American Political Science Review (2025) 119, 4, 2002–2010
doi:10.1017/S0003055424000935 © US Department of Defense and the Author(s), 2024. To the extent this is a work of the US Government, it is not subject to copyright protection within the United States. Published by Cambridge University Press on behalf of American Political Science Association. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Letter

Regularized Regression Can Reintroduce Backdoor Confounding: The Case of Mass Polarization

JONATHAN MELLON *West Point, United States*
CHRISTOPHER PROSSER *Royal Holloway, University of London, United Kingdom*

Regularization can improve statistical estimates made with highly correlated data. However, any regularization procedure embeds assumptions about the data generating process that can have counterintuitive consequences when those assumptions are untenable. We show that rather than simply shrinking estimates, regularization can reopen backdoor causal paths, inflating the estimates of some effects, and in the wrong circumstances, even reversing their direction. Recently, Cavari and Freedman (2023), argued that declining cooperation rates in surveys have inflated measures of mass polarization. We show that this finding is driven by large penalty terms in their regularized regressions, which leads to the estimates being confounded with time. Alternative methods do not show a clear positive or negative effect of declining cooperation on estimated levels of mass polarization.

INTRODUCTION

Regression models are ubiquitous in political science, but standard estimators can produce extreme and unstable inferences, particularly with highly correlated predictors (Gelman et al. 2008). These extreme results overinflate effect size estimates, especially when filtered on statistical significance. One solution is to apply regularization techniques, either through Bayesian priors or frequentist methods such as ridge regression or least absolute shrinkage and selection operator (LASSO). Regularization techniques have become popular in the last decade, with at least 102 articles in the top three political science journals (see Appendix R5 for details).

Regularization reduces the variance of estimates by deliberately biasing them toward zero. In repeated samples, unregularized estimators converge on the population value of a parameter, but individual estimates may be far from this value. By contrast, regularized estimators are biased in expectation (i.e., closer to zero than the population value), but there is less variation in estimates between different samples. The degree of bias is determined by making assumptions about the likely distribution of effects. Different methods make different assumptions. Faced with a set of correlated predictors, a researcher who thought there was likely to be a large number of small effects would choose ridge regression, but one who thought there was likely to be a small number of large effects—while other variables had no effect at all—would choose LASSO. Bayesian methods are more flexible, allowing researchers to specify ridge, LASSO, and elastic net equivalent priors, but also to use a range of other distributions (Carvalho, Polson, and Scott 2009). Used judiciously, regularization stabilizes estimates and improves out-of-sample predictions.

However, regularization has underappreciated consequences for making causal inferences: with strong regularization, as some coefficients are shrunk, the coefficients of correlated variables can be *inflated*. This happens because regularized coefficients do not account for all the variance associated with a variable, leaving residual variance open to be modeled by other variables. In other words, regularization partially reopens backdoor causal paths that were previously blocked by covariate adjustment.

In this letter, we first demonstrate how regularization can reopen backdoor casual paths. We then discuss this problem with respect to an *American Political Science Review* letter by Cavari and Freedman (2023, hereafter CF2), which employs regularized regression to argue that falling survey cooperation rates have exaggerated mass polarization in the United States. CF2 include a linear time control in their model to guard against spurious correlations due to trending variables. However, their use of regularization entails the assumption that the effect of time is small. We argue that *a priori* we should have strong expectations that any two (causally

Corresponding author: Jonathan Mellon, Associate Professor, Department of Systems Engineering, West Point, United States, jonathan.mellon@westpoint.edu
Christopher Prosser, Senior Lecturer, Department of Politics, International Relations and Philosophy, Royal Holloway, University of London, United Kingdom, chris.prosser@rhul.ac.uk

Received: August 22, 2022; revised: January 08, 2024; accepted: August 02, 2024. First published online: October 31, 2024.

https://doi.org/10.1017/S0003055424000935 Published online by Cambridge University Press

unrelated) variables measured over time are likely to be correlated with one another. Shrinking the time effect reopens the backdoor path between response rates and polarization via time. We reanalyze CF2's data in light of these problems and show it does not support the claim that declining response rates have inflated estimates of mass polarization.

REGULARIZED REGRESSION

Ordinary Least Squares (OLS) and regularized regression methods are both approaches for estimating the model $y = X\beta + \epsilon$. Frequentist regularized methods shrink coefficient estimates by adding a penalty term to the OLS cost function, but differ in how they do so. We examine three common frequentist regularization methods: ridge regression, LASSO, and elastic net. We also contextualize these methods against the Bayesian approach of specifying explicit priors.

