Winkelmann, Rainer

Working Paper

Neglected heterogeneity, Simpson's paradox, and the anatomy of least squares

Working Paper, No. 426

Provided in Cooperation with:
Department of Economics, University of Zurich

Suggested Citation: Winkelmann, Rainer (2023) : Neglected heterogeneity, Simpson's paradox, and the anatomy of least squares, Working Paper, No. 426, University of Zurich, Department of Economics, Zurich,
https://doi.org/10.5167/uzh-229123

This Version is available at:
https://hdl.handle.net/10419/276270

University of Zurich

Department of Economics

Working Paper Series

ISSN 1664-7041 (print)
ISSN 1664-705X (online)

Working Paper No. 426

Neglected Heterogeneity, Simpson's Paradox, and the Anatomy of Least Squares

Rainer Winkelmann

Revised version, May 2023

Neglected heterogeneity, Simpson's paradox, and the anatomy of least squares*

Rainer Winkelmann
University of Zurich

May 2023

Abstract

This paper explores an algebraic relationship between two types of coefficients for a regression with several predictors and a group structure. In a general regression, the regression coefficients are allowed to be group-specific, the restricted regression imposes constant coefficients. The key result is that the restricted coefficients are not necessarily a convex average of the group-specific coefficients. In the context of regression with two independent variables and two groups, I show that the coefficient of a regressor estimated from pooled data can be negative, even though the separately estimated coefficients are positive in each group, providing an additional example of Simpson's paradox.

Keywords: Covariance-weighting, non-convex average, average treatment effect.

JEL classification: C21

* I am grateful to Joshua Angrist, Joao Santos Silva, Julius Schäper and Kaspar Wüthrich for helpful comments on an earlier draft of the paper. Department of Economics, University of Zurich, E-mail: rainer.winkelmann@econ.uzh.ch.

1 Introduction

It is well understood that adding a third variable to a bivariate regression can change the sign of the slope coefficient. Indeed this is a common definition of Simpson's paradox Pearl (2014).[1] An alternative definition is broader: Simpson's paradox refers then to a correlation that appears in several groups of data but disappears or reverses when the groups are combined (Simpson, 1951). The two definitions are closely related, since we can declare each specific value of the third variable a group and "control for confounding" by running a separate bivariate regression for each of them. The paradox emerges if the signs are opposite from that of a single bivariate regression on the entire sample.

The purpose of this paper is to extend the scope of the paradox and provide an additional explanation for sign reversal when comparing combined and group-level results, for the case where there is more than a single regressor of interest, as in most practical applications: here, the focus will be on group level heterogeneity, unimportant in the conventional argument above, but important in a multivariate context. Heterogeneity manifests itself by varying coefficients in separate group-level regressions. If, instead, a single regression is used for the pooled data, without allowing for group-level interactions, coefficients can display sign reversal, simply due to the way multivariate regression combines group-level coefficients. As it turns out, one of the surprising consequences is that adding a variable to a bivariate regression can lead to a sign reversal even if that regressor is unrelated to the outcome. This is impossible under standard omitted variable bias.

I provide general conditions under which neglected heterogeneity can give rise to sign reversal, or other less extreme forms of "heterogeneity bias", such as pooled coefficients that are a non-convex combination of the group specific ones. When talking about "bias", I treat the group specific coefficients as meaningful, perhaps also from a causal perspective, and similarly the benchmark of an average effect obtained by weighting the group-level effects by the respective group shares. This seems reasonable in many but not all applications. In the language of causal modeling, it requires the group variable to be exogenous, as would be the case if it is a pre-treatment variable (see Pearl, 2014, for further discussion). However, the results I derive are based on algebraic properties of regression and as such remain valid independently of any notion of causality and

[1] As Pearl (2014) points out, the notion of paradox comes from the fact that people tend to interpret any correlation as a causal relationship, whereas a serious consideration of the two involved regressions would show that at most one of them can be causal.

"correct specification".

To fix ideas, I consider a stylized regression with two regressors, x and z and a group indicator w that, for simplicity of exposition, takes only two values, 0 and 1. Specifically, consider the following three regressions: The pooled, or aggregate, regression without heterogeneity is

$$y_i = b_0 + b_x x_i + b_z z_i + b_w w_i + e_i \tag{1}$$

where $w_i \in \{0, 1\}$ and x_i and z_i can be scaled arbitrarily. The two group-specific regressions are

$$y_i^0 = b_0^0 + b_x^0 x_i^0 + b_z^0 z_i^0 + e_i^0 \tag{2}$$

for $w_i = 0$, and

$$y_i^1 = b_0^1 + b_x^1 x_i^1 + b_z^1 z_i^1 + e_i^1 \tag{3}$$

for $w_i = 1$.[2] Below, I establish algebraic results for the relation between the coefficient of x in the pooled regression, b_x in (1), and the four group-specific coefficients b_x^0, b_x^1, b_z^0, and b_z^1.

Many applications in empirical economics can be condensed to this framework. For example, in a panel data context, assume that there are two years of data, and let a year corresponds to a group. Estimation is possible year-by-year, or by pooling the data over the two years.[3] If a few individual units, say firms, are observed for many time periods, (2) and (3) can be seen as a set of seemingly unrelated regressions (e.g. Zellner, 1962) that are estimated separately for each unit. Coefficient heterogeneity simply means that there is variation in firm level coefficients, and one could be tempted to run a pooled firm level fixed effects regression in order to obtain average effects (see for example Campello et al., 2019, and Breitung and Salish, 2021). Another application is the estimation of treatment effects in a randomized controlled trial with several treatment arms, when treatment takes place at several sites and site-specific factors both confound and modify the treatment effects (e.g., Goldsmith-Pinkham et al., 2022).

The problem addressed in this paper is related to, but different from a sizeable literature on heterogeneous partial effects and the question what linear constant-coefficients regression estimates

[2] As is well known, an equivalent representation is obtained from a fully interacted regression $y_i = a_0 + a_1 x_i + a_2 z_i + a_3 w_i + a_4 x_i w_i + a_5 z_i w_i + u_i$, where, $b_0^0 = a_0$, $b_x^0 = a_1$, $b_z^0 = a_2$, $b_0^1 = a_0 + a_3$, $b_x^1 = a_1 + a_4$, and $b_z^1 = a_2 + a_5$.

[3] Pooling is not limited to genuine panel data, it can also be applied to repeated cross-sections. The regression properties of aggregation, when repeated observations on the same unit are collapsed to individual means in order to run a between regression, or when going from monthly to annual data, are not considered in this paper.

in such a case. For example, if regressors are normally distributed, the regression slope coincides with the average partial effect for a general class of non-linear conditional expectation functions (Stoker, 1986). Angrist (1998) shows that with effect heterogeneity in a model with a scalar binary treatment and a single discrete confounder, OLS gives a variance-weighted average, thus precluding a Simpson's reversal (see also Yitzhaki, 1996).

