Masking Identification of
Discrete Choice Models under Simulation Methods

Lesley Chiou[1] and Joan L. Walker[2]

We present examples based on actual and synthetic datasets to illustrate how simulation methods can mask identification problems in the estimation of discrete choice models such as mixed logit. Simulation methods approximate an integral (without a closed form) by taking draws from the underlying distribution of the random variable of integration. Our examples reveal how a low number of draws can generate estimates that appear identified, but in fact, are either not theoretically identified by the model or not empirically identified by the data. For the particular case of maximum simulated likelihood estimation, we investigate the underlying source of the problem by focusing on the shape of the simulated log-likelihood function under different conditions.

JEL classification: C15, C25
Keywords: simulation methods, discrete choice, identification

---

* The authors would like to thank Moshe Ben-Akiva, Denis Bolduc, Victor Chernozhukov, Glenn Ellison,
John Geweke, Nancy Rose, and anonymous r4.7(m si)-6:e*(nderlying t(4.7(m om)8.6e*(nder.6(a)4.5(k.6(aT4.342(e)-1.5(,)6(.Ap)

Over the past decade, simulation methods have grown in popularity as advancements in computational speed have allowed researchers to estimate increasingly richer models of consumer and firm behavior. In particular, the literature on consumer choice theory has spread rapidly with the development of numerical techniques such as

We present examples of mixed logit models where employing a “low” number of draws to construct the simulated integral can generate estimates that are not identified by the model or the data. In each of the examples, we estimate the model under different types of simulation draws: random, Halton, and shuffled Halton. We consider two types of identification problems: theoretical and empirical unidentification. Theoretical unidentification occurs when the model cannot be estimated in principle (regardless of the data at hand). Empirical unidentification occurs when the data cannot support the model even though the model may be estimable in principle.

The first example utilizes a dataset on consumers’ choices of telephone plans. It demonstrates how a model that is not theoretically identified can appear to result in identified estimates at a low number of draws. A classic symptom of unidentification, a singular Hessian, does not emerge until a much higher number of draws is employed. In the second set of examples, we generate synthetic datasets to investigate the source of empirical unidentification by examining the shape of the simulated log-likelihood functions under varying numbers of draws and identification conditions. The last two examples use an actual dataset on consumer choices across retail stores and a synthetic dataset to consider empirical unidentification under more complex specifications.

Limited research exists on the empirical identification of discrete choice models

models and support their findings with empirical examples. In contrast, this paper focuses primarily on the issue of empi

A consumer chooses the alternative that gives her the highest utility. More specifically, the set of values A_{ni} of the idiosyncratic error ε_{ni} that induces consumer n to choose alternative i is given by:

$$A_{ni} \quad \{ \ {}_{ni} : U_{ni}(X_{ni} , \ {}_{n} , \ {}_{ni}) \quad \max_{j \quad 1,\ldots,J_n} U_{nj}(X_{nj} , \ {}_{n} , \ {}_{nj})\}$$

To estimate the mixed logit model under maximum simulated likelihood, we construct the log-likelihood by calculating each individual's probability P_{ni}

We focus on three common procedures for generating draws from a density. The most straightforward approach obtains draws through a pseudo-random number generator available in most statistical software.

An alternative approach creates draws based on a deterministic Halton sequence (Halton, 1960). Train (1999, 2003) provide an explanation and an example of the construction of a Halton sequence. In general, a Halton sequence can be created from any prime number p. The unit interval [0,1] is divided into p equally-sized segments, and the endpoints or “breaks” of these segments form the first p numbers in the Halton sequence. Successive numbers in sequence are generated by further subdividing each segment into p equally-sized segments and adding the breaks in a particular order.

The resulting Halton draws achieve greater precision and coverage for a given number of draws than random draws, since successive Halton draws are negatively correlated and therefore tend to be “self-correcting” (Train, 2003). In fact, Bhat (2001) demonstrates that for a mixed logit model, 100 Halton draws provided results that were more accurate than 1000 random draws.

Since each Halton sequence is constructed from a prime number, each dimension of simulation corresponds to a different

X_{ni} is a (1x5) vector consisting of four alternative-specific constants and the log of the cost of the service, and α is the *(5x1)* vector of associated taste parameters. The variable $NEST_i$ is a *(1x2)* vector of dummies for each nest. The parameter β_n is a (2x1) vector consisting of $\beta 1_n$ distributed $N(0, \sigma_1^2)$ and $\beta 2_n$ distributed $N(0, \sigma_2^2)$ where $\beta 1_n$ and $\beta 2_n$ are independent. More detail on the dataset and model can be found in Train, McFadden, and Ben-Akiva (1987) and Walker (2001).

