IDEA-Aligned Estimates of Racial Disproportionality in Special Education versus Conventional Approaches: A cautionary note on included-variable bias when achievement and socioeconomic status proxy for special education need

February 14, 2026

Abstract

Racial disproportionality in special education is a contested policy space. Federal oversight has traditionally focused on minority over-representation through IDEA's significant disproportionality framework. However, observational studies report that Black students appear under-identified based on a canonical model that regresses special education receipt on race and controls, notably prior achievement and socioeconomic status (SES). Drawing on this evidence, recent federal proposals would scale back oversight tied to significant disproportionality determinations. We formalize an IDEA-aligned estimand and show analytically that the canonical model recovers it only under strong assumptions. Most notably, it requires that the residual Black–White achievement gap net of SES reflects a gap in disability-related need rather than residual differences in opportunity to learn and other exclusionary factors. In simulations calibrated to empirically plausible environments, the canonical model overstates Black under-representation to an extent sufficient to fully account for previously reported levels. This bias stems from over-controlling for components of social inequality IDEA regards as distinct from disability. We further show popular sensitivity analyses perform poorly in addressing this issue. The implication is that negative coefficients from standard adjustments should not be interpreted as evidence of IDEA-aligned under-service without designs that better separate disability from opportunity.

Keywords: special education policy; IDEA implementation; racial disproportionality; classification practices

1 Introduction

Descriptive analyses of US schools consistently show that racially minoritized students, and Black students in particular, receive special education services at higher rates than White students, where service receipt is typically measured as the presence of an Individualized Education Program (IEP) (Artiles 2013; Donovan and Cross 2002; Dunn 1968). These disparities have concerned researchers and policymakers for decades, and federal policy has required states to monitor and address racial disproportionality in special education since 2004 (IDEA-Sec.300.309 2006).

To interpret these disparities, it is important to distinguish among the sources of racial differences in IEP receipt. Two main sources stand out. First, pre-existing racial inequalities in health, neighborhood conditions, and schooling opportunities may mean that Black students, on average, have higher special education needs at the time of classification decisions (Artiles et al. 2002; Donovan and Cross 2002; Shifrer 2018). Second, school-based classification practices themselves may systematically favor particular racial groups, in the sense that Black and White students with similar underlying special education need may be treated differently by teachers, evaluators, or eligibility teams. If racial favoritism contributes to observed disparities, then reforming identification processes is essential for equity. In contrast, if disparities can be fully explained by pre-existing inequalities, then policy attention should focus on those earlier stages rather than the classification system itself.

Distinguishing between these explanations requires clarity about what constitutes a special education need. Federal law explicitly differentiates disability from disadvantage. Consider, for instance, eligibility for specific learning disabilities (SLDs), the most common disability category. Under the Individuals with Disabilities Education Act (IDEA), an SLD classification requires both low achievement in one or more academic domains and evidence that this low achievement cannot be explained by exclusionary factors such as inadequate instruction, limited English proficiency, or socioeconomic disadvantage (IDEA-Sec.300.309 2006). School teams are therefore asked to decide whether a child's low achievement sig-

nals disability-related need for special education, or instead reflects unequal opportunities to learn.

Traditionally, quantitative studies have tried to differentiate these sources of disparities through a regression of IEP receipt on race and other student characteristics, most prominently prior achievement (measured through test scores) and socioeconomic status (SES) indicators (e.g., Fish 2019; Hibel et al. 2010; Kincaid and Sullivan 2017; Morgan et al. 2012, 2015, 2020; Shifrer 2018; Sullivan and Bal 2013) — what we call the "canonical" regression. These studies generally find that, after controlling for achievement and SES, Black students are less likely than White students to receive special education services. This pattern has been widely interpreted to mean that Black students are under-identified relative to White students with similar special education need (Morgan and Farkas 2015; Morgan et al. 2015).

In this paper, we show that this common regression framework is unlikely to estimate what researchers intend: racial differences in IEP receipt net of special education need. By estimating racial differences conditional on prior achievement and socioeconomic status, the canonical comparison can depart from IDEA's exclusionary-clause understanding of special education need. The problem arises because achievement and socioeconomic status each reflect both disability-related need and unequal opportunities to learn. Conditioning on these variables therefore conflates disability and disadvantage in ways that can bias estimates of racial differences in classification practices.

This paper explains and formalizes this argument. Our central point is that, once these assumptions are made explicit, the standard regression framework does not, in general, estimate Black–White differences in IEP receipt among students with the same special education need. Our analysis yields the following contributions.

1. We formalize an IDEA-aligned estimand τ^* that compares Black and White students with the same underlying special education need, D_i^*.
2. We show how an oracle regression would identify this estimand if D_i^* were observed. Relative to this benchmark, we derive a simple decomposition of the canonical race co-

efficient α_1 into the oracle effect plus a bias term governed by how SES and achievement proxy for special education need.

3. We show, analytically, that the canonical model recovers the target estimand only under strong assumptions. Most notably, that *the residual Black–White achievement gap conditional on SES must fully and only reflects racial differences in special education need.* This reveals that under IDEA's understanding of special education need, SES and prior test performance play different roles in the canonical regression. Achievement is intended to operationalize Black–White differences in learning difficulties while SES is meant to absorb exclusionary factors
4. Our derivations demonstrate that, under plausible empirical configurations, canonical estimates can understate IDEA-aligned disparities by construction. This pattern arises because achievement covariates likely pick up more Black-White differences in exclusionary factors than SES controls can reliably account for. Thus, conditioning on these variables can result in included-variable bias (Ayres 2005; Jung et al. 2024; Souto-Maior and Shroff 2024), over-controlling for components of social inequality many would regard as distinct from disability.
5. Using Monte Carlo simulations, we show that in a wide range of empirically plausible data-generating environments the canonical regression implies greater under-representation of Black students even when the IDEA-aligned estimand is assumed to be zero. Further, the magnitude of this bias can be sufficient to fully account for previously reported levels of under-representation.
6. We show that popular sensitivity analyses designed for omitted-variable bias, such as E-values and Oster bounds, perform poorly in addressing the issue we identify here. In our simulations, these diagnostics often look most reassuring in the scenarios where the canonical coefficient is furthest from the IDEA-aligned estimand, giving applied researchers a false sense of robustness about estimates that in fact mischaracterize Black–White disproportionality in IEP receipt.

From a policy standpoint, the research of Morgan et al. (2012) has been cited by the Heritage Foundation's *Project 2025* to argue against the "Equity in IDEA" framework (Burke 2023). Recent executive orders discourage the use of statistical disparities as a trigger for civil-rights intervention and direct agencies to curtail disparate-impact style enforcement. For example, Executive Order 14173 (January 21, 2025) frames race-conscious compliance initiatives as unlawful preferences, and the April 23, 2025 Executive Order, "Restoring Equality of Opportunity and Meritocracy," announces a federal policy to eliminate disparate-impact liability and instructs agencies to deprioritize enforcement and revisit Title VI implementing regulations. While prior critiques of this research highlighted measurement challenges and the conceptual risks of oversimplification (Fish et al. 2025; Skiba et al. 2016), our contribution is to formally interrogate and deconstruct the regression estimand itself. Consequently, regression evidence of "under-representation after controls" offers little warrant to roll back disproportionality monitoring. Rather than providing a quantitative estimate that Black students are being neglected relative to similarly situated White students, such results highlight that standard adjustments cannot reliably distinguish disability-related need from opportunity-driven low achievement, motivating oversight approaches that better separate need from opportunity to learn.

The remainder of the paper proceeds as follows. In the next section we formalize an IDEA-aligned estimand, τ^*, and derive the oracle regression that would identify it if special education need were observed. We then introduce a simple structural framework and its graphical representation to relate this oracle model to the canonical (SES-and-achievement) regression, making clear how the latter departs from the former and how it reallocates racial differences in achievement into special education need versus unequal opportunity. We then define and bound the bias from the canonical regression in relation to the proposed oracle specification. Finally, we present the simulation study, which uses the same structural framework to quantify how bias varies across empirically plausible environments and to illustrate why common sensitivity analyses can fail in this setting.

2 Oracle versus Canonical Models for Estimating Disproportionality in IEP Receipt

In this section, we formalize the policy-relevant estimand that corresponds to IDEA's distinction between special education need and exclusionary sources of low achievement. We then introduce an oracle regression that would identify this estimand if special education need were observed, and we show how the common practice of conditioning on achievement and socioeconomic status departs from this benchmark when achievement is also shaped by unequal opportunity to learn. Finally, we derive the conditions under which the canonical regression can approximate the oracle comparison and clarify what is implicitly assumed when residual Black–White achievement gaps are interpreted as differences in special education need.

2.1 The estimand and its identification through an oracle regression

Let Y_i be an indicator for whether student i receives an IEP, and let Black_i indicate Black students. We use D_i^* to denote the student's underlying *special education need*, understood as the disability-related need that IDEA intends eligibility: learning difficulties that cannot be attributed to exclusionary factors such as inadequate instruction, limited English proficiency, or socioeconomic disadvantage. By defining D_i^* in this way, we treat exclusionary factors as operating outside of special education need, so that holding D_i^* fixed corresponds to comparing students whom IDEA would regard as having the same special education need even if their opportunities to learn differ. Then, to assess whether special education classification practices themselves are racially biased, the descriptive estimand of interest is the difference in the probability of IEP receipt between Black and White students who share the

same level of underlying special education need:[1]

$$\tau^*(d) \equiv \mathbb{E}\big[Y_i \mid \text{Black}_i = 1, D_i^* = d\big] - \mathbb{E}\big[Y_i \mid \text{Black}_i = 0, D_i^* = d\big].$$

If D_i^* were observed, a natural benchmark would be an *oracle regression* that conditions directly on special education need:

$$Y_i = \gamma_0 + \gamma_1 \text{Black}_i + \gamma_2 D_i^* + u_i, \tag{1}$$

with $\mathbb{E}[u_i \mid \text{Black}_i, D_i^*] = 0$. Under (1), the conditional expectation of Y_i given race and special education need is

$$\mathbb{E}[Y_i \mid \text{Black}_i = b, D_i^* = d] = \gamma_0 + \gamma_1 b + \gamma_2 d, \quad b \in \{0, 1\}.$$

It follows immediately that for any d,

$$\mathbb{E}[Y_i \mid \text{Black}_i = 1, D_i^* = d] - \mathbb{E}[Y_i \mid \text{Black}_i = 0, D_i^* = d] = \gamma_1.$$

Hence the oracle regression identifies the estimand:

$$\tau^*(d) = \gamma_1.$$

In other words, if special education need D_i^* were observed and the linear specification (1) held, the coefficient on Black_i would summarize the racial difference in IEP receipt among students with the same special education need. This oracle model directly operationalizes the question at the heart of disproportionality debates: whether classification practices differ by race when special education need (as IDEA defines it) is held constant.

2.2 Canonical model with achievement only

Because D_i^* is unobserved, empirical studies replace it with observable covariates. Let us first consider the case where D_i^* is proxied only by prior achievement. Let A_i denote prior

[1] We explicitly note that this is a descriptive estimand (Lundberg et al. 2021): the motivating question concerns a comparison of conditional expectations (whether Black and White students with the same special education need receive IEPs at different rates) rather than the estimation of a causal effect. See Online Appendix A for discussion of the relation between this descriptive specification and causal studies.

achievement (e.g., a standardized test score). A common starting point is a regression of the form

$$Y_i = \beta_0 + \beta_1 \text{Black}_i + \beta_2 A_i + \varepsilon_i. \tag{2}$$

The intended interpretation is that A_i serves as a proxy for D_i^*, so that conditioning on A_i approximates conditioning on special education need.

To see why (2) does not generally recover τ^*, it is useful to introduce an explicit opportunity to learn variable O_i that summarizes instructional quality, school resources, and other aspects of educational opportunity (all of which are exclusionary factors under IDEA). Suppose that achievement is generated by

$$A_i = \delta_0 + \delta_1 D_i^* + \delta_2 O_i + e_i, \tag{3}$$

and that IEP receipt depends on both special education need and opportunity:

$$Y_i = \theta_0 + \theta_1 \text{Black}_i + \theta_2 D_i^* + \theta_3 O_i + v_i. \tag{4}$$

If O_i is correlated with Black_i (for example because Black students have systematically fewer opportunities to learn), then in the regression (2) the coefficient β_1 mixes any direct race component in (4) with an opportunity-driven term involving θ_3 and $\text{Cov}(\text{Black}_i, O_i \mid A_i)$. In particular, unless $\theta_3 = 0$ and O_i is unrelated to race conditional on A_i, β_1 will not equal the oracle estimand τ^* defined below.

Intuitively, A_i in (2) contains both the special education need component $\delta_1 D_i^*$ and an opportunity component $\delta_2 O_i$. Conditioning on A_i therefore conditions on a mixture of special education need and unequal opportunity. The resulting regression coefficient on Black_i does not isolate how the classification process treats Black and White students with the same D_i^*.

2.3 Canonical model with achievement and SES

To address this concern, the canonical regression in the special education disproportionality literature typically augments (2) with measures of socioeconomic status. Let SES_i summarize socioeconomic background. The complete canonical regression then becomes:

$$Y_i = \alpha_0 + \alpha_1 \text{Black}_i + \alpha_2 \text{SES}_i + \alpha_3 A_i + \varepsilon_i, \tag{5}$$

To examine this model, let us complement structural equations (3) and (4) above to introduce SES in the data-generating process (DGP). One might imagine that opportunity O_i depends on SES and race via

$$O_i = \rho_0 + \rho_1 \text{SES}_i + \rho_2 \text{Black}_i + \omega_i, \tag{6}$$

and that SES itself is racially stratified,

$$\text{SES}_i = \sigma_0 + \sigma_1 \text{Black}_i + \eta_i. \tag{7}$$

Together, the structural equations (3), (4), (6), and (7) constitute the structural equation model (SEM)[2] in Figure 1.

Note that this SEM considers the simplified case where special education need D_i^* is assumed to be fully exogenous, having no parents in the structural system. We introduce this assumption to keep the main derivations focused on the over-controlling mechanism created by conditioning on achievement. Section 2.6 helps understand how this assumption affects our results and Online Appendix B formally relaxes this assumption by allowing D_i^* to covary with SES_i, showing that differences between the canonical and oracle models remain present throughout.

Our framework, by explicitly defining the target estimand in terms of IDEA's definition of special education need, clarifies the roles of the achievement and SES covariates in the complete canonical regression (5). While prior studies often motivate these covariates by appealing to the need to account for confounders, we clarify that achievement and SES play different roles: prior achievement provides a measure of learning difficulties while SES indicators hold constant those social disadvantages considered exclusionary factors by IDEA (such as poverty-related impediments to learning), effectively attempting to isolate

[2]We intentionally refer to it as a structural equation model (SEM) and not a directed acyclic graph (DAG) to emphasize these represent conditional associations and not causal relationships.

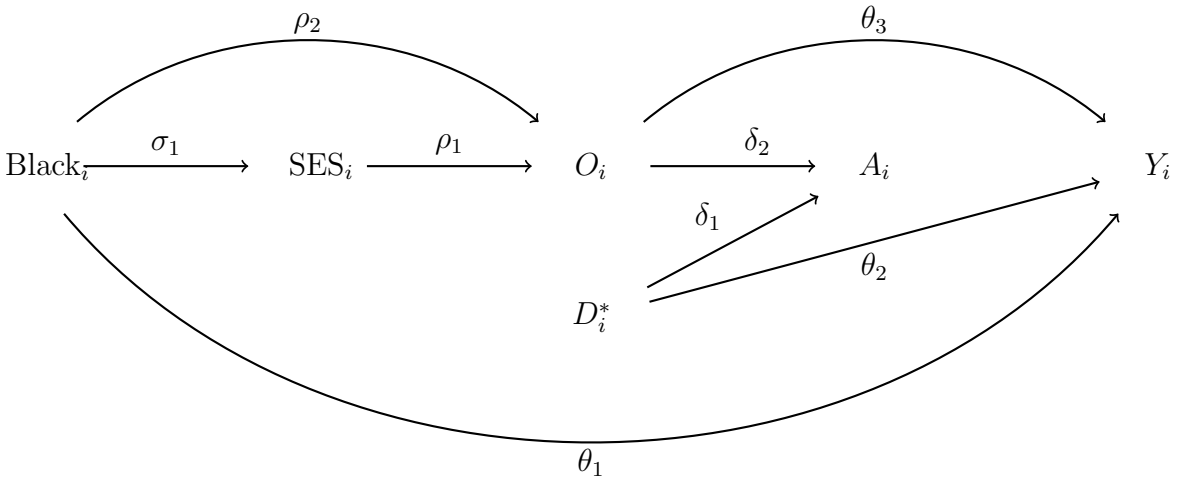

Figure 1: Data-generating process proposed by our structural equations (3), (4), (6), and (7).

the portion of achievement that reflects special education need rather than these exclusionary factors. More formally, including SES_i in (5) can help soak up the exclusionary factors that operate through SES_i and O_i, so that the remaining variation in A_i is a cleaner proxy for D_i^*. The SEM helps visualize this argument: adjusting for SES helps account for the $\text{Black}_i \to \text{SES}_i \to O_i \to A_i$ pathway. Under this rationale, achievement together with selected SES controls might provide a more reliable proxy for D_i^* relative to controlling for achievement alone.

