# Efficient Online Estimation of Causal Effects by Deciding What to Observe

**Anonymous Author(s)**
Affiliation
Address
email

## Abstract

Researchers often face data fusion problems, where multiple data sources are available, each capturing a distinct subset of variables. While problem formulations typically take the data as given, in practice, data acquisition can be an ongoing process. In this paper, we introduce the problem of deciding, at each time, which data source to sample from. Our goal is to estimate a given functional of the parameters of a probabilistic model as efficiently as possible. We propose *online moment selection* (OMS), a framework in which structural assumptions are encoded as moment conditions. The optimal action at each step depends, in part, on the very moments that identify the functional of interest. Our algorithms balance exploration with choosing the best action as suggested by estimated moments. We propose two selection strategies: (1) explore-then-commit (OMS-ETC) and (2) explore-then-greedy (OMS-ETG), proving that both achieve zero asymptotic regret as assessed by MSE. We instantiate our setup for average treatment effect estimation, where structural assumptions are given by a causal graph and data sources include subsets of mediators, confounders, and instrumental variables.

## 1 Introduction

Statistical and causal modeling typically proceed from the assumption that we already know which variables are (and are not) observed. However, this perspective fails to address the difficult data collection decisions that precede such modeling efforts: Doctors must decide which tests to run. Survey designers must select a slate of questions to ask. Companies must choose which among many relevant datasets to purchase. Whether or not we model these decisions, they pervade practice and influence both what questions we can ask and how accurately we can answer them.

One might ask, *why not collect everything?* The answers are two-fold: First, data acquisition can be expensive. In a medical setting, blood tests can cost anywhere from tens to thousands of dollars. Running every test for every patient is infeasible. Likewise, space on surveys is limited, and asking every conceivable question of every respondent is infeasible. Second, in many settings, we lack complete control over the set of variables observed. Instead, we might have access to multiple data sources, each capturing a different subset of variables. Such *data fusion problems* pervade economic modeling and public health, and present interesting challenges: (i) estimating a population parameter of interest requires intelligently combining data from multiple sources; (ii) data collection is often iterative, with tentative conclusions at each stage informing choices about what data to collect next.

In this paper, we formalize the sequential problem of deciding, at each time, which data source to query (i.e., *what to observe*) in order to efficiently estimate a target parameter. We propose online moment selection (OMS), a framework that applies generalized method of moments (GMM) [15] both to estimate the parameter and to decide which data sources to query. This framework can be applied to estimate any statistical parameter that can be identified by a set of moment conditions—e.g.,

Submitted to 35th Conference on Neural Information Processing Systems (NeurIPS 2021). Do not distribute.

in maximum likelihood estimation (MLE) and estimating average treatment effects (ATEs) using instrumental variables (IVs), backdoor adjustment sets, mediators, or other identification strategies involving structural causal models.

Our strategy requires only that the agent has sufficient structural knowledge to formulate the set of moment conditions and that each moment can be estimated using the variables returned by at least one of the data sources. Interestingly, the optimal decisions, leading to the lowest mean squared error (MSE) estimates, depend on the (unknown) model parameters. This motivates our adaptive strategy: as we collect more data, we better estimate the underlying parameters, thus informing our decisions about which data sources to query in the future.

We first address the setting where the cost per instance is equal across data sources (Sec. 4). First, we show that any fixed policy that differs from the oracle suffers constant asymptotic regret, as assessed by MSE. Thus, we instantiate two adaptive strategies that choose data sources based on the *estimated* asymptotic variance of the target parameter: explore-then-commit (OMS-ETC) and explore-then-greedy (OMS-ETG). Under OMS-ETC, we use some fraction of the sample budget to explore randomly, using the collected data to estimate the model parameters. We then collect the remaining samples according to the ratio that the current estimated model suggests will minimize the asymptotic variance of the target parameter (exploitation). In OMS-ETG, we continue to update our parameter estimates after every step as we collect new data. We prove that both policies achieve zero asymptotic regret. To overcome the non-i.i.d. nature of the sample moments, we draw upon martingale theory. To derive zero asymptotic regret, we show uniform concentration of sample moments and a finite-sample concentration inequality for the GMM estimator with dependent data. Next, we generalize our setup to handle heterogeneous costs over the data sources (Sec. 5). We adapt OMS-ETC and OMS-ETG to this setting and prove that they still have zero asymptotic regret.

Finally, we validate our findings experimentally (Sec. 6). Motivated by ATE estimation in causal models encoded as directed acyclic graphs (DAGs), we generate synthetic data from a variety of causal graphs and show that the regret of our proposed methods converges to zero. Furthermore, we see that despite being asymptotically equivalent, OMS-ETG outperforms OMS-ETC in finite samples. Finally, we demonstrate the effectiveness of our methods on two semi-synthetic datasets: the Infant Health Development Program (IHDP) dataset [17] and a Vietnam era draft lottery dataset [2].

## 2 Related Work

Many works attempt to identify and estimate causal effects from multiple datasets. [4, 19] study the problem of combining multiple heterogeneous datasets and propose methods for dealing with various biases. [26] study causal identification when joint and interventional distributions involving different sets of variables are available. [35] provide a search-based algorithm for causal identification from multiple datasets, each covering a subset of variables. [11] introduce efficient estimators of the ATE that efficiently combine two datasets, one where confounders are observed (enabling the backdoor adjustment) and another where mediators are observed (enabling the frontdoor adjustment).

Another related line of work addresses finding optimal adjustment sets for covariate adjustment [16, 29, 39, 32]. While these works take for granted the collection of available datasets, we focus on the problem of deciding which data to collect. Our work shares motivation with active learning, where the learner strategically chooses which (unlabeled) samples to label in order to learn most efficiently [31, 24]. [7] design algorithms for actively collecting samples in a manner that minimizes the learner's variance. In settings where there is a cost associated with collecting each feature, active feature acquisition methods incrementally query feature values to improve a predictive model [20, 30, 18]. [41] propose an active learning criterion to find the most informative questions to ask each respondent in a survey. In the context of causal inference, [10] demonstrate that for ATE estimation, actively deconfounding data can improve sample complexity. [38] propose strategies for acquiring missing confounders to efficiently estimate the ATE.

Others have studied moment selection and IV selection from batch data. [1] introduce consistent moment selection procedures for the GMM setting with some incorrect moments. [6] propose an information-based lasso method for excluding invalid or redundant moment conditions. [36] propose a variable selection framework that uses lasso regression to decide which covariates to include. [12] propose four statistical criteria—including estimation efficiency and non-redundancy—for selecting among a set of candidate IVs. [8] develop an IV selection criteria based on asymptotic MSE and

develop it for the GMM and generalized empirical likelihood estimators. By contrast, we are interested in selecting the data sources for the moments in an online setting.