Ridge regression minimizes the cost function:

$$\sum_{i=1}^{M}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}\beta_j^2. \quad (1)$$

The ridge cost function is the OLS cost function, which minimizes the sum of squared residuals, plus an additional penalty term $\lambda\sum_{j=1}^{p}\beta_j^2$. Regularization penalties shrink coefficients because any reduction in the sum of squared residuals produced by larger coefficients is partially offset by an increase in the size of the penalty term. For example, while a model with two coefficients 2 and 1 might result in a lower sum of squares compared to a model with coefficients 1.9 and 0.9, the latter has a smaller penalty term, and with a sufficiently large λ will have a smaller overall sum. Importantly, the size of the ridge penalty is proportional to the square of the coefficients, so if coefficient β_1 is twice the magnitude of coefficient β_2, β_1 contributes four times as large a penalty as β_2. Consequently, for a marginal increase in the sum of squared residuals, ridge regression prefers two coefficients of similar sizes to one large and one small coefficient.

LASSO minimizes the cost function:

$$\sum_{i=1}^{M}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j|. \quad (2)$$

The difference between Equation 2 and Equation 1 being that the LASSO loss function regularizes coefficients proportional to their absolute size rather than their square, and so does not produce the effect of ridge regression of preferring many small coefficients to a smaller set of larger coefficients.

The elastic net incorporates both of these penalty terms, adding an additional term α, which controls the relative weight between the two. When $\alpha = 1$, the elastic net is equivalent to LASSO, and when $\alpha = 0$, it is equivalent to ridge regression:

$$\sum_{i=1}^{M}\left(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij}\right)^2 + \lambda\sum_{j=1}^{p}(\alpha|\beta_j| + (1-\alpha)\beta_j^2). \quad (3)$$

These penalty terms encode assumptions about the likely size of the β_j effects, which can be interpreted as Bayesian priors (Hastie, Tibshirani, and Friedman 2009). In Bayesian terms, the ridge regression penalty expects that β_j are normally distributed around zero, which implies that coefficients are increasingly unlikely the further away from zero they are. As a Bayesian prior the LASSO penalty follows a Laplace distribution, which is sharply peaked at zero, entailing the expectation that large values of β_j are unlikely, but furthermore that many effects are exactly zero. The elastic net prior is a mixture of these two distributions. Larger λs imply stronger versions of these assumptions, or in Bayesian terms, prior distributions with smaller standard deviations and/or scale parameters.

Choosing λ

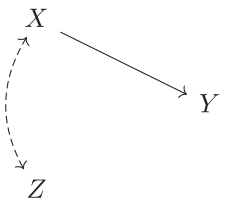
Given the role of the penalty term in regularized estimators, λ must be chosen with care. The most common approach is to choose λ based on out-of-sample predictive performance, typically using cross-validation (CV). For ridge regression, an alternative approach is based on a transformation of sample variance and covariance. CF2 take this latter approach, using the "KM4" method (Muniz and Kibria 2009). Different methods for choosing λ can result in radically different values. In our simulations below, KM4 tends to choose considerably higher λs (*mean* = 4.1) compared to CV (*mean* = 0.9), and so a KM4 approach to choosing λ would result in much stronger regularization.

Thinking about penalty terms as being equivalent to Bayesian priors suggests an alternative approach. In Bayesian statistics, a common choice is to use weakly informative priors, such as N(0, 1) when all variables are standardized to have mean 0 and SD 1 (Betancourt 2017). For a regression where all variables are unit-scaled, a ridge regression with penalty λ is equivalent to a Bayesian estimate with a normally distributed prior with mean 0 and standard deviation $\sqrt{\frac{Var(\hat{\epsilon})}{\lambda}}$ (Hastie, Tibshirani, and Friedman 2009, 64). In practice, the conversion between λ and priors can be further complicated by rescaling procedures applied by particular software, which we discuss in Section S5 of the Supplementary Material.

Regularized Regressions Can Reopen Backdoor Causal Paths

In order to demonstrate the counter-intuitive effects that regularization can have, we run simulations in which X causes Y, and X and Z are correlated, but Z has no causal effect on Y, illustrated in the directed acyclic graph (DAG) (4) below.

https://doi.org/10.1017/S0003055424000935 Published online by Cambridge University Press



(4)

Where X and Z are multivariate standard normal variables with a correlation ρ, and Y is a linear function of $\beta_X X$ and a standard normal error term (in CF2's data, if the response rate effect on polarization was actually zero, then X would be equivalent to time, Z to response rate, and Y to polarization):

$$\begin{aligned} \begin{bmatrix} X \\ Z \end{bmatrix} &\sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right), \\ Y &\sim \beta_X X + \mathcal{N}(0,1). \end{aligned} \tag{5}$$

From this DGP, we simulate ten thousand datasets with $\beta_X \sim U(0, 1)$ and $\rho = -0.9$ (approximately the correlation between time and response rate in CF2's data).

We then fit the following model to the simulated data:

$$y_i = \beta_0 + \beta_X X_i + \beta_Z Z_i + \epsilon_i \tag{6}$$

using four estimators: OLS, ridge regression, LASSO, and elastic net (with $\alpha = 0.5$, an equal mix of the two penalties). For each regularized estimator, we set four λ penalties: 0.00615, 0.154, 0.615, and 2.46, which correspond approximately to a weakly informative prior for the average simulated case, and priors that are 20%, 10%, and 5% the width of the standard deviation of that weakly informative prior for ridge regression.

