

A Model Selection Approach to Interference*

Gustavo Diaz[†]

This version: April 23, 2021

[Link to most recent version](#)

Abstract

In many social science applications, units interfere with each other in substantively interesting ways. Current methods to analyze interference or spillovers from a causal inference perspective assume that the researcher knows the relevant pathways through which interference occurs and can use that information to identify which units are exposed to treatment. In most cases, this assumption is not verifiable, which means researchers must make modeling decisions that are only indirectly informed by theory. Moreover, the discipline currently lacks standards to evaluate the appropriateness of alternative operationalizations. Focusing on the case of converting distance metrics into treatment exposures, this paper suggests approaching the problem as a supervised learning model selection task. Researchers can use the desired performance criterion, algorithm, and resampling strategy to choose the most appropriate model of interference given the data and estimation target. This paper focuses on inverse probability weighting estimators and the lasso family of methods, but the ideas are general enough to fit researchers' theoretical and analytical needs. I illustrate the effectiveness of this approach with simulations and an application to the study of voter registration irregularities in Ghana.

*I thank audiences at the 2019 Latin American Political Methodology Meeting, the 2020 Annual Meeting of the Society for Political Methodology, and the 2021 Rebecca B. Morton Conference on Experimental Political Science for valuable feedback.

[†]PhD Candidate. Department of Political Science. University of Illinois at Urbana-Champaign. E-mail: diazdia2@illinois.edu

1 Introduction

In the causal inference framework, interference or spillovers occur when a unit’s potential outcome depends on the treatment assignment of other units (Cox 1958). This is a violation of the stable unit treatment value assumption (SUTVA) used to justify the implementation of common design-based estimators (Rubin 1990). Some scholarship aims at detecting interference (Aronow 2012) or recovering the ability to identify causal effects in its presence (Sävje, Aronow, and Hudgens 2019). However, in the social sciences, interference can be a phenomenon of interest by itself. For example, research on electoral fraud suggests that the presence of election observers reduces irregularities in treated units, but increases them in nearby localities (Ichino and Schündeln 2012; Asunka et al. 2017).

Detecting and estimating interference in the causal inference framework is an active research agenda (see Halloran and Hudgens 2016; Aronow et al. 2021; Ogburn and VanderWeele 2014 for overviews). Current approaches assume the researcher knows the pathway(s) through which interference occurs. For example, Aronow and Samii (2017) propose a general estimator of the average unit-level causal effect of experiencing different exposure regimes, which requires knowledge of what units are exposed to treatment.¹ Similarly, Hudgens and Halloran (2008) develop estimators for several quantities of interest under the assumption that interference occurs within, but not across, groups or strata. Beyond estimation, Bowers, Fredrickson, and Panagopoulos (2013) introduce a framework to test hypotheses of counterfactual causal effects under a theoretically informed model of interference. As with any application of mathematical models, this approach assumes a correct model specification.² These type of assumptions are not verifiable, but researchers can justify them with research designs meant to capture interference, such as multilevel experiments (Sinclair, McConnell, and Green 2012), saturation designs (Baird et al. 2018), or natural experiments (Keele and Titiumik 2018).

¹I use “exposure to treatment” as a shorthand for “being exposed to a treated unit.”

²See also Toulis and Kao (2013), Athey, Eckles, and Imbens (2018), and Base, Feller, and Toulis (2019) for alternative approaches with similar assumptions.

In many social science applications, these research designs are not feasible. If the known pathways assumption is not met, the resulting estimates may be biased (Aronow et al. 2021). This paper focuses on estimating interference effects in the case in which researchers' domain expertise suggests a pathway for interference, but does not indicate how units connect to each other. For example, we may know units' geographic location, but not how far away a treated unit interferes with the outcome of others. To make a statement of what units are exposed to treatment, we must convert observed pairwise distances into exposures. This conversion is a decision that follows from theory, but is not directly informed by it, and the discipline currently lacks standards to determine whether a specific conversion is more appropriate than alternative operationalizations.

I argue that we can approach this challenge as a model selection task in supervised learning. We can think of alternative conversions from pairwise distances to exposures as a set of alternative models, or variables within a model, among which the researcher can select the one that best satisfies a performance criterion, given an algorithm and resampling strategy of choice. While this paper focuses primarily on inverse probability weighted difference in means (Aronow and Middleton 2013; Aronow and Samii 2017) and the lasso-family methods for model selection (see Ratkovic and Tingley 2017 for an overview), I propose a general protocol that researchers can adapt to fit their theoretical and estimation needs.

The choice of algorithm depends primarily on two other decisions. First, the researcher must decide whether the purpose is to estimate a categorical (e.g. Aronow and Samii 2017) or a marginal exposure effect (e.g. Hudgens and Halloran 2008). Second, whether the task is to estimate a single global effect or multiple local effects at different distance ranges. I illustrate how to navigate these decisions, as well as the protocol's effectiveness, with simulations based on hypothetical experiments on a road network in Ghana (based on Bowers et al. 2018) and a reproduction of the analysis of an experiment designed to capture the spillover effect of election observers on voter registration irregularities in the same country (Ichino and

Schündeln 2012).

The main contribution of this paper is to develop standards to assist researchers in making informed decisions while modeling interference in experiments and observational studies that use experimental logic when the known pathways assumption is not satisfied by design. The model selection approach described here is helpful in identifying the most appropriate model among alternatives stemming from one or more theoretically relevant pathways. It extends recent work on analyzing interference when the assumption of known pathways is not sustainable. For example, Egami (2020) develops sensitivity analysis for interference in the presence of unobserved networks and Sävje (2019) replaces the assumption of known exposure in Aronow and Samii (2017) with the weaker condition of sufficiently controlled (and hence estimable) specification errors. It also extends on recent literature using statistical learning to improve causal inference (see Blakely et al. 2019 for a general treatment), especially in the context of using supervised learning model selection algorithms prior to estimation to increase efficiency (e.g. Belloni, Chernozhukov, and Hansen 2013; Bloniarz et al. 2016).

2 Causal Inference Approaches to Interference

2.1 Setting

The technical aspects of this section follow Aronow et al. (2021) closely. Consider an experiment conducted in a sample of N units indexed by $i = \{1, 2, \dots, N\}$. Let Z denote a transposed treatment assignment vector so that $Z = \{Z_1, \dots, Z_N\}^\top$. For simplicity, assume a binary treatment variable, which implies unit i 's possible treatment status is $Z_i = \{0, 1\}$. The logic in this paper still applies to multiple and continuous treatments. Based on the experimental design, the probability of treatment assignment $Pr(Z = z)$ is known for all possible treatment vectors $z \in \{0, 1\}^N$.

Assuming no interference, unit i 's potential outcome $Y_i(z)$ depends on its treatment assignment

only, which means its observed outcome is $Y_i = z_i Y_i(1) + (1 - z_i) Y_i(0)$, where z_i is the observed treatment status of unit i . To make the absence of interference more explicit, let z_{-i} and z'_{-i} denote two different treatment vectors that exclude unit i 's treatment status. If potential outcomes depend on treatment assignment only, we can claim that $Y_i(z_i, z_{-i}) = Y_i(z_i, z'_{-i})$.

Interference implies that $Y_i(z_i, z_{-i}) \neq Y_i(z_i, z'_{-i})$, so we need to account for treatment status of other units to fully characterize unit i 's potential outcome. Accounting for every unit's treatment status leads unit i to have 2^N possible potential outcomes. Since this is intractable in most applications, researchers need to impose some structure. When manipulation is feasible, researchers can implement saturation (Baird et al. 2018) or multilevel (Sinclair, McConnell, and Green 2012) designs so that both treatment assignment and exposure to treatment are known. Natural experiments can also provide such a setting (e.g. Keele and Titiunik 2018). When manipulation is not feasible, structure can come from theoretically informed assumptions. For example, we could assume that interference is more likely among legislators with similar ideology (Coppock 2014), or that the presence of election observers affects electoral irregularities within a given distance radius (Ichino and Schündeln 2012).

Different assumptions inform different empirical strategies. If the researcher observes how units connect with each other in a network and assumes that said network is sufficient to capture all relevant information about how treatments propagate, then they can estimate interference as the contrast across different exposures. Alternatively, if the researcher is willing to assume that interference happens within observed groups only, they can estimate the marginal or saturation effect of being exposed to treatment within the group.

