Dirichlet-Multinomial Regression*

Paulo Guimarães

Medical University of South Carolina

Richard Lindrooth

Medical University of South Carolina

August 22, 2005

Abstract

In this paper we provide a Random-Utility based derivation of the Dirichlet-Multinomial regression and posit it as a convenient alternative for dealing with overdispersed multinomial data. We show that this model is a natural extension of McFadden's conditional logit for grouped data and show how it relates with count models. Finally, we use a data set on patient choice of hospitals to illustrate an application of the Dirichlet-Multinomial regression.

JEL Codes: C25, C21, I11.

*Corresponding Author: Paulo Guimarães, Department of Biostatistics, Bioinformatics and Epidemiology, 135 Cannon Street, Suite 303, Charleston, SC, 49425. Phone: 843-876-1593. E-mail: guimarap@musc.edu.

1 Introduction

McFadden's (1974) conditional logit is the econometric tool of choice for modeling individuals' choice behavior. The attractiveness of this approach stems from its direct link to microeconomic theory. When faced with competing choices, individuals attribute a level of utility to each choice and select that which provides the highest utility. From the perspective of the modeler there are unobservable components, specific to the individual or to the choice, that introduce a random element into the decision process. Researchers observe actual choices and the factors likely to affect the indirect utility associated with the available choices, and use this information to understand how these factors impact the decision making process. The popularity of this approach extends beyond economics into other disciplines such as marketing, psychology and transportation, *inter alia*.

In this paper we focus on the particular situation when the information on actual choices may be grouped into vectors of counts without any loss of information. This will occur if, from the perspective of the modeler, there are groups of individuals facing the same choice set and same choice characteristics. Many examples could be provided, but we select a few that help establish the argument. Consider the problem of identification of the relevant regional factors that affect industrial firm location. Typically, researchers view these individual location decisions as profit (utility) maximizing actions. Firms from diverse industrial sectors evaluate the regional characteristics of different regions (e.g. counties, states) and, idiosyncrasies apart, choose to locate in the region that maximizes potential profits. In this case it is common to assume that all firms face the same choice set and the relevant characteristics of the regional choices are identical for firms belonging to the same industrial sector. The available information consists of regional counts of investments by industrial sector and variables that reflect the characteristics of the regions. A similar situation applies when modeling the locational choices of immigrants. The available information may be summarized by the number of individuals by ethnic group (or country of origin) and the characteristics of regions. Consider another example taken from the literature on political science. There is substantial spatial variation in electoral results, and researchers often devote some effort to understanding what factors impact the choice of a political candidate in an election. In this situation, the choices are the candidates (possibly different by precinct), and the available data are the number of votes for each candidate as well as the characteristics of the candidates and the precincts. A final example, the patient-hospital choice model, is one that we use in our application. Patients with the same diagnosis in the same location (i.e. zip code) will face the same choice set and will have the similar idiosyncratic preferences of a hospital. All patients will be faced with similar travel times and, at least *ex ante*, be subject to similar medical procedures. The information about the quality of each hospital will also be highly correlated within each zipcode if neighbors consult with each other prior to making a decision (Pauly & Satterthwaite 1981). Thus aggregation to the zip code-disease level can be done with minimal loss of information, while at the same time, making analysis of large urban markets computational feasible on a personal computer. All of the above examples share a common feature. Despite that the data consist of individual level choices, the true level of variation of the data is at the group level. Thus, data for the dependent variable may be summarized by vectors of counts.

Nevertheless, we are interested in modeling these data as resulting from McFadden's discrete choice Random Utility Maximization (RUM) framework. This means that inference is based on the multinomial distribution because our interest lies in studying the impact that covariates have on choice probabilities, treating the number of individuals in each group as given. In all of the above examples, groups share some common characteristics: firms share industrial sector characteristics; immigrants share ethnic characteristics; voters share characteristics with neighbors in the same precinct; and patients share location and disease characteristics. This introduces the possibility that there exist some unobservable group specific effects that are likely to equally influence all individuals belonging to the same group. If this happens, then the individual choices will be correlated and the vectors of counts will exhibit extra multinomial variation (overdispersion). Much like what happens with count models, the statistical properties of the parameter estimates will be affected [see McCullagh & Nelder (1989)]. One approach to deal with this problem is through the use of quasi-likelihood (robust) estimators [eg. Mebane & Sekhon (2004)]. Here we present a fully parametric alternative based on the Dirichlet-Multinomial distribution. We use McFadden's RUM framework to explicitly derive a discrete-choice model that is appropriate for grouped data and that naturally accounts for extra multinomial variation. The presentation emphasizes the connection with count data models. The paper is organized as follows. In section 2 we present McFadden's conditional logit model. In section 3 we present a detailed derivation of the Dirichlet-Multinomial regression and highlight its connections with count data models. In section 4 we provide an application of the Dirichlet-Multinomial to the choice of hospital by a sample of patients from the Tampa-St Petersburg Statistical Metropolitan Area (SMSA). Section 5 concludes.

