A Monte Carlo Sampling Approach to Testing Separate Families of Hypotheses:

Monte Carlo Results

by

Nouhoun Coulibaly and B. Wade Brorsen

* Nouhoun Coulibaly is a Ph.D. candidate and B. Wade Brorsen is a regents professor in the Department of Agricultural Economics, Oklahoma State University, Stillwater, OK 74078.
Abstract

Monte Carlo experiments are designed to compare the finite sample performances of two Monte Carlo hypothesis tests with Pesaran and Pesaran's Cox-type test. The size of the Pesaran and Pesaran test is generally incorrect. The Monte Carlo tests perform equally well and are both preferred to Pesaran and Pesaran's test.

Key words: Cox test, Monte Carlo test, nonnested hypotheses

A Monte Carlo Sampling Approach to Testing Separate Families of Hypotheses:

Monte Carlo Results
The Cox statistic for testing separate families of hypotheses is the difference between the log
likelihood ratio and its expected value (Cox, 1961; 1962). While the log likelihood ratio is
straightforward to obtain, the computation of its expected value is generally intractable (Cox,
1962; Pesaran and Pesaran, 1993). Thus, a number of Cox-type tests that are easier to
compute have been proposed in the econometric literature (Pesaran; Pesaran and Deaton;
Aneuryn-Evans and Deaton). These tests, however, were developed for specific problems.

Simulation approaches to conducting Cox’s test that are applicable to many problems
have been proposed. One approach, due to Pesaran and Pesaran (1993; 1995), uses stochastic
simulation to calculate the numerator of the Cox test statistic. A second approach to
conducting Cox’s test is to use Monte Carlo hypothesis testing procedures based on the log-
likelihood ratio statistic. Cox-type tests using this approach have been developed by Williams
and Lee and Brorsen. Only the Lee and Brorsen test uses Noreen’s Monte Carlo hypothesis
testing procedures. These procedures allow directly computing the significance level (p-value)
of the test statistic being used. The Lee and Brorsen test is considered here.

Since introduced by Barnard, Monte Carlo tests have received considerable attention.
Monte Carlo tests are very useful when the distribution of the test statistic being used is
unknown or difficult to obtain analytically (Hope; Noreen). Monte Carlo tests have excellent
size and power properties (Hall and Titterington; Hope). In particular, a Monte Carlo test that
is based on an asymptotically pivotal statistic has better size properties than the corresponding
asymptotic test (Hall and Titterington). Hope showed that Monte Carlo tests have powers that
are very close to those of most uniformly powerful tests provided that sufficient random
samples are used. Hall and Titterington proved that the excellent power properties of Monte Carlo tests hold even if the test statistic used is not asymptotically pivotal, but the same is not true for the size properties.

The log-likelihood ratio statistic used in previous Monte Carlo tests is not asymptotically pivotal. An alternative Cox-type test is proposed here. This test uses Monte Carlo hypothesis testing procedures based on an asymptotically pivotal statistic. The finite sample performances of this test are compared with those of Pesaran and Pesaran (1993; 1995) and Lee and Brorsen test procedures through Monte Carlo experiments.

**Pesaran and Pesaran and Lee and Brorsen Test Procedures**

Consider the following two nonnested hypotheses:

\[ H_0^* : \ y \sim f(x, \theta_0) \]  
\[ H_1 : \ y \sim g(z, \theta_1) \]  

where \( y \) is a \( T \times 1 \) vector of dependent variables, \( x \) and \( z \) are \( T \times (K_0 + 1) \) and \( T \times (K_1 + 1) \) matrices of independent variables, \( \theta_0 \) and \( \theta_1 \) are unknown vectors of parameters, \( f \) and \( g \) are density functions, \( K_0 \) and \( K_1 \) are the number of independent variables under \( H_0 \) and \( H_1 \), respectively, and \( T \) is the number of observations. For the test of \( H_0 \) against the alternative hypothesis \( H_1 \), Pesaran and Pesaran and Lee and Brorsen proposed the following test procedures.
Pesaran and Pesaran's Test (PP)

