THE UNIVERSITY OF SYDNEY

Economics Working Paper Series

2012 - 9

Estimation of Public Goods Game Data

Danielle Merrett

April 2012

Estimation of Public Goods Game Data*

Danielle Merrett†

April 2012

Abstract

This paper compares the performance of alternative estimation approaches for Public Goods Game data. A leave-one-out cross validation was applied to test the performance of five estimation approaches. Random effects is revealed as the best estimation approach because of its un-biased and precise estimates and its ability to estimate time-invariant demographics. Surprisingly, approaches that treat the choice variable as continuous out-perform those that treat the choice variable as discrete. Correcting for censoring is shown to induce biased estimates. A finite Poisson mixture model produced relatively un-biased estimates however lacked the precision of fixed and random effects estimation.

Keywords: public goods, voluntary contributions mechanism, economic experiments, random effects, fixed effects, ordered logit, finite mixture models

*I am grateful to David Drukker, Director of Econometrics at StataCorp, to Professors Adrian Pagan and Robert Slonim for their valuable suggestions and comments and to Professor Simon Gächter for providing the data cited in this paper.

†School of Economics, University of Sydney, NSW 2006

1 Introduction

The Public Goods Game is extensively used by experimental economists[1] as a tool to study social dilemmas and cooperation. However, even though it has been nearly 30 years since the first laboratory Public Goods Game experiments were published (Isaac, Walker, and Thomas, 1984; Kim and Walker, 1984; Isaac, McCue, and Plott, 1985) the empirical analysis of the game choice data has still not moved beyond descriptive statistics in most papers. The likely reason for this is that the distribution of the choice data for this game is highly non-standard and is complicated by its discrete, censored and often panelled nature. With little known about the preciseness or extent of biasedness of estimates for this data under these conditions, many authors have avoided model estimation entirely.

There have been a few exceptions though. Carpenter (2004) for example, used Tobit random effects estimation to account for data censoring to model contribution choice in a 10 period public goods game. Bardsley and Moffatt (2007) made a clear attempt at advancing the analytical toolbox for public goods experiments by proposing that public goods data be modelled using a finite mixture model to incorporate heterogeneity of types within a population with Tobit components to address censoring, and a tremble term to model decision error. Despite the sophistication of the model and compelling rational for the approach, the finite mixture modelling approach was never taken up in the Public Goods experimental literature, probably due to its complexity.

Random effects estimation have been used by (Tan and Bolle, 2007; Nikiforakis, 2010) and more recently, Breitmoser (2010) estimates a nested ordered logit using Public Goods data in order to compare the internal and external validity of different structural models. This paper differs from Breitmoser (2010) in that this paper is specific to Public Goods experiments only and in this paper the structural model is held constant and the performance of the estimates are compared for different estimation approaches. In contrast, Breitmoser (2010) holds neither the structural model nor estimation approach constant. Different specified models with different estimation approaches are compared using Bayes Information Criteria (BIC) and log likelihoods producing somewhat idiosyncratic results. In line with the results of this paper, Poen (2009) finds evidence of bias in Tobit random effects estimates from simulated public goods game data. However, Poen (2009) suggests the bias is likely due to the inclusion of a feedback variable that may introduce endogeneity. This

[1] For example, a review paper by Chaudhuri (2011) cites 146 Public Goods experiment publications.

paper decisively shows, by estimating a model including a feedback variable with and without using a tobit approach, that the source of the bias is not endogeneity but instead from the use of Tobit estimation.

This paper provides Public Goods Game experimentalists with a clear evidence-based prescription for the best estimation approach for Public Goods Game choices. With greater knowledge and certainty as to how different estimators will perform with Public Goods Game data, it is hoped that authors will be more confident in generating inferences from Public Goods Game models.

2 Distribution of Contributions

2.1 Public Goods Game

I examine a typical public goods situation found in experimental economics literature, a standard Voluntary Contributions Mechanism (VCM) (Davis and Holt, 1993; Ledyard, 1995). Participants have the same endowment w and are in groups of N. Each individual has to decide how much of his endowment to allocate to a public account y_i and how much to keep for himself $w - y_i$. For each group, the sum of the individual allocations to the public good $\sum_{j=1}^{N} y_j$ is then multiplied by a factor $a(N > a > 1)$, to model the additional value generated from the public nature of the good. The final value of the public account is then shared equally among the group members. The payoff therefore of player i under a VCM is given by:

$$\pi_i = (w - y_i) + \frac{a}{N} \sum_{j=1}^{N} y_j$$

The VCM is primarily used to model social dilemmas because the dominant strategy for each player is to free ride by allocating nothing to the public account (assuming players maximize their own monetary payoff and rationality is common knowledge). However, maximum efficiency is achieved when all members allocate their entire endowment to the public account $y_i = w$.

2.2 Data

Data was sourced from a previous study that used a 10 period public goods experiment (Guillen, Merrett, and Slonim, 2012). This study used procedures and instructions that closely resembling those from previous literature[2] The data set is a panel of 4000 observations from 400 subjects

[2]The instructions used in the study were adapted from Herrmann, Thöni, and Gächter (2008)

with each subject making 10 contribution decisions. The Guillen, Merrett, and Slonim (2012) study recruited undergraduates from the University of Sydney, Australia (undergraduates are typically recruited as subjects for Public Goods Game experiments in the literature) and involved two stages. In the first stage all subjects played a standard 10 period VCM game and in the second stage subjects were re-matched into different groups and played a variety of different 10 period public goods games. In this paper only the data from the first stage standard 10 period VCM game is used. Subjects played in groups of $N = 4$ and were given an endowment $w = 100$ cents in which to make a contribution decision $y = \{0, 1, 2...100\}$. The experimenters multiplied contributions by a factor of $a = 2$ thereby giving a marginal per capita return (MPCR) equal to 0.5 for every cent contributed.

