

WORKING PAPER SERIES

2018/002

Revisiting the Antisocial Punishment across
Societies Experiment

Pablo Lucas

*University College Dublin, School of Sociology and Geary
Institute, Ireland*

Issam Malki

*University of Westminster, Westminster Business
School, England*

All rights reserved

Pablo Lucas¹ and Issam Malki^{2, 1}

¹University College Dublin, School of Sociology and Geary Institute, Ireland; ²University of Westminster, Westminster Business School, Department of Economics and Quantitative Methods, England.

ARTICLE HISTORY:

Compiled January 2018.

ABSTRACT

We offer, in this paper, an alternative interpretation of an experimental public goods game dataset, particularly on the dynamics of the observed antisocial behaviour phenomenon between subjects of a public goods experiment in different cities around the world. The anonymous nature of contributions and punishments in this experiment are taken into account to interpret results. This is done by analysing dynamic behaviour in terms of mean contributions across societies and their association with antisocial punishment. By taking into account the heterogeneity between the cities in which the public goods experiment has been performed, this analysis show a contrasting interpretation. Instead of one trend across cities, two opposite trends are seen across different cities.

KEY WORDS

Convergence, experiments, panel data, public goods games and punishment.

¹ CONTACT: P. Lucas. pablo.lucas@ucd.ie. I. Malki. Email: i.malki@westminster.ac.uk.

1. Introduction

The behavioural and experimental economics literature has been steadily providing new insights on the phenomenon of punishment assigned to freeloaders and also to cooperators in public goods experiments. Hermann *et al* (2008a), for example, discuss an experimental setup allowing two types of treatment: with and without punishment amongst individuals. The main argument in the aforementioned reference is that “comparable social groups from complex developed societies with the widest possible range of cultural and economic backgrounds” present similar tendencies of antisocial punishment. The paper’s accompanying dataset is comprised of 16 different cities across the world where the experiment took place; each presenting a level of punishment towards those who behave pro-socially (i.e. individual investment in the public good).

Although Hermann *et al* (2008a) aimed to study antisocial punishment behaviour across different cities, it is unclear how such heterogeneity has been accounted for and harnessed in the experiment. This is as the experimental method implicitly assumed every society would be comparable (i.e. homogenous) in terms of societal characteristics including social, economic and political background. Indeed, Hermann *et al* (2008b)² reports country-level indicators for societies including aspects such as social capital, economic prosperity, law enforcement and democracy, cultural dimensions and value orientations. Yet the wide differences across these cities are apparent and these group and individual-level characteristics have not been accounted for.

The aim of this paper is to provide an alternative interpretation of the antisocial punishment phenomenon in light of the public goods game experiments conducted by Hermann *et al* (2008a). This goal is achieved by explicitly accounting for heterogeneity across cities and identifying clusters of those with similar antisocial behaviour, which form homogenous clusters (or blocks). This approach allows the investigation as to whether the relationship between antisocial punishment and mean contributions, as described in the original paper, holds for every different cluster in the dataset, or if it is a specific feature of certain groups of cities.

The paper is organised as follows. Section two discusses the experimental public goods game setup and the analysis of the resulting experimental data. After that, the econometric model and empirical

² See Table S1, page 6.

approach is introduced and discussed in detail. Findings are discussed in section 3 and section 4 concludes the paper.

2. Experimental setup and analysed data

The data associated with Hermann et al (2008a, 2008b) consists mainly of individual contributions and individual punishment levels, according to the public goods game experiments run with and without the possibility of punishment³. The setup consisted of a 10-round game administered in 16 different cities across the world, with varying number of participants. Individuals were anonymously assigned to groups of 4, with the aim of recording the numeric levels of observed cooperation and punishment among them. This was done using a zTree Fischbacher (2007) interface designed to collect data from participants regarding contributions and punishments using one desktop computer per recruited subject. Each participant received 20 tokens of which they needed to decide how many, per round, would be kept for themselves or otherwise invested in the public good. Each participant earned 0.4 tokens for each token invested, regardless of having contributed or not. Individual decisions were recorded independently and only the anonymised contributions of other group members could be seen after all participants having committed to a contribution amount. If punishments had been enabled in the experimental setup, each individual participant would also see the total amount of assigned punishments, but not who in their group has decided to do so. Punishments vary from 0 to 10 points, of which each does reduce the punished member's earnings by 3 tokens. Furthermore, the individual cost of punishing another group member was 1 token, per round in the corresponding experimental setup.

2.1. The econometric model and statistical test

For the purpose of this paper, we use mean contributions with and without punishment over the ten periods for all 16 cities. We begin by applying the approach proposed by Phillips and Sul (2007), PS hereafter. The approach proposes a simple econometric model, which can be used to cluster the individuals' cross-sections in clubs⁴ of convergence. The econometric model can be defined as a nonlinear transition factor model as:

$$C_{it} = \delta_{it} \mu_t \tag{1}$$

³ Further information is available on Hermann *et al* (2008b).

