Choosing Controls in Regression Analyses Involving Equity

Note: The Equity Evaluation Memo Series is intended to guide OES' commitment to equity in our evaluation process and efforts toward understanding and reducing barriers to equitable access to federal programs. This series is intended to be an internal guidance document for OES team members.

Summary: When you are interested in the question of whether a demographic group receives access to a program at different rates compared to other demographic groups, irrespective of the reason, then a bivariate model will answer that question. However, if you are interested in whether otherwise similarly situated individuals of different demographics access a benefit at the same rate, then including pre-treatment controls that are correlated with the demographic characteristic and the outcome is typically appropriate, depending on the functional form of the model.

Selecting Control Variables in Regression Analysis for Equity Evaluation

Multiple regression estimation can be a powerful tool for understanding equity in evaluations.

As analysts, when we want to know whether some group accessed a program or received some benefit at a different rate than other groups, and we have a dataset that records whether individuals accessed a program alongside their demographics and other contextual factors, many of us may reach for multiple regression to estimate group-level differences while controlling for confounding variables. However, we must be very careful when choosing which controls to include or exclude in studies of this kind. Incorrect inclusion and exclusion of control variables can lead to the wrong inference about equity in program access.

This guidance memo describes how you should specify your regression model to match your research question. You can also try out the Stata and R code that create the data and run the regressions. In the appendix, we also translate all the examples into "directed acyclic graphs" (DAGs), as an additional way of illustrating why different approaches give you the wrong answer.

We are particularly interested in the concept of statistical "bias" — statistical bias arises when you use an estimator that, if you could run the same study lots of times, would, on average, give you an answer that differs from the true answer you'd like to know (your estimand). No single study ever tells us the "true" answer, but we want to use estimation strategies that we expect to give us the right answer *on average*.

Post-attribute bias – we falsely infer there is (in)equity in funding when we include a "bad control" variable[1]

One factor that business relief programs often use to score applications for funding — with the aim to support financially viable businesses — is profitability (e.g., not to prop up businesses that would have failed anyway, see here for an example). However, women-owned businesses (WOBs) typically have lower profitability than non-WOBs. In these cases, bias may arise when WOB is measured after the demographic attribute is introduced.

Let's imagine we had a dataset with six businesses—three WOBs and three non-WOBs—and the quantity we want to know (the "estimand") was: what is the difference in the probability of WOBs and non-WOBs getting funded?

Table 1. An Imaginary Dataset on Small Business Relief Applications

Business ID	Women-Owned	Earnings (tens of thousands)	Funded
1	0	50	1
2	0	50	1
3	0	50	1
4	1	10	0
5	1	10	0
6	1	50	1

Many analysts would be tempted to control for different situational factors in order to get a more precise estimate. In other words, we might be tempted to run the "full" model (Table 2) below, as opposed to the very simple "bivariate" model. Table 2 shows the code for the two models and what the estimated difference in the probability of funding for WOBs and non-WOBs would be.

Table 2. Regression Model to Estimate Inequity in Probability of Funding for WOBs

Regression Model	Stata code	R code	Estimated inequity in probability of funding:
Full	reg funded women_owned earnings	lm(funded ~ women_owned + earnings)	0 percentage points
Bivariate	reg funded women_owned	lm(funded ~ women_owned)	-67 percentage points

If you ran a regression of *funded* on *women-owned* controlling for *earnings* using the dataset above, the coefficient on *women-owned* would be zero. So, you might infer from your regression that there is no inequity between WOB and non-WOBs. But in this conveniently small and simple dataset, it's plain to see that WOBs are underrepresented among the funded pool (50% versus 33%)! If you were to run a bivariate regression of *funded* on *women-owned* you would detect the inequity: the coefficient implies WOBs had a 67 percentage point lower chance of being funded than non-WOBs. So what's going on?

[1] We borrow the term "bad controls" from Cinelli, et al., "A Crash Course in Good and Bad Controls." See https://ftp.cs.ucla.edu/pub/stat_ser/r493.pdf.

Whether or not a business gets funded is perfectly predicted by their earnings – every business with earnings of 500,000 USD got funded and no business with 100,000 USD in earnings got funded. Bias arises because earnings are *also* negatively correlated with being a WOB: while the non-WOBs have an average earnings of 500,000 USD, the WOBs have average earnings of 233,333 USD. Because earnings is both a good predictor of whether a business is women-owned *and* perfectly predictive of whether it gets funding, there is no variation in the funding status left for the women-owned variable to account for. The bivariate regression does not suffer from this bias.

