PA

The Consequences of Model Misspecification for the Estimation of Nonlinear Interaction Effects

Janina Beiser-McGrath[1] and Liam F. Beiser-McGrath [ID][2]

[1] Lecturer (Assistant Professor) in Politics and International Relations (Quantitative Methods), Department of Politics, International Relations, and Philosophy, Royal Holloway, University of London, Egham, UK. E-mail: janina.beiser-mcgrath@rhul.ac.uk

[2] Assistant Professor in International Social and Public Policy, Department of Social Policy, London School of Economics and Political Science, London, UK. E-mail: liam@liambeisermcgrath.com

Abstract

Recent research has shown that interaction effects may often be nonlinear (Hainmueller, Mummolo, and Xu [2019, *Political Analysis* 27, 163–192]). As standard interaction effect specifications assume a linear interaction effect, that is, the moderator conditions the effect at a constant rate, this can lead to bias. However, allowing nonlinear interaction effects, without accounting for other nonlinearities and nonlinear interaction effects, can also lead to biased estimates. Specifically, researchers can infer nonlinear interaction effects, even though the true interaction effect is linear, when variables used for covariate adjustment that are correlated with the moderator have a nonlinear effect upon the outcome of interest. We illustrate this bias with simulations and show how diagnostic tools recommended in the literature are unable to uncover the issue. We show how using the adaptive Lasso to identify relevant nonlinearities among variables used for covariate adjustment can avoid this issue. Moreover, the use of regularized estimators, which allow for a fuller set of nonlinearities, both independent and interactive, is more generally shown to avoid this bias and more general forms of omitted interaction bias.

Keywords: Interactions, Non-linear, regression, machine learning, misspecification, lasso

1 Introduction

Many relationships that political scientists are interested in are conditional. Political scientists frequently specify models to test whether the effect of a variable of interest, the treatment, is conditional upon a moderating variable. Starting with Brambor, Clark, and Golder (2006), a number of publications have focused on improving practice and inference in the use of interaction models (e.g., Beiser-McGrath and Beiser-McGrath 2020; Berry, DeMeritt, and Esarey 2010, 2016; Blackwell and Olson 2021; Esarey and Sumner 2018; Hainmueller, Mummolo, and Xu 2019).

Hainmueller *et al.* (2019) identify a key issue stemming from the uncritical application of the assumption that the marginal effect of the treatment linearly changes with changing values of the moderating variable (they refer to this as the linear interaction effect [LIE] assumption). When this assumption does not hold, inferences based on a linear interaction model will be biased. Hainmueller *et al.* (2019) introduce a binning estimator and a kernel estimator that relax the LIE assumption and allow researchers to uncover nonlinear interactive relationships. The authors suggest two ways of diagnosing when the LIE assumption does not hold and scholars should use one of those estimators instead of a linear interaction model: (a) a Wald test based on the comparison of the binning estimator and a linear interaction specification and (b) graphical methods investigating the raw data.

However, the methods Hainmueller *et al.* (2019) propose for diagnosing violations of the LIE assumption do not prevent other forms of bias to enter. In particular, the binning estimator for modeling nonlinear interactions can pick up unmodeled nonlinearities in control variables which are correlated with the moderator. When relevant nonlinear terms of control variables are omitted

from the model, truly linear interactions can be misidentified as nonlinear. Moreover, graphical diagnostic tools are of limited use for uncovering the true relationship in these cases.

In this letter, we demonstrate this problem using simulations. The simulations show that the Wald test based on the binning estimator misdiagnoses violations of the LIE assumption when relevant quadratic terms of control variables are omitted. Moreover, we demonstrate the inability of graphical diagnostics to detect this problem.

The results suggest that while moving beyond linearity as a default in the estimation of interaction effects can expand our knowledge of political processes, it needs to be coupled with a broader consideration of nonlinearities, both independent and interactive, in model specification (Beiser-McGrath and Beiser-McGrath 2020). Allowing for a nonlinear interaction effect, without accounting for other potential nonlinearities, leaves this interaction effect open to absorbing these unmodeled effects, thus causing bias.

