OLS-estimates with different controls are not comparable unless effect heterogeneity is modelled correctly or ruled out*

Felix Weinhardt§

PRELIMINARY AND INCOMPLETE
May 27, 2024

Abstract

While obvious to theoretical econometricians, applied researchers often overlook that the standard interpretation of multivariate OLS estimates requires assuming homogeneous treatment effects. When effects are heterogeneous and not explicitly modeled, assuming constant variance of the explanatory variable across strata is critical. This 'identically distributed' regressor assumption is particularly unlikely when controls are added, making coefficient changes across specifications uninformative. Moreover, model selection based on statistical significance or mean squared error fails the applied researcher.

Keywords: ordinary least squares, heterogeneous treatment effects
JEL Codes: I23, I24, I26

*We thank Jan Bietenbeck and Jan Marcus for comments. All remaining errors are mine.
§European University Viadrina, Berlin School of Economics, CEP (LSE), CESifo, IZA; weinhardt@europa-uni.de

1 Introduction

The typical way in that applied researchers present and interpret estimates roughly follows the following sequence, when simple OLS cannot be trusted:

1. Estimates from simple OLS or from OLS with only a handful of controls are shown.
2. More control variables or different layers of fixed effects are added to account for endogeneity.
3. Last but not least, alternative estimators such as IV or experimental estimates are presented, if the setting allows this.

This note makes the point that comparisons of estimates across such as sequence of specifications are not informative when treatment effects are heterogeneous (except in a very special case). This is because with heterogeneous treatment effects, each OLS specification (that does not explicitly model the effect heterogeneity) with a different sets of controls estimates a differently re-weighted population parameter.

Table 1: Irrelevant controls become relevant in pooled OLS

	Sample 1		Sample 2		Pooled Sample	
	(1)	(2)	(3)	(4)	(5)	(6)
Estimated effects of X	0.957**	0.957**	5.017**	5.017**	3.000**	2.466**
	(0.046)	(0.046)	(0.044)	(0.098)	(0.113)	(0.089)
Estimates effect of W		-0.008		-0.055	-	1.061**
		(0.043)		(0.091)		(0.073)
Controls included		✓		✓		✓
N	500	500	500	500	1000	1000

Notes: This table presents OLS estimates of a generated dataset where treatment effects are heterogeneous across two strata where the variance in the regressor constant across groups. The exact Stata code for replication can be found here. Hereroskedasticity-robust standard errors are in parenthesis. ** denotes significance at the 1%-level.

Table 1 illustrates this point. We see estimates of a single realisation of a data generating process where Y depends on X only. The researcher has access to an additional control variable W that can correlate with X but here does not have independent effects on Y. Columns (1) to (4) show the OLS estimates for two samples that have true βs equal to one and five respectively. In these samples W represents what is routinely refereed to as "irrelevant control". Adding W indeed has little effects on the estimates, which are close to their population parameters.

In practice, substrata, i.e. groups that have different effect sizes, are rarely known and instead the applied researcher estimates a single β coefficient in a pooled sample. In our example, the average effect in the population of a one unit change in X equals

three. Column (5) shows the OLS estimate over the pooled sample that is very close to this population parameter. As we will discuss below, this is the case because of identical group sizes and i.i.d. regressors, which are not innocent assumptions.

Unfortunately, adding the previously irrelevant control W to this pooled sample in column (6) changes the estimate. The estimated effect of a one-unit change in X on Y is now 2.466, highly significant at conventional levels and different to the previous estimate and from the population paramater. As will become clear, OLS is behaving correctly but the interpretation of this estimate as average effect is no longer valid. As an immediate implication, the estimates in columns (5) and (6) are not directly comparable.

Are such coefficient movements a real or just a statistical problem? As we will see below, adding "irrelevant" controls can re-weight the pooled β estimate to any value between the largest and smallest effect-strata in the data. For comparison, this range of possible values is identical to that of IV-LATE estimates for compilers. In other words: ignoring this property of OLS, in terms of making mistakes, is similar to interpreting all IV settings as providing estimates for the average treatment effect for the treated (ATT). Since treatment heterogeneity is *a priori* usually not known to the researcher, large coefficient movements are thus a real possibility. In the case of the example considered in Table 1, the pooled estimate depending on the precise nature of the included "irrelevant controls" relates to population effects between the values of one and five.

