Earning While Learning

How to Run Batched Bandit Experiments

Davud Rostam-Afschar
University of Mannheim

2025 Oceania Stata Conference

Based on joint work with Jan Kemper; applied in Gaul et al. (2024)

Adaptive experimental designs

- ► Randomized controlled trials gold standard of causal inference

- ► Adaptive experiments allow "earning while learning"

- ► Push to replace non-adaptive randomized trials with bandits

  - ► In medicine, economics, political science, survey methods research, education, psychology, ...

  - ► Practitioners use bandit algorithms

  - ► Can improve outcomes for participants (optimize regret)

  - ► Can improve policies learned at the end of trial (best-arm identification)

- ► Some popular algorithms

  - ► ε-first

  - ► ε-greedy

  - ► Thompson sampling

Recent "exploding" growth of papers

- ▶ In medicine (Lei et al., 2022)
- ▶ economics and finance (Hirano and Porter, 2023; Chen and Andrews, 2023; Kasy and Sautmann, 2021; Hadad et al., 2021; Avivi et al., 2021)
- ▶ political science (Offer-Westort et al., 2021)
- ▶ survey methods research (Gaul et al., 2024)
- ▶ education (Rafferty et al., 2019)
- ▶ psychology (Schulz et al., 2020)
- ▶ ...
- ▶ Practitioners use bandit algorithms
  (Hill et al., 2017; Scott, 2015; Agarwal et al., 2014; Chapelle and Li, 2011; Scott, 2010; Graepel et al., 2010)
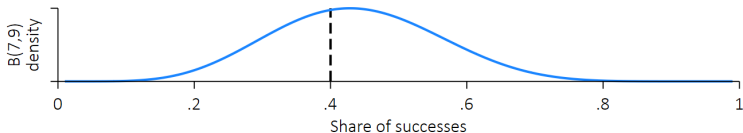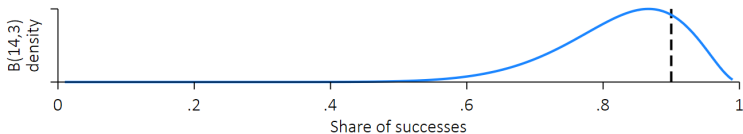
Stylized data structure

Obs	Selected arm	Reward
1	A	0
2	B	0
3	A	1
4	B	0
5	A	0
6	B	1
7	A	1
8	B	0
9	A	0
10	A	1
11	A	1
12	B	0
13	A	1
14	A	0
15	A	1
16	B	0

► Does arm A or arm B perform better?

► Which arm to play in next trial (round 17)?

Thompson (1933, 1935) sampling
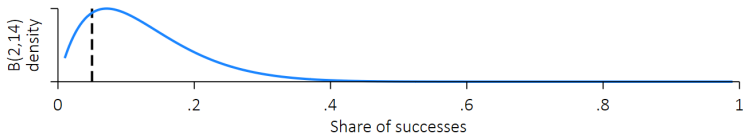
- ▶ Beta-Bernoulli Thompson sampling
- ▶ Models uncertainty about the shape of the distribution and the expected outcome R explicitly

Click to watch!

Thompson sampling

- Beta distribution $B(R_{k,t}|\alpha_k, \beta_k)$ denotes the density of the beta distribution for random variable R_t with parameters α_k and β_k
- Posterior distribution $P(\theta|R_t)$ is also beta with parameters that can be updated according to a simple rule:

$$(\alpha_k, \beta_k) = \begin{cases} (\alpha_k, \beta_k) & \text{if chosen arm} \neq k, \\ (\alpha_k, \beta_k) + (R_t, 1 - R_t) & \text{if chosen arm} = k. \end{cases}$$

- α_k or β_k increases by one with each observed success or failure
- Distribution more concentrated as $\alpha_k + \beta_k$ grows
- Mean $\alpha_k/(\alpha_k + \beta_k)$ and variance $\frac{\alpha_k\beta_k}{(\alpha_k+\beta_k)^2(\alpha_k+\beta_k+1)}$

Bandits $>>$ A/B Tests

- ▶ Push to replace non-adaptive randomized trials with bandits
    - ▶ In development and labor economics, finance, biostats, health, ...
    - ▶ Can improve outcomes for participants (optimize regret)
    - ▶ Can improve policies learned at the end of trial (best-arm identification)
- ▶ **Problem:**
    - ▶ Bandits are not easy to implement
      Not available in statistical software like Stata
    - ▶ Bandits break inference
      Adaptive arm allocations
      $\rightarrow$ breaks asymptotics of usual estimators
      $\rightarrow$ wrong confidence intervals
- ▶ **Solution: Batched OLS (BOLS) for Batched Bandits**

