CONSIRT Cross-National Studies: Interdisciplinary Research and Training Program

Working Papers Series

9 Costs and Benefits of Including or Omitting Interaction Terms: A Monte Carlo Simulation

Małgorzata Mikucka
Université catholique de Louvain, Belgium and HSE-LCSR, Russia
mikucka.m@gmail.com
Francesco Sarracino
Statistical Office of Luxembourg and HSE-LCSR, Russia
Joshua Kjerulf Dubrow
Institute of Philosophy and Sociology, Polish Academy of Sciences

CONSIRT, The Ohio State University and the Polish Academy of Sciences
July 2015

Mission

CONSIRT Working Papers are high quality, cross-national, comparative, English language scholarly work that communicates new ideas and has clear contributions to the social sciences. We encourage empirical and methodological papers with deep theoretical roots. Interdisciplinary work in this regard is also encouraged. Our mission is to promote these works in an open, electronic forum for the benefit of the international social science community.

What CONSIRT Working Papers Are

Papers in the CONSIRT Working Papers Series are pre-publication versions of the author(s)' academic work. Working Papers posted on consirt.osu.edu are in progress, under submission, or forthcoming. Working Papers are produced on this site by the author(s) for the consumption of the international scientific community and others interested in this work. CONSIRT Working Papers are not subject to double-blind peer review. CONSIRT reviewed the Working Paper submission for general suitability, including criteria for professional scholarly work, but the quality of the Working Paper itself is the responsibility of the author(s). The form and content of papers are the responsibility of individual authors. The cover sheet is standard across all papers (subject to change), but the format of the rest of the work is not standardized. Comments on papers or questions about their content should be sent directly to the author(s), at their email address.

Citation of CONSIRT Working Papers

Working Papers may be cited without seeking prior permission from the author(s). The proper form for citing CONSIRT Working Papers is:

> Author(s). Year. "Title." CONSIRT Working Papers Series [number in the series] at consirt.osu.edu.

Once a CONSIRT Working Paper has been published elsewhere, it is customary that it be cited in its final, published version, rather than in its Working Paper version.

Copyright Statement

Copyright to CONSIRT Working Papers remains with the author(s). Posting a Working Paper on consirt.osu.edu does not preclude simultaneous or subsequent publication elsewhere, including other working papers series. It is the current policy of CONSIRT to maintain all Working Papers posted on the CONSIRT website unless otherwise notified by the author(s). Users of this website may consume papers for their own personal use, but downloading of papers for any other activity, including but not limited to electronic reposting to the internet in any way, may only be done with the written consent of the author(s). It is the author(s)'s responsibility to know if copyright has been transferred upon publication elsewhere and to ask CONSIRT to have it removed from the site, if necessary.

Costs and Benefits of Including or Omitting Interaction Terms: A Monte Carlo Simulation

Małgorzata Mikucka
Université Catholique de Louvain, Belgium and HSE-LCSR, Russia

Francesco Sarracino
Statistical Office of Luxembourg and HSE-LCSR, Russia

Joshua Kjerulf Dubrow
Institute of Philosophy and Sociology, Polish Academy of Sciences

July 1, 2015

Abstract

What costs or benefits might accrue from including or omitting interaction terms? Failing to account for theoretically important interaction creates the risk of omitted variable bias. Yet, introducing many interaction terms whose effects are small may unnecessarily complicate the model and its interpretation. This analysis uses a Monte Carlo simulation to investigate how wrongly omitting or including an interaction term affects the predictive power of models and the bias of coefficients. Wrongly including an interaction term has little effect on the bias of estimates or on the loss of predictive power. Wrongly omitting an interaction term creates larger biases and a loss of adjusted R^2. The bias of estimates is largest in the case of data with dichotomous variables and depends on the size of the interaction effect, whereas the loss of predictive power is largest in the case of the interaction of two strongly correlated continuous variables, and in data-sets with small error variance.

Keywords: bias, interaction term, intersectionality, Monte Carlo simulation

1 Introduction

What costs or benefits might accrue from including or omitting interaction terms? Researchers motivated to include interactions in their model face the risk of mis-specification. Omitting theoretically justified interaction terms whose effect is substantial may bias the estimates and lead to wrong conclusions. Including an interaction term whose effect size is small may unnecessarily complicate the model and may not improve the quality of the estimate or the predictive power of the model. Our aim is to test whether, and under which conditions, a model with interaction terms is preferable over a model accounting only for main effects. To test the validity of the model with interactions and to compare it with the non-interaction-term model, we use a Monte Carlo simulation of an artificial data-set containing 2000 observations and three variables. This strategy enables us to know beforehand the true nature of the relationship among variables, thus providing a benchmark against which to evaluate the performance of the interaction-term model versus the one without the interaction terms.

All social science researchers face the dilemma of whether to include interaction terms, but those working in the intersectionality paradigm, whose models require the intersection of two or more demographic categories, cannot avoid the problem. In the social sciences, the theoretical idea of intersectionality, i.e. that people are comprised of multiple, overlapping demographics rooted in social structure, is commonly accepted (McCall, 2005). In the quantitative analysis of intersectionality featuring social surveys, researchers face a situation in which interaction terms are required (Dubrow, 2008, 2013, Hughes, 2015, Weldon, 2006). Additive models assume that each individual social category - by itself or in combination with others - contributes to the outcome variable. In contrast, interaction models explicitly account for intersectionality, in that the influence of demographics on a social outcome is conditional on the intersections of the demographic categories. The intersectional approach advocates for interaction terms of categories that identify demographic groups (Dubrow, 2008, 2013, Weldon, 2006). There are few methodological illustrations that show whether intersectional models - those with interaction terms between demographic categories - are an improvement over models that do not employ interaction terms. We provide a methodological illustration of this.

2 Intersections and interactions

Accounting for interaction effects in statistical analyses is important. Researchers agree that failing to account for them if they exist has direct consequences in increasing the risk of coming to false conclusions (Brambor et al., 2006, Landsheer and Van Den Wittenboer, 2005, Miller and Farmer, 1988).

