Manual: How to compare two repeated public good experiments

by

Donggyu Sul
And
Jianning Kong
(UT at Dallas)

This version: October, 2012

Comments Welcomed
Reference:

Donggyu Sul (2012), Estimation of Treatment Effects in Repeated Public Good Experiments, UT at Dallas

Jianning Kong and Donggyu Sul (2012), Estimation of Treatment Effects under Multiple Equilibriums in Repeated Public Good Experiments, UT at Dallas

Acknowledgement:

Special thanks to go to Rachel Croson, Catherine Eckel, Daniel Houser and Sherry Li for providing helpful comments on the previous version of the papers, Ryan Greenaway-McGrevy for editorial help, and many graduate students who have asked many empirical questions related to repeated public good experiments.
**Do and Don't List**

<table>
<thead>
<tr>
<th>Don't</th>
<th>Do</th>
</tr>
</thead>
<tbody>
<tr>
<td>Don't run Wilcoxon-Mann-Whitney' rank sum test, Dynamic Panel Censored (Tobit) Regression, or a standard t-test.</td>
<td>1. Plot cross sectional averages and variances over round</td>
</tr>
<tr>
<td></td>
<td>2. See if both of them, especially variances, are converging to a constant or zero.</td>
</tr>
</tbody>
</table>

**Objective:**

You want to compare the **overall treatment effects** (the contributions to the public account) *robustly and accurately*.

**Issues:**

Both estimation and testing are depending on whether or not all subjects may be free riders in the long run.
What do we (experimental economists) want to estimate?
Average or distribution?

Example 1: Andreoni (1995, AER)’s game

Andreoni wants to test whether or not the two games provide the same average contributions.

In fact, almost all of experimental studies care about average contributions.

However, most of studies test equal distribution.

Q1: Do you want to test whether or not two outcomes are same in terms of central tendency or come from the same distribution?
A1: Test the same central tendency or average.

1 Of course, if the two experiments share the same distribution, then this implies automatically the central tendency.
Then How to measure the average contribution?
It is not easy to measure. Why?

Example 2:

Conclusion: We need to measure the outcomes robustly (independent from the total number of rounds).

How?: We need to know the true data generating process (or estimate the unknown decay function).
Then How to measure the average contribution?
It is not easy to measure. Why?
Example 3:

Suppose that in both Game A and B, there are two types of subjects: Confused and free riders.
- The free riders in Game A have slower decay rates that those in Game B.
- The confused subjects contribute 0.5 always.
- In Game A, the fraction of the confused group is 0.1 but that in Game B is 0.3.
- Hence in the long run, Game B is always more efficient. (donate more)

However,....
Example 3:

- The average contribution in A is always less than that in B (if round ≤ 10). Looks Game A is more efficient but the truth is that Game A is less efficient.
- If \( t=20 \) (long enough), then the contribution in B > that in A.
Lesson from Example 2 & 3,

1. To measure the average contribution, we have to know whether or not subjects’ outcomes are converging (single equilibrium) or diverging (multiple equilibriums).
2. Under single equilibrium (convergence case), the unknown decay rate should be obtained to measure the average contribution.
3. Under multiple equilibrium (divergence case), the fraction(s) of the confused and (or) Pareto groups should be obtained. This requires clustering analysis.
Why Convergence Test?

- When there are multiple equilbriums (under somewhat restrictive conditions), the gaps (or cross sectional variance) are increasing over round.
First Check: Convergence

Purpose:

If all subjects are free riders, then the estimation of the treatment effect becomes much simpler. Otherwise, you have to cluster subjects into Altruists, Warm-glow givers and Free-riders.

Symptoms for Divergence

1. Cross sectional averages show sometimes `jerky' behavior over round. (but not always)
2. Cross sectional variances increase over round (usually).

Causes for Divergence

1. There are possibly more than two long run equilibriums. Typical one is Nash but under divergence, some subjects form Pareto or Warm-glow group.
2. The existence of conditional co-operators does not mean that there are multiple equilibriums. If there are only pure free riders and conditional co-operators, then the cross sectional variances do not diverge over round.
Example 4: Andreoni (1995) and IWW (1994)

Consider two experimental games: Regular games by Andreoni (1995, AER) and Isaac, Walker and Williams (1994, JPE). -- Andreoni's game: Group size 5, MPCR=0.5; IWW's game: Group size 40, MPCR=0.4. Other conditions are the same.