Pesaran and Pesaran (1993; 1995) used the standardized version of the Cox statistic. The expected value of the log-likelihood ratio is approximated by the Kullback-Leibler measure of closeness of $H_0$ with respect to $H_1$. The simulated estimator of the closeness measure, $C_K(\hat{\theta}_0, \hat{\theta}_1(R))$, is obtained as:

$$C_K(\hat{\theta}_0, \hat{\theta}_1(R)) = \frac{1}{R} \sum_{j=1}^{R} (L_0(\hat{\theta}_0, y_j) - L_1(\hat{\theta}_1(R), y_j)), \quad (3)$$

where $R$ is the number of random samples generated using the maximum likelihood parameter estimates $\hat{\theta}_0$ under $H_0$, $\hat{\theta}_1(R)$ is a consistent estimator of the probability limit of $\Theta_1$, and $L_0(\hat{\theta}_0, y_j)$ and $L_1(\hat{\theta}_1(R), y_j)$ are the log-likelihoods functions for the $j^{th}$ random sample under $H_0$ and $H_1$, respectively. The estimators $\hat{\theta}_0$ and $\hat{\theta}_1(R)$ are thus treated as fixed (Pesaran and Pesaran, 1993). Note that $y_j$ is the $j^{th}$ vector of the $T$ artificially generated observations on the dependent variable $y$.

Pesaran and Pesaran (1995) indicated three asymptotically equivalent methods for computing the variance of the Cox's statistic. The method based on the "outer-product" expression of the information matrix will not be used here since it often yields negative values for the variance (Pesaran and Pesaran, 1993; 1995). The variance obtained using the "inner-product" expression of the information matrix will be referred to as $\hat{\vartheta}_{ab}$, The simplified version of the variance will be noted $\hat{\vartheta}_{ab}$. 


These results allow consistently estimating the standardized Cox statistic as:

\[ S_0(R) = \frac{\sqrt{n}T_0(R)}{\sqrt{\hat{\nu}_d^2}} \]  

(4)

where \( T_0(R) = L_{01} - C_R(\hat{\theta}_0, \hat{\theta}_1(R)) \) under \( H_0 \). \( \hat{\nu}_d^2 \) could also be used in the denominator of \( S_0(R) \).

Lee and Brorsen’s Test (MC(LB))

Lee and Brorsen used the log-likelihood ratio as the test statistic. Using Noreen’s approach, the significance level of the test is calculated as (Lee and Brorsen):

\[ p - value = \frac{(\text{numb}(L_{0j} - L_{ij}) \leq L_{0j}) + 1}{R + 1} \]  

(5)

where \( L_{0j} \) and \( L_{ij} \) are the log-likelihoods obtained from the jth random sample under \( H_0 \) and \( H_1 \), respectively, and numb[] means the number of elements of the set for which the specified relationship is true.

An Alternative Test Procedure

An alternative test procedure for conducting Cox’s test is proposed in this section. The test is implemented using Noreen’s Monte Carlo hypothesis testing approach. Unlike MC(LB), however, the test statistic used here is asymptotically pivotal. The test statistic is the standardized version of the Cox test. The variance can be computed using any of the methods
discussed above. The expected value of \( L_{01} \), however, is calculated by simulation as follows:

\[
\hat{E}_0( L_{01} ) = \frac{1}{R} \sum_{j=1}^{R} L_{01j},
\]

where \( L_{01j} = L_0( \hat{\theta}_0, y_j ) - L_i( \hat{\theta}_{ij}, y_j ) \). \( \hat{\theta}_0 \) is the maximum likelihood estimator of \( \theta \) for the \( j \)th random sample, and all other parameters and variables are defined as previously. Note that, contrary to Pesaran and Pesaran, the parameter estimates used here to calculate the expected value of the log-likelihood ratio are not treated as fixed. The simulated log-likelihood ratios computed here are independently distributed random variables. Thus, by the strong law of large numbers, \( \hat{E}_0( L_{01} ) \) converges to the true value of \( E_0( L_{01} ) \) as the number of random samples \( R \) and the number of observation \( T \) increase.

Under \( H_0 \), the standardized Cox test statistic (ST) can be consistently estimated as:

\[
ST_0 = \frac{\sqrt{n} [ L_{01} - \hat{E}_0( L_{01} )]}{\sqrt{\hat{\sigma}^2_{d0}}},
\]

(6)

\( ST_0 \) can be computed using \( \hat{\sigma}^2_{d0} \) as well.