We propose two approaches to assess the robustness of a specified nonlinear interaction effect to misspecification bias. First, we suggest using methods for variable selection to identify nonlinearities and interactions among the variables used for covariate adjustment (Z) that have the potential to cause bias in the Hainmueller *et al.* (2019) binning estimator, and thus should be included. Second, we propose more general machine learning methods that allow for a full set of nonlinearities, interactions, and nonlinear interactions while avoiding overfitting, in order to minimize this problem (Beiser-McGrath and Beiser-McGrath 2020; Blackwell and Olson 2021; Hainmueller and Hazlett 2014; Kenkel and Signorino 2013).

2 The Problem

In this section, we show how the approach to modeling nonlinear interaction effects proposed by Hainmueller *et al.* (2019) can suffer from misspecification bias. For diagnosis and as one option for analysis, Hainmueller *et al.* (2019) introduce a binning estimator[1] where G_J are dummies associated with different bins across the range of X

$$y = \sum_{j=1}^{3} (\mu_j + \alpha_j D + \eta_j(X - x_j) + \beta_j(X - x_j)D)G_j + \gamma Z + \epsilon. \tag{1}$$

This specification allows for the relationship between y and a treatment D to vary nonlinearly with the values of a moderator X. Moreover, the `interflex` R package that implements this estimator allows researchers to specify whether to use a functional form referred to as fully moderated that interacts all control variables with the moderator X (but not the treatment D). Importantly, however, the binning estimator, fully moderated or not, does not account for nonlinearities in variables used for covariate adjustment (Z). This is problematic as unmodeled nonlinear functions of Z can be picked up by the more flexible functional form allowed for X when moderating D, thus potentially incorrectly implying that X moderates D nonlinearly.

2.1 Simulation Evidence

We show the potential for this to occur with a Monte Carlo simulation. In our data generating process, the true effect of D is conditional on X with a linear functional form (i.e., a linear and not a nonlinear interaction effect), that is,

$$y = -0.5 \times D + 0.5 \times X + 0.5 \times Z - 1.5 \times Z^2 + 0.5 \times DX + \epsilon. \tag{2}$$

We draw X and Z jointly from a multivariate normal distribution with mean zero, variance one, and covariance $\rho = 0.5$ and D and ϵ from a standard normal distribution.

---

1 We focus on the binning estimator due to its importance in diagnosing violations of the LIE assumption and predominant use by applied researchers. As we show in our simulated example, the kernel estimator is also not immune to these issues.
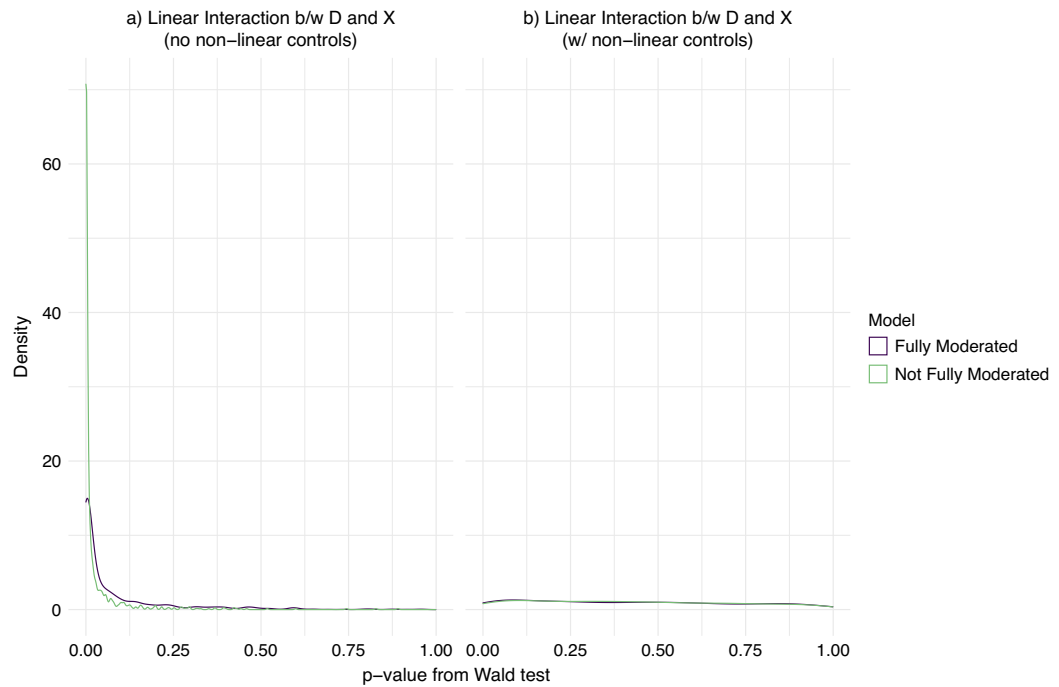
Figure 1. Distribution of *p*-values from the Wald test of whether the linear interaction model can be rejected in favor of the more flexible binning estimator model that relaxes the linear interaction effect assumption.

