Title
Simulation Methods for Probit and Related Models Based on Convenient Error Partitioning

Author
Train, Kenneth E.

Publication Date
1995-06-01

Peer reviewed

UNIVERSITY OF CALIFORNIA AT BERKELEY

Department of Economics

Berkeley, California 94720-3880

Working Paper No. 95-237

Simulation Methods for Probit and Related Models Based on Convenient Error Partitioning

Kenneth E. Train

Department of Economics
University of California, Berkeley

June 1995

Key words: Probit, Simulation

JEL Classification: C13, C15, C25

Abstract

Two probit simulators are described that are conceptually and computationally simple. The first is based on simulating the utilities of the non-chosen alternatives and calculating the probability that the chosen alternative's utility exceeds this maximum. This simulator is apparently new. The second, which is implicit in the discussions of McFadden (1989) and Bolduc (1992), is applicable when the covariance among utilities arises from random parameters and/or error components that are common across alternatives. The parameters and common error components are simulated, and then the probability that the observed event occurs is calculated conditional on these values. Both simulators are unbiased, strictly positive, and continuous. The second is twice-differentiable, while the first has points of non-differentiability. Both are easy to program and can be expected to be very fast computationally.

Simulation Methods for Probit and Related Models Based on Convenient Error Partitioning

Kenneth E. Train[1]
Department of Economics
University of California, Berkeley

June 1, 1995

I. Introduction

Probit models take the form $P=\int 1(\varepsilon \in A) f(\varepsilon) d\varepsilon$ where P is the probability of the observed event, ε is a vector with a multivariate normal distribution denoted f, A is the set of ε's for which the observed event occurs, and 1(.) is an indicator that ε is in A. The difficulty with probit models is that the integration cannot be performed explicitly to obtain a closed form for P. Numerical integration is required. Various simulation methods (discussed below) have been proposed that provide numerical approximations to the integral. In this paper, we present two simulators that might prove attractive in certain circumstances. The two simulators utilize the concept of partitioning the random terms, though in different ways. The general approach is as follows. Partition ε into two subvectors ε_1 and ε_2 with marginal and conditional distributions $f_1(\varepsilon_1)$ and $f_2(\varepsilon_2|\varepsilon_1)$, respectively, where the partitioning is chosen to provide desirable properties of the resulting simulator. Then $P=\int p(\varepsilon_1) f_1(\varepsilon_1) d\varepsilon_1$ where $p(\varepsilon_1) \equiv \int 1((\varepsilon_1,\varepsilon_2) \in A) f_2(\varepsilon_2|\varepsilon_1) d\varepsilon_2$ is the probability that the observed event occurs conditional on ε_1. If drawing ε_1 from its marginal distribution is easy, and $p(\varepsilon_1)$ is easy to calculate, then P can be readily simulated with this error partitioning. The simulation is performed by (1) drawing a value of ε_1 from its distribution, (2) calculating $p(\varepsilon_1)$ for this value of ε_1, (3) repeating this process for repeated draws of ε_1 and averaging the results. The simulator is unbiased by construction. Its other properties are determined by the characteristics of $p(\varepsilon_1)$. Ideally, a partitioning will provide a $p(\varepsilon_1)$ that is (a) strictly positive, so

---

[1]I have benefited from comments by Paul Ruud and Daniel McFadden on earlier drafts of this paper. I am particularly grateful to Paul Ruud for (a) pointing out how, in section II, the conditional distribution for the chosen alternative's utility can be determined without performing a separate factorization for each agent, and (b) identifying an important error in an earlier version of section III.

that the simulated probability can be inserted into the log-likelihood function for parameter estimation, (b) continuous in underlying parameters and variables, to avoid "jumps" in the simulated log-likelihood function and the calculation of arc elasticities, and (c) differentiable, so that derivatives of the simulated probability can be used in the search for the maximum of the simulated log-likelihood function and in the calculation of point elasticities.

The two simulators described below differ in their properties. Both are strictly positive, continuous, and very easy to calculate. The first simulator has points of non-differentiability, while the second is twice-differentiable at all points. Each is applicable to particular types of models. The first is applicable to cross-sectional probit with a general covariance structure for the errors. The second is applicable to random-parameters, error-components probits. It can be used with cross-sectional or panel data and can be easily adapted to non-normal distributions. Both methods are simple conceptually and easy to program.

