ECONOMICS BULLETIN

A monte carlo analysis of the type II tobit maximum likelihood estimator when the true model is the type I tobit model

Kazumitsu Nawata
University of Tokyo

Abstract

Type I (censored regression) and Type II Tobit (sample selection) models are widely used in the various fields of economics. The Type I Tobit model is a special case of the Type II Tobit model. However, the dimension of the error terms decreases and the distribution of the error terms degenerates in the Type I Tobit Model. Therefore, we cannot use the standard asymptotic theorems for the Type II Tobit Maximum Likelihood Estimator (MLE) when the sample is obtained from the Type I Tobit model. Results of Monte Carlo experiments show strange behavior that has never been reported before for the Type II MLE.

I would like to thank Prof. Yoshiyuki Takeuchi of Osaka University for his helpful comments.
Citation: Nawata, Kazumitsu, (2007) "A monte carlo analysis of the type II tobit maximum likelihood estimator when the true model is the type I tobit model." *Economics Bulletin,* Vol. 3, No. 54 pp. 1-10
Submitted: November 3, 2007. **Accepted:** November 3, 2007.
URL: http://economicsbulletin.vanderbilt.edu/2007/volume3/EB-07C20123A.pdf

1. Introduction

Type I (censored regression) and Type II Tobit (sample selection) models are widely used in the various fields of economics. For details of the models, see Amemiya (1985). The Type I Tobit model is a special case of the Type II Tobit model. The former model is obtained from the latter with some restrictions to the parameters. However, the dimension of the error terms decreases from two (Type II Tobit model) to one (Type I Tobit model); that is, the distribution of the error terms degenerates in the Type I Tobit model. Therefore, the standard asymptotic theorems cannot be used with it. In this paper, I use Monte Carlo experiments to analyze the behavior of the Type II Tobit Maximum Likelihood Estimator (MLE) when the sample is obtained from the Type I Tobit model. Results of the Monte Carlo experiments show strange behavior that has never been reported before for the Type II Tobit MLE.

2. Type I and Type II Tobit Models

The Type I Tobit model is given by

(2.1) $Y_i^* = x_i'\beta + u_i$, and

$$Y_i = \begin{cases} Y_i^* & Y_i^* > 0, \\ 0 & Y_i^* \le 0. \end{cases} \qquad i = 1,2,\ldots,n,$$

The value of Y_i^* is not observable if it is negative. x_i is a vector of explanatory variables. u_i follows the normal distribution with a mean of 0 and a variance of σ^2.

On the other hand, the Type II Tobit model is given by

(2.2) $Y_{1i}^* = x_{1i}'\beta_1 + u_{1i}$,

$Y_{2i}^* = x_{2i}'\beta_2 + u_{2i}$, and

$$Y_{2i} = \begin{cases} Y_{2i}^* & Y_{1i}^* > 0, \\ 0 & Y_{1i}^* \le 0. \end{cases} \qquad i = 1,2,\ldots,n,$$

Y_{1i}^* is not observable and only its sign is observable. Y_{2i}^* is observable if and only if $Y_{1i}^* > 0$.

x_{1i} and x_{2i} are vectors of explanatory variables. u_{1i} and u_{2i} are jointly normal with means of 0, variances of 1 and σ_2^2, respectively, and a covariance of σ_{12}. The likelihood function is given by

(2.3) $$L(\beta_1,\beta_2,\sigma_2^2,\rho)=\prod_{d_i=0}\{1-\Phi(x_{1i}'\beta_1)\}$$
$$\cdot\prod_{d_i=1}\Phi[\{x_{1i}'\beta_1+\rho(Y_{2i}-x_{2i}'\beta_2)/\sigma_2\}/\sqrt{1-\rho^2}]\cdot\sigma_2^{-1}\phi\{(Y_{2i}-x_{2i}'\beta_2)/\sigma_2\}$$

Suppose that $x_{1i}=x_{2i}=x_i$ in (2.2). In this case, if

(2.4) $\beta_1=\beta_2/\sigma_2$ and $u_{1i}=u_{2i}/\sigma_2$, $i=1,2,...,n$,

the Type II Tobit model becomes the Type I Tobit model. The second condition of (2.4) gives

(2.5) $\rho=1.0$,

where ρ is the correlation coefficient between u_{1i} and u_{2i}. Therefore, when $x_{1i}=x_{2i}=x_i$, the Type I Tobit model is a special case of the Type II Tobit model satisfying

(2.6) $\beta_1=\beta_2/\sigma_2$ and $\rho=1.0$.