2.3 Distributions

The contribution data replicate the temporal results of earlier VCM experiments (Ledyard, 1995) where mean contributions start between 40 to 60 percent of the endowment and decline to close to zero (Figure 1). The decay in contributions in the standard VCM game has been replicated many times by different authors and is observed across different cultures (Gächter, Herrmann, and Thöni, 2010).

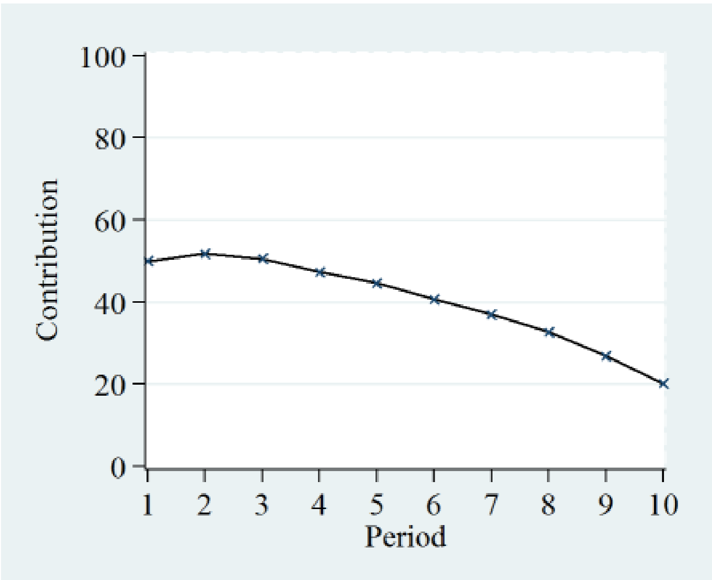

Figure 1: Average contributions by period replicate the results of previous experiments ($N = 4000$)

The distributions of contributions across all 10 periods is given in Figure 2. The distribution of contributions is similar to those obtained from other VCM studies, see for example, Gächter, Renner, and Sefton (2008)[3] and Herrmann, Thöni, and Gächter (2008)[4] (See Fig. 3). The distributions in Fig. 2 and Fig. 3 are both highly truncated (in Fig. 2 almost 40 percent of observations are at one of the two limits), more so at the zero end than the 100 end, and display a flat, almost uniform, distribution in between the two limits with a noticeable node at the midpoint. The distributions in Fig. 3 also show noticeable modes between the midpoint and zero and the midpoint and upper bound however the distributions are comparable for the most part.

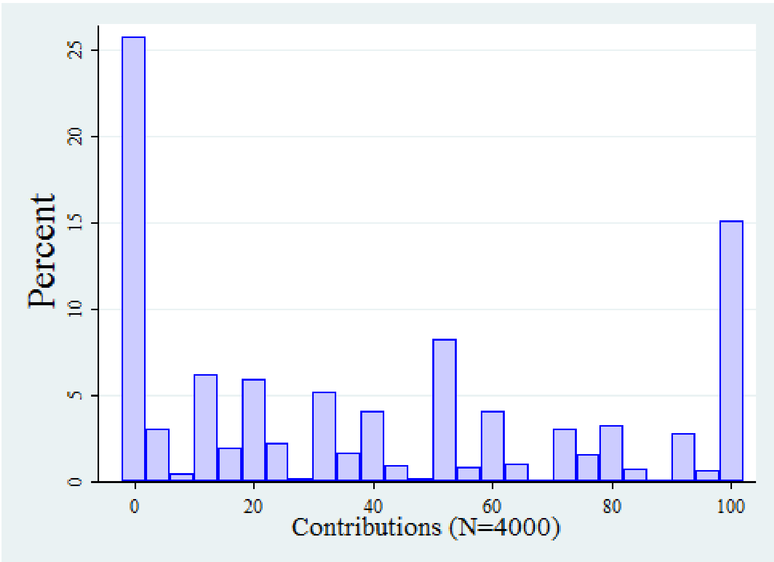

Figure 2: Histogram of contribution data used in this paper (all 10 periods)

Given that standard parameters and standard experimental methods were used and the contribution distribution replicates previous literature, the dataset used in this paper is representative of standard VCM experimental data. The dataset is also substantial involving 400 subjects and 4000 observations. These attributes make this dataset a good candidate to test the performance of different VCM estimation approaches.

The correct identification of the distribution becomes particularly important when using maximum likelihood estimation (MLE). Whereas

[3]In this study subjects played in groups of 3, were given an endowment of 20 tokens and received a MPCR of 0.5.

[4]This study collected contribution data from subjects in several different countries around the world including Melbourne, Australia. Subjects played in groups of 4, were given an endowment of 20 tokens and received a MPCR of 0.4.

