# Size And Power Property Of The Asymptotic Tests And The Bootstrap Tests In The Probit Model: Simulation Results

2006

Xiaobin Shen

*University of Central Florida*

# SIZE AND POWER PROPERTY OF THE ASYMPTOTIC TESTS AND THE BOOTSTRAP TESTS IN THE PROBIT MODEL: SIMULATION RESULTS

by

XIAOBIN SHEN
B.S. Tongji University, 1998

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science
in the Department of Economics
in the College of Business Administration
at the University of Central Florida
Orlando, Florida

Summer Term
2006

## ABSTRACT

This paper compares the size and power properties of the asymptotic tests based on the asymptotic standard errors with the bootstrap tests based on the bootstrap confidence interval in the Probit model. The asymptotic tests work surprisingly well even when the sample size is quite small (e.g., $n = 30$) for the test of exclusion hypothesis $\beta = 0$. The bootstrap tests work similarly well. It shares essentially the same size and power property of the asymptotic tests when the null hypothesis is $\beta = 0$. However, the small sample probit estimators can be seriously biased when $\beta/\sigma$ is large. Consequently, when we are interested in the non-exclusion hypothesis such as $\beta/\sigma = 1$, the conventional asymptotic tests can suffer size distortion and low power. But, following our simulation results, the size of the bootstrap tests is quite robust to the presence of the bias and the power is much better. Therefore, the bootstrap approach has some limited usefulness in practice when we are interested in the non-exclusion tests such as $\beta/\sigma = 1$ in the probit model.

# TABLE OF CONTENTS

## LIST OF TABLES

# CHAPTER ONE: INTRODUCTION

Bootstrap was introduced in 1979 as a computer-based method for estimating the standard error of $\hat{\theta}$, which is an estimator of a parameter of interest $\theta$. The reason for using bootstrap inference is that the hypothesis tests and the confidence intervals based on the asymptotic theory can be seriously misleading when the sample size is small. The bootstrap can often, but does not always, lead to much more accurate inferences than traditional approaches (MacKinnon, 2002). The bootstrap sometimes relieves the burden of complex mathematical derivations, or in some instances does provide alternative solution where no analytical answer can be obtained.

Unlike the linear regression models, asymptotic standard error is based on the first order approximation in the probit model. Consequently, the standard error might be biased when the sample size is small, which can result in the size of the test very different from the nominal size (e.g., 0.05 level). It is worth emphasizing that the probit model is fully parametric in the sense that the error is assumed to be standard normal, which provides an excellent environment for the generation of bootstrap estimators. As it shows in the following sections, the bootstrap tests work very well, producing the size of the tests very close to the nominal size. However, the asymptotic tests which are based on the asymptotic standard errors work just as good when the null hypothesis is exclusion hypothesis such as $\beta = 0$. It turns out that the size of the tests is influenced more by the bias of the estimators than by the bias in the standard errors in the probit model. The bias depends on the true parameter value of $\beta$. The bias is very close to zero when $\beta = 0$, and the size of the asymptotic tests is very close to the nominal size when the null hypothesis is $\beta = 0$. But, the bias of the probit estimator is quite serious when $\beta/\sigma = 1$ when the sample size is smaller than 50, and the tests suffer some serious size distortion and power loss. The bootstrap based tests work much better than the asymptotic tests in this environment. However, the hypothesis such as $\beta/\sigma = 1$ may not be very practical.

This paper is organized as follows. In Chapter two we review the binary probit model and the maximum likelihood estimation method, and provide a review on the existing literature on the probit simulation results. Chapter three describes the data generation process for simulation and the details of bootstrap procedure. Findings of this paper are presented in Chapter four. Chapter five concludes.

# CHAPTER TWO: LITERATURE REVIEW

## §2.1 Binary Probit Model

A binary response model is a regression model in which the dependent variable $y$ is a binary random variable that takes on only the values zero and one. Application of the binary response model is extensive in practice. For example, a commuter chooses to drive a car to work or to take public transport. Another example is the choice of a worker between taking a job or not. Driving to work and taking a job are choices that correspond to $y = 1$, and taking public transport and not taking a job to $y = 0$. The model gives the probability that $y = 1$ is chosen conditional on a set of explanatory variables. In the transportation example, common explanatory variables include the time and the cost of travel; in the worker example, common explanatory variables include age, education and experience (Horowitz and Savin, 2001).

The econometric problem is to estimate the conditional probability $y = 1$ as a function of the explanatory variables. In a binary response model, our interest lies primarily in the response probability

$$P(y=1|\mathrm{X}) = G(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k) = G(\mathrm{X}\beta), \qquad (1)$$

where we use $\mathrm{X}$ to denote the full set of explanatory variables $(x_1, x_2, \ldots, x_k)$. For simplicity, we absorb the intercept into the vector $\mathrm{X}$.