2.2 Interference as contrast across exposures

Consider the case of estimating interference in randomized experiments where the researcher knows how units connect with each other. Aronow and Samii (2017) introduce the concept of exposure mapping to characterize all the possible treatment-exposure combinations that may

emerge depending on what the researcher assumes about interference. Under this setting, $Y_i(d_k)$ denotes the potential outcome of unit i under exposure k .³ With this, we can express the unit-level average exposure effect as:

$$\tau(d_k, d_{k'}) = \frac{1}{N} \sum_{i=1}^N Y_i(d_k) - \frac{1}{N} \sum_{i=1}^N Y_i(d_{k'}) \quad (1)$$

Which is the difference in means between the potential outcomes under exposures k and k' , with k' denoting any exposure different from k . For example, if we want to express the effect of being exposed to treatment among control units, we would write:

$$\tau(d_{01}, d_{00}) = \frac{1}{N} \sum_{i=1}^N Y_i(d_{01}) - \frac{1}{N} \sum_{i=1}^N Y_i(d_{00}) \quad (2)$$

With $Y_i(d_{01})$ denoting the potential outcome of unit i when assigned to control and exposed to treatment, and $Y_i(d_{00})$ denoting the potential outcome of i when assigned to control and not exposed to treatment.

Per the fundamental problem of causal inference (Holland 1986), we cannot observe unit i 's potential outcome in more than one exposure at a time. Moreover, the probability of experiencing different exposures varies across units. Aronow and Samii (2017) propose a Horvitz-Thompson inverse probability estimator of $\tau(d_k, d_{k'})$ that accounts for both issues:

$$\widehat{\tau}_{HT}(d_k, d_{k'}) = \frac{1}{N} \left[\sum_{i=1}^N \mathbf{I}(D_i = d_k) \frac{Y_i}{\pi_i(d_k)} - \sum_{i=1}^N \mathbf{I}(D_i = d_{k'}) \frac{Y_i}{\pi_i(d_{k'})} \right] \quad (3)$$

Which is the difference in inverse probability weighted means between two exposures, with $\mathbf{I}(D_i = d_k)$ denoting the units for which we observe exposure k , and $\pi_i(d_k)$ denoting the probability of unit i experiencing exposure k . The notation is analogous for k' . For a randomized experiment, $\pi_i(d_k)$ is the expected proportion of treatment assignments that

³See Aronow and Samii (2017) for an extensive formalization of this approach.

induce unit i to experience exposure k , which can be computed exactly in relatively small samples or approximated through simulation otherwise.

See Aronow and Samii (2017) for a subsequent discussion of variance estimation. For the purposes of this paper, the important part is that the estimator $\widehat{\tau}_{HT}(d_k, d_{k'})$ is unbiased under the known pathways assumption. Therefore, when this assumption is not satisfied by design, any operationalization of interference, even if informed by theory, may inadvertently introduce bias. From a model selection perspective, the task to find an operationalization that minimizes this bias.

2.3 Stratified interference

Another way to impose structure under interference is to assume that it occurs only within observed groups or strata. For example, Get Out the Vote campaigns may affect the voting behavior of untreated individuals living in treated households (Nickerson 2008), or that of individuals in other households in the neighborhood (Sinclair, McConnell, and Green 2012). With this restriction, unit i does not have 2^N different potential outcomes, but rather 2^{n_g} , with n_g denoting the number of units in group g .⁴ Capturing exposure-specific effects under this assumption may still be intractable in applications with large groups, so Hudgens and Halloran (2008) focus on marginal exposure effects instead.

They define four quantities of interest at the individual, group, and population level: the direct, indirect, total, and overall causal effects. For simplicity, I only describe individual level effects.⁵ The average direct causal effect for individual i in group g is:

$$\tau_{ig}^D(\psi) = \bar{Y}_{ig}(0; \psi) - \bar{Y}_{ig}(1; \psi) \tag{4}$$

⁴This notation is different from the one in Hudgens and Halloran (2008). They denote the group as i and the individual as j .

⁵In general, the group and population level causal effects are the aggregation of their individual level counterparts. See Hudgens and Halloran (2008) for details.

Which is the difference in means between treatment conditions, holding exposure regime ψ constant. The individual average indirect causal effect is:

$$\tau_{ij}^I(\phi, \psi) = \bar{Y}_{ij}(0; \phi) - \bar{Y}_{ij}(0; \psi) \quad (5)$$

Which is the difference in means between exposures regimes ϕ and ψ under the control condition. The individual average total causal effect is:

$$\tau_{ig}^T(\phi, \psi) = \bar{Y}_{ig}(0; \phi) - \bar{Y}_{ig}(1; \psi) \quad (6)$$

Which is the same as the sum of direct and indirect effects on individual i . Finally, the individual average overall causal effect is:

$$\tau_{ig}^O(\phi, \psi) = \bar{Y}_{ig}(\phi) - \bar{Y}_{ig}(\psi) \quad (7)$$

Which is equivalent to the difference in means across two different treatment regimes.

Under Bernoulli random assignment with probability ψ ,⁶ an unbiased estimator of unit i 's outcome in group g under treatment assignment $z = \{0, 1\}$ and exposure regime ψ is:

$$\hat{Y}_g(z; \psi) = \frac{\sum_{i=1}^{n_g} Y_{ig}(Z_g) I[Z_{ig} = z]}{\sum_{i=1}^{n_g} I[Z_{ig} = z]} \quad (8)$$

Which is the average of observed outcomes in a group under the same treatment condition and treatment regime Z_g . This informs the population-level estimators of direct, indirect, total, and overall causal effects. For example, the estimator for the population average indirect effect is $\hat{\tau}^I(\phi, \psi) = \hat{Y}(0; \phi) - \hat{Y}(0; \psi)$.

⁶Hudgens and Halloran (2008) suggest that the estimator is unbiased with any randomization schedule, Sävje, Aronow, and Hudgens (2019) clarify that this is only true for Bernoulli random assignment.

These estimators are unbiased under the partial interference assumption, which states that unit i 's potential outcomes depend only on the treatment assignment of units within, but not across, groups (Sobel 2006). The assumption of partial interference, which states that the potential outcomes of unit i depend on the proportion or number of treated units in group g , but not on which units within the group are treated, is sufficient but not necessary to identify causal effects.⁷

Aronow et al. (2021) show that violations of partial interference introduce bias in marginal exposure effect estimates. This violation is plausible in social science settings since individuals can interact through unobserved networks. Egami (2020) considers sensitivity analysis for the case in which the researcher observes a causally relevant network, but has concerns over whether interaction in unobserved networks may violate partial interference. For example, the researcher may observe how treatments propagate in online social networks but may not be able to account for treatment propagation through face to face interaction.

More generally, bias may emerge in this setting because groups themselves are not properly defined. In experiments with individuals as the unit of analysis, the challenge is to identify the right reference group. For example, in an experiment that assigns treatments to individuals in their households, interference can occur within a household, or across households in a neighborhood. In experiments with administrative units, the task is to identify an upper bound that determines which localities can be considered part of the same group. This paper focuses on the latter case.

3 A Supervised Learning Model Selection Protocol

Both approaches to interference, contrast across exposures and stratified interference, may suffer bias when the known pathways assumption is not satisfied. Also in both cases, a

⁷However, Hudgens and Halloran (2008) show that stratified interference is necessary for variance estimation.

supervised learning model selection approach involves identifying which of the many plausible theoretically informed operationalizations minimizes this bias. Table 1 summarizes the proposed protocol to accomplish this task. The first step is to identify one or more relevant pathways through which interference occurs. This choice comes from theory or domain expertise. In the social sciences, common pathways include connections between peers (Paluck, Shepherd, and Aronow 2016), ideological similarity (Coppock 2014), or geographical proximity (Ichino and Schündeln 2012).

Second, the researcher expresses the chosen pathway as pairwise distances between units along a distance metric. These pairwise distances are straightforward when the theoretically relevant pathway is a continuous measure indicating units' placement in multidimensional space (e.g. location, ideology scores), in which case we can express distance as the difference between the values of any two units. When pathways indicate how units connect to each other in a discrete manner, as in a peer network, the distance can be expressed as the geodesic (the lowest number of edges connecting two nodes).

As an illustration, let $W \in [0, +\infty)$ be an $M \times T$ distance matrix where M is the number plausibly exposed units and T is the number of treated units:

$$W = \begin{bmatrix} \infty & w_{12} & \cdots & w_{1T} \\ w_{21} & \infty & \cdots & w_{2T} \\ \vdots & \vdots & \ddots & \\ w_{M1} & w_{M2} & \cdots & \infty \end{bmatrix} \quad (9)$$

The dimensions of W depend on the assumptions we make about the nature of interference. If exposure to treatment affects both treated and control units, then $M = N$, if only control units can be affected by the treatment assignment of others, then $M = (N - T)$.

The task is to convert the matrix W into a matrix K of binary indicators of whether unit i can be considered as exposed to treated unit j . Let κ denote a threshold so that:

1. Identify relevant pathway(s) through which interference occurs
2. Express the pathway as pairwise distances between (plausibly) exposed and treated units along a distance metric
3. Use pairwise distances to inform plausible models of interference
4. Use supervised learning algorithm of choice to perform model/variable selection
5. Estimate interference using the model(s) that satisfy the performance criterion

Table 1: A supervised learning model selection protocol for interference

$$k_{ij} = \begin{cases} 1 & \text{if } w_{ij} \leq \kappa \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

When $k_{ij} = 1$, we say that unit i is exposed to unit j . With this information, the researcher can compute the predictors of interest based on the desired estimation target. If the goal is to estimate categorical exposure effects, then they should record unit i is exposed to at least one treated unit. If the goal is marginal exposure effects, the predictors should indicate the number of proportion of treated units i is exposed to.