Previous works address learning from adaptively collected data. [23] propose methods for adaptive experimental design, where at each step, the experimenter must decide the treatment probability using past data in order to efficiently estimate the ATE. [21, 22] propose a doubly-robust estimator for off-policy evaluation with dependent samples. [40] provide regret bounds for learning an optimal policy using adaptively collected data, where the probability of selecting an action is a function of past data. [42, 43] study statistical inference for OLS and M-estimation with non-i.i.d. bandit data. While these settings are different from ours, some of the theoretical tools (e.g., martingale asymptotics and uniform martingale concentration bounds) are similar.

## 3 Preliminaries

In the GMM framework [15, 27], we estimate model parameters by leveraging moment conditions that are satisfied at the true parameters $\theta^*$. A moment condition is a vector $g(X_t, \theta)$ such that $\mathbf{E}_{X_t}[g(X_t, \theta^*)] = 0$. We estimate $\theta^*$ by minimizing the objective $\widehat{Q}_T$:

$$\widehat{\theta} = \arg\min_{\theta \in \Theta} \widehat{Q}_T(\theta), \quad \text{where} \quad \widehat{Q}_T(\theta) = \left[\frac{1}{T}\sum_{t=1}^{T} g_t(\theta)\right]^\top \widehat{W} \left[\frac{1}{T}\sum_{t=1}^{T} g_t(\theta)\right],$$

$\Theta$ is the parameter space, $g_t(\theta) = g(X_t, \theta)$, and $\widehat{W}$ is some (possibly data dependent) positive definite matrix. In this work, we use the two-step GMM estimator, where the first-step estimator $\tilde{\theta}$ is computed by using $\widehat{W} = I$ (identity) and for the second step, we use $\widehat{W} = \left[\frac{1}{T}\sum_{t=1}^{T} g_t(\tilde{\theta}) g_t(\tilde{\theta})^\top\right]^{-1}$.

Let $\mathbf{V}$ be the set of variables of interest and $\psi$ a collection of subsets of $\mathbf{V}$, each corresponding to the specific variables observable via one of the available data sources. Our methods are applicable whenever the target parameter can be identified by a set of moment conditions such that each moment relies on variables simultaneously observable in at least one data source. The *selection vector*, denoted by $s_t \in \{0,1\}^{|\psi|}$, is the binary vector indicating the data source selected at time $t$.

**Assumption 1.** *The agent queries one data source at each step:* $\sum_{i=1}^{|\psi|} s_{t,i} = 1$ *($s_t$ is one-hot).*

We can handle the querying of multiple sources by adding the union of their variables to $\psi$. We write the moment conditions as follows: $g_t(\theta) = m(s_t) \otimes \tilde{g}_t(\theta) \in \mathbf{R}^M$, where $\otimes$ is the Hadamard product, $m : \{0,1\}^{|\psi|} \to \{0,1\}^M$ is a fixed known function that determines which moments get selected, and $\tilde{g}_t(\theta)$ are i.i.d. across $t$. For concreteness, we instantiate our setup with a simple example:

**Example 1** (Instrumental Variable (IV) graph)**.** *Consider a linear IV causal model (shown in Fig. 2a) with instrument $Z$, treatment $X$, outcome $Y$, and the following data-generating process:*

$$X := \alpha Z + \eta, \;\; Y := \beta X + \epsilon, \;\; \epsilon \not\perp\!\!\!\perp \eta, \epsilon \perp\!\!\!\perp Z, \eta \perp\!\!\!\perp Z.$$

*The target parameter is the ATE $\beta$. For $\psi = \{\{Z, X\}, \{Z, Y\}\}$, the moment conditions are*

$$g_t(\theta) = \underbrace{\begin{bmatrix} s_{t,1} \\ s_{t,2} \end{bmatrix}}_{=m(s_t)} \otimes \underbrace{\begin{bmatrix} Z_t(X_t - \alpha Z_t) \\ Z_t(Y_t - \alpha\beta Z_t) \end{bmatrix}}_{=\tilde{g}_t(\theta)} = \begin{bmatrix} s_{t,1} Z_t(X_t - \alpha Z_t) \\ (1 - s_{t,1}) Z_t(Y_t - \alpha\beta Z_t) \end{bmatrix},$$

*where $\theta = [\beta, \alpha]$ and $\{Z_t, X_t, Y_t\}$ are i.i.d.*

For some known function $f_{\text{tar}} : \Theta \to \mathbf{R}$, let $\beta^* := f_{\text{tar}}(\theta^*)$ be the target (e.g., the ATE). In practice, we estimate the target parameter by plugging-in the GMM estimate: $\widehat{\beta} = f_{\text{tar}}(\widehat{\theta})$. Let $H_t$ represent the *history* or the data collected until time $t$ with $H_0 = \{\}$ and space $\mathcal{H}_t$. A data collection policy $\pi$ consists of a sequence of functions $\pi_t : \mathcal{H}_{t-1} \to \{0,1\}^{|\psi|}$ with $s_t = \pi_t(H_{t-1})$. Thus $s_t$ can be dependent on data collected until time $(t-1)$ and so the sample moments $g_t(\theta)$ are *not* i.i.d.

**Definition 1** (Selection ratio)**.** *The selection ratio, denoted by $\kappa_t^{(\pi)}$, is the fraction of samples collected from each data source until time $t$:* $\kappa_t^{(\pi)} = \frac{1}{t}\sum_{i=1}^{t} s_t \in \Delta^{|\psi|-1}$ *(standard simplex).*

$\widehat{\theta}_t^{(\pi)}$ denotes the two-step GMM estimator that uses the data $H_t$. To reduce clutter, we use $\Delta_\psi := \Delta^{|\psi|-1}$, $\text{ctr}(\Delta_\psi) = \left[\frac{1}{|\psi|}, \frac{1}{|\psi|}, \ldots, \frac{1}{|\psi|}\right]$ (center of simplex), and might drop the superscript $\pi$ from $\kappa_t, \widehat{\theta}_t$, and $\widehat{\beta}_t$. $\|.\|$ denotes the spectral and $l_2$ norms for matrices and vectors, respectively.

# 4 Adaptive Data Collection

The central challenge in this work is to make strategic decisions about which data to observe so that we most efficiently estimate the target functional. In this section, we present three policies: (i) *fixed*: query the data sources according to a pre-specified ratio; (ii) OMS-ETC: select uniformly for a specified exploration period, estimate the optimal ratio based on the inferred parameters and thereafter aim for that ratio; and (iii) OMS-ETG: same as OMS-ETC but continue to update parameter (and thus ratio) estimates after the exploration period. For now, we analyze these policies for the case when the cost to query is identical across data sources. By $T \in \mathbb{N}$, we denote the (known) *horizon*, which can be thought of as the agent's data acquisition budget. We defer all proofs to Appendix A.

To begin, we present sufficient conditions for consistency and asymptotic normality of the GMM estimator under adaptively collected data. Using these results we derive the regret of our policies.

