On the Estimation and Inference of Global Categorical Effects in General Parametric Models: with a Logit Regression Application

Paulo Saraiva[*, †] Somdutta Basu[‡]

April 20, 2017

Abstract

In this paper we propose a method for estimating and conducting inference on categorical effects of random variables that are characterize by more that two categories. We focus on a class of parametric asymptotically normal estimators in deriving the the properties which allows for inference on the proposed categorical effects. Existing methods prescribe pairwise within categories or sub-sampling comparisons. The proposed method allows for global comparison between each mutually exclusive category relative to a global evaluation of the other categories, while allowing the remainder of the regressors in the model to be evaluated unconditional on the category of interest. We also provide an example and a Monte Carlo simulation in the context of the Logit model.

JEL classification: C1, C5, C12, C13

Key words: categorical effect, Logit, M-estimators, marginal effect

1 Introduction

There are many statistical packages available that calculate marginal effects for both linear and nonlinear models. In this paper we use Stata and R as examples in the calculation of "marginal effects." In both software there are two predominant approaches used, namely the Marginal Effects at Means (MEM) and Average Marginal Effect (AME). The commands that compute both MEM and AME are based on the partial derivative of the nonlinear model, whenever the regressor of interest is a continuous variable. If, however, the effect of interest is related to categorical variables, a partial derivative will not make conceptual

[*]Corresponding author.

[†]Financial Services and Practice, Econ One Research, 1625 Shattuck Ave, Suite 200. Berkeley, CA 94709, USA, and Department of Economics, University of Colorado, Boulder, CO 80309, USA. Email: `psaraiva@econone.com`, `paulo.saraiva@colorado.edu`.

[‡]Financial Services and Practice, Econ One Research, 1625 Shattuck Ave, Suite 200. Berkeley, CA, USA, 94709. Email: `sbasu@econone.com`.

sense, since infinitesimal changes are not possible for discrete variables. In which case, a difference between the function that characterizes the model evaluated at two different points of interest is calculated as the effect of a particular category.

Calculation of categorical effect is straight forward for regressors that characterize only two categories (Anderson and Newell, 2003). For instance, if the variable of interest is gender and the data contains only two categories for gender, namely male and female, a dummy variable is included to the model. There are two possible dummies that can be included in the model. A dummy that takes value one for males and zero otherwise, or a dummy that takes value one for females and zero otherwise, but, for identification purposes, not both. In this case, the effect of gender obtained through the difference between the function that characterizes the model evaluated at the two different gender categories, *i.e.* the difference of the estimated function of interest evaluated at the dummy equal to one and zero. For linear models without iteration terms, the aforementioned difference equals to the parameter associated with the dummy variable characterizing gender.

If, however, one is interested in the categorical effects of a variable that takes multiple discrete values, existing software do not compare the difference between one category and the rest, but only one to one comparisons between categories. For instance, in the context of a race variable that characterize six categories, Native, Asian, Hispanic, Black, White, and Other, existing algorithms only conduct pairwise comparisons. In other words, they only directly estimate and conduct inference related to the moving from one category to another, for instance the categorical effect of Hispanic relative to Asian. The current version of Stata, however, has an option that compares the model evaluated, for instance, at the Hispanic sub-sample versus the non Hispanic sub-sample. In other words, current methods either compare a category A to a category B, while other regressors evaluated at their respective estimated means, or they compare the sub-subsample of category A to the sub-sample of non A categories, while other regressors are evaluated at the estimated mean conditional on category A and non A respectively (*e.g.,* Williams, 2012). We propose a comparison method of comparing category A to non A categories, while other regressors are evaluated at their respective estimated unconditional mean.

Mathematically, consider a situation in which researcher is interested in estimation and inference related to the categorical effect on a function:

$$m(X, Z)$$

where $Z = (Z_1, \ldots, Z_{K_Z})$ is a $1 \times K_Z$ random variable with mutually exclusive columns of zeros and ones related to some set of categories and X is a $1 \times K_X$ vector of other regressors. When m is such that $E(Y|X, Z) = m(X, Z)$ almost surely, then m is said to be a regression function. Current statistical packages give researcher two options, namely the Categorical Effects at Means (CEM) and the Average Categorical Effect (ACE), which consist of the categorical variable equivalent of MEM and AME respectively. The premise of CEM and ACE entails evaluating m at estimated expected values of the regressors and estimating the expectation of m evaluated at (X, Z) respectively. Because both CEM and ACE is based on the difference of m evaluated at different points, some or all the expectations considered are conditional on two different values of Z. For instance, letting $z_1 = \mathbf{i}'_{1,K_Z}$ and $z_0 = \mathbf{i}'_{2,K_Z}$, where $\mathbf{i}_{r,s}$ is the r-th column of the s identity matrix $\mathbf{I}_s$, two of the CEM calculations in Stata consists in estimating on of the following differences,