The third step is to use the information about pairwise distances to identify multiple plausible models of interference. Different values of κ imply different versions of K . Once again, this decision depends on the estimation target. If the researcher seeks to identify a single global effect, then different values of κ inform alternative upper bounds, and hence different models, for which units can be considered as exposed to treatment. For example, Paluck, Shepherd, and Aronow (2016) argue that the effects of anti-conflict interventions in school spread to those students who spend time with treated students. However, the intervention can also reach those who spend time with those who spend time with treated students. In this case, κ can record network degrees as plausible upper bounds for interference.⁸

⁸Aronow and Samii (2017) overcome this challenge theoretically by allowing an arbitrary number of exposure mappings. In their approach, the possibility of second degree effects implies an increase in the number of distinct exposures.

If the goal is to identify local exposure effects at different distance ranges, then different values of κ inform the construction of different predictors among which an algorithm will select to identify the most appropriate model. For example, Ichino and Schündeln (2012) estimate the marginal effect of election observer visits on voter registration irregularities within the 0-5 km and 5-10 km distance ranges. From a model selection perspective, each one of this brackets is a different predictor that responds to a different value of κ .

In the fourth step, the researcher uses the supervised learning algorithm of choice to perform variable selection. As in the previous step, the choice of algorithm depends on the estimation target. Off the shelf variable selection algorithms are appropriate when the task is to select among multiple plausible local exposure effects. This paper focuses on the lasso-family of algorithms (see Ratkovic and Tingley 2017 for a review) since they are a straightforward extension of the regression models that researchers usually estimate in this setting. However, the researcher may also use tree-based methods (Breiman 2001; Bleich et al. 2014; Speiser et al. 2019) or any other feature selection algorithm of choice.

When the target is a single global effect, different values of κ inform a different version of the estimator, so the algorithm is the estimator itself, and the research can select the most appropriate model by searching over range of values for κ using the resampling strategy and performance criterion of choice. For regression, a common strategy is to compute mean squared error via k-fold cross-validation (Hastie, Tibshirani, and Friedman 2009). In this context, κ is a hyper-parameter, which value the researcher can tune searching over a manual grid. If the application suggests a computationally restrictive grid, random search is also an option (Bergstra and Bengio 2021).

The final step is to estimate the optimal model(s) suggested by the supervised learning protocol. The researcher can use the estimator of choice, including those discussed in the previous two sub-sections. Depending on the performance metric and resampling strategy, the protocol may suggest multiple models that satisfy the performance criteria. For example,

several models can be within one standard deviation from the model that minimizes mean squared error, in which case the researcher can consider all of them as plausible definitions of nearby.⁹

4 Simulation

4.1 Setup

This simulation study illustrates the properties of the supervised learning model selection approach in the context of estimating interference as the contrast across exposures, with the goal of identifying a single global effect.

Consider a hypothetical experiment conducted in a network. I use the network in Bowers et al. (2018) since it reflects a realistic layout. Each node in this network is an electoral area in Ghana, a subdivision of a legislative district ($N = 868$). Edges in this network denote whether a direct road connects two areas. The median number of direct connections is 22 (mean: 38, standard deviation: 41).

Bowers et al. (2018) focus on the case where interference occurs between adjacent nodes only. In this simulation exercise, interference may also occur at higher degrees. Therefore, the distance metric is the geodesic, or the number of edges in the shortest path between two nodes.

Z is a treatment assignment vector. $Z_i \sim \text{Bernoulli}(\alpha)$, which implies two possible treatment conditions $Z_i = \{0, 1\}$, which I refer to as control and treatment, respectively. One of the main conclusions in Bowers et al. (2018) is that experiments have higher power to detect interference when the proportion of units assigned to treatment is smaller than the usual even split across conditions. With this in mind, I set $\alpha \in [0.1, 0.4]$. Figure 1 shows one realization

⁹When supervised learning algorithms use current data to predict new data, the common practice to reduce overfitting is to select the model with the highest mean squared error within one standard deviation from the minimum (Hastie, Tibshirani, and Friedman 2009). Since the task here is within-sample model selection, any model within a reasonable range away from the minimum can be considered viable.

of the treatment assignment in the network with $\alpha = 0.1$.

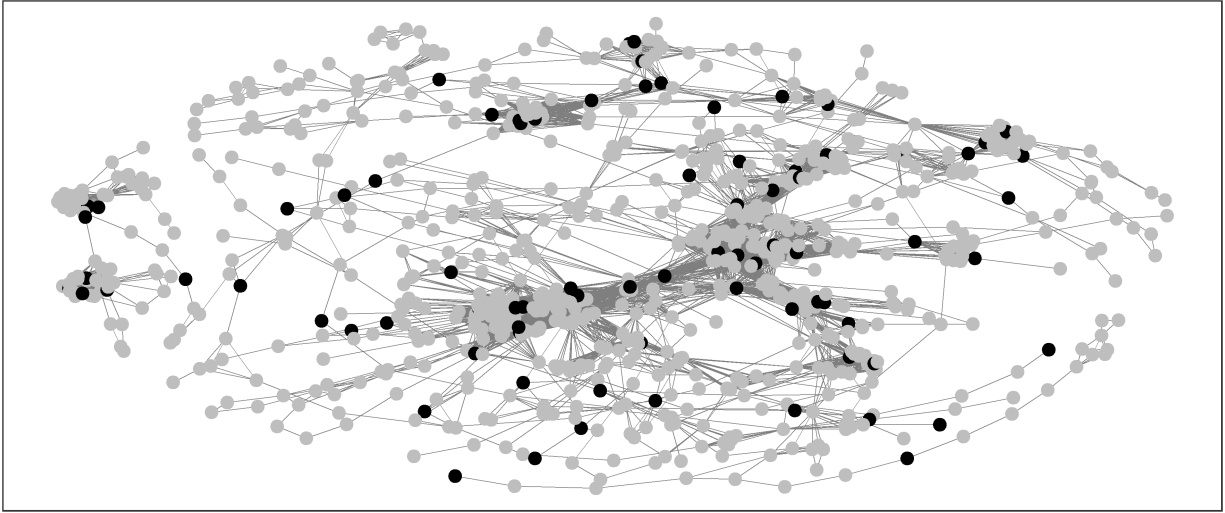
$Y(d_{00})$ is the vector of potential outcomes under control and without exposure, $Y(d_{10})$ under treatment without exposure, $Y(d_{01})$ under control with exposure, and $Y(d_{11})$ under both treatment and exposure. I assume $Y(d_{00}) \sim U(0, 1)$ and $Y(d_{10}) = Y(d_{01}) = Y(d_{11}) = \lambda Y(d_{00})$ with the multiplicative treatment effect $\lambda \in \{0.26, 0.63\}$, which is about one and two standard deviations of the uniform distribution of the outcome, respectively. Under this setup, we can restrict our attention the effect of being exposed to treatment within control units, which is the contrast between $Y(d_{00})$ and $Y(d_{01})$.

To capture the idea that interference may occur within different distance ranges, I set $\kappa = \{1, 2, 3\}$ as the true interference upper bound. For example, with $\kappa = 2$ a node is exposed to treatment if it is within two or fewer edges away from a treated unit. Within this range, a node is infected with probability $\gamma \in [0.5, 1]$. Infected units exhibit the outcome under exposure $Y(d_{01})$, otherwise they exhibit $Y(d_{00})$. This reflects the scenario where treatments propagate at different rates.

For each combination of the parameters α , λ , κ , and γ , I consider plausible upper bounds $k = \{1, 2, 3, 4\}$. k is the researcher's guess of what the value of the true κ may be, with higher values reflecting a broader range within which units are considered exposed to treatment. For each value of k , I consider the inverse probability weighted difference in means described in Equation 3 as both the algorithm and the estimator. I use 5-fold cross validation to compute the root mean squared error (RMSE) for each model and choose the one with the largest RMSE within one standard deviation of the mean. I denote the value of k in the chosen model as \bar{k} , and estimate the corresponding model.¹⁰

I repeat this process 1,000 times for each parameter combination and compare the protocol's performance with an oracle that knows the value of κ ahead of time. I assess performance with bias, mean absolute deviation (as a measure of consistency), power at a 0.05 false

¹⁰This decision rule implies erring on the side of selecting a smaller \bar{k} .