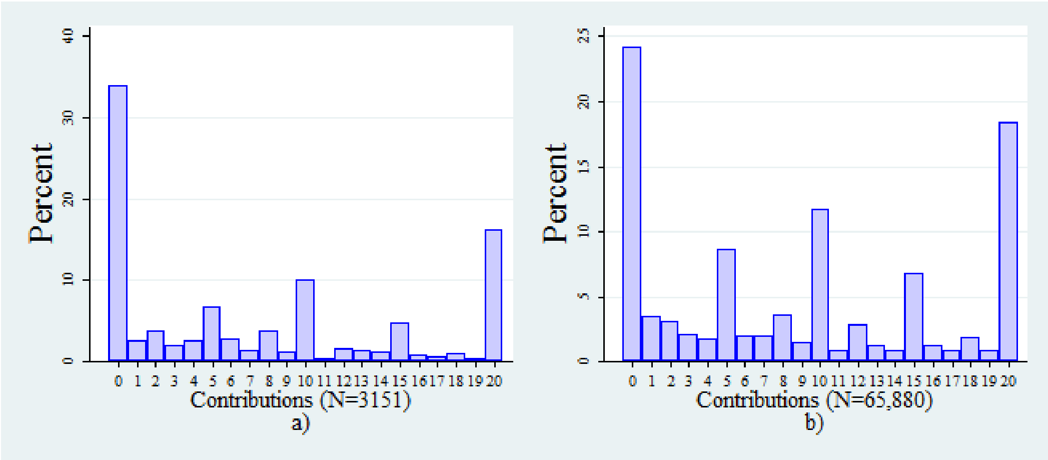

Figure 3: Histogram of contributions in the baseline VCM game condition from a) Gächter, Renner, and Sefton (2008) (all 50 periods); and b) Herrmann, Thöni, and Gächter (2008) (all 10 periods).

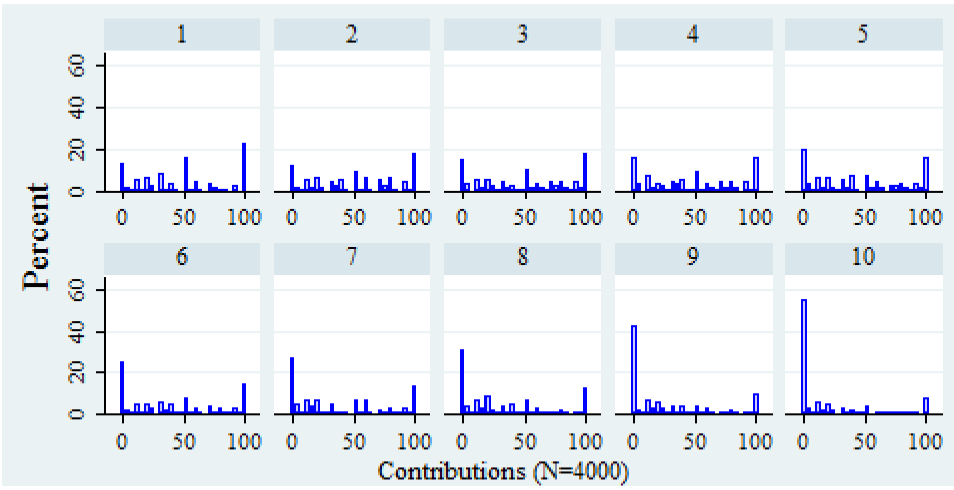

Figure 4: Histogram of contributions by period

least squares estimation only requires that the distribution of the errors be known, MLE requires that the distribution of the dependent variable is known. If the distribution is misspecified then MLE estimates can be inconsistent invalidating standard inference techniques White (1982).

A reasonable assumption might be that contributions are distributed according to a Tobit distribution with lower limit censoring occurring at zero and upper censoring occurring at 100. A simulated Tobit distribution with a mean and variance comparable to that of the dataset illustrates how different the Tobit distribution is compared to the distribution of contributions (Figure 5). Further examination of contribution distributions for each period shows that none of the periods demonstrate a distribution similar to the simulated Tobit distribution (Figure 4).

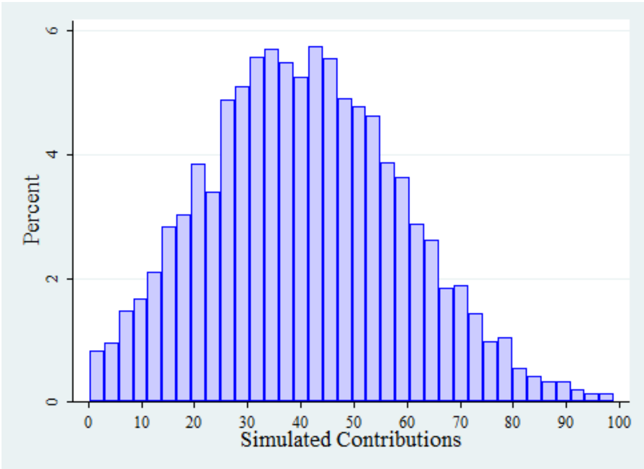

Figure 5: Tobit Simulation N=4000 ($\mu = 40, \sigma^2 = 400$)

3 Estimation Approaches

Given the unique nature of VCM data (discrete, panel, with a non-standard distribution changing over time) many papers avoid the difficult task of estimating models using VCM data, instead choosing descriptive analysis over regression analysis. From those authors who have, there have been a number of different estimation approaches. The estimation approaches can be grouped into two main categories: the continuous approach (as with generalized least squares for panel data models) and the discrete approach (as with logit and finite mixture models). The estimation approaches chosen for comparison are those used in the literature which include, random effects estimation (Tan and Bolle, 2007; Nikiforakis, 2010), tobit random effects estimation (Carpenter, 2004), ordered logit (Breitmoser, 2010), and finite mixture models (Bardsley and Moffatt, 2007).

In order to compare the performance of different approaches, the same model covariates should be estimated in each approach. To be comparable to the literature, the covariates must also reflect the previous findings. A good model is one that is also parsimonious and incorporates the temporal nature of the data. The following model was created with this criteria in mind and is used to compare the performance of different estimation approaches in sections (3.1-3.5).

$$y_{ip} = \beta_0 + \beta_1 p + \beta_2 p9 + \beta_3 p10 + \beta_1 AV_{ip-1} + u_{ip} \tag{1}$$

The dependent variable in Eq. (1) is an individual's i's contribution y in period p. The covariates are lagged Average Group Contribution (AV) which *excludes* individual i's contribution, period (p) and period 9 and 10 dummies (*p9* and *p10* respectively). The intercept is only estimated when appropriate for that approach.