![Graph showing cross sectional averages for Andreoni (1995) and Isaac, Walker and Williams (1994)](image)

Observation:

1. Cross sectional averages in Andreoni's game decrease over round (almost monotonically)

2. Cross sectional averages in IWW's game don't seem to converge, and more importantly fluctuate widely over round.
Example 4: Andreoni (1995) and IWW (1994)

Observation:
1. Cross sectional variances in Andreoni's game seems to decrease over round
2. Cross sectional variances in IWW's game seems to increase definitely.

Conclusion:
Andreoni's game seems to converge but IWW's game diverges definitely!

From the eyeball examinations, we conclude that

A. All subjects in Andreoni's game may form the single equilibrium in the long run: Free-riders
B. Subjects in IWW's game may form at least two equilibriums in the long run.
C. An increase in the group size results in more contributions to the public account! Don't need any formal test.
A Formal Convergence (No-divergence) Test

Notation: $H_{nt} = \text{Cross sectional variance of } y_{it}$

- Run the following simple trend regression
  $$H_{nt} = a + bt + u_t$$
  The null and alternative hypotheses become
  $$H_0: b \leq 0 \text{ (no divergence)} \quad v.s. \quad H_A: b > 0 \text{ (divergence)}$$

- How to test:
  - Construct the standard $t$-ratio. If $t$-statistic is greater than 1.65, the null is rejected at the 5% level. (One side test)
  - The rejection of the null hypothesis does not imply that all subjects are diverging but simply imply that there are multiple equilibriums.

- You can skip this by using the eyeball exam. If the cross sectional variance is overall decreasing over round, then this implies that there is overall convergence. If not, there is divergence.

- Statistically, as $t \to \infty$, the cross sectional variance becomes a constant under convergence. However under mild conditions, the cross sectional variance becomes diverging under multiple equilibriums.
Under Single (Nash) Equilibrium

Need to estimate the unknown decay rate: Have to set up `econometric model’ to measure average contribution. Revisit Example 1 (Andreoni, 1995). Sul (2012) suggests the exponential decay model to fit the experimental outcomes. Looks very reasonable and does fit very well.

Example 1:
Convergence Test

- All Convergent to Nash
  - Run Trend Regression
    - Let’s Exam it Now

- A converges but B diverges
  - B dominates A always
    - Done

- A & B diverge both
  - Cluster subjects into Nash, Pareto and confused groups
  - Fraction of the confused and Pareto groups are key variables

Done
Econometric Model under Single (Nash) Equilibrium

- Sul (2012) uses exponential decay model. Here are assumptions and definitions in his model
  - Let the long run equilibrium be $a_i: \lim_{T \to \infty} y_{it} = a_i$
    - Of course, if a game is repeated infinitely, the outcome of the game must be different.
    - However, here we need to define the long run equilibrium. Hence we are letting $T$ to go to infinity.
  - First outcome and its expectation: $y_{i1} = \mu_i + e_{i1}, \ E(y_{i1}|\mu_i) = \mu_i$
    - Treat $\mu_i$ as a constant here. Individual specific.
    - Unconditional mean becomes $\mu$, which represents the average of the initial outcomes.
    - $e_{i1}$ is the expectation error. Individual specific.
  - Transitional path is approximated as
    $$E(y_{it}|a_i, \mu_i) = a_i + (\mu_i - a_i)\exp(-\beta [t - 1])$$
    - The decay rate, $\rho$, is defined as $\exp(-\beta)$.
    - The decay rate represents the speed of learning. An easy game leads to faster convergence.
  - Define the expectation error
    $$u_{it} = y_{it} - E(y_{it}|a_i, \mu_i)$$
    - If subjects are rational, then the errors are not serially correlated.
    - Also the error may include specification and approximation errors.
Econometric Model under Single (Nash) Equilibrium

- Under the above assumptions, Sul (2012) derives the following important results.
  