Condition ● Treatment ● Control

Figure 1: Hypothetical experiment in a road network in Ghana

Note: Colors denote treatment assignment. Units are assigned to treatment with probability $\alpha = 0.1$.

positive rate, and the mean value of \bar{k} .¹¹

4.2 Results

To facilitate exposition, Figures 2-5 report performance at select parameter values $\alpha = \{0.1, 0.4\}$, $\gamma = \{0.5, 1\}$, and $\tau = 0.26$. Section A of the appendix reports simulation results in full. Figure 2 shows the bias of the estimator selected by the proposed supervised learning protocol against an oracle that knows the true value of the upper bound κ . Each value in the figure is based on 1,000 simulations. The protocol has similar bias than the oracle when treatment assignment probability α is low. When α is close to a coin flip ($\alpha = 0.4$), the protocol increases in bias with κ . In general, bias is higher when the infection probability γ is lower, but the difference between the protocol and oracle as κ increases under high α is more pronounced when the infection probability γ is deterministic ($\gamma = 1$).

Figure 3 shows mean absolute deviation as a measure of consistency, with higher values

¹¹In some settings, treatment spreads too fast, which leads to implausible or rank-deficient estimates. This occurs in roughly 3% of the simulations (out of a total of 288,000) and happens more often as α and κ increase. I exclude these from the analysis.

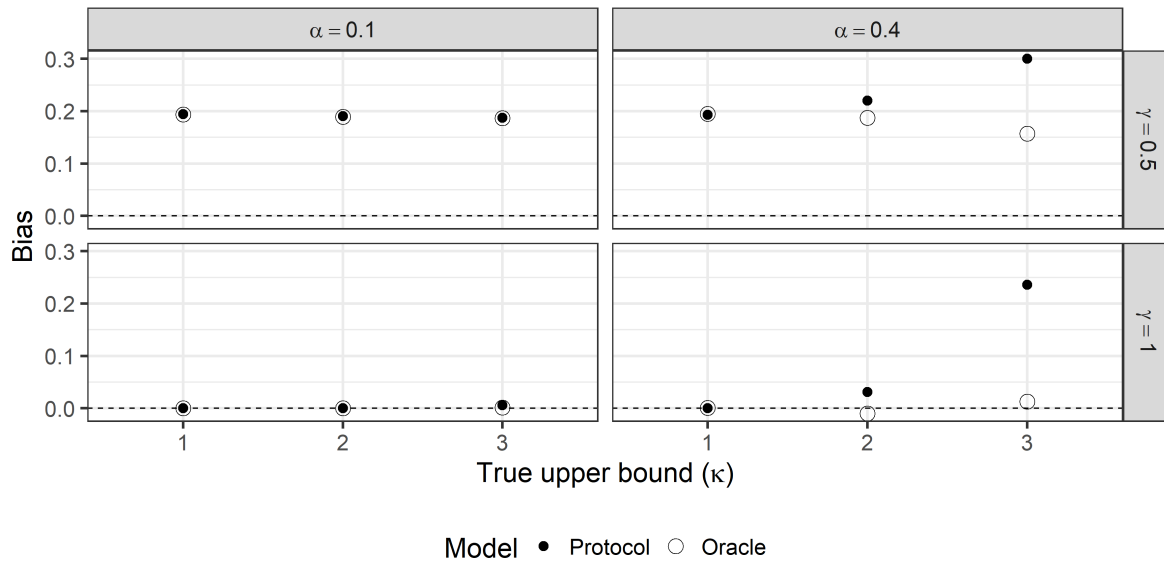


Figure 2: Comparing bias of the supervised learning model selection protocol against the oracle

Note: Each value is based on 1,000 simulations. When the treatment assignment probability α is low, the protocol and oracle have similar bias. When α is high, the bias of the protocol increases with the true upper bound κ . Bias is higher when the infection rate γ is a coin toss.

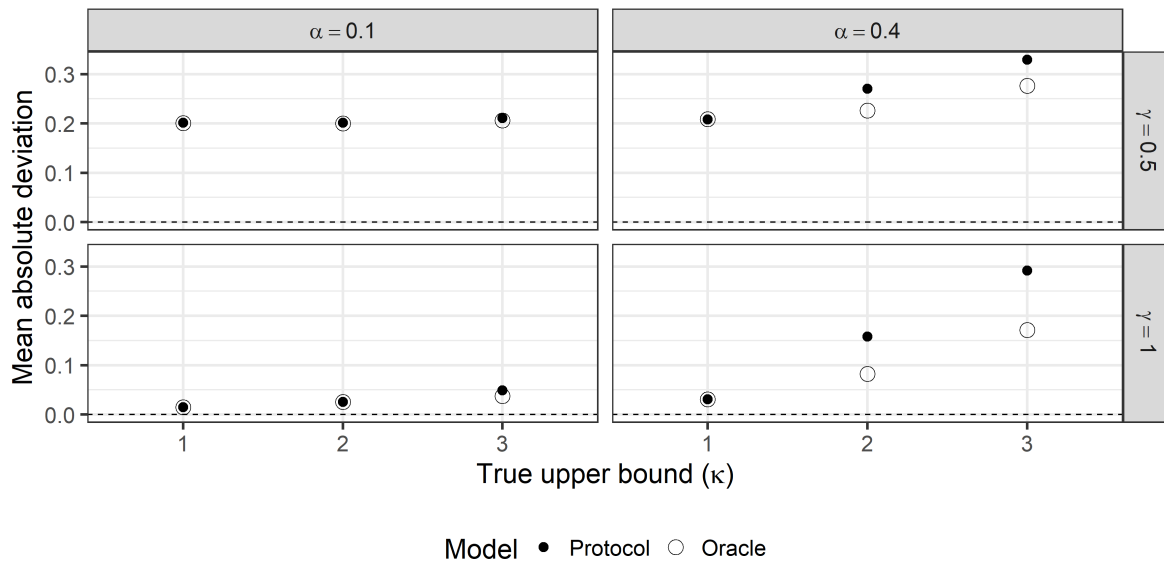


Figure 3: Comparing the consistency of the supervised learning model selection protocol against the oracle

Note: Each value is based on 1,000 simulations. When the treatment assignment probability α is low, the protocol and oracle have similar mean absolute deviation. When α is high, the mean absolute deviation of the protocol increases with the true upper bound κ faster than the oracle. Mean absolute deviation is higher when the infection rate γ is a coin toss.

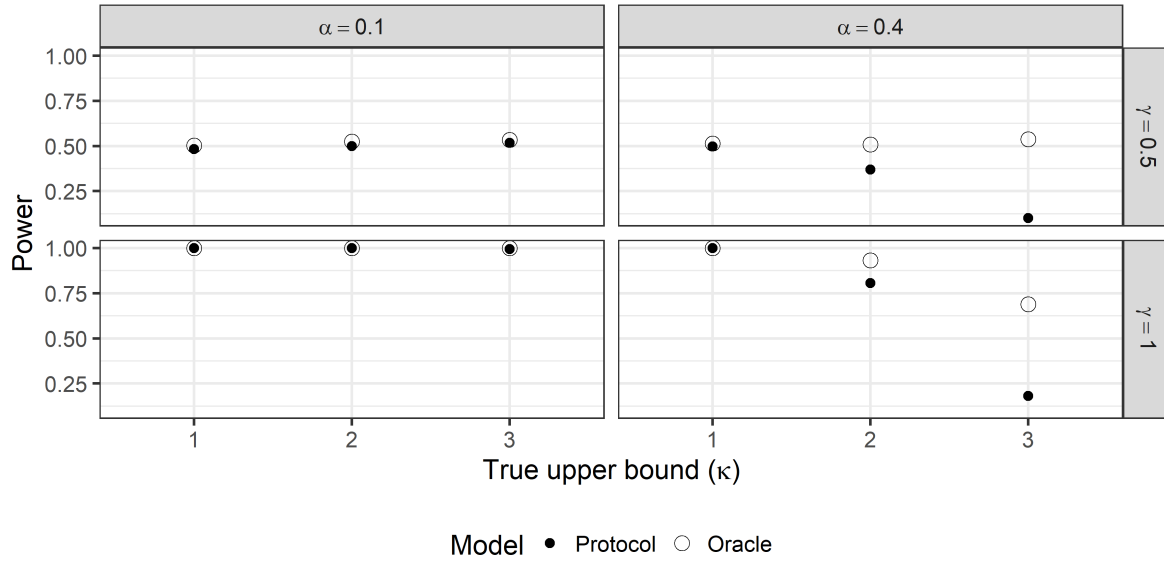


Figure 4: Comparing the power of the supervised learning model selection protocol against the oracle

Note: Each value is based on 1,000 simulations. Power is the proportion of the simulations in which the test rejects the null of no effect at a significance level of 0.05 when $\tau = 0.26$. Power is high when the infection rate γ is deterministic and the treatment assignment probability α or the true upper bound κ are low. As κ increases, the power of the protocol decreases faster than the oracle.