The covariate AV was included in Eq. (1) in order to model the effect of conditional cooperation. There is evidence that people demonstrate conditional cooperation behavior in public goods games (Urs Fischbacher and Fehr, 2001; Keser and van Winden, 2000; Fischbacher and Gächter, 2010). That is, people will contribute to the public account if the others in their group contribute as well. Their contributions therefore, are dependent on the contributions of the others in their group. The covariate p was used to control for the declining trend in contributions over time. To explicitly model any possible end-game effects (Andreoni, 1988) a dummy for each of the last two periods was included.

3.1 Fixed Effects

The fixed effect approach is a method of removing the unobserved individual specific effects from panel data by applying a transformation (2) to the data prior to estimation (Wooldridge, 2009). The transformation uses the mean of the dependent variable $\bar{y}$ and mean of the regressors $\bar{x}$ to time de-mean the data.

$$y_{ip} - \bar{y}_i = \beta(x_{ip} - \bar{x}_i) + u_{ip} - \bar{u}_i \tag{2}$$

The time-demeaned data (time is denoted as period p here) is then regressed using ordinary least squares (OLS). Because the fixed effects approach effectively removes the individual effects it does not require the stricter conditions that random effects estimation (3.2) imposes. One drawback though is that the transformation not only removes the individual effects from the intercept, but removes any time invariant variables, for example, gender and race.

Table 2 reports the estimation results of model (1) using fixed effects estimation. All covariates are significant except *Period 9* and all coefficients have a sign that we would expect. The coefficients are interpreted in the same way as an OLS estimation. *R-Squares* are typically lower in panel models and are less meaningful than those from cross-sectional data. *Rho* reports the correlation coefficient. A correlation coefficient equal to one suggests perfect correlation of contribution choices within the same individual. A *Rho* of 0.498 here suggests suggests that 50 percent of the variance is due to differences between the individuals.

Table 1: Estimation results : Fixed Effects

Variable	**Coefficient**	**(Std. Err.)**
Dependent variable: Contribution		
Period	-2.326***	(0.295)
Period 9	-2.193	(1.395)
Period 10	-4.408**	(1.915)
Lag Average Group Contribution	0.385***	(0.037)
Intercept	37.454***	(2.447)
R^2 (overall)	0.226	
Rho	0.498	

Significance: $***1\%, **5\%, *10\%$
Notes: Robust standard errors clustered at subject id level.

3.2 Random Effects

Random effects estimation does not remove the individual effects but instead allows each individual to have their own random intercept ie, individual effect, Demidenko (2004). The main advantage of using Random effects over fixed effects estimation is that it allows for covariates that are constant over time. However random effects requires the stricter assumptions that the individual effects are uncorrelated with the covariates and that the individual effects are normally distributed in the population. Allowing for individual effects in the data does create serial correlation which is solved using generalized least squares (GLS). GLS eliminates serial correlation by a similar transformation to (2) except that only a proportion of the transformation (λ) is applied in random effects estimation.

$$y_{ip} - \lambda\bar{y}_i = \beta_0(1-\lambda) + \beta_1(x_{ip} - \lambda\bar{x}_i) + (u_{ip} - \lambda\bar{u}_i) \tag{3}$$

The proportion is determined by the strength if the individual effects Eq. (4). If observations within the same individual are highly correlated, then the within individual variation σ_u^2 will be low relative to the between individual variation σ_b^2 and a greater proportion (λ) of the transformation will applied. In the extreme case where $\lambda = 1$, fixed effects estimation is obtained and when $\lambda = 0$, pooled OLS estimation is obtained.

$$\lambda = 1 - \frac{\sigma_u^2}{\sigma_u^2 + P\sigma_b^2} \tag{4}$$

A Hausman test (Hausman, 1978) can be used to help determine whether fixed effects or random effects is the more appropriate estimator to use. The Hausman test involves regressing the model using fixed effects estimation then random effects estimation and the compares whether the estimates are significantly different. If the null hypothesis of no systematic difference is rejected then fixed effects estimation should be used. A rejection of the null suggests that some of the assumptions of random effects estimation have been violated leading to very different results. An LM test (Breusch and Pagan, 1980) can be used to help decide between random effects estimation and a simple OLS regression. If the null hypothesis of no significant differences across individuals is rejected, then OLS estimation should be used.

A Hausman test was applied to model (1) obtaining ($\chi^2 = 40.93, P = 0.000$) suggesting that fixed effects is the appropriate estimator for this model and data. This is congruent with the observed distribution of contributions in section 2 in which is appears that the random effects are distributed non-normally. In the interest of understanding how dire such a violation may be to estimation performance when estimating public goods data, I have included random effects estimation in the presence of violations in this paper. The performance of random effects estimation here can alert experimentalists as to the importance of such a mispecification on estimation results.

Table 2 reports the estimation results of model (1) using random effects estimation. The coefficients are different than those obtained from fixed effects estimation (Table 1). However the difference do not seem dramatic and all covariates display the same signs as those obtained from fixed effect estimation.

Table 2: Estimation results : Random Effects

Variable	**Coefficient**	**(Std. Err.)**
Dependent variable: Contribution		
Period	-2.193***	(0.298)
Period 9 Dummy	-1.986	(1.392)
Period 10 Dummy	-4.013**	(1.902)
Lag Average Group Contribution	0.440***	(0.033)
Intercept	34.233***	(2.438)
R^2 (overall)	0.231	
Rho	0.452	

Significance: $***1\%, **5\%, *10\%$. $N = 3600$

Notes: Robust standard errors clustered at subject id level.