The fact that neither of these simulators is applicable to all latent variables models with normal errors is, in a sense, a characteristic of simulators created through error partitioning. A successful partitioning (that is, a partitioning that results in a simulator with desirable properties) exploits the structure of the model. By utilizing the structure, the simulator has the potential to be simple and fast; however, its reliance on the structure prevents it from being used in situations without that structure. In this regard, the approach of error partitioning differs from the standard approach to developing simulators, which is to define simulators that operate for all types of probit models.

II. Simulating the Maximum Utility for the Non-Chosen Alternatives

In standard probit, the agent chooses the alternative with the highest utility. The chosen alternative's utility is therefore higher than the maximum of the utilities of the non-chosen alternatives. Using this fact, the probabilities can be simulated as follows. Simulate the utilities for the non-chosen alternatives, drawing from their unrestricted normal distribution. Conditional on these utilities for the non-chosen alternatives, calculate explicitly the probability that the utility

for the chosen alternative exceeds the maximum utility for the non-chosen alternatives. This probability is a one-dimensional cumulative normal, which is easy to calculate. With only one repetition, the simulated probability is strictly positive, unbiased for the true choice probability, and continuous in the underlying parameters. The method is described below and then compared with other simulators.

An agent faces a choice among J alternatives and chooses the one which provides the greatest utility. From the researcher's perspective, utility vector U is distributed $N(V,\Omega)$. Denote the chosen alternative as i with utility U_i. The utility vector for the non-chosen alternatives is denoted U_N, which is U with the i-th element removed. The chosen alternative's utility is higher than the maximum of the utilities of the non-chosen alternatives: $U_i > K(U_N)$ where $K(U_N) \equiv \max\{U_j, j=1,...,J; j\neq i\}$. The probability that alternative i is chosen is therefore:

$$P_i = \int [\text{Prob}(U_i > K(U_N) \mid U_N\}]\, f(U_N)\, dU_N.$$

where $f(U_N)$ is the distribution of U_N marginal over U_i. The term in brackets is a one-dimensional cumulative normal, which is easy to calculate, and the integral over U_N can be approximated by simulation. Conditional on U_N, the distribution of U_i is normal with mean $\alpha(U_N) \equiv V_i + \omega_{i,N}(\Omega_N)^{-1}(U_N - V_N)$ and variance $\sigma^2 = \omega_{ii} - \omega_{i,N}(\Omega_N)^{-1}(\omega_{i,N})'$, where $\omega_{i,N}$ is the i-th row of Ω with the i-th element removed, Ω_N is Ω with the i-th row and column removed, and ϖ_{ii} is the i-th element of the i-th row of Ω. Therefore

$$\text{Prob}(U_i > K(U_N) \mid U_N) = 1 - \Phi((K(U_N) - \alpha(U_N))/\sigma) = \Phi((\alpha(U_N) - K(U_N))/\sigma)$$

The simulation is performed as follows. Draw J independent standard normal deviates, and label the vector of these as ε^1 to denote that this is the first simulation. Calculate the associated utilities: $U^1 = V + \Gamma\varepsilon^1$ where Γ is the lower triangular Choleski factor of Ω. (Note that all utilities are simulated, though only the simulated utilities for the non-chosen alternatives will be used; instead, the utilities of only the non-chosen alternatives could be simulated. Computationally, it is probably faster to simulate all utilities than to locate the chosen alternative and not

simulate it. Also, conceptually, simulating all utilities and ignoring the chosen alternative's utility is a direct way of drawing utilities for the non-chosen alternatives from their marginal distribution.) Observe the largest of the simulated utilities for the non-chosen alternatives and label it K^1. Calculate $\alpha^1 = V_i + \omega_{iN}(\Omega_N)^{-1}(U_N^1 - V_N)$, and then calculate $\Phi((\alpha^1 - K^1)/\sigma)$. Repeat this process R times, labeling each repetition appropriately. The simulated approximation to P_i is the average of the R cumulative normal probabilities:

$$SP_i = (1/R)\Sigma\Phi((\alpha^r - K^r)/\sigma) .$$

SP_i is an unbiased estimator of P_i for a finite number of repetitions; it is sufficient to show for R=1:

$$E(SP_i) = E\{\Phi((\alpha^1 - K^1)/\sigma)\} = \int\Phi((\alpha^1 - K^1)/\sigma)\, f(U_N)\, dU_N = P_i.$$

