An alterantive approach to ML estimation of multinomial choice models

Alexey V. Zakharov

May 15, 2009

The principal assumption of a multinomial choice model is that there are a number of individuals whose properties are known; I also observe the choice of each individual, and the set of alternatives available to him. I can then construct a utility function that is associated with each of the outcomes, and estimate its parameters — usually through the maximim likelihood method (although sometimes different methods, such as Bayesean estimation, are used). This methodology ignores additional information that is sometimes available in the problem context.

Consider, for example, a problem of modeling a voter's choice in an election. On the input, one usually has data from a pre-election survey, where each respondent indicates his socio-economic characteristics, his policy preferences on a set of issues, and the political party (or candidate) he intends to vote for. In a "spatial" voting model it is assumed that the voter's utility toward a party is a function of the distances between the voter's preferences on each issue, and the party's stated positions on those issues (Poole and Rosenthal, 1984; Schofield and Sened, 2006). All previous research treated party positions as exogenous and arbitrary. However, from the problem context I know that each party cares about the share of vote that it receives. I know that the observed policy positions of each party is rational, conditional on the party's utility function, the set of alternative policy positions available to the party, and the party's knowledge of the voters' decision model. That information is ignored in the traditional multinomial choice models, since the set of alternatives available to each party is not observable.

This short paper proposes an approach to incorporate this additional information in the likelihood function. In the first section, I postulate the assumptions used in this approach, and define the modified likelihood function. In the second section, I use the approach to re-evaluate a voting model estimated in Schofield (2007).

1 The multinomial choice model and game definition.

I conside a problem of estimating a model of individual choice. I have a dataset with $i = 1, \ldots, N$ observations, each corresponding to an individual. For each observation I have a vector of personal characteristics $x_i \in \mathbf{R}^{M_1}$, and a choice variable $d_i \in \{1, \ldots, J\}$. I assume that the utility of individual i choosing an alterantive j is

$$u_{ij} = u(x_i, \alpha_j, \beta, j) + \epsilon_{ij} = \bar{u}_{ij} + \epsilon_{ij}, \tag{1}$$

where $\alpha_j \in \mathbf{R}^{M_2}$ is a vector of choice-specific parameters, and $\beta \in \mathbf{R}^{M_3}$ is a vector of choice-independent parameters. I make some assumptions about the distribution of the random variables ϵ_{ij} — usually independence for different values of i. Let $d \in J^N$ denote the choices of all individuals, and $x \in \mathbf{R}^{M_1 N}$ the personal characteristics of all individuals. Our goal is to estimate the values of the parameters $\alpha = (\alpha_j) \in \mathbf{R}^{M_2 J}$, β given our observations (x, d).

A common way to solve this problem is through the maximum-likelihood method. For example, assume that ϵ_{ij} are distributed independently with a Type 1 extreme value distribution:

$$P(\epsilon_{ij} \leq h) = e^{-e^{-h}}. \tag{2}$$

Then, the likelihood of observation i would be

$$P_i = \frac{e^{\bar{u}_{id_i}}}{\sum_{k=1}^{J} e^{\bar{u}_{ik}}}, \tag{3}$$

and of the whole sample —

$$L(x, d, \alpha, \beta) = \prod_{i=1}^{N} P_i. \tag{4}$$

Maximizing L will give us the maximum-likelihood estimates of α and β. These are the parameter variables that maximize the probability that the observed outcome will arise in nature, given our assumptions about the form of the utility function, and the distribution of the error terms.[1]

I now introduce a new concept that will require a slight modification of the model.

Assumption 1 There exist K player agents. Each player agent k can choose some action y_k from a finite strategy set S_k before the individuals make their choices.

Put $S = \times S_k$. Let $y \in S$ denote an action profile for the player agents. For any k and $y \in S$, let y_{-k} be the actions of all player agents other than k.

Assumption 2 The are N *individual* (or *non-player*) agents. The payoff to an individual i choosing an alternative j depends on the actions of the player agents:

$$u_{ij} = u(x_i, \alpha_j, \beta, y, j) + \epsilon_{ij} = \bar{u}_{ij} + \epsilon_{ij} \tag{5}$$

Assumption 3 Every realization of d defines a payoff $U_k(d, y)$ to every player agent k, for every y. The player agents know the true values of the parameters (α, β) and x, but cannot observe ϵ_{ij}s.

The choices of the individuals d and hence the payoffs $U_k(d, y)$ are stochastic. However, I can define the expeted payoffs, based on what I know about the utilities of individual agents (5). Assuming that ϵ_{ij} are independent, the expected payoff of player agent k is

$$\bar{U}_k(x, \alpha, \beta, y) = \sum_{\delta \in J^N} \left(\prod_{i=1}^{N} p_{i\delta_i}(x_i, \alpha, \beta, y) \right) U_k(\delta, y). \tag{6}$$

where δ runs through all possible choice profiles, and $p_{i\delta_i}$ is the probability that individual i chooses alternative δ_i.