2 The Grouped Conditional Logit Model

Following McFadden's (1974) *Random Utility Hypothesis* it is assumed that each individual (consumer, firm, etc.) i faces an exhaustive set of J_i mutually exclusive alternatives. Each alternative j in his choice set has utility (profit) given by:

$$U_{ij} = V_{ij} + \epsilon_{ij} \ , \tag{1}$$

where the first term in the right-hand side is a function of observable components (the systematic component) and ϵ_{ij} is a random variable. Assuming that the ϵ_{ij} are independent and identically distributed as Type I Extreme Value and that individual i selects the choice for which U_{ij} is maximum, then it can be shown that the probability that the individual selects choice j among the set of J_i alternatives is given by

$$p_{ij} = \frac{\exp(V_{ij})}{\sum_{j=1}^{J_i} \exp(V_{ij})} = \frac{\exp(\beta' \mathbf{x}_{ij})}{\sum_{j=1}^{J_i} \exp(\beta' \mathbf{x}_{ij})} \ , \tag{2}$$

where, as usually done, we are assuming that V_{ij} is a linear combination of observable variables. Thus, β is a vector of unknown parameters and the $\mathbf{x}_{ij}$ are covariates that may change with individual, choice, or both. This logit formulation is quite general, and it contains as a special case the multinomial logit model for the situation when covariates are restricted to characteristics of the individual. To estimate the model by maximum likelihood, we define the variable $d_{ij} = 1$ if individual i picks choice j, and $d_{ij} = 0$ otherwise. Hence, the likelihood function for the conditional logit model may be expressed as,

$$L_{CL} = \prod_{i=1}^{N} \prod_{j=1}^{J_i} p_{ij}^{d_{ij}} \ . \tag{3}$$

The above presentation of the conditional logit model is quite general and admits the (possible) situation where the number of choices and their characteristics differ across all individuals. But, as argued earlier, there are many occasions where the p_{ij} are identical for groups (clusters) of individuals. This will happen when a set of individuals is presented with the same choices and vectors of (choice) characteristics meaning that covariates change across groups and/or choices but not across individuals within a group. If we index the different groups by g and let G denote the total number of groups, then the likelihood in (3) becomes that of the grouped conditional logit model (without loss of generality and to simplify notation we will henceforth assume that all individuals face choice sets with the same number of alternatives),

$$L_{GL} = \prod_{g=1}^{G} \prod_{j=1}^{J} p_{jg}^{n_{jg}} \ , \tag{4}$$

where the n_{jg} are the number of individuals from group g that select choice j. Within this context the utility of the choice faced by individual i belonging to group g may be expressed as

$$U_{ijg} = \beta' \mathbf{x}_{jg} + \varepsilon_{ijg} \ , \tag{5}$$

where, the $\mathbf{x}_{jg}$ are characteristics of the group and/or choice that affect individual decisions. The other random term, ε_{ijg}, is as defined earlier. Thus to estimate the above model, all that is required is information on the vectors of counts by group, the n_{jg}, and the corresponding information on the $\mathbf{x}_{jg}$.

It would have been possible to model n_{jg} directly as a count variable. To see this let,

$$E(n_{jg}) = \lambda_{jg} = \exp(\alpha_g + \beta^{'}\mathbf{x}_{jg}) \ ,$$

and assume a Poisson distribution for n_{jg},

$$f_{Poi}(n_{jg}) = \frac{\lambda_{jg}^{n_{jg}} e^{-\lambda_{jg}}}{n_{jg}!} \ . \tag{6}$$

This implies that n_g, the sum of counts for group g, also follows a Poisson law with parameter $\lambda_g = \sum_{j=1}^{J} \lambda_{jg}$. It is now straightforward to verify that if we construct the likelihood function by conditioning on the sum of counts for each group,

$$L_{PoiC} = \prod_{g=1}^{G} \prod_{j=1}^{J} f_{Poi}(n_{jg}) f_{Poi}^{-1}(n_g) \ , \tag{7}$$

then the group level constants, α_g, cancel out and we will obtain (ignoring multiplication constants in the likelihood) the maximum likelihood function of the grouped conditional logit shown in (4). As shown in Guimarães, Figueiredo & Woodward (2003) the grouped conditional logit and the Poisson regression will yield identical estimates for β and its variance-covariance matrix, i.e., the same estimates will result whether or not the likelihood for the Poisson distribution is conditioned in the group totals.