However, under this data-generating process, this model also fails to generally recover τ^*, as the SES control leaves in A_i a residual combination of D_i^* and any remaining opportunity differences such as those flowing through the direct $\text{Black}_i \to O_i$ arrow. Conditioning on A_i opens the collider and induces the canonical model to identify a race–D_i^* association even when D_i^* is independent of race. Below, we provide a formal discussion of this issue.

2.4 Relating the canonical model to the oracle model

We can formalize the connection between the oracle regression (1) and the canonical regression (5) by writing special education need as its population linear projection on achieve-

ment and SES plus a residual. Let the linear projection of D_i^* onto (A_i, SES_i) be

$$D_i^* = \lambda_0 + \lambda_A A_i + \lambda_S \text{SES}_i + r_i, \tag{8}$$

where $(\lambda_0, \lambda_A, \lambda_S)$ minimize $\mathbb{E}[(D_i^* - \lambda_0 - \lambda_A A_i - \lambda_S \text{SES}_i)^2]$. By construction, the projection error satisfies

$$\mathbb{E}[r_i \mid A_i, \text{SES}_i] = 0.$$

Equation (8) does not assume that special education need is literally determined by achievement and SES. It is simply the population linear projection of D_i^* on these covariates, which exists for any joint distribution of $(D_i^*, A_i, \text{SES}_i)$.

Substituting (8) into the oracle model (1) yields

$$\begin{aligned} Y_i &= \gamma_0 + \gamma_1 \text{Black}_i + \gamma_2 D_i^* + u_i \\ &= \gamma_0 + \gamma_1 \text{Black}_i + \gamma_2 \left(\lambda_0 + \lambda_A A_i + \lambda_S \text{SES}_i + r_i\right) + u_i \\ &= (\gamma_0 + \gamma_2 \lambda_0) + \gamma_1 \text{Black}_i + \gamma_2 \lambda_S \text{SES}_i + \gamma_2 \lambda_A A_i + (\gamma_2 r_i + u_i). \end{aligned} \tag{9}$$

Comparing (9) to the canonical regression (5), the canonical coefficients match the oracle-implied coefficients,

$$\alpha_1 = \gamma_1, \quad \alpha_2 = \gamma_2 \lambda_S, \quad \alpha_3 = \gamma_2 \lambda_A,$$

whenever the composite term $(\gamma_2 r_i + u_i)$ behaves like a regression error with respect to the canonical regressors. A sufficient (and standard) condition is mean independence:

$$\mathbb{E}[\gamma_2 r_i + u_i \mid \text{Black}_i, \text{SES}_i, A_i] = 0.$$

Under this condition, the canonical regression error

$$\varepsilon_i \equiv \gamma_2 r_i + u_i$$

is mean independent of $(\text{Black}_i, \text{SES}_i, A_i)$, so (5) recovers the oracle-implied coefficients.

This restriction places requirements on both u_i and r_i. In particular, it requires that the projection residual r_i carry no systematic information about race once achievement and SES

are held fixed:

$$\mathbb{E}[r_i \mid \text{Black}_i, \text{SES}_i, A_i] = 0.$$

This is a very strong requirement. *Put differently, it requires that any systematic racial differences in special education need not captured by SES$_i$ be fully mediated by achievement A_i once SES is held fixed.* If achievement reflects racial differences that cannot be traced to special education need, then the canonical regression effectively controls for more racial differences than researchers may intend.

2.5 Residual achievement gaps as disability differences

To see the implications of the canonical restriction more clearly, recall the linear projection of special education need on achievement and SES:

$$D_i^* = \lambda_0 + \lambda_A A_i + \lambda_S \text{SES}_i + r_i, \tag{10}$$

where r_i is the projection error and, by construction,

$$\mathbb{E}[r_i \mid A_i, \text{SES}_i] = 0.$$

We now take expectations of (10) conditional on race and SES. For any $b \in \{0, 1\}$ and SES level s,

$$\begin{aligned}\mathbb{E}[D_i^* \mid \text{Black}_i = b, \text{SES}_i = s] = \lambda_0 + \lambda_A \, \mathbb{E}[A_i \mid \text{Black}_i = b, \text{SES}_i = s] \\ + \lambda_S s + \mathbb{E}[r_i \mid \text{Black}_i = b, \text{SES}_i = s].\end{aligned}$$

The first three terms depend on constants, SES, and the conditional mean of achievement within SES strata. The final term, $\mathbb{E}[r_i \mid \text{Black}_i = b, \text{SES}_i = s]$, captures any remaining systematic racial difference in special education need at SES level s that is not explained by achievement.

Subtracting the non-Black expression from the Black expression yields the conditional

racial difference in special education need at SES level s:

$$\begin{aligned}
&\mathbb{E}[D_i^* \mid \text{Black}_i = 1, \text{SES}_i = s] - \mathbb{E}[D_i^* \mid \text{Black}_i = 0, \text{SES}_i = s] \\
&= \lambda_A \Big(\mathbb{E}[A_i \mid \text{Black}_i = 1, \text{SES}_i = s] - \mathbb{E}[A_i \mid \text{Black}_i = 0, \text{SES}_i = s] \Big) \\
&\quad + \Big(\mathbb{E}[r_i \mid \text{Black}_i = 1, \text{SES}_i = s] - \mathbb{E}[r_i \mid \text{Black}_i = 0, \text{SES}_i = s] \Big).
\end{aligned}$$

Thus, within SES strata, racial differences in expected need decompose into an "achievement channel" scaled by λ_A and a "residual-need channel" governed by differences in the projection error r_i.

The key identification restriction in Section 2.4 requires that the canonical regression error ε_i in (5) be mean independent of the regressors. Under the mapping implied by (9), this corresponds to

$$\mathbb{E}[\gamma_2 r_i + u_i \mid \text{Black}_i, \text{SES}_i, A_i] = 0,$$

which in particular implies

$$\mathbb{E}[r_i \mid \text{Black}_i, \text{SES}_i, A_i] = 0.$$

By the law of iterated expectations, this stronger condition implies the weaker within-SES restriction used in the decomposition above:

$$\begin{aligned}
\mathbb{E}[r_i \mid \text{Black}_i = b, \text{SES}_i = s] &= \mathbb{E}[\mathbb{E}[r_i \mid \text{Black}_i = b, \text{SES}_i = s, A_i] \,|\, \text{Black}_i = b, \text{SES}_i = s] \\
&= 0 \quad \text{for all } (b, s).
\end{aligned}$$

Therefore, under the canonical identification restriction,

$$\begin{aligned}
&\mathbb{E}[D_i^* \mid \text{Black}_i = 1, \text{SES}_i = s] - \mathbb{E}[D_i^* \mid \text{Black}_i = 0, \text{SES}_i = s] \\
=&\lambda_A \Big(\mathbb{E}[A_i \mid \text{Black}_i = 1, \text{SES}_i = s] - \mathbb{E}[A_i \mid \text{Black}_i = 0, \text{SES}_i = s] \Big).
\end{aligned}$$

In words, once one imposes the canonical restriction that there is no systematic racial variation in the residual-need component r_i after conditioning on achievement and SES, racial differences in special education need within SES strata become fully explained by

residual racial achievement gaps within those same SES strata. *Equivalently, after conditioning on SES, any remaining Black–White achievement gap is treated as arising entirely from differences in special education need, rather than from other unmeasured differences in educational opportunity, instruction, or other exclusionary factors.*

For researchers who rely on the canonical regression, the conclusion that Black students are under-identified is therefore obtained only after the model has, by assumption, mapped residual achievement gaps into disability differences. Many applied researchers may accept the under-identification empirical statement while being uncomfortable with the strength of the assumption that delivers this mapping. Whether one revises conclusions about the empirical statement or about the maintained assumption is an interpretive choice that the canonical framework forces the reader to confront.

2.6 Dual roles of SES covariates

Section 2.4 argued that the SES covariate in the canonical model helps absorb exclusionary factors introduced via the achievement control, allowing residual achievement gaps to serve as a better proxy for disability. However, attentive readers will note that, as the projection (10) reveals, the role of SES in the canonical model can be more nuanced: SES_i is itself informative about D_i^*.

In the projection in (10),

$$D_i^* = \lambda_0 + \lambda_A A_i + \lambda_S \text{SES}_i + r_i,$$

$\lambda_S \neq 0$ means that, in population, SES predicts disability-related functioning even after achievement is held fixed. Substantively, this is plausible even under our definition of D_i^* as "net of exclusionary factors" because real-world measures of SES are proxies for a bundle of upstream conditions that can shape disability-related functioning, measurement, and access to evaluation. Once $\lambda_S \neq 0$, conditioning on SES_i does not merely remove exclusionary factors from A_i, it also conditions away a component of the variation in D_i^* itself.

This makes the interpretation proposed in Section 2.5 — that the canonical model would

recover the oracle estimand if one was willing to interpret the residual Black–White achievement gap after conditioning on SES as a disability gap — even more stringent. SES loading onto D_i^* implies that some of the disability gap has already been partialled out by the SES controls. Put differently, if SES both (i) soaks up opportunity-related determinants of achievement and (ii) captures a component of special education need, then the remaining achievement gap net of SES must shoulder an even stronger burden. It must be interpreted as disability-related difference despite the fact that the model has already used SES to remove some disability-related variance.

Note that our maintained assumption that D_i^* is exogenous limits this second role of SES in the structural system. In Figure 1, there is no direct path from SES_i to D_i^*. This simplification keeps the main text focused on the over-controlling problem that arises when researchers condition on achievement as a proxy for need, even though achievement also reflects unequal opportunity. Online Appendix Online Appendix B relaxes this simplification by allowing special education need to covary with socioeconomic background. Doing so makes explicit the second role highlighted above, namely that SES can be informative about D_i^* in addition to absorbing opportunity-related determinants of achievement, and it shows how the canonical and oracle estimands compare once SES plays both roles simultaneously.

Overall, we summarize these results as follows:

- *Canonical estimates can understate IDEA-aligned disparities by construction.* When $\lambda_S \neq 0$, the canonical regression can reduce the apparent role of special education need by conditioning on SES and then requires the remaining achievement gap to stand in for disability. This combination can mechanically attenuate the race coefficient in settings where many readers would already find it implausible to attribute the residual achievement gap to disability alone.
- *Residual achievement gaps net of SES are not interpretable as "pure" disability differences.* Even if one were willing to treat all remaining achievement differences as reflecting D_i^*, this would still describe only the part of D_i^* that is not already captured

by SES, rather than the full racial difference in special education need.

These results — and the central role of the residual Black-White achievement gap net of SES they reveal — inform our proposed diagnostics, which we describe and formalize below. A simple and informative summary is the fraction of the Black–White achievement gap that persists after conditioning on SES. When this SES-adjusted achievement gap remains large, interpreting the canonical race coefficient as an under-identification effect requires especially strong assumptions about what the residual achievement gap represents.

3 Bounding the bias from using achievement and SES as proxies for special education need

The previous section showed that the canonical regression recovers the oracle parameter $\gamma_1 = \tau^*$ only under strong conditions on how well achievement and SES proxy for D_i^* and on how their residual components relate to race. Here we use the same setup to characterize the bias in the canonical race coefficient and to describe empirical quantities that can be used to bound its magnitude. We first derive a decomposition and bound, then use polar cases for intuition, and finally introduce empirical summaries that position applications between these extremes and guide the calibration and sensitivity analysis that follows.

3.1 General expression for the bias: decomposing the race coefficient in the canonical model

Let $\tilde{\text{Black}}_i$ denote the residual from regressing Black_i on (A_i, SES_i), so that the coefficient on race in (5) can be written as

$$\alpha_1 = \frac{\text{Cov}(\tilde{\text{Black}}_i, Y_i)}{\text{Var}(\tilde{\text{Black}}_i)}.$$

Substituting (9) into this expression yields

$$\alpha_1 = \gamma_1 + \frac{\text{Cov}(\tilde{\text{Black}}_i, \gamma_2 r_i + u_i)}{\text{Var}(\tilde{\text{Black}}_i)}.$$

Thus the difference between the canonical and oracle race coefficients is

$$\alpha_1 - \gamma_1 = \frac{\mathrm{Cov}(\tilde{\mathrm{Black}}_i, \gamma_2 r_i + u_i)}{\mathrm{Var}(\tilde{\mathrm{Black}}_i)}.$$

Under the condition in the previous subsection that $\gamma_2 r_i + u_i$ is mean independent of $(\mathrm{Black}_i, A_i, \mathrm{SES}_i)$, this bias term is zero and $\alpha_1 = \gamma_1 = \tau^*$. When, as we argue is substantively plausible, the residual component r_i remains correlated with race even after conditioning on achievement and SES, the bias term will generally be nonzero. In that case the canonical race coefficient cannot be interpreted as an estimate of τ^* and may substantially understate the contribution of the classification process to racial disproportionality.

Recall the linear projection of special education need on achievement and SES,

$$D_i^* = \lambda_0 + \lambda_A A_i + \lambda_S \mathrm{SES}_i + r_i, \tag{11}$$

with $\mathbb{E}[r_i \mid A_i, \mathrm{SES}_i] = 0$. Substituting (11) into the oracle model yields

$$Y_i = (\gamma_0 + \gamma_2\lambda_0) + \gamma_1 \mathrm{Black}_i + \gamma_2\lambda_A A_i + \gamma_2\lambda_S \mathrm{SES}_i + (\gamma_2 r_i + u_i). \tag{12}$$

Let $\tilde{\mathrm{Black}}_i$ denote the residual from regressing Black_i on (A_i, SES_i), so that by the Frisch–Waugh–Lovell theorem the race coefficient in the canonical regression can be written as

$$\alpha_1 = \frac{\mathrm{Cov}(\tilde{\mathrm{Black}}_i, Y_i)}{\mathrm{Var}(\tilde{\mathrm{Black}}_i)}.$$

Substituting (12) into this expression gives

$$\begin{aligned}\alpha_1 &= \gamma_1 + \frac{\mathrm{Cov}(\tilde{\mathrm{Black}}_i, \gamma_2 r_i + u_i)}{\mathrm{Var}(\tilde{\mathrm{Black}}_i)} \\ &= \gamma_1 + \gamma_2 \frac{\mathrm{Cov}(\tilde{\mathrm{Black}}_i, r_i)}{\mathrm{Var}(\tilde{\mathrm{Black}}_i)} + \frac{\mathrm{Cov}(\tilde{\mathrm{Black}}_i, u_i)}{\mathrm{Var}(\tilde{\mathrm{Black}}_i)}.\end{aligned} \tag{13}$$

The difference between the canonical race coefficient and the oracle race coefficient is therefore

$$\alpha_1 - \gamma_1 = \gamma_2 \frac{\mathrm{Cov}(\tilde{\mathrm{Black}}_i, r_i)}{\mathrm{Var}(\tilde{\mathrm{Black}}_i)} + \frac{\mathrm{Cov}(\tilde{\mathrm{Black}}_i, u_i)}{\mathrm{Var}(\tilde{\mathrm{Black}}_i)}. \tag{14}$$

The second term, $\mathrm{Cov}(\tilde{\mathrm{Black}}_i, u_i)/\mathrm{Var}(\tilde{\mathrm{Black}}_i)$, captures any remaining association between the oracle disturbance u_i and the component of race that is orthogonal to (A_i, SES_i).

This component can arise if the oracle mean-independence condition fails (i.e., $\mathbb{E}[u_i \mid \text{Black}_i, D_i^*] \neq 0$). More generally, because $\tilde{\text{Black}}_i$ is defined by residualizing Black_i on (A_i, SES_i) rather than on D_i^*, this term also bundles any correlation between u_i and the proxy covariates used to construct $\tilde{\text{Black}}_i$. In what follows, we impose the standard oracle orthogonality condition $\mathbb{E}[u_i \mid \text{Black}_i, D_i^*] = 0$ and abstract from this component in order to focus on the first term, which is the additional bias that arises specifically from using achievement and SES as imperfect proxies for special education need. This proxy-bias term depends on two quantities. The variance of the projection error r_i summarizes how much of D_i^* is not captured by (A_i, SES_i). The covariance between r_i and the residualized race indicator $\tilde{\text{Black}}_i$ summarizes how strongly that unmeasured component of special education need remains racially patterned after conditioning on achievement and SES.