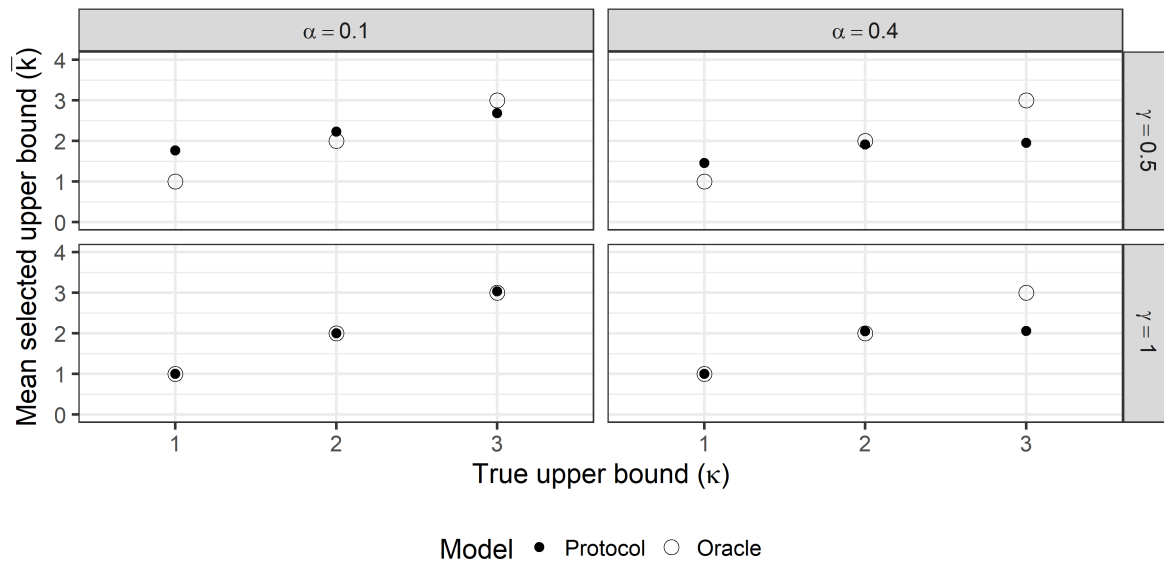


Figure 5: Comparing the mean selected upper bound of the supervised learning model selection protocol against the oracle

Note: Each value is based on 1,000 simulations. By definition the oracle always selects the right value of the true upper bound κ . In average, the protocol chooses the correct upper bound when the treatment assignment probability α is low and the infection rate γ is deterministic. Otherwise, the protocol tends to select a broader upper bound when κ is low and a narrower upper bound when κ is high.

suggesting more variation in the distribution of estimates. As with bias, the protocol has similar consistency to the oracle with low α , and in general estimates are more inconsistent when γ is a coin toss. When alpha is high, mean absolute deviation increases with κ for both the protocol and the oracle, but the change is more pronounced for the protocol.

Figure 4 shows power as the proportion of the simulations in which the test that follows from the inverse probability weighted difference in means rejects the null hypothesis of no effect at a significance level of 0.05. Power is high when γ is deterministic and either α or κ are low. Similar to bias and consistency, power tends to decrease as κ increases under $\alpha = 0.4$ and the change is more pronounced for the protocol than the oracle. This is more evident in the case of $\gamma = 1$.

Finally, Figure 5 shows the mean selected upper bound \bar{k} , this statistic reflects the protocol's average guess for κ across simulations. By definition, the oracle always selects the correct interference upper bound, so the protocol performs better when it resembles the oracle. This is the case under $\alpha = 0.1$ and $\gamma = 1$. In other scenarios, the protocol tends to overestimate the upper bound when κ is low, and underestimate it when κ is high.

To summarize, the simulation exercise suggests that the protocol performs better when a relatively small proportion of units are assigned to treatment. This finding also appears in the simulations in Bowers et al. (2018). The underlying intuition is that when too many units are treated there is not enough information about the outcome of control units to assess the difference between those exposed to and isolated from treatment. For the same reason, performance tends to be worse when treatments travel too far, which is captured by κ . If too many units are exposed, then the researcher does not have enough information about those isolated.

Finally, performance also suffers when treatments spread in a probabilistic way. This is true for both the protocol and the oracle, which suggests that the problem does not lie in the supervised learning model selection approach, but rather in the theoretical decision to

commit to the same upper bound across units. Future work may alleviate this problem by implementing more flexible distance metrics and algorithms that allow the interference upper bound to vary across units or clusters thereof.

5 Application

5.1 Election observers and voting registration irregularities in Ghana

The simulation exercise in the previous section illustrates the supervised learning model selection approach to interference in the case of a single global categorical effect. This section illustrates the case of multiple marginal effects by reanalyzing an experiment on the spillover effect of election observers on voter registration irregularities in Ghana (Ichino and Schündeln 2012). Table 2 summarizes the two-stage research design. Before randomization, the authors group constituencies (legislative districts) into blocks according to the difference in vote shares between the two main parties in the country in the 2004 legislative election. At the first stage, one constituency per block was randomly assigned to treatment and two were assigned to control. In the second stage, roughly 25% of the electoral areas (ELAs), subdivisions of constituencies, were assigned to receive the visit of an election registration observer in the wake of the 2008 election.

The main finding in this experiment is that election observers do not deter but rather displace irregularities, one additional ELA assigned to treatment within 5 kilometers leads roughly to a 3% increase in irregularities, measured as the percent change in voter registration between 2004 and 2008, in control ELAs.¹² The effect is indistinguishable from zero for the number of ELAs assigned to treatment in the 5-10 kilometer range.

¹²This is the intent-to-treat effect, which is the focus of this paper, the local average treatment effect for those ELAs that were visited by an election observer is similar. See Ichino and Schündeln (2012) for details.

Constituencies	Electoral areas (ELAs)	Observations
Control	Control	592
Treatment	Control	199
Treatment	Treatment	77

Table 2: Research design in Ichino and Schündeln (2012)

Note: Two-stage design. Constituencies are randomly assigned to treatment and control, then ELAs within treatment constituencies are randomly assigned to the visit of an election observer.

5.2 Lasso model selection

The exercise from a model selection perspective is to determine whether the proposed protocol can recover the same findings as an experiment designed to capture interference. I assume that the researcher believes that geographic distance is a relevant distance metric and, along with the original authors' decision, count the number ELAs assigned to treatment at 5 kilometer intervals up to 50 kilometers. This translates to 10 predictors among which to select.

I select among these predictors using lasso-family algorithms. The logic of lasso (Least Absolute Shrinkage and Selection Operator) methods is to add a penalty term to the the objective function of OLS regression, which is the residual sum of squares (see Tibshirani 1996 for mathematical details). As the penalty term increases, the resulting regression coefficients shrink toward zero, a coefficient that reaches zero is excluded from the estimation step. The size of the penalty term is determined by a tuning parameter λ , which for the frequentist lasso is tuned via cross-validation or other resampling strategy.

A shortcoming of the standard lasso is that it tends to select irrelevant predictors that correlate with relevant predictors. Figure 6 shows the correlation matrix for the variables involved in the reanalysis. Many predictors correlate positively with each other, which may create complications for variable selection in the lasso. Subsequent refinements to the lasso attempt to minimize the tendency to over-select predictors (see Ratkovic and Tingley 2017 for a review).

In this application, I focus on the following variants of the lasso:

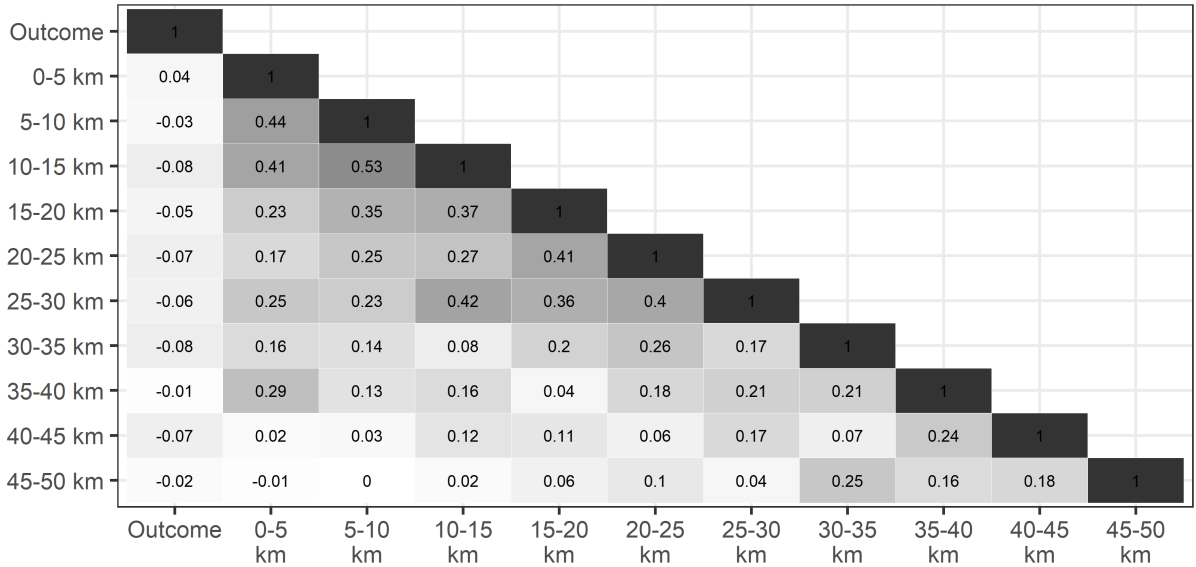


Figure 6: Correlation matrix for outcome and predictors

Note: The outcome is the 2004-2008 percent change in voter registration. The predictors count the number of ELAs assigned to treatment in the corresponding range. Numbers indicate the Pearson correlation coefficient between the intersecting variables. Darker shades of gray indicate more extreme values.