**Assumption 2.** *(a) (Identification)* $\forall \theta \neq \theta^*, \mathbf{P}\left(\liminf_{T\to\infty} \bar{Q}_T^{(\pi)}(\theta) > 0\right) = 1$*, where* $\bar{Q}_T^{(\pi)}(\theta) = (g_T^*)^\top W g_T^*$*, and* $g_T^* = \left[\frac{1}{T}\sum_{t=1}^T m(s_t) \otimes \mathbf{E}\left[\tilde{g}_t(\theta)\right]\right]$*; (b)* $\Theta$ *is a compact subset of* $\mathbf{R}^D$*; (c)* $\forall \theta \in \Theta,\ \tilde{g}_i(\theta)$ *is twice continuously differentiable (c.d.),* $\mathbf{E}[\tilde{g}_i(\theta)]$ *is continuous; and (d)* $f_{tar}$ *is c.d. at* $\theta^*$*.*

By Assumption 2(a), the GMM objective is uniquely minimized at $\theta^*$. Sufficient conditions for this when $M = D$ (just-identified case) are (i) each moment is collected enough times: $\forall j,\ \mathbf{P}\left(\liminf_{T\to\infty} \frac{1}{T}\sum_{t=1}^T m_j(s_t) \neq 0\right) = 1$; and (ii) $\forall \theta \neq \theta^*,\ W\mathbf{E}\left[\tilde{g}_t(\theta)\right] \neq 0$.

**Condition 1** (ULLN)**.** *Let* $a_i(\theta) := a(X_i;\theta) \in \mathbf{R}$ *be a function with* $X_i$ *sampled i.i.d.* $a_i(\theta)$ *satisfies the ULLN condition if (i)* $\forall \theta,\ \mathbf{E}\left[a_i(\theta)^2\right] < \infty$*; (ii)* $\forall \theta, |a_i(\theta)| \leq A(X_i)$*; and (iii)* $\mathbf{E}[A(X_i)] < \infty$*.*

**Proposition 1** (Consistency)**.** *Suppose that (i) Assumption 2 holds, and (ii)* $\forall j \in [M],\ \tilde{g}_{i,j}(\theta)$ *satisfies Condition 1. Then, for any policy* $\pi$*,* $\widehat{\theta}_T^{(\pi)} \xrightarrow[T\to\infty]{p} \theta^*$*.*

**Proposition 2** (Asymptotic normality)**.** *Suppose that (i)* $\widehat{\theta}_T^{(\pi)} \xrightarrow{p} \theta^*$*; (ii)* $\forall (i,j) \in [M] \times [D],\ \left[\frac{\partial \tilde{g}_t}{\partial \theta}(\theta)\right]_{i,j}$ *satisfies Condition 1; (iii)* $\forall (i,j) \in [M]^2,\ \left[\tilde{g}_t(\theta)\tilde{g}_t(\theta)^\top\right]_{i,j}$ *satisfies Condition 1; (iv)* $\exists \delta > 0$ *s.t.* $\mathbf{E}\left[\|\tilde{g}_i(\theta^*)\|^{2+\delta}\right] < \infty$*, and (v) (Selection ratio convergence)* $\kappa_T^{(\pi)} \xrightarrow{p} k$ *for some constant* $k \in \Delta_\psi$*. Then* $\widehat{\theta}_T$ *is asymptotically normal:*

$$\sqrt{T}(\widehat{\theta}_T - \theta^*) \xrightarrow{d} \mathcal{N}\left(0, \Sigma(\theta^*, k)\right),$$

*where* $\Sigma(\theta^*, k)$ *is a constant matrix that depends only* $\theta^*$ *and* $k$ *(see Appendix A.2 for the complete expression). By the Delta method,* $\widehat{\beta}_T$ *is also asymptotically normal:*

$$\sqrt{T}(\widehat{\beta}_T - \beta^*) \xrightarrow{d} \mathcal{N}\left(0, V(\theta^*, k)\right), \text{ where } V(\theta^*, k) = \nabla_\theta f_{tar}(\theta^*)^\top [\Sigma(\theta^*, k)] \nabla_\theta f_{tar}(\theta^*).$$

Proposition 2 shows that for a policy under which the selection ratio $\kappa_T$ converges in probability to a constant, the GMM estimator $\widehat{\theta}_T$ will be asymptotically normal.

A fixed (or non-adaptive) policy $\pi_k$ has $\kappa_T = k$ for some constant $k$. Here, the collection decisions do not depend on the data and each data source is queried a fixed fraction of the time. By Prop. 2, $\widehat{\theta}_T^{(\pi_k)}$ will be asymptotically normal. The *oracle policy*, denoted by $\pi^*$, is the fixed policy with the lowest asymptotic variance. Thus for $\pi^*$, $\kappa_T^{(\pi^*)} = \kappa^*$, where $\kappa^* = \arg\min_\kappa V(\theta^*, \kappa)$. We call $\kappa^*$ the *oracle selection ratio*. For the oracle policy, we have $\sqrt{T}(\widehat{\beta}_T^{(\pi^*)} - \beta) \xrightarrow{d} \mathcal{N}\left(0, V(\theta^*, \kappa^*)\right)$.

**Definition 2** (Asymptotic regret)**.** *The asymptotic regret captures how close the scaled asymptotic error of a given policy is to the oracle policy. We define the asymptotic regret of a policy* $\pi$ *as*

$$R_\infty(\pi) = \textit{AMSE}\left(\sqrt{T}\left(\widehat{\beta}_T^{(\pi)} - \beta^*\right)\right) - V(\theta^*, \kappa^*), \tag{1}$$

*where AMSE is the asymptotic MSE (i.e., the MSE of the limiting distribution).*

**Input:** $T \in \mathbb{N}$, $e \in (0,1)$

Collect $Te$ samples s.t. $\kappa_{Te} = \text{ctr}\,(\Delta_\psi)$;

$\widehat{\theta}_{Te} = \text{GMM}(H_{Te})$;

$\widehat{k} = \arg\min_{\kappa \in \Delta_\psi} V(\widehat{\theta}_{Te}, \kappa)$;

Collect remaining samples such that $\kappa_T = \text{proj}(\widehat{k}, \tilde{\Delta})$;

$\widehat{\theta}_T = \text{GMM}(H_T)$;

**Output:** $f_{\text{tar}}\left(\widehat{\theta}_T\right)$

(a) The Explore-then-Commit policy.

**Input:** $T \in \mathbb{N}$, $s \in (0,1)$

$\widehat{k} = \text{ctr}\,(\Delta_\psi)$;

$J = \frac{1}{s}$;

**for** $j \in [1, 2, \ldots, J]$ **do**

  $t = Tsj$;

  Collect $Ts$ samples s.t. $\kappa_t = \widehat{k}$;

  $\widehat{\theta}_t = \text{GMM}(H_t)$;

  $\widehat{k}_t = \arg\min_{\kappa \in \Delta_\psi} V(\widehat{\theta}_t, \kappa)$;

  $\widehat{k} = \text{proj}(\widehat{k}_t, \tilde{\Delta}_{j+1}(\kappa_t))$;

**end**

$\widehat{\theta}_T = \text{GMM}(H_T)$;

**Output:** $f_{\text{tar}}\left(\widehat{\theta}_T\right)$

(b) The Explore-then-Greedy policy.

