CENTRE FOR DECISION RESEARCH & EXPERIMENTAL ECONOMICS

The University of Nottingham

Discussion Paper No. 2009-14

Eva Poen

July 2009

The Tobit Model with Feedback and Random Effects: A Monte-Carlo Study

CeDEx Discussion Paper Series
ISSN 1749 - 3293

Karina Terry
Centre for Decision Research and Experimental Economics
School of Economics
University of Nottingham
University Park
Nottingham
NG7 2RD
Tel: +44 (0) 115 95 15620
Fax: +44 (0) 115 95 14159
karina.terry@nottingham.ac.uk

The Tobit model with feedback and random effects: A Monte-Carlo study

Eva Poen*

July 2009

We study a random effects censored regression model in the context of repeated games. Introducing a feedback variable into the model leads to violation of the strict exogeneity assumption, thus rendering the random effects estimator inconsistent. Using the example of contributions to a public good, we investigate the size of this bias in a Monte-Carlo study. We find that the magnitude of the bias is around one per cent when initial values and individual effects are correlated. The rate of censoring, as well as the size of the groups in which subjects interact, both have an effect on the magnitude of the bias. The coefficients of strictly exogenous, continuous regressors remain unaffected by the endogeneity bias. The size of the endogeneity bias in our model is very small compared to the size of the heterogeneity bias, which occurs when individual heterogeneity is not accounted for in estimation of nonlinear models.

JEL classifications: C15, C24, C92

Keywords: Monte-Carlo, Simulation, Random Effects, Censored Regression Model, Public Goods, Heterogeneity, Endogeneity.

Acknowledgements: Many thanks to Simon Gächter and Richard Upward for helpful comments. Tony Wheatman and Ruslan Kabalin provided valuable assistance with the IT setup. Thanks to all of my colleagues who donated valuable computation time for the simulations.

*CeDEx, School of Economics, University of Nottingham, Sir Clive Granger Building, University Park, Nottingham NG7 2RD, United Kingdom (e-mail eva.poen@gmail.com).

1 Introduction

In economic experiments, subjects often interact in groups and receive some form of feedback. The repeated public goods game is a prime example: Subjects take their decision and then learn about the aggregate outcome in their group before the next round starts. The same applies to other games, for example the gift exchange game or the trust game when they are repeated. While experimenters can control many aspects of the environment, the feedback presented to subjects in repeated play is typically determined by the subjects' actions and therefore endogenous.[1] Often, researchers believe that past feedback is a meaningful predictor for current choices, thus including feedback as dependent variable into their models. For econometric analysis, this poses two problems: the observations are no longer independent between subjects, and the strict exogeneity assumption fails. The consequences of this violation of the independence and strict exogeneity assumptions depend on the type of model estimated. In *pooled estimation*, all parameters can be consistently estimated as long as the model is dynamically complete; see Wooldridge (2002, p. 256). If *individual effects* feature in the model, and panel models are estimated, matters are more complicated. In the remainder of this article we study the properties of a specific estimator when feedback as well as individual heterogeneity are included in the model. In our empirical application, we focus on contributions to an experimental public good under the voluntary contribution mechanism (VCM), but none of our results is theoretically linked to public goods.

Experimental researchers have become increasingly interested in learning more about the heterogeneity of behaviour in their experiments. The discussion about types of players in the public goods literature is one example of this interest in heterogeneity. Some experimental methods have been employed to study individual heterogeneity, for example the strategy method in Fischbacher, Gächter et al. (2001). This effort has been complemented by econometric methods like the Two-Limit p-Tobit model by Bardsley and Moffatt (2007). The latter study employed a sequential design and special randomisation technique to identify four types of players in their data. However, a very large number of public goods experiments have been conducted with a simultaneous design and repeated interaction; our own dataset comprises more than 3500 individuals

[1] Some authors have manipulated feedback in public good experiments. Bardsley and Moffatt (2007) employ a random mechanism to simulate other group members. Sell (1997) uses (manipulated) feedback as a treatment variable (high cooperation versus low cooperation).

who participated in studies with this design.[2] Given the evidence that individuals are indeed heterogeneous, the econometric analysis of this type of data must take heterogeneity into account. It is for this reason that we concentrate on models with individual effects.

More precisely, we want to design a model that allows for individual-specific intercepts as well as slopes (mixed model); this allows us to capture very different types of contribution behaviour in a single model. One subject, for example, may be highly sensitive to the feedback received, while another might contribute the same amount no matter what. A third subject might provide contributions that decline over time for strategic reasons. By allowing the intercepts and slopes to vary between subjects, we are able to account for this multi-dimensional form of heterogeneity that we believe is present in public goods data.

The participants of an experiment can be considered a random sample from a much larger population. We are not interested in the actual values of the individual effects for our subjects (since these values are random), but rather in the properties of the estimated distribution of these effects. Therefore, the random effects framework is appropriate (see Baltagi, 2001, p.15). Another reason to employ random effects lies in the importance of time-invariant regressors in experimental studies; often, dummy variables for treatments or sessions are added to the model, and the coefficients on these cannot be identified in a fixed effects model.

As is true for many experimental settings, the outcome variable in VCM public goods experiments is bounded. Contributions are limited to lie in between zero and the endowment, ω. Typically, there is a sizable share of observations at either end of this range, and especially at zero. We therefore concentrate on the random effects Tobit framework to account for censoring. This model can be consistently estimated given that there is a single individual random effect (random intercept), and all regressors satisfy the strict exogeneity assumption; it is implemented in Stata in the `xttobit` command.[3] Here, we will add two complications to the model: more than one individual random effect, and a feedback variable as regressor, which violates strict exogeneity.

Arellano and Honoré stated in 2001: '[...] very little is known about the estimation of nonlinear panel data models with predetermined explanatory variables.' (Arellano and Honoré, 2001, p.3266). Some progress has been made since then; Honoré and Hu

[2] The dataset mentioned is a collection of data from several studies, all of which were co-authored by Simon Gächter (CeDEx, University of Nottingham).

[3] We used version 10.1 of the software for our simulation.

(2004) prove that the GMM estimator in the case of a lagged dependent variable is consistent under a set of regularity conditions. They also generalise the findings to predetermined variables without limiting the nature of the feedback. However, their model does not include random slopes.

To our knowledge there is no estimator for the censored regression model that allows for random slopes and endogenous regressors. The aim of this study is to investigate the random effects estimator under feedback, as it is present in repeated experiments. We hope that, although we include a regressor that violates strict exogeneity, the bias will be small due to the nature of feedback and censoring in the model.

In order to proceed, we need to be more specific about what exactly we mean by feedback. In the current study we assume that N individuals are observed over T time periods. T is regarded as fixed and small in comparison to N. There are P groups, made up of K individuals each (groups are of an equal size, therefore $K\,P = N$). Let $\overline{y}_{t-1,p-i}$ denote the average outcome y in the previous round of group p, excluding individual i, i.e.

$$\overline{y}_{t-1,p-i} = \frac{1}{K-1} \sum_{j=1, j\neq i, j\in p}^{K} y_{j,t-1}. \tag{1}$$

If the true data generating process is, generally speaking,

$$E(y_{i,t}) = g(\overline{y}_{t-1,p-i},\, X, \beta) \tag{2}$$

then $\overline{y}_{t-1,p-i}$ is considered a feedback variable.[4] The effect is twofold: First, $y_{i,t-1}$ feeds into $y_{j\neq i, j\in p,t}$, creating a correlation between observations in a group. One period on, it feeds back into $y_{i,t+1}$ via the same process (2). Feedback effects similar to (1) can occur in many experimental settings where subjects repeatedly interact in groups.

In the context of a linear model and $K = 2$, we can see how the feedback variable violates strict exogeneity. Assume the simple linear model

$$y_{i,t} = \beta_0 y_{j,t-1} + \beta_1 x_{i,t} + u_i + \varepsilon_{i,t} \tag{3}$$

where x is a strictly exogenous regressor, ε is Gaussian white noise, and u is an unobserved individual effect which is uncorrelated with x and ε. Here, the feedback variable

[4] We assume that group composition is constant over time ('partner' matching). This does not mean that feedback effects are not present in 'stranger' matching, but we would have to adopt a more general notation to cater for it.

$\overline{y}_{t-1,p-i}$ is simply equal to the lagged contribution of group member j since $K = 2$. (3) can be rewritten as

$$\begin{aligned} y_{i,t} &= \beta_0 \left(\beta_0 y_{i,t-2} + \beta_1 x_{j,t-1} + u_j + \varepsilon_{j,t-1}\right) + \beta_1 x_{i,t} + u_i + \varepsilon_{i,t} \\ &= \beta_0 \left[\beta_0 \left(\beta_0 y_{j,t-3} + \beta_1 x_{i,t-2} + u_i + \varepsilon_{i,t-2}\right) + \beta_1 x_{j,t-1} + u_j + \varepsilon_{j,t-1}\right] \\ &\quad + \beta_1 x_{i,t} + u_i + \varepsilon_{i,t} \end{aligned} \tag{4}$$

From (4) we can see the two effects of including feedback into the model: Firstly, $y_{i,t}$ is correlated with u_j, which means that all observations of subject i are correlated with all observations of subject j. Secondly, $y_{i,t}$ depends on $y_{i,t-2}$, creating a dependence over time and introducing endogeneity. The feedback regressor $y_{j,t-1}$ is correlated with the individual effect u_i; we therefore have a similar situation to a lagged dependent variable in the presence of unobserved effects, and the strict exogeneity assumption never holds (see Wooldridge (2002, p. 256)). $y_{j,t-1}$ is not exogenous but it is predetermined.

This linear model cannot be consistently estimated by OLS due to $E(y'_{j,t-1}u_i) > 0$. The random effects estimator, which relies on an even stricter exogeneity assumption, is also inconsistent. The research question at hand is the extent of the inconsistency in the parameter estimates, given the unusual structure of feedback in groups. We will be focussing on the Tobit model for censored outcomes. We do not consider any other type of misspecification, such as heteroscedasticity, non-normality of errors, or contemporaneous endogeneity.