A simple bound follows from Cauchy–Schwarz (i.e., a covariance cannot be larger in magnitude than the product of the standard deviations) when we temporarily set $\text{Cov}(\tilde{\text{Black}}_i, u_i) = 0$ to isolate the contribution of r_i:

$$\begin{aligned}
\left|\alpha_1 - \gamma_1\right| &= \left|\gamma_2 \frac{\text{Cov}(\tilde{\text{Black}}_i, r_i)}{\text{Var}(\tilde{\text{Black}}_i)}\right| \\
&\leq |\gamma_2| \frac{\sqrt{\text{Var}(\tilde{\text{Black}}_i)}\sqrt{\text{Var}(r_i)}}{\text{Var}(\tilde{\text{Black}}_i)} \\
&= |\gamma_2| \frac{\sqrt{\text{Var}(r_i)}}{\sqrt{\text{Var}(\tilde{\text{Black}}_i)}}.
\end{aligned} \tag{15}$$

Define

$$R^2_{D|A,\text{SES}} \equiv \frac{\text{Var}\left(\lambda_0 + \lambda_A A_i + \lambda_S \text{SES}_i\right)}{\text{Var}(D_i^*)},$$

the proportion of variance in special education need explained by achievement and SES. Then

$$\text{Var}(r_i) = \text{Var}(D_i^*)\left(1 - R^2_{D|A,\text{SES}}\right),$$

and the bound (15) can be rewritten as

$$\left|\alpha_1 - \gamma_1\right| \leq |\gamma_2| \frac{\sqrt{\text{Var}(D_i^*)}}{\sqrt{\text{Var}(\tilde{\text{Black}}_i)}} \sqrt{1 - R^2_{D|A,\text{SES}}}. \tag{16}$$

This expression shows that the worst-case bias from using achievement and SES as proxies for special education need shrinks as (A_i, SES_i) become more informative about D_i^* (higher $R^2_{D|A,\mathrm{SES}}$) and as there is more residual variation in race after conditioning on (A_i, SES_i) (larger $\mathrm{Var}(\widetilde{\mathrm{Black}_i})$).

3.2 Two polar cases

The general bound in (16) does not require any assumption about the relative roles of SES and achievement in explaining racial disparities in achievement. It is useful, however, to consider two polar cases that bracket the behavior of the canonical model.

Polar case 1: SES fully explains the Black–White achievement gap. First consider the case in which SES fully explains the Black–White achievement gap in the sense that the conditional mean of achievement depends on SES but not on race:

$$\mathbb{E}[A_i \mid \mathrm{Black}_i = 1, \mathrm{SES}_i = s] = \mathbb{E}[A_i \mid \mathrm{Black}_i = 0, \mathrm{SES}_i = s] \quad \text{for all } s.$$

Equivalently, there exists some function $m(\cdot)$ such that

$$\mathbb{E}[A_i \mid \mathrm{SES}_i, \mathrm{Black}_i] = m(\mathrm{SES}_i) = \mathbb{E}[A_i \mid \mathrm{SES}_i].$$

Under a linear approximation we can write

$$A_i = \psi_0 + \psi_1 \mathrm{SES}_i + \eta_i, \quad \mathbb{E}[\eta_i \mid \mathrm{SES}_i, \mathrm{Black}_i] = 0,$$

so that achievement residuals carry no information about race once SES is held fixed. In this case, the auxiliary regression of Black_i on (SES_i, A_i) assigns zero weight to A_i in population, and residualizing race on (SES_i, A_i) is equivalent to residualizing race on SES_i alone. The race coefficient in the canonical regression therefore satisfies, in population,

$$\alpha_1^{(\mathrm{race+SES+Ach})} = \alpha_1^{(\mathrm{race+SES})},$$

so achievement has a purely prognostic role for predicting Y_i and does not affect the estimated racial disparity. In this limit the problematic feature of the canonical model, if any, is not

the inclusion of achievement but the use of SES to absorb racialized pathways that may themselves be part of the disparity in access to special education.

To understand when the SES regression itself converges to the oracle estimand, let D_i^* denote underlying disability need and let the outcome equation satisfy

$$Y_i = \gamma_0 + \gamma_1 \text{Black}_i + \gamma_2 D_i^* + u_i,$$

with $\mathbb{E}[u_i \mid D_i^*, \text{Black}_i, \text{SES}_i] = 0$. Define the linear projection of D_i^* on SES alone,

$$P[D_i^* \mid \text{SES}_i] = a_0 + a_1 \text{SES}_i,$$

and write

$$D_i^* = P[D_i^* \mid \text{SES}_i] + r_i^{(S)},$$

so $r_i^{(S)}$ is the component of disability need not explained by SES.[3] Substituting into the outcome equation yields

$$Y_i = (\gamma_0 + \gamma_2 a_0) + \gamma_1 \text{Black}_i + \gamma_2 a_1 \text{SES}_i + \gamma_2 r_i^{(S)} + u_i.$$

Regressing Y_i on $(\text{Black}_i, \text{SES}_i)$ therefore yields a race coefficient that can be decomposed into the oracle race coefficient γ_1 plus a component coming from the correlation between $(\gamma_2 r_i^{(S)} + u_i)$ and race after conditioning on SES. In particular, if $\widetilde{\text{Black}}_i^{(\text{SES})}$ denotes the residual from regressing Black_i on SES_i, the race coefficient in the SES-only regression can be written as

$$\alpha_1^{(\text{race+SES})} = \gamma_1 + \gamma_2 \frac{\text{Cov}\big(\widetilde{\text{Black}}_i^{(\text{SES})}, r_i^{(S)}\big)}{\text{Var}\big(\widetilde{\text{Black}}_i^{(\text{SES})}\big)} + \frac{\text{Cov}\big(\widetilde{\text{Black}}_i^{(\text{SES})}, u_i\big)}{\text{Var}\big(\widetilde{\text{Black}}_i^{(\text{SES})}\big)}. \tag{17}$$

If the assignment rule is race neutral given disability (so $\gamma_1 = 0$ and u_i is mean independent of race conditional on D_i^*), the SES regression converges to the oracle effect if and only if

$$\mathbb{E}[D_i^* \mid \text{SES}_i, \text{Black}_i] = \mathbb{E}[D_i^* \mid \text{SES}_i] \quad \text{for all SES}_i,$$

[3] The notation $r_i^{(S)}$ is used here to distinguish the SES-only projection residual from the projection residual r_i defined elsewhere when projecting D_i^* on (A_i, SES_i).

so that $r_i^{(S)}$ is mean independent of race conditional on SES and the covariance term in (17) is zero. Under this strong condition, controlling for SES recovers the oracle estimand, and because achievement is redundant in this polar case, the canonical model with (Black, SES, A) also converges to the oracle.

In many applications, however, SES has heterogeneous effects on disability by race. For example, neighborhood SES may generate different mixtures of opportunity and disabling exposures for Black and White students, so that

$$\mathbb{E}[D_i^* \mid \text{SES}_i, \text{Black}_i = 1] \neq \mathbb{E}[D_i^* \mid \text{SES}_i, \text{Black}_i = 0] \quad \text{for some SES}_i.$$

In this case the residual $r_i^{(S)} = D_i^* - P[D_i^* \mid \text{SES}_i]$ necessarily remains correlated with race given SES, the covariance term in (17) generally does not vanish, and a regression that controls only for SES need not approximate the oracle comparison even if SES fully closes the Black–White achievement gap. Gap closure tells us that $\mathbb{E}[A_i \mid \text{SES}_i, \text{Black}_i]$ is equalized across race. It does not guarantee that $\mathbb{E}[D_i^* \mid \text{SES}_i, \text{Black}_i]$ is equalized. In such environments, a near zero race coefficient in the SES (or SES+achievement) regression can mask differences in IEP receipt between Black and White students with the same level of underlying disability need.

Polar case 2: SES explains none of the Black–White achievement gap. At the other extreme, suppose SES explains none of the Black–White achievement gap in achievement, so that conditioning on SES does not narrow the racial achievement gap. In the notation used elsewhere, this corresponds to $\kappa_A = 1$.[4]

To isolate the intuition in this polar case, it is useful to momentarily consider a stylized specification that omits SES and regresses IEP receipt on race and achievement alone,

$$Y_i = \beta_0 + \beta_1 \text{Black}_i + \beta_2 A_i + \varepsilon_i. \tag{18}$$

[4]Here $\kappa_A = 1$ is a statement about achievement gaps, not about whether SES predicts Y_i conditional on race and achievement. In particular, SES could remain an important predictor of IEP receipt even if it does not reduce the Black–White achievement gap in A_i.

The goal of this simplification is purely expositional: it makes transparent what happens when achievement must do essentially all of the proxying for latent need because SES provides no "equalizing" reduction in racial achievement gaps.

Let

$$D_i^* = \lambda_0^{(A)} + \lambda_A^{(A)} A_i + r_i^{(A)}, \quad \mathbb{E}[r_i^{(A)} \mid A_i] = 0,$$

be the linear projection of special education need on achievement. Substituting this into the oracle model yields

$$Y_i = (\gamma_0 + \gamma_2 \lambda_0^{(A)}) + \gamma_1 \text{Black}_i + \gamma_2 \lambda_A^{(A)} A_i + (\gamma_2 r_i^{(A)} + u_i).$$

If $\tilde{\text{Black}}_i^{(A)}$ denotes the residual from regressing Black_i on A_i alone, then the race coefficient in (18) satisfies

$$\beta_1 = \gamma_1 + \gamma_2 \frac{\text{Cov}(\tilde{\text{Black}}_i^{(A)}, r_i^{(A)})}{\text{Var}(\tilde{\text{Black}}_i^{(A)})} + \frac{\text{Cov}(\tilde{\text{Black}}_i^{(A)}, u_i)}{\text{Var}(\tilde{\text{Black}}_i^{(A)})}.$$

In this second polar case, achievement bears the entire burden of approximating D_i^* and SES does no equalizing work on racial achievement differences. The regression therefore interprets the full Black–White achievement gap as relevant for proxying special education need, subject only to what achievement fails to explain about D_i^* in $r_i^{(A)}$. Bias can therefore be large whenever opportunity to learn and other exclusionary factors contribute substantially to achievement gaps.

Empirical settings will lie between these extremes. SES will typically explain a nontrivial but incomplete share of the Black–White achievement gap, and achievement and SES will explain a nontrivial but incomplete share of the variance in special education need. The canonical model can then be viewed as implicitly reassigning a fraction of the residual achievement gap after SES to differences in D_i^* while discarding the residual variation in D_i^* that is not captured by (A_i, SES_i). The first polar case shows that convergence to the oracle requires more than closing the achievement gap with SES; it also requires that disability need be equalized across race conditional on SES. The second polar case shows that when

SES plays no equalizing role in achievement, the burden of proxying for need falls entirely on achievement, and resulting bias can be substantial. In that polar case we analyzed the A-only regression purely for expositional clarity, and the qualitative conclusion applies to the full canonical estimator α_1 because adding SES does not reduce the racial achievement gap and therefore does not materially relieve the proxy burden placed on A_i.

These polar cases clarify the extremes. To position a given application between them, we next introduce two empirical summaries, one capturing how well achievement and SES proxy disability related functioning, and the other capturing how much of the achievement gap remains after conditioning on SES.

3.3 Empirical ingredients for bounding the bias

Expression (16) shows that two empirical features of the data are central to bounding the bias from using test scores as proxies for special education need.

First, one needs to know how informative achievement and SES are about special education need. This is summarized by $R^2_{D|A,\mathrm{SES}}$, the proportion of variance in D_i^* explained by (A_i, SES_i). In practice D_i^* is unobserved, but in rich longitudinal surveys — such as the ECLS ones — can construct an empirical proxy $\hat{D}_i$ from multiple indicators that are closer to IDEA style special education need than achievement alone. Examples include teacher ratings, behavior scales, early developmental assessments, and clinical diagnoses. Regressing such a proxy on achievement and SES,

$$\hat{D}_i = \lambda_0 + \lambda_A A_i + \lambda_S \mathrm{SES}_i + \text{residual},$$

provides an empirical estimate of how much of the variation in disability related functioning is captured by the canonical covariates. A low R^2 implies a large $\mathrm{Var}(r_i)$ and therefore a wider potential gap between α_1 and γ_1.

Second, one needs to know how much of the Black–White achievement gap remains after conditioning on SES, which we summarize by κ_A in (19). To obtain this parameter, define

the unconditional gap

$$\Delta_A^{\text{tot}} \equiv \mathbb{E}[A_i \mid \text{Black}_i = 1] - \mathbb{E}[A_i \mid \text{Black}_i = 0],$$

and the SES adjusted gap as the coefficient on Black_i in the regression of achievement on race and SES,

$$A_i = \phi_0 + \phi_1 \text{Black}_i + \phi_2 \text{SES}_i + \zeta_i,$$

so that ϕ_1 is the residual Black–White achievement difference after conditioning on SES. The corresponding summary statistic is

$$\kappa_A \equiv \frac{\phi_1}{\Delta_A^{\text{tot}}}. \tag{19}$$

summarizes the fraction of the unconditional achievement gap that persists after SES is held fixed. In the first polar case above, $\kappa_A = 0$ and SES fully explains the gap. In the second polar case, $\kappa_A = 1$ and SES explains none of the gap. For intermediate values of κ_A, the canonical model implicitly treats a κ_A fraction of the achievement disparity as a disparity in special education need at a given SES level, while assigning the remaining $1 - \kappa_A$ fraction to SES.

Combining information about $R^2_{D|A,\text{SES}}$ and κ_A therefore allows one to position a given empirical application along the continuum between the two polar cases. A high $R^2_{D|A,\text{SES}}$ and a small κ_A would support the view that achievement and SES provide reasonably good proxies for special education need and that little of the residual achievement gap is being recoded as disability. A low $R^2_{D|A,\text{SES}}$ and a large κ_A would instead imply that the canonical regression discards a substantial amount of disability related variation and attributes most of the residual Black–White achievement disparity to differences in D_i^*, suggesting the potential for large included variable bias in estimates of racial disproportionality.

3.4 What SES adjustment can and cannot tell us about the IEP gap

The polar cases in the previous subsections describe limits in which SES either fully explains or does not explain the Black–White achievement gap. Most empirical studies of

racial disproportionality in special education sit between these extremes. We compiled a subset of studies that have used commonly available nationally representative datasets to estimate either achievement or IEP gaps. The general pattern from these studies as shown in Table 1 is that:

1. there is a sizable unconditional Black–White achievement gap, $\Delta_A^{\text{tot}} < 0$,
2. conditioning on an observed SES vector X reduces the gap but does not eliminate it, so that the residual gap after SES adjustment $\phi_1 = \kappa_A \Delta_A^{\text{tot}}$ satisfies $0 < \kappa_A < 1$, and
3. in parallel, the race coefficient in an SES-adjusted regression of IEP receipt,

$$Y_i = \alpha_0^{(race+SES)} + \alpha_1^{(race+SES)} \text{Black}_i + \alpha_2^{(race+SES)\prime} X_i + \varepsilon_i,$$

is often small in magnitude and frequently statistically indistinguishable from zero.

This configuration is sometimes read as evidence that observed SES differences account for most racial differences in special education identification, and that the remaining achievement gap reflects factors not captured by available SES measures. In the language of Section 3.3, $\kappa_A > 0$ indicates that the observed SES vector X is an incomplete proxy for the forces that shape achievement. Descriptively, a sizable residual achievement gap is therefore consistent with residual between-group differences within SES strata in determinants of achievement, including racialized exposure to resources and risks. The key question is what this implies for interpreting an SES-adjusted IEP gap, and what we would learn if we had richer SES covariates.