Figure 1: Algorithms for OMS-ETC and OMS-ETG.

For any fixed policy $\pi_k$ such that $\kappa_T^{(\pi_k)} = k$ for some constant $k \neq \kappa^*$, we have $R_\infty(\pi_k) = [V(\theta^*, k) - V(\theta^*, \kappa^*)] = \mathcal{O}(1)$. This shows that a fixed policy suffers constant regret. This motivates the design of adaptive policies, where the regret asymptotically converges to zero.

### 4.1 Online Moment Selection via Explore-then-Commit (OMS-ETC)

OMS-ETC is inspired by the ETC strategy in the multi-armed bandits (MAB) literature [25, Ch. 6]. Under OMS-ETC, we first explore by collecting a fixed number of samples for each choice. Then, we use these samples to estimate the oracle selection ratio $\kappa^*$. Finally, we commit to this ratio for the remaining time steps. We denote the ETC policy by $\pi_{\text{ETC}}$ (Fig. 1a).

Importantly, $\pi_{\text{ETC}}$ is characterized by an exploration fraction $e \in (0,1)$. We first collect $Te$ samples by querying each data source equally so that $\kappa_{Te} = \text{ctr}\,(\Delta_\psi)$. We then estimate $\widehat{\theta}_{Te}$ and obtain the plugin estimate of $\kappa^*$ as $\widehat{k} = \arg\min_{\kappa \in \Delta_\psi} V(\widehat{\theta}_{Te}, \kappa)$. The *feasible region* for $\kappa_T$ is the set of values that $\kappa_T$ can take after we have devoted $Te$ samples to exploration and is given by $\tilde{\Delta} = \{e\kappa_{Te} + (1-e)\kappa : \kappa \in \Delta_\psi\}$ (proof in Appendix D). We collect the remaining $T(1-e)$ samples such that $\kappa_T$ is as close to $\widehat{k}$ as possible: $\kappa_T = \text{proj}(\widehat{k}, \tilde{\Delta})$, where $\text{proj}\left(k, \tilde{\Delta}\right) = \min_{k' \in \tilde{\Delta}} \|k - k'\|$.

**Assumption 3.** *$\kappa^*$ uniquely minimizes $V(\theta^*, \kappa)$, i.e., $\forall \kappa \in \Delta_\psi, \kappa \neq \kappa^*,\ V(\theta^*, \kappa) > V(\theta^*, \kappa^*)$.*

**Theorem 1** (Regret of OMS-ETC)**.** *Suppose that (i) conditions (i)-(iv) of Proposition 2 hold and (ii) Assumption 3 holds. For a fixed $e \in (0,1)$, if $\kappa^* \in \tilde{\Delta}$, then the regret converges to zero: $R_\infty(\pi_{ETC}) = 0$. If $\kappa^* \notin \tilde{\Delta}$, then $\pi_{ETC}$ suffers constant regret: $R_\infty(\pi_{ETC}) = \mathcal{O}(1)$. If $e$ depends on $T$ such that $e = o(1)$ and $Te \to \infty$ as $T \to \infty$, then $\forall \theta^* \in \Theta$, we have $R_\infty(\pi_{ETC}) = 0$.*

The main idea in the proof is to show that $\kappa_T \xrightarrow{p} \kappa^*$ and apply Prop. 2. The theorem shows that if we explore for a fixed fraction of the horizon $T$, there would be certain parameter values for which $\kappa^*$ would be outside the feasible region and the regret will not converge to zero. To achieve zero asymptotic regret for every $\theta^* \in \Theta$, we can scale $e$ with $T$ such that the $e$ becomes asymptotically negligible (e.g. $e = \frac{1}{\sqrt{T}}$). With this scaling, the feasible region asymptotically covers the entire simplex, i.e., $\tilde{\Delta} \to \Delta_\psi$ as $T \to \infty$ and that is sufficient to show that $\kappa_T \xrightarrow{p} \kappa^*$.

### 4.2 Online Moment Selection via Explore-then-Greedy (OMS-ETG)

We extend OMS-ETC by periodically updating our estimate of $\kappa^*$ as we collect additional samples instead of committing to a value after exploration. Importantly, the data is collected in batches.

OMS-ETG (Fig. 1b) is characterized by a batch fraction $s \in (0,1)$. The algorithm runs for $J = \frac{1}{s}$ rounds and we collect $b = Ts$ samples in each round. In the first round, we explore and thus

$\kappa_b = \text{ctr}(\Delta_\psi)$. After every round $j \in [J]$, we estimate $\widehat{\theta}_{bj}$ and the oracle selection ratio: $\widehat{k}_{bj} = \arg\min_{\kappa \in \Delta_\psi} V(\widehat{\theta}_{bj}, \kappa)$. The feasible region for round $j+1$ (the set of values of $\kappa_{b(j+1)}$ can take) is $\tilde{\Delta}_{j+1}(\kappa_{bj}) = \left\{ \frac{j\kappa_{bj} + \kappa}{j+1} : \kappa \in \Delta_\psi \right\}$ (proof in Appendix D). In round $(j+1)$, we collect samples such that $\kappa_{b(j+1)}$ is as close to $\widehat{k}_{j+1}$ as possible: $\kappa_{b(j+1)} = \text{proj}\left(\widehat{k}_j, \tilde{\Delta}_{j+1}(\kappa_{bj})\right)$.

Next, we state a finite sample concentration inequality for the two-step GMM estimator under adaptively collected (non-i.i.d.) data and use it to derive the asymptotic regret of OMS-ETG.

**Condition 2** (Concentration condition)**.** *Let $\tilde{a}_i(\theta) := \tilde{a}(X_i; \theta) \in \mathbf{R}$ with $X_i$ sampled i.i.d., $\tilde{a}_*(\theta) = \mathbf{E}[\tilde{a}(X_i; \theta)]$, $u_i(\eta) = \sup_{\theta, \theta' \in \Theta, \|\theta - \theta'\| \le \eta} |\tilde{a}_i(\theta) - \tilde{a}_i(\theta')|$, and $u_*(\eta) = \mathbf{E}[u_i(\eta)]$. Let $L_1, L_2, \nu_1, \nu_2, \alpha_1, \alpha_2, \eta_0$ be some positive constants. $\tilde{a}(\theta)$ satisfies the concentration condition if (i) $\tilde{a}_*(\theta)$ is $L_1$-Lipschitz, (ii) $\forall \theta \in \Theta$, $[\tilde{a}_i(\theta) - \tilde{a}_*(\theta)]$ is $(\nu_1^2, \alpha_1)$-sub-Exponential, (iii) $u_*(\eta) \le L_2 \eta$, and (iv) $\forall \eta \in (0, \eta_0)$, $[u_i(\eta) - u_*(\eta)]$ is $(\nu_2^2, \alpha_2)$-sub-Exponential.*

**Remark.** *Condition 2(iv) might be hard to check but is satisfied for (i) bounded function classes, i.e., when $\|\tilde{a}_i\|_\infty < A < \infty$; and (ii) linear function classes when the data is sub-Exponential.*