A simple example

► OLS and BOLS under Beta-Bernoulli two-arm Thompson Sampling with batch size $N_t = 100$ at batch $t = 10$

► All simulations are with no margin ($\beta_1 = \beta_0 = 0$)
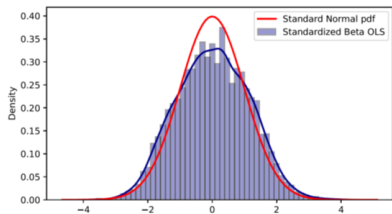
(a) Empirical distribution of standardized OLS estimator for the margin
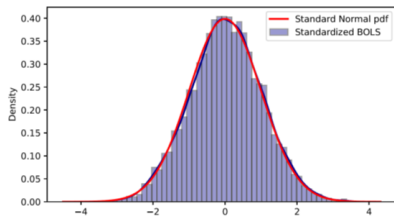
(b) Empirical distribution of standardized BOLS estimator for the margin

Stylized data structure

Obs	Selected arm	Batch	Reward
1	A	0	.
2	B	0	.
3	A	0	.
4	B	0	.
5	.	1	.
6	.	1	.
7	.	1	.
8	.	1	.
9	.	2	.
10	.	2	.
11	.	2	.
12	.	2	.
13	.	3	.
14	.	3	.
15	.	3	.
16	.	3	.

Stylized data structure

Obs	Selected arm	Batch	Reward
1	A	0	0
2	B	0	0
3	A	0	1
4	B	0	0
5	.	1	.
6	.	1	.
7	.	1	.
8	.	1	.
9	.	2	.
10	.	2	.
11	.	2	.
12	.	2	.
13	.	3	.
14	.	3	.
15	.	3	.
16	.	3	.

Stylized data structure

Obs	Selected arm	Batch	Reward
1	A	0	0
2	B	0	0
3	A	0	1
4	B	0	0
5	A	1	.
6	B	1	.
7	A	1	.
8	B	1	.
9	.	2	.
10	.	2	.
11	.	2	.
12	.	2	.
13	.	3	.
14	.	3	.
15	.	3	.
16	.	3	.

Stylized data structure

Obs	Selected arm	Batch	Reward
1	A	0	0
2	B	0	0
3	A	0	1
4	B	0	0
5	A	1	0
6	B	1	1
7	A	1	1
8	B	1	0
9	.	2	.
10	.	2	.
11	.	2	.
12	.	2	.
13	.	3	.
14	.	3	.
15	.	3	.
16	.	3	.

Stylized data structure

Obs	Selected arm	Batch	Reward
1	A	0	0
2	B	0	0
3	A	0	1
4	B	0	0
5	A	1	0
6	B	1	1
7	A	1	1
8	B	1	0
9	A	2	.
10	A	2	.
11	A	2	.
12	B	2	.
13	.	3	.
14	.	3	.
15	.	3	.
16	.	3	.

Stylized data structure

Obs	Selected arm	Batch	Reward
1	A	0	0
2	B	0	0
3	A	0	1
4	B	0	0
5	A	1	0
6	B	1	1
7	A	1	1
8	B	1	0
9	A	2	0
10	A	2	1
11	A	2	1
12	B	2	0
13	.	3	.
14	.	3	.
15	.	3	.
16	.	3	.

Stylized data structure

Obs	Selected arm	Batch	Reward
1	A	0	0
2	B	0	0
3	A	0	1
4	B	0	0
5	A	1	0
6	B	1	1
7	A	1	1
8	B	1	0
9	A	2	0
10	A	2	1
11	A	2	1
12	B	2	0
13	A	3	.
14	A	3	.
15	A	3	.
16	B	3	.

Stylized data structure

Obs	Selected arm	Batch	Reward
1	A	0	0
2	B	0	0
3	A	0	1
4	B	0	0
5	A	1	0
6	B	1	1
7	A	1	1
8	B	1	0
9	A	2	0
10	A	2	1
11	A	2	1
12	B	2	0
13	A	3	1
14	A	3	0
15	A	3	1
16	B	3	0

Stylized data structure

Obs	Selected arm	Batch	Reward	True Expected Reward
1	A	0	0	0.5
2	B	0	0	0.2
3	A	0	1	0.5
4	B	0	0	0.2
5	A	1	0	0.5
6	B	1	1	0.2
7	A	1	1	0.5
8	B	1	0	0.2
9	A	2	0	0.5
10	A	2	1	0.2
11	A	2	1	0.5
12	B	2	0	0.2
13	A	3	1	0.5
14	A	3	0	0.2
15	A	3	1	0.5
16	B	3	0	0.2