To fix ideas, consider the SES-adjusted regression of IEP receipt on race and X. Using the same Frisch–Waugh logic developed earlier, we can write

$$\alpha_1^{(race+SES)} = \tau^* + \gamma_2 \frac{\text{Cov}(\widetilde{\text{Black}}_i^{(X)}, r_i^{(X)})}{\text{Var}(\widetilde{\text{Black}}_i^{(X)})} + \frac{\text{Cov}(\widetilde{\text{Black}}_i^{(X)}, u_i)}{\text{Var}(\widetilde{\text{Black}}_i^{(X)})},$$

where $\widetilde{\text{Black}}_i^{(X)}$ is the residual from regressing Black_i on X_i, and $r_i^{(X)}$ is the residual from projecting D_i^* on X_i alone, $D_i^* = P[D_i^* \mid X_i] + r_i^{(X)}$. This expression clarifies what SES adjustment buys us. Even when the IEP assignment rule is race neutral given true need,

Table 1: Black–White gaps in achievement and special education identification

Study	Outcome	Model	Covar	Data	Grade	Gap	$\widehat{\kappa}_A$
Panel A. Achievement							
FL04	Math	U	0	ECLS-K:1998	K(F)	-0.638 SD	0.577
FL04	Math	C	1	ECLS-K:1998	K(F)	-0.368 SD	
FL04	Read	U	0	ECLS-K:1998	K(F)	-0.401 SD	0.334
FL04	Read	C	1	ECLS-K:1998	K(F)	-0.134 SD	
BL13[a]	Edu	U	0	CNLSY:CYA	K	-0.82 Yrs	0.573
BL13[a]	Edu	C	2	CNLSY:CYA	K	-0.47 Yrs	
BL13[a]	Edu	U	0	CNLSY:CYA	G5	-0.79 Yrs	0.709
BL13[a]	Edu	C	2	CNLSY:CYA	G5	-0.56 Yrs	
HM24[b]	Read	U	0	ECLS-K:2010	G1	-0.45 SD	0.356
HM24[b]	Read	C	8	ECLS-K:2010	G1	-0.16 SD	
HM24	Math	U	0	ECLS-K:2010	G5	-0.82 SD	0.610
HM24	Math	C	8	ECLS-K:2010	G5	-0.50 SD	
HM24	Read	U	0	ECLS-K:2010	G5	-0.64 SD	0.516
HM24	Read	C	8	ECLS-K:2010	G5	-0.33 SD	
Panel B. IEP identification							
EL21	Any	U	0	FL-BirthSch	K	-0.025 PP	
EL21	Any	C	3	FL-BirthSch	K	-0.034 PP	
EL21	Any	U	0	FL-BirthSch	G4	0.005 PP	
EL21	Any	C	3	FL-BirthSch	G4	-0.023 PP	
HFM10	Any	U	4	ECLS-K:1998	$\leq$G5	1.25 OR	
HFM10	Any	C	5	ECLS-K:1998	$\leq$G5	0.91 OR	
M17	Any	U	0	NAEP-R:2013	G4	1.03 OR	
M17	Any	C	6	NAEP-R:2013	G4	0.84 OR	

Notes. White is the reference group. Units: SD = test-score SD, Yrs = education-scaled years, PP = proportion points, OR = odds ratio. Mod: U = unconditional, C = conditional on household SES or early-childhood environment covariates (no prior achievement). NR = not reported. Study codes: FL04 = Fryer & Levitt (2004), BL13 = Bond & Lang (2013; NBER w19243), HM24 = Hu & Morgan (2024; Fordham), EL21 = Elder et al. (2021), HFM10 = Hibel et al. (2010), M17 = Morgan et al. (2017). Data codes: ECLS-K:1998 = ECLS-K 1998–99 cohort; ECLS-K:2010 = ECLS-K 2010–11 cohort; CNLSY:CYA = CNLSY Children & Young Adults; FL-BirthSch = Florida linked birth + school records; NAEP-R:2013 = NAEP Reading 2013 (public); ELS:2002 = Education Longitudinal Study 2002. X sets: 0 = none; 1 = ECLS SES composite; 2 = CNLSY early environment bundle; 3 = birth-record endowments; 4 = sex only; 5 = sex + family SES; 6 = sex + FRL + ELL; 7 = sex + family income + parents' education; 8 = SES+ bundle (parent education, occupational prestige, household income, household structure, and home-opportunity factors). $\widehat{\kappa}_A$ is computed within each outcome pair in Panel A as (conditional gap)/(unconditional gap), i.e., $\widehat{\kappa}_A \equiv \widehat{\Delta}_A^{\text{cond}}/\widehat{\Delta}_A^{\text{uncond}}$.

[a] BL13 tables report White minus Black. Signs are reversed here. Unconditional gaps from Table 8 (IV-adjusted). Conditional gaps from Table 10 (Math, early-environment column).

[b] HM24 conditional gaps computed as $\widehat{\Delta}_C = (1-r)\widehat{\Delta}_U$, where r is the "percent explained by SES+" from Figure 5. HM24 reports the grade 1 unconditional reading gap in text; the grade 1 conditional reading gap uses $r = 0.64$.

the SES-adjusted coefficient need not equal τ^* unless X is also an adequate proxy for the components of D_i^* relevant for identification and for any classification disturbance u_i.

Now consider enriching SES from X to an augmented vector $X^\star$ that adds family, neighborhood, and institutional measures. The empirical value of richer SES covariates can be organized around two diagnostic comparisons.

First, if adding $X^\star$ substantially reduces κ_A and also meaningfully shifts $\alpha_1^{(race+SES)}$, this suggests that the additional covariates are altering the residualized race component relevant for identification, either by improving proxying of the need component in D_i^* or by changing how classification noise loads onto race after conditioning on SES. In this case, richer SES contains information that is directly relevant for interpreting the IEP gap.

Second, if κ_A shrinks with $X^\star$ but $\alpha_1^{(race+SES)}$ changes little, this indicates that the additional covariates are informative about determinants of achievement that were previously omitted, while leaving largely unchanged the residual race association with the components of need or classification relevant for identification. In that case, richer SES clarifies why achievement gaps persist within SES strata, but it does not by itself resolve the interpretation of the SES-adjusted IEP coefficient.

Perhaps most practically, the preceding results mean that researchers using observational data to fit canonical regressions, where achievement may absorb residual opportunity, should report (i) the unconditional and SES-adjusted Black–White achievement gaps (or, equivalently, $\widehat{\kappa}_A$), (ii) the corresponding unconditional and SES-adjusted IEP gaps, and (iii) how both sets of quantities change as the SES vector is enriched, including sensitivity to alternative SES measurement choices.

This prior discussion assumed SES related to opportunity and special education need similarly across race. However, once SES can affect opportunity or disability differently by race, the mapping from "closing the achievement gap with SES" to "recovering the oracle IEP gap" becomes unreliable. To see this, allow disability need to vary with SES differently

by race,

$$D_i^* = h(X_i) + \theta \, \text{Black}_i \cdot s(X_i) + \xi_i,$$

where $s(X_i)$ picks out the relevant SES component. In this case the same observed SES value can correspond to different distributions of D_i^* by race, so conditioning on X_i equalizes measured SES but need not align underlying need.

Moreover, even if richer SES covariates reduce the residual achievement gap, this need not imply that need has been equalized within SES strata. Because achievement is jointly determined by opportunity and disability-related difficulty in learning, conditional parity in A_i can arise from different combinations of O_i and D_i^* across race within the same X_i stratum. In that environment, a small SES-adjusted IEP coefficient can coexist with substantively important differences in underlying need or in opportunity components relevant for identification.

In short, the dominant empirical configuration sits between the stylized limits. SES explains a substantial but incomplete share of the achievement gap, and the SES-adjusted IEP coefficient is often small. Our framework implies that this pattern is informative about how the available SES measures relate jointly to achievement and identification, but it is not a direct test of racial neutrality in special education identification. Richer SES covariates are most informative when changes in X produce coupled movements in κ_A and the SES-adjusted IEP coefficient. When they only reduce κ_A, the implications for the IEP gap remain theoretically ambiguous.

3.5 A plausibility check for interpreting residual achievement gaps as differences in special education need

The canonical interpretation of (5) requires that, after conditioning on SES, any remaining Black–White achievement difference is attributable to differences in underlying special education need D_i^* rather than to residual differences in opportunity to learn or other exclusionary factors. The statistic κ_A in (19) makes the strength of this requirement transparent.

Since the residual gap after SES adjustment is $\phi_1 = \kappa_A \Delta_A^{\text{tot}}$, the canonical model effectively treats an amount $\kappa_A \Delta_A^{\text{tot}}$ of the achievement gap as disability-related.

A simple calibration highlights how demanding this is. Let $Z_i \in \{0, 1\}$ denote an indicator for whether a student has special education need (for example, whether D_i^* exceeds an IDEA-relevant threshold). Let $p_b \equiv \Pr(Z_i = 1 \mid \text{Black}_i = b)$ denote disability prevalence by race, and let

$$\Delta_A^Z \equiv \mathbb{E}[A_i \mid Z_i = 1] - \mathbb{E}[A_i \mid Z_i = 0]$$

denote the mean achievement difference between students with and without special education need, measured in SD units so that $\Delta_A^Z < 0$ corresponds to lower achievement among students with need. In the most favorable case for the canonical interpretation, assume that (i) the only source of the Black–White achievement difference that remains after conditioning on SES is the difference in disability prevalence, and (ii) Δ_A^Z does not vary by race. Under these assumptions, the disability-attributable portion of the achievement gap satisfies

$$\kappa_A \Delta_A^{\text{tot}} \approx (p_1 - p_0)\, \Delta_A^Z. \tag{20}$$

Equation (20) provides two equivalent interpretations. Given an assumed prevalence gap $(p_1 - p_0)$, it yields the *implied* disability-related achievement gap

$$\Delta_A^Z \approx \frac{\kappa_A \Delta_A^{\text{tot}}}{p_1 - p_0}.$$

Given an assumed disability-related achievement gap Δ_A^Z, it yields the *implied* prevalence gap

$$p_1 - p_0 \approx \frac{\kappa_A \Delta_A^{\text{tot}}}{\Delta_A^Z}.$$

To illustrate magnitudes, take the unconditional Black–White achievement gap in grade 5 from Table 1 ($\widehat{\Delta}_A^{\text{tot}} \approx -0.79$ SD) and the residual gap after SES adjustment ($\widehat{\phi}_1 \approx -0.56$ SD), implying $\widehat{\kappa}_A \equiv \widehat{\phi}_1 / \widehat{\Delta}_A^{\text{tot}} \approx 0.71$. For a benchmark on the size of plausible prevalence differences, national IDEA Part B service rates differ by only a few percentage points across groups. For example, in 2022–23 the fraction served was 18.9% for Black students and

15.6% for White students, a gap of about 3.3 percentage points.[5] Using $(p_1 - p_0) = 0.033$ and $\widehat{\phi}_1 \approx -0.56$, (20) implies

$$\Delta_A^{D^*} \approx \frac{-0.56}{0.033} \approx -17 \text{ SD},$$

which is implausibly large. Alternatively, fixing a typical benchmark for the SWD–non-SWD achievement gap, meta-analytic summaries place the reading difference on the order of 1.2 SD (Gilmour et al. 2025). Plugging $\Delta_A^{D^*} \approx -1.2$ into (20) implies

$$p_1 - p_0 \approx \frac{-0.56}{-1.2} \approx 0.47,$$

so a disability-prevalence difference of roughly 47 percentage points, far larger than population service-rate contrasts (National Center for Education Statistics 2024a). Even if one relaxed the canonical interpretation so that only 25% of the total achievement gap were disability-related, (20) would still imply either $|\Delta_A^{D^*}| \approx 6.1$ SD when $(p_1 - p_0) = 0.033$, or a prevalence gap of about 17 percentage points when $\Delta_A^{D^*} \approx -1.2$, neither of which is remotely observed in empirical data.

This plausibility check reinforces the conceptual point in Sections 2.5 and 2.6. When SES covariates are allowed to predict D_i^* directly, conditioning on SES can remove some disability-related variation before achievement is used as a proxy for D_i^*. In that case, interpreting the remaining achievement gap after SES adjustment as disability-related requires that gap to be disability-related despite the fact that some disability-related variance has already been partialled out, which strengthens the required assumptions. Put simply, *in empirically typical regimes with $\kappa_A > 0$, interpreting the residual achievement gap net of SES as a disability gap demands either implausibly large disability-related achievement differences or implausibly large racial differences in disability prevalence.*

[5] These IDEA Part B service rates are not available by grade, nor are they a measure of latent special education need D_i^*. We use them only as an order-of-magnitude benchmark for how large race differences in disability prevalence can plausibly be in population data. If the true prevalence gap $(p_1 - p_0)$ in special education need is smaller than the service-rate gap, then (20) implies even larger required values of $|\Delta_A^{D^*}|$.

3.6 Sensitivity analysis and why e-value approaches miss the problem

The bounds in (16) are expressed in terms of $R^2_{D|A,\mathrm{SES}}$ and the covariance between the projection error r_i and residualized race. These quantities are defined with respect to the latent disability construct D_i^* and are therefore not directly observable. A natural reaction is to consider generic sensitivity analyses that do not require observing D_i^*, such as e-value methods (Morgan 2021) or Oster style coefficient stability (Oster 2019). Recent work in the special education literature has followed exactly this route, applying e-values to regressions of the form

$$Y_i = \alpha_0 + \alpha_1 \mathrm{Black}_i + \alpha_2 \mathrm{SES}_i + \alpha_3 A_i + \varepsilon_i, \tag{21}$$

and asking how strong an unobserved confounder would have to be to move α_1 from a negative value toward zero (e.g., Morgan 2021).

The key assumption behind these e-value calculations is that the covariates included in (21) form an appropriate adjustment set for the estimand of interest. That is, SES and achievement are treated as valid confounders for the race coefficient, and the only remaining threat is omitted-variable bias from some additional unobserved factor U_i that affects both race and IEP receipt. In this framework, sensitivity analysis asks how strong U_i would have to be, relative to the observed covariates, in order to change the sign or magnitude of α_1.

Our analysis shows that the main problem with (21) is different. The central threat is not primarily omitted-variable confounding by an unobserved U_i, but included-variable bias arising from the way SES and achievement enter the regression. To see this, recall the representation of special education need as its projection on (A_i, SES_i)

$$D_i^* = \lambda_0 + \lambda_A A_i + \lambda_S \mathrm{SES}_i + r_i,$$

and the oracle model

$$Y_i = \gamma_0 + \gamma_1 \mathrm{Black}_i + \gamma_2 D_i^* + u_i,$$

which identifies the IDEA motivated estimand $\tau^* = \gamma_1$ when D_i^* is observed. Substituting

the projection of D_i^* into the oracle regression yields

$$Y_i = (\gamma_0 + \gamma_2\lambda_0) + \gamma_1 \text{Black}_i + \gamma_2\lambda_A A_i + \gamma_2\lambda_S \text{SES}_i + (\gamma_2 r_i + u_i). \tag{22}$$

Let $\tilde{\text{Black}}_i$ denote the residual from regressing Black_i on (A_i, SES_i). By the Frisch–Waugh–Lovell theorem, the race coefficient in the canonical regression (21) can be written as

$$\alpha_1 = \gamma_1 + \frac{\text{Cov}(\tilde{\text{Black}}_i, \gamma_2 r_i + u_i)}{\text{Var}(\tilde{\text{Black}}_i)}. \tag{23}$$

Expression (23) makes two points clear. First, even in a world with no additional unobserved confounder beyond $(D_i^*, A_i, \text{SES}_i)$, the race coefficient from the canonical regression need not equal the oracle parameter γ_1. Setting u_i to be mean independent of $(\text{Black}_i, D_i^*, A_i, \text{SES}_i)$ removes traditional omitted-variable bias in the oracle model, but the term involving r_i generally remains:

$$\alpha_1 - \gamma_1 = \gamma_2 \frac{\text{Cov}(\tilde{\text{Black}}_i, r_i)}{\text{Var}(\tilde{\text{Black}}_i)}.$$

This term captures the fact that achievement and SES are *contaminated proxies*. They mix special education need and opportunity, so residual racial differences in D_i^* that are not captured by (A_i, SES_i) remain embedded in r_i and, under plausible conditions, stay correlated with race after conditioning on (A_i, SES_i). As a result, the canonical race coefficient can be attenuated toward zero or even negative purely because of how achievement and SES are included, even when there is no omitted confounder U_i in the usual sense.

Second, standard e-value calculations conflate these two very different sources of discrepancy between α_1 and τ^*. In the e-value setup, one conceptually augments (21) with an omitted confounder, writing

$$Y_i = \alpha_0 + \alpha_1 \text{Black}_i + \boldsymbol{\alpha}_2' X_i + \gamma_3 U_i + \tilde{u}_i,$$

where α_1 denotes the race coefficient from the canonical model that omits U_i (i.e., the estimand whose robustness is being probed), X_i collects SES, achievement, and other observed controls, and U_i is a generic unobserved confounder. The e-value then quantifies how large

the associations between U_i and race, and between U_i and IEP receipt, would need to be to move α_1 to zero. This exercise takes the adjustment for (A_i, SES_i) as appropriate for the estimand and asks only about further omissions.

