

# Linear Regression Estimation of Discrete Choice Models with Nonparametric Distributions of Random Coefficients

By PATRICK BAJARI, JEREMY T. FOX, AND STEPHEN P. RYAN\*

In a discrete choice demand model, consumer  $i$  chooses product  $j$  out of the set  $J$  if  $u_{i,j} > u_{i,k} \forall k = 1, \dots, J, k \neq j$ . A standard setup has a vector of  $D$  characteristics  $\mathbf{x}_{i,j}$  for product  $j$ , and utility is parametrized as  $u_{i,j} = \mathbf{x}'_{i,j}\boldsymbol{\beta} + \varepsilon_{i,j}$ , where  $\boldsymbol{\beta}$  is a vector of parameters to estimate that reflect the marginal utility of the product characteristics, and  $\varepsilon_{i,j}$  is a product-specific error term. Typically  $\varepsilon_{i,j}$  is assumed to have a logit or normal marginal distribution.

In industrial organization, marketing, and transportation economics, hundreds of papers use random coefficient models to estimate both individual and aggregate demand. For some examples, see Hayden J. Boyd and Robert E. Mellman (1980); N. Scott Cardell and Frederick C. Dunbar (1980); Dean A. Follmann and Diane Lambert (1989); Pradeep K. Chintagunta, Dipak C. Jain, and Naufel J. Vilcassim (1991); Steven Berry, James Levinsohn, and Ariel Pakes (1995); Aviv Nevo (2001); Amil Petrin (2002); and Kenneth E. Train (2003).

Random coefficients generalize the model so that  $u_{i,j} = \mathbf{x}'_{i,j}\boldsymbol{\beta}_i + \varepsilon_{i,j}$ , where the  $D$ -vector  $\boldsymbol{\beta}_i$  is specific to consumer  $i$ . Adding random coefficients allows consumers to substitute (as prices in  $\mathbf{x}_{i,j}$  change) between products with similar nonprice observables in  $\mathbf{x}_{i,j}$ . Daniel McFadden and Train (2000) show the more general mixed logit can flexibly approximate choice patterns.

Random coefficient estimators are difficult to compute. The likelihood for data on the choices  $j_i$  of consumers  $i = 1, \dots, N$  is

$$L(\gamma) = \prod_{i=1}^N \int_{\boldsymbol{\beta}} \frac{\exp(\mathbf{x}'_{i,j_i}\boldsymbol{\beta})}{\sum_{k=1}^J \exp(\mathbf{x}'_{i,k}\boldsymbol{\beta})} f(\boldsymbol{\beta} | \gamma) d\boldsymbol{\beta}.$$

\*Bajari: Department of Economics, University of Minnesota, Twin Cities, 1035 Heller Hall, 271 19th Ave. South, Minneapolis, MN 55455, and National Bureau of Economic Research (e-mail: bajari@econ.umn.edu); Fox: Department of Economics, University of Chicago, 1126 E. 59th St., Chicago, IL 60637 (e-mail: fox@uchicago.edu); Ryan: Massachusetts Institute of Technology, 50 Memorial Drive, E52-262C, Cambridge, MA 02142, and NBER (e-mail: sryan@mit.edu).

The parametric density  $f(\boldsymbol{\beta} | \gamma)$  reflects the distribution of the unobserved heterogeneity, the tastes  $\boldsymbol{\beta}_i$ . The goal is to estimate the parameters  $\gamma$  in this density.

Estimation usually proceeds by simulation: maximum likelihood or the method of moments. The consumer  $i$ -specific numerical integral is of dimension  $D$ . The likelihood must be evaluated repeatedly at trial guesses of  $\gamma$ . The nonlinear search over  $\gamma$  can suffer from multiple local maxes, resulting in the need to try many starting values. The dimension of  $\gamma$  can be large— $\gamma$  often contains variance matrices of multivariate normal distributions.