$$m(E(X), E(Z|Z = z_1)) - m(E(X), E(Z|Z = z_0)), \text{ or}$$

$$m(E(X, Z|Z = z_1)) - m(E(X, Z|Z \neq z_1)).$$

In this paper we argue that another useful method of measuring CEM ought to be estimating

$$m(E(X), E(Z|Z = z_1)) - m(E(X), E(Z|Z \neq z_1))$$

which, to the best of our knowledge, is not directly implemented with currently available statistical software packages. The proposed method makes strong economic sense and allows researchers to filter the effect of z_1 alone relative to an aggregation in the different coordinates of Z, while allowing other regressors X to be evaluated at their unconditional means.

We consider a generic parametric m in order to incorporate many different statistical models. For instance, in a Logit model, $m(X, Z) = P(Y = 1|X, Z) = \Lambda(X\beta_0 + Z\gamma_0)$, where Λ is the logistic distribution function. In a system of simultaneous equations, m can refer to one of the several equations that characterize the model. In multinomial models, where Y takes value on a finite $\mathcal{Y} \subset \mathbb{N}$, the m of interest might be $m(X, Z) = P(Y = y|X, Z)$, for some $y \in \mathcal{Y}$.

2 Estimation of the Categorical Effect

As previously noted, we consider a parametric function of interest m that relate the random variable Y to the set of regressors X and Z. The random variable Z takes value in $\mathcal{Z}$, where

$$\mathcal{Z} = \bigcup_{j=1}^{K_Z} \{\mathbf{i}'_{j,K_Z}\}.$$

Intuitively this means that Z is a random variable that characterizes a set of categories potentially consisting of more than two categories. The definition of $\mathcal{Z}$ implies two things, at least one category must occur, and two categories cannot concurrently occur, *i.e.* categories in Z are mutually exclusive.

For $z_1 \in \mathcal{Z}$, $\mu_X = E(X)$, and $\mu_{z_{-1}} = E(Z|Z \neq z_1)$ the categorical effect is given by

$$\text{CE}(z_1) = m(\mu_X, E(Z|Z = z_1)) - m(\mu_X, \mu_{z_{-1}}) = m(\mu_X, z_1) - m(\mu_X, \mu_{z_{-1}}).$$

Denote the parameters of the model by $\theta_0 = (\theta_{0,1}, \dots, \theta_{0,K})'$ with $K \geq K_X + K_Z$. We henceforth explicitly write CE and m as functions of the parameters of the model, in particular $\text{CE}(\cdot; \theta_0)$ and $m(\cdot, \cdot; \theta_0)$.

Estimation of $\text{CE}(z_1; \theta_0)$ is obtained by substituting estimated parameters $\hat{\theta}$ into the CE function, $\text{CE}(z_1; \hat{\theta})$. A feasible version of $\text{CE}(z_1; \hat{\theta})$ is obtained by substituting μ_X and $\mu_{z_{-1}}$ with appropriate estimators. More precisely,

$$\widehat{\text{CE}}(z_1; \hat{\theta}) = m(\hat{\mu}_X, z_1; \hat{\theta}) - m(\hat{\mu}_X, \hat{\mu}_{z_{-1}}; \hat{\theta}). \tag{1}$$

Finally consistent estimators for μ_X and $\mu_{z_{-1}}$ are obtained as follows

$$\hat{\mu}_X = \frac{1}{n} \sum_{i=1}^{n} X_i'$$

and

$$\hat{\mu}_{z_{-1}} = \frac{1}{\sum_{i=1}^{n} \mathbb{1}(Z_i \neq z_1)} \sum_{i=1}^{n} \mathbb{1}(Z_i \neq z_1) Z_{i,-1}'$$

where $\mathbb{1}$ is the indicator function and $Z_{i,-1} = (Z_{i,2}, \ldots, Z_{i,K_Z})$.

3 List of Assumptions

For ease of reference, in this section we summarize the assumption used throughout this paper.