Our derivations show that, for the IDEA motivated estimand τ^*, this conditioning set is itself the source of the distortion. The problematic feature of the canonical model is that SES and achievement are not neutral controls that simply remove confounding. They reallocate racially structured differences in special education need and opportunity between the covariates and the error term, and thereby change the estimand from τ^* to a different quantity. E-values computed for α_1 in (21) are therefore assessing the robustness of the wrong object. They can show that the conditional association between race and IEP receipt given SES and achievement is not easily explained away by an additional omitted confounder, but they cannot tell us whether that conditional association is a good approximation to τ^*.

In short, the main concern raised by our oracle versus canonical analysis is not that the canonical race coefficient is fragile to omitted variables in the usual sense. It is that the canonical specification itself embeds strong and largely unacknowledged assumptions about how achievement and SES decompose racial inequality into special education need versus unequal opportunity. Sensitivity analysis tools that treat these covariates as valid controls, such as e-value calculations applied to (21), do not address this included-variable bias and can give a misleading sense of robustness to claims of under representation.

4 Simulation study

Here we provide a Monte Carlo simulation that illustrates how the canonical regression of IEP receipt on race, SES, and prior achievement can deviate from the IDEA-aligned estimand. To better vizualize the canonical bias, we impose a data-generating process with race neutrality in IEP receipt conditional on special education need, so the oracle comparison is zero, meaning that $\tau^*(d) = 0$ for all d, or, equivalently, that $\gamma_1 = 0$ in (1). In our proposed SEM, Figure 1, this corresponds to setting the direct race effect in (4) to $\theta_1 = 0$. In this

data-generating process, any racial difference estimated by the canonical regression reflects bias introduced by using (A_i, SES_i) as imperfect proxies for D_i^*.

Our simulation focuses on two aspects: the conditions under which the canonical regression is biased relative to the oracle regression and the performance of confoundedness checks, such as Oster bounds and e-values, for diagnosing model fit with the oracle regression.

4.1 Simulation setup

For each Monte Carlo replication, we simulate a sample of $N = 20{,}000$ students indexed by i. Following the structural equations in Section 2 as well as the general expressions for the bias in the canonical model derived in Section 3.1, we have the following structural model of interest:

$$\begin{aligned}
\text{Black}_i &\sim \text{Bernoulli}(p_B), \\
\text{SES}_i &= \sigma_1 \text{Black}_i + \eta_i, && \eta_i \sim \mathcal{N}(0, \sigma_\eta^2), \\
O_i &= \rho_1 \text{SES}_i + \rho_2 \text{Black}_i + \omega_i, && \omega_i \sim \mathcal{N}(0, \sigma_\omega^2), \\
D_i^* &\sim \text{Bernoulli}(p_D), \\
A_i^{raw} &= \delta_1 D_i^* + \delta_2 O_i + e_{Ai}, && e_{Ai} \sim \mathcal{N}(0, \sigma_{e_A}^2), \\
A_i &= \frac{A_i^{raw} - \mathbb{E}[A_i^{raw}]}{\sqrt{\text{Var}(A_i^{raw})}}, \\
Y_i &= \gamma_0 + \gamma_2 D_i^* + e_{Yi}, && e_{Yi} \sim \mathcal{N}(0, \sigma_{e_Y}^2).
\end{aligned} \tag{24}$$

For analytical purposes, we let Y_i be a continuous index of IEP propensity rather than a literal receipt indicator in $\{0, 1\}$. It is therefore not constrained to lie in $[0, 1]$ and may take negative values; our focus is on oracle and canonical comparisons based on conditional means and linear regression coefficients.

This system maintains the simplifying assumption described in Section 2: we take special education need D_i^* to be exogenous and independent of race and socioeconomic status, so that $D_i^* \perp (\text{Black}_i, \text{SES}_i)$. As discussed in Section 2, this restriction allows the simulation to focus on the first role of SES — to control for exclusionary factors captured by achievement — and minimize its second role — of being directly informative about D_i. This isolates

the over-controlling mechanism created by conditioning on achievement, which is a collider between opportunity and need. Online Appendix B considers the case where the second role of SES is stronger, relaxing the exogeneity assumption and allowing it to directly shape D_i. The appendix illustrates that strengthening this second role does not mechanically ensure convergence of the canonical model to the oracle. The canonical race coefficient remains biased throughout and, in our calibration, becomes more negative as ρ_D increases, even though the oracle estimand is fixed at zero.

Further, under our maintained restriction that Y_i depends on race and opportunity only through disability need D_i^*, the oracle race effect

$$\tau^*(d) \equiv \mathbb{E}[Y_i \mid \text{Black}_i = 1, D_i^* = d] - \mathbb{E}[Y_i \mid \text{Black}_i = 0, D_i^* = d]$$

is zero by construction for all $d \in \{0, 1\}$.

Table 2 summarizes the baseline calibration. Our goal is not to match any single dataset exactly, but to choose transparent magnitudes that yield empirically plausible gaps and signal-to-noise ratios while preserving the mechanisms highlighted by the theory. The share of Black students ($p_B = 0.25$) and the prevalence of special education need ($p_D = 0.15$) are set to reflect national orders of magnitude in U.S. public schools (National Center for Education Statistics 2024b; Office for Civil Rights 2024).[6] We normalize latent SES in standard deviation units and set $\sigma_1 = -1$ so that the Black–White difference in socioeconomic background is large but consistent with standardized SES composites used in related work (e.g., Reardon et al. 2019). We choose $\rho_1 = 0.7$ so that SES explains a substantial, but incomplete, share of variation in opportunity (e.g., Dee 2004; Lewis and Diamond 2015). We set $\rho_2 = -0.5$ to represent a meaningful residual disparity in educational opportunity by race that persists even among students with similar measured SES. In the DGP, this term is a reduced-form representation of race-patterned structural conditions not captured by typical SES measures but nevertheless affect opportunity, such as segregation, differences in school

[6]Under the assumption that mean IEP receipt is a reasonable proxy for mean prevalence of special education need.

resources, neighborhood environments, exposure to discrimination, and other exclusionary factors (Duncan and Magnuson 2005; Fryer and Levitt 2004). Because opportunity enters achievement directly in (3), a nonzero ρ_2 implies that achievement can reflect both need and unequal opportunity even after conditioning on SES.[7] The disability effect on achievement ($\delta_1 = -1$) corresponds to a one standard deviation gap between students with and without disabilities, consistent with observed achievement differences by disability status (Gilmour et al. 2019). Finally, we treat Y_i as an index of IEP receipt that is increasing in need, setting $\gamma_2 = 1$ and choosing γ_0 so that $\mathbb{E}[Y_i] \approx \bar{Y}$ with $\bar{Y} = 0.15$. All residual standard deviations are set to unity ($\sigma_\eta = \sigma_\omega = \sigma_{e_A} = \sigma_{e_Y} = 1$) as a scale normalization, so coefficients can be read as standard-deviation-sized effects.

For the logit-based sensitivity calculations (e-values), we additionally construct a binary outcome $Y_i^{bin} = \mathbf{1}\{Y_i > c\}$ by thresholding the index within each replication (we set c to the 85th percentile of Y_i), and estimate logit models using Y_i^{bin} solely for that purpose.

Table 2: Simulation parameters (baseline specification)

Parameter	Value	Description
p_B	0.25	Share of students who are Black
p_D	0.15	Prevalence of special education need
σ_1	-1	SES penalty associated with being Black
σ_η	1	SD of SES shock η_i
ρ_1	0.7	Effect of SES on opportunity
ρ_2	-0.5	Residual Black penalty in opportunity
σ_ω	1	SD of opportunity shock ω_i
δ_1	-1	Effect of disability on achievement
δ_2	1	Effect of opportunity on achievement
σ_{e_A}	1	SD of achievement shock e_{Ai}
γ_2	1	Effect of disability on IEP receipt
σ_{e_Y}	1	SD of IEP shock e_{Yi}
$\bar{Y}$	0.15	Target mean of Y_i
N	20,000	Sample size per replication
R	500	Number of Monte Carlo replications

[7] With $\delta_2 = 1$, a residual opportunity gap of roughly half a standard deviation maps into a comparably sized contribution to achievement differences before within-replication standardization.

4.2 Parameter variation and estimators

The theoretical results in equation (13) show that the canonical race coefficient can be written as the sum of the oracle effect τ^* and a term that depends on how strongly true disability need loads onto the conditioning variables and how those variables correlate with race. In our DGP $\tau^* = 0$ and D_i^* is independent of race and SES, so any nonzero canonical race effect arises purely from the way achievement and SES act as noisy proxies for opportunity and disability.

To study these forces, we vary six parameters in turn while keeping the remaining simulation paramters described above intact.

- ρ_2 (residual Black–White opportunity gap). This parameter captures the extent to which Black students receive lower opportunity than White students even after conditioning on SES. When $\rho_2 = 0$, all systematic racial differences in O_i operate through SES. More negative values of ρ_2 represent larger opportunity gaps that the SES covariates cannot explain.
- δ_1 (effect of disability on achievement). This governs how informative achievement is about true disability need, holding opportunity fixed. When $\delta_1 = 0$, achievement reflects only opportunity. As δ_1 becomes more negative, achievement becomes a stronger proxy for D_i^*.
- σ_{e_A} (noise in achievement). This controls measurement quality of achievement. Smaller σ_{e_A} means that achievement is a very precise proxy for the latent index $\delta_1 D_i^* + \delta_2 O_i$. Larger values make it noisier.
- p_D (disability prevalence). This controls how common disability is in the population. For fixed γ_2, higher prevalence means that a larger share of the variance in IEP receipt is explained by true disability rather than noise.
- σ_1 (Black–White SES gap). This governs the size of the racial SES disparity. Larger negative values of σ_1 produce larger SES gaps, which then feed into racial gaps in opportunity and achievement.

- ρ_1 (effect of SES on opportunity). This controls how strongly SES maps into opportunity. Higher values make SES a cleaner proxy for opportunity.

In each exercise we vary one of these parameters across a grid, holding the others at their baseline values, and recompute Monte Carlo averages. For example, we vary

$$\rho_2 \in \{-0.5, -0.45, \ldots, 0\}, \qquad \delta_1 \in \{0, -0.1, -0.2, \ldots, -1.6\},$$

and use analogous grids for the remaining parameters over empirically reasonable ranges. The grids are chosen so that the Black–White achievement gap remains large and negative and SES reductions in the gap remain empirically plausible. The baseline case reported in Table 2 corresponds exactly to the parameter values in that table.

For each replication $r = 1, \ldots, R$, we draw a new sample and compute three linear regressions of Y_i.

1. Naive regression on race only

$$Y_i = \beta_0^{naive} + \beta_1^{naive}\text{Black}_i + u_i^{naive},$$

   and record the coefficient on Black_i as $\hat{\tau}_r^{naive} = \hat{\beta}_1^{naive}$.

2. Oracle regression that conditions on true disability need

$$Y_i = \gamma_0 + \gamma_1\text{Black}_i + \gamma_2 D_i^* + u_i^{oracle},$$

   and record $\hat{\tau}_r^{oracle} = \hat{\gamma}_1$. The population value of this coefficient satisfies $\tau^* = 0$ in the DGP.

3. Canonical regression that conditions on SES and achievement

$$Y_i = \alpha_0 + \alpha_1\text{Black}_i + \alpha_2\text{SES}_i + \alpha_3 A_i + u_i^{canon},$$

   and record $\hat{\tau}_r^{canon} = \hat{\alpha}_1$. This specification mirrors the canonical model in equation (5) that conditions on SES and prior achievement but does not directly observe D_i^*.

We then report $\hat{\alpha}^{canon}$, $\hat{\alpha}^{oracle}$, and bias, defined as $\hat{\alpha}^{canon} - \hat{\alpha}^{oracle}$.

To benchmark confounder-based sensitivity tools, we apply two commonly used diagnostics that ask how strongly an omitted variable would need to relate to both race and the outcome, conditional on observed controls, to move an estimated race effect toward a target value (Oster 2019; VanderWeele and Ding 2017). For Oster's coefficient-stability measure, we use the continuous outcome index Y_i and OLS, estimating within each replication a "short" regression of Y_i on Black_i alone and a "long" regression that matches the canonical adjustment set $(\text{Black}_i, \text{SES}_i, A_i)$; we report the implied $\widehat{\delta}$ needed to move the long-model race coefficient to $\beta^* = 0$ under $R^2_{\max} = 1.3 \times R^2_{\text{long}}$ capped at one. For e-values (e.g., as applied by Morgan 2021), we threshold the index to form $Y_i^{bin} = \mathbf{1}\{Y_i > c\}$ using the within-replication 85th percentile c, estimate the corresponding canonical logit model, and compute the e-value for the resulting race odds ratio. We include these diagnostics to illustrate that tools designed to assess sensitivity to *omitted confounding* can look reassuring in our simulations even when the canonical estimate is biased for a different reason, namely *over-control* arising from conditioning on contaminated proxies (A_i, SES_i) for D_i^*.

4.3 Simulation results

To interpret Figure 2, it is helpful to start from the region of the parameter space that most closely matches typical empirical applications. In most settings, there are residual racialized opportunity differences that are not fully absorbed by available SES covariates ($\rho_2 < 0$), disability need depresses achievement ($\delta_1 < 0$), and achievement is measured with some but not overwhelming noise (intermediate σ_{e_A}). In that region, the canonical estimate is consistently negative while the oracle estimate remains near zero. The canonical–oracle gap is on the order of a few percentage points, roughly 0.02 to 0.04 in magnitude. See the panels for ρ_2, δ_1, σ_{e_A}, and p_D in Figure 2.

More concretely, the panels that encode the most empirically plausible environments show bias magnitudes in the same range. Along the residual opportunity-gap panel (ρ_2), moving away from $\rho_2 \approx 0$ quickly produces canonical coefficients around -0.02 to -0.03. Along

the disability-penalty panel (δ_1), values away from $\delta_1 = 0$ produce canonical coefficients approaching -0.03 to -0.04. Along the achievement-noise panel (σ_{e_A}), bias is largest when achievement is relatively precise and remains material across the intermediate range. Finally, along the prevalence panel (p_D), bias is modest when disability is very rare but rises toward -0.03 to -0.04 as prevalence increases.

In contrast, low bias arises only under three restrictive scenarios: (i) SES nearly exhausts the Black–White achievement gap, which is typically empirically untenable, (ii) achievement ceases to predict disability need, in which case conditioning on achievement is no longer substantively informative, and (iii) achievement is so noisy that it barely predicts IEP receipt at all, which is itself generally empirically implausible and also makes achievement a weak covariate for adjustment. These cases correspond to limiting parameter values that shut down the bias term highlighted by the decomposition in equation (13). In panel (a), when $\rho_2 \approx 0$ there is essentially no residual opportunity gap after conditioning on SES, and the canonical coefficient is nearly unbiased because the SES covariates capture the IDEA-aligned opportunity path. In panel (b), when $\delta_1 = 0$ achievement reflects opportunity and noise but carries little information about disability, so the canonical coefficient again lies close to the oracle value. Likewise, when achievement is very noisy (large σ_{e_A}), conditioning on A_i weakens the induced race–need association and bias shrinks.

We now briefly summarize the panel patterns in Figure 2 and relate them to the bias decomposition in equation (13).

Residual opportunity gaps (ρ_2). Panel (a) shows that bias is smallest when $\rho_2 \approx 0$, where SES nearly eliminates residual opportunity differences by race, and grows steadily as ρ_2 becomes more negative. When Black students face worse opportunity than White students even at the same SES, achievement embeds this residual gap, so conditioning on (SES_i, A_i) induces a stronger race and special education need association and pushes the canonical coefficient away from the oracle.

