Problems with Products?
Control Strategies for Models with Interactive and Quadratic Effects

Supplementary Materials

Janina Beiser-McGrath[*] Liam F. Beiser-McGrath[†]

This document contains supplementary information related to the main text.

[*]Senior Research Fellow, Department of Politics and Public Administration, Universität Konstanz. Contact e-mail: janina.beiser@uni-konstanz.de

[†]Senior Research Fellow, Universität Konstanz. Research Associate, ETH Zürich. Contact e-mail: liam.beiser-mcgrath@uni-konstanz.de, liam.mcgrath@ir.gess.ethz.ch

Contents

1 *Additional example illustrating the logic of the problem of omitted product terms*

We provide a second example for the problems that can occur when product terms amongst the control variables are omitted based on Williams and Whitten (2015), the paper we replicate in the article. The authors hypothesize that the state of the economy in a country will affect the vote share of the party of the prime minister more than the vote share of other parties. Another factor that has the potential to affect different parties' vote shares is the timing of an election, that is whether elections are called earlier than constitutionally required. The authors expect that early elections increase the vote share of the prime minister's party compared to opposition parties. There are several reasons to believe that the state of the economy and early elections are linked. Firstly, governments have incentives to call early elections if the economy is in good shape. Secondly, economic crises may reveal irreconcilable ideological differences between coalition partners and require early elections. Thirdly, political instability may stifle economic growth. Even though the authors are interested in the moderating effect of the state of the economy on the party of the prime minister, they include two interaction terms in their model: one between the variable on whether a party is the party of the prime minister and the state of the economy and one between the measure on prime minister's party and the time left until elections are constitutionally required.[1] If they had failed to do so, their finding on the interactive relationship between economic performance and being

[1] They also include interactions between a variable on coalition partners and those two potentially moderating variables.

the party of the prime minister could reflect a potentially moderating effect of early elections on the variable on the prime minister's party and a correlation between the state of the economy and early elections.

2 *Bias Analysis for Omitted Product Terms*

2.1 *Omitting an Interaction Term That Shares a Constitutive Term with an Included Interaction Term*

In a first step, let us consider the following data generating process in scalar notation

$$y_i = \beta_0 + \beta_1 a_i + \beta_2 b_i + \beta_3 c_i + \beta_4 d_i + \beta_5 a_i b_i + \epsilon_i. \tag{1}$$

For the sake of simplicity, we assume that a_i, b_i, c_i, and d_i are mean-centered and have unit variance. Also, let a_i, b_i, c_i, and d_i follow a multivariate normal distribution and the standard OLS assumptions about ϵ_i apply.[2]

Now consider a situation where we estimate y_i from (1) with a misspecified model in which the relevant interaction term ab_i is omitted, and an alternative ("false") term, ac_i, is used instead:

$$y_i = \beta_0^* + \beta_1^* a_i + \beta_2^* b_i + \beta_3^* c_i + \beta_4^* d_i + \beta_5^* a_i c_i + u_i. \tag{2}$$

Knowing the true data generating process, it is clear that $\beta_5^* = 0$ and $u_i =$

[2]Note that while the assumptions about the scaling of the regressors have no substantial impact on the presented results, the consequences of relaxing the assumption of multivariate normality are discussed in the next section.

$\beta_5 a_i b_i + \epsilon_i$. Now let us switch to matrix notation and define

$$\mathbf{X}_{n\times 6} = \begin{pmatrix} 1 & a_1 & b_1 & c_1 & d_1 & a_1c_1 \\ 1 & a_2 & b_2 & c_2 & d_2 & a_2c_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & a_n & b_n & c_n & d_n & a_nc_n \end{pmatrix}, \ \mathbf{Y}_{n\times 1} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

$$\mathbf{ab}_{n\times 1} = \begin{pmatrix} a_1b_1 \\ a_2b_2 \\ \vdots \\ a_nb_n \end{pmatrix}, \ \boldsymbol{\epsilon}_{n\times 1} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix},$$

and $\boldsymbol{\beta^*}_{6\times 1} = (\beta_0^*, \beta_1^*, \beta_2^*, \beta_3^*, \beta_4^*, 0)'$. Then, using the information about the true data generating process, the estimated model (2) may be written as $\boldsymbol{Y} = \boldsymbol{X\beta^*} + \beta_5 \boldsymbol{ab} + \boldsymbol{\epsilon}$. Finally, plugging the latter expression into the OLS estimator for $\boldsymbol{\beta^*}$ and taking expectations yields

$$E(\hat{\boldsymbol{\beta}}^*) = \boldsymbol{\beta^*} + \beta_5(\boldsymbol{X'X})^{-1}\boldsymbol{X'ab} = \begin{pmatrix} \beta_0^* + \hat{\gamma}_0\beta_5 \\ \beta_1^* + \hat{\gamma}_1\beta_5 \\ \beta_2^* + \hat{\gamma}_2\beta_5 \\ \beta_3^* + \hat{\gamma}_3\beta_5 \\ \beta_4^* + \hat{\gamma}_4\beta_5 \\ \hat{\gamma}_5\beta_5 \end{pmatrix}, \tag{3}$$

with the γs being the estimated coefficients from the auxiliary regression

$$a_i b_i = \gamma_0 + \gamma_1 a_i + \gamma_2 b_i + \gamma_3 c_i + \gamma_4 d_i + \gamma_5 a_i c_i + \eta_i. \tag{4}$$

Expression (3) shows that the estimated coefficients of the misspecified model will be biased if 1) the effect of the omitted interaction term on the dependent variable, β_5, is non-zero; *and* 2) the respective OLS coefficient of regressing the regressors of the misspecified model on the omitted interaction term, $\hat{\gamma}_k$, is non-zero.

The analytic result of (3) follows the same logic as the well known equation for omitted variable bias (Greene, 2003): The coefficients for the constitutive terms of the omitted, true, interaction term, a_i and b_i, will be biased in the misspecified model as long as β_5, the coefficient of the true interaction term, is non-zero. After all, they have an interactive effect that cannot be captured without including an interaction term. In our context, that means that the estimates for the constitutive term of the included interaction, a_i, and additive control variable b_i will be biased. This is a rather unsurprising result as it has become a well known fact that testing conditional hypotheses with purely additive models will lead to biased coefficients.

What is more interesting, however, is that the estimated coefficient for the interaction $a_i c_i$ is also biased. This coefficient is of primary interest because researchers including interaction terms are usually interested in testing whether there is a conditional effect of a_i and c_i on the response. Hence, any bias in the estimated coefficient for $a_i c_i$ relates directly to the risk of drawing wrong conclusions about the presence of a presumed conditional effect of these variables on the outcome. According to expression (3), $\hat{\beta}_5^*$ will be biased if both β_5 and γ_5 are non-zero. Let us assume that this is the case for the former, and focus on the latter. If we define $\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5)'$, we can derive the OLS estimate $\hat{\gamma}_5$ by solving the linear equation system $(\boldsymbol{X'X})\gamma = \boldsymbol{X'ab}$. Under the above made distributional assumptions, doing so yields that the bias associated with the OLS estimate of the "false" multiplicative term $a_i c_i$ is[3]

$$\hat{\gamma}_5 \beta_5 = \beta_5 * \frac{var(a_i)cov(b_i, c_i) + cov(a_i, b_i)cov(a_i, c_i)}{var(a_i)var(c_i) + cov(a_i, c_i)^2}. \tag{5}$$

Expression (5) shows under which conditions $\hat{\beta}_5^*$ will be biased. If the constitutive terms of the "false" interaction term, a_i and c_i, are each correlated with one of the constitutive terms of the omitted "true" interaction term, a_i and b_i, then the OLS estimator of the "false" interaction term will be biased. Moreover, for bias it is sufficient for c_i, the false moderator of a_i, and b_i, the true moderator of a_i to be correlated. In addition, it is also sufficient for c_i and b_i to both be correlated with a_i. It is very important to note that these biases occur *even though* the constitutive terms of the omitted interaction term have been included in the estimated model as additive regressors.

2.2 *Omitting a Squared Term That Shares a Constitutive Term with an Included Interaction term*

While the previous section has established that omitted interaction terms may lead to biased coefficients for included interaction terms, in this section, we

[3]For the full derivation of this result, see section 3.1.

show that the same argument is applicable to quadratic terms. An interaction term included in a regression specification can pick up the non-linear effects of the variables included in the interaction term, suggesting conditional effects even when they do not exist.[4][5]

Consider a data generating process similar to (1), but including a squared term instead of a multiplicative term

$$y_i = \beta_0 + \beta_1 a_i + \beta_2 b_i + \beta_3 c_i + \beta_4 d_i + \beta_5 a_i^2 + \epsilon_i, \tag{6}$$

and let the same distributional assumptions apply as above. In this data generating process a has a non-linear effect upon the outcome, the shape of which is determined by β_1 and β_5. Suppose, that we estimate the response y_i with the same misspecified model used previously

$$y_i = \beta_0^* + \beta_1^* a_i + \beta_2^* b_i + \beta_3^* c_i + \beta_4^* d_i + \beta_5^* a_i c_i + u_i. \tag{7}$$

Hence, we specify that the effect of a_i is conditional upon c_i. It is easy to see that the effects of a_i and c_i (as well as all other variables) will be biased to the degree that they are partially correlated with the omitted squared term. Moreover, if we proceed analogously to the previous section, we can derive that the OLS

[4]Naturally a similar logic to the previous subsection holds, an included non-linear effect can pick up an omitted non-linear effect. In this paper we focus exclusively on quadratic terms, and not higher powers, because testing a second-order polynomial is usually the first step in assessing curvilinear effects, and hypothesised relationships with third- or higher-order polynomials are extremely rare in political science.

[5]A similar case of misspecification is shown in Hainmueller and Hazlett (2014) in a Monte Carlo experiment.

estimator for the "false" product term $a_i c_i$ will incur the following bias:[6]

$$\hat{\gamma}_5 \beta_5 = \beta_5 * \frac{2 * var(a_i) cov(a_i, c_i)}{var(a_i) var(c_i) + cov(a_i, c_i)^2}. \tag{8}$$

Expression (8) shows that the OLS estimator of the unnecessary interaction effect $a_i c_i$ is biased to the extent that c_i is correlated with a_i as it partly captures the non-linear effect of a_i.

Once we consider the fact that interaction terms and squared terms are mathematically equivalent to the extent that they are both products of explanatory variables, it becomes clear that the cases discussed in this and the previous section may be subsumed under a more general result: Omitting relevant product terms biases the estimated parameters for included product terms if their respective constitutive variables covary, regardless of whether these product terms represent curvilinear or conditional effects[7]. After all, the issue we discuss here is an incorrect conclusion about the degree to which a variable moderates another one or indeed itself.

2.3 *Omitting Product Terms That Do Not Share a Constitutive Term with an Included Product Term*

So far we have shown how omitting relevant interactions or squared terms can bias included interaction terms that they share a constitutive term with. While

[6] For a formal derivation, see section 3.1.

[7] For the purpose of this paper, we use the term "product terms" when referring to interaction- and squared terms. We are certain that the results presented in this paper also apply to higher-order products, but for the sake of parsimony and current specifications within the literature, we leave formal proof of this proposition to future research.

these results are quite intuitive, we show now that omitting a relevant product term can bias a product interaction term even if the two products do not share any constitutive terms.

Suppose we estimate a model where

$$y_i = \beta_0^* + \beta_1^* a_i + \beta_2^* b_i + \beta_3^* c_i + \beta_4^* d_i + \beta_5^* cd_i + u_i. \tag{9}$$

Following the previous logic, if the true model[8] contains an interaction term ab_i, omitting this term introduces the following bias to the coefficient of the included interaction term cd_i

$$\hat{\gamma}_5 \beta_5 = \beta_5 * \frac{cov(a_i, c_i)cov(b_i, d_i) + cov(a_i, d_i)cov(b_i, c_i)}{var(c_i)var(d_i) + cov(c_i, d_i)^2}. \tag{10}$$

Therefore, as long as there is a non-zero covariance between each interaction term in the included interaction and one of the interaction terms in the excluded interaction, conclusions on the effect of the included interaction will be biased. Specifically, here, β_5 is biased if either a and c and b and d are correlated or if a and d and b and c are correlated.

A similar problem can occur for the case of omitted quadratic terms. For example if the true model[9] contains quadratic term a_i^2, omitting this term biases the included interaction term cd_i in the following way

[8] i.e. $y_i = \beta_0 + \beta_1 a_i + \beta_2 b_i + \beta_3 c_i + \beta_4 d_i + \beta_5 a_i b_i + \epsilon_i$

[9] i.e. $y_i = \beta_0 + \beta_1 a_i + \beta_2 b_i + \beta_3 c_i + \beta_4 d_i + \beta_5 a_i^2 + \epsilon_i$

$$\hat{\gamma}_5\beta_5 = \beta_5 * \frac{2 * cov(a_i, c_i)cov(a_i, d_i)}{var(c_i)var(d_i) + cov(c_i, d_i)^2}. \quad (11)$$

Thus, as long as the constitutive term of the squared term is correlated with both constitutive terms of the included interaction, the inference about the interaction will be biased.