Apart from least squares regression, numerous covariate adjustments methods exist that actually recover the average treatment effect under the conditional independence assumption (see Wooldridge and Imbens, 2009, for a survey of such methods). For continuous, and potentially multiple, regressors of interest, methods for consistent estimation of average partial effects are discussed in Wooldridge (2004) and Graham and Pinto (2022), among others. Goldsmith-Pinkham et al. (2022) show how neglected heterogeneity in the effect of one regressor can "contaminate" the effect estimation of another one.

While most of this literature is model-based, the current paper solely exploits algebraic properties of ordinary least squares. Thus, results hold regardless of model assumptions, which may or may not be valid. Also, they hold for purely descriptive regressions, and for any sample size as they do not rely on asymptotic properties. On the other hand, this is not a framework to address questions of causality, population estimands and efficiency.

2 Linear regression with group-specific heterogeneity

2.1 Matrix-weighted averaging

Before considering the two-regressors two-groups case in detail, it is useful to recall a general result on pooled regressions, and its representation as a matrix-weighted average of subset regressions. Suppose that there are K regressors and two groups of size N_0 and N_1, respectively. Let y_0 be the $(N_0 \times 1)$ vector of outcomes in group 0, and y_1 be the $(N_1 \times 1)$ vector of outcomes in group 1. The regression coefficient in group 0 is given by

$$b^0 = (X_0'X_0)^{-1}X_0y_0$$

where the group-specific constant has been partialed out, and therefore X_0 is an $(N_0 \times K)$ matrix with typical element $\{x_{ik} - \bar{x}_k^0\}$, $i = 1, \dots N_0$, $k = 1, \dots K$. Similarly, for group 1,

$$b^1 = (X_1'X_1)^{-1}X_1y_1$$

with typical X_1 element $\{x_{ik} - \bar{x}_k^1\}$. Heterogeneity means that $b^0 \neq b^1$. For the pooled regression, assume that data are sorted, such that the first N_0 observations pertain to group 0 and observations $N_0 + 1, \ldots, N_0 + N_1$ to group 1. Correspondingly, we can define the vertically stacked outcome vector and the vertically stacked regressor matrix as

$$y = \begin{pmatrix} y_0 \\ y_1 \end{pmatrix} \text{ and } X = \begin{pmatrix} X_0 \\ X_1 \end{pmatrix},$$

where again, regressors are expressed as deviations from group specific means, which corresponds to a pooled regression that includes an indicator variable for group membership and thus allows for a group-specific constant.[4] The pooled coefficient vector for the slopes of such a regression can then be written as

$$\begin{aligned} b &= (X'X)^{-1}X'y \\ &= (X_0'X_0 + X_1'X_1)^{-1}(X_0'\,y_0 + X_1'\,y_1) \\ &= (H^0 + H^1)^{-1}H^0b^0 + (H^0 + H^1)^{-1}H^1b^1 \end{aligned} \tag{4}$$

where $H^0 = X_0'X_0$, $H^1 = X_1'X_1$, $b^0 = (H^0)^{-1}X_0'\,y_0$ and $b^1 = (H^1)^{-1}X_1'\,y_1$. Hence, the overall vector of regression coefficients is a matrix-weighted average of subset specific regression coefficients. The weights are proportional to the group-specific variance-covariance matrices of the regressors.

This is a mechanical property of ordinary least squares. In the context of Bayesian updating for the normal linear model with homoskedastic errors, H^0 is referred to as the precision matrix of the prior location vector b^0, while H^1 is the precision matrix of the likelihood location vector b^1. Chamberlain and Leamer (1976) study the properties of this type of information pooling, with the goal of deriving bounds for the posterior location parameter when the prior precision matrix is unknown, and thus a kind of sensitivity analysis. Their key result is that without further restrictions on the prior precision matrix (such as being diagonal, or proportional to H_1), the posterior coefficient can lie essentially anywhere. The reason is that a matrix-weighted average of vectors does not mean that element-by-element, the pooled coefficients lie algebraically between b^0 and b^1.

In the present context, H_0 is not arbitrary but rather an observed matrix, and therefore, it becomes possible to study the specific properties of the element-by-element weighting and thus the

[4]The full regression therefore has $2+K$ coefficients, but the focus here is on the slopes, as obtained after partialing out the two constants.

relationship between elements of b and the heterogeneous coefficients collected in b^0 and b^1 for a given dataset. Based on the results by Chamberlain and Leamer (1976), it is to be expected that depending on H_0 and H_1, there can be situations where elements of b are arbitrarily far away from a simple convexly weighted average of the corresponding elements in b^0 and b^1. The objective of this paper is to derive conditions for such non-convex weighting.

I proceed in two steps. First, I consider the case of a single regressor and thereby replicate results known from the literature (e.g. Angrist, 1998). In a second step, I extend the analysis to the case of two regressors, and I will show that counterintuitive results are possible, including Simpson's reversals as defined above.

2.2 Special case: A single regressor

With a single x and a binary group variable w, the pooled model can be written as

$$y_i = b_0 + b_x x_i + b_w w_i + e_i \tag{5}$$

Directly applying (4) from the previous section, we obtain b_x as a scalar average

$$\begin{aligned} b_x &= \frac{(x_0 - \bar{x}_0)'(x_0 - \bar{x}_0)\, b_x^0 + (x_1 - \bar{x}_1)'(x_1 - \bar{x}_1)\, b_x^1}{(x_0 - \bar{x}_0)'(x_0 - \bar{x}_0) + (x_1 - \bar{x}_1)'(x_1 - \bar{x}_1)} \\ &= \frac{N_0 \hat{\sigma}^2_{x,w=0}}{N_0 \hat{\sigma}^2_{x,w=0} + N_1 \hat{\sigma}^2_{x,w=1}} b_x^0 + \frac{N_1 \hat{\sigma}^2_{x,w=1}}{N_0 \hat{\sigma}^2_{x,w=0} + N_1 \hat{\sigma}^2_{x,w=1}} b_x^1 \end{aligned}$$

where b_x^0 is the coefficient in a regression of y_i on x_i in group $w_i = 0$, b_x^1 is the corresponding coefficient for group $w_i = 1$, and

$$\hat{\sigma}^2_{x,w} = \frac{1}{N_w} \sum_{i=1}^{N_w} (x_i - \bar{x}^w)^2 \qquad w \in \{0, 1\}$$

are the within group variances of the regressor.

Hence, the pooled coefficient b_x is not the "average treatment effect", defined as $N_0/(N_0+N_1)b_x^0 + N_1/(N_0+N_1)b_x^1$. However, it is a convex average of b_x^0 and b_x^1. The weights depend on relative group sizes, N_0 and N_1, as well as on the within-group variances, $\hat{\sigma}^2_{x,w=0}$ and $\hat{\sigma}^2_{x,w=1}$. If the within-group variances are identical, the pooled model indeed estimates the group-size weighted average. A closely related population version of this result is given in Angrist (1998) who considers the reverse situation where the regression coefficient of interest is that of a binary treatment indicator and a multivalued, discrete or continuous, confounder is partialed out. Unfortunately, as the next subsection will show, this simple weighting does not generalize when there are two or more regressors.