However, under these conditions, the error terms change from (u_{1i},u_{2i}) to u_i. The dimension of the error terms decreases from 2 to 1 and the distribution degenerates. Since the likelihood function of the Type II Tobit model is a function of $1/(1-\rho^2)$, we cannot analyze the likelihood function in the neighborhood of $\rho=1.0$ by the standard asymptotic theorems. Therefore, I have analyzed the MLE by the Monte Carlo experiments in the following section.

3. Monte Carlo Experiment

The true model is the Type I Tobit Model, and the data is generated by

(3.1) $Y_{1i}^*=\beta_{11}+\beta_{12}\mathrm{x}_i+u_i$,

$Y_{2i}^*=\beta_{21}+\beta_{22}\mathrm{x}_i+u_i$, $i=1,2,...,n$, and

$\beta_{11}=\beta_{21},\beta_{12}=\beta_{22}$.

x_i follows a uniform distribution over (0,2) and u_i follows a standard normal distribution. For the values of β_{ij}, two cases such that i) $\beta_{11}=\beta_{21}=0.0$ and $\beta_{12}=\beta_{22}=0.0$, and ii) $\beta_{11}=\beta_{21}=-1.0$ and $\beta_{12}=\beta_{22}=1.0$ are considered. The sample sizes are $n=100$, 200, 400, 800, and 1600. The number of trials is 1,000 for each case. Since the data used in this study is a special case of the Type II Tobit Model, the standard algorithms such as the ones used in LIMDEP and STATA seldom converge. Therefore, the scanning method proposed by Nawata (1994 and 1995), Nawata and Nagase (1996) and Nawata and McAleer (2001) is used in the estimation. The program is written in the C-language, and all calculations are done using double precision. The maximum value of ρ is set to be 0.999999 due to the accuracy of the calculations. Note that the program guarantees to calculate the global maximum up to ρ = 0.999999.

The results of $\hat{\rho}$, the estimator of ρ, are given in Table 1. $\hat{\rho}$ becomes 0.999999, the maximum of ρ, in the majority of trials. Among 1,000 trials, $\hat{\rho}$ becomes 0.999999 in 949 and 955 trials when n =100, 903 and 913 trials when n=200, 935 and 923 trials when n=400, 948 and 949 trials when n=800, and 912 and 920 trials when n=1600 (the former numbers are in the $\beta_{11} = \beta_{21} = 0.0$ and $\beta_{12} = \beta_{22} = 0.0$ case and the latter numbers are in the $\beta_{11} = \beta_{21} = -1.0$ and $\beta_{12} = \beta_{22} = 1.0$ case). Figure 1 shows the graph of $\log L$ given by (2.3) in the trial where $\hat{\rho}$ = 0.999999 (n=100, $\beta_{11} = \beta_{21} = -1.0$ and $\beta_{12} = \beta_{22} = 1.0$). log*L* is calculated by maximizing the conditional likelihood function for a given value of ρ. Since the shape of the likelihood function in the neighborhood of ρ=1.0 is important, the horizontal axis is set to be $-\log_{10}(1-\rho) = \log_{10}\{1/(1-\rho)\}$. Although log*L* increases as ρ approaches 1.0, it does not diverge but rather converges at a certain value.

Although $\hat{\rho}$ becomes 0.999999 in the majority of trials, the interesting finding is that $\hat{\rho}$ is not 0.999999 in some (5-10%) trials. Let $\gamma = \beta_2 / \sigma_2 - \beta_1$. Then

$$\text{(3.2)} \qquad \frac{d \log L}{d\rho} = (1-\rho^2)^{-3/2} \sum_{d_i=1} \lambda(\psi_i)\{Y_i / \sigma - x_i \gamma + (1-\rho)(Y_i - x_i' \beta_2)/\sigma_2\},$$

where $\psi_i(\beta_2, \gamma, \sigma_2, \rho) = \{Y_{2i}/\sigma_2 - x_i'\gamma - (1-\rho)(Y_{2i} - x_i'\beta_2)/\sigma_2\} / \sqrt{1-\rho^2}$ and $\lambda(z) = \phi(z)/\Phi(z)$.

