Pretest Estimation in the Random Parameters Logit Model

June 15, 2010

Tong Zeng[1] and R. Carter Hill[2]

Abstract

In this paper we use quasi-Monte Carlo sampling experiments to examine the properties of pretest estimators in the random parameters logit model. The pretests are for the presence of random parameters. We study the Lagrange Multiplier (LM), Likelihood Ratio (LR) and Wald tests, using the conditional logit model as the restricted model. The LM test is the fastest test to implement among these three test procedures since it only uses restricted, conditional logit, estimates. However, the LM-based pretest estimator provides poor risk properties in our experiment. The ratio of LM-based pretest estimator RMSE to the random parameters logit model estimator RMSE diverges from one with increases in the standard deviation of the parameter distribution. The LR and Wald tests exhibit properties of consistent tests, with the power approaching one as the specification error increases, so that the pretest estimator is consistent. We explore the power of these three tests for the random parameters by calculating the empirical percentile values, size and rejection rates of the test statistics. We find the power of LR and Wald tests decreases with increases in the mean of the coefficient distribution. The LM test has the weakest power for presence of the random coefficient in the RPL model.

Keywords: pretest estimator, Lagrange Multiplier, Likelihood Ratio and Wald tests

[1] Department of Economics, Louisiana State University, Baton Rouge, LA 70803 Email: tzeng1@lsu.edu
[2] Department of Economics, Louisiana State University, Baton Rouge, LA, 70803 Email: eohill@lsu.edu
We thank Professors Greene, Fomby and two referees for their comments.

1 Introduction

In this paper, we use quasi-Monte Carlo sampling experiments to examine the properties of pretest estimators in the random parameters logit model or also called mixed logit model. The pretests are for the presence of random parameters. We study the Lagrange Multiplier (LM), Likelihood Ratio (LR) and Wald tests, using the conditional logit model as the restricted model. Unlike the conditional logit model, the mixed logit model does not impose the *Independence from Irrelevant Alternatives* (IIA) assumption. The mixed logit model can capture random taste variation among individuals and allow the unobserved factors of utility to be correlated over time as well.

The choice probabilities in mixed logit cannot be calculated exactly because they involve a multi-dimensional integral which does not have closed form. In applications, the integral can be approximated through simulation using pseudo-random numbers. The requirement of a large number of draws during the simulation leads to long computational times. We are interested in testing the randomness of the mixed logit coefficients and the properties of pretest estimators in the mixed logit. If the coefficients are not random, then the mixed logit model reduces to the simpler conditional logit model. The most commonly used test procedures for this purpose are the Wald (or t-) test and the Likelihood ratio test for the significance of the coefficient random components. The problem is that in order to implement these tests the mixed logit model must be estimated. It would be much faster to implement a Lagrange multiplier test, as the restricted estimates, from the model with non-random coefficients, come from the conditional logit model, which is easily estimated.

We use quasi-Monte Carlo experiments in the context of one and two parameters choice models with four alternatives to examine the risk properties of pretest estimator based on LM, LR and Wald tests. We explore the power of the three tests for the random parameters by calculating the empirical 90th and 95th percentile values of the three tests statistics and examine rejection rates of the three tests by using the empirical 90th and 95th percentile values as the critical values for 10% and 5% significance level. We find the pretest parameters estimators based on the LR and Wald statistics have RMSE that is less than that of the random parameters logit model when the parameter variance is small, but that RMSE is worse than that of the random parameters logit model over the remaining parameter space. The LR and Wald tests exhibit properties of consistent tests, with the power approaching one as the specification error increases. However, the power of LR and Wald tests decreases with increases in the mean of the coefficient distribution. The ratios of LM-based pretest estimator RMSE to that RMSE of the random parameters logit model rise and become further away from one with increases in the standard deviation of the parameter variance.

The plan of the paper is as follows. In the following section, we review the conditional logit model and introduce the mixed logit specification. In Section 3, we introduce quasi-random numbers and describe our Monte Carlo experiments. We also show the efficiency of the quasi-random numbers in this section. Section 4 summarizes the mean square error properties of the pretest estimator based on LM, LR and Wald tests, and the size corrected rejection rates of these three tests. Some conclusions and recommendations are given in Section 5.

2 Conditional and Mixed Logit Models

The conditional logit model is frequently used in applied econometrics. The related choice probability can be computed conveniently without multivariate integration. The *Independence from Irrelevant Alternatives* (IIA) assumption of the conditional logit model is inappropriate in many choice situations, especially for the choices that are close substitutes, which was first pointed out by Chipman (1960) and Debreu (1960). The IIA assumption arises because in logit models the unobserved components of utility are independent and identically Type I extreme value distributions. This is violated in many cases, such as when unobserved factors that affect the choice persist over time.

The mixed logit model was first applied by Body and Mellman (1980) and Cardell and Dunbar (1980) to forecast automobile choices by individuals. Unlike the probit models, it is fully flexible because its unobserved utility is not limited to the normal distribution. It decomposes the random parts of utility into two parts. One has the independent, identical type I extreme value distribution, and the other representing individual tastes can be any distribution. The related utility associated with alternative i as evaluated by individual n in the mixed logit model is written as:

(1) $$U_{ni} = \beta_n' x_{ni} + \varepsilon_{ni}$$

Where x_{ni} are observed variables for alternative i and individual n, β_n is a vector of coefficients for individual n varying over individuals in the population with density function $f(\beta)$, and ε_{ni} is iid extreme value, which is independent of β_n and x_{ni}. If β_n is fixed, the mixed logit becomes conditional logit model and the choice probability $L_{ni}(\beta)$ for individual n choosing alternative i is:

(2) $$L_{ni}(\beta) = \frac{e^{\beta' x_{ni}}}{\sum_j e^{\beta' x_{nj}}}$$

In the mixed logit model we specify a distribution for the random coefficients $f(\beta\,|\,\theta)$, where θ indicates the parameters' distribution, such the mean and variances. These are the parameters to be ultimately estimated. The choice probability executing the log likelihood function is:

(3) $$P_{ni} = \int \frac{e^{\beta' x_{ni}}}{\sum_i e^{\beta' x_{ni}}} f(\beta\,|\,\theta) d\beta = \int L_{ni}(\beta) f(\beta\,|\,\theta) d\beta$$

Hensher and Greene (2001) discuss how to choose an appropriate distribution for random coefficients. In the following section, we will describe how to estimate the unknown parameters (θ) and introduce the quasi-Monte Carlo methods.

3 Quasi-Monte Carlo Methods

3.1 Simulated Log-likelihood Function

Unlike the conditional logit model, the mixed logit probability cannot be calculated exactly, since the related integral does not have a closed form. The choice probability can be estimated through simulation and the unknown parameters (θ) can be estimated by maximizing the simulated log-likelihood function. With simulation, a value of β labeled as β^r, representing the rth draw, is drawn randomly from a previously specified distribution. The standard logit $L_{ni}(\beta)$ in equation (2) can be calculated given β^r. Repeating this process R times, and the simulated probability of individual n choosing alternative i is obtained by averaging $L_{ni}(\beta^r)$:

(4) $$\breve{P}_{ni} = \frac{1}{R} \sum_{r=1}^{R} L_{ni}(\beta_n^r)$$

The simulated log-likelihood function is:

(5) $$SLL(\theta) = \sum_{n=1}^{N} \sum_{i=1}^{J} d_{ni} \ln \breve{P}_{ni}$$

where $d_{ni} = 1$ if individual n chooses alternative i and $d_{ni} = 0$ otherwise. Each individual is assumed to make choices independently and faces choice once.

3.2 The Halton Sequences

The classical Monte Carlo method is used above to estimate the probability P_{ni}. It reduces the integration problem to the problem of estimating an expected value on the basis of the strong law of large numbers. In general terms, the classical Monte Carlo method is described as a numerical method based on random sampling. The random sampling here uses pseudo-random numbers. In terms of the number of pseudo-random numbers N, it gives us a probabilistic error bound, since there is never any guarantee that the expected accuracy is achieved in a concrete calculation (Niederreiter, 1992, page7), also called convergence rate, $O(N^{-1/2})$ for numerical integration. It represents the stochastic character of the classical-Monte Carlo method. A wonderful feature of the classical Monte Carlo method is that the convergence rate of the numerical integration does not depend on the dimension of the integration. Good estimates, however, require a large number of pseudo-random numbers, which leads to long computational times. To reduce the cost of long run times, we can replace the pseudo-random numbers with a constructed set of points. The same or even higher estimation accuracy can be reached with fewer points. The essence of the number theoretic method (NTM) is to find a set of uniformly scattered points over an s-dimensional unit cube. Such set of points obtained by NTM is usually called a set of quasi-random numbers or a number theoretic net. Sometimes it can be used in the classical Monte Carlo method to achieve a significantly higher accuracy. The difference between the quasi-Monte Carlo method and the classical Monte Carlo method is the quasi-Monte Carlo method uses quasi-random numbers instead of pseudo-random numbers. In fact, there are several classical methods to construct the quasi-random numbers. Here we use the Halton sequences proposed by Halton (1960). Bhat (2001) found the error measures of the estimated parameters was smaller using 100 Halton draws than 1000 random numbers in mixed logit model.

The Halton sequences are based on the base- p number system which implies that any integer n can be written as:

(6) $$n \equiv n_M n_{M-1} \cdots n_2 n_1 n_0 = n_0 + n_1 p + n_2 p^2 + \cdots + n_M p^M$$

where $M = [\log_p^n] = [\ln n / \ln p]$, square brackets denoting the integral part. The base is p which can be any integer except 1; n_i is the digit at position i, $0 \le i \le M$; $0 \le n_i \le p-1$ and p^i is the weight of position i. With the base p =10, the integer $n = 468$ has $n_0 = 8, n_1 = 6, n_2 = 4$.

Using the base- p number system, we can construct one and only one fraction φ that is smaller than 1 by writing n with a different base number system and reversing the order of the digits in n. It is called *the radical inverse function* defined as the follows:

(7) $$\phi = \phi_p(n) = 0.n_0 n_1 n_2 \cdots n_M = n_0 p^{-1} + n_1 p^{-2} + \cdots + n_M p^{-M-1}$$

The Halton sequence of length N is developed from *the radical inverse function* and the points of the Halton sequence are $\varphi_p(n)$ for n =1, $2 \cdots N$ where p is a prime number. The k -dimensional sequence is defined as

(8) $$\phi_n = (\varphi_{p_1}(n), \varphi_{p_2}(n), \cdots \varphi_{p_k}(n))$$

where $p_1, p_2, \cdots p_k$ are prime to each other and are always chosen from the first k primes to achieve the smaller discrepancy.

In applications, Halton sequences are used to replace random number generators to produce points in the interval [0, 1]. The points of the Halton sequence are generated iteratively. A one-dimensional Halton sequence based on prime p divides 0-1 interval into p segments. It systematically fills in the empty space by iteratively dividing each segment into smaller p segments. The position of the points is determined by the base which is used to construct iteration. A large base implies more points in each iteration, or a long cycle. Due to the high correlation among the initial points of Halton sequence, the first ten points of the sequences are usually discarded in applications (Morokoff and Caflisch, 1995; Bratley et al., 1992).