  - **Time varying relationship**

    \[ E(y_{it}|a_i, \mu_i) = a_i(1 - \rho) + \rho E(y_{i,t-1}|a_i, \mu_i) \]

    - The current expected outcome for each subject is serially correlated with the past outcomes.
    - Also the current expected outcome for each subject is dependent on the past outcome of all groups. To see this, let \( t = 2 \)

      \[ E(y_{i2}|a_i, \mu_i) = a_i(1 - \rho) + \rho E(y_{i1}|\mu_i) = a_i(1 - \rho) + \frac{\mu_i}{\mu} \mu \rho \]

      = discounted LR outcome + Relative individual response × group mean at round 2

    - At the initial round, the subjects form unconditional expectation, but when \( t \geq 2 \), the subjects behave conditional on the previous outcomes.
    - Such different expectation outcome is called ``non-stationary initial condition’’ in Econometrics, which generates nonlinear decay function, \( \rho^{t-1} \).
    - That’s why the cross sectional averages have a nonlinear trend.
**Latent Structure (or Data generating process) and Cross sectional dependence**

\[ y_{it} = a_i + (\mu_i - a_i)\rho^{t-1} + u_{it}, \quad u_{it} = \phi u_{it-1} + \xi_{it} \]

- Subjects’ outcomes are written as ‘approximate common factor’ representation.
- The decay function, \( \rho^{t-1} \), becomes a common factor, and the difference between the initial and long run outcomes becomes factor loadings.
- The approximate common factor model has been used for modeling cross sectional dependence in many areas. In other words, subjects start to depend on each other after \( t=2 \).
- Note that at \( t=1 \), subjects are not cross sectionally dependent since they are recruited randomly. However when \( t >1 \), subjects form dependence due to the same learning.
- Within and across (a) groups, subjects are dependent since they are playing the same game.
- If subjects form rational expectation, then \( \phi = 0 \). Otherwise, \( \phi \neq 0 \).
- When the dominant strategy becomes Nash, the long run value, \( a_i = 0 \). Then we have
  \[ y_{it} = \mu_i \rho^{t-1} + u_{it} \]
  - The cross sectional average has a downward trend temporarily.
- When the dominant strategy becomes Pareto, the long run value, \( a_i = 1 \). Then we have
  \[ y_{it} = 1 + (\mu_i - 1)\rho^{t-1} + u_{it} \]
  - The cross sectional average increases over round
Revisit Example 1: How to calculate the overall contribution

- The nonlinear factor model fits the data very well.
- From the fitted model, we can calculate the overall fitted treatment outcome (the area under the fitted curve), which can be written as $\int_0^{\infty} f(t)dt$ where $f(t) = \mu \rho^{t-1}$.
- Note that $\int_1^{\infty} f(t)dt = \frac{\mu}{1-\rho}$, so we need to estimate $\mu$ and $\rho$. 
Estimation of Average Contribution under Single (Nash) Equilibrium

Notation: $y_{nt} = $ cross sectional average.

1. Take logarithm of $y_{nt}$, and then run the following trend regression

$$\log y_{nt} = \alpha + \gamma (t - 1) + e_t$$

(1)

2. Take exponential of $\hat{\alpha}$ and $\hat{\gamma}$. Also get the sample variances of $\hat{\alpha}$ and $\hat{\gamma}$.
   - $\exp(\hat{\alpha})$ estimates the unknown mean of the first outcomes, $\mu$.
   - $\exp(\hat{\gamma})$ estimates the unknown decay rate, $\rho$.

3. Construct the following measure

$$\hat{\Pi} = \hat{\mu} / \{1 - \hat{\rho}\}$$

4. $\Pi$ estimates the asymptotic treatment effects which measures the robust and overall contributions to the public account.\(^2\)

5. Repeat Step 1 through Step 4 with the other game. Let $\Pi_A$ and $\Pi_B$ be the asymptotic treatments for Game A and B. Then the overall treatment effects becomes $\Pi_A - \Pi_B$ and its variance becomes $\hat{\Omega}_{\Pi_A} + \hat{\Omega}_{\Pi_B}$.