Interaction effects are not only important, but also more challenging than coefficients of additive models. This is so for several reasons. First, estimation of interaction effects requires more statistical power than the estimation of main effects (Aguinis, 1995, Landsheer and Van Den Wittenboer, 2005, Wahlsten, 1990), especially in non-experimental data (McClelland and Judd, 1993). What is more, some researchers consider a problem the multi-collinearity resulting from including interaction terms in regression models (for a review of the debate, see: Friedrich, 1982, Miller and Farmer, 1988). Second, interpretation of interaction effects may pose further difficulties. This concerns both assessing their significance (Braumoeller, 2004), and interpreting the size of coefficients of the main effects (Brambor et al., 2006).

In general, failing to include significant predictors in regression model may produce biased estimates (omitted variable bias). Consequently, omitting an interaction term may produce a bias because constraining its value to zero introduces restrictions on the main effects (Hargens, 2006). This may affect the sizes and significance of remaining coefficients (Brambor et al., 2006). Up to now, the literature devoted little attention to the bias of the main effects resulting from wrongly omitting interaction term. Among the rare exceptions, Brambor et al. (2006) noticed that the bias depends on the distribution of the moderating variable and its relationship with the variable of interest. However, we do not know under which conditions the bias resulting from omitting an interaction effects is negligible. Also we do not know the conditions under which the introduction of unnecessary interaction terms may influence the precision of the model and bias the estimates.

Our work investigates these issues and analyzes a range of conditions which may affect the bias resulting from wrongly omitting or including interaction terms in linear regression models. We consider various scenarios with both dichotomous and continuous predictors, varying correlations between predictors, and varying size of the interaction effect. We assess the performance of additive and intersectional models with respect to their predictive power and the bias of estimates.

3 Monte Carlo simulation

By using Monte Carlo simulation technique, we generate sample data with known parameters. This allows us to evaluate the performance of various models. Specifically, we analyze our data with "correct" and "incorrect" models. We assess the performance of the models by investigating their adjusted R^2 and by comparing the estimated coefficients with the true data.

Our empirical strategy consists of four main steps. The first three steps are schematically presented in Figure 1.

1. First, we generate the data-sets containing two independent variables and their interaction.

2. Second, we add to these data-sets two dependent variables: the first one is a linear function of both variables and their interaction, and the second one is a linear function of the two variables but not of their interaction.

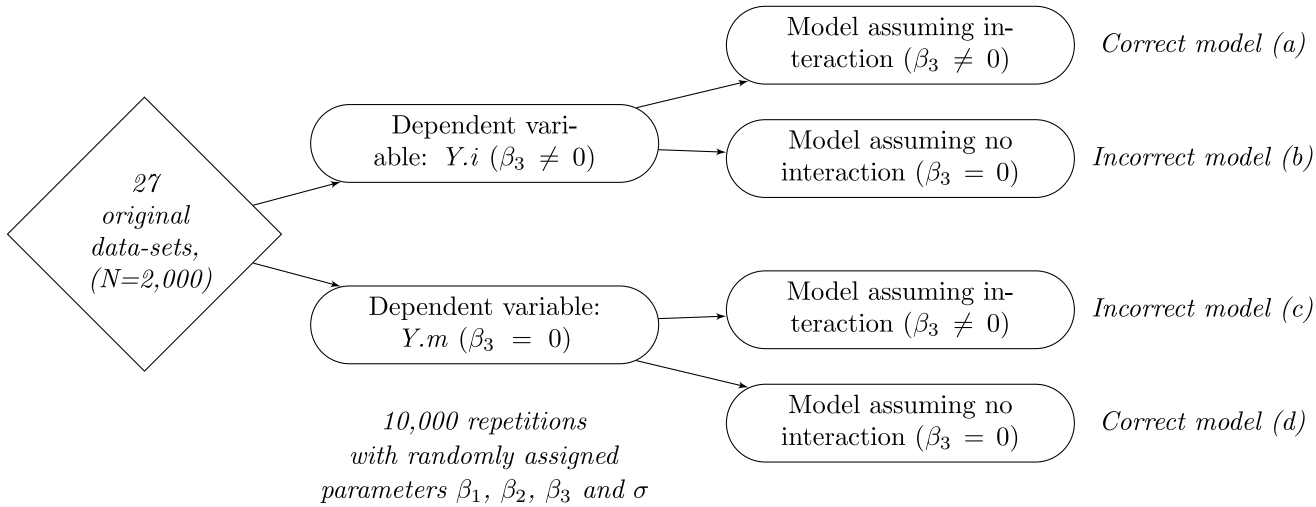
3. Third, we estimate our models on these data. For each of the two dependent variables we estimate a model including the interaction of the two predictors and a model including only the main effects of the predictors. Thus, we estimate two "correct" and two "incorrect" models.

4. Finally, we collect information on the predictive power of the models and analyze it by comparing the adjusted R^2 of the correct and incorrect models for the same dependent variable. We also calculate and analyze the bias of the estimated main effects by comparing them with the true values used to generate the data.

3.1 Data generation

We focus on the simplest case when intersectionality can be observed, i.e. the case with two independent variables which may or may not interact. Overall, we build 27 different data-sets of 2,000 observations each. The data-sets are defined by the measurement level of the independent variables, sample composition, and correlation between the independent variables. Specifically, we generate 27 data-sets, in which:

Figure 1: Diagram summarizing the empirical strategy employed in the paper.



- the two independent variables are dichotomous, and the correlation between the variables ranges from -0.60 to 0.60 (9 data-sets shown in Table 1); we manipulate the distribution of the two variables, generating data-sets where the variables are equally distributed (row 4 in Table 1), highly polarized (rows 2 and 8 in Table 1), as well as intermediate situations;

- one of the independent variables (X1) is continuous and the other independent variable (X2) is dichotomous, and the correlation between them ranges between -0.58 and 0.57 (9 data-sets, Table 2);

- the two independent variables are continuous, and the correlation between them ranges between -0.59 and 0.6 (9 data-sets, Table 3).

Table 1: Characteristics of the data sets in which both predictors are dichotomous (data-sets numbered consecutively from DD1 to DD9).

	ρ(X1, X_2)	N(X1=0,X_2=0)	N(X1=0,X_2=1)	N(X1=1,X_2=0)	N(X1=1,X_2=1)
DD1	−0.60	200	800	800	200
DD2	−0.38	200	200	1,400	200
DD3	−0.25	200	600	600	600
DD4	0	500	500	500	500
DD5	0	200	200	800	800
DD6	0	800	800	200	200
DD7	0.25	600	600	200	600
DD8	0.38	1,400	200	200	200
DD9	0.60	800	200	200	800

Note: ρ(X1, X_2) shows the collinearity between the predictors.