Of course, there is only bias in an estimator to the extent that there is an estimand that it is incorrectly estimating. If the question were not "do WOBs access funding at different rates than non-WOBs?", but were instead "do WOBs access funding at different rates to *similarly situated* non-WOBs?", then the answer that the full regression provides might in fact be right: WOBs and non-WOBs *with the same profitability* do not access funding at different rates.

Whether or not we should use the full or bivariate model in the above example depends on whether we want to understand inequity in access *for any reason* or *for reasons unrelated to profitability*. There is an important related point. As stated above, if you are interested in evaluating equity in access to funding for WOBs and similarly situated non-WOBs, you will find that WOBs and non-WOBs *with the same profitability* do not access funding at different rates and conclude that there is no inequity in funding decisions. However, the fact that WOBs have lower profitability is indicative of pre-existing differences before the funding decision was made; profitability is only measured after the introduction of the woman owner. Although being a WOB may not affect the funding decision, it may affect profitability, and by extension the funding outcome. In such cases, answering the question of whether there is inequity in access for any reason, using a bivariate model, will give you a more accurate picture of equity in funding *outcomes* than the question of whether there is inequity in access for similarly situated WOBs and non-WOBs.

Omitted variable bias – we falsely infer there is (in)equity in funding when we exclude a control variable

This next example is inspired by the fact that, in the wake of the pandemic, early rounds of federal relief for small businesses were not available to non-employer businesses (often referred to as sole proprietorships). To make up for this gap in access, many local governments created programs directly targeted toward sole proprietors. Importantly, Black-owned businesses (BOBs) are more likely to be sole proprietorships (see here).

Table 3. A Second Imaginary Dataset on Small Business Relief Applications

Business ID	Black-Owned	Sole proprietorship	Funded
1	0	0	0
2	0	0	0
3	0	1	1
4	0	1	1
5	1	0	0
6	1	1	1
7	1	1	0
8	1	1	1

Suppose that a researcher understood equity to mean that similarly situated BOBs and non-BOBs get funding at the same rate. Should that researcher control for whether businesses are sole proprietors or not?

Table 4. Regression Model to Estimate Inequity in Probability of Funding for BOBs

Regression Model	Stata code	R code	Estimated inequity in probability of funding:
Full	reg funded black_owned sole_prop	lm(funded ~ black_owned + sole_prop)	-21 percentage points
Bivariate	reg funded black_owned	lm(funded ~ black_owned)	0 percentage points

If you ran a bivariate regression of *funded* on *black-owned*, the coefficient on *black-owned* will be zero. However, it would be incorrect to infer that similar BOBs and non-BOBs have an equal likelihood of getting funding. If you ran a regression of *funded* on *black-owned* and *sole proprietorship*, you would find that BOBs were 20 percentage points less likely than non-BOBs to receive funding. So what's going on?

In this example, sole proprietors are more likely to get funded — there's only one sole proprietorship that is not funded, and no non-sole proprietorship that is funded. Because BOBs are more likely than non-BOBs to be sole proprietors (¾ vs. ½), that also means that BOBs are *unconditionally* more likely to be funded. Note, however, that the only unfunded sole proprietorship is Black-owned. *Conditional* on being a sole proprietor, ⅔ of BOBs got funding versus 100% of non-BOBs. Thus, if we don't account for the fact that BOBs are more likely to be sole proprietors, we are overly optimistic about their rates of funding compared to non-BOBs.

Again, whether the bivariate regression estimator is biased here—i.e., whether the answer it gives us is systematically different from the true answer to the question we are posing—depends on our question. If our question were instead more similar to the first question, such as "did Black-owned businesses receive less funding than non Black-owned businesses (for whatever reason)?", the answer we get from the bivariate model would be, correctly, no. However, if we are concerned about whether similarly situated BOBs and non-BOBs are funded at the same or similar rates, then the bivariate model will mask potential discrimination or other barriers that Black sole proprietors face compared to non-Black sole proprietors.

Using Regression to Detect Discrimination

Caution should be taken when using this approach to detect *discrimination*—unequal outcomes that arise *due to* demographic differences and not other factors. Even if you control for all *available* contextual variables that should predict whether a business gets funded and still find WOBs got less funding than non-WOBs, for example, that does not necessarily mean WOBs were discriminated against in the funding decision.