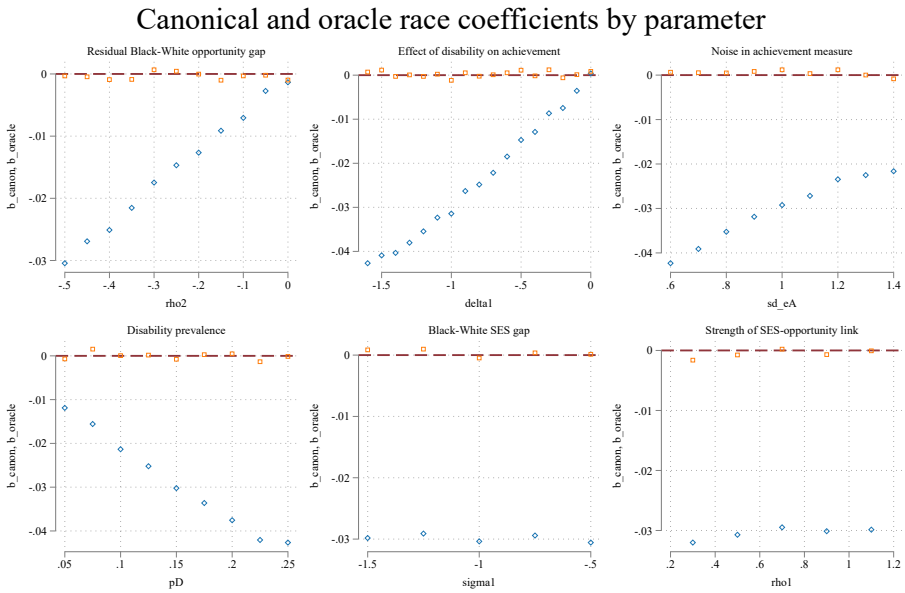

Figure 2: Each panel plots the Monte Carlo average of the canonical race coefficient, $\hat{\alpha}_{\text{canon}}$ (diamond) and the oracle race coefficient, $\hat{\alpha}_{\text{oracle}}$ (square), under the assumption that $\tau = 0$. Panels vary, in order, the residual Black–White opportunity gap (ρ_2), the effect of disability on achievement (δ_1), the noise in achievement (σ_{e_A}), disability prevalence (p_D), the Black–White SES gap (σ_1), and the strength of the SES–opportunity link (ρ_1).

How strongly achievement loads on special education need (δ_1) and how precisely it is measured (σ_{e_A}). Panels (b) and (c) show that the canonical bias increases when achievement becomes a sharper proxy for the latent index combining opportunity and need. Larger $|\delta_1|$ strengthens the collider pathway through A_i, and low σ_{e_A} makes A_i more informative about that latent index, both of which expand the scope for over-control bias. As $\delta_1 \to 0$ or σ_{e_A} grows, conditioning on A_i becomes less informative about need and the canonical coefficient moves back toward the oracle.

Disability prevalence (p_D). Panel (d) shows that bias is modest when disability is rare, since D_i^* explains little of the variation in achievement or IEP receipt, and increases as p_D rises and disability accounts for more variation in both A_i and Y_i.

SES distributions and scaling (σ_1 and ρ_1). Panels (e) and (f) show little movement in bias as the Black–White SES gap or the SES–opportunity slope changes. Once SES is included as a control, these parameters primarily shift marginal distributions rather than

the conditional relationship between race and need given (SES_i, A_i), so they do little to the bias term in equation (13).

Overall, Figure 2 indicates that the canonical race coefficient departs most from the oracle when there are meaningful residual opportunity gaps not captured by SES and when achievement is a relatively precise proxy for need and opportunity, while changes in the marginal SES distribution by race have comparatively little effect once SES is conditioned on.

4.4 Sensitivity to confoundedness based on E-values and Oster bounds

Applied studies sometimes appeal to sensitivity analyses to argue that remaining unobserved confounding is unlikely to overturn the estimated effect. Two popular approaches are the E-value and the Oster (2019) selection-on-unobservables bound. We use our data-generating process to examine how these diagnostics behave when the main threat is included-variable bias that arises from over-controlling for achievement and SES.

Figures 3 and 4 plot these diagnostics against the bias in the canonical race coefficient, $\hat{\alpha}_{\text{canon}} - \hat{\alpha}_{\text{oracle}}$, as we vary each structural parameter in turn. Recall that the true race effect is zero, so the oracle race coefficient is near zero throughout, and the bias is therefore close to the canonical race coefficient in our designs. In the figure, E-values tend to increase with the oracle gap, meaning they are largest in precisely those designs where the canonical and oracle estimands are furthest apart. Oster's $\widehat{\delta}$ behaves similarly but less orderly. Using the short regression of Y on race and the long regression adding (SES_i, A_i), targeting a zero race effect and setting $R^2_{\text{max}} = 1.3 \times R^2_{\text{long}}$ (capped at one), the implied $\widehat{\delta}$ values are often well above common benchmarks and are not reliably ordered by the bias induced by the oracle model. Overall, Oster bounds produce large δ values across almost all simulated designs, including those with substantial included-variable bias, and the magnitude of δ bears no simple relation to the true bias.

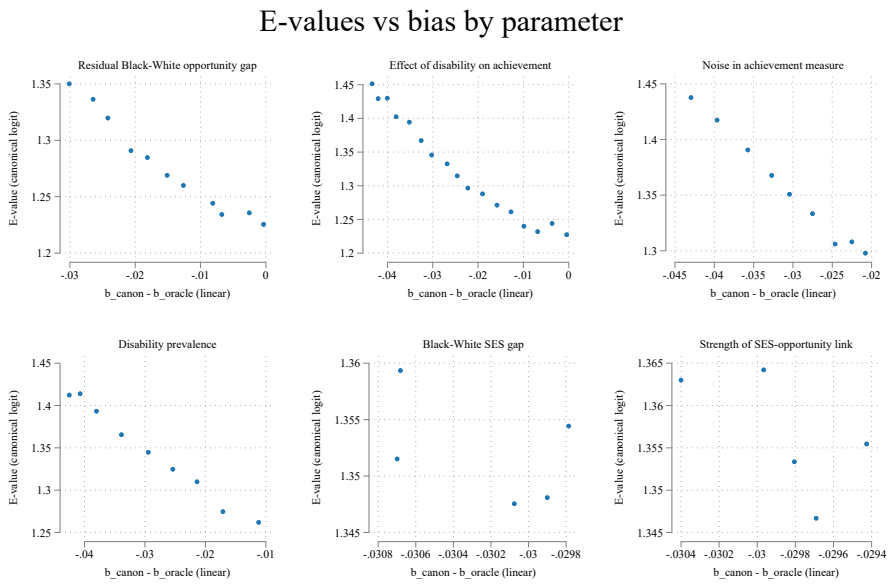

Figure 3: Each panel plots the canonical E-value from a logit model of IEP receipt against the bias in the canonical race coefficient, $\hat{\alpha}_{\text{canon}} - \hat{\alpha}_{\text{oracle}}$, for the same grids used in Figure 2. Panel titles describe the parameter that is varied. Points further to the right correspond to data-generating processes with larger included-variable bias from conditioning on (SES_i, A_i).

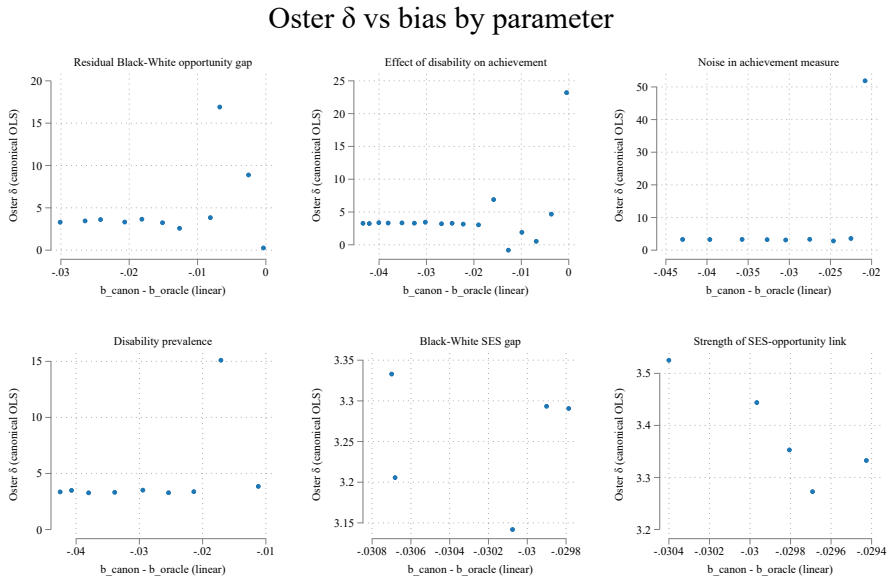

Figure 4: Each panel plots Oster's selection-on-unobservables parameter, δ, computed from the canonical OLS regression, against the bias in the canonical race coefficient, $\hat{\alpha}_{\text{canon}} - \hat{\alpha}_{\text{oracle}}$, for the same grids used in Figure 2. Very large δ values often occur in simulations that still feature substantial included-variable bias, and designs with similar bias can yield quite different δ values.

Together, these figures show that standard "robustness" summaries can appear most reassuring precisely when the canonical specification is most biased, because the core issue here is not omitted-variable bias from an unmeasured confounder but bias induced by conditioning on imperfect proxies and a collider.

5 Discussion

We have argued that, under IDEA, the policy-relevant comparison concerns differential identification among students with the same special education need. Our results show that the canonical model (conditioning on achievement and SES controls) recovers that IDEA-aligned estimand only under very strong conditions. Most notably, it requires the assumption that, after conditioning on SES, the remaining Black–White achievement gap can be fully treated as a gap in disability-related conditions rather than as a residual opportunity gap. When the SES-adjusted achievement gap remains sizable, this requirement is demanding. Given this statistical problem, should we still read existing observational evidence derived from the canonical model (e.g., Morgan et al. (2012)) as providing evidence of under-representation?

To address this question, we highlight two key results from our exercises. First, under common empirical configurations, the canonical model exaggerates under-representation. Table 1 documents a common pattern: SES covariates explain a substantial but incomplete share of Black–White achievement gaps. Such a sizable residual achievement gap after SES adjustment strongly suggests that the available SES measures are incomplete proxies for racialized exposure to resources and risks, or that the mapping from measured SES to opportunity differs across groups. Under this condition, achievement covariates pick up more Black–White differences in exclusionary factors than SES controls can reliably account for, and the canonical model mechanically generates negative race coefficients, suggesting under-representation even under a neutral benchmark.

Second, under specific and empirically plausible data-generating processes, the canonical

model pushes the coefficient toward negative values to an extent sufficient to fully account for reported levels of under-representation. Under our base parameters (Table 2), the canonical model produces a race coefficient of approximately negative 3 percentage points even when the IDEA-aligned effect is zero by construction. This simulated bias matches the magnitude of many published under-representation estimates (Table 1), suggesting that canonical bias can fully account for reported under-representation. In this sense, findings interpreted as evidence of identification barriers facing Black students could instead reflect the mismatch between what IDEA requires schools to consider and what researchers observe.

Regarding these two points, it is important to note that our simulations are likely conservative with respect to the magnitude of the bias. In empirically plausible settings where SES both shapes opportunity and also causally affects special education need, the residual Black–White achievement gap after conditioning on SES would have to be interpreted as differences in special education need *even though SES has already partialled out some need-related variation*. As discussed in Section 2.6, this implication stems from the fact that SES covariates can play two conceptually distinct roles in this setting. First, they proxy for educational opportunity and other exclusionary factors that IDEA instructs evaluators not to treat as disability-related need. Second, SES may itself causally affect special education need through biological and environmental pathways, for example via differential exposure to lead paint, air pollution, chronic stress, and other neurotoxic or health shocks that can impair learning. For clarity, our simulations isolate only the first role and impose that SES affects achievement through the opportunity channel alone (but see Online Appendix B where we relax this assumption by allowing D_i^* to covary with socioeconomic status). Thus, given the large unexplained portion of Black–White achievement gaps, the canonical assumption that the remaining SES-adjusted gap reflects differences in special education need becomes increasingly untenable.

These two points, coupled with our conservative modeling choices, lead us to conclude that existing observational evidence does not definitively establish under-representation. The

data remain consistent with parity, and over-representation cannot be excluded on empirical grounds.

Importantly, we emphasize that our exercise is not conclusive about the sign and magnitude of the IDEA-aligned estimand and that its key contribution is to help illuminate the empirical assumptions needed for given conclusions to hold. The exact sign and magnitude of estimand remains an empirical challenge, as the canonical bias is a function of two unknown factors. First, it depends on the residual achievement gap net of opportunity. The better we believe existing SES controls capture exclusionary factors, the smaller the bias. Second, it depends on how much SES already captures determinants of need. The more we think this to be true, the more demanding the interpretation of the residual achievement gap becomes.

In fact, in addition to clarifying the assumptions required for the under-representation conclusion (as discussed extensively throughout the paper), our framework also helps us understand what would be needed to reconcile existing evidence with an *over-representation* conclusion. Under the conservative assumption that SES covariates serve only their first role (proxying exclusionary determinants of achievement), over-representation would generally require that an important omitted factor remains racially stratified and depresses Black students' achievement, yet differs from the observed SES proxies in how it relates to IEP receipt. We can see this in Table 1, which shows that commonly used SES proxies are positively correlated with Black racial category and negatively correlated with IEP receipt; all else equal, omitted factors that behave like these proxies tend to push canonical estimates toward under-representation rather than over-representation. For the IDEA-aligned estimand to reflect over-representation, an omitted factor would need to disadvantage Black students in achievement (like conventional SES measures) while also being positively associated with IEP receipt.

This is possible. Environmental lead exposure provides one example: it is racially stratified due to segregation and housing disparities, it can depress achievement through neurodevelopmental harms, and it may increase disability-related need (and thus identification)

through its effects on cognition and behavior. While not impossible, this assumption — that unobserved opportunity operates in opposite directions with respect to race and special education identification —nonetheless represents a demanding one. The key point is not that such factors cannot exist, but that reconciling over-representation with canonical regression patterns requires omitted channels that are not only missing from typical SES measures but also operate differently with respect to identification than the SES proxies commonly used in observational work.

These conclusions suggest two implications for estimations of differences in IEP receipt using observational data. First, empirical work should be explicit about the estimand and about what is being conditioned on, and should avoid interpreting the canonical coefficient as the policy-relevant comparison unless the required assumptions about residual achievement gaps are defended. Researchers and policymakers should view the canonical regression not as a "correction" for poverty, but as a specific, likely biased lower-bound estimate that conflates opportunity gaps with disability-related need. More generally, standard omitted-variable sensitivity analyses are not designed to diagnose this correlated-proxy problem, so reassuring sensitivity metrics should not be taken as evidence that canonical adjustments are IDEA-aligned. Second, progress on the policy-relevant question requires richer measurement of disability-related functioning that is less entangled with opportunity to learn. To that end, methodological innovations that directly attempt to estimate the latent construct of interest can be promising (Grossman et al. 2023, 2024; Jung et al. 2024; Souto-Maior and Shroff 2024). Further, alternative observational approaches such as the threshold test (Pierson et al. 2018; Simoiu et al. 2017) and variations of well-known benchmark and outcome tests (Gaebler et al. 2022; Gaebler and Goel 2024) can offer additional promising avenues beyond traditional regression approaches.

That said, given the measurement and data challenges of identifying the estimand in observational data, another conclusion from this exercise is that experimental designs provide a promising and relatively underexplored approach within this literature. For example, evi-

dence from audits of principals suggests that access to schools and responsiveness to family requests are jointly shaped by disability status and race in ways consistent with diminished responsiveness to Black families seeking accommodations or support (Rivera and Tilcsik 2023). Similarly, in audit studies of teachers, White students are more likely to be recommended for academic disability diagnostic evaluation than Black students, whereas Black students are more likely to receive referrals for behavioral problems (Fish 2017). Taken together, these designs illuminate mechanisms of discretionary bias, showing that discrimination can shape access to evaluation and support through in which needs are recognized—especially for academic concerns—that vary by disability and race. However, these are small-scale studies in selected contexts, and improved large-scale observational evidence remains essential for establishing the population-level direction of representation patterns.

Overall, our framework helps assess how far the canonical coefficient can drift from the IDEA-aligned comparison across empirically plausible environments, and clarifies why common omitted-variable sensitivity tools can appear reassuring in a proxy-control setting even when the canonical estimate is substantially misaligned with the policy-relevant estimand. The substantive takeaway is that negative race coefficients from canonical adjustments are unlikely to provide reliable quantitative estimates of IDEA-aligned under-service. In our framework, such negative coefficients can plausibly be generated by proxy-induced bias even when the IDEA-aligned estimand implies parity or even over-representation. Accordingly, findings of substantial under-representation that rest primarily on canonical adjustments should be interpreted with caution and should not be treated as direct quantitative evidence of under-service. This insight should give pause to policy prescriptions which, based on prior evidence, encourage greater identification of Black students for special education services. Whether the true estimand reflects modest under-representation, parity, or over-representation remains an unresolved empirical question. Academic progress requires designs and measurements that better separate disability-related functioning from opportunity to learn to better inform this central policy question.

References

Artiles, Alfredo J. 2013. "Untangling the racialization of disabilities: An intersectionality critique across disability models." *Du Bois Review: Social Science Research on Race* 10:329–347.