SP_i is strictly positive for any R including 1, and is continuous in the underlying parameters and variables. It is not globally differentiable in parameters and variables. In particular, it has kinks at the points where the identity of the non-chosen alternative that has highest utility changes: that is, where $K(U_N)$ goes from being the utility of one non-chosen alternative to a different non-chosen alternative. However, these are kink points, not jumps, and consequently do not appreciably hamper numerical procedures for parameter estimation and forecasting.

Essentially, the simulated probability is based on the recognition that P_i is actually a three dimensional integral, rather than, as usually expressed, a J-1 dimensional integral. That is, P_i is the value of $\Phi((\alpha - K)/\sigma)$, which is itself an integral, integrated over all possible values of the maximum utility for the non-chosen alternatives, K, and the conditional mean of the utility of the chosen alternative, α:

$$P_i = \iint\Phi((\alpha - K)/\sigma)\, f(\alpha, K)\, d\alpha\, dK.$$

The joint distribution of K and α is not normal and, to my knowledge, has not been derived. As a result, this expression cannot be integrated analytically, However, it can serve as the basis of simulation without knowing the distribution explicitly, since the distribution is obtained from the distribution of utilities for the non-chosen alternatives. That is, values of α and K are drawn from their distribution by drawing values of the utilities of the non-chosen alternatives from their joint distribution and calculating the associated values of K and α.

The proposed simulator is closely aligned to accept/reject methods (Lerman and Manski, 1981) and maintains their conceptual simplicity. Like accept/reject, the utility of each non-chosen alternative is simulated; however, instead of simulating the utility of the chosen alternative and determining whether it exceeds that of the non-chosen alternatives (as in accept/reject), the probability of this event is calculated explicitly. The accept/reject probability can be zero for a finite number of draws, if no "accept" is obtained. Also, the relation between the accept/reject probability and the underlying parameters and variables is a step function: as a parameter or variable changes, there is no change in the accept/reject probability until a simulation moves from being a "reject" to an "accept" or vice versa. Numerical procedures for parameter estimation and for forecasting the impact of changes in explanatory variables (such as calculating arc elasticities numerically) are hampered as a result. These limitations are avoided in the proposed procedure by recognizing that, for any values of the utilities of the non-chosen alternatives, there is a strictly positive and continuous probability that the utility for the chosen alternative will be higher than those for the non-chosen alternatives.

The procedure has a flavor of the GHK probability simulator, in that a truncated normal is calculated for one unobserved term conditional on simulated values for other unobserved terms (Hajivassiliou and McFadden, 1992; Borsch-Supan and Hajivassiliou, 1993) However, the recursive triangulation of GHK is not followed. GHK has been found in Monte Carlo studies to perform better than other simulators including accept/reject (Hajivassiliou, McFadden, and Ruud, 1992; Borsch and Hajivassiliou, 1993.) Two computation simplifications to GHK are provided by the proposed approach: (1) utilities are drawn simultaneously from an unrestricted normal rather than sequentially from truncated normals, and (2) only one factorization is required rather

than for each chosen alternative. For GHK, the utility of the chosen alternative is subtracted from that of each non-chosen alternative, such that the covariance matrix for these utility-differences, and hence the factorization that is used in drawing them, depends on which alternative is chosen. The utility-differences are drawn sequentially, with each drawn from a truncated normal whose truncation point depends on the previously drawn utility-differences. The truncation point assures that the utility-difference is negative (meaning that the non-chosen alternative has lower utility than the chosen alternative), given that each previously simulated utility-difference is also negative. (This last consideration captures the correlations among utility-differences.) By contrast, the proposed procedure operates on utility rather than utility-differences and draws simultaneously from an unrestricted normal. The distribution from which utilities are drawn does not depend on which alternative is chosen, so that one Choleski factorization is applicable for all agents.[2,3]

Finally, the procedure is conceptually similar to the Clark algorithm (Clark, 1961; Daganzo, Bouthelier, and Sheffi, 1977) in that both emphasize the comparison of the chosen alternative's utility with the maximum utility for the non-chosen alternatives. For the Clark algorithm, the

---

[2] If the covariance matrix depends on agent-specific data, then Ω varies over agents and agent-specific factorization is required with both GHK and the proposed approach. However, rearrangement of Ω based on the agent's chosen alternative is not required in the proposed approach.