Last but not least, in the pooled sample W itself becomes significant. As we discuss below, the model including W is strictly preferable from the statistical perspective. As a result, model selection tools would routinely prefer the specification in column (6) over estimate from column (5).

We discuss implications for applied resarch in the conclusions. To highlight a few already here: the standard interpretation of coefficient movements in the context of omitted variable bias is not longer valid. Similarly, if estimates do not change when more control variables are added, this cannot be interpreted as "good news" regarding endogeneity concerns. Comparisons of OLS with IV LATE estimates are invalid, as these allow for heterogeneity by interpretation of the IV estimate as LATE. As demonstrated below, multivariate OLS estimates that falsely include an instrument as control are weighted away from the IV complier population.

This note relates to a recent literature that examines sample properties of OLS estimators. Angrist and Pischke (2008) in chapter 3 discuss weighting when comparing OLS to matching but do not elaborate on implications for the interpretation of plain OLS. A recent literature highlights weighting problems in difference in differences estimation with effect heterogeneity, i.e. Goodman-Bacon (2021), or when group sizes vary for binary treatments (Słoczyński, 2022). This note equally applies to all diff-in-diff estimators as long as these are estimated using OLS.

2 Simple OLS and heterogeneous effects

2.1 Identically distributed regressors

We start with a simplest model with heterogeneous treatment effects: outcome Y for individual i, where the subscripts i are suppressed in the following, is regressed on the variable of interest X.

$$Y = \alpha + \beta_g * X + \epsilon \tag{1}$$

β_g denotes the average effects for strata g. OLS that ignores this heterogeneity provides a single pooled estimate (β^P) that averages over β_g, which if X is binary is often interpreted as the Population Average Effect (AE), or Population Average Treatment Effect (ATE) if the treatment is binary.

Unfortunately, this interpretation is only correct if the variance of X is identical for all strata, taking into account the number of observations within each strata.

This can be seen by constructing the average effect step-wise. In a first step, take all β_g that can be obtained from G separate specifications for sub-samples of size N_g, with homogeneous effects within each strata. In these, $\beta_g = \frac{COV(X_g,Y_g)}{VAR(X_g)}$[1]. In a second step, the average effect is obtained by averaging over these effects and taking into account the strata size so that $\frac{N_g}{N}$ gives the population share for each strata: $\beta^{AE} = \sum_{g=1}^{G} \frac{N_g}{N} \frac{COV(X_g,Y_g)}{VAR(X_g)}$.

The pooled OLS that ignores heterogeneity estimates instead $\beta^P = \frac{COV(X,Y)}{VAR(X)}$. How does this estimator weight different observations/strata? The denominator is identical for all observations, so the weighting comes from subgroup combinations within the numerator. OLS provides the sample-size-variance-weighted average across strata:

$$\beta^P = \sum_{g=1}^{G} w_g \beta_g = \sum_{g=1}^{G} \frac{N_g VAR(X|g)}{VAR(X)} \beta_g \tag{2}$$

This is not a new result, e.g. Rao and Precht (1985). But this weighting of OLS is rarely discussed in the context of heterogeneous treatment effects, possibly because OLS properties were well understood in the 1980s but effect heterogeneity only gained traction in main stream discussions much later. Abstracting from differences in the number of observations across strata, from this weighting it follows directly that if differences in $VAR(X)|g$ are related to treatment effect heterogeneity, i.e. differences in β_g, β^{AE} is unlikely to equal β^P. Whenever this is the case, the pooled OLS β^p returns a re-weighted average treatment effect, which differs from the target parameter. In terms of classical assumptions, differences in $VAR(X)|g$ across strata are routinely assumed away requiring not only indepedent but also identically distributed (i.i.d.) data generating processes of the regressors. Next, we use a simple illustrative example to argue that this assumption

[1] For ease of notation I am not using the expectation operator.

will often not be met in practice.

