Posterior Summarization for the Causal Linear Model

Spencer Woody[1]⋆, Carlos M. Carvalho[2], & Jared S. Murray[2]

The University of Texas at Austin

[1] Department of Statistics and Data Sciences

[2] Department of Information, Risk, and Operations Management

⋆ spencerwoody.github.io — spencer.woody@utexas.edu

The University of Texas at Austin

Abstract

In the causal linear model, it is common practice to include many control variables to meet the assumption of strong ignorability. However, this results in high standard errors for the treatment effect and leads to difficulty in understanding the causal mechanism, as we do not know which controls are most important. In this work we propose to a method to give valid inference on a range of plausible explanations of model behavior through the use of posterior summarization. Following other works in this line of research, we follow a two-stage approach. In the first stage, a high-dimensional model is fit to the data. Then, in the second stage a lower-dimensional summary model is fit to the predictions of the original model. After this, it is determined whether the summary model is adequately representative of the original model in terms of estimating the treatment effect. We apply our method to a dataset measuring the impact of abortion on murder rates.

Introduction

Consider the linear model for the effect of continuous treatment Z on continuous outcome Y,

$$\begin{aligned} \text{Treatment equation:} &\quad Z = X\gamma + \epsilon \\ \text{Response equation:} &\quad Y = \alpha Z + X\beta + \nu \end{aligned} \tag{1}$$

with normal errors $\epsilon \sim \mathcal{N}(0, \sigma_z^2 \mathcal{I})$, $\nu \sim \mathcal{N}(0, \sigma_y^2 \mathcal{I})$, carefully specified shrinkage priors (see [3]) for γ (control effects) and β (prognostic effects), and noninformative priors for α (treatment effect), and variance components σ_z^2 and σ_y^2.

For identification of the treatment effect α, it is required that X contains the proper controls so that

$$\text{cov}(Z_i, \nu_i \mid X_i) = 0. \tag{2}$$

Because such an assumption is untestable, it is standard practice to **include as many controls as is deemed feasible** (i.e., p is large). However, this has several downsides, including **inflation of standard errors** and **difficulty in understandably interpreting causal mechanism**. We introduce a method for ranking the most significant controls by summarizing the posterior for the full model.

Posterior summarization

We work with the framework of posterior summarization [2, 4] to **separate modeling and inference**.

(i) Fit a large, complex model $f(x, z)$ to account for complications in the regression model (confounders, nonlinearities, interactions, etc.).

(ii) Find a lower-dimensional summary model $\gamma(x, z)$ suitable for easier interpretation to explain the original model's predictions. This summary can either be user-defined, or one which minimizes a loss function.

(iii) Compute a posterior for the summary by projecting Monte Carlo draws of original model fit onto the predictive space of the summary.

Once the summary model is computed, it is determined whether it is adequately representative of the full model.

Toy example: summarizing predictive nonparametric regression

We simulate data from the model

$$y_i = f(x_1, x_2) + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

centered on the bivariate nonadditive function defined by

$$f(x_1, x_2) = 1/\{1 + \exp(-2x_1 - 2x_2)\} + 1/\{1 + \exp(-x_1 + 4x_2)\},$$

with $\sigma^2 = 0.25$. In Figure 1 we estimate the function with a Gaussian process with the squared-exponential kernel, and produce a linear summary $\gamma_1(x_1, x_2) = \alpha_1 + \beta_1 x_1 + \beta_2 x_2$, and an additive summary $\gamma_2(x_1, x_2) = \alpha_2 + f_1(x_1) + f_2(x_2)$ to estimate the partial effects of each variable, by *fitting these summary models to the output of the estimated regression function*. See [4] for more details.

