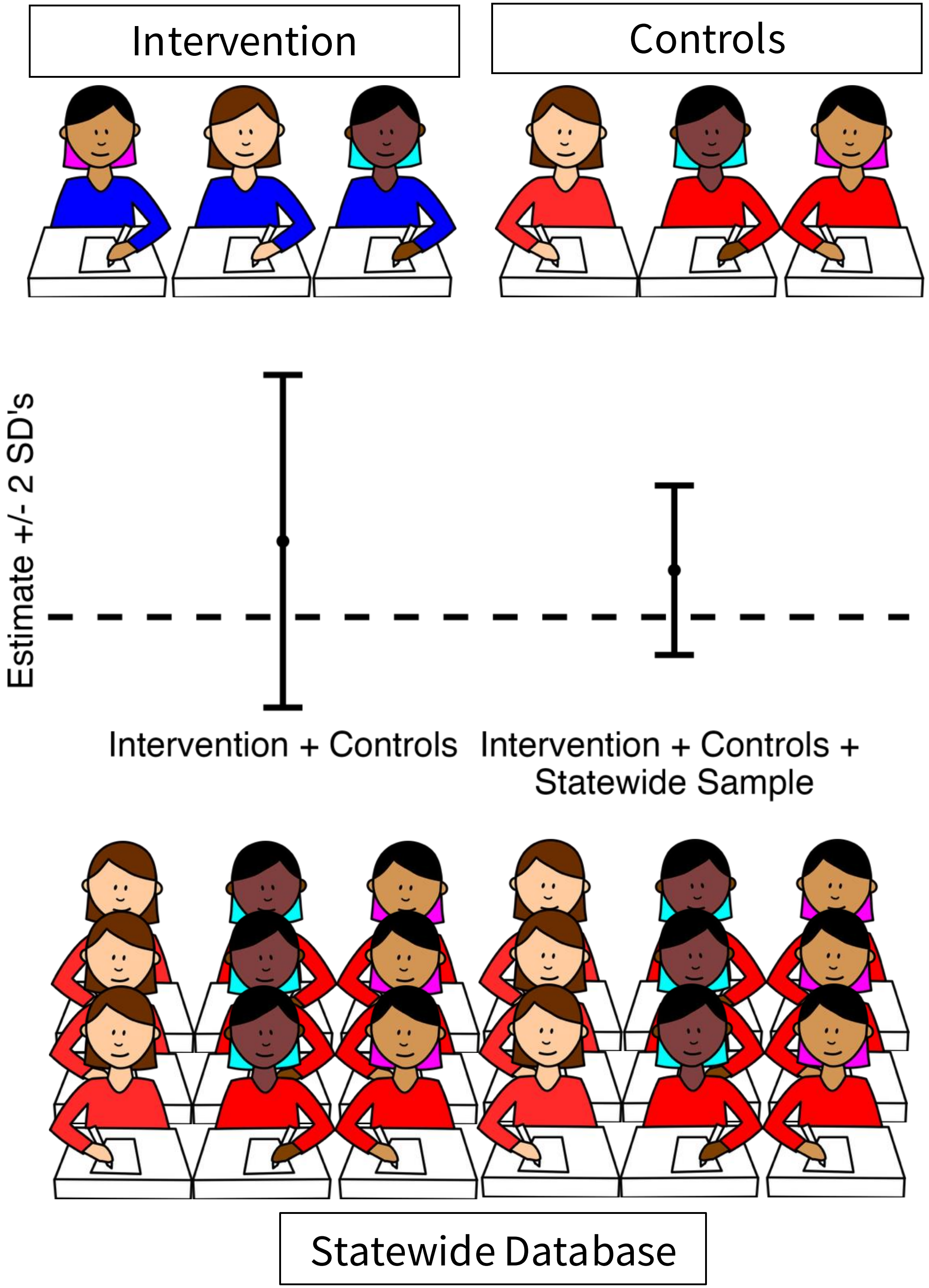
Separating covariance adjustment from causal effect estimation to leverage auxiliary datasets

Joshua Wasserman and Ben B. Hansen

Department of Statistics, University of Michigan



How does one typically adjust for covariates?

