The Hausman-Taylor Estimator

Estimating Time-Invariant Effects with Endogeneity

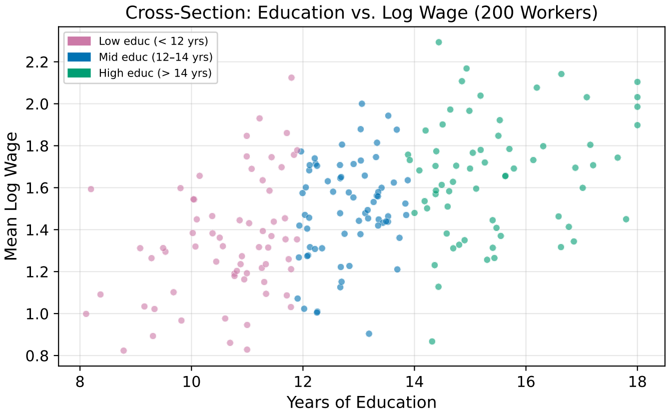
Jake Anderson

March 3, 2026

Outline

The Data

A policymaker wants to know the return to one more year of schooling. We have panel data on **200 workers** over **5 years**. How hard can it be?

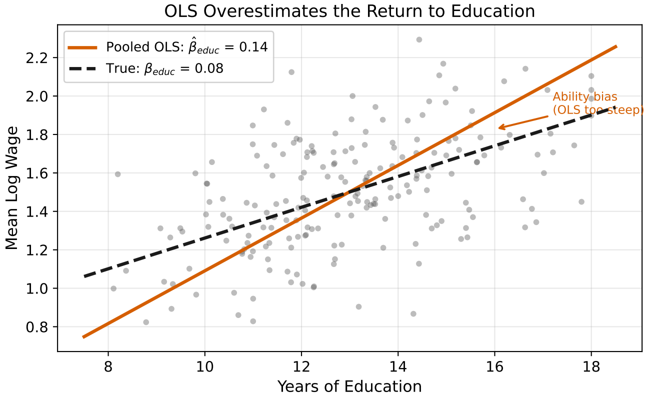


Higher education is associated with higher wages. But is the relationship **causal**?

Pooled OLS: A First Pass

Run pooled OLS of log wages on education, experience, union, and race:

$$\log(\text{wage}_{it}) = \beta_0 + \beta_1 \text{educ}_i + \beta_2 \text{exper}_{it} + \beta_3 \text{union}_{it} + \beta_4 \text{black}_i + \varepsilon_{it}$$



OLS estimates $\hat{\beta}_{\text{educ}} \approx 0.14$, but the true return is 0.08. Why the overestimate?

The Endogeneity Problem: Omitted Ability

Unobserved **ability** (α_i) affects both wages and education:

- High-ability workers earn more $\implies$ α_i belongs in the wage equation
- High-ability workers get more education $\implies$ $\text{Cov}(\alpha_i, \text{educ}_i) > 0$

OVB formula (single-regressor version):

$$\hat{\beta}_{\text{educ}}^{\text{OLS}} = \underbrace{\beta_{\text{educ}}}_{\text{true}} + \underbrace{\beta_{\alpha}}_{\substack{\text{effect of ability} \\ \text{on wages}}} \times \underbrace{\delta_1}_{\substack{\text{relationship of} \\ \text{ability to educ}}}$$

Both $\beta_\alpha > 0$ and $\delta_1 > 0$ $\implies$ OLS is **biased upward**.

Fixed Effects to the Rescue?