**Lemma 1** (GMM concentration inequality)**.** *Suppose that (i) Assumption 2 holds; (ii) $\forall j$, $\tilde{g}_{i,j}(\theta)$ satisfies Condition 2; (iii) $\forall (j,k)$, $[\tilde{g}_{i,j}(\theta)\tilde{g}_{i,k}(\theta)]$ satisfies Condition 2; there exists an $\eta$-ball $N_\eta(\theta^*)$ and $h, c, L, \delta_0 > 0$ s.t., (iv) (Local curvature) $\forall \theta \in N_\eta(\theta^*)$, $\mathbf{P}\left(\left\|\frac{\partial^2 \bar{Q}}{\partial \theta^2}(\theta)^{-1}\right\| \le h\right) = 1$ ($\bar{Q}$ is defined in Assumption 2); (v) (Strict minimization) $\forall \theta \in N_\eta(\theta^*)$, there is a unique minimizer $\kappa(\theta)$ s.t. $\|\kappa - \kappa(\theta)\| > \delta \implies V(\theta, \kappa) - V(\theta, \kappa(\theta)) > c\delta^2$; and (vi) $\sup_\kappa |V(\theta, \kappa) - V(\theta', \kappa)| \le L\|\theta - \theta'\|$. Then, for $\widehat{k}_t = \arg\min_{\kappa \in \Delta_\psi} V(\widehat{\theta}_t^{(\pi)}, \kappa)$, any policy $\pi$, and $\forall \delta \in (0, \delta_0)$,*

$$\mathbf{P}\left(\left\|\widehat{\theta}_t^{(\pi)} - \theta^*\right\| > \delta\right) < \frac{1}{\delta^{2D}} \exp\left\{-\mathcal{O}\left(t\delta^4\right)\right\}, \mathbf{P}\left(\left\|\widehat{k}_t - \kappa^*\right\| > \delta\right) < \frac{1}{\delta^{4D}} \exp\left\{-\mathcal{O}\left(t\delta^8\right)\right\}.$$

*Better rates for $\widehat{k}_t$ are applicable under additional restrictions on $\theta^*$ (details in Appendix A.4).*

**Theorem 2** (Regret of OMS-ETG)**.** *Suppose that conditions (i)-(iv) of Prop. 2 hold. Let $\tilde{\Delta}(s) = \{s\kappa_b + (1-s)\kappa : \kappa \in \Delta_\psi\}$. For a fixed $s \in (0,1)$, if the oracle selection ratio $\kappa^* \in \tilde{\Delta}(s)$, then the regret converges to zero: $R_\infty(\pi_{ETG}) = 0$. If $\kappa^* \notin \tilde{\Delta}(s)$, then $R_\infty(\pi_{ETG}) = \mathcal{O}(1)$. Now also suppose that the conditions for Lemma 1 hold. If $s$ depends on $T$ s.t. $s = CT^{\eta - \frac{1}{4}}$ for some constant $C$ and any $\eta \in \left(0, \frac{1}{4}\right)$, then $\forall \theta^* \in \Theta$, the regret converges to zero: $R_\infty(\pi_{ETG}) = 0$.*

In the proof, we show that $\kappa_T \xrightarrow{p} \kappa^*$ and apply Prop. 2. To do so, we prove that $\forall j \in [J], \kappa_{bj} = \widehat{k}_{b(j-1)}$ (using Lemma 1) and thus show that $\kappa_T = \widehat{k}_{b(J-1)} \xrightarrow{p} \kappa^*$. Like OMS-ETC, to get zero asymptotic regret for every $\theta^* \in \Theta$, $s$ must depend on $T$. By Thm. 2, this requires $b = Ts \asymp T^{\eta + \frac{3}{4}}$. This constrains how small the batch size can be relative to $T$. However, empirically, we noticed that much smaller batch sizes work well in practice too (including a fully online policy with $s \asymp \frac{1}{T}$).

Both OMS-ETC and OMS-ETG are asymptotically equivalent as they can achieve zero regret for every $\theta^* \in \Theta$. But in our experiments, the regret of OMS-ETG appears to converge faster than OMS-ETC and OMS-ETG outperforms OMS-ETC for small sample sizes. With OMS-ETG, the estimate of $\kappa^*$ keeps improving as more samples are collected instead of being fixed after exploration. This suggests that better estimates of $\kappa^*$ may lead to higher-order reductions in MSE.

**Additional exploration.** MAB algorithms usually require additional exploration to perform well (e.g. $\epsilon$-greedy and upper confidence bound strategies [25][Ch. 7]). However, we empirically notice that additional exploration hurts performance in our setup. We believe that this is because unlike typical bandit setups where pulling one arm does not improve the estimates of another arm, querying any data source improves the estimates of the model parameter $\theta^*$ in our setup.

## 5 Incorporating a Cost Structure

In many real-world settings, the agent has a budget constraint and must pay a different cost to query each data source. We adapt OMS-ETC and OMS-ETG to this setting where a cost structure is associated with the data sources in $\psi$. We prove that they still have zero asymptotic regret for every

$\theta^* \in \Theta$. Let $(\psi_j)_{j\in[|\psi|]}$ be an indexed family. By $B \in \mathbb{N}$, we denote the (known) budget and by $c \in \mathbf{R}^{|\psi|}$, a cost vector where $0 < c_j < \infty$ is the cost of querying data source $\psi_j$. Due to the cost structure, the horizon $T$ is a random variable dependent on $\pi$ with $T = \left\lfloor \frac{B}{\kappa_T^\top c} \right\rfloor$. The setting in Sec. 4 is a special case of this formulation when $\forall i, c_i = 1$ and $B = T$. We defer proofs to Appendix B.

For a fixed policy $\pi_k$, we have $\kappa_T^{(\pi_k)} = k$, for some constant $k \in \Delta_\psi$. By Prop. 2, as $B \to \infty$, we have $\sqrt{B}\left(\widehat{\beta}_T^{(\pi_k)} - \beta^*\right) \xrightarrow{d} \mathcal{N}\left(0, V(\theta^*, k)\left(k^\top c\right)\right)$. Here we scale by $\sqrt{B}$ instead of $\sqrt{T}$ to make comparisons across policies meaningful. The *oracle selection ratio* is now defined as $\kappa^* = \arg\min_{\kappa\in\Delta_\psi}\left[V(\theta^*, \kappa)\left(\kappa^\top c\right)\right]$ and the *asymptotic regret* of policy $\pi$ now is

$$R_\infty(\pi) = \text{AMSE}\left(\sqrt{B}\left(\widehat{\beta}_T^{(\pi)} - \beta^*\right)\right) - V(\theta^*, \kappa^*)\left((\kappa^*)^\top c\right).$$

OMS-ETC-CS (*OMS-ETC with cost structure*) is an adaptation of OMS-ETC for this setting. We use $Be$ budget to explore and estimate $\kappa^*$ by $\widehat{k} = \arg\min_{\kappa\in\Delta_\psi}\left[V(\widehat{\theta}_{T_e}, \kappa)\left(\kappa^\top c\right)\right]$, where $T_e = \left\lfloor \frac{eB}{\kappa_{T_e}^\top c} \right\rfloor$ and $\kappa_{T_e} = \text{ctr}\left(\Delta_\psi\right)$. Exploration strategies that utilize the cost structure can also be used (e.g., evenly dividing the budget across data sources while exploring). With the remaining budget, we collect samples such that $\kappa_T = \text{proj}\left(\widehat{k}, \tilde{\Delta}\right)$, where $\tilde{\Delta}$ is the feasibility region (expression in Appendix D).