Stylized data structure

Obs	Selected arm	Batch	Reward	True Expected Reward	OLS
1	A	0	0	0.5	0.600
2	B	0	0	0.2	0.167
3	A	0	1	0.5	0.600
4	B	0	0	0.2	0.167
5	A	1	0	0.5	0.600
6	B	1	1	0.2	0.167
7	A	1	1	0.5	0.600
8	B	1	0	0.2	0.167
9	A	2	0	0.5	0.600
10	A	2	1	0.2	0.600
11	A	2	1	0.5	0.600
12	B	2	0	0.2	0.167
13	A	3	1	0.5	0.600
14	A	3	0	0.2	0.600
15	A	3	1	0.5	0.600
16	B	3	0	0.2	0.167

Stylized data structure

Obs	Selected arm	Batch	Reward	True Expected Reward	OLS	Batch-Wise OLS
1	A	0	0	0.5	0.600	0.500
2	B	0	0	0.2	0.167	0.000
3	A	0	1	0.5	0.600	0.500
4	B	0	0	0.2	0.167	0.000
5	A	1	0	0.5	0.600	0.500
6	B	1	1	0.2	0.167	0.500
7	A	1	1	0.5	0.600	0.500
8	B	1	0	0.2	0.167	0.500
9	A	2	0	0.5	0.600	0.667
10	A	2	1	0.2	0.600	0.667
11	A	2	1	0.5	0.600	0.667
12	B	2	0	0.2	0.167	0.000
13	A	3	1	0.5	0.600	0.667
14	A	3	0	0.2	0.600	0.667
15	A	3	1	0.5	0.600	0.667
16	B	3	0	0.2	0.167	0.000

Stylized data structure

Obs	Selected arm	Batch	Reward	True Expected Reward	OLS	Batch-Wise OLS	ω_t
1	A	0	0	0.5	0.600	0.500	$\sqrt{\frac{2 \times 2}{2+2}}$
2	B	0	0	0.2	0.167	0.000	$\sqrt{\frac{2 \times 2}{2+2}}$
3	A	0	1	0.5	0.600	0.500	$\sqrt{\frac{2 \times 2}{2+2}}$
4	B	0	0	0.2	0.167	0.000	$\sqrt{\frac{2 \times 2}{2+2}}$
5	A	1	0	0.5	0.600	0.500	$\sqrt{\frac{2 \times 2}{2+2}}$
6	B	1	1	0.2	0.167	0.500	$\sqrt{\frac{2 \times 2}{2+2}}$
7	A	1	1	0.5	0.600	0.500	$\sqrt{\frac{2 \times 2}{2+2}}$
8	B	1	0	0.2	0.167	0.500	$\sqrt{\frac{2 \times 2}{2+2}}$
9	A	2	0	0.5	0.600	0.667	$\sqrt{\frac{1 \times 3}{1+3}}$
10	A	2	1	0.2	0.600	0.667	$\sqrt{\frac{1 \times 3}{1+3}}$
11	A	2	1	0.5	0.600	0.667	$\sqrt{\frac{1 \times 3}{1+3}}$
12	B	2	0	0.2	0.167	0.000	$\sqrt{\frac{1 \times 3}{1+3}}$
13	A	3	1	0.5	0.600	0.667	$\sqrt{\frac{1 \times 3}{1+3}}$
14	A	3	0	0.2	0.600	0.667	$\sqrt{\frac{1 \times 3}{1+3}}$
15	A	3	1	0.5	0.600	0.667	$\sqrt{\frac{1 \times 3}{1+3}}$
16	B	3	0	0.2	0.167	0.000	$\sqrt{\frac{1 \times 3}{1+3}}$

Point estimates OLS vs. BOLS

Aggregate or batched OLS (BOLS) estimator

$$\Delta^{\text{BOLS}} = \frac{\sum_t^T \omega_t \times \Delta_t^{BOLS}}{\sum_t^T \omega_t},$$

where $\omega_t = \sqrt{\frac{N_{t,k} \times N_{t,b}}{N_{t,k} + N_{t,b}}}$.