To assess whether a nonlinear interaction effect is supported by the estimator, we examine the *p*-value from the Wald test proposed by Hainmueller *et al.* (2019) that is based on the binning estimator. Under the null hypothesis, there is no nonlinear interaction effect, and thus the *p*-values of the Wald test from our simulations should be distributed uniformly given the null is true. In the simulations, the binning estimator splits *X* into three bins at the terciles using the default settings of the `interflex` *R* package, as this is how we expect applied researchers typically engage with modeling nonlinear interaction effects.[2]

We estimate four models 1,000 times each using the binning estimator. The first set of models includes *D*, *X*, and *Z* as proposed by Hainmueller *et al.* (2019). The second also adjusts for the inclusion of D^2, X^2, and Z^2. By doing so, we can assess how model misspecification, in terms of failure to account for other possible nonlinearities, affects inferences about the presence of nonlinear interaction effects. Within each set of models, we use the fully moderated version of the binning estimator that interacts *X* with all control variables (*Z*) (Blackwell and Olson 2021) and a version that is not fully moderated, that is, does not include any interactions between *X* and control variables.

Figure 1 displays the results of our simulations.[3] The models in each panel use the binning estimator, with one curve showing results from the fully moderated version that interacts *X* with all control variables and one curve showing the not fully moderated version. In addition, the models in panel (b) include squared terms of all variables in the model: *D*, *X*, and *Z*. Thus, the models in panel (b) include the term Z^2 from the data generating process, whereas the models in panel (a) do not.

The models in panel (b) that include additional squared terms correctly identify that there is no nonlinear interaction in the data, with a uniform distribution of *p*-values as expected if the null were true. In panel (a), on the other hand, both specifications identify a nonlinear interaction in

---

2 We also use the default robust standard errors.
3 Replication code for this article is available in Beiser-McGrath and Beiser-McGrath (2022).

(a) Results using binning estimator,
controlling for Z and using
fully moderated specification.

(b) Results using binning estimator,
controlling for Z but not using
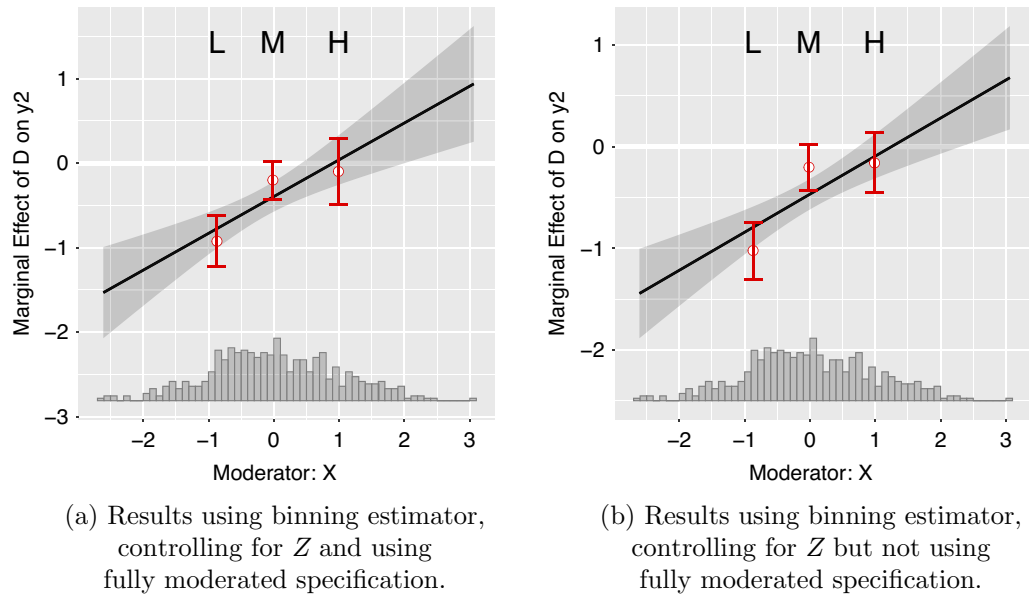fully moderated specification.

