Twin Peaks: Expressive Voting and Soccer Hooliganism

David K. Levine[1], Salvatore Modica[2] & Junze Sun[1,2]

[1]European University Institute

[2]University of Palermo

EEA-ESEM, 22 August 2022

Introduction

- The "paradox of voting" (Downs, 1957): the pivotal chances are too small in large elections to justify rational individuals to turn out to vote.

- State-of-Art models to address this puzzle:
  - Peer pressure / monitoring (Levine and Mattozzi, 2020AER).
  - Expressive utility (Pons and Tricaud, 2018ECMA).

Introduction

- The "paradox of voting" (Downs, 1957): the pivotal chances are too small in large elections to justify rational individuals to turn out to vote.
- State-of-Art models to address this puzzle:
  - Peer pressure / monitoring (Levine and Mattozzi, 2020AER).
  - Expressive utility (Pons and Tricaud, 2018ECMA).
- In this paper we
  - combine peer pressure and expressive utility in a unified model.
  - introduce the **expressive externality** – people prefer cheering a match at the pub with friends to watching at home alone on television.
- Our main result is the possibility of "**tipping**" – group participation may jump up discontinuously when the expressive externality is strong enough.
- We present applications of tipping in sports ticket pricing and voting.

Outline

1. Model and the Tipping Theorem

2. Application I: Tickets Pricing in Sports

3. Application II: Voter Turnout

4. Conclusion

Model

- Group $k = 1, 2$ with size η_k. Each member i can choose to *participate* (root for the team) or *not participate* (not root for the team).
- Utility of non-participant: $p_k v_k + \lambda h_k \phi_k$
- Utility of participant: $p_k v_k + \lambda h_k \phi_k + h_k - (c_0 + y_i)$
- Interpretations of parameters:
  - p_k: Probability of group k winning (may or may not depend on ϕ_k and ϕ_{-k}).
  - v_k: Per-capita payoff from group k winning.
  - $\phi_k \in [0, 1]$: participation rate of group k (endogenous).
  - h_k: Expressive payoff (non-material).
  - **Key element**: λ – the positive externality of expression.
  - $c_0 + y_i$: participation costs of member i, with $c_0 < 0$ and $y_i \sim U[0, 1]$.

Model

- Group $k = 1, 2$ with size η_k. Each member i can choose to *participate* (root for the team) or *not participate* (not root for the team).
- Utility of non-participant: $p_k v_k + \lambda h_k \phi_k$
- Utility of participant: $p_k v_k + \lambda h_k \phi_k + h_k - (c_0 + y_i)$
- Interpretations of parameters:
  - p_k: Probability of group k winning (may or may not depend on ϕ_k and ϕ_{-k}).
  - v_k: Per-capita payoff from group k winning.
  - $\phi_k \in [0, 1]$: participation rate of group k (endogenous).
  - h_k: Expressive payoff (non-material).
  - **Key element**: λ – the positive externality of expression.
  - $c_0 + y_i$: participation costs of member i, with $c_0 < 0$ and $y_i \sim U[0, 1]$.
- Without social norm and assume $p_k v_k$ is constant, then participation is individually rational iff

$$y_i \leq h_k - c_0 \equiv \underline{\varphi}$$

$\underline{\varphi}$ is the fraction of "**committed members**" who always participate.

Social Norm: Inducing Participation via Peer Monitoring

- **Social norm**: an incentive compatible mechanism to induce a *targeted participation rate* $\varphi_k \in [\underline{\varphi}, 1]$; that is, *participate iff* $y_i \leq \varphi_k$.

Social Norm: Inducing Participation via Peer Monitoring

- **Social norm**: an incentive compatible mechanism to induce a *targeted participation rate* $\varphi_k \in \left[\underline{\varphi}, 1\right]$; that is, *participate iff* $y_i \leq \varphi_k$.
- Implementation and imperfect peer monitoring:
  - If member i participated, no punishment exerted on i.
  - If member i did not participate, neighbors of i receives a binary signal $s_i \in \{G, B\}$, and exert **punishment** $P_k \geq 0$ on i if $s_i = B$.