Figure 1 shows the average coefficient estimates for β_X and β_Z for each estimator, which illustrates the interaction between effect size and λ. We first consider bias (the average signed difference between the estimate and simulated coefficient). The OLS estimate is unbiased for both β_X and β_Z across the range of the simulated β_X terms. With a penalty equivalent to a weakly informative prior ($\lambda = 0.00615$, row A), the regularized estimates are only marginally biased. With a larger penalty ($\lambda = 0.154$, row B), the bias in the estimated β_x is substantially larger. For ridge regression, this bias increases linearly as the simulated β_X increases. For LASSO (and to a lesser extent elastic net), the level of bias flattens off with sufficiently large β_X values. As λ gets larger, the bias in the β_X estimates also increases (rows C and D). These biases are strongest for LASSO—which with the highest penalty almost always shrinks the estimates to zero—and smallest for ridge regression, with elastic net in between.

That regularized regression induces bias and shrinks β_X toward zero is well understood—this is the price these estimators pay for reducing variance. What is less appreciated is that regularized estimators also bias β_Z *away* from zero. This occurs because some of the $X \rightarrow Y$ effect is left unaccounted for by the biased β_X estimate, and so is transferred to Z via the $Z \leftrightarrow X \rightarrow Y$ backdoor path (for a demonstration, see Section S3 of the Supplementary Material). The β_Z bias is most pronounced for ridge regression: as the λ penalty increases, the bias in β_Z initially *increases* (compare the ridge regression lines in rows B/C). As λ increases further, the ridge β_Z estimates are themselves regularized, and move back toward zero. Although this means the *absolute* bias in β_Z is reduced with the largest λ terms, the effect of β_Z *relative* to β_X continues to increase, with estimators with large penalties assigning almost equal values to β_X and β_Z. For LASSO, the bias in β_Z is only noticeable at the smaller levels of λ (i.e., row B); with higher penalties β_Z is almost always pulled to zero. Elastic net is again between the ridge and LASSO estimates.

Making Statistical Inferences with Regularized Regressions

To make inferences about underlying population parameters (e.g., the true effect of response rate on estimates of mass polarization), we typically use estimates of sampling variance. With unbiased estimators this is straightforward; in repeated samples, we would expect 95% of our confidence intervals (CIs) to overlap the population value. With regularized estimators, this is more complicated because (1) the sampling variance of some regularized estimators is poorly defined, and (2) even where it is well defined (such as for ridge regression) the resulting CIs may have no overlap with the underlying population parameter.

A particular challenge for the LASSO and elastic net estimators is that "we still do not have a general, statistically valid method of obtaining standard errors of LASSO estimates" (Kyung et al. 2010, 377) even using bootstrap methods. LASSO and elastic net are therefore poorly suited to inference.

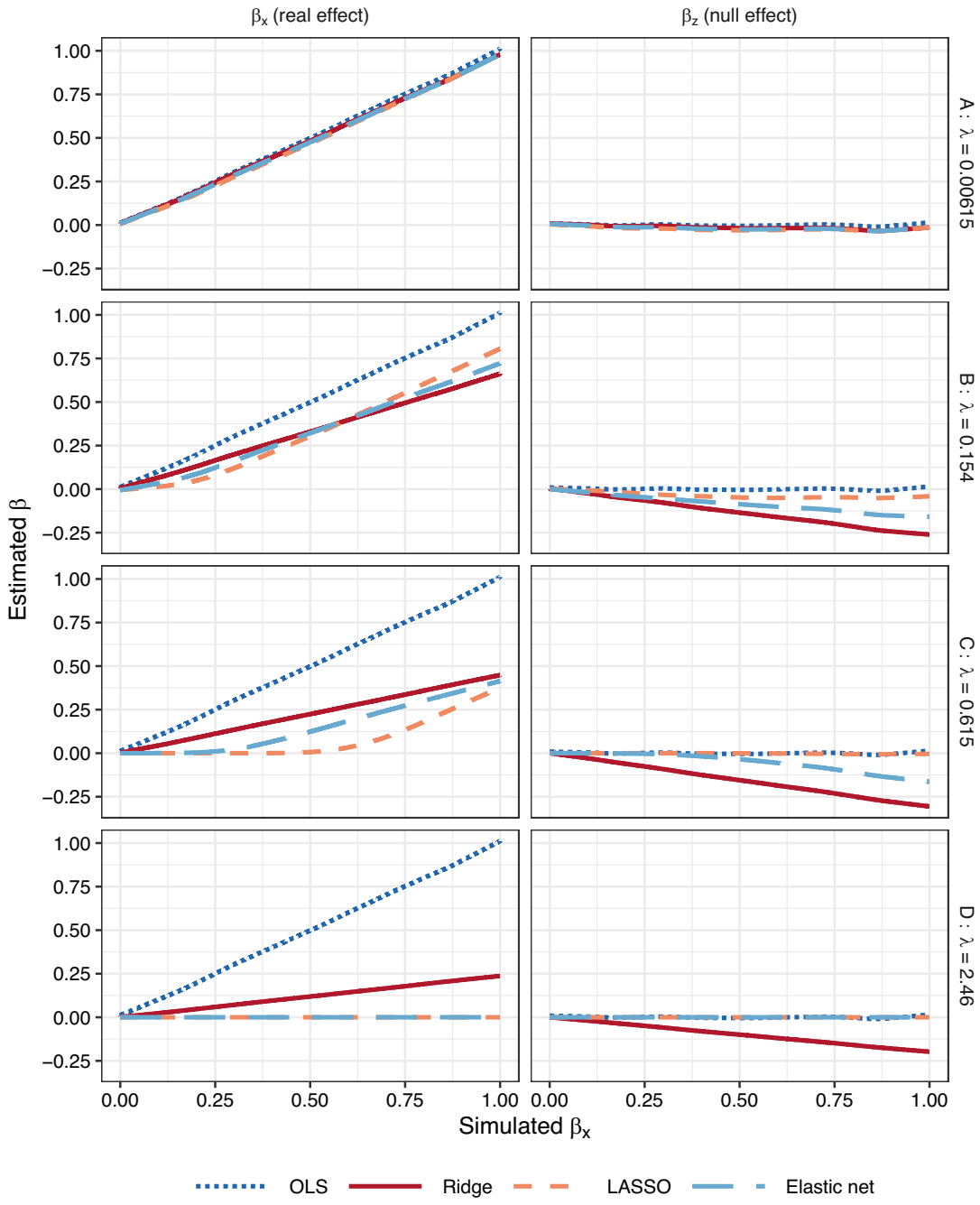
This problem does not apply to ridge regression, where the variance of ridge estimates is