3 The Dirichlet-Multinomial Model

3.1 The Model

In the following we admit that the utility ascribed to each choice is also influenced by an additional unobservable factor specific to each group. This factor, which we will treat as a random variable, accounts for omitted variables that exert their influence at the group level but that are not observed by the modeler. To account for this type of group specific unobserved heterogeneity we modify (5) and let it become,

$$U_{ijg} = \beta' \mathbf{x}_{jg} + \eta_{jg} + \varepsilon_{ijg} \ , \tag{8}$$

where the η_{jg} are random effects that affect identically all individuals belonging to group g and the ε_{ijg} are assumed to be independent conditional on the group random effects. The existence of these group specific random variates will induce some correlation across the choices of individuals in the same group. As we will see, this correlation will translate into overdispersion of the n_{jg} count variables. Conditional on the group level random effects, $\eta_{jg}s$, and drawing again on McFadden's (1974) result, we can express the probability that an individual from group g selects choice j as,

$$\widetilde{p}_{jg} = \frac{\exp(\beta' \mathbf{x}_{jg} + \eta_{jg})}{\sum_{j=1}^{J} \exp(\beta' \mathbf{x}_{jg} + \eta_{jg})} = \frac{\widetilde{\lambda}_{jg} \exp(\eta_{jg})}{\sum_{j=1}^{J} \widetilde{\lambda}_{jg} \exp(\eta_{jg})} \ , \tag{9}$$

where $\widetilde{\lambda}_{jg} = \exp(\beta' \mathbf{x}_{jg})$. Now, the conditional likelihood function (conditional on the η_{jg}) is given by:

$$L = \prod_{g=1}^{G} \prod_{j=1}^{J} \widetilde{p}_{jg}^{n_{jg}}. \tag{10}$$

Assume that the random cluster effects, $\exp(\eta_{jg})s$, are i.i.d. gamma distributed with parameters $(\delta_g^{-1}\widetilde{\lambda}_{jg}, \delta_g^{-1}\widetilde{\lambda}_{jg})$ where $\delta_g > 0$ is a group specific parameter. Under this assumption it follows that the $\exp(\eta_{jg})$ have an expected value of unity and a variance equal to $\delta_g \widetilde{\lambda}_{jg}^{-1}$. Moreover, the variables defined by the product $\widetilde{\lambda}_{jg} \exp(\eta_{jg})$ also follow independent gamma distributions with parameters $(\delta_g^{-1}\widetilde{\lambda}_{jg}, \delta_g^{-1})$. Given that all these variables follow independent gamma distributions with the same scale parameter, we can directly apply a theorem demonstrated in Mosimann (1962) (Theorem 1, pg 74) to conclude that the vector $(\widetilde{p}_{1g}, \widetilde{p}_{2g}, ..., \widetilde{p}_{Jg})$ follows a multivariate beta distribution (Dirichlet distribution) with parameters $(\delta_g^{-1}\widetilde{\lambda}_{1g}, \delta_g^{-1}\widetilde{\lambda}_{2g}, ..., \delta_g^{-1}\widetilde{\lambda}_{Jg})$, that is,

$$f_{DM}(\widetilde{p}_{1g}, ..., \widetilde{p}_{J-1g}) = \frac{\Gamma(\delta_g^{-1}\widetilde{\lambda}_g)}{\prod_{j=1}^{J} \Gamma(\delta_g^{-1}\widetilde{\lambda}_{jg})} \prod_{j=1}^{J} \widetilde{p}_{jg}^{\delta_g^{-1}\widetilde{\lambda}_{jg}-1}. \tag{11}$$

with $\widetilde{p}_{Jg} = 1 - \sum_{j=1}^{J-1} \widetilde{p}_{jg}$. From the properties of the Dirichlet distribution it follows that

$$E(\widetilde{p}_{jg}) = \frac{\delta_g^{-1}\widetilde{\lambda}_{jg}}{\sum_{j=1}^{J} \delta_g^{-1}\widetilde{\lambda}_{jg}} = \frac{\widetilde{\lambda}_{jg}}{\sum_{j=1}^{J} \widetilde{\lambda}_{jg}} . \tag{12}$$