[3] As stated above, the proposed simulator is not directly applicable to panel data. However, the proposed simulator can be combined with the concepts that motivate the GHK simulator to obtain a simulator that operates on panel data. In particular, for the first time period, simulate with the proposed method the probability of the alternative that is chosen in that period. For this simulation, the utilities of the non-chosen alternatives are simulated, and the range within which the utility of the chosen alternative must fall is calculated conditional on the simulated non-chosen utilities. Draw a value of the chosen alternative's utility from this range (that is, from the appropriate truncated univariate normal.) For the second period, use the simulated values of the utilities for all alternatives in the first period to calculate the expected values of the utilities in the second period, based on the covariance among utilities over time. Continue this process for each time period. Essentially, the proposed procedure is used to capture correlations across alternatives within each time period, and the GHK method of drawing recursively utilities for which the observed event occurs is used to capture correlation over time.

concept that the maximum of two normal variables is approximately normal is applied sequentially for each non-chosen alternative to obtain a normal distribution that approximates the true distribution of the maximum utility for the non-chosen alternatives. The probability that the chosen alternative's utility is higher than the maximum non-chosen utility is then calculated explicitly by integrating over this one-dimensional normal distribution. The proposed approach uses simulation of the maximum non-chosen utility rather than the Clark algorithm to reduce the dimension of integration to one. The simulations are drawn from the true distribution for the maximum non-chosen utility, but for any finite number of draws the distribution of simulated values for the maximum non-chosen utility is an approximation to the true distribution. In effect, the two procedures are simply approximating the distribution for the maximum non-chosen utility in different ways. Horowitz, Sparmonn and Daganzo (1984) show that the Clark algorithm can provide a poor approximation under various circumstances. With enough draws, the proposed procedure is necessarily more accurate.

III. Simulating Random Parameters and Error Components

In many situations, the correlation in the unobserved portion of utility arises from random parameters, which Hausman and Wise (1978) give as a motivation for probit over logit. The correlations might also reflect error components that are common to all alternatives in a group, a specification that provides a probit analog to nested logit (McFadden, 1978). In these situations, probit probabilities can be simulated in a straightforward and computationally attractive way. The procedure has been recognized for some time but does not seem to have been written down.[4] Given the multitude of empirical situations that can be meaningfully represented by random-parameter/error-component probits, and the simplicity of the simulator, an explicit description seems warranted.

---

[4] For example, Buldoc (1992) implicitly utilizes the concepts in his presentation of probits with a more complex error structure, called a first-order generalized autoregressive (GAR(1)) error process.

Let there be J alternatives with utility vector $U=W\alpha+D\mu+\varepsilon$ where α is a normally distributed vector of L random parameters associated with variables W; D is a matrix composed of Q indicator variables whose element $d_{jq}=1$ if alternative j is a member of group q; μ is a vector of length Q with zero mean whose element μ_q is the common unobserved term that enters the utility for all alternatives in group q; and ε is a vector of J independent normal deviates with possibly unequal variance representing the unobserved factors that are specific to each alternative separately.[5] Correlations across alternatives in the unobserved portion of utility are captured in (a) the common effects μ for alternatives within a group and (b) the correlations in explanatory variables across alternatives, which enter the unobserved portion of utility because the coefficients α are random.

The error components are actually random coefficients of dummy variables that identify groups of alternatives, with the means of the coefficients constrained to zero. The utility vector can therefore be expressed more succinctly as $U=X\beta + \varepsilon$, where X is a matrix of K explanatory variables for J alternatives, including dummies for groups of alternatives, such that $K=L+Q$; and β is distributed normal with mean b (with some elements perhaps constrained to zero) and covariance Λ. The proposed procedure is to simulate β (which is the sources of correlation among unobserved utility) as well as the element of ε for the chosen alternative, and then explicitly integrate over the remaining elements of ε. Since these elements of ε are independent normals, this remaining integration is computationally simple. The simulated probabilities are strictly positive and smooth (continuous and twice-differentiable) in underlying parameters and variables.