Hierarchical Bayesian estimation is an alternative (Peter E. Rossi, Greg M. Allenby, and Robert McCulloch 2005). Computationally efficient Gibbs sampling requires training in conjugate, family relationships that are needed for efficient random number generation. These conjugate families require restrictive model assumptions that can break down with small model perturbations. Gibbs sampling itself requires training and monitoring by the user.

This paper describes a method for estimating random coefficient discrete choice models that is both flexible and simple to compute. We demonstrate that, with a finite number of types, choice probabilities are a linear function of the model parameters. Because of this linearity, our model can be estimated using linear regression, subject to inequality constraints. We can approximate an arbitrary distribution of random coefficients by allowing the number of types to be sufficiently large. Therefore, we say our estimator is nonparametric for the distribution of heterogeneity.

## I. Review of Series Estimators

Let  $y_{j,t}$  be the market share of product  $j$  in market  $t$ ,  $\mathbf{x}_t$  the characteristics of all  $J$  products in market  $t$ , and  $\eta_{j,t}$  measurement error in market shares. Let  $y_{j,t} = g_j(\mathbf{x}_t) + \eta_{j,t}$ . A series estimator approximates an unknown function  $g_j(\mathbf{x}_t)$  with

the approximation  $g_j(\mathbf{x}) \approx \sum_{r=1}^R h_r(\mathbf{x})\theta_r$ . Here,  $\{h_r(\mathbf{x})\}_{r=1}^R$  is a known basis of  $R$  functions chosen to ease mathematical approximation, and  $\theta_r$  is an approximation weight on the function  $r$ .

The key behind series estimation is that the unknown parameters  $\theta_r$  enter the market share approximation linearly. Estimation just regresses  $y_{j,t}$  on  $\{h_r(\mathbf{x}_t)\}_{r=1}^R$  for various markets  $\mathbf{x}_t$ . Donald W. K. Andrews (1991) shows that estimators of  $g_j(\mathbf{x})$  for a given  $\mathbf{x}$  and some functions of  $g_j(\mathbf{x})$  are asymptotically normal.

### II. Estimating Random Coefficient Logit Models

We now show how linear regression can estimate the random coefficients logit model for market share data. Assume, for now, that there really are  $r$  known, discrete consumer types. Each type  $r$  is distinguished by a known, random coefficient vector  $\beta^r$ . Let  $\theta^r$  be the fraction of consumers of type  $r$  in the population.

Let  $P(j|t)$  be the no-measurement error market share of product  $j$  in market  $t$ . Market shares are the sum of the individual choice probabilities of each type in the marketplace,

$$P(j|t) = \sum_{r=1}^R \left( \frac{\exp(\beta^r \mathbf{x}_{j,t})}{\sum_{k=1}^J \exp(\beta^r \mathbf{x}_{k,t})} \right) \theta^r.$$

Type  $r$ 's logit choice probabilities are weighted by its frequency  $\theta^r$ . The basis functions are not the flexible mathematical functions from traditional series estimators, but the predictions of an individual choice model for consumer type  $r$ . No unknown parameters enter the logit choice probabilities. Each  $\beta^r$  represents all the utility parameters for type  $r$ . The unknown frequencies  $\theta^r$  are structural objects, not just the approximation weights from series estimation.

The key idea is that the type frequencies  $\theta^r$  enter the market shares linearly and can be estimated from a linear regression of shares on logit choice probabilities for all types.

With actual data  $y_{j,t}$  on market shares, we estimate the regression equation

$$y_{j,t} = \sum_{r=1}^R \left( \frac{\exp(\beta^r \mathbf{x}_{j,t})}{\sum_{k=1}^J \exp(\beta^r \mathbf{x}_{k,t})} \right) \theta^r + (y_{j,t} - P(j|t))$$

to estimate the  $R$   $\theta^r$ 's. Let  $T$  be the number of markets. There is one regression observation for

each product and each product or  $T \cdot J$  regression observations. The number of unknown parameters is the number of types,  $R$ . The term  $(y_{j,t} - P(j|t))$  reflects the measurement error in market shares.

Once the  $\theta^r$ 's are estimated, we can predict out-of-sample market shares by varying the product covariates  $\mathbf{x}_{j,t}$  in the logit choice probabilities that enter the equation for  $P(j|t)$ .