We conduct a Monte-Carlo study to examine the properties of the maximum likelihood estimator implemented in `gllamm` under certain conditions. The user-written software `gllamm` (Generalized Linear and Latent Mixed Models; see Rabe-Hesketh, Skrondal et al. (2004)) allows the estimation of a wide range of models, including nonlinear hierarchical models with random intercepts and random slopes. With some adjustments to the data it can also be employed to fit Tobit models. While there is a Tobit random effects estimator `xttobit` in official Stata, `gllamm` can handle more than one random effect on more than one level. This estimator in conjunction with the Tobit model, however, has not been widely discussed in the literature. Another goal of the current study is therefore to establish the consistency of the estimator in `gllamm` when estimating two-limit Tobit models with one or more random effects.[5]

[5] The authors of `gllamm` explain how to fit Tobit models in Rabe-Hesketh, Skrondal et al. (2005) and Rabe-Hesketh and Skrondal (2007). The method was also discussed on the discussion list Statalist earlier than that. `gllamm`'s capability of fitting Tobit models seems to have gone largely unnoticed in the literature, with many authors relying on either linear mixed models or simpler Tobit models for estimation (see for example Carpenter (2007)). Some authors do employ `gllamm`

The Monte-Carlo method is instrumental in learning about the impact of feedback effects on an estimator. With real world data, the true data-generating process is unknown, as is the true distribution of the data. It is therefore very difficult, if not impossible, to make reliable conclusions about the statistical properties of an estimator using real world data. All our analyses are based on data that are randomly generated from a well-defined distribution, using a data generating process that we select according to our design.

2 Design

Table 1: Design matrix

	No feedback	Uncorrelated feedback	Correlated feedback
No random effects		Experiment 3	
Random intercept only	Experiment 1	Experiment 4	Experiment 6
Random intercept and random slopes	Experiment 2	Experiment 5	Experiment 7

We implemented a three-by-three design for our simulation study; see table 1. The treatment variables are the number of random effects (none, one, or three random effects) and the presence of feedback (no feedback, uncorrelated feedback and correlated feedback). The top left cell of the design matrix is trivial (the standard Tobit model), and no experiments have been performed for this configuration. The top right cell of the matrix empty because it is infeasible. The correlation is between the individual effect and the feedback; a model without random effects cannot be constructed.

In order to be able to distinguish between the two features of our feedback process (2)

to fit Tobit models, but only specify a single random intercept (e. g. Godin (2008) and Shephard, Falcaro et al. (2003)) – a model which can be estimated by the `xttobit` command in official Stata. However, the adaptive quadrature implemented in `gllamm` is superior in situations involving large cluster sizes or high intraclass correlation; see Rabe-Hesketh, Skrondal et al. (2005). The downside of adaptive quadrature is that convergence is slow.

– correlation between observations in a group and correlation between the regressor affected by feedback and the individual effect – we also conduct experiments in which we introduce a lagged dependent variable instead of feedback from other group members. A lagged dependent variable induces correlation with the unobserved effect, but preserves independence of observations between cross-sectional units.

3 The model and parameters

Our data-generating process is a linear panel model with N cross-sectional units and T time periods. We then apply censoring from below and above. The basic model is

$$
\begin{aligned}
y^*_{i,t} &= \beta_0 + \beta_1 x1_{i,t} + \beta_2 x2_{i,t} + \beta_3 x3_{i,t} + \beta_{dT} d(t = T) \\
&\qquad\qquad\qquad + \beta_{d(T-1)} d(t = T-1) + \beta_t t + \varepsilon_{i,t} \\
y_{i,t} &= \begin{cases} 0 & \text{if } y^*_{i,t} < 0, \\ y^*_{i,t} & \text{if } 0 \leq y^*_{i,t} \leq 20, \\ 20 & \text{if } y^*_{i,t} > 20 \end{cases}
\end{aligned}
\tag{5}
$$

where $x1$, $x2$ and $x3$ are exogenous predictors, t is a linear time trend, $d(t = T)$ is an indicator variable (dummy) for the final time period, $d(t = T - 1)$ is a dummy variable for the penultimate time period, and $\varepsilon_{i,t}$ is an *iid* normally distributed disturbance. Depending on the experiment, we introduce one or more random effects into the equation for $y^*_{i,t}$.

Our choice of model is closely related to features of public goods data that are generally found in experimental studies (see Ledyard (1995) for an overview of the main findings). We include a linear time trend, and allow for an 'end-game effect' with additional time dummies for the ultimate and penultimate round. The censoring points of zero and 20 reflect our own data; the endowment ω was equal to 20 tokens in all our experiments. We chose the parameter values β such that the censoring rate is approximately 60 per cent, which is fairly high; the highest rate of censoring in our own data is 68 per cent.

The values of N and T remain constant for the experiment; we selected $N = 200$ and $T = 10$. The choice of 200 cross-sectional units stems from the experimental background of our data: in an experimental study, 200 subjects would constitute a very good sample size. Since we develop our model for experimental data, we did not want to assume unrealistically large sample sizes. We also did not want to use less than 200 units in order to give `gllamm` a reasonable chance of identifying the model (200 units mean 200 realisations of the random effects). Investigating the asymptotic properties of the estimator as N grows large, as well as the (very) small sample properties, exceeds the scope of the current study and is left for future research. The choice of 10 time periods is again motivated by our own data, which predominantly consists of experimental sessions with 10 rounds.

In all simulation experiments, our pseudo-population is composed of a vector of variables ($x1$, $x2$, $x3$, $u0$, $u1$, ut), where the x variables are regressors and the u variables are individual random effects. The variables follow a multivariate normal distribution with expectation μ and variance-covariance matrix Σ:

$$\mu \begin{pmatrix} x1 \\ x2 \\ x3 \\ u0 \\ u1 \\ ut \end{pmatrix} = \begin{pmatrix} 19 \\ 6 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \qquad \Sigma \begin{pmatrix} x1 \\ x2 \\ x3 \\ u0 \\ u1 \\ ut \end{pmatrix} = \begin{pmatrix} 64 & & & & & \\ 12 & 25 & & & & \\ -2.4 & 0.75 & 2.25 & & & \\ 0 & 0 & 0 & 36 & & \\ 0 & 0 & 0 & -0.75 & 0.25 & \\ 0 & 0 & 0 & 0.48 & -0.06 & 0.16 \end{pmatrix} \tag{6}$$

The covariates are correlated with each other, as are the random effects. All random effects are uncorrelated with the covariates, satisfying the random effects assumption. We also introduce a random disturbance $\varepsilon_{i,t}$ which is uncorrelated with both x_i and u_i and normally distributed with $\mu_\varepsilon = 0$ and $\sigma_\varepsilon = 6$. The true parameters of our linear model are as follows:

$$\begin{pmatrix} \beta_{x1} \\ \beta_{x2} \\ \beta_{x3} \\ \beta_t \\ \beta_{d(t=10)} \\ \beta_{d(t=9)} \\ \beta_{\text{cons}} \end{pmatrix} = \begin{pmatrix} 0.4 \\ -3 \\ -2.5 \\ -0.7 \\ -7 \\ -3 \\ 22 \end{pmatrix} \tag{7}$$

In experiments with feedback, the first time period is not included in the estimations. $x1$ is the transmitter of feedback effects in our feedback experiments, via a mechanism explained later on.

Since the type of models that can be estimated by `xttobit` are a subset of those that can be estimated by `gllamm`, we can directly compare the two estimators in some cases. We begin by simulating a model with a random intercept, which can be estimated by both commands, in order to establish whether or not `gllamm` can estimate censored models under very regular conditions. The official `xttobit` command will be our benchmark.

Convergence with `gllamm` in the case of three random effects can take a very long time (approximately four hours for a model with three random effects and 2000 observations). For this reason, the experiments with three random effects are carried out with a low number of repetitions of between 600 and 1000 each. One would usually conduct Monte-Carlo studies with 10000 repetitions or more, but this is infeasible due to computing time.[6]

4 Experiments without feedback effects

4.1 Random intercept only: Experiment 1

The first experiment is for a model with a random intercept $u0$. There are no feedback effects, and the data come from the well-behaved pseudo-population described in (6); these are optimal conditions for both `xttobit` and `gllamm` to estimate this model. The data-generating process is

$$\begin{aligned} y^*_{i,t} &= \beta_0 + \beta_1 x1_{i,t} + \beta_2 x2_{i,t} + \beta_3 x3_{i,t} + \beta_{d10} d(t=10) \\ &\qquad + \beta_{d9} d(t=9) + \beta_t t + u0_i + \varepsilon_{i,t} \\ y_{i,t} &= \begin{cases} 0 & \text{if } y^*_{i,t} < 0, \\ y^*_{i,t} & \text{if } 0 \leq y^*_{i,t} \leq 20, \\ 20 & \text{if } y^*_{i,t} > 20 \end{cases} \end{aligned} \tag{8}$$

The results are listed in table 2 and 3. We can see that the averages of the estimated parameters are very close to the true values, and biases are well below one per cent. Both `xttobit` and `gllamm` perform very well. The normality tests and a visual inspection suggests that the estimates are normally distributed. The average estimated

[6] In order to see whether the smaller number of repetitions for `gllamm` influences our results, we cross-check the findings from experiments 1, 4 and 6 using only a random sample of 800 repetitions each. While there is some variation in the extent of the biases, all results continue to hold. The average bias across all coefficients in these experiments increases from 0.28 per cent (10000 repetitions) to 0.35 per cent (800 repetitions).

Table 2: Results for experiment 1: `xttobit`

Parameter	true value	average estimate	bias in %	standard deviation	average estimated s.e.	p-value t-test	p-value normality test
β_1	0.4	0.40	0.01	0.03	0.03	0.82	0.17
β_2	-3	-3.00	0.07	0.07	0.07	0.00	0.00
β_3	-2.5	-2.50	0.09	0.14	0.14	0.10	0.49
β_t	-0.7	-0.70	0.04	0.09	0.09	0.74	0.59
β_{d10}	-7	-7.02	0.24	0.80	0.80	0.04	0.36
β_{d9}	-3	-3.01	0.23	0.71	0.72	0.32	0.04
β_0	22	22.02	0.08	0.79	0.79	0.03	0.27
σ_ε	6	5.99	0.24	0.17	0.17	0.00	0.01
σ_{u0}	6	5.98	0.38	0.37	0.37	0.00	0.00

$N = 200$, $T = 10$, estimated by `xttobit`, 10000 repetitions.
The t-test is for equality of the average estimated parameter and the true value (two-sided test).

Table 3: Results for experiment 1: `gllamm`

Parameter	true value	average estimate	bias in %	standard deviation	average estimated s.e.	p-value t-test	p-value normality test
β_1	0.4	0.40	0.07	0.03	0.03	0.26	0.59
β_2	-3	-3.00	0.06	0.07	0.07	0.01	0.00
β_3	-2.5	-2.50	0.00	0.14	0.14	0.99	0.00
β_t	-0.7	-0.70	0.17	0.09	0.09	0.16	0.14
β_{d10}	-7	-7.01	0.09	0.80	0.80	0.44	0.61
β_{d9}	-3	-3.02	0.58	0.72	0.72	0.02	0.92
β_0	22	22.00	0.00	0.79	0.79	0.97	0.52
$\log(\sigma_\varepsilon)$	1.79	1.79	0.16	0.03	0.03	0.00	0.23
σ_{u0}	6	5.98	0.37	0.37	0.37	0.00	0.07

$N = 200$, $T = 10$, estimated by `gllamm`, 10000 repetitions.
The t-test is for equality of the average estimated parameter and the true value (two-sided test).

standard errors are very close to the actual standard deviations of the estimates for both estimators.