Let M be a Choleski factor of Λ. (If Λ is diagonal, representing no covariances among random parameters and error components, then the elements of M are the square root of the elements of Λ; of course, even with diagonal Λ, the unobserved portion of utility is correlated over

[5] If ε has a non-diagonal covariance matrix, then the simulator in this section can be combined with that in the previous section. However, usually the purpose of specifying random parameters and error components is to capture the correlation among the unobserved portion of utility structurally through these terms.

alternatives). Define T as the diagonal matrix whose elements are the standard deviations of the elements of ε . The utility vector can be expressed as:

$$U = Xb + XM\eta + T\mu$$

where η and μ are vectors of iid standard normal deviates, with lengths K and J, respectively. Subscripts are used to denote rows of these arrays, such that the utility of alternative j is $U_j = X_jb + X_jM\eta + t_j\mu_j$, where t_j is the standard deviation of ε_j.

The probability that alternative i is chosen is:

$$P_i = \text{Prob}(U_i > U_j \ \forall j \neq i)$$

$$= \int\int 1((X_ib + X_iM\eta + t_i\mu_i > X_jb + X_jM\eta + t_j\mu_j) \ \forall j \neq i \ ) \ f(\mu) \ f(\eta) \ d\mu \ d\eta$$

$$= \int\int \text{Prob}[\ t_j\mu_j < (X_ib + X_iM\eta + t_i\mu_i \ ) - (X_jb + X_jM\eta) \ \forall j \neq i \mid \eta, \ \mu_i] \ f(\mu_i) \ f(\eta) \ d\mu_i \ d\eta$$

$$= \int\int \Pi\Phi(ARG_{i\text{-}j}) \ f(\mu_i) \ f(\eta) \ d\mu_i \ d\eta$$

where $ARG_{i\text{-}j} = ((X_ib + X_iM\eta + t_i\mu_i) - (X_jb + X_jM\eta))/ \ t_j$ and the product Π is over all $j \neq i$. The partitioning of errors occurs in the third equality: the probability Prob[.] is the explicit integral over μ_j for all $j \neq i$, conditional on the values of η and μ_i. As given in the last line, this integral is easy to calculate and, since η and μ_i are independent standard normal, drawing from their distributions is easy.

This expression for P_i is simulated as follows. (1) Draw a value of η and μ_i. That is, draw K+1 independent standard normal deviates. (2) Calculate $\Phi(ARG_{i\text{-}j})$ for all $j \neq i$, and take their product. (3) Repeat this process for numerous draws and average the results. The simulated probability obtained in this way is, with as few as one draw, unbiased for the true probability, strictly positive, and continuous and twice-differentiable in the underlying parameters and variables.

Denote with a superscript r the value of ARG_{i-j} that is obtained with the r-th draw of η and μ_i, and define $SR_i^r = \Pi\Phi(ARG_{i-j}^r)$. Then the simulated probability is $SP_i = (1/R)\sum [\Pi\Phi(ARG_{i-j}^r)] = (1/R)\sum SR_i^r$. The derivatives of the simulated probability with respect to underlying parameters are:

$$\delta SP_i/\delta b = (1/R) \sum_r SR_i^r \sum_{j\neq i} [\phi(ARG_{i-j}^r) /\Phi(ARG_{i-j}^r)] [(X_i-X_j)/t_j]$$

$$\delta SP_i/\delta M = (1/R) \sum_r SR_i^r \sum_{j\neq i} [\phi(ARG_{i-j}^r) /\Phi(ARG_{i-j}^r)] [(X_i-X_j)/t_j] \eta^r$$

$$\delta SP_i/\delta t_j = - (1/R) \sum_r SR_i^r [\phi(ARG_{i-j}^r) /\Phi(ARG_{i-j}^r)] [(ARG_{i-j})/t_j] \quad \text{for } j\neq i$$

$$\delta SP_i/\delta t_i = (1/R) \sum_r SR_i^r \sum_{j\neq i} [\phi(ARG_{i-j}^r) /\Phi(ARG_{i-j}^r)] [\mu_i^r/t_j] .$$

To aid numerical search, the second derivatives can also be calculated.