3.3 Tobit Random Effects

Contribution data from VCM public goods games are highly censored (Section 2). Greene (1981) demonstrates that ignoring censoring and proceeding with Least Squares estimation leads to inconsistent and downward biased parameter estimates. A Tobit estimation is sometimes used by authors to address this concern. Statistical packages such as Stata can fit a random effects tobit model by MLE however no statistic exists for tobit fixed effects that would produce un-biased estimates.

Tobit random effects estimation assumes that the random effects, α, are normally distributed. The joint density function is a nested function. The normally distributed random effects nests the tobit distribution of the contributions. The individual level likelihood function is given by

$$l_i = \int_{-\infty}^{\infty} \frac{e^{-\alpha_i^2/2\sigma_\alpha^2}}{\sqrt{2\pi\sigma^2}} \{\prod_{p=1}^{n_i} F(y_{ip}, x_{ip}\beta + \alpha_i)\} d\alpha_i \tag{5}$$

where:

$$F(y_{ip}, \mathbf{x}\beta) = \begin{cases} (\sqrt{2\pi\sigma^2})^{-1} e^{-(y_{ip}-\mathbf{x}\beta)^2/(2\sigma_\epsilon^2)} & \text{if } 0 < y_{ip} < 100 \\ \Phi\left(\dfrac{0-\mathbf{x}\beta}{\sigma_\epsilon}\right) & \text{if } y_{ip} = 0 \\ 1 - \Phi\left(\dfrac{100-\mathbf{x}\beta}{\sigma_\epsilon}\right) & \text{if } y_{ip} = 100 \end{cases}$$

Table 3 reports the estimates of model (1) using Tobit random effects estimation. All variables are reported significant and have the signs that we would expect. Tobit estimates predict the average marginal impact of covariates on the dependent variable in its theoretically true uncensored state. For this reason the estimates of a tobit regression are not directly comparable to an OLS regression, which estimates the marginal effects of the covariates only on the *observed* outcomes. The magnitude of Tobit coefficient estimates are often slightly inflated because of this subtle difference. The significance and signs though are directly comparable. If one wished to directly compare Tobit estimates to OLS estimates this can be done by multiplying the Tobit estimate by the adjustment factor $n^{-1} \sum \Phi(x_i\hat{\beta}/\hat{\sigma})$. Tobit regression does not have an R-squared that can be calculated in the same way as those of OLS regression.

Table 3: Estimation results : Random Effects Tobit

Variable	Coefficient	(Std. Err.)
Dependent variable : Contribution		
Period	-2.854***	(0.357)
Period 9 Dummy	-5.585**	(2.460)
Period 10 Dummy	-12.526***	(2.738)
Lag Average Group Contribution	0.664***	(0.040)
Intercept	25.976***	(3.466)
Rho 0.539	left-censored observations	900
	right-censored observations	508

Significance: $***1\%, **5\%, *10\%$. $N = 3600$
Notes: Robust standard errors clustered at subject id level.

3.4 Ordered Logit Regression

An ordered logit model fits an ordinal categorical dependent variable on a set of independent variables. This estimation approach allows us to compare the performance of an estimation technique that treats the dependent variable as discrete as apposed to continuous. All the other approaches presented in this paper have assumed the dependent variable as continuous. The results of which approach, discrete or continuous, provides better estimates for VCM data may resolve some debates within the experimental economics community on the issue.

An ordered logit was used on the contribution data instead of multinomial logit regression (MLR) because ordinal information is lost in MLRs which disregards the ordinal nature of the categories. Even though there are 101 possible choices in the contribution set $y = \{0, 1, 2...100\}$, we only observe 75 different contribution choices in the dataset. Therefore a model of 75 categories representing each observed contribution is fit from the data. I chose not to reduce the number of contribution categories into contribution intervals because this would be difficult to compare the predictive performance of the ordered logit to the other estimation approaches.

Table 4 shows the estimation results of fitted ordered logistic model. A standard interpretation for the *Lag Average Group Contribution* coefficient is that for every one unit increase in the *Lag Average Group Contribution*, the ordered log-odds of being in a higher contribution category would increase by 0.031 on average, holding other variables constant. The estimated cut off points can be used to find the probability of an individual choosing a particular contribution category. These were estimated (output excluded) and were used to predict contribution choices from the model (in

Section 4).

Table 4: Estimation results : Ordered Logit

Variable	Coefficient	(Std. Err.)
Dependent variable: Contribution choice		
Period	-0.093***	(0.017)
Period 9 Dummy	-0.186**	(0.082)
Period 10 Dummy	-0.483***	(0.116)
Lag Average Group Contribution	0.031***	(0.003)

Significance: $***1\%, **5\%, *10\%$
Notes: 75 cut-points were estimated (output excluded). Robust standard errors clustered at subject id level.

3.5 Finite Mixture Models

The previous estimation approaches assumed that contributions were generated from the same decision making process. Finite mixture models can be used to relax this assumption and explicitly model a finite number of different decision making process (McLachlan and Peel, 2000; Harrison and Rutstrom, 2009; Bardsley and Moffatt, 2007). Under a finite mixture model, agents can be categorized into one of a finite number of groups. A mixture density function is formed by aggregating the category k densities so that:

$$f(y_i; \Psi) = \sum_{k=1}^{g} \pi_k f_i(y_i; \theta_i)$$

With the constraint that $\sum_{k=1}^{g} \pi_k = 1$.

To demonstrate how well a mixture density can fit the observed distribution of the data, an adhoc mixture of a uniform (rounded to the nearest 10) and a discrete distribution was simulated (Figure 6). In the adhoc mixture the parameters were not estimated but simply calibrated to reflect the observed distribution. As you can see, it is easy to find an adhoc finite mixture that fits the data very well. The challenge though, is to estimate a finite mixture model that nests a predictive structural model of contributions.