Figure 2. Estimated nonlinear interaction effects from the Binning estimator.

the data even though there is only a linear interaction in the data generating process. The not fully moderated specification that is the default option in the most recent version of the `interflex` R package (Version 1.2.6) is particularly prone to do so and returns a Wald test with a p-value less than 0.05 in 86.5% of simulated datasets, whereas the moderated version does so in 66% of simulated datasets. Thus, researchers using the binning estimator in its default specification are likely to incorrectly mistake a linear interaction for a nonlinear one if a quadratic term of a control variable that is correlated with the moderator is omitted from the model.[4]

2.2 Binning Estimator Results

We now turn to examining the substantive inferences from the binning estimator applied to this case. Figure 2 shows results from the fully moderated binning estimator on one of the datasets simulated under the data generating process described above.[5] Even though the relationship between D and Y is moderated linearly by X, the results based on the binning estimator suggest that X non-linearly moderates the relationship between D and Y, both in the fully moderated form in panel (a) and in the unmoderated form in panel (b).

3 Solutions

Our simulation evidence of the best practice for diagnosing nonlinear interactions as suggested by Hainmueller *et al.* (2019) shows that researchers can incorrectly identify a nonlinear interaction when a relevant squared term in a control variable correlated with its constitutive terms is omitted. How should researchers avoid this issue? As this issue occurs when relevant squared terms are omitted (see Figure 1), researchers need to make sure that all relevant nonlinearities are included in the model.

In the first step, this means that when considering whether there are nonlinear interaction effects in a model, scholars also need to think carefully about whether there are theoretical reasons that suggest that one of the control variables may have a nonlinear effect. In a first instance, scholars can include polynomials of those variables to model their expectations.

---

4 In Section 1 of the Supplementary Material, we also discuss how the graphical diagnostics suggested by Hainmueller *et al.* (2019) also suffer in their ability to diagnose nonlinear interactions in the presence of other unmodeled nonlinearities.

5 This is the same simulated dataset we use to illustrate issues with the graphical diagnostics in Section 1 of the Supplementary Material.

In the second step, we suggest using methods for variable selection to identify nonlinearities and interactions in the set of variables used for covariate adjustment (Z) with the Hainmueller *et al.* (2019) binning estimator. A common method for doing so is the use of the adaptive Lasso, which explicitly sets small parameter estimates to zero. Such an approach has been used by applied researchers when faced with specification uncertainty generally (e.g., Bosancianu *et al.* 2020) but also for the estimation of interaction effects (e.g., Belloni, Chernozhukov, and Hansen 2013; Blackwell and Olson 2021). This approach retains the ease of interpretation, post-estimation, provided by the Hainmueller *et al.* (2019) binning estimator. Using the adaptive Lasso to select control terms will help researchers identify relevant nonlinearities among control variables that they did not consider in the first step.

In the third step, we suggest basing inference on machine learning methods that allow for flexible functional forms and complex interactions while at the same time penalizing against unnecessary complexity. This approach also has the advantage of easily incorporating additional nonlinear interactions between the D and X terms with the covariates Z, reducing the possibility of the specified nonlinear interaction picking up an alternative unspecified nonlinear interaction. Beiser-McGrath and Beiser-McGrath (2020) show in Monte Carlo analyses that the adaptive Lasso, kernel regularized least squares (KRLS), and Bayesian additive regression trees (BARTs) are good at identifying linear interactions in the presence of additional nonlinearities and interactions among other correlated variables in the data generating process. However, there is the potential drawback of these methods being more complex, both computationally and in terms of post-estimation inference, thereby placing a greater burden on applied researchers. Additionally, while such methods do not result in "false positives" (Beiser-McGrath and Beiser-McGrath 2020), that is, incorrectly suggesting the presence of interaction effects, they are conservative as penalization can lead to "false negatives", that is, setting true interaction effects to zero (Blackwell and Olson 2021).[6] Despite these potential drawbacks, the third step is necessary to identify nonlinear interactions between variables of interest and other control variables in the model.

In sum, we suggest researchers follow the following three steps. We consider these three steps as providing an increasing degree of robustness when evaluating nonlinear interaction effects. This process allows researchers to assess how confident they can be about the presence of a nonlinear interaction effect originally identified using the Hainmueller *et al.* (2019) binning estimator, as each step increases the robustness of their estimation to nonlinearities and unmodeled nonlinear interaction effects.