Therefore, it is possible to obtain $\partial \log L / \partial \rho$=0 at a value $\rho < 1.0$, and such trials are actually observed. Figure 2 is the graph of *LogL* in the trial where $\hat{\rho}$=0.9374 (n=100, $\beta_{11} = \beta_{21} = -1.0$ and $\beta_{12} = \beta_{22} = 1.0$). In this case, $\hat{\rho}$ is considerably different from 1.0. *logL* reaches its maximum value, -100.531, at $\rho = 0.9374$ (value of the horizontal axis $= -\log_{10}(1-\rho)$ =1.2037), decreases after that and reaches a local minimum, -103.11, at $\rho = 0.9966$（value of the horizontal axis=2.4684）. *LogL* increases for $\rho > 0.9966$, but the increment rate decreases, and *LogL* converges to a certain value.

In other problems where the dimension decreases and the distribution degenerates [1], we always get the true parameter values for some parameters. For example, let us consider the principle component analysis when some eigenvalues are 0. If eigenvalues are 0, the dimension of the sample space decreases and eigenvalues calculated from the sample data always become 0. This means that the null hypothesis, that eigenvalues are 0, is rejected unless sample eigenvalues are exactly equal to 0. Considering observation errors and other factors, this test is meaningless in practice. We can interpret the cointegration methods such as those of Johansen (1988 and 1991) as making practically meaningful tests possible by introducing I(0) and I(1) processes. Although the ratios of eigenvalues approach 0 as the number of observations increases, they are not 0 in finite sample sizes.

In the Type II Tobit model, we can also consider a test of the Type I Tobit model using the null hypothesis given by [2]

(3.3) $\beta_{11} = \beta_{21} / \sigma_2$ and $\beta_{12} = \beta_{22} / \sigma_2$.

The distributions of $\hat{\beta}_{11} - \hat{\beta}_{21} / \hat{\sigma}_2$ and $\hat{\beta}_{12} - \hat{\beta}_{22} / \hat{\sigma}_2$ are given in Table 2. The biases are small. For the $\beta_{11} = \beta_{21} = 0.0$ and $\beta_{12} = \beta_{22} = 0.0$ case, the standard deviations are 0.1694 and 0.09983 when n=100, 0.0530 and 0.0422 when n=200, 0.0239 and 0.0208 when n=400, 0.0136 and 0.0112 when n=800, and 0.00679 and 0.0061 when n=1600 (the former values are standard deviations of $\hat{\beta}_{11} - \hat{\beta}_{21} / \hat{\sigma}_2$ and the latter values are those of $\hat{\beta}_{12} - \hat{\beta}_{22} / \hat{\sigma}_2$). For the $\beta_{11} = \beta_{21} = -1.0$ and $\beta_{12} = \beta_{22} = 1.0$ case, they are 0.1355 and 0.1090 when n=100, 0.0752 and 0.0422 when n= 200, 0.0306 and 0.0256 when n=400, 0.0163 and 0.0142 when n=800, and 0.0085 and 0.0072 when n=1600. The decreasing rates of the standard deviations are much faster than those of the individual estimators. Although it is not as notable as the $\hat{\rho}$ cases, this fact suggests the possibility of super-consistency; that is $\hat{\beta}_{1j} - \hat{\beta}_{2j} / \hat{\sigma}_2, j = 1,2$ converges to zero faster than $n^{-1/2}$.

Let $(MSE_j)^{1/2}$ be the square root of the mean squared error of $\hat{\beta}_{1j} - \hat{\beta}_{2j} / \hat{\sigma}_2, j = 1,2$. $(MSE_2)^{1/2}$ is approximately proportional to n^{-1}. Estimating the equation

(3.4) $\log(MSE_j) = \alpha_1 + \alpha_2 \log(n) + \varepsilon_i$

by the least squares method, we get $\hat{\alpha}_2$ = -1.0312 (0.1043), -0.9938 (0.0413), -0.9727

(0.0452) and -0.9407 (0.0578) for each case (standard errors are in parentheses). $\hat{\alpha}_2$ is close to 1.0 for all cases, and it is suggested that $\hat{\beta}_{1i} - \hat{\beta}_{2i} / \hat{\sigma}_2, i = 1,2$ is of order n^{-1}. For the test of $H_0 : \beta_1 = \beta_2 / \sigma_2$, Fin and Schmidt (1984) proposed a Lagrange multiplier test, and Greene (2000, pp. 915) proposed a test based on the probit and truncated regression models. However, the likelihood function of the Type II Tobit model is not considered in these tests, and the tests are of order $n^{-1/2}$. Considering the fact that the dimension of the error terms decreases and the distribution of the error terms degenerates, it may be possible that we can perform a super-consistent test where the rate of convergence of the test statistic is faster than $n^{-1/2}$.