The Linear Probability Model (LPM) specifies that the conditional probability is a linear function of $\mathrm{X}$:

$$P(y=1|\mathrm{X}) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k = \mathrm{X}\beta\,. \tag{2}$$

As we know, LPM has the defect that the conditional probability is not constrained to lie between zero and one. This defect can be corrected by replacing the linear function with one with a lower kink that keeps the conditional probability from being less than zero and an upper kink that keeps it from being greater than one such as

$$P(y=1|\mathrm{X}) = G(\beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k) = G(\mathrm{X}\beta)$$

where $G(\cdot)$ denotes the any assumed cumulative distribution function (cdf). In the probit model, $G(\cdot)$ is the standard normal cdf, which is expressed as the integral:

$$G(z) = \Phi(z) \equiv \int_{-\infty}^{z} \phi(v)dv\,, \tag{3}$$

where $\phi(z)$ is the standard normal density

$$\phi(z) = (2\pi)^{-1/2}\exp(-z^2/2)\,. \tag{4}$$

This choice of $G$ ensures that (1) is strictly between zero and one for all values of the parameters and the $x_j$. The $G$ in (3) is an increasing function. It increases most quickly at $z = 0$. $G(z) \to 0$ as $z \to -\infty$, and $G(z) \to 1$ as $z \to +\infty$.

Probit model can be derived from an underlying latent variable model. Let $y*$ be an unobserved, or latent, variable, determined by

$$y* = \mathrm{X}\beta + e, \; y = I(y* > 0), \tag{5}$$

where $I(\cdot)$ is the indicator function which takes on the value one if the event in brackets is true, and zero otherwise. Therefore, $y$ is one if $y* > 0$, and $y$ is zero if $y* \leq 0$. We assume that $e$ is independent of $\mathrm{X}$ and that $e$ follows the standard normal distribution with unit variance. Since $e$ is symmetrically distributed about zero, $1 - G(-z) = G(z)$ for all real number $z$. Economists tend to favor the normality assumption for $e$, which is why the probit model is more popular than logit in econometrics. In addition, several specification problems are most easily analyzed using probit because of properties of the normal distribution. From (5) and the assumptions given, we can derive the response probability for $y$:

$$\begin{aligned} P(y = 1 \mid \mathrm{X}) &= P(y* > 0 \mid \mathrm{X}) = P[e > -\mathrm{X}\beta \mid \mathrm{X}] \\ &= 1 - G(-\mathrm{X}\beta) = G(\mathrm{X}\beta) \end{aligned},$$

which is exactly the same as (1).

We need to notice that the assumption of known variance of $e$ is an innocent normalization. Suppose the variance of $e$ is scaled by an unrestricted parameter σ. The latent variable model will be $y^* = X\beta + \sigma * e$. But, $(y^*/\sigma) = X(\beta/\sigma) + e$ is the same model with the same data. The observed data will be unchanged; $y$ is still 0 or 1, depending only on the sign of $y^*$ not on its scale. This means that there is no information about $\sigma$ in the data so it cannot be estimated.

## §2.2 MLE Procedure of Parameter Estimation

Because of the nonlinear nature of $E(y \mid X)$, ordinary least squares (OLS) and weighted least squares (WLS) cannot be used to estimate the probit models. The maximum likelihood estimation (MLE) method is indispensable. Because the maximum likelihood estimation is based on the distribution of $y$ given $X$, the heteroskedasticity in $Var(y \mid X)$ is automatically accounted for.

Assume that we have a random sample of size $n$. To obtain the maximum likelihood estimator, conditional on the explanatory variables, we need the density of $y_i$ given $X_i$. We can write this as

$$f(y \mid X_i; \beta) = [G(X_i\beta)]^y [1 - G(X_i\beta)]^{1-y}, \; y = 0, 1, \qquad (6)$$

When $y = 1$, we get $G(X_i\beta)$ and when $y = 0$, we get $1-G(X_i\beta)$. The log-likelihood function for observation $i$ is a function of the parameters and the data $(X_i, y_i)$ and is obtained by taking the log of (6):

$$\ell_i(\beta) = y_i \log[G(X_i\beta)] + (1-y_i)\log[1-G(X_i\beta)]. \quad (7)$$

Because $G(\cdot)$ is strictly between zero and one, $\ell_i(\beta)$ is well-defined for all the values of $\beta$.

The log-likelihood for a sample size of $n$ is obtained by summing (7) across all observations as

$$L(\beta) = \sum_{i=1}^{n} \ell_i(\beta).$$

The MLE of $\beta$, denoted by $\hat{\beta}$, maximizes this log-likelihood. The resulting estimator $\hat{\beta}$ is the probit estimator.