Following these results, we have now confidence in the performance of `gllamm` when estimating Tobit random effects models under regular conditions.

4.2 Random intercept and random slopes: Experiment 2

We now add two random slopes to the model: one for the time trend t (which we call ut), and one for $x1$ (which we call $u1$). The new model is

$$\begin{aligned} y^*_{i,t} &= \beta_0 + \beta_1 x1_{i,t} + \beta_2 x2_{i,t} + \beta_3 x3_{i,t} + \beta_{d10} d(t=10) \\ &\qquad + \beta_{d9} d(t=9) + \beta_t t + u1_i x1_{i,t} + ut_i t + u0_i + \varepsilon_{i,t} \\ y_{i,t} &= \begin{cases} 0 & \text{if } y^*_{i,t} < 0, \\ y^*_{i,t} & \text{if } 0 \leq y^*_{i,t} \leq 20, \\ 20 & \text{if } y^*_{i,t} > 20 \end{cases} \end{aligned} \tag{9}$$

Again, we run simulations using both `xttobit` and `gllamm`. We do not expect `xttobit` to be consistent any longer, because the model it estimates is misspecified and the random slopes are correlated with the random intercept; the estimates will suffer from *heterogeneity bias*.

Table 4: Results for experiment 2: `xttobit`

Parameter	true value	average estimate	bias in %	standard deviation	average estimated s.e.	p-value t-test	p-value normality test
β_1	0.4	0.42	5.74	0.05	0.03	0.00	0.00
β_2	-3	-3.00	0.03	0.08	0.08	0.25	0.00
β_3	-2.5	-2.50	0.12	0.16	0.16	0.07	0.67
β_t	-0.7	-0.70	0.08	0.11	0.10	0.62	0.88
β_{d10}	-7	-6.95	0.71	0.95	0.95	0.00	0.03
β_{d9}	-3	-2.99	0.31	0.86	0.86	0.27	0.25
β_0	22	21.60	1.83	0.96	1.04	0.00	0.14
σ_ε	6	7.11	18.50	0.23	0.21	0.00	0.00
σ_{u0}	6	9.50	58.40	0.57	0.57	0.00	0.00

Three random effects. $N = 200$, $T = 10$, estimated by `xttobit`, 10000 repetitions.
The t-test is for equality of the average estimated parameter and the true value (two-sided test).

Table 5: Results for experiment 2: `gllamm`

Parameter	true value	average estimate	bias in %	standard deviation	average estimated s.e.	p-value t-test	p-value normality test
β_1	0.4	0.40	0.59	0.05	0.05	0.18	0.32
β_2	-3	-3.01	0.23	0.07	0.08	0.01	0.47
β_3	-2.5	-2.51	0.50	0.16	0.15	0.03	0.46
β_t	-0.7	-0.69	0.88	0.10	0.10	0.08	0.13
β_{d10}	-7	-7.08	1.17	0.85	0.88	0.01	0.90
β_{d9}	-3	-3.04	1.24	0.79	0.78	0.20	0.00
β_0	22	22.08	0.37	0.86	0.85	0.01	0.49
$\log(\sigma_\varepsilon)$	1.79	1.79	0.18	0.03	0.03	0.01	0.33
Cholesky(1,1)	6	5.98	0.34	0.97	0.97	0.56	0.01
Cholesky(2,2)	0.484	0.48	1.82	0.04	0.04	0.00	0.05
Cholesky(3,3)	0.378	0.30	21.90	0.19	0.28	0.00	0.00
Cholesky(2,1)	-0.125	-0.12	4.41	0.08	0.08	0.07	0.00
Cholesky(3,1)	0.08	0.08	3.75	0.14	0.14	0.55	0.22
Cholesky(3,2)	-0.103	-0.10	0.44	0.10	0.10	0.90	0.02
Var(u0)	36	36.70	1.94	11.52	(11.56)	0.10	0.00
Var(u1)	0.25	0.25	0.56	0.04	(0.04)	0.38	0.00
Cov(u0,u1)	-0.75	-0.75	0.22	0.55	(0.54)	0.94	0.15
Var(ut)	0.16	0.17	5.60	0.10	(0.10)	0.02	0.00
Cov(u0,ut)	0.48	0.39	18.70	0.77	(0.80)	0.00	0.00
Cov(u1,ut)	-0.06	-0.06	2.66	0.05	(0.05)	0.38	0.28

Standard deviations in parentheses are calculated from the estimated variance-covariance matrix using the delta method.

$N = 200$, $T = 10$, estimated by `gllamm`, 645 repetitions.

The t-test is for equality of the average estimated parameter and the true value (two-sided test).

Table 4 lists the results for `xttobit` (10000 repetitions). As expected, `xttobit` no longer consistently estimates most parameters. For all but one parameter (that of x2), the bias is now at least 30 per cent larger than it was before, with the parameter of x1 as well as σ_ε and the random effect being the worst affected. The standard deviations of all estimated parameters are up by at least 17 per cent, and `xttobit` underestimates these standard deviations by up to 33 per cent.

`gllamm` however (see table 5) estimates the main parameters of the model with very little bias. We can see some bias (of less than 1.3 per cent) on the indicator variables, which may be due to the small number of repetitions. The estimated standard errors are very close to the actual standard deviations.

The variances and covariances of the random effects[7] are estimated reasonably well. The biases that we see are presumably due to a twofold small sample effect: we have only 200 realisations of the random effects per repetition, and we also only have a relatively small number of repetitions. Especially the element $[3, 3]$ of the Cholesky decomposition does not seem to be very well estimated by `gllamm`, and this leads to biases in the variance and covariances of *ut*. However, we chose a very small value for the variance of *ut* (0.16), and `gllamm` only missed out by 0.009 on average. We would expect this bias to disappear when increasing the number of cross-sectional units; such an experiment is outside the scope of the current study.

We are now reasonably confident that `gllamm` consistently estimates the Tobit model with random intercept and random slopes, given a sufficient number of cross-sectional units.

5 Feedback effects

We now introduce feedback into the model. Since the other group members' behaviour plays an important part in contribution theory and evidence, we want to use this variable as a predictor. The average contribution of the other group members in the previous round violates the strict exogeneity assumption, as explained in the introduction.

[7] In `gllamm`, the variances and covariances of the random effects are not estimated directly; instead, the software estimates the elements of the Cholesky decomposition of the variancecovariance matrix. The last six rows of table 5 therefore depict the values derived from the estimated quantities.

We stipulate the following model:

$$\begin{aligned} y^*_{i,t} &= \beta_0 + \beta_1 \overline{y}_{t-1,p-i} + \beta_2 x2_{i,t} + \beta_3 x3_{i,t} + \beta_{d10} d(t=10) \\ &\qquad + \beta_{d9} d(t=9) + \beta_t t + u0_i + \varepsilon_{i,t} \\ y_{i,t} &= \begin{cases} 0 & \text{if } y^*_{i,t} < 0, \\ y^*_{i,t} & \text{if } 0 \leq y^*_{i,t} \leq 20, \\ 20 & \text{if } y^*_{i,t} > 20 \end{cases} \end{aligned} \tag{10}$$

β_1 is now the coefficient of the feedback variable. The feedback is blurred by averaging and censoring, and we hope that, if this is the only transmitter of feedback, the estimator will be relatively robust towards it. We apply censoring to the feedback variable since $\overline{y^*}_{t-1,p-i}$ is unobservable; subjects only observed $\overline{y}_{t-1,p-i}$, and the latent variable cannot have influenced the decision making process.

For the practical implementation, we draw the initial (i. e. time period 1) values of y from $x1$ in (6), alongside the other variables. We assign all units to groups of four. We then compute $\overline{y}_{1,p-i}$ for each unit, which we censor according to our censoring rule at zero and 20. This allows us to compute $y_{i,2}$ according to (10). This process is then repeated for all time periods.

Groups remain constant in our design (i. e. partner matching), since we believe that the feedback effects will be strongest in this setting. Changing the groups randomly would therefore not constitute a good test of robustness of the estimator.

The observations are no longer independent between cross-sectional units; strictly speaking, we are applying a quasi-likelihood approach in the following sections.

5.1 Without random effects: Experiment 3

In a first step, we want to confirm that a pooled Tobit model can be consistently estimated in the presence of feedback effects, but without any random effects. The specification in (10) is dynamically complete (i.e. all relevant lags appear in the model); thus, inference is the same as in the standard Tobit model (Wooldridge, 2002, p.539). The results of 10000 repetitions are listed in table 6. The coefficients and standard errors being estimated very accurately. If there is a small sample bias, it is negligible.

Table 6: Results for experiment 3: `tobit`

Parameter	true value	average estimate	bias in %	standard deviation	average estimated s.e.	p-value t-test	p-value normality test
β_1	0.4	0.40	0.19	0.03	0.03	0.01	0.48
β_2	-3	-3.00	0.03	0.07	0.07	0.22	0.00
β_3	-2.5	-2.50	0.07	0.14	0.14	0.21	0.02
β_t	-0.7	-0.70	0.47	0.12	0.12	0.01	0.04
β_{d10}	-7	-7.01	0.09	0.90	0.90	0.48	0.04
β_{d9}	-3	-3.00	0.04	0.78	0.78	0.89	0.18
β_0	22	22.03	0.13	0.83	0.82	0.00	0.00
σ_ε	6	5.98	0.36	0.17	0.17	0.00	0.03

$N = 200$, $T = 9$, estimated by `tobit`, 10000 repetitions.
The t-test is for equality of the average estimated parameter and the true value (two-sided test).

5.2 Random intercept only: Experiment 4

We now add a random intercept $u0$ to the model, and estimate using `xttobit` and `gllamm`. The results are displayed in tables 7 and 8.

Table 7: Results for experiment 4: `xttobit`

Parameter	true value	average estimate	bias in %	standard deviation	average estimated s.e.	p-value t-test	p-value normality test
β_1	0.4	0.40	0.39	0.03	0.03	0.00	0.48
β_2	-3	-3.00	0.06	0.08	0.08	0.03	0.00
β_3	-2.5	-2.50	0.06	0.15	0.15	0.35	0.08
β_t	-0.7	-0.70	0.18	0.13	0.12	0.32	0.02
β_{d10}	-7	-7.01	0.09	0.94	0.94	0.51	0.16
β_{d9}	-3	-3.00	0.11	0.81	0.81	0.70	0.06
β_0	22	21.99	0.05	0.98	0.98	0.25	0.24
σ_ε	6	5.98	0.34	0.19	0.19	0.00	0.00
σ_{u0}	6	5.97	0.47	0.40	0.39	0.00	0.00

$N = 200$, $T = 9$, estimated by `xttobit`, 10000 repetitions.
The t-test is for equality of the average estimated parameter and the true value (two-sided test).

