Unraveling the Interplay between Carryover Effects and Reward Autocorrelations in Switchback Experiments

Qianglin Wen[1], *Chengchun Shi*[2*], *Yang Ying*[3], *Niansheng Tang*[1], *Hongtu Zhu*[4†]

[1]YNU – Yunnan University
[2]LSE – London School of Economics
[3]FDU – Fudan University

[4]UNC – University of North Carolina at Chapel Hill

[*]Equal contribution   [†]Corresponding author: htzhu@email.unc.edu

A/B testing

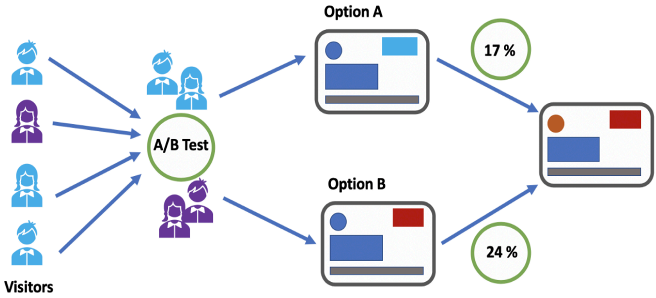


Figure 1: **An example of A/B testing setup. Taken from** towardsdatascience.com.

Average Treatment Effect (ATE)= the averaged difference in expected rewards (denoted by $R_t \in \mathbb{R}$) between the new and old policies over all time steps t:

$$\text{ATE} = \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}^1(R_t) - \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}^0(R_t),$$

Ridesharing

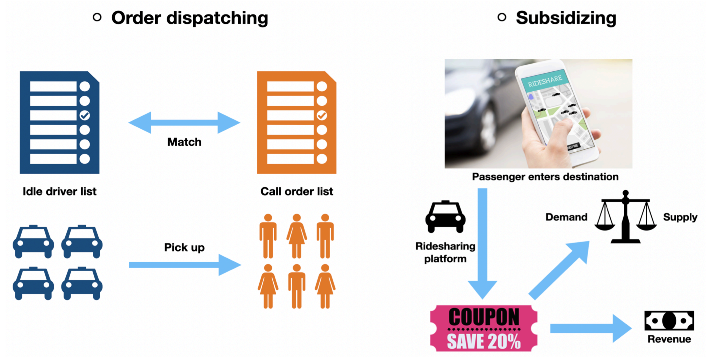


Figure 2: **An illustration of a ridesharing platform. Taken from** callme-spring.github.

Switchback Experiments

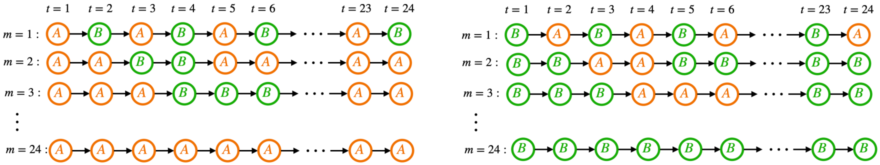


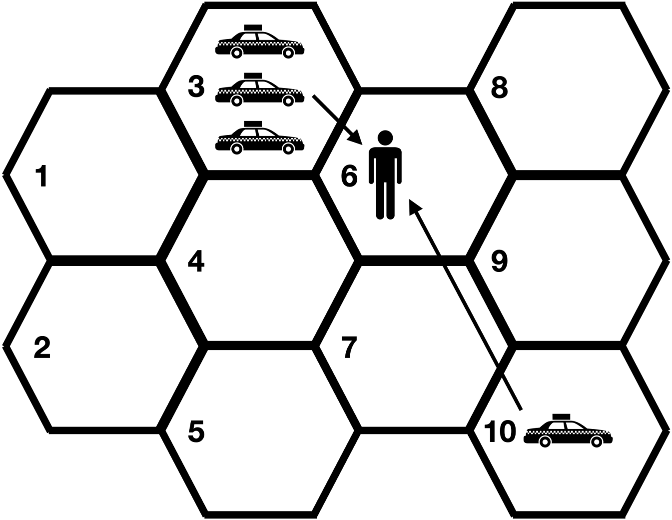
Figure 3: **Orange blocks represent control group assignments, and green blocks represent treatment assignments. The initial policy is control in the left plot and treatment in the right plot.**