\(^2\) The variance of $\hat{\Pi}$, $\hat{\Omega}_{\Pi}$, can obtained by $\hat{\Omega}_{\Pi} = R\hat{\Omega}HR'$, where $R = [(1 - \hat{\rho})^{-1} , \hat{\mu}(1 - \hat{\rho})^{-2}]$, $H = \begin{bmatrix} \hat{\mu} & 0 \\ 0 & \hat{\rho} \end{bmatrix}$, and $\hat{\Omega}$ is the sample variance of $\hat{\alpha}$ and $\hat{\gamma}$. $\hat{\Omega} = \hat{\Omega}_0 + \begin{bmatrix} \hat{\sigma}_\mu^2 & 0 \\ 0 & 0 \end{bmatrix}$ where $\hat{\Omega}_0$ is the sample covariance matrix from the eq. (1). $\hat{\sigma}_\mu^2$ can be obtained by $Var(y_{it}) - Var(\hat{e}_t) \times N$. 

Pitfalls of Existent Methods

1. Eyeball Exam (intra-ocular Trauma Test):
2. Nonparametric rank-sum test (Wilcoxon-Mann-Whitney Test)
3. T-test (or Z-score test)
4. Panel Dynamic Tobit (censored) Regression
5. Panel Dynamic (unconditional) Regression

Don’t use them:
Why?
1,3 => The null hypothesis is the same mean. But 1 ignores statistical inference and 2 uses wrong variance.

2,4,5 => The null hypothesis is different.

Let’s investigate the problems of each test.
Eyeball Exam

- How to do it: Plot cross sectional averages and see if there are permanent differences.

- Problem: Ignore statistical inference.

- Example: Isaac, Walker and Williams (1994, JPE). They plotted 90% confidence intervals. No difference between two games.

Wilcoxon-Mann-Whitney Test

- How to do it: compare ranks of individuals between two games.

- Problem:
  
  - WMW examines the null hypothesis of the equal distribution (same mean, variance and density). Not equal mean: Example: Consider two games A and B where the initial observations are given by $y_{i1}^A \sim iidN(0.5,1)$, $y_{i1}^B \sim iidN(0.5,1.2)$. Then WMW test will reject the null of the same CDF since the variance of $y_{i1}^A$ is smaller than that of $y_{i1}^B$.
  
  - With nonstationary initial condition, the asymptotic properties of WMW test are unknown. Should not use WMW test at all even with $t=1$ (in one shot non-repeated games)

6. Example: I will show the example shortly.
Standard t-test
• How to do it: Compare two means.
• Problem:
  o Only for t=1, a standard t-test is valid: Subjects are randomly selected → no cross sectional dependence.
  o However with t ≥ 2, t-test becomes invalid due to cross sectional and time series dependence. Panel robust covariance matrix option (STATA: Rob) does not cure this problem.
7. Example: Kong and Sul (2012). Also I will show the examples shortly.

Panel Censored Regression
• How to do it (Censored regression: Not panel censored): Enough show the problem of the censored regression. Panel censored regressions are more problematic.
  o Run Tobit regression and test significance of the dummy variable.
  \[ y_i^* = a + b d_s + e_i \text{ if } 0 < y_i^* < 1, s = A,B. \]
  \[ y_i = 0 \text{ if } y_i^* < 0, \text{ and } y_i = 1 \text{ if } y_i^* > 1. \]
  o \( y_i^* \) is called `latent' variable and \( y_i \) is actual observation.
  o If \( b = 0 \), then experimentalists interpret this evidence as no treatment effect.
• Problem: The null hypothesis of \( b = 0 \) implies that \( E(y_{iA}^*) = E(y_{iB}^*) \), not \( E(y_{iA}) = E(y_{iB}) \). The latter hypothesis of \( E(y_{iA}) = E(y_{iB}) \) is what you want to test.
  o Does make any difference? Yes, a lot.
  o Example: Assume that \( y_{iA}^* \sim iidN(0.5,2) \) but \( y_{iB}^* \sim iidN(0.5,1) \). Then we have
<table>
<thead>
<tr>
<th>Case</th>
<th>$\sigma^2$</th>
<th>$E(y^*)$</th>
<th>$E(y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2</td>
<td>0.5</td>
<td>0.421</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>0.5</td>
<td>0.382</td>
</tr>
</tbody>
</table>

- $E(y_{iA}) > E(y_{iB})$ so that Game A is more effective than Game B.
- However you can’t reject the null of no treatment effect since $E(y^*_{iA}) = E(y^*_{iB})$.