Each $\hat{\beta}_j$ comes with an asymptotic standard error. Given the binary response model $P(y=1|X) = G(X\beta)$, where $G(\cdot)$ is the probit function, and $\beta$ is the $k \times 1$ vector of parameters, the asymptotic variance matrix of $\hat{\beta}$ is estimated as

$$A\,\mathrm{var}(\hat{\beta}) \equiv \left( \sum_{i=1}^{n} \frac{[g(X_i\hat{\beta})]^2 X_i' X_i}{G(X_i\hat{\beta})[1-G(X_i\hat{\beta})]} \right)^{-1}. \quad (8)$$

It is a $k \times k$ matrix. The expression in (8) accounts for the nonlinear nature of the response probability – that is, the nonlinear nature of $G(\cdot)$ – as well as the particular form of heteroskedasticity in a binary response model: $Var(y \mid \mathrm{X}) = G(\mathrm{X}\beta)[1 - G(\mathrm{X}\beta)]$.

The square roots of the diagonal elements of (8) are the asymptotic standard errors of the $\hat{\beta}_j$, j = 1,2,…,k. They are routinely reported by econometrics software that supports the probit analysis. Once we have these, (asymptotic) *t* statistics and confidence intervals are obtained in the usual ways. See, for example, Wooldridge (2003).

## §2.3 Bootstrap Method

Classical statistical inference is predicated upon the use of a statistic *T* (using an iid sample from a population) to make inference about an unknown population characteristic. However, in a large number of instances it has been demonstrated that the use of asymptotic arguments in testing hypothesis are significantly inaccurate in finite sample.

Bootstrap was introduced in 1979 as a computer-based method for estimating the standard error of $\hat{\theta}$, which is an estimator of a parameter of interest $\theta$. The bootstrap estimate of standard error requires no theoretical calculations, and is available no matter how mathematically complicated the estimator $\hat{\theta}$ may be (Efron and Tibshirani 1993). This line of research is to base statistical inferences on the distributions that are calculated by simulation rather than on ones that are suggested by asymptotic theory and are strictly valid only when the sample size is infinitely large. In this approach, the parameter estimates and test statistics are calculated in fairly conventional way, but the p-values and the confidence intervals are computed using "bootstrap" distributions obtained by simulation. This bootstrap approach can often, but not always, lead to much more accurate inferences than the traditional approaches. See, for example, MacKinnon (2002).

Bootstrap methods depend on the notion of a bootstrap sample. Let $\hat{F}$ be the empirical distribution, putting probability 1/*n* on each of the observed values $x_i$, $i = 1, 2, \ldots, n$. A bootstrap sample is defined to be a random sample of size *n* drawn from $\hat{F}$, say

$$\mathbf{x}^* = (x_1^*, x_2^*, \ldots, x_n^*),$$

$$\hat{F} \rightarrow (x_1^*, x_2^*, \ldots, x_n^*). \tag{9}$$

The star notation indicates that $\mathbf{x}^*$ is not the actual data set **x**, but rather a randomized, or resampled, version of **x**. In other words, the bootstrap data points $x_1^*, x_2^*, \ldots, x_n^*$ are a random sample of size *n* drawn with replacement from the population of *n* objects $(x_1, x_2, \ldots, x_n)$. Thus we might have $x_1^* = x_7$, $x_2^* = x_3$, $x_3^* = x_3$, $x_4^* = x_{22}$, $\ldots$, $x_n^* = x_7$. The bootstrap data set $(x_1^*, x_2^*, \ldots, x_n^*)$ consists of members of the original data set $(x_1, x_2, \ldots, x_n)$, some appearing zero times, some appearing once, some appearing twice, etc.

There are several ways in which the bootstrap can be used for hypothesis testing. One approach, which in our view is the simplest and most satisfactory, is to use the bootstrap to compute the p-values. We first compute a test statistic, say $\hat{\tau}$, in the usual way. Using the estimates of the model under the null hypothesis, we then draw *B* bootstrap samples. Each of these is used to compute a bootstrap test statistic $\tau_j^*$ in exactly the same way that $\hat{\tau}$ was computed from the real sample. For a one-tailed test with a rejection region in the upper tail, the bootstrap p-value may then be estimated by

$$\hat{p}^*(\hat{\tau}) = \frac{1}{B}\sum_{j=1}^{B} I(\tau_j \geq \hat{\tau}), \tag{10}$$

As $B \to \infty$, it is clear that the estimated bootstrap *P* value $\hat{p}^*(\hat{\tau})$ will tend to the (true) bootstrap p-value $p^*(\hat{\tau})$, which is defined as

$$p^*(\hat{\tau}) \equiv \Pr_{\hat{\mu}}(\tau_j \geq \hat{\tau}), \tag{11}$$

where $\hat{\mu}$ denotes the bootstrap DGP (Data Generation Process) that is used to generate the bootstrap samples.

In many situations, the bootstrap can be used to perform hypothesis tests that are more reliable in finite samples than the tests based on asymptotic theory (Davidson and MacKinnon, 2001). In econometrics, the use of the bootstrap for this purpose has been advocated by Horowitz (1994), Hall and Horowitz (1996), and others. For the bootstrap to work well the original test statistic must be asymptotically pivotal. In other words, its asymptotic distribution must not depend on any unknown features of the process that generated the data. For asymptotically pivotal test statistics, the bootstrap will yield more accurate inferences than the tests based on the asymptotic theory in the sense that the errors it makes will be of lower order in the sample size *n*.

