 | 3.9 | 3.9 | 3.9 | 3.9 | 3.9 |
| -.50 | 3.9 | 3.9 | 3.9 | 3.9 | 3.9 |
| .00 | 3.9 | 3.9 | 3.9 | 3.9 | 3.9 |
| | HS LM | | | | |
| -1.5 | 3.5 | 3.5 | 3.5 | 3.6 | 3.6 |
| -1.0 | 3.7 | 3.8 | 3.8 | 3.8 | 3.8 |
| -.50 | 3.9 | 3.9 | 3.9 | 3.9 | 3.9 |
| .00 | 3.9 | 3.9 | 3.9 | 3.9 | 3.9 |
| | OP LM | | | | |
| -1.5 | 23 | 22 | 19 | 16 | 12 |
| -1.0 | 7.4 | 7.1 | 6.3 | 5.5 | 4.9 |
| -.50 | 4.2 | 4.2 | 4.1 | 4.1 | 4.1 |
| .00 | 3.9 | 3.9 | 3.9 | 3.9 | 3.9 |
| | EX WALD | | | | |
| -1.5 | 3.0 | 3.0 | 3.2 | 3.3 | 3.4 |
| -1.0 | 3.6 | 3.6 | 3.6 | 3.7 | 3.7 |
| -.50 | 3.8 | 3.8 | 3.8 | 3.8 | 3.8 |
| .00 | 3.8 | 3.8 | 3.8 | 3.8 | 3.8 |
| | HS WALD | | | | |
| -1.5 | 2.9 | 2.9 | 3.1 | 3.2 | 3.2 |
| -1.0 | 3.5 | 3.5 | 3.6 | 3.6 | 3.7 |
| -.50 | 3.8 | 3.8 | 3.8 | 3.8 | 3.8 |
| .00 | 3.8 | 3.8 | 3.8 | 3.8 | 3.8 |
| | OP WALD | | | | |
| -1.5 | 2.3 | 2.3 | 2.5 | 2.7 | 2.8 |
| -1.0 | 3.1 | 3.2 | 3.2 | 3.3 | 3.4 |
| -.50 | 3.6 | 3.6 | 3.6 | 3.6 | 3.6 |
| .00 | 3.7 | 3.7 | 3.7 | 3.7 | 3.7 |
| | LR | | | | |
| -1.5 | 4.3 | 4.3 | 4.3 | 4.3 | 4.2 |
| -1.0 | 4.1 | 4.1 | 4.1 | 4.1 | 4.1 |
| -.50 | 4.0 | 4.0 | 4.0 | 4.0 | 4.0 |
| .00 | 4.0 | 4.0 | 4.0 | 4.0 | 4.0 |

*Notes*: See Table 1.

**Table 3**. Empirical rejection probabilities (percent): probit with uniform regressor $X_2$ (Design 2)

| $\beta_1$ | $\beta_2$ | | | | | |
|---|---|---|---|---|---|---|
| | .00 | | .20 | .40 | .60 | .80 |
| | EX LM | | | | | |
| -1.5 | ( 4.0) | 4.9 | 8.1 | 27 | 49 | 63 |
| -1.0 | ( 5.4) | 5.6 | 11 | 26 | 42 | 42 |
| -.50 | ( 5.7) | 5.1 | 11 | 23 | 28 | 27 |
| .00 | ( 5.0) | 5.0 | 9.1 | 18 | 17 | 13 |
| | HS LM | | | | | |
| -1.5 | ( 3.5) | 5.3 | 8.1 | 27 | 49 | 63 |
| -1.0 | ( 5.6) | 5.8 | 11 | 26 | 43 | 42 |
| -.50 | ( 5.7) | 5.1 | 11 | 23 | 28 | 27 |
| .00 | ( 5.1) | 4.8 | 9.1 | 19 | 17 | 12 |
| | OP LM | | | | | |
| -1.5 | (28*) | 2.0* | 7.2 | 26 | 49 | 59 |
| -1.0 | (9.9*) | 4.1 | 11 | 25 | 40 | 26 |
| -.50 | (7.1*) | 4.9 | 11 | 22 | 17 | 11 |
| .00 | (5.0) | 5.1 | 8.0 | 10 | 6.1 | 12 |
| | EX WALD | | | | | |
| -1.5 | ( 1.0*) | 4.6 | 8.8 | 27 | 49 | 64 |
| -1.0 | ( 5.1) | 6.1 | 11 | 25 | 43 | 43 |
| -.50 | ( 5.7) | 5.3 | 11 | 23 | 29 | 22 |
| .00 | ( 4.8) | 4.9 | 8.9 | 20 | 13 | 5.3 |
| | HS WALD | | | | | |
| -1.5 | ( .93*) | 4.4 | 8.8 | 27 | 49 | 64 |
| -1.0 | ( 4.5) | 6.3 | 11 | 26 | 44 | 42 |
| -.50 | ( 5.5) | 5.1 | 11 | 24 | 29 | 21 |
| .00 | ( 4.9) | 4.8 | 9.0 | 20 | 13 | 4.7 |
| | OP WALD | | | | | |
| -1.5 | ( .41*) | 2.8* | 9.6 | 28 | 50 | 58 |
| -1.0 | (3.2) | 6.0 | 11 | 26 | 40 | 27 |
| -.50 | ( 4.9) | 5.6 | 12 | 22 | 21 | 11 |
| .00 | ( 4.9) | 5.2 | 9.7 | 14 | 7.8 | 2.8 |
| | LR | | | | | |
| -1.5 | ( 8.0*) | 5.5 | 7.6 | 26 | 49 | 62 |
| -1.0 | ( 6.5) | 5.5 | 11 | 26 | 42 | 40 |
| -.50 | ( 5.9) | 5.0 | 11 | 23 | 26 | 24 |
| .00 | ( 5.3) | 4.8 | 9.1 | 18 | 14 | 11 |

*Notes*: See Table 1.