• Example: See Sul (2012) for more examples.

**Dynamic Panel Regressions**

• How to do it: Running the following dynamic panel regression without thinking.

$$y_{it} = \alpha_i + bd_s + \beta y_{it-1} + \delta t + e_{it}$$

and test $H_0: b = 0$.

• Problem: Two misspecifications.
  - Including common trend: If the decay rates are different across games, then homogeneity restriction on the trend term gives wrong result.
  - Including lagged dependent variable: If the degrees of serial dependence are different across games, then the null of $b=0$ does not imply the same means across games.

• See Sul (2012) for detailed discussion.

• Example: See shortly.

I use Andreoni (1995a,1995b, 1988)’s data and re-estimate the asymptotic treatment effects. Here is the result.

Table 4: Empirical Results based on Trend Regressions

<table>
<thead>
<tr>
<th>Experiments</th>
<th>$\hat{\mu}$ (s.e)</th>
<th>$\hat{\rho}$ (s.e)</th>
<th>$\hat{\sigma}_{\mu}^2$</th>
<th>$\hat{\phi}$</th>
<th>$\hat{\theta}$ (s.e)</th>
<th>$\hat{\phi}$ (s.e)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reg1</td>
<td>0.549 (0.018)</td>
<td>0.922 (0.006)</td>
<td>0.120</td>
<td>0.031</td>
<td>7.070 (0.710)</td>
<td>0.021 (0.225)</td>
</tr>
<tr>
<td>Reg2</td>
<td>0.572 (0.035)</td>
<td>0.925 (0.011)</td>
<td>0.064</td>
<td>0.059</td>
<td>7.641 (1.487)</td>
<td>0.050 (0.306)</td>
</tr>
<tr>
<td>RegR</td>
<td>0.494 (0.038)</td>
<td>0.807 (0.011)</td>
<td>0.065</td>
<td>0.025</td>
<td>2.561 (0.333)</td>
<td>-0.286 (0.296)</td>
</tr>
<tr>
<td>Neg</td>
<td>0.254 (0.014)</td>
<td>0.907 (0.010)</td>
<td>0.109</td>
<td>0.010</td>
<td>2.748 (0.435)</td>
<td>-0.164 (0.322)</td>
</tr>
<tr>
<td>Part</td>
<td>0.546 (0.063)</td>
<td>0.881 (0.019)</td>
<td>0.077</td>
<td>0.052</td>
<td>4.608 (1.214)</td>
<td>-0.089 (0.331)</td>
</tr>
</tbody>
</table>

Note: The numbers in parentheses are standard errors based on Andrews’ HAC estimation.

1) The samples in the last round are not used since its cross sectional average is near to zero.

- Reg1 and Reg2 are regular games (MPCR=0.5, group size = 5), RegR is regular and ranked game, Neg is negative frame game. These four games are the games with strangers. Part is the partner game.
- Subjects form rational expectation since $\phi$ is not significantly different from zero.
- The decay rates are affected by the game nature. The decay rate in the regular-ranked game has the fastest rate. Meanwhile the regular and Negative frame games have similar rates.
Empirical Example of Trend Regressions in Sul (2012) - Continue

From Sul (2012).