As the contribution data is comprised of discrete non-negative integers, the Poisson distribution was chosen as the component densities for the finite mixture model because it models the probability of an occurrence of a discrete non-negative integer. The component distributions need not have the same variance but do need to belong to the same family of distributions. In a finite mixture model (FMM), the number of

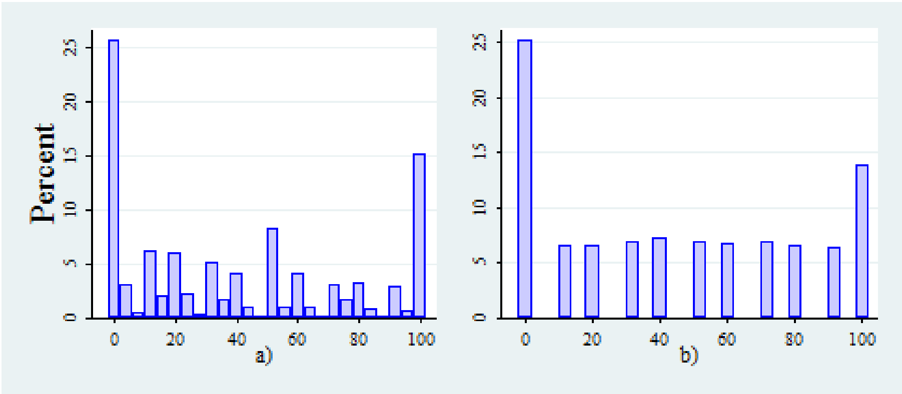

Figure 6: a) Distribution of average contributions in the 10 period Public Goods Game. b) Simulated distribution (N=500) containing components of Binomial ($p_1 = 0.30$) and Uniform distributions with ad-hoc component parameters ($\pi_1 = 0.25, \pi_2 = 0.75$) respectively.

components are chosen *a priori*. Theory can help guide the choice of components however the choice is largely subjective. Once the number of components is decided the component moments and proportions π_i can then be estimated simultaneously with the coefficients of the structural model. Four components were chosen for the FMM estimated in this paper which replicates the number of components estimated by Bardsley and Moffatt (2007). They chose four types based on the theory that there are four contributor types: Free Riders, Altruists, Strategists and Reciprocators.

Because it is easy to over-fit FMMs it is useful to use the Bayes information criteria (BIC) (Schwarz, 1978) to ensure you haven't over-fit the number of components. These statistics are available post-estimation and is a criteria that incorporates a tradeoff between fit and parsimony. Smaller BICs are preferred as they are associated with higher log likelihoods. BIC is used over Akaike information criteria (AIC) here because there is a tendency for AIC to over-fit models in large sample sizes (Hurvich and ling Tsai, 1989). Table 5 reports the BIC values for FMM of 2 to four components. The four component model has the lowest BIC and is therefore preferred over the other two.

Table 5: Goodness-of-fit criteria

Model	**Obs**	**df**	**BIC**
2-Components	3600	9	52897.36
3-Components	3600	14	35934.83
4-Components	3600	19	30913.12

Estimation of FMM can be achieved through GLS, MLE or the Expectation Maximization algorithm. The FMM estimated in this paper was done in Stata through a user-written `fmm` command by Deb (2007) that enables MLE of FMMs using many of the standard distributions. To understand the construction of the grand log likelihood function for the FMM estimated in this paper first consider the likelihood function for a single Poisson distribution:

$$\ell = \prod_{j=1}^{m} \frac{e^{-y_j(\theta' x_j)} e^{-e^{\theta' x_j}}}{y_j!}$$

The log likelihood is

$$\ln\ell = \sum_{j=1}^{m} \Big(y_j(\theta' x_j) - e^{\theta' x_j} - \ln y_j! \Big)$$

The grand likelihood is constructed from four component Poisson likelihood functions

$$\text{Grand } \ell = \prod_{i=1}^{N} (\pi_1\ell_1 + \pi_2\ell_2 + \pi_3\ell_3 + \pi_4\ell_4)$$

Taking the log

$$\text{Grand } \ln\ell = \sum_{i=1}^{N} \ln(\pi_1\ell_1 + \pi_2\ell_2 + \pi_3\ell_3 + \pi_4\ell_4) \tag{6}$$

The estimates for each component probability π_k are constrained to be between 0 and 1 and sum to 1, using a post-estimation log-odds transformation (Harrison, 2007).

The mean contributions and standard deviations for the four estimated components of Equation 6 are reported in Table 6. The first component, or group of contributors, contribute an average of almost 0 cents out of 100 per period. These members could be classified as "free riders". The second group of members contribute on average 90 cents out of 100, these could be classified as "altruists". The third group contribute on average 50 cents out of 100 and their contributions vary during the 10 periods (Std. Deviation is 5.6). The fourth group contributes an average of 16 cents out of 100 per period.