⁴ In this paper, the terms 'cluster' and 'club' are used interchangeably.

with C_{it} is contribution in the i -th city over the t periods. The model contains a common factor μ_t and an idiosyncratic element δ_{it} , which measure the deviation of contribution of the i -th individual trend over period t from the common factor μ_t . The latter can be interpreted as a common steady state of contribution in the panel. If individuals in every studied city behave similarly, then their contribution over each of the experimental periods would converge to a common steady state. This assumption can be tested using the convergence concept introduced by PS. The statistical test is based on the individual's contribution, which is measured by δ_{it} . This latter captures the individual contribution of the total as μ_t .

The idiosyncratic element is defined as:

$$\delta_{it} = \delta_i + \sigma_i \zeta_{it} L(t)^{-1} t^{-\alpha} \quad \delta_i \quad \zeta_{it} \quad L(t) \quad t^{-\alpha} \quad (2)$$

where δ_i fixed, $\zeta_{it} \sim iid(0,1)$ across individuals $i = 1, 2, \dots, N$ and dependent over time t . Finally, $L(t)$ is a slowly varying function of time, in which $L(t) \rightarrow \infty$ as $t \rightarrow \infty$.

Based on the formulation above, the null hypothesis of convergence is accepted if for all $\alpha \geq 0$ if $\delta_{it} \rightarrow \delta_i$.

The analysis of convergence and steady state equilibrium is based on studying the dynamic behaviour of a relative transition coefficient, h_{it} , which measures and captures the divergent behaviour of individuals' contribution from the steady state level of contribution μ_t . The relative transition coefficient is defined formally as:

$$h_{it} = \frac{X_{it}}{\frac{1}{N} \sum_{i=1}^N X_{it}} = \frac{\delta_{it}}{\frac{1}{N} \sum_{i=1}^N \delta_{it}} \quad (3)$$

Using this parameter, along with the loading coefficient δ_{it} , we assess the convergence. That is, namely, if $\delta_{it} \rightarrow \delta_i$, then $h_{it} \rightarrow 1$, so that the cross-sectional variance of h_{it} converges to zero and the result is:

$$\sigma_t^2 = \frac{1}{N} \sum_{i=1}^N (h_{it} - 1)^2 \rightarrow 0 \text{ as } t \rightarrow \infty \quad (4)$$

The property shown in (A.4) is essential in testing the null hypothesis of convergence and to cluster individuals into convergence clubs. The coefficient h_{it} , in addition to displaying the relative transition path for individual subjects in the experimental public goods panel data, measures and captures the divergent behaviour of individual subjects from what is the common stochastic trend or the long-run growth path μ_t . The procedure is implemented in two stages. The first stage is a test as to whether there is convergence. The second stage is a test as to whether there are different clubs of convergence.

In the first stage, the null hypothesis of overall convergence⁵ is tested, $H_0 : \delta_i = \delta$ and $\alpha \geq 0$, against the alternative hypothesis of no convergence, $H_0 : \delta_i \neq \delta$ for all i , or $\alpha < 0$. In order to test this hypothesis, a *logt* regression is used as proposed by PS, which is based on the cross-sectional variance ratio H_1/H_t where:

$$H_t = \frac{1}{N} \sum_{i=1}^N (h_{it} - 1)^2 \quad (5)$$

The regression is defined as follows:

$$\log\left(\frac{H_1}{H_t}\right) - 2 \log L(t) = \hat{a} + \hat{b} \log t + \hat{u}_t \quad (6)$$

where $t = [rT], [rT] + \dots, T$, with $r > 0$, $L(t) = \log(t + 1)$, $\hat{b} = 2\hat{\alpha}$ and $\hat{\alpha}$ is the estimated value of α in the null above. The regression is run starting at $t = [rT]$, which is the integer part of rT for some fraction $r > 0$. PS recommends using $r = 0.3$. Once the regression is run, the null is accepted if the autocorrelation heteroskedasticity robust one tail t_b statistic is above the critical value, c (e.g. at 5% level of significance, accept the null if $t_b \geq -1.65$).

⁵ Overall convergence refers to testing the convergence hypothesis across all states together.

If the null is rejected, this does not imply no convergence. It may in fact imply that relative convergence test should be performed; that is: the test is then about whether there are different clusters of convergence.

2.2 Clustering algorithm

PS propose the following steps to perform the clustering procedure:

Step 1 (Last observation ordering): individual series in the panel are ordered in descending fashion according to the most recent observation. In practice, however, PS put forward a recommendation of ordering the series according to its average over the last fraction of the sample for $r \in [0,1]$.

Step 2 (Core group formation): this stage identifies the initial club of series, consisting of at least two series, with the highest average in the panel (as described in Step 1). This group forms a *core* group, G_k for $2 \leq k < N$, against which the remaining series is to be compared. A sequence of *logt* regression is run and the test statistic $t_k = t(G_k)$ is estimated for this subgroup to choose the core group size k^* . The size k^* is determined according the following criterion:

$$k^* = \arg \max_k \{t_k\} \text{ subject to } \min\{t_k\} > -1.65 \quad (7)$$

This criterion ensures that the null of convergence is maintained for each k . If the condition (7) does not hold for all sequential pairs involved in the testing procedure, then we conclude that there is no convergence subgroup; otherwise, we have found the core subgroup, which is denoted by G_{k^*} .