This simulator is essentially an example of Stern's (1987) decomposition simulator that utilizes the special structure of a random-parameters/error-components model. Stern suggests, for utility vector U distributed normal with covariance matrix Ω, to find a λ such that U can be decomposed into Y+W where Y is normal with covariance $\lambda^2 I$ and W is normal with covariance $\Omega-\lambda^2 I$. That is, the utility vector is decomposed into a part that carries the correlation and a part that is uncorrelated over alternatives. Given this decomposition, the values of W are simulated through repeated draws from its distribution, and the integration over Y given W is performed explicitly. This integration is easy to perform numerically since the elements of Y are independent. For general Ω, the Stern decomposition can be difficult, since finding a λ that provides a positive definite $\Omega-\lambda^2 I$ is computationally burdensome, and the task must be performed for each trial value of the parameters. As pointed out by Hajivassiliou, McFadden, and Ruud (1992), the accuracy of the general Stern simulator fails when the number of alternatives is large and the eigenvalues of Ω are uneven. This difficulty is most likely the reason that the Stern simulator performed poorly in the Monte Carlo experiments of these authors (since, given λ, the simulator is very fast to compute.) However, when Ω takes the form implied by a random-

parameters/error-components model, calculation of λ is not needed: the structure of the model provides a natural decomposition. The computationally difficult part of the Stern simulator is avoided while the extremely fast aspects of it are retained. Since many empirical situations can be adequately represented by random parameters and error components, the numerical and conceptual ease of the approach in these situations is particularly appealing.

The simulator allows straightforward generalization to other distributions for the error terms. The parameters β can have any distribution that allows β to be expressed as $b + M\eta$ with b and M being parameters and η being a vector of random variables whose distribution (1) does not depend on unknown parameters and (2) is easy to draw from. Multivariate normal β can obviously be expressed in this way, with the elements of η being standard normal deviates. Other possibilities for η are gamma, which is useful when the parameters are known to take only one sign, and uniform, which is useful primarily because of simplicity. The independent random elements, ε, can have other distributions that allow for simple integration (either numerical or analytic) conditional on β. A prime candidate is iid extreme value, for which simulation is even easier than with iid normal ε. Conditional on η, the probability that $U_i > U_j \ \forall\ j \neq i$ is the logit formula, such that:

$$P_i = \int [\exp(V_i)/\sum\exp(V_j)]\ f(\eta)\ d\eta$$

where $V_j = X_j b + X_j M\eta$. Simulation is performed by drawing values of η, calculating the logit formula in brackets for each draw, and averaging the results. This simulation can be expected to be faster than with normal ε because: (1) the i-th element does not need to be simulated, and (2) the logit formula is easier to calculate than the product of cumulative normals.

Non-normal distributions for the random parameters and error components do not change the derivatives, since their distribution is captured by drawing η from a parameter-free distribution and constructing β the same as when η is normal. (That is, the distribution of η affects the values of the simulated probability but not its formula conditional on η.) The derivatives are different, of course, when the distribution of ε is changed.

The simulator is applicable to panel data when the dependence over time in the unobserved portion of utility is due to parameters and/or error components being constant over time for each agent (time-dependence due to customer heterogeneity; Heckman, 1981.) Let subscript n denote the agent and t denote the time period. The utility vector is specified as: $U_{nt} = X_{nt}\beta_n + \varepsilon_{nt} = X_{nt}b + X_{nt}M\eta_n + T\mu_{nt}$. That is, the variables vary over time and agents; the parameters associated with the variables and the error components that are common to groups of parameters (which represent the agent's "taste" for that group) vary over agents but are constant over time for each agent; finally, the alternative-specific errors vary independently over time and agents. Let i(nt) be the chosen alternative of agent n in time t, and let P_n^* be the probability of the sequence of agent n's choices. Since η is constant over time for each agent, the choices in each period conditional on η depend only on the value of μ, which is independent over time and agents. We have

$$P_n^* = \int\int \Pi_t \left[\Pi_{j \neq i(nt)} \Phi(ARG_{nt,i(nt)-j}) \right] f(\mu_{i(nt)})\, f(\eta)\, d\eta\, d\mu_{i(nt)}$$