For both estimators, the biases are well below one per cent, and the average estimated standard errors are very close to the actual standard deviations in the simulations. In fact, the t-test of $\overline{\hat{\beta}} = \beta_{\text{true}}$ cannot reject the null hypothesis for the majority of parameters. Standard deviations have gone up by 19 per cent on average for both estimators, compared to the situation without feedback effects; the likely cause is that the effective sample size has been reduced due to intra-group correlation, while the

Table 8: Results for experiment 4: `gllamm`

Parameter	true value	average estimate	bias in %	standard deviation	average estimated s.e.	p-value t-test	p-value normality test
β_1	0.4	0.40	0.39	0.03	0.03	0.00	0.38
β_2	-3	-3.00	0.07	0.08	0.08	0.01	0.00
β_3	-2.5	-2.50	0.08	0.15	0.15	0.18	0.04
β_t	-0.7	-0.70	0.36	0.13	0.12	0.04	0.62
β_{d10}	-7	-7.02	0.34	0.94	0.94	0.01	0.01
β_{d9}	-3	-3.01	0.45	0.81	0.81	0.10	0.09
β_0	22	21.98	0.10	0.98	0.98	0.02	0.85
$\log(\sigma_\varepsilon)$	1.79	1.79	0.22	0.03	0.03	0.00	0.57
σ_{u0}	6	5.98	0.42	0.39	0.39	0.00	0.00

$N = 200$, $T = 9$, estimated by `gllamm`, 10000 repetitions.
The t-test is for equality of the average estimated parameter and the true value (two-sided test).

number of observations stays the same. Overall, the estimators prove fairly robust to the type of feedback we have introduced in this section. Although strict exogeneity is violated by construction, we have so far only implemented a very mild, and probably unrealistic, form of feedback: the initial values of y_i in time period 1 are drawn from the same distribution as $x1$ in (7), and are not correlated with the individual effect. We are going to relax this assumption later on.

5.3 Random intercept and random slopes: Experiment 5

Next we introduce random slopes for $x1$ and t into the model, as we have done before in section 4.2. Again, we expect `xttobit` to be inconsistent.

As expected, `xttobit` (see table 9) produces biased estimates of many coefficients. It also substantially underestimates the standard error of the coefficient of the feedback variable $x1$ by 26 per cent. The t-test is now being rejected on a significance level of five per cent for seven coefficients. Comparing these results to the results without feedback in table 4, we can see that the bias has increased for some coefficients and decreased for others. Most notably, there is now a bias of more than three per cent on the parameter β_t.

`gllamm`, on the other hand, estimates coefficients and standard errors very well. For all coefficients of the main model, except the standard deviation of the residual, the t-test cannot reject the null hypothesis. The normality test fares equally well. The standard errors are out by at most 6.6 per cent. Overall, the introduction of feedback seems to

Table 9: Results for experiment 5: `xttobit`

Parameter	true value	average estimate	bias in %	standard deviation	average estimated s.e.	p-value t-test	p-value normality test
β_1	0.4	0.42	3.86	0.05	0.04	0.00	0.18
β_2	-3	-3.00	0.13	0.09	0.09	0.00	0.00
β_3	-2.5	-2.50	0.09	0.17	0.17	0.19	0.00
β_t	-0.7	-0.67	3.73	0.14	0.14	0.00	0.23
β_{d10}	-7	-6.99	0.19	1.06	1.08	0.21	0.07
β_{d9}	-3	-3.05	1.70	0.92	0.94	0.00	0.15
β_0	22	21.74	1.20	1.08	1.12	0.00	0.47
σ_ε	6	7.14	19.00	0.26	0.23	0.00	0.00
σ_{u0}	6	6.73	12.10	0.47	0.45	0.00	0.00

Three random effects. $N = 200$, $T = 9$, estimated by `xttobit`, 10000 repetitions.
The t-test is for equality of the average estimated parameter and the true value (two-sided test).

Table 10: Results for experiment 5: `gllamm`

Parameter	true value	average estimate	bias in %	standard deviation	average estimated s.e.	p-value t-test	p-value normality test
β_1	0.4	0.40	0.04	0.05	0.05	0.94	0.83
β_2	-3	-3.00	0.05	0.08	0.08	0.59	0.14
β_3	-2.5	-2.50	0.10	0.16	0.16	0.69	0.46
β_t	-0.7	-0.70	0.41	0.14	0.14	0.61	0.17
β_{d10}	-7	-7.01	0.07	1.05	1.00	0.91	0.95
β_{d9}	-3	-3.04	1.47	0.90	0.86	0.20	0.03
β_0	22	22.02	0.08	1.05	1.03	0.66	0.60
$\log(\sigma_\varepsilon)$	1.79	1.78	0.42	0.04	0.04	0.00	0.54
Cholesky(1,1)	6	6.09	1.57	1.05	1.08	0.02	0.26
Cholesky(2,2)	0.484	0.47	2.65	0.04	0.04	0.00	0.07
Cholesky(3,3)	0.378	0.26	30.60	0.24	0.46	0.00	
Cholesky(2,1)	-0.125	-0.13	1.59	0.10	0.10	0.60	0.38
Cholesky(3,1)	0.08	0.06	29.40	0.18	0.19	0.00	0.00
Cholesky(3,2)	-0.103	-0.10	1.16	0.15	0.14	0.83	0.46
Var(u0)	36	38.24	6.21	12.93	(13.20)	0.00	0.00
Var(u1)	0.25	0.25	0.12	0.05	(0.05)	0.88	0.01
Cov(u0,u1)	-0.75	-0.81	7.96	0.65	(0.65)	0.02	0.00
Var(ut)	0.16	0.19	20.20	0.14	(0.15)	0.00	0.00
Cov(u0,ut)	0.48	0.19	59.80	1.13	(1.11)	0.00	0.00
Cov(u1,ut)	-0.06	-0.05	11.90	0.07	(0.07)	0.01	0.01

Standard deviations in parentheses are calculated from the estimated variance-covariance matrix using the delta method.

$N = 200$, $T = 9$, estimated by `gllamm`, 662 repetitions.

The t-test is for equality of the average estimated parameter and the true value (two-sided test).

have had a very small effect on the consistency of the estimates. The standard errors are higher under feedback, due to the reduced effective sample size.

The estimates of the random effects variancecovariance matrix suffer from similar problems as in section 4.2, and again we suspect small sample size and few repetitions as the underlying reasons for the biases. These biases appear to have increased with the introduction of feedback.

6 Correlated feedback

In the previous section, the initial values $y_{i,1}$ for each unit i were simply drawn from a normal distribution according to $x1$ in (6). This implies that the initial values are uncorrelated with the individual effects. The assumption of zero correlation is very strict, and probably unrealistic. We would expect an individual with a high $u0$ to have a higher initial value of y than another individual with a low value of $u0$. Introducing correlation between $y_{i,1}$ and $u0_i$ is likely to amplify the violation of the strict exogeneity assumption underlying the random effects estimator. To implement correlated feedback, we modify the way the initial values are constructed: we again draw random values according to $x1$ in (6), but then we add $u0 - 9$ to them. The term -9 is included in order to bring the average values of the new feedback variable in line with the old feedback variable.[8] Figure 1 shows the simulated average feedback variable over time for both the uncorrelated and the correlated case (10000 repetitions each). The two patterns are very similar to each other. They are also very similar to the actual trend of average contributions observed in our own data. By adding the random effect $u0$ to the initial values, we induce a correlation between $y_{i,1}$ and $u0_i$ of approximately 60 per cent.

6.1 Random intercept only: Experiment 6

We now study the properties of the two estimators under correlated feedback. Of special interest is the bias of the parameter β_1 (the coefficient of the feedback variable). Under uncorrelated feedback (experiment 4), the bias was 0.38 per cent for both `gllamm` and

[8] To be able to observe the pure effect of adding correlation to the feedback, we must ensure that all other aspects of the environment remain constant between experiments. This applies not only to the distribution of the data but also to the censoring rate and the range and trend of the resulting feedback.

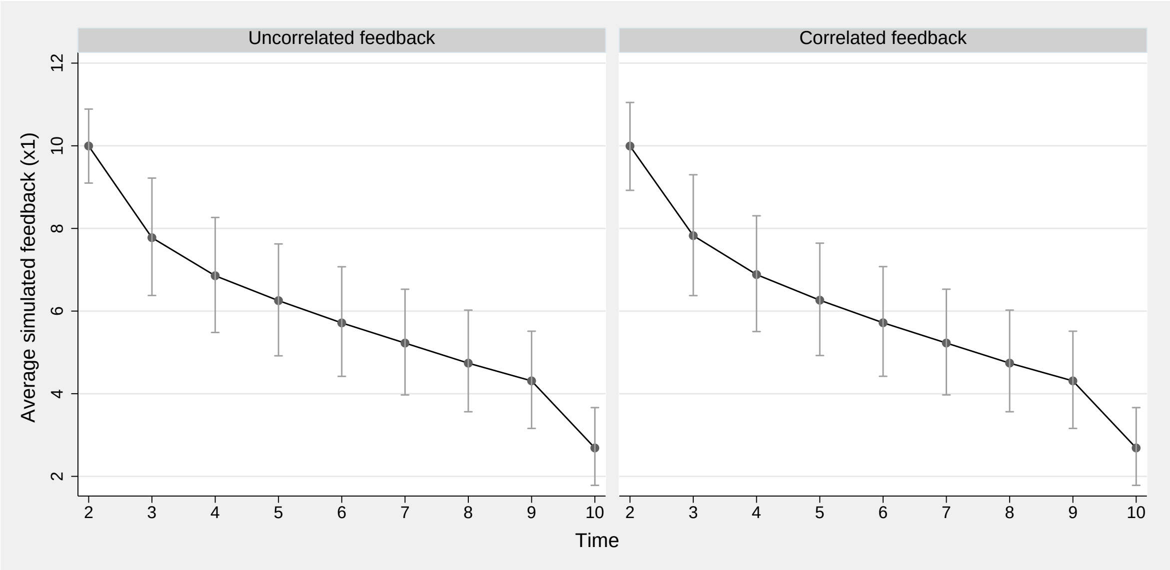

Figure 1: Comparison of simulated feedback effects over time

`xttobit`; the standard deviation of the estimates was ~ 0.034. Results for correlated feedback are listed in table 11 and 12.