Social Norm: Inducing Participation via Peer Monitoring

- **Social norm**: an incentive compatible mechanism to induce a *targeted participation rate* $\varphi_k \in \left[\underline{\varphi}, 1\right]$; that is, *participate iff* $y_i \leq \varphi_k$.
- Implementation and imperfect peer monitoring:
  - If member i participated, no punishment exerted on i.
  - If member i did not participate, neighbors of i receives a binary signal $s_i \in \{G, B\}$, and exert **punishment** $P_k \geq 0$ on i if $s_i = B$.
  - Imperfect monitoring:

$$Pr[s_i = B] = \begin{cases} 0, & \text{if } i \text{ participated} \\ 1, & \text{if } i \text{ not participated and } y_i \leq \varphi_k \\ \pi, & \text{if } i \text{ not participated and } y_i > \varphi_k \end{cases}$$

$\pi \in [0, 1]$ equals the probability of misplacing punishment to a norm-follower.

Social Norm: Inducing Participation via Peer Monitoring

- **Social norm**: an incentive compatible mechanism to induce a *targeted participation rate* $\varphi_k \in [\underline{\varphi}, 1]$; that is, *participate iff* $y_i \leq \varphi_k$.
- Implementation and imperfect peer monitoring:
  - If member i participated, no punishment exerted on i.
  - If member i did not participate, neighbors of i receives a binary signal $s_i \in \{G, B\}$, and exert **punishment** $P_k \geq 0$ on i if $s_i = B$.
  - Imperfect monitoring:

$$Pr[s_i = B] = \begin{cases} 0, & \text{if } i \text{ participated} \\ 1, & \text{if } i \text{ not participated and } y_i \leq \varphi_k \\ \pi, & \text{if } i \text{ not participated and } y_i > \varphi_k \end{cases}$$

$\pi \in [0, 1]$ equals the probability of misplacing punishment to a norm-follower.

- Incentive compatibility: Indifference in participation at $y_i = \varphi_k$, that is,

$$P_k = \varphi_k - h_k + c_0$$

The Costs of Inducing Social Norm $\varphi_k > \underline{\varphi}$

Costs of implementing social norm $\varphi_k > \underline{\varphi}$:

- Direct participation costs:

$$T(\varphi_k) = \int_0^{\varphi_k} (c_0 + y_i) dy_i = \varphi_k(\varphi_k + 2c_0)/2$$

- Monitoring costs (misplaced punishment) for any $\varphi_k > \underline{\varphi}$:

$$M(\varphi_k) = \int_{\varphi_k}^{1} \pi P_k dy_i = \pi(1 - \varphi_k)(\varphi_k - h_k + c_0)$$

- Total cost: $C(\varphi_k) = T(\varphi_k) + M(\varphi_k)$.

The Costs of Inducing Social Norm $\varphi_k > \underline{\varphi}$

Costs of implementing social norm $\varphi_k > \underline{\varphi}$:

- Direct participation costs:

$$T(\varphi_k) = \int_0^{\varphi_k} (c_0 + y_i) dy_i = \varphi_k(\varphi_k + 2c_0)/2$$

- Monitoring costs (misplaced punishment) for any $\varphi_k > \underline{\varphi}$:

$$M(\varphi_k) = \int_{\varphi_k}^{1} \pi P_k dy_i = \pi(1 - \varphi_k)(\varphi_k - h_k + c_0)$$

- Total cost: $C(\varphi_k) = T(\varphi_k) + M(\varphi_k)$.