3 Results for two regressors and two groups

With two regressors of interest, from now on labeled x and z, and the same group indicator as before, the pooled regression takes the form:

$$y_i = b_0 + b_x x_i + b_z w_i + b_w w_i + e_i \qquad \text{for } i = 1, \ldots, N \tag{6}$$

where e_i is a regression residual such that $Cov(e_i, x_i) = Cov(e_i, z_i) = Cov(e_i, w_i) = 0$. The regressors x and z can be binary, discrete, or continuous, and it is assumed that there are two groups only. Regression (6) allows the constant to shift depending on w_i but imposes homogeneous slopes b_x and b_z. After partialing out the constant and w_i, we obtain the trivariate regression

$$y_i = b_x(x_i - \bar{x}^w) + b_z(z_i - \bar{z}^w) + u_i$$

where $\bar{x}^0$, $\bar{x}^1$, $\bar{z}^0$ and $\bar{z}^1$ are group-specific means. Define the following quantities:

$$S^0_{xx} = \sum_{i=1}^{N_0}(x_i - \bar{x}^0)^2 \qquad S^0_{zz} = \sum_{i=1}^{N_0}(z_i - \bar{z}^0)^2$$

$$S^0_{xy} = \sum_{i=1}^{N_0}(x_i - \bar{x}^0)y_i \qquad S^0_{zy} = \sum_{i=1}^{N_0}(z_i - \bar{z}^0)y_i$$

$$S^0_{zx} = \sum_{i=1}^{N_0}(z_i - \bar{z}^0)(x_i - \bar{x}^0)$$

and same for S^1_{xx}, S^1_{zz}, etc. The pooled least squares coefficients $b = (b_x, b_z)'$ in (6) are obtained as

$$b = \begin{pmatrix} S^0_{xx} + S^1_{xx} & S^0_{xz} + S^1_{xz} \\ S^0_{zx} + S^1_{zx} & S^0_{zz} + S^1_{zz} \end{pmatrix}^{-1} \begin{pmatrix} S^0_{xy} + S^1_{xy} \\ S^0_{zy} + S^1_{zy} \end{pmatrix} \tag{7}$$

As in (4), we can substitute the two identities

$$\begin{pmatrix} S^w_{xy} \\ S^w_{zy} \end{pmatrix} = \begin{pmatrix} S^w_{xx} & S^w_{xz} \\ S^w_{zx} & S^w_{zz} \end{pmatrix} b^w \qquad \text{for } w \in \{0, 1\} \tag{8}$$

defining the group-specific, potentially heterogeneous, coefficients b^0 and b^1 to express b as a matrix-weighted average of the group-level coefficients:

$$b = (H^0 + H^1)^{-1}(H^0\, b^0 + H^1\, b^1) \qquad \text{where} \qquad H^w = \begin{pmatrix} S^w_{xx} & S^w_{xz} \\ S^w_{zx} & S^w_{zz} \end{pmatrix}$$

In order to derive the element-by-element relationship between the two elements of b, b_x and b_z, and the four group-level coefficients b^0_x, b^1_x, b^0_z and b^1_z, we need to solve the three systems of equations

described by (7) and (8). For example, it is straightforward to show that the first element of the pooled regression vector, b_x is given by

$$b_x = \frac{(S^0_{zz} + S^1_{zz})(S^0_{xy} + S^1_{xy}) - (S^0_{xz} + S^1_{xz})(S^0_{zy} + S^1_{zy})}{(S^0_{xx} + S^1_{xx})(S^0_{zz} + S^1_{zz}) - (S^0_{xz} + S^1_{xz})^2} \tag{9}$$

Similarly, the group-level coefficients can be written as

$$b^w_x = \frac{S^w_{zz}S^w_{xy} - S^w_{xz}S^w_{zy}}{S^w_{xx}S^w_{zz} - S^w_{xz}S^w_{xz}} \tag{10}$$

and results for b_z, b^0_z and b^1_z follow from symmetry.

3.1 Decomposing the regression coefficient b_x

In the Appendix, I show how to express b_x as a relatively simple function of the four group-level coefficients. In particular, the numerator of (9) is equal to

$$(S^0_{xx}S^0_{zz} - S^0_{xz}S^0_{xz} + S^1_{zz}S^0_{xx} - S^1_{xz}S^0_{xz})b^0_x + (S^1_{xx}S^1_{zz} - S^1_{xz}S^1_{xz} + S^0_{zz}S^1_{xx} - S^0_{xz}S^1_{xz})b^1_x$$

$$+(S^1_{zz}S^0_{xz} - S^1_{xz}S^0_{zz})b^0_z + (S^0_{zz}S^1_{xz} - S^0_{xz}S^1_{zz})b^1_z$$

To obtain the aggregate b_x coefficient, we need to divide this numerator by the original denominator $(S^0_{xx} + S^1_{xx})(S^0_{zz} + S^1_{zz}) - (S^0_{xz} + S^1_{xz})^2$. Comparing coefficients, we can see that

$$b_x = \frac{\mathcal{A}b^0_x + \mathcal{B}b^1_x + \mathcal{C}(b^1_z - b^0_z)}{\mathcal{A} + \mathcal{B}} \tag{11}$$

where

$$\mathcal{A} = S^0_{xx}(S^0_{zz} + S^1_{zz}) - S^0_{xz}(S^0_{xz} + S^1_{xz})$$

$$\mathcal{B} = S^1_{xx}(S^0_{zz} + S^1_{zz}) - S^1_{xz}(S^0_{xz} + S^1_{xz})$$

and

$$\mathcal{C} = S^0_{zz}S^1_{xz} - S^0_{xz}S^1_{zz} = S^0_{zz}S^1_{zz}(g_1 - g_0)$$

and where g_1 and g_0 are the slopes of a regression of x on z in group 1 and group 0, respectively.

The decomposition (11) makes it clear that b_x is in general not a simple convex average of b^0_x and b^1_x. The aggregation of heterogeneous group-specific coefficients depends on three weights, $\mathcal{A}/(\mathcal{A}+\mathcal{B})$ and $\mathcal{B}/(\mathcal{A}+\mathcal{B})$ for the own coefficients b^0_x and b^1_x, and $\mathcal{C}/(\mathcal{A}+\mathcal{B})$ for heterogeneity in the z-coefficients. The weights in turn are functions of products of variances and covariances of x

and z in the two groups. Thus they do not depend on the outcome at all, and the weights can in principle be computed from the "design matrix" of regressors, before the data are collected. Of course, the heterogeneous coefficients themselves depend on the outcomes, and thus are not known before those data are available.

As I will show, each of the weights can be negative or greater than one. For reasons, that will become clear below, I refer to the first two weights as *covariance weights*. $\mathcal{C}/(\mathcal{A}+\mathcal{B})$ on the other hand is a *heterogeneity spillover weight*. As (11) shows, non-convex weighting can occur even if there is no heterogeneity in the z-coefficients, and, more surprisingly, even if the coefficient of z is zero in both groups.

