Swiss Tournaments

CIS700/04: Machine Learning and Econometrics

Chris Hua

April 20, 2017

University of Pennsylvania

Motivation

- Tournament structure and design
- Do they work?

Swiss Tournaments

- Widely used, including chess, policy debate, Hearthstone
- Random start + power matched rounds
- In debate: preseason tournaments identify top-k debaters
    - Reaching eliminations earns a bid for the postseason tournament

Do Swiss tournaments find the top-k competitors?

Simulation: Bradley-Terry

- Tournaments are sets of pairwise comparisons
- Assume each team has an underlying strength θ
    - Simulated using lognormal distribution
- Find winner by doing a random draw

$$\Pr(Y_{i,j} = 1) = \frac{\theta_i}{\theta_i + \theta_j}$$

Simulation: Pairings?

- 2 rounds of random pairings
- 4 rounds of power-matched pairings
- Teams cannot be paired with teams they've already faced
- Prefer teams with same # of wins, otherwise, max difference of 1
- Run 500 simulations

Simulation: Pairings!

Maximum-weight perfect-matching

- Treat pairings as a graph problem
- Teams = nodes (n), possible pairings = edges (m)
- Complexity of $O(nm \log n) \sim O(n^3)$

Metrics

- Champion: Top-team is undefeated (Copeland champion)
- Top-k: Percent of the top-k teams by strength which meet selection criteria
- Spearman's ρ
- Kendalls τ

Simulations

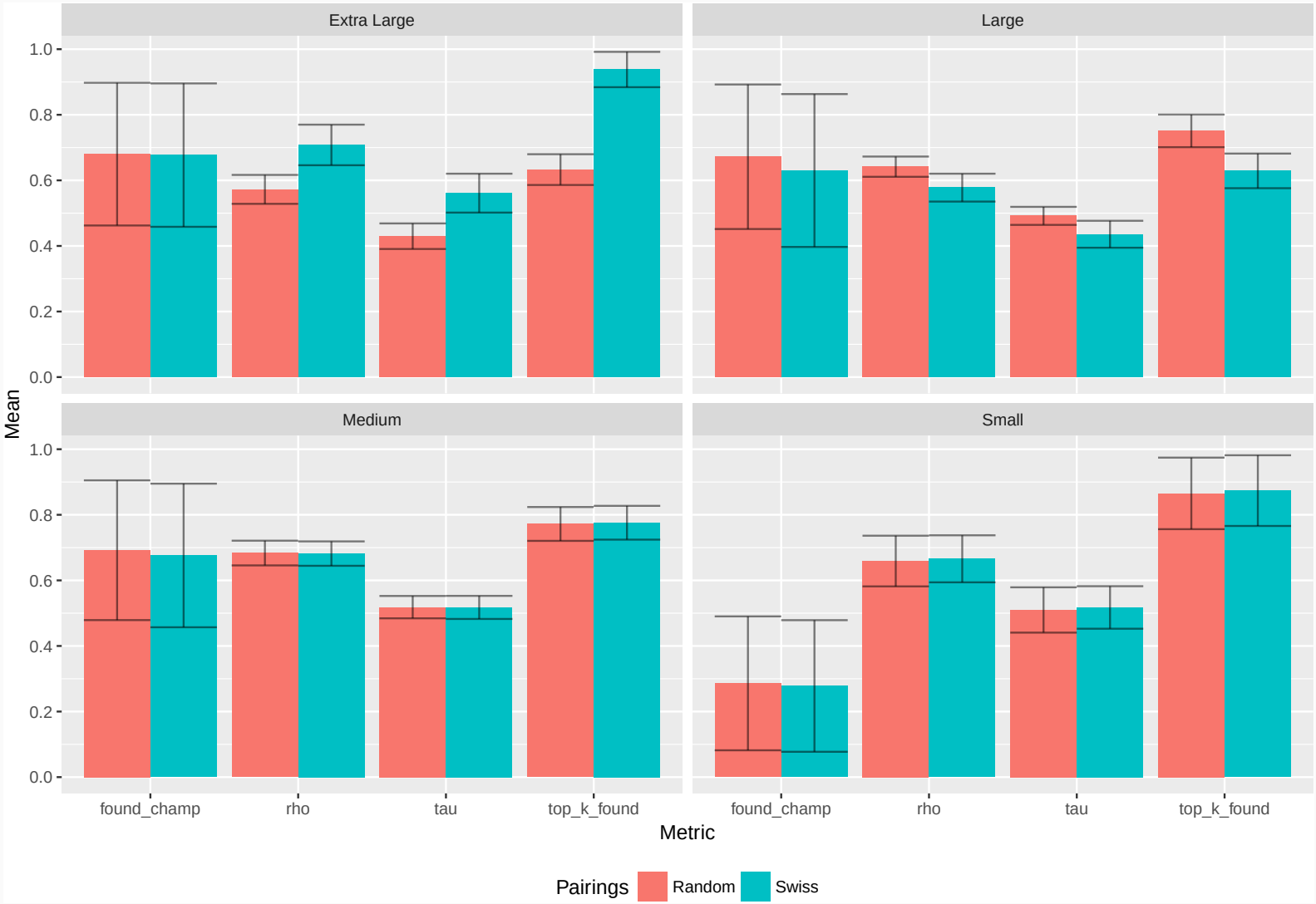
500 trials each, recorded mean and standard deviation