For $\pi = 1/2$ and $\varphi_k > \underline{\varphi}$, we obtain a linear total cost function

$$C(\varphi_k) = \frac{1 + h_k + c_0}{2}\varphi_k - \frac{h_k - c_0}{2}$$

Main Result: Tipping Theorem

- Per capita utility of members in group k under participation rate φ_k:

$$\mathcal{U}_k(\varphi_k) = p_k v_k + (1+\lambda) h_k \varphi_k - C(\varphi_k)$$

The **optimal social norm** φ_k^* solves $\max_{\varphi_k \in [\underline{\varphi},1]} \mathcal{U}_k(\varphi_k)$.

Main Result: Tipping Theorem

- Per capita utility of members in group k under participation rate φ_k:

$$\mathcal{U}_k(\varphi_k) = p_k v_k + (1+\lambda) h_k \varphi_k - C(\varphi_k)$$

The **optimal social norm** φ_k^* solves $\max_{\varphi_k \in [\underline{\varphi}, 1]} \mathcal{U}_k(\varphi_k)$.

- If $\pi = 1/2$ and $p_k v_k$ is constant, we get $\mathcal{U}_k(\varphi_k) = p_k v_k + \frac{h_k - c_0}{2} - \xi \varphi_k$, where

$$\xi = \frac{1 - h_k + c_0}{2} - \lambda h_k = \frac{1 - \underline{\varphi}}{2} - \lambda h_k$$

is the **marginal cost** of mobilizing participation.

Main Result: Tipping Theorem

- Per capita utility of members in group k under participation rate φ_k:

$$\mathcal{U}_k(\varphi_k) = p_k v_k + (1+\lambda) h_k \varphi_k - C(\varphi_k)$$

The **optimal social norm** φ_k^* solves $\max_{\varphi_k \in [\underline{\varphi}, 1]} \mathcal{U}_k(\varphi_k)$.

- If $\pi = 1/2$ and $p_k v_k$ is constant, we get $\mathcal{U}_k(\varphi_k) = p_k v_k + \frac{h_k - c_0}{2} - \xi \varphi_k$, where

$$\xi = \frac{1 - h_k + c_0}{2} - \lambda h_k = \frac{1 - \underline{\varphi}}{2} - \lambda h_k$$

is the **marginal cost** of mobilizing participation.

- **Tipping Theorem**: Suppose $\pi = 1/2$, $0 \le \underline{\varphi} < 1$, and $p_k v_k$ is constant. Let $\lambda^* = \left(1 - \underline{\varphi}\right)/(2h_k) > 0$. Then the optimal social norm is

$$\varphi_k^* = \begin{cases} \underline{\varphi}, & \text{if } \lambda < \lambda^* \\ 1, & \text{if } \lambda > \lambda^* \end{cases}.$$

$\Longrightarrow$ Without expressive externality (i.e. $\lambda = 0$) mobilization is not worthy.

Outline

Ticket Pricing in Sports

- A sports team owns a stadium with capacity $Q \in (0, 1)$, interpreted as the fraction of fans it can hold. It charges ticket price $r > 0$.
- Fans are a self-organizing group. Given price r, each fan i's participation cost is $\zeta_0 + r + y_i$ with some $\zeta_0 < 0$.
- The marginal cost of mobilizing participation is increasing in r:

$$\xi = \frac{1 - h_k + \zeta_0 + r}{2} - \lambda h_k$$

If price r too high $\Longrightarrow$ self-organization not worthy and attendance rate low.

Ticket Pricing in Sports

- A sports team owns a stadium with capacity $Q \in (0, 1)$, interpreted as the fraction of fans it can hold. It charges ticket price $r > 0$.

- Fans are a self-organizing group. Given price r, each fan i's participation cost is $\zeta_0 + r + y_i$ with some $\zeta_0 < 0$.

- The marginal cost of mobilizing participation is increasing in r:

$$\xi = \frac{1 - h_k + \zeta_0 + r}{2} - \lambda h_k$$

  If price r too high $\Longrightarrow$ self-organization not worthy and attendance rate low.