Table 5: Estimation of Asymptotic Treatment Effects and Comparison to Other Tests

<table>
<thead>
<tr>
<th>Panel A: Summary of Other Tests</th>
<th>Controlled Treated</th>
<th>Reg1 v.s.</th>
<th>Reg2 v.s.</th>
<th>RegR v.s.</th>
<th>Neg v.s.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Reg2 Neg Part</td>
<td>RegR Neg Part</td>
<td>RegR Neg Part</td>
<td>Neg Part</td>
<td>Neg v.s. Part</td>
</tr>
<tr>
<td>Eyeball</td>
<td>? Yes Yes Yes Yes</td>
<td>Yes Yes Yes</td>
<td>? Yes Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>WMW</td>
<td>Yes Yes Yes Yes</td>
<td>Yes Yes Yes</td>
<td>Yes No Yes</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>Tobit</td>
<td>No Yes Yes No</td>
<td>Yes Yes No</td>
<td>No No Yes</td>
<td>Yes</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trend Regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\beta} = \hat{\beta}_c - \hat{\beta}_r )</td>
</tr>
<tr>
<td>( t - )ratio</td>
</tr>
</tbody>
</table>

- The eyeball exam (inter-ocular trauma test) used by early experimental studies ignored statistical inference. Hence they are inaccurate.
- Wilcoxon-Mann-Whitney (WMW)'s rank-sum test rejects too often.
- Panel Tobit regression is rather more accurate than the two (the eyeball and WMW tests) but does not have statistical justification. See Sul (2012) for detailed discussion.
- Trend regression is accurate. Particularly, with trend regression, the outcomes of Partner game are not different from those of Stranger game statistically.
Convergence Test

All Convergent to Nash

A converges but B diverges

A & B diverge both

Run Trend Regression

B dominates A always

Cluster subjects into Nash, Pareto and confused groups

Fraction of the confused and Pareto groups are key variables

Done

Done

Let’s Exam it Now
Under Divergence

- Before we discuss how to measure (relative) average contributions between two games, we need to think about why there are possibly multiple equilibriums in a game.

- Consider environments such as Group Size, MPCR, Communication and Penalty.
  - Group Size: as it increases, more opportunity for free riding but at the same time, the return from the public account is increasing (the return from the public account to the subject is constant, I suggest “the investment in public account can benefit more subjects”). So possibly diverging.
  - MPCR: as it increases, more return from the public account. So some subjects may contribute more.
  - Communication: More discussion leads to more donations.
  - Penalty: If subjects are risk averse, they co-operate.


- Need to develop theory further.

Let’s discuss now how to measure average contribution under divergence.
Relative contributions between two divergent games

Revisit Example 3: Game A (10% are confused) v.s. Game B (30% are confused)

- Game A: Less confused. Converges to 0.05
- Game B: More Confused. Converges to 0.15
Game B is definitely more effective in the *long run*. To find out this effect, need very large ROUNDS: Expansive and time consuming.

**Relative contributions between two divergent games**

There are possibly three convergent sub-groups: Nash, Pareto and confused. From the exponential decay models, the outcome of the $i$th subject at round $t$ can be written as

$$y_{it} = \begin{cases} 
\mu_i \rho^{t-1} + e_{it} & \text{if } i \in \text{Nash} \quad n_1 = N_1/N \\
\mu_i + e_{it} & \text{if } i \in \text{Warm-glow or confused} \quad n_2 = N_2/N \\
1 + (\mu_i - 1) \rho^{t-1} + e_{it} & \text{if } i \in \text{Pareto} \quad n_3 = N_3/N
\end{cases}$$

Note that the fractions of free riders, altruists and confused subjects are non-zero constant (that is, independent from the total number of subjects, $N$.)

Under such multiple equilibriums, the overall average outcome can be written as

$$\text{TE} = \text{plim}_{N \to \infty} \frac{1}{TN} \sum_{i=1}^{N} \sum_{t=1}^{T} y_{it} = \frac{1-\rho^T}{1-\rho} \frac{n_1 \mu}{T} + n_2 \mu + n_3 - \frac{n_3 (1-\mu)}{T} \left( \frac{1-\rho^T}{1-\rho} \right)$$

Since this measure depends on $T$, we take the sample time series mean and then let $T$ goes to infinity to get

$$\text{Asym ATE} = \text{plim}_{N,T \to \infty} \frac{1}{TN} \sum_{i=1}^{N} \sum_{t=1}^{T} y_{it} = n_2 \mu + n_3 = \tau$$
Before we discuss how to estimate $\tau$, we will explain how to estimate $n_1, n_2,$ and $n_3$.