Walker (2001) discusses conditions for identification of the model and shows that only the value $(\sigma_1^2 + \sigma_2^2)$ is identified. That is, when exactly two nests exist, only one nesting parameter is identified and to estimate the model an identifying constraint must be imposed, e.g., $\sigma_1 = \sigma_2$, $\sigma_1 = 0$, *or* $\sigma_2 = 0$.

Table 1 shows the estimation results[1] for a specification that does not include the necessary identifying constraint. Even without a necessary identifying restriction, the estimation procedure generates estimates that appear identified under a low number of draws (1000 Halton, 5000 pseudo-random). A large number of simulation draws (in this case, 2000 Halton draws) are necessary before resulting in a singular Hessian.

Not realizing the identification condition can lead to incorrect conclusions drawn from hypothesis tests based on standard errors. Since the parameter estimates and standard errors are poorly approximated under a low number of draws, they are a function of the specific draws and the starting values that are used. For example, the 2000 pseudo-random draw results for the telephone dataset would lead the modeler to incorrectly conclude that there is no correlation within the first nest (measured), but there

[1] All estimation results in this paper were estimated using either (1) BIOGEME using the DONLP2 optimization routine (see Bierlaire, Bolduc and Godbout, 2004, and http://roso.epfl.ch/biogeme) or (2) a MATLAB implementation of Kenneth Train's GAUSS code using a BHHH-algorithm (Berndt, et al., 1974) to compute the Hessian (for further information, see Chiou, 2005). Both estimation programs were shown to produce similar results.

is correlation within the second nest (flat). However, the estimation results under 5000 pseudo-random draws lead to the opposite conclusion (correlation among the measured alternatives but not among the flat alternatives).

In this section, we use extremely simple, synthetic datasets to examine the source of the empirical unidentification by investigating the properties of the simulated log-likelihood function as the number of draws increase. We show that regardless of whether the model is empirically identified, the simulated log-likelihood for mixed logit is always globally concave under only one draw because the model is analogous to a standard logit. When a model is not empirically identified, the simulated log-likelihood function begins to flatten and exhibit a singular Hessian only as the number of draws increases. The obfuscation of the identification problem occurs in the intermediate cases where the log-likelihood still exhibits the concavity as when only one draw is used. For the discussion, we consider two examples of the most common ways in which mixed logit is applied: first error components with a nesting formulation and then random coefficients on continuous explanatory variables.

We consider a simplified case where only one parameter is estimated. The discrete choice model consists of five alternatives that are divided into two nests, and the first three alternatives comprise Nest 1. Consumer *n*'s utility of choosing alternative *i* is given by:

$$U_{ni} = NEST1_i \beta_n + \varepsilon_{ni} \tag{9}$$

where *NEST1* is a dummy for whether alternative *i* lies in Nest 1. We specify the random coefficient β_n with a normal distribution, $N(0, \sigma^2)$; thus only one parameter σ must be estimated. Strictly speaking, the resulting standard deviation is calculated as $\sqrt{\sigma^2}$, and therefore the sign of the estimated coefficient σ is irrelevant. As the number of draws approaches infinity, the simulated log-likelihood will be symmetric about zero.

We generate the synthetic data according to the true value $\sigma = 2.0$ by creating observations for *N* consumers. For each consumer, we calculate the utility of each alternative by taking a single draw of β_n

the number of draws R is equal to one, then the simulated probability of consumer n choosing alternative i is:

$$\hat{P}_{ni} = \frac{exp(X_{ni}\beta_n)}{\sum_{j=1}^{J_n} exp(X_{nj}\beta_n)} \tag{10}$$