- **Main result**: If externality λ is high enough, then it is optimal to sell at full capacity Q and the revenue-maximizing price $r^*(Q)$ is increasing in Q.

- **Intuition**: Larger stadium capacity implies higher payoff from participation due to externality λ. This allows the sports team to charge higher prices without discouraging full participation by self-organization of fans.

Outline

Setup (based on Levine and Mattozzi (2020AER))

- A unit mass of voters divided into two parties of size η_k, where $k \in \{S, L\}$ (*small* and *large*) and $\eta_S < \eta_L$.
- Parties compete in a "winner-take-all" election:
  - the side that mobilizes more voters wins a prize of value V.
  - the per capita payoff is thus $v_k = V/\eta_k$ for the winning party k.
  - expressive payoff is $h_L = h_S = hV$.
- Each party $k \in \{S, L\}$ simultaneously chooses an incentive compatible social norm to implement the targeted participation rate φ_k for their voters.

Main Result: Tipping for Voter Turnout

- **Assumptions**:
  - Low expressive payoff: $h \to 0$.
  - Positive fraction of committed voters: $\underline{\varphi} = h - c_0 \to -c_0 \in (0, 1)$.
  - **Strong positive externality**: $\lambda h \to \kappa$ for some $\kappa > 0$.
  - Nontrivial competition: $\eta_S > \eta_L \underline{\varphi}$.

Main Result: Tipping for Voter Turnout

- **Assumptions**:
  - Low expressive payoff: $h \to 0$.
  - Positive fraction of committed voters: $\underline{\varphi} = h - c_0 \to -c_0 \in (0, 1)$.
  - **Strong positive externality**: $\lambda h \to \kappa$ for some $\kappa > 0$.
  - Nontrivial competition: $\eta_S > \eta_L \underline{\varphi}$.

- **Main result**: Tipping in turnout takes place at $\overline{V} := (1 + c_0)/(2\kappa)$.
  - For $V > \overline{V}$ the aggregate turnout of the two parties is $\overline{b} = \eta_S + \eta_L$.
  - For $V < \overline{V}$ but close the aggregate turnout is approximately $\underline{b} = \eta_S(1 + \underline{\varphi})$.

  Note: Tipping never occurs if externality λ is zero or too low.

- **Implication**: Voter turnout may discontinuously jump up even if the model is continuous in parameters (e.g., V increases only continuously).

Empirical Predictions and Estimations

- **Predictions**:
  - With tipping: Voter turnout follows a distribution with two peaks.
  - Without tipping: Voter turnout follows a single-peaked distribution.
- **Data**: U.S. presidential elections (1920-2020) and U.K. general elections (1918-2019).[1]
- **Estimation Method**: (Partial) maximum likelihood estimation
  - The time series of voter turnout data exhibits strong evidence of positive serial correlation and stationarity. Details
  - Building on Levine (1983 JoE): we can obtain consistent estimates by treating data as if they are independently drawn from the stationary distribution.

[1] For both we focus on elections where women were permitted to vote.

The Bernoulli-Normal mixture model for turnout

- In period t, turnout τ_t is given by[2]

$$\tau_t = b_t + \epsilon_t$$

where b_t is a Bernoulli random variable on $\{\mu, \mu + g\}$ with $g > 0$ and $Pr[b_t = \mu] = Q \in (0, 1)$, and $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$.

  - Note: $g\sqrt{Q(1-Q)}$ is the standard deviation of the Bernoulli component.

- Estimate parameters $\vartheta = \left(Q, g, \mu, \sigma^2\right)$ using (partial) MLE. Details

- Hypotheses:
  - No tipping (null): $g\sqrt{Q(1-Q)} = 0$ (i.e., either $g = 0$ or $Q \in \{0, 1\}$)
  - Tipping: $g\sqrt{Q(1-Q)} > 0$ (i.e., both $g > 0$ and $Q \in (0, 1)$)

[2] A micro-foundation for this mixture model is in paper.