**Step by Step Clustering Method**

1. Sorting out obvious ones first: If $y_{it} = 0$ for all $t$, then this subject is a free rider definitely. If $y_{it} = 1$ for all $t$, then this subject becomes an altruist.
2. For the rest of subjects, Run the following trend regression for each subject’s outcome.
   
   $$y_{it} = a + bt + \varepsilon_{it}$$

3. Construct the conventional $t$-statistic.
4. Sorting subjects into the confused group if only if $|t_{ib}| < 1.28$. Count $N_2$ and calculate $n_2$.
5. Count the total cases where $t_{ib} > 1.28$ and $t_{ib} < -1.28$. They become $N_3$ and $N_1$.
   
   Note that the critical value is not tight enough to identify whether or not a subject is a free rider but the fraction of free riders is rather accurately estimated.

Do the next step to choose truly (core) Nash and Pareto subjects:

6. Sorting subjects into the Pareto and Nash groups if only if $t_{ib} > 3.02$ or $< -3.02$, respectively.
   o Note that if you use higher critical value, then you will under-estimate $N_1$ and $N_3$, and over-estimate $N_2$.
   o These core Nash and Pareto subjects will be used for the estimation of the long run treatment effects later.
Actual Example from IWW (1994): MPCR=0.75, Group Size = 40.

Obvious: Red filled circle (P), Empty circle (N)

Obvious also

Not Obvious? Filled (t-stat: 2.68), Empty (-1.28)

Confused ones: ( |t-stat| < 1.28)
Econometric Theory on the Crude Clustering Method

- In Psychology, latent growth curve mixture model is popularly used for clustering subjects. We can’t use this method since the decay rate goes to zero as the round increases.

- In Economic Growth literature, Phillips and Sul (2007)’s clustering algorithm has been popularly used. If the total number of rounds is very big, their method can be used here. However again, as the round increases, the decay rate goes to zero. So this method is also ruled out.

- The simple trend regression is effective enough to cluster subjects into three groups. See Kong and Sul (2012) for detailed discussion.

- The fraction of each group must be a constant even when the total number of subjects, $N$, goes to infinity. Otherwise, there is no divergence in a game: For an example, in Andreoni (1995)’s regular games (regular 1 and 2), two subjects chooses all ones. In regular 1 game, $N=60$ but in regular 2, $N = 40$. The fraction of the Pareto group as $N$ is getting larger goes to zero. In this case, the Pareto subjects can be treated as outliers and they are not influenced the average contribution asymptotically.
Estimation of the Asymptotic Average Treatment Effect (AATE)

Step 1: Estimate $\mu$ by using the first observations.

$$
\hat{\mu} = N^{-1} \sum y_{i,1}
$$

Step 2: Use the core Pareto and Nash subjects. Run the following pooled regression

$$
\log(y_{n,t}^{cn}) - \log \hat{\mu} = \gamma(t - 1) + v_{1t},
$$

$$
\log(1 - y_{n,t}^{cp}) - \log(1 - \hat{\mu}) = \gamma(t - 1) + v_{3t},
$$

where $y_{n,t}^{cn}$ and $y_{n,t}^{cp}$ are the cored Nash and Pareto subjects, respectively. Take exponential of $\hat{\gamma}$, which becomes the estimate of $\hat{\rho}$.

Step 3: Run the following time series regression

$$
y_{nt} = \tau + \alpha \hat{\rho}^{t-1} + e_t.
$$

1. The estimate of $\hat{\tau}$ becomes AATE. $y_{nt}$ is the cross sectional averages of $y_{it}$.
2. Construct the ordinary sample variance of $\hat{\tau}$, $\hat{\sigma}^2$. This estimator is inconsistent. See the next step.

Step 4: Use the confused subjects (see step 3 in the previous page).

1. Take time series average first, and then take cross sectional variance of the time series averages. This gives a consistent and efficient estimate of $\hat{\sigma}_\mu^2$ which is the variance of the unknown initial mean.
2. Calculate the following quantity: $n_2 \times \hat{\sigma}_\mu^2 / N$.
3. Construct the consistent variance of $\tau$: $\hat{\sigma}_\tau^2 = \hat{\sigma}^2 + n_2 \times \hat{\sigma}_\mu^2 / N$. 
Empirical Example from Kong and Sul (2012)

Data: IWW (1994), Regular Games
Game A: MPCR = 0.75, Group Size = 100; Game B: MPCR = 0.30, Group Size = 40.