3 Derivation of Analytical Results

3.1 Linear Models and Multivariate Normal Data

In this section we will briefly derive the bias expressions in the previous section. We have established that the total bias associated with the estimated term $\hat{\beta}_5^*$ is given by $\hat{\gamma}_5\beta_5$. Whereas β_5 is determined by the data generating process, we know that $\hat{\gamma}_5$ is the OLS estimator of γ_5 in the linear model

$$a_i b_i = \gamma_0 + \gamma_1 a_i + \gamma_2 b_i + \gamma_3 c_i + \gamma_4 d_i + \gamma_5 c_i d_i + \eta_i. \tag{12}$$

Writing the OLS estimator of model (12) in matrix notation (and using the same definitions as in section 2.1) yields $\hat{\boldsymbol{\gamma}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{ab}$. Rearranging this expression yields the linear equation system $(\boldsymbol{X}'\boldsymbol{X})^{-1}\hat{\boldsymbol{\gamma}} = \boldsymbol{X}'\boldsymbol{ab}$. Taking expectations on both sides, we can write this system in scalar notation with the following six equations:

$$\begin{aligned}
E(a_ib_i) &= \hat{\gamma}_0 &&+ \hat{\gamma}_1 E(a_i) &&+ \hat{\gamma}_2 E(b_i) &&+ \hat{\gamma}_3 E(c_i) &&+ \hat{\gamma}_4 E(d_i) &&+ \hat{\gamma}_5 E(c_id_i)\\
E(a_i^2b_i) &= \hat{\gamma}_0 E(a_i) &&+ \hat{\gamma}_1 E(a_i^2) &&+ \hat{\gamma}_2 E(a_ib_i) &&+ \hat{\gamma}_3 E(a_ic_i) &&+ \hat{\gamma}_4 E(a_id_i) &&+ \hat{\gamma}_5 E(a_ic_id_i)\\
E(a_ib_i^2) &= \hat{\gamma}_0 E(b_i) &&+ \hat{\gamma}_1 E(a_ib_i) &&+ \hat{\gamma}_2 E(b_i^2) &&+ \hat{\gamma}_3 E(b_ic_i) &&+ \hat{\gamma}_4 E(b_id_i) &&+ \hat{\gamma}_5 E(b_ic_id_i)\\
E(a_ib_ic_i) &= \hat{\gamma}_0 E(c_i) &&+ \hat{\gamma}_1 E(a_ic_i) &&+ \hat{\gamma}_2 E(b_ic_i) &&+ \hat{\gamma}_3 E(c_i^2) &&+ \hat{\gamma}_4 E(c_id_i) &&+ \hat{\gamma}_5 E(c_i^2d_i)\\
E(a_ib_id_i) &= \hat{\gamma}_0 E(d_i) &&+ \hat{\gamma}_1 E(a_id_i) &&+ \hat{\gamma}_2 E(b_id_i) &&+ \hat{\gamma}_3 E(c_id_i) &&+ \hat{\gamma}_4 E(d_i^2) &&+ \hat{\gamma}_5 E(c_id_i^2)\\
E(a_ib_ic_id_i) &= \hat{\gamma}_0 E(c_id_i) &&+ \hat{\gamma}_1 E(a_ic_id_i) &&+ \hat{\gamma}_2 E(b_ic_id_i) &&+ \hat{\gamma}_3 E(c_i^2d_i) &&+ \hat{\gamma}_4 E(c_id_i^2) &&+ \hat{\gamma}_5 E(c_i^2d_i^2)
\end{aligned}$$

Since we assume all explanatory variables to be mean centered, all expectations in the above equations equal corresponding central moments. Hence, since we assume unit variance, all expectations of squared terms equal unity. Furthermore, because under multivariate normality all odd central moments equal zero, all expectations involving three factors vanish. Thus, the linear equation system simplifies to the following expressions:

$$\begin{array}{lllllll}
E(a_ib_i) & = \hat{\gamma}_0 & + 0 & + 0 & + 0 & + 0 & + \hat{\gamma}_5 E(c_id_i) \\
0 & = 0 & + \hat{\gamma}_1 & + \hat{\gamma}_2 E(a_ib_i) & + \hat{\gamma}_3 E(a_ic_i) & + \hat{\gamma}_4 E(a_id_i) & + 0 \\
0 & = 0 & + \hat{\gamma}_1 E(a_ib_i) & + \hat{\gamma}_2 & + \hat{\gamma}_3 E(b_ic_i) & + \hat{\gamma}_4 E(b_id_i) & + 0 \\
0 & = 0 & + \hat{\gamma}_1 E(a_ic_i) & + \hat{\gamma}_2 E(b_ic_i) & + \hat{\gamma}_3 & + \hat{\gamma}_4 E(c_id_i) & + 0 \\
0 & = 0 & + \hat{\gamma}_1 E(a_id_i) & + \hat{\gamma}_2 E(b_id_i) & + \hat{\gamma}_3 E(c_id_i) & + \hat{\gamma}_4 & + 0 \\
E(a_ib_ic_id_i) & = \hat{\gamma}_0 E(c_id_i) & + 0 & + 0 & + 0 & + 0 & + \hat{\gamma}_5 E(c_i^2d_i^2)
\end{array} \tag{13}$$

Fortunately, this system allows to derive $\hat{\gamma}_5$ through elimination using only the first and last equation. Doing so yields

$$\hat{\gamma}_5 = \frac{E(a_ib_ic_id_i) - E(a_ib_i)E(c_id_i)}{E(c_i^2d_i^2) - E(c_id_i)^2}. \tag{14}$$

Attentive readers will recognize that the numerator of expression (14) is equal to $cov(a_ib_i, c_id_i)$, and the denominator equals $var(c_id_i)$. It is more instructive, however, to simplify (14) to covariances between the explanatory variables. For the numerator, we may do so using Isserlis' Theorem on higher order moments in multivariate normal distributions (see Isserlis 1918), which implies that $E(a_ib_ic_id_i) = E(a_ib_i)E(c_id_i) + E(a_ic_i)E(b_id_i) + E(a_id_i)E(b_ic_i)$. For the denominator, we may use the result by (Bohrnstedt and Goldberger, 1969) for the variance of products under multivariate normality, according to which

$$\begin{aligned} var(c_i d_i) &= E(c_i)^2 var(d_i) + E(d_i)^2 var(c_i) \\ &\qquad + 2E(c_i)E(d_i)cov(c_i d_i) + var(c_i)var(d_i) + cov(c_i d_i)^2 \\ &= var(c_i)var(d_i) + cov(c_i d_i)^2. \end{aligned}$$

Using these two results, we may write the total bias associated with $\hat{\beta}_5^*$ as

$$\hat{\gamma}_5 \beta_5 = \beta_5 * \frac{cov(a_i, c_i)cov(b_i, d_i) + cov(a_i, d_i)cov(b_i, c_i)}{var(c_i)var(d_i) + cov(c_i, d_i)^2}. \tag{15}$$

To derive expression 11 from section 2.4, we may proceed analogously. In this case, we solve the linear equation system $(\boldsymbol{X'X})^{-1}\hat{\boldsymbol{\gamma}} = \boldsymbol{X'a^2}$ to receive

$$\hat{\gamma}_5 = \frac{E(a_i^2 c_i d_i) - E(a_i^2)E(c_i d_i)}{E(c_i^2 d_i^2) - E(c_i d_i)^2},$$

which, using the same sources as above, simplifies to

$$\frac{2 * cov(a_i, c_i)cov(a_i, d_i)}{var(c_i)var(d_i) + cov(c_i, d_i)^2}. \tag{16}$$

3.2 *Linear Models and Non-Normal Data*

In section 3.1, all presented derivations have rested on the assumption of multivariate normally distributed explanatory variables. Now we will show that in the case of non-normal data, the illustrated bias may even occur if the respective constitutive variables do not covary, but have non-zero higher order central cross-moments. Reconsider the data generating process:

$$y_i = \beta_0 + \beta_1 a_i + \beta_2 b_i + \beta_3 c_i + \beta_4 d_i + \beta_5 a_i b_i + \epsilon_i. \tag{17}$$

As in the previous section, let a_i, b_i, c_i, and d_i be mean-centered and have unit variance. Now, however, we do *not* assume multivariate normality. Instead, we assume all pairwise covariances between the explanatory variables equal zero: $cov(x, y) = 0$ for all pairs $(x, y)|x, y \in (a_i, b_i, c_i, d_i)$ except $x = y$. Suppose, again, we estimate the response y_i with a misspecified model in which $a_i b_i$ is omitted, and the product term $c_i d_i$ is included as a regressor instead

$$y_i = \beta_0^* + \beta_1^* a_i + \beta_2^* b_i + \beta_3^* c_i + \beta_4^* d_i + \beta_5^* c_i d_i + u_i. \tag{18}$$

It is easy to see from expression (3) that in the multivariate normal case, under the present assumption of zero covariance between the explanatory variables the OLS estimators for β_5^* would be unbiased.

Using the same definitions as in section 2.1 and proceeding analogously, we may again obtain the bias associated with OLS estimator for the "false" interaction term $c_i d_i$, $\hat{\beta}_5{}^*$, by solving the linear equation system $(\boldsymbol{X'X})\gamma = \boldsymbol{X'ab}$.

Since the data generating process and the estimated model are the same as in section 2.2, finding the bias associated with $\hat{\beta}_5^*$ again requires solving the linear equation system given in (13). Now, however, we are operating under different distributional assumptions, namely that the explanatory variables are *not* multivariate normally distributed, but instead all covariances equal zero. Thus, all expectations of products involving only two factors vanish. Moreover, as above, expectations of the explanatory variables equal zero, and, since we assume unit variance, expectations of squared terms equal unity. Therefore, the linear equation system simplifies to

$$\begin{aligned}
0 &= \hat{\gamma}_0 + 0 && + 0 && + 0 && + 0 && + 0 \\
E(a_i^2 b_i) &= 0 \ + \hat{\gamma}_1 && + 0 && + 0 && + 0 && + \hat{\gamma}_5 E(a_i c_i d_i) \\
E(a_i b_i^2) &= 0 \ + 0 && + \hat{\gamma}_2 && + 0 && + 0 && + \hat{\gamma}_5 E(b_i c_i d_i) \\
E(a_i b_i c_i) &= 0 \ + 0 && + 0 && + \hat{\gamma}_3 && + 0 && + \hat{\gamma}_5 E(c_i^2 d_i) \\
E(a_i b_i d_i) &= 0 \ + 0 && + 0 && + 0 && + \hat{\gamma}_4 && + \hat{\gamma}_5 E(c_i d_i^2) \\
E(a_i b_i c_i d_i) &= 0 \ + \hat{\gamma}_1 E(a_i c_i d_i) && + \hat{\gamma}_2 E(b_i c_i d_i) && + \hat{\gamma}_3 E(c_i^2 d_i) && + \hat{\gamma}_4 E(c_i d_i^2) && + \hat{\gamma}_5 E(c_i^2 d_i^2).
\end{aligned}$$

We can solve this linear equation system for $\hat{\gamma}_5$ using sequential elimination, which yields

$$\begin{aligned}
\hat{\gamma}_5 = & \frac{E(a_i b_i c_i d_i) - E(a_i^2 b_i)E(a_i c_i d_i) - E(a_i b_i^2)E(b_i c_i d_i)}{E(c_i^2 d_i^2) - E(a_i c_i d_i)^2 - E(b_i c_i d_i)^2 - E(c_i^2 d_i)^2 - E(c_i d_i^2)^2} \\
& - \frac{E(c_i^2 d_i)E(a_i b_i c_i) - E(c_i d_i^2)E(a_i b_i d_i)}{E(c_i^2 d_i^2) - E(a_i c_i d_i)^2 - E(b_i c_i d_i)^2 - E(c_i^2 d_i)^2 - E(c_i d_i^2)^2}.
\end{aligned} \tag{19}$$

Note that since we have assumed all explanatory variables to be mean centered, all expected values in the numerator and denominator are third and fourth central cross-moments of the explanatory variables.[10] The relationship between this expression and its equivalent in section 2.2 is that in the latter case, the assumption of multivariate normality ensures that third and fourth central moments equal zero.

Even without analyzing the expression in detail, it is clear that it will generally

[10] E.g., if $E(x) = E(y) = E(z) = 0$, then $E(xyz) = E((x - E(x))(y - E(y))(z - E(z)))$.

be non-zero. Hence, with non-normally distributed explanatory variables, the discussed bias may occur even if the latter are pairwise uncorrelated. While it is the case that misspecified control variables will lead to biased estimates for product terms if the explanatory variables are appropriately correlated[11], the *argumentum e contrario* that uncorrelated explanatory variables prevent the discussed bias from occurring is not valid. In fact, as far as the analysis in this paper goes, pairwise uncorrelatedness only ensures the absence of the discussed bias if either the explanatory variables are multivariate normal (as shown above), or if we can be certain that our explanatory variables also comply with stricter forms of independence than zero covariance[12], for instance in randomized experiments. Consequently, in political research, where most empirical analysis relies on observational data that is neither normally distributed nor statistically independent, controlling for pairwise uncorrelatedness cannot replace correctly specifying the funtional form of our control variables when testing conditional or curvilinear effects.

[11]To be exact, it must be mentioned that we have only provided formal proof for the proposition that covariant explanatory variables produce the illustrated bias for the case of normally distributed data. There is, however, no reason to believe that this should be any different with non-normally distributed data, since mathematically, the only difference between the two cases is that in the latter, third and fourth central cross-moments do not simplify, and thus make a closed-form representation of the bias significantly more complex.

[12]Specifically, Bohrnstedt and Goldberger (1969) show that pairwise conditional independence and homoskedasticity are sufficient for third and fourth cross-moments to reduce to zero.

4 Examples of omitted product terms from the literature

An example of the practice can be found in Wegenast and Basedau (2014). A previous article of one of the authors, Basedau and Lay (2009), argues that states with high oil revenues will be able to avoid civil conflict, for example by repressing or paying off opposition. In a model that includes a variable on oil production per capita as well as its squared term, the authors find some evidence suggesting that oil revenue has an inversely u-shaped effect on the probability of civil conflict breaking out.[13] Wegenast and Basedau (2014) argue that the effect of ethnic fractionalization on ethnic conflict should be higher in resource-rich states. In their empirical test, they split the sample at different levels of oil production per capita and test the effect of ethnic fractionalization on ethnic conflict in each subsample. Despite the previous finding by one of the authors, they do not consider the possibility of a curvilinear effect of oil-production in their models.[14] They also do not consider the possibility of a quadratic effect of ethnic fractionalization, despite stating that previous research has found a curvilinear effect of the variable on conflict. Similarly, the authors do not control for the possibility of other factors moderating any of their variables of interest, despite describing previous research that finds that the effect of ethnic fractionalization on conflict depends on a state's regime type.

In Wood (2014), the author tests whether rebel groups in civil wars kill larger numbers of civilians after experiencing battlefield losses. The author hypoth-

[13] They find this effect only in some specifications, depending on the set of control variables and on which outliers they drop.