**Lemma 2** (Regret of OMS-ETC-CS)**.** *Suppose that the conditions of Theorem 1 hold. For $e = o(1)$ s.t. $Be \to \infty$ as $B \to \infty$, $\forall\theta^* \in \Theta$, we have $R_\infty(\pi_{\textit{ETC-CS}}) = 0$.*

We propose two ways of adapting the OMS-ETG to this setting. In the first variant—OMS-ETG-FS (*OMS-ETG with fixed samples*)—we collect a fixed number of samples in every round. In the second variant—OMS-ETG-FB (*OMS-ETG with fixed budget*)—we spend a fixed fraction of the budget in each round. Let $c_{\max} = \max_{i\in[|\psi|]} c_i$ and $c_{\min} = \min_{i\in[|\psi|]} c_i$. In OMS-ETC-FS, we collect $b = \frac{Bs}{c_{\max}}$ samples in every round except the last one (the last batch can be smaller as we may not have enough budget left for a full batch). The number of rounds $J$ is random since, in every round, depending on what we collect, we use up a different amount of the budget. Like OMS-ETG, after each round, we estimate $\kappa^*$ and greedily collect samples to get as close to it as possible.

**Lemma 3** (Regret of OMS-ETG-FS)**.** *Suppose that the conditions of Theorem 2 hold. For $s \asymp B^{\eta-\frac{1}{4}}$ and any $\eta \in \left(0, \frac{1}{4}\right)$, $\forall\theta^* \in \Theta$, we have $R_\infty\left(\pi_{\textit{ETG-FS}}\right) = 0$.*

In OMS-ETG-FB, we spend $Bs$ budget in each round. Thus the number of rounds $J$ is fixed with $J = \frac{1}{s}$. But the number of samples collected per round is now random. Like OMS-ETG, after each round, we estimate $\kappa^*$ and collect samples to get as close to the estimate as possible.

**Lemma 4** (Regret of OMS-ETG-FB)**.** *Suppose that the conditions of Theorem 2 hold. For $s \asymp B^{\eta-\frac{1}{4}}$ and any $\eta \in \left(0, \frac{1}{4}\right)$, $\forall\theta^* \in \Theta$, we have $R_\infty\left(\pi_{\textit{ETG-FB}}\right) = 0$.*

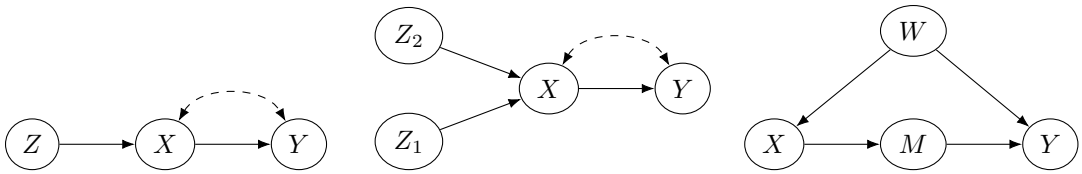
# 6 Experiments

## 6.1 Synthetic data

To begin, we validate our methods on synthetic data generated from known causal graphs (see Appendix C for parameter values and moment conditions used). We first simulate data from a linear IV graph (Fig. 2a) and compare the performance of different polices (Fig. 3a). We use $\psi = \{\{Z, X\}, \{Z, Y\}\}$ and assume that both choices cost the same. We set parameter values such that $\kappa^* \approx [0.36, 0.64]^\top$. We compare policies based on *relative regret (RR)* w.r.t. the oracle policy:

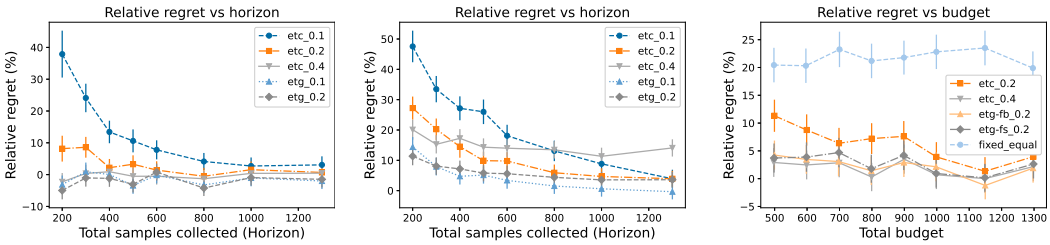
$$\text{Relative regret} = \text{RR}(\pi) = \frac{\text{MSE}^{(\pi)} - \text{MSE}^{(\text{oracle})}}{\text{MSE}^{(\text{oracle})}} \times 100\%.$$

The MSE values are computed across $12,000$ runs. The label *etc_*$\{x\}$ in the plot refers to OMS-ETC with exploration fraction $e = x$ (e.g. *etc_0.1* means $e = 10\%$). Similarly, *etg_*$\{x\}$ refers to OMS-ETG with $s = x$. We see that as the horizon increases, the RR of all policies converges to zero. This supports the claim that both OMS-ETC and OMS-ETG have zero asymptotic regret. However, when the horizon is small ($T = 200$, $T = 300$), both *etc_0.1* and *etc_0.2* perform poorly due to insufficient




(a) Instrumental variable graph. (b) Two IVs graph. (c) Confounder-mediator graph.

Figure 2: Examples of causal models—with treatment $X$ and outcome $Y$—where the ATE can be identified by different data sources returning a different subsets of variables.




(a) IV model. (b) Two IVs model. (c) Confounder-mediator model.

Figure 3: Relative regret (RR) across horizons/budgets for different policies (error bars denote 95% CIs). (a) RR for the IV model (Fig. 2a) with $\psi = \{\{Z, X\}, \{Z, Y\}\}$. (b) RR for the two IVs model (Fig. 2b) with $\psi = \{\{X, Y, Z_1\}, \{X, Y, Z_2\}\}$. (c) RR for the confounder-mediator model (Fig. 2c) with $\psi = \{\{X, Y, W\}, \{X, Y, M\}\}$ and a cost structure.