Introducing the varying level of correlation between the independent variables (i.e. collinearity between the independent variables) allows us to verify if the level of collinearity exacerbates or attenuates the consequences of wrongly omitting or including an interaction term.

Table 2: Characteristics of the data sets in which one predictor (X1) is continous and the other (X_2) is a dichotomous variable (data-sets numbered consecutively from DC1 to DC9).

	ρ(X1, X_2)	mean of X_1 if X_2=0	mean od X_1 if X_2=1
DC1	−0.58	−0.01	−0.82
DC2	−0.44	−0.01	−0.71
DC3	−0.32	0.04	−0.53
DC4	−0.19	0.02	−0.34
DC5	0.01	0.01	0.03
DC6	0.15	0.02	0.31
DC7	0.31	−0.04	0.50
DC8	0.47	−0.02	0.72
DC9	0.57	0.00	0.82

Note: $\rho(X_1, X_2)$ shows the collinearity between the predictors.

Table 3: Characteristics of the data sets in which both predictors are continous variables (data-sets numbered consecutively from CC1 to CC9).

	$\rho(X_1, X_2)$	mean(X_1)	sd(X_1)	min(X_1)	max(X_1)	mean(X_2)	sd(X_2)	min(X_2)	max(X_2)
CC1	−0.59	0.00	1.00	−3.40	3.20	−0.01	0.99	−3.40	2.90
CC2	−0.44	0.00	0.96	−3.90	3.10	−0.01	0.99	−3.40	2.90
CC3	−0.31	−0.05	0.99	−3.00	3.60	−0.01	0.99	−3.40	2.90
CC4	−0.13	0.01	1.00	−3.60	3.20	−0.01	0.99	−3.40	2.90
CC5	−0.01	0.02	0.96	−2.90	3.40	−0.01	0.99	−3.40	2.90
CC6	0.17	−0.00	0.99	−3.40	3.50	−0.01	0.99	−3.40	2.90
CC7	0.33	0.03	1.00	−3.10	3.40	−0.01	0.99	−3.40	2.90
CC8	0.44	−0.01	1.00	−3.10	3.40	−0.01	0.99	−3.40	2.90
CC9	0.60	−0.01	0.99	−3.40	3.50	−0.01	0.99	−3.40	2.90

Note: $\rho(X_1, X_2)$ shows the collinearity between the predictors.

3.2 Generating the dependent variables

In each of the 27 data-sets both dependent variables are continuous. This choice allows us to use the OLS regression, whose coefficients can be easily compared across the models. The first dependent variable (labeled $Y.m$) is a linear function of the two predictors, but not of their interaction (see Equation 1 below). The second dependent variable (labeled $Y.i$) is a linear function of the two predictors and of their interaction (see Equation 2).

$$Y.m = \alpha + \beta_1 \cdot X1 + \beta_2 \cdot X2 + \epsilon \tag{1}$$

$$Y.i = \alpha + \beta_1 \cdot X1 + \beta_2 \cdot X2 + \beta_3 \cdot X1 \cdot X2 + \epsilon \tag{2}$$

3.3 Repetitions

We repeat the process of generating each of the 27 data-sets 10,000 times. In each repetition we randomly draw two types of parameters. First, we randomly draw the coefficients α, β_1, β_2, and β_3, which define the relationship between the independent and the dependent variables. We draw them from a uniform distribution of numbers between -100 and 100. The random variation of coefficients β_1, β_2 and β_3 allow us to draw general conclusions, and allows us to verify if the error resulting from wrongly including or omitting interaction terms varies with the effect of the interaction term (β_3) and with the sizes of the main effects (β_1 and β_2).

Second, we randomly draw the parameter σ from a uniform distribution of numbers between 1 and 100. The parameter σ is the standard deviation of the normal distribution from which we draw the individual error terms ϵ; thus σ^2 is the variance of the error terms. The elements of ϵ are drawn from a normal distribution with mean equal to 0 and standard deviation equal to σ. Small values of σ result in more precise models, while large σ results in models with larger random components. We allow σ to vary in order to assess if the error resulting from wrongly including or omitting interaction terms varies with the precision of the model.

3.4 Estimation of the "correct" and "incorrect" models

In the third step we take the perspective of a potential data user, assuming no a priori knowledge about the relationships between the variables. For each data-set we estimate four regression models. We model each of the two dependent variables, $Y.m$ and $Y.i$ as (1) functions of the two predictors but not of their interaction, and (2) functions of the two predictors and of their interaction (see Figure 1). In other words, for each dependent variable we estimate one correct model and one incorrect model. For $Y.i$, the correct model ((a) on Figure 1) assumes an interaction term ($\beta_3 \neq 0$), and the incorrect model assumes no interaction between the predictors ($\beta_3 = 0$, (b) on Figure 1). For $Y.m$, the incorrect model allows for an interaction between the two predictors ($\beta_3 \neq 0$, (c) on Figure 1), and the correct model restricts $\beta_3 = 0$ ((d) on Figure 1).

As mentioned before, we start with 27 initial data-sets (9 for two dichotomous predictors, 9 for one dichotomous and one continuous predictor, and 9 data-sets for two continuous predictors). For each data-set we randomly draw the parameters α, β_1, β_2, β_3, and σ 10,000 times to generate the variables $Y.m$ and $Y.i$. This results in 270,000 datasets and for each one of them we estimate four models (as shown in Figure 1), giving a total of 1,080,000 estimations.

3.5 Collecting information on the performance of models

To asses the risk associated with wrongly omitting or including an interaction term, we collect two types of information. First, we collect the adjusted R^2 of all estimates. Subsequently, we calculate the difference of the adjusted R^2 between the correct and the incorrect models for each dependent variable ($Y.m$ and $Y.i$). We investigate this difference to assess the extent to which wrongly omitting an iteraction term reduces the model's fit. We also test how the model's fit is affected in case of wrongly including an iteraction term.

Second, we collect the estimated coefficients β_1 and β_2, and we compare them with the true values that were used to construct the data. This allows us to assess the bias of the estimated coefficients and, subsequently, to investigate the conditions which affect the size of the bias.