[14] They note, however, that they consider this mechanism not to be a threat to their argument as previous research did not find coercive capacity to play an important role in reducing conflict and as only few states have large enough oil revenues for the payoff mechanism to apply.

esizes that several factors moderate the effect of losses: control over territory, support by third states, financing via natural resource wealth and central control. To test these hypotheses at the macro-level, the author includes an interaction term between losses and one of the moderating variables of interest in the model while including the other potentially moderating variables as linear control variables. It is easy to imagine that territorial control and external sources of finance or a financing via resources are correlated.

Another example of the issue can be found in Koga (2011). Here, the author tests whether a country's regime type moderates the effect of other variables on the likelihood of intervening in foreign civil wars on behalf of either the opposition or the government or in general. In a first set of models, the author finds that a country's Polity score moderates the effect of other variables such as for example rebel strength on the government's likelihood to intervene elsewhere on behalf of the opposition. As the argument is specifically about the size of the winning coalition as opposed to the regime type per se, the author replicates these models using the size of the winning coalition instead of the Polity score in the interactions of interest. She notes that as the size of the winning coalition is only one aspect of a state's regime type she controls for the Polity score in the models using the winning coalition's size in order to be able to distinguish effects. For this, she adds the Polity score in a linear fashion *despite* finding in previous models that the Polity score moderates the effects of some of the variables of interest that are now interacted with the selectorate size.

Our final example is Danneman and Ritter (2014). One of the authors' hypotheses is that higher levels of civil conflict in neighbour states increase a state's level of repression. As they want to test whether repression is used before domestic conflict breaks out, they interact the variable on neighbourhood civil war with variables on civil war or dissent in the country in question. They also include a control variable on regime type in a linear fashion, despite citing previous papers that have found non-linearities in the effect of regime type on government repression. Non-linearity in the effect of regime type could be picked up by the interaction terms, especially as civil conflict in a state and its neighbour are likely to be a function of regime type and regime types tend to cluster geographically as well.

5 *Simulation-based illustrations of inferential errors stemming from omitted product terms*

We illustrate the problem of omitted product terms using simulated data with 1000 observations. Consider the first case: Here, the true data generating process is $y_i = \beta_x x_i + \beta_{z_1} z_{1i} + \beta_{xz_1} x_i z_{1i} + \beta_{z_2} z_{2i} + \epsilon_i$.[15] We set all coefficients to the value 1. z_1 and z_2 have means 0, variances 1 and a covariance of .5. Figure 1 in the Manuscript shows results from OLS models using once the correct model specification (top row) and once an incorrect specification that includes an irrelevant interaction term xz_2 and omits the relevant interaction term xz_1 (bottom row).[16] Using the incorrect specification, a researcher draws incorrect conclusions about the moderating variable of x. Naturally, as the interaction term xz_1 is not included in the incorrect model, she finds x to have a constant effect across values of z_1 (bottom left panel). Instead, she finds a significantly positive interaction between x and z_2 with a coefficient of magnitude .49 (bottom right panel) even though z_2 does not moderate the effect of x in the underlying data generating process.

If both xz_1 and xz_2 are relevant in the data generating process ($y_i = \beta_x x_i + \beta_{z_1} z_{1i} + \beta_{xz_1} x_i z_{1i} + \beta_{z_2} z_{2i} + \beta_{xz_2} x_i z_{2i} + \epsilon_i$), the issue of incorrect inference occurs as well. For this illustration we also draw on a simulated dataset of 1000 observations. We set β_{xz_2} to -1 and all other coefficients to 1. z_1 and z_2 have means 0, variances 1

[15]In all examples presented in this section we draw ϵ_i from a normal distribution with mean 0 and variance 1. The same is the case for all variables that are not specified to be correlated with other variables.

[16]Throughout this section, marginal effects of variables with more than one moderating variable have been calculated setting the respective other moderating variable to 0.

and a covariance of .5. The researcher erroneously omits interaction term xz_1 in the incorrectly specified linear model. As the bottom row of Figure A.1 shows, and expectedly, the researcher is again unable to uncover the moderating effect of z_1 on x as no interaction term is included. In addition, the conclusion about the moderating effect of z_2 on x is incorrect: the estimated moderating effect of z_2 also picks up the positive moderating effect of omitted moderator z_1 and the researcher finds the interaction term xz_2 to have a coefficient of -.51 even though the coefficient in the data generating process is -1.

The logic also applies in a case where a relevant squared term that shares a constitutive term with an included interaction is omitted – and vice versa. Consider the following case, again based on 1000 simulated observations. Now the data generating process is $y_i = b_x x_i + b_{z_1} z_{1i} + b_{xx} x_i^2 + \epsilon_i$, so now x has a u-shaped effect. Again, all coefficients are set to 1. Now, x and z_1 have means 0, variances 1 and a covariance of .5. Figure A.2 contrasts OLS results from the correct model specification (top row) with results using an incorrect specification where the researcher erroneously omits the squared term xx and instead includes interaction term xz_1 (bottom row). The incorrectly specified model is unable to uncover the u-shaped effect of x and instead attributes a moderating effect on x to z_1 (interaction coefficient of .89) that does not exist in the true data generating process.

An omitted product term – squared or interactive – can even bias an included product term if it does not share a constitutive term with it. We illustrate this issue in a final example using simulated data with n=1000. Here, the data generating process is $y_i = \beta_{x_1} x_{1i} + \beta_{z_1} z_{1i} + \beta_{x_1 z_1} x_{1i} z_{1i} + \beta_{x_2} x_{2i} + \beta_{z_2} z_{2i} + \beta_{x_2 z_2} x_{2i} z_{2i} + \epsilon_i$

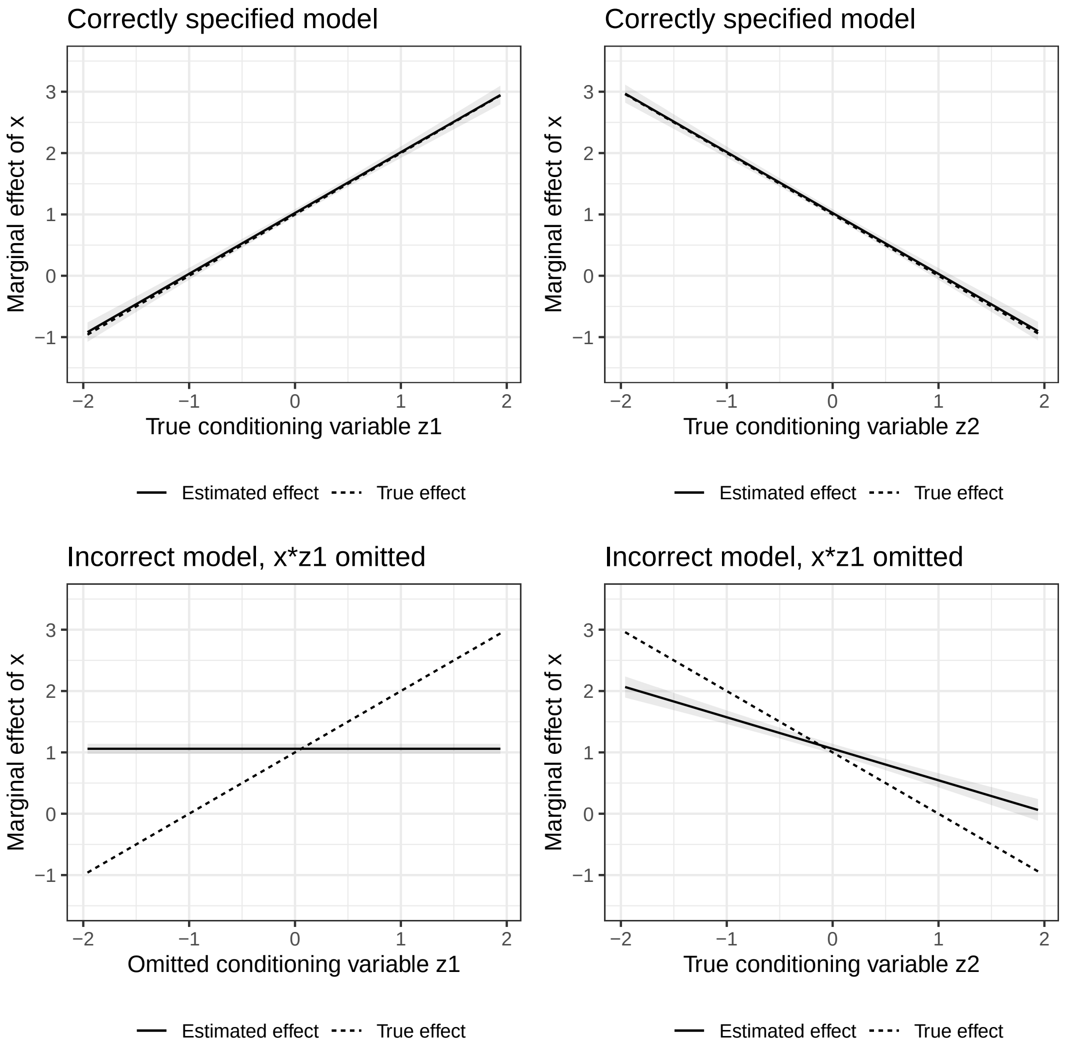

Figure A.1: Conclusions about moderating variables of x based on correct ($y = \beta_x x + \beta_{z_1} z_1 + \beta_{xz_1} x z_1 + \beta_{z_2} z_2 + \beta_{xz_2} x z_2$) and incorrect model specification ($y = \beta_x x + \beta_{z_1} z_1 + \beta_{z_2} z_2 + \beta_{xz_2} x z_2$).

with all coefficients set to 1. Now the two interaction terms do not share any constitutive term. Instead, x_1 and x_2 have means 0, variances 1 and a covariance of .5. The same is the case for z_1 and z_2. As before, the researcher omits the interaction between x_1 and z_1 in the incorrect specification. Figure 2 in the Manuscript compares OLS results using the incorrect specification (bottom row)

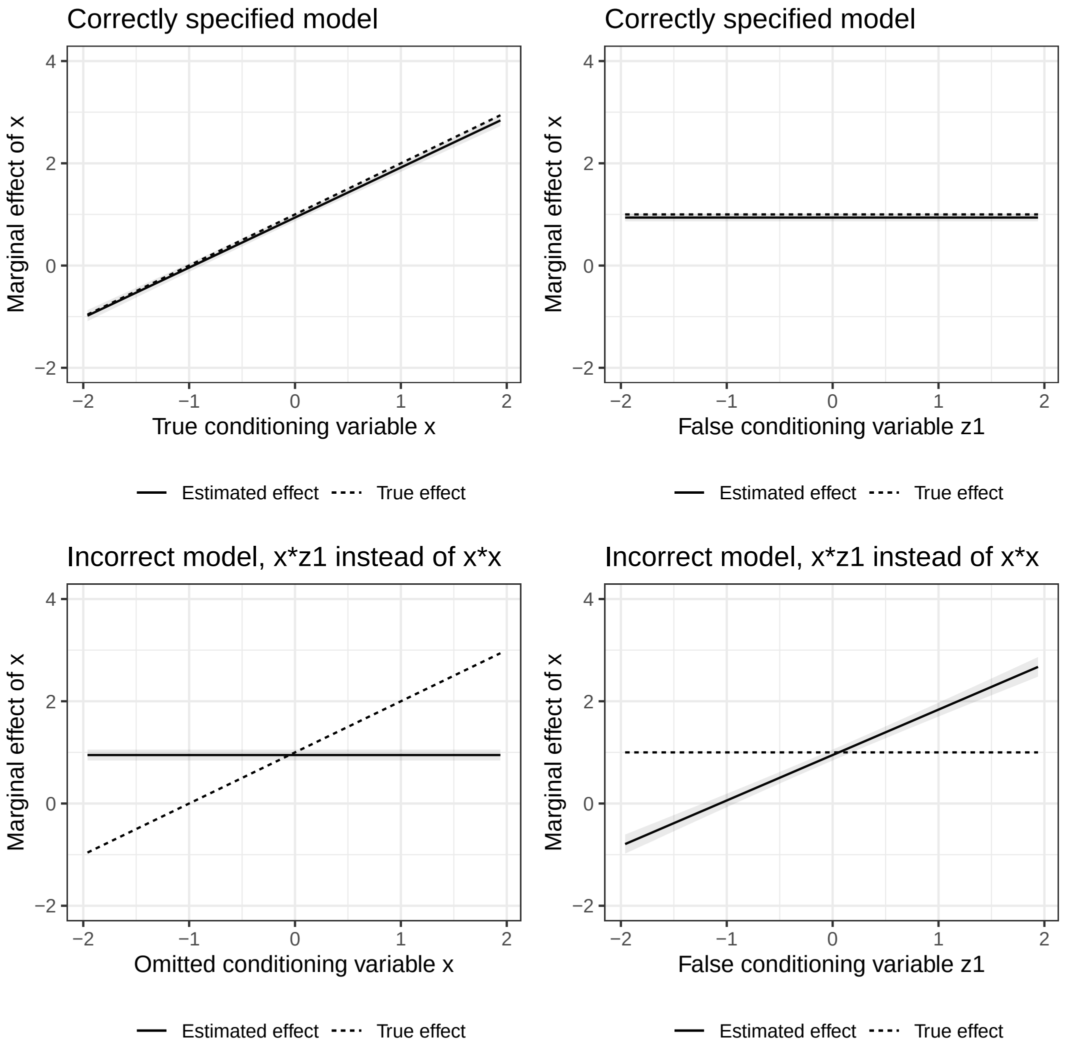

Figure A.2: Conclusions about moderating variables of x based on correct ($y = \beta_x x + \beta_{z1} z_1 + \beta_{xx} xx$) and incorrect model specification ($y = \beta_x x + \beta_{z1} z_1 + \beta_{xz1} x z_1$).

with results from the correctly specified model. Again, the researcher misses that variable x_1 is moderated by variable z_1 as no interaction term has been included. In addition, the moderating effect of variable z_2 on x_2 is overestimated (coefficient of 1.22 when the coefficient in the data generating process is 1) as it also reflects the omitted interaction between x_1 and z_1.