Typically the  $R$  random coefficient vectors  $\beta^r$  are unknown. We view our estimator for the  $\theta^r$ 's as a nonparametric approximation to an underlying, possibly continuous density of random coefficients. So before running the regression, we first draw or deterministically choose  $R$  random coefficients  $\beta^r$ . The estimator is not particularly sensitive to the scheme used to pick the  $\beta^r$ 's, as the  $\theta^r$ 's are completely flexible parameters to be estimated. We do require that we span the domain of the underlying true, random coefficient distribution in the limit as  $R$  grows.

There is no way to impose that the types have some restrictive parametric distribution. Given that we know of no empirical applications where researchers would actually know the types have some distribution, we see no reason why a researcher would not want to be nonparametric on the weights over the  $R$  types.

$R$  does not have to be too large. For two dimensions of  $\beta$  ( $D = 2$ ),  $R = 200$  works well.

One option imposes the constraints  $\sum_{r=1}^R \theta^r = 1$  and  $\theta^r \geq 0$  for  $r = 1, \dots, R$ . If so, the closed-form regression becomes an inequality constrained least squares (ICLS) estimator, which requires numerical optimization. The ICLS minimization problem is convex, so a standard least squares algorithm will find a global optimum. The minimization problem is much simpler than minimizing over both the  $R$   $\beta^r$ 's and the  $R$   $\theta^r$ 's (James J. Heckman and Burton H. Singer 1984).

Our approach shares the intuition of spanning the space of economic models with the importance sampling estimator of Daniel A. Akerberg (2001). His importance sampling estimator requires parametric type densities, a change of variables assumption, and numerical optimization, however.

### III. Extensions

A type  $r$  could involve values for the choice-specific errors of the form  $\varepsilon_j^r$  in addition to

random coefficients  $\beta^r$ . Shares are still the sum of decisions of  $R$  types of consumers,

$$y_{j,t} \approx \sum_{r=1}^R 1\{\text{type } r \text{ buys } j | \mathbf{x}_t\} \theta_r,$$

where the indicator  $1\{\text{type } r \text{ buys } j | \mathbf{x}_t\}$  says that consumer type  $r$  would buy product  $j$  when faced with the choice characteristics  $\mathbf{x}_t$ . The symbol “ $\approx$ ” abstracts away from approximation and measurement errors.

If the data are individual, choice  $j_i$  for consumer  $i$ , then the regression becomes a linear probability model. Each consumer has  $J$  observations of the form, for observation  $j$ ,

$$1\{j = j_i\} \approx \sum_{r=1}^R 1\{\text{type } r \text{ buys } j | \{\mathbf{x}_{i,k}\}_{k=1}^J\} \theta_r,$$

where the dependent variable is 1 if consumer  $i$  bought  $j$ , and 0 otherwise. If the data are individual panels of strings of choices of the form  $j_{i,1}, j_{i,2}, j_{i,3}, \dots, j_{i,T}$ , then the linear probability model becomes

$$1\{j^T = (j_{i,1}, j_{i,2}, j_{i,3}, \dots, j_{i,T})\} \approx$$

$$\sum_{r=1}^R 1\{\text{type } r \text{ chooses string } j^T | \{\{\mathbf{x}'_{i,k}\}_{k=1}^J\}_{t=1}^T\} \theta_r.$$

With  $T$  periods of data,  $i$  has  $J^T$  regression observations. The computational burden of linear regression is the number of parameters  $R$ . Regressions can have millions of observations.

The estimation of a forward-looking dynamic programming model is similar. Estimate the regression equation

$$1\{j^T = (j_{i,1}, j_{i,2}, j_{i,3}, \dots, j_{i,T})\} \approx$$

$$\sum_{r=1}^R L(\text{type } r \text{ chooses string } j^T | \beta^r, \{\{\mathbf{x}'_{i,k}\}_{k=1}^J\}_{t=1}^T) \theta_r,$$

where the likelihood  $L$  for consumer  $i$  with preferences  $\beta^r$  requires solving a dynamic programming problem and integrating over unobserved actions and states, such as a consumer’s unobserved inventory of a storable good. The dynamic programming problem must be solved  $R$  times before the regression, but only  $R$  times. Under the assumptions in John Rust (1987),  $L$

has a closed form once the choice-specific value functions are computed.