2.2 Numerical example simple OLS

The Stata code in the attached do-file (also available here) generates a single dataset with $N = 1000$, where Y depends only on X and a normally distributed error term.[2] To abstract from weighting due to group size, there exist two equally sized strata ($g = 1$ for obs. 1-500, $g = 2$ for obs. 501-1000). Moreover, the variance in X is not constant across strata, we have $VAR(X_g|1) < VAR(X_g|2)$, so the regressor values are independent but not identically distributed. The outcome is defined as $Y = \beta_g X + \epsilon$ and treatment effects are heterogeneous with $\beta_1 = 1$ and $\beta_2 = 5$. Note that in this example it is the strata with the larger β_g that also has the higher variance in X_g. This means OLS will put excessive weight onto the strata with the higher treatment effect, so that $\beta^P > \beta^{AE}$.

Table 2 shows in columns (1) the OLS estimate only considering the observations for the first strata, in column (2) only for the second strata. In column (3) we have $\hat{\beta}^{AE}$ as constructed above. Last but not least, column (4) shows $\hat{\beta}^P$.

Table 2: Estimates with heterogeneous effects and heteroskedastic strata that are positively related

	(1) $\hat{\beta}_1$	(2) $\hat{\beta}_2$	(3) $\hat{\beta}^{AE}$	(4) $\hat{\beta}^P$
Estimated effects of X	0.957**	4.979**	2.975	4.004**
	(0.046)	(0.025)	-	(0.082)
N	500	500	1000	1000

Notes: This table presents simple OLS estimates of a generated dataset where treatment effects are heterogeneous across two strata and the variance in the regressor is not constant across groups. The exact Stata code for replication can be found here. Hereroskedasticity-robust standard errors are in parenthesis. ** denotes significance at the 1%-level.

The estimates in columns (1) and (2), where treatment effects are homogeneous, behave as expected. The "hand-constructed" average effect $\hat{\beta}^{AE}$ in column (3) is close to three, too. However, $\hat{\beta}^P$ is far away from this average effect of a one-unit change in X on Y. This is because observations from the second group (with larger effects) are over-weighted due to their higher variance.

Notice that in this numerical example we set the $VAR(X)|1 = \frac{1}{3}VAR(X)|2$. Such a difference in treatment variation can occur for a binary treatment for example if half of the individuals who would really benefit from a treatment (strata 2) take the treatment, but only about ten percent of individuals from strata 1 (who benefit much less) do so. Such differences in treatment variation across strata, violating the assumption of identically

[2]Of course, this exercise can be repeated for distributions of estimates. Here, for simplicity and illustration only a single sample realisation of the data generating process is discussed.

distributed regressors, are not implausible.

In practice, the applied researcher rarely knows the sizes or identities of strata, so that only β^P can be estimated. This should not be interpreted as estimate of an average population parameter without the explicit assumption of homogeneous treatment effects or heterogeneous effects with constant variance in X across (unknown) strata.

3 Multivariate OLS and heterogeneous effects

3.1 When 'irrelevant' controls become relevant

The applied researcher rarely considers simple OLS as sufficient to estimate causal effects. The first remedy to endogeneity problems is to include control variables. In fact, Oster (2019), building on Altonji et al. (2005), established a method to compare coefficient movements between specifications with different numbers of control variables in order to bound omitted variable bias. We have just reminded ourselves of the importance of identically distributed regressors. Unfortunately, this assumption will be even harder to meet when control variables are included. As we will see below, one implication of this is that coefficient movements between specifications cease being informative when effects are heterogeneous and pooled estimates are being considered. This is because the inclusion of controls, except in a special case, will change the variation in the regressor that is being used for estimation, violating i.i.d..

To see this, assume that there exist variables $\boldsymbol{W}$ for that $COV(\boldsymbol{W}, \epsilon) = 0$ and $COV(\boldsymbol{W}, X) \neq 0$. Notice already here that $\boldsymbol{W}$ happens to fulfil the IV relevance and exclusion restrictions. But now let us consider $\boldsymbol{W}$ as controls. In the homogeneous treatment world, $\boldsymbol{W}$ would be called "irrelevant controls" that should not be included because of the resulting loss of degrees of freedom. Wrongly including $\boldsymbol{W}$ would not result in inconsistencies. Our focus here is on coefficient movements, which do not occur with homogeneity. Now, let us consider that happens when treatment effects are heterogeneous. The researcher believes in the following model, which includes W:

$$Y_i = \alpha + \beta_g * X + \boldsymbol{W}'\boldsymbol{\gamma} + \epsilon \tag{3}$$