We know FE eliminates time-invariant unobservables. Write the model as:

$$\log(\text{wage}_{it}) = \alpha_i + \beta_1 \text{educ}_i + \beta_2 \text{exper}_{it} + \beta_3 \text{union}_{it} + \beta_4 \text{black}_i + \varepsilon_{it}$$

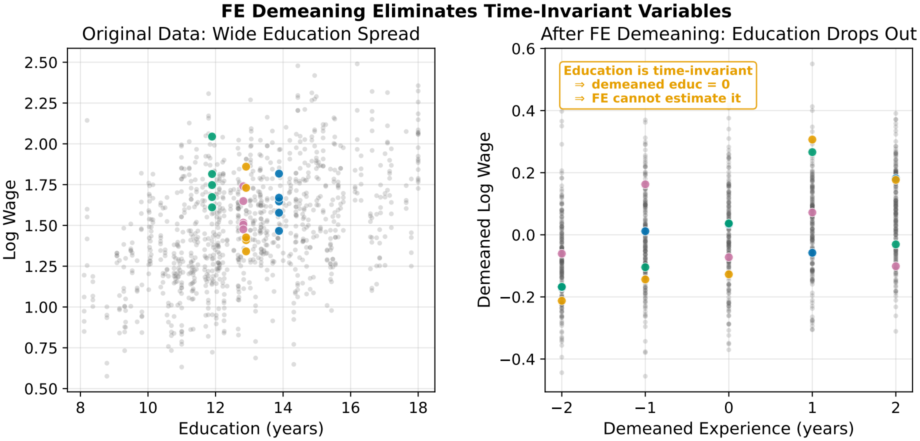
If we demean the model, which variables survive?

FE subtracts the worker mean from each variable:

$$\log(\text{wage}_{it}) - \overline{\log(\text{wage})}_i = \beta_2(\text{exper}_{it} - \overline{\text{exper}}_i) + \beta_3(\text{union}_{it} - \overline{\text{union}}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

α_i is gone. But so are $\beta_1 \text{educ}_i$ and $\beta_4 \text{black}_i$ (since educ_i does not vary over time, $\text{educ}_i - \overline{\text{educ}}_i = 0$).

$\implies$ **Education is time-invariant.** Demeaning removes it entirely.



Each worker has **one education level** across all 5 years. After demeaning, the within-worker variation in education is exactly zero. FE cannot estimate β_{educ}.

Random Effects: Can It Estimate Education?

RE treats α_i as a random draw, uncorrelated with the regressors:

$$\log(\text{wage}_{it}) = \beta_0 + \beta_1 \text{educ}_i + \beta_2 \text{exper}_{it} + \beta_3 \text{union}_{it} + \beta_4 \text{black}_i + \alpha_i + \varepsilon_{it}$$

RE uses both within and between variation $\implies$ it **can** estimate β_{educ}.

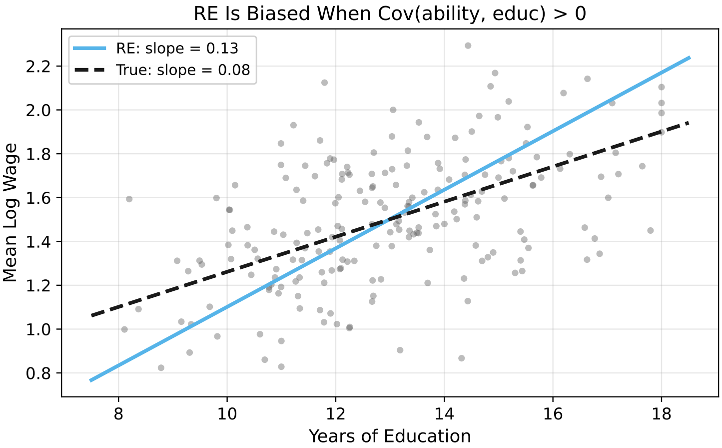
Will RE give us a better estimate than OLS?

RE requires:

$$\text{Cov}(\alpha_i, \text{educ}_i) = 0$$

We just argued this is **violated** (ability correlates with education). So the RE estimate is **biased**.

RE Is Biased



RE estimates $\hat{\beta}_{\text{educ}} \approx 0.13$ (true $= 0.08$). The ability bias persists because RE does not fully eliminate α_i from the estimation.

The Dilemma

Abbreviations: TV = time-varying; TI = time-invariant.