Analysis of covariance (ANCOVA):

$$Y_i(j) = \vec{x}_i\beta + a_i\tau_j + \epsilon_i$$

$$\begin{aligned}
Y_i(j) &= \text{potential outcome when assigned to group } j \\
a_i &= \text{indicator for assignment to group } j \\
\vec{x}_i &= \text{vector of exogenous pre-assignment covariates} \\
\epsilon_i &= \text{mean-zero independent noise}
\end{aligned}$$

Imputation estimators for average intent-to-treat (ITT) effects

$$\boxed{\hat{\tau}_j = \overline{\tilde{Y}(j)}_{\{a_i=j\}} - \overline{\tilde{Y}(0)}_{\{a_i=0\}}}$$

$$\overline{\tilde{Y}(j)}_{\{a_i=j\}} = \frac{\sum_{a_i=j} w_i \tilde{Y}_i(j)}{\sum_{a_i=j} w_i}$$

$$\tilde{Y}_i(j) = Y_i(j) - f(\vec{x}_i; \hat{\beta})$$

$$w_i = \text{design weight, i.e. inverse probability of assignment weight}$$

$\hat{\tau}_j$ is the difference in average difference of outcomes and predictions from a **covariance model** f

Proposed average ITT effect estimator

1. **Fit covariance model** to a sample comprising study units, $\mathcal{Q}$, and auxiliary units, $\mathcal{C}$, using weights if needed
2. **Generate model predictions** for units in $\mathcal{Q}$
3. **Difference outcomes and predictions** for units in $\mathcal{Q}$
4. **Difference the average difference** between intervention group j and the control group using inverse probability of assignment weights to produce $\hat{\tau}_j$

Theorem 1. *Under regularity conditions,* $\hat{\tau}_j \xrightarrow{P} \tau_j^*$*, where* $\tau_j^* = \mathbb{E}[Y(j) - Y(0)]$.

Conventional variance estimation for average ITT effect estimates

$$\boxed{\widehat{\mathrm{Var}}_{\mathrm{ney}}(\hat{\tau}_j) = \widehat{\mathrm{Var}}(\overline{\tilde{Y}}_j) + \widehat{\mathrm{Var}}(\overline{\tilde{Y}}_0) = \frac{\widehat{\mathrm{Var}}(\tilde{Y}_i(j))}{n_j} + \frac{\widehat{\mathrm{Var}}(\tilde{Y}_i(0))}{n_0}}$$

$$\widehat{\mathrm{Var}}(\tilde{Y}_i(j)) = \frac{\sum_{i \in \mathcal{Q}, a_i = j} (Y_i(j) - f(\vec{x}_i; \hat{\beta}))^2}{n_j - 1}$$

$$n_j = \sum_{i \in \mathcal{Q}} \mathbf{1}(a_i = j)$$

Proposed estimator for $\mathrm{Var}(\hat{\tau}_j)$

Let $\tilde{\tau}_j$ be the linear approximation to $\hat{\tau}_j$ at β^*, the true coefficients of the covariance model (which includes an intercept).

$$
\begin{aligned}
\mathrm{Var}(\tilde{\tau}_j) = \mathrm{Var}(\hat{\tau}_j) \\
&+ \nabla_\beta \mathbb{E}\left[\overline{f(\vec{x}_i;\beta)}_j - \overline{f(\vec{x}_i;\beta)}_0\right]\Big|_{\beta=\beta^*} \mathrm{Cov}\left(\hat{\beta}-\beta^*\right) \nabla_\beta \mathbb{E}\left[\overline{f(\vec{x}_i;\beta)}_j - \overline{f(\vec{x}_i;\beta)}_0\right]\Big|_{\beta=\beta^*}^T \\
&- 2 \cdot \mathrm{Cov}\left(\bar{\tilde{Y}}_j - \bar{\tilde{Y}}_0, \nabla_\beta \mathbb{E}\left[\overline{f(\vec{x}_i;\beta)}_j - \overline{f(\vec{x}_i;\beta)}_0\right]\Big|_{\beta=\beta^*} (\hat{\beta}-\beta^*)\right)
\end{aligned}
$$