4. Conclusion

In this paper, I compare the Type I and Type II Tobit models. The former model is a special case of the latter model. However, since the dimension of the error terms decreases and the distribution degenerates, the standard asymptotic theorems cannot be used for the Type II Tobit MLE when the sample is obtained from the Type I Tobit model. This paper is the first attempt to analyze this problem. The results of Monte Carlo experiments show the strange behavior of the MLE. Although $\hat{\rho}$ becomes 0.9999999 (the maximum value of the calculation) in majority of trials, it becomes other values in 5-10% of the trials. It is also considered that the rate of convergence of $\hat{\gamma} = \hat{\beta}_1 - \hat{\beta}_2 / \hat{\sigma}_2$ is n^{-1}. The asymptotic distributions of $\hat{\rho}$ and $\hat{\gamma}$ are not yet known. It is necessary to find out their asymptotic distributions in future studies. Note that the dimension of the error terms decreases and the distribution of the error terms degenerates in multinomial probit and ordered probit models. An analysis of multinomial probit and ordered probit models is another subject that should be studied in the future.

Notes

1) There exit several studies of hypothesis testing when a parameter is on the boundary of the maintained hypothesis (for details, see Andrews (2001)). However, problems where the dimension of error terms decreases and the distribution degenerates are not considered

in these studies.

2) $\hat{\beta}_{ij}$ do not converge quickly and they are considered to be ordinal estimators of $O_p(n^{-1/2})$.

References

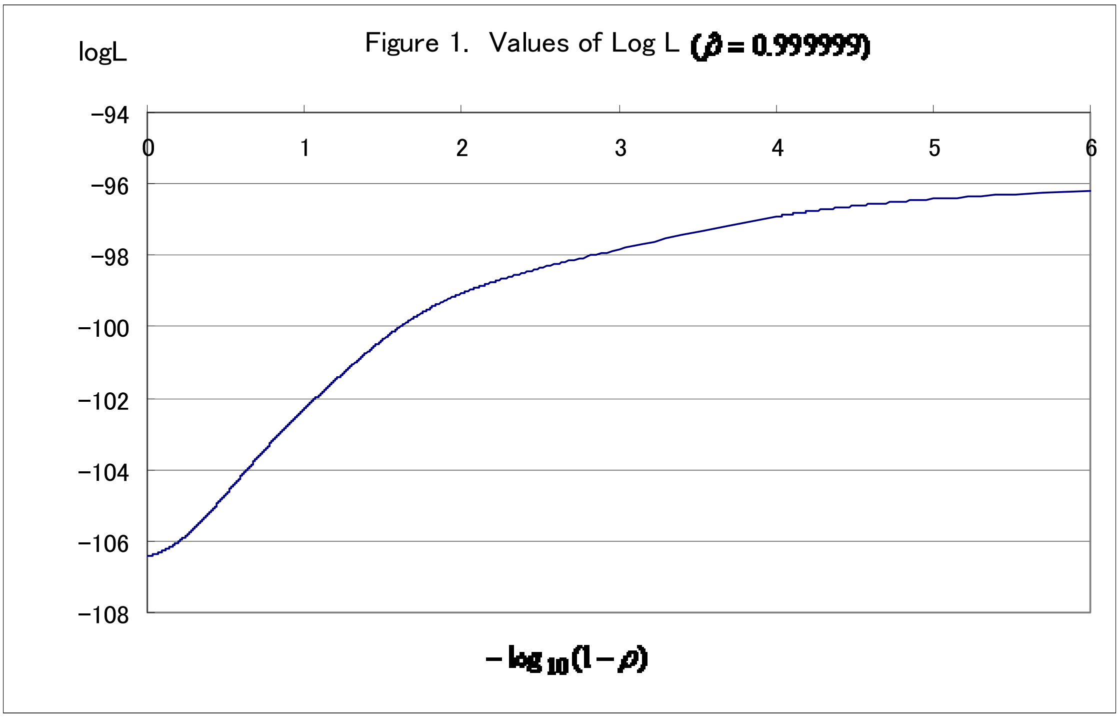
1. Amemiya, T., 1985, *Advanced Econometrics,* Harvard University Press.
2. Andrews, D.W.K., 2001, “Testing when a parameter is on the boundary of the maintained hypothesis,” *Econometrica* 69, 683-734.
3. Fin, T. and P. Schmidt, 1984, “A Test of the Tobit Specifications Against an Alternative Suggested by Cragg,” *Review of Economics and Statistics*, 66, 174-177.
4. Greene, W., 2000, *Econometric Analysis (Fourth Edition)*, Prentice Hall, Englewood Cliffs, New Jersey.
5. Johansen, S., 1988, “Statistical Analysis of Cointegeration Vectors,” *Journal of Economic Dynamics and Control*, 12, 231-254.
6. _____, 1991, “Estimation and Hypothesis of Cointegration Vectors in Gaussian Vector Autoregressive Model,” *Econometrica*, 59, 1551-1580.
7. Nawata, K., 1994, "Estimation of Sample Selection Bias Models by the Maximum Likelihood Estimator and Heckman's Two-Step Estimator," *Economic Letters*, 1994, Vol. 45, 33-40.
8. _____, 1995, "Estimation of Sample-Selection Models by the Maximum Likelihood Method," *Mathematics and Computers in Simulation*, Vol. 39, 299-303.
9. Nawata, K., and N. Nagase, 1996, "Estimation of Sample Selection Biases Models," 1996, *Econometric Reviews*, Vol. 15, 387-400.
10. Nawata, K. and M. McAleer, 2001, “Size Characteristics of Test for Sample Selection Biases: A Monte Carlo Comparison and Empirical Examples,” *Econometric Reviews*, Vol. 20, 105-112.