- Step 1: Think theoretically about whether any nonlineariti among control variables are plausible and if so, include terms to model them as control variables when using the binning estimator.
- Step 2: Use the adaptive Lasso to select nonlinearities among control variables and include those as control variables in the binning estimator.
- Step 3: Use the adaptive Lasso to estimate a fully interactive and nonlinear model.

To demonstrate the relevance of the proposed technical solutions we propose in Steps 2 and 3, we reanalyze the previous example using the solutions we propose. Figure 3 finds that both of our proposed solutions suggest a linear interaction effect over the majority of the support for X. In contrast, the nonlinear interaction binning estimator finds a significant nonlinear interaction effect. While the marginal effect increases comparing the first and second bin estimates, from this point, it remains constant when comparing the second and third bin estimats. Additionally, the kernel estimator proposed by Hainmueller *et al.* (2019) also displays nonlinearities, even though the true marginal effect is linear. This example shows that by allowing for a broader range of nonlinearities, both independent and interactive, than the Hainmueller *et al.* (2019) estimators,

---

6 Beiser-McGrath and Beiser-McGrath (2020) find that BART is more conservative in such settings, when compared with the adaptive Lasso and KRLS.
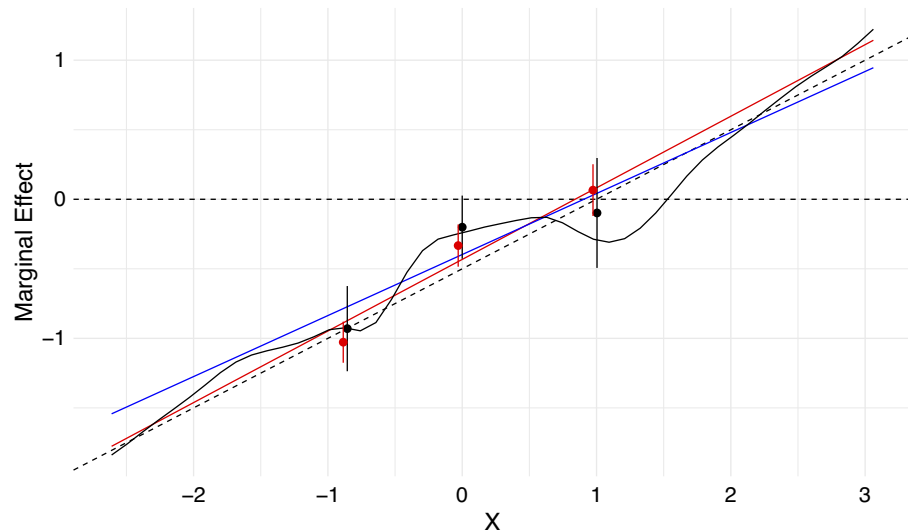

Figure 3. Nonlinear interaction effect estimators incorrectly identify a nonlinear interaction effect, when the true interaction effect is linear and covariates have nonlinear effects. The dashed line presents the true marginal effect. The blue line presents the marginal effect from a standard linear interaction model. Point estimates with 95% confidence intervals from the Hainmueller *et al.* (2019) binning estimator are also displayed in black, and from the binning estimator with adaptive Lasso selected nonlinearities and interactions for covariate adjustment in red. The black line presents the marginal effect from the Hainmueller *et al.* (2019) kernel estimator. The red line presents the marginal effect from a fully nonlinear interactive adaptive Lasso.

the methods we propose are able to avoid the bias that arises from model misspecification due to unmodeled nonlinearities.

4 Reanalysis of Previous Studies

We also provide a broader replication of previous studies, to demonstrate how our proposals for increasing the robustness of nonlinear interaction effects affect inferences more generally. To do so, we use the replications of previous research presented in Hainmueller *et al.* (2019). From this set of studies, we focus on those that do not suffer from the issue of common support, which was also discussed by Hainmueller *et al.* (2019). Additionally, we do not replicate studies that do not have additional covariates (Z) beyond the moderator (X), and that suffer from computational issues.[7] This leaves us with 17 studies with 23 estimated interaction effects that we reanalyze.[8] Figure 4 illustrates the findings of our reanalysis. In each panel, the black line shows the marginal effect from a standard linear interaction model, the black point estimates derive from the Hainmueller *et al.* (2019) binning estimator, the red point estimates derive from the binning estimator with adaptive Lasso selected nonlinearities and interactions for covariate adjustment, and the red line presents the marginal effect from a fully nonlinear interactive adaptive Lasso.