Proposed estimator for $\mathrm{Var}(\hat{\tau}_j)$

Let $\tilde{\tau}_j$ be the linear approximation to $\hat{\tau}_j$ at β^*, the true coefficients of the covariance model (which includes an intercept).

$$\begin{aligned}\mathrm{Var}(\tilde{\tau}_j) &= \mathrm{Var}(\hat{\tau}_j) \\ &+ \nabla_\beta \mathbb{E}\left[\overline{f(\vec{x}_i;\beta)}_j - \overline{f(\vec{x}_i;\beta)}_0\right]\Big|_{\beta=\beta^*} \mathrm{Cov}\left(\hat{\beta}-\beta^*\right) \nabla_\beta \mathbb{E}\left[\overline{f(\vec{x}_i;\beta)}_j - \overline{f(\vec{x}_i;\beta)}_0\right]\Big|_{\beta=\beta^*}^T \\ &- 2 \cdot \mathrm{Cov}\left(\overline{\tilde{Y}}_j - \overline{\tilde{Y}}_0, \nabla_\beta \mathbb{E}\left[\overline{f(\vec{x}_i;\beta)}_j - \overline{f(\vec{x}_i;\beta)}_0\right]\Big|_{\beta=\beta^*} (\hat{\beta}-\beta^*)\right)\end{aligned}$$

Theorem 2. *Let $\theta = (\beta \quad \tau)$, where $\beta \in \mathbb{R}^p$ and $\tau = (\tau_1 \quad \dots \tau_K)$. Under regularity conditions, $\sqrt{n}(\hat{\theta} - \theta^*) \xrightarrow{d} \mathcal{N}(0, \Sigma)$, where $\Sigma_{p+j,p+j} = Var(\tilde{\tau}_j)$.*

Proposed estimator for $\mathrm{Var}(\hat{\tau}_j)$

Let $\tilde{\tau}_j$ be the linear approximation to $\hat{\tau}_j$ at β^*, the true coefficients of the covariance model (which includes an intercept).

$$
\begin{aligned}
\mathrm{Var}(\tilde{\tau}_j) = \mathrm{Var}(\hat{\tau}_j) & \\
& + \nabla_\beta \mathbb{E}\left[\overline{f(\vec{x}_i;\beta)}_j - \overline{f(\vec{x}_i;\beta)}_0\right]\Big|_{\beta=\beta^*} \mathrm{Cov}\left(\hat{\beta}-\beta^*\right) \nabla_\beta \mathbb{E}\left[\overline{f(\vec{x}_i;\beta)}_j - \overline{f(\vec{x}_i;\beta)}_0\right]\Big|_{\beta=\beta^*}^T \\
& - 2 \cdot \mathrm{Cov}\left(\bar{\tilde{Y}}_j - \bar{\tilde{Y}}_0, \nabla_\beta \mathbb{E}\left[\overline{f(\vec{x}_i;\beta)}_j - \overline{f(\vec{x}_i;\beta)}_0\right]\Big|_{\beta=\beta^*} (\hat{\beta}-\beta^*)\right)
\end{aligned}
$$

Theorem 3. *Under the conditions of Theorems 1 and 2 and the assumptions that* $f(\vec{x}_i;\beta) = \mathbb{E}[Y_i(0)|\vec{x}_i]$ *and* $\mathbb{E}[Y_i(j) - Y_i(0)] = \tau_j$ *for all* $i \in \mathcal{Q}$, $\widehat{Var}(\tilde{\tau}_j) \xrightarrow{P} Var(\tilde{\tau}_j)$.