Results and Interpretation:

<table>
<thead>
<tr>
<th>Estimation From Clustering</th>
<th>MPCR=0.75</th>
<th>MPCR=0.30</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{\mu} )</td>
<td>0.466</td>
<td>0.484</td>
</tr>
<tr>
<td>( \hat{\rho} )</td>
<td>0.760</td>
<td>0.816</td>
</tr>
<tr>
<td>( \hat{n}_1 )</td>
<td>0.290</td>
<td>0.387</td>
</tr>
<tr>
<td>( \hat{n}_2 )</td>
<td>0.430</td>
<td>0.450</td>
</tr>
<tr>
<td>( \hat{n}_3 )</td>
<td>0.280</td>
<td>0.163</td>
</tr>
</tbody>
</table>

Estimation of Asymptotic Average Treatment Effect (AATE)

| \( t \)                  | 0.426     | 0.339     | \( \Rightarrow \) 43% tokens allocated to the public account with MPCR=0.75. |
|                          |           |           | \( \Rightarrow \) 34% tokens allocated to the public account with MPCR=0.30. |

\[ \hat{\sigma}_t^2 \times 10^3 \]

| 1.226                     | 0.596     | \( \Rightarrow \) Will be used for a formal test |

Difference between AATE

\[ \hat{\pi} = 0.426 - 0.339 = 0.087 \]

\[ V(\hat{\pi}) = 0.001226 + 0.000596 = 0.00182 \]

\[ t_{\hat{\pi}} = \frac{0.086}{\sqrt{0.002}} = 2.015 \]

Significant at the 5% level (1.96).
Combining IWW (1994) and Andreoni (1995)

- Note that for convergence case, the AATE becomes zero. However for the comparison, we report the asymptotic treatment effects rather than AATE.
- Roughly mountain or bell shape: Too much MPCR or many Group Sizes don’t influence on the level of contributions.
- High MPCR or Group Size makes divergence.
STATA CODES

- Unzipped file into the same directory where your data are.
- Preparing Data Set: Data set must be TxN (long panel) type without having any heading.

Example: There are 10 rounds and 240 subjects.

The maximum token is 100.
• Edit `example.ado` file: Open example.ado and edit as follows:

```stata
insheet using 1ec40.csv
forvalues i=1/240{
  replace v`i'=v`i'/100
}
convt v1-v240
```

File name. TxN matrix format

Total number of subjects (N). If N =20, change it to 20.

Endowed Token =100. Change this number along with endowed token.

`convt` is the new commend for convergence test. `v1-v240` implies that you are using 240 subjects. If your N is 20, then change it v1-v20.

This program reports the coefficient and t-ratio of the convergence test.

• If the t-ratio is greater than 1.65, the null hypothesis is rejected at 5% significance level, which provides evidence of existence of multiple equilibriums. Then add

```stata
  treatment v1-v240
```

The estimated fractions of the three groups, the decay rate, the treatment effect and variance of treatment effect will be reported.
• If the t-ratio is less than 1.65, then add

\texttt{asymtreat v1-v240}

This program reports the estimate and variance of treatment effect under the convergence.
Concluding Remarks

- To estimate the treatment effect (overall contributions to the public account), examine convergent behavior (with cross sectional variance).
- For repeated strategy games (prisoners’ dilemma), the transition probabilities become of interest.
  - Probabilities are either converging or diverging.
  - If probabilities followed Sul’s decay model, then the methods considered in this manual could be directly used.
  - However, we didn’t study this case yet in detail.
- The suggested methods work even under nonlinear payoff function, or penalty schemes.
- Read the two reference papers for more detailed discussions about econometric procedures and justifications.
- If you have questions, please contact us: Donggyu Sul (d.sul@utdallas.edu) or Jianning Kong (jianningkong@gmail.com)