We now summarize the findings from Figure 4.[9] First, we examine the results from the original binning estimator (black point estimates in Figure 4). Using this estimator, we find a clear nonlinear relationship with the point estimates of the marginal effects changing nonmonotonically between terciles for 13 of the interactions. In addition, five of the interactions where marginal effects do not change nonmonotonically are visibly nonlinear.[10]

---

7 There are two interactions from a study where only two bins are identified by the Hainmueller *et al.* (2019) binning estimator, and two studies that each take longer than 36 hours to estimate on a personal computer.

8 Section 2 of the Supplementary Material lists all the articles included in the reanalysis.

9 A full table summarizing these results is located in Section 3 of the Supplementary Material.

10 To assess whether findings based on the binning estimator (black and red point estimates) are visually nonlinear, we compare the slope between the marginal effect in the first and second terciles to the slope between the marginal effect in
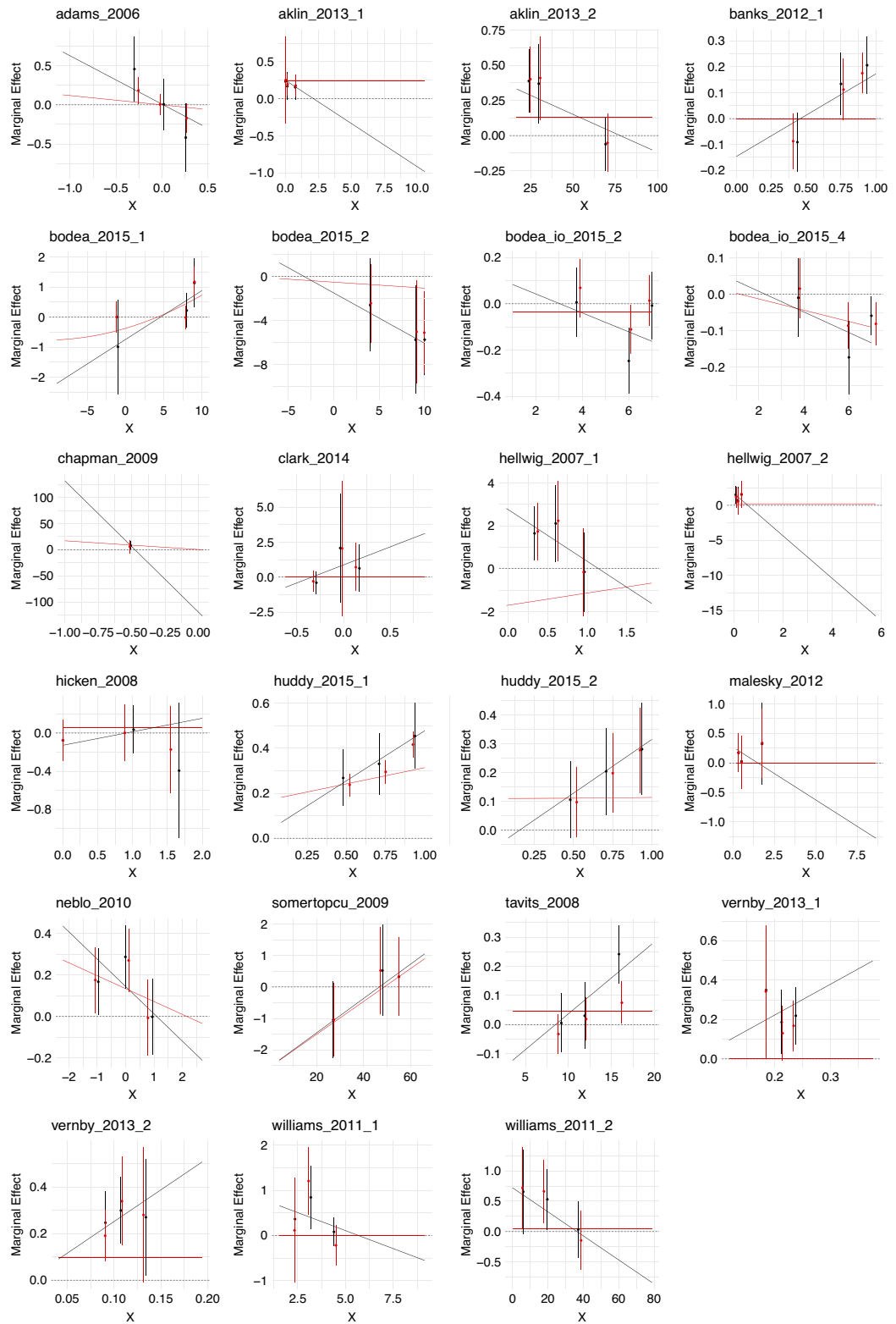