Challenges

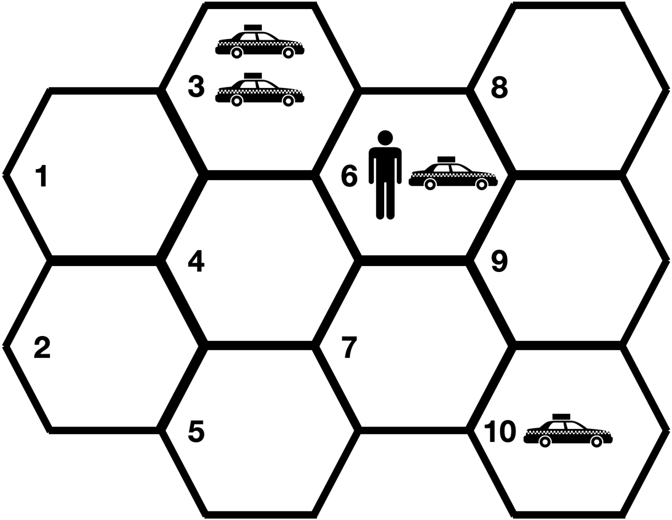
1. **Carryover Effects & Switchback Experiments**
   - Past treatments influence future observations (Li et al., 2024, Figure 2).
   - Carryover biases lead to biased estimates or flawed statistical inference procedures(Bojinov, Simchi-Levi, and Zhao, 2023; Xiong, Chin, and Taylor, 2023; Hu and Wager, 2022; Shi, Wang, et al., 2023).
2. **Auto-correlated Errors**(see the Figure 4)
   - Autoregressive, moving average, and exchangeable covariance structures are widely used in statistical modeling(Williams, 1952; Berenblut and Webb, 1974; Zeger, 1988).

To the best of our knowledge, **no prior work** has systematically examined the effectiveness of different switchback designs in Reinforcement Learning (RL) while accounting **for these two key factors**.