#### IV. Monte Carlo

The following Monte Carlo explores the performance of three estimators on three different random coefficient logit fake data designs. Each choice set has ten choices and an outside option with utility 0.  $D = 2$  and each of the two  $\mathbf{x}_j$  components are generated by exponentiating uniform draws from  $[0, 3]$ .

In the first design, the tastes for the two characteristics are independent:  $\beta_1 \sim N(0, 1)$  and  $\beta_2 \sim N(1, 2)$ . In the second design, the tastes keep the same marginals and add a negative covariance of  $-0.9$ . In the third design, the taste parameters are drawn from a mixture of multivariate normals:

$$0.7 \cdot N\left(\begin{bmatrix} 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 0.1 & -0.1 \\ -0.1 & 0.5 \end{bmatrix}\right) + 0.3 \cdot N\left(\begin{bmatrix} 0 \\ 3 \end{bmatrix}, \begin{bmatrix} 0.3 & 0.1 \\ 0.1 & 0.3 \end{bmatrix}\right).$$

This bimodal, two-dimensional distribution of tastes is plotted in Figure 1.

Our estimator uses Matlab’s ICLS minimizer. To obtain the  $\beta^r$ ’s, we drew  $R$  different coefficients. Each coefficient is independent normal, with the mean the estimate from the standard logit, and the variance 3. We set  $R = n/5$ .

We estimate each of the three models on our fake data. Figure 2 is our estimate of Figure 1 for a case of  $n = 1,000$  and  $R = 200$ . Only 17 points have positive mass, and those capture the mass points in the underlying continuous density.

Table 1 presents the results for the root mean squared prediction error (RMSE) for a market share prediction exercise where we sample new  $\mathbf{x}$  characteristics for all the products. The prediction exercise tests the structural use of demand models. An RMSE of 0.01 corresponds to true market shares of 10 percent for the ten products and prediction errors of 1 percent in each.

In the first design, the RMSE is low and decreases with the sample size for the two consistent estimators: regression and the random coefficients (RC) logit. The pure logit is inconsistent. In the second design, only our estimator