3.2 Covariance weighting

To understand the interpretation of the terms $\mathcal{A}$ and $\mathcal{B}$, consider the auxiliary regression of x on z and w for the pooled data. The slope after partialing out of w is equal to S_{xz}/S_{zz}, and the covariance between residuals $u_i = (x_i - \bar{x}^w) - S_{xz}/S_{zz}(z_i - \bar{z}^w)$ and x_i for group 0 can be written as

$$\begin{aligned} Cov_0(u_i, x_i) &= \frac{1}{N_0}\sum_{i=1}^{N_0}\left((x_i - \bar{x}^0) - \frac{S_{xz}}{S_{zz}}(z_i - \bar{z}^0)\right)X_i \\ &= \frac{S^0_{xx}(S^0_{zz}+S^1_{zz}) - (S^0_{xz}+S^1_{xz})S^0_{xz}}{N_0(S^0_{zz}+S^1_{zz})} = \frac{\mathcal{A}}{N_0 S_{zz}} \end{aligned}$$

Thus, $\mathcal{A}$ is proportional to a group-specific covariance. This is a fundamental difference to the single regressor case, discussed in Section 2.2, where $\mathcal{A}$ was proportional to a conditional variance, and thus necessarily positive.

The weighting in the two-regressor case depends on covariances rather than variances, because $u_i = x_i - \hat{x}_i$ is not orthogonal to $\hat{x}_i$ in the $w_i = 0$ subset, and hence $Cov(x_i - \hat{x}_i, x_i | w_i = 0) \neq Cov(x_i - \hat{x}_i, x_i - \hat{x}_i | w_i = 0)$. Subset-orthogonality of u_i and $\hat{x}_i$ fails, because the auxiliary regression omits the interaction between z and w and is thus not "saturated" in W.[5] Adding the interaction would be equivalent to separate group-wise regressions of x on z, which would restore orthogonality of u_i and $\hat{x}_i$ in each group. But this does not correspond to the regression equation (6) under consideration.

We can now state conditions under which the covariance, and hence $\mathcal{A}$ and the weight $\mathcal{A}/(\mathcal{A}+\mathcal{B})$,

[5] Angrist (1998) also noted that the variance-weighting result for the case of a single predictor requires a partialing out regression saturated in the confounder. By definition, a binary confounder as in (5) satisfies this requirement.

are negative. Since $\mathcal{A} = S^0_{xx}S_{zz} - S^0_{xz}S_{xz}$, we obtain

$$\mathcal{A} < 0 \qquad \Longleftrightarrow \qquad 1 - \frac{S^0_{xz}S_{xz}}{S^0_{xx}S_{zz}} < 0 \qquad \Longleftrightarrow \qquad \frac{S_{xz}}{S_{zz}}\frac{S^0_{xz}}{S^0_{xx}} > 1 \tag{12}$$

Similarly

$$\mathcal{B} < 0 \qquad \Longleftrightarrow \qquad 1 - \frac{S^1_{xz}S_{xz}}{S^1_{xx}S_{zz}} < 0 \qquad \Longleftrightarrow \qquad \frac{S_{xz}}{S_{zz}}\frac{S^1_{xz}}{S^1_{xx}} > 1 \tag{13}$$

For instance, $\mathcal{A}$ is negative if the product of the full-sample x-on-z-and-w regression slope and the group 0 (inverse) z-on-x regression slope exceeds 1. This can be easily checked in the data. A negative $\mathcal{A}$ implies that the weight $\mathcal{A}/(\mathcal{A}+\mathcal{B})$ is negative, since $\mathcal{A}+\mathcal{B} = S_{xx}S_{zz} - S^2_{xz} > 0$ due to the Cauchy-Schwarz inequality. It also follows that either $\mathcal{A}$ or $\mathcal{B}$ can be negative, but not both, as their sum must be positive.

A necessary and sufficient conditions for convex weighting is that both $\mathcal{A}$ and $\mathcal{B}$ are positive. This is for example the case if the reverse regression slopes S^w_{xz}/S^w_{xx} are equal in the two groups. The argument is by contradiction: suppose S^0_{xz}/S^0_{xx} is such that (12) holds. With equal slopes, (13) holds as well, but this is not possible because $\mathcal{A}$ and $\mathcal{B}$ cannot be both negative. Hence, with equal slopes, $\mathcal{A}$ and $\mathcal{B}$ must be both positive. If x and z are uncorrelated in both groups, $\mathcal{A}/(\mathcal{A}+\mathcal{B}) = S^0_{xx}/(S^0_{xx}+S^1_{xx})$ and the weighting is proportional to the within group variances and hence always convex.

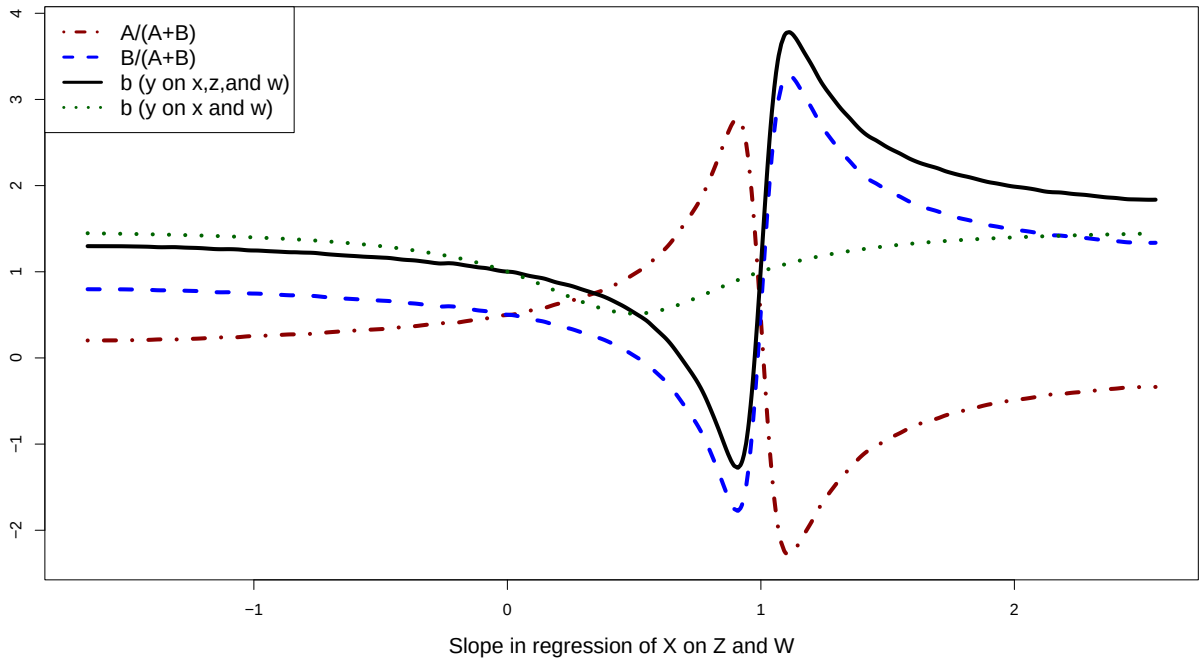
An illustration using simulated data

Consider the following generation of pseudo random numbers: There are two groups of equal size, $N_0 = N_1 = 1000$; z is drawn from a standard normal distribution in both groups. In group 0, $x^0_i = z^0_i + 0.1 \times Normal(0,1)$; in group 1, $x^1_i = \theta_r z^1_i + 0.1 \times Normal(0,1)$, where θ_r increases in steps of 0.01 from -4 to +4. Figure 1 plots, for each value of θ_r, the implied covariance weights $\mathcal{A}/(\mathcal{A}+\mathcal{B})$ and $\mathcal{B}/(\mathcal{A}+\mathcal{B})$ against the implied slope of the pooled partialing out regression of x on z and w. For example, for $\theta_r = 1.1$, we obtain $S^0_{xz}/S^0_{xx} = 0.997$, $S_{xz}/S_{zz} = 1.048$, and $\mathcal{A}/(\mathcal{A}+\mathcal{B}) = -1.780$. Figure 1 also shows the pooled slope coefficient b_x (solid black line), based on outcome data generated from the process $y = 0.5x + xw$. For $\theta_r = 1.1$, it is equal to 3.280.