Method	**Estimate $\hat{\beta}_{\text{educ}}$**	**Consistent?**	**Problem**
OLS	≈ 0.14	No	Omits ability
FE	N/A		Cannot estimate TI vars
RE	≈ 0.13	No	Requires $\text{Cov}(\alpha_i, x) = 0$
True	**0.08**		

- FE is consistent for time-varying coefficients, but **drops** education
- RE can estimate education, but is **biased** when $\text{Cov}(\alpha_i, \text{educ}_i) \neq 0$

The paradox: the only estimator that controls for ability (FE) is the one that cannot estimate the coefficient we care about.

$\Longrightarrow$ We need an estimator that handles **time-invariant endogenous** variables.

Can We Recycle What FE Gave Us?

FE gave us consistent estimates of the time-varying coefficients ($\hat{\beta}_{\text{exper}}$, $\hat{\beta}_{\text{union}}$). Those estimates produced residuals. And the variables themselves vary across individuals.

What if we could *reuse* those time-varying exogenous variables as instruments for the time-invariant endogenous variable?

Think about it:

- Experience and union status are uncorrelated with ability (exogenous)
- But their **worker-level averages** are correlated with education (workers who studied longer entered the workforce later $\implies$ less experience)

$\implies$ We already have instruments sitting inside the panel. We just need a procedure that uses them.

Outline

The Hausman-Taylor Idea

Hausman and Taylor (1981) proposed using information *already in the panel* to construct instruments for the endogenous time-invariant variables.

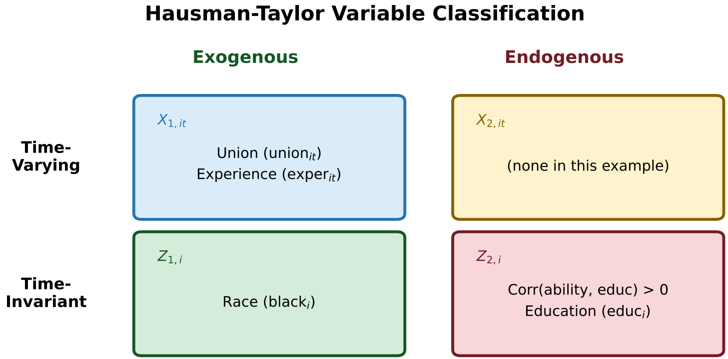
The insight:

- Some time-varying variables (like experience) are **exogenous**: $\text{Cov}(\alpha_i, \text{exper}_{it}) = 0$
- Their **within-group means** ($\overline{\text{exper}}_i$) vary across individuals
- These means are correlated with education (workers who studied longer entered the workforce later $\implies$ less experience)
- But they are **uncorrelated with ability** (by assumption)

$\implies$ The within-group means of time-varying exogenous variables are **valid instruments** for education.

Variable Classification: Four Categories

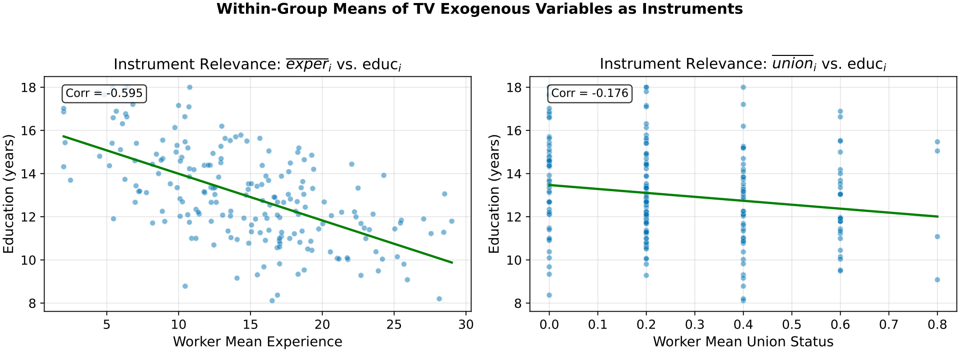
HT requires classifying every variable into one of four groups:



The classification determines which variables need instruments and which variables *provide* instruments.