Step 3 (Sieve individuals for membership): in this step, we define whether there are other individuals to be added to the core subgroup defined in step 2 from the remaining individuals, which form the complementary subgroup G_k^c . One individual from G_k^c is added one at a time to the core subgroup and the *logt* regression is implemented. If the *t statistic* estimated from this regression is above the critical value, then the individual chosen is to be added to the G_k^c , otherwise it is dropped. This procedure is repeated on the remaining individuals and the first convergence club is constructed. Next, estimate the *logt* regression using the data in this first convergence club and confirm that $t_b > -1.65$ for the whole cluster.

Step 4 (Stopping rule): in this step, the remaining individuals from step 3 are used to form a subgroup. The *logt* test is run and if $t_b > -1.65$ then the cluster convergence is formed. This implies there are two convergent clubs in the panel. Otherwise, steps 1-3 are repeated to search for smaller subgroups of convergent series according to the dataset. If the alternative hypothesis is accepted for all size of subgroups (i.e. no k in step 2 for which $t_b > -1.65$), then we conclude that the remaining individuals are divergent.

3. Discussion of findings

Figures 1 and 2 illustrate the general trends of contributions with and without punishment and the statistical association between contribution and punishment, respectively. Figures 1A and 1B are replications of the Figures 2A and 3 in Hermann et al (2008a) and illustrate the mean contribution before and after punishment to the public good over the 10 periods.

These graphs show that (i), in the experiment without punishment, individual contributions decline over the period in all cities and (ii), in the experiment with punishment, contributions tend to stabilise around the rather constant averages in every city recording higher contributions (i.e. conceptually grouped as Western societies) in comparison to contributions observed in other cities.

Figure 2 shows a statistically significant negative correlation between mean contributions and mean antisocial punishments (-0.90 [p -value= 0.00]⁶). This confirms the two observations above, as deduced from figures 1A and 1B.

[Figures 1 and 2 about here]

Although the aim of the graphs above is to compare the contribution across cities, these cannot be interpreted as a reflection of the typical behaviour of a society towards their contribution to the public goods, as the baseline is the average contribution across non-homogenous cities. In other words, the dynamic of contributions presented by these figures are not independently determined by each sample but rather by the aggregation effect caused by the sample selection bias in the form of omitting other cities in the analysis. This is illustrated in Figures 1C and 1D, where the concept of relative mean contribution is used. Figure 1C shows a different pattern from its counterpart, Figure 1A. Namely that (i) contribution does not decline in every experimented society and that (ii) those cities found to be contributing less (including Athens, Dnipropetrovs'k, Istanbul, Minsk, Muscat, Riyadh and Samara) have contributed relatively more –particularly towards the end rounds of the public good experiment. Figure 1D shows a different dynamic, clustered in patterns, when compared to Figure 1B. The original depiction of contribution after punishment seems to start at lower levels, yet there is an increase of contributions before declining again. Hence the relative mean contribution concept depicts those contributions that start at higher levels before declining.

The correlation coefficient and plot shown in Figure 2 is not strongly consistent across all cities. Various outlier cities are located far from the main cluster (these being Athens, Dnipropetrovs'k, Istanbul, Minsk, Muscat, Riyadh and Samara). Once the dataset is analysed separately according to these groups, one can see contrasting Spearman's correlation coefficients. These values differ from the analysis where all cities are included and thus provide the insight that different behavioural patterns are clustered in two different groups in this experimental dataset. These are: -0.58 [p -value= 0.09] when excluding outliers and -0.36 [p -value= 0.43] when outliers alone are considered.

Table 1 reports clustering based on PS algorithm. In both cases, with and without punishment, contribution does not converge. There are four estimated clusters in the case of contribution without punishment, while two are estimated when punishment is allowed. Figures 3A and 3B illustrate these clusters. Contributions with and without punishment are not identical across cities once society-

⁶ The estimate is taken from Hermann et al (2008a).

specific characteristics are accounted for. This suggests reviewing the interpretation of a uniform association between mean contributions and antisocial punishments in this dataset.

[Table 1 and Figures 3A-3B about here]

Figures 4A and 4B illustrate the correlation between mean contribution in clusters 1 and 2 and corresponding punishment, respectively. The result is different and the association depends on the location of cities in either cluster 1 or 2. Figure 4A shows that mean contribution declines with higher punishment. This is evidence of antisocial punishment in cities located in cluster 1, which is consistent with Hermann et al (2008a) (spearman=-0.62 [p -value=0.05]). Nevertheless, the association is reversed when considering cities of the second cluster with positive correlation (spearman=0.64 [p -value=0.05]), as illustrated in figure 4B.

[Figures 4A-4B about here]

Table 2 reports the Ordinary Least Square (OLS) estimates of the relationship between the dependent variable (i.e. group average contributions in period 2 to 10) and the set of independent variables including: group average contributions in period 1, group average punishment of free riding and group average antisocial punishment. The purpose of this empirical exercise is to confirm the extent to which the level of cooperation and punishment has on the overall performance of individual subjects in this public goods games experiment. It is within this context that the presence and effect of clubs of convergence is investigated. Column (1) in Table 2 replicates the original results by Hermann et al (2008), while Column (2) reports OLS estimates of the same empirical model but accounting for the clubs of contribution. According to Hermann et al (2008) findings, all independent variables are found to be statistically significant. While contribution in period 1 and free-riding punishment have a positive effect on the average contribution in periods 2 to 10, the antisocial punishment does substantially decrease the average contribution in periods 2 to 10.