Size	Teams	Rounds	K
Small	32	5	8
Medium	64	6	16
Large	128	6	32
Very large	256	7	64

Real-world data

- Scraped 2009-2010 and 2010-2011 policy debate tournament results
    - 2009-2010: 13310 debated rounds by 1424 teams, in 67 tournaments.
    - Did actual MLE estimates - but hard to estimate results
- Hearthstone @ Dreamhack 2016: 190 players to pick top 8 for playoffs, 9 rounds

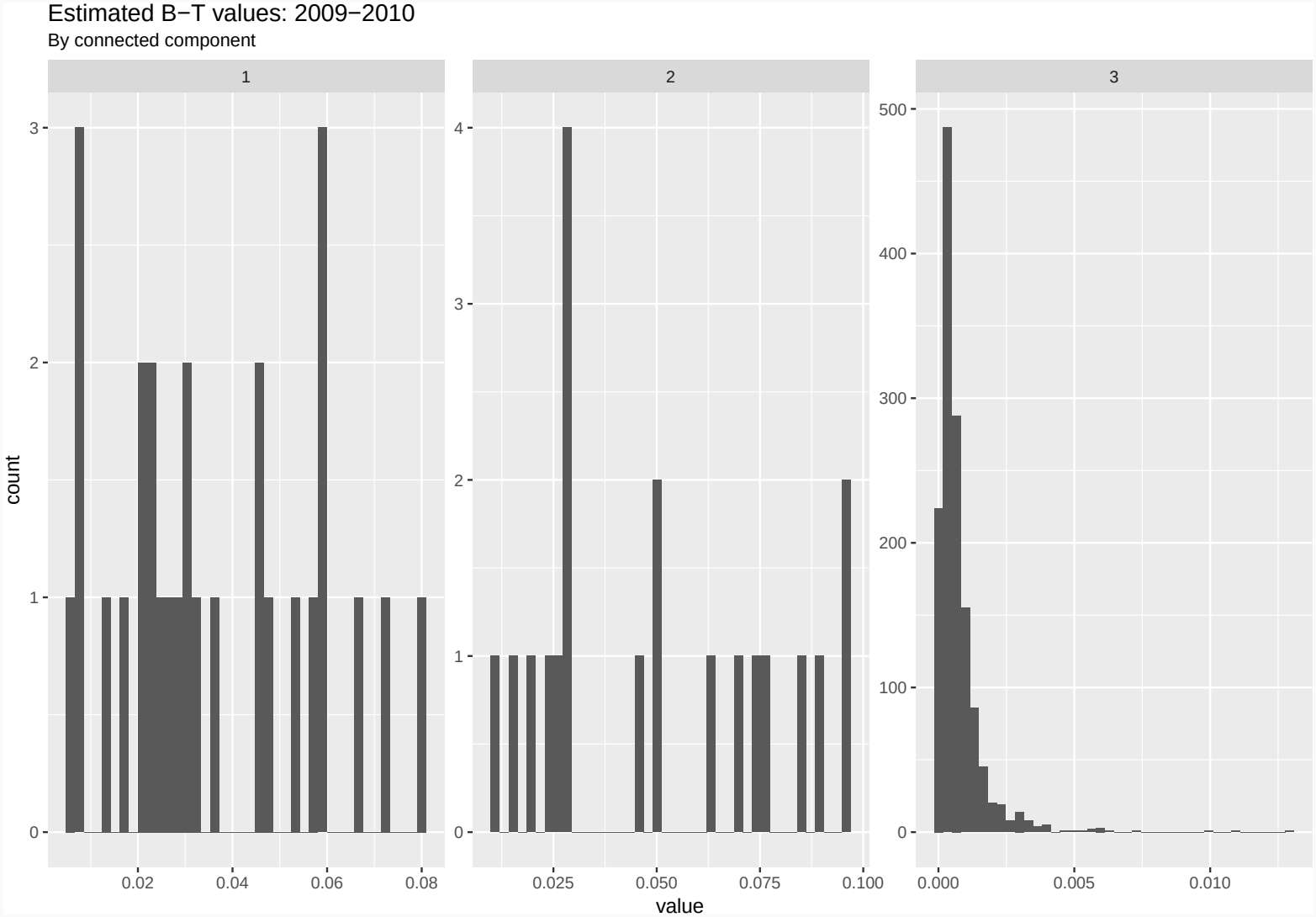
Results - synthetic data



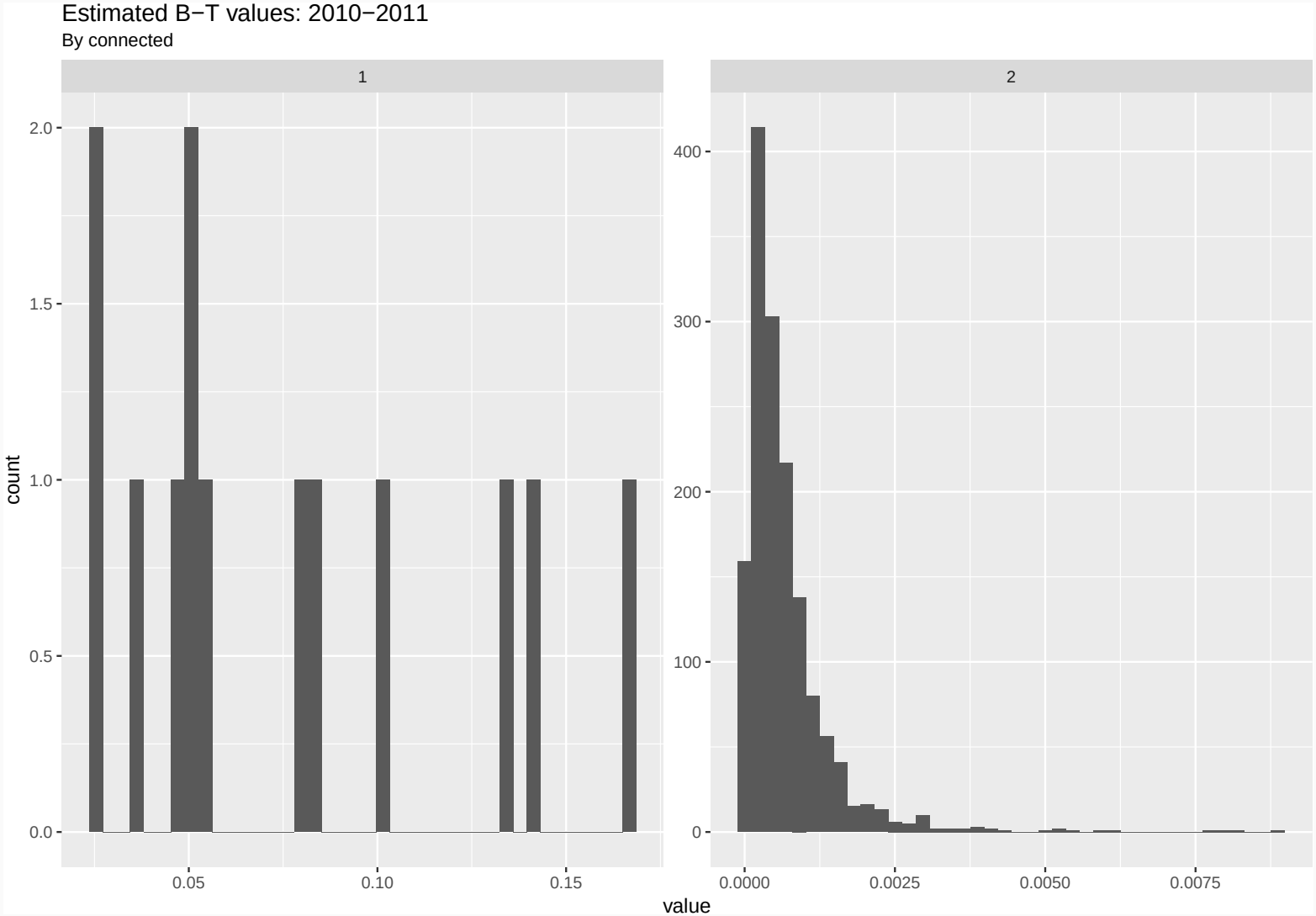
Results - Analysis

- Surprisingly, Swiss doesn't do significantly better than random pairings
- Swiss is worse (probably) at having top team go undefeated
- Swiss underperformed in large specification, overperformed in extra large tournament.