Application: Cognitive Tutor study (Pane et al., 2014)

The data we have

- 14 high schools from one Michigan county matched into 4 pairs and 2 triplets
- 1-2 schools per block randomized to teach algebra course with additional computerized tutoring sessions
- Effects measured using school-level aggregates of standardized test scores
- School-level covariates for adjustment:
  - Demographics: breakdowns by race/ethnicity, gender, and FRPL eligibility; locale/urbanicity
  - Education: school type; Title I status; charter status; magnet status; pupil/teacher ratio; one year of prior average scaled scores and proficiency rates on math and reading standardized tests

Application: Cognitive Tutor study (Pane et al., 2014)

6 inverse probability of assignment-weighted (IPW) ITT effect estimates:

1. No covariance adjustment
2. ANCOVA
3. Imputation estimator fitting covariance model to the study sample
4. Imputation estimator fitting covariance model to high schools in the county
5. Imputation estimator fitting covariance model to high schools in the state
6. Imputation estimator fitting high-dimensional covariance model to high schools in the state

Application: Cognitive Tutor study (Pane et al., 2014)

Variance estimators:

- For estimator 1: Neyman variance estimate
- For estimator 2: HC2-corrected sandwich estimate
- For estimators 3-6:
    - $\widehat{\mathrm{Var}}_{\mathrm{ney}}(\hat{\tau}_j)$
    - $\widehat{Var}(\tilde{\tau}_j)$ with HC1 correction
    - $\widehat{Var}(\tilde{\tau}_j)$ with HC2 correction

Application: Cognitive Tutor study (Pane et al., 2014)

Covariance Sample	Model Complexity	Variance Estimator	SE($\hat{\tau}$)
Statewide	High	HC1	1.817
		HC2	1.976
		$\widehat{\text{Var}}_{\text{ney}}(\hat{\tau})$	1.888
Statewide	Low	HC1	4.806
		HC2	5.248
		$\widehat{\text{Var}}_{\text{ney}}(\hat{\tau})$	5.253
County	Low	HC1	2.021
		HC2	2.179
		$\widehat{\text{Var}}_{\text{ney}}(\hat{\tau})$	2.207
Assignment Sample	Low	HC1	0.988
		HC2	1.017
		$\widehat{\text{Var}}_{\text{ney}}(\hat{\tau})$	1.397
ANCOVA	Low	HC2	2.188
No Covariance Adjustment	-	$\widehat{\text{Var}}_{\text{ney}}(\hat{\tau})$	8.043

TABLE 1. Standard error estimates for various estimates of the PAITT of Cognitive Tutor after the first study year.

Look for the “propertee” package on CRAN soon!

Use this method with the R package “propertee”

1. Fit a covariance model using lm, glm, lmer, lmrob, or glmrob:

```
cmod <- model_fitting_fn(model_form, cmod_data, weights = optional_weights)
```

2. Specify information about the study design

```
des <- rct_design(assignment_col ~ unitid(assignment_unit_col) + block(block_col),
                  study_data)
```

3. Combine covariate adjustment and IPW weights to estimate ITT effects

```
itt_est <- lmitt(outcome_col ~ 1, study_data, design = des,
                 weights = ate(des), offset = cov_adj(cmod))
```

Application: Cognitive Tutor study (Pane et al., 2014)

Original study

- 74 middle and 73 high schools matched into pairs and triplets
- 1-2 schools per block randomized to teach algebra course with additional computerized tutoring sessions
- Effects measured using algebra-specific pre- and post-tests
- Student-level covariates for adjustment:
  - Demographics: race/ethnicity; gender; socioeconomic status; free- or reduced-price lunch (FRPL) eligibility
  - Education: English-language learner; special education/gifted program participant; two years of prior standardized test scores

Application: Cognitive Tutor study (Pane et al., 2014)