The errors committed by using the bootstrap are generally lower by a factor of either $n^{-1/2}$ or $n^{-1}$ than the errors committed by relying on asymptotic tests.

In summary, the reason for using bootstrap inference is that the hypothesis tests and confidence intervals based on the asymptotic theory can be seriously misleading when the sample size is not large. Bootstrap allows the data analyst to assess the statistical accuracy of complicated procedures, by exploiting the power of the computer. The use of the bootstrap either relieves the analyst from dealing with complex mathematical derivations, or in some instances provides an alternative solution when no analytical answer can be obtained. Of course, asymptotic tests are not always misleading. In many cases, a bootstrap test will yield essentially the same inferences as an asymptotic test based on the same test statistic. Although this does not necessarily imply that the asymptotic test is reliable, the investigator may reasonably feel greater confidence in the results of asymptotic tests that have been confirmed in this way.

## §2.4 Bootstrap in Probit Model

In the probit model, unlike in popular linear regression models, asymptotic standard error is based on the first order approximation. Consequently, the standard errors might be biased when the sample size is small and the size of the test may be very different from the nominal size. Also note that the probit model is fully parametric in the sense that the error is assumed to be standard normal, which gives an excellent environment for the generation of bootstrap estimators.

Estimated asymptotic variances for the estimates of the parameters in a logit-probit model for binary response data are unreliable for moderate sized samples. Albanese and Knott (2001) show how bootstrapping gives a better idea of the sampling distribution of the estimators, and can also allow an assessment of the reliability of the scoring of individuals on the latent scale. Their bootstrap results suggest there is bias in the ML estimates and although the bootstrap distributions must underestimate the variation which would be present in the true sampling distribution for the estimators, the authors believe that they give a better guide to sampling variation than the usual first order normal approximation. Bootstrapping methods seem to be very useful for investigating the adequacy of the normal approximation in doubtful cases. When the discrimination parameters are small the asymptotic theory works well, but when they get large it can be inadequate.

# CHAPTER THREE: METHODOLOGY

In this chapter, we discuss the data generation process for simulation and the details of bootstrap procedure. For simplicity, consider the case where the probability that *y* takes on the value zero or one is conditional on a single explanatory variable $x_1$.

## §3.1 Data Generation Process for Simulation

We have the latent variable model as $y^* = \beta_0 + \beta_1 x_1 + e$, $y = I(y^* > 0)$. Given the assumption, error term *e* is generated as standard normal random numbers. $x_1$ is also generated by a standard normal random distribution. $\beta_0$ and $\beta_1$ are the true value of the intercept and the slope. In Table 1-3, the null hypothesis is $\beta_1 = 0$, and in Table 4, we consider the case when the null hypothesis is $\beta_1 = 1$. Five values are given to $\beta_1$ used to in data generation for each Table. For Tables 1-3, $\beta_1 = 0$ to calculate the size of the tests, $\beta_1 = -1, -0.5, 0.5, 1$ to calculate the power of the tests. The intercept, $\beta_0$, is adjusted to make the P(Y=1) = 0.5 for Table 1, P(Y=1) = 0.7 for Table 2, P(Y=1) = 0.3 for Table 3, and P(Y=1) = 0.5 for Table 4. The dependent variable *y* is generated as

$$\begin{cases} y = 1, & if\ \beta_0 + \beta_1 x_1 + e > 0 \\ y = 0, & if\ \beta_0 + \beta_1 x_1 + e \le 0 \end{cases}$$

MLE method which is described above is applied to estimate $\beta_0$ and $\beta_1$. The asymptotic standard error is calculated by using (8). The usual *t* statistics (asymptotic t-statistic) is then computed and total number of rejection has been counted out of 5,000 replications. They are used as the size and the power in the tables for three sample cases for *n* = 30, 50, 100. All the size and the power of tests are based on the nominal 5% critical value. We also report the bias and the variance of the maximum likelihood estimator of the slope $\beta_1$.

## §3.2 Bootstrap Procedure

Since the probit model is fully parametric in the sense that the error distribution is standard normal, it makes perfect sense to generate $n \times 1$ random numbers from standard normal to generate new y as:

$$\begin{cases} y = 1, & if \ b_0 + \beta_s x_1 + e > 0 \\ y = 0, & if \ b_0 + \beta_s x_1 + e \le 0 \end{cases}$$

where $b_0$ is the MLE intercept estimator, $\beta_s$ is the assumed value of the slope under the null hypothesis, and *e* is newly generated standard normal random numbers for each bootstrapping. Next, conduct the probit MLE procedure using the same $x_1$ and newly generated *y*. Repeat the entire procedure 1000 times (number of Bootstrap samples), sort the 1000 bootstrapped estimators to find the low tail 2.5% critical value and the upper tail 2.5% critical value to count the number of rejections. We reject the null hypothesis if the maximum likelihood estimator is in the rejection region built on the 100 bootstrapped estimators. The size and power of the bootstrap-based tests are reported along with the size and power of the asymptotic tests.