Estimation Results (in table)

Table: (Partial) ML Estimation Results for US and UK

	US presidential elections (1920-2020)		UK general elections (1918-2019)	
Parameters	Single-peaked Normal	Bernoulli-Normal mixture	Single-peaked Normal	Bernoulli-Normal mixture
$\hat{g}$	–	0.066	–	0.127
		(0.014)		(0.034)
$\hat{Q}$	–	0.626	–	0.177
		(0.233)		(0.159)
$\hat{g}\sqrt{\hat{Q}(1-\hat{Q})}$	–	0.032	–	0.049
		(0.006)		(0.006)
$\hat{\mu}$	0.554	0.529	0.727	0.622
	(0.014)	(0.009)	(0.023)	(0.031)
$\hat{\sigma}$	0.042	0.027	0.064	0.041
	(0.002)	(0.005)	(0.007)	(0.006)
Partial Log-likelihood	45.691	46.683	37.459	39.788
#.Observations	26	26	28	28

Note: Robust standard errors are reported in parentheses and they are computed following the method in Levine (1983JoE) with lag $k = 4$. The choice $k = 4$ is made based on a tradeoff between bias and precision of estimates.

Implications of the Point Estimates

1. Effects of tipping are large in size:
   - For US, tipping increases turnout by 6.6%-points.
   - For UK, tipping increases turnout by 12.7%-points.
2. Tipping in UK is more substantial than that in US: ▸ Figure
   - The jump in turnout due to tipping is higher in UK.
   - Voter turnout is more likely to reach the higher peak in UK.[3]
   - One possible explanation based on our model is that elections have higher values (and thus more likely to trigger tipping) in UK than in US.

[3] In our model the probability of reaching the higher peak is $Pr[V > \overline{V}]$.

Is Tipping due to Sampling Error?

- Recall our hypotheses:
  - No tipping (null): $g\sqrt{Q(1-Q)} = 0$ (i.e., either $g = 0$ or $Q \in \{0, 1\}$)
  - Tipping: $g\sqrt{Q(1-Q)} > 0$ (i.e., both $g > 0$ and $Q \in (0, 1)$)
- Standard t-test does not work here because the null hypothesis is at the boundary of parameter space.

Is Tipping due to Sampling Error?

- Recall our hypotheses:
  - No tipping (null): $g\sqrt{Q(1-Q)} = 0$ (i.e., either $g = 0$ or $Q \in \{0, 1\}$)
  - Tipping: $g\sqrt{Q(1-Q)} > 0$ (i.e., both $g > 0$ and $Q \in (0, 1)$)
- Standard t-test does not work here because the null hypothesis is at the boundary of parameter space.
- The procedure of our test is conceptually similar to a permutation test (Young, 2019QJE). ▸ Procedure
  - **Key idea**: how likely it is that our obtained estimates could be generated from an underlying model with no tipping (i.e., under the null hypothesis)?

	$\hat{g}$	$\sqrt{\hat{Q}(1-\hat{Q})\hat{g}^2}$
UK data	0.068	0.088
US data	0.469	0.215

  - **Result**: Tipping in US is likely a statistical fluke, in UK this is not the case.

Outline

Conclusion

- We develop a model of group behavior that combines peer pressure with expressive externality.
- Our key result is the possibility of “tipping” – group participation may jump up discontinuously when the expressive externality becomes strong enough.
- We derive implications of tipping for sports ticket pricing and voter turnout.
- We test our tipping prediction using voter turnout data from the US and the UK, and found evidence for tipping in the latter.

Thank you!