The argument in this paper is that the dynamics of punishment and contribution is not consistent across all cities and societies. As a result of this, the role of punishment and cooperation is distorted if all cities and groups are pooled in the same sample with the assumption of a homogeneous contribution trend across the dataset. Hence, different conclusions regarding the role of free-riding and antisocial punishment are obtained when accounting for differences in contribution across cities according to the clubs of convergence. Further to this, column 2 in Table 2 suggests that when accounting for heterogeneous behaviour of contribution and punishment, free-riding and antisocial punishment become both statistically significant in explaining the contribution dynamics in those cities belonging to Club 2 (Athens, Dnipropetrovs'k, Minsk, Muscat, Riyadh, Samara). Conversely,

unlike Club 2, members of Club 1 (Bonn, Boston, Chengdu, Copenhagen, Istanbul, Melbourne, Nottingham, Seoul, St. Gallen, Zurich) are found to behave in the same way as originally reported in Hermann et al (2008).

[Table 2 about here]

4. Concluding remarks

In this paper, the extent in which the conclusions discussed in Hermann *et al* (2008) are investigated in terms of robustness when a different data analysis approach is deployed. We find the dynamics of contributions and its relationship with antisocial punishment being dependant on (i) the behaviour in each clustered group of experimented cities and (ii) the particular characteristics of each city. When simultaneously accounting for these, the antisocial punishment behaviour can be interpreted differently from the original discussion related to this experimental dataset. The experimental setup has no parallel with an actual public goods setup, so the results of analysing this dataset are limited regarding empirical insights (Weisberg (2004) and Guala and Salanti (2001)). This is as nowhere individuals pay for anonymous and direct punishments of other public good beneficiaries. Hence one cannot determine whether the original interpretation of the experimental dataset depends on the analytical methodology or on the experimental design assumptions (Levins, 1996).

As such, this paper provides an alternative analysis with a different interpretation to the experimental data, bearing in mind the limitations of the experiment design for observing antisocial punishment in it. In the context of alternative interpretation, we account for differences in the dynamics of contribution and punishment in public good games across cities. Unlike the original interpretation by Hermann et al (2008), we relax the homogeneity assumption about the dynamics of contribution across cities. This resulted in different dynamic paths of contribution, which show that contribution and antisocial punishment does not behave uniformly across cities. The main implication of the presence of heterogenous dynamics of contribution is the significant role of prosocial punishment and antisocial punishment. While the original work, by Hermann et al (2008), suggests that antisocial punishment is a prominent feature of contribution across cities, our results suggest that antisocial punishment is not existent in all cities and only affects a group cities. Antisocial punishment is found to have negative and significant role in cities in the first club (including Bonn, Boston, Chengdu, Copenhagen, Istanbul, Melbourne, Nottingham, Seoul, St. Gallen, Zurich) and found to have no effect on contribution dynamics in the cities identified in the second club (including Athens, Dnipropetrovs'k, Minsk, Muscat, Riyadh, Samara).

