RURAL ECONOMY

Dollars or Cents: Impacts of Rescaling Data on a Mixed Logit Model with Normally and Lognormally Distributed Coefficients

Wuyang Hu

Staff Paper 04-01

Staff Paper

Department of Rural Economy
Faculty of Agriculture, Forestry
and Home Economics
University of Alberta
Edmonton, Canada

Dollars or Cents: Impacts of Rescaling Data on a Mixed Logit Model with Normally and Lognormally Distributed Coefficients

Wuyang Hu

Staff Paper 04-01

November 2004

The author is Postdoctoral Fellow, Department of Rural Economy, University of Alberta.

The author thanks Wiktor Adamowicz, Kenneth Train, and David Hensher for helpful comments. The author is responsible for any remaining errors.

Abstract

It is a common practice to rescale data to assist the model estimation process. This paper describes a case where care is called upon when interpreting the results after rescaling. The case is shown as associated with the lognormally distributed coefficient in a mixed logit model. Implications of rescaling data on a normally distributed coefficient are also given for comparison.

Introduction

The mixed logit (ML) model, sometimes also referred to as the random parameter logit or error component logit model (Train 2003), has proved to be flexible and useful in modeling taste heterogeneity in random utility models. Also due to its simplicity compared with other flexible models (e.g., probit or mixed probit model), it has drawn a significant amount of attention in the economics, marketing, and transportation literature. This paper describes a problem in interpretation of the model results after rescaling the data that is specified to be associated with a lognormally distributed random utility coefficient. Rescaling is a common practice used to alleviate numerical problems in estimation or simply for ease of interpretation. However, in this paper we show that the t-ratio associated with the estimate of one particular parameter in a lognormally distributed coefficient is not independent from the weighting factor used in rescaling although the likelihood function and model fit should be identical for the mixed logit model before and after rescaling the data. This finding prevents the attempt to interpret that estimated parameter directly.

Mixed Logit Model

Choice models can be used to describe under what conditions consumers are more likely to make a purchasing/participating decision versus not. In these models, researchers often wish to assume, for example, the coefficient associated with the price variable to be non-positive. If the price coefficient is to be specified as random corresponding to the belief that consumers may be different in their “taste” or sensitivity towards price, the density function of the coefficient cannot have any mass on negative values. In the literature researchers have applied a series of different types of distributions to satisfy this restriction, including lognormal, gamma, Rayleigh, and various truncated symmetric distributions. Among these the lognormal distribution is by far the most commonly used distribution (e.g., Train 1998; Hensher and Greene 200).

In a random utility model (RUM), the utility of consumer *n* choosing alternative *i* among *J* total available alternatives can be written as:

$$U_{ni} = V_{ni} + e_{ni}, \qquad i = 1,2,\ldots,J \tag{1}$$

where V_{ni} and e_{ni} are respectively the deterministic and stochastic portion of the utility from the perspective of the researcher. Suppose three variables X_1, X_2, and X_3 are used to directly describe V_{ni}, then in a parametric context, one can write:

$$V_{ni} = \beta_1 X_1 + \beta_{2n} X_2 + \beta_{3n} X_3 \tag{2}$$

X_1 to X_3 are respectively assumed to be an alternative specific constant (ASC), a continuous quality index, and the price variable multiplied by negative one. To allow the model to be representative and incorporate both fixed and random coefficients, we specify β_1 as a fixed coefficient and β_{2n} and β_{3n} as random coefficients with normal and lognormal distributions respectively. We further assume that a variable *Z* (e.g., a demographic characteristic) can be used to explain the heterogeneity in the "average" of each of the two random coefficients across the sampled individuals. The most popular parameterization for β_{2n} and β_{3n} in the literature are as follows:

$$\beta_{2n} = a_0 + a_1 Z + \sigma_a \varepsilon_a, \ \varepsilon_a \sim Normal(0,1)$$

and (3)

$$\beta_{3n} = \exp(b_0 + b_1 Z + \sigma_b \varepsilon_b), \ \varepsilon_b \sim Normal(0,1)$$