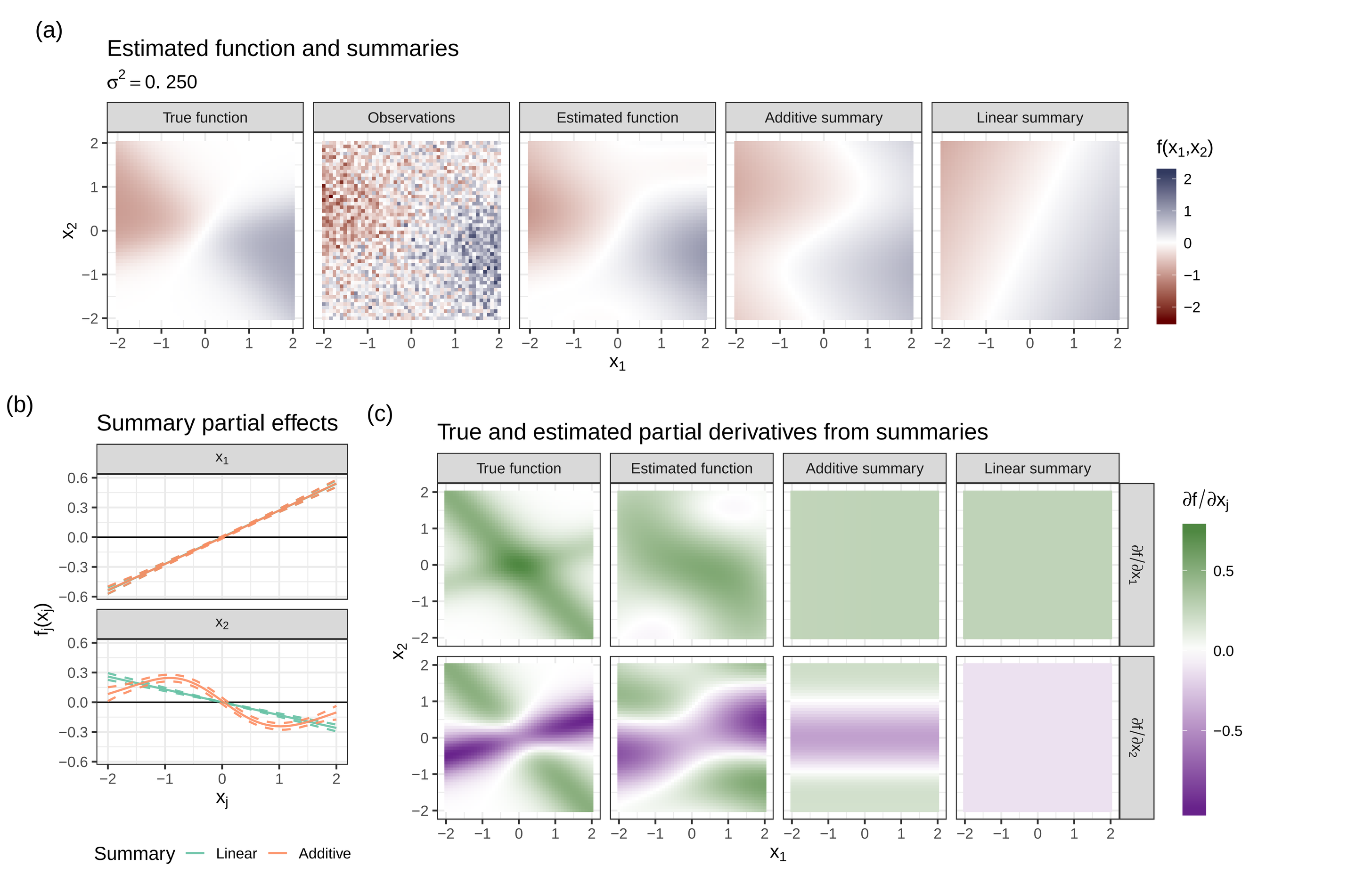

Figure 1: Toy example summarizing a nonparametric regression estimate with a linear and additive summary model.

Summarizing the causal linear model

Let $\mathcal{M}$ denote the full model in Eq. (1). Consider **the restricted model only including a subset** $\phi \subseteq \{1, \ldots, p\}$ of the controls X,

$$\begin{aligned} \text{Treatment equation:} &\quad Z = X_\phi \gamma_\phi + \epsilon \\ \text{Response equation:} &\quad Y = \alpha Z + X_\phi \beta_\phi + \nu \end{aligned} \tag{3}$$

and denote this model as $\mathcal{M}_\phi$. Let $D = [Z \;\; X]$ and $D_\phi = [Z \;\; X_\phi]$. Once we have computed the posterior for $\mathcal{M}$, we can compute the projected posterior for the summary model $\mathcal{M}_\phi$ using Monte Carlo draws $k = 1, \ldots, K$ from the original posterior,

$$\begin{aligned} \gamma_\phi^{(k)} &= (X_\phi^\intercal X_\phi)^{-1} X_\phi^\intercal X \gamma^{(k)} \\ \begin{bmatrix} \alpha \\ \beta_\phi \end{bmatrix}^{(k)} &= (D_\phi^\intercal D_\phi)^{-1} D_\phi^\intercal D \begin{bmatrix} \alpha \\ \beta \end{bmatrix}^{(k)} \end{aligned} \tag{4}$$

that is, *projecting each prediction from the full model onto the restricted column space*.