A1: The parameter θ_0 belongs to a convex and compact parameter space $\boldsymbol{\Theta}$;

A2: The model characterized by m is identified;

A3: The estimator $\hat{\theta}$ is consistent and $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{\text{d}} \text{N}(0, \mathbf{V})$;

A4: The function of interest, m is differentiable with respect to θ; and

A5: The first derivative of m with respect to θ is a continuous function of $\theta \in \boldsymbol{\Theta}$, $x \in \mathbb{R}^{K_X}$ and $z \in \mathcal{Z} = \bigcup_{j=1}^{K_Z} \{\mathbf{i}'_{j,K_Z}\}$.

Assumptions **A1** and **A5** can be relaxed to weaker assumptions, specifically

A1*: The parameter θ_0 belongs to a parameter space $\boldsymbol{\Theta}$, such that, there exits some open and convex set $\mathcal{O} \subset \boldsymbol{\Theta}$ with $\theta_0 \in \mathcal{O}$; and

A5*: The first derivative of m with respect to θ is a measurable function of $\theta \in \boldsymbol{\Theta}$, $x \in \mathbb{R}^{K_X}$ and $z \in \mathcal{Z} = \bigcup_{j=1}^{K_Z} \{\mathbf{i}'_{j,K_Z}\}$.

The convexity assumption in **A1** or local convexity in **A1*** together with consistency of $\hat{\theta}$ in **A3** imply that, for large enough sample size n, the derivative of m with respect to θ is well defined at $\alpha\hat{\theta} + (1-\alpha)\theta_0$ for all $\alpha \in [0, 1]$. The compactness in **A1** has to do with the existence of a maximum and a minimum, for the cases where $\hat{\theta}$ belong to the class M-estimators. Assumptions **A1** or **A1*** are usually already implied in the case of M-estimators.

Assumption **A1** or **A1*** combined with **A4** is what allow us to use the mean value theorem on CE. **A3** is usually the main result of most papers proposing parametric estimators. It is also what allow us to conclude asymptotic normality of the leading term of CE when appropriately normalized and centered. **A5** or **A5*** enables us to conclude that the derivative of m, when evaluated at consistent estimators of θ_0, μ_X, and $\mu_{z_{-1}}$ converge in probability. That, together with **A3** is essential to establishing asymptotic normality of $\widehat{\mathrm{CE}}$ when appropriately normalized and centered.

4 Inference and Hypothesis Test

The asymptotic distribution of CE rely heavily on the mean value theorem, sometimes referred to as the delta method. Provided that **A1*** holds, then for any $\theta \in \mathcal{O} \subset \mathbf{\Theta}$,

$$\frac{\partial}{\partial \theta'}\mathrm{CE}(z_1;\theta) = \frac{\partial}{\partial \theta'}m(\mu_X, z_1;\theta) - \frac{\partial}{\partial \theta'}m(\mu_X, \mu_{z_{-1}};\theta).$$

Assumption **A4** enables use of the mean value theorem, in particular for any $\theta \in \mathcal{O} \subset \mathbf{\Theta}$,

$$\mathrm{CE}(z_1;\theta) - \mathrm{CE}(z_1;\theta_0) = \frac{\partial}{\partial \theta'}\mathrm{CE}(z_1;\tilde{\theta})(\theta - \theta_0) \tag{2}$$

where $\tilde{\theta} = \alpha\theta + (1-\alpha)\theta_0$ for some $\alpha \in [0,1]$.

Newey and McFadden (1994) provide sufficient conditions for M-estimators to be consistent and root n asymptotically normal. M-estimators a wide class of estimators, such as maximum likelihood (ML) and generalized method of moments (GMM) estimators. Thus assumption **A3** entails that $\hat{\theta}$ is a consistent estimator of θ_0,

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{\mathrm{d}} \mathrm{N}(0, \mathbf{\Sigma}).$$

Assumption **A5*** implies that

$$\frac{\partial}{\partial \theta'}\widehat{\mathrm{CE}}(z_1;\hat{\theta}) \xrightarrow{\mathrm{p}} \frac{\partial}{\partial \theta'}\mathrm{CE}(z_1;\theta_0)$$

provided that $\hat{\mu}_X$ and $\hat{\mu}_{z_{-1}}$ are consistent estimators (*e.g.*, Davidson, 1994; White, 2001). Consistency of $\hat{\theta}$ also implies that $\alpha\hat{\theta} + (1-\alpha)\theta_0$ is consistent for all $\alpha \in [0,1]$. Thus plugging $\hat{\theta}$ into equation (2) leads to the following asymptotic result,

$$\sqrt{n}(\widehat{\mathrm{CE}}(z_1;\hat{\theta}) - \mathrm{CE}(z_1;\theta_0)) \xrightarrow{\mathrm{d}} \mathrm{N}(0, \mathbf{V}) \tag{3}$$

where

$$\mathbf{V} = \frac{\partial}{\partial \theta'}\mathrm{CE}(z_1;\theta_0)\mathbf{\Sigma}\frac{\partial}{\partial \theta}\mathrm{CE}(z_1;\theta_0).$$