Suppose that your dataset does not measure whether businesses have staff accountants, that WOBs are less likely to have on-staff accountants than non-WOBs, and that businesses with on-staff accountants are more likely to get funding because they can submit complete applications more quickly. In that case, even if program staff treat WOBs equally to non-WOBs, after controlling for all situational factors you will still find a funding gap that is due to the on-staff accountant variable that you cannot measure.

Note that this issue also means you cannot *rule out* discrimination, even if you find no gaps. To see this point, return to the example of excluded variable bias for BOBs above, and imagine that your dataset did not measure whether a business was a sole proprietorship—your inability to control for this variable might mean you fail to detect discrimination against black-owned sole proprietorships.

So what *can* we say about equal treatment, based on available data? In some contexts, your dataset might actually measure *all* available data that program staff used to make allocation decisions. If, for example, program staff make decisions on how to allocate some benefit based entirely on the same dataset to which you have access – they never contact potential beneficiaries directly or conduct internet searches to gather information beyond what you have in your dataset – and after controlling for all variables in the dataset, you still find that demographic characteristics are predictive of decision-making, then you may have grounds to infer those demographic characteristics were used to make unequal decisions.[2]

In most cases, however, establishing discrimination will require designing and defending a research strategy that can credibly claim that other factors besides discrimination could not account for the estimated differences in the outcomes. This can be done in a number of ways. Tuttle (2019), for example, estimates racial disparities in federal sentencing by comparing cases sentenced before and after the Fair Sentencing Act, a 2010 law that changed the 10-year mandatory minimum threshold for crack-cocaine, and finds a disproportionately large increase in sentencing for Black and Hispanic offenders at the point that now triggers a 10-year mandatory minimum.[3] Pierson et al. (2020) find evidence of racial disparities in police stop decisions by analyzing nearly 100 million traffic stops across the U.S..[4] They found that Black drivers were

---

[2] One example of a context in which all available data used to make allocation decisions is available to program staff could be the refugee resettlement process. See: Bansak et al., "Improving refugee integration through data-driven algorithmic assignment" *Science*, Vol 359, Issue 6373 (2018): 325-329.
[3] Tuttle, Cody. "Racial disparities in federal sentencing: Evidence from drug mandatory minimums." *Available at SSRN* (2019).
[4] Pierson, E., Simoiu, C., Overgoor, J. *et al.* A large-scale analysis of racial disparities in police stops across the United States. *Nat Hum Behav* 4, 736–745 (2020). https://doi.org/10.1038/s41562-020-0858-1

less likely to be stopped after sunset, when a 'veil of darkness' masks drivers' race. Interested readers can see Bertrand and Duflo (2016) for a review of experimental methods that have been employed to measure the prevalence of discrimination, including audit and correspondence studies.[5]

If no research design to detect discrimination is possible, then prudent interpretation of the results is important. For example, one might state that WOBs that are similarly situated on all measurable dimensions experience funding at lower rates than non-WOBs, while acknowledging that discrimination is only one among many potential omitted variables that could explain the (lack of) gaps.

Including Uncorrelated Variables and Intersecting Equity Concerns

There are two additional scenarios that you are likely to encounter when evaluating equity using regression. The first is inclusion of control variables that are not correlated with the demographic variable of interest but *are* correlated with the outcome in which you're interested. Whether the business is located in a priority zone in Table 5 below does predict whether a business gets funded but is not predictive of the business-owner's race. If you were to control for being in a *priority zone* when regressing whether an applicant is *funded* on whether they are *Black-owned* you will find that your coefficient on *Black-owned* is unchanged (see the second regression in Table 6 below). Inclusion of such variables will reduce noise, by making your standard errors smaller and your inferences more precise. Therefore, analysts should be less concerned about inclusion of uncorrelated variables as they will not bias their estimates. However, if you were to also include a control for *sole proprietorship* in this model you would find that your coefficient on *Black-owned* changes (from -21 to -24 percentage points, see the third regression in Table 6), because of the correlation of B*lack-owned* with *funded*, *sole proprietorship*, and *priority zone*. The difference is relatively small and is likely driven by the small sample size. But the example reveals the point that relationships between variables can get more complex with the inclusion of more variables in your regression model. One simple way to examine how and whether such relationships matter for the inferences you draw is to see how the estimated coefficient on the demographic variable of interest (here, the BOB indicator) changes as you add or subtract controls.