Finally, we explore the factors contributing to the loss of adjusted R^2 and to the bias of coefficients. To this aim we regress the latters on the following characteristics of the data-sets and models:

1. Type of predictors. We distinguish between data-sets with two dichotomous predictors ("DD data-sets"), one dichotomous and one continuous predictor ("CD data-sets"), and two continuous predictors ("CC data-sets".)

2. Composition of DD data-sets. We include a dummy for each DD data-set to account for the various distributions of variables. The reference category is the DD4 data-set, i.e. the data-set with equally distributed X1 and X2 variables. This allows us to check if the distribution of the two dichotomous predictors affects the loss of R^2 or the bias of the estimates.

3. Correlation between the predictors. We include the measure of the correlation coefficient (rho, as in Tables 1-3). As we include data-sets both with negatively and positively correlated predictors, we also include the absolute value of the correlation between the predictors.

4. Variance of the error term. We include σ^2 to investigate how the precision of the models correlates with the loss of predictive power and with the bias of the estimated coefficients.

5. Main effects. We include β_1 and β_2, i.e. the true correlation coefficients between the independent and dependent variables $Y.m$ and $Y.i$ as possible factors affecting the loss of predictive power and the estimation bias.

6. Finally, we control also for the true size of the interaction term, i.e. β_3.

4 Results

4.1 Loss of predictive power of models

This section describes the consequences of wrongly omitting or including interaction terms for the predictive power of models. We calculate the difference between adjusted R^2 of the correct and incorrect models, separately for $Y.m$ and $Y.i$. Boxplots in Figure 2 present the distribution of the R^2 difference between the correct and incorrect models estimated for the data-set DD4 (see Table 1), in which the independent variables are uncorrelated and equally distributed. The result for $Y.i$ is shown in the left panel of Figure 2; it allows understanding how strongly does omitting an interaction term affect the predictive power of a model. The result for $Y.m$ is shown in the right panel of Figure 2, and it demonstrates the consequences of wrongly including an interaction term for the predictive power of the model.

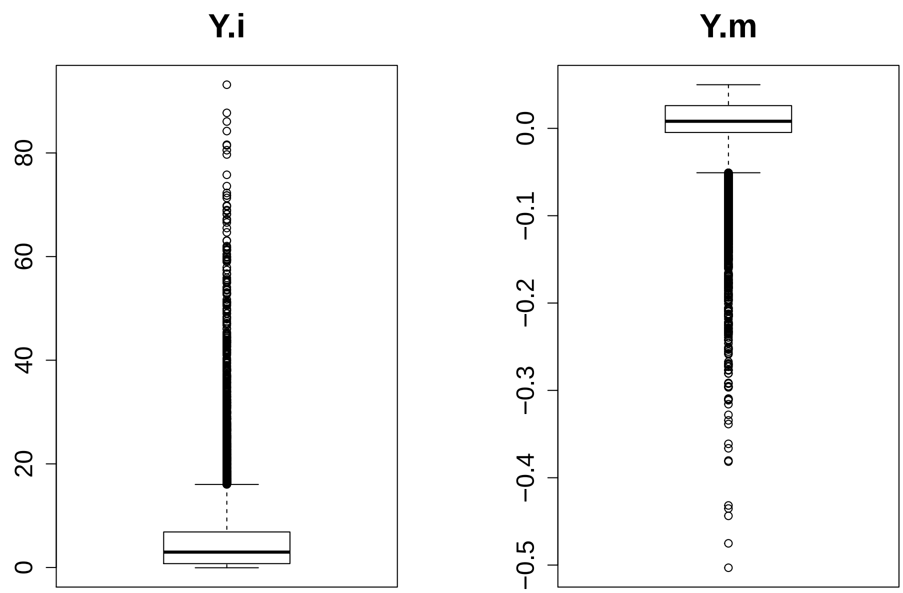


Figure 2: The loss of predictive power (i.e. the decline of adjusted R^2) associated with wrongly omitting (left panel) and wrongly including (right panel) the interaction term.
Note : the boxplots show the distribution of the difference $[R^2(\textit{correct model}) - R^2(\textit{incorrect model})]$ for the data-set DD4. The box depicts the 25^{th}, 50^{th}, and 75^{th} percentile, and the whiskers show the 1^{st} and the 99^{th} percentile.

For $Y.i$ (left panel), wrongly omitting an interaction term lowers the adjusted R^2, i.e. it decreases the predictive power of the model. In 25% of cases the difference of adjusted R^2 exceeds 6.9%, and in 5% of cases it exceeds 21.1%, which is a considerable loss of predictive power for the values of R^2 usually obtained in social sciences.

In case of $Y.m$ (right panel), wrongly including an interaction term may even increase the model's predictive power. The differences of R^2 between the correct and the incorrect model are much smaller, in 99% of cases not exceeding 0.047%. This means that in data where the dependent variable is not associated with the interaction term, wrongly including an interaction term is unlikely to reduce the predictive power of the model.

Table 4 summarizes the same results for all the 27 data-sets, by showing the 25th, 50th and 75th percentile of the distribution of the R^2 differences between correct and incorrect models. Overall, wrongly omitting an interaction term lowers the predictive power of the model.

Table 4: The loss of predictive power (i.e. the decline of adjusted R^2: $R^2(correct\ model) - R^2(incorrect\ model)$) associated with wrongly omitting (columns 2-4) and wrongly including (columns 5-7) the interaction term.