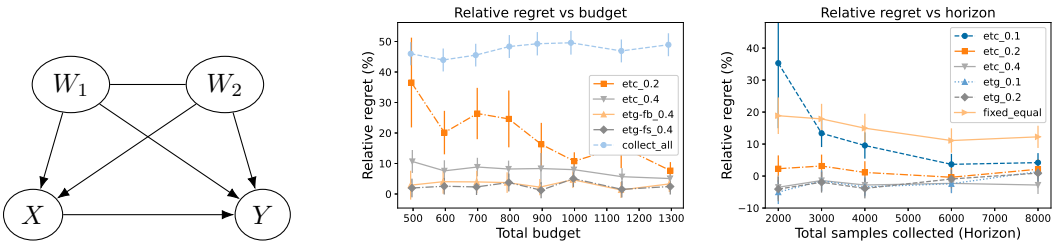
exploration. In contrast, OMS-ETG has close to zero RR even for small horizons which shows that repeatedly improving the parameter estimates can lead to faster convergence of regret.

Next, we simulate data from a linear graph with two IVs (Fig. 2b). Here $\psi = \{\{X, Y, Z_1\}, \{X, Y, Z_2\}\}$ and both choices cost the same. We set the parameters such that $\kappa^* = [0, 1]^\top$ ($\kappa^*$ is on the simplex corner). We compare the RR across various horizons (Fig. 3b). We see the OMS-ETC performs worse than OMS-ETG for small horizons. One difference from the previous case (Fig. 3a) is that *etc_0.4* performs poorly even for large horizons. This is because after using $40\%$ of the samples for exploration, the feasibility region is not large enough to get close to the corner of the simplex. This shows another benefit of the OMS-ETG over OMS-ETC: as long as the batch fraction $s$ is sufficiently small, OMS-ETG has close to zero regret in both cases for all horizons.

Finally, we simulate data from a linear confounder-mediator graph (Fig. 2c). Here, both the backdoor (using $\{X, Y, W\}$) and frontdoor (using $\{X, Y, M\}$) adjustments are applicable [28, Ch. 3.3]. We use $\psi = \{\{X, Y, W\}, \{X, Y, M\}\}$ with cost structure $c = [1.8, 1]^\top$ (confounders $W$ cost more than the mediators $M$). We set the parameters such that $\kappa^* \approx [0.15, 0.85]^\top$. We see similar conclusions as the previous cases. OMS-ETC with low exploration performs poorly but converges for large horizons. Both OMS-ETG variants—OMS-ETG-FS and OMS-ETG-FB—have close to zero RR for all horizons. We see no significant difference between the regret of OMS-ETG-FS and OMS-ETG-FB. *fixed_equal* shows the performance of a fixed policy that collects an equal fraction of both subsets. Its RR does not converge and is substantially higher than the oracle ($\approx 20\%$). This demonstrates that adaptive policies can lead to significant gains in MSE over fixed policies and that our methods remain applicable even with an associated cost structure.

### 6.2 Semi-synthetic data

**IHDP.** Hill [17] constructed a dataset based on the Infant Health and Development Program (IHDP). The data is from a randomized experiment studying the effect of home visits by a trained provider on future cognitive test scores of children (we use samples from Dorie [9]). Following Hill [17], we create an observational dataset by removing a non-random subset of the data. The treatment $X$ is binary. The dataset contains pre-treatment covariates which are measurements on the mother and the child. For simplicity, we only use two covariates: birth weight (continuous) ($W_1$) and whether the mother smoked (binary) ($W_2$). For each sample of the generated semi-synthetic data, $(X, W_1, W_2)$




(a) IHDP data causal graph (b) RR for IHDP data (c) RR for Earnings data

Figure 4: Relative regret (RR) on semi-synthetic data (error bars denote 95% CIs). For both IHDP (b) and Earnings data (c), adaptive policies converge to zero RR but fixed policies (*collect_all*, *fixed_equal*) suffer constant regret. OMS-ETG outperforms OMS-ETC for small horizons.

are sampled uniformly at random from the real data. The outcome $Y$ (continuous) is simulated: $Y = \alpha_1 W_1 + \alpha_2 W_2 + \beta X + \epsilon_y$, where $\epsilon_y \sim \mathcal{N}(0, \sigma_y^2)$, $\alpha_1, \alpha_2, \beta \in \mathbf{R}$, and $\sigma_y^2 \in \mathbf{R}^+$ (see Fig. 4a). Here $\alpha_1, \alpha_2, \beta$, and $\sigma_y$ are model parameters with $\beta$ being the ATE (target paramter).

For this experiment, we use $\psi = \{\{X, Y, W_1\}, \{X, Y, W_2\}, \{X, Y, W_1, W_2\}\}$ with cost structure $c = [1, 3, 3.5]^\top$. Thus, at each step, the agent can collect either or both of the covariates, and each choice has a distinct cost. Setting model parameters such that $\kappa^* \approx [0.59, 0, 0.41]^\top$, we compare performance across policies for various total budgets (Fig. 4b). *collect_all* is a fixed policy that collects $\{X, Y, W_1, W_2\}$ at every step. This policy has higher RR ($\approx 50\%$ higher MSE than the oracle) for all budgets demonstrating that *collecting all covariates for every sample can be sub-optimal*. *etc_0.2* does poorly with a small budget whereas *etc_0.4* has close to zero RR. We note the effectiveness of greedy strategies: both ETG-FB and ETG-FS have close to zero RR for all horizons.

**The Vietnam draft and future earnings.** Angrist [2] computed the effect of veteran status on future earnings from the Vietnam draft lottery data using an IV (Fig. 2a). The binary IV $Z$ indicates whether an individual was eligible for the draft based on a random lottery. The treatment $X$ (binary) indicates whether they actually served. The outcome $Y$ (continuous) represents their future earnings. The IV removes bias caused by certain types of men being more likely to serve. In this dataset, $\{Z, X\}$ and $\{Z, Y\}$ were collected using different data sources, which suits our framework. We construct a semi-synthetic dataset that closely matches the real data [3] so that we know the ground-truth causal effect (for computing the MSE). For each instance, we sample $Z$ uniformly at random from the real data. $X$ is sampled from the empirical distribution $\widehat{\mathbf{P}}(X|Z)$ (with values from [2, Table 2]). We generate $Y$ using $Y = \alpha Z + c_0 + \epsilon_y$, where $\epsilon \sim \mathcal{N}(0, \sigma_y^2)$. $\alpha, c_0$, and $\sigma_y^2$ are estimated by fitting a linear regression model of $Y$ on $Z$. We compare the RR of our policies on this dataset (Fig.4c). The adaptive policies converge to near zero RR as the horizon gets large. OMS-ETC does poorly with low exploration while OMS-ETG policies have significantly lower RR for smaller horizons. By contrast, *fixed_equal* (the fixed policy that samples from both data sources equally) suffers constant regret and has $\approx 20\%$ higher regret than the oracle even for a large horizon. This demonstrates that adaptively collecting data can lead to substantial MSE gains in a real-world setting.