showing that on average the choice probabilities are identical to those obtained from the grouped conditional logit model. Mosimann (1962) has also shown that the multivariate beta distribution is a prior conjugate to the multinomial distribution. Given that the contribution of group g to (10) amounts to the kernel of a multinomial distribution with parameters $(n_g; \widetilde{p}_{1g}, \widetilde{p}_{2g}, ..., \widetilde{p}_{Jg})$ we can use Mosimann's (1962) result to arrive at a closed form expression for the unconditional likelihood distribution. Adding the necessary constants to transform (10) into a product of multinomial distributions, and computing the unconditional likelihood by integrating with respect to the $\widetilde{p}_{jg}$, we obtain,

$$L_{DM} = \prod_{g=1}^{G} \int \prod_{j=1}^{J} n_g! \frac{\widetilde{p}_{jg}^{n_{jg}}}{n_{jg}!} f_{DM}(\widetilde{p}_{1g}, \widetilde{p}_{2g}, ..., \widetilde{p}_{J-1g}) d\widetilde{p}_1 d\widetilde{p}_2, ..., d\widetilde{p}_{J-1} \ . \tag{13}$$

The expression under the integral results in the Dirichlet-Multinomial multivariate distribution (also known as the Compound Multinomial) and, in its closed form, equals:

$$L_{DM} = \prod_{g=1}^{G} \frac{n_g! \Gamma(\delta_g^{-1} \widetilde{\lambda}_g)}{\Gamma(\delta_g^{-1} \widetilde{\lambda}_g + n_g)} \prod_{j=1}^{J} \frac{\Gamma(\delta_g^{-1} \widetilde{\lambda}_{jg} + n_{jg})}{\Gamma(\delta_g^{-1} \widetilde{\lambda}_{jg}) n_{jg}!} \ . \tag{14}$$

Maximization of the above likelihood function provides estimates for the β in (8).

We have shown earlier that it is possible to obtain the likelihood for the grouped conditional logit model by letting the n_{jg} follow a Poisson law and conditioning on the total sum for each group. A similar relationship exists for the Dirichlet-Multinomial model. To show this suppose that we model n_{jg} directly as an overdispersed count variable by assuming that

$$E(n_{jg}/\eta_{jg}) = \exp(\alpha_g + \beta' \mathbf{x}_{jg} + \eta_{jg}) \ .$$

Now, as we did earlier, let the $\exp(\eta_{jg})$ be independently gamma distributed with parameters $(\theta_g^{-1} \lambda_{jg}, \theta_g^{-1} \lambda_{jg})$ and admit that conditional on the random effect the

n_{jg} follow a Poisson law. Under these assumptions, it is known that the n_{jg} are distributed according to the negative binomial law with a probability density function given by

$$f_{NB}(n_{jg}) = \frac{\Gamma(\theta_g^{-1}\lambda_{jg} + n_{jg})}{\Gamma(\theta_g^{-1}\lambda_{jg})n_{jg}!} \left(\frac{1}{1+\theta_g^{-1}}\right)^{n_{jg}} \left(\frac{\theta_g^{-1}}{1+\theta_g^{-1}}\right)^{\theta_g^{-1}\lambda_{jg}} , \tag{15}$$

and with expected value and variance of,

$$\begin{aligned} E(n_{jg}) &= \lambda_{jg} , \\ V(n_{jg}) &= \lambda_{jg}(1+\theta_g) . \end{aligned}$$

This type of parametrization of the negative binomial is known in the econometric literature as NEGBIN type 1, (Cameron & Trivedi 1998). Under this particular parametrization the total sum of counts for each group, n_g, also follow a negative binomial distribution,

$$f_{NB}(n_g) = \frac{\Gamma(\theta_g^{-1}\lambda_g + n_g)}{\Gamma(\theta_g^{-1}\lambda_g)n_g!} \left(\frac{1}{1+\theta_g^{-1}}\right)^{n_g} \left(\frac{\theta_g^{-1}}{1+\theta_g^{-1}}\right)^{\theta_g^{-1}\lambda_g} . \tag{16}$$

Now, it is fairly evident that constructing the likelihood by conditioning on the sum of counts for each group as in

$$L_{NBC} = \prod_{g=1}^{G}\prod_{j=1}^{J} f_{NB}(n_{jg}) f_{NB}^{-1}(n_g) , \tag{17}$$

and letting $\theta_g^{-1}\exp(\alpha_g) = \delta_g^{-1}$ results in the likelihood for the Dirichlet-Multinomial regression. The above derivation also sheds some light into what happens to the group level intercepts (the α_gs). These constants drop out when we condition in the group totals for the Poisson case but that does not happen for the negative binomial case. In this latter situation the α_gs are not identifiable and are absorbed into the δ_g.