6 *Monte Carlo Experiments - Interactions Between Binary Independent Variables*

Figures A.3 to A.6 displays the results for the Monte Carlo experiments where binary-binary and binary continuous interactions are included. The specification for the relevant variables is:

$$y_i = \beta_a a_i + \beta_b b_i + \beta_c c_i + \beta_{ab} a_i b_i + \beta_{ac} a_i c_i + \epsilon_i \tag{20}$$

However now a and b are binary independent variables. This means that the interaction ab is a binary-binary interaction, and the interaction ac is a binary-continuous interaction. Specifically we make a and b binary variables, such that they equal 1 if the draw from the multivariate normal distribution was above 0, and 0 otherwise. This means that we include binary-binary and binary-continuous interactions within the scenario. The resulting binary variables have a correlation of approximately 0.4 with the continuous variables, when the draws are from a multivariate normal distribution where $\rho = 0.5$. The results for this scenario echo those in the main text. For the scenario the sample size is 500, and the correlation between variables is 0.5.

In general the Monte Carlos find that KRLS and the adaptive Lasso perform well at recovering the conditional effects of the binary variables. However in this circumstance BART does not perform well, with considerable uncertainty and the relevant effects being estimated close to zero. Therefore in such circumstances we would recommend the use of KRLS or the adaptive Lasso over

BART.

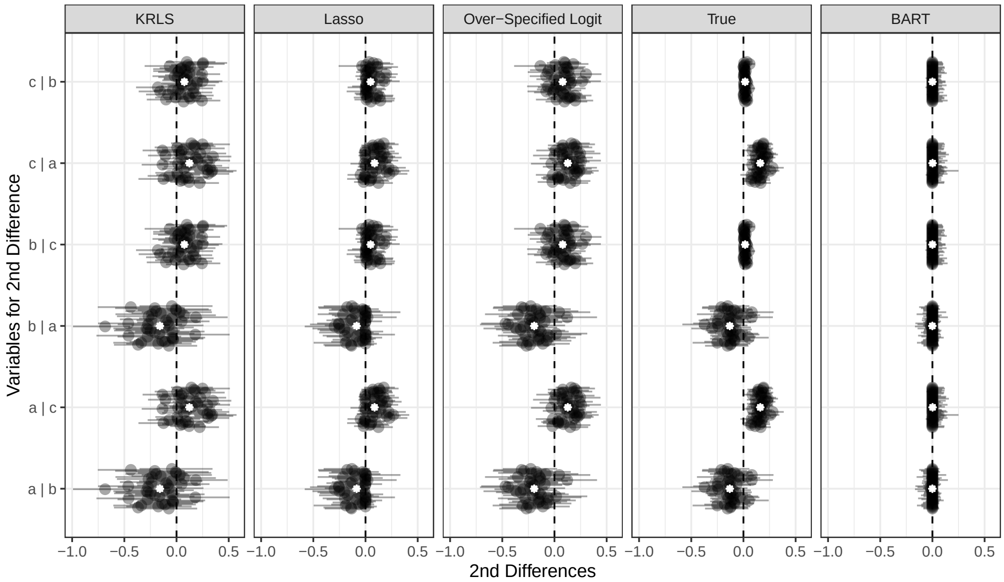

Figure A.3: Estimated second differences for relevant variables. The confidence intervals generated by the adaptive Lasso, KRLS, BART, an over-specified non-regularized parametric model (Logit) and the correctly specified model for each iteration are displayed in grey, while the mean of the second differences for all 240 simulations is displayed in white. Columns indicate the statistical model used, and rows indicate the level of correlation between the independent variables.

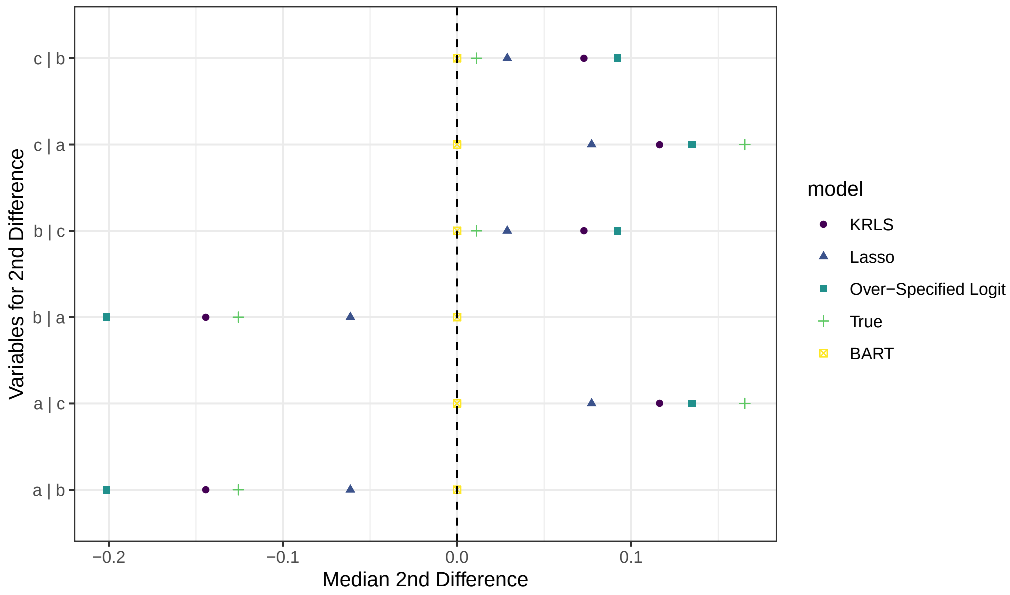

Figure A.4: The median estimated second differences for relevant variables by statistical model used. Columns indicate the level of correlation between the independent variables, while colour and shape indicate the statistical model used.

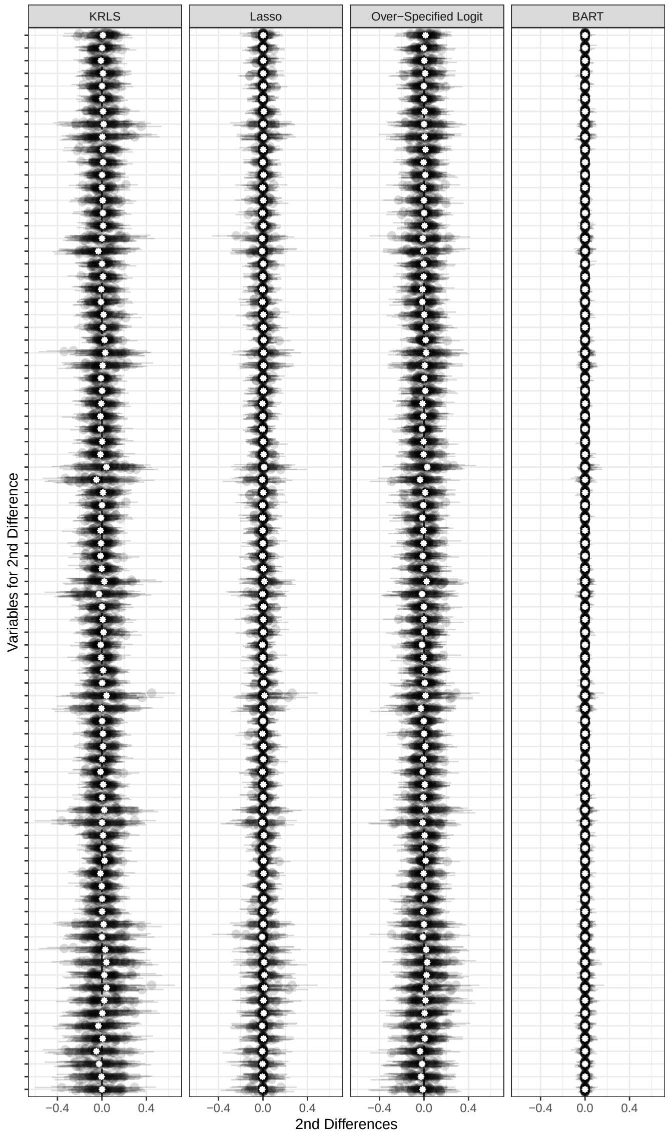

Figure A.5: Estimated second differences for irrelevant product terms. Each white point corresponds to the mean second difference estimated for every combination of variables, that have no conditional effect. All of these second differences should be equal to zero. The confidence/credible intervals generated by the adaptive Lasso, KRLS, BART, and an over-specified non-regularized parametric model (Logit) for each iteration are displayed in grey. Columns indicate the statistical model used, and rows indicate the level of correlation between the independent variables.

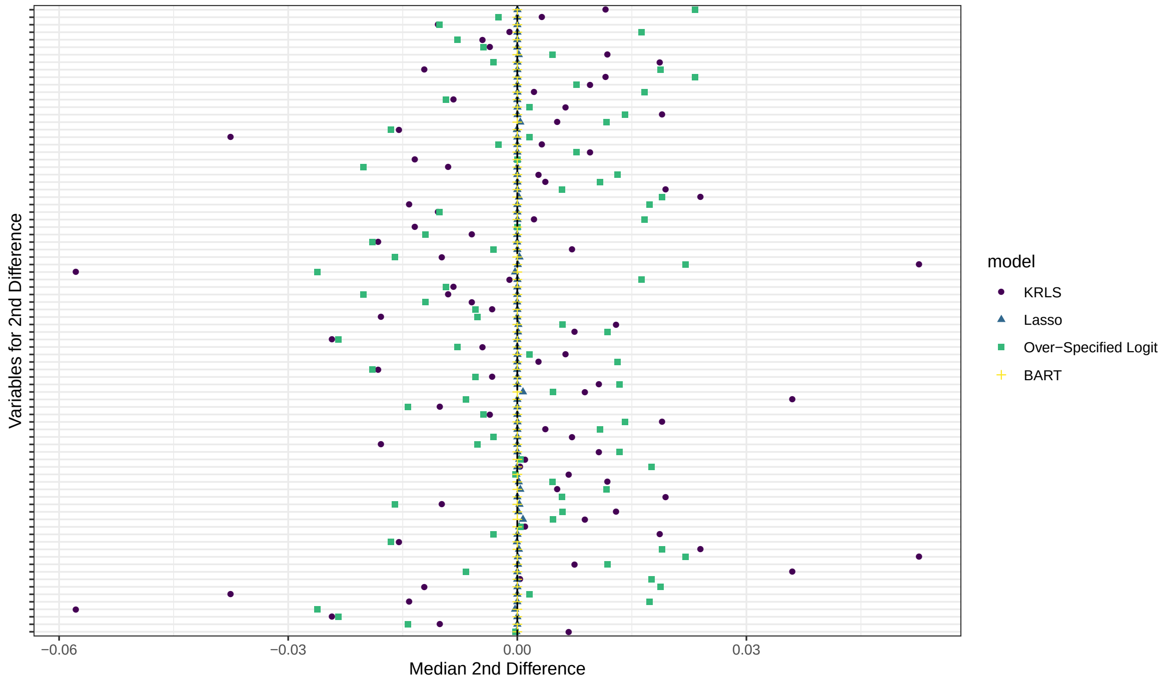

Figure A.6: The median estimated second differences for irrelevant product terms by statistical model used. Each point corresponds to the median second difference estimated for every combination of variables that have no conditional effect. Columns indicate the level of correlation between the independent variables, while colour and shape indicate the statistical model used.

7 *Monte Carlo Experiments - How Inferences Are Sensitive to Number of Covariates Included*

7.1 *Second Differences for Relevant Variables by Number of Covariates*

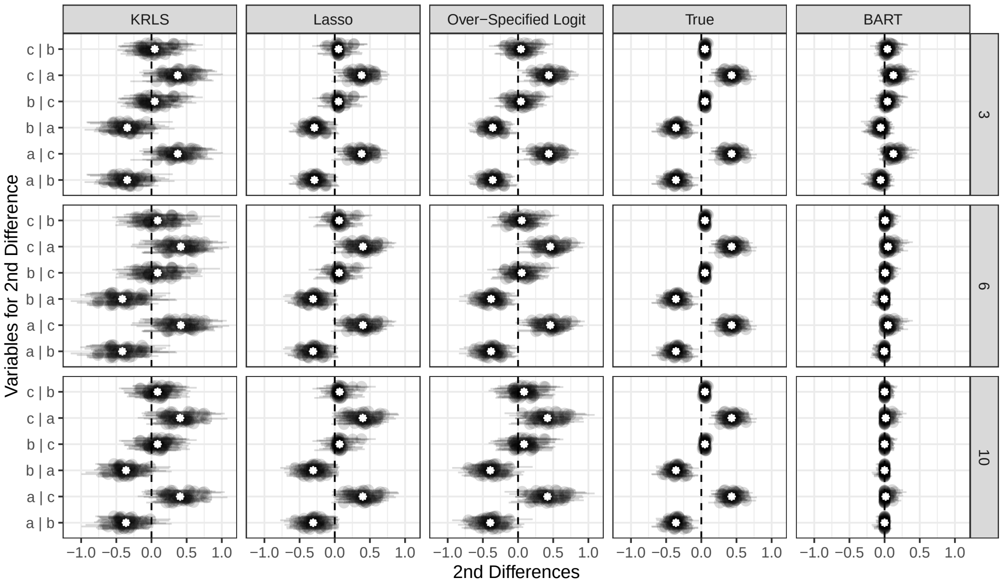

Figure A.7: Estimated second differences for relevant variables. The confidence intervals generated by the adaptive Lasso, KRLS, BART, an over-specified non-regularized parametric model (Logit) and the correctly specified model for each iteration are displayed in grey, while the mean of the second differences for all 240 simulations is displayed in white. Columns indicate the statistical model used, and rows indicate the level of correlation between the independent variables.