- ▶ $N_{t,k}$ is the number of times that comparison arm k was played

- ▶ $N_{t,b}$ is the number of times that baseline arm b was played

- ▶ weights batchwise estimates

- ▶ such that the aggregate margins are consistent and asymptotically **normally** distributed (Zhang et al., 2020)

Point estimates OLS vs. BOLS

Example from stylized data structure

OLS $\qquad \widehat{\text{Reward}} = 0.6 - 0.433 \times \mathbb{1}_{\text{arm B}}$

BOLS $-0.443 = \dfrac{1 \times 0.5 + 1 \times 0 + \sqrt{\frac{1 \times 3}{1+3}} \times 0.667 + \sqrt{\frac{1 \times 3}{1+3}} \times 0.667}{1 + 1 + \sqrt{\frac{1 \times 3}{1+3}} + \sqrt{\frac{1 \times 3}{1+3}}}$

$\qquad\qquad \widehat{\text{Reward}} = 0.6 - 0.443 \times \mathbb{1}_{\text{arm B}}$

Inference OLS vs. BOLS

$$Pr\left(\Delta^{\text{BOLS}} - c\sigma w_t \leq \mu \leq \Delta^{\text{BOLS}} + c\sigma w_t \right) = 1 - \alpha,$$

► where Δ^{BOLS} is the weighted estimated marginal effect
► μ is the hypothesized difference between means of the arms
► c is a critical value, e.g., the $1 - \alpha/2 = 97.5$th percentile of a normal
► σ reflects the sampling error
► w_t is a weight correcting the bias due to adaptive sampling

$$w_t = \sqrt{T}/\sum_{t=1}^{T} \omega_t.$$

► T is the total number of batches

The bbandits command

Syntax & Options   **Click to download!**

▶ <u>bbandits</u> *reward assignedarm batch, options*

Returned results
▶ OLS margins
▶ BOLS margins
▶ *z* statistics
▶ p-values
▶ BOLS 95% confidence intervals
▶ observations of the reference arm
▶ observations of the treatment arm

Empirical application (Kasy and Sautmann, 2021)

Six call methods to enroll rice farmers

- ▶ Kasy and Sautmann (2021) designed an experiment using <u>exploration sampling</u> for Precision Agriculture for Development

- ▶ NGO that works with government partners to provide a phone-based personalized agricultural extension service to farmers in India

- ▶ Aim is to choose best call methods to enroll rice farmers in one state

Six call methods to enroll rice farmers

▶ The outcome (reward) is a binary variable for call completion:

  ▶ = 1 if call recipient answered five questions asked during call

  ▶ = 0 otherwise

```
                           Voice Call Treatments


SMS alerts 1h ahead     SMS alerts 24h ahead        No SMS alert

10 am    6:30 pm        10 am    6:30 pm          10 am    6:30 pm
```
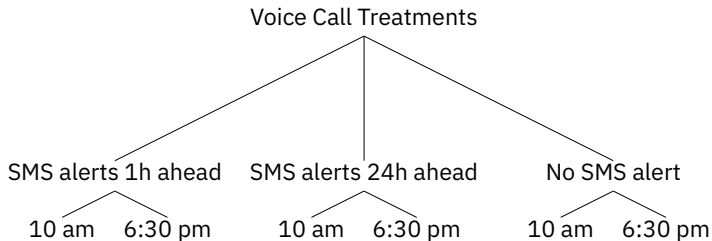
Empirical application (Kasy and Sautmann, 2021)

▶ <u>Exploration sampling</u> replaces the Thompson assignment shares

▶ modification shifts weight away from the best performing option to competing treatments

▶ 10,000 valid phone numbers randomly assigned to one of 16 batches

▶ batch size was 600 numbers each (and one with 400)

▶ From June 3, 2019 batches run every other day, completed next day

Empirical application (Kasy and Sautmann, 2021)

```
.
. use "example data\kasy_sautmann_2021.dta", clear
. bbandits outcome treatment date

Number of obs                          =       10000
Est. Rewards only best arm             =        1926    Mean reward best arm    =      0.1926
Actual total reward                    =        1804    Actual mean reward      =      0.1804
Est. reward uniformly chosen arms      =        1709    Mean reward uniform     =      0.1709
```

Arm b	Mean Reward						Share arm b
	0.1606						0.0903
k v. b	Margin OLS	Margin BOLS	z	P>|z|	[95% Conf. Interval]		Share arm k
1-0	0.0320	0.0406	2.61	0.009	0.0101	0.0711	0.3931
2-0	0.0185	0.0249	1.51	0.132	-0.0075	0.0572	0.2234
3-0	-0.0158	-0.0289	-1.12	0.262	-0.0795	0.0216	0.0366
4-0	0.0078	0.0188	0.97	0.330	-0.0191	0.0568	0.1081
5-0	0.0192	0.0243	1.40	0.161	-0.0097	0.0582	0.1485