$$Var(\hat{\beta}_{ridge}|X) = \sigma^2 (X^T X + \lambda I)^{-1} X^T X (X^T X + \lambda I)^{-1}. \tag{7}$$

We must be clear what this variance captures: how much a ridge estimate $\hat{\beta}_{ridge}$ is expected to differ across repeated samples (when $\lambda = 0$, this is the same as OLS). By design, the λ penalty decreases variance. However, λ also *increases bias* (indeed, it reduces variance *by* increasing bias). This has the pernicious effect of shrinking the ridge CIs at the same time as increasing the distance between the ridge estimates and the

https://doi.org/10.1017/S0003055424000935 Published online by Cambridge University Press

FIGURE 1. Estimated OLS, Ridge, LASSO, and Elastic Net Coefficients for Simulated Values of β_X and λ (Local Linear Smooths)



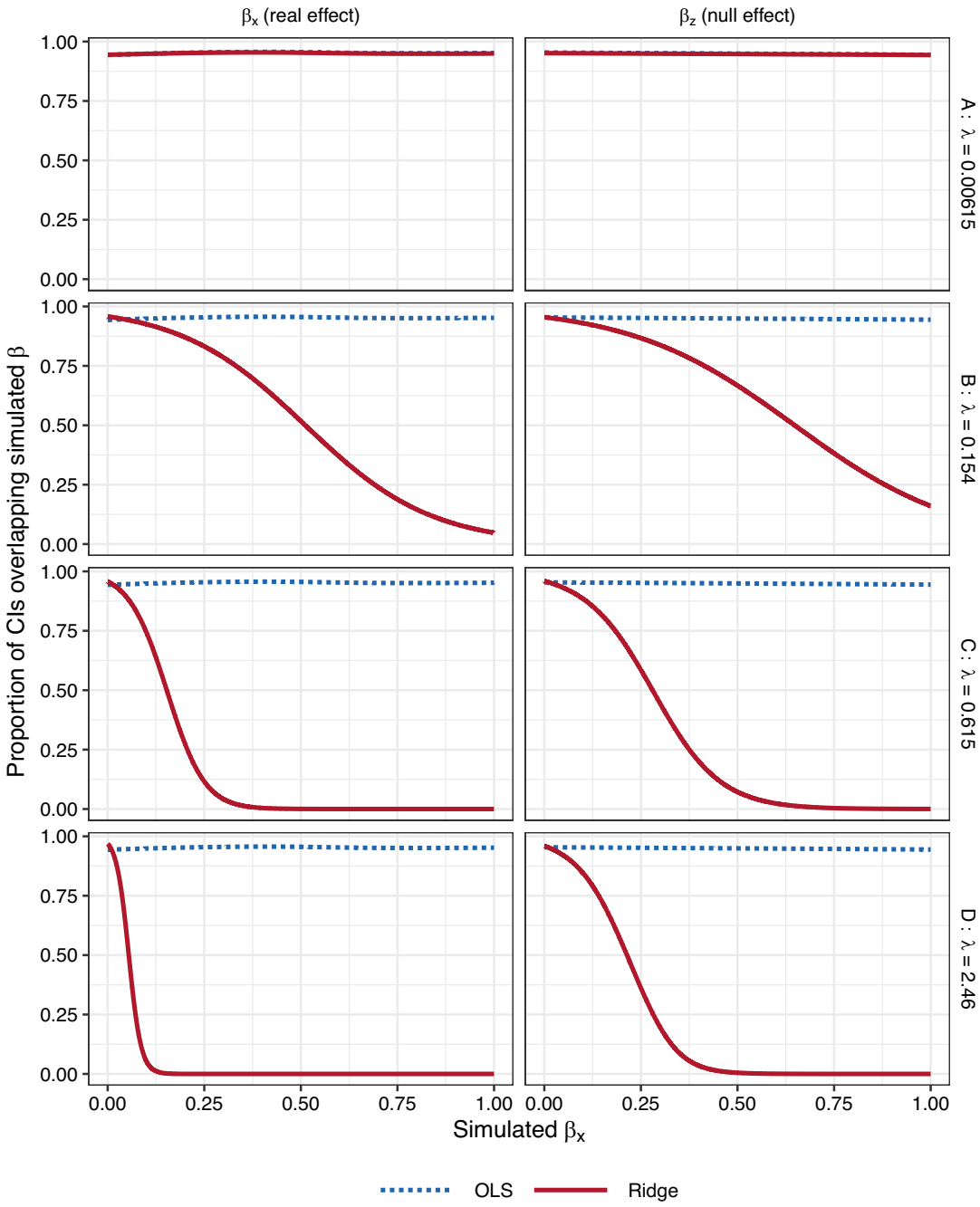
Note: β_Z always simulated as zero.