7.2 *Median Second Differences for Relevant Variables by Number of Covariates*

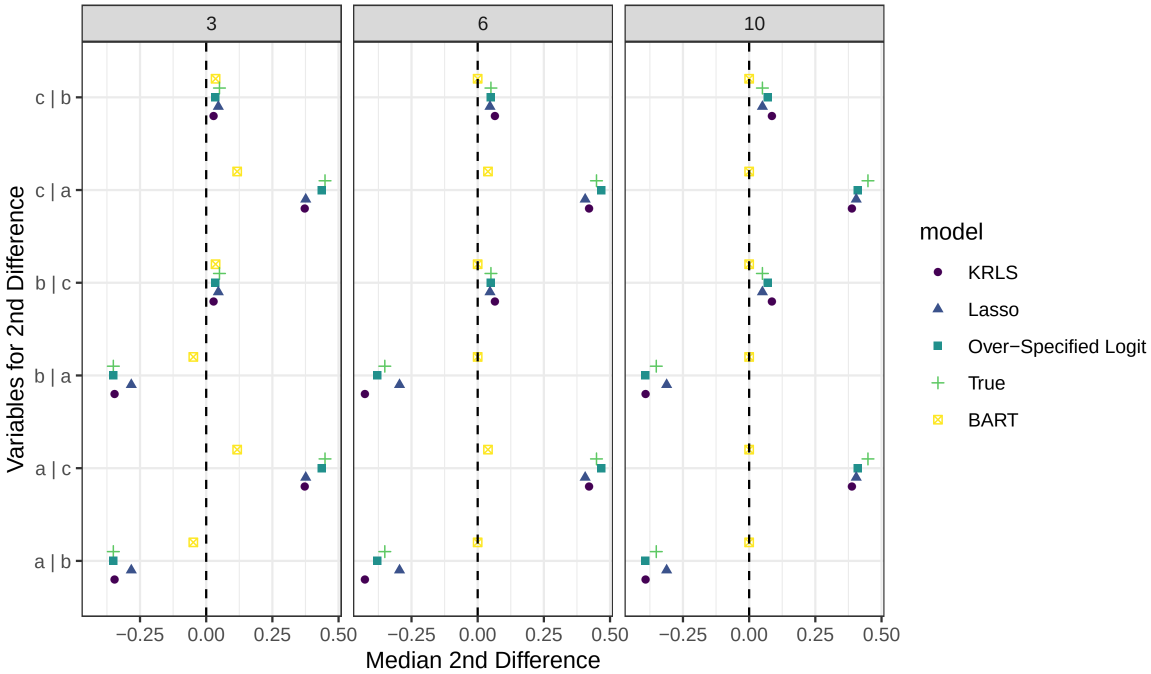

Figure A.8: The median estimated second differences for relevant variables by statistical model used. Columns indicate the level of correlation between the independent variables, while colour and shape indicate the statistical model used.

7.3 *Second Differences for Irrelevant Variables by Number of Covariates*

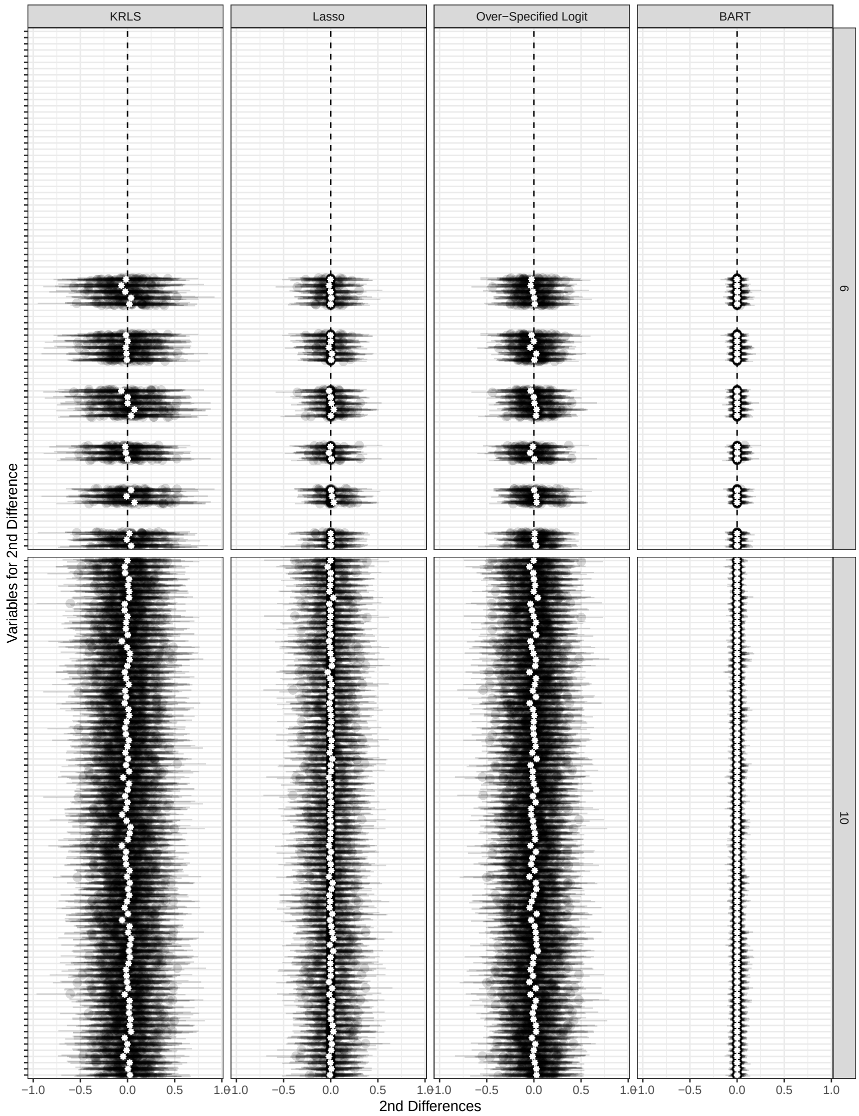

Figure A.9: Estimated second differences for irrelevant variables. The confidence intervals generated by the adaptive Lasso, KRLS, BART, an over-specified non-regularized parametric model (Logit) and the correctly specified model for each iteration are displayed in grey, while the mean of the second differences for all 240 simulations is displayed in white. Columns indicate the statistical model used, and rows indicate the level of correlation between the independent variables.

7.4 Median Second Differences for Irrelevant Variables by Number of Covariates

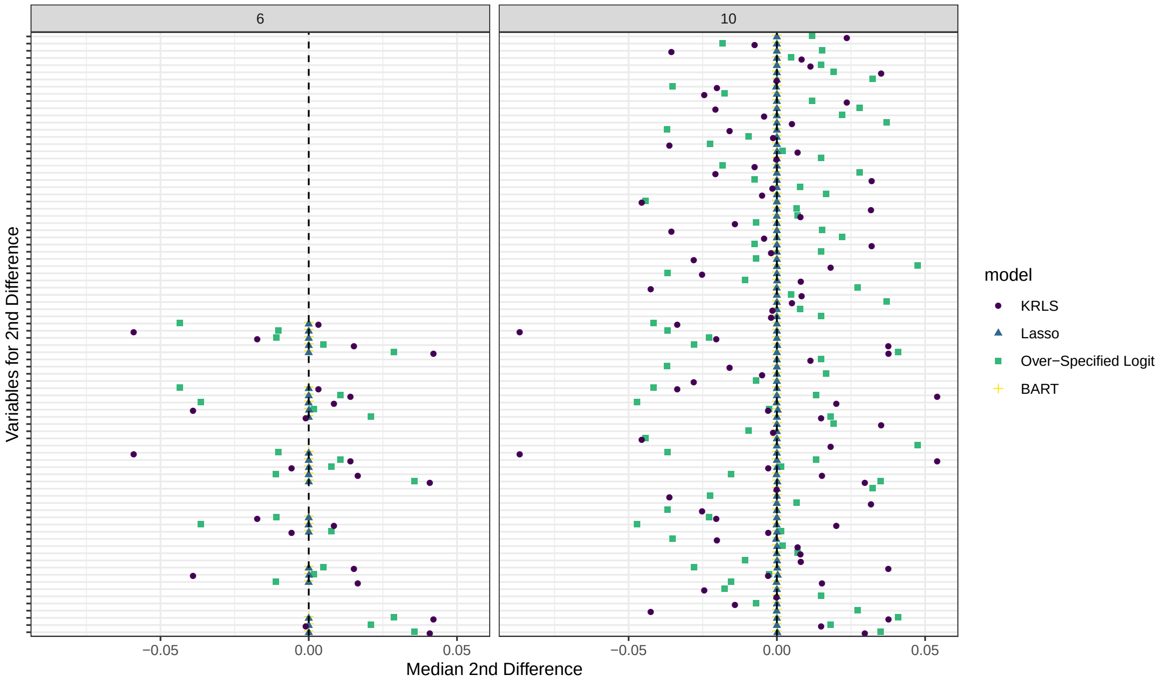

Figure A.10: The median estimated second differences for irrelevant variables by statistical model used. Columns indicate the level of correlation between the independent variables, while colour and shape indicate the statistical model used.

8 *Monte Carlo Experiments - The Size of Uncertainty Associated with Different Estimators*

8.1 *2.5th and 97.5th Percentiles of Second Differences for Irrelevant Variables*

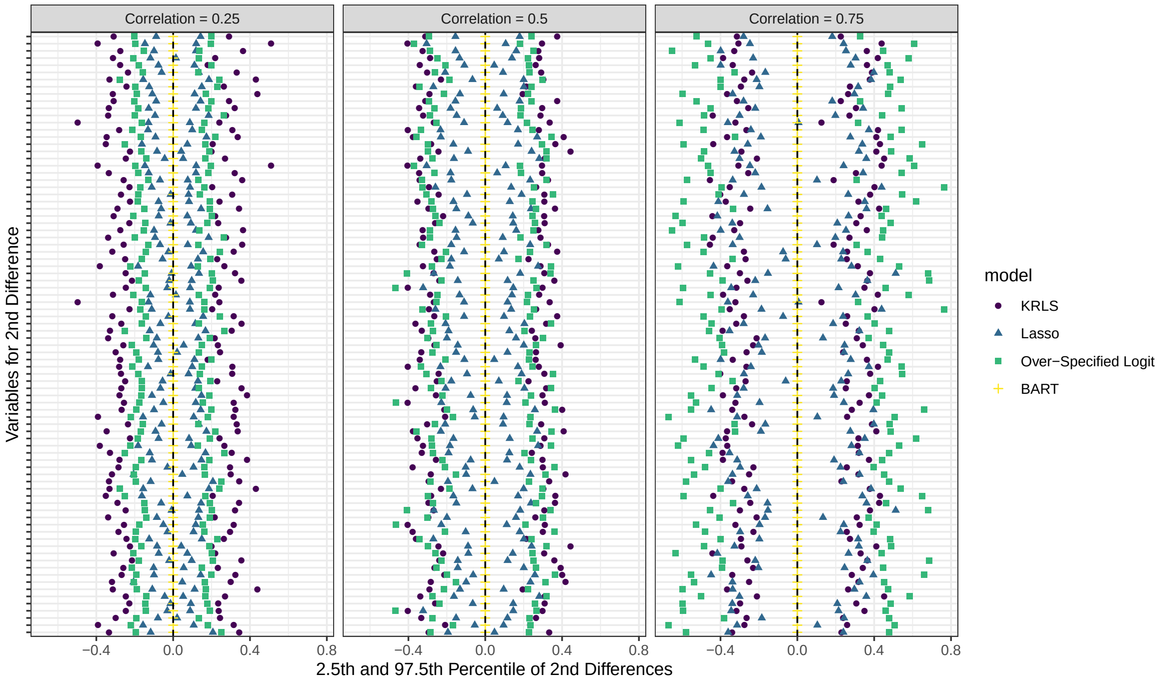

Figure A.11: The 2.5th and 97.5th percentiles of the estimated second differences for irrelevant variables by statistical model used. Columns indicate the level of correlation between the independent variables, while colour and shape indicate the statistical model used.

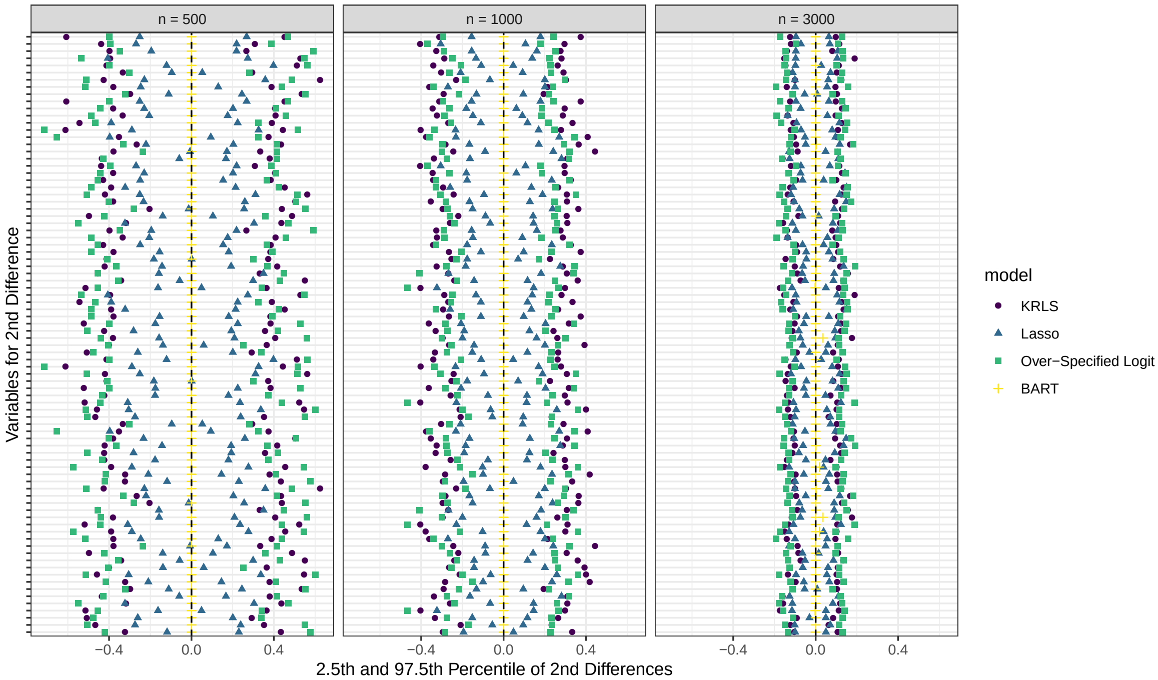

Figure A.12: The 2.5th and 97.5th percentiles of the estimated second differences for irrelevant variables by statistical model used. Columns indicate the number of observations, while colour and shape indicate the statistical model used.