	Wrongly omitting an interaction term			Wrongly including an interaction term		
	25%	50%	75%	25%	50%	75%
	(2)	(3)	(4)	(5)	(6)	(7)
DD1	0.5	2.0	5.0	0.0	0.0	0.0
DD2	0.5	1.8	4.7	0.0	0.0	0.0
DD3	0.6	2.4	5.5	0.0	0.0	0.0
DD4	0.8	3.0	6.9	0.0	0.0	0.0
DD5	0.5	1.9	4.5	0.0	0.0	0.0
DD6	0.5	2.2	5.3	0.0	0.0	0.0
DD7	0.6	2.3	5.6	0.0	0.0	0.0
DD8	0.5	1.9	4.7	0.0	0.0	0.0
DD9	0.5	1.9	4.4	0.0	0.0	0.0
CD1	0.8	3.1	7.2	0.0	0.0	0.0
CD2	1.2	4.6	10.0	0.0	0.0	0.0
CD3	1.6	5.8	13.0	0.0	0.0	0.0
CD4	1.8	6.7	15.0	0.0	0.0	0.0
CD5	1.9	7.0	15.0	0.0	0.0	0.0
CD6	1.9	7.1	16.0	0.0	0.0	0.0
CD7	1.6	5.7	13.0	0.0	0.0	0.0
CD8	1.2	4.3	9.8	0.0	0.0	0.0
CD9	0.9	3.2	7.5	0.0	0.0	0.0
CC1	7.2	25.0	46.0	0.0	0.0	0.0
CC2	6.5	23.0	42.0	0.0	0.0	0.0
CC3	6.1	21.0	40.0	0.0	0.0	0.0
CC4	6.4	22.0	40.0	0.0	0.0	0.0
CC5	5.4	19.0	36.0	0.0	0.0	0.0
CC6	5.9	20.0	38.0	0.0	0.0	0.0
CC7	6.5	22.0	41.0	0.0	0.0	0.0
CC8	6.8	23.0	42.0	0.0	0.0	0.0
CC9	8.2	27.0	48.0	0.0	0.0	0.0

Note: 25^{th}, 50^{th}, and 75^{th} percentile of the distribution of the loss of R^2.

In data-sets with two independent dichotomous variables (rows 1-9) the median loss of adjusted R^2 ranges between 1.8% and 3%; in data-sets with one dichotomous and one continuous predictor (rows 10-18) the loss of adjusted R^2 is larger and it ranges between 3.1% and 7.1%. Finally in models with two continuous independent variables (rows 19-27) the loss of R^2 is overall greatest and it ranges between 19% and 27%.

On the other hand, wrongly including interactions does not reduce the predictive power of the models (see columns 5-7 in Table 4).

These results suggest that wrongly omitting interaction terms, especially if both predictors are continuous variables, may considerably reduce the predictive power of models. On the

Table 5: Regression of the loss of predictive power of the model (i.e. the decline of adjusted R2 in case of wrongly omitting an interaction term) on data and model's characteristics.

	Dependent variable: Loss of predictive power				
	All data-sets	DD data-sets (1)	DD data-sets (2)	DC data-sets	CC data-sets
DC data-sets	4.70***				
	(0.07)				
CC data-sets	22.00***				
	(0.07)				
DD1 data		−1.30***			
		(0.10)			
DD2 data		−1.60***			
		(0.10)			
DD3 data		−1.00***			
		(0.10)			
DD5 data		−1.60***			
		(0.10)			
DD6 data		−1.30***			
		(0.10)			
DD7 data		−0.98***			
		(0.10)			
DD8 data		−1.60***			
		(0.10)			
DD9 data		−1.60***			
		(0.10)			
10% change of $\rho(X_1, X_2)$ (abs)	−0.05***		−0.10***	−1.00***	0.95***
	(0.01)		(0.01)	(0.02)	(0.04)
10% change of $\rho(X_1, X_2)$	0.02**		−0.02**	0.01	0.04**
	(0.01)		(0.01)	(0.01)	(0.02)
10% change of error variance (σ^2)	−1.40***	−0.90***	−0.90***	−1.40***	−2.00***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
10% change of β_1 (main effect)	0.00	−0.02***	−0.02***	−0.01	0.04
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
10% change of β_2 (main effect)	0.02	−0.03***	−0.03***	−0.02	0.09***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
10% change of β_3 (interaction effect)	0.02*	0.04***	0.04***	0.01	−0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Constant	9.30***	8.70***	7.80***	17.00***	30.00***
	(0.07)	(0.07)	(0.04)	(0.09)	(0.16)
Observations	270,000	90,000	90,000	90,000	90,000
Adjusted R^2	0.33	0.13	0.13	0.14	0.08

Note: *p<0.1; **p<0.05; ***p<0.01; The dependent variable is the percentage change of adjusted R^2.

contrary, wrongly including interaction terms is not associated with the loss of predictive power. Hence, we focus on the determinants of the loss of predictive power in case of wrongly omitting interaction terms. To this aim we regress the loss of adjusted R^2 on characteristics of the data-sets and of models as listed in section 3.5. Results of these estimations are presented in Table 5.

The results confirm our previous observation that the loss of predictive power associated with wrongly omitting interaction terms reduces more the predictive power of model if both predictors are continuous. In such cases the decrease of adjusted R^2 is on average 30%, whereas in data-sets with one continuous and one dichotomous predictor it is on average 17%, and in data-sets with two dichotomous predictors it is on average 8.7%.

The collinearity between the two predictors plays a role in data-sets with at least one continuous predictor. In data-sets with two continuous predictors a stronger (in absolute terms) correlation between predictors increases the loss of R^2: an increase of correlation by 10% corresponds to 1% higher loss of predictive power. However, in data-sets with one continuous and one dichotomous predictor stronger (absolute) correlation between predictors corresponds to lower loss of predictive power. 10% increase of correlation results in 1% lower loss of predictive power.

In data-sets with two independent dichotomous variables the relationship between the loss of predictive power and the collinearity of the predictors is not clear. However, the loss of R^2 is largest in the data-set DD4, i.e. the one where the predictors are uncorrelated and evenly distributed.

Furthermore, wrongly omitting an interaction term comes with a smaller R^2 loss if the error variance (i.e. σ^2) is larger. In other words, wrongly omitting an interaction term in precise models is more costly in terms of the predictive power than in less precise models. The effect is strongest for data-sets with two continuous predictors (2% lower loss of R^2 per 10% increase of the error term variance), and weakest in data-sets with two dichotomous predictors (0.9% lower loss of R^2 per 10% increase of the error term variance).

Finally, in data-sets with two dichotomous predictors the R^2 loss is larger when the β_3, i.e. the coefficient of the interaction term, is larger, and when the main effects of the predictors (β_1 and β_2) are smaller. However, these effects are negligible in terms of size: the 10% change of a coefficient corresponds to 0.01%-0.02% lower loss of R^2.

4.2 Bias of the estimates

In this section we investigate the consequences of wrongly omitting or wrongly including interaction terms for the bias of the estimated coefficients. We define the bias as the difference between the estimated coefficient and the true values used to generate the data.

Boxplots in Figures 3 and 4 represent the distributions of the bias of coefficients α, β_1, β_2, and β_3, for the dependent variables $Y.i$ and $Y.m$ respectively. The figures show the bias of both correct and incorrect models estimated for the data-set DD4, in which the two dichotomous independent variables are equally distributed.