underlying parameter of interest. As bias grows, CI coverage falls because more of the discrepancy between $\hat{\beta}_{ridge}$ and β is due to bias (which the CIs do not account for) rather than sampling variation (which the CIs do account for). If the bias is large enough, CIs routinely exclude the true value of β. Narrow ridge regression CIs do not necessarily indicate a precise estimate of the parameter of interest, but simply that ridge regression would produce similar estimates with new samples. The resulting CIs may lead to misleading inferences about population parameters.

We can again demonstrate this problem using simulations. Returning to the simulated data we used in Figure 1, we calculate CIs for OLS and ridge regression. For each simulation, we record whether the CI overlaps the simulated parameters for β_X and β_Z,

https://doi.org/10.1017/S0003055424000935 Published online by Cambridge University Press

FIGURE 2. Estimated OLS and Ridge CI Coverage for Simulated Values of β_X and λ (Local Logistic Smooths)



Note: β_Z always simulated as zero.

illustrated in Figure 2. The OLS CIs have the appropriate coverage (95%) across all simulated values of β_X. For the ridge CIs, however, we again see an interaction between λ and the size of β_X. When λ is small and β_X is small, the CIs still have coverage close to the 95% benchmark. As β_X gets larger the coverage rate begins to fall for both β_X and β_Z. This problem is exacerbated by larger values of λ—the coverage rate falls precipitously and approaches zero. At the highest levels of λ and β_X, this means that the CIs for β_X and β_Z *never* overlap the population values for these parameters. In all but the most favorable of circumstances, ridge regression is simply not an appropriate tool for statistical inference.

https://doi.org/10.1017/S0003055424000935 Published online by Cambridge University Press

AN EXAMPLE OF RIDGE REGRESSION ESTIMATES IMPLYING UNREALISTIC PRIORS ABOUT THE EFFECT OF TIME

In light of the issues we discuss above, we turn to CF2's analysis. CF2 analyze the relationship between survey response/cooperation rate and partisan polarization for survey questions asked by Pew between 2004 and 2018. They operationalize polarization as the absolute Cohen's D of mean differences between Republican and Democrat identifiers' responses. They estimate models for six issue areas: economy, energy, immigration, civil rights, welfare, and foreign policy, specified as

$$y_{it} = \beta_0 + \beta_1 rr_{it} + \beta_2 congress_t + \beta_3 year_t + \epsilon_{it}, \quad (8)$$

where rr_{it} is survey i's response rate at time t, $congress_t$ is congressional polarization at time t, and $year_t$ is the year in which the survey was conducted. For the contact and cooperation rate versions of these models, the $\beta_1 rr_{it}$ term is replaced with $\beta_4 contact_{it}$ and $\beta_5 cooperation_{it}$. Equation 8, which adds a year term to an earlier model estimated by Cavari and Freedman (2018, henceforth CF1) was suggested by Mellon and Prosser (2021, hereafter MP) in a critique of CF1 as one of several ways they could deal with the problem of time trends.[1] CF2 estimate their models using ridge regression (unlike MP, who use OLS) with λ chosen by the KM4 method. CF2 reports statistically significant negative effects of response and cooperation rates (i.e., lower response rates increase the estimated level of polarization) for three areas: economy, energy, and immigration.