With known strata, we could again first estimate separately for each group the β_gs and then average these to obtain β^{AEM}, just as above in the case of the simple regression. From Frisch-Waugh we know that β_g can also be estimated in a simple regression on $\tilde{X}_i$, the residual from the auxiliary regression of X on $\boldsymbol{W}$. Of course, this is only possible if the strata can be identified.

$$Y_i = \alpha + \beta^{PM} * X + \boldsymbol{W}'\boldsymbol{\gamma} + \epsilon \tag{4}$$

Since strata are most likely unknown, the researcher usually considers a model that

estimates a single parameter, denoted by β^{PM} instead, that pools over all strata (equation 4). Again using Frisch-Waugh, this pooled OLS-estimate in this multivariate regression is given by $\beta^{PM} = \frac{COV(\tilde{X},Y)}{VAR(\tilde{X})}$. How does this weight strata? Again, OLS provides the sample-size-variance weighted average:

$$\beta^{PM} = \sum_{g=1}^{G} \frac{N_g VAR(\tilde{X}|g)}{VAR(\tilde{X})} \beta_g \tag{5}$$

Compared to β^P, this OLS-estimate uses only a subset of the available variation in X to estimate the effect, namely the subset that is not predicted by $\boldsymbol{W}$ through the auxiliary regression. The variation in X that can be predicted with $\boldsymbol{W}$ is not used. Notice that the denominator of these weights is identical across all strata, so that weighting again comes from differences in the numerator.

Recall that for $\beta^P = \beta^{AE}$ we require $VAR(X)|g$ constant across strata. Similarly, for $\beta^{PM} = \beta^{AEM}$ we require that the netting out of variation in X through Frisch-Waugh does not affect differently the remaining variation across strata, so that $VAR(\tilde{X})|g$ remains constant across strata. Otherwise the conditional treatment variation is not identically distributed across strata.

In particular, if $\boldsymbol{W}$ does not predict X in the same way across strata, the pooled OLS estimate that includes $\boldsymbol{W}$ as controls weights strata differently into an overall effect so that $\beta^P \neq \beta^{PM}$. The same line of arguments holds across different specifications that use a different set of control variables. Each time, a different weighting across strata occurs, unless the control variables correlate identically across strata, and therefore do not violate the assumption of constant variation in X across strata.[3]

It is informative to compare this to a specification that uses $\boldsymbol{W}$ as instrument for X in a 2SLS estimation procedure. Here, Y is regressed on $\hat{X}$ in the second stage and the resulting IV-estimate gives the effect for the complier population. OLS that includes the instrument as control, in contrast, is equivalent to regressing Y on $\tilde{X}$, where $\tilde{X}$ is obtained from the same "first stage" regression. OLS only considers $\tilde{X}$, i.e. variation from the residual of the frist stage. IV only considers the predicted variation $\hat{X}$. Accordingly, in OLS the complier variation in X is netted out, and OLS estimates the average effect of a one-unit change in X for everyone except the complier population of that particular instrument, so that β^{PM} is weighted away from β^{AE} in the opposite direction compared to β^{IV}. Both estimates are bound by the larges and smallest β_g, which we will also derive below for the case of the pooled OLS with controls.

[3]There is a special case where coefficients would not move, when differences in weighting due to changes in the variation in the conditional treatment and differently signed treatment effects across strata cancel each other out coincidentally.

3.2 Implications for model selection

So far, the discussion has focused on the interpretation of a pooled OLS-estimate β^{PM} when treatment effects are heterogeneous. As discussed, the inclusion of "irrelevant controls" changes this pooled estimate even in the absence of classical omitted variable bias, except in the special case of constant conditional treatment variance across strata. One might conclude from this that the best solution would be not to include $\boldsymbol{W}$ as controls in the first place.

Unfortunately, the applied researcher will not get much guidance on model selection from comparing models with and without $\boldsymbol{W}$ included. This is because statistically, the model that includes $\boldsymbol{W}$ dominates the simple model.

The reason for this is that the estimated model is mis-specified and that the inclusion of $\boldsymbol{W}$ can partly remedy this. To see this, consider the correct model from specification 3 denoted in the following way:

$$Y_i = \alpha + \delta_1 X + \sum_{g=2}^{G} \delta_g D_g X + \epsilon \tag{6}$$

where D_g are dummy variables for the strata. In this fully interacted model, the effect of a one unit change in X on Y is given through $\frac{dY}{dX}$. Of course, δ_1 alone cannot be interpreted as the average effect.