3.2 Additional Considerations

Under the Dirichlet-Multinomial model the marginal distributions of the n_{jg} follow a beta-binomial distribution - a mixture of the beta and binomial distributions [see Johnson, Kotz & Balakrishnan (1997)] with expected value and variance given by

$$\begin{aligned} E(n_{jg}/n_g) &= n_g E(\widetilde{p}_{jg}) \, , \\ V(n_{jg}/n_g) &= n_g E(\widetilde{p}_{jg}) \, (1 - E(\widetilde{p}_{jg})) \, \frac{\widetilde{\lambda}_g + n_g \delta_g}{\widetilde{\lambda}_g + \delta_g} \, . \end{aligned}$$

It is now obvious how the Dirichlet-Multinomial model accommodates overdispersion. The variance of n_{jg} is increased by a constant (by group) factor when compared to the variance that would attain under the multinomial distribution. As δ_g (or θ_g) goes to zero (the variance of the group random effects tend to zero), the variance of n_{jg} collapses to that of the binomial distribution (the marginal distribution for the multinomial). As mentioned earlier, the introduction of a group random effect induces a parallel phenomena of correlation across the choices. It is well known that a variable with a beta-binomial distribution may be interpreted as the sum of equicorrelated Bernoulli variables [see, for example, McCulloch & Searle (2001)].

This means that we can interpret each n_{jg} as resulting from the sum of n_g Bernoulli variables, each reflecting the individual decisions to select choice j with probability p_{jg}. The intragroup correlation coefficient between these Bernoulli variables is given by

$$\gamma_g = \frac{1}{\delta_g^{-1}\widetilde{\lambda}_g + 1} = \frac{\delta_g}{\widetilde{\lambda}_g + \delta_g} = \frac{\theta_g}{\lambda_g + \theta_g} \ . \tag{18}$$

As expected, this correlation coefficient tends to zero as δ_g approaches zero. When implementing the Dirichlet-Multinomial one should be aware of the fact that the δ_g are allowed to vary by group. It could be possible to absorb all that variability by adding to the specification a constant specific to each one of the groups. However, this may be impractical in applied work, particularly if we are dealing with a large number of groups. One possible option (which we designate by Option 1) is to let the δ_g be a function of explanatory variables that characterize the different groups (and consequently that do not affect the choice probabilities). If only a constant is used then we are assuming that the group specific coefficients are all identical (i.e. $\delta_g = \delta$). Another option (Option 2) is to parameterize the Dirichlet-Multinomial likelihood in terms of the intragroup correlation coefficient (γ_g) and let it be a function of group level covariates. If those covariates are restricted to a constant then it is assumed that all groups share a common correlation coefficient. This approach was used in Paul, Liang & Self (1989) and Shonkwiler & Hanley (2003). When parameterized in terms of the intraclass correlation coefficient, the likelihood function for the Dirichlet-

Multinomial becomes

$$L_{DMeq} = \prod_{g=1}^{G} \frac{n_g! \Gamma(\gamma_g^{-1} - 1)}{\Gamma((\gamma_g^{-1} - 1) + n_g)} \prod_{j=1}^{J} \frac{\Gamma((\gamma_g^{-1} - 1)p_{jg} + n_{jg})}{\Gamma((\gamma_g^{-1} - 1)p_{jg})n_{jg}!} \ . \tag{19}$$

When restricted to just two choices the Dirichlet-Multinomial distribution simplifies to the Beta-Binomial distribution becoming a version of the binomial logit regression that allows for overdispersion (or correlation across choices). Applications of the Beta-Binomial logit regression model have relied on different parameterizations. Heckman & Willis (1977), who apparently were the first to propose the Beta-Binomial regression model (the authors called it a "Beta-Logistic regression"), used a parametrization which is equivalent to our Option 1 without group covariates. Applications in Biostatistics [eg. Kupper, Portier, Hogan & Yamamoto (1986), Prentice (1986)] parameterized the Beta-Binomial in terms of the correlation coefficient (Option 2).