# CHAPTER FOUR: FINDINGS

Table 1-3 report the results for the exclusion restriction $\beta_1 = 0$. Table 1 reports the case when P(Y=1) = 0.5. In all the tables, the shade part is the size of the tests. Let's review the results reported in Table 1 first. The first thing to note is that the probit MLE can be biased seriously when the sample size is small, and the bias depends on the value of $\beta_1$. In the first panel of Table 1 where $n = 30$, bias of $\beta_1$ estimator is (-0.204, -0.072, 0.000, 0.080, 0.217) when $\beta_1$ = (-1, -0.5, 0, 0.5, 1). It is very important to note that the bias is very small when $\beta_1 = 0$, because this is what that matters for the size of the test for exclusion test of $\beta_1 = 0$. The size of the asymptotic test is 0.051 for the asymptotic test and 0.048 for the bootstrap test for $n = 30$. Both are very close to the nominal 5% size.

It is interesting to note that the direction of the bias is the same as the sign of $\beta_1$, and the magnitude of the bias gets larger for larger value of $\beta_1$. In another words the bias is positive when $\beta_1$ is positive, and it is negative when $\beta_1$ is negative. This direction of the bias is favorable for the power of the tests. For example, consider the power of the test when $\beta_1 = 1$ when the null hypothesis $\beta_1 = 0$. Since the estimators are positively biased, the distribution of the test statistic is skewed to the right, and it gives higher probability of rejection (or high power). When $\beta_1$ is negative, it works in the same way with exactly opposite direction. Note that the results reported in Table 1 suggest that the powers of the tests are compared very closely between the asymptotic test and the bootstrap test. The difference does not exceed 1% for the results reported in Table 1.

The bias quickly disappears as the sample size gets larger. When $\beta_1 = 1$, bias shrinks from 0.217 to 0.108 and 0.046 as the sample size rises from 30 to 50 and 100. But, the bias is virtually zero regardless of the sample sizes considered in the simulation, and size of the test are all close to the nominal 5%.

Table 2 reports the results when P(Y=1) = 0.7, and Table 3 reports the simulation results when P(Y=1) = 0.3. Not only the variance of the estimators, but the bias get larger to compare with the results in Table 1 where P(Y=1) = 0.5. The biases and the variances reported in Table 2 and 3 are similar to each other. Consequently, the power of the tests is lower in Table 2 and 3. However, the sizes of the tests are all fine in Tables 2 and 3, mainly because the bias is negligible when the slope is zero.

If we are ever interested in the null hypothesis $\beta_1 = 1$, we have very different results (see Table 4). But, we have to note that this sort of non-exclusion hypothesis show up rarely in practice. Since what we really have in the model to estimate is $\beta_1/\sigma$ whenever we assume the standard normal error in the latent variable model. $\beta$ and $\sigma$ cannot be identified separately in the binary response model. The test of $\beta_1 = 1$ in fact means that $\beta_1/\sigma = 1$, or $\beta_1 = \sigma$, which would be hardly interesting hypothesis in practice.

Since the probit MLE is biased when the sample size is small and $\beta_1 = 1$, the size of the test for the null hypothesis $\beta_1 = 1$ can be quite different from the nominal size. When $n = 30$, bias is 0.217, the size of the asymptotic test is 2.5%, and the size of the bootstrap test is 4.4%. As the sample size increases to 50, the bias decreases to 0.108. The sizes of the asymptotic test and the bootstrap tests are 3.8% and 4.8%, respectively. The bias is even bigger when the proportion of $y = 1$ away from 0.5. But, we do not report for these cases. In summary, the bias results in the smaller size than the nominal 5%.

The most serious problem with the smaller size than the nominal size is in the low power property of the test. It is not exception in the probit model. However, it is interesting to see that the tests have more power against $\beta_1 < 1$ where the bias tends to be smaller, and becomes less power against the alternative of $\beta_1 > 1$ where the bias tends to be smaller. We think the variance of the estimators plays out important role here. The power of the test is smaller than the size of the test when $n = 30$, null hypothesis is $\beta_1 = 1$, and when in fact the true value of $\beta_1 = 0.5$. As the sample size grows, the power increases quickly. When $n = 50$, the power is 18.6%, and when $n = 100$, the power is 50.4%.

The size of the bootstrap test is closer to the nominal size, and the power is much better. The test still is favorable power property against $\beta_1 < 1$ just like the asymptotic tests. But, the overall power property is much better than that of the asymptotic tests, which is not surprising when we consider that the size was bigger and closer to the nominal size for the bootstrap tests to compare with the asymptotic tests.

In summary, the tests based on the asymptotic standard errors are fine when the tests are for exclusion of $\beta_1 = 0$. Bootstrapping size is quite robust to the presence of the bias, so it has some limited usefulness when the null hypothesis is, for example, $\beta_1 = 1$.