Table 5. Third Imaginary Dataset on Small Business Relief Applications

Business ID	Black-Owned	Sole proprietorship	Funded	In Priority Zone
1	0	0	0	0
2	0	0	0	1
3	0	1	1	1
4	0	1	1	1
5	1	0	0	0
6	1	1	1	1
7	1	1	0	1
8	1	1	1	1

[5] Bertrand, M., & Duflo, E. (2017). Field experiments on discrimination. *Handbook of economic field experiments*, *1*, 309-393.

Table 6. Regression Model to Estimate Inequity in Probability of Funding for BOBs

Regression Model	**Stata code**	**R code**	**Estimated inequity in probability of funding:**
Bivariate regression	reg funded black_owned	lm(funded ~ black_owned)	0 percentage points
Including Priority Zone	reg funded black_owned priority_zone	lm(funded ~ black_owned + priority_zone)	0 percentage points
Including Priority Zone and Sole Proprietorship	reg funded black_owned sole_prop priority_zone	lm(funded ~ black_owned + sole_prop + priority_zone)	-24 percentage points

The second issue is one where there are intersecting equity concerns. Consider a case where WOBs and BOBs are each treated equally as other businesses, however, a business that is both a WOB and BOB is not (see Table 6). In such a case, whether you are trying to answer the question, "does a business which is Black-women owned access funding at different rates to non-Black-women-owned businesses, for whatever reason?" or "do Black-women owned businesses access funding at different rates to non-Black-women-owned businesses, because they are Black-women-owned?" you will need to include both *women-owned*, *Black-owned*, and *Black-women-owned* indicators in your model. Failure to include the intersection would lead to wrong inferences about whether inequity in access to funding or inequitable treatment existed for this group. In the example below, the funding rate of women-owned vs. non-women owned businesses is 2/4 vs. 3/6, and the funding rate of Black-owned vs. non-Black-owned businesses is also 2/4 vs. 3/6. So, on average, these two groups fare similarly. However, when we look at intersections, things are different: whereas 1/4 of non-Black-non-womened owned businesses were funded, and all non-Black, women-owned businesses and non-women, Black-owned businesses were funded, no single business that was Black-owned and women-owned was funded.

Table 7. Fourth Imaginary Dataset on Small Business Relief Applications

Business ID	**Black-Owned**	**Women-Owned**	**Funded**
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	1
5	0	1	1
6	0	1	1
7	1	0	1
8	1	0	1
9	1	1	0
10	1	1	0

Table 8. Regression Model to Estimate Inequity in Probability of Funding for Black-Women Owned Businesses

Regression Model	**Stata code**	**R code**	**Estimated inequity in probability of funding:**
Including Intersection	reg funded women_owned black_owned women_owned##black_owned	lm(funded ~ women_owned + black_owned + women_owned:black_owned)	0.75 percentage points (women owned, Black owned), -1.75 percentage points (Black-women owned)
Not Including Intersection	reg funded women_owned black_owned	lm(funded ~ women_owned + black_owned)	0 percentage points (women owned, Black owned)

Appendix: Interpreting the Examples Through the Lens of DAGs

Directed acyclic graphs (DAGs) are graphical representations of one-directional causal relationships between variables. They are often used as a heuristic device to decide what controls to include given: a) a specific type of causal relationship one seeks to estimate, and b) assumptions about how variables depend on other variables.[6] Below, we use the web interface "DAGitty" to represent the examples above in DAG format. Each variable is represented by a circle, and a causal relationship running from X to Y is represented by an arrow pointing from X to Y. The green circle with a triangle in it represents the cause of interest, and the blue circle with a bar in it represents the outcome. Blue circles without bars are variables that we can observe but we have omitted from the regression (we are not using them to "adjust" our estimates in any way). Gray circles with Black outlines represent variables that we are adjusting for, by including them in the regression. If an arrow between two variables is Black, that means that we have blocked the pathway between those variables. If it is green, it means the pathway is open and we can estimate it in an unbiased manner. If it is red (not the case in any examples below), then the causal pathway is confounded and our estimates of that relationship will be biased.