Artiles, Alfredo J, Beth Harry, Daniel J Reschly, and Philip C Chinn. 2002. "Over-identification of students of color in special education: A critical overview." *Multicultural Perspectives* 4:3–10.

Ayres, Ian. 2005. "Three tests for measuring unjustified disparate impacts in organ transplantation: The problem of 'included variable' bias." *Perspectives in Biology and Medicine* 48:68–S87.

Burke, Lindsey M. 2023. "Department of Education." In *Mandate for Leadership: The Conservative Promise*, edited by Paul Dans and Steven Groves, chapter 11, pp. 319–362. Washington, DC: The Heritage Foundation. Project 2025 Presidential Transition Project.

Dee, Thomas S. 2004. "Teachers, race, and student achievement in a randomized experiment." *Review of Economics and Statistics* 86:195–210.

Donovan, M Suzanne and Christopher T Cross. 2002. *Minority students in special and gifted education*. National Academies Press.

Duncan, Greg J and Katherine A Magnuson. 2005. "Can family socioeconomic resources account for racial and ethnic test score gaps?" *The Future of Children* pp. 35–54.

Dunn, Lloyd M. 1968. "Special education for the mildly retarded—Is much of it justifiable?" *Exceptional children* 35:5–22.

Fish, Rachel Elizabeth. 2017. "The racialized construction of exceptionality: Experimental evidence of race/ethnicity effects on teachers' interventions." *Social Science Research* 62:317–334.

Fish, Rachel Elizabeth. 2019. "Standing out and sorting in: Exploring the role of racial composition in racial disparities in special education." *American Educational Research Journal* 56:2573–2608.

Fish, Rachel Elizabeth, Kenneth Shores, and João M Souto Maior. 2025. "A Critical Appraisal of the Evidence on Racial Disproportionality in Special Education." *Exceptional Children* p. 00144029251350094.

Fryer, Roland G and Steven D Levitt. 2004. "Understanding the black-white test score gap in the first two years of school." *Review of economics and statistics* 86:447–464.

Gaebler, Johann, William Cai, Guillaume Basse, Ravi Shroff, Sharad Goel, and Jennifer Hill. 2022. "A causal framework for observational studies of discrimination." *Statistics and Public Policy* 9:26–48.

Gaebler, Johann D and Sharad Goel. 2024. "A Simple, Statistically Robust Test of Discrimination." *arXiv preprint arXiv:2407.06539* .

Gilmour, Allison F., Douglas Fuchs, and Joseph Wehby. 2025. "Reconsidering the achievement gap for students with disabilities." *Education Next* Accessed 2026-01-05.

Gilmour, Allison F, Douglas Fuchs, and Joseph H Wehby. 2019. "Are students with disabilities accessing the curriculum? A meta-analysis of the reading achievement gap between students with and without disabilities." *Exceptional Children* 85:329–346.

Grossman, Joshua, Julian Nyarko, and Sharad Goel. 2023. "Racial bias as a multi-stage, multi-actor problem: An analysis of pretrial detention." *Journal of Empirical Legal Studies* 20:86–133.

Grossman, Joshua, Julian Nyarko, and Sharad Goel. 2024. "Reconciling Legal and Empirical Conceptions of Disparate Impact: An Analysis of Police Stops Across California." *Journal of Law and Empirical Analysis* 1:2755323X241243168.

Hibel, Jacob, George Farkas, and Paul L Morgan. 2010. "Who is placed into special education?" *Sociology of Education* 83:312–332.

IDEA-Sec.300.309. 2006. "Individuals with Disabilities Education Act of 2004. Sec. 300.309: Determining the existence of a specific learning disability." FR 46753, Aug. 14, 2006, as amended at 82 FR 31912, July 11, 2017. https://sites.ed.gov/idea/regs/b/d/300.309.

Jung, Jongbin, Sam Corbett-Davies, Ravi Shroff, and Sharad Goel. 2024. "Omitted and included variable bias in tests for disparate impact." *arXiv preprint arXiv:1809.05651* .

Kincaid, Aleksis P and Amanda L Sullivan. 2017. "Parsing the relations of race and socioeconomic status in special education disproportionality." *Remedial and Special Education* 38:159–170.

Lewis, Amanda E and John B Diamond. 2015. *Despite the best intentions: How racial inequality thrives in good schools*. Oxford University Press.

Lundberg, Ian, Rebecca Johnson, and Brandon M Stewart. 2021. "What is your estimand? Defining the target quantity connects statistical evidence to theory." *American Sociological Review* 86:532–565.

Morgan, Paul L. 2021. "Unmeasured confounding and racial or ethnic disparities in disability identification." *Educational Evaluation and Policy Analysis* 43:351–361.

Morgan, Paul L and George Farkas. 2015. "Is Special Education Racist?" *New York Times* June 24, 2015. https://www.nytimes.com/2015/06/24/opinion/is-special-education-racist.html.

Morgan, Paul L, George Farkas, Marianne M Hillemeier, and Steve Maczuga. 2012. "Are minority children disproportionately represented in early intervention and early childhood special education?" *Educational Researcher* 41:339–351.

Morgan, Paul L, George Farkas, Marianne M Hillemeier, Richard Mattison, Steve Maczuga, Hui Li, and Michael Cook. 2015. "Minorities are disproportionately underrepresented in special education: Longitudinal evidence across five disability conditions." *Educational Researcher* 44:278–292.

Morgan, Paul L, Adrienne D Woods, Yangyang Wang, Marianne M Hillemeier, George Farkas, and Cynthia Mitchell. 2020. "Are schools in the US South using special education to segregate students by race?" *Exceptional Children* 86:255–275.

National Center for Education Statistics. 2024a. "Digest of Education Statistics, Table 204.40. Children 3 to 21 years old served under IDEA, Part B, by race/ethnicity." Accessed 2026-01-05.

National Center for Education Statistics. 2024b. "Racial/Ethnic Enrollment in Public Schools." Technical report, Institute of Education Sciences, U.S. Department of Education.

Office for Civil Rights. 2024. "Profile of Students with Disabilities in U.S. Public Schools During the 2020-21 School Year." Technical report, U.S. Department of Education.

Oster, Emily. 2019. "Unobservable selection and coefficient stability: Theory and evidence." *Journal of Business & Economic Statistics* 37:187–204.

Pierson, Emma, Sam Corbett-Davies, and Sharad Goel. 2018. "Fast threshold tests for detecting discrimination." *International conference on artificial intelligence and statistics* pp. 96–105.

Reardon, Sean F, Demetra Kalogrides, and Kenneth Shores. 2019. "The geography of racial/ethnic test score gaps." *American Journal of Sociology* 124:1164–1221.

Rivera, Lauren A. and Andras Tilcsik. 2023. "Not in my schoolyard: Disability discrimination in educational access." *American Sociological Review* 88:284–321. Accessed 2026-01-05.

Shifrer, Dara. 2018. "Clarifying the social roots of the disproportionate classification of racial minorities and males with learning disabilities." *The Sociological Quarterly* 59:384–406.

Simoiu, Camelia, Sam Corbett-Davies, and Sharad Goel. 2017. "The problem of infra-marginality in outcome tests for discrimination." *The Annals of Applied Statistics* .

Skiba, Russell J, Alfredo J Artiles, Elizabeth B Kozleski, Daniel J Losen, and Elizabeth G Harry. 2016. "Risks and consequences of oversimplifying educational inequities: A response to Morgan et al.(2015)." *Educational Researcher* 45:221–225.

Souto-Maior, João M and Ravi Shroff. 2024. "Differences in Academic Preparedness Do Not Fully Explain Black–White Enrollment Disparities in Advanced High School Coursework." *Sociological Science* 11:138–163.

Sullivan, Amanda L and Aydin Bal. 2013. "Disproportionality in special education: Effects of individual and school variables on disability risk." *Exceptional Children* 79:475–494.

VanderWeele, Tyler J and Peng Ding. 2017. "Sensitivity analysis in observational research: introducing the E-value." *Annals of internal medicine* 167:268–274.

Online Appendix A Audit Studies and the Canonical Regression

A.1 Relationship between the oracle model and audit studies

Our IDEA-aligned estimand compares IEP receipt rates between Black and White students holding fixed underlying special education need, D_i^*:

$$\tau^*(d) \equiv \mathbb{E}[Y_i \mid Black_i = 1, D_i^* = d] - \mathbb{E}[Y_i \mid Black_i = 0, D_i^* = d]\,. \tag{A.1}$$

As discussed above, if D_i^* were observed, an oracle regression that conditions directly on D_i^*,

$$Y_i = \gamma_0 + \gamma_1 Black_i + \gamma_2 D_i^* + u_i, \qquad \mathbb{E}[u_i \mid Black_i, D_i^*] = 0, \tag{A.2}$$

identifies (A.1) via $\tau^*(d) = \gamma_1$ for all d under the maintained linear specification.

Audit studies. An audit study can be written in similar notation by distinguishing actual race from a randomized racial *signal*. Let $Z_i \in \{0, 1\}$ indicate whether an inquiry is assigned a Black-sounding (versus White-sounding) name, where Z_i is randomly assigned by the researcher. Let D_i^* denote the underlying special education need implied by the inquiry (or the child's described profile) and let Y_i denote the binary response of interest (e.g., whether the school replies, encourages evaluation, offers an appointment, etc.). Define potential outcomes $Y_i(z)$ for $z \in \{0, 1\}$ as the response that would be observed if the inquiry were sent under racial signal z.

Random assignment implies $(Y_i(0), Y_i(1), D_i^*) \perp Z_i$, and by consistency $Y_i = Y_i(Z_i)$, so for any need level d,

$$\mathbb{E}[Y_i(1) - Y_i(0) \mid D_i^* = d] = \mathbb{E}[Y_i \mid Z_i = 1, D_i^* = d] - \mathbb{E}[Y_i \mid Z_i = 0, D_i^* = d]\,. \tag{A.3}$$

Thus, the causal estimand in an audit study is numerically equal to a conditional mean gap, with the key difference that the gap is causally interpretable because Z_i is randomized.

Consider estimating the audit study via the regression

$$Y_i = \beta_0 + \beta_1 Z_i + \beta_2 D_i^* + \varepsilon_i. \tag{A.4}$$

Because Z_i is randomized, $\mathbb{E}[\varepsilon_i \mid Z_i, D_i^*] = 0$ holds by design. Moreover, if the effect of the race signal does not vary with need level (or if one estimates the effect flexibly by interacting Z_i with D_i^*), then the implied conditional mean gap at any fixed d equals

$$\mathbb{E}[Y_i \mid Z_i = 1, D_i^* = d] - \mathbb{E}[Y_i \mid Z_i = 0, D_i^* = d] = \beta_1, \tag{A.5}$$

and combining (A.5) with (A.3) yields

$$\beta_1 = \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i^* = d]\,. \tag{A.6}$$

Connecting the two comparisons. Equation (A.1) defines $\tau^*(d)$ as an observational conditional disparity using the actual race indicator $Black_i$. By contrast, audit studies identify a causal effect of a randomized *race signal* Z_i. To compare like with like, define an *audit-analogue* oracle gap by replacing $Black_i$ with the manipulable signal Z_i:

$$\tau_Z(d) \equiv \mathbb{E}[Y_i \mid Z_i = 1, D_i^* = d] - \mathbb{E}[Y_i \mid Z_i = 0, D_i^* = d]\,. \tag{A.7}$$

Under random assignment, this audit-analogue gap coincides with the audit CATE:

$$\tau_Z(d) = \mathbb{E}[Y_i \mid Z_i = 1, D_i^* = d] - \mathbb{E}[Y_i \mid Z_i = 0, D_i^* = d] = \mathbb{E}[Y_i(1) - Y_i(0) \mid D_i^* = d]\,. \tag{A.8}$$

In short, audits identify the same *functional form* as an oracle-style conditional mean gap, but with a manipulable signal and design-based justification. Randomization makes the "error term unrelated to race" (or a racial signifier) condition automatic, whereas in observational settings the oracle regression (A.2) requires the maintained mean-independence assumption $\mathbb{E}[u_i \mid Black_i, D_i^*] = 0$.

It is important that $\tau^*(d)$ is defined as a descriptive conditional disparity in the population, while audit studies are designed to identify a causal effect of perceived race. The equivalence in (A.8) therefore does not assert that $\tau^*(d)$ equals an audit intent to treat effect. Rather, it clarifies that both objects are conditional mean differences, and they differ in which "race" variable is used and in how the contrast is interpreted and justified.

This equivalence has two implications. First, audit evidence can be read as a direct, design-based benchmark for the magnitude of differential treatment that would generate a given conditional gap when race is construed as a signal. Second, it sharpens what is required to interpret observational estimates of $\tau^*(d)$ as treatment effects. Observational designs must argue that conditioning on D_i^* suffices to render race orthogonal to unobserved determinants of Y_i, while audits obtain that orthogonality by random assignment.

Finally, the equivalence also highlights external-validity and construct-validity limits. Audit studies manipulate a specific racial cue (often names) and typically capture an early-stage decision (e.g., reply, encouragement, referral), while IEP receipt is a multi-stage process and represents the realized outcome among an enrolled population. Thus, even when the estimands take the form of conditional mean gaps, the setting and the race cue being manipulated can change which part of the broader identification process the estimate speaks to.

Online Appendix B Allowing SES and D^* to covary: simulation design and diagnostics for proxy and collider bias

B.1 Overview and procedure

The simulation presented in Section 4 in the main text imposes that underlying disability need is exogenous and independent of race and SES. Here, we relax this simplification by allowing special education need, D_i^*, to covary with latent socioeconomic status. We place this extension in an online appendix because it adds an additional layer of uncertainty — specifically, the need to model differences between true and observed SES — without substantially influencing our argument that the canonical model can systematically deviate from the oracle model. In fact, we show here that introducing SES to shape D_i^* can further magnify the canonical bias.

We replicate our Monte Carlo simulations (Section 4 in the main text) to study when a

canonical observational adjustment strategy can misstate a race effect even though the true outcome model is race neutral conditional on disability need. Each replication follows four steps. First, we draw race and a latent socioeconomic construct, then generate an observed socioeconomic proxy that can be measured with error. Second, we generate educational opportunity as a function of latent socioeconomic status and race. Third, we generate underlying disability need from a latent index that is allowed to be correlated with latent socioeconomic status. Fourth, we generate achievement and the outcome, then estimate three regression specifications that mirror a naive comparison, an oracle regression that conditions on disability need, and the canonical regression that conditions only on observed proxies. Across replications, we average coefficients and diagnostic statistics. We then vary selected structural parameters one at a time, treating each as a sensitivity knob, and trace how the canonical estimand diverges from the oracle estimand.

B.2 Data-generating process

We modify the structural system described in Section 4.3 in the main text to allow SES to influence special education need. To do so, we acknowledge that SES is often measured with error, creating variables for true (SES_i^*) and observed (SES_i) SES. Further, we introduce parameter ρ_D to capture the influence of true SES on special education need.

Specifically, we draw

$$\text{SES}_i^* = \sigma_1 B_i + \eta_i, \qquad \eta_i \sim \mathcal{N}(0, \sigma_\eta^2),$$

and

$$\text{SES}_i = \text{SES}_i^* + u_i, \qquad u_i \sim \mathcal{N}(0, \sigma_{\text{SES,me}}^2),$$

where $\sigma_{\text{SES,me}}$ captures the variance in the SES error.

Educational opportunity is generated from true socioeconomic status and race,

$$O_i = \rho_1 \text{SES}_i^* + \rho_2 B_i + \omega_i, \qquad \omega_i \sim \mathcal{N}(0, 1).$$

To generate underlying disability in terms of latent SES, let

$$Z_i = \frac{\text{SES}_i^* - \mathbb{E}(\text{SES}^*)}{\text{SD}(\text{SES}^*)}$$

denote the within-replication standardized version of SES_i^*. Then we let

$$D_i^{\text{index}} = \rho_D Z_i + \sqrt{1-\rho_D^2}\, v_i, \qquad v_i \sim \mathcal{N}(0,1),$$

and we set $D_i^* = \mathbb{1}\{D_i^{\text{index}} > c_D\}$ where c_D is chosen within each replication so that $\Pr(D_i^* = 1) = p_D$. By construction, ρ_D equals $\text{Corr}(D^{\text{index}}, \text{SES}^*)$. The correlation between the binary D_i^* and SES_i^* is monotone in ρ_D but need not equal ρ_D.