In the estimation procedure, the random coefficient for a single draw is decomposed as $\beta_n = \overline{\beta} + \sigma\,\upsilon_n$ where υ_n, $n=1,...,N$, are independent draws from a standard normal distribution. The coefficients to be estimated are $\overline{\beta}$ and σ, which are the population mean and standard deviation for the random taste. Substituting this expression into (10), we obtain:

$$\hat{P}_{ni} = \frac{exp(X_{ni}(\overline{\beta} + \sigma\upsilon_n))}{\sum_{j=1}^{J_n} exp(X_{nj}(\overline{\beta} + \sigma\upsilon_n))} = \frac{exp(X_{ni}\overline{\beta} + X_{ni}\upsilon_n\sigma)}{\sum_{j=1}^{J_n} exp(X_{nj}\overline{\beta} + X_{nj}\upsilon_n\sigma)} = \frac{exp(X_{ni}\overline{\beta} + W_{ni}\sigma)}{\sum_{j=1}^{J_n} exp(X_{nj}\overline{\beta} + W_{nj}\sigma)}. \tag{11}$$

where $W_{ni} = X_{ni}\upsilon_n$ is defined as a "new" variable created from X_{ni} and the particular draw for the consumer n. This is the standard logit formula where fixed parameters $\overline{\beta}$ and σ are estimated over the variables X_{ni} and W_{ni}. Since the standard logit model is globally concave (Train, 2003), estimation will always return a non-singular Hessian.

In our example, a draw υ_n is taken from a $N(0,1)$ distribution, and β_n is calculated as $\beta_n = \sigma\upsilon_n$. The "new" variable is $\upsilon_n*NEST1_i$. In other words, we can reinterpret the utility function as:

$$U_{ni} = \sigma(NEST1_i * \upsilon_n) + \varepsilon_{ni} \tag{12}$$

where σ is a fixed coefficient on the variable $\upsilon_n*NEST1_i$. Not surprisingly, the local maximum when 1 draw is used occurs near the origin, since the purely random draw has no explanatory power.

On the other hand, the singularity of the Hessian is evident under 1000 random draws. In Figure 1, the simulated log-likelihood function rises away from 0, reflecting that the true value of σ is not zero, but the log-likelihood function flattens at higher magnitudes of σ. The data cannot empirically distinguish among the higher magnitudes of σ. In the intermediate case of 10 random draws, the simulated log-likelihood still exhibits a single peak as in the case of 1 draw. The local concavity gives rise to a convergence of the maximization routine.

Due to the efficiency of Halton draws relative to random draws, the Halton draws achieve the same unidentification properties at a lower threshold. In Table 2, the singularity of the Hessian occurs at 35 Halton draws whereas 35 random draws still generate a local maximum. Moreover, the large standard error of 101.142 under 10 Halton draws suggest the presence of an identification problem.

Table 3 reports the estimates of the standard deviation σ when the dataset contains N = 10000 observations. In contrast to the previous case of only 50 observations, the dataset of 10000 observations is sufficient to empirically identify the model. The parameter estimates stabilize and approach the true value of $\sigma = 2.0$ as the number of draws increases. Figures 3 and 4 graph the simulated log-likelihood as a function of the parameter σ for a varying number of random and Halton draws, indicating a unique maximum (disregarding the sign) even for large number of draws.

The dataset generated for this example uses three alternatives ($i = 1,2,3$) and two explanatory variables (*X1* and *X2*). Unlike the previous example with a nest dummy, the

parameter estimates for 1000 and 2000 draws; these estimates are also close to the true values used to generate the synthetic data.

The simulated log-likelihood functions for mixed logit models are not as well-behaved as a standard logit model; mixed logit likelihood functions are not globally concave and are sensitive to poor approximations of the simulated probabilities for low number of draws. The results from these simple examples show that for a small number of draws, an empirically unidentified model can appear identified. It is only after a sufficient number of draws is used that the shape of the log-likelihood reveals the singularity of the Hessian.

The simple examples from the previous section were used to explore the source of empirical unidentification and to demonstrate the behavior of the log-likelihood function under different numbers of draws and under different conditions of identification. This section considers estimation results for more realistic datasets, including one real dataset regarding households' purchases of DVDs and one synthetic dataset.

In the first example, we apply data from Chiou (2005) to examine a household's choice of retail store to purchase a DVD. We estimate a consumer's choice of store conditional on the purchase of a DVD. The consumer's choice set consists of retail stores from the top 15 chains that sell DVDs, and each retail store is classified under one of five store types: mass merchant, video specialty, electronics, music, and online. Consumer *n*'s utility from traveling to store *i* to purchase her chosen video is given by:

$$U_{ni} = X_{ni}\alpha + \sum_{k=1}^{5} TYPE_{ik} * \beta_{nk} + \varepsilon_{ni}$$

Table 6 presents the results under 1, 100, 200, and 1000 draws with pseudo-random, Halton, and shuffled Halton draws.[4] Although the model specification is theoretically identified, the results with high numbers of draws indicate that the parameters are not empirically identified by the data. Nonetheless, optimizations from 200 random or 100 Halton draws converge and generate estimates that appear identified. Without checking the robustness of the estimates to varying number of draws, the unidentification issue would not be apparent. For instance, under 200 random draws, the mean and standard deviation of the population distribution of tastes over video specialists are 19.391 (6.381) and 1.501 (0.446), and the coefficients are significant at the 1% level. Similarly, under 100 Halton draws, the mean and standard deviation of the random coefficient on video specialists are 19.576 (8.817) and 1.244 (0.507).