We judge the adequacy of the summary model by how much the estimated treatment effect is altered. Letting $\bar{\alpha}$ and $\bar{\alpha}_\phi$ be the posterior means of α from the full and restricted models, one possible loss function is

$$\mathcal{L}(\mathcal{M}, \mathcal{M}_\phi) = (\bar{\alpha} - \bar{\alpha}_\phi)^2. \tag{5}$$

If this difference is large, then the restricted model is "too simple," and there is induced confounding from summarizing too aggressively. If it is small, then this restricted model is a suitable summary for explaining the causal mechanism in the original model.

Application to crime data from Donohue III & Levitt (2001)

We apply our method to data from [1], measuring the effect of abortion rate Z on the per capita murder rate Y in each of the contiguous U.S. states, from 1985 to 1997, so there are $N = 624$ observations. In the original study, the authors considered 8 state-level control variables (unemployment, police level, etc.). After including state and year dummy variables variables, there are $p_1 = 66$ total control variables. Using OLS, they found a significant negative estimate of the treatment effect at the 5% level.

Following [3], we construct a much more elaborate model, allowing for *quadratic trends for the 8 per-state controls over time* (interacting the controls with year and year-squared), and *quadratic state-level trends over time* (interacting state dummy variables year and year-squared). This model with augmented controls includes $p_2 = 176$ total control variables.

In Figure 2, *we consider estimates of the treatment effect coming from three different models*:

- **Full model** with all augmented controls ($p_2 = 176$) using the prior from [3].
- **A summary model** using only the original ($p_1 = 66$) control variables, found by *projecting the augmented control model* using Eq. (4).
- **Refitted model** using prior from [3] with only original ($p_1 = 66$) controls.

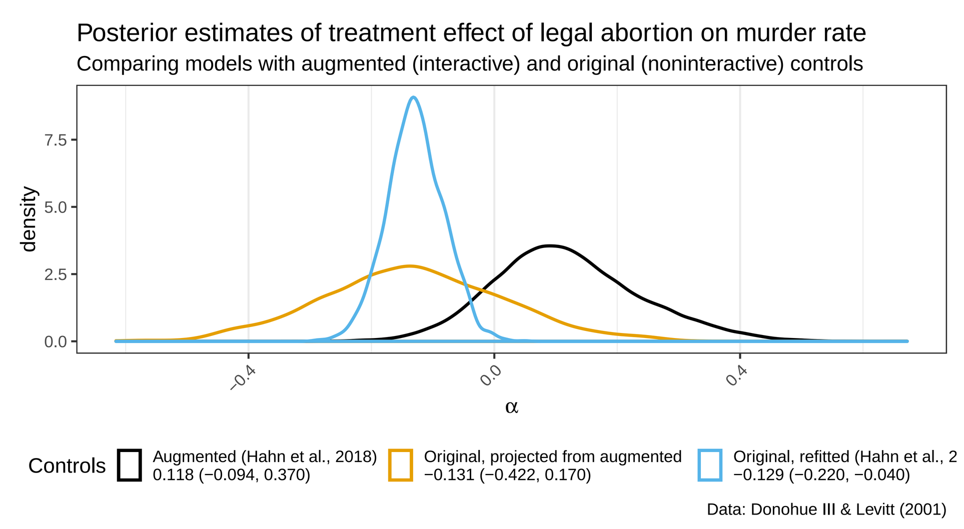

Figure 2: Estimates of causal effect of abortion rate on murder rate from considered methods.

The estimate of the treatment effect under the projected summary model is markedly different from that from the full model with all the augmented controls, implying that some of the omitted controls are significant confounders. However, this summary model still propagates uncertainty from the more complex model in a way that the refitted model does not.

Figure 3 shows an alternative approach. Here, we start with the full model with all augmented controls, and take a stepwise approach to create a sparser summary. *We iteratively remove one of the higher-order augmented controls at a time by finding the restricted model that minimizes the loss* (5), and compute the projected posterior of the treatment effect.

At first we see tighter posterior credible intervals. Eventually, however, moving from model size 132 to 131, the posterior variance suddenly surges, suggesting that we have summarized too far and removed an important prognostic variable which decreases the precision of the treatment effect estimate.