Compared to the pseudo-random numbers, the coverage of the points of Halton sequence are more uniform, since the pseudo-random numbers may cluster in some areas and leave some areas uncovered. It can be easily seen from Figure 1, which is based on a figure from Bhat (2001). Figure 1 (a) is a plot of 200 points taken from uniform distribution of two dimensions using pseudo-random numbers. Figure 1 (b) is a plot of 200 points obtained by the Halton sequence. The latter scatters more uniformly on the unit square than the former. Since the points generated from the Halton sequences are deterministic points, unlike the classical-Monte Carlo method, quasi-Monte Carlo provides a deterministic error bound instead of probabilistic error bound. It is also called the discrepancy in the literature of number theoretic methods. The smaller the discrepancy, the more evenly the quasi-random numbers spread over the domain. The deterministic error bound of quasi-Monte Carlo method with the Halton sequences is $O(N^{-1}(\ln N)^k)$ (Halton, 1960), which is smaller than the probabilistic error bound of classical-Monte Carlo method $O(N^{-1/2})$. The shortcoming of the Halton sequences is it needs a large number of points to insure a uniform scattering on the domain for large dimensions, usually $k \geq 10$. It increases the computational time and also leads to high correlation among higher coordinates of the Halton sequences. Sometimes, Halton sequences are refined by scrambling the points of Halton sequence points (Bhat 2003).

Monte Carlo simulation methods require random samples from various distributions. A *Discrepancy-preserving transformation* is often applied in quasi-Monte Carlo simulation to transform a set of n quasi-random number $\{Y_k = (Y_{k1}, \cdots Y_{ks}), k = 1, \cdots, n\}$ generated from the s-dimensional unit cube, with discrepancy d, to a random variable x with another statistical distributions by solving $x_k = (F_1^{-1}(Y_{k1}), \cdots, F_s^{-1}(Y_{ks})), k = 1, \cdots, n$. We achieve the same discrepancy d with respect to $F(x)$. $F(x)$ is an increasing continuous multivariate distribution function $F(x) = F(x_1, \cdots, x_s) = \prod_{i=1}^{s} F_i(x_i)$ and $F_i(x_i)$ is the marginal distribution function of x (Fang and Wang, 1994, Chapter 4). Due to the faster convergence rate and fewer draws, less computational time is needed. We apply the Halton sequences with maximum simulated likelihood method to estimate mixed logit model. How to choose the number of Halton draws is an issue in application of the Halton sequences. Different researchers provide different suggestions. To determine the number of Halton draws in our experiments, we compare the results of estimated mixed logit parameters with different sets of Halton draws and pseudo-random numbers on the basis of the bias, Monte Carlo standard deviations, the average nominal standard errors, the ratio of average nominal standard errors to the Monte Carlo standard deviations and the root mean square error (RMSE) of random coefficient estimates.

3.3 Monte Carlo Experiment Design

Our experiments are based on a mixed logit model which has no intercept term, with one or two coefficients that are independent of each other. In our experiments, each individual faces four mutually exclusive alternatives on one choice occasion. The utility of individual n choosing alternative i is:

(9) $$U_{ni} = \beta_n' x_{ni} + \varepsilon_{ni}$$

The explanatory variables for each individual and each alternative x_{ni} are generated from independent standard normals. The coefficients for each individual β_n are generated from normal distribution $N(\overline{\beta}, \overline{\sigma}_{\beta}^{2})$. These values of x_{ni} and β_n are held fixed over each experiment design. The choice probability for each individual is generated by comparing the utility of each alternative:

(10) $$I_{ni}^{r} = \begin{cases} 1 & \beta_n' x_{ni} + \varepsilon_{ni}^{r} > \beta_n' x_{nj} + \varepsilon_{nj}^{r} \\ 0 & Otherwise \end{cases} \qquad \forall i \neq j$$

The indicator function I_{ni}^{r} represents whether individual n chooses alternative i or not based on utility function. The values of errors are generated from iid extreme value type I distribution, ε_{ni}^{r} representing the *r*th draw. We calculate and compare the utility of each alternative using these random errors. This process is repeated 1000 times. The simulated choice probability P_{ni} for each individual n choosing alternative i is

(11) $$P_{ni} = \frac{1}{1000} \sum_{r=1}^{1000} I_{ni}^{r}$$

The dependent variables y_{ni} are determined by these values of simulated choice probabilities. In our experiments, we choose the estimation sample size $N = 200$ and generate 999 Monte Carlo samples with specific mean and variance that we set for the coefficient distribution. To test the efficiency of the mixed logit estimators with the Halton sequence, we use 25, 100, and 250 Halton draws and 1000 random draws to estimate the mean and variance of the coefficient distribution respectively.

3.4 Findings in Simulation Efficiency

In the one coefficient case, the two parameters of interest are $\bar{\beta}$ and $\bar{\sigma}_{\beta}$. We denote the estimates of these parameters as $\hat{\beta}$ and $\hat{\sigma}_{\beta}$. Table1 reports for the $NSAM = 999$ Monte Carlo samples. The Monte Carlo average of the estimated mixed logit parameters and the error measures of mixed logit estimates with one random parameter are calculated as follows:

MC average $\bar{\hat{\beta}} = \sum \hat{\beta}_i / NSAM$

MC standard deviation (s.d.) of $\hat{\beta} = \sqrt{\sum (\hat{\beta}_i - \bar{\hat{\beta}})^2 / (NSAM - 1)}$

Average nominal standard error (s.e.) of $\hat{\beta} = \sum \sqrt{\hat{\text{var}}(\hat{\beta}_i)} / NSAM$

Root mean square error (RMSE) of $\hat{\beta} = \sqrt{\sum (\hat{\beta}_i - \bar{\beta})^2 / NSAM}$

From Table 1, increases in the number of Halton draws changes error measures only by small amounts. The number of Halton draws influences the RMSE of mixed logit estimators slightly. The Monte Carlo average value of $\hat{\beta}_i$ underestimates the true parameter $\bar{\beta} = 1.5$ by less than 2%.

With Halton draws, the average nominal standard errors of $\hat{\beta}$ are only 1% larger than the Monte Carlo standard deviations. The average nominal standard errors for $\hat{\sigma}_{\beta}$ are 20% larger than the Monte Carlo averages. With 100 Halton draws, the ratios of average nominal standard errors to Monte Carlo standard errors are closest to 1. Compared to the classical-Monte Carlo estimation, our results confirm the findings of Bhat (2001, *p*.691). We can reach almost the same RMSE of estimated parameters with only 100 Halton draws as compared with 1000 random draws, and computational time is considerably reduced with 100 Halton draws. Considering the relative accurate estimation of the standard errors of $\hat{\beta}$ and $\hat{\sigma}_{\beta}$ which are used to construct t-test, and the acceptable estimation time, we use 100 Halton draws to estimate the mixed logit parameters with only one random coefficient.

In Table 2, we use the same error measures and show the Monte Carlo average values of the estimated mixed logit parameters with two random coefficients. The true mean and standard deviation of new independent random coefficient distribution are 2.5 and 0.3, respectively. In Table 2, with increases in the number of Halton draws, the percentage changes of the Monte Carlo average values of the estimated mixed logit parameters are no more than 1%. Unlike the one random coefficient case, the Monte Carlo average values of estimated means of two independent random coefficient distributions are overestimated by 10%. However, the biases are stable and not sensitive to the number of Halton draws. From Table 2, the average nominal standard errors of $\hat{\beta}_{1i}$ and $\hat{\beta}_{2i}$ are underestimated and further away from the Monte Carlo standard deviations than in one random coefficient case. The ratios of the average nominal standard deviations to the Monte Carlo standard deviations of the estimated parameters are slightly closer to one with 100 Halton draws. Using 100 Halton draws also provides smaller root mean square errors of the estimated parameters. Based on these results, 100 Halton draws are also used in our two independent random coefficients mixed logit model. All of these factors lead us to conclude that increasing the number of Halton draws in our experiments will not greatly improve the RMSE of estimated mixed logit parameters. Since the convergence rate of the quasi-Monte Carlo method with Halton sequences in theory is mainly determined by the structure of the sequences, the simulation error will not considerably decline with increases in the number of Halton draws for each individual.

4 Pretest Estimators

Even though mixed logit model is a highly flexible model, it requires the use of simulation to obtain empirical estimates. It is desirable to have a specification test to determine whether the mixed logit is needed or not. The likelihood ratio (LR) and Wald tests are the most popular test procedures used for testing the significance of coefficient estimates. The problem is that in order to implement these tests the mixed logit model must be estimated. It is much faster to implement a Lagrange Multiplier (LM) test. It is interesting and important to examine the power of these three tests for the presence of the random coefficients in the mixed logit model. We use quasi-Monte Carlo experiments in the context of a one and two parameters choice model with four alternatives to examine the properties of pretest estimators in the random parameters logit model with LR, LM and Wald tests.

In our experiments, the LR, Wald and LM tests are constructed based on the null hypothesis $H_0 : \sigma_\beta = 0$ against the alternative hypothesis $H_1 : \sigma_\beta > 0$. With the one-tailed tests, when the null hypothesis is true, the true estimator will lie on the boundary of the parameter space. The asymptotic distribution of the estimator is complex, since it does not include a neighborhood of zero. The standard theory of the LR, Wald and LM tests is violated, which assumes that the true estimator lies inside an open set in the parameter space. Thus, the LR, Wald and LM statistics do not have usual chi-square asymptotic distribution. To solve this problem, some researchers such as Shapiro (1985), Hillier (1986), Gourieroux, Holly and Monfort (1982) and Andrews (2001) studied the asymptotic distribution of the LR and Wald statistics and found they have a mixture of chi-squared distributions. In our experiments, we use their results to analyze the power of the LR, LM and Wald tests.

In the one random parameter model, we set four different values for the parameter mean, $\bar{\beta} = \{0.5, 1.5, 2.5, 3.0\}$. Corresponding to each value of the mean $\bar{\beta}$, we set six different values for the standard deviation of the parameter distribution, $\bar{\sigma}_\beta = \{0, 0.15, 0.3, 0.8, 1.2, 1.8\}$. The restricted and unrestricted estimates come from the conditional logit and mixed logit model respectively. The LR, Wald and LM tests are constructed based on the null hypothesis $H_0 : \sigma_\beta = 0$ against the alternative hypothesis $H_1 : \sigma_\beta > 0$. The inverse of information matrix in the Wald and LM tests is estimated using BHHH (outer product of gradients).