Treatment	0	1	2	3	4	5
SMS	—	1h ahead	24h ahead	—	1h ahead	24h ahead
Call time	10 am	10 am	10 am	6:30 pm	6:30 pm	6:30 pm

Empirical application (Kasy and Sautmann, 2021)
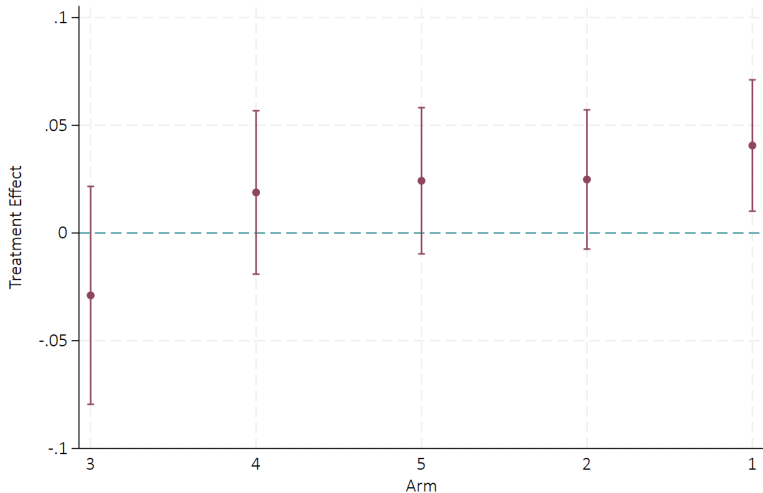
The figure was generated using kasy_sautmann_2021.dta and running
bbandits outcome treatment date

Empirical application (Kasy and Sautmann, 2021)
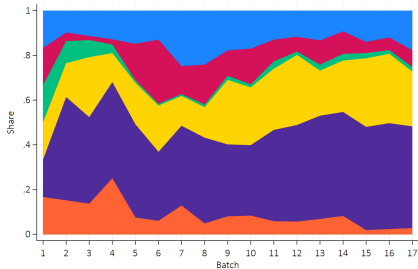
(a) Batchwise shares
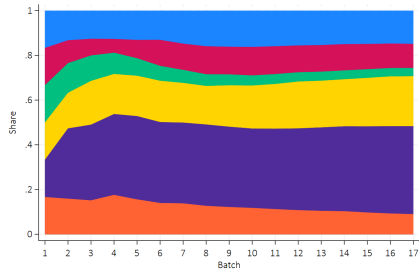
(b) Cumulative shares

The figure was generated using `kasy_sautmann_2021.dta` and running `bbandits outcome treatment date`

Empirical application (Kasy and Sautmann, 2021)

Clear best and worst arms

▶ Best: Calling farmers at 10 am after a message an hour ahead of time

▶ Worst: Calling at 6:30 pm without a text message alert

Improvement of success rate

▶ 18.04% success rates within the experiment

▶ 17.15% success rate with equal assignment

32 invitation messages for business survey

► Gaul et al. (2024) designed an experiment using <u>Thompson sampling</u> to support the German Business Panel (GBP)

► Aim is to select among a variety of different invitation messages to survey firm decision makers in Germany

► The GBP is a web-based survey study of firm decision makers in Germany that invites participants each work day
(see Bischof et al. (2024); Hack and Rostam-Afschar (2024))

► The outcome (reward) is a binary variable for the start of the survey:
  ► = 1 if email invitation recipient started the survey
  ► = 0 otherwise

Empirical application (Gaul et al., 2024)
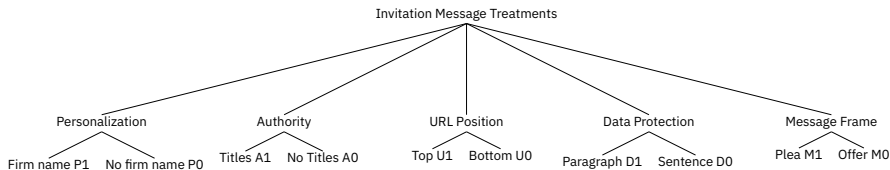
- ▶ Five components of invitation letters and their full interactions
  - → $2^5 = 32$ treatments
    - ▶ **personalization** by mentioning or not mentioning the firm name
    - ▶ **authority** of the sender by listing the official full academic titles along with the senders' names or their names only
    - ▶ **URL position** to start the survey at the top or bottom of the invitation
    - ▶ **data protection** in a separate paragraph with two strongly phrased sentences or in a single sentence
    - ▶ **message frame** by including phrases that plea for support in the survey's cause or to simply offer to participate

Empirical application (Gaul et al., 2024)