Positive Serial Correlation and Stationarity

Back

Assume no tipping and the turnout follows an $AR(1)$ process:

$$\tau_t = \rho_0 + \rho_1 \tau_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_\varepsilon^2)$$

Table: OLS Estimation Results

	$\hat{\rho}_0$	$\hat{\rho}_1$	$\hat{\sigma}_\varepsilon$	$\hat{\mu}_{stationary}$	$\hat{\sigma}_{stationary}$
US data	0.314	0.439	0.037	0.560	0.041
	(0.091)	(0.165)			
UK data	0.349	0.526	0.046	0.737	0.054
	(0.131)	(0.172)			

Note: Newey-West standard errors robust to heteroskedasticity and first-order autocorrelation are reported in parentheses below the estimates $\hat{\rho}_0$ and $\hat{\rho}_1$.

On stationarity: Augmented Dickey-Fuller tests reject the unit root hypothesis (i.e., $\rho_1 = 1$) for both U.S. and U.K. data at 5% significance level.

(Partial) Log-likelihood Function

Back

- In period t, turnout τ_t is given by

$$\tau_t = b_t + \epsilon_t$$

where b_t is a Bernoulli random variable on $\{\mu, \mu + g\}$ with $g > 0$ and $Pr[b_t = \mu] = Q \in (0, 1)$, and $\epsilon_t \sim \mathcal{N}(0, \sigma^2)$.

- The (partial) likelihood function for dataset $\{\tau_t\}_{t=1}^T$ is then

$$\mathcal{L}(\vartheta; \boldsymbol{\tau}) = \sum_{t=1}^{T} \log \left(Qe^{-\frac{(\tau_t - \mu)^2}{2\sigma^2}} + (1 - Q)e^{-\frac{(\tau_t - \mu - g)^2}{2\sigma^2}} \right) - \frac{T}{2} \log 2\pi\sigma^2$$

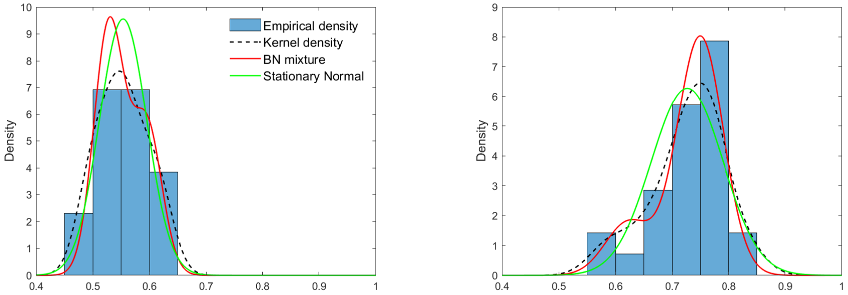
(Partial) ML estimator $\hat{\vartheta}$ maximizes $\mathcal{L}(\vartheta; \boldsymbol{\tau})$.[4]

- Remarks:
  - This estimation procedure treats turnout data as if they are independently drawn from the stationary distribution.
  - The partial ML estimator obtained this way is consistent for stationary time series (Levine, 1983JoE).

[4] These are parameters for the stationary distribution of turnout.

Estimation Results (in figure) Back

Figure: Estimated densities of turnout distribution for U.S. (left) and U.K. (right)



Note: The red lines plot the densities implied the estimated Bernoulli-Normal mixture model. The green lines plot the densities implied by the stationary normal distribution under AR(1) estimated by OLS. The black dashed curves are the estimated kernel densities (the optimal bandwidth are 0.0254 for US and 0.0323 for UK). Blue bars are the empirical density of data.

Procedure of Monte Carlo Experiment

Back

- The number of Monte Carlo trials: $M = 10000$.
- In each trial m, we
  1. simulate a sample of serially correlated turnout data using the estimated AR(1) model without tipping.
  2. obtain the partial ML estimator $\hat{\vartheta}_m$ for the simulated sample.
- These produce collections $\{\hat{g}_m\}_{m=1}^{10000}$ and $\left\{\hat{g}_m\sqrt{\hat{Q}_m(1-\hat{Q}_m)}\right\}_{m=1}^{10000}$.
- We calculate the fractions of these simulated estimates being larger than the actual estimates obtained from real turnout data.