The data generating process (DGP) is quite simple. There is no effect of z on the outcome in either group, and all that is changing is the covariance between x and z in the two groups. Yet the relationship between S_{xz}/S_{zz} and b_x is non-monotonic, and asymmetrical around the value of 1, which is approximately the value of the reverse regression of z on x in group 0. On the left of this

point, $\mathcal{A}/(\mathcal{A}+\mathcal{B})$ is always positive, and $\mathcal{B}/(\mathcal{A}+\mathcal{B})$ always below 1; on the right, $\mathcal{A}/(\mathcal{A}+\mathcal{B})$ is always negative, and $\mathcal{B}/(\mathcal{A}+\mathcal{B})$ above 1. The pooled regression coefficients reflect the movements of the weights and vary between a minimum of -1.3 to a maximum of +3.8. Thus, the pooled regression of y on x, z and w produces highly misleading estimates of b_x that can be far away from the group size-weighted averages of the heterogenous x coefficients, here $1/2\times 0.5+1/2\times 1.5 = 1$. Moreover, in about 7% of simulated datasets, there is full Simpson's reversal, i.e., a negative regression coefficient when the two groups are combined, although both separately estimated coefficients are positive.

Figure 1: Covariance weighting of heterogeneous coefficients.



In this DGP, z is not part of the outcome equation, so it would be much better to just drop it from the regression. The result would be a variance weighted combination of the two group-specific coefficients, as described in Section 2.2. This average is depicted by the green, dotted line in Figure 1. Values are bounded between the estimated values of b_x^0 and b_x^1, so approximately between 0.5 and 1.5, the two group-specific coefficients. The within-group-0 variance of x is always around 1 in this DGP, whereas the within-group-1 variance moves from $(-4)^2 + 0.1^2$ to a minimum of 0.1^2 and back to $4^2 + 0.1^2$. Hence, in the far left and right areas of the plot, almost all weight goes to the group-1 coefficient of 1.5 based on the much higher within-group variance, and the minimum is observed when $\theta_r = 0$, which corresponds to a residual slope value of $S_{xz}/S_{zz} = 0.5$.

Different x-coefficients would have been obtained for a regression of y on x, z, w and wz. In

this case the partialing out equation is fully saturated in w, and weights are proportional to the conditional-on-z-and-w covariances. For this specific DGP, conditioning on z means that variances in both groups are equal to the variance of the error term, 0.1[1], and this symmetry implies that the least squares coefficient, not shown in Figure 1, are close to 1, the average of 0.5 and 1.5, in all cases.

The above example is admittedly somewhat unrealistic since x and z are highy collinear in both subgroups. This gives coefficients for the reverse regressions of z^w on x^w that are close to the inverse regression coefficients of x^w on z^w and, the conditions for negative values of $\mathcal{A}$ or $\mathcal{B}$ in (12) and (13) are more likely satisfied.[6] For example, if the errors were standard normal instead, all covariance weights are between zero and one. In applications, the problem of non-convex weighting due to negative covariances will therefore tend to arise less frequently than is suggested by the illustration above, unless the degree of multicollinearity is high.

3.3 Heterogeneity spillover

From (11), the heterogeneity spill-over term is given by

$$\frac{\mathcal{C}(b_z^1 - b_z^0)}{\mathcal{A}+\mathcal{B}} = \frac{S_{zz}^0 S_{zz}^1 (g_1 - g_0)(b_z^1 - b_z^0)}{(S_{xx}^0 + S_{xx}^1)(S_{zz}^0 + S_{zz}^1) - (S_{xz}^0 + S_{xz}^1)^2}$$

where g_1 and g_0 are the slopes of a regression of x on z in group 1 and group 0, respectively, and b_z^1 and b_z^0 are the coefficients of z in the group-wise trivariate regrssion of y^w on x^w and z^w.

Since $\mathcal{A}+\mathcal{B} > 0$, the sign of the spillover depends on the sign of the product $(g_1 - g_0)(b_z^1 - b_z^0)$. It is positive if the group with the greater x-on-z slope has also the greater slope of z in the group-specific outcome equation. Only the difference in slopes, i.e., the heterogeneity, matters for the spillover effect, not the absolute values. Clearly, there is no spillover effect if either $b_z^0 = b_z^1$ (no heterogeneity in the z-coefficient) or $g_1 = g_0$, or both.[7]

The two leading examples for $g_0 = g_1$ are as follows. First, x and z could be uncorrelated in both groups, such that $S_{xz}^0 = S_{xz}^1 = 0$. This assumption is unlikely to hold in applications. For example, in the case of two mutually exclusive treatment arms (when x and z are both dummy

[6]It does not matter whether the correlation between x^0 and z^0 is close to +1 or close to -1. For $x_i^0 = -z_i^0 + 0.1 \times Normal(0,1)$, the graphs in Figure 1 are mirrored at 0, such that extreme values of b_x are observed for partialing-out regression slopes of around -1.

[7]In the previous section, equality of the reverse regression slopes S_{xz}^w / S_{xx}^w was shown to be a sufficient condition for the absence of negative covariance weighting.

variables), $z = 0$ whenever $x = 1$ and vice versa, so x and z cannot be uncorrelated. Moreover, in many applications z is included exactly because it is a confounder, and thus necessarily correlated with x.

A second possibility is that the regressors are the same in both groups, for example $S^0_{zz} = S^1_{zz}$ etc. A leading example is that of a stratified randomized controlled trial, where each group represents a stratum, there are multiple treatment arms, and the proportion treated is equal in each stratum. In this case, there is no heterogeneity spillover effect, and in fact, one can show that b_x in a regression of the outcome on the treatment indicators and strata fixed effects is equal to the average treatment effect, i.e., the weighted mean of b^0_x and b^1_x, where weights are proportional to the strata shares.

In general, however, the aggregate x-coefficient in the regression of y on x, z and w depends directly on the heterogeneous coefficients of the *other* regressor z. Thus, there can be a spill-over of neglected heterogeneity, or a "contamination" as described recently by Goldsmith-Pinkham et al. (2022). Even if both sub-sample coefficients of x are zero, the estimated overall effect b_x can be non-zero, because the z-coefficients matter as well, another possible reason for a Simpson's reversal.