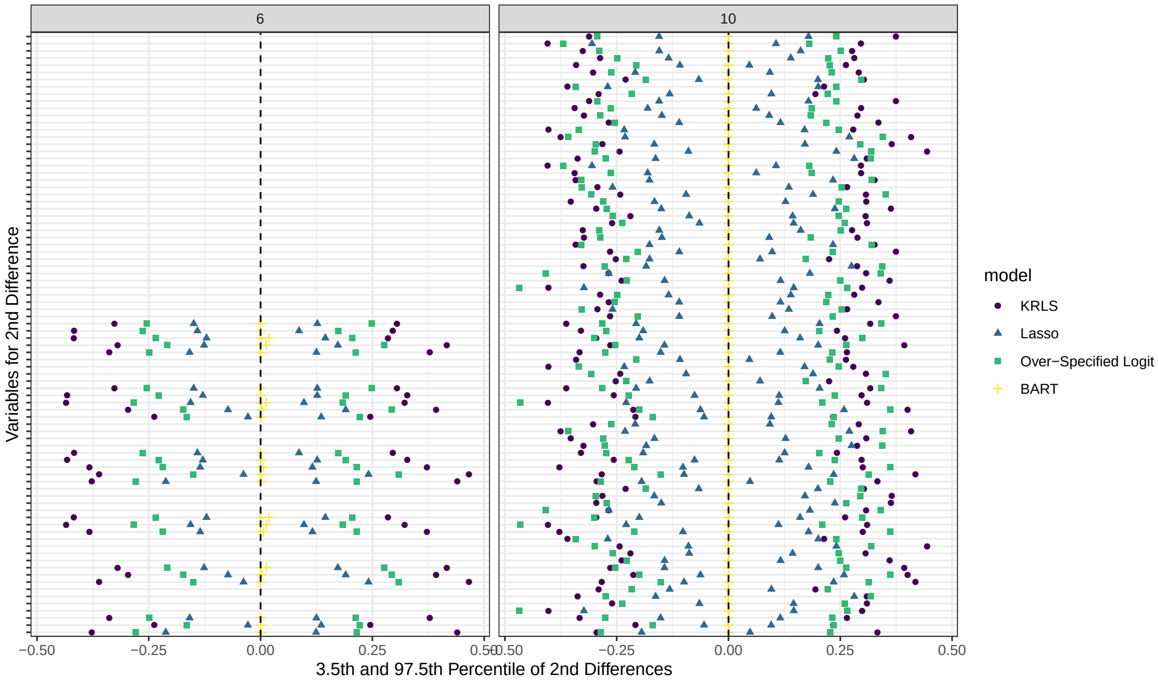

Figure A.13: The 2.5th and 97.5th percentiles of the estimated second differences for irrelevant variables by statistical model used. Columns indicate the number of independent variables included, while colour and shape indicate the statistical model used.

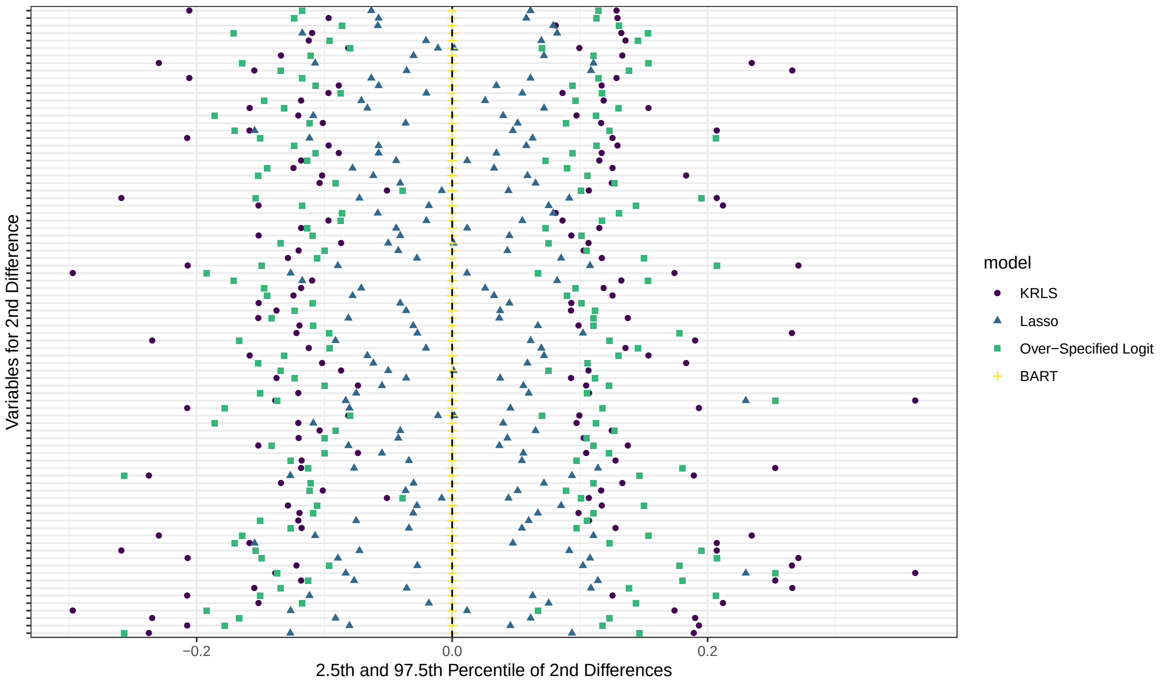

Figure A.14: The 2.5th and 97.5th percentiles of the estimated second differences for irrelevant variables by statistical model used. These models include binary independent variables, as in section 6 of the appendix. Colour and shape indicate the statistical model used.

8.2 *Mean Confidence/Credible Interval Length for Irrelevant Variables*

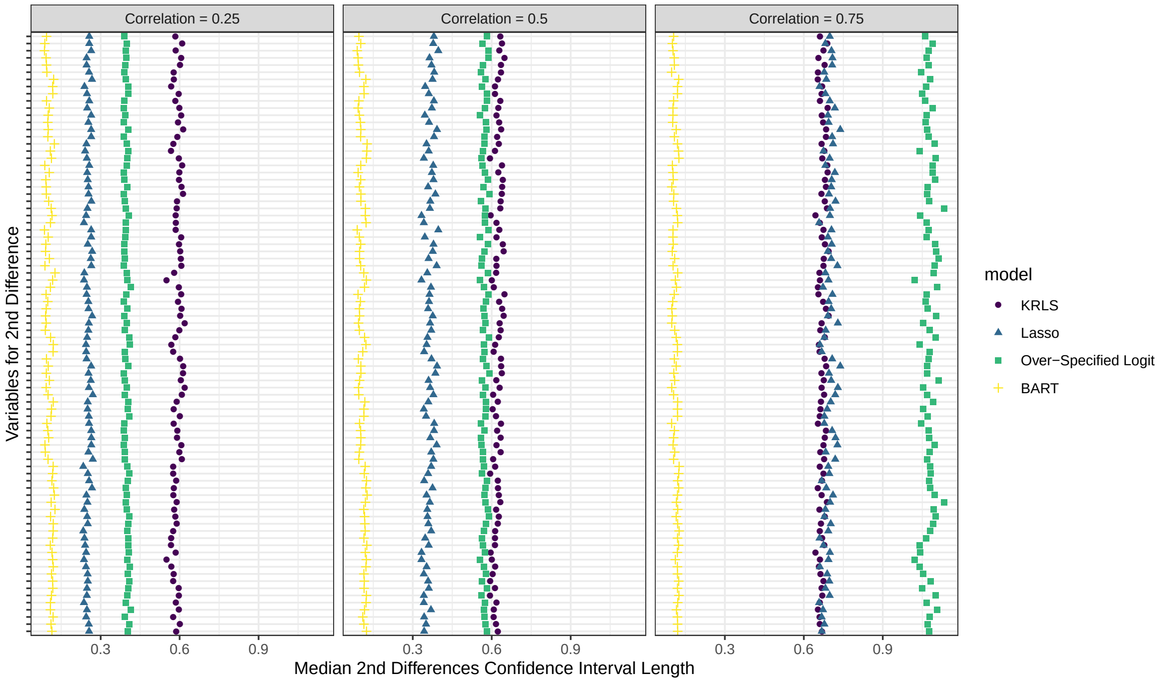

Figure A.15: The mean confidence/credible length for irrelevant variables by statistical model used. Columns indicate the level of correlation between the independent variables, while colour and shape indicate the statistical model used.

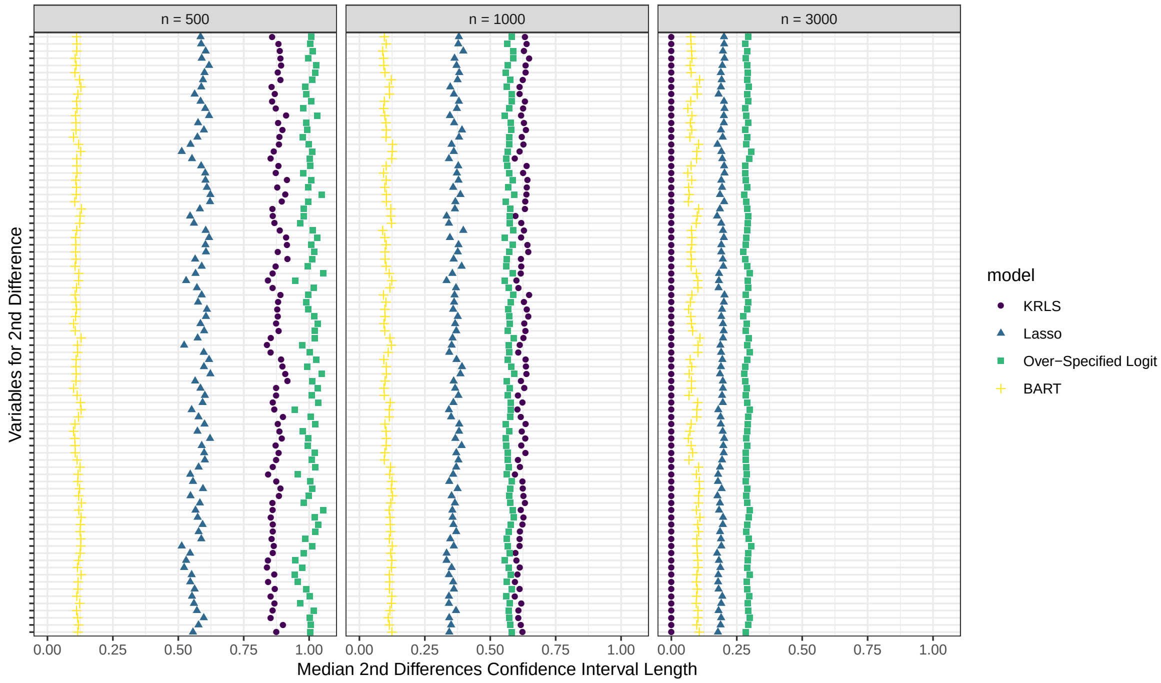

Figure A.16: The mean confidence/credible length for irrelevant variables by statistical model used. Columns indicate the number of observations, while colour and shape indicate the statistical model used.

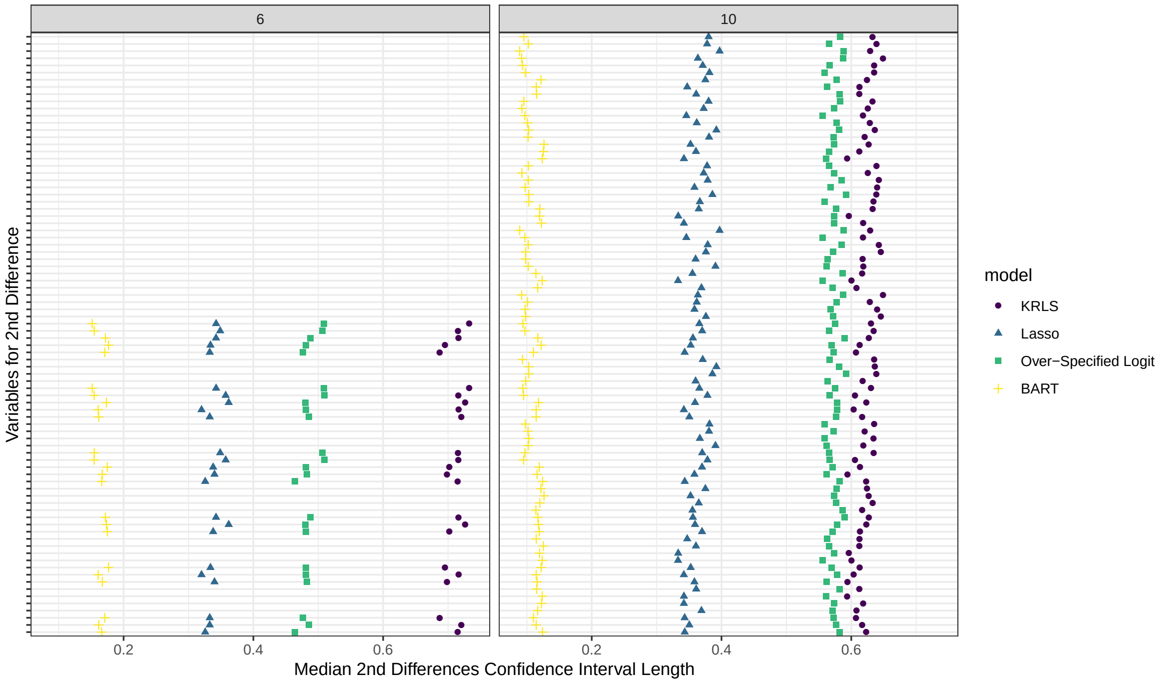

Figure A.17: The mean confidence/credible length for irrelevant variables by statistical model used. Columns indicate the number of independent variables included, while colour and shape indicate the statistical model used.