Now I can formulate the final assumption necessary in our approach.

[1] The maximum likelihood method has its well-known advantages and disadvantages. It is consistent and asymptotically unbiased; however, small-sample bias can be substatial, especially for those α_j where the number of individuals who have chosen alternative j is small. Comparing different models and estimating condifence intervals can also be a problem. An alternative approach is to treat the parametes as random variables, and estimate their joint distributions through a repeated Bayesian updating process.

Assumption 4 The observed actions y are a Nash equilibrium in a game with players $1, \ldots, K$, strategy sets S_k, and utilities

$$\tilde{U}_k = \bar{U}_k + \epsilon_k, \tag{7}$$

where ϵ_k are independent random variables. The values ϵ_k are known to all player agents, but not to the observer.

Speaking in the terms of this model, the traditional maximum likelihood method is about finding the parameters (α, β) that best explain the actions of non-player agents, given their characteristics x and observed choices d. A similar method can be used to find (α, β) that best explain the actions of player agents, conditional on the observed charateristics x of non-player agents, and the non-observed (but assumed) set of actions S available to the player agents.

Suppose that $y, y' \in S$ are two strategy profiles for the player agents, and I observe y. From Assumption 4 it follows that for every k, I have

$$\tilde{U}_k(x, \alpha, \beta, y) \geq \tilde{U}_k(x, \alpha, \beta, (y'_k, y_{-k})). \tag{8}$$

I want to find the estimates of α, β that best reflect the fact that every agent k chose action y_k, given the actions y_{-k} of all other players. In order to find those estimates given existing data, I need to specify what alternative actions did each agent k consider, before choosing y_k.

Consider two agent action profiles, y and some y'. Denote by

$$P_j(x, \alpha, \beta, y, y') = P(\tilde{U}_k(x, \alpha, \beta, y) \geq \tilde{U}_k(x, \alpha, \beta, (y'_k, y_{-k})) \tag{9}$$

the probability that agent k choses action y_k over action y'_k, given that all other agents choose y_{-k}.

Suppose that $S = \times S_k$ is a set of action profiles, with $y \in S$. Suppose that agent k knows that all other agents will choose y_{-k}. As ϵ_k are independent, the likelihood that he chooses action y_k from S_k is the probability that any pairwise comparison between y_k and any other action $y'_k \in S_k$ is in favor of y_k. The likelihood of observing $y \in S$ is then

$$L_P(x, \alpha, \beta, y, S) = \prod_{k=1}^{K} \prod_{y'_k \in S_k - \{y_k\}} P_j(x, \alpha, \beta, y, (y'_k, y_{-k})). \tag{10}$$

Now, I can give the definition of a new estimator for (α, β).

Definition 1 Let $0 < \gamma \leq 1$, and S be a set of alternatives for player agents. The *weighted Nash equilibrium maximum likelihood estimator* of (α, β) maximizes the weighted likelihood function

$$L = L_P(x, \alpha, \beta, y, S)^{\gamma} L(x, d, \alpha, \beta)^{1-\gamma}. \tag{11}$$

2 Example — a probabilistic voting model.

Suppose that individuals represent a representative sample of voters. Each individual is characterized by a vector of personal characteristics (such as age or income) $x_i \in \mathbf{R}^{M_1}$, and by value $v_i \in \mathbf{R}^{M_4}$ that reflects the individual's preferences with respect to the policies that will be carried out by the winning party in the election. Note that as $K = J$, I will use subscript j to index player agents. The choice variable d_i represents the index of the political party that the individual intends to vote for in the upcoming election. Let the utility functions of the individuals be given by

$$u_{ij} = a_j + \alpha_j^T x_i - \beta \|v_i - y_j\| + \epsilon_{ij} = \bar{u}_{ij} + \epsilon_{ij}, \tag{12}$$

where a_j is a party-specific constant, $\alpha_j \in \mathbf{R}^{M_1}$ is a party-specific vector of parameters, β is a parameter, $\|\cdot\|$ is the Eucledian norm, $y_j \in \mathbf{R}^{M_4}$ is the policy program of party j, and ϵ_{ij} is an independent random variable. Values x_i are usually the socio-economic characteristics of the voter (age, religion, etc). Assuming the distribution (2), I have the probability of individual i voting for party j given by

$$p_{ij} = \frac{e^{\bar{u}_{ij}}}{\sum_{h=1}^{J} e^{\bar{u}_{ih}}}. \tag{13}$$

If I assume that the payoff of a political party is equal to the expected number of votes that it will receive in the elections times a constant μ_j, I have

$$U_j(x, \alpha, \beta, y) = \mu_j \sum_{i=1}^{N} p_{ij}. \tag{14}$$

For each j, this value is a function of individual characteristics x, the parameters α, β, and the policy platforms y. I can define a game between the J parties, where the strategy of party j is $y_j \in \mathbf{R}^{M_4}$, and the payoff is (6).