For simplicity, we assume that ε_a and ε_b are independent. Parameters $\alpha_0, \alpha_1, b_0, b_1, \sigma_a$ and σ_b are to be estimated. Substituting (3) into (2) and then (2) into (1), one generates the complete indirect utility function. If e_{ni} is assumed to be iid Gumbel distributed, the implied choice probability can be represented by a mixed logit model with choice probability:

$$\overline{P}_{ni} = \int P_{ni} f(\beta_{2n}) f(\beta_{3n}) d\beta_2 d\beta_3 \qquad (4)$$

where $P_{ni} = \dfrac{\exp(V_{ni})}{\sum_{j=1}^{J} \exp(V_{nj})}$

This probability does not have a closed form but it can be simulated. Suppose β^d is the *d*-th draw of β_{2n} and β_{3n} from their corresponding density functions, $\beta^d = (\beta_{2n}^d, \beta_{3n}^d)$ (draws of β_{2n}^d and β_{3n}^d are independent). Given the total *D* numbers of draws, the simulated probability can be written as:

$$\tilde{P}_{ni} = \frac{1}{D} \sum_{d=1}^{D} (P_{ni} \mid \beta^d) \qquad (5)$$

The simulated log-likelihood function for individual *n* is:

$$SLL = \sum_{i=1}^{J} c_{ni} \ln(\tilde{P}_{ni}) \qquad (6)$$

where $c_{ni} = 1$ if alternative *i* is chosen by individual *n*.

The Problem of Rescaling

It is well known that when the magnitudes of the variables in a model differ very significantly (say by a factor of 1000), the numerical procedure for maximization may have difficulty achieving convergence. In this case, researchers usually proceed by scaling up (down) the excessively small (large) variables to help the maximization algorithm. The requirement for scaling may come from the data directly. For example, for a study on consumers' choices of food items, the price variable may be measured by a small unit in the original data, such as cents rather than dollars, or in a small currency unit such as the Japanese Yen. These indicate that the magnitude of the measures in the price data will be relatively large. However, in the same model, all other variables may be relative small (e.g., dummy variables). In these cases, scaling is often necessary.

The model fit (LL value and ρ^2 statistic) should be identical before and after rescaling and if any X_1 to X_3 is scaled up (down) by a certain factor, the associated coefficient β_1, β_{2n} or β_{3n} will be scaled down (up) by the same factor. It is simple to show that the impact of rescaling X_1 on β_1 is the same as if in a linear regression model. To examine the impact of scaling on the "deeper" parameters of the random coefficients, one can write:

$$\beta_{2n} X_2 = a_0 X_2 + a_1 Z X_2 + \sigma_a \varepsilon_a X_2$$

and (7)

$$\beta_{3n} X_3 = [\exp(b_0 + b_1 Z + \sigma_b \varepsilon_b)] X_3 = \exp(b_0)(\exp(b_1))^Z (\exp(\sigma_b))^{\varepsilon_b} X_3$$

For the normally distributed coefficient, it can be easily seen that holding Z unchanged, if X_2 is scaled up by a factor f, then the estimated parameters a_0, a_1, and σ_a will all be scaled down by f. However, this is not obvious in the case of lognormally distributed coefficient. Suppose X_3 is scaled up by a factor f, then the expression $\exp(b_0)(\exp(b_1))^Z(\exp(\sigma_b))^{\varepsilon_b}$ will be scaled down by f. In contrast to the case of the normal distribution, these terms are multiplicative rather than additive. By further observing the three terms involved, it can be seen that parameter b_1 is directly associated with variable Z which is not affected by rescaling, and σ_b is directly associated with the random variable ε_b, which follows a standard normal distribution. The only "free" parameter that can be scaled down in order for β_{3n} to be scaled down by f is b_0. In other words, the estimated b_0 after scaling must satisfy

$$\frac{\exp\left(b_0^{before\ \ scaling}\right)}{\exp\left(b_0^{after\ \ scaling}\right)} = f \tag{8}$$

In simulated maximum likelihood estimation, the asymptotic variance of parameters is the corresponding diagonal of the inverse Hessian matrix. It is known that the exact Hessian for a mixed logit model often cannot be inverted due to the complexity of the model. Usually the approximated Hessian is computed instead and by far the most commonly used method is the BHHH algorithm. Under this algorithm, the Hessian matrix is approximated by the outer product of the first-order derivative of the SLL with respect to the parameters:

$$H' = \frac{\partial SLL}{\partial \theta}\frac{\partial SLL}{\partial \theta}' \tag{9}$$

where *H'* is the approximated Hessian and θ stands for all the parameters involved in the model. For parameters defining the normally distributed coefficient, it is not difficult to show that

$$\left.\frac{\partial SLL}{\partial \theta}\right|_{\beta_{3n}} = \frac{\partial SLL}{\partial (a_0, a_1, \sigma_a)} = \frac{1}{\tilde{P}_{ni}} \frac{1}{D} \sum_{d=1}^{D} P_{nid} \left(X_2 - \sum_{j=1}^{J} P_{njd} X_2 \right) \frac{\partial \beta_{2n}}{\partial (a_0, a_1, \sigma_a)}$$

$$= \frac{1}{\tilde{P}_{ni}} \frac{1}{D} \sum_{d=1}^{D} P_{nid} \left(X_2 - \sum_{j=1}^{J} P_{njd} X_2 \right) \begin{bmatrix} 1 \\ Z \\ \varepsilon_a \end{bmatrix} \tag{10}$$