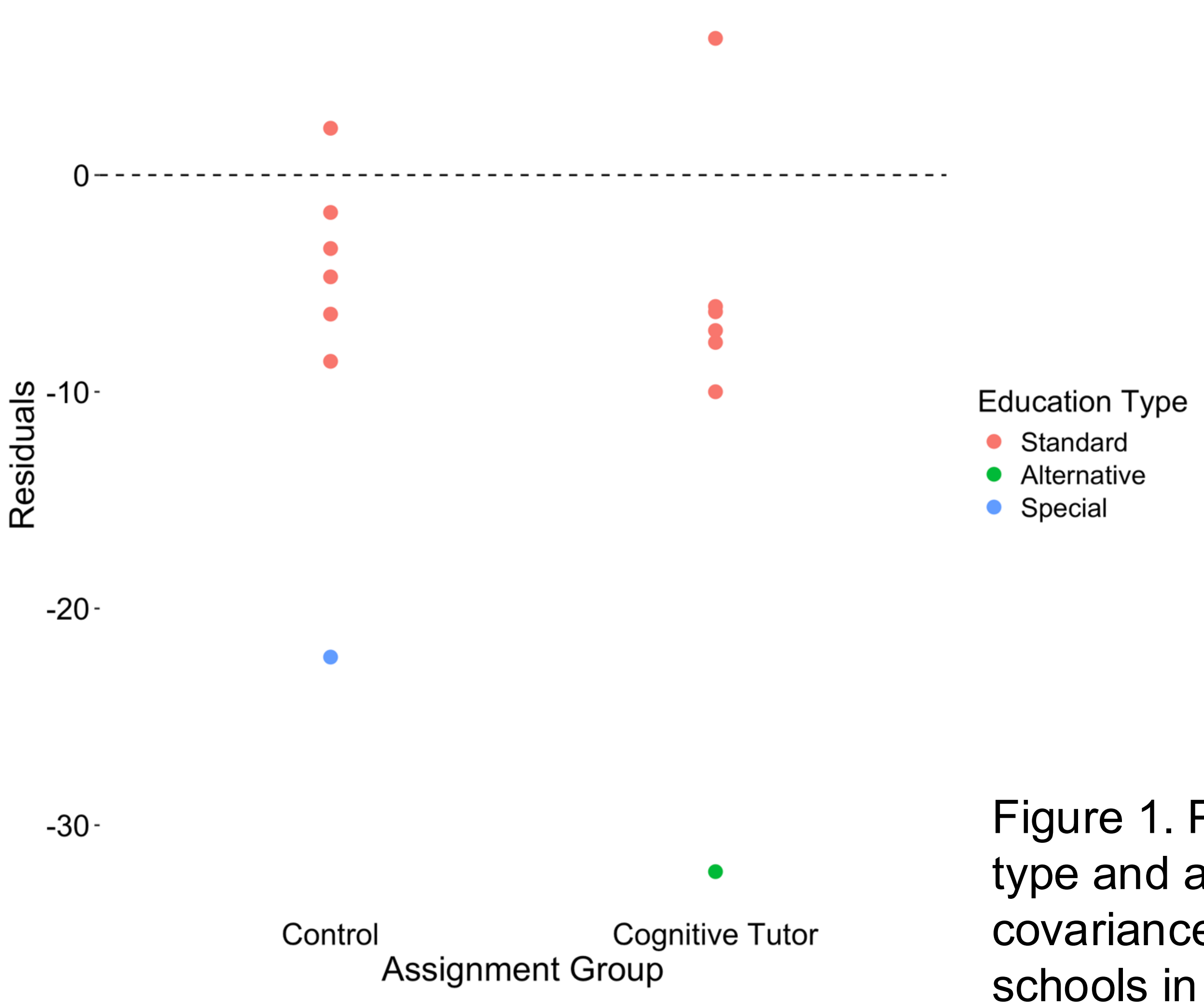


Figure 1. Partial residuals by school type and assignment group when covariance model is fit to high schools in the state.

Application: Cognitive Tutor study (Pane et al., 2014)

Covariance Sample	Variance Estimator	$\mathrm{SE}(\hat{\tau})$
Statewide	HC1	2.495
	HC2	2.69
	$\widehat{\mathrm{Var}}_{\mathrm{ney}}(\hat{\tau})$	2.562
County	HC1	1.823
	HC2	1.929
	$\widehat{\mathrm{Var}}_{\mathrm{ney}}(\hat{\tau})$	2.207
Assignment Sample	HC1	0.424
	HC2	0.401
	$\widehat{\mathrm{Var}}_{\mathrm{ney}}(\hat{\tau})$	0.908
ANCOVA	HC2	N/A

TABLE 2. Standard error estimates when the parsimonious covariance adjustment model includes an indicator variable for education type.