FIGURES

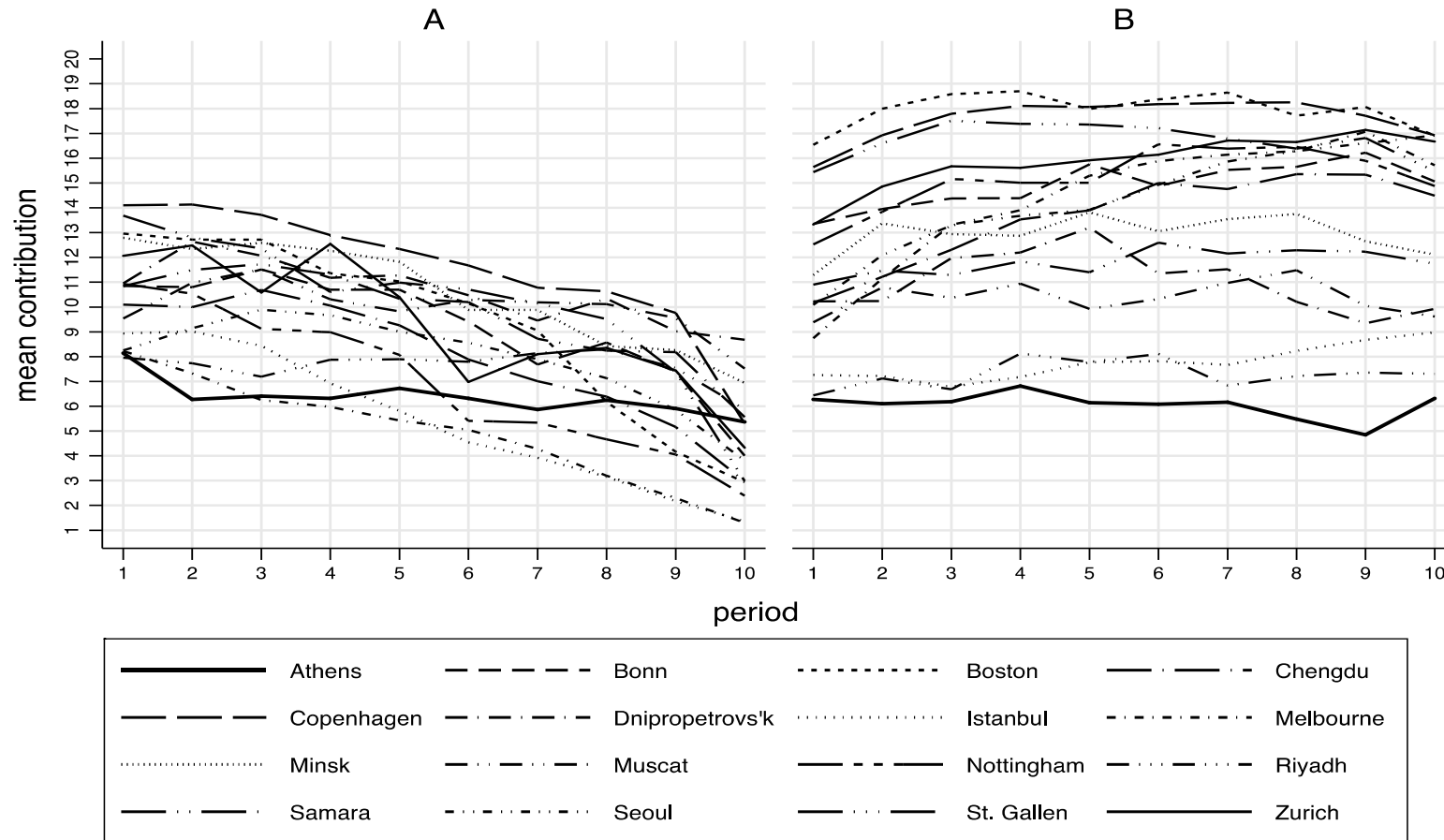


Figure 1: (A) mean contribution before and (B) after punishment

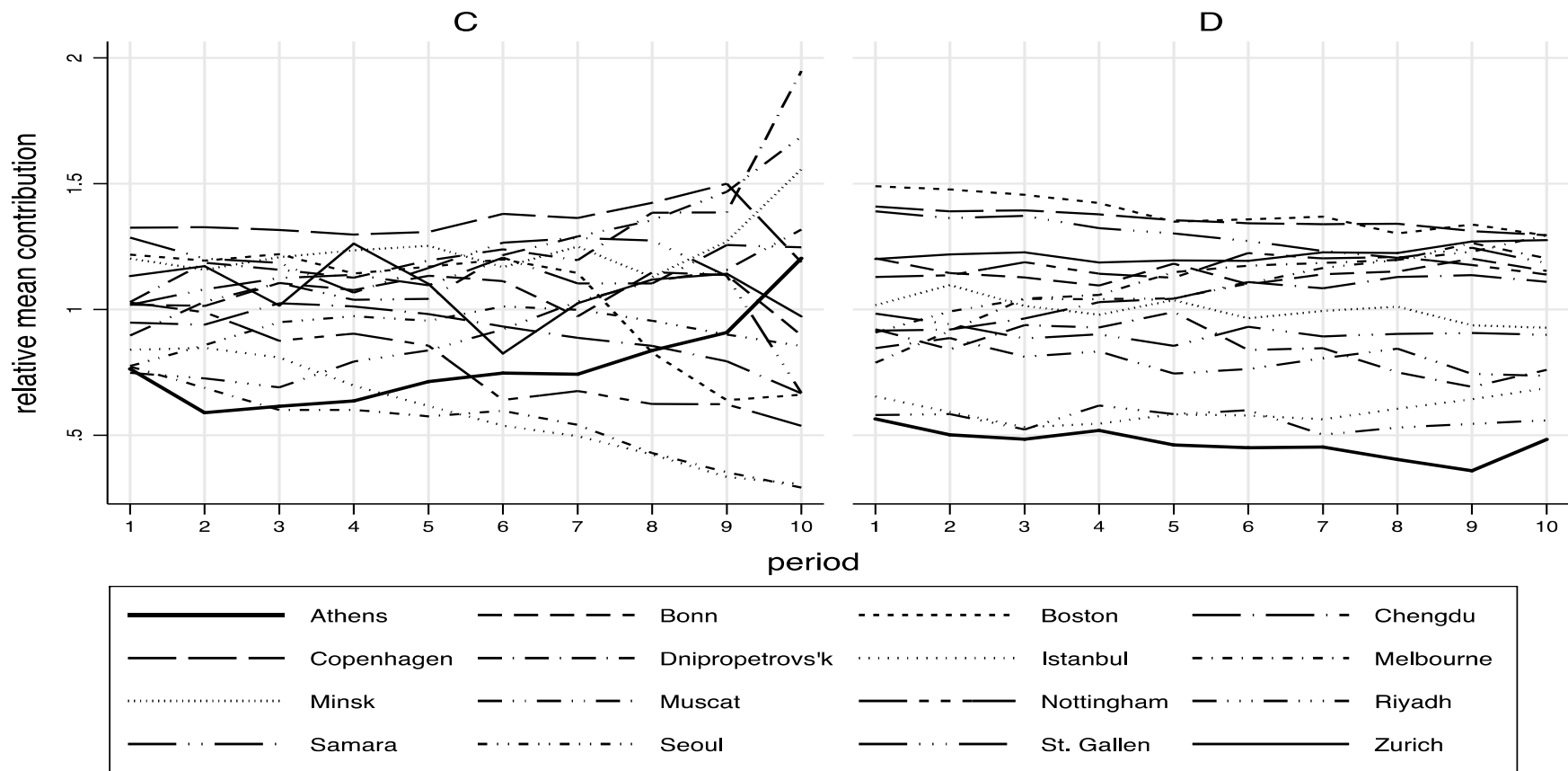