Figure 2 shows the ratio of pretest estimator RMSE of $\bar{\beta}$ relative to the random parameters logit model estimator RMSE of $\bar{\beta}$ using the LR, Wald and LM tests at a 25% significance level. We choose a 25% significance level because 5% pretests are not optimal in many settings, and this is also true in our experiments. Under a one-tailed alternative hypothesis, the distribution of LR and Wald $\chi^2 -$ test statistics has a mixture of chi-square distributions (Gourieroux, Holly and Monfort 1982, Shapiro 1985, and Gourieroux and Monfort 1989). In the one parameter case, the $1-2\alpha$ quantile of standard chi-square is the critical value for significance level α (Andrews, 2001, *p*.713). Thus 0.455 is the critical value for a pretest estimator at 25% significance level. Figure 2 shows that the pretest estimators based on the LR and Wald statistics have RMSE that is less than that of the random parameters logit model when the parameter variance is small, but that RMSE is worse than that of the random parameters logit model over the remaining parameter space. The LR and Wald tests exhibit properties of consistent tests, with the power approaching one as the specification error increases, so that the pretest estimator is consistent. But the ratios of LM-based pretest estimator RMSE of $\bar{\beta}$ to that RMSE of the random parameters logit model rise and become further away from one with increases in the standard deviation of the parameter distribution. The poor properties of the LM-based pretest estimator arise from the poor power of the LM test in our experiments. It is interesting that even though the pretest estimator based on the LR and Wald statistics are consistent, the maximum risk ratio based on the LR and Wald tests increases in the parameter mean $\bar{\beta}$. The range over which the risk ratio is less than one also increases in the mean of the parameter distribution $\bar{\beta}$.

To explore the power of the three tests for the presence of the random coefficient in the mixed logit model further, we calculate the empirical 90th and 95th percentile value of the LR, Wald and LM statistics given the different combinations of means and standard deviations of the parameter distribution in the one random parameter model. The results in Table 3 show that the Monte Carlo 90th and 95th percentile values of the three tests change with the changes in the mean and standard deviation of parameter distribution. In general, the Monte Carlo critical values with different parameter means are neither close to 1.64 and 2.71 (the $1-2\alpha$ quantile of standard chi-square statistics for 10% and 5% significance level respectively) nor to the usual critical values 2.71 and 3.84. When $\bar{\beta} = 0.5$ and $\bar{\sigma}_{\beta} = 0$, the 90th and 95th empirical percentiles of LR, Wald and LM in our experiments both are greater than the asymptotic critical values 1.64 and 2.71. With increases in the true standard deviation of the coefficient distribution, the 90th and 95th empirical percentiles increase for the LR and Wald statistics, indicating that these tests will have some power in choosing the correct model with random coefficients. The corresponding percentile values based on the LM statistics decline, meaning that the LM test has declining power. An interesting feature of Table 3 is that most empirical percentile values based on the LR and Wald statistics decrease in the mean of coefficient distribution $\bar{\beta}$, at the same time, most of them based on the LM test increase in the parameter mean $\bar{\beta}$.

The results on the empirical percentiles of the LR, Wald and LM statistics imply the rejection rates of the three tests will vary depending on the mean and standard deviation of the parameter distribution. To get the rejection rate for the three tests, we choose the "corrected" chi-square critical values 1.64 and 2.71 for 10% and 5% significance levels with one degree of freedom. Table 4 provides the percentage of rejecting the null hypothesis $\sigma_\beta = 0$, using critical value 1.64 and 2.71. When the null hypothesis is true, most empirical percentage rates of LR test rejecting the true null hypothesis are less than the nominal rejection rates 10% and 5%, and become further away from the nominal rejection rates with increases in the parameter mean $\overline{\beta}$. All empirical rejection rates of Wald and LM tests given a true null hypothesis are greater than the related expected percentage rates. The size of the LR test is too large, and the size of LM and Wald tests is too small.

Figure 3 contains graphs based on the results of Table 4. From Figure 3, we can see the changes in the rejection rates of these three test statistics with increases in the mean and standard deviation of the parameter distribution respectively. We find the rejection frequency of the LR and Wald statistics declines in the mean of the parameter distribution.

Due to the different sizes of the three tests, power comparisons are invalid. We use the Monte Carlo percentile values for each combination of parameter mean and standard deviation as the critical value to correct the size of the three tests. Table 5 provides the size corrected rejection rates for the three tests. The size corrected rejection rates for the LR and Wald tests increase in the standard deviation of the coefficient distribution as expected. Based on the results, there is not too much difference between these two size corrected tests. But the power of these two tests still declines with increases in the parameter mean. In our experiments, at the 10% and 5% significance levels, the LM test shows the weakest power for the presence of the random coefficient among the three tests. Graphs in Figure 4 are based on the results of Table 5. After adjusting the size of the test, the power of LR test declines slowly in the parameter mean. The results of the power of these three tests are consistent with the results of pretest estimators based on these three tests.

We expand the model to two parameters. The mean and standard deviation of the added random parameter $\bar{\beta}_2$ are set as 1.5 and 0.8 respectively. We still set four different values for the first parameter mean, $\bar{\beta}_1$={0.5, 1.5, 2.5, 3.0}. For each value of the mean $\bar{\beta}_1$, we set six different values for the standard deviation, $\bar{\sigma}_{\beta_1}$ ={0, 0.15, 0.3, 0.8, 1.2, 1.8}. In the two parameters model, the LR, Wald and LM tests are constructed based on the joint null hypothesis $H_0 : \sigma_{\beta_1} = 0$ and $\sigma_{\beta_2} = 0$ against the alternative hypothesis $H_1 : \sigma_{\beta_1} > 0$ or $\sigma_{\beta_2} > 0$ or $\sigma_{\beta_1} > 0$ and $\sigma_{\beta_2} > 0$. Figure 5 shows the ratios of the pretest estimator RMSE of $\bar{\beta}_1$ and $\bar{\beta}_2$ to the random parameters logit model estimator RMSE of $\bar{\beta}_1$ and $\bar{\beta}_2$ based on the joint LR, Wald and LM tests at a 25% significance level. Here we use the standard chi-square as the critical value for 25% significance level, 2.773. The joint LR and Wald tests show properties of consistent tests. The maximum risk ratio based on the joint LR and Wald tests still increases in the parameter mean $\bar{\beta}_1$. In two parameters model, the pretest estimators based on the joint LR and Wald statistics have larger RMSE than that of the random parameters logit model. The properties of the joint LM-based pretest estimator are also poor in two parameters model. Table 6 reports the 90th and 95th empirical percentiles of the joint LR, Wald and LM tests. They are different with different combinations of means and standard deviations of one random parameter distribution. All the empirical 90th and 95th percentile values of the joint LR and Wald tests are much greater than the related standard chi-square statistics 4.605 and 5.991. The Monte Carlo empirical percentiles of the joint LM test are also not close to the standard chi-square statistics. Since the weighted chi-square statistics are even smaller than the standard chi-square statistics, we choose the standard ones to find the rejection rate of the three tests. Table 7 shows the rejection rates of the three joint tests based on the standard chi-square statistics for 10% and 5% significance level. The results are consistent with the Table 6. When the null hypothesis is true, the joint LR and Wald tests reject the true null hypothesis more frequently than the nominal rejection rates 10% and 5%. They become closer to the nominal rejection rates with increases in the parameter mean $\bar{\beta}_1$. When $\bar{\beta}_1$=0.5 and 3.0, the joint LM test rejecting the true null hypothesis are less than the nominal rejection rates. However, with $\bar{\beta}_1$=1.5 and 2.5, it rejects more frequently than the nominal rejection rates 10% and 5%. Figure 6 shows the graphs based on the results of Table 7. They almost have the same trends as in the one parameter case. The rejection frequency of the joint LR and Wald statistics decreases in the mean of the parameter distribution $\bar{\beta}_1$.

To compare the power of the three joint tests in the two parameters case, we also correct the size of the three joint tests using the Monte Carlo empirical critical values for 10% and 5% significance level. Table 8 provides the size corrected rejection rates for the three joint tests. Figure 7 presents the graphs based on the Table 8. As in the one parameter case, the joint LM test shows the weakest power for the presence of the random coefficient. The power of the joint LR and Wald tests decreases when the parameter mean $\bar{\beta}_1$ increases from 0.5 to 1.5. However, the power of these two joint tests increases when the parameter mean $\bar{\beta}_1$ increases further to 3.0.

An interesting question is why the power of LR and Wald tests for the presence of the random coefficient declines in the parameter mean and how to refine the LM test in the setting of the random parameters logit model. The Lagrange Multiplier test is developed by Aitchison and

Silvey (1958) and Silvey (1959), which associated with the constrained optimization problem. In our setting, the Lagrangian function is:

$$\ln L(\theta) + \lambda'(c(\theta) - q)$$

where $\ln L(\theta)$ is the log-likelihood function, which subject to the constraints $(c(\theta) - q) = 0$. The related first-order conditions are:

$$\begin{cases} \dfrac{\partial \ln L(\theta)}{\partial \theta} + \dfrac{\partial c(\theta)}{\partial \theta}\lambda = 0 \\ c(\theta) - q = 0 \end{cases}$$

Under the standard assumptions of the LM test, we know

$$\sqrt{n}(\hat{\theta} - \theta) \sim N(0, I(\theta)^{-1})$$

and

$$n^{-1/2}\hat{\lambda} \sim N\left(0, \left(\frac{\partial c(\theta)}{\partial \theta'} I(\theta)^{-1} \frac{\partial c'(\theta)}{\partial \theta}\right)\right).$$

Based on the Lagrangian function, we have

$$\hat{\lambda}' \frac{\partial c(\hat{\theta})}{\partial \hat{\theta}'} I(\hat{\theta})^{-1} \frac{\partial c'(\hat{\theta})}{\partial \hat{\theta}} \hat{\lambda} = \frac{\partial \ln L(\hat{\theta})}{\partial \hat{\theta}'} I(\hat{\theta})^{-1} \frac{\partial \ln L(\hat{\theta})}{\partial \theta}$$

From the above results, the LM statistic has the asymptotic χ^2-distribution. The asymptotic distribution of the LM statistic is derived from the distribution of Lagrangian multiplier, which essentially based on the asymptotic normality of the log-likelihood estimators. In the Lagrangian function, the log-likelihood function is subject to the equality constraints. The weak power of the LM test for the presence of the random coefficient is caused by the failure of taking into account the properties of the one-tailed alternative hypothesis. Gourieroux, Holly and Monfort (1982) extended the LM test to the Kuhn-Tucker multiplier test and showed that it is asymptotically equivalent to the LR and Wald tests. However, computing the Kuhn-Tucker multiplier test is complicated. How to refine the LM test in the random parameters logit model is our future research.