The Instruments: Within-Group Means

Why do $\bar{\text{exper}}_i$ and $\bar{\text{union}}_i$ work as instruments for educ_i?



Relevance: Workers with more education have less experience (delayed entry) and lower unionization.
Validity: These means are uncorrelated with unobserved ability (exogenous by classification).

Instrument Requirements

Recall: X_1 = TV exogenous (exper, union); Z_2 = TI endogenous (educ).

For $\bar{X}_{1,i}$ (worker-level means of TV exogenous vars) to be valid instruments for $Z_{2,i}$:

1. Relevance:

$$\text{Corr}(\bar{X}_{1,i},\ Z_{2,i}) \neq 0$$

In our data: $\text{Corr}(\overline{\text{exper}}_i, \text{educ}_i) \approx -0.60$. Workers who went to school longer started working later.

2. Validity (exclusion restriction):

$$\text{Cov}(\bar{X}_{1,i},\ \alpha_i) = 0$$

This is an **assumption**, not something we can verify directly. It follows from classifying $X_{1,it}$ as exogenous: if $\text{Cov}(\alpha_i, \text{exper}_{it}) = 0$ for all t, then $\text{Cov}(\alpha_i, \overline{\text{exper}}_i) = 0$.

When might this fail? If high-ability workers accumulate more experience through promotions or job mobility, then $\text{Cov}(\alpha_i, \text{exper}_{it}) \neq 0$ and the instruments are invalid. The maintained assumption in this example rules that out: experience here reflects labor market entry timing, not ability-driven career advancement.

$\Longrightarrow$ Same logic as any IV estimation, but the instruments come from **within the panel**.

HT Procedure: Step 1 (Within Estimation)

We need consistent time-varying coefficients so we can strip them out and isolate the between-individual variation.

Run FE on the full model. FE wipes out both α_i (good) and the time-invariant regressors (the cost):

$$\ddot{y}_{it} = \hat{\beta}^{FE}_{\text{exper}}\,\ddot{\text{exper}}_{it} + \hat{\beta}^{FE}_{\text{union}}\,\ddot{\text{union}}_{it} + \ddot{\varepsilon}_{it}$$

where $\ddot{x}_{it} = x_{it} - \bar{x}_i$ denotes the demeaned variable.

$\implies$ We now have consistent $\hat{\beta}^{FE}_{\text{exper}}$ and $\hat{\beta}^{FE}_{\text{union}}$. We still need $\hat{\beta}_{\text{educ}}$ and $\hat{\beta}_{\text{black}}$.

HT Procedure: Step 2a (Between Equation)

Now we have a cross-sectional equation where education is still present but ability is in the error.

Take worker-level averages and subtract the FE-estimated time-varying effects:

$$\underbrace{\bar{y}_i - \hat{\beta}^{FE}_{\text{exper}}\overline{\text{exper}}_i - \hat{\beta}^{FE}_{\text{union}}\overline{\text{union}}_i}_{d_i} = \beta_0 + \beta_{\text{educ}}\,\text{educ}_i + \beta_{\text{black}}\,\text{black}_i + \alpha_i + \bar{\varepsilon}_i$$

This is a **cross-sectional equation** with one observation per worker. Education is back, but the problem remains: α_i is in the error and $\text{Cov}(\alpha_i, \text{educ}_i) \neq 0$.

$\Longrightarrow$ OLS on this equation would still be biased. We need instruments.

HT Procedure: Step 2b (IV Estimation)

This is a standard IV problem: endogenous regressor (educ_i), error contains α_i.

Run 2SLS on the between equation, using $\overline{\text{exper}}_i$ and $\overline{\text{union}}_i$ as instruments for educ_i.

Why does this work?

- **Relevance:** $\overline{\text{exper}}_i$ and $\overline{\text{union}}_i$ are correlated with educ_i (through labor market timing)
- **Validity:** They are uncorrelated with α_i (by the TV exogeneity assumption)

$\