▶ 11,000 randomly selected contacts from firms in Germany

▶ Assigned to each of 15 batches from a list of 176,000 contacts

▶ Each batch corresponds to a week between
August 16, 2022 and November 25, 2022

▶ First four batches used fixed and balanced burn-in phase
with treatment probability $1/32$

▶ From batch 5, Thompson assignment rule for each consecutive batch

```
. use "example data\gaul_et_al_2024.dta", clear

. bbandits reward selected trial
```

Number of obs	=	176000			
Est. Rewards only best arm	=	8623	Mean reward best arm	=	0.0490
Actual total reward	=	7833	Actual mean reward	=	0.0445
Est. reward uniformly chosen arms	=	7430	Mean reward uniform	=	0.0422

Arm b	Mean Reward						Share arm b
	0.0417						0.0181

| k v. b	Margin OLS	Margin BOLS	z	P>|z|	[95% Conf. Interval]		Share arm k
1-0	−0.0014	0.0023	0.37	0.712	−0.0100	0.0146	0.0220
2-0	−0.0003	0.0097	1.50	0.135	−0.0030	0.0224	0.0288
3-0	−0.0028	−0.0023	−0.37	0.710	−0.0144	0.0098	0.0137
4-0	−0.0030	0.0039	0.57	0.567	−0.0095	0.0174	0.0125
5-0	0.0022	0.0008	0.14	0.888	−0.0103	0.0119	0.0312
6-0	−0.0060	−0.0086	−1.32	0.187	−0.0214	0.0042	0.0121
7-0	0.0018	0.0025	0.47	0.638	−0.0078	0.0127	0.0284
8-0	−0.0003	0.0036	0.57	0.570	−0.0087	0.0158	0.0141
9-0	0.0048	0.0030	0.50	0.619	−0.0088	0.0147	0.0444
10-0	−0.0003	0.0011	0.19	0.851	−0.0105	0.0128	0.0162
11-0	−0.0000	0.0029	0.55	0.583	−0.0075	0.0134	0.0308
12-0	−0.0077	−0.0058	−0.73	0.466	−0.0213	0.0097	0.0097
13-0	−0.0009	0.0022	0.40	0.692	−0.0088	0.0132	0.0186
14-0	0.0068	0.0078	1.51	0.130	−0.0023	0.0180	0.0715
15-0	0.0029	0.0066	1.16	0.245	−0.0045	0.0177	0.0331
16-0	0.0011	0.0017	0.30	0.762	−0.0093	0.0127	0.0219
17-0	0.0042	0.0064	1.23	0.218	−0.0038	0.0165	0.0527
18-0	−0.0005	0.0025	0.45	0.650	−0.0084	0.0135	0.0255
19-0	0.0015	0.0045	0.78	0.438	−0.0068	0.0158	0.0256
20-0	0.0061	0.0107	2.02	0.044	0.0003	0.0210	0.0571
21-0	0.0060	0.0084	1.53	0.126	−0.0024	0.0191	0.0271
22-0	0.0072	0.0194	3.61	0.000	0.0089	0.0300	0.1840
23-0	−0.0020	−0.0011	−0.19	0.847	−0.0126	0.0103	0.0166
24-0	−0.0012	0.0021	0.36	0.718	−0.0094	0.0137	0.0190
25-0	0.0022	0.0079	1.52	0.129	−0.0023	0.0182	0.0308
26-0	−0.0039	−0.0018	−0.27	0.787	−0.0147	0.0111	0.0119
27-0	−0.0037	−0.0031	−0.48	0.628	−0.0155	0.0094	0.0155
28-0	−0.0028	0.0015	0.26	0.797	−0.0100	0.0130	0.0183
29-0	0.0020	0.0077	1.38	0.168	−0.0032	0.0186	0.0400
30-0	0.0007	0.0048	0.83	0.405	−0.0064	0.0160	0.0210
31-0	0.0024	0.0024	0.44	0.658	−0.0081	0.0128	0.0278