Figure 1: (C) relative mean contribution before and (D) after punishment

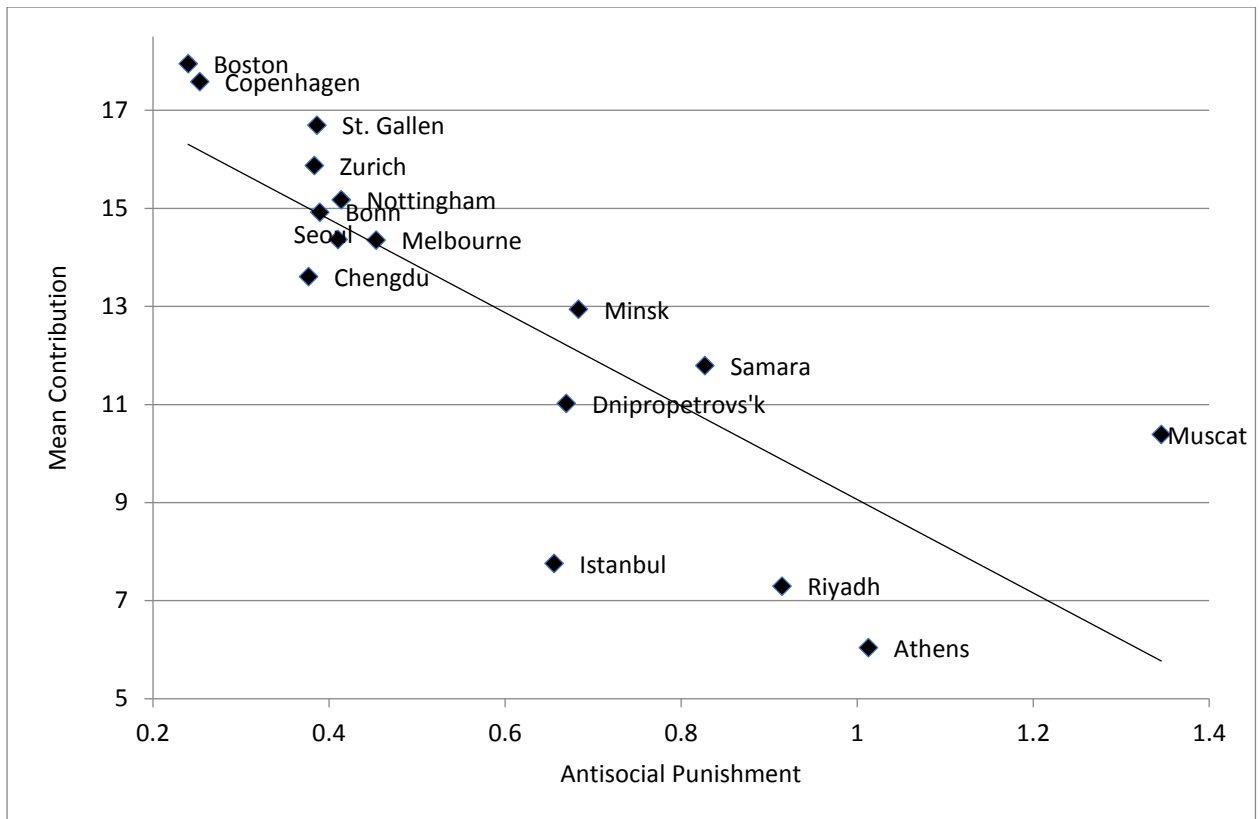


Figure 2: mean contribution against mean antisocial punishment (Spearman= -0.90 [0.00])