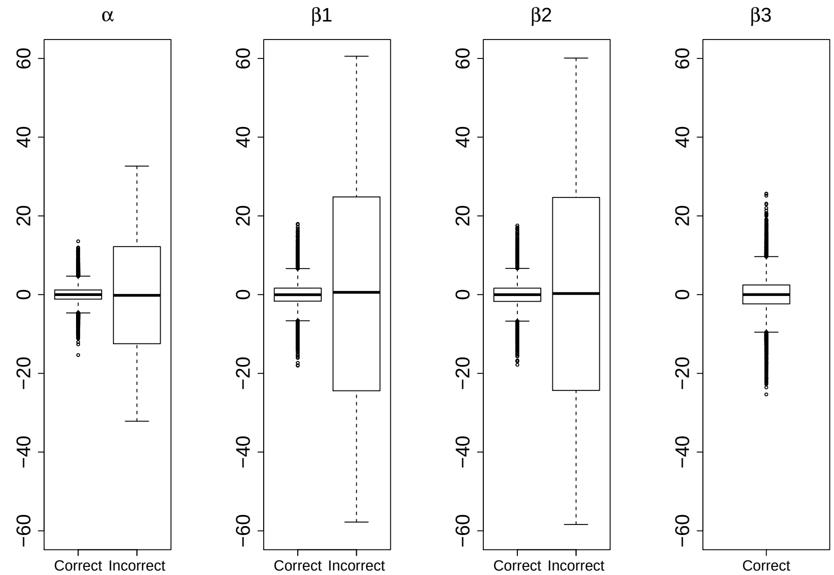


Figure 3: Comparison of the models estimated for $Y.i$. Each boxplot shows the bias of coefficients for the correct model (i.e. with the interaction, on the left) and for the incorrect model (i.e. omitting the interaction, on the right).

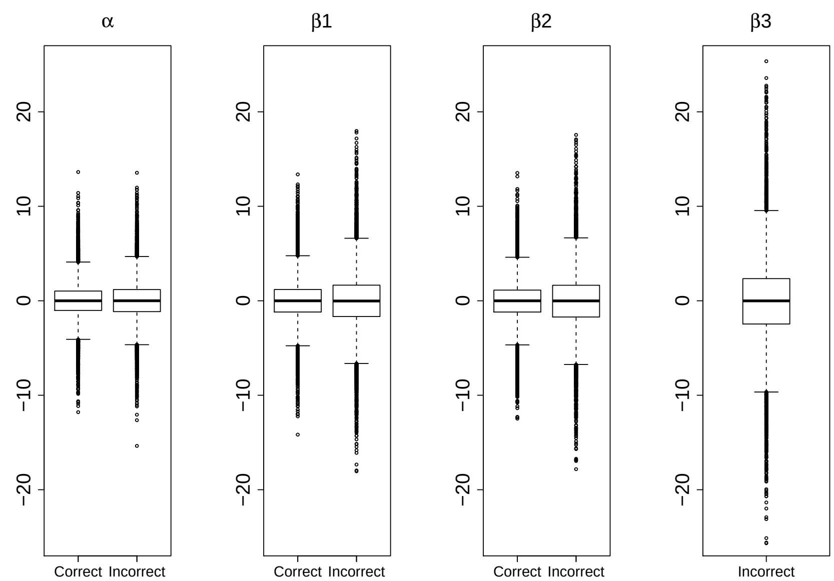


Figure 4: Comparison of the models estimated for $Y.m$. Each boxplot shows the bias of coefficients for the correct model (i.e. without the interaction, on the left) and for the incorrect model (i.e. including the interaction, on the right).

Figures 3 and 4 unsurprisingly show that using a correct model creates a lower risk of estimates' bias than using an incorrect model. Most importantly, the figures document that wrongly omitting an interaction term (Figure 3) can bias the estimates much more than wrongly including an interaction term (Figure 4).

Table 6: The bias of coefficients associated with correctly including an interaction term (columns 2-4), wrongly omitting an interaction term (columns 5-6), correctly omitting an interaction term (columns 7-8), and wrongly including an interaction term (columns 9-11).

	Dependent variable: Y.i					Dependent variable: Y.m				
	correct model (i.e. with interaction)			incorrect model (i.e. without interaction)		correct model (i.e. without interaction)		incorrect model (i.e. with interaction)		
	bias of β_1	bias of β_2	bias of β_3	bias of β_1	bias of β_2	bias of β_1	bias of β_2	bias of β_1	bias of β_2	bias of β_3
	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
DD1	4.2	4.3	5.9	49.0	49.0	2.9	2.9	4.2	4.3	5.9
DD2	5.2	4.1	6.7	63.0	36.0	3.2	3.2	5.2	4.1	6.7
DD3	4.3	4.4	5.3	66.0	66.0	2.5	2.5	4.3	4.4	5.3
DD4	3.3	3.4	4.8	49.0	49.0	2.4	2.3	3.3	3.4	4.8
DD5	5.2	4.2	5.8	79.0	49.0	2.4	3.0	5.2	4.2	5.8
DD6	2.7	4.3	6.0	20.0	50.0	2.4	3.0	2.7	4.3	6.0
DD7	3.0	4.4	5.4	33.0	66.0	2.5	2.6	3.0	4.4	5.4
DD8	4.0	4.1	6.8	36.0	36.0	3.2	3.2	4.0	4.1	6.8
DD9	4.3	4.3	6.1	49.0	49.0	3.0	2.9	4.3	4.3	6.1
CD1	3.3	2.9	4.0	41.0	49.0	2.9	2.1	3.3	2.9	4.0
CD2	2.9	2.4	3.4	35.0	48.0	2.7	1.7	2.9	2.4	3.4
CD3	2.6	2.0	2.9	24.0	48.0	2.5	1.4	2.6	2.0	2.9
CD4	2.4	1.7	2.5	17.0	47.0	2.4	1.2	2.4	1.7	2.5
CD5	2.4	1.8	2.5	3.4	50.0	2.4	1.2	2.4	1.8	2.5
CD6	2.5	1.8	2.6	16.0	50.0	2.4	1.3	2.5	1.8	2.6
CD7	2.6	2.0	2.8	24.0	51.0	2.5	1.4	2.6	2.0	2.8
CD8	2.9	2.3	3.3	36.0	48.0	2.7	1.7	2.9	2.3	3.3
CD9	3.3	3.0	4.3	40.0	50.0	2.9	2.1	3.3	3.0	4.3
CC1	1.5	1.5	1.0	1.7	3.1	1.5	1.5	1.5	1.5	1.0
CC2	1.3	1.4	1.1	1.7	3.7	1.3	1.4	1.3	1.4	1.1
CC3	1.2	1.3	1.1	4.9	2.4	1.2	1.3	1.2	1.3	1.1
CC4	1.2	1.2	1.2	5.7	1.4	1.2	1.2	1.2	1.2	1.2
CC5	1.2	1.2	1.2	2.8	5.0	1.2	1.2	1.2	1.2	1.2
CC6	1.2	1.2	1.1	2.3	1.6	1.2	1.2	1.2	1.2	1.1
CC7	1.2	1.3	1.2	1.4	4.5	1.2	1.3	1.2	1.3	1.2
CC8	1.4	1.3	1.1	2.4	1.7	1.4	1.3	1.4	1.3	1.1
CC9	1.5	1.5	1.1	2.2	2.0	1.5	1.5	1.5	1.5	1.1