5 Conclusions

In our quasi-Monte Carlo experiments, the LM test for the significance of the standard deviation of the coefficient distribution is not reliable. It has low power and the resulting pretest estimator has poor risk properties. At the same time, the LR and Wald tests are more reliable and provide more predictable pretest estimation performance. The empirical critical values of the LR, Wald and LM tests change with changes in the mean and standard deviation of the parameter distribution, which implies that the distribution of test statistics may change in the mean and standard deviation of the parameter distribution. The distribution of the simulated maximum likelihood estimator of the random parameters logit model is an interesting topic to further study. In the maximum likelihood based tests how to construct the efficient one-tailed LM test is also very interesting.

i

Figure1(a) 200 points pseduo–random numbers in two–dimensions

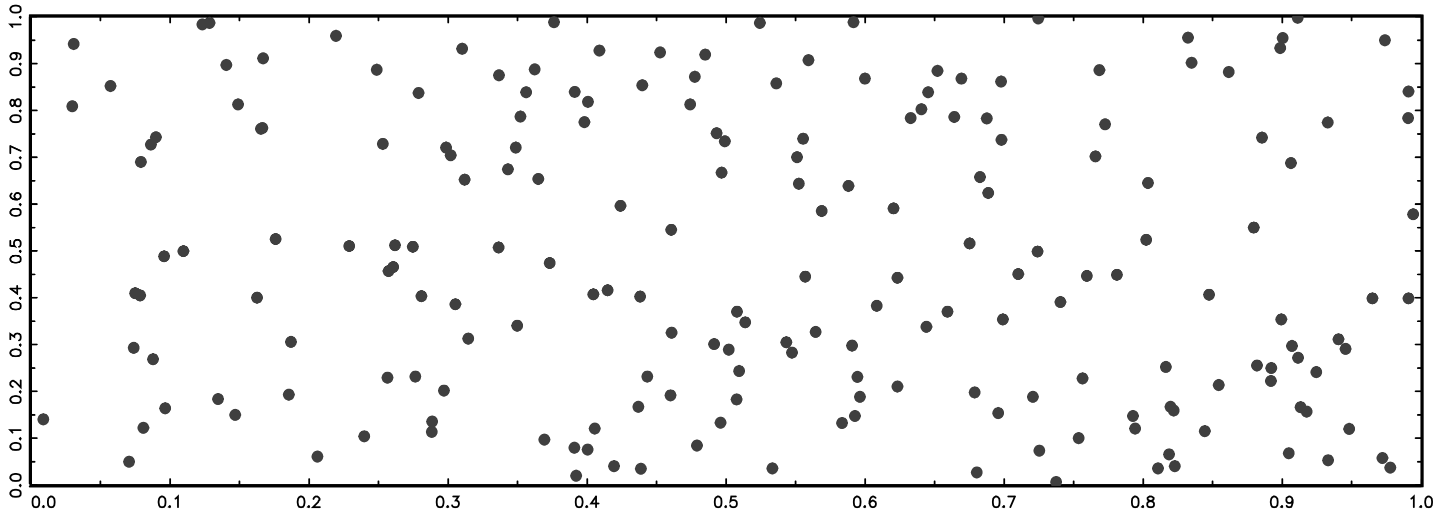


Figure1 (b) 200 points of two–dimension Halton sequence generated with prime 2 and 3

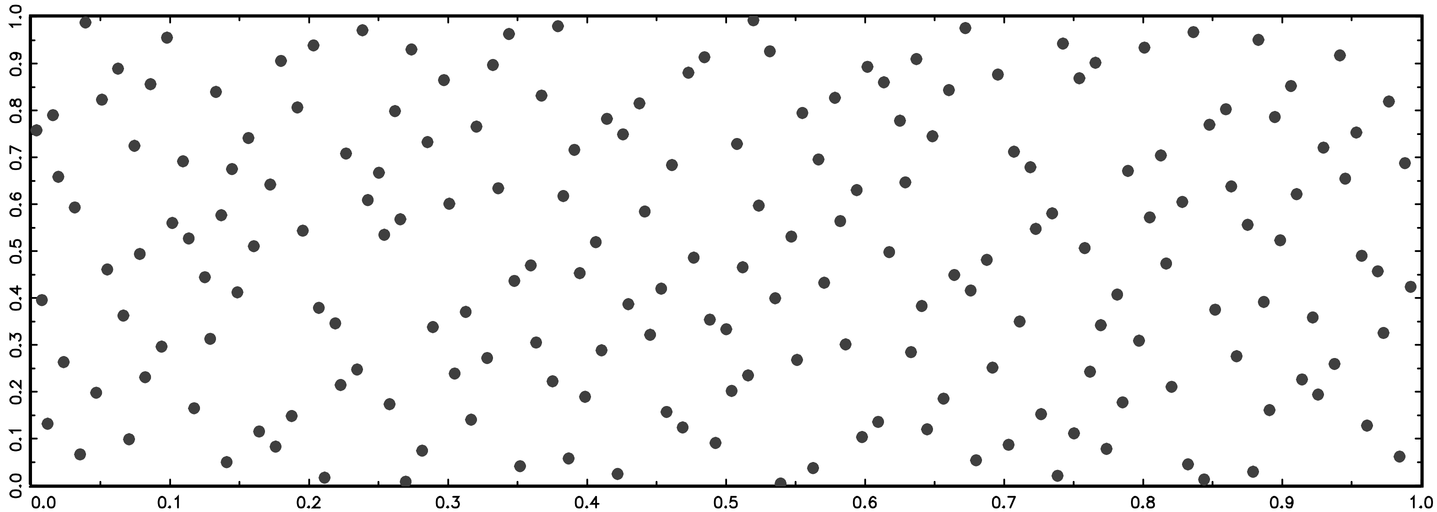


Table 1: The Mixed Logit Model Estimated with Classical-Monte Carlo and Quasi-Monte Carlo Estimation

The mixed logit model with one random parameter				
The parameters	**Classical- Monte Carlo estimation**	**Quasi-Monte Carlo estimation**		
$\bar{\beta}$ **=1.5***	**Number of random draws**	**Number of Halton Draws**		
$\bar{\sigma}_{\beta}$ **=0.8***	**1000**	**25**	**100**	**250**
Monte Carlo average value of $\hat{\beta}_i$	1.486	1.468	1.477	1.477
Monte Carlo average value of $\hat{\sigma}_{\beta_i}$	0.625	0.594	0.606	0.602
Monte Carlo s.d. of $\hat{\beta}_i$	0.234	0.226	0.233	0.232
Monte Carlo s.d. of $\hat{\sigma}_{\beta_i}$	0.363	0.337	0.372	0.375
average nominal s.e. of $\hat{\beta}_i$	0.240	0.236	0.237	0.237
average nominal s.e. of $\hat{\sigma}_{\beta_i}$	0.434	0.417	0.447	0.465
ratio of average nominal s.e. to MC s.d. of $\hat{\beta}_i$	1.026	1.044	1.017	1.021
ratio of average nominal s.e. to MC s.d. of $\hat{\sigma}_{\beta_i}$	1.196	1.238	1.202	1.241
RMSE of $\hat{\beta}_i$	0.234	0.228	0.234	0.233
RMSE of $\hat{\sigma}_{\beta_i}$	0.402	0.395	0.419	0.424

*the mean and the standard deviation of the distribution of random coefficient β_n

Table 2: The Mixed Logit Model Estimated with Classical-Monte Carlo and Quasi-Monte Carlo Estimation

The mixed logit model with two random parameters				
The parameters	**Classical-Monte Carlo estimation**	**Quasi-Monte Carlo estimation**		
$\overline{\beta}_1$ **=2.5**	**Number of Random Draws**	**Number of Halton Draws**		
$\overline{\sigma}_{\beta_1}$ **=0.3**	**1000**	**25**	**100**	**250**
Monte Carlo(MC) average value of $\widehat{\beta}_{1i}$	2.733	2.754	2.732	2.728
Monte Carlo average value of $\widehat{\sigma}_{\beta_{1i}}$	0.332	0.401	0.318	0.302
Monte Carlo average value of $\widehat{\beta}_{2i}$	1.674	1.680	1.676	1.672
Monte Carlo average value of $\widehat{\sigma}_{\beta_{2i}}$	0.601	0.615	0.605	0.592
Monte Carlo s.d. of $\widehat{\beta}_{1i}$	0.491	0.497	0.477	0.490
Monte Carlo s.d. of $\widehat{\sigma}_{\beta_{1i}}$	0.428	0.438	0.435	0.448
Monte Carlo s.d. of $\widehat{\beta}_{2i}$	0.327	0.325	0.316	0.323
Monte Carlo s.d. of $\widehat{\sigma}_{\beta_{2i}}$	0.439	0.423	0.430	0.447
average nominal s.e. of $\widehat{\beta}_{1i}$	0.445	0.450	0.445	0.443
average nominal s.e. of $\widehat{\sigma}_{\beta_{1i}}$	0.737	0.678	0.772	0.833
average nominal s.e. of $\widehat{\beta}_{2i}$	0.298	0.300	0.297	0.297
average nominal s.e. of $\widehat{\sigma}_{\beta_{2i}}$	0.512	0.494	0.499	0.537
ratio of average nominal s.e. to MC s.d. of $\widehat{\beta}_{1i}$	0.907	0.906	0.933	0.904
ratio of average nominal s.e. to MC s.d. o $\widehat{\sigma}_{\beta_{1i}}$	1.721	1.548	1.776	1.859
ratio of average nominal s.e. to MC s.d. of $\widehat{\beta}_{2i}$	0.912	0.923	0.940	0.919
ratio of average nominal s.e. to MC s.d. of $\widehat{\sigma}_{\beta_{2i}}$	1.167	1.168	1.160	1.202
RMSE of $\widehat{\beta}_{1i}$	0.543	0.558	0.531	0.540
RMSE of $\widehat{\sigma}_{\beta_{1i}}$	0.429	0.449	0.435	0.448
RMSE of $\widehat{\beta}_{2i}$	0.370	0.371	0.361	0.366
RMSE of $\widehat{\sigma}_{\beta_{2i}}$	0.481	0.461	0.472	0.492

Figure 2: The ratio of the Pre-test Estimator RMSE of $\overline{\beta}$ to the Mixed Logit Estimator RMSE of $\overline{\beta}$