Table 1: Probit Size and Power: $\beta_1$ = 0, P(Y=1)=0.5

| n | $\beta_1$ | Bias | Variance | Rejections (Asymptotic Standard Errors) | Rejections (Bootstrap Confidence Interval) |
|---|---|---|---|---|---|
| 30 | -1 | -0.204 | 0.386 | 0.943 | 0.942 |
| 30 | -0.5 | -0.072 | 0.113 | 0.507 | 0.491 |
| 30 | 0 | 0.000 | 0.076 | 0.051 | 0.048 |
| 30 | 0.5 | 0.080 | 0.114 | 0.516 | 0.509 |
| 30 | 1 | 0.217 | 0.362 | 0.949 | 0.948 |
| | | | | | |
| 50 | -1 | -0.102 | 0.124 | 0.997 | 0.997 |
| 50 | -0.5 | -0.039 | 0.054 | 0.749 | 0.741 |
| 50 | 0 | -0.002 | 0.040 | 0.057 | 0.054 |
| 50 | 0.5 | 0.045 | 0.055 | 0.759 | 0.752 |
| 50 | 1 | 0.108 | 0.125 | 0.998 | 0.998 |
| | | | | | |
| 100 | -1 | -0.042 | 0.047 | 1.000 | 1.000 |
| 100 | -0.5 | -0.018 | 0.024 | 0.967 | 0.964 |
| 100 | 0 | -0.002 | 0.018 | 0.055 | 0.054 |
| 100 | 0.5 | 0.018 | 0.023 | 0.967 | 0.964 |
| 100 | 1 | 0.046 | 0.046 | 1.000 | 1.000 |

Table 2: Probit Size and Power: $\beta_1$ = 0, P(Y=1)=0.7

| n | $B_1$ | Bias | Variance | Rejections (Asymptotic Standard Errors) | Rejections (Bootstrap Confidence Interval) |
|---|---|---|---|---|---|
| 30 | -1 | -0.262 | 0.971 | 0.884 | 0.823 |
| 30 | -0.5 | -0.102 | 0.547 | 0.445 | 0.390 |
| 30 | 0 | 0.001 | 0.095 | 0.048 | 0.038 |
| 30 | 0.5 | 0.109 | 0.596 | 0.454 | 0.409 |
| 30 | 1 | 0.268 | 0.658 | 0.880 | 0.835 |
| | | | | | |
| 50 | -1 | -0.126 | 0.170 | 0.992 | 0.989 |
| 50 | -0.5 | -0.051 | 0.067 | 0.692 | 0.668 |
| 50 | 0 | 0.000 | 0.047 | 0.052 | 0.048 |
| 50 | 0.5 | 0.048 | 0.063 | 0.694 | 0.680 |
| 50 | 1 | 0.127 | 0.171 | 0.992 | 0.988 |
| | | | | | |
| 100 | -1 | -0.054 | 0.057 | 1.000 | 1.000 |
| 100 | -0.5 | -0.023 | 0.027 | 0.945 | 0.939 |
| 100 | 0 | 0.000 | 0.020 | 0.054 | 0.053 |
| 100 | 0.5 | 0.023 | 0.027 | 0.945 | 0.944 |
| 100 | 1 | 0.052 | 0.056 | 1.000 | 1.000 |

Table 3: Probit Size and Power: $\beta_1$ = 0, P(Y=1)=0.3

| n | $\beta_1$ | Bias | Variance | Rejections (Asymptotic Standard Errors) | Rejections (Bootstrap Confidence Interval) |
|---|---|---|---|---|---|
| 30 | -1 | -0.247 | 0.554 | 0.885 | 0.854 |
| 30 | -0.5 | -0.093 | 0.167 | 0.445 | 0.394 |
| 30 | 0 | 0.001 | 0.088 | 0.043 | 0.032 |
| 30 | 0.5 | 0.098 | 0.174 | 0.443 | 0.400 |
| 30 | 1 | 0.261 | 0.543 | 0.882 | 0.855 |
| | | | | | |
| 50 | -1 | -0.119 | 0.158 | 0.991 | 0.989 |
| 50 | -0.5 | -0.047 | 0.063 | 0.691 | 0.666 |
| 50 | 0 | 0.000 | 0.044 | 0.048 | 0.045 |
| 50 | 0.5 | 0.053 | 0.068 | 0.693 | 0.679 |
| 50 | 1 | 0.125 | 0.168 | 0.989 | 0.989 |
| | | | | | |
| 100 | -1 | -0.050 | 0.056 | 1.000 | 1.000 |
| 100 | -0.5 | -0.021 | 0.027 | 0.947 | 0.942 |
| 100 | 0 | -0.001 | 0.019 | 0.049 | 0.048 |
| 100 | 0.5 | 0.020 | 0.027 | 0.939 | 0.936 |
| 100 | 1 | 0.054 | 0.056 | 1.000 | 1.000 |