Empirical application (Gaul et al., 2024)
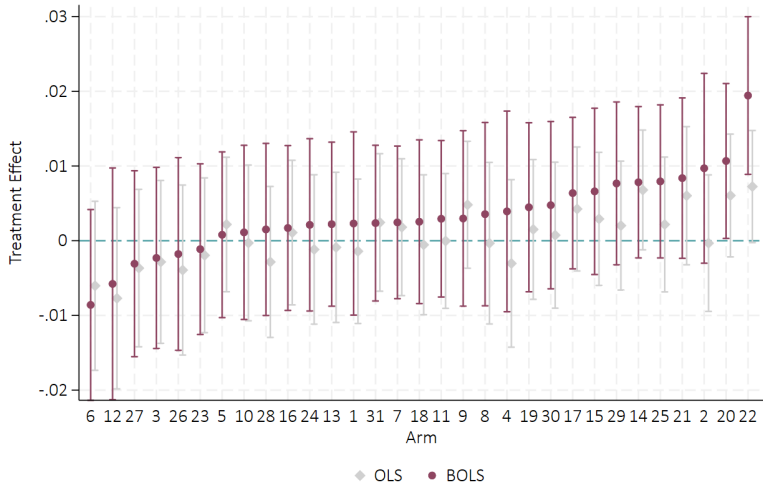
The figure was generated using `gaul_et_al_2024.dta` and running
`bbandits reward selected trial`

Empirical application (Gaul et al., 2024)
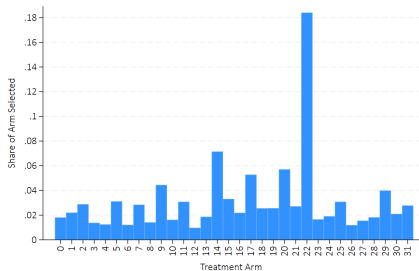
(a) Total frequency of treatment assignment
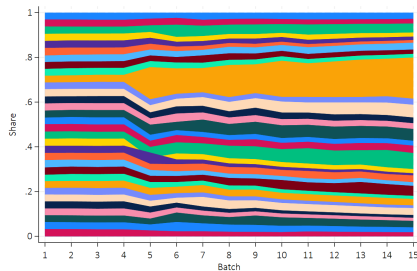
(b) Cumulative shares of arms played

The figure was generated using `gaul_et_al_2024.dta` and running
`bbandits reward selected trial`

Empirical application (Gaul et al., 2024)

Clear best and worst arms

► Using personalization, authority, and pleading for support has greatest success

Interaction effects important, too

More results in the paper...

Monte Carlo Simulations

► *Click to watch*: OLS fails normality when margin is small

► *Click to watch*: BOLS normal even when margin is small

► Run own simulations with

```
bbandit_sim 0.5 0.4 0.3, size(200) batch(10) clipping(0.1)
Thompson plot_Thompson
```

Running own bandit experiments interactively

- ▶ . bbandit_initialize, batch(10) arms(3)
  exploration_phase(2)

	ID	reward	chosen_arm	batch
1	school_1	.	3	1
2	school_2	.	3	1
3	school_3	.	3	1
4	school_4	.	2	1
5	school_5	.	3	1
6	school_6	.	1	1
7	school_7	.	1	1
8	school_8	.	2	1
9	school_9	.	1	1
10	school_10	.	3	1

Running own bandit experiments interactively

- ► . bbandit_update reward chosen_arm batch, thompson clipping(0.2) excel("mypath")

	A	B	C	D	E	F
1	ID	rand	reward	chosen_arm	batch	chosen_arm_numeric
192	school_19	1	0	1	2	2
193	school_19	1	1	3	2	0
194	school_19	0	1	3	2	0
195	school_19	1	0	3	2	0
196	school_19	0	1	2	2	1
197	school_19	0	1	1	2	2
198	school_19	1	0	2	2	1
199	school_19	1	0	3	2	0
200	school_19	0	1	2	2	1
201	school_20	0	1	1	2	2
202	school_20	1		3	3	0
203	school_20	1		3	3	0
204	school_20	1		3	3	0
205	school_20	1		2	3	1
206	school_20	0		2	3	1

Running own bandit experiments interactively

- ► `. bbandit_update reward chosen_arm batch, thompson`
  `clipping(0.2) excel("mypath")`

Running own bandit experiments interactively

- `. bbandit_update reward chosen_arm batch, thompson clipping(0.2) excel("mypath")`
- …

Running own bandit experiments interactively

- `. bbandit_update reward chosen_arm batch, thompson clipping(0.2) excel("mypath")`
- …
- analyse with

  `. bbandits reward chosen_arm batch`

Best Practices

- ▶ Report OLS and BOLS

- ▶ BOLS inference in the small margin case **correct** but...

- ▶ OLS inference in the large margin case **more precise**

- ▶ Check batch-wise OLS estimates

- ▶ **At least 50 observations** per batch and arm

- ▶ From *statistical testing* perspective:

  more observations per batch and arm better

- ▶ from *regret optimization* perspective

  fewer observations and thus fails are better

- ▶ use **bbandits** to simulate, visualize, and analyse bandit experiments

Conclusions

► Bandits may improve learning and exploitation

► There is a push to use more bandits in real experiments in development and labor econ, biostats, health, ...and accounting!