If ϵ_k is distributed with distribution (2), then I must have

$$L_P(x, \alpha, \beta, y, S) = \prod_{k=1}^{K} \frac{e^{U_j(x,\alpha,\beta,y)}}{\sum_{y'_j \in S_j - \{y_j\}} e^{U_j(x,\alpha,\beta,(y'_j, y_{-j}))}}. \tag{15}$$

Let the weighted log likelihood function for this problem be

$$L = w_V L_V + w_P L_P, \tag{16}$$

where L_V is the likelihood of the observed voting profile, and w_V, w_P are weights.

Consider a dataset for 1996 Israel Knesset elections and a two-dimensional spatial model analyzed in Schofield (2007). To simplify exposition, I take $\alpha_j = 0$; this corresponds to the "valence-only" model, where the voter's utility toward a party does not depend on his socio-economic characteristics. I consider the two largest parties (Likud and Avoda) to be player agents. For each party, the strategy set has five elements: the observed policy position, and four deviations (plus or minus 1 on each dimension). I take $\mu_1 = \mu_2 = 5/N$. Let the weights be $w_V = (1 - \gamma)$ and $w_P = 600\gamma$. Hence, $\gamma = 0$ corresponds to the traditional maximum-likelihood approach; for $\gamma = 1$, only the likelihood of player agents (political parties in this case) is considered.

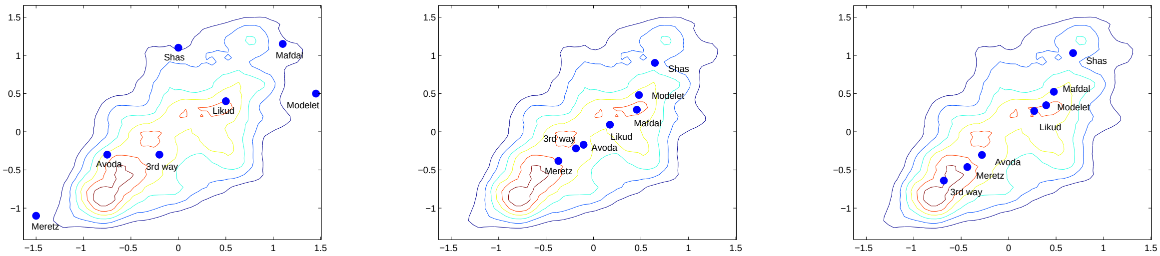
Table 1 shows the result of model estimation for different values of γ.

	$\gamma = 0$	$\gamma = 0.5$	$\gamma = 0.8$
Likud	0.7778	0.6135	2.3978
Labor	0.9901	0.6552	1.8475
Mafdal	-0.6270	-1.0018	-0.8998
Modelet	-1.2595	-0.8874	1.7995
Third Way	-2.2916	-2.4721	-0.3101
Shas	-2.0239	-2.5701	-3.0521
β	-1.2075	-1.9050	-3.6245
Log likelihood (voters)	-776.95	-823.0	-1,204.5
Log likelihood (parties)	-1,444.1	-1,338.0	-1,172.8

Table 1: Estimation of valence only model with one β parameter

In this model, a voter has seven choices, each corresponding to a political party; the first six numbers in each column correspond to the parameter a_j for each of the choice options (the seventh parameter is fixed at zero to identify the model). One can see that the more weight one puts on the likelihood of party policy positions, the higher is the estemated value of β, which reflects the importance of party policy positions in voter's decision.

The voter model estimated by the new methodology can produce greater consistency between observed and predicted behavior of political parties. Consider Figure 1. The first subfigure shows the actual positions of the seven political parties, superimposed on the density plot of voter policy preferences. The second and third figures show the simulated local Nash equilibria[2], where the utility funcitons of the parties were assumed to be party vote shares.



(a) Actual positions (b) Nash equilibrium, $\gamma = 0$ (c) Nash equilibrium, $\gamma = 0.5$

Figure 1: Comparison of simulated Nash equilibria

The equilibrium that used the voting model estimated for $\gamma = 0.5$ was more in line with the observed policy positions. In particular, the distance between the policy positions of the two largest parties (Likud and Avoda) is much larger in the third figure than in the second figure, although it is still smaller than the observed policy distance.