Note: The table shows interquartile range (75^{th} percentile - 25^{th} percentile) of estimates' bias.

Table 6 summarizes the same results for all the 27 data-sets, by showing the interquartile range of the bias, i.e. the difference between the 75^{th} and 25^{th} percentile, of β coefficients estimated for $Y.i$ and $Y.m$ using the correct and the incorrect models. Interpretation of these values is quite straightforward. For example, the value 4.2 in the first column and first row of Table 6 indicates that the bias of β_1 if $Y.i$ is estimated with a correct model ranges between -2.1 and 2.1 in 50% of cases, in 25% of cases it exceeds -2.1, and in 25% of cases it exceeds 2.1.[1]

The interquartile range of the bias is visibly largest if the interaction term is wrongly omitted (see columns 4 and 5). In this case the interquartile range of the bias of β_1 ranges between 19.6 and 79 for data-sets with two dichotomous predictors; between 3.36 and 40.7 for data-sets with a continuous and a dichotomous predictor; and between 1.4 and 5.7 for data-sets with two continuous predictors. The bias of β_2 takes similar values: 35.6 - 66, 47 - 50.6 , and 1.4 - 5, respectively.

In case of using a correct model for estimating $Y.m$ (columns 6-7) the interquartile range of the bias is smaller and takes the values between 1.17 and 3.25. The bias is smaller also if the interaction term is included in the model (between 1.03 and 6.81). Interestingly, the bias of coefficients if the interaction term is included in the model is the same for $Y.i$ (columns 1-3) and for $Y.m$ (columns 8-10). In other words, including an interaction term in a model produces on average the same bias no matter if the true interaction equals zero ($\beta_3 = 0$) or not ($\beta_3 \neq 0$).

These results show that wrongly omitting interaction terms typically produces most biased coefficients, and that the bias is largest if at least one of the predictors is a dichotomous variable.

Subsequently, we explore the determinants of the estimation bias associated with wrongly omitting an interaction term. In Tables 7 and 8 we report the result of four regressions in which the estimation bias associated with wrongly omitting an interaction them is regressed over a set of characteristics of the data-sets and models. Tables 7 and 8 show these results for β_1 and β_2 respectively.

Not surprisingly, the strongest predictor of the β_2 and β_3 biases is the size of β_3, i.e. the coefficient of the interaction term. A 10% increase of β_3 corresponds to a 10-unit (i.e. 5%) increase of the bias of β_1 and β_2 (which take the values between -100 and 100). However, this result holds only for dichotomous variables: β_1 and β_2 in DD data-sets, and β_2 in CD data-sets. In case of continuous predictors, the bias of coefficients (β_1 in CD data-sets and both β_1 and β_2 in CC data-sets) remained largely unaffected by the true size of β_3.

Even though other factors may statistically significantly correlate with the bias of β_1 and β_2, the size of these effects are very small.

[1]The value 2.1 comes from dividing by two the value of interquartile range (4.2).

Table 7: Regression of the bias of β_1 associated with wrongly omitting an interaction term on data and model's characteristics.

	Dependent variable: Bias of β_1				
	All data-sets	DD data-sets (1)	DD data-sets (2)	DC data-sets	CC data-sets
DC data-sets	-0.18^{**}				
	(0.08)				
CC data-sets	-0.20^{**}				
	(0.08)				
DD1 data		0.01			
		(0.15)			
DD2 data		0.07			
		(0.15)			
DD3 data		0.05			
		(0.15)			
DD5 data		0.09			
		(0.15)			
DD6 data		-0.13			
		(0.15)			
DD7 data		-0.09			
		(0.15)			
DD8 data		-0.07			
		(0.15)			
DD9 data		-0.08			
		(0.15)			
10% change of $\rho(X_1, X_2)$ (abs)	-0.00		-0.00	-0.00	-0.00
	(0.02)		(0.02)	(0.03)	(0.00)
10% change of $\rho(X_1, X_2)$	0.01		-0.01	0.03^{**}	0.00
	(0.01)		(0.01)	(0.01)	(0.00)
10% change of error variance (σ^2)	-0.00	-0.00	-0.00	-0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.02)	(0.00)
10% change of β_1 (main effect)	-0.01	-0.01	-0.01	-0.01	0.00
	(0.01)	(0.01)	(0.01)	(0.02)	(0.00)
10% change of β_2 (main effect)	0.00	0.00	0.00	0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.02)	(0.00)
10% change of β_3 (interaction effect)	3.30^{***}	10.00^{***}	10.00^{***}	0.08^{***}	-0.31^{***}
	(0.01)	(0.01)	(0.01)	(0.02)	(0.00)
Constant	0.15^{*}	0.03	0.02	0.03	0.01
	(0.08)	(0.11)	(0.07)	(0.13)	(0.02)
Observations	270,000	90,000	90,000	90,000	90,000
Adjusted R^2	0.21	0.88	0.88	0.00	0.17

Note: *p<0.1; **p<0.05; ***p<0.01; The dependent variable is the bias of coefficient β_1; β_1 takes the values between -100 and 100.

Table 8: Regression of the bias of β_1 associated with wrongly omitting an interaction term on data and model's characteristics.