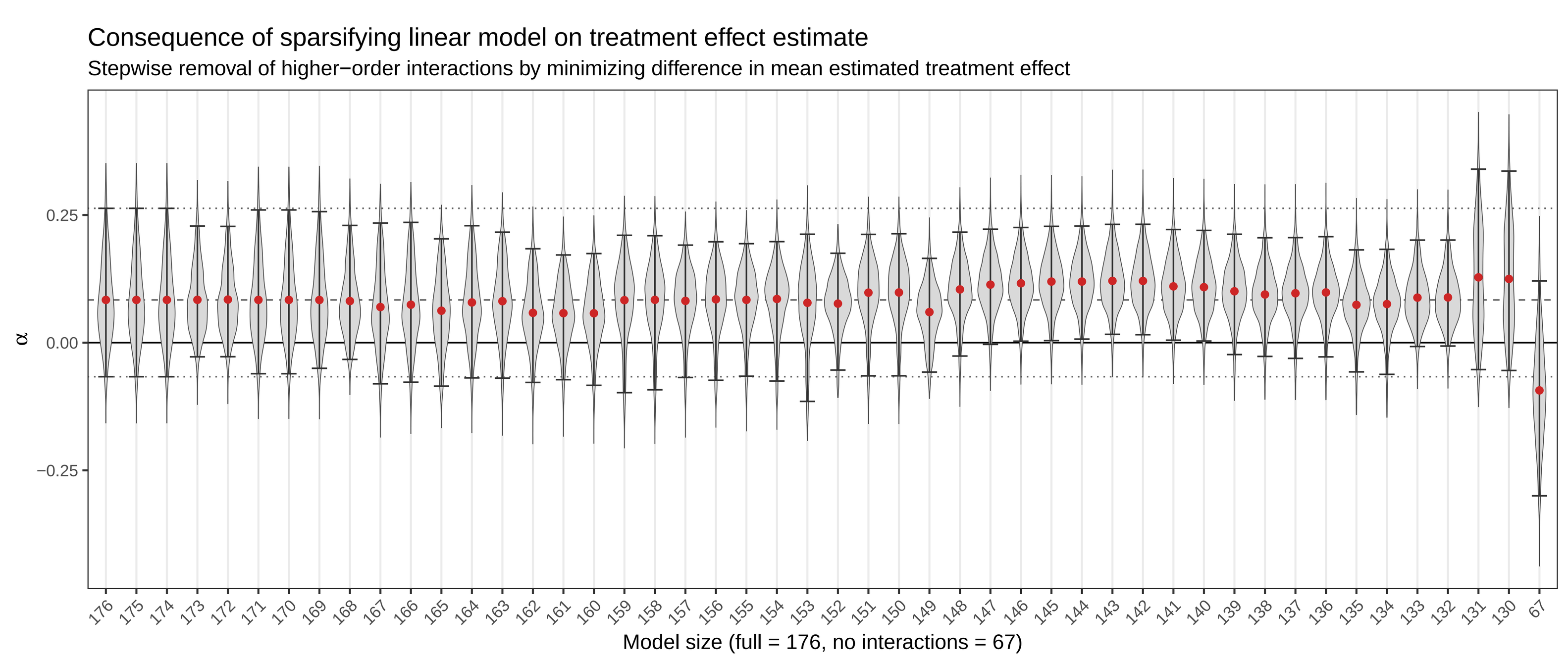

Figure 3: Stepwise summary model search, showing projected posteriors of the treatment effect.

Forthcoming Research

We plan to continue this line of work, defining a *summary search solution path* to minimizing the loss function (5) for a range of complexity of summary models. [2, 4] show how to do this in the setting of purely predictive modeling. When choosing which variables to exclude, there is a bias-variance trade-off; eliminating variables which are confounders induces bias in the treatment effect estimate, while eliminating prognostic variables increases variance of the treatment effect estimate.

Next, we plan to give *interpretable summaries of nonparametric models for heterogenious treatment effect estimation*. This is an extension of work on summarizing nonparametric regression models for prediction [4].

References

[1] John J. Donohue III and Steven D. Levitt. The Impact of Legalized Abortion on Crime. *The Quarterly Journal of Economics*, 116(2):379–420, 05 2001.

[2] P. Richard Hahn and Carlos M. Carvalho. Decoupling shrinkage and selection in Bayesian linear models: A posterior summary perspective. *Journal of the American Statistical Association*, 110(509):435–448, 2015.

[3] P. Richard Hahn, Carlos M. Carvalho, David Puelz, and Jingyu He. Regularization and confounding in linear regression for treatment effect estimation. *Bayesian Analysis*, 13(1):163–182, 03 2018.

[4] Spencer Woody, Calos M. Carvalho, and Jared S. Murray. Model interpretation through lower-dimensional posterior summarization. *arXiv preprint*, 2019.