**Table 4**. Average bootstrap critical values: probit with uniform regressor $X_2$ (Design 2)

| $\beta_1$ | $\beta_2$ | | | | |
|---|---|---|---|---|---|
| | 0.0 | .20 | .40 | .60 | .80 |
| | EX LM | | | | |
| -1.5 | 3.5 | 3.9 | 3.9 | 3.9 | 3.9 |
| -1.0 | 3.8 | 3.9 | 3.9 | 3.9 | 3.7 |
| -.50 | 3.9 | 3.9 | 3.9 | 3.7 | 3.3 |
| .00 | 3.9 | 3.9 | 3.8 | 3.3 | 3.0 |
| | HS LM | | | | |
| -1.5 | 3.4 | 3.8 | 3.9 | 3.9 | 3.8 |
| -1.0 | 3.8 | 3.9 | 3.9 | 3.9 | 3.6 |
| -.50 | 3.9 | 3.9 | 3.9 | 3.6 | 3.1 |
| .00 | 3.9 | 3.9 | 3.7 | 3.1 | 3.8 |
| | OP LM | | | | |
| -1.5 | 25 | 6.6 | 4.0 | 4.0 | 5.5 |
| -1.0 | 8.6 | 4.1 | 4.0 | 5.1 | 17 |
| -.50 | 4.2 | 3.9 | 4.7 | 15 | 31 |
| .00 | 3.9 | 4.4 | 13 | 31 | 35 |
| | EX WALD | | | | |
| -1.5 | 2.7 | 3.6 | 3.8 | 3.8 | 3.7 |
| -1.0 | 3.5 | 3.8 | 3.8 | 3.7 | 3.1 |
| -.50 | 3.8 | 3.8 | 3.8 | 3.2 | 2.4 |
| .00 | 3.9 | 3.8 | 3.3 | 2.4 | 2.1 |
| | HS WALD | | | | |
| -1.5 | 2.6 | 3.6 | 3.8 | 3.8 | 3.7 |
| -1.0 | 3.4 | 3.8 | 3.8 | 3.7 | 3.0 |
| -.50 | 3.8 | 3.8 | 3.7 | 3.1 | 2.3 |
| .00 | 3.8 | 3.8 | 3.2 | 2.3 | 2.0 |
| | OP WALD | | | | |
| -1.5 | 2.0 | 3.3 | 3.8 | 3.8 | 3.4 |
| -1.0 | 3.1 | 3.7 | 3.8 | 3.5 | 2.5 |
| -.50 | 3.7 | 3.8 | 3.5 | 2.6 | 1.6 |
| .00 | 3.8 | 3.6 | 2.7 | 1.6 | 1.2 |
| | LR | | | | |
| -1.5 | 4.6 | 4.2 | 4.0 | 4.0 | 4.1 |
| -1.0 | 4.2 | 4.0 | 4.0 | 4.1 | 4.5 |
| -.50 | 4.0 | 4.0 | 4.1 | 4.4 | 4.7 |
| .00 | 4.0 | 4.0 | 4.4 | 4.7 | 4.6 |

*Notes*: See Table 1.

**Table 5**. Empirical rejection probabilities (percent): censored model with $\sigma^2$ = 1(Design 3)

| $\beta_1$ | $\beta_2$ | | | | | |
|---|---|---|---|---|---|---|
| | .00 | | .20 | .40 | .60 | .80 |
| | EX LM | | | | | |
| -.75 | (5.1) | 5.4 | 16 | 46 | 81 | 97 |
| -.50 | (5.9) | 6.1 | 20 | 53 | 87 | 99 |
| .00 | (5.9) | 5.7 | 23 | 63 | 94 | 100 |
| .50 | (6.7*) | 6.2 | 24 | 70 | 96 | 100 |
| .75 | (5.8) | 5.7 | 25 | 70 | 97 | 100 |
| | OP LM | | | | | |
| -.75 | (10*) | 5.2 | 14 | 43 | 76 | 96 |
| -.50 | (8.5*) | 5.7 | 18 | 51 | 84 | 98 |
| 00 | (7.8*) | 5.2 | 22 | 61 | 92 | 100 |
| .50 | (7.9*) | 6.2 | 23 | 67 | 95 | 100 |
| .75 | (7.5*) | 6.0 | 23 | 68 | 96 | 100 |
| | EX WALD | | | | | |
| -.75 | (6.8*) | 5.7 | 11 | 32 | 65 | 90 |
| -.50 | (4.9) | 5.7 | 16 | 50 | 83 | 97 |
| .00 | (5.9) | 5.7 | 23 | 63 | 94 | 100 |
| .50 | (6.6*) | 6.2 | 24 | 69 | 96 | 100 |
| .75 | (6.6*) | 6.0 | 25 | 70 | 97 | 100 |
| | HS WALD | | | | | |
| -.75 | (4.0) | 5.5 | 16 | 46 | 81 | 97 |
| -.50 | (4.9) | 6.1 | 19 | 53 | 86 | 98 |
| .00 | (5.9) | 5.6 | 23 | 63 | 94 | 100 |
| .50 | (6.7*) | 6.2 | 24 | 69 | 96 | 100 |
| .75 | (6.4) | 6.0 | 25 | 70 | 97 | 100 |
| | OP WALD | | | | | |
| -.75 | (2.7*) | 6.0 | 14 | 39 | 68 | 90 |
| -.50 | (3.2*) | 6.6* | 18 | 48 | 80 | 96 |
| .00 | (4.5) | 5.7 | 23 | 59 | 92 | 99 |
| .50 | (4.9) | 5.3 | 23 | 67 | 96 | 100 |
| .75 | (5.7) | 5.8 | 24 | 69 | 96 | 100 |
| | LR | | | | | |
| -.75 | (6.3) | 5.1 | 16 | 47 | 81 | 97 |
| -.50 | (6.4) | 5.8 | 20 | 53 | 86 | 98 |
| .00 | (6.0) | 5.8 | 23 | 63 | 94 | 100 |
| .50 | (6.7*) | 6.2 | 24 | 69 | 96 | 100 |
| .75 | (6.1) | 5.8 | 24 | 70 | 97 | 100 |

*Notes*: See Table 1.