Example 1. Controlling for profitability leads you to miss inequity between women-owned and non-women-owned businesses

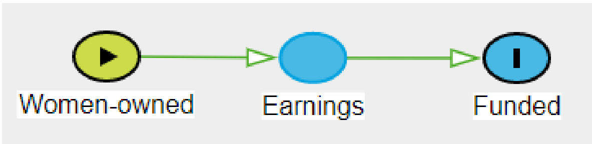
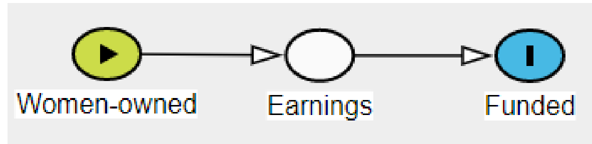
In this example, we are interested in whether women-owned businesses accessed funding at similar rates to non-women-owned businesses, for any reason, not just because they are women-owned. One way to interpret this is as a *descriptive* question: we simply want to know how the funding distribution looks according to the gender of business owners, without assuming any causality. In this case, we don't need DAGs: simply comparing the average rate of funding among businesses that are and are not women-owned.

But we could also express this as a causal question, namely: what is the *total* effect of a business being women-owned on its probability of being funded. The DAGs on Table A1 below represent a scenario consistent with the first example.

There is no discrimination in the program, in that there is no direct effect of a business being women-owned on the probability of that business being funded. However, a business being women-owned does affect earnings, and earnings affect the probability of being funded. If we control for earnings, we block the path between the causal variable we care about (women-owned) and the outcome (funded). If we do not control, we are able to estimate the total effect.

[6] For a foundational text in this field, see: Pearl, J. *Causality: models, reasoning and inference.* 2000, Cambridge University of Cambridge. For a lighter introduction, see here.

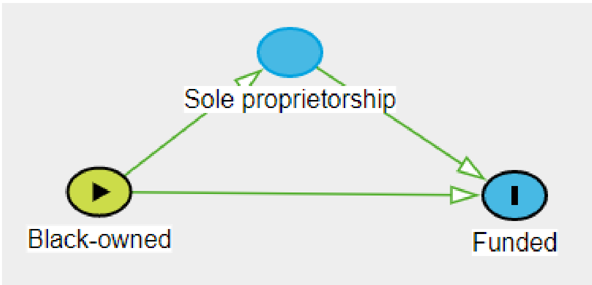
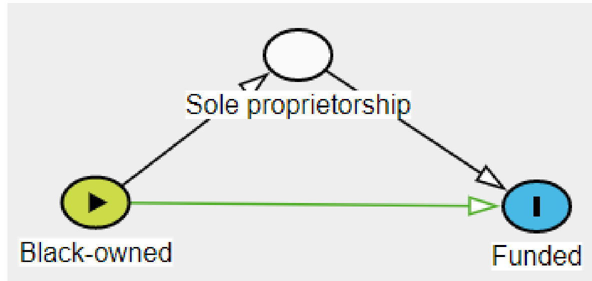
Table A1. Example 1 Expressed as DAGs

<table>
<tr><td>Bivariate regression</td><td>Full regression</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>No adjustment is necessary to estimate the total effect of women-owned on Funded. By leaving the path open, we are able to estimate that being women-owned decreases the probability of being funded (through the effect on earnings).</td><td>The total effect cannot be estimated due to adjustment of the causal variable we care about. Since there is no direct effect of being women-owned on being funded, and we have blocked the indirect path, we estimate no relationship between the causal variable we care about and the outcome.</td></tr>
<tr><td colspan="2">Enter the following code into the http://www.dagitty.net/ visualizer to reconstruct these DAGs:</td></tr>
<tr><td>dag {
bb="-2.312,-2.713,2.361,2.672"
"Women-owned" [exposure,pos="-0.477,0.036"]
Earnings [pos="-0.049,0.041"]
Funded [outcome,pos="0.373,0.042"]
"Women-owned" -> Earnings
Earnings -> Funded
}</td><td>dag {
bb="-2.312,-2.713,2.361,2.672"
"Women-owned" [exposure,pos="-0.477,0.036"]
Earnings [adjusted,pos="-0.049,0.041"]
Funded [outcome,pos="0.373,0.042"]
"Women-owned" -> Earnings
Earnings -> Funded
}</td></tr>
</table>

Example 2. Not controlling for a business being a sole proprietorship leads you to miss discrimination against Black-owned businesses

In the second example, we want to know whether similarly situated Black-owned and non-Black-owned businesses get funding at the same rate. One way of stating this is that we want to know the direct effect of a business being Black-owned on its probability of being funded. The DAGs on Table A2 represent this example. Recall that the effect of being Black-owned on the probability of being a sole proprietorship was positive, and the effect of being a sole proprietor on being funded was also positive, but that the direct effect of being a Black-owned business on being funded was negative. In this example, not controlling for sole proprietorship by running the bivariate regression means we estimate the total effect of these relationships, which cancel each other out. However, if we control for sole proprietorship status, we are able to estimate the negative direct effect of Black ownership on funding, and detect the discrimination.