Table 11: Results for experiment 6: `xttobit`

Parameter	true value	average estimate	bias in %	standard deviation	average estimated s.e.	p-value t-test	p-value normality test
β_1	0.4	0.40	1.03	0.04	0.04	0.00	0.49
β_2	-3	-3.00	0.07	0.08	0.08	0.01	0.00
β_3	-2.5	-2.50	0.15	0.15	0.15	0.01	0.02
β_t	-0.7	-0.70	0.40	0.11	0.11	0.01	0.03
β_{d10}	-7	-7.03	0.43	0.93	0.93	0.00	0.34
β_{d9}	-3	-3.02	0.50	0.81	0.80	0.07	0.78
β_0	22	21.97	0.16	0.88	0.88	0.00	0.00
σ_ε	6	5.98	0.33	0.19	0.19	0.00	0.00
σ_{u0}	6	5.98	0.37	0.39	0.39	0.00	0.00

$N = 200$, $T = 9$, estimated by `xttobit`, 10000 repetitions.
The t-test is for equality of the average estimated parameter and the true value (two-sided test).

The bias of β_1 has more than doubled to $\sim 1.05 percent$. The standard deviation remains almost unchanged at 0.036. `xttobit` and `gllamm` produce very similar results. The estimates of σ_{u0} are not badly affected by the introduction of correlation to the initial conditions. Generally, standard deviations have not increased significantly, and they are on average well estimated by both estimators.

Table 12: Results for experiment 6: `gllamm`

Parameter	true value	average estimate	bias in %	standard deviation	average estimated s.e.	p-value t-test	p-value normality test
β_1	0.4	0.40	1.05	0.04	0.04	0.00	0.10
β_2	-3	-3.00	0.08	0.08	0.08	0.00	0.00
β_3	-2.5	-2.50	0.07	0.15	0.15	0.23	0.05
β_t	-0.7	-0.70	0.58	0.11	0.11	0.00	0.07
β_{d10}	-7	-7.03	0.48	0.92	0.93	0.00	0.40
β_{d9}	-3	-3.03	0.86	0.81	0.80	0.00	0.67
β_0	22	21.96	0.17	0.89	0.88	0.00	0.00
$\log(\sigma_\varepsilon)$	1.79	1.79	0.23	0.03	0.03	0.00	0.91
σ_{u0}	6	5.97	0.50	0.39	0.39	0.00	0.00

$N = 200$, $T = 9$, estimated by `gllamm`, 10000 repetitions.
The t-test is for equality of the average estimated parameter and the true value (two-sided test).

So far, the worst news in terms of consistency is the bias of ca. 1 per cent on the coefficient of the feedback variable. In the next section, we will examine if this changes with the introduction of random slopes.

6.2 Random intercept and random slopes: Experiment 7

We introduce random slopes into the model, in the same way as we have done in previous sections. Since all three random effects are correlated with each other according to (6), we indirectly induce correlation not only between the initial conditions and $u0$ but also between the initial conditions and $u1$ and ut. For the reason that this model is the most complicated (and possibly demanding in terms of estimation), we have collected slightly more observations for `gllamm` for this model (962 observations). Results for `xttobit` are listed in table 13, and those for `gllamm` in table 14.

Compared to experiment 5, the biases for `xttobit` have largely remained the same, with only the bias on the time trend increasing by 40 per cent to 5.14 per cent. For `gllamm`, biases on the parameters of the main model are all below one per cent. For the elements of the Cholesky decomposition, average estimates are even a little closer to the true values than they were in experiment 5. Worst affected is the individual effect $u0$, which does not come as a surprise, given our implementation of correlated feedback.

Overall, the results are encouraging. From a theoretical viewpoint, the estimator implemented in `gllamm` is inconsistent for our model; however, even under correlated feed-

Table 13: Results for experiment 7: `xttobit`

Parameter	true value	average estimate	bias in %	standard deviation	average estimated s.e.	p-value t-test	p-value normality test
β_1	0.4	0.42	3.90	0.06	0.04	0.00	0.42
β_2	-3	-3.00	0.04	0.09	0.08	0.18	0.00
β_3	-2.5	-2.50	0.08	0.17	0.17	0.24	0.04
β_t	-0.7	-0.66	5.14	0.13	0.13	0.00	0.27
β_{d10}	-7	-6.91	1.35	1.03	1.04	0.00	0.05
β_{d9}	-3	-3.00	0.12	0.89	0.90	0.69	0.50
β_0	22	21.69	1.40	0.96	0.98	0.00	0.85
σ_ε	6	6.83	13.80	0.24	0.22	0.00	0.00
σ_{u0}	6	6.65	10.80	0.46	0.44	0.00	0.00

Three random effects. $N = 200$, $T = 9$, estimated by `xttobit`, 10000 repetitions.
The t-test is for equality of the average estimated parameter and the true value (two-sided test).

Table 14: Results for experiment 7: `gllamm`

Parameter	true value	average estimate	bias in %	standard deviation	average estimated s.e.	p-value t-test	p-value normality test
β_1	0.4	0.40	0.79	0.06	0.05	0.08	0.30
β_2	-3	-3.00	0.09	0.08	0.08	0.29	0.04
β_3	-2.5	-2.49	0.24	0.15	0.16	0.23	0.84
β_t	-0.7	-0.70	0.35	0.13	0.12	0.57	0.46
β_{d10}	-7	-7.00	0.05	1.02	0.99	0.92	0.98
β_{d9}	-3	-3.00	0.12	0.85	0.85	0.90	0.35
β_0	22	21.97	0.12	0.94	0.93	0.39	0.19
$\log(\sigma_\varepsilon)$	1.79	1.79	0.40	0.04	0.04	0.00	0.73
Cholesky(1,1)	6	6.07	1.16	0.91	0.93	0.02	0.15
Cholesky(2,2)	0.484	0.47	2.80	0.05	0.05	0.00	0.16
Cholesky(3,3)	0.378	0.28	26.60	0.23	0.38	0.00	
Cholesky(2,1)	-0.125	-0.12	2.69	0.10	0.09	0.28	0.05
Cholesky(3,1)	0.08	0.06	24.50	0.16	0.17	0.00	0.01
Cholesky(3,2)	-0.103	-0.11	1.63	0.14	0.14	0.72	0.01
Var(u0)	36	37.66	4.60	10.96	(11.38)	0.00	0.00
Var(u1)	0.25	0.25	0.91	0.06	(0.06)	0.20	0.00
Cov(u0,u1)	-0.75	-0.77	2.26	0.62	(0.62)	0.40	0.02
Var(ut)	0.16	0.19	18.30	0.13	(0.13)	0.00	0.00
Cov(u0,ut)	0.48	0.26	46.40	0.97	(0.97)	0.00	0.00
Cov(u1,ut)	-0.06	-0.06	7.47	0.07	(0.07)	0.03	0.04

Standard deviations in parentheses are calculated from the estimated variance-covariance matrix using the delta method.

$N = 200$, $T = 9$, estimated by `gllamm`, 962 repetitions.
The t-test is for equality of the average estimated parameter and the true value (two-sided test).

back and with a sample of just 200 cross-sectional units, biases on the main parameters are below one per cent. Even the standard errors are estimated well by `gllamm` (only the conventional maximum likelihood variance estimator was used in the experiments, without adjustment for clustering).

7 Heterogeneity bias

In this section we address the question of how large the *endogeneity bias* is in our model relative to the size of the *heterogeneity bias*. We encounter heterogeneity bias when the true model features individual heterogeneity but we estimate a simpler model without accounting for heterogeneity. In nonlinear models, the population-averaged approach (i. e. pooled estimation) generally leads to inconsistent estimates (Cameron and Trivedi, 2009, p.603). We have already seen this bias in experiments 2, 5, and 7, where we reported (misspecified) `xttobit` estimates for a model with random slopes. In what follows, we take a closer look at the magnitude of the heterogeneity bias and its relation with the endogeneity bias in our model.[9] Figure 2 shows the biases in per cent for each of the main parameters under the various feedback conditions. The black bars represent the biases when estimating our model, which has three individual random effects, using a simple tobit estimation, thus ignoring all heterogeneity. The grey bars represent the estimates from a random intercept tobit model (using `xttobit`), and the white bars represent the estimates from `gllamm` where all three random effects have been specified correctly.[10]

The `gllamm` estimates do not suffer from heterogeneity bias in any of the three feedback conditions. The `xttobit` estimates and `tobit` estimates do suffer from this bias. The heterogeneity bias ought to be strongest in the `tobit` estimates, and this is confirmed by figure 2.

The endogeneity bias occurs whenever we have individual heterogeneity *and* feedback. The top left panel in figure 2 therefore depicts a situation where no endogeneity bias is present; the data generating process is described in equation (9). The bias on the coefficient β_1 gives us an estimate of the pure heterogeneity bias in our model: 7 per

[9] In a different context, Wilcox (2006) studies the heterogeneity bias in learning models. He finds that the bias is very large, but can be greatly reduced by applying random coefficients estimators even if these are misspecified.

[10] Note that the estimates for `xttobit` and `gllamm` (the grey and white bars) have been discussed in detail in the previous sections.