Estimation of the Dirichlet-Multinomial model offers no particular challenge, and numerical optimization routines based on the Newton-Raphson algorithm converge rapidly to a global maximum. Functionally, the likelihood of the Dirichlet-Multinomial model presented in (14) is identical to that of the "negative binomial with fixed effects," a model proposed by Hausman, Hall & Griliches (1984) to deal with count panel data. Thus, existing routines for estimation of the "negative binomial with fixed effects" available in econometric packages (eg. LIMDEP, Stata, etc.) may be readily employed to estimate the Dirichlet-Multinomial (Option 1).

It is intuitive to see that the Dirichlet-Multinomial will collapse to the conditional logit model if: a) The variances of the random effects are zero; b) There is a single individual per group. The first of these assertions is quite obvious. The second one follows directly from inspection of the likelihood function of the Dirichlet-Multinomial model. With one individual per group, $n_g = 1$ and $n_{jg} = 1$ for the choice selected by the individual and 0 otherwise. Applying the recursive property of the gamma function, $\Gamma(x) = (x-1)\Gamma(x-1)$, it is immediate to verify that (14) collapses to (4). As just mentioned, in the absence of overdispersion the Dirichlet-Multinomial distribution collapses to a standard multinomial distribution. Hence, it is possible to implement a likelihood ratio test for overdispersion based on the comparison of the likelihoods of the Dirichlet-Multinomial and the grouped conditional logit [but note that the null hypothesis for the test is in the boundary of the parameter space (see Self & Liang (1987))].

4 Application

To illustrate the application of the Dirichlet-Multinomial regression we model the choice of hospital by patient in the Tampa-St. Petersburg market. Our original data consists of 1998 inpatient claims from the State Inpatient Database of the Healthcare Cost and Utilization Project (HCUP-SID) for the general hospitals in Hillsborough, Pasco and Pinellas counties. All patients in these counties, as well as some of those from surrounding zip codes, were included (surrounding zip codes are included only if more than 60 percent of residents sought care in the Tampa-St. Petersburg market). We restricted our analysis to non-emergency admissions of patients in the 5 most frequent diagnosis related groups (DRGs) (patients with the same DRGs have a clinically similar condition). The 5 most frequent DRGs were: 373 - Vaginal Delivery w/o Complicating Diagnoses; 209 - Total joint replacement or Major Joint and Limb Reattachment Procedures of Lower Extremity; 116 - Other Permanent Cardiac Pacemaker Implant or PTCA with Coronary Artery Stent Implant; 127 - Heart Failure and Shock and 88 - Chronic Obstructive Pulmonary Disease. Additionally, we only considered patients who had Medicare, fee-for-service or PPO insurance coverage because they were likely to have access to all of the hospitals. We also excluded patients aged 18 or less years. Our final sample consisted of 13,079 patients.

We used data from the 1998 American Hospital Association's Annual Survey of Hospitals to identify for-profit hospitals, teaching hospitals, and each hospital's nursing intensity. We define these as indicator variables showing whether or not the hospital is a teaching institution (TEACH), as well as a measure of nursing intensity calculated as full time equivalent (FTE) nursed per inpatient day (NURSE). We also used *MAPQUEST.COM* to measure the drive time from the epicenter of each zipcode to each hospital in the choice set, defined as DVTIME. We interact DVTIME with the other patient and hospital characteristics to fully account for spatial preferences across hospitals and diagnoses. We also merged in zipcode level median income from the Census Bureau. Other variables include the approximate drive time between the patient's residence zip code and the hospital (DVTIME), and interactions between drive time and hospital characteristics. For patients in DRGs 116 and 127 we added an additional hospital indicator variable (CIRC) indicating whether or not the hospital had specialized services in circulatory diseases.

For our example the level of variation of the data is at the zipcode $\times$ DRG level.

Given that patients in our sample originate from a total of 133 different zipcode areas, this results in a potential maximum number of 665 groups. After excluding groups with a zero number of patients, the number of groups dropped to 598. We assumed that all the patients could choose between any of the 25 hospitals in our sample.