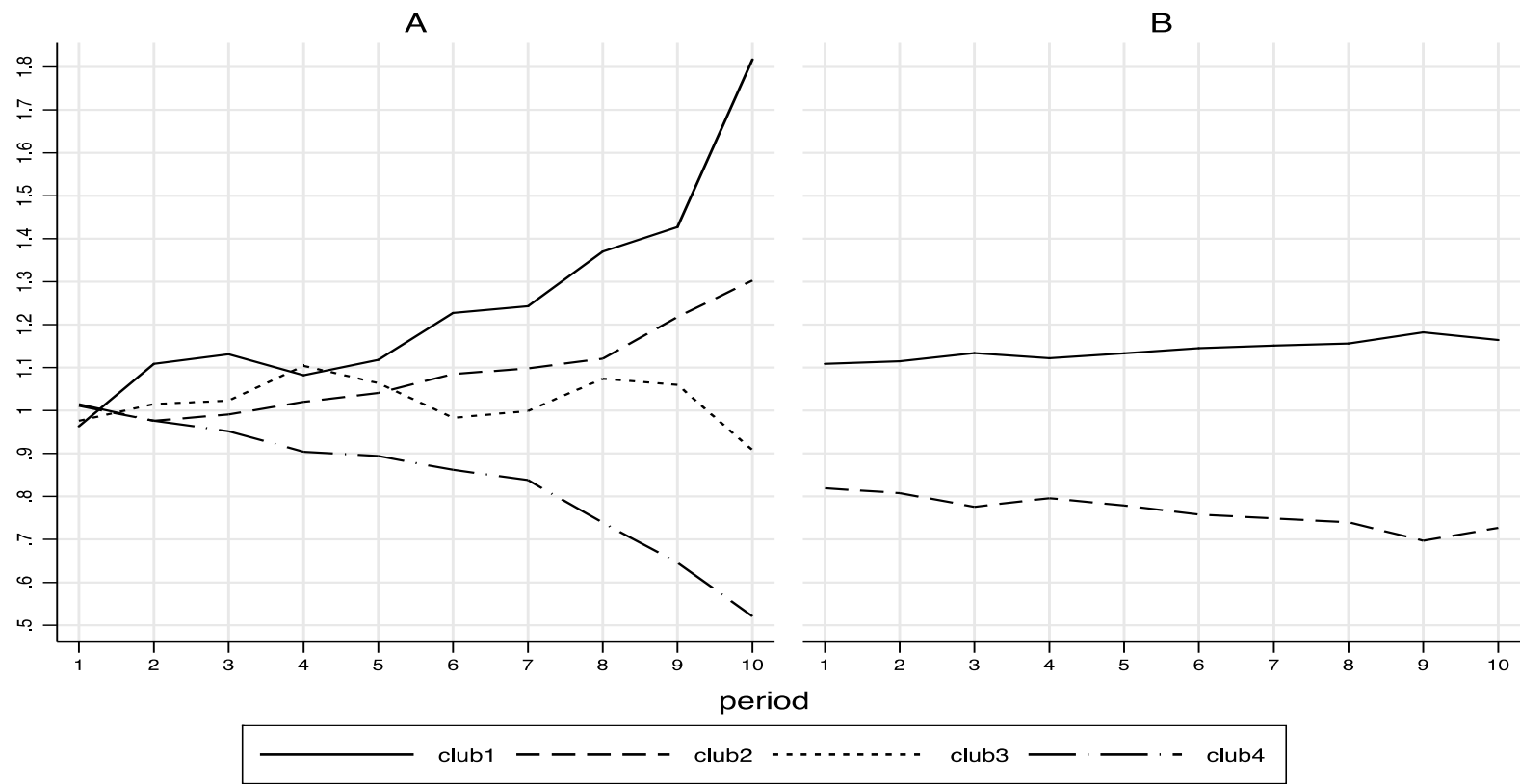


Figure 3: (A) clustering before and (B) after punishment

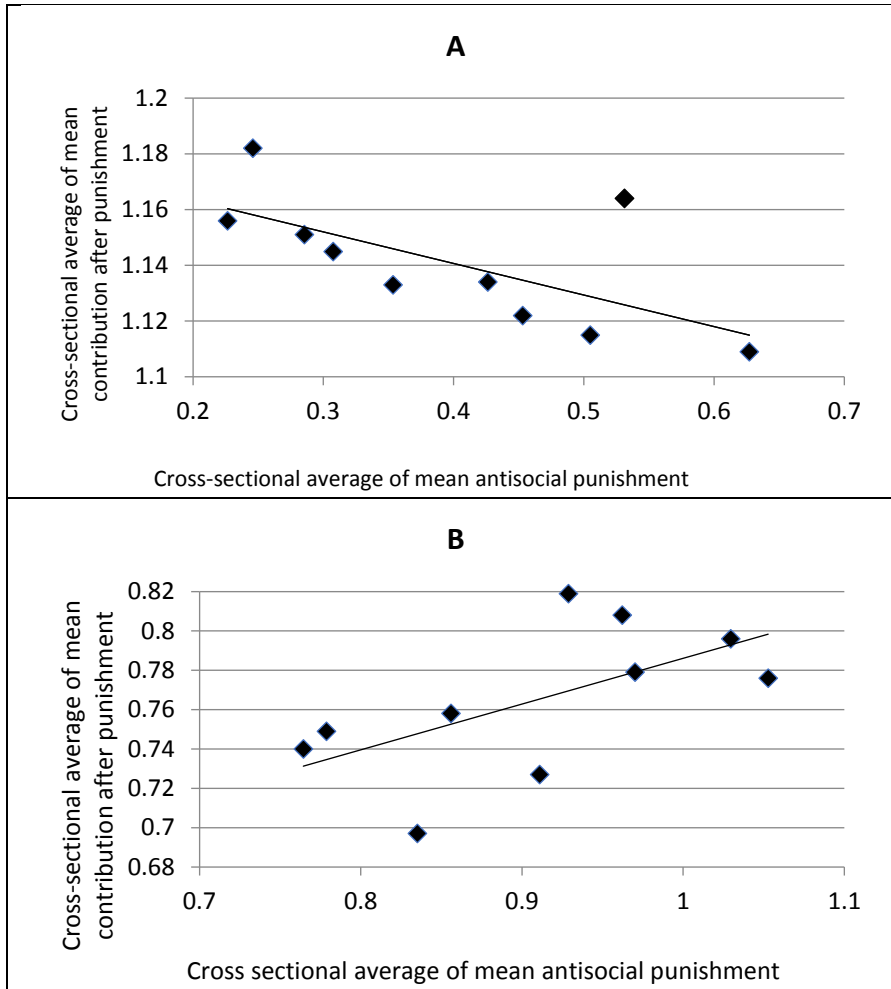


Figure 4: (A) the association between contribution and punishment of Cluster 1 (spearman=-0.62 [0.05]) and (B) Cluster 2 (spearman=0.66 [0.05])

TABLES

Table 1: Clusters of contribution

	Mean Contribution	
	Without Punishment	With Punishment
Overall Convergence	-11.542*	-2.275*
Cluster 1	-1.112 [Dnipropetrovs'k, Muscat]	4.572 [Bonn, Boston, Chengdu, Copenhagen, Istanbul, Melbourne, Nottingham, Seoul, St. Gallen, Zurich]
Cluster 2	2.667 [Athens, Copenhagen, Minsk, Riyadh, Samara]	-0.113 [Athens, Dnipropetrovs'k, Minsk, Muscat, Riyadh, Samara]
Cluster 3	-1.505 [Bonn, Seoul, Zurich]	N.A.
Cluster 4	-0.539 [Boston, Chengdu, Istanbul, Melbourne, Nottingham, St. Gallen]	N.A.

Overall test: The null hypothesis states the convergence to the common steady state. All numbers reported are the estimated t statistics. The null cannot be rejected as long as estimated one tail t statistic is larger than the critical value $t_c = -1.65$ at 5% level of significance * indicates significance at 5% level, or the rejection of the null. The term 'Cluster' refers to the club of convergence within which a subset of cities' contribution are convergent. In other words, each cluster represents a club-based steady state. If the overall test shows failure to reject the alternative hypothesis, the procedure moves to test the presence of clubs of convergence. Thus, the above output show failure to reject the null hypothesis and we conclude that there is club of convergence.

Table 2: OLS Estimates of the determinants of Group Average Contribution in periods 2 to 10.

Dependent variable: Group average contribution in periods 2 to 10	(1)	(2)	
		Cluster 1	Cluster 2
Group average contributions in period 1	0.779** (0.052)	0.794*** (0.047)	0.901*** (0.079)