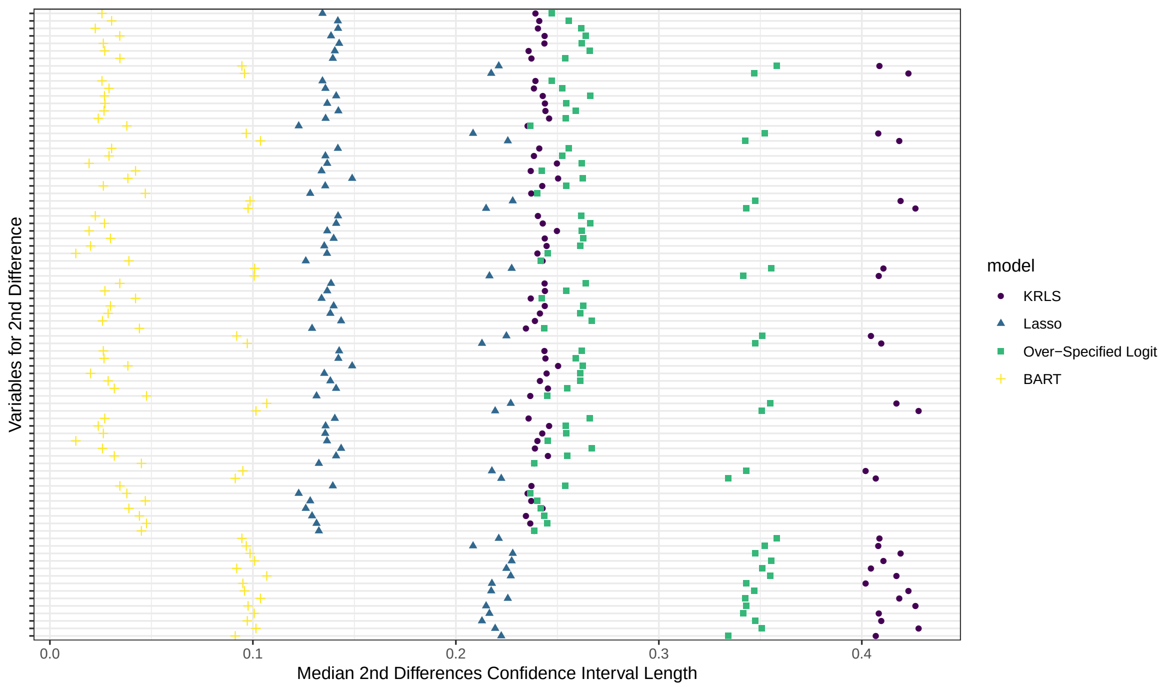

Figure A.18: The mean confidence/credible length for irrelevant variables by statistical model used. These models include binary variables, as in section 6 of the appendix. Colour and shape indicate the statistical model used.

9 *Further Results from the Empirical Illustration using Williams and Whitten (2015)*

9.1 *Restricted Models*

Tables A.1 and A.2, display the results from the original Williams and Whitten specification and the restricted version.[17] If the authors had not included the additional interactive terms alongside the two interactions between the dummies on prime minister and coalition party and economic growth and instead just added all control variables in a non-interactive fashion (Models 2 in Tables A.1 and A.2), results on the marginal effect of growth on different types of parties would not have changed.[18] But the authors' inclusion of additional interactive terms affects conclusions about the effects of the dummy variables on being a coalition party and the party of the prime minister at different levels of GDP per capita growth in the high clarity setting.

[17]Tables are produced using the `stargazer` package in R (Hlavac, 2018).

[18]Results comparing the restricted and the original model refer to a specification with a spatial lag. In the original model, we average over the model observations' values of all additional variables that are interacted with the variables of interest.

Results from the original and a restricted model that only includes interactions between growth and the two dummies on party type are illustrated in Figure A.19. Consider for example the high clarity setting where the degree to which the prime minister's party is punished for low and rewarded for high growth depends on the inclusion of additional interaction terms. Here, the original model finds that prime minister's parties have a general advantage over coalition partners that becomes larger as economic growth increases. The restricted model, on the other hand, finds that in fact coalition parties have an advantage over prime minister's parties but also that this effect diminishes with growth. This difference in overall advantage also affects the degree to which the prime minister's party is punished for low and rewarded for high growth: in the original model, the marginal effect of being the prime minister's party is negative but insignificant when growth is low and becomes positive at a growth rate of about 1.7 per cent, a value between the 33th and the 34th percentile of the sample distribution. The effect becomes significantly positive at a growth rate of about 6.7 per cent. In the restricted model, on the other hand, the dummy on being a prime minister's party has a significantly negative effect when growth is low and only becomes positive at a growth rate of about 9.6 percent, a value that lies between the 97th and the 98th percentile of the sample distribution. Here, the positive effect never reaches statistical significance. Thus, while in the original model economic voting directed towards the prime minister's party manifests itself in rewarding this party for already moderate levels of growth, in a model without additional interactions economic voting with respect to the prime minister's party means punishing the prime minister's party for low levels of growth. In both the restricted and the sophisticated model the coalition

Table A.1: Williams and Whitten (2015): Low Clarity

	Dependent variable: change						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Rho	−0.012***	−0.013***	−0.013***	−0.013***	−0.013***	−0.012***	−0.012***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
GDP Growth	−0.014	−0.001	0.023	0.023	0.023	0.022	0.022
	(0.058)	(0.059)	(0.043)	(0.043)	(0.043)	(0.043)	(0.043)
PMs Party	1.204	−0.469	2.323**	0.057	−0.070	−1.071**	0.717
	(1.802)	(0.480)	(0.964)	(0.389)	(0.686)	(0.476)	(0.513)
Coalition Party	−1.416***	−0.583	−0.775**	−0.842**	−0.817**	−1.636***	−0.935***
	(0.516)	(0.430)	(0.339)	(0.339)	(0.346)	(0.417)	(0.343)
ENP	−0.177	−0.166	−0.174	−0.164	−0.162	−0.160	−0.171
	(0.125)	(0.126)	(0.126)	(0.126)	(0.126)	(0.125)	(0.126)
Vote, t-1	−0.031**	−0.051***	−0.039***	−0.047***	−0.049***	−0.048***	−0.051***
	(0.013)	(0.011)	(0.012)	(0.012)	(0.012)	(0.011)	(0.012)
Niche Party	0.443	−0.365	−0.275	0.128	−0.355	−0.329	−0.360
	(0.541)	(0.345)	(0.347)	(0.541)	(0.347)	(0.343)	(0.346)
# Government Parties	0.153	0.152	0.094	0.149	0.126	0.171	0.149
	(0.150)	(0.143)	(0.144)	(0.144)	(0.151)	(0.142)	(0.143)
Time Left in CIEP	−0.011*	0.005	0.006	0.006	0.006	−0.010*	0.005
	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.006)	(0.005)
Party Shift	0.015*	0.015	0.017*	0.017*	0.017*	0.016*	0.017*
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Party Shift, $t-1$	0.031***	0.027***	0.028***	0.028***	0.027***	0.030***	0.028***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Majority	0.119	0.063	0.101	0.081	0.081	0.096	0.305
	(0.319)	(0.292)	(0.291)	(0.292)	(0.293)	(0.290)	(0.317)
GDP growth x PMs party	0.256**	0.222**					
	(0.109)	(0.109)					
GDP Growth x Coalition Party	−0.076	−0.112					
	(0.106)	(0.106)					
PMs Party x Vote, t-1	−0.084**		−0.069**				
	(0.037)		(0.028)				
Vote, t-1 x Niche party	−0.072*			−0.052			
	(0.044)			(0.044)			
PMs Party x # Government Parties	−0.111				0.085		
	(0.356)				(0.250)		
PMs Party x Time Left in CIEP	0.047***					0.048***	
	(0.013)					(0.012)	
Coalition Party x Time Left in CIEP	0.034***					0.036***	
	(0.013)					(0.012)	
PMs Party x Majority	−0.133						−1.059*
	(0.729)						(0.606)
Constant	1.172**	1.012*	0.883*	0.867	0.942*	1.226**	0.880*
	(0.548)	(0.541)	(0.527)	(0.531)	(0.528)	(0.527)	(0.528)
Observations	1,030	1,030	1,030	1,030	1,030	1,030	1,030
R^2	0.175	0.150	0.150	0.146	0.144	0.161	0.147
Note:							*p<0.1; **p<0.05; ***p<0.01

Table A.2: Williams and Whitten (2015): High Clarity

	Dependent variable:					
	change					
	(1)	(2)	(3)	(4)	(5)	(6)
Rho	−0.004***	−0.004***	−0.004***	−0.004***	−0.004***	−0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
GDP Growth	−0.257**	−0.255**	−0.027	−0.030	−0.025	−0.026
	(0.100)	(0.104)	(0.076)	(0.078)	(0.078)	(0.077)
PMs Party	6.937**	−5.082***	5.447***	−3.915***	−4.881***	−5.345***
	(3.274)	(0.890)	(2.092)	(0.744)	(1.067)	(0.858)
Coalition Party	−4.092***	−3.500***	−1.524**	−2.068***	−1.870***	−2.795***
	(1.008)	(0.870)	(0.668)	(0.675)	(0.691)	(0.795)
ENP	−0.242	0.073	−0.294	0.167	0.060	0.123
	(0.294)	(0.288)	(0.295)	(0.295)	(0.291)	(0.286)
Vote, t-1	0.001	−0.043*	−0.005	−0.029	−0.033	−0.033
	(0.023)	(0.022)	(0.022)	(0.023)	(0.022)	(0.022)
Niche Party	−0.120	−2.079***	−1.501**	−0.686	−1.905**	−1.925***
	(1.216)	(0.734)	(0.731)	(1.270)	(0.746)	(0.732)
# Government Parties	0.155	−0.005	0.024	−0.024	−0.097	−0.040
	(0.215)	(0.205)	(0.201)	(0.207)	(0.216)	(0.204)
Time Left in CIEP	−0.031*	0.011	0.005	0.009	0.008	−0.032*
	(0.017)	(0.012)	(0.012)	(0.012)	(0.012)	(0.018)
Party Shift	0.004	0.006	0.004	0.001	0.003	0.002
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Party Shift, $t-1$	0.009	0.013	0.012	0.010	0.014	0.014
	(0.014)	(0.014)	(0.014)	(0.015)	(0.015)	(0.015)
GDP growth x PMs party	0.526***	0.525***				
	(0.195)	(0.199)				
GDP Growth x Coalition Party	0.460***	0.445**				
	(0.170)	(0.173)				
PMs Party x Vote, t-1	−0.314***		−0.254***			
	(0.068)		(0.054)			
Vote, t-1 x Niche party	−0.134			−0.126		
	(0.091)			(0.096)		
PMs Party x # Government Parties	−0.961**				0.500	
	(0.430)				(0.350)	
PMs Party x Time Left in CIEP	0.092***					0.105***
	(0.029)					(0.030)
Coalition Party x Time Left in CIEP	0.058**					0.051*
	(0.026)					(0.027)
Constant	2.638**	2.276**	2.000**	1.115	1.687*	1.925*
	(1.040)	(1.030)	(0.989)	(1.039)	(1.021)	(1.007)
Observations	398	398	398	398	398	398
R^2	0.277	0.202	0.225	0.184	0.185	0.207
Note:						*p<0.1; **p<0.05; ***p<0.01

partner is punished until growth reaches about 6.7 and 7.7 per cent respectively, both above the 90th percentile of the distribution. This means that while in the restricted model there is not much difference in how economic voting affects prime minister's and coalition parties, the sophisticated model suggests that coalition partners are punished for low growth while prime minister's parties benefit from high growth.

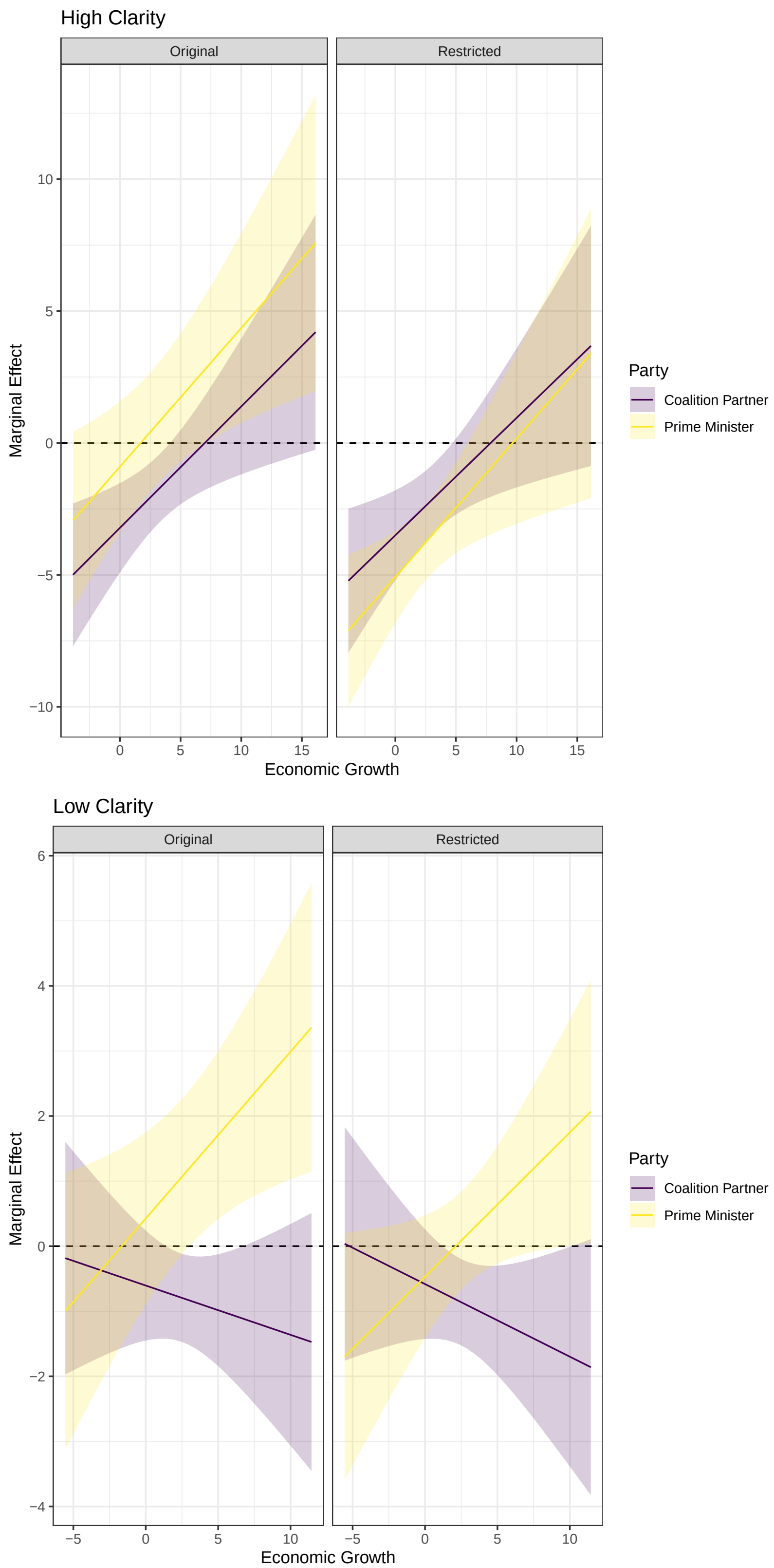

Figure A.19: Marginal effect of party type at different levels of growth. Comparison category is opposition party.