The identification issue becomes readily apparent under 100 shuffled Halton draws; the optimization routine does not converge as parameter estimates explode. The results suggest that shuffled Halton draws expose the identification issue at a lower number of draws relative to other types of draws.

The above retail and synthetic datasets consist of relatively simple specifications. In this section, we present an example to illustrate how simulation difficulties become more apparent as additional complexities are introduced into the model. The example uses a dataset that consists of 2000

With advancements in computational speed, simulation techniques have vastly improved the ability to estimate complex models to answer a myriad of questions. However, the implementation of simulation can often mask problems of identification. A low number of draws can result in estimates that appear identified, but in fact are not identified either theoretically by the model or empirically by the data.

To highlight the issue, we present examples of maximum simulated likelihood estimation of mixed logit models under actual and synthetic datasets. Although each of

In addition to the empirical results, the underlying source of these issues was investigated by examining the shape of the log-likelihood function under varying numbers of draws and different identifyi

Ben-Akiva, M. and D. Bolduc, 1996, Multinomial probit with a logit kernel and a general parametric specification of the covariance structure. Working paper, Massachusetts Institute of Technology.

Berndt, E., B. Hall, R. Hall, and J. Hausman, 1974, Estimation and inference in nonlinear

Chiou, L., 2005, Empirical analysis of retail

Train, K., 1999, Halton sequences for mixed logit. Working paper, University of California, Berkeley.

Train, K., 2003, Discrete choice methods with simulation. Cambridge University Press, New York.

Train, K., D. McFadden, and M. Ben-Akiva, 1987, The demand for local telephone service: A fully discrete model of residential calling patterns and service choices. RAND Journal of Economics 18(1), 109-123.

Walker, J. L., 2001, Extended discrete choice models: Integrated framework, flexible

Table 1. The Demand for Telephone Service

σ_1 σ_2				
$(\sigma_1^2+\sigma_2^2)^{1/2}$				

Table 2. Empirical Unidentification with a Random Coefficient on a Nest Dummy

σ			

Notes: Standard error in parentheses. Uses non-robust standard errors.

Table 3. Empirical Identification with a Random Coefficient on a Nest Dummy

Table 4. Empirical Unidentification with Random Coefficients on Continuous Variables

σ1		
σ2		

Table 5. Empirical Identification with Random Coefficients on Continuous Variables

$\sigma 1$		
$\sigma 2$		

Notes: Standard error in parentheses. Uses robust standard errors.

Table 6. Retail Stores

Note: Uses non-robust standard errors.

Table 7. Synthetic Results

Figure 1. Unidentified Model with Random Coefficient on a Nest Dummy (50 observations, Random Draws)