Achievement is affected by disability need and opportunity and then standardized within replication,

$$A_i^{\text{raw}} = \delta_1 D_i^* + O_i + \varepsilon_{Ai}, \qquad \varepsilon_{Ai} \sim \mathcal{N}(0, \sigma_A^2), \qquad A_i = \frac{A_i^{\text{raw}} - \mathbb{E}(A^{\text{raw}})}{\text{SD}(A^{\text{raw}})}.$$

The main outcome is generated to be race neutral conditional on disability need,

$$Y_i = \gamma_0 + \gamma_2 D_i^* + \varepsilon_{Yi}, \qquad \varepsilon_{Yi} \sim \mathcal{N}(0,1), \qquad \gamma_0 \text{ chosen so that } \mathbb{E}(Y) = 0.15.$$

As in Section 4 in the main text, Y_i is a continuous IEP propensity index rather than a literal receipt indicator. In the DGP, IEP assignment conditional on special education receipt remains race neutral ($\tau_{\text{true}} = 0$).

B.3 Estimators

In each replication we estimate the three linear regressions detailed in Section 4.3 in the main text.

B.4 Baseline calibration and Monte Carlo design

Unless otherwise stated, we use the baseline parameter values defined in 2 in the main text. We set the new error variance parameter, $\sigma^2_{\text{SES,me}}$, to 4 and vary the new parameter ρ_D, which is the correlation between the latent index D_i^* and SES_i. When $\rho_D = 0$ the

design returns to the baseline maintained assumption that D_i^* is independent of SES in the population, so any canonical bias that remains is attributable to the contaminated-proxy mechanism operating through achievement and residual racial differences in opportunity.

B.5 Sensitivity knobs

We vary one parameter at a time over a grid while holding the others at baseline.

- $\rho_D \in [-1, 1]$ controls the coupling between the disability propensity index and latent socioeconomic status, where $\rho_D = \mathrm{Corr}(D^{\mathrm{index}}, \mathrm{SES}^*)$ by construction. In the mechanism figure we emphasize $\rho_D \geq 0$ for compactness. Negative values simply reverse the direction of the SES–disability association.
- ρ_2 varies the residual Black–White opportunity gap after conditioning on latent SES.
- δ_1 varies the causal effect of disability on achievement.
- σ_A varies noise in the achievement measure.
- p_D varies disability prevalence.
- σ_1 varies the Black–White gap in latent SES.
- ρ_1 varies the strength of the SES–opportunity relationship.

B.6 Analytical link between canonical bias and residual race–need association

Throughout the paper we define bias as the canonical–oracle gap in the race coefficient. In this appendix, because the outcome is generated to depend on race only through D_i^*, that same gap admits a convenient representation: it equals γ_2 times the residual race coefficient from the linear projection of D_i^* on the canonical adjustment set. This equality results from the following.

First, this appendix generates outcomes from

$$Y_i = \gamma_0 + \gamma_2 D_i^* + \varepsilon_{Yi}, \tag{B.1}$$

and estimates the canonical regression of Y_i on race and the observed proxies,

$$Y_i = \beta_0^{\text{canon}} + \beta_B^{\text{canon}} B_i + \beta_S^{\text{canon}} \text{SES}_i + \beta_A^{\text{canon}} A_i + u_i. \tag{B.2}$$

We now show that the population canonical race coefficient β_B^{canon} is proportional to the population race coefficient from the linear projection of D_i^* on the same covariates that appear in the canonical regression. That is, we show that

$$\beta_B^{\text{canon}} = \gamma_2 \, \pi_B,$$

where π_B is the coefficient on B_i in the population regression of D_i^* on (B_i, SES_i, A_i) and γ_2 is the effect of disability need on Y_i in (B.1).

Second, let $X_i \equiv (\text{SES}_i, A_i)$ collect the non-race controls. Let $\mathcal{L}(\cdot \mid X_i)$ denote the population linear projection onto $(1, X_i)$. Define residualized variables

$$\widetilde{Y}_i \equiv Y_i - \mathcal{L}(Y_i \mid X_i), \qquad \widetilde{B}_i \equiv B_i - \mathcal{L}(B_i \mid X_i), \qquad \widetilde{D}_i^* \equiv D_i^* - \mathcal{L}(D_i^* \mid X_i).$$

By the Frisch–Waugh–Lovell theorem, the population coefficient on B_i in (B.2) equals the slope from the bivariate regression of $\widetilde{Y}_i$ on $\widetilde{B}_i$,

$$\beta_B^{\text{canon}} = \frac{\text{Cov}(\widetilde{B}_i, \widetilde{Y}_i)}{\text{Var}(\widetilde{B}_i)}. \tag{B.3}$$

Next substitute the outcome model (B.1) into $\widetilde{Y}_i$. Using linearity of the projection operator,

$$\mathbb{E}[Y_i \mid X_i] = \gamma_0 + \gamma_2 \mathbb{E}[D_i^* \mid X_i] + \mathbb{E}[\varepsilon_{Yi} \mid X_i],$$

so

$$\widetilde{Y}_i = Y_i - \mathbb{E}[Y_i \mid X_i] = \gamma_2 \big(D_i^* - \mathbb{E}[D_i^* \mid X_i]\big) + \big(\varepsilon_{Yi} - \mathbb{E}[\varepsilon_{Yi} \mid X_i]\big) = \gamma_2 \widetilde{D}_i^* + \widetilde{\varepsilon}_{Yi},$$

where $\widetilde{\varepsilon}_{Yi} \equiv \varepsilon_{Yi} - \mathbb{E}[\varepsilon_{Yi} \mid X_i]$.

Plug this decomposition into (B.3) to obtain

$$\beta_B^{\text{canon}} = \frac{\text{Cov}(\widetilde{B}_i, \, \gamma_2 \widetilde{D}_i^* + \widetilde{\varepsilon}_{Yi})}{\text{Var}(\widetilde{B}_i)} = \gamma_2 \frac{\text{Cov}(\widetilde{B}_i, \widetilde{D}_i^*)}{\text{Var}(\widetilde{B}_i)} + \frac{\text{Cov}(\widetilde{B}_i, \widetilde{\varepsilon}_{Yi})}{\text{Var}(\widetilde{B}_i)}. \tag{B.4}$$

In our Monte Carlo design, ε_{Yi} is generated independently of (B_i, X_i, D_i^*), so $\mathrm{Cov}(\widetilde{B}_i, \widetilde{\varepsilon}_{Yi}) = 0$. Under this condition,

$$\beta_B^{\text{canon}} = \gamma_2 \frac{\mathrm{Cov}(\widetilde{B}_i, \widetilde{D}_i^*)}{\mathrm{Var}(\widetilde{B}_i)}. \tag{B.5}$$

This result shows that the race coefficient from the canonical regression can be expressed as the direct effect of special education need D_i^* weighted by (or proportional to) the $\frac{\mathrm{Cov}(\widetilde{B}_i, \widetilde{D}_i^*)}{\mathrm{Var}(\widetilde{B}_i)}$. To understand this final component, consider the population linear projection of disability need on the same covariates,

$$D_i^* = \pi_0 + \pi_B B_i + \pi_X' X_i + \nu_i. \tag{B.6}$$

Applying Frisch–Waugh–Lovell again, the coefficient on B_i in (B.6) satisfies

$$\pi_B = \frac{\mathrm{Cov}(\widetilde{B}_i, \widetilde{D}_i^*)}{\mathrm{Var}(\widetilde{B}_i)}.$$

Combining this identity with (B.5) yields the key proportionality,

$$\beta_B^{\text{canon}} = \gamma_2 \pi_B. \tag{B.7}$$

Because the oracle regression conditions on D_i^* directly and the outcome model (B.1) contains no direct race term, the oracle race coefficient is $\beta_B^{\text{oracle}} = 0$ in this appendix. Therefore the canonical–oracle gap reduces to

$$\beta_B^{\text{canon}} - \beta_B^{\text{oracle}} = \gamma_2 \pi_B.$$

This result shows that, in this design, canonical bias is equivalent to residual race–need association remaining after conditioning on the observed proxies (SES_i, A_i).

B.7 Diagnostics for interpreting ρ_D variation

Equation (B.7) implies that the bias plot and the "residual racial sorting into disability need" plot are the same object up to scale. With $\gamma_2 > 0$ fixed, the sign and shape of $\beta_B^{\text{canon}} - \beta_B^{\text{oracle}}$ over ρ_D variation are governed by π_B, the coefficient on B_i in the projection $D_i^* \sim B_i + \mathrm{SES}_i + A_i$. The remaining diagnostics explain why π_B changes with ρ_D. The

proxy-strength panels summarize how much of the variation in D_i^* is absorbed by (SES_i, A_i), which in turn determines how much race-related variation in D_i^* remains after conditioning. The collider panel quantifies how conditioning on A_i induces dependence between D_i^* and opportunity O_i, which can shift π_B whenever opportunity retains residual racial differences through ρ_2.

To interpret why this residual association changes, we report four mechanism diagnostics.

(i) Proxy strength for need. We report R^2 from the linear probability projections $D_i^* \sim \mathrm{SES}_i$ and $D_i^* \sim \mathrm{SES}_i + A_i$. These quantify how much of the variation in D_i^* is captured by the observed proxies, which is central because ν_i in the projection above is precisely the component of need that remains uncontrolled in the canonical regression.

(ii) Collider-induced coupling of need and opportunity. In the structural system, achievement is a common effect of need and opportunity. Conditioning on A_i therefore changes the conditional dependence between D_i^* and O_i. We measure this by comparing $\mathrm{pcorr}(D_i^*, O_i \mid B_i, \mathrm{SES}_i)$ to $\mathrm{pcorr}(D_i^*, O_i \mid B_i, \mathrm{SES}_i, A_i)$, computed from residualized versions of D_i^* and O_i.

(iii) Residual race–need association after conditioning. We track the coefficient on B_i in $D_i^* \sim B_i + \mathrm{SES}_i$ and in $D_i^* \sim B_i + \mathrm{SES}_i + A_i$. By the identity above, the latter coefficient is proportional to the canonical race coefficient in the Y regression.

(iv) Canonical bias. We report $\beta^{\mathrm{canon}} - \beta^{\mathrm{oracle}}$ and compare it to diagnostics (i)–(iii).

B.8 Interpreting how ρ_D moves bias

Figure B.1 varies $\rho_D = \mathrm{Corr}(D^{\mathrm{index}}, \mathrm{SES}^*)$ while holding fixed the remaining structure. Recall from Equation (B.7) that, in population, $\beta_B^{\mathrm{canon}} - \beta_B^{\mathrm{oracle}} = \gamma_2 \pi_B$, so movement in the canonical–oracle gap over changes to ρ_D must operate through π_B, the residual race–need association after conditioning on the canonical proxy variables (SES, A).

Two distinct roles of SES discussed in Section 2.6 are relevant in this extension. First, SES continues to proxy for exclusionary factors because it predicts opportunity through $O_i = \rho_1 \text{SES}_i^* + \rho_2 B_i + \omega_i$, and opportunity loads into achievement. Conditioning on observed SES is therefore intended to absorb part of the opportunity component of achievement, reducing the extent to which residual achievement gaps reflect unequal opportunity rather than need.

Second, $\rho_D > 0$ activates the role of SES as being directly informative about need through the latent disability index. Because SES is observed with error, conditioning on SES_i does not fully absorb SES_i^*. As ρ_D rises, the remaining variation in SES_i^* continues to predict D_i^*, and since SES_i^* is racially stratified, this shows up as a non-zero coefficient on B_i in $D_i^* \sim B_i + \text{SES}_i$.

The diagnostic panels illustrate how the mechanisms translate into the canonical–oracle gap. Panel A (top left) shows proxy strength. As ρ_D increases, observed SES becomes more informative about D_i^*, so $R^2(D_i^* \sim \text{SES}_i)$ rises, while the incremental gain from adding achievement shrinks.

Panel B (top right) summarizes the collider mechanism. Holding fixed (B_i, SES_i), conditioning on achievement raises $\text{pcorr}(D_i^*, O_i \mid B_i, \text{SES}_i, A_i)$ relative to $\text{pcorr}(D_i^*, O_i \mid B_i, \text{SES}_i)$, consistent with A_i being a common effect of need and opportunity. Because opportunity retains a race component through ρ_2, this collider-induced coupling can change the residual race–need association after adjustment, and therefore shift π_B.

Panel C (bottom left) reports the race coefficient in disability-need projections. In particular, the "SES+Ach" series is π_B from $D_i^* \sim B_i + \text{SES}_i + A_i$. By Equation (B.7), this coefficient maps one-for-one (up to the scale factor γ_2) into the population canonical race coefficient in the Y regression.

Panel D (bottom right) then shows the implied result for Y. As ρ_D increases, Panel C shows π_B becoming more negative in this calibration, so Panel D shows the canonical race coefficient becoming more negative as well, while the oracle coefficient remains near zero.

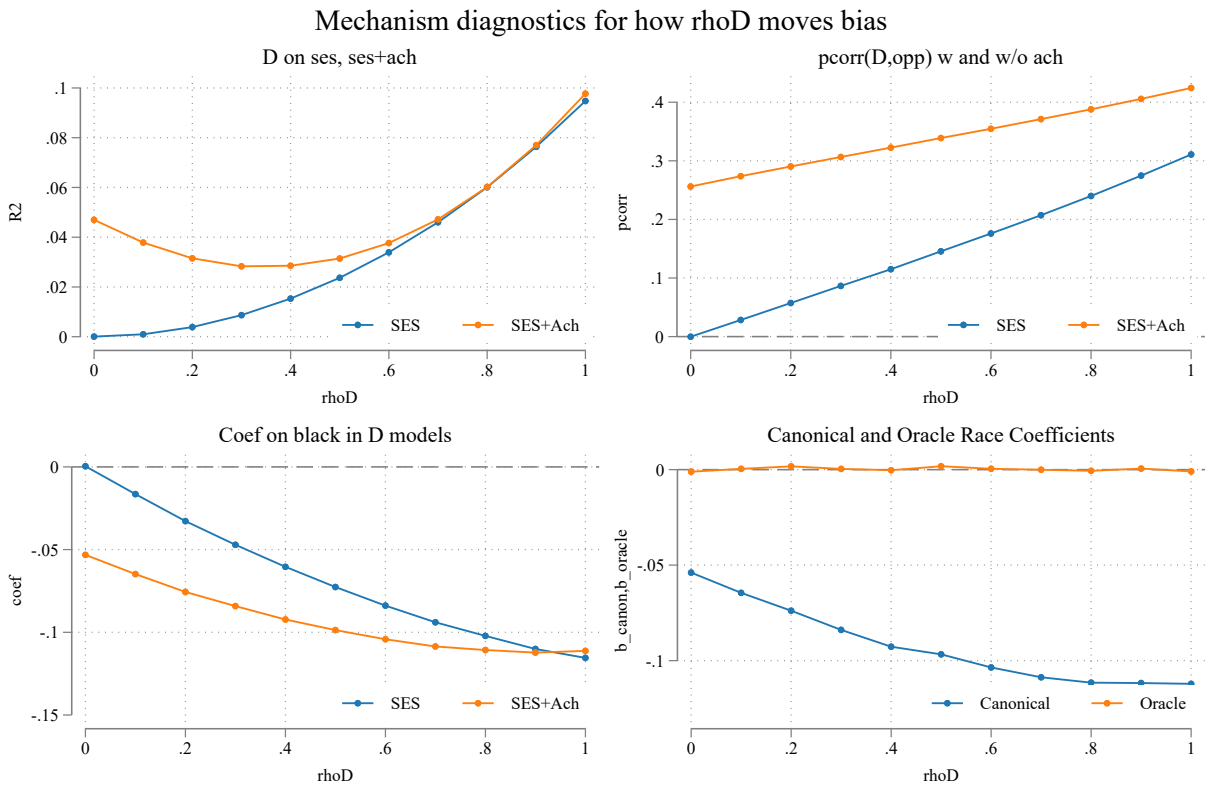

Figure B.1: Mechanism diagnostics for how ρ_D moves bias over $\rho_D \in [0, 1]$. Top left reports proxy strength for predicting disability need from SES and from SES plus achievement. Top right reports the partial correlation between disability need and opportunity residuals with and without conditioning on achievement. Bottom left reports the coefficient on race in disability-need regressions net of observed SES and net of observed SES plus achievement. Bottom right reports the canonical and oracle race coefficients, whose difference is the bias metric $\beta^{\text{canon}} - \beta^{\text{oracle}}$.

Takeaway. Allowing need to covary with socioeconomic background does not remove the contaminated-proxy problem. Increasing ρ_D strengthens SES as a predictor of need, but with measurement error in SES and residual racial differences in opportunity, conditioning on achievement can still generate residual race–need association, and therefore a non-zero canonical race coefficient. Indeed, this exercise demonstrates that introducing SES to shape D_i^* can further magnify the canonical bias.