Since scaling the data does not affect the overall model result, $\tilde{P}_{ni}$ and P_{nid} will remain unchanged after scaling. It can be seen that if X_2 is scaled up by *f*, then $\frac{\partial SLL}{\partial (a_0, a_1, \sigma_a)}$ will be scaled up by *f* as well. This implies that the inverse of the *H'* matrix will be scaled *down* by f^2, which in turn implies that the estimated standard errors of the parameters in β_{2n} will be scaled down by *f*. Combining the result we have on the impact of scaling on the parameter estimates themselves, the conclusion is that the t-ratios of these parameters will remain unchanged before and after rescaling the data.

Similarly, for the lognormally distributed coefficient β_{3n}, the first-order derivative of the SLL with respect to the parameters is:

$$\frac{\partial SLL}{\partial (b_0, b_1, \sigma_b)} = \frac{1}{\tilde{P}_{ni}} \frac{1}{D} \sum_{d=1}^{D} P_{nid} \left(X_3 - \sum_{j=1}^{J} P_{njd} X_3 \right) \frac{\partial \beta_{3n}}{\partial (b_0, b_1, \sigma_b)}$$

$$= \Gamma \frac{\partial \beta_{3n}}{\partial (b_0, b_1, \sigma_b)} \tag{11}$$

However, given the expression of β_{3n} in (3), it follows that:

$$\frac{\partial \beta_{3n}}{\partial(b_0, b_1, \sigma_b)} = \begin{bmatrix} \beta_{3n} \\ \beta_{3n} Z \\ \beta_{3n} \sigma_b \end{bmatrix}_{\beta_{3n} = \hat{\beta}_{3n}} \tag{12}$$

If X_3 is scaled up by f, then Γ will be scaled up by f as well. However, since $\frac{\partial \beta_{3n}}{\partial(b_0, b_1, \sigma_b)}$ is a function of $\hat{\beta}_{3n}$, $\hat{\beta}_{3n}$ is scaled down by f due to the scaling in X_3. Therefore, the expression $\frac{\partial SLL}{\partial(b_0, b_1, \sigma_b)}$ will not be affected by the scaling in X_3. This further implies that the estimated standard errors for the parameters in β_{3n} will remain unchanged as well. However, it is noted previously that the estimated parameter b_0 *will* be changed to maintain the relationship in (8). The resulting effect is that the t-ratio for parameter b_0 will not be constant before and after rescaling of the data. In other words, the t-ratio for b_0 is dependent on the scaling factor *f*. To complicate the situation, there is no "correct" scaling factor *f* one can use. Any arbitrary rescaling may be used, dollars or cents for example, and still generate identical model fit but with different t-ratios associated with the parameter b_0.

Concluding Remarks

The findings of this paper can be summarized as follows:

1. Scaling data will not affect the overall model fit and estimation.
2. Scaling variable with a normally distributed random coefficient changes the magnitude of the estimated parameters of the random coefficient but does not change their corresponding t-ratios.

3. Scaling variables with a lognormally distributed random coefficient will not change the magnitude of the estimated parameters of the random coefficient except the constant term in the distribution.
4. Scaling variables with a lognormally distributed random coefficient will not change the standard errors associated with the parameters associated with the random coefficient. However, the implied t-ratio of the constant term will be affected.

It is noticeable that these results hold regardless of whether variables explaining heterogeneity in mean random coefficients are specified (the *Z* variable). Researchers who wish to estimate a lognormally distributed coefficient in a mixed logit model should be aware of these impacts of scaling and not interpret the constant term separately but rather interpret the overall mean or median of the lognormal term, which in this study are given as $\exp\left(b_0 + b_1 Z + \left(\sigma_b^2/2\right)\right)$ and $\exp(b_0)$. The standard errors of these two measures can be obtained by the delta method or simulation.

References

Hensher, D. A. and W. H. Greene (2003) “Mixed Logit Models: State of Practice” *Transportation* 30(2): 133-176.

Train, K. (2003) *Discrete Choice Models with Simulation* Cambridge University Press, Cambridge.

Train, K. (1998) “Recreation Demand Models with Taste Variation over People” *Land Economics* 74(2): 230-239.