\( ST_0 \) is the value of the test statistic for the actual data. To implement the Monte Carlo test, the values of the Cox statistic for each \( j \)th random sample are computed as:

\[
ST_{0j} = \frac{\sqrt{n} [ L_{01j} - \hat{E}_0( L_{01j} )]}{\sqrt{\hat{\sigma}^2_{d0j}}},
\]

(8)

where \( L_{01j} \) is defined as above, \( \hat{\sigma}^2_{d0j} \) is the variance of the simulated log-likelihood ratio for the \( j \)th
sample, and $\hat{E}_0( L_{0ij})$ is obtained by simulation. R values of $S_{T0j}$ are computed and the p-value of the Cox test is obtained as:

$$p\cdot value = \frac{\text{numb}[S_{T0j} \leq S_{T0}] + 1}{R + 1}.$$  

(9)

Monte Carlo Experiments

The Monte Carlo experiments are conducted using data from a real world problem. The design matrix contains weekly data on hamburger prices and advertising expenditures. These data are taken from Griffiths, Hill, and Judge (pp. 295). The following two nonnested models are considered:

\[ H_0: \quad tr_t = \alpha_0 + \alpha_1 \log(a_t) + \alpha_2 \log(p_t) + e_{0t} \]  

(10)

\[ H_1: \quad \log(tr_t) = \beta_0 + \beta_1 \log(a_t) + \beta_2 \log(p_t) + e_{1t}, \]  

(11)

where $tr_t$ is weekly hamburger chain’s total receipts, $p_t$ is price, $a_t$ is advertising expenditures, and the $e_t$’s are normally distributed with means equal zeros and variances $\sigma_0$ and $\sigma_1$, respectively. These two functional forms closely approximate each other. We purposely selected a case where it would be difficult to discriminate between the two hypotheses.

When the semi-logarithmic model is the true model ($H_0$), the dependent variable is generated according to:

\[ tr_t = 82.514 + 24.841 \log(a_t) - 21.509 \log(p_t) + \hat{e}_0. \]  

(12)
When the log-linear model is true ($H_1$), $\log(tr_t)$ is generated as:

$$\log(tr_t) = 4.466 + 0.206 \log(a_t) - 0.177 \log(P_t) + \hat{e}_t. \quad (13)$$

The parameter estimates of these data generating processes are obtained using 78 observations on $tr_t$, $a_t$, and $p_t$. Both $\hat{e}_0$ and $\hat{e}_1$ are generated using the RNDN command of GAUSS and
standard errors $\sigma_0 = (2.327, 6.984)$ and $\sigma_1 = (0.020, 0.055)$, respectively. Note that 6.984 and 0.055 are the actual estimates of the standard errors of the error terms under $H_0$ and $H_1$, respectively. The standard errors are varied to determine the effects of the variances on the Monte Carlo results. A different seed is used only when $\sigma_0$ and $\sigma_1$ are varied.

The experiments are conducted using samples of 20, 50, 100, and 200 observations. The design matrix is duplicated when the samples of 100 and 200 observations are used. The number of replications is 1000 for samples sizes 20 and 50, and 500 for samples sizes 100 and 200. For Pesaran and Pesaran test procedure, the measure of closeness is calculated using 100 random samples. The number of random samples (R) used in the Monte Carlo tests is 99. Conducting the experiments with larger numbers of random samples did not substantially change the results. The log-likelihood functions of the log-linear model are adjusted for Jacobian terms.

Monte Carlo Results

The sizes and powers of the Pesaran and Pesaran (PP) test, MC(LB), and ST are reported in tables 1 and 2 along with their standard errors in parentheses. The standard errors were
obtained as the square root of $N^{-1} \alpha (1 - \alpha)$, where $N$ is the number of replications and $\alpha$ is the estimated size or power. The nominal significance level selected is 0.05.

All of the tests have high power, which make them good candidates for discriminating among nonnested regression models. There is, however, a clear difference between the sizes of the Monte Carlo tests and the PP test. Consider the case where the inner product of the information matrix is used to calculate the variance of the Cox test. The size of the PP test is too high, except for samples of sizes 100 and 200 in table 2. Pesaran and Pesaran (1995) found similar results. Similarly, when the simplified version of the variance is used, the PP test over-rejects for all sample sizes but sample size 200 in table 1. In table 2, the size of the PP test is also incorrect in small samples, but sometimes the PP test under-rejects.

As expected, ST has correct size for all samples, irrespective of which version of the variance is used. Interestingly, the MC(LB) test also has correct size for all samples. The excellent size properties of the MC(LB) test could not be guaranteed a priori since the log-likelihood ratio is not an asymptotically pivotal statistic.

The ST and MC(LB) tests have very similar powers for all sample sizes. Since the PP test tends to reject too often, it is not surprising that it often has the largest power. In the few cases that the size of the PP test is correct, its power is roughly equal to the powers of the ST and MC(LB) tests.

**Conclusions**
This paper has determined the finite sample performances of three simulated Cox-type tests. The first test is not a true Monte Carlo test and is due to Pesaran and Pesaran. It uses stochastic simulation to compute the numerator of the Cox test statistic and tests are conducted based on asymptotic normality. The second test uses Monte Carlo hypothesis testing procedures to discriminate between two separate families of hypotheses. In this approach, the log-likelihood ratio is considered as the test statistic. The third Cox-type test is a new Monte Carlo test. Unlike the second test, however, the test statistic used in the third approach is asymptotically pivotal. Pivotalness assures that the excellent size properties of Monte Carlo tests hold.