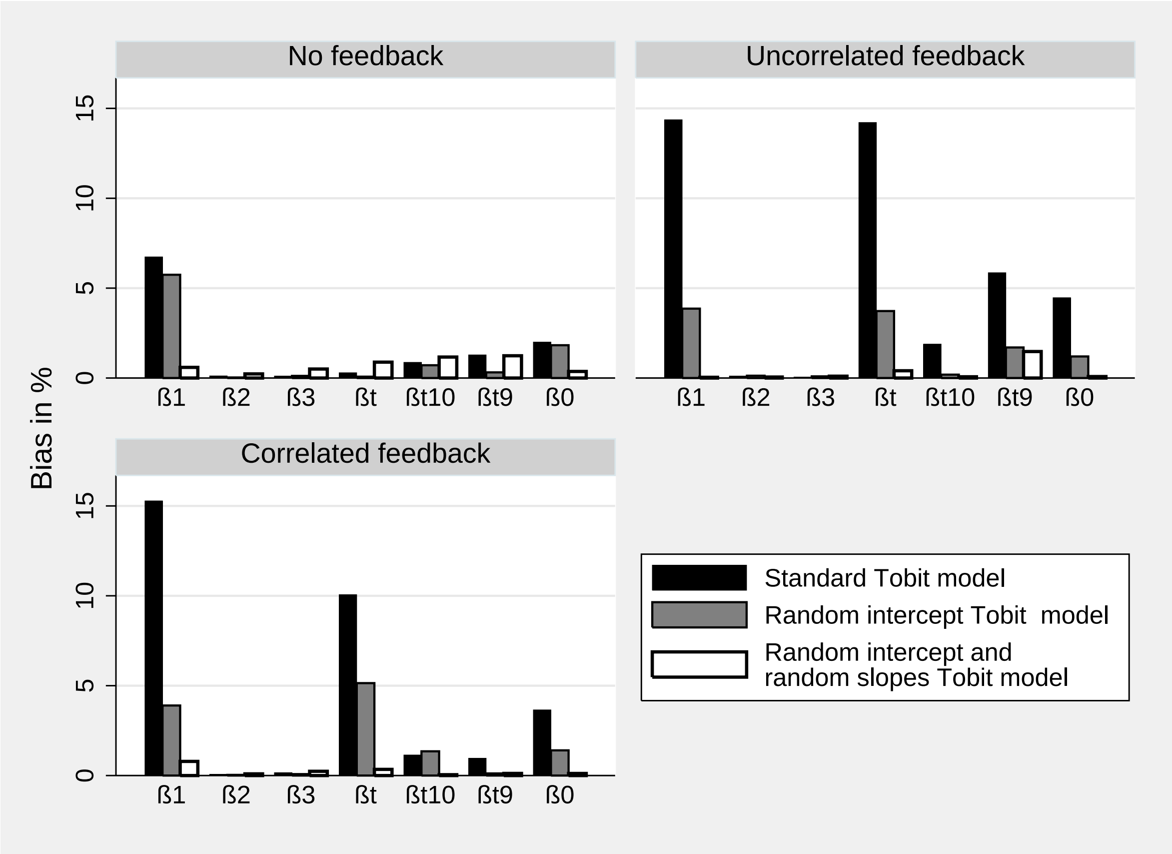

Figure 2: Heterogeneity and endogeneity bias

cent for the simple tobit model, and 6 per cent for the random intercept model. We also have an estimate of the pure endogeneity bias in figure 2: the white bars in the bottom left and right hand side panels. They depict a situation where endogeneity is present, but individual heterogeneity has been modelled accurately. The endogeneity bias on β_1 is 0.8 per cent.

When adding endogeneity to incorrectly modelled heterogeneity, it appears that the two types of biases reinforce each other, at least for the case of the simple tobit estimates. In the right hand side panel, we have the situation of endogeneity through feedback, but without correlation of initial values and individual effects (see equation (10)). Biases on β_1 and β_t are substantial when estimating by simple tobit. Upon adding correlation to the feedback process (bottom left panel in the graph), the bias on β_1 tops 15 per cent, showing a very poor performance of the simple tobit model under heterogeneity and endogeneity. The random intercept model performs better, but still produces biases up to 5 per cent; more than six times as much as the pure endogeneity bias that we encounter with the `gllamm` estimates. The *direction* of the biases is such that we overestimate the feedback effect β_1 and underestimate the (negative) time effect β_t.

This leads to the conclusion that using a simple tobit model where individual heterogeneity is present results in significant biases in the parameter estimates, and potential endogeneity only makes things worse. If such a model was estimated in the presence of heterogeneity and endogeneity, the researcher would find a feedback effect which is higher than the true average feedback effect, and a time effect which is lower than the true average time effect, thus shifting the interpretation in favour of conditional cooperation and away from strategic contributions.[11] The `gllamm` estimates, on the other hand, perform very well, with a low endogeneity bias of ~ 0.8 per cent. The simple tobit model is only appropriate in a situation without any individual heterogeneity; in this case it is robust to feedback effects, as we demonstrated in table 6.

While the small endogeneity bias is good news for our proposed model, it is unclear whether our results will still hold under different censoring rates. We suspect that the censoring helps to reduce the correlation between unobserved effects and the feedback variable in the 'Correlated Feedback' condition. While it is infeasible to repeat all experiments with different rates of censoring, we attempt to shed light on the issue in the next section where we investigate the effect of the censoring of the feedback variable.

8 The effect of feedback censoring on the bias

In our experiments with feedback we applied censoring to the feedback variable, since $\overline{y^*}_{t-1,p-i}$ is unobservable. This procedure is likely to reduce the correlation between the regressor and the individual effect $u0_i$ (the correlation will be zero in the censored domain). Given our relatively high censoring rate of 60 per cent, this dampening effect might be substantial. In this section, we repeat experiments 4 and 6, this time with uncensored feedback. While uncensored feedback is an unrealistic data generating process in our case, it will help understand how much the censoring affects the feedback bias. If removing the censoring on the feedback substantially increases the bias, the results provide evidence that with a lower censoring rate the bias might be worse.

[11] We mention conditional cooperation and strategic contributions merely as examples of how one might interpret the two coefficients in the context of a public good. We do not deal with the issue of conditional cooperation, and how to measure it, in this work; our research question is purely related to the statistical performance of an estimator.

All parameters of the model remain the same. Estimation is done with the `xttobit` routine. The only change to the data generating process is that $\overline{y}_{t-1,p-i}$ is replaced by $\overline{y^*}_{t-1,p-i}$:

$$
\begin{aligned}
y^*_{i,t} &= \beta_0 + \beta_1 \overline{y^*}_{t-1,p-i} + \beta_2 x2_{i,t} + \beta_3 x3_{i,t} + \beta_{d10} d(t=10) \\
&\qquad + \beta_{d9} d(t=9) + \beta_t t + u0_i + \varepsilon_{i,t} \\
y_{i,t} &= \begin{cases} 0 & \text{if } y^*_{i,t} < 0, \\ y^*_{i,t} & \text{if } 0 \leq y^*_{i,t} \leq 20, \\ 20 & \text{if } y^*_{i,t} > 20 \end{cases}
\end{aligned}
\tag{11}
$$

We restrict our analysis to the case of a single random intercept $u0$, for reasons of computational costs.

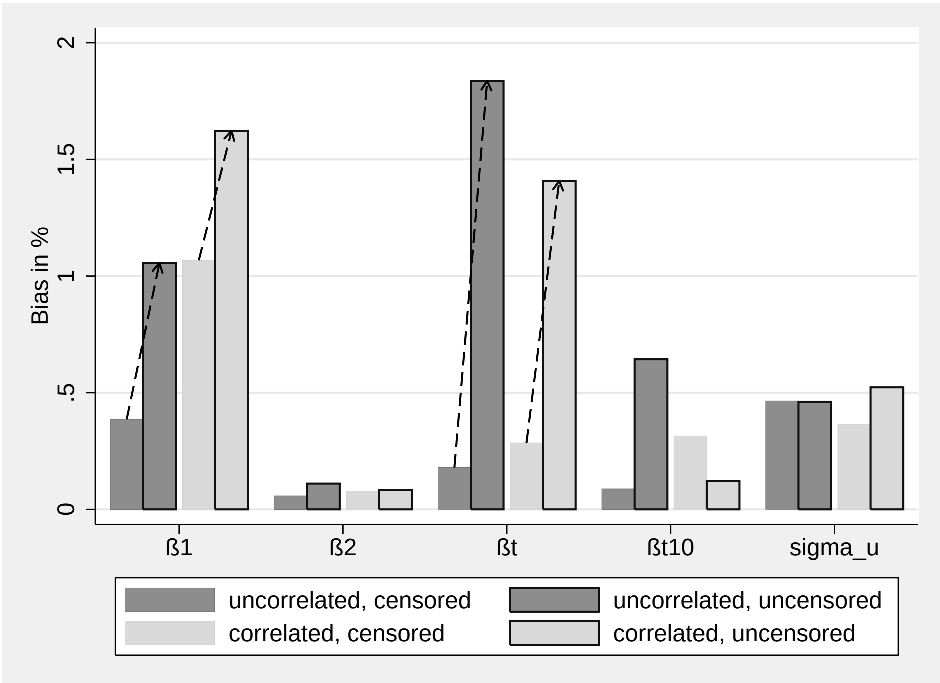

Figure 3: Effect of removing the censoring from the feedback variable

Figure 3 depicts the effect of removing the censoring of the feedback variable on the bias of selected coefficients. Generally, biases increase when removing the censoring; the strongest effects have been highlighted with arrows in the graph. For β_1, for example, the bias more than doubles in the case with uncorrelated feedback. Interestingly, the time trend is now estimated with a much larger bias as well. On the other hand, the bias on the coefficient of $x2$ (which is strictly exogenous in our model) is hardly affected at all. The exercise confirms our suspicion that the size of the feedback bias in the tobit model is related to censoring. We would expect a similar effect from reducing the overall censoring rate; however, to confirm this, more simulations are required. This is left to future research. It is worth pointing out that even with uncensored feedback, biases are still much lower than the heterogeneity biases we found in the previous section. Full results of the experiments in this section are reported in tables 17 and 18 in the appendix.

9 Lagged dependent variable

Building on our results so far, we attempt to understand how the correlation between observations within a group affects the estimation. Our feedback mechanism (1) has two effects: on the one hand, it creates a situation similar to a lagged dependent variable, and on the other hand it gives rise to correlations between observations in a group. To investigate, we create experiments 8 and 9, where we remove the feedback and add a lagged dependent variable. Observations within a group will be independent in this setup.

Panel tobit models with a lagged dependent variable have been studied in the literature; see for example Arellano, Bover et al. (1997) and Honoré (1993). As explained in the introduction, OLS and the random effects estimator will be biased. Our aim is to study the magnitude of the random effects estimator bias in a situation where one explanatory variable is predetermined, but observations are independent between cross-sectional units. We can then compare the findings to our results from experiments 4/4a and 6/6a. For the same reasons as in the previous section, we restrict our analysis to the case of a single random intercept.

9.1 Uncensored lag: Experiment 8

In the spirit of the previous section, we conduct our lagged dependent variable experiment both with and without censoring of the lag. We expect the bias to be worse in the uncensored case. The data generating process for experiment 8 is as follows:

$$\begin{aligned} y^*_{i,t} &= \beta_0 + \beta_1 y^*_{i,t-1} + \beta_2 x2_{i,t} + \beta_3 x3_{i,t} + \beta_{d10} d(t=10) \\ &\qquad + \beta_{d9} d(t=9) + \beta_t t + u0_i + \varepsilon_{i,t} \\ y_{i,t} &= \begin{cases} 0 & \text{if } y^*_{i,t} < 0, \\ y^*_{i,t} & \text{if } 0 \leq y^*_{i,t} \leq 20, \\ 20 & \text{if } y^*_{i,t} > 20 \end{cases} \end{aligned} \tag{12}$$

All parameters stay exactly the same as in the previous sections. The latent, unobservable y^* appears in lagged form as regressor. From table 15 we can see that biases

Table 15: Results for experiment 8: `xttobit`

Parameter	true value	average estimate	bias in %	standard deviation	average estimated s.e.	p-value t-test	p-value normality test
β_1	0.4	0.42	3.78	0.02	0.02	0.00	0.00
β_2	-3	-3.01	0.37	0.08	0.08	0.00	0.00
β_3	-2.5	-2.51	0.33	0.16	0.16	0.00	0.10
β_t	-0.7	-0.65	7.14	0.12	0.12	0.00	0.00
β_{d10}	-7	-7.10	1.43	1.02	1.02	0.00	0.40
β_{d9}	-3	-3.10	3.23	0.88	0.87	0.00	0.22
β_0	22	21.76	1.09	0.86	0.86	0.00	0.01
σ_ε	6	6.01	0.15	0.20	0.20	0.00	0.00
σ_{u0}	6	5.80	3.37	0.42	0.42	0.00	0.24

Uncensored lagged dependent variable. $N = 200$, $T = 9$, estimated by `xttobit`, 10000 repetitions. The t-test is for equality of the average estimated parameter and the true value (two-sided test).

go up significantly for most variables. Especially the coefficient on the lagged dependent variable, β_1, and the linear time trend are affected, the latter having the largest bias of 7.14 per cent. It becomes clear why different estimation strategies need to be considered for lagged dependent variables. $y^*_{i,t-1}$ is unobservable, though, and it will be interesting to see how the biases react to censoring of the lagged variable. This is being done in the next experiment.