An illustration using simulated data

The following numerical example illustrates the potential for contamination bias. As in the simulation results of the previous section, data are generated for two groups with $N = 1000$ observations each. In the context of heterogeneity spillovers, it is important that the association between x^w and z^w differs between the two groups. Therefore, I generate z^0, z^1, and z^0 as i.i.d standard normal random variables, whereas in group 1, x^1 and z^1 are related by the equation

$$x^1 = g_1 z^1 + Normal(0, 1)$$

where g_1 increases stepwise from -2 to $+2$. Since x^0 and z^0 are uncorrelated, $g_1 = g_1 - g_0$ indicates the differential "response" of x to z in group 1 relative to group 0.[8] Outcome data are generated using the equation

$$y = 1 + w + 0.05\, x + 2\, z\, w$$

This means that the true coefficient of x, equal to 0.05, is the same in both groups, whereas there is heterogeneity in the z coefficient, since $b^1_z - b^1_z = 2$.

[8] By construction, an increase in the absolute value of g_1 also affects the group-1 variance of x, since $\mathrm{Var}(x^1) = g_1^2 + 1$. This explains the parabolic shape of the covarianc weights $\mathcal{A}/(\mathcal{A}+\mathcal{B})$ and $\mathcal{B}/(\mathcal{A}+\mathcal{B})$ in Figure 2.

Figure 2: Heterogeneity spillover.

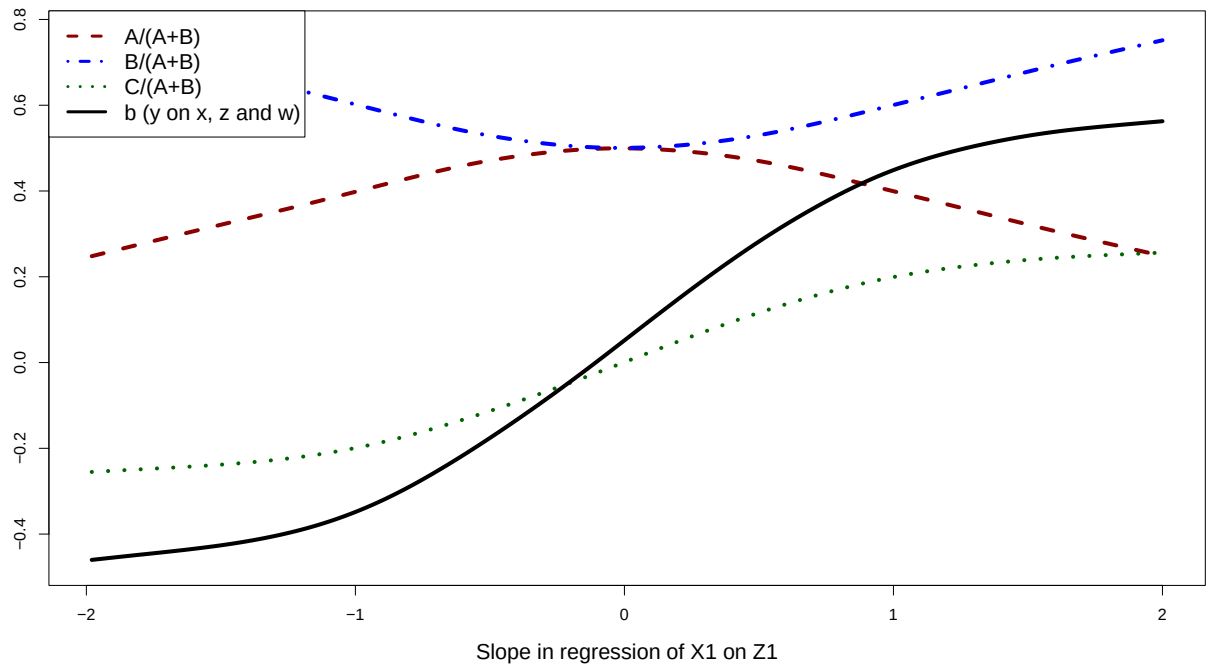


Figure 2 shows the three aggregation weights, together with the pooled regression coefficient b_x (solid, back line), as functions of $g_1 - g_0$. We know that there is no heterogeneity in the effect of x on y, so the covariance weights $\mathcal{A}/(\mathcal{A}+\mathcal{B})$ and $\mathcal{B}/(\mathcal{A}+\mathcal{B})$ do not contribute to departures of b_x from the true average coefficient of 0.05. Nevertheless, observed values of b_x take values between -0.45 and +0.55 which is entirely due to heterogeneity spillovers. The large group-1 specific effect of z (in combination with no effect in group 0) enters the estimation of b_x with positive or negative weight, depending on the sign of g_1. In many instances, this leads to Simpson's reversal: the coefficient of x in separate groupwise regressions is positive in both groups, but negative, when the two groups are combined.

4 Generalizations

In practice, one will rarely encounter an application with two regressors and two groups only, and the question arises whether any algebraic results can be derived for more complex regressions. In particular, what happens as the number of regressors or the number of groups is increased? Increasing the number of groups does not change the nature of the argument. For instance, with

K groups, where $K \geq 2$, (4) generalizes to

$$b = \left(\sum_{k=1}^{K} H^k \right)^{-1} \sum_{k=1}^{K} H^k \, b^k$$

where the weight matrices H^k and the subset regression coefficients b^k, $k \in \{1, \ldots, K\}$ are defined as before. It becomes, however, practically impossible to derive closed form results on the relation between single elements of b and the single elements of the subset coefficients b^k, as the number of terms in the numerator and denominator of the element-wise equation increases quadratically in the number of groups.

Similar issues arise if the number of groups is kept at two but the dimensionality of the regressor vector is increased. While the trivariate problem studied above was manageable, higher order regressions are less so. One exception is the case where the additional variables are group-wise orthogonal to the included ones. In this case, the H^k matrices are block-diagonal, and the above results still apply, as heterogeneity of the additional regressors cannot "spill over" to estimation of the other coefficients when groups are combined in a single pooled regression. This consideration can apply, for example, in a randomized controlled trial with multiple treatment arms, where additional regressors are included simply in order to soak up residual variation and increase precision.

5 Real data examples

While it is straightforward to produce numerical examples of Simpson's reversal due to neglected heterogeneity in artificially generated data, it turns out that finding examples in actual applied work is more difficult. In the end, such an extreme form of heterogeneity bias is perhaps more of theoretical than of practical importance. However, non-convex weighting of heterogeneous coefficients occurs quite regularly when examining concrete dataset used in economics. In this section, I provide two examples, both based on publicly available textbook datasets.

The first one is the estimation of an investment equation, using times series data collected for two firms. This application fits exactly the theoretical decomposition results derived in this paper since there are two regressors, capital stock and market value, and two groups, in this case the two firms. Hence, it is possible to compute the covariance weights as well as the heterogeneity spillover weights in order to trace out, why, as will be the case, the pooled `value` coefficient is not a convex average of the firm-specific `value` coefficients.

The second example follows the spirit of the neglected heterogeneity problem, but applies it

to a wage regression with many covariates, where either two subset regressions are run using male/female as the group variable, or alternatively a combined regression with a male dummy only, using data from the 1976 Current Population Survey. In order to document the combined effect of potentially negative covariance weightings and heterogeneity spillovers, I report the implicit overall own-coefficient weights, by comparing for each regressor the three coefficients b_k^{pooled}, b_k^{male} and b_k^{female}. It turns out that the implicit weights are outside the (0,1) interval for 6 out of 21 regression coefficients.