Figure 4. Reanalysis of 23 nonlinear interaction effects from 17 studies. The black line shows the marginal effect from a standard linear interaction model, the black point estimates derive from the Hainmueller *et al.* (2019) binning estimator, the red point estimates from the binning estimator with adaptive Lasso selected nonlinearities and interactions for covariate adjustment, and the red line presents the marginal effect from a fully nonlinear interactive adaptive Lasso.

the second and third terciles. Specifically, we class an interaction as nonlinear if the slope that is less steep is less than a third of the steeper slope.

We now compare these inferences from the Hainmueller *et al.* (2019) binning estimator to our proposed solutions. First, comparing these findings to a version of the binning estimator which accounts for nonlinearities and interactions among the set of variables used for covariate adjustment (*Z*), selected by the adaptive Lasso, the findings (red point estimates in Figure 4) are very similar. In one of the cases where the marginal effects between terciles change nonmonotonically when using the original binning estimator this is no longer the case using Lasso-selected control variables (bodea_2015_2), even though the interactive relationship is still visibly nonlinear. In addition, only one of the interactions that were found to be visibly nonlinear without changing nonmonotonically using the original binning estimator is no longer at least visibly nonlinear when including Lasso-selected controls (tavits_2008).

Second, using the more severe robustness check of the fully nonlinear interactive adaptive Lasso, we find that the adaptive Lasso only finds a weakly nonlinear interaction in one of the 18 cases that were identified by the binning estimator as substantively nonlinear interactions (bodea_2015_1). Instead, in 12 of these cases, the Lasso suggests that there is no interaction between the two variables of interest at all (the slope of the marginal effect is zero, i.e., a horizontal line). In the remaining five cases that were identified as substantively nonlinear interactions by the binning estimator, the Lasso suggests a linear interaction.

In the Supplementary Material, we conduct additional analyses for those interaction effects where the fully specified adaptive Lasso returns a noninteractive effect, that is, a constant marginal effect. For those cases we also estimate BARTs (Green and Kern 2012), to ensure that these results are not purely a function of potential penalization bias from the Lasso. For 10 of the 14 interaction effects, we find that BART also leads to the inference that there is no substantive interaction effect, and for all studies, no statistically significant interaction effect, increasing our confidence that these inferences are not due to the choice of estimator.

One concern may be that the lack of interaction effects identified is a product of regularization bias, where meaningful interactions are over-regularized and set to zero in this high-dimensional setting. However, we do find many nonlinearities and nonlinear interactions for other variables in Section 5 of the Supplementary Material. This suggests that regularization bias is not a problem *per se*, rather the nonlinear interactions specified are absorbing other important nonlinearities and nonlinear interactions that are not specified in the typical application of the Hainmueller *et al.* (2019) estimator. Thus, in practice, omitted interaction bias is a larger concern than regularization bias.

In summary, our results suggest that the estimation of nonlinear interaction effects is sensitive to the problem of unmodeled nonlinearities discussed in this paper as well as the broader issue of omitted interaction bias discussed in previous research (Beiser-McGrath and Beiser-McGrath 2020; Blackwell and Olson 2021). The solutions we propose in this paper thus provide researchers with approaches to evaluate the sensitivity of their chosen nonlinear interaction effects which can help demonstrate the strength and robustness of their findings.

5 Conclusion

Interaction effects are commonly used in Applied Political Science (Brambor *et al.* 2006). Hainmueller *et al.* (2019) have convincingly demonstrated that an incorrect assumption of a linear functional form of moderating effects can lead to bias. They have also provided essential tools for researches to estimate a nonlinear interaction effect and provided a code of the best practice to enable researchers to diagnose when the linear interaction effect assumption does not hold.