**Table 6.** Average bootstrap critical values: censored model with $\sigma^2 = 1$(Design 3)

| $\beta_1$ | $\beta_2$ | | | | |
|---|---|---|---|---|---|
| | .00 | .20 | .40 | .60 | .80 |
| | EX LM | | | | |
| -.75 | 3.8 | 3.8 | 3.8 | 3.8 | 3.9 |
| -.50 | 3.9 | 3.9 | 3.9 | 3.9 | 3.9 |
| .00 | 4.0 | 4.0 | 4.0 | 4.0 | 4.0 |
| .50 | 4.0 | 4.0 | 4.0 | 4.0 | 4.0 |
| .75 | 4.0 | 4.0 | 4.0 | 4.0 | 4.0 |
| | OP LM | | | | |
| -.75 | 6.3 | 6.2 | 5.8 | 5.5 | 5.2 |
| -.50 | 5.0 | 5.0 | 4.9 | 4.9 | 4.8 |
| .00 | 4.5 | 4.5 | 4.5 | 4.5 | 4.5 |
| .50 | 4.4 | 4.4 | 4.4 | 4.4 | 4.4 |
| .75 | 4.5 | 4.5 | 4.5 | 4.5 | 4.5 |
| | EX WALD | | | | |
| -.75 | 4.3 | 4.3 | 4.1 | 3.9 | 3.8 |
| -.50 | 3.7 | 3.7 | 3.7 | 3.7 | 3.7 |
| .00 | 4.0 | 4.0 | 4.0 | 4.0 | 4.0 |
| .50 | 4.1 | 4.1 | 4.1 | 4.1 | 4.1 |
| .75 | 4.1 | 4.1 | 4.1 | 4.1 | 4.1 |
| | HS WALD | | | | |
| -.75 | 3.4 | 3.4 | 3.4 | 3.5 | 3.5 |
| -.50 | 3.6 | 3.6 | 3.7 | 3.7 | 3.7 |
| .00 | 3.9 | 3.9 | 3.9 | 3.9 | 3.9 |
| .50 | 4.1 | 4.1 | 4.1 | 4.1 | 4.1 |
| .75 | 4.1 | 4.1 | 4.1 | 4.1 | 4.1 |
| | OP WALD | | | | |
| -.75 | 2.8 | 2.8 | 2.9 | 2.9 | 3.0 |
| -.50 | 3.1 | 3.1 | 3.2 | 3.2 | 3.3 |
| .00 | 3.5 | 3.5 | 3.6 | 3.6 | 3.6 |
| .50 | 3.7 | 3.7 | 3.7 | 3.7 | 3.7 |
| .75 | 3.8 | 3.8 | 3.8 | 3.8 | 3.8 |
| | LR | | | | |
| -.75 | 4.1 | 4.1 | 4.1 | 4.1 | 4.1 |
| -.50 | 4.1 | 4.1 | 4.1 | 4.1 | 4.1 |
| .00 | 4.0 | 4.0 | 4.0 | 4.0 | 4.0 |
| .50 | 4.0 | 4.0 | 4.0 | 4.0 | 4.0 |
| .75 | 4.0 | 4.0 | 4.0 | 4.0 | 4.0 |

*Notes*: See Table 1.

**Table 7**. Empirical rejection probabilities (percent): censored model with $\beta_1$ = -0.25 (Design 4)

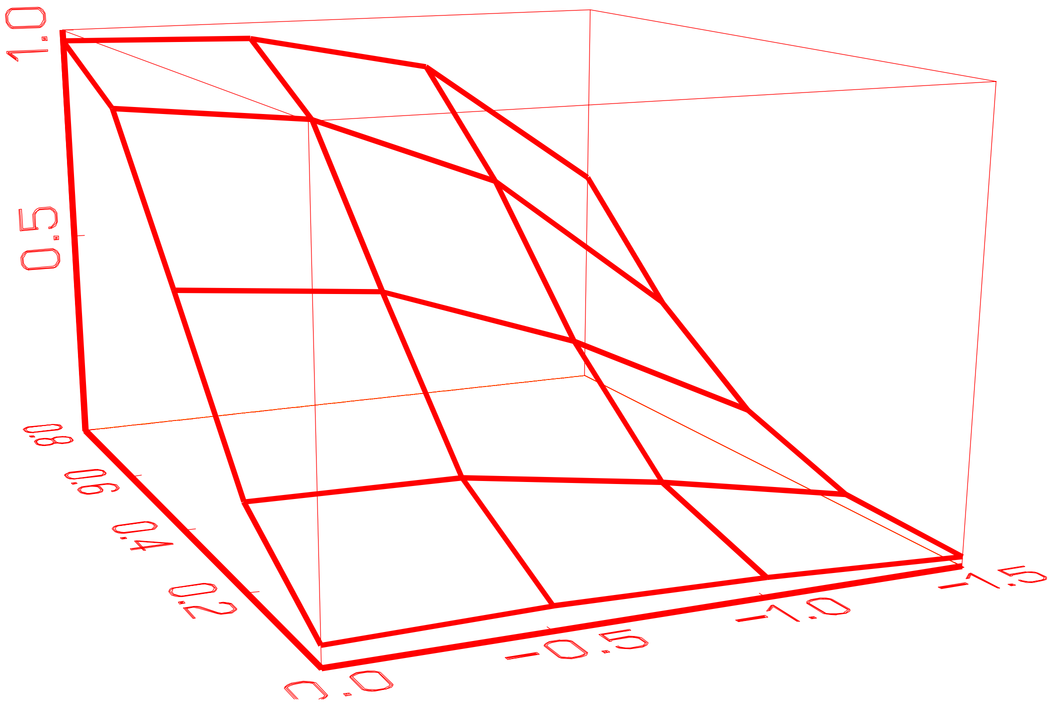
| $\sigma^2$ | $\beta_2$ | | | | | |
|---|---|---|---|---|---|---|
| | .00 | | .20 | .40 | .60 | .80 |
| | EX LM | | | | | |
| .25 | (5.0) | 5.7 | 98 | 100 | 100 | 100 |
| .50 | (5.9) | 6.1 | 58 | 100 | 100 | 100 |
| 1.0 | (6.9*) | 5.9 | 22 | 62 | 94 | 100 |
| 2.0 | (6.9*) | 6.7* | 9.3 | 23 | 42 | 64 |
| | OP LM | | | | | |
| .25 | (14*) | 3.9 | 98 | 100 | 100 | 100 |
| .50 | (8.5*) | 5.7 | 56 | 100 | 100 | 100 |
| 1.0 | (8.1*) | 5.9 | 21 | 60 | 92 | 100 |
| 2.0 | (8.6*) | 5.8 | 9.6 | 22 | 41 | 63 |
| | EX WALD | | | | | |
| .25 | (6.9*) | 4.6 | 96 | 100 | 100 | 100 |
| .50 | (4.9) | 5.7 | 58 | 100 | 100 | 100 |
| 1.0 | (6.2) | 6.1 | 23 | 62 | 94 | 100 |
| 2.0 | (6.6*) | 6.5* | 9.8 | 23 | 42 | 64 |
| | HS WALD | | | | | |
| .25 | (2.6*) | 6.4 | 98 | 100 | 100 | 100 |
| .50 | (4.9) | 6.1 | 58 | 100 | 100 | 100 |
| 1.0 | (6.1) | 6.1 | 22 | 62 | 94 | 100 |
| 2.0 | (6.4) | 6.3 | 9.8 | 23 | 41 | 64 |
| | OP WALD | | | | | |
| .25 | (1.6*) | 5.6 | 94 | 100 | 100 | 100 |
| .50 | (3.2*) | 6.6* | 53 | 99 | 100 | 100 |
| 1.0 | (4.6) | 6.0 | 21 | 58 | 92 | 100 |
| 2.0 | (5.0) | 6.5* | 9.1 | 22 | 38 | 62 |
| | LR | | | | | |
| .25 | (6.4) | 5.4 | 98 | 100 | 100 | 100 |
| .50 | (6.4) | 5.8 | 58 | 100 | 100 | 100 |
| 1.0 | (7.0*) | 6.1 | 22 | 62 | 94 | 100 |
| 2.0 | (7.2*) | 6.5* | 9.4 | 23 | 42 | 64 |

*Notes*: See Table 1.