9.2 Censored lag: Experiment 9

We now repeat experiment 8 with the modification of censoring of the lag. The new data generating process is

$$y^*_{i,t} = \beta_0 + \beta_1 y_{i,t-1} + \beta_2 x2_{i,t} + \beta_3 x3_{i,t} + \beta_{d10} d(t=10) + \beta_{d9} d(t=9) + \beta_t t + u0_i + \varepsilon_{i,t}$$

$$y_{i,t} = \begin{cases} 0 & \text{if } y^*_{i,t} < 0, \\ y^*_{i,t} & \text{if } 0 \leq y^*_{i,t} \leq 20, \\ 20 & \text{if } y^*_{i,t} > 20 \end{cases} \tag{13}$$

with all the parameters remaining at their previous values. Instead of the unobservable $y^*_{i,t-1}$, we include $y_{i,t-1}$ which can be observed. Comparing table 16 with table 15, we

Table 16: Results for experiment 9: `xttobit`

Parameter	true value	average estimate	bias in %	standard deviation	average estimated s.e.	p-value t-test	p-value normality test
β_1	0.4	0.40	0.14	0.03	0.03	0.05	0.84
β_2	-3	-3.00	0.08	0.08	0.08	0.00	0.00
β_3	-2.5	-2.50	0.05	0.15	0.15	0.40	0.01
β_t	-0.7	-0.70	0.36	0.12	0.12	0.03	0.05
β_{d10}	-7	-7.01	0.07	0.92	0.92	0.58	0.47
β_{d9}	-3	-3.01	0.41	0.81	0.80	0.13	0.47
β_0	22	22.03	0.14	0.90	0.89	0.00	0.18
σ_ε	6	5.98	0.30	0.19	0.19	0.00	0.01
σ_{u0}	6	5.98	0.31	0.41	0.40	0.00	0.01

Censored lagged dependent variable. $N = 200$, $T = 9$, estimated by `xttobit`, 10000 repetitions.
The t-test is for equality of the average estimated parameter and the true value (two-sided test).

can see that the biases are on average dramatically smaller. Biases are now well below 1 per cent, and for four parameters, the null hypothesis of the t-test cannot be rejected at 5 per cent level. This confirms the dampening effect of censoring on the endogeneity bias that we already discovered for the feedback case.

How big is the bias in the case of experiment 9 (endogeneity bias, but independent observations between cross-sectional units) compared to experiment 6 (endogeneity bias and correlated observations within a group; see table 11)? In both experiments, the variable causing the endogeneity bias is censored. Biases in the correlated feedback case are up to 7 times as big as those in the lagged dependent variable case. This provides evidence that a large part of the problem is caused by the intragroup correlation.

However, in the experiments without censoring of the feedback or lagged dependent variable, the converse is true. The biases under correlated feedback are significantly lower (by as much as factor 11) than those under lagged dependent variable. This shows that the two effects - censoring and intragroup correlation - interact with each other. Intragroup correlation appears to make the problem worse, while censoring of the dependent variable has a dampening effect on bias.

10 Larger groups, bigger problem?

We have decided on a group size of four in this study, reflecting the situation of our own data. Given our results so far, it is likely that group size has a significant effect on bias in the random effects tobit model with feedback.[12] The direction of the effect is a priori unclear: if, for a given censoring rate, the bias stems mainly from the induced correlation between observations within a group, then it should increase with group size. If the bias originates mainly from the correlation between a regressor and the (own) individual effect, we would not expect it to move around a lot; since the feedback effect is formulated as an average, a larger groups means that the feedback concerning u_0 is nominally smaller, but it comes from a larger number of individuals. In the previous section, we found some evidence that intragroup correlation may be a big part of the problem; if this is the case, we should find now that biases increase with group size. But we also found evidence that the dampening effect of censoring takes over.

We repeated experiment 6 (correlated feedback) for group sizes two to ten,[13] and estimated with `xttobit`. All parameters of the model remain the same, except the number of cross-sectional units which has to be slightly adjusted to allow for different group sizes. Figures 4 summarises the results; detailed results are listed in the appendix.

Figure 4 shows the average bias in per cent versus group size for four coefficients in the model: β_1 (the coefficient of the feedback variable), β_2 (the coefficient of a strictly

[12] This is a very different aspect to the effect of group size on contributions, which has been examined in several economic experiments (see for example Isaac and Walker (1988) and Isaac, Walker et al. (1994)). Here we are concerned with the statistical properties of an estimator.

[13] Especially for the larger group sizes, a higher number of repetitions was needed in order to achieve stable results. The experiments in this section were conducted with in between 19400 and 30000 repetitions.

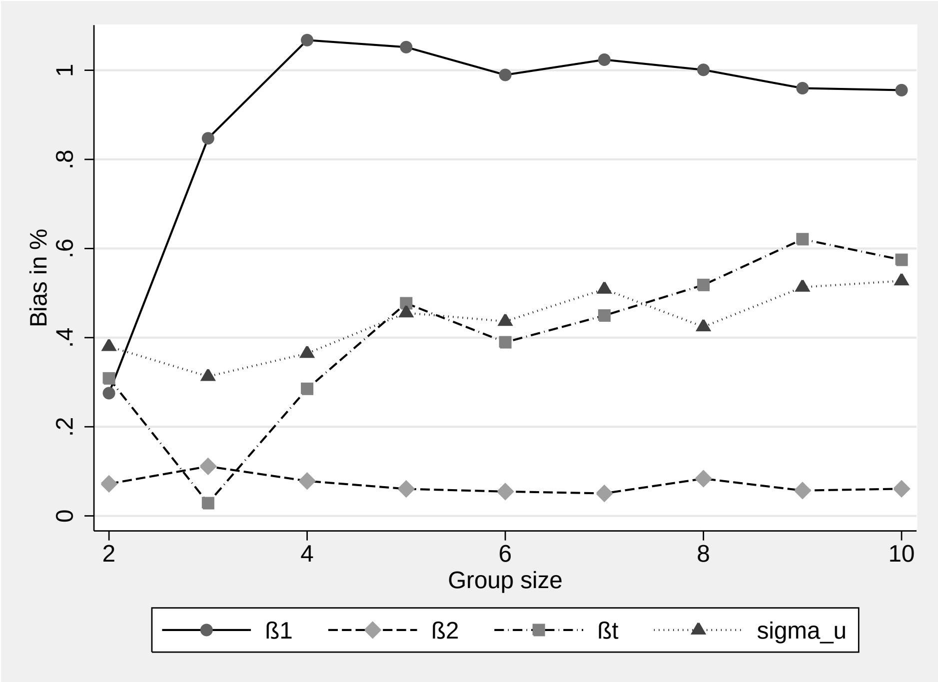

Figure 4: Average bias of β_1, β_2, β_t and σ_{u0} for various group sizes

exogenous regressor), β_t (the coefficient of the linear time trend), and the standard deviation of the random effect, σ_u. The bias of β_1 starts out very low for a group size of two, and increases quickly until it reaches 1.05 per cent for a group size of four (this is experiment 6). For all group sizes larger than four, the bias does not increase any more but remains at around one per cent. At the same time, the standard deviation of the estimates of β_1 increases from 0.028 for a group size of two to 0.053 for a group size of ten (not shown). This is to be expected: variation in $x1$ is taken away by averaging. The larger the groups, the less variation there is, and the less precise is the estimate.

Since $x2$ is exogenous, its coefficient should be affected much less by any endogeneity problem in $x1$ than β_1. This is confirmed in our experiments, where the bias of this coefficient remains low for all group sizes. For the standard deviation of the random effect, σ_u, and the coefficient of the time trend β_t, the bias increases only slightly with group size.

Overall, we find some support for biases increasing with group size, although the effect appears to flatten as groups grow larger. Here, we may have identified a 'turning point' where the adverse effect of larger groups on correlation is outweighed by the dampening of censoring. Reassuringly, the bias on parameters of strictly exogenous variables appear to be small and unaffected by group size.

11 Summary and conclusions

We study the random effects estimator under a specific feedback setup that is frequently encountered in economic experiments. We stipulated a simple feedback mechanism which, in theory, renders the random effects estimator inconsistent by introducing an endogeneity bias. We proceed to make the mechanism more realistic by introducing correlation between the individual effect and the initial value of the dependent variable.

11.1 Performance of `gllamm`

One goal of this study was to evaluate the performance of `gllamm` when estimating Tobit models in Stata. Our benchmark case is experiment 1, where we implemented only a single random effect and no feedback. `gllamm` produces estimates of both the coefficients and standard errors that are very similar to `xttobit`. We conclude that the estimator implemented in `gllamm` consistently estimates Tobit models with a random intercept under regular conditions. In experiment 2, we added random slopes to the setup. The main coefficients of the model were still estimated well, while some elements of the Cholesky decomposition of the covariance matrix were biased. We think that four factors are likely to contribute towards the bias: a) small number of repetitions (due to computational cost); b) small number of cross-sectional units; c) small values in the true covariance matrix; d) high censoring rate. The question of whether or not the bias disappears under different conditions has to be left for future research.

When introducing feedback into the model, `gllamm`'s performance remained very similar to `xttobit`. The problems in estimating the Cholesky decomposition got worse with the introduction with feedback; this may have to do with the reduction in the effective sample size due to feedback. However, the main parameters are estimated well by `gllamm`. Simply ignoring the random slopes and estimating a random intercept model leads to significant biases on the main parameters of the model.

11.2 Uncorrelated feedback

Introducing uncorrelated feedback into the Tobit model in the presence of random effects leads to some bias in the main coefficients. The size of the bias is small (below 0.5 per cent). While researchers always prefer unbiased estimators, our results are encouraging in the sense that the problem may not be as bad as originally thought.

11.3 Correlated feedback

In the next setup, we induce a correlation of about 60 per cent between the individual effect and the initial value $y_{i,1}$. This leads, as expected, to an increase in bias, most notably for the coefficient of the feedback variable β_1; its bias is now ~ 1.05 per cent. Again, in increase in bias is certainly bad news; however, the bias is still very small, and certainly small compared to the bias resulting from ignoring the random slopes.