► need for valid inference to support conclusions

  ► bandits break inference

  ► researchers want valid confidence intervals

► **Batched bandit inference (BBandit)**

  ► First Stata routine for adaptive experiments

  ► allows valid statistical inference & correct coverage for batched bandits

  ► easy illustrations for statistical learning from adaptively collected data

Thank you!
https://rostam-afschar.de/

References I

Agarwal, D., B. Long, J. Traupman, D. Xin, and L. Zhang (2014): "LASER: A Scalable Response Prediction Platform for Online Advertising," in Proceedings of the 7th ACM International Conference on Web Search and Data Mining, WSDM '14, 173–182, New York, NY, USA. Association for Computing Machinery.

Avivi, H., P. Kline, E. Rose, and C. Walters (2021): "Adaptive Correspondence Experiments," AEA Papers and Proceedings, 111, 43–48.

Bischof, J., P. Doerrenberg, D. Rostam-Afschar, D. Simons, and J. Voget (2024): "The German Business Panel: Firm-Level Data for Accounting and Taxation Research," European Accounting Review.

Chapelle, O., and L. Li (2011): "An Empirical Evaluation of Thompson Sampling," in Advances in Neural Information Processing Systems, ed. by J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Weinberger, vol. 24. Curran Associates, Inc.

Chen, J., and I. Andrews (2023): "Optimal Conditional Inference in Adaptive Experiments," .

Gaul, J. J., F. Keusch, D. Rostam-Afschar, and T. Simon (2024): "Invitation Messages for Business Surveys: A Multi-Armed Bandit Experiment," Discussion Paper 170, TRR 266 Accounting for Transparency, cond. accepted Survey Research Methhods.

References II

Graepel, T., J. Q. Candela, T. Borchert, and R. Herbrich (2010): "Web-scale Bayesian click-through rate prediction for sponsored search advertising in Microsoft's Bing search engine," in Proceedings of the 27th International Conference on International Conference on Machine Learning (ICML '10), 13–20. Omnipress, Madison, WI, USA,.

Hack, L., and D. Rostam-Afschar (2024): "Understanding Firm Dynamics with Daily Data," CRC TR 224 Discussion Paper Series 593, University of Bonn and University of Mannheim, Germany.

Hadad, V., D. A. Hirshberg, R. Zhan, S. Wager, and S. Athey (2021): "Confidence intervals for policy evaluation in adaptive experiments," Proceedings of the National Academy of Sciences, 118(15), e2014602118.

Hill, D. N., H. Nassif, Y. Liu, A. Iyer, and S. Vishwanathan (2017): "An Efficient Bandit Algorithm for Realtime Multivariate Optimization," in Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '17. ACM.

Hirano, K., and J. R. Porter (2023): "Asymptotic Representations for Sequential Decisions, Adaptive Experiments, and Batched Bandits," Discussion paper.

Kasy, M., and A. Sautmann (2021): "Adaptive Treatment Assignment in Experiments for Policy Choice," Econometrica, 89(1), 113–132.

Lei, H., Y. Lu, A. Tewari, and S. A. Murphy (2022): "An Actor-Critic Contextual Bandit Algorithm for Personalized Mobile Health Interventions," .

Offer-Westort, M., A. Coppock, and D. P. Green (2021): "Adaptive Experimental Design: Prospects and Applications in Political Science," American Journal of Political Science, 65(4), 826–844.

Rafferty, A., H. Ying, and J. Williams (2019): "Statistical Consequences of using Multi-armed Bandits to Conduct Adaptive Educational Experiments," Journal of Educational Data Mining, 11(1), 47–79.

Schulz, E., N. T. Franklin, and S. J. Gershman (2020): "Finding structure in multi-armed bandits," Cognitive Psychology, 119, 101261.

Scott, S. L. (2010): "A modern Bayesian look at the multi-armed bandit," Applied Stochastic Models in Business and Industry, 26(6), 639–658.

Scott, S. L. (2015): "Multi-armed bandit experiments in the online service economy," Applied Stochastic Models in Business and Industry, 31, 37–49, Special issue on actual impact and future perspectives on stochastic modelling in business and industry.

Thompson, W. R. (1933): "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples," Biometrika, 25(3-4), 285–294.

References IV

Thompson, W. R. (1935): "On the Theory of Apportionment," _American Journal of Mathematics_, 57(2), 450–456.

Zhang, K., L. Janson, and S. Murphy (2020): "Inference for batched bandits," _Advances in neural information processing systems_, 33, 9818–9829.