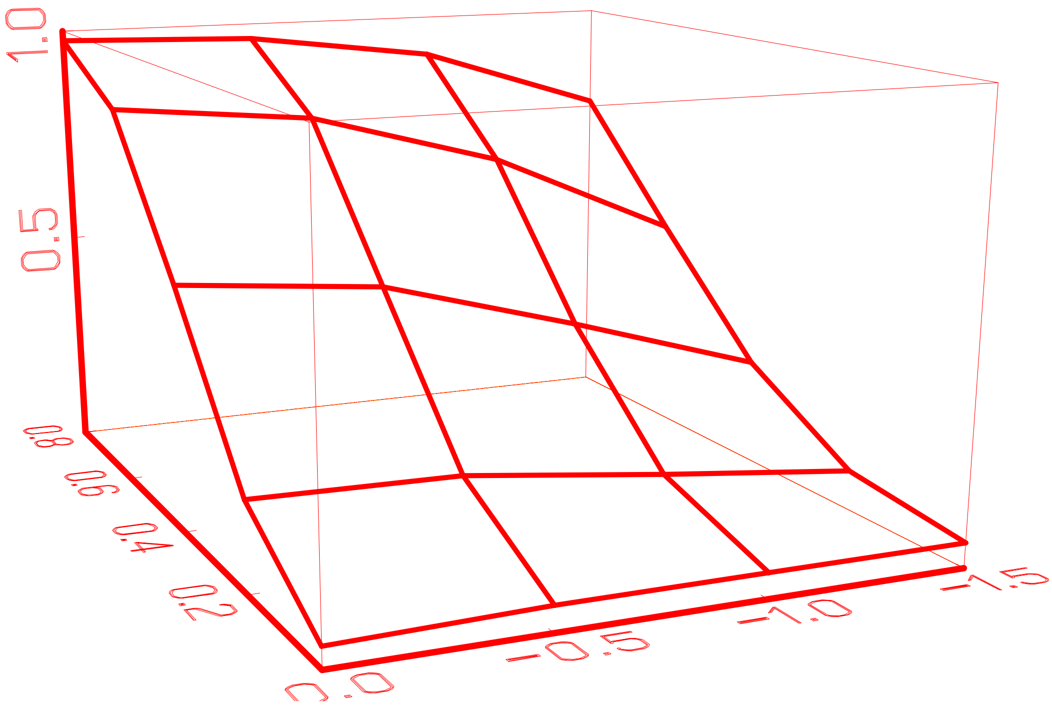
**Table 8**. Average bootstrap critical values: censored model with $\beta_1$ = -0.25(Design 4)

| $\sigma^2$ | $\beta_2$ | | | | |
|---|---|---|---|---|---|
| | 0.0 | .20 | .40 | .60 | .80 |
| | EX LM | | | | |
| .25 | 3.7 | 3.9 | 4.0 | 4.0 | 4.0 |
| .50 | 3.9 | 3.9 | 4.0 | 4.0 | 4.0 |
| 1.0 | 4.0 | 4.0 | 4.0 | 4.0 | 4.0 |
| 2.0 | 4.0 | 4.0 | 4.0 | 4.0 | 4.0 |
| | OP LM | | | | |
| .25 | 11 | 4.9 | 4.5 | 4.4 | 4.4 |
| .50 | 5.0 | 4.7 | 4.5 | 4.5 | 4.4 |
| 1.0 | 4.6 | 4.5 | 4.5 | 4.5 | 4.4 |
| 2.0 | 4.5 | 4.5 | 4.5 | 4.5 | 4.4 |
| | EX WALD | | | | |
| .25 | 5.6 | 3.7 | 3.9 | 4.0 | 4.1 |
| .50 | 3.7 | 3.8 | 3.9 | 4.0 | 4.1 |
| 1.0 | 3.8 | 3.9 | 4.0 | 4.0 | 4.0 |
| 2.0 | 3.9 | 3.9 | 4.0 | 4.0 | 4.0 |
| | HS WALD | | | | |
| .25 | 3.0 | 3.7 | 3.9 | 4.0 | 4.0 |
| .50 | 3.6 | 3.8 | 3.9 | 4.0 | 4.0 |
| 1.0 | 3.8 | 3.9 | 3.9 | 3.9 | 4.0 |
| 2.0 | 3.9 | 3.9 | 3.9 | 3.9 | 4.0 |
| | OP WALD | | | | |
| .25 | 2.3 | 3.2 | 3.5 | 3.6 | 3.7 |
| .50 | 3.1 | 3.3 | 3.5 | 3.6 | 3.6 |
| 1.0 | 3.4 | 3.4 | 3.5 | 3.6 | 3.6 |
| 2.0 | 3.5 | 3.5 | 3.5 | 3.6 | 3.6 |
| | LR | | | | |
| .25 | 4.2 | 4.1 | 4.0 | 4.0 | 4.0 |
| .50 | 4.1 | 4.0 | 4.0 | 4.0 | 4.0 |
| 1.0 | 4.0 | 4.0 | 4.0 | 4.0 | 4.0 |
| 2.0 | 4.0 | 4.0 | 4.0 | 4.0 | 4.0 |

*Notes*: See Table 1.