Table A.3: Variable Labels

Variable Label	Variable Name
Outcome variable	change
GDP Growth	rgdppc_growth
Rho	w_change
PM's Party	prime_dummy
Coalition Party	xregbet
ENP	eff_par
$Vote_{t-1}$	lag_pervote
Niche Party	niche
# Government Parties	gparties
Time Left in CIEP	ciep_perc
$PartyShift_t$	party_shift_t
$PartyShift_{t-1}$	party_shift_t1
Majority	majority

9.2 Adaptive Lasso

Tables A.4 and A.5 display the results from using the adaptive Lasso on the variables included by Williams and Whitten. The variable names are located in A.3

Figure A.20 shows results on party type depending on growth using the adaptive Lasso at a polynomial expansion of degree two.

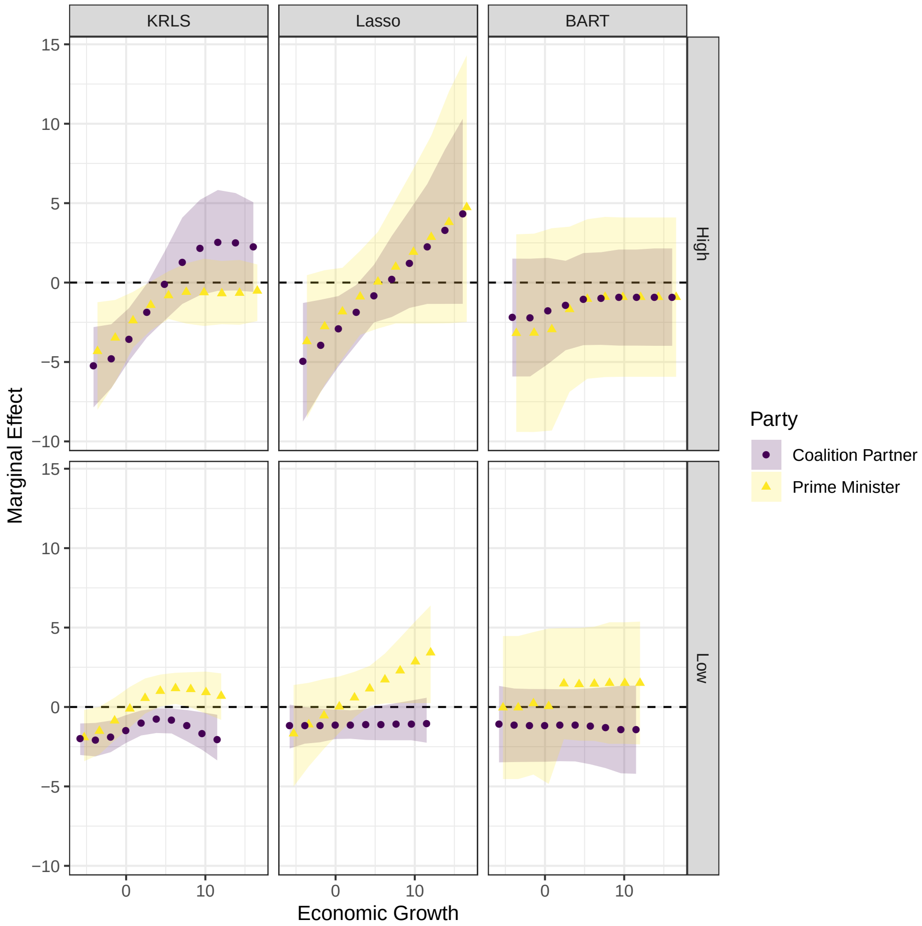

Figure A.20: Effect of party type at different levels of growth in high and low clarity settings. Comparison category is opposition party. Lasso degree of polynomial expansion: 2.

10 *Review of the Existing Literature*

We have surveyed 40 recent papers in top journals and top subfield journals of Political Science that cite Brambor, Clark and Golder's seminal paper. More specifically, the papers surveyed were published in the American Journal of Political Science, the American Political Science Review, Comparative Political Studies, International Organization, the Journal of Conflict Resolution, and the Journal of Politics between May 2014 and April 2016. One paper was excluded as it did not include a product term in its specification, but instead cites Brambor, Clark and Golder for the use of simulation to generate predicted probabilities. The following table displays the results from doing so.

Table A.4: Williams and Whitten (2015) - Adaptive Lasso for Low Clarity Sample

	Estimate	Std. Error	2.5%	97.5%
(Intercept)	1.34	1.13	-0.12	4.25
w_change	-0.01	0.01	-0.02	0.00
xregbet	-2.38	1.11	-4.68	-0.57
eff_par	-0.14	0.44	-1.53	0.16
lag_pervote	0.02	0.06	-0.01	0.20
ciep_perc	-0.02	0.02	-0.06	0.01
party_shift_t	0.01	0.03	-0.01	0.11
w_change.eff_par	-0.00	0.00	-0.01	0.00
w_change.lag_pervote	-0.00	0.00	-0.00	-0.00
w_change.niche	0.01	0.00	0.00	0.01
w_change.gparties	0.00	0.00	0.00	0.01
w_change.party_shift_t	-0.00	0.00	-0.00	0.00
w_change.party_shift_t1	-0.00	0.00	-0.00	0.00
w_change.majority	0.00	0.00	-0.00	0.01
rgdppc_growth.prime_dummy	0.27	0.17	0.00	0.64
rgdppc_growth.lag_pervote	-0.00	0.00	-0.01	0.00
rgdppc_growth.gparties	-0.04	0.04	-0.12	0.00
rgdppc_growth.party_shift_t	-0.00	0.00	-0.01	0.00
rgdppc_growth.majority	0.10	0.11	0.00	0.33
prime_dummy.eff_par	0.05	0.33	0.00	1.01
prime_dummy.lag_pervote	-0.05	0.04	-0.12	0.00
prime_dummy.gparties	-0.02	0.43	-1.38	0.25
prime_dummy.ciep_perc	0.04	0.02	0.00	0.06
prime_dummy.party_shift_t	0.03	0.03	0.00	0.11
xregbet.eff_par	0.15	0.20	0.00	0.61
xregbet.lag_pervote	-0.04	0.03	-0.10	0.00
xregbet.gparties	0.24	0.30	0.00	0.92
xregbet.ciep_perc	0.04	0.02	0.00	0.08
eff_par.lag_pervote	-0.01	0.01	-0.04	0.00
eff_par.party_shift_t1	0.01	0.01	0.00	0.03
lag_pervote^2	-0.00	0.00	-0.00	0.00
lag_pervote.niche	-0.06	0.04	-0.14	0.00
lag_pervote.gparties	0.00	0.01	0.00	0.04
lag_pervote.ciep_perc	0.00	0.00	0.00	0.00
niche.ciep_perc	0.01	0.01	0.00	0.04
niche.party_shift_t1	0.00	0.02	0.00	0.06
gparties^2	0.03	0.04	-0.05	0.13
gparties.party_shift_t1	-0.00	0.01	-0.03	0.00
ciep_perc.majority	-0.01	0.01	-0.04	0.00
party_shift_t.party_shift_t1	0.00	0.00	0.00	0.00
party_shift_t1^2	0.00	0.00	0.00	0.00

Table A.5: Williams and Whitten (2015) - Adaptive Lasso for High Clarity Sample

	Estimate	Std. Error	2.5%	97.5%
(Intercept)	1.13	2.71	-0.44	9.11
eff_par	-0.14	1.68	-4.91	1.77
w_change.lag_pervote	-0.00	0.00	-0.00	0.00
rgdppc_growth.prime_dummy	0.08	0.34	0.00	1.14
prime_dummy.lag_pervote	-0.13	0.09	-0.39	0.00
prime_dummy.ciep_perc	0.05	0.04	0.03	0.18
xregbet.lag_pervote	-0.05	0.07	-0.22	0.00
eff_par.party_shift_t1	0.00	0.01	-0.01	0.03
lag_pervote.niche	-0.08	0.10	-0.36	0.00

Table A.6: Coding of Existing Literature

Author(s)	Year	doi	# Hyp. Product Terms	All in Model	Some in Model	Product Term Controls
Holtermann, H	2016	10.1177/0022002714540470	6	0	1	0
Goplerud, M; Schleiter, P	2016	10.1177/0010414015612393	1			0
Fraga, BL	2016	10.1111/ajps.12172	1			1
Bodea, C; Hicks, R	2015	10.1017/S0020818314000277	3	0	0	0
Poast, P	2015	10.1017/S0020818314000265	1			0
Owen, E	2015	10.1177/0010414015592641	1			0
Schumacher, G; van de Wardt, M; Vis, B; Klitgaard, MB	2015	10.1111/ajps.12174	1			1
Hanegraaff, M; Braun, C; De Bievre, D; Beyers, J	2015	10.1177/0010414015591363	3	1		0
Cheon, A; Urpelainen, J	2015	10.1177/0022002713520529	1			1
Hobolt, SB; de Vries, CE	2015	10.1177/0010414015575030	1			0
Hinkle, RK	2015	10.1086/681059	1			0
Hall, MEK; Ura, JD	2015	10.1086/681437	1			0
Ennser-Jedenastik, L	2015	10.1177/0010414014558259	2	1		0
Williams, LK; Whitten, GD	2015	10.1111/ajps.12124	2	1		1
Branton, R; Martinez-Ebers, V; Carey, TE; Matsubayashi, T	2015	10.1111/ajps.12159	3	0	0	0
DeMeritt, JHR	2015	10.1177/0022002713515406	1			0
Andre, A; Depauw, S; Martin, S	2015	10.1177/0010414014545512	3	0	0	0
Leiras, M; Tunon, G; Giraudy, A	2015	10.1086/678975	1			0
Bodea, C; Hicks, R	2015	10.1086/678987	1			0
Tenorio, BZ	2014	10.1177/0010414013519409	2	1		0
Barabas, J; Jerit, J; Pollock, W; Rainey, C	2014	10.1017/S0003055414000392	1			0
De Sio, L; Weber, T	2014	10.1017/S0003055414000379	1			1
Neudorfer, NS; Theuerkauf, UG	2014	10.1177/0010414013516919	1			1
van de Wardt, M; De Vries, CE; Hobolt, SB	2014	10.1017/S0022381614000565	4	1		1

Wood, RM	2014	10.1017/S0020818314000204	4	0	0	0
Touchton, M; Wampler, B	2014	10.1177/0010414013512601	1			0
Fair, CC; Malhotra, N; Shapiro, JN	2014	10.1177/0022002713478564	3	0	1	1
Shea, PE	2014	10.1177/0022002713478567	1			0
Getmansky, A; Zeitzoff, T	2014	10.1017/S0003055414000288	1			0
Lehoucq, F; Perez-Linan, A	2014	10.1177/0010414013488561	2	1		0
Testa, PF; Hibbing, MV; Ritchie, M	2014	10.1017/S0022381614000255	1			0
Horowitz, MC; Stam, AC	2014	10.1017/S0020818314000046	3	1		0
Gunitsky, S	2014	10.1017/S0020818314000113	1			0
Salehyan, I; Siroky, D; Wood, RM	2014	10.1017/S002081831400006X	1			0
Peterson, TM	2014	10.1177/0022002713478794	1			1
Avdan, N	2014	10.1177/0022002713478795	3	0	0	0
Zeigler, S; Pierskalla, JH; Mazumder, S	2014	10.1177/0022002713478561	1			0
Crespo-Tenorio, A; Jensen, NM; Rosas, G	2014	10.1177/0010414013488559	2	0	1	1
Truex, R	2014	10.1017/S0003055414000112	2	0	0	0

References

Basedau, Matthias and Jann Lay. 2009. "Resource Curse or Rentier Peace? The Ambiguous Effects of Oil Wealth and Oil Dependence on Violent Conflict." *Journal of Peace Research* 46(6):757–776.

Bohrnstedt, George W. and Arthur S. Goldberger. 1969. "On the Exact Covariance of Products of Random Variables." *Journal of the American Statistical Association* 64(328):1439–1442.

Danneman, Nathan and Emily Hencken Ritter. 2014. "Contagious Rebellion and Preemptive Repression." *Journal of Conflict Resolution* 58(2):254–279.

Greene, William H. 2003. *Econometric Analysis*. 5. ed. Upper Saddle River, NJ: Prentice Hall.

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

Hlavac, Marek. 2018. *stargazer: Well-Formatted Regression and Summary Statistics Tables*. Bratislava, Slovakia: Central European Labour Studies Institute (CELSI).

Isserlis, L. 1918. "On a Formula for the Product-Moment Coefficient of any Order of a Normal Frequency Distribution in any Number of Variables." *Biometrika* 12(1/2):134–139.

Koga, Jun. 2011. "Where Do Third Parties Intervene? Third Parties' Domestic

Institutions and Military Interventions in Civil Conflicts." *International Studies Quarterly* 55(4):1143–1166.

Wegenast, Tim C and Matthias Basedau. 2014. "Ethnic fractionalization, natural resources and armed conflict." *Conflict Management and Peace Science* 31(4):432–457.

Williams, Laron K and Guy D. Whitten. 2015. "Don't Stand So Close to Me: Spatial Contagion Effects and Party Competition." *American Journal of Political Science* 59(2):309–325.

Wood, Reed M. 2014. "From Loss to Looting? Battlefield Costs and Rebel Incentives for Violence." *International Organization* 68(4):979–999.