Table 7 reports the estimates of the FMM which nests model (1). The first thing to note is that 28 percent of the individuals in the dataset could be classified as free riders ($\pi_1 = 0.278$). The other other components are fairly evenly proportioned. The end game effect only seems to be significant

Table 6: Summary statistics

Variable	Mean	Std. Dev.
Component 1	0.123	0.042
Component 2	89.645	2.683
Component 3	46.354	5.609
Component 4	16.021	2.841
Total N=3600		

for component 1 members (the free riders) with the *Period 10* dummy significant at the 5 percent level. The *Lag Average Group Contributions* significantly affect the contributions of individuals in all contributor groups except the free riders (component 1). However the altruists (component 2) are the only group to have a signifcant downward trend in contributions. This result is interesting as the aggregate distributions demonstrate a clear downward trend in contributions. It is suprising to observe the temporal decline is insignificant for most individuals when *Lag Average Group Contributions* are controlled for. One explanation is that the altruists are the instigators of the decline which is magnified by *Lag Average Group Contributions.*

4 Comparing the Performance of Different Estimation Approaches

A cross validation method (Stone, 1974) was used to measure the predictive accuracy of each model. In the leave-one-out cross validation (LOOCV) method, one observation is removed from the data set and used as the test observation. The model is then fit from the remaining data. The value of the test observation is predicted from the fitted model and the predicted residual is calculated from the fit. This is repeated for each observation in the data set and the Mean Squared Error (MSE) Eqn. (7) is calculated from the resulting residuals. The model that has the greatest predictive accuracy is the one with the lowest MSE. The LOOCV method is used as it is a more efficient use of the data than a leave-k-out cross-validation in which k observations are left out at each step.

$$MSE = \sum_{i=1}^{n} (y - \hat{y})^2 / n \qquad (7)$$

Where $(y - \hat{y})$ is the difference between the observed and predicted contribution also known as the residual.

Table 7: Estimation results : Finite Mixture Poisson Model

Variable	Coefficient	(Std. Err.)
Dependent Variable: Contribution		
Equation 1 : component 1		
Period	0.009	(0.087)
Period 9	-0.063	(0.328)
Period 10	-0.975**	(0.454)
Lag Average Group Contribution	-0.011	(0.008)
Intercept	-1.617**	(0.652)
Equation 2 : component 2		
Period	0.007**	(0.003)
Period 9	0.001	(0.017)
Period 10	-0.021	(0.021)
Lag Average Group Contribution	0.001***	(0.000)
Intercept	4.408***	(0.029)
Equation 3 : component 3		
Period	0.002	(0.008)
Period 9	0.055	(0.051)
Period 10	-0.005	(0.060)
Lag Average Group Contribution	0.005***	(0.001)
Intercept	3.617***	(0.062)
Equation 4 : component 4		
Period	0.004	(0.015)
Period 9	0.096	(0.079)
Period 10	-0.095	(0.102)
Lag Average Group Contribution	0.006***	(0.001)
Intercept	2.464***	(0.114)
Proportion: π_1		
	0.278***	(0.012)
Proportion: π_2		
	0.262***	(0.016)
Proportion: π_3		
	0.251***	(0.013)
Proportion: π_4		
	0.204***	(0.012)

Significance: $***1\%, **5\%, *10\%$
Notes: Robust standard errors clustered at subject id level.

The predicted contributions for the ordered logit, are given by the contribution category with the highest probability, conditional on the leave one out covariates. There are some contribution categories that are only observed once in the dataset. These observations were dropped before the LOOCV for the ordered logit approach because if they were used as a test observation, their contribution category would not be represented in the fitted set, thus inducing zero probability of an accurate prediction. There were 25 contribution choices that were only observed once reducing the number of categories for which the ordered logit was fit to 51 and the number of total observations to 3575.

The results of the LOOCV for each estimation approach are displayed in Table 8. Random effects estimation had the lowest MSE therefore has the highest predictive performance of the approaches examined here. However its performance was only negligibly better than the fixed effects and tobit random effects estimation. The worst performer (by far) was the ordered logit. Surprisingly, the finite mixture model, whose Poisson distributions most closely resembled the distribution of the data, came second last. The mean error (ME) measures estimation bias by sign and magnitude and is the mean of the residuals. MEs suggest that fixed effect estimates are un-biased and random effects estimates are infinitesimally biased. Tobit random effects estimation produces the most biased estimates. The positive bias is most likely due to the mis-specification of the distribution of contributions (Fig. 2) as a Tobit distribution and to tobit estimates predicting the latent un-censored variable as apposed to the observed censored variable (as discussed in Section 3.3).

The estimation approaches that treat the dependent variable as continuous (random effects, fixed effects and tobit random effects) clearly out-perform the estimation approaches that treat the dependent variable as discrete. One explanation could be that the larger MSEs in the discrete approaches are simply due to rounding to the nearest integer and that these rounding errors are magnified by the square. To test this explanation I ran a second LOOCV on the continuous approaches that rounded predicted contributions to the nearest integer. The MSE was then calculated using the rounded prediction (MSE-integer Table 8) making MSE exactly comparable to the discrete MSEs. There is little difference between the MSE and MSE-integer values. Rounding does not explain the poorer performance of the discrete approaches. To see whether dropping the 25 uniquely observed observations might have aversely affected the ordered logit's predictive power, I tested its *in-sample* predictive power by running a cross validation for each variable fitting the model from every observation to give it its best chance at accurately predicting contribution choices. The MSE was just as large (2279.282) suggesting that this was not the cause of its poor performance.

Table 8: Leave-One-Out Cross Validation

Model	**Obs**	**MSE** (precision)	**MSE-integer**	**ME** (bias)
Random Effects	3600	1039.959	1040.367	0.001
Fixed Effects	3600	1051.498	1051.799	0.000
Tobit Random Effects	3600	1086.326	1086.209	-4.079
4-Component Poisson Mixture	3600	1267.854		-0.270
Ordered Logit	3575	2285.079		-1.885

5 Conclusion

The continuous estimators convincingly outperformed the discrete approaches in both precision (MSE) and un-biassedness (fixed and random effects). The difference in predictive performance between fixed effects, random effects and tobit random effects are negligible. However Tobit random effects estimates are biased. Given that there is no substantial tradeoff in performance and un-biasedness, Random effects estimation is preferred over fixed effects for VCM model estimations as it has the advantage of being able to estimate time in-variant demographic variables.