## 7 Conclusion

This paper takes some initial strides towards endogenizing decisions about which variables to solicit into the modeling process. Addressing the problem of deciding, sequentially, which data sources to query in order to efficiently estimate a parameter, we developed the online moment selection (OMS) framework and two instantiations: OMS-ETC and OMS-ETG. We prove that over the entire parameter space, adaptive collection with either method can provide substantial MSE gains. While our study focuses on ATE estimation, our framework is more broadly applicable to any parameters identified by moment conditions. In future work, we hope to apply our framework in more general prediction problems, addressing practical considerations including high-dimensional data and complex model classes (e.g., neural networks). Moreover, in real-world settings, common assumptions like ignorability rarely hold. We hope to extend our framework to overcome issues such as model misspecification, or to overcome biases present in some, but not all data sources.

# References


[1] D. W. Andrews. Consistent moment selection procedures for generalized method of moments estimation. *Econometrica*, 1999.

[2] J. D. Angrist. Lifetime earnings and the vietnam era draft lottery: evidence from social security administrative records. *The American Economic Review*, 1990.

[3] J. D. Angrist. Replication data for: Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records, 2009. URL `https://doi.org/10.7910/DVN/PLF0YL`.

[4] E. Bareinboim and J. Pearl. Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, 2016.

[5] J. Bell. The lindeberg central limit theorem, 2015.

[6] X. Cheng and Z. Liao. Select the valid and relevant moments: An information-based lasso for gmm with many moments. *Journal of Econometrics*, 2015.

[7] D. A. Cohn, Z. Ghahramani, and M. I. Jordan. Active learning with statistical models. *Journal of artificial intelligence research*, 1996.

[8] S. G. Donald, G. W. Imbens, and W. K. Newey. Choosing instrumental variables in conditional moment restriction models. *Journal of Econometrics*, 2009.

[9] V. Dorie. Non-parametrics for causal inference. `https://github.com/vdorie/npci`, 2016.

[10] K. Gan, A. A. Li, Z. C. Lipton, and S. Tayur. Causal inference with selectively-deconfounded data. *arXiv preprint arXiv:2002.11096*, 2020.

[11] S. Gupta, Z. C. Lipton, and D. Childers. Estimating treatment effects with observed confounders and mediators. *arXiv preprint arXiv:2003.11991*, 2020.

[12] A. R. Hall and F. P. Peixe. A consistent method for the selection of relevant instruments. *Econometric reviews*, 2003.

[13] P. Hall and C. C. Heyde. *Martingale limit theory and its application*. Academic press, 1980.

[14] J. D. Hamilton. *Time series analysis*. Princeton university press, 1994.

[15] L. P. Hansen. Large sample properties of generalized method of moments estimators. *Econometrica: Journal of the Econometric Society*, 1982.

[16] L. Henckel, E. Perković, and M. H. Maathuis. Graphical criteria for efficient total effect estimation via adjustment in causal linear models. *arXiv preprint arXiv:1907.02435*, 2019.

[17] J. L. Hill. Bayesian nonparametric modeling for causal inference. *Journal of Computational and Graphical Statistics*, 2011.

[18] S.-J. Huang, M. Xu, M.-K. Xie, M. Sugiyama, G. Niu, and S. Chen. Active feature acquisition with supervised matrix completion. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018.

[19] P. Hünermund and E. Bareinboim. Causal inference and data-fusion in econometrics. *arXiv preprint arXiv:1912.09104*, 2019.

[20] S. Ji and L. Carin. Cost-sensitive feature acquisition and classification. *Pattern Recognition*, 2007.

[21] M. Kato. Adaptive doubly robust estimator and paradox concerning logging policy: Off-policy evaluation from dependent samples, 2021.

[22] M. Kato. Adaptive doubly robust estimator from non-stationary logging policy under a convergence of average probability, 2021.

[23] M. Kato, T. Ishihara, J. Honda, and Y. Narita. Adaptive experimental design for efficient treatment effect estimation, 2020.

[24] P. Kumar and A. Gupta. Active learning query strategies for classification, regression, and clustering: A survey. *Journal of Computer Science and Technology*, 2020.

[25] T. Lattimore and C. Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

[26] S. Lee and E. Bareinboim. Causal effect identifiability under partial-observability. In *International Conference on Machine Learning*. PMLR, 2020.

[27] W. K. Newey and D. McFadden. Large sample estimation and hypothesis testing. *Handbook of econometrics*, 1994.

[28] J. Pearl. *Causality*. Cambridge university press, 2009.

[29] A. Rotnitzky and E. Smucler. Efficient adjustment sets for population average treatment effect estimation in non-parametric causal graphical models. *arXiv preprint arXiv:1912.00306*, 2019.

[30] M. Saar-Tsechansky, P. Melville, and F. Provost. Active feature-value acquisition. *Management Science*, 2009.

[31] B. Settles. Active learning literature survey. *Computer Science Technical Report 1648*, 2009.

[32] E. Smucler, F. Sapienza, and A. Rotnitzky. Efficient adjustment sets in causal graphical models with hidden variables. *arXiv preprint arXiv:2004.10521*, 2020.

[33] T. Tao. *Topics in random matrix theory*. American Mathematical Soc., 2012.

[34] G. Tauchen. Diagnostic testing and evaluation of maximum likelihood models. *Journal of Econometrics*, 1985.

[35] S. Tikka, A. Hyttinen, and J. Karvanen. Causal effect identification from multiple incomplete data sources: A general search-based approach. *arXiv preprint arXiv:1902.01073*, 2019.

[36] O. Urminsky, C. Hansen, and V. Chernozhukov. Using double-lasso regression for principled variable selection. *Available at SSRN 2733374*, 2016.

[37] M. J. Wainwright. *High-dimensional statistics: A non-asymptotic viewpoint*. Cambridge University Press, 2019.

[38] S. Wang, S. E. Yi, S. Joshi, and M. Ghassemi. Confounding feature acquisition for causal effect estimation. In *Machine Learning for Health*. PMLR, 2020.

[39] J. Witte, L. Henckel, M. H. Maathuis, and V. Didelez. On efficient adjustment in causal graphs. *Journal of Machine Learning Research*, 2020.

[40] R. Zhan, Z. Ren, S. Athey, and Z. Zhou. Policy learning with adaptively collected data. *arXiv preprint arXiv:2105.02344*, 2021.

[41] C. Zhang, S. J. Taylor, C. Cobb, J. Sekhon, et al. Active matrix factorization for surveys. *Annals of Applied Statistics*, 2020.

[42] K. W. Zhang, L. Janson, and S. A. Murphy. Inference for batched bandits, 2021.

[43] K. W. Zhang, L. Janson, and S. A. Murphy. Statistical inference with m-estimators on bandit data, 2021.

## Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes]
   (b) Did you describe the limitations of your work? [Yes] We precisely state conditions when our framework is applicable.
   (c) Did you discuss any potential negative societal impacts of your work? [N/A]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes] See Sections 4and 5.
   (b) Did you include complete proofs of all theoretical results? [Yes] Proofs are in the Appendix.
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] Code is provided in the supplementary material.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] See Section 6
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A] Our code uses standard software like Python, numpy, pandas, etc.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] We use two datasets and we cite them in Section 6.
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [Yes] We include a copy of the two datasets we used.
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [No] The datasets are publicly available and we cite the datasets.
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]