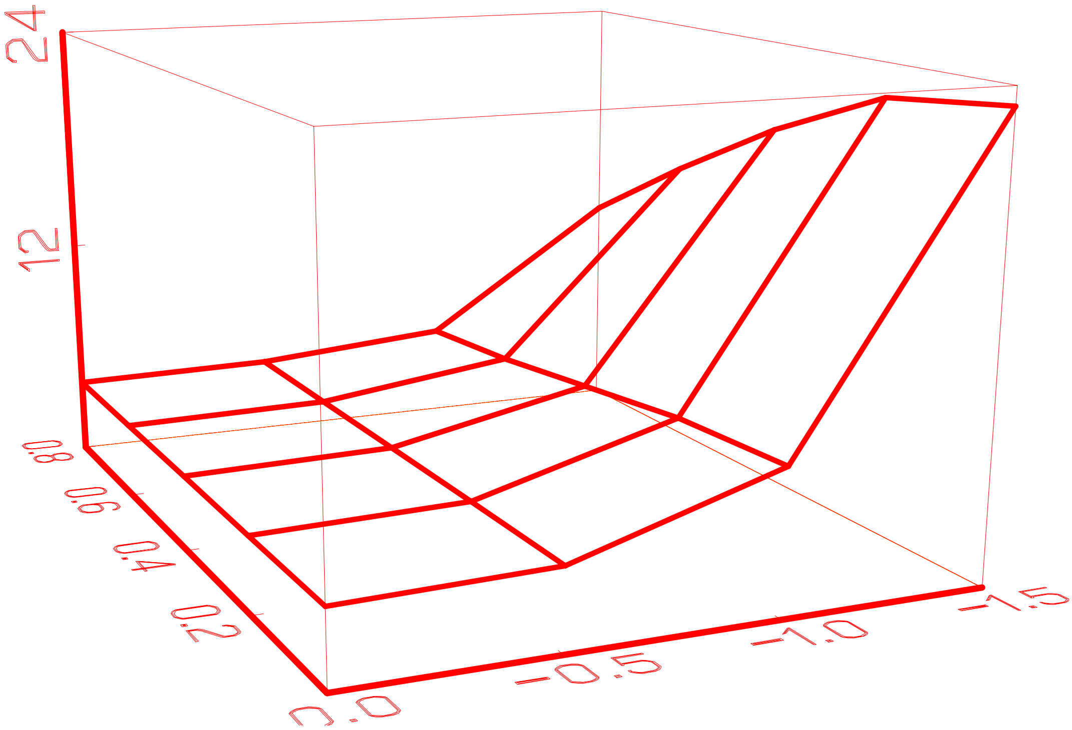



OP LM test

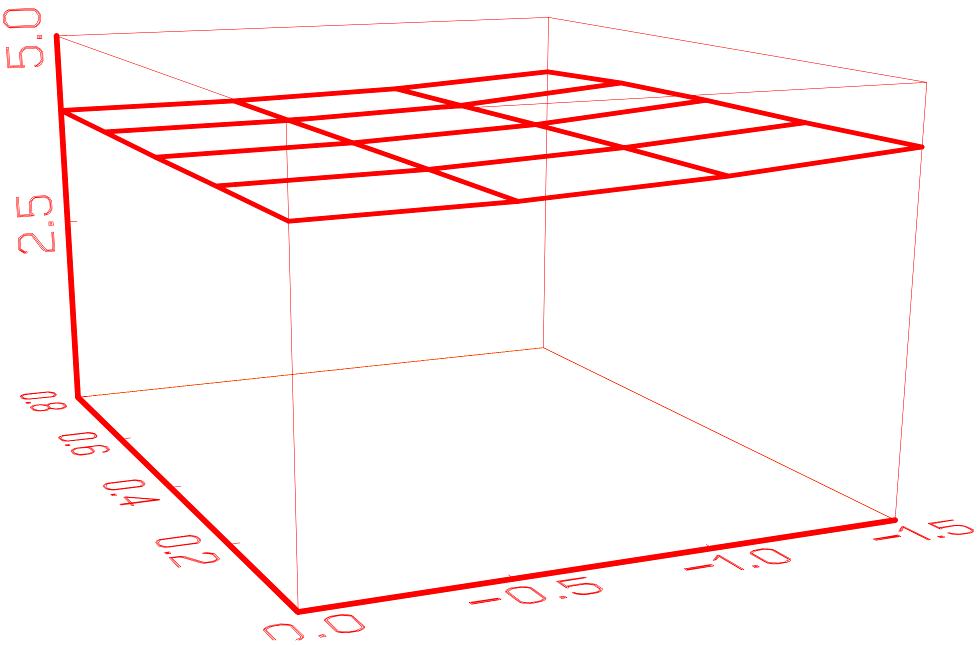



LR test

Figure 1. Powers of nominal 0.05 OP LM and LR tests in the probit model: Design 1.