Table 4: Probit Size and Power: $\beta_1$ = 1, P(Y=1)=0.5

| n | $\beta_1$ | Bias | Variance | Rejections (Asymptotic Standard Errors) | Rejections (Bootstrap Confidence Interval) |
|---|---|---|---|---|---|
| 30 | 0 | 0.000 | 0.076 | 0.953 | 0.954 |
| 30 | 0.5 | 0.080 | 0.114 | 0.409 | 0.413 |
| 30 | 1 | 0.217 | 0.362 | 0.025 | 0.044 |
| 30 | 1.5 | 0.417 | 0.960 | 0.014 | 0.120 |
| 30 | 2 | 0.665 | 2.233 | 0.036 | 0.315 |
| | | | | | |
| 50 | 0 | -0.002 | 0.040 | 0.996 | 0.996 |
| 50 | 0.5 | 0.045 | 0.055 | 0.593 | 0.598 |
| 50 | 1 | 0.108 | 0.125 | 0.038 | 0.048 |
| 50 | 1.5 | 0.213 | 0.324 | 0.186 | 0.257 |
| 50 | 2 | 0.386 | 0.888 | 0.542 | 0.656 |
| | | | | | |
| 100 | 0 | -0.002 | 0.018 | 1.000 | 1.000 |
| 100 | 0.5 | 0.018 | 0.023 | 0.875 | 0.871 |
| 100 | 1 | 0.046 | 0.046 | 0.045 | 0.055 |
| 100 | 1.5 | 0.094 | 0.106 | 0.504 | 0.533 |
| 100 | 2 | 0.161 | 0.225 | 0.955 | 0.961 |

# CHAPTER FIVE: CONCLUSION

Overall, the tests based on the asymptotic standard errors are fine when the tests are tests for exclusion ($\beta = 0$). Bootstrap tests yield essentially the same inferences as an asymptotic test based on the same test statistic, the investigator may reasonably feel greater confidence in the results of asymptotic tests that have been confirmed by bootstrap method. This result is consistent with MacKinnon (2002). Bootstrapping size is quite robust to the presence of the bias, so it has some limited usefulness when the null hypothesis is, for example, $\beta = 1$.

We expanded the program to the case when there are more than one regressor, so the model changes to $P(y=1|\mathrm{X}) = G(\beta_0 + \beta_1 x_1 + \beta_2 x_2)$. When $x_1$ and $x_2$ are correlated, and $\beta_2$ is not zero, we found that there is no significant impact on the bias of $\beta_1$ estimator.

In this paper, only the probabilities of $y = 1$ and $y = 0$ were controlled. In future study, we can conduct an alternative simulation, where the proportion of $y = 1$ and $y = 0$ are fixed as it is given in the sample.

Bias in the small sample seems too high to rely upon the estimators. Bootstrap might be an alternative avenue for bias correction. MacKinnon and Smith (1998) discussed the methods for reducing the bias of consistent estimators that are biased in finite samples. Further work may shed light on this issue.