One Random Parameter Model

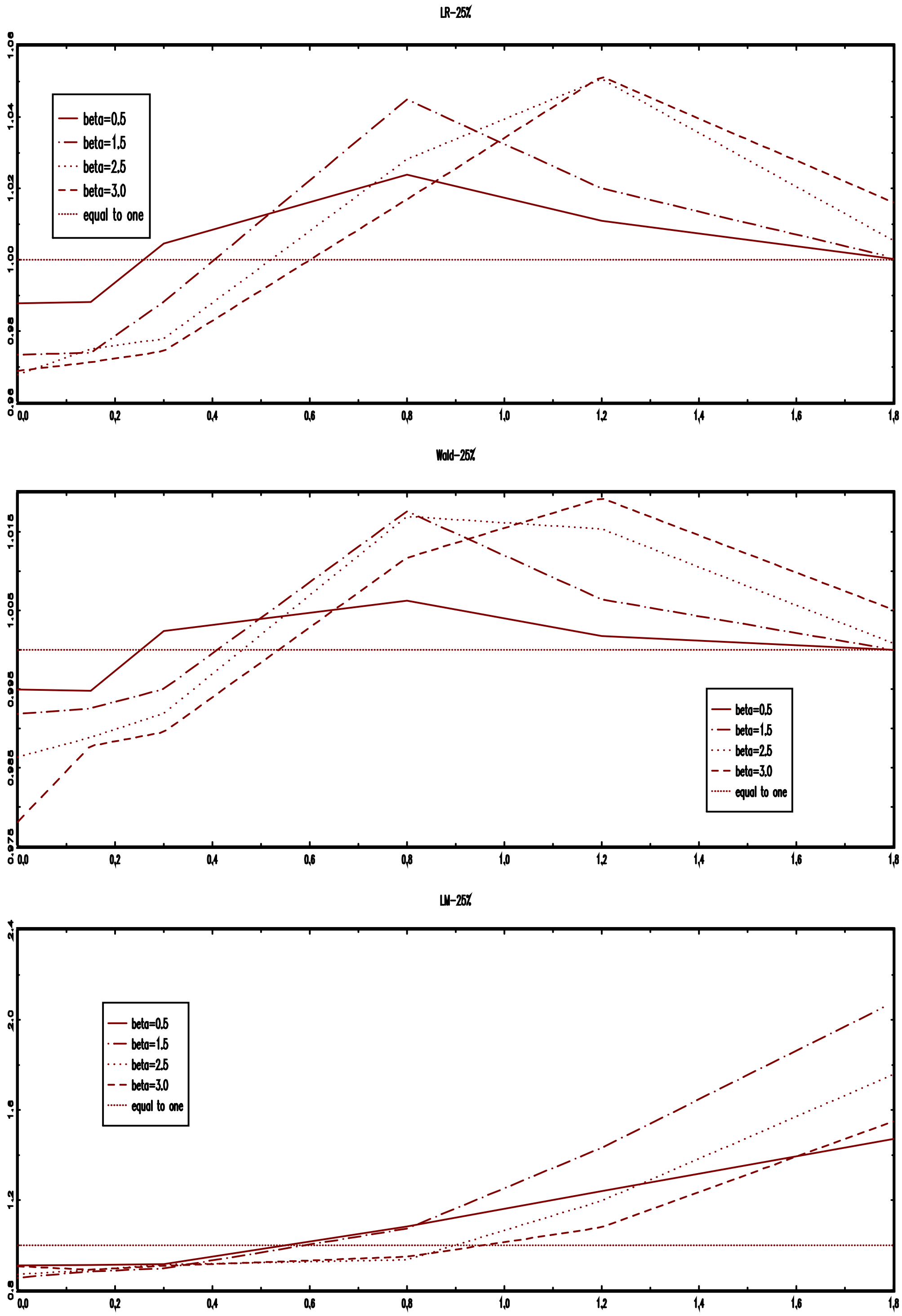


Table 3: 90th and 95th Empirical Percentiles of Likelihood Ratio, Wald and Lagrange Multiplier Tests
One Random Parameter Model

$\bar{\beta}$	$\bar{\sigma}_{\beta}$	LR-90th	LR-95th	Wald-90th	Wald-95th	LM-90th	LM-95th
0.5	0.00	1.927	3.267	4.006	5.917	2.628	3.576
0.5	0.15	1.749	2.755	3.850	5.425	2.749	3.862
0.5	0.30	2.239	3.420	4.722	6.210	2.594	3.544
0.5	0.80	6.044	7.779	9.605	11.014	2.155	3.043
0.5	1.20	12.940	15.684	14.472	15.574	1.712	2.344
0.5	1.80	26.703	31.347	19.225	19.950	1.494	2.041
1.5	0.00	1.518	2.668	3.671	5.672	2.762	3.972
1.5	0.15	1.541	2.414	3.661	5.443	3.020	4.158
1.5	0.30	1.837	3.364	4.361	6.578	3.048	4.308
1.5	0.80	5.753	7.451	8.603	10.424	2.496	3.489
1.5	1.20	11.604	13.953	12.930	13.974	1.825	2.376
1.5	1.80	24.684	28.374	17.680	18.455	1.346	1.947
2.5	0.00	0.980	1.727	2.581	4.017	2.978	4.147
2.5	0.15	1.020	1.858	2.598	4.256	2.976	4.317
2.5	0.30	1.217	2.235	2.751	4.616	3.035	4.429
2.5	0.80	2.766	4.667	6.387	8.407	3.119	4.315
2.5	1.20	6.321	8.643	9.700	11.598	2.714	3.832
2.5	1.80	18.018	20.828	14.895	15.822	2.189	3.275
3.0	0.00	1.042	1.720	2.691	4.264	3.455	4.594
3.0	0.15	1.040	1.941	2.548	4.878	3.285	4.441
3.0	0.30	1.260	2.114	3.068	5.124	3.164	4.324
3.0	0.80	2.356	3.167	4.915	7.106	3.073	4.198
3.0	1.20	4.610	6.570	8.086	10.296	2.917	4.224
3.0	1.80	13.261	15.622	12.960	14.052	2.579	3.478

*Testing $H_0 : \sigma_\beta = 0$; One tail critical values are 1.64 (10%) and 2.71 (5%), compared to the usual values 2.71 and 3.84 respectively.

Table 4: Rejection Rate of Likelihood Ratio, Wald and Lagrange Multiplier Tests
One Random Parameter Model

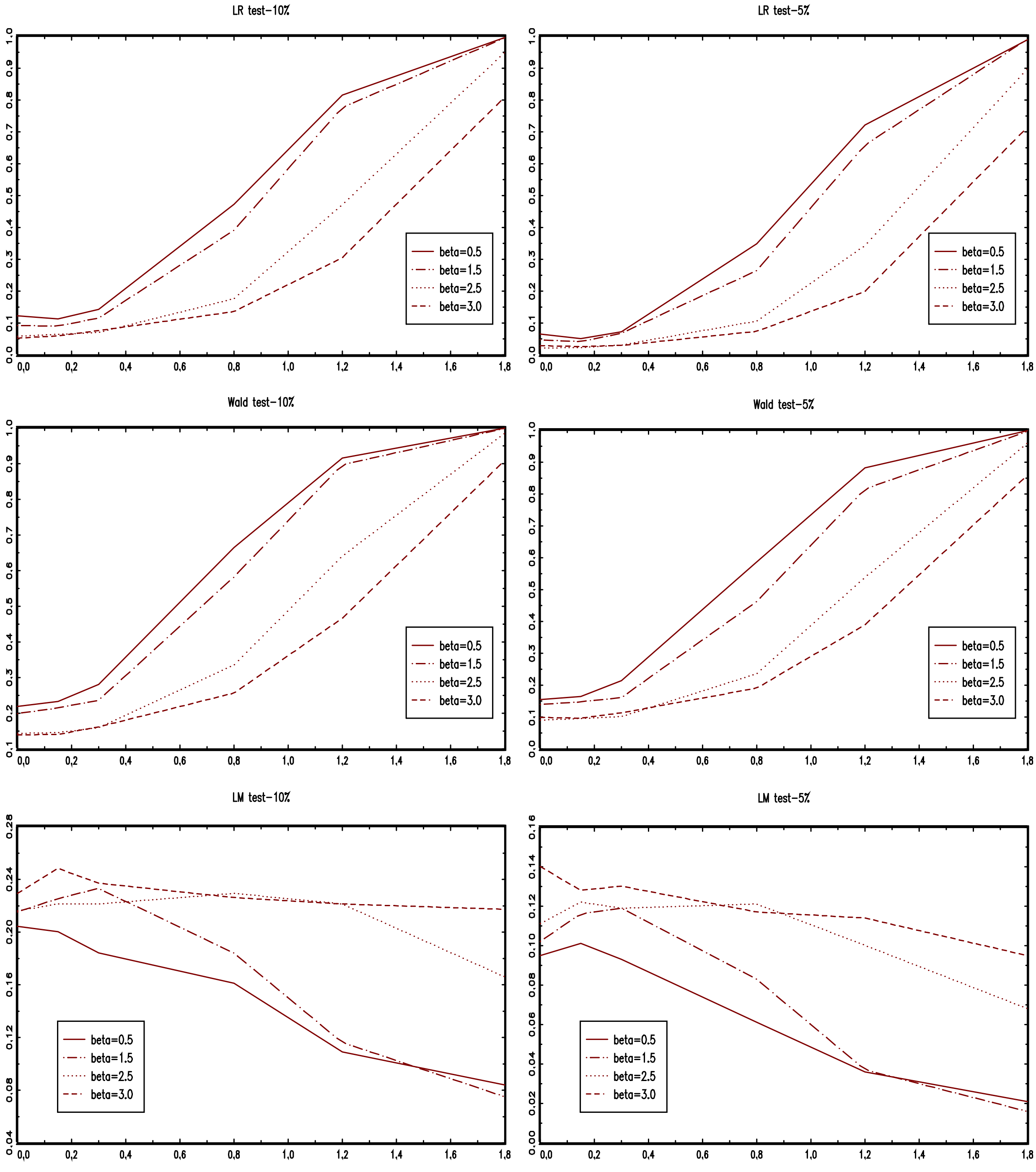
$\hat{\beta}$	$\hat{\sigma}_\beta$	$se(\hat{\beta})^*$	$se(\hat{\sigma}_\beta)^*$	LR-10%**	LR-5%**	Wald-10%**	Wald-5%**	LM-10%**	LM-5%**
0.5	0.00	0.123	0.454	0.122	0.065	0.219	0.155	0.204	0.095
0.5	0.15	0.125	0.461	0.113	0.051	0.233	0.164	0.200	0.101
0.5	0.30	0.125	0.460	0.143	0.072	0.281	0.214	0.184	0.093
0.5	0.80	0.135	0.416	0.472	0.348	0.665	0.587	0.161	0.061
0.5	1.20	0.153	0.391	0.816	0.722	0.916	0.882	0.109	0.036
0.5	1.80	0.195	0.438	0.996	0.989	1.000	0.999	0.084	0.021
1.5	0.00	0.242	0.593	0.092	0.048	0.199	0.139	0.215	0.102
1.5	0.15	0.243	0.586	0.090	0.042	0.215	0.148	0.225	0.116
1.5	0.30	0.243	0.567	0.115	0.068	0.236	0.160	0.233	0.119
1.5	0.80	0.247	0.439	0.390	0.264	0.582	0.461	0.184	0.083
1.5	1.20	0.261	0.391	0.777	0.659	0.897	0.816	0.116	0.037
1.5	1.80	0.291	0.443	0.995	0.990	0.999	0.996	0.075	0.016
2.5	0.00	0.416	0.910	0.058	0.022	0.143	0.090	0.216	0.111
2.5	0.15	0.416	0.889	0.064	0.023	0.146	0.095	0.221	0.122
2.5	0.30	0.410	0.853	0.070	0.031	0.159	0.101	0.221	0.119
2.5	0.80	0.392	0.714	0.176	0.106	0.335	0.235	0.229	0.121
2.5	1.20	0.392	0.537	0.471	0.342	0.641	0.539	0.221	0.100
2.5	1.80	0.412	0.453	0.949	0.898	0.985	0.959	0.166	0.068
3.0	0.00	0.519	1.131	0.052	0.028	0.139	0.099	0.229	0.140
3.0	0.15	0.508	1.062	0.060	0.026	0.140	0.096	0.248	0.128
3.0	0.30	0.514	0.975	0.076	0.030	0.162	0.113	0.237	0.130
3.0	0.80	0.489	0.910	0.135	0.074	0.256	0.190	0.226	0.117
3.0	1.20	0.478	0.701	0.304	0.199	0.465	0.389	0.221	0.114
3.0	1.80	0.479	0.505	0.808	0.714	0.909	0.858	0.217	0.095