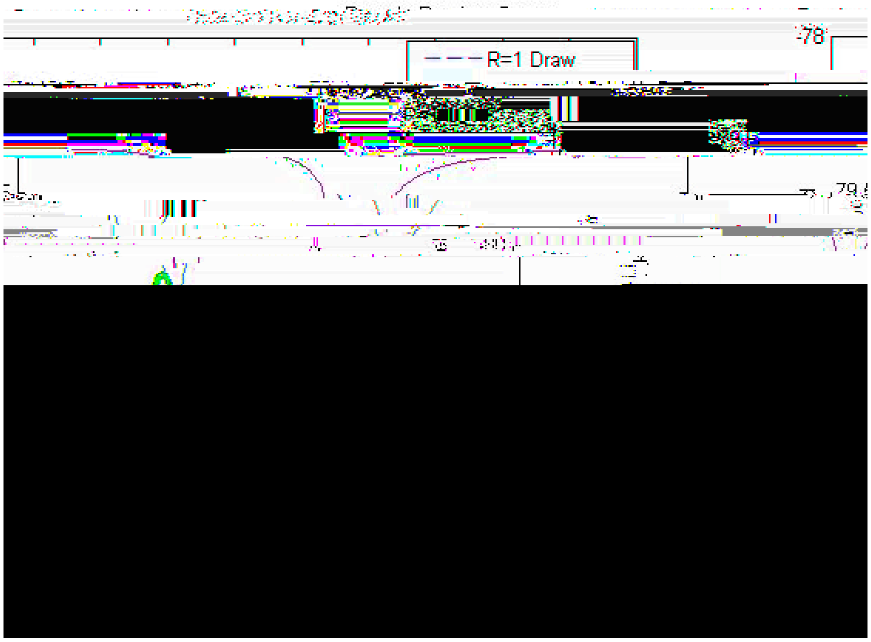


Figure 2. Unidentified Model with Random Coefficient on a Nest Dummy (50 observations, Halton Draws)

Figure 3. Identified Model with Random Coefficient on a Nest Dummy (10000 observations, Random Draws)

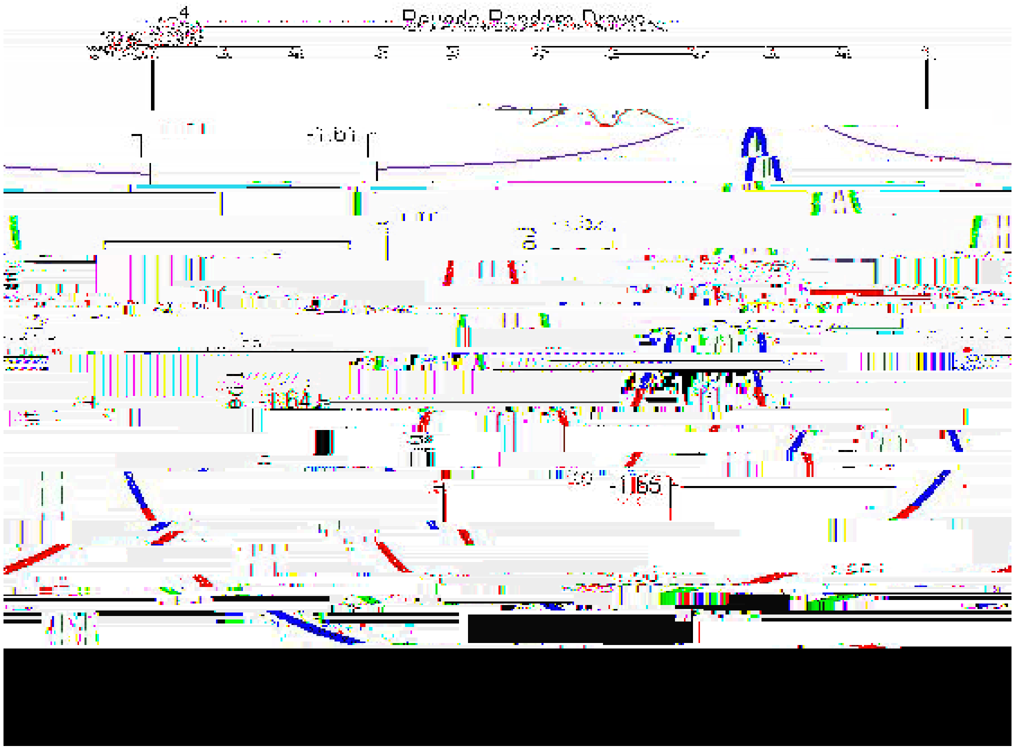


Figure 4. Identified Model with Random Coefficient on a Nest Dummy (10000 observations, Halton Draws)