Given the result in equation (3), researchers may conduct both one tailed or two tailed hypothesis test. In other words, researchers may test, for instance, that the categorical effect of z_1 is positive relative to $\mu_{z_{-1}}$. Alternatively, inference can be conducted to test if categorical effect of z_1 is significant relative to $\mu_{z_{-1}}$.

We have, therefore shown that in terms of econometric asymptotic theory, the choice of where categorical effect is evaluated is researchers' prerogative. They should make the decision of where to evaluate the measurement of categorical effect based on the economics of their research questions. Our proposed points of evaluation allow researchers to estimate the effect of a category relative to the others, globally. The existing sub-sample approach that compared the funtional of interest at $E((X, Z)|Z = z_1)$ to the same function at $E((X, Z)|Z \neq z_1)$ eliminates some of what applied research sometimes call the "controlling" aspect of the calculation of categorical effect.

For instance, suppose X is a random variable describing years of education, and Z a K_Z dimensional row vector of dummy variables describing race. Let, for instance, z_1 be the row vector that is one in the coordinate associated with white and zero for all remaining races. If whites receive on average higher income, then the effect calculated by the difference of the functional of interest at $E((X, Z)|Z = z_1)$ and $E((X, Z)|Z \neq z_1)$ respectively, will not only be capturing the effect of white versus non whites, but also the effect of an average higher years of education.

The aforementioned existing approach is not wrong, nor is it inadequate. There are research questions that are best answered by the existing approach. However, if a researcher wishes to gauge the effect of one category in relation to the others while everything else is left constant, our proposed approach allows researchers to do so. In the aforementioned race and years of education example, our approach allows researcher to gauge what is the categorical effect of white relative to non-white, while leaving years of education constant across races, thus comparing two individuals, a white versus an average non-white individual that are otherwise identical.

5 A logistic regression numerical example

We generate a hypothetical simulated data set with five hundred observations and four variables namely, admission status of students (*admit*) in graduate school, GRE scores of the students (*gre*), their GPA (*gpa*), and the rank (rank) of the undergraduate institution the student went to.[1] In particular, we follow Buis (2007) to generate logit data. The data, a detail description of the data simulation, and a small Monte Carlo simulation study can be found in the supplementary material. Admission status is a binary variable, which is also the dependent variable. GRE scores and GPA of the students are treated as continuous variables, while rank is a categorical variable taking on values 1 through 4. Institutions with a rank of 1 have the highest prestige, while those with a rank of 4 have the lowest. We are interested in how variables, such as GRE (Graduate Record Exam scores), GPA (grade point average) and prestige of the undergraduate institution, effect admission into graduate school. We

[1] Simulation was inspired by `http://stats.idre.ucla.edu/r/dae/logit-regression/`.

estimate the following model:

$$\log\left(\frac{\Lambda(X,Z)}{1-\Lambda(X,Z)}\right) = \alpha_0 + \beta_{1,0} gre + \beta_{2,0} gpa + \gamma_{2,0} Z_2 + \gamma_{3,0} Z_3 + \gamma_{3,0} Z_4,$$

where $\Lambda(X,Z) = P(admit = 1|X,Z)$, $X = (1, gre, gpa)$, $Z = (Z_1, Z_2, Z_3, Z_4)$, and Z_1, Z_2, Z_3 and Z_4 are dummy variables for ranks 1, 2, 3 and 4 respectively. The Logit output parameter estimates are summarized in Table 1 below.

Table 1: Logit Output

	Estimate	Std. Error	Z statistic	P-Value
gre	0.004	0.001	3.870	0.000
gpa	0.701	0.324	2.160	0.031
Rank 2	-0.817	0.302	-2.710	0.007
Rank 3	-1.990	0.378	-5.270	0.000
Rank 4	-1.645	0.505	-3.250	0.001

After estimating the model using logistic regression, we calculate categorical effects for each category of the variable rank using the built in packages in Stata and R. Table 2 below summarizes the estimated means of each variable of the model conditional on each rank. Unconditional mean estimates are reported in the row of Table 2 labeled "Overall."