1. Lasso (Tibshirani 1996)
2. Adaptive lasso (Zou 2006)
3. Lasso + OLS (Belloni and Chernozhukov 2013)
4. LASSOplus (Ratkovic and Tingley 2017)

The logic of the original lasso is described above. The adaptive lasso introduces more flexibility by allowing predictor-specific weights, resulting on a penalty term that varies across predictors (Zou 2006). Following standard practice, I weight each predictor by the multiplicative inverse of the absolute value of its coefficient in a multivariate OLS regression of the outcome against the predictors, so for predictor p the weight is $1/|\beta_p|$.

The LASSO + OLS variant seeks to improve upon the initial lasso fit by excluding non-zero predictors that exceed a threshold. In this application, I focus on the goodness of fit thresholding (also known as OLS post-fit Lasso) procedure proposed by Belloni and Chernozhukov (2013), which consists of fitting an OLS regression for all the subsets of the

model suggested by the lasso, and then choosing the subset with a residual variance that resembles the residual variance of the suggested lasso fit.

Lastly, the LASSOplus proposed by Ratkovic and Tingley (2017) is a Bayesian procedure that streamlines model selection and estimation. This method has three advantages. First, as a Bayesian method, it produces credible intervals analogous to the confidence intervals obtained by OLS regression, so it does not need an additional step for inference like the frequentist variants do. The second advantage of LASSOplus is a thresholding function that zeroes out small coefficients. Bayesian variants of the lasso tend to have better predictive performance (Casella et al. 2010), but are not sparse estimators in the sense that they do not automatically zero-out coefficients. Third, unlike standard lasso methods, LASSOplus can incorporate prior knowledge about the data generating process, which in the context of this paper translates to research design features. For example, LASSOplus can automatically incorporate the blocked structure of the Ichino and Schündeln (2012) experiment and adjust the posterior distribution accordingly.

For the frequentist variants of the lasso (standard, adaptive, and lasso + OLS), I tune λ via 10-fold cross-validation using the `glmnet` R package and choose the most generous model among those implied by the range within one standard deviation from the value of λ that minimizes RMSE.¹³ I follow the default parameters of the `sparsereg` R package for LASSOplus, which estimates the model via MCMC via Gibbs sampling with 200 burn-in iterations, 200 posterior samples, and thinning every 10 samples.¹⁴

5.3 Results

Figure 7 summarizes the variables selected by each lasso variant. Figures B1-B2 in the appendix show additional details for the frequentist lasso variants. The horizontal axis

¹³In this case, more generous implies selecting the model with most predictors.

¹⁴This is a relatively small number of iterations yet, as the next sub-section shows, sufficient to yield satisfactory results.

lists every predictor and the dark blocks denote whether a predictor was selected by each also variant. For reference, the first row denotes the original model selection in Ichino and Schündeln (2012), including the number of ELAs assigned to treatment within 0-5 kilometers and 5-10 kilometers, of which only the first has a non-zero effect. All the frequentist variants of the lasso correctly select this predictor, but also overselect every other predictor. This pattern emerges because continuous predictors tend to correlate with each other, and the lasso chooses to penalize one heavily and to leave the other untouched. LASSOplus, in turn, correctly selects the number of ELAs assigned to treatment within 0-5 kilometers as the only non-zero predictor, which aligns with the original findings.

Figure 8 shows the estimates from the models suggested by each lasso variant in separate frames. The first frame reproduces the original result in Ichino and Schündeln (2012). For the frequentist variants, estimation occurs after model selection following the original specification, which includes block fixed effects, a control for whether an ELA is in a treatment constituency, and clustered standard errors at the block level. Although all of the frequentist variants select irrelevant predictors, the estimation returns an effect similar to the original for the number of ELAs assigned to treatment in the 0-5 kilometer range. The remaining estimates are indistinguishable from zero, which still aligns with the conclusion that spillover effects do not occur beyond five kilometers.

For LASSOplus, the point estimate is the posterior mean of the same model used for variable selection, which includes the original research design features as prior information. Unlike the frequentist variants, LASSOplus only selects the number of ELAs assigned to treatment in the 0-5 kilometer range, and while the estimate is smaller than the original, the 95% credible interval is narrower than in the frequentist counterparts.

To summarize, both the frequentist variants of the lasso and LASSOplus arrive at the same substantive conclusion than an experiment intended to capture interference. As documented in previous work, the frequentist variants tend to select irrelevant effects that correlate with

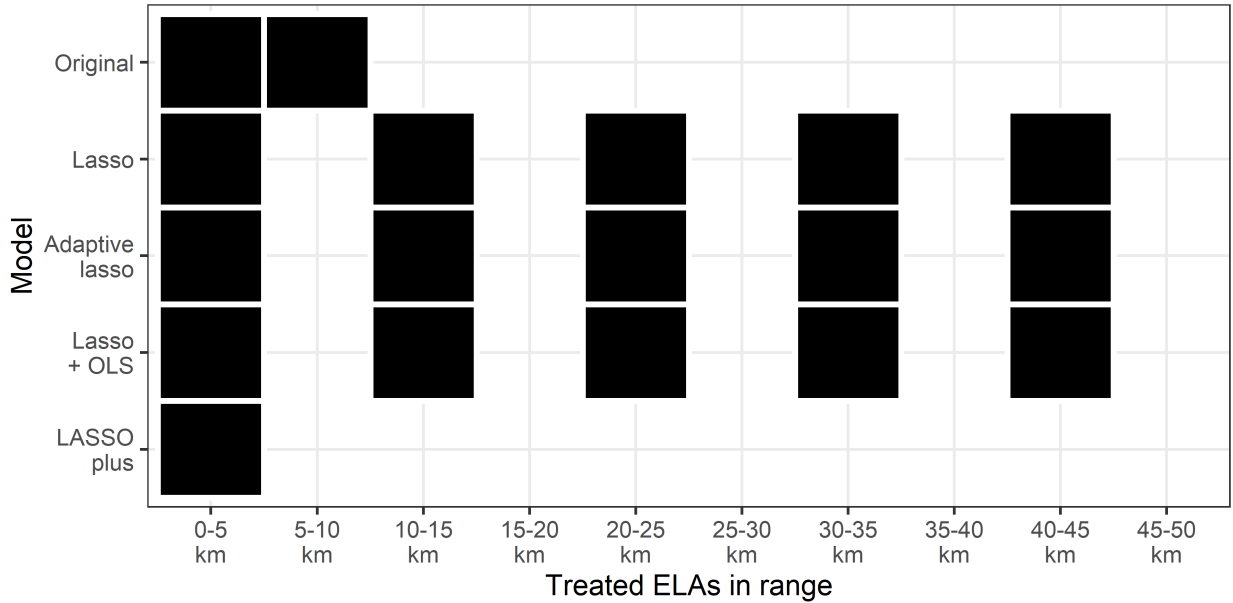


Figure 7: Predictors selected by each lasso variant

Note: Each column in the horizontal axis denotes a different predictor counting the number of ELAs assigned to treatment in the corresponding range. Blocks denote whether each predictor has a nonzero coefficient in the model chosen by the corresponding lasso method. For reference, the first row denotes the model selection in Ichino and Schündeln (2012).

true non-zero effects (Ratkovic and Tingley 2017), a problem that LASSOplus overcomes in this application. Frequentist lasso-family methods arrive at the correct substantive conclusion despite over-selecting because, under regular conditions, the frequentist lasso and its variants satisfy the oracle inequality, which is the idea that, asymptotically, the lasso selects at least a subset of the true model (Candès 2006).¹⁵ Therefore, both frequentist and Bayesian sparse estimation models are suitable for a model selection approach to interference.

6 Conclusion

This paper proposes a supervised learning model selection approach to study interference in contexts in which the identification assumptions of current design-based causal inference approaches cannot be satisfied. In such case, theory still guides modeling decisions, but it

¹⁵The oracle inequality exists in contrast with the oracle property, which requires a variable selection procedure to asymptotically select the right model (Fan and Li 2001).