Since the strata are unknown to the researcher estimating specification 4, the interaction terms $\sum_{g=2}^{G} \gamma_g D_g X$ that capture the effect heterogeneity remain in the error term. This becomes problematic because variables $\boldsymbol{W}$ can correlate with the interaction terms (in the error term) and thus serve as proxy. Mirroring the above discussion, they will do so whenever the correlation of $\boldsymbol{W}$ with X is not constant across strata and therefore not fully captured by the included X. Statistically, including a valid proxy for the omitted interaction term is preferable to not including the proxy, and thus improves model fit. As a result, estimates for γ in specification 4 can turn significant in their own right, despite being seemingly "irrelevant controls". Model selection routines based on statistical significance or model fit criteria such as the mean squared error thus favour the inclusion of $\boldsymbol{W}$. Unfortunately, the better $\boldsymbol{W}$ proxies for the omitted interactions, the closer β^{PM} moves to δ_1 from specification 6, which cannot be interpreted as average effect. Since the applied researcher usually has no information about which strata δ_1 relates to, this bounds the coefficient movements in β^{ATE} from including $\boldsymbol{W}$ to values between the largest and smallest effect strata β_g in the population.

3.3 Numerical example: multivariate OLS

The Stata code in the attached do-file (also available here) generates a dataset with $N = 1000$, where Y depends only on X and a normally distributed error term (and not

on W). Again, there exist two equally sized strata ($g = 1$ for obs. 1-500, $g = 2$ for obs. 501-1000). As above, the outcome is defined as $Y = \beta_g X + \epsilon$ and treatment effects are heterogeneous with $\beta_1 = 1$ and $\beta_2 = 5$.

In contrast to the previous example, we here set the variance in X as constant across strata, we have $VAR(X)|1 = VAR(X)|2$. This means that a simple OLS in this setting returns a valid estimate for the average treatment effect and $\beta^P = \beta^{AE}$ as regressors are i.i.d..

We now want to understand what happens if control variables are added to this specification. For this, we define a single variable W that correlates with X in the following way: $COV(W, X)|1 = 0$ and $COV(W, X)|2 > 0$. In words, W correlates with observations in strata 2 but not in strata 1. Notice further that W has no independent effects on Y. Due to the non-constant correlations with X across strata, we expect that the pooled estimate β^{PM} differs between specifications that include or exclude W as controls.

Specifically, including W as control reduces the variation in X that is used to estimate β^{PM} from the second group. Recall that we defined the true $\beta_2 > \beta_1$. We therefore expect $\beta^{PM} < \beta^P$ due to the inclusion of W as control.

Table 3 shows in columns (1) and (2) the pooled OLS estimates known from Table 1. Column (1) shows the simple OLS, column (2) includes W as a control (in the full sample). In column (3) a different control variable V is added, which equals X for observations of the second strata and is zero otherwise. Last but not least, column (4) shows the IV-estimate that uses W as instrument for X.

Table 3: Simple OLS with heterogeneous effects and heteroskedastic strata: how controls change the estimates

	(1) β^P	(2) β^{PM}	(3) β^{PM}	(4) β^{IV}
Estimated effects of X	3.000**	2.466**	0.955**	4.914**
	(0.113)	(0.089)	(0.046)	(0.191)
Estimates effect of W	-	1.061**	-	
		(0.073)		
Estimates effect of V		-	4.010**	-
			(0.044)	
Controls included		✓	✓	
N	1000	1000	1000	1000

Notes: This table presents OLS estimates of a generated dataset where treatment effects are heterogeneous across two strata where the variance in the regressor constant across groups. W is a valid instruments and therefore "irrelevant" control. V equals X for observations of the second strata and is zero otherwise. The exact Stata code for replication can be found here. Hereroskedasticity-robust standard errors are in parenthesis. ** denotes significance at the 1%-level.

Using the full sample with two effect-strata, the estimate in column (1) that gives equal weight to both strata is a valid estimate of the population average effect, now that the regressor is constructed to be i.i.d. In this case, the interpretation of the pooled estiamte as average effect remains valid.