The results of the Monte Carlo experiments show that, in general, the Pesaran and Pesaran test has incorrect size. As expected, the test proposed here has excellent size properties for all sample sizes, irrespective of which version of the variance is being used. Interestingly, the Monte Carlo test based on the log-likelihood ratio also has excellent size and power properties for all sample sizes, even though the log-likelihood ratio statistic is not asymptotically pivotal.

On the basis of their sizes, the Monte Carlo tests are clearly preferred to the Pesaran and Pesaran test. When the size of the Pesaran and Pesaran test is correct, its power is close or even equal to the powers of the Monte Carlo tests. Thus, we would recommend against using the Pesaran and Pesaran test. The Monte Carlo tests (ST and MC(LB)) have similar powers. The MC(LB) test is by far the simplest to compute and therefore we recommend that it be used in applied work.

Table 1. Monte Carlo Results. The Semi-logarithmic Model (H_0) is the True Model. The Log-linear Model is the Alternative Model.
<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Test</th>
<th>( \sigma_0 = 2.327 )</th>
<th>( \sigma_0 = 6.984 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Size</td>
<td>Power</td>
</tr>
<tr>
<td>20</td>
<td>PP1</td>
<td>0.121* (0.010)</td>
<td>0.657 (0.015)</td>
</tr>
<tr>
<td></td>
<td>PP2</td>
<td>0.144* (0.011)</td>
<td>0.738 (0.014)</td>
</tr>
<tr>
<td></td>
<td>ST1</td>
<td>0.042 (0.006)</td>
<td>0.487 (0.016)</td>
</tr>
<tr>
<td></td>
<td>MC(LB)</td>
<td>0.059 (0.007)</td>
<td>0.529 (0.016)</td>
</tr>
<tr>
<td>50</td>
<td>PP1</td>
<td>0.098* (0.009)</td>
<td>0.935 (0.008)</td>
</tr>
<tr>
<td></td>
<td>PP2</td>
<td>0.140* (0.011)</td>
<td>0.953 (0.007)</td>
</tr>
<tr>
<td></td>
<td>ST1</td>
<td>0.052 (0.007)</td>
<td>0.892 (0.010)</td>
</tr>
<tr>
<td></td>
<td>MC(LB)</td>
<td>0.047 (0.007)</td>
<td>0.946 (0.007)</td>
</tr>
<tr>
<td>100</td>
<td>PP1</td>
<td>0.080* (0.012)</td>
<td>0.998 (0.002)</td>
</tr>
<tr>
<td></td>
<td>PP2</td>
<td>0.104* (0.014)</td>
<td>1.000 (0.000)</td>
</tr>
<tr>
<td></td>
<td>ST1</td>
<td>0.052 (0.010)</td>
<td>0.998 (0.002)</td>
</tr>
<tr>
<td></td>
<td>MC(LB)</td>
<td>0.050 (0.010)</td>
<td>0.996 (0.003)</td>
</tr>
<tr>
<td>200</td>
<td>PP1</td>
<td>0.070 (0.011)</td>
<td>1.000 (0.000)</td>
</tr>
<tr>
<td></td>
<td>PP2</td>
<td>0.082* (0.012)</td>
<td>1.000 (0.000)</td>
</tr>
<tr>
<td></td>
<td>ST1</td>
<td>0.052 (0.010)</td>
<td>1.000 (0.000)</td>
</tr>
<tr>
<td></td>
<td>MC(LB)</td>
<td>0.054 (0.010)</td>
<td>1.000 (0.000)</td>
</tr>
</tbody>
</table>

Note: An asterisk means the estimated size is significantly different from 0.05. Subscripts 1 and 2 refer to the inner product and simplified versions of the variance, respectively.
Table 2. Monte Carlo Results. The Log-linear Model ($H_1$) is the True Model. The Semi-Logarithmic Model is the Alternative Model.