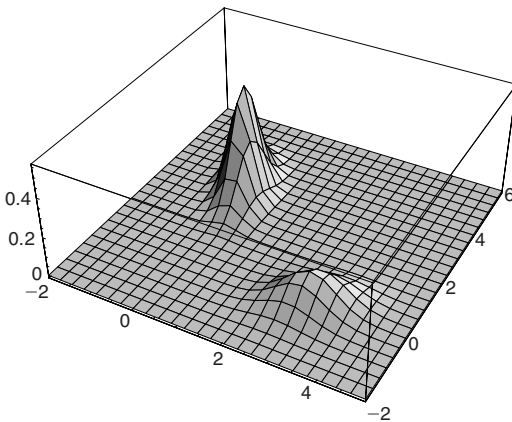


FIGURE 1. TRUE RANDOM COEFFICIENT DENSITY IN THE THIRD EXPERIMENT

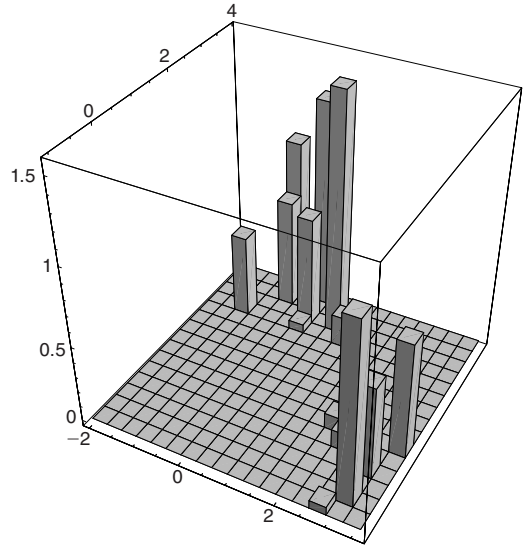


FIGURE 2. ESTIMATED RANDOM COEFFICIENT DENSITY IN THE THIRD EXPERIMENT

TABLE 1—MONTE CARLO RMSEs FOR NEW MARKET SHARE PREDICTIONS

Design	<i>n</i>	Logit	Logit+RC	OLS Logit
Indep. RC	500	0.080	0.015	0.015
	1000	0.081	0.010	0.010
	2000	0.081	0.008	0.008
Corr. RC	500	0.095	0.015	0.015
	1000	0.103	0.011	0.011
	2000	0.099	0.008	0.008
Mixed Corr. RC	500	0.121	0.016	0.016
	1000	0.131	0.012	0.012
	2000	0.127	0.008	0.008

is likely consistent because the RC logit assumes independent random coefficients. The prediction error is low and decreases with the sample size only for our estimator. In the third design, the mixed multivariate normal in Figure 1 is a strong test of the nonparametric ability of our estimator. Our estimator performs well. The RMSE is low and decreases with the sample size. Recall that we use at most  $R = 200$ .

Overall, our estimator has much better RMSE than the inconsistent estimators, and the loss in prediction RMSE is low in the first experiment, where the RC logit is efficient.

V. Conclusion

Random coefficients models have been thought to be computational demanding. We

show that this is not the case, by introducing a linear regression estimator that is nonparametric on the density of random coefficients. We discuss extensions to panel data and forward-looking dynamic programming models.

REFERENCES

Ackerberg, Daniel A. 2001. "A New Use of Importance Sampling to Reduce Computational Burden in Simulation Estimation." Unpublished.

▶ Andrews, Donald W. K. 1991. "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models." *Econometrica*, 59(2): 307–45.

▶ Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica*, 63(4): 841–90.

▶ Boyd, J. Hayden, and Robert E. Mellman. 1980. "The Effect of Fuel Economy Standards on the U.S. Automobile Market: An Hedonic Demand Analysis." *Transportation Research Part A*, 14: 367–378.

▶ Cardell, N. Scott, and Frederick C. Dunbar. 1980. "Measuring the Societal Impacts of Automobile Downsizing." *Transportation Research Part A*, 14: 423–34.

- ▶ **Chintagunta, Pradeep K., Dipak C. Jain, and Naufel J. Vilcassim.** 1991. "Investigating Heterogeneity in Brand Preferences in Logit Models for Panel Data." *Journal of Marketing Research*, 28(4): 417-28.
- ▶ **Follmann, Dean A., and Diane Lambert.** 1989. "Generalizing Logistic Regression by Non-parametric Mixing." *Journal of the American Statistical Association*, 84(405): 295-300.
- ▶ **Heckman, James J., and Burton H. Singer.** 1984. "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Methods for Duration Data." *Econometrica*, 52(2): 271-320.
- ▶ **McFadden, Daniel, and Kenneth Train.** 2000. "Mixed MNL Models for Discrete Response." *Journal of Applied Econometrics*, 15(5): 447-70.
- ▶ **Nevo, Aviv.** 2001. "Measuring Market Power in the Ready-to-Eat Cereal Industry." *Econometrica*, 69(2): 307-42.
- ▶ **Petrin, Amil.** 2002. "Quantifying the Benefits of New Products: The Case of the Minivan." *Journal of Political Economy*, 110(4): 705-29.
- ▶ **Rossi, Peter E., Greg M. Allenby, and Robert McCulloch.** 2005. *Bayesian Statistics and Marketing*. West Sussex: John Wiley & Sons.
- ▶ **Rust, John.** 1987. "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher." *Econometrica*, 55(5): 999-1033.
- ▶ **Train, Kenneth E.** 2003. *Discrete Choice Methods with Simulation*. Cambridge: Cambridge University Press.