In Table 1 we present the results of our estimation. Column 1 shows the estimates from the conditional logit model. Overall, the sign and magnitude of the coefficients are consistent with theory. First, patients choose hospitals based on various attributes. The negative coefficient on drive time is consistent with the expectation that patients are more likely to go to a closer hospital. In addition, the coefficient on nurse FTE per day is positive and significant, implying that hospitals with high nurse staffing ratios are preferable. Teaching hospitals are less attractive to patients holding drive time constant but are more likely to be selected as drive time is increased. Teaching hospitals generally have the most advanced technology and the capability of treating the most complicated cases. Thus, patients that live far from a teaching hospital with relatively complicated cases are likely to travel to a teaching hospital. Thus the coefficient on the teach / drive time interaction is positive and beyond about 30 minutes larger than the teach variable This can be seen by noting that the coefficient on teach/drivetime is positive and about 0.050, whereas the coefficient on Teach is about -1.80. Thus, patients that live beyond 30-40 minutes from the teaching hospitals are actually more likely to visit them because $40 * 0.05 > 1.80$. For profit hospitals are less attractive, *ceteris paribus* and the interaction between profit and drive time is not significant. Finally, the coefficient on CIRC is positive and significant, as expected. However, most variables present high z-statistics a sign that overdispersion may be a problem.

In column 2 we estimate the Dirichlet-Multinomial model imposing the restriction that $\delta_g = \delta$ (Option 1). The large improvement in the log-likelihood reinforces the idea that overdispersion is a problem with this data. As expected, the higher z-statistics were substantially deflated, but overall there were not considerable changes in the estimated coefficients. The variables more affected are the DVTIME interaction variables. One of them (DVTIME*PROFIT) becomes non-significant, and DVTIME*NURSE reverses sign and significance suggesting now that importance of nurse staffing ratios declines as drive time is increased. As discussed earlier we can account for some of the potential variability in δ_g by introducing covariates that change with group. The analysis of the impact of these covariates may be of secondary interest (they do not affect choice probabilities), but its introduction may help provide a better fit to the data. In line with this idea we estimated a second version of the Dirichlet-Multinomial, introducing as covariates dummy variables for each of the DRGs (the omited category was DRG 88) as well as a variable containing average household income at the zipcode level. All of these variables prove statistically significant, suggesting that there is substantial variation across groups in the δ_g. However, the results for the estimated coefficients affecting probabilities remain practically unchanged.

The next version of the Dirichlet-Multinomial is parameterized in terms of the correlation coefficient (Option 2). Curiously the model assuming identical correlation coefficient provides a better fit then any of the other estimations. Nevertheless, we obtain results that do not differ much from the previous models. The estimated intragroup correlation, $\widehat{\gamma}$, is significantly different from zero.

Our final specification admits the possibility that the intragroup correlation coefficient is linearly related with group level covariates (income and dummies for DRG). Again, except for small changes in magnitude, we do not observe any change in the sign and significance of the coefficients associated with the choice probabilities. An interesting result is that median income is linked to stronger intragroup correlations. Note that the Medicare patients do not face different prices across hospitals and most private patients in our sample do not face differences in out-of-pocket payments across hospitals. Thus for these patients this could reflect better information/education which is likely to be correlated with income. In addition, DRG 373, Vaginal Delivery w/o complicating diagnoses, also exhibits strong intragroup correlation. This is likely due to the fact that expectant mothers gather information regarding hospitals prior to delivery. This is possible with normal deliveries because it is predictable in advance, leaving plenty of time to shop around. Clearly a major source of information is their neighbors. DRG 209, total joint replacement or reattachment of the lower extremity exhibits relatively low correlation, possibly due to the fact that patients in these DRGs are more likely to be elderly, and have less time to shop before the procedure is done. Thus there may be less information sharing amongst people in the same zip code.

5 Conclusion

In this paper we showed that Dirichlet-Multinomial regression is a natural extension of McFadden's conditional logit model. The relationship of the Dirichlet-Multinomial regression to the grouped conditional logit regression is much like that of the negative binomial regression to the Poisson regression. It provides a viable parametric alternative to deal with the problem of overdispersed data that may arise when the conditional logit model is applied to grouped data. Moreover, because the Dirichlet-Multinomial regression allows for parameterizing of the intra-class correlation coefficient in terms of group specific covariates, it may reveal additional information which may not be captured by the grouped conditional logit model.