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Test</th>
<th>$\sigma_1 = 0.020$</th>
<th>Power</th>
<th>$\sigma_1 = 0.055$</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Size</td>
<td></td>
<td>Size</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>PP$_1$</td>
<td>0.108* (0.010)</td>
<td>0.608 (0.015)</td>
<td>0.117* (0.010)</td>
<td>0.441 (0.016)</td>
</tr>
<tr>
<td></td>
<td>PP$_2$</td>
<td>0.137* (0.011)</td>
<td>0.674 (0.015)</td>
<td>0.152* (0.011)</td>
<td>0.503 (0.016)</td>
</tr>
<tr>
<td></td>
<td>ST$_1$</td>
<td>0.044 (0.006)</td>
<td>0.372 (0.015)</td>
<td>0.040 (0.006)</td>
<td>0.161 (0.012)</td>
</tr>
<tr>
<td></td>
<td>ST$_2$</td>
<td>0.053 (0.007)</td>
<td>0.442 (0.016)</td>
<td>0.050 (0.007)</td>
<td>0.201 (0.013)</td>
</tr>
<tr>
<td></td>
<td>MC(LB)</td>
<td>0.059 (0.007)</td>
<td>0.439 (0.016)</td>
<td>0.058 (0.007)</td>
<td>0.180 (0.012)</td>
</tr>
<tr>
<td>50</td>
<td>PP$_1$</td>
<td>0.053 (0.007)</td>
<td>0.936 (0.008)</td>
<td>0.075* (0.008)</td>
<td>0.630 (0.015)</td>
</tr>
<tr>
<td></td>
<td>PP$_2$</td>
<td>0.082* (0.009)</td>
<td>0.958 (0.006)</td>
<td>0.095* (0.010)</td>
<td>0.698 (0.015)</td>
</tr>
<tr>
<td></td>
<td>ST$_1$</td>
<td>0.051 (0.007)</td>
<td>0.853 (0.011)</td>
<td>0.049 (0.007)</td>
<td>0.353 (0.015)</td>
</tr>
<tr>
<td></td>
<td>ST$_2$</td>
<td>0.062 (0.007)</td>
<td>0.893 (0.010)</td>
<td>0.057 (0.007)</td>
<td>0.425 (0.016)</td>
</tr>
<tr>
<td></td>
<td>MC(LB)</td>
<td>0.044 (0.006)</td>
<td>0.897 (0.010)</td>
<td>0.047 (0.007)</td>
<td>0.445 (0.016)</td>
</tr>
<tr>
<td>100</td>
<td>PP$_1$</td>
<td>0.036 (0.008)</td>
<td>0.996 (0.003)</td>
<td>0.056 (0.010)</td>
<td>0.858 (0.016)</td>
</tr>
<tr>
<td></td>
<td>PP$_2$</td>
<td>0.052 (0.010)</td>
<td>0.998 (0.002)</td>
<td>0.074 (0.012)</td>
<td>0.886 (0.014)</td>
</tr>
<tr>
<td></td>
<td>ST$_1$</td>
<td>0.056 (0.010)</td>
<td>0.994 (0.003)</td>
<td>0.062 (0.011)</td>
<td>0.648 (0.021)</td>
</tr>
<tr>
<td></td>
<td>ST$_2$</td>
<td>0.062 (0.011)</td>
<td>0.996 (0.003)</td>
<td>0.072 (0.012)</td>
<td>0.696 (0.021)</td>
</tr>
<tr>
<td></td>
<td>MC(LB)</td>
<td>0.058 (0.010)</td>
<td>0.996 (0.003)</td>
<td>0.056 (0.010)</td>
<td>0.722 (0.020)</td>
</tr>
<tr>
<td>200</td>
<td>PP$_1$</td>
<td>0.034* (0.008)</td>
<td>1.000 (0.000)</td>
<td>0.058 (0.010)</td>
<td>0.974 (0.007)</td>
</tr>
<tr>
<td></td>
<td>PP$_2$</td>
<td>0.042 (0.009)</td>
<td>1.000 (0.000)</td>
<td>0.064 (0.011)</td>
<td>0.976 (0.007)</td>
</tr>
<tr>
<td></td>
<td>ST$_1$</td>
<td>0.058 (0.010)</td>
<td>1.000 (0.000)</td>
<td>0.060 (0.011)</td>
<td>0.932 (0.011)</td>
</tr>
<tr>
<td></td>
<td>ST$_2$</td>
<td>0.062 (0.011)</td>
<td>1.000 (0.000)</td>
<td>0.058 (0.010)</td>
<td>0.958 (0.009)</td>
</tr>
<tr>
<td></td>
<td>MC(LB)</td>
<td>0.054 (0.010)</td>
<td>1.000 (0.000)</td>
<td>0.040 (0.009)</td>
<td>0.936 (0.011)</td>
</tr>
</tbody>
</table>

Note: An asterisk means the estimated size is significantly different from 0.05. Subscripts 1 and 2 refer to the inner product and simplified versions of the variance, respectively.
References