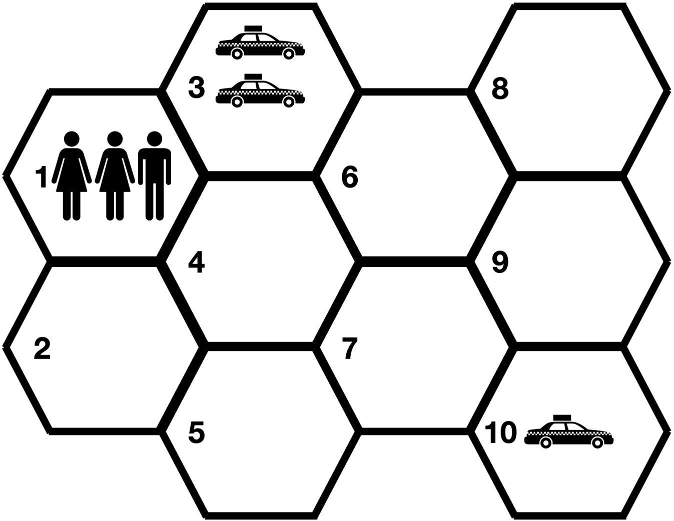
Challenge I: Carryover Effects



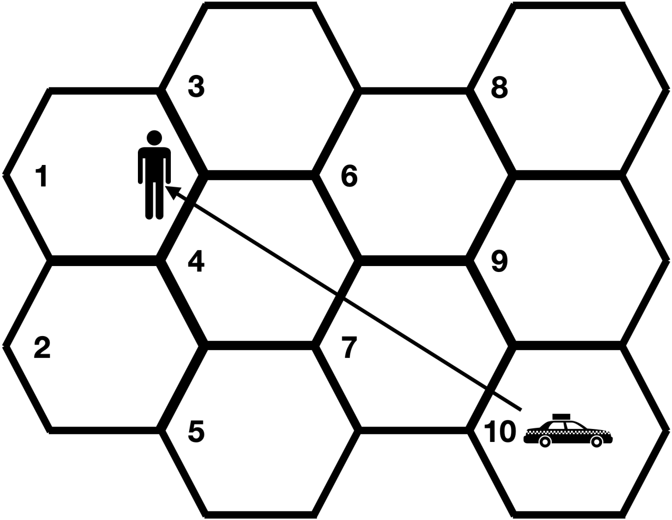
Adopting the Closest Driver Policy



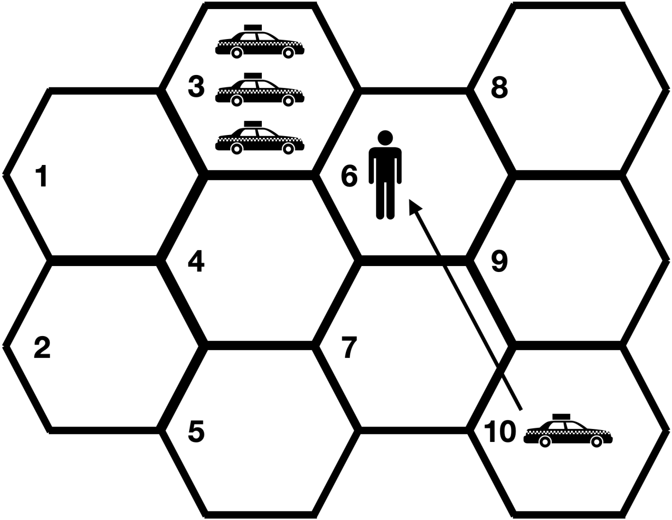
Some Time Later · · ·



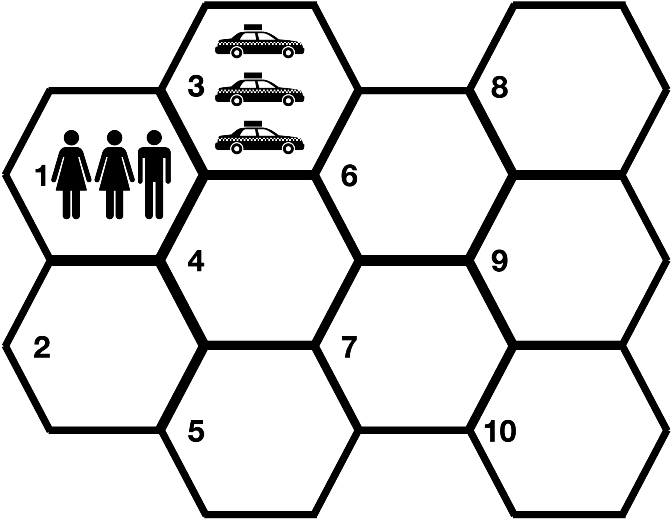
Miss One Order



Consider a Different Action



Able to Match All Orders



Challenge I: Carryover Effects (Cont'd)

$$\textit{past treatments} \rightarrow \textit{distribution of drivers} \rightarrow$$

$$\textit{future outcomes}$$

Challenge II: Real-data based autocorrelations

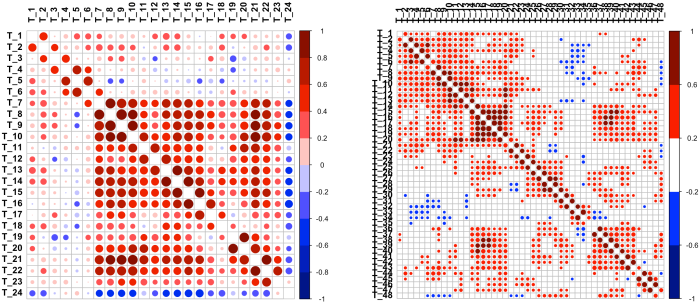


Figure 4: **The estimated correlation coefficients between pairs of fitted reward residuals, based on two datasets provided by a ridesharing company. Most residual pairs are non-negatively correlated, with a large proportion exhibiting positive correlation. The diagonal components have been omitted to enhance clarity.**

Our Contributions

The analysis unravels the interplay between carryover effects and reward autocorrelations in determining the optimal switchback experiments. In particular, **when the carryover effect is weak**, we show that:

- **With predominantly positively correlated reward errors**, the precision of the ATE estimator tends to improve with more frequent alternations between policies.
- **With predominantly negatively correlated reward errors**, the precision of the ATE estimator tends to improve with less frequent alternations between policies.
- **With predominantly uncorrelated reward errors**, all designs become asymptotically equivalent in theory. Our numerical studies indicate that the Alternating-Day (AD, i.e., $m = T$) design generally exhibits superior performance in finite samples.

Additionally, **when the carryover effect is large**, AD or Switchback designs with less frequent switches tend to perform the best.

Finally, these findings are **estimator-agnostic**, i.e. they apply to most RL estimators.

MDP with autocorrelated errors

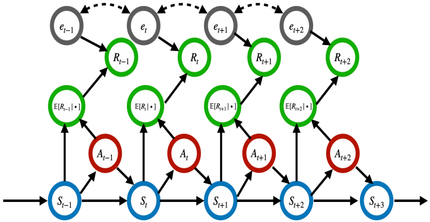


Figure 5: **Visualization of our Markov Decision Process (MDP) with autocorrelated reward errors. The solid lines represent the causal relationships. The dash lines imply that the reward errors are potentially correlated.**

Theory: Main Theorem

Notations:

- n is the number of experimental days, T is the number of time intervals, and $R_{\max}$ bounds the absolute rewards: $\max_t |R_t| \le R_{\max}$.
- $\sigma_e(t_1, t_2)$ denotes the covariance between reward errors e_{t_1} and e_{t_2}.
- δ measures the impact of the new policy on state transition functions $p_t(s'|a,s)$, where $s, s' \in \mathbb{R}^d$ and $a \in \{0,1\}$. Specifically, $\delta = \max_{s,t} \sum_{s'} |p_t(s'|1,s) - p_t(s'|0,s)|$.