*The average nominal standard error of estimated mean and standard deviation of the random coefficient distribution

**Testing $H_0 : \sigma_\beta = 0$; One tail critical values are 1.64 (10%) and 2.71 (5%)

Figure 3: The Rejection Rate of LR, Wald and LM Tests

One Random Parameter Model



*Testing $H_0 : \sigma_\beta = 0$; One tail critical values are 1.64 (10%) and 2.71 (5%)

Table 5: Size Corrected Rejection rates of LR, Wald and LM Tests:
One Random Parameter Model

$\overline{\beta}$	$\overline{\sigma}_\beta$	LR -10%	LR -5%	Wald -10%	Wald -5%	LM -10%	LM -5%
0.5	0.00	0.100	0.050	0.100	0.050	0.100	0.050
0.5	0.15	0.094	0.035	0.093	0.036	0.108	0.060
0.5	0.30	0.121	0.055	0.123	0.056	0.099	0.049
0.5	0.80	0.431	0.287	0.498	0.336	0.066	0.028
0.5	1.20	0.792	0.676	0.834	0.746	0.040	0.016
0.5	1.80	0.995	0.980	0.999	0.991	0.022	0.005
1.5	0.00	0.100	0.050	0.100	0.050	0.100	0.050
1.5	0.15	0.100	0.043	0.098	0.047	0.112	0.056
1.5	0.30	0.124	0.068	0.124	0.067	0.115	0.058
1.5	0.80	0.407	0.269	0.383	0.240	0.078	0.031
1.5	1.20	0.788	0.663	0.758	0.616	0.035	0.014
1.5	1.80	0.995	0.990	0.995	0.988	0.011	0.005
2.5	0.00	0.100	0.050	0.100	0.050	0.100	0.050
2.5	0.15	0.101	0.06	0.100	0.056	0.099	0.052
2.5	0.30	0.119	0.069	0.110	0.065	0.103	0.057
2.5	0.80	0.256	0.166	0.242	0.173	0.104	0.051
2.5	1.20	0.565	0.460	0.544	0.444	0.082	0.037
2.5	1.80	0.971	0.942	0.961	0.931	0.062	0.022
3.0	0.00	0.100	0.050	0.100	0.050	0.100	0.050
3.0	0.15	0.099	0.058	0.096	0.059	0.089	0.046
3.0	0.30	0.120	0.071	0.114	0.080	0.083	0.042
3.0	0.80	0.197	0.133	0.192	0.121	0.079	0.042
3.0	1.20	0.403	0.294	0.392	0.282	0.072	0.041
3.0	1.80	0.873	0.803	0.859	0.764	0.051	0.031

Testing $H_0 : \sigma_\beta = 0$; using Monte Carlo percentile values as the critical values to adjust the size the LR, Wald and LM tests

Figure 4: The size Corrected Rejection Rates
One Random Parameter Model

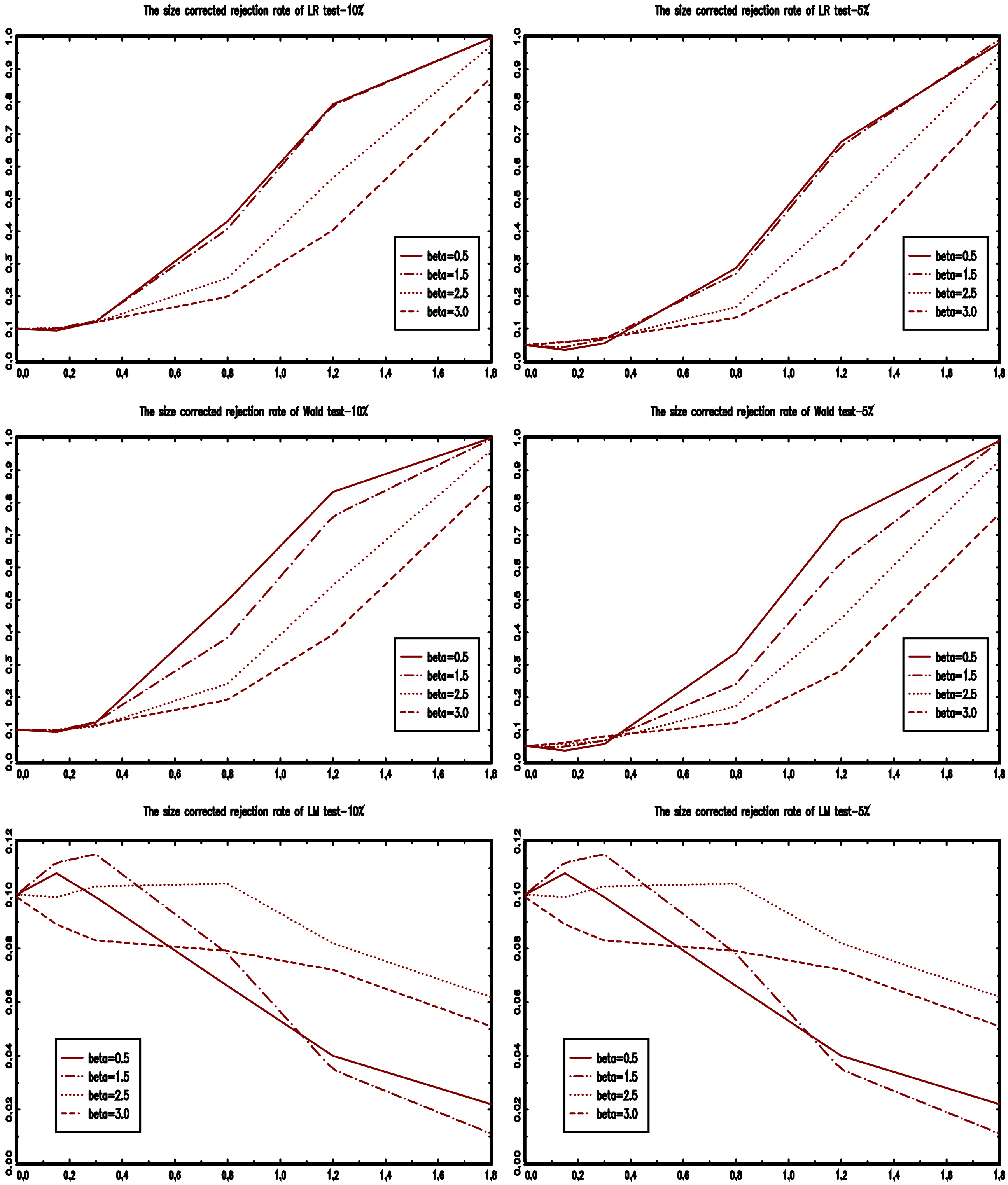
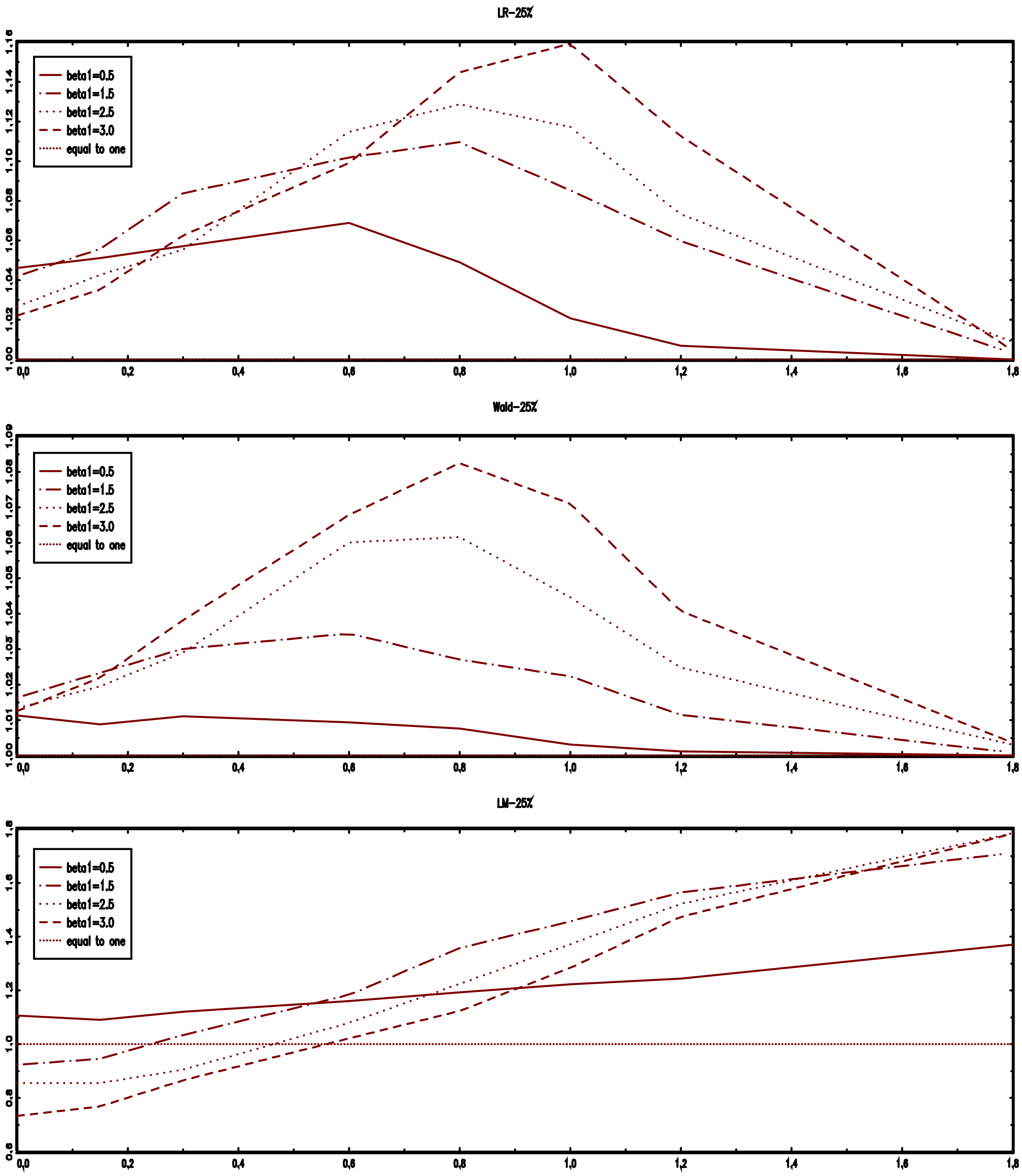