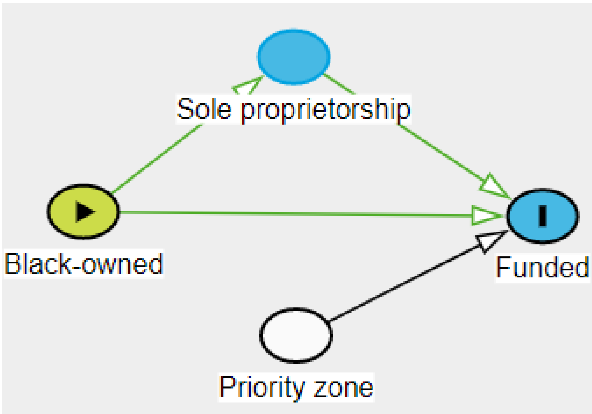
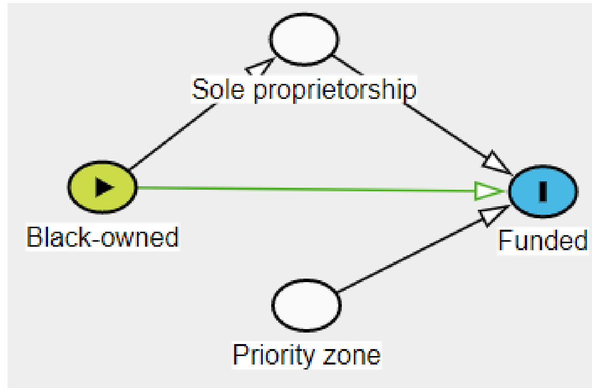
Table A2. Example 2 Expressed as DAGs

<table>
<tr><td>Bivariate regression</td><td>Full regression</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>By leaving the indirect path running from Black-owned to funded via sole proprietorship open, the bivariate regression of funded on Black-owned leads us to estimate the total effect.</td><td>Adjusting for sole proprietorship blocks the indirect causal pathway from Black-owned to funded, which is positive, and leaves open only the direct effect, which is negative. We now detect the discrimination against Black-owned firms.</td></tr>
<tr><td colspan="2">Enter the following code into the http://www.dagitty.net/ visualizer to reconstruct these DAGs:</td></tr>
<tr><td>dag {
bb="0,0,1,1"
"Black-owned" [exposure,pos="0.325,0.430"]
"Sole proprietorship" [pos="0.412,0.328"]
Funded [outcome,pos="0.515,0.433"]
"Black-owned" -> "Sole proprietorship"
"Black-owned" -> Funded
"Sole proprietorship" -> Funded
}</td><td>dag {
bb="0,0,1,1"
"Black-owned" [exposure,pos="0.325,0.430"]
"Sole proprietorship"
[adjusted,pos="0.412,0.328"]
Funded [outcome,pos="0.515,0.433"]
"Black-owned" -> "Sole proprietorship"
"Black-owned" -> Funded
"Sole proprietorship" -> Funded
}</td></tr>
</table>

Example 3. Controlling for an unconfounded parent of the outcome increases precision but does not change inferences

Example 3 augments example 2 by adding a variable, being in a priority zone, that is causally unrelated to the demographics of the business owner(s), but is a cause of the outcome (funding). The DAGs below make clear why controlling for this variable just changes the standard errors, but does not affect the inference about the direct or total effect of being Black-owned on funding: the variable we are controlling for in this case does not lie on the path from Black-owned to funded.