Group average punishment of free riding	0.521** (0.201)	1.608* (0.841)	0.416 (0.849)
Group average antisocial punishment	-2.247*** (0.350)	-4.522*** (1.362)	-0.883 (0.871)
Constant		3.073*** (0.599)	1.184 (0.881)
\bar{R}^2	0.60	0.66	0.59
F test	136.9 [0.0]	110.19 [0.0]	49.96 [0.0]
N	273	168	105

\bar{R}^2 refers to the adjusted R-squared. N : the sample size. The F test tests the null of the joint insignificance of the model against the alternative joint significance. Both models are jointly significant. Value in (.) and [.] are standard errors and p-values respectively. (***), (**), (*) refer to significance at 1%, 5% and 10% respectively.

References

- Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10, pp 171-178.
- Guala, F. and Salanti, A. (2001). Theory, experiments, and explanation in economics. *Revue Internationale de Philosophie*, 217, pp 327 – 349.
- Herrmann, B., Thoni, C., Gächter, S. (2008a). Antisocial punishment across societies, *Science*, 319, pp 62-66.
- Herrmann, B., Thoni, C., Gächter, S. (2008b). Supporting online material for antisocial punishment across societies, *Science Supporting Online Material*, 319, <http://www.sciencemag.org/content/suppl/2008/03/06/319.5868.1362.DC1/Herrmann.SOM.pdf> [25/04/2013].
- Levins, R. (1966), “The Strategy of Model Building in Population Biology,” in E. Sober (ed.), *Conceptual Issues in Evolutionary Biology*. 1st ed. Cambridge, MA: MIT Press.
- Phillips, P.C.B., & Sul, D. (2007) Transition modelling and econometric convergence tests, *Econometrica*, 75 , pp 1771–1855.
- Weisberg, M. (2004). Robustness analysis, *Philosophy of Science*, 73 pp 730-742.

Acknowledgements

The authors would like to thank Peter Phillips and Donggyu Sul, for having shared the Aptech's GAUSS code for convergence testing and Benedikt Herrmann, for having shared the dataset.