We have shown in this research note, however, that by relaxing this functional form assumption, there is the risk of unmodeled nonlinearities among variables used for covariate adjustment (*Z*) biasing interaction effect estimates. This can result in researchers finding evidence for a nonlinear interaction effect, even though the true interaction effect is likely linear. Moreover, the diagnostic

tools suggested by Hainmueller *et al.* (2019) are not able to identify this potential bias, leaving researchers vulnerable to omitted nonlinearities in correlated control variables. Additionally, allowing for a fuller set of nonlinear interactions, including D and X, finds that many interaction effects are likely zero, as they are absorbing other nonmodeled nonlinear additive and interaction effects.

The results suggest that while moving beyond linearity as a default in the estimation of interaction effects is crucial for ensuring robust inferences, it needs to be coupled with a broader consideration of nonlinearities, both independent and interactive, in model specification (Beiser-McGrath and Beiser-McGrath 2020). Methods such as the adaptive Lasso (Blackwell and Olson 2021; Kenkel and Signorino 2013), KRLS (Hainmueller and Hazlett 2014), and BARTs (Green and Kern 2012) thus serve as an important step in establishing the robustness of nonlinear interaction effects identified through the Hainmueller *et al.* (2019) estimator.

Acknowledgments

Data Availability Statement

Replication code for this article is available in Beiser-McGrath and Beiser-McGrath (2022) at https://doi.org/10.7910/DVN/S44D0E.

Supplementary Material

For supplementary material accompanying this paper, please visit https://doi.org/10.1017/pan.2022.25.

References

Beiser-McGrath, J., and L. F. Beiser-McGrath. 2020. "Problems with Products? Control Strategies for Models with Interactive and Quadratic Effects." *Political Science Research and Methods* 8 (4): 707–730. https://doi.org/10.1017/psrm.2020.17

Beiser-McGrath, J., and L. F. Beiser-McGrath. 2022. "The Consequences of Model Misspecification for the Estimation of Non-Linear Interaction Effects." https://doi.org/10.7910/DVN/S44D0E

Belloni, A., V. Chernozhukov, and C. Hansen. 2013. "Inference on Treatment Effects after Selection among High-Dimensional Controls." *The Review of Economic Studies* 81 (2): 608–650. https://doi.org/10.1093/restud/rdt044

Berry, W. D., J. H. R. DeMeritt, and J. Esarey. 2010. "Testing for Interaction in Binary Logit and Probit Models: Is a Product Term Essential?" *American Journal of Political Science* 54 (1): 248–266. https://doi.org/10.1111/j.1540-5907.2009.00429.x

Berry, W. D., J. H. R. DeMeritt, and J. Esarey. 2016. "Bias and Overconfidence in Parametric Models of Interactive Processes." *American Journal of Political Science* 60 (2): 521–539. https://onlinelibrary.wiley.com/doi/abs/10.1111/ajps.12123

Blackwell, M., and M. P. Olson. 2021. "Reducing Model Misspecification and Bias in the Estimation of Interactions." *Political Analysis* 30 (4): 1–20.

Bosancianu, C. M., H. Hilbig, M. Humphreys, K. C. Sampada, N. Lieber, and A. Scacco. 2020. "Political and Social Correlates of Covid-19 Mortality." osf.io/preprints/socarxiv/ub3zd.

Brambor, T., W. R. Clark, and M. Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14: 63–82.

Esarey, J., and J. L. Sumner. 2018. "Marginal Effects in Interaction Models: Determining and Controlling the False Positive Rate." *Comparative Political Studies* 51 (9): 1144–1176. https://doi.org/10.1177/0010414017730080

Green, D. P., and H. L. Kern. 2012. "Modeling Heterogeneous Treatment Effects in Survey Experiments with Bayesian Additive Regression Trees." *Public Opinion Quarterly* 76 (3): 491–511. http://doi.org/10.1093/poq/nfs036

Hainmueller, J., and C. Hazlett. 2014. "Kernel Regularized Least Squares: Reducing Misspecification Bias with a Flexible and Interpretable Machine Learning Approach." *Political Analysis* 22 (2): 143–168.

Hainmueller, J., J. Mummolo, and Y. Xu. 2019. "How Much Should We Trust Estimates from Multiplicative Interaction Models? Simple Tools to Improve Empirical Practice." *Political Analysis* 27 (2): 163–192.

Kenkel, B., and C. S. Signorino. 2013. "Bootstrapped Basis Regression with Variable Selection—A New Method for Flexible Functional Form Estimation." Working Paper. http://bkenkel.com/data/basics.pdf.