Table 1: Choice of Hospital: Estimates for Different Parametrizations of the Dirichlet-Multinomial

	CLM	Dirichlet-Multinomial			
		$\delta_g = \delta$	$\delta_g = f(\mathbf{x}_g)$	$\gamma_g = \gamma$	$\gamma_g = f(\mathbf{x}_g)$
PROFIT	-0.654	-0.656	-0.626	-0.660	-0.604
	(-16.77)	(-7.51)	(-7.21)	(-7.87)	(-7.20)
TEACH	-1.875	-1.868	-1.860	-1.754	-1.655
	(-20.28)	(-10.37)	(-10.27)	(-10.27)	(-9.69)
NURSE	0.068	0.198	0.165	0.191	0.139
	(3.19)	(4.33)	(3.57)	(4.26)	(3.09)
DVTIME	-0.151	-0.085	-0.089	-0.100	-0.105
	(-40.72)	(-12.72)	(-13.10)	(-15.37)	(-15.93)
DVTIME * PROFIT	-0.009	0.004	0.004	0.003	0.002
	(-4.46)	(1.06)	(0.99)	(0.75)	(0.42)
DVTIME * TEACH	0.038	0.059	0.059	0.050	0.050
	(9.83)	(8.62)	(8.59)	(8.01)	(7.90)
DVTIME* NURSE	0.002	-0.008	-0.007	-0.006	-0.006
	(2.18)	(-3.76)	(-3.38)	(-3.32)	(-2.83)
CIRC	0.896	0.643	1.097	1.081	1.038
	(22.55)	(10.80)	(12.70)	(12.68)	(12.17)
ZIPINC			-0.016		0.003
			(-2.37)		(2.19)
DRG116			-0.867		0.043
			(-6.19)		(1.94)
DRG127			-0.721		-0.013
			(-4.92)		(-0.58)
DRG209			0.708		-0.083
			(5.82)		(-4.43)
DRG373			-0.352		0.089
			(-2.96)		(4.01)
Constant		0.006	0.505	0.232	0.186
		(0.04)	(2.39)	(36.680)	(7.77)
Log-Likelihood	-14559.2	-7465.9	-7360.2	-7226.6	-7167.6

Note: z-statistics in parentheses.

References

Cameron, A. C. & Trivedi, P. K. (1998), *Regression Analysis of Count Data*, Econometric Society Monographs, Cambridge University Press, Cambridge, UK.

Guimarães, P., Figueiredo, O. & Woodward, D. (2003), 'A tractable approach to the firm location decision problem', *The Review of Economic and Statistics* **85**(1), 201–204.

Hausman, J., Hall, B. & Griliches, Z. (1984), 'Economic models for count data with an application to the patents-R&D relationship', *Econometrica* **52**, 909–938.

Heckman, J. & Willis, R. (1977), 'A beta-logistic model for the analysis of sequential labor force participation by married women', *Journal of Political Economy* **85**, 27–58.

Johnson, N., Kotz, S. & Balakrishnan, N. (1997), *Discrete Multivariate Distributions*, Wiley Series in Probability and Statistics, John Wiley and Sons, NY.

Kupper, L., Portier, C., Hogan, M. & Yamamoto, E. (1986), 'The impact of litter effects on dose-response modeling in teratology', *Biometrics* **42**, 85–98.

McCullagh, P. & Nelder, J. (1989), *Generalized Linear Models*, Monographs on Statistics and Applied Probability, Chapman & Hall, New York.

McCulloch, C. & Searle, S. (2001), *Generalized, Linear, and Mixed Models*, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., New York.

McFadden, D. (1974), Conditional logit analysis of qualitative choice behavior, *in* P. Zarembka, ed., 'Frontiers in Econometrics', Academic Press, NY, chapter 4, pp. 105–142.

Mebane, W. & Sekhon, J. (2004), 'Robust estimation and outlier detection for overdispersed multinomial models of count data', *American Journal of Political Science* **48**, 392–411.

Mosimann, J. (1962), 'On the compound multinomial distribution, the multivariate beta-distribution, and correlations among proportions', *Biometrika* **49**, 65–82.

Paul, S. R., Liang, K. Y. & Self, S. G. (1989), 'On testing departure from the binomial and multinomial assumptions', *Biometrics* **45**, 231–236.

Pauly, M. V. & Satterthwaite, M. A. (1981), 'The pricing of primary care physicians services: A test of the role of consumer information', *Bell Journal of Economics* **12**, 488–506.

Prentice, R. (1986), 'Binary regression using an extended beta-binomial distribution, with discussion of correlation induced by covariate measurement errors', *Journal of the American Statistical Association* **81**, 321–327.

Self, S. & Liang, K.-Y. (1987), 'Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions', *Journal of the American Statistical Association* **82**, 605–610.

Shonkwiler, J. & Hanley, N. (2003), 'A new approach to random utility modeling using the dirichlet multinomial distribution', *Environmental and Resource Economics* **26**, 401–416.