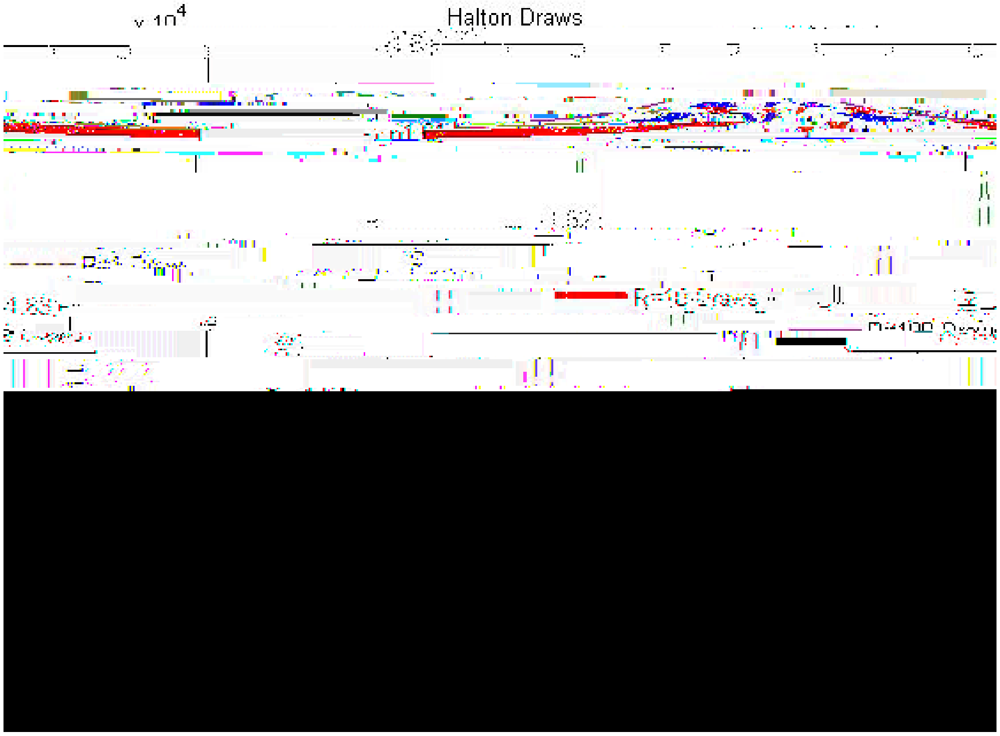


Figure 5. Unidentified Model with Random Coefficient on Normally Distributed Variable (100 observations, Random Draws)

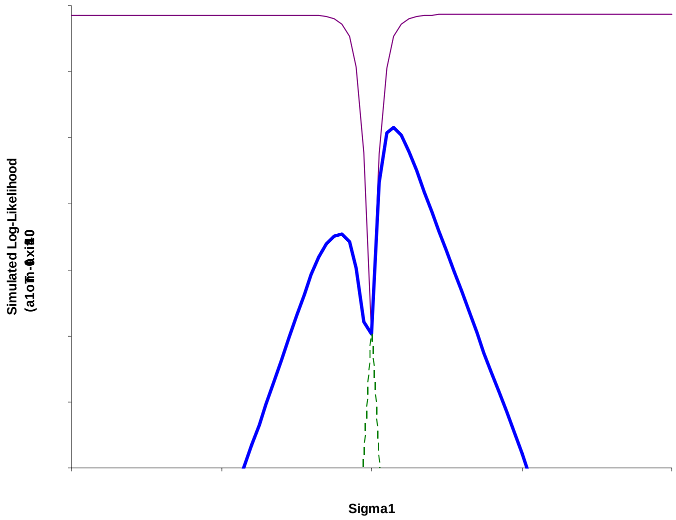
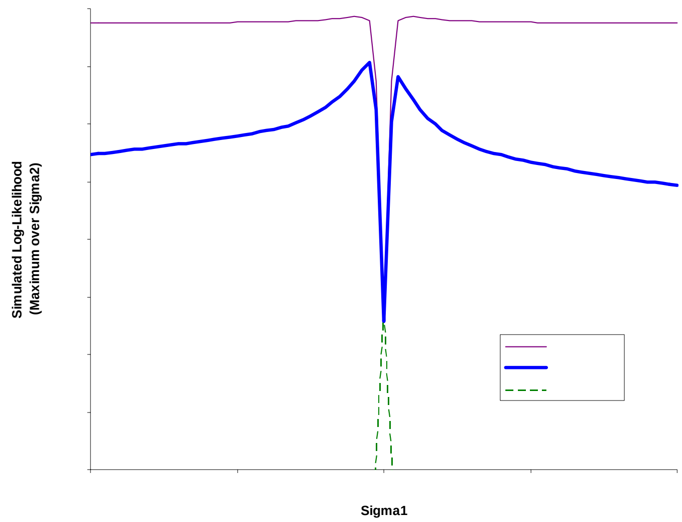


Figure 6. Identified Model with Random Coefficient on Normally Distributed Variable (10000 observations, Random Draws)



Note: This figure plots the maximum of the Simulated Log-Likelihood over Sigma2 for a given value of Sigma1.