In column (2) the "irrelevant" controls W are added to the full sample. Recall, these correlate only with the X in the second strata and therefore give higher weight to the first. As a result, the pooled estimate is indeed smaller and cannot any more be interpreted as a valid estimate for an average population effect. This is because the conditional regressor variance is not identically distributed across strata.

Notice again that the estimate of W is positive and significant in column (2). As discussed, this is because W predicts group membership which correlates with effect size and therefore becomes informative, despite not having independent effects on the outcome. W act as proxy for the omitted (and unobserved) interaction term from specification 6.

In column (3) the included control V proxies perfectly for the omitted treatment heterogeneity: because we have set V to equal X for the second strata only, and to zero otherwise. This is equivalent to the omitted interaction term $\delta_2 D_2 X$. As a result, the model in column (3) is correctly specified and the combination of the estimates for X and V fully capture the treatment effect heterogeneity. The estimate for X now represents the effect for strata 1 only, given by δ_1 from the interacted specification 6. This is not an estimate of the average effect, i.e. β^{AET}.

Last but not least, in column (5) W is used as instrument for X. As expected, the estimate for the LATE-compiler population, shown in column (5) of Table 3, is close to five.

4 Discussions and conclusions

Whenever treatment effects are heterogeneous OLS that pools over strata can only be interpreted as estimate for an average population effect if the variation in the explanatory variable is constant across strata. This is a special case and there is no clear reason to assume this assumption is met in practice. Moreover, whenever control variables do not correlate in the same way with the explanatory variable across strata, every time a (different) control variable is added, a different re-weighting occurs. In practice, this means that coefficient movements across specifications with different (sets of) control variables do not have a clear interpretation: these can occur because of the re-weighting of strata, as argued and demonstrated in this note, or because of changes in "traditional" omitted variable bias.

Notice that this is a very general property of OLS that pools over strata and so equally applies to coefficient movements from the inclusion of controls in DiD estimates that are obtained from OLS. Notably, experimental estimates are not affected since with experimental assignment the treatment variation is orthogonal to controls. Similarly, RDD

estimates already estimate local effects across a cut-off where controls remain constant so that including more controls cannot lead to re-weighting.

How likely is it that control variables correlated differently with regressors across strata? The IV literature is full of examples of variables that do not predict regressors similarly across strata, giving rise to the LATE interpretation. There is no reason to believe that other controls that are included to alleviate potential endogeneity would behave any different. In our simple example, adding a control that correlates highly with strata of larger treatment effects reduces the OLS coefficient, away from the LATE-IV estimate. This replicates a very common pattern in the applied literature of smaller and smaller OLS estimates (when more controls are added), and then (surprisingly) large IV estimates, without endogeneity.

A further side-effect of the re-weighting from the inclusion of controls is that even control variables that are irrelevant in the homogeneous world become predictive and can show up as being significant at conventional levels. This is because controls that do not correlate uniformly with strata proxy for the omitted interaction terms that capture the effect heterogeneity. From a purely statistical perspective adding such proxies is indeed preferable and model selection routines that are based on statistical significance or other criteria such as the Mean Squared Error (which falls from the inclusion of the "irrelevant" control), including many ML and lasso selection routines, will opt for models that include controls. Unfortunately, in these models the main estimate of interest cannot any more be interpreted as average effect. As a result, whenever heterogeneity is not ruled out or modelled, estimates between specifications that include different sets of controls are not comparable.

References

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber, "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 2005, *113* (1), 151–184.

Angrist, Joshua D. and Jörn-Steffen Pischke, "Mostly Harmless Econometrics," *Princeton University Press*, December 2008.

Goodman-Bacon, Andrew, "Difference-in-differences with variation in treatment timing," *Journal of Econometrics*, 2021, *225* (2), 254–277. Themed Issue: Treatment Effect 1.

Oster, Emily, "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business & Economic Statistics*, 2019, *37* (2), 187–204.

Rao, P. S. S. N. V. P. and M. Precht, "On a Conjecture of Hoerl and Kennard on a Property of Least Squares Estimators of Regression Coefficients," *Life Sciences*, 1985.

Słoczyński, Tymon, "Interpreting OLS Estimands When Treatment Effects Are Heterogeneous: Smaller Groups Get Larger Weights," *The Review of Economics and Statistics*, 05 2022, *104* (3), 501–509.