Table 2: Conditional Mean Estimates

Rank	*gpa*	*gre*	*admit*
1	3.589	643.671	0.658
2	3.345	594.359	0.382
3	3.017	572.201	0.128
4	2.715	506.931	0.115
Overall	3.166	579.670	0.296

For illustration purposes, we specifically focus on the categorical effect of "rank 2" on "*admit*." The packages in these software calculate categorical effects by estimating the following expressions,

$$\Lambda(E(X), \mathbf{i}'_{2,4}) - \Lambda(E(X), \mathbf{i}'_{1,4}), \text{ and} \tag{4}$$

$$\Lambda(E(X), E(Z_1), 1, E(Z_3, Z_4)) - \Lambda(E(X), E(Z_1), 0, E(Z_3, Z_4)) \tag{5}$$

respectively. Henceforth, we refer to method 1 and 2 as estimation of equations (4) and (5) respectively. Table 3 reports the categorical effects reported by the Stata and R built in functions,[2] which estimate (4) and (5), as well as the

[2]For Stata the "margins" command calculates the categorical effects. For R, the package we used is called "mfx".

restricted model characterized by

$$\log\left(\frac{\Lambda(X,Z)}{1-\Lambda(X,Z)}\right) = \alpha_0 + \beta_{1,0}gre + \beta_{2,0}gpa + \gamma_{2,0}Z_2,$$

and our proposed methodology. As is clear from the table, the different approaches produce different results for the categorical effect.

Table 3: Categorical Effect Estimates

		Categorical Effects	Standard Error	Z Statistic	P-Val
Proposed	Rank 1	0.320	0.075	4.288	0.000
	Rank 2	0.119	0.048	2.463	0.014
	Rank 3	-0.194	0.041	-4.679	0.000
	Rank 4	-0.089	0.063	-1.405	0.160
Method 1	Rank 2	-0.198	0.073	-2.691	0.007
	Rank 3	-0.396	0.077	-5.155	0.000
	Rank 4	-0.351	0.097	-3.627	0.000
Method 2	Rank 2	-0.087	0.028	-3.101	0.002
	Rank 3	-0.170	0.022	-7.820	0.000
	Rank 4	-0.171	0.032	-5.318	0.000
Restricted	Rank 1	0.249	0.071	3.513	0.000
	Rank 2	0.050	0.044	1.142	0.253
	Rank 3	-0.191	0.040	-4.727	0.000
	Rank 4	-0.030	0.072	-0.419	0.675

A procedure based on equation (4) estimates the effect of "rank 2" relative to "rank 1" on admittance likelihood of an average applicant. Procedures based on the restricted model have been advocated informally in many graduate econometric courses. However, both estimation and inference is potentially negatively affected by misspecification. Finally, our approach of estimating

$$\Lambda(E(X), E(Z|Z = \mathbf{i}'_{2,4})) - \Lambda(E(X), E(Z|Z \neq \mathbf{i}'_{2,4}))$$

allows researcher measure the global effect of "rank 2" on admittance likelihood of an average applicant. In this context, global effect means it is relative to the average ranks different from 2.

As reported in Table 3, the categorical effect calculated using our proposed method is not only approximately 58% higher than that obtained from the restricted model, the P-Values are substantially different too. Therefore, not only the coefficients are different for the two models, the researcher can potentially end up with completely different inferences based on which approach is used.

6 Monte Carlo

We conduct a small Monte Carlo study to shed light on the finite sample performance of each method. In particular we calculate the mean Categorical Effect estimate, Bias, and Mean Squared Error (MSE) based on 1,000 replications of our numerical example. We consider two data generating processes (DGP), namely DGP 1 and 2 respectively. In DGP 1 we simulate "*gpa*" and "*gre*" with means dependent on "rank." Alternatively, DGP 2 was simulated with independent regressors. More detailed description of our Monte Carlo study can be found in the supplementary material.

We summarize our Monte Carlo results in Table 4. It is important to note that method 1 and 2 estimate different objects to the proposed and restricted model methods, hence the significantly different mean Categorical Effect across them. However, using methods 1 and 2 as performance benchmarks, note that our proposed approach performs similar to methods 1 and 2. Finally, our proposed method and the restricted model method attempt to estimate the same object. Our proposed method outperforms the restricted model approach, especially when the regressors in the model are dependent. With independent regressors the performance of the restricted model method considerably improved, although still underperformed when compared to our proposed approach. In summary, our method outperforms and is more robust than the restricted model method.