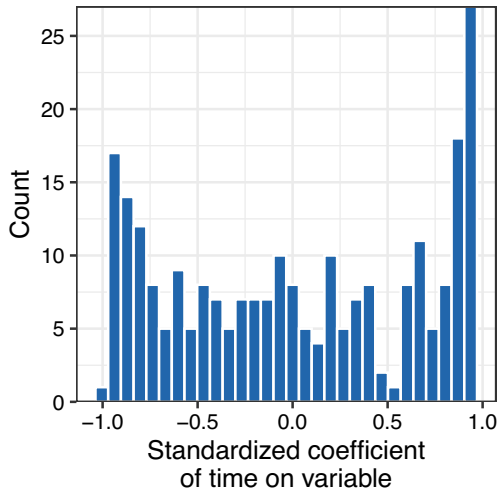
CF2's KM4 ridge regression estimates imply strong priors that are between 1.5% and 5.9% the width of a weakly informative prior. Are these priors plausible when considering time trends? *A priori*, there are good reasons to think not. Many social processes can be modeled as random walks or similar processes. Independent random walks will tend to correlate with time (Granger and Newbold 1974) because their variance increases as a function of time (see Section S2 of the Supplementary Material).

Moreover, we can show that many social science variables are correlated with time in practice. We calculate the annual mean of binary, ordinal and continuous variables asked in 15 or more years of the GSS (Smith et al. 2019) and estimate a linear model of each variable on time using OLS. Figure 3 shows a histogram of the standardized coefficients for the effect of time on each variable. The distribution has a far higher dispersion than CF2's priors imply, with a standard deviation of 0.657 and heavy tails close to 1 and −1.

To show how different these priors are from CF2's, we can consider how likely we are to observe a 0.5 or larger magnitude unit-scale coefficient. The *weakest* of CF2's priors implies the probability of observing a 0.5 or larger unit-scale coefficient (considerably smaller than the mean OLS-year estimate on their data) is 2.8×10^{-17}. By contrast, a N(0, 0.657) prior gives a probability of 0.447 of observing an effect size magnitude of 0.5 or larger.

CF2's narrow priors mean they strongly shrink the year coefficient in their models, which risks reopening the backdoor causal paths that the inclusion of a time control is designed to block.

FIGURE 3. Two Hundred Forty Nine Standardized Coefficients of Time on GSS Variable Annual Averages



AN EXAMPLE OF HOW A LARGE SHRINKAGE PARAMETER AFFECTS EMPIRICAL RESULTS

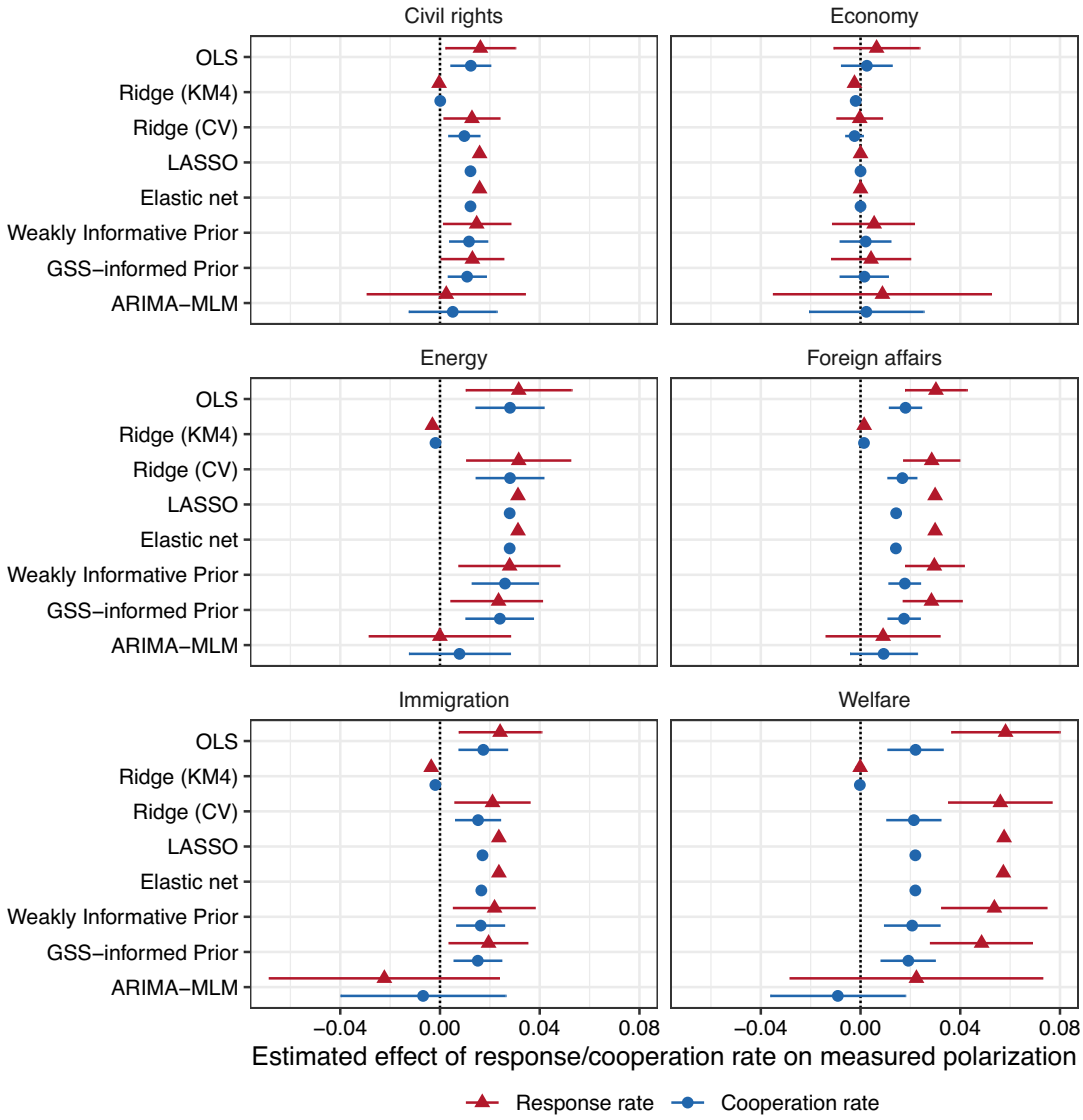
To demonstrate the impact of CF2's regularization approach, we reestimate their models using a variety of approaches: their original KM4 ridge regressions, OLS, ridge regression with λ chosen by CV, a weakly informative N(0, 1) unit-scaled prior, a prior informed by the GSS distribution of time effect sizes N (0, 0.657), LASSO with λ chosen by CV, and elastic net ($\alpha = 0.5$) with λ again chosen by CV. These estimators show the impact of different regularization approaches, but they all share a common approach to controlling for time (a linear trend), which may not fully account for time-series problems. Given this, we benchmark them against a more sophisticated repeated cross-section model suggested by Lebo and Weber (2015) (ARIMA-MLM—we elaborate on this approach in Section S1 of the Supplementary Material). These estimates for the key response and cooperation rate variables are shown in Figure 4.[2]