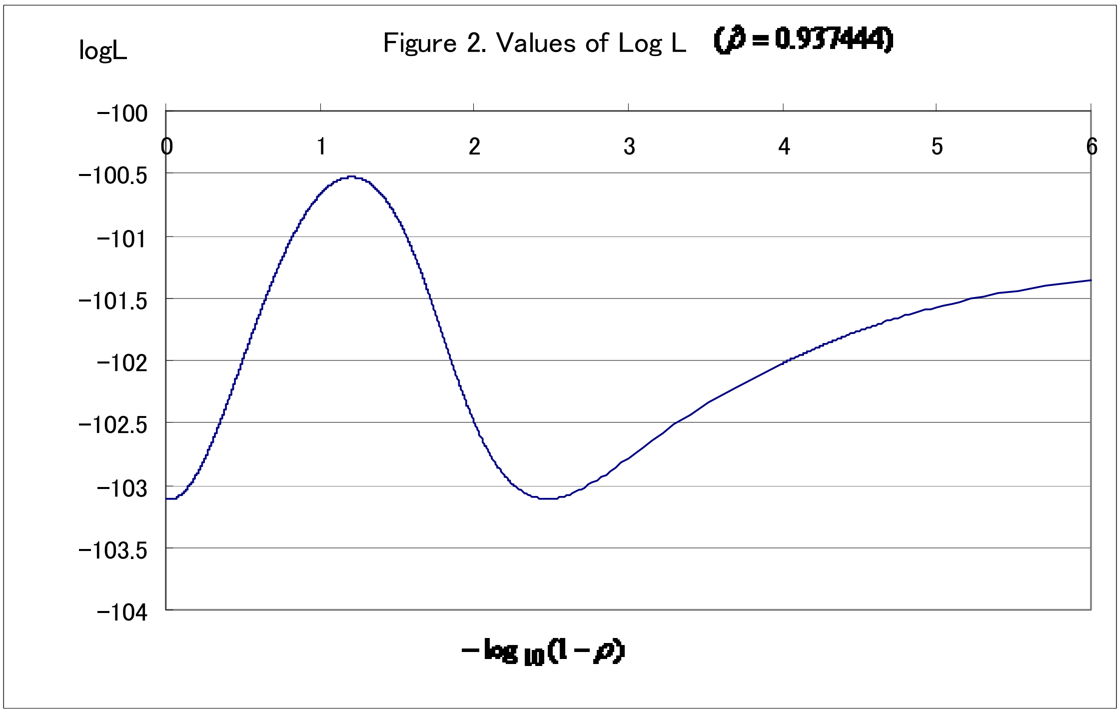
Table 1. Distributions of $\hat{\rho}$

$\beta_{11} = \beta_{21} = 0.0$ and $\beta_{12} = \beta_{22} = 0.0$					
	n=100	n=200	n=400	n=800	n=1600
Mean	0.993041	0.999508	0.999952	0.999980	0.999994
Standard Deviation	0.075229	0.004740	0.000243	0.000120	0.000043
Frequency					
0.8 or less	7	0	0	0	0
0.8-0.9	0	0	0	0	0
0.9-0.95	5	1	0	0	0
0.95-0.99	12	7	0	0	0
0.99-0.999	25	45	9	6	1
0.999-0.9999	2	44	55	34	12
0.9999-0.999998	0	0	1	12	75
0.999999	949	903	935	948	912
$\beta_{11} = \beta_{21} = -1.0$ and $\beta_{12} = \beta_{22} = 1.0$					
	n=100	n=200	n=400	n=800	n=1600
Mean	0.998823	0.998309	0.999922	0.999974	0.999993
Standard Deviation	0.011143	0.025277	0.000395	0.000167	0.000041
Frequency					
0.8 or less	1	2	0	0	0
0.8-0.9	0	1	0	0	0
0.9-0.95	5	2	0	0	0
0.95-0.99	20	10	0	0	0
0.99-0.999	18	41	22	5	0
0.999-0.9999	1	29	54	36	15
0.9999-0.999998	0	0	1	10	65
0.999999	955	915	923	949	920

Table 2. Estimates of $\hat{\beta}_{1j} - \hat{\beta}_{2j} / \hat{\sigma}_2, j = 1,2$

n	Mean	Standard Deviation	25% Percentile	Median	75% Percentile	(MSE)^0.5
$\beta_{11} = \beta_{21} = 0.0$ and $\beta_{12} = \beta_{22} = 0.0$						
$\hat{\beta}_{11} - \hat{\beta}_{21} / \hat{\sigma}_2$						
100	-0.04190	0.16944	-0.07771	-0.01899	0.01354	0.17454
200	-0.02335	0.05299	-0.03964	-0.01154	0.00732	0.05790
400	-0.01102	0.02395	-0.02059	-0.00682	0.00351	0.02637
800	-0.00830	0.01363	-0.01471	-0.00587	0.00040	0.01596