5.1 An investment equation (Grunfeld data)

This example considers time-series regressions of investment on market value and capital stock for two major U.S. corporations at the time, Union Oil and General Motors, using historical data for the years 1935-1954 obtained from the Grunfeld data distribution.[9]. The first two columns of Table 1 show the heterogenous coefficients obtained from separate, firm-level regressions. The coefficients of the combined regression that includes the two regressors as well as a firm dummy (General Motors yes/no), are shown in the third column of the Table.

Table 1: Grunfeld investment equation

Dependent Variable: Investment	Union Oil b^0	General Motors b^1	Pooled b
Market value/100	8.75	11.93	12.27
Capital stock/100	12.38	37.14	35.98
General Motors (yes/no)			-73.07
Constant	-4.50	-149.78	-84.09
Number of observations	20	20	40

Generally speaking, Union Oil's investment behavior seems to be less responsive to market value and capital stock than that of General Motors. Here, $b_x = 12.27$ correspond to the coefficient of market capitalization in the combined regression. We see that it is a non-convex combination of the firm level coefficients of $b_x^0 = 8.75$ for Union Oil and $b_x^1 = 11.93$ for General Motors.

We can now compute the two covariance weights as well as the heterogeneity spill-over weight for this example. Specifically, $\mathcal{A}/(\mathcal{A}+\mathcal{B}) = 0.002$, $\mathcal{B}/(\mathcal{A}+\mathcal{B}) = 0.998$, and $\mathcal{C}/(\mathcal{A}+\mathcal{B}) = 0.014$. So the pooled coefficient puts almost all its weight on the General Motors coefficient for market

[9]See Kleiber and Zeileis (2010) for further information, and https://www.zeileis.org/grunfeld/ for a download link.

capitalization. This in itself is an interesting result, since both firms contribute the same number of observations to the combined sample, but the regression weights are highly unequal. If it was not for heterogeneity spillover, the pooled coefficient would be a convex combination of the firm-level coefficients, since both covariance weights are positive. However, there is heterogeneity spillover, and it can be computed as $\mathcal{C}(b_z^1 - b_z^0)/(\mathcal{A}+\mathcal{B}) = 0.014(37.14 - 12.38) = 0.347$. Adding 0.347 to the otherwise convex combination of 11.923 yields the combined coefficient of 12.27 that is displayed in Table 1.

5.2 Wages of men and women (CPS)

In the second example, I use a textbook dataset on wages and worker's characteristics from Wooldridge (2012), an extract from the 1976 Current Population Survey.[10] The dependent variable is the logarithm of average hourly earnings. Explanatory variables include years of education, years of potential experience and its square, years with current employer and its square, number of dependents, indicators for being nonwhite, married, living in a metropolitan area, as well as three regional and nine industry dummies. Thus, there is a total of 21 regressors and the dataset provides 526 observations. As group variable of interest, I consider here the gender of the worker. This choice is of course somewhat arbitrary, but wage related regression analyses that do not stratify by gender have been conducted in the literature (e.g. Oreopolous, 2006).

To estimate the group-specific, heterogeneous coefficients, the sample is split and two regressions are conducted, one using the subset of 274 men and one the using the subset of 252 women. Results are shown in Table 2. For 6 out of 21 coefficients, the aggregation weights are non-convex. This is seen in the last column of Table 2, where the equation $b_x = \alpha \times b_x^{men} + (1-\alpha) \times b_x^{women}$ is solved for α. In three instances, this weight is negative (for tenure, for the number of dependents and for the wholesale dummy). In another three instances, it is greater than one (for tenure squared, services and professional occupations). This can lead to quite counterintuitive results. For instance, if one were to use these results to rank industries by their wage differentials, some reversals would occur. For example, for both genders, services pays higher wages than trade, ceteris paribus. Yet, in the aggregate, trade wages are estimated to lie above those of service workers.

[10] The data file "wage1" can be obtained from the R-repository, package name "wooldridge": https://cran.r-project.org/web/packages/wooldridge/wooldridge.pdf.

Table 2: Wages of U.S. workers

Dependent variable: *logarithmic hourly wage*				
	b	b^{men}	b^{women}	weight
years of education	0.047	0.052	0.043	0.41
experience	0.025	0.032	0.020	0.44
experience squared	-0.001	-0.001	0.000	0.48
tenure	0.022	0.025	0.024	-4.82
tenure squared	0.000	0.000	-0.001	1.06
nonwhite	-0.004	0.051	-0.093	0.62
married	0.056	0.159	-0.054	0.52
number of dependents	-0.022	-0.032	-0.022	-0.09
lives in SMSA	0.139	0.142	0.101	0.91
lives in north central U.S	-0.058	-0.118	-0.023	0.37
lives in southern region	-0.044	-0.112	0.013	0.46
lives in western region	0.055	0.018	0.067	0.26
construc. indus.	-0.053	0.026	-0.081	0.26
nondur. manuf. indus.	-0.107	-0.060	-0.109	0.03
trans, commun, pub ut	-0.096	-0.073	-0.142	0.67
trade (wholesale or retail)	-0.303	-0.271	-0.271	-771.9
services indus.	-0.309	-0.255	-0.236	3.83
prof. serv. indus.	-0.095	-0.172	0.013	0.58
profess. occupation	0.225	0.215	0.193	1.44
clerical occupation	0.038	0.115	0.022	0.18
service occupation	-0.094	-0.087	-0.149	0.88
female	-0.268			

Note that this crude assessment of non-convex weighting cannot discriminate between the two sources of non-convexity. So we do not know whether it is primarily due to heterogeneity spillover, or to covariance-weighting of own heterogeneous coefficient contrasts, or both. Since the formulae derived in this paper only dealt with the two-regressor-case, and not with a high-dimensional regressor vector as presently, such a decomposition is not feasible.

6 Discussion

The purpose of this paper was pointing out an algebraic relationship for heterogeneous effects, not to address inference. Estimated sub-group coefficients will never be identical in practice, which begs the question whether their difference is "statistically significant". Such tests are widely available and straightforward to implement (e.g., Chow, 1960, Breitung and Salish, 2021). But in terms of point estimates, the presented results hold regardless of the outcomes of such a test: estimated coefficients in a regression with multiple regressors are not necessarily convex combinations of the

group-level coefficients, and sign reversal is possible.

The non-convexity result follows directly from regression algebra: subgroup regression coefficients are aggregated using matrix level variance-covariance weighting, but this does not imply element by element convex aggregation. With several regressors of interest, non-convex weighting can arise due to two reasons. The first one is technical, since residuals in the partialing out equation cannot be mean independent, implying covariance-weighted averaging; the second reason is substantive, as coefficients in general suffer from spill-overs, or contamination, from heterogeneous coefficients of *other* regressors. In theory, such non-convex weighting can easily lead to a Simpson's reversal: a group-level association may switch its sign once the two groups are combined and a single regression is performed.