	Dependent variable: Bias of β_2				
	All data-sets	DD data-sets (1)	DD data-sets (2)	DC data-sets	CC data-sets
DC data-sets	-0.02				
	(0.07)				
CC data-sets	-0.26^{***}				
	(0.07)				
DD1 data		0.05			
		(0.09)			
DD2 data		-0.01			
		(0.09)			
DD3 data		0.08			
		(0.09)			
DD5 data		0.04			
		(0.09)			
DD6 data		0.04			
		(0.09)			
DD7 data		0.08			
		(0.09)			
DD8 data		-0.02			
		(0.09)			
DD9 data		0.08			
		(0.09)			
10% change of $\rho(X_1, X_2)$ (abs)	-0.00		0.00	-0.00	-0.00
	(0.01)		(0.01)	(0.00)	(0.00)
10% change of $\rho(X_1, X_2)$	-0.00		0.00	-0.00	-0.00
	(0.01)		(0.01)	(0.00)	(0.00)
10% change of error variance (σ^2)	-0.00	0.01	0.01	0.00	-0.02^{***}
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
10% change of β_1 (main effect)	-0.00	-0.01	-0.01	0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
10% change of β_2 (main effect)	0.00	0.01	0.01	-0.00	-0.00
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
10% change of β_3 (interaction effect)	6.70^{***}	10.00^{***}	10.00^{***}	9.90^{***}	-0.05^{***}
	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
Constant	0.10	-0.05	-0.02	0.01	0.04^{***}
	(0.07)	(0.07)	(0.04)	(0.01)	(0.02)
Observations	270,000	90,000	90,000	90,000	90,000
Adjusted R^2	0.64	0.95	0.95	1.00	0.00

Note: *p<0.1; **p<0.05; ***p<0.01; The dependent variable is the bias of coefficient β_2; β_2 takes the values between -100 and 100.

5 Conclusions

Failing to account for interaction effects when they are theoretically necessary bears the risk of coming to wrong conclusions. Yet, in multivariate regression, introducing many interaction terms may unnecessarily complicate the model and its interpretation, especially if the effect sizes of the interaction terms are small.

Our results show that wrongly omitting an interaction term affects both the predictive power of the model (adjusted R^2) and the bias of β coefficients. Both loss of adjusted R^2 and bias of βs are in this case much stronger than in case of wrongly including interaction terms or of using a correct model.

Concerning adjusted R^2, the loss of predictive power due to omitting an interaction term is largest in data-sets with two continuous predictors: here the median loss of adjusted R^2 ranges between 19.12 and 26.78. In data-sets with at least one dichotomous variable the loss of adjusted R^2 is much smaller (between 1.8 and 7.1). The loss of adjusted R^2 is also higher in data-sets where the predictors are strongly correlated (in absolute terms), and in data-sets with small error variance (i.e. in data-sets where it would be possible to estimate precise models.) On the other hand, the loss of predictive power is smallest in data-sets with two dichotomous predictors which are, in absolute terms, strongly correlated, and in data with large random component, i.e. those for which it is unlikely to design a precise predictive model.

The bias of β coefficients in case of wrongly omitting interaction terms is considerable (values of interquartile range between 1.38 and 79.01) and it concerns all investigated data-sets. The bias is systematically higher in data-sets with two dichotomous predictors (19.58 - 79.01), and systematically lower in data-sets with two continuous predictors (1.38 - 5.7). The size of the bias of the effects of dichotomous predictors is strongly determined by the true effect of the interaction term (β_3). The explanatory variables used in our analysis do not fully explain the bias of the coefficients of continuous predictors.

This article contributes to both literatures on the proper use of interaction terms and the methodology of intersectionality by analysing simulated data. Our results suggest that if researchers are interested in intersections, they can safely adopt a model with interaction terms as there are no risks to get biased results associated with such choice. The implication for the growing ranks of quantitatively-minded, intersectionality scholars is clear: they should continue to pursue their theories as to why intersectionality matters through careful attention to the use of interaction terms. The intersectional model is, under many conditions faced by social scientists, a safe choice.

References

Aguinis, H. (1995). Statistical power with moderated multiple regression in management research. *Journal of Management*, 21(6):1141–1158.

Brambor, T., Clark, W. R., and Golder, M. (2006). Understanding interaction models: Improving empirical analyses. *Political analysis*, 14(1):63–82.

Braumoeller, B. F. (2004). Hypothesis testing and multiplicative interaction terms. *International organization*, 58(04):807–820.

Dubrow, J. K. (2008). How can we account for intersectionality in quantitative analysis of survey data? Empirical illustration for Central and Eastern Europe. *ASK. Research & Methods*, (17):85–102.

Dubrow, J. K. (2013). Why should social scientists account for intersectionality in quantitative analysis of survey data? In Vera Kallenberg, J. M. and Müller, J. M., editors, *Intersectionality und Kritik*, pages 161 – 177. Springer.

Friedrich, R. J. (1982). In defense of multiplicative terms in multiple regression equations. *American Journal of Political Science*, pages 797–833.

Hargens, L. (2006). Interpreting product–variable models of interaction effects. *Centre for Statistics and Social Sciences Working Paper No 7*.

Hughes, M. M. (2015). Crossing intersections: Overcoming the challenges of cross-national research on the legislative representation of women from marginalized groups. In Dubrow, J. K., editor, *Political Inequality in an Age of Democracy: Cross-national Perspectives*, pages 51 – 66. Routledge.

Landsheer, J. A. and Van Den Wittenboer, G. (2005). Main effects as a by-product in a fixed effect 2x2 design, analyzed with anova. *Quality and Quantity*, 38(5):653 – 673.

McCall, L. (2005). The complexity of intersectionality. *Signs: Journal of Women in Culture and Society*, 30(3):1771–1800.

McClelland, G. H. and Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological bulletin*, 114(2):376.

Miller, M. K. and Farmer, F. L. (1988). Substantive nonadditivity in social science research a note on induced collinearity and measurement and testing of effects. *Quality and Quantity*, 22(3):221–237.

Wahlsten, D. (1990). Insensitivity of the analysis of variance to heredity-environment interaction. *Behavioral and Brain Sciences*, 13(1):109–120.

Weldon, S. L. (2006). The structure of intersectionality: A comparative politics of gender. *Politics & Gender*, 2(2):235–248.