1600	-0.00638	0.00680	-0.01092	-0.00556	-0.00112	0.00932
$\hat{\beta}_{12} - \hat{\beta}_{22} / \hat{\sigma}_2$						
100	0.00049	0.09831	-0.04032	-0.00019	0.03930	0.09831
200	0.00100	0.04218	-0.01710	0.00072	0.01893	0.04220
400	-0.00050	0.02077	-0.01118	0.00026	0.01067	0.02077
800	0.00060	0.01115	-0.00627	0.00053	0.00707	0.01117
1600	0.00012	0.00610	-0.00400	0.00047	0.00432	0.00610
$\beta_{11} = \beta_{21} = -1.0$ and $\beta_{12} = \beta_{22} = 1.0$						
$\hat{\beta}_{11} - \hat{\beta}_{21} / \hat{\sigma}_2$						
100	-0.05319	0.13545	-0.09332	-0.02602	0.01632	0.14552
200	-0.02717	0.07522	-0.04754	-0.01186	0.00897	0.07998
400	-0.01394	0.03063	-0.02651	-0.00741	0.00457	0.03366
800	-0.00732	0.01634	-0.01416	-0.00418	0.00264	0.01790
1600	-0.00624	0.00852	-0.01165	-0.00509	0.00024	0.01056
$\hat{\beta}_{12} - \hat{\beta}_{22} / \hat{\sigma}_2$						
100	0.00461	0.10898	-0.04258	0.00374	0.04792	0.10907
200	0.00100	0.04218	-0.01710	0.00072	0.01893	0.04220
400	0.00106	0.02556	-0.01280	0.00014	0.01323	0.02559
800	-0.00080	0.01424	-0.00783	-0.00048	0.00668	0.01426
1600	-0.00004	0.00720	-0.00493	-0.00031	0.00500	0.00720


logL
Figure 1. Values of Log L $(\hat{\rho} = 0.999999)$
-94
-96
-98
-100
-102
-104
-106
-108
0
1
2
3
4
5
6
$-\log_{10}(1-\rho)$


logL
Figure 2. Values of Log L $(\hat{\rho} = 0.937444)$
-100
-100.5
-101
-101.5
-102
-102.5
-103
-103.5
-104
0
1
2
3
4
5
6
$-\log_{10}(1-\rho)$