An application to estimating a wage regression illustrated that non-convex weights arise quite commonly in practice. When enforcing identical coefficients on 21 regressors, rather than letting them vary by gender, six out of these 21 estimates do not lie between the female and the male estimates.

As a remedy to these problems, one should better conduct group-wise (i.e. fully interacted) regressions, from where one can obtain average coefficients, for example by weighting the heterogeneous coefficients by their relative group sizes. In the context of panel data, this is straightforward as long as the number of years or the number of individual units is small. If the regression adjustment is made in order to satisfy a conditional independence assumption, the situation is more complicated, as there are usually many confounders that may interact with the treatment effect in complicated ways. For example, regarding the model considered in this paper, the treatment effect of x could also vary with z, not only with w. In such cases, matching estimators can be preferable as they avoid the weighting issues inherent to regression (see, e.g. Imbens and Wooldridge, 2009).

7 References

Angrist, J.D. (1998) Estimating the labor market impact of voluntary military service using social security data on military applicants, *Econometrica* 66, 249-288.

Breitung, J. and N. Salish (2021) Estimation of heterogeneous panels with systematic slope variations, *Journal of Econometrics* 220, 399-415.

Campello, M., A. Galvao and T. Juhl (2019) Testing for slope heterogeneity bias in panel data models, *Journal of Business Economics and Statistics*, 2019 (37), 749-760.

Chamberlain, G. and E.E. Leamer (1976) Matrix weighted averages and posterior bounds, *Journal of the Royal Statistical Society B*, 38, 73-84.

Chow, G. (1960) Tests of equality between sets of coefficients in two linear regressions, *Econometrica* 28, 591-605.

Goldsmith-Pinkham, P., P. Hull and M. Kolesar (2022) Contamination bias in linear regressions, Working Paper, https://arxiv.org/abs/2106.05024

Graham, B.S., and C.C. Pinto (2022) Semi-parametrically efficient estimation of the average linear regression function, *Journal of Econometrics* 226 (1), 115-138.

Imbens, G.W. and J.M. Wooldridge (2009) Recent developments in the econometrics of program evaluation, *Journal of Economic Literature*, 47, 5-86.

Kleiber C. and A. Zeileis (2010) The Grunfeld Data at 50, *German Economic Review*, 11(4), 404-417.

Oreopoulos, P. (2006) Estimating average and local average treatment effects of education when compulsory schooling laws really matter, *American Economic Review* 96, 152-175.

Pearl, J. (2014) Comment: Understanding Simpson's Paradox, *The American Statistician*, 68, 8-13.

Simpson, E. (1951) The interpretation of interaction in contingency tables, *Journal of the Royal Statistical Society, Series B* 13, 238-241.

Stoker, T.M. (1986) Consistent estimation of scaled coefficients, *Econometrica* 54, 1461-1481.

Wooldridge, J.M. (2004) Estimating average partial effects under conditional moment independence assumptions, CeMMAP working papers CWP03/04.

Wooldridge, J.M. (2012) *Introductory Econometrics: A Modern Approach*, 5th edition.

Yitzhaki, S. (1996) On using linear regressions in welfare economics, *Journal of Business & Economic Statistics*, 14(4), 478-486.

Zellner, A. (1962) An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias, *Journal of the American Statistical Association*, 57, 348-368.

Appendix

Deriving equation (11)

All notation is as defined in the main text. It holds that b_x is defined by the following fraction.

$$b_x = \frac{(S^0_{zz} + S^1_{zz})(S^0_{xy} + S^1_{xy}) - (S^0_{xz} + S^1_{xz})(S^0_{zy} + S^1_{zy})}{(S^0_{xx} + S^1_{xx})(S^0_{zz} + S^1_{zz}) - (S^0_{xz} + S^1_{xz})^2} \tag{14}$$

The numerator can be re-written by multiplying out and re-ordering terms first into those involving group 0 and group 1 only, followed by all mixed terms:

$$S^0_{zz}S^0_{xy} - S^0_{xz}S^0_{zy} + S^1_{zz}S^1_{xy} - S^1_{xz}S^1_{zy} + S^0_{zz}S^1_{xy} + S^1_{zz}S^0_{xy} - S^0_{xz}S^1_{zy} - S^1_{xz}S^0_{zy}$$
$$= (S^0_{xx}S^0_{zz} - S^0_{xz}S^0_{xz})b^0_x + (S^1_{xx}S^1_{zz} - S^1_{xz}S^1_{xz})b^1_x + S^0_{zz}S^1_{xy} + S^1_{zz}S^0_{xy} - S^0_{xz}S^1_{zy} - S^1_{xz}S^0_{zy}$$

where we have substituted

$$S^w_{zz}S^w_{xy} - S^w_{xz}S^w_{zy} = (S^w_{xx}S^w_{zz} - S^w_{xz}S^w_{xz})b^w_x$$

using equation (10). Next, consider the mixed terms in the numerator of (14):

$$S^0_{zz}S^1_{xy} + S^1_{zz}S^0_{xy} - S^0_{xz}S^1_{zy} - S^1_{xz}S^0_{zy}$$

We can substitute all covariance terms involving y using short-long regression algebra. For instance

$$S^1_{xy} = S^1_{xx}(b^1_x + S^1_{xz}/S^1_{xx}b^1_z)$$

where the term in parentheses is the coefficient of the bivariate subset regression of y^1 on x^1 expressed in terms of the direct effect of x^1 plus the effect of ^{1} on z^1 times the direct effect of z^1 in the trivariate regression. Hence, for instance,

$$S^0_{zz}S^1_{xy} = S^0_{zz}S^1_{xx}b^1_x + S^0_{zz}S^1_{xz}b^1_z$$

etc.; In conclusion, the mixed terms can be written as

$$S^0_{zz}S^1_{xx}b^1_x + S^0_{zz}S^1_{xz}b^1_z + S^1_{zz}S^0_{xx}b^0_x + S^1_{zz}S^0_{xz}b^0_z - S^0_{xz}S^1_{zz}b^1_z - S^0_{xz}S^1_{xz}b^1_x - S^1_{xz}S^0_{zz}b^0_z - S^1_{xz}S^0_{xz}b^0_x$$

Putting things back into (14) and collecting terms, we can write the numerator as an explicit function of the four group-specific, heterogeneous effects of X and Z:

$$(S^0_{xx}S^0_{zz} - S^0_{xz}S^0_{xz} + S^1_{zz}S^0_{xx} - S^1_{xz}S^0_{xz})b^0_x + (S^1_{xx}S^1_{zz} - S^1_{xz}S^1_{xz} + S^0_{zz}S^1_{xx} - S^0_{xz}S^1_{xz})b^1_x \tag{15}$$
$$+(S^1_{zz}S^0_{xz} - S^1_{xz}S^0_{zz})b^0_z + (S^0_{zz}S^1_{xz} - S^0_{xz}S^1_{zz})b^1_z$$

To obtain the aggregate b_x coefficient, simply divide the numerator (15) by the original denominator $(S^0_{xx} + S^1_{xx})(S^0_{zz} + S^1_{zz}) - (S^0_{xz} + S^1_{xz})^2$.