The outcome of estimation depends on our assumptions about the strategy sets available to the political parties. With a voter utility function (12) one should expect the estimates of β to be closer to the baseline estimate for $\gamma = 0$ if the variance of the elements of the policy sets S_k is smaller (as choosing a more proximate alternative policy will have less effect on the voteshare).

To check that hypothesis, I re-estimated the probabilistic voting model with different strategy sets. Again, I assumed that there are two parties that are player agents; the strategy set of each party consisted of its observed position, plus a number of normal, independent, zero-mean perturbations to that position. I repeated the estimation 100 times for each value of the standard deviation of the perturbation (0.5 or 1), and each size of stategy sets (6 or 11). Figure 2 shows the resulting densities of estimated parameters β.

There are two observations to make. First, the variance of elements in the strategy set has the predicted effect on the estimated value of β. Second, the variance of the estimates of β is smaller as the size of the strategy set increases.

3 Conclusion

I propose a methodology to improve maximum likelihood estimates of multinomial choice models, by assuming the existence of additional utility-maximizing agents. One cannot observe the choice sets of those agents, only the choices that are actually made. I postulate that the

[2] A more general version of model (12) was used.

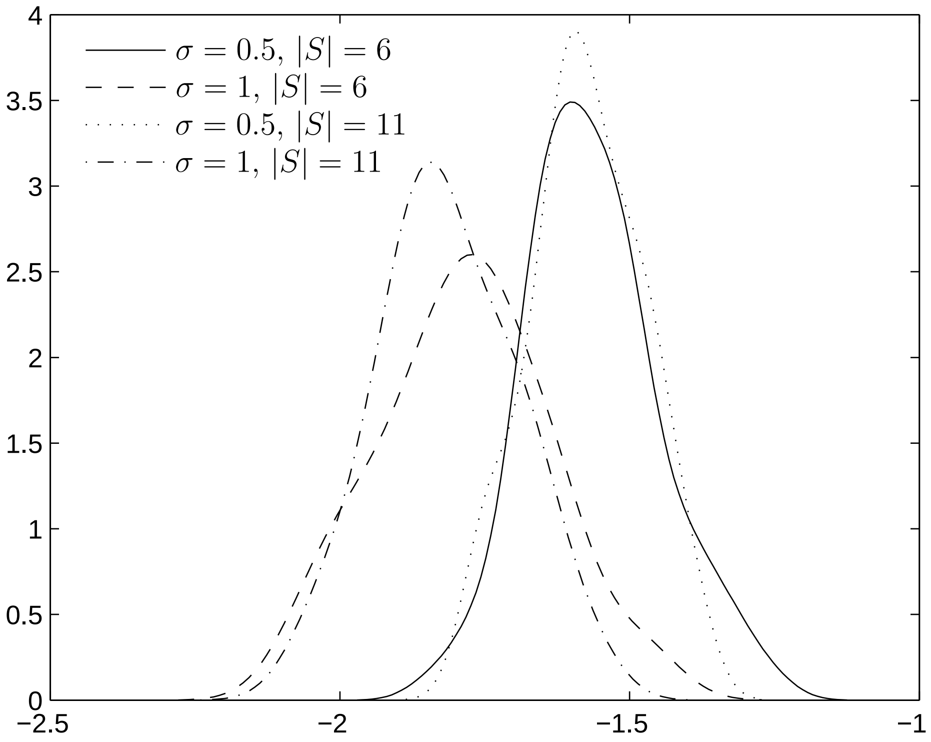


Figure 2: Distributions of β estimates, $\gamma = 0.5$.

observed choices maximize the utility of each player agent; by assuming some strategy set for each player agent, one can construct a modified maximum likelihood estimator.

This approach is applied to estimation of a spatial voting model using the data from a survey conducted prior to 1996 Israel Knesset election. The common problem with all previously estimated spatial voting models is the inconsistency between the observed policy positions of political agents (parties or candidates) and numerically simulated Nash equilibria (given the assumption that the political agents maximize votes). It is always the case that the policy platforms in a simulated Nash equilibrium are located much closer to the center and to each other than the actual policy positions. My research shows that a large part of this discrepancy arises because the traditional multinomial choice approach (either using maximum likelihood or Bayesean methods) underestimates the effect of policy position on vote choice.

References

[1] Poole, Keith T., and Howard Rosenthal. 1984. "U.S. Presidential Elections 168-1980: A Spatial Analysis." *American Journal of Political Science* 28(2): 282–312

[2] Schofield, Norman. 2007. "The Mean Voter Theorem: Necessary and Sufficient Conditions for Convergent Equilibrium." *Review Of Economic Studies* 42: 27–50 .