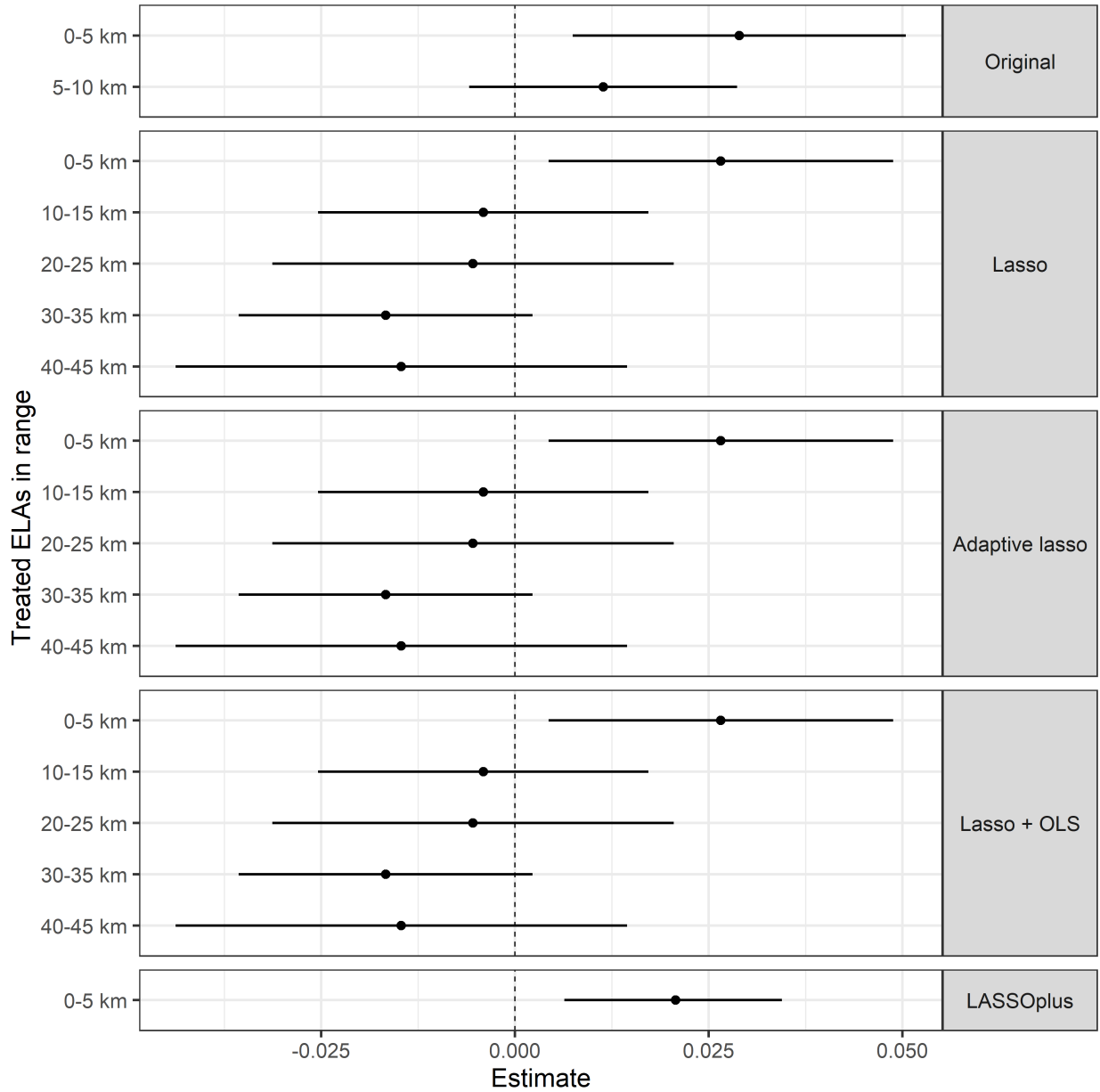


Figure 8: Estimates for the marginal effect of the number of ELAs assigned to treatment within selected distance ranges on voter registration irregularities

Note: Each frame denotes a different model. The first frame reproduces the original results. For the frequentist variants of lasso, the estimates come from post-selection OLS regression with block fixed effects and a control for whether an ELA is in a treatment constituency, with bars representing 95% confidence intervals from block clustered standard errors. For LASSOplus, the point estimate is the posterior mean and the bars denote the 95% credible interval.

often implies multiple ways to operationalize interference, and the discipline currently lacks a standard to determine which model is more appropriate. This approach still puts the weight of justifying causal inference on what the researcher is willing to assume about the pathway through which interference occurs, but it provides tools to connect theoretical justification with modeling decisions.

The key intuition the supervised learning model based approach is that, if the researcher can express a theory about interference in terms of a distance metric, then the metric can help in identifying plausible alternative operationalizations, among which the researcher can select using the algorithm, resampling strategy, and performance metric of choice.

Context-specific considerations aside, the choice of model selection approach depends primarily on two decisions about the estimation target. First, the researcher must determine whether the goal is to estimate categorical or marginal exposure effects. This informs whether to record if a unit is exposed to a treatment condition or not, or to count the number of units assigned to a treatment condition within a certain range.

Second, the researcher must determine whether the task is to estimate a single global effect, or multiple local effects at different distance ranges. This informs the type of algorithm. In the case of a single global effect, the distance metric informs the different values of the upper bound of a range within which interference occurs, which in turns assists in identify alternative plausible operationalizations. For multiple local effects, the distance metric informs the construction of predictors among which the researcher can select using off-the-shelf algorithms for variable selection.

This paper illustrates the supervised learning model selection approach with simulations and the reanalysis of an experiment intended to capture interference. Focusing on the case of estimating the contrast across two exposures in a road network in Ghana, the simulations suggest that the proposed protocol performs better when a relatively small proportion of units is treated, when the probability of infection conditional on exposure is deterministic (as

opposed to probabilistic), and when treatments do not spread too far along the theoretically relevant pathways. These conclusions align with previous simulations exercises (e.g. Bowers et al. 2018), which highlights the appropriateness of the approach. Similarly, the reanalysis of an experiment on the spillover effect of election observers on voter registration irregularities in Ghana (Ichino and Schündeln 2012) shows how the lasso-family of algorithms select models that lead to estimates that resemble the original findings, which also validates the usefulness of the approach.

The simulations and reanalysis focus on model selection using cross-validation and the lasso family of methods, respectively. Similarly, the focus on the estimation side is on inverse probability weighted difference-in-means (Aronow and Samii 2017) and conventional design-based estimators for experimental data. However, the ideas in this paper are general enough to accommodate for other model selection algorithms and estimation approaches.

A limitation of this approach is that it does not suit hypothesis testing approaches to interference. For example, the key intuition in Bowers, Fredrickson, and Panagopoulos (2013) is that we can test hypotheses about a range of parameter values in a theoretical model of interference. Supervised learning algorithms for variable selection operate under the assumption that a true effect exists and can be detected, even if unknown to the researcher. Testing approaches follow the idea that we can compute uncertainty for a range of hypothetical parameter values.

Future work should acknowledge this different and develop standards to select among competing models in the context of hypothesis testing approaches to interference. Another area for future development is to develop guidelines on how to choose the appropriate distance metric. For example, the reanalysis of an experiment on the effect of election observers on voter registration irregularities in Ghana follows the authors' decision to count the number of treated units within 5 kilometer bins. However, in novel applications the choice of the number and size of bins (or whether to use bins at all) might be consequential for both the

applicability and performance of a supervised learning algorithm.

Finally, future work should apply the ideas of the supervised learning model selection approach to the task of comparing the appropriateness of multiple plausible theoretical pathways. In such a case, the challenge is not only to determine which one is more appropriate, but also to assess and interpret whether alternative pathways are complementary.

References

- Aronow, Peter M. 2012. “A General Method for Detecting Interference Between Units in Randomized Experiments.” *Sociological Methods & Research* 41 (1): 3–16. <https://doi.org/10.1177/0049124112437535>.
- Aronow, Peter M., Dean Eckles, Cyrus Samii, and Stephanie Zonszein. 2021. “Spillover Effects in Experimental Data.” In *Advances in Experimental Political Science*, 289–319. Cambridge University Press. <https://doi.org/10.1017/9781108777919.021>.
- Aronow, Peter M., and Joel A. Middleton. 2013. “A Class of Unbiased Estimators of the Average Treatment Effect in Randomized Experiments.” *Journal of Causal Inference* 1 (1): 135–54. <https://doi.org/10.1515/jci-2012-0009>.
- Aronow, Peter M., and Cyrus Samii. 2017. “Estimating Average Causal Effects Under General Interference, with Application to a Social Network Experiment.” *The Annals of Applied Statistics* 11 (4): 1912–47. <https://doi.org/10.1214/16-aos1005>.
- Asunka, Joseph, Sarah Brierley, Miriam Golden, Eric Kramon, and George Oforu. 2017. “Electoral Fraud or Violence: The Effect of Observers on Party Manipulation Strategies.” *British Journal of Political Science* 49 (1): 129–51. <https://doi.org/10.1017/s0007123416000491>.
- Athey, Susan, Dean Eckles, and Guido W. Imbens. 2018. “Exact p-Values for Network Interference.” *Journal of the American Statistical Association* 113 (521): 230–40. <https://doi.org/10.1080/01621459.2018.1512745>.