where $ARG_{nt,i(nt)-j} = ((X_{nt,i(nt)}b + X_{nt,i(nt)}M\eta + t_{i(nt)}\mu_{i(nt)}) - (X_{nt,j}b + X_{nt,j}M\eta))/ t_j$. The probability is simulated for each agent as follows. (1) Draw a value of η and one $\mu_{i(nt)}$ for each time period. (2) Calculate $\Phi(ARG_{nt,i(nt)-j})$ for each time period, using the agent's drawn value of $\mu_{i(nt)}$ for that period and the value of η for all the time periods. (3) Take the product of these cumulative normals over all non-chosen alternatives in all time periods. (4) Repeat the process for numerous draws and average the results.

References

Bolduc, D., 1992, "Generalized Autoregressive Errors in the Multinomial Probit Model," Transportation Research, Vol. 26B, No. 2, pp. 155-170.

Borsch-Supan, A., and V. Hajivassiliou, 1993, "Smooth Unbiased Multivariate Probability Simulators for Maximum Likelihood Estimation of Limited Dependent Variable Models," Journal of Econometrics, Vol. 58, pp. 347-368.

Clark, C., 1961, "The Greatest of a Finite Set of Random Variables," Operations Research, Vol. 9, pp. 145-162.

Daganzo, C., F. Bouthelier, and Y. Sheffi, 1977, "Multinomial Probit and Qualitative Choice: A Computationally Efficient Algorithm," Transportation Science, Vol. 11, pp. 338-358.

Hajivassiliou, V., and D. McFadden, 1992, "The Method of Simulated Scores for the Estimation of LDV Models," forthcoming, Econometrica.

Hajivassiliou, V., D. McFadden, and P. Ruud, 1992, "Simulation of Multivariate Normal Rectangle Probabilities and their Derivatives: Theoretical and Computational Results," forthcoming, Journal of Econometrics.

Hausman, J., and D. Wise, 1978, "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences," Econometrica, Vol. 48, No. 2, March, pp. 403-426.

Heckman, J., 1981, "Statistical Models for Discrete Panel Data," in C. Manski and D. McFadden, eds., Structural Analysis of Discrete Data with Econometric Applications, MIT Press.

Horowitz, J., J. Sparmonn, and C. Daganzo, 1984, "An Investigation of the Accuracy of the Clark Approximation for the Multinomial Probit Model," Transportation Science, Vol. 16, pp. 383-401.

Lerman, S. and C. Manski, 1981, "On the Use of Simulated Frequencies to Approximate Choice Probabilities," in C. Manski and D. McFadden, eds., Structural Analysis of Discrete Data with Econometric Applications, MIT Press.

McFadden, D., 1978, "Modelling the Choice of Residential Location," in A. Karquist, et al, eds., Spatial Interaction Theory and Planning Models, Amsterdam: North-Holland Publishing Company.

McFadden, D., 1989, "A Method of Simulated Moments for Estimation of the Multinomial Probit Without Numerical Integration," Econometrica, Vol. 57, pp. 995-1026.

Stern, S., 1992, "A Method for Smoothing Simulated Moments of Discrete Probabilities in Multinomial Probit Models," Econometrica, Vol. 60, pp. 943-952.

June 13, 1995

Working Paper Series
Department of Economics
University of California, Berkeley

Individual copies are available for $3.50 within the USA and Canada; $6.00 for Europe and South America; and $7.00 for all other areas. Papers may be obtained from the Institute of Business and Economic Research: send requests to IBER, 156 Barrows Hall, University of California, Berkeley CA 94720-1922. Prepayment is required. Make checks or money orders payable to "The Regents of the University of California."

92-203 "The Impact of 1989 California Major Anti-Smoking Legislation on Cigarette Consumption: Three Years Later." Teh-wei Hu, Jushan Bai, Theodore E. Keeler, Paul G. Barnett. September 1992.

92-204 "A Dynamic Simultaneous-Equations Model for Cigarette Consumption in the Western States." Hai-Yen Sung, Teh-Wei Hu, and Theodore E. Keeler. November 1992.

92-205 "Maastricht: Prospect and Retrospect." John M. Letiche. December 1992.