Figure 5: The Ratio of the Pre-test Estimator RMSE of $\bar{\beta}$ to the Mixed Logit Estimator RMSE of $\bar{\beta}$
Two Random Parameters Model



RMSE of $\hat{\beta} = \sqrt{(\sum(\hat{\beta}_1 - \beta)^2 + \sum(\hat{\beta}_2 - \beta)^2)/NSAM}$ where NSAM=999

Table 6: 90th and 95th Empirical Percentiles of Likelihood Ratio, Wald and Lagrange Multiplier Tests
Two Random Parameter Model

$\bar{\beta}_1$	$\bar{\sigma}_{\beta_1}$	$\bar{\beta}_2$	$\bar{\sigma}_{\beta_2}$	LR-90th	LR-95th	Wald-90th	Wald-95th	LM-90th	LM-95th
0.5	0.00	1.5	0.8	14.634	17.899	13.493	14.647	4.033	5.196
0.5	0.15	1.5	0.8	13.583	17.001	13.148	14.118	4.164	5.242
0.5	0.30	1.5	0.8	13.504	16.043	13.060	14.156	4.208	5.420
0.5	0.80	1.5	0.8	14.961	17.867	12.496	13.157	4.052	5.062
0.5	1.20	1.5	0.8	19.940	23.966	13.536	14.305	4.168	5.215
0.5	1.80	1.5	0.8	29.429	32.083	15.208	16.081	3.989	5.218
1.5	0.00	1.5	0.8	14.109	16.844	12.638	14.074	6.329	8.105
1.5	0.15	1.5	0.8	12.645	15.466	11.961	13.448	5.991	7.689
1.5	0.30	1.5	0.8	11.955	14.415	11.498	12.641	5.881	7.444
1.5	0.80	1.5	0.8	12.341	14.569	11.022	12.017	4.480	5.601
1.5	1.20	1.5	0.8	15.529	17.472	11.760	12.860	4.478	5.699
1.5	1.80	1.5	0.8	22.300	25.700	13.321	14.155	4.682	5.639
2.5	0.00	1.5	0.8	11.315	13.966	10.161	11.439	5.094	6.275
2.5	0.15	1.5	0.8	10.449	13.120	9.820	11.137	4.920	6.368
2.5	0.30	1.5	0.8	9.998	12.437	9.707	10.986	5.051	6.230
2.5	0.80	1.5	0.8	10.388	12.690	9.554	10.657	4.714	6.092
2.5	1.20	1.5	0.8	14.168	17.001	10.527	11.433	4.552	5.829
2.5	1.80	1.5	0.8	21.625	24.694	12.815	13.704	4.994	6.248
3.0	0.00	1.5	0.8	9.713	12.354	8.905	10.552	4.528	5.729
3.0	0.15	1.5	0.8	9.185	11.450	8.493	10.215	4.434	5.923
3.0	0.30	1.5	0.8	8.384	10.388	8.262	9.7540	4.245	5.418
3.0	0.80	1.5	0.8	8.219	10.083	8.499	10.010	4.486	5.716
3.0	1.20	1.5	0.8	13.704	15.917	10.058	10.967	4.972	6.353
3.0	1.80	1.5	0.8	20.939	23.476	12.454	13.282	5.273	6.544

Table 7: Rejection Rate of Likelihood Ratio, Wald and Lagrange Multiplier Tests
Two Random Parameters Model

$\bar{\beta}_1$	$\bar{\sigma}_{\beta_1}$	$\bar{\beta}_2$	$\bar{\sigma}_{\beta_2}$	LR-10%	LR-5%	Wald-10%	Wald-5%	LM-10%	LM-5%
0.5	0.00	1.5	0.8	0.719	0.594	0.890	0.824	0.069	0.031
0.5	0.15	1.5	0.8	0.681	0.563	0.880	0.781	0.077	0.033
0.5	0.30	1.5	0.8	0.668	0.534	0.876	0.779	0.083	0.031
0.5	0.80	1.5	0.8	0.749	0.631	0.920	0.823	0.070	0.032
0.5	1.20	1.5	0.8	0.949	0.892	0.985	0.960	0.077	0.033
0.5	1.80	1.5	0.8	0.999	0.992	1.000	0.997	0.077	0.030
1.5	0.00	1.5	0.8	0.600	0.476	0.762	0.632	0.205	0.114
1.5	0.15	1.5	0.8	0.563	0.430	0.728	0.603	0.191	0.099
1.5	0.30	1.5	0.8	0.520	0.386	0.705	0.561	0.176	0.096
1.5	0.80	1.5	0.8	0.620	0.482	0.783	0.640	0.092	0.039
1.5	1.20	1.5	0.8	0.796	0.672	0.914	0.816	0.093	0.035
1.5	1.80	1.5	0.8	0.969	0.936	0.995	0.980	0.105	0.035
2.5	0.00	1.5	0.8	0.492	0.381	0.631	0.462	0.127	0.059
2.5	0.15	1.5	0.8	0.451	0.327	0.589	0.427	0.116	0.059
2.5	0.30	1.5	0.8	0.388	0.284	0.540	0.352	0.126	0.061
2.5	0.80	1.5	0.8	0.428	0.317	0.576	0.429	0.105	0.053
2.5	1.20	1.5	0.8	0.755	0.628	0.837	0.718	0.097	0.044
2.5	1.80	1.5	0.8	0.963	0.928	0.982	0.954	0.131	0.058
3.0	0.00	1.5	0.8	0.411	0.291	0.502	0.332	0.094	0.039
3.0	0.15	1.5	0.8	0.374	0.253	0.481	0.293	0.092	0.048
3.0	0.30	1.5	0.8	0.333	0.223	0.436	0.272	0.078	0.032
3.0	0.80	1.5	0.8	0.284	0.188	0.400	0.244	0.088	0.042
3.0	1.20	1.5	0.8	0.623	0.528	0.747	0.608	0.119	0.059
3.0	1.80	1.5	0.8	0.965	0.913	0.982	0.939	0.137	0.067