11.4 Heterogeneity bias

We find strong evidence for the presence of heterogeneity biases in our model. Especially pooled tobit estimation of the heterogeneous slopes model results in a large upwards bias on the feedback effect β_1 and a large downwards bias on the (negative) time effect β_t, thus distorting the interpretation of results away from strategic effects.

11.5 The effect of censoring

We examine if the small size of the bias may be due to the high rate of censoring, which eradicates some of the endogeneity problem. While it is infeasible to repeat all experiments with different rates of censoring, we repeat two experiments where we remove the censoring on the feedback variable only. The results of this exercise are very clear: biases increase dramatically, especially for the feedback variable and the linear time trend. We conclude that censoring has a dampening effect on the bias by 'swallowing' some of the correlation induced through the feedback.

11.6 The effect of intragroup correlation

Another important aspect of the feedback process is intra-group correlation. We attempt to single out the effect of intra-group correlation by conducting two experiments with lagged dependent variables instead of feedback variables. In this scenario, the endogeneity problem is still present while the observations across subjects are independent. For the censored case, we find that the biases are smaller when the intra-group correlation is removed, providing evidence for an adverse effect of correlation within groups on the properties of the estimator. However, for the uncensored case, we find the opposite result: biases are larger under independence. We conclude that there is some interaction between the dampening effect of feedback and the adverse effect of intra-group correlation, which needs to be examined in detail in further research.

11.7 The effect of group size

Finally, we investigate the effect of group size. The amount of intra-group correlation increases with group size, while the endogeneity problem due to the correlation between a regressor and the (own) individual effect remains at the same level. Unsurprisingly, after what we learned in the previous section, we find a nonlinear effect of group size. The bias of β_1 increases until it reaches 1.05 per cent at a group size of four, and then remains at around one per cent as group size increases to ten. Our results hint at an interaction of censoring, intra-group correlation and endogeneity.

11.8 Conclusion

Our results show that there is an endogeneity bias when estimating the random effects censored regression model under feedback. The size of the bias, however, is very small. This is encouraging news for those studies that have applied this estimator under feedback in the past (for example Carpenter, Bowles et al. (2009)). The user-written software `gllamm` offers the possibility to estimate multi-level censored regression models with random intercepts and random slopes; models which have not been employed very much in the public goods literature. Given our results, we find that such models with feedback can be estimated approximately consistently in `gllamm`. Simply ignoring the random slopes if they are present in the true data-generating process results in very significant heterogeneity biases. Further studies are needed to examine the effect of the censoring rate on the endogeneity bias, as well as the effect of having larger groups.

References

Arellano, M., O. Bover and J. M. Labeaga (1997): Autoregressive models with sample selectivity for panel data. Unpublished manuscript.

Arellano, M. and B. Honoré (2001): Panel data models: some recent developments. In: J. Heckman and E. Leamer (Editors), *Handbook of Econometrics*, volume 5, Elsevier Science.

Baltagi, B. H. (2001): *Econometric Analysis of Panel Data*. John Wiley & Sons, Ltd.

Bardsley, N. and P. G. Moffatt (2007): The experimentrics of public goods: Inferring motivations from contributions. *Theory and Decision*, 62, pp. 161–193.

Cameron, A. C. and P. K. Trivedi (2009): *Microeconometrics using Stata*. Stata Press, College Station, Texas.

Carpenter, J. P. (2007): The demand for punishment. *Journal of Economic Behaviour & Organization*, 62, pp. 522–542.

Carpenter, J. P., S. Bowles, H. Gintis and S.-H. Hwang (2009): Strong reciprocity and team production: Theory and evidence. *Journal of Economic Behaviour and Organization*, in press.

Fischbacher, U., S. Gächter and E. Fehr (2001): Are people conditionally cooperative? evidence from a public goods experiment. *Economics Letters*, 71, pp. 397–404.

Godin, J.-F. (2008): Immigrant economic integration: A prospective analysis over 10 years of settlement. *International Migration*, 46(2), pp. 135–165.

Honoré, B. (1993): Orthogonality conditions for tobit models with fixed effects and lagged dependent variables. *Journal of Econometrics*, 59, pp. 35–61.

Honoré, B. and L. Hu (2004): Estimation of cross sectional and panel data censored regression models with endogeneity. *Journal of Econometrics*, 122, pp. 293–316.

Isaac, R. M. and J. M. Walker (1988): Group size effects in public goods provision: The voluntary contribution mechanism. *The Quarterly Journal of Economics*, 103(1), pp. 179–199.

Isaac, R. M., J. M. Walker and A. W. Williams (1994): Group size and the voluntary provision of public goods : Experimental evidence utilizing large groups. *Journal of Public Economics*, 54(1), pp. 1–36.

Ledyard, J. O. (1995): Public goods: A survey of experimental research. In: J. H. Kagel and A. E. Roth (Editors), *Handbook of Experimental Economics*, chapter 2, pp. 111–194, Princeton University Press.

Rabe-Hesketh, S. and A. Skrondal (2007): Multilevel and latent variable modeling with composite links and exploded likelihoods. *Psychoometrika*, 72(2), pp. 123–140.

Rabe-Hesketh, S., A. Skrondal and A. Pickles (2004): Gllamm manual. Technical Report 16, U.C. Berkeley Division of Biostatistics Working Paper Series.

—— (2005): Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects. *Journal of Econometrics*, 128(2), pp. 301–323.

Sell, J. (1997): Gender, strategies, and contributions to public goods. *Social Psychology Quarterly*, 60(3), pp. 252–265.

Shephard, N., M. Falcaro, E. Zeggini, P. Chapman, A. Hinks, A. Barton, J. Worthington, A. Pickles and S. John (2003): Linkage analysis of cross-sectional and longitudinally derived phenotypic measures to identify loci influencing blood pressure. *BMC Genetics*, 4(Suppl 1), p. S26.

Wilcox, N. T. (2006): Theories of learning in games and heterogeneity bias. *Econometrica*, 74(5), pp. 1271–1292.

Wooldridge, J. M. (2002): *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, Cambridge, Massachusetts.

Appendix

A Censoring and feedback

Table 17: Results for experiment 4a: `xttobit`

Parameter	true value	average estimate	bias in %	standard deviation	average estimated s.e.	p-value t-test	p-value normality test
β_1	0.4	0.40	1.06	0.02	0.02	0.00	0.22
β_2	-3	-3.00	0.11	0.08	0.08	0.00	0.00
β_3	-2.5	-2.50	0.06	0.16	0.15	0.37	0.57
β_t	-0.7	-0.69	1.84	0.13	0.13	0.00	0.63
β_{d10}	-7	-7.05	0.64	1.00	1.00	0.00	0.06
β_{d9}	-3	-3.03	0.86	0.84	0.84	0.00	0.47
β_0	22	21.93	0.32	0.92	0.92	0.00	0.05
σ_ε	6	5.98	0.35	0.19	0.19	0.00	0.01
σ_{u0}	6	5.97	0.46	0.40	0.40	0.00	0.00

Uncensored, uncorrelated feedback. $N = 200$, $T = 9$, estimated by `xttobit`, 10000 repetitions.
The t-test is for equality of the average estimated parameter and the true value (two-sided test).

Table 18: Results for experiment 6a: `xttobit`

Parameter	true value	average estimate	bias in %	standard deviation	average estimated s.e.	p-value t-test	p-value normality test
β_1	0.4	0.41	1.62	0.02	0.02	0.00	0.00
β_2	-3	-3.00	0.08	0.08	0.08	0.00	0.00
β_3	-2.5	-2.50	0.11	0.16	0.15	0.07	0.00
β_t	-0.7	-0.69	1.41	0.12	0.12	0.00	0.28
β_{d10}	-7	-7.01	0.12	1.00	0.99	0.40	0.01
β_{d9}	-3	-3.03	0.82	0.85	0.84	0.00	0.10
β_0	22	21.94	0.29	0.86	0.86	0.00	0.79
σ_ε	6	5.98	0.36	0.20	0.19	0.00	0.00
σ_{u0}	6	5.97	0.52	0.41	0.40	0.00	0.02

Uncensored, correlated feedback. $N = 200$, $T = 9$, estimated by `xttobit`, 10000 repetitions.
The t-test is for equality of the average estimated parameter and the true value (two-sided test).

B Group size: detailed results

Table 19: Bias and standard deviation for the experiments with varying group size

				Bias in %					
Group size	2	3	4	5	6	7	8	9	10
Coefficient									
β_1	0.28	0.85	1.07	1.05	0.99	1.02	1.00	0.96	0.96
β_2	0.07	0.11	0.08	0.06	0.05	0.05	0.08	0.06	0.06
β_3	0.07	0.14	0.14	0.11	0.07	0.04	0.08	0.07	0.14
β_t	0.31	0.03	0.29	0.48	0.39	0.45	0.52	0.62	0.58
β_{d10}	0.09	0.20	0.32	0.29	0.21	0.24	0.24	0.22	0.27
β_{d9}	0.03	0.16	0.32	0.42	0.40	0.16	0.52	0.31	0.53
β_0	0.08	0.03	0.12	0.14	0.13	0.15	0.14	0.17	0.15
σ_ε	0.35	0.29	0.33	0.36	0.32	0.33	0.29	0.32	0.35
σ_{u0}	0.38	0.31	0.37	0.46	0.44	0.51	0.42	0.51	0.53

				Standard deviation					
Group size	2	3	4	5	6	7	8	9	10
Coefficient									
β_1	0.028	0.032	0.036	0.040	0.042	0.046	0.048	0.052	0.055
β_2	0.077	0.077	0.077	0.078	0.077	0.078	0.077	0.079	0.078
β_3	0.151	0.149	0.150	0.149	0.149	0.148	0.150	0.151	0.151
β_t	0.110	0.112	0.113	0.115	0.116	0.119	0.122	0.124	0.126
β_{d10}	0.915	0.926	0.928	0.937	0.933	0.941	0.943	0.950	0.947
β_{d9}	0.797	0.809	0.806	0.804	0.802	0.810	0.807	0.820	0.817
β_0	0.844	0.864	0.885	0.905	0.911	0.939	0.956	0.987	0.993
σ_ε	0.187	0.187	0.190	0.188	0.187	0.188	0.188	0.192	0.191
σ_{u0}	0.391	0.391	0.395	0.393	0.388	0.392	0.390	0.395	0.393

				Parameters					
Group size	2	3	4	5	6	7	8	9	10
Parameter									
# of repetitions	21280	21664	21843	19406	30045	23230	21565	21111	20866
N	200	201	200	200	204	203	200	198	200
T	9	9	9	9	9	9	9	9	9