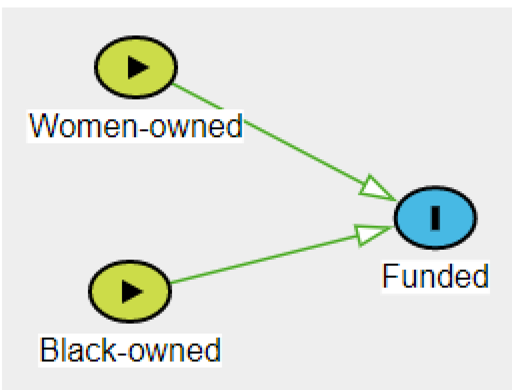
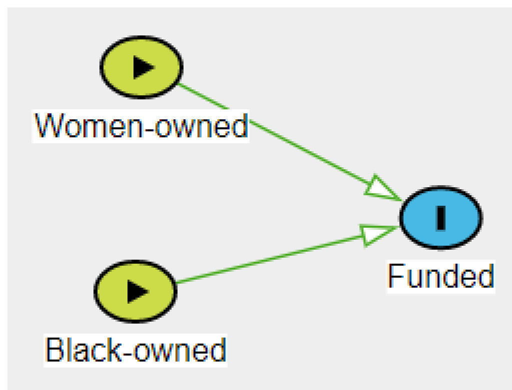
Table A3. Example 3 Expressed as DAGs

<table>
<tr><td>Including priority zone</td><td>Including priority zone and sole proprietorship</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td>By leaving the indirect path running from Black-owned to funded via sole proprietorship open, the regression of funded on Black-owned controlling for priority zone only leads us to estimate the total effect of Black-owned on funded (which here is close to zero due to the canceling out of the direct and indirect paths).</td><td>Adjusting for sole proprietorship blocks the indirect causal pathway from Black-owned to funded, which is positive, and leaves open only the direct effect, which is negative. We now detect the discrimination against Black-owned firms. Controlling or not controlling for priority zone does not affect our estimate of the relationship between Black-owned and funded.</td></tr>
<tr><td colspan="2">Enter the following code into the http://www.dagitty.net/ visualizer to reconstruct these DAGs:</td></tr>
<tr><td>dag {
bb="0,0,1,1"
"Black-owned" [exposure,pos="0.325,0.430"]
"Priority zone" [adjusted,pos="0.413,0.512"]
"Sole proprietorship" [pos="0.412,0.328"]
Funded [outcome,pos="0.515,0.433"]
"Black-owned" -> "Sole proprietorship"
"Black-owned" -> Funded
"Priority zone" -> Funded
"Sole proprietorship" -> Funded
}</td><td>dag {
bb="0,0,1,1"
"Black-owned" [exposure,pos="0.325,0.430"]
"Priority zone" [adjusted,pos="0.413,0.512"]
"Sole proprietorship"
[adjusted,pos="0.412,0.328"]
Funded [outcome,pos="0.515,0.433"]
"Black-owned" -> "Sole proprietorship"
"Black-owned" -> Funded
"Priority zone" -> Funded
"Sole proprietorship" -> Funded
}</td></tr>
</table>

Example 4. Including an Interaction Reveals Intersectional Inequity

In example four, we only detect the inequity in the program by looking at the intersection of race and gender. This example helps to show the limits of DAGs. Traditionally, DAGs are non-parametric: in a graph such as X->Y<-Z, the DAG only specifies that Y is produced by a combination of X and Z, but it does not tell us the specific form of this relationship. For example, the following functional equations are all consistent with this DAG: Y = X*Z; Y = X + Z; Y = X^2*Z^2; and so on. A DAG cannot tell us whether we *should* include an interaction (see [here]() for a proposal for inclusion of interactions in DAGs) in a regression, even though in this case it matters a lot whether we do. In general, therefore, a good rule of thumb is to investigate interactions between demographic categories.

Table A4. Example 4 Expressed as DAGs

<table>
<tr><td>Including intersection</td><td>Not including intersection</td></tr>
<tr><td>
</td><td>
</td></tr>
<tr><td colspan="2">These graphs are the same, because a DAG encodes non-parametric dependencies between variables, it does not encode the nature of the functional relationships between variables. The fact that the effect of being women-owned (black-owned) is different when the business is black-owned (women-owned) cannot be represented using a conventional DAG.</td></tr>
<tr><td colspan="2">Enter the following code into the http://www.dagitty.net/ visualizer to reconstruct these DAGs:</td></tr>
<tr><td colspan="2"><pre>dag {
bb="0,0,1,1"
"Black-owned" [exposure,pos="0.404,0.476"]
"Women-owned" [exposure,pos="0.407,0.346"]
Funded [outcome,pos="0.515,0.433"]
"Black-owned" -> Funded
"Women-owned" -> Funded
}</pre></td></tr>
</table>