92-206 "Nonparametric Methods to Measure Efficiency: A Comparison of Methods." Richard F. Garbaccio, Benjamin E. Hermalin, Nancy E. Wallace. December 1992.

93-207 "The Value of Intangible Corporate Assets: An Empirical Study of the Components of Tobin's Q." Bronwyn H. Hall. January 1993.

93-208 "R&D Tax Policy During the 1980s: Success or Failure?" Bronwyn H. Hall. January 1993.

93-209 "Hospital Costs and Excess Bed Capacity: A Statistical Analysis." Theodore E. Keeler and John S. Ying. April 1993.

93-210 "Homothetic Preferences, Homothetic Transformations, and the Law of Demand in Exchange Economies." John K.-H. Quah. April 1993.

93-211 "Deviations, Dynamics and Equilibrium Refinements." Matthew Rabin and Joel Sobel. May 1993.

93-212 "Loss Aversion in a Savings Model." David Bowman, Debby Minehart, and Matthew Rabin. May 1993.

93-213 "Nonparametric Multivariate Regression Subject to Constraint." Steven M. Goldman and Paul A. Ruud. May 1993.

93-214 "Asset Prices and the Fundamentals: A Q Test." Roger Craine. May 1993.

93-215 "A Note on the Finiteness of the Set of Equilibria in an Exchange Economy with Constrained Endowments." Deborah Minehart. September 1993.

93-216 "Institutional Prerequisites for Economic Growth: Europe After World War II." Barry Eichengreen. September 1993.

93-217 "Industrial Research During the 1980s: Did the Rate of Return Fall?" Bronwyn H. Hall. September 1993.

93-218 "Efficient Standards of Due Care: Should Courts Find More Parties Negligent Under Comparative Negligence?" Aaron S. Edlin. October 1993.

93-219 "Classical Estimation Methods for LDV Models Using Simulation." Vassilis A. Hajivassiliou and Paul A. Ruud. October 1993.

93-220 "Perspectives on the Borchardt Debate." Barry Eichengreen. November 1993.

93-221 "Explicit Tests of Contingent Claims Models of Mortgage Defaults." John M. Quigley. November 1993.

93-222 "Rivalrous Benefit Taxation: The Independent Viability of Separate Agencies or Firms." Aaron S. Edlin and Mario Epelbaum. December 1993.

94-223 "Convergence of the Aumann-Davis-Maschler and Geanakoplos Bargaining Sets." Robert M. Anderson. January 1994.

94-224 "Nonconvergence of the Mas-Colell and Zhou Bargaining Sets." Robert M. Anderson and Walter Trockel and Lin Zhou. January 1994.

94-225 "Tobacco Taxes and the Anti-Smoking Media Campaign: The California Experience." Teh-wei Hu, Hai-yen Sung and Theodore E. Keeler. January 1994.

94-226 "Walrasian Comparative Studies." Donald J. Brown and Rosa L. Matzkin. January 1994.

94-227 "Corporate Diversification and Agency." Benjamin E. Hermalin and Michael L. Katz. February 1994.

94-228 "Density Weighted Linear Least Squares." Whitney K. Newey and Paul A. Ruud. August 1994.

94-229 "Unemployment and the Structure of Labor Markets: The Long View." Barry Eichengreen. October 1994.

94-230 "Financing Infrastructure in Developing Countries: Lessons from the Railway Age." Barry Eichengreen. October 1994.

94-231 "Debt Deflation and Financial Instability: Two Historical Explorations." Barry Eichengreen and Richard S. Grossman. October 1994.

94-232 "Increasing Returns in Infinite Horizon Economies." Chris Shannon. October 1994.

94-233 "Determinacy in Infinite Horizon Exchange Economies." Chris Shannon. December 1994.

95-234 "Fairly Priced Deposit Insurance and Bank Charter Policy." Roger Craine. May 1995.

95-235 "Edgeworth's Conjecture with Infinitely Many Commodities." Robert M. Anderson and William R. Zame. May 1995.

95-236 "Regulation, Competition and Cross-subsidization in Hospital Care: Lessons from the Economics of Regulation." Stephen Earl Foreman and Theodore E. Keeler. May 1995.

95-237 "Simulation Methods for Probit and Related Models Based on Convenient Error Partitioning." Kenneth E. Train. June 1995.