[1] Section S1 of the Supplementary Material examines the other MP models.

[2] CF2 report scaled coefficients. We report unscaled coefficients to facilitate comparison between estimators.

https://doi.org/10.1017/S0003055424000935 Published online by Cambridge University Press

FIGURE 4. Estimated Effects of Response and Cooperation Rates from Replication of CF2's Analysis Using Six Estimators: OLS, Ridge Regression (λ Chosen Using KM4), Ridge Regression (λ Chosen by CV), LASSO (λ Chosen by CV), Elastic Net (λ Chosen by CV), Bayesian Regression (Weakly Informative Priors), Bayesian Regression (Informed Priors Based on the GSS Distribution of Time Trends), and ARIMA-MLM



Note: CIs omitted from LASSO and elastic net estimates for the reasons discussed in Section "Making Statistical Inferences with Regularized Regressions." CIs for the ridge (KM4) estimates are drawn, but are smaller than the size of point estimate marker. See Appendix R6 for full tables.

The results in Figure 4 present a striking pattern—six of the seven estimators of the linear-time-trend model produce similar results, and one does not. With the exception of the economy issue area, CF2's KM4 results are considerably closer to zero than either the OLS estimate or the other regularized estimates. All of the unregularized estimates, and nearly all of the other regularized estimates apart from KM4, have estimates in the *opposite* direction to those reported by CF2. Those estimates that CF2 report as negative have *swapped directions* compared to their unregularized equivalents —the same effect we demonstrated earlier in our simulations—and are driven not by the data, but by high λ penalties.[3] Only in one other case—ridge regression with λ chosen by CV for the economy—are the estimates of response and cooperation rate effects negative. Even these estimates do not support an effect of response rates on mass polarization: the CIs for both overlap zero.

Appendix R3 presents further simulations that examine the likelihood of false positives using CF2's approach if the true effect of response/cooperation

[3] Appendix R1 analyses CF2's data across a range of λ ridge penalties.

https://doi.org/10.1017/S0003055424000935 Published online by Cambridge University Press

rates was zero. The results are stark. We observe false positives more than 98% of the time for the effect of response/cooperation rates for energy, immigration, and foreign policy and more than 83% for the economy. In other words, CF2's results are what we would expect to see if there were no true effect of response/cooperation rate on mass polarization.

Should we conclude that the effects of response/cooperation rates on mass polarization are null or even positive (i.e., the opposite of CF2's claims)? While there are plausible theoretical models that lower response rates could *de*flate measures of mass polarization (see Appendix R4), the evidence is insufficient to make that claim. While some OLS estimates of the linear-time-trend model suggest positive effects, the ARIMA-MLM approach (and several other alternate model specifications) never find significant effects in either direction, suggesting that the linear-time-trend specification does not fully account for structural properties such as autocorrelation. We believe that ARIMA-MLM is a better model specification for removing spurious time effects, so prioritize its result over the linear specification used by CF2.

We also refrain from claiming a precise null effect. As Section S4 of the Supplementary Material shows, the data are underpowered even if the linear time trend specification is correct. We suspect that matched individual-level and administrative data (e.g., Clinton, Lapinski, and Trussler 2022) may be a more fruitful avenue for exploring this question. Our claim is merely that the apparent evidence of an effect of response rates on mass polarization is driven by an inappropriate estimator and that the data are insufficient to test for effects of the claimed size.