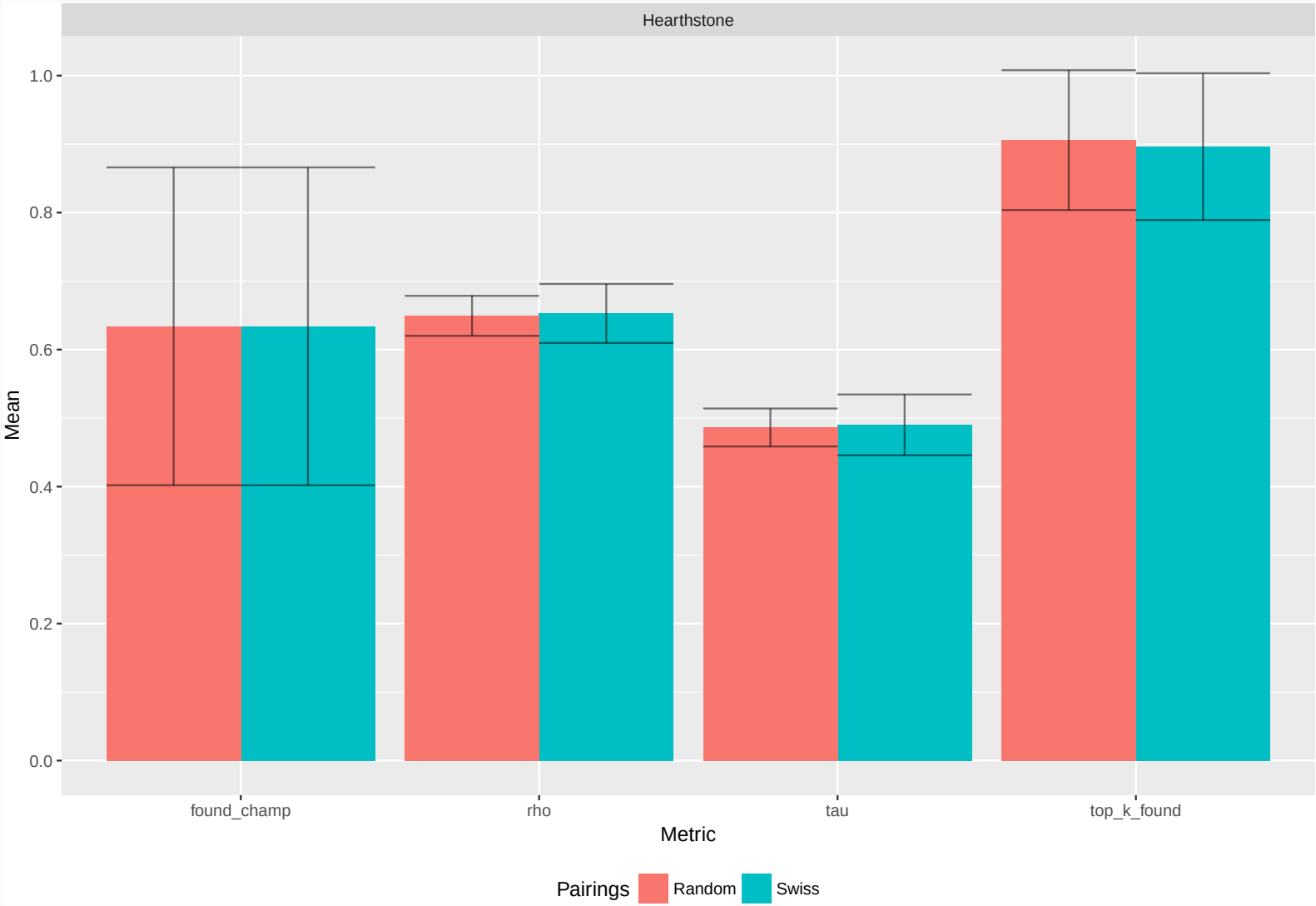
Results - BT distribution



Results - BT distribution



Results - Hearthstone



Conclusion

- Variety of real world settings tested
- Swiss rarely outperforms random pairings, and usually does very similarly
- Further work:
    - Different pairing strategies
    - Further investigation of effect of size