[//doi.org/10.1080/01621459.2016.1241178](https://doi.org/10.1080/01621459.2016.1241178).

- Baird, Sarah, J. Aislinn Bohren, Craig McIntosh, and Berk Özler. 2018. “Optimal Design of Experiments in the Presence of Interference.” *The Review of Economics and Statistics* 100 (5): 844–60. https://doi.org/10.1162/rest_a_00716.
- Basse, Guillaume W., Avi Feller, and Panos Toulis. 2019. “Randomization Tests of Causal Effects Under Interference.” *Biometrika* 106 (2): 487–94. <https://doi.org/10.1093/biomet/asy072>.
- Belloni, Alexandre, and Victor Chernozhukov. 2013. “Least Squares After Model Selection in High-Dimensional Sparse Models.” *Bernoulli* 19 (2): 521–47. <https://doi.org/10.3150/11-bej410>.
- Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2013. “Inference on Treatment Effects After Selection Among High-Dimensional Controls.” *The Review of Economic Studies* 81 (2): 608–50. <https://doi.org/10.1093/restud/rdt044>.
- Bergstra, James, and Yoshua Bengio. 2021. “Random Search for Hyper-Parameter Optimization.” *Journal of Machine Learning Research* 13 (10): 281–305. <http://jmlr.org/papers/v13/bergstra12a.html>.
- Blakely, Tony, John Lynch, Koen Simons, Rebecca Bentley, and Sherri Rose. 2019. “Reflection on Modern Methods: When Worlds Collide - Prediction, Machine Learning and Causal Inference.” *International Journal of Epidemiology*, July. <https://doi.org/10.1093/ije/dyz132>.
- Bleich, Justin, Adam Kapelner, Edward I. George, and Shane T. Jensen. 2014. “Variable Selection for BART: An Application to Gene Regulation.” *The Annals of Applied Statistics* 8 (3): 1750–81. <https://doi.org/10.1214/14-aos755>.
- Bloniarczyk, Adam, Hanzhong Liu, Cun-Hui Zhang, Jasjeet S. Sekhon, and Bin Yu. 2016. “Lasso Adjustments of Treatment Effect Estimates in Randomized Experiments.” *Proceedings of*

- the National Academy of Sciences* 113 (27): 7383–90. <https://doi.org/10.1073/pnas.1510506113>.
- Bowers, Jake, Bruce A. Desmarais, Mark Frederickson, Nahomi Ichino, Hsuan-Wei Lee, and Simi Wang. 2018. “Models, Methods and Network Topology: Experimental Design for the Study of Interference.” *Social Networks* 54 (July): 196–208. <https://doi.org/10.1016/j.socnet.2018.01.010>.
- Bowers, Jake, Mark M. Fredrickson, and Costas Panagopoulos. 2013. “Reasoning about Interference Between Units: A General Framework.” *Political Analysis* 21 (1): 97–124. <https://doi.org/10.1093/pan/mps038>.
- Breiman, Leo. 2001. “Random Forests.” *Machine Learning* 45 (1): 5–32. <https://doi.org/10.1023/a:1010933404324>.
- Candès, Emmanuel J. 2006. “Modern Statistical Estimation via Oracle Inequalities.” *Acta Numerica* 15 (May): 257–325. <https://doi.org/10.1017/s0962492906230010>.
- Casella, George, Malay Ghosh, Jeff Gill, and Minjung Kyung. 2010. “Penalized Regression, Standard Errors, and Bayesian Lasso.” *Bayesian Analysis* 5 (2). <https://doi.org/10.1214/10-ba607>.
- Coppock, Alexander. 2014. “Information Spillovers: Another Look at Experimental Estimates of Legislator Responsiveness.” *Journal of Experimental Political Science* 1 (2): 159–69. <https://doi.org/10.1017/xps.2014.9>.
- Cox, David R. 1958. *Planning of Experiments*. New York: Wiley.
- Egami, Naoki. 2020. “Spillover Effects in the Presence of Unobserved Networks.” *Political Analysis*, November, 1–30. <https://doi.org/10.1017/pan.2020.28>.
- Fan, Jianqing, and Runze Li. 2001. “Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties.” *Journal of the American Statistical Association* 96 (456):

- 1348–60. <https://doi.org/10.1198/016214501753382273>.
- Halloran, M. Elizabeth, and Michael G. Hudgens. 2016. “Dependent Happenings: A Recent Methodological Review.” *Current Epidemiology Reports* 3 (4): 297–305. <https://doi.org/10.1007/s40471-016-0086-4>.
- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning*. New York: Springer.
- Holland, Paul W. 1986. “Statistics and Causal Inference.” *Journal of the American Statistical Association* 81 (396): 945–60. <https://doi.org/10.1080/01621459.1986.10478354>.
- Hudgens, Michael G, and M. Elizabeth Halloran. 2008. “Toward Causal Inference with Interference.” *Journal of the American Statistical Association* 103 (482): 832–42. <https://doi.org/10.1198/016214508000000292>.
- Ichino, Nahomi, and Matthias Schündeln. 2012. “Deterring or Displacing Electoral Irregularities? Spillover Effects of Observers in a Randomized Field Experiment in Ghana.” *The Journal of Politics* 74 (1): 292–307. <https://doi.org/10.1017/s0022381611001368>.
- Keele, Luke, and Rocío Titiunik. 2018. “Geographic Natural Experiments with Interference: The Effect of All-Mail Voting on Turnout in Colorado.” *CESifo Economic Studies* 64 (2): 127–49. <https://doi.org/10.1093/cesifo/ify004>.
- Nickerson, David W. 2008. “Is Voting Contagious? Evidence from Two Field Experiments.” *American Political Science Review* 102 (1): 49–57. <https://doi.org/10.1017/s0003055408080039>.
- Ogburn, Elizabeth L., and Tyler J. VanderWeele. 2014. “Causal Diagrams for Interference.” *Statistical Science* 29 (4). <https://doi.org/10.1214/14-sts501>.
- Paluck, Elizabeth Levy, Hana Shepherd, and Peter M. Aronow. 2016. “Changing Climates of Conflict: A Social Network Experiment in 56 Schools.” *Proceedings of the National*

- Academy of Sciences* 113 (3): 566–71. <https://doi.org/10.1073/pnas.1514483113>.
- Ratkovic, Marc, and Dustin Tingley. 2017. “Sparse Estimation and Uncertainty with Application to Subgroup Analysis.” *Political Analysis* 25 (1): 1–40. <https://doi.org/10.1017/pan.2016.14>.
- Rubin, Donald B. 1990. “Formal Models of Statistical Inference for Causal Effects.” *Journal of Statistical Planning and Inference* 25 (3): 279–92. [https://doi.org/10.1016/0378-3758\(90\)90077-8](https://doi.org/10.1016/0378-3758(90)90077-8).
- Sävje, Fredrik. 2019. “Causal Inference with Misspecified Exposure Mappings.” *Working Paper*. <https://fredriksavje.com/papers/misspecified-exposures.pdf>.
- Sävje, Fredrik, Peter M. Aronow, and Michael G. Hudgens. 2019. “Average Treatment Effects in the Presence of Unknown Interference.” <http://arxiv.org/abs/1711.06399>.
- Sinclair, Betsy, Margaret McConnell, and Donald P. Green. 2012. “Detecting Spillover Effects: Design and Analysis of Multilevel Experiments.” *American Journal of Political Science* 56 (4): 1055–69. <https://doi.org/10.1111/j.1540-5907.2012.00592.x>.
- Sobel, Michael E. 2006. “What Do Randomized Studies of Housing Mobility Demonstrate?” *Journal of the American Statistical Association* 101 (476): 1398–1407. <https://doi.org/10.1198/016214506000000636>.
- Speiser, Jaime Lynn, Michael E. Miller, Janet Tooze, and Edward Ip. 2019. “A Comparison of Random Forest Variable Selection Methods for Classification Prediction Modeling.” *Expert Systems with Applications* 134 (November): 93–101. <https://doi.org/10.1016/j.eswa.2019.05.028>.
- Tibshirani, Robert. 1996. “Regression Shrinkage and Selection via the Lasso.” *Journal of the Royal Statistical Society: Series B (Methodological)* 58 (1): 267–88. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>.

Toulis, Panos, and Edward Kao. 2013. “Estimation of Causal Peer Influence Effects.” *Proceedings of the 30th International Conference on Machine Learning* 3 (28): 1489–97. <http://proceedings.mlr.press/v28/toulis13.html>.

Zou, Hui. 2006. “The Adaptive Lasso and Its Oracle Properties.” *Journal of the American Statistical Association* 101 (476): 1418–29. <https://doi.org/10.1198/016214506000000735>.