# APPENDIX: GAUSS PROGRAM

```
new;
output file = probit2.out reset;

/* This program is to see the size of
   (1) the standard probit estimation and asymptotic variance and
   (2) bootstrap probit */

pp=1;
do until pp > 3;

bs=0;          /* Assumed Slope Coefficient under the null */
repl=5000;     /* number of replications */
bsp=1000;               /* Number of resampling in bootstrapping */
tol=1e-06;
maxiter=25;    /* Maximum iteration for Probit Estimation */

nn={30, 50, 100};
case=1;
do until case > rows(nn);

n=nn[case];
cst=ones(n,1);

ggap=seqa(-1,0.5,5);
sp=1;
do until sp > rows(ggap);

gap=ggap[sp];

bt=bs+gap;      /* True Slope Coefficient */

if pp==1;
   b0=0;          /* Intercept for P(Y=1)=0.5 */
elseif pp==2;
   b0=0.5244*sqrt(bt^2+1);   /* Intercept for P(Y=1)=0.7 */
elseif pp==3;
   b0=-0.5244*sqrt(bt^2+1);   /* Intercept for P(Y=1)=0.3 */
endif;

/* 1. Bias, Variance of Probit Estimation and Rejection Rate based on the Asymptotic Varaince */

seed1 = 123586;
seed2 = 345334;
seed3 = 675646;

bsum=0;
bsqsum=0;
h1sum=0;
h2sum=0;
j=1;
do until j > repl;     /* Beginning of replication */

   here1:
```

```
/* Probit Estimation */

x1=rndns(n,1,seed1);   /* regressor from N(0,1) */
e=rndns(n,1,seed2);       /* error term from N(0,1) */
y=b0.*cst + bt*x1 + e .>0;   /*True value of the slope = bt*/

x=cst~x1;
b=invpd(x'x)*x'y;   /* OLS Initial Value */

crit=1;
i=1;
do until (crit < tol) or (i ge maxiter);

   bn=b;
   pdf=pdfn(x*bn);                /** Normal pdf **/
   cdf=cdfn(x*bn);                /** Normal cdf **/

   g = y.*(pdf./cdf).*x-(1-y).*(pdf./(1-cdf)).*x; /* Gradient vector */
   g=sumc(g);

   D = pdf.*((y.*(pdf+(x*bn).*cdf)./cdf^2)
      +((1-y).*(pdf-(x*bn).*(1-cdf))./(1-cdf)^2));

   H=-(x.*D)'x;     /* Hessian */

   if rank(H) < rows(H);   /* If Hessian is singular, start all over again from generation of y */
      goto here1;
   else;
   endif;

   db = -inv(H)*g;    /** Newton step **/

   crit=maxc(abs(db));

   b=bn+db;  /* Update the estimator */

i=i+1;
endo;

/* Computation of asymptotic standard error, t-statistic, and h1=1 if rejected */

pdf=pdfn(x*b);
cdf=cdfn(x*b);
x0=pdf.*x;

se=sqrt(diag(invpd(x0'(x0./(cdf.*(1-cdf))))));

b1=b[1];      /* will be used for bootstrap data generation */
b2=b[2];
t=(b2-bs)/se[2];
h1=(abs(t) > 1.96);

bsum = bsum+b2;
bsqsum = bsqsum+b2^2;
```

```
h1sum=h1sum+h1;


     /* 2. Bootstrapping:  y will be Regenerated based on the newly generated error terms */

     bb=zeros(bsp,1);              /* Room to generate the distribution of bootstrap estimators of the slope */
     jj=1;
     do until jj > bsp;

  here2:

  y = b1 + bs.*x1 + rndns(n,1,seed3) .> 0;   /* Bootstrap y based on the assume value of b2=bs and MLE b1 */
  b=invpd(x'x)*x'y;   /* OLS Initial Value */

  /* Probit Estimation using the bootstrap y. X is the same */

  crit=1;
  i=1;
  do until (crit < tol) or (i ge maxiter);

    bn=b;
    pdf=pdfn(x*bn);               /** Normal pdf **/
    cdf=cdfn(x*bn);               /** Normal cdf **/

    g = y.*(pdf./cdf).*x-(1-y).*(pdf./(1-cdf)).*x; /* Gradient vector */
    g=sumc(g);

    D = pdf.*((y.*(pdf+(x*bn).*cdf)./cdf^2)
       +((1-y).*(pdf-(x*bn).*(1-cdf))./(1-cdf)^2));

    H=-(x.*D)'x;     /* Hessian */

    if rank(H) < rows(H);   /* If Hessian is singular, start all over again from generation of y */
      goto here2;
    else;
    endif;

    db = -inv(H)*g;    /** Newton step **/

    crit=maxc(abs(db));

    b=bn+db;  /* Updating the estimator */

  i=i+1;
  endo;

  bb[jj] = b[2];

jj=jj+1;
endo;

bb=sortc(bb,1);                    /* Sorting the bootstrapped estimators */
lcr=bb[bsp*.025];                  /* Low tail 2.5% critical value */
ucr=bb[bsp*.975];                  /* Upper tail 2.5% critical value */
```

```
   h2=(b2 < lcr) + (ucr < b2);          /* 1 if the MLE b2 is in the rejection region built upon bootstrap estimators */
   h2sum = h2sum + h2;

j=j+1;
endo;

meanb=bsum/repl;
bias=meanb-bt;

varb=bsqsum/repl-meanb^2;
rej1=h1sum/repl;
rej2=h2sum/repl;

format /rd 9,4;
n~bt~bias~varb~rej1~rej2;

sp=sp+1;
endo;

case=case+1;
endo;"";"";

pp=pp+1;
endo;

end;
```

## LIST OF REFERENCES


Albanese, M.T. and M. Knott (2001), "Bootstrapping Latent Variable Models for Binary Response," London School of Economics working paper.

Davidson, Russell and James G. MacKinnon (2004), "The Power of Bootstrap and Asymptotic tests," Queens University working paper 1035.

Efron, Bradley and Robert J. Tibshirani (1993), *An Introduction to the Bootstrap*, Chapman & Hall, New York.

Hall, P. and Joel L. Horowitz (1996), "Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators," *Econometrica*, 64, 891-916.

Horowitz, Joel L. (1994), "Bootstrap-based Critical Values for the Information Matrix Test," *Journal of Econometrics*, 61, 395-411.

Horowitz, Joel L. and N.E. Savin (2001), "Binary Response Models: Logits, Probits and Semiparametrics," *Journal of Economic Perspectives*, 15, (4), 43-56.

MacKinnon, James G. (2002), "Bootstrap Inference in Econometrics," *Canadian Journal of Economics*, 35, (4), 615-645.

MacKinnon, James G. and Anthony A. Smith (1998), "Approximate Bias Correction in Econometrics," *Journal of Econometrics*, 85, 205-230.

Wooldridge, Jeffrey M. (2003), *Introductory Econometrics: A Modern Approach*, 2e, South Western, Ohio.