Table 4: Monte Carlo Simulation Results

		DGP 1			DGP 2		
	Rank	Mean Categorical Effect	Bias	MSE	Mean Categorical Effect	Bias	MSE
Proposed	Rank 1	0.243	0.006	0.006	0.246	0.002	0.004
	Rank 2	0.092	0.005	0.003	0.094	0.003	0.002
	Rank 3	-0.117	0.001	0.002	-0.124	0.001	0.002
	Rank 4	-0.132	-0.002	0.004	-0.138	0.000	0.003
Method 1	Rank 2	-0.145	-0.001	0.005	-0.145	0.001	0.005
	Rank 3	-0.282	-0.004	0.006	-0.288	-0.001	0.004
	Rank 4	-0.314	-0.005	0.010	-0.321	-0.001	0.005
Method 2	Rank 2	-0.070	0.002	0.001	-0.074	0.002	0.001
	Rank 3	-0.136	0.002	0.001	-0.145	0.001	0.001
	Rank 4	-0.173	0.003	0.002	-0.187	0.002	0.001
Restricted	Rank 1	0.173	-0.064	0.052	0.238	-0.006	0.058
	Rank 2	0.035	-0.052	0.003	0.080	-0.011	0.002
	Rank 3	-0.111	0.007	0.022	-0.135	-0.010	0.023
	Rank 4	-0.086	0.044	0.035	-0.147	-0.009	0.029

7 Final Remarks

In this paper we extend Anderson and Newell (2003) by proposing a method of globally estimating marginal effects of categorical variables with potentially more than two categories for general parametric models. Prior to this paper, existing methods focused on pairwise categorical or sub-sample comparison. Both pairwise and sub-sample comparison methods have many valid research applications. However, our proposed method will be more adequate whenever researchers wish to globally gauge the categorical effect of one category *ceteris-paribus*.

Simple algebra show that in some models, such as the logit specification without iterations associated with the categorical variables, the sign of the pairwise categorical effects will coincide with the sign of the estimated coefficients. However, in our approach, that is not necessarily the case. In particular, in our empirical example, "rank 2" had negative coefficient estimate larger than the coefficients of "ranks 3" and "rank 4" respectively. Consequently, "rank 2" categorical effect is negative relative to "rank 1," but positive categorical effect relative to "ranks 3" and "rank 4." Using our approach we conclude that "rank 2" had positive categorical effect globally, *i.e.* relative to "rank 1," "ranks 3" and "rank 4" combined. Thus illustrating that our method and the existing methods are suitable to answering different research questions.

Finally, we compare our approach to the informally advocated restricted model approach. Because of the variation of conditional mean estimates of *gpa* and *gre* across ranks, results associated with the restricted model were considerably different from our approach. In particular, for certain confidence levels, significance differs between the proposed and the restricted model approach. Furthermore, the magnitude of the proposed categorical effect of "rank 2" was approximately 58% higher than the restrictive model method. In light of the Frisch-Waugh-Lovell theorem associated with the classical linear regression model (*e.g.*, Davidson and MacKinnon, 2004), we conduct a small Monte Carlo study with two distinct DGPs. One DGP was simulated such that data displays *gpa* and *gre* mean dependence across ranks, *i.e.* conditional mean variation across ranks. In the second DGP we simulate *gpa* and *gre* independent of rank. Simulation shows that our method outperform the restricted model approach. Furthermore, performance of the restricted model approach is considerably worse when regressors are dependent, whereas our proposed method is robust to dependence across regressors.

References

Anderson, S. and Newell, R. G. (2003). Simplified marginal effects in discrete choice models, *Economics Letters* **81**(3): 321–326.

Buis, M. (2007). Stata tip 48: Discrete uses for `uniform ()`, *The Stata Journal* **7**(3): 434–435.

Davidson, J. (1994). *Stochastic limit theory: An introduction for econometricians*, Oxford University Press, New York, NY.

Davidson, R. and MacKinnon, J. G. (2004). *Econometric theory and methods*, Oxford University Press, New York, NY.

Newey, W. and McFadden, D. (1994). Large sample estimation and hypothesis testing, *in* D. McFadden and R. F. Engle (eds), *Handbook of Econometrics*, Vol. IV, Elsevier, North-Holland, Amsterdam, pp. 2111–2245.

White, H. (2001). *Asymptotic Theory for Econometricians: Revised Edition*, Academic Press, San Diego, CA.

Williams, R. (2012). Using the margins command to estimate and interpret adjusted predictions and marginal effects, *Stata Journal* **12**(2): 308–331.