

Scoring Works in Crowdsourcing: How Many Control Questions are Enough?

Qiang Liu, Mark Steyvers, Alexander Ihler (UC Irvine)

Crowdsourcing



amazon mechanical turk™
Artificial Artificial Intelligence



Machine learning
datasets



Knowledge Base



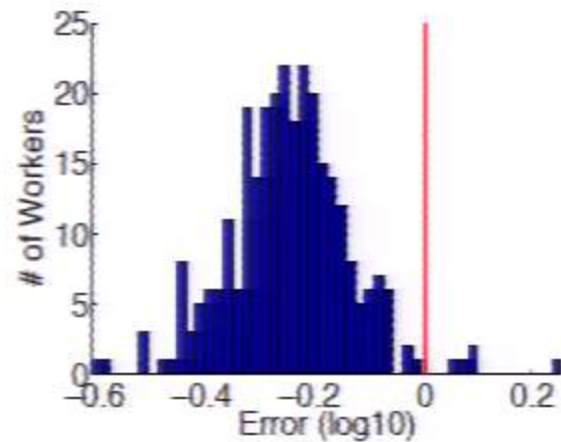
Web Search Relevance



Forecasting: Economics,
sports, politics

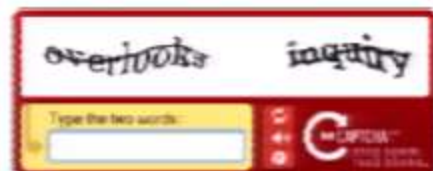
Biases and Uncertainties in Workers

Guessing prices:

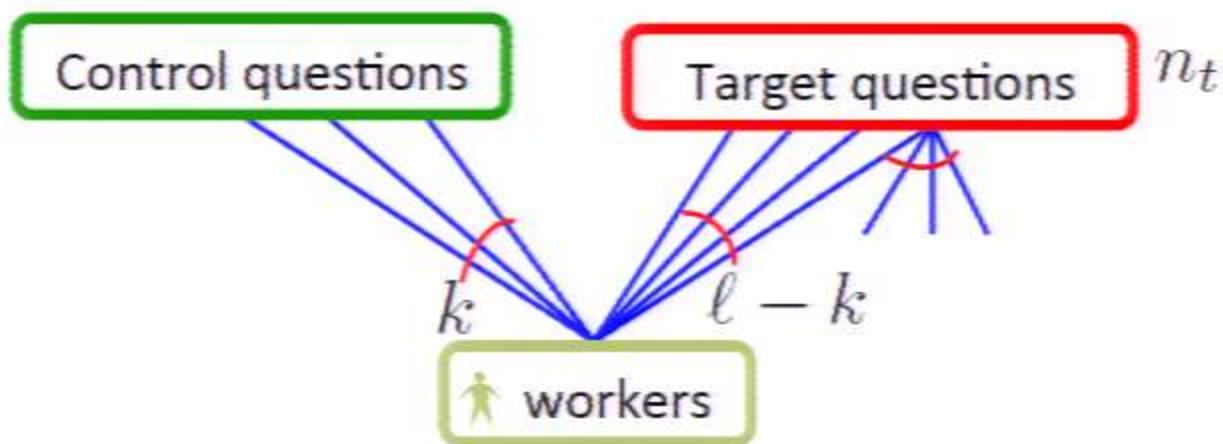


Humans under-estimate prices

- Solution: Seed “control questions” to score / correct the workers



Question: How many control items to use?



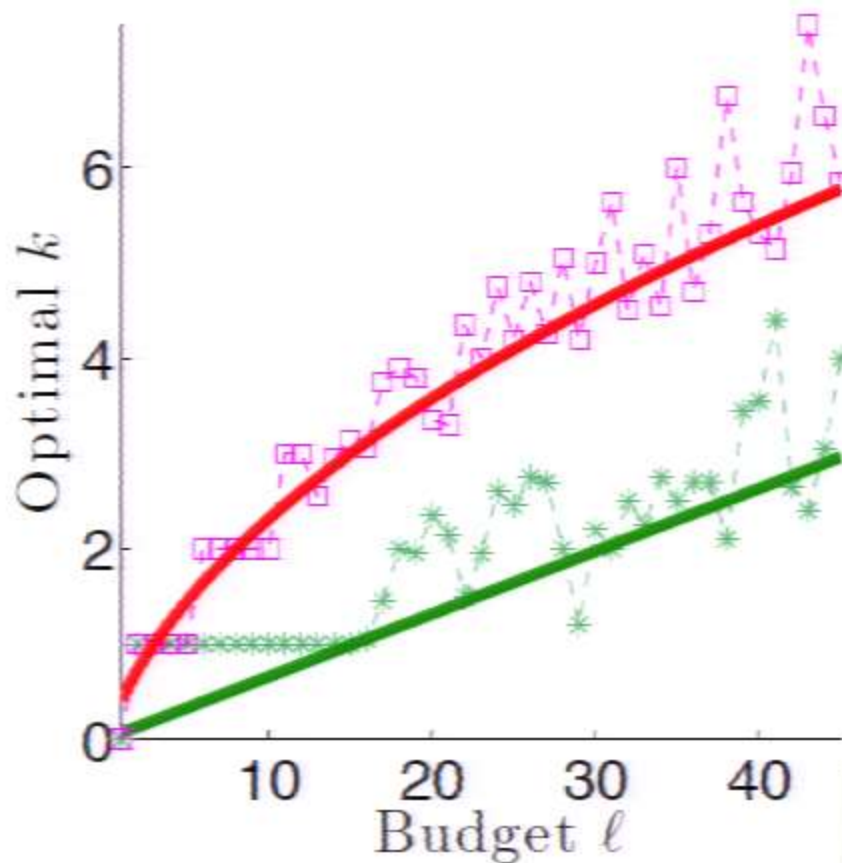
ℓ : # of all questions asked workers (budget)

k : # of control questions asked workers

- Goal: optimal k to minimize error on the target questions
- Analyze two methods: two-stage vs. joint estimation

Theoretical & Empirical Results

- Point Spreads of NFL Games
- 245 NFL games, 386 workers



Two-stage estimator:

$$k^* = \sqrt{\ell}$$

Joint estimator:

$$k^* = \ell / \sqrt{n_t}$$

- See more details in poster Sat72
- Crowdsourcing Workshop (Monday)