CONCLUSIONS

The problem of correlated predictors is a real concern for scholars. Regularization through weakly informative priors or equivalent frequentist regularized estimators may be helpful for reducing variance in these cases with only a limited increase in bias. However, we urge scholars to avoid using data-driven regularization procedures without assessing their substantive implications. With overly strong penalty terms, regularization can reintroduce confounding through backdoor causal paths that are blocked by covariate adjustment with an unregularized estimator.[4] We recommend either directly using Bayesian priors, or translating the implicit assumptions of frequentist regularized regressions into priors so that their plausibility and impact can be assessed.

There are also broader lessons that can be taken from our analysis. First, correlation with time is a serious threat to valid statistical inference. Many—perhaps *most*—social science variables are correlated with time. Any statistical model that uses longitudinal data needs to account for time. Second, social scientists frequently use complex methods to try to extract the maximum available information from limited data, which often entail opaque assumptions about the world. These assumptions should be made explicit and justified. Third, simulation should be part of any social scientist's toolkit. Simulations make clear the impact that different modeling choice can have on our estimates and require little expertise beyond that required to estimate the same statistical models on observed data.

[4] Section S6 of the Supplementary Material discusses how using regularized methods such as LASSO for variable selection reintroduces backdoor confounding for slightly different reasons.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit https://doi.org/10.1017/S0003055424000935. Further supplementary material (indicated with by appendices labeled with R) can be found with the replication material.

DATA AVAILABILITY STATEMENT

Research documentation and data that support the findings of this study are openly available at the American Political Science Review Dataverse: https://doi.org/10.7910/DVN/VEFZXI.

ACKNOWLEDGMENTS

The views expressed herein are those of the authors and do not reflect the position of the United States Military Academy, the Department of the Army, or the Department of Defense. Thank you to Jack Bailey and Rex Douglas for their insightful comments on earlier drafts of this paper.

CONFLICT OF INTEREST

The authors declare no ethical issues or conflicts of interest in this research.

ETHICAL STANDARDS

The authors affirm that this research did not involve human participants.

REFERENCES

Betancourt, Michael. 2017. "How the Shape of a Weakly Informative Prior Affects Inferences." https://mc-stan.org/users/documentation/case-studies/weakly_informative_shapes.html.

Carvalho, Carlos M., Nicholas G. Polson, and James G. Scott. 2009. "Handling Sparsity via the Horseshoe." *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics in Proceedings of Machine Learning Research* 5: 73–80.

Cavari, Amnon, and Guy Freedman. 2018. "Polarized Mass or Polarized Few? Assessing the Parallel Rise of Survey Nonresponse and Measures of Polarization." *Journal of Politics* 80 (2): 719–25.

https://doi.org/10.1017/S0003055424000935 Published online by Cambridge University Press

Cavari, Amnon, and Guy Freedman. 2023. "Survey Nonresponse and Mass Polarization: The Consequences of Declining Contact and Cooperation Rates." *American Political Science Review* 117 (1): 332–9.
Clinton, Joshua, John Lapinski, and Marc Trussler. 2022. "Reluctant Republicans, Eager Democrats?" *Public Opinion Quarterly* 86 (2): 247–69.
Gelman, Andrew, Aleks Jakulin, Maria Pittau, and Yu-Sung Su. 2008. "A Weakly Informative Default Prior Distribution for Logistic and Other Regression Models." *The Annals of Applied Statistics* 2 (4): 1360–83.
Granger, C. W. J., and P. Newbold. 1974. "Spurious Regressions in Econometrics." *Journal of Econometrics* 2 (2): 111–20.
Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning*. New York: Springer.
Kyung, Minjung, Jeff Gill, Malay Ghosh, and George Casella. 2010. "Penalized Regression, Standard Errors, and Bayesian Lassos." *Bayesian Analysis* 5 (2): 369–411.
Lebo, Matthew, and Christopher Weber. 2015. "An Effective Approach to the Repeated Cross-Sectional Design." *American Journal of Political Science* 59 (1): 242–58.
Mellon, Jonathan, and Christopher Prosser. 2021. "Correlation with Time Explains the Relationship between Survey Nonresponse and Mass Polarization." *Journal of Politics* 83 (1): 390–5.
Mellon, Jonathan, and Christopher Prosser. 2024. "Replication Data for: Regularized Regression Can Reintroduce Backdoor Confounding: The Case of Mass Polarization." Harvard Dataverse. Dataset. https://doi.org/10.7910/DVN/VEFZXI.
Muniz, Gisela, and B. M. Golam Kibria. 2009. "On Some Ridge Regression Estimators: An Empirical Comparisons." *Communications in Statistics - Simulation and Computation* 38 (3): 621–30.
Smith, Tom W., Michael Davern, Jeremy Freese, and Stephen L. Morgan. 2019. "General Social Surveys, 1972–2018." National Opinion Research Center.

https://doi.org/10.1017/S0003055424000935 Published online by Cambridge University Press