Figure 6: The Rejection Rate of LR, Wald and LM Tests

Two Random Parameters Model

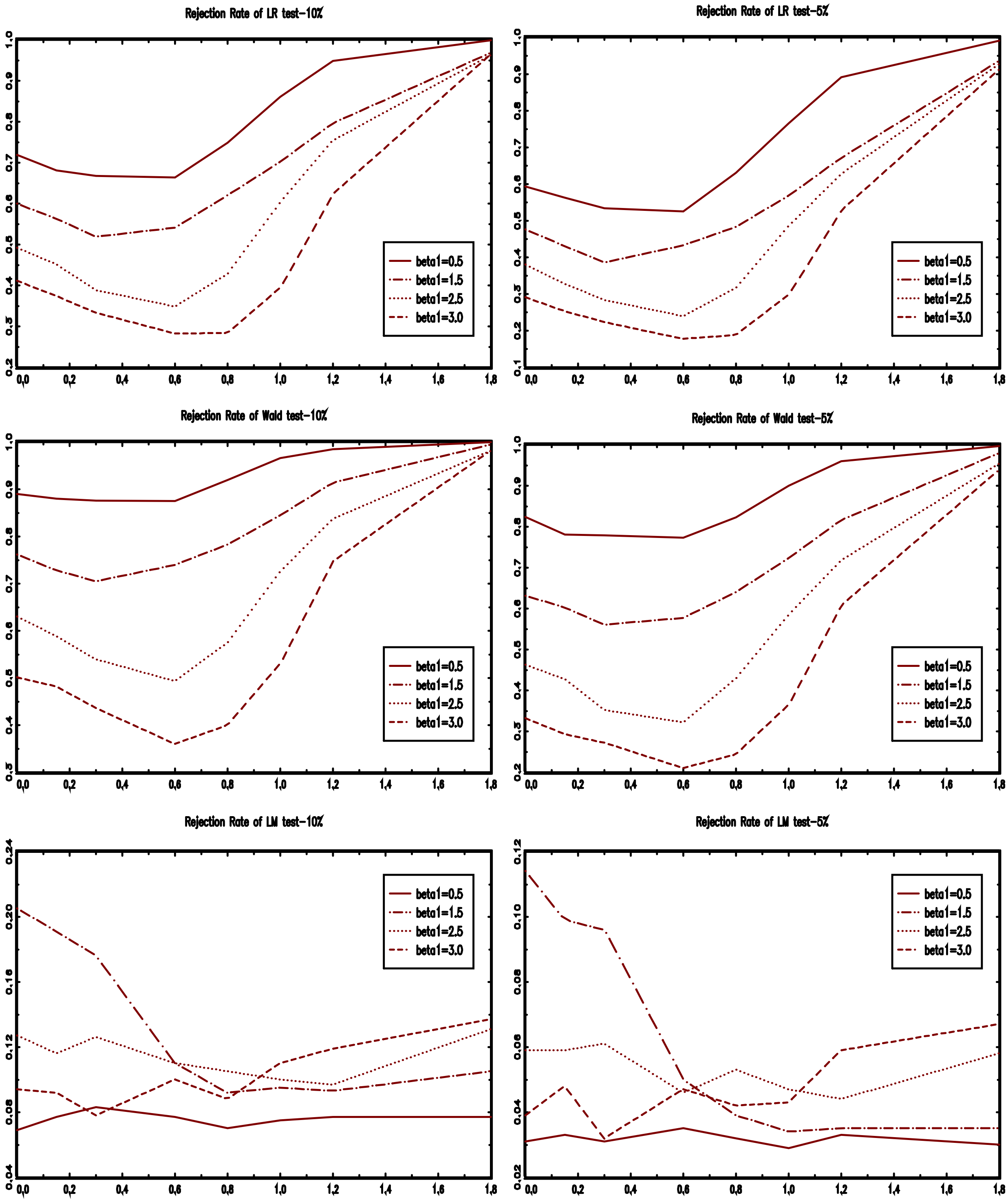


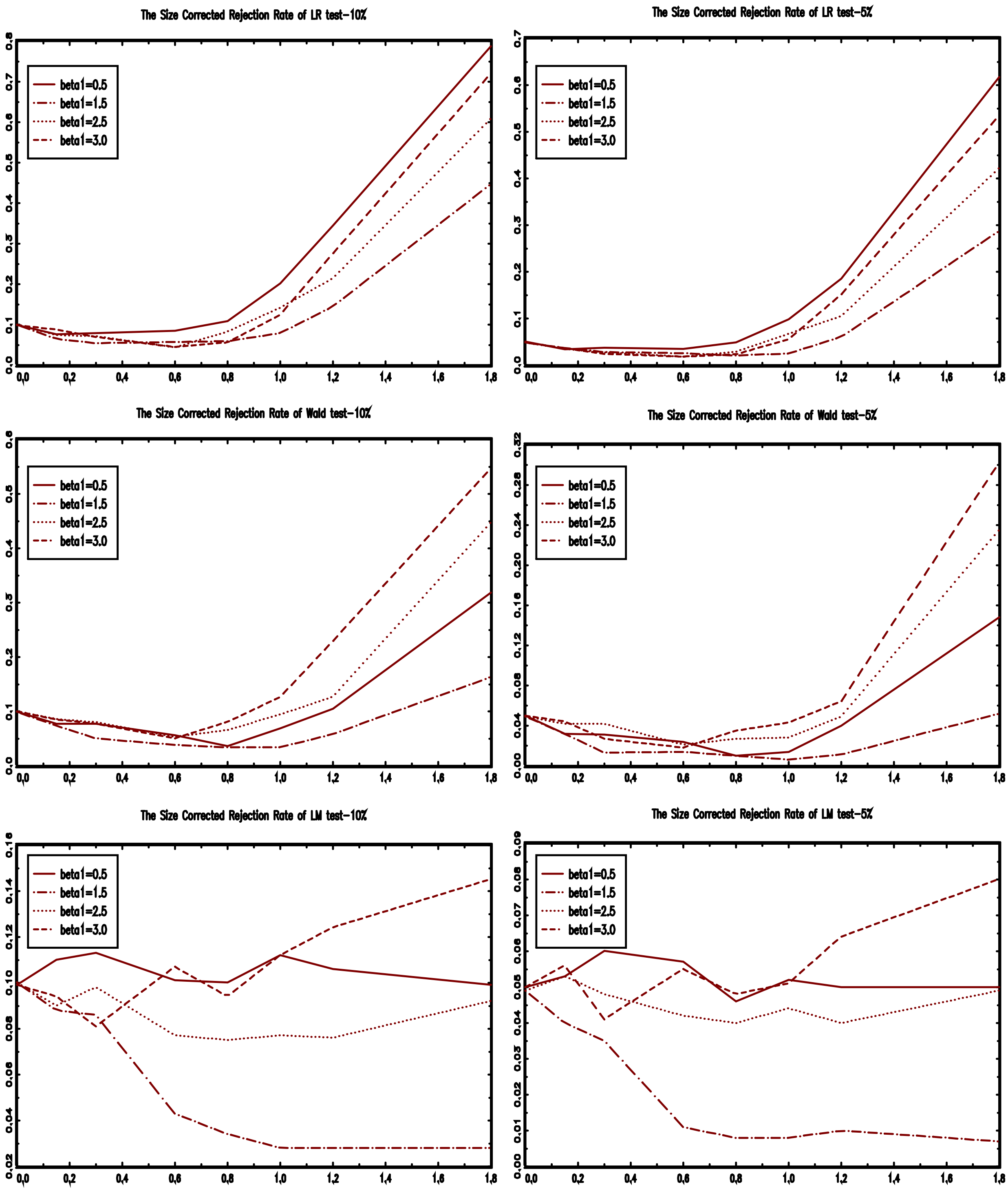
Table 8: The Size Corrected Rejection Rates of LR, Wald and LM Tests

Two Random Parameters Model

$\bar{\beta}_1$	$\bar{\sigma}_{\beta_1}$	$\bar{\beta}_2$	$\bar{\sigma}_{\beta_2}$	LR-10%	LR-5%	Wald-10%	Wald-5%	LM-10%	LM-5%
0.5	0.00	1.5	0.8	0.100	0.050	0.100	0.050	0.100	0.050
0.5	0.15	1.5	0.8	0.076	0.034	0.077	0.032	0.110	0.053
0.5	0.30	1.5	0.8	0.079	0.037	0.077	0.031	0.113	0.060
0.5	0.80	1.5	0.8	0.108	0.049	0.036	0.010	0.100	0.046
0.5	1.20	1.5	0.8	0.344	0.185	0.105	0.040	0.106	0.050
0.5	1.80	1.5	0.8	0.788	0.618	0.318	0.148	0.099	0.050
1.5	0.00	1.5	0.8	0.100	0.050	0.100	0.050	0.100	0.050
1.5	0.15	1.5	0.8	0.065	0.036	0.074	0.032	0.088	0.040
1.5	0.30	1.5	0.8	0.054	0.028	0.050	0.013	0.086	0.035
1.5	0.80	1.5	0.8	0.059	0.021	0.034	0.010	0.034	0.008
1.5	1.20	1.5	0.8	0.145	0.060	0.058	0.011	0.028	0.010
1.5	1.80	1.5	0.8	0.446	0.287	0.163	0.052	0.028	0.007
2.5	0.00	1.5	0.8	0.100	0.050	0.100	0.050	0.100	0.050
2.5	0.15	1.5	0.8	0.074	0.035	0.086	0.042	0.090	0.053
2.5	0.30	1.5	0.8	0.071	0.027	0.080	0.042	0.098	0.048
2.5	0.80	1.5	0.8	0.083	0.029	0.066	0.027	0.075	0.040
2.5	1.20	1.5	0.8	0.214	0.105	0.127	0.049	0.076	0.040
2.5	1.80	1.5	0.8	0.609	0.422	0.447	0.235	0.092	0.049
3.0	0.00	1.5	0.8	0.100	0.050	0.100	0.050	0.100	0.050
3.0	0.15	1.5	0.8	0.088	0.036	0.085	0.044	0.094	0.056
3.0	0.30	1.5	0.8	0.069	0.024	0.077	0.027	0.081	0.041
3.0	0.80	1.5	0.8	0.056	0.023	0.081	0.035	0.094	0.048
3.0	1.20	1.5	0.8	0.275	0.151	0.229	0.064	0.124	0.064
3.0	1.80	1.5	0.8	0.720	0.535	0.547	0.302	0.145	0.080

Figure 7: The Size Corrected Rejection Rates

Two Random Parameters Model



References:

Aitchison, J. and Silvey, S. D. (1958) "Maximum Likelihood Estimation of Parameters Subject to Restraints," *Annals of Mathematical Statistics*, Vol. 29, 813-828

Andrews, Donald W. K. (2001) "Testing When A Parameter is on the Boundary of the Maintained Hypothesis," *Econometrica*, Vol 69, No.3 (May, 2001), 683-734

Bhat, Chandra R. (2001) "Quasi-random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model," *Transportation Research* PartB

Bhat, Chandra R. (2003) "Simulation Estimation of Mixed Discrete Choice Models Using Randomized and Scrambled Halton Sequences," Transportation Research Part B, 37(9), 837-855

Bockenholt, Ulf (2007) "Thurstonian-Based Analyses: Past, Present, and Future Utilities," *Psychometrika* Vol.71, No. 4, 615-629

Boyd, J. Hayden and Mellman, Robert E. (1980) "The Effect of Fuel Economy Standards on the U.S. Automotive Market: An Hedonic Demand Analysis," *Transportation Research* Vol. 14A, 367-378

Bratley, P., Fox, B.L. and Niederreiter, H. (1992) "Implementation and tests of low-discrepancy sequences" *ACM Transactions on Modeling and Computer Simulation* 2, 195-213

Brownstone, David and Train, Kenneth (1996), "Forecasting New Product Penetration with Flexible Substitution Patterns,"

Buse, A. "The Likelihood Ratio,Wald, and Lagrange Multiplier Tests: an Expository Note," *The American Statistician*, Vol 36,.No. 3, Part 1.153-157

Cardell, N. Scott and Dunbar, Frederick C. (1980) "Measuring the Societal Impacts of Automobile Downsizing," *Transportation Research* Vol. 14A, 423-434

Fang and Wang (1994) *Number-theoretic Methods in Statistics*

Gourieroux, Christian, Holly, Alberto and Monfort, Alain (1982) "Likelihood Ratio Test, Wald Test, and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters," *Econometrica*, Vol 50, No. 1, 63-88

Halton, J. H. (1960) "On the Efficiency of Certain Quasi-random Sequences of Points in Evaluating Multi-dimensional Integrals," *Numerishe Mathematik* 2, 84-90

Hausman, Jerry A. and Wise, David A. (1978) “A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences,” *Econometrica*, Vol. 46, No. 2, 403-426

Hillier, G.H. (1986) “Joint Tests for Zero Restrictions on Non-negative Regression Coefficients,” *Biometrica*, 73, 657-669

Lancaster, Kelvin J. (1966) “A New Approach to Consumer Theory,” *The Journal of Political Economy*, Vol. 74, No. 2, 132-157

Luce, R. Duncan *Individual Choice Behavior*, New York: Wiley, 1959

Luce, R. Duncan and Suppes, Pat *Handbook of Mathematical Psychology* New York: Wiley, 1966

Manski, Charles F. (2001) “Daniel McFadden and the Econometric Analysis of Discrete Choice,” *The Scandinavian Journal of Economics*, Vol.103, No. 2, 217-229

Marschak, Jacob (1959). “Binary-Choice Constraints and Random Utility Indicators,” Stanford Symposium on Mathematical Methods in the Social Sciences, Stanford, CA: Stanford University Press, 1960, 312-329

McFadden, Daniel (1974) “Conditional Logit Analysis of Qualitative Choice Behavior,” *Frontiers in Econometrics*, New York Academic Press, 105-142

McFadden, Daniel (2001). “Economic Choices,” *The American Economic Review*, Vol. 91, No. 3, 351-378

Morokoff, W.J. and Caflisch, R.E. C (1995) “Quasi-Monte Carlo Integration” *Journal of Computational Physics* 122, 218-230

Neumann, John Von and Morgenstern, Oskar (1953) Theory of Games and Economic Behavior, Princeton University Press

Niederreiter, Harald (1992) Random Number Generation and Quasi-Monte Carlo Mehods

Revelt, David and Train, Kenneth E. (1997) “Mixed Logit with Repeated Choices: Households’ Choices of Appliance Efficiency Level,”

Thurstone, L.L. (1927) “A Law of Comparative Judgment,” *Psychological Review*, 34, 273-286

Train, Kenneth E. (2003) *Discrete Choice Methods with Simulation*

Shapiro, Alexander (1985) “Asymptotic Distribution of Test Statistics in the Analysis of Moment Structures under Inequality Constraints,” *Biometrika*, 72, 1, 133-144

Silvey, S. D. (1959) “The Lagrangian Multiplier Test,” *Annals of Mathematical Statistics*, Vol 30, 389-407

Simon, Herbert A. (1959). “Theories of Decision-Making in Economics and Behavioral Science,” *The American Economic Review*, Vol. 49, No. 3, 253-283

[i] We use SAS 9.2 and NLOGIT version 4.0 to check and compare the results of our GAUSS program. The Monte Carlo experiments were programmed in Gauss 9.0 using portions of Ken Train’s posted Gauss code.