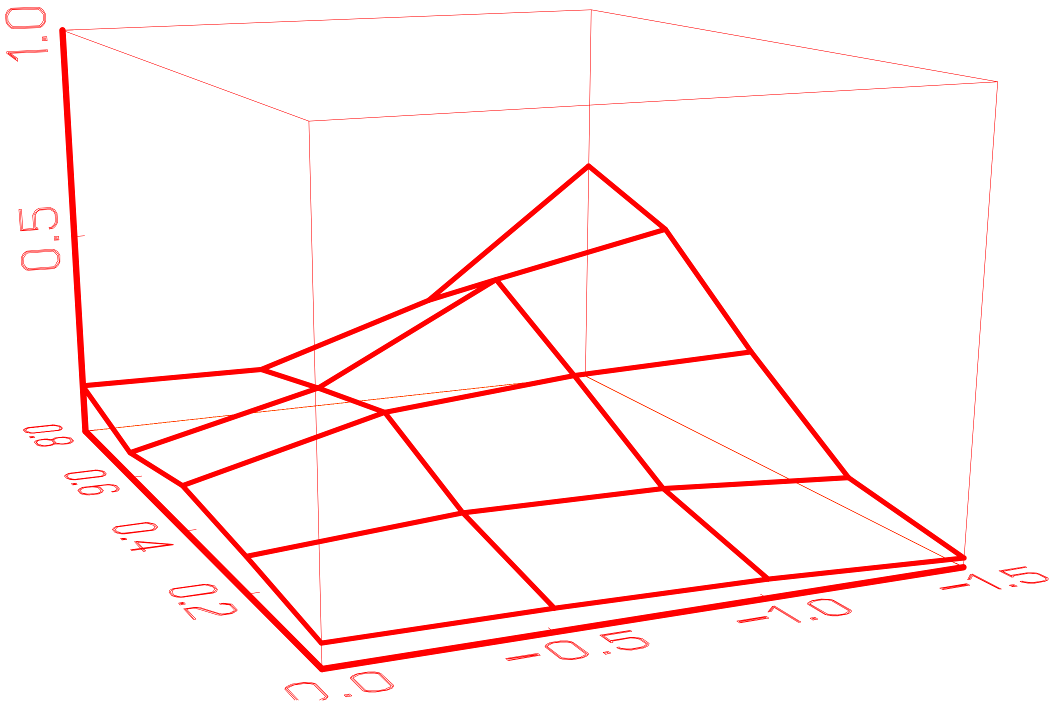



OP LM test

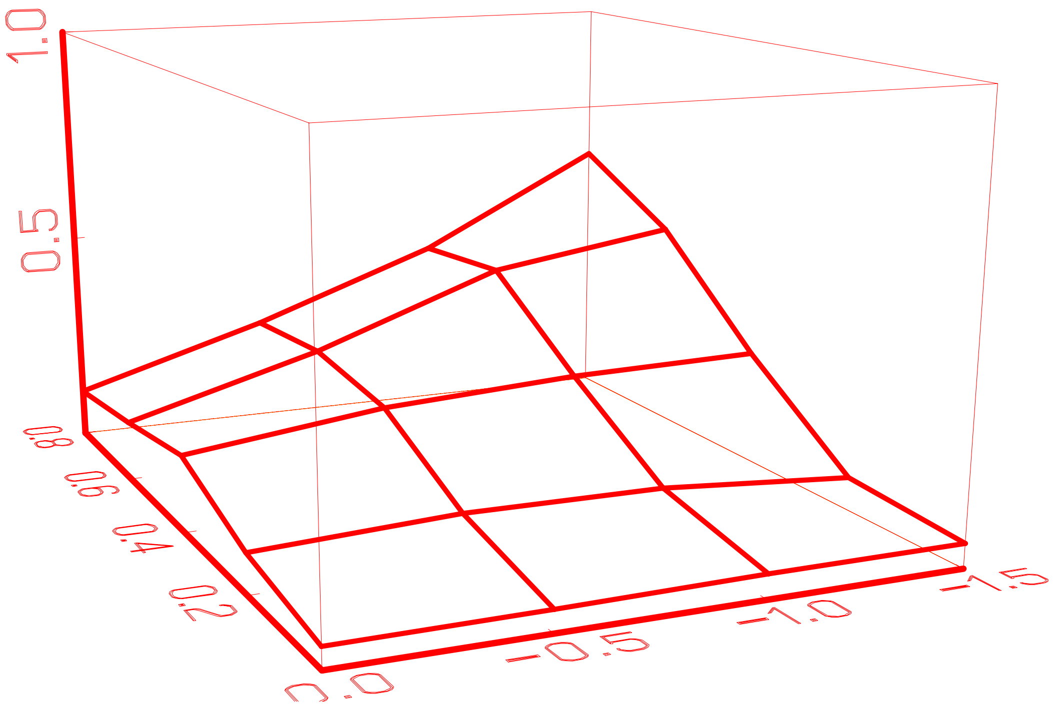



LR test

Figure 2. Average bootstrap critical values of nominal 0.05 OP LM and LR tests in the probit model: Design 1.

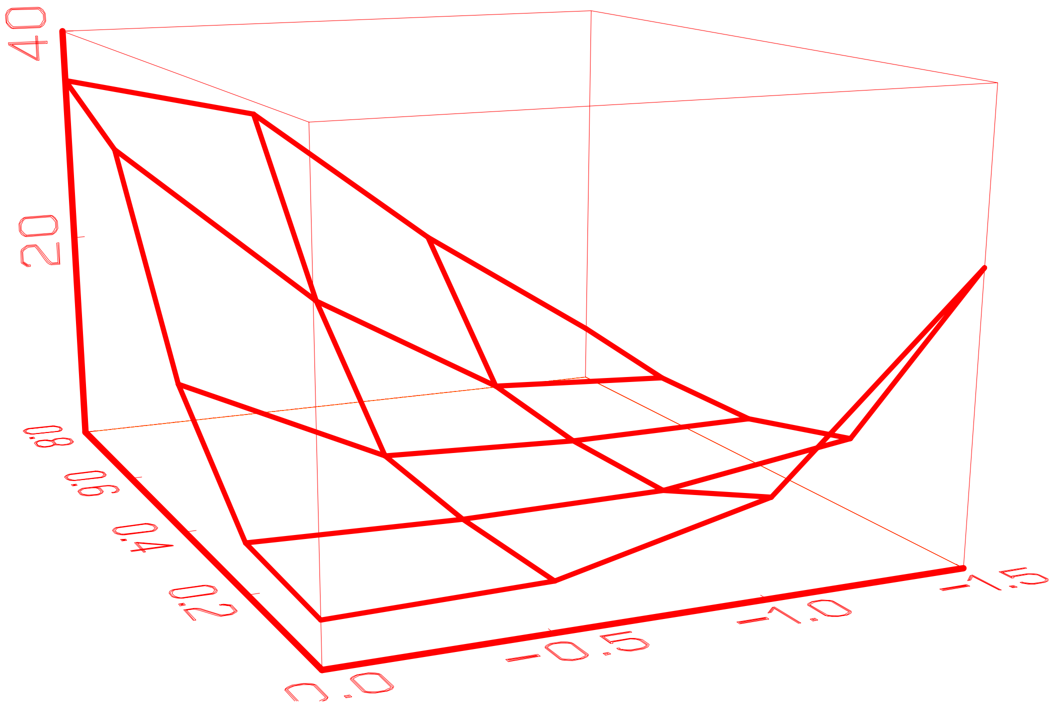



OP LM test

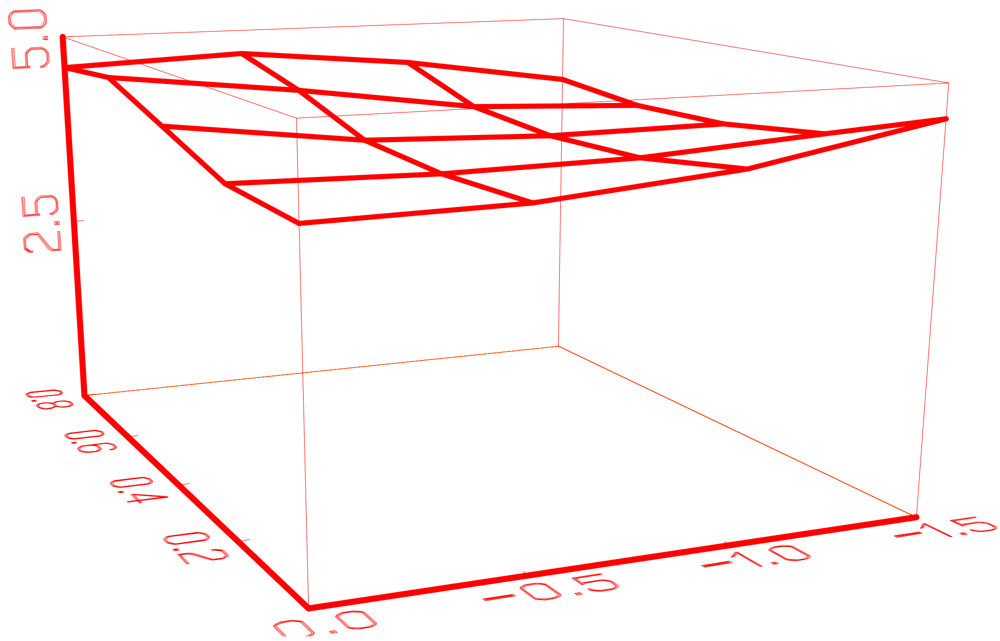



LR test

Figure 3. Powers of nominal 0.05 OP LM and LR tests in the probit model: Design 2.