The MEs suggest that as long as a reasonable model is specified, authors should not be too concerned about the possible biases induced by censoring. In fact trying to correct for censoring will likely induce bias. Greene (1981) even concedes that estimation bias can become negligible even in the presence of severe censoring as the fit of the model increases.

The question raised from these results is why do the discrete estimation approaches perform so badly? For the FMM approach it may be because there are too many points to cluster around. This problem is one of identification. Its performance might be considerably improved by adding more components. If this were the case though, one must then question whether the FMM is reflecting a finite number of contributor types, or is instead clustering around the VCM game groups exogenously randomly determined by the experimenter. In this circumstance, the FMM would not be modeling heterogeneity in contribution preferences but simply reflecting random clustering by experimental design. Further research could be done to investigate the number of mixture components needed to outperform random effects estimation and whether the estimates are clustering on group membership.

The poor performance of the ordered logit might be because the logit is predicting the mode where as the continuous estimators are predicting the mean. If this is the case, we might expect a discrete approach to out-perform a continuous approach under a unimodal data generating process. VCM data though is characterized by multiple modes which is the likely reason the logit estimator performed so badly. In such a circumstance continuous estimators are preferred.

Occam's Razor appears to win the final debate when estimating models using VCM data. As with many things in life, the simplest solution is often the best.

References

Andreoni, J. (1988): "Why free ride?: Strategies and learning in public goods experiments," *Journal of Public Economics*, 37(3), 291304.

Bardsley, N., and P. G. Moffatt (2007): "The experimetrics of public goods: inferring motivations from contributions," *Theory and Decision*, 62, 161193.

Breitmoser, Y. (2010): "Structural modeling of altruistic giving," *MPRA Working Paper*, Paper No. 24262, 1–22.

Breusch, T. S., and A. R. Pagan (1980): "The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics," *The Review of Economic Studies*, 47(1), 239–253.

Carpenter, J. (2004): "When in Rome: Conformity and the provision of public goods," *Journal of Socio-Economics*.

Chaudhuri, A. (2011): "Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature," *Experimental Economics*, 14(47-83).

Davis, D., and C. Holt (1993): *Experimental Economics*. New Jersey: Princeton University Press.

Deb, P. (2007): "FMM: Stata module to estimate finite mixture models," Statistical Software Components S456895.

Demidenko, E. (2004): *Mixed Models: Theory and Applications*. New Jersey: Wiley.

Fischbacher, U., and S. Gächter (2010): "Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments," *American Economic Review*, 100(1), 541–556.

Gächter, S., B. Herrmann, and C. Thöni (2010): "Culture and Coooperation," 365(1553), 2651–2661.

Gächter, S., E. Renner, and M. Sefton (2008): "The Long-Run Benefits of Punishment," *Science*, 322(5907), 1510.

Greene, W. H. (1981): "On the Asymptotic Bias of the Ordinary Least Squares Estimator of the Tobit Model," *Econometrica*, 49(2), 505–513.

Guillen, P., D. Merrett, and R. Slonim (2012): "Inter-group competition and the efficient provision of public goods," *Working Paper*.

Harrison, G. W. (2007): "Maximum Likelihood Estimation of Utility Functions Using *Stata*," *Working Paper 06-12, University of Central Florida*.

Harrison, G. W., and E. Rutstrom (2009): "Expected utility theory and prospect theory: one wedding and a decent funeral," *Experimental Economics*, 12, 133–158.

Hausman, J. A. (1978): "Specification Tests in Econometrics," *Econometrica*, 46(6), 12511271.

Herrmann, B., C. Thöni, and S. Gächter (2008): "Antisocial punishment across societies," *Science*, 319, 1362–1367.

Hurvich, C. M., and C. ling Tsai (1989): "Regression and time series model selection in small samples," *Biometrika*, 76(2), 297–307.

Isaac, R. M., K. F. McCue, and C. R. Plott (1985): "Public goods provision in an experimental environment," *Journal of Public Economics*, 26(1), 51–74.

Isaac, R. M., J. M. Walker, and S. H. Thomas (1984): "Divergent evidence on free riding: An experimental examination of possible explanations," *Public Choice*, 43, 113–149.

Keser, C., and F. van Winden (2000): "Conditional Cooperation and Voluntary Contributions to Public Good," *The Scandinavian Journal of Economics*, 102(1), 23–39.

Kim, O., and M. Walker (1984): "The free rider problem: Experimental evidence," *Public Choice*, 43, 3–24.

Ledyard, J. O. (1995): *The Handbook of Experimental Economics*. New Jersey: Princeton University Press.

McLachlan, G., and D. Peel (2000): *Finite Mixture Models*. John Wiley & Sons.

NIKIFORAKIS, N. (2010): "Feedback; Punishment and Cooperation in Public Good Experiments," *Games and Economic Behavior*, 68, 689.

POEN, E. (2009): "The Tobit Model with Feedback and Random Effects: A Monte-Carlo Study," *CeDEx Discussion Paper No. 2009-14*, pp. 1–37.

SCHWARZ, G. (1978): "Estimating the Dimension of a Model," *Annals of Statistics*, 6(2), 461–464.

STONE, M. (1974): "Cross-Validatory Choice and Assessment of Statistical Predictions," *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2), 111–147.

TAN, J. H., AND F. BOLLE (2007): "Team Competition and the Public Goods Game," *Economic Letters*, 96, 133–139.

URS FISCHBACHER, S. G., AND E. FEHR (2001): "Are people conditionally cooperative? Evidence from a public goods experiment," *Economics Letters*, 71, 397–404.

WHITE, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50(1), 1–25.

WOOLDRIDGE, J. M. (2009): *Introductory Econometrics*. USA: South-Western Cengage Learning.