The Excess Mean Square Errors (MSEs) Theorem:

- Under the certain conditions: bounded rewards (i.e. $\max_t |R_t| \le R_{\max}$), estimators, states and transition functions, Non-singular covariance matrix, sieve basis functions, nuisance functions, the difference in the MSE of the ATE estimator (i.e. OLS, LSTD, DRL) between **the AD design** and **an m-switchback design** (where each switch duration equals m) **is lower bounded by**

$$\underbrace{\frac{16}{nT^2} \sum_{\substack{k_2-k_1=1,3,5,\ldots \\ 0\le k_1<k_2<T/m}} \sum_{l_1,l_2=1}^{m} \sigma_e(l_1+k_1m, l_2+k_2m)}_{\text{Autocorrelated term}} - \underbrace{O\left(\frac{\delta R_{\max}^2}{n}\right)}_{\text{Carryover effects term}} - \underbrace{o(n^{-1})}_{\text{Estimator-dependent reminder term}},$$

for some constant $c > 0$.

Practical workflow

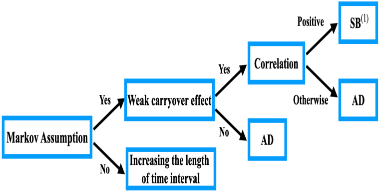


Figure 6: The proposed workflow guideline.

To summary: Two key factors that influence the efficiency of Switchback experiments are: **the autocorrelation structure** and **the magnitude of the carryover effect**.

Thank you for your attention!

Berenblut, II and GI Webb (1974). “Experimental design in the presence of autocorrelated errors”. In: *Biometrika* 61.3, pp. 427–437.

Bojinov, Iavor, David Simchi-Levi, and Jinglong Zhao (2023). “Design and analysis of switchback experiments”. In: *Management Science* 69.7, pp. 3759–3777.

Grenander, Ulf (1981). *Abstract inference*. Wiley Series, New York.

Hu, Yuchen and Stefan Wager (2022). “Switchback experiments under geometric mixing”. In: *arXiv preprint arXiv:2209.00197*.

Kallus, Nathan and Masatoshi Uehara (2020). “Double reinforcement learning for efficient off-policy evaluation in markov decision processes”. In: *The Journal of Machine Learning Research* 21.1, pp. 6742–6804.

- Kallus, Nathan and Masatoshi Uehara (2022). "Efficiently breaking the curse of horizon in off-policy evaluation with double reinforcement learning". In: *Operations Research* 70.6, pp. 3282–3302.
- Li, Ting et al. (2024). "Evaluating Dynamic Conditional Quantile Treatment Effects with Applications in Ridesharing". In: *Journal of the American Statistical Association* just-accepted, pp. 1–26.
- Luo, Shikai et al. (2024). "Policy evaluation for temporal and/or spatial dependent experiments". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology*, pp. 1–27.
- Qin, Zhiwei Tony, Hongtu Zhu, and Jieping Ye (2022). "Reinforcement learning for ridesharing: An extended survey". In: *Transportation Research Part C: Emerging Technologies* 144, p. 103852.
- Shi, Chengchun (2025). "Statistical inference in reinforcement learning: A selective survey". In: *arXiv preprint arXiv:2502.16195*.

- Shi, Chengchun, Rui Song, et al. (2021). "Statistical inference for high-dimensional models via recursive online-score estimation". In: *Journal of the American Statistical Association* 116.535, pp. 1307–1318.
- Shi, Chengchun, Xiaoyu Wang, et al. (2023). "Dynamic causal effects evaluation in a/b testing with a reinforcement learning framework". In: *Journal of the American Statistical Association* 118.543, pp. 2059–2071.
- Uehara, Masatoshi, Chengchun Shi, and Nathan Kallus (2022). "A review of off-policy evaluation in reinforcement learning". In: *arXiv preprint arXiv:2212.06355.*
- Williams, RM (1952). "Experimental designs for serially correlated observations". In: *Biometrika* 39.1/2, pp. 151–167.

Xiong, Ruoxuan, Alex Chin, and Sean J Taylor (2023). "Data-Driven Switchback Designs: Theoretical Tradeoffs and Empirical Calibration". In: *Available at SSRN.*

Zeger, Scott L (1988). "A regression model for time series of counts". In: *Biometrika* 75.4, pp. 621–629.