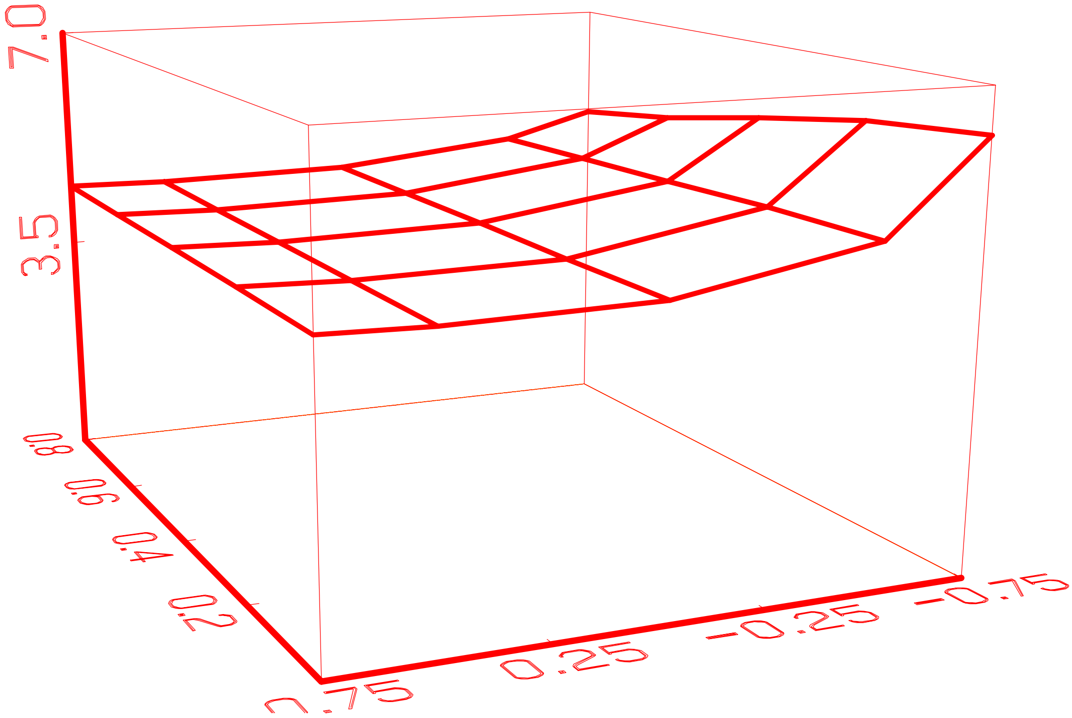



OP LM test

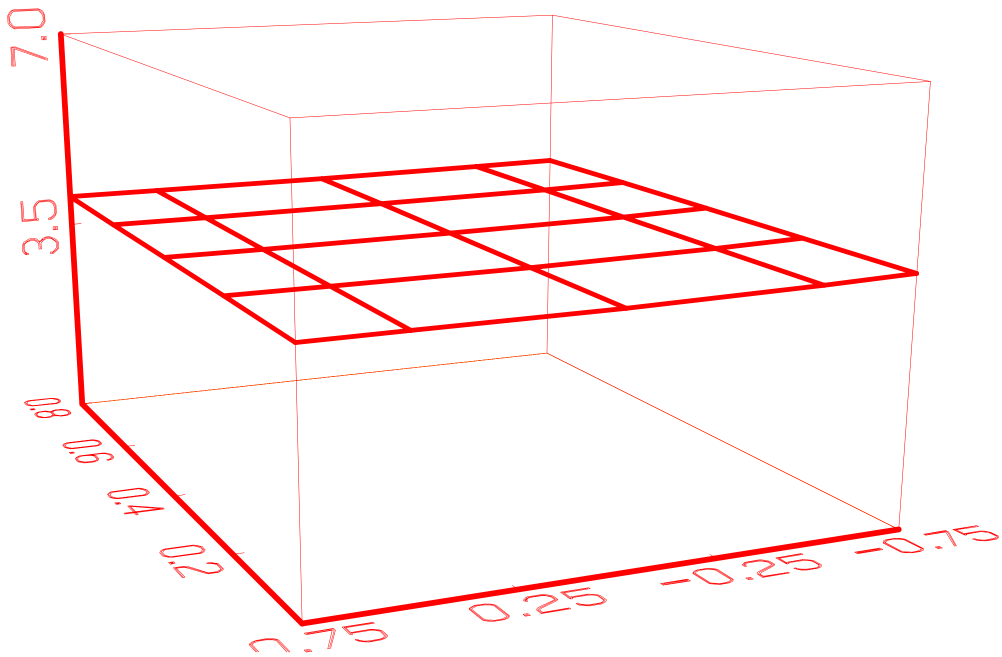



LR test

Figure 4. Average bootstrap critical values of nominal 0.05 OP LM and LR tests in the probit model: Design 2.

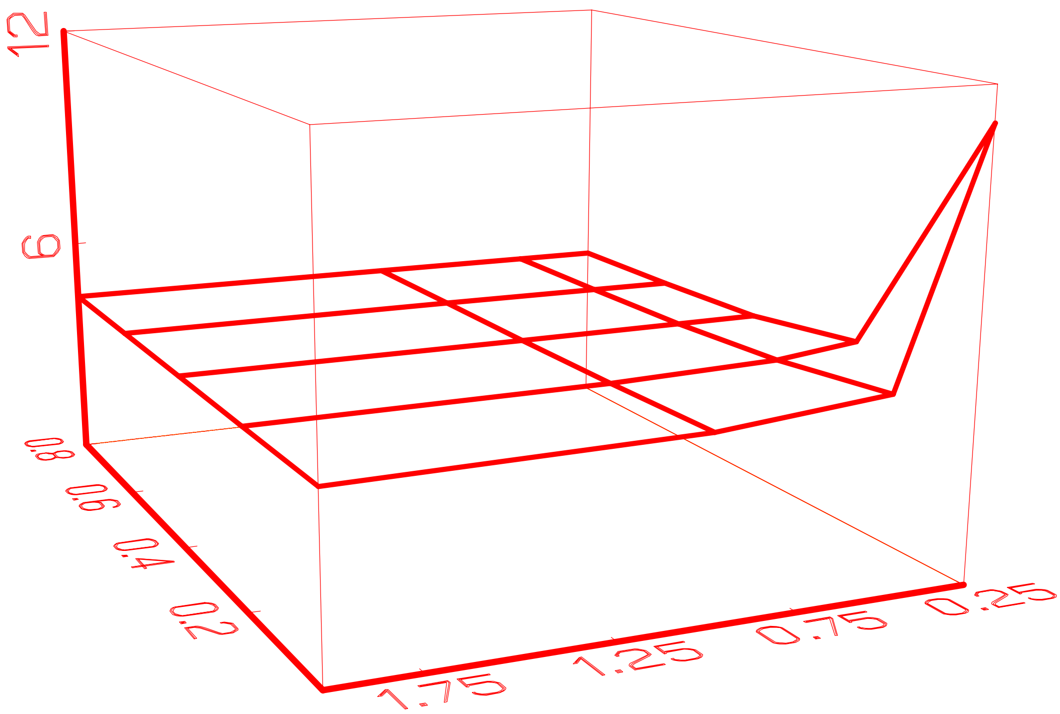



OP LM test

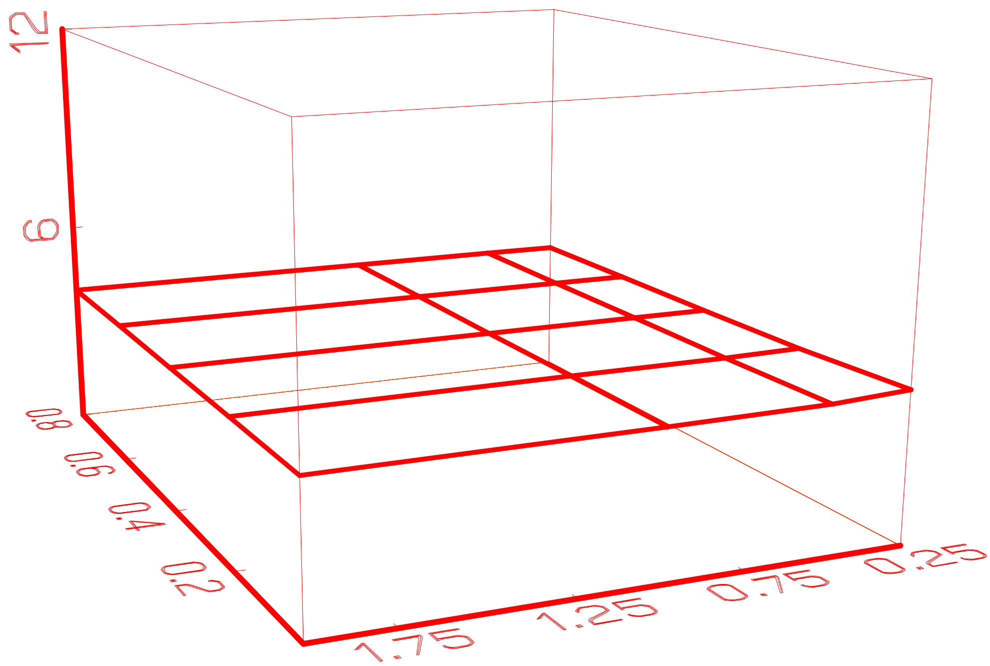



LR test

Figure 5. Average bootstrap critical values of nominal 0.05 OP LM and LR tests in the censored model: Design 3.




OP LM test




LR test

Figure 6. Average bootstrap critical values of nominal 0.05 OP LM and LR tests in the censored model: Design 4.

**Working Paper**

| | |
|---|---|
| 1999-6 | Torben M. Andersen and Eric Toulemonde: Adapting Prices or Quantities in the Presence of Adjustment Costs? |
| 1999-7 | Torben M. Andersen and Eric Toulemonde: Imperfectively Competitive Labour Markets and the Productivity Puzzle. |
| 1999-8 | Christian M. Dahl: An Investigation of Tests for Linearity and the Accurracy of Flexible Nonlinear Inference. |
| 1999-9 | Peter Nannestad and Martin Paldam: The Cost of Ruling. A Foundation Stone for Two Theories. |
| 1999-10 | Torben M. Andersen and Niels C. Beier: Propagation of Nominal Shocks in Open Economies. |
| 1999-11 | Martin Paldam: The Big Pattern of Corruption. Economics, Culture and the Seesaw Dynamics. |
| 1999-12 | Jakob Roland Munch: The Location of Firms in Unionized Countries. |
| 1999-13 | Goran Peskir and Jamsheed Shorish: Market Forces and Dynamic Asset Pricing. |
| 1999-14 | Jakob Roland Munch and Jan Rose Sørensen: Product Market Integration and Wages in Unionized Countries. |
| 1999-15 | Christian M. Dahl and Niels L. Hansen: The Formation of Inflation Expectations under Changing Inflation Regimes. |
| 1999-16 | Ebbe Yndgaard: Some Reflections on Time and Equilibrium. |
| 1999-17 | Eric Toulemonde: Deterring Entry through High Wages. |
| 1999-18 | Pierre M. Picard and Eric Toulemonde: On the Equivalence of Taxes Paid by Employers and Employees. |
| 1999-19 | Eric Toulemonde: Wage Bargaining: Reconciling Theory and Evidence. |
| 1999-20 | N.E. Savin and Allan H. Würtz: Empirically Relevant Power Comparisons for Limited Dependent Variable 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:**

1997-1 Tom Engsted, Jesus Gonzalo and Niels Haldrup: Testing for Multi-cointegration.

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 Popoulation 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.

1997-18 Tom Engsted and Niels Haldrup: Multicointegration in Stock-Flow Models.

1998-6 Svend Hylleberg and Rikke Willemoes Jørgensen: A Note on the Estimation of Markup Pricing in Manufacturing.

1998-15 Henning Bunzel, Bent Jesper Christensen, Niels Haldrup, Svend Hylleberg, Viggo Høst, Peter Jensen og Allan Würtz: Udviklingslinier i Økonometrien.

1999-3 Niels Haldrup and Michael Jansson: Spurious Regression, Cointegration, and Near Cointegration: A Unifying Approach.

1999-4 Christian M. Dahl and Svend Hylleberg: Specifying Nonlinear Econometric Models by Flexible Regression Models and Relative Forecast Performance.

1999-8 Christian M. Dahl: An Investigation of Tests for Linearity and the Accuracy of Flexible Nonlinear Inference.

1999-13 Goran Peskir and Jamsheed Shorish: Market Forces and Dynamic Asset Pricing.

1999-15 Christian M. Dahl and Niels L. Hansen: The Formation of Inflation Expectations under Changing Inflation Regimes.

1999-20 N.E. Savin and Allan H. Würtz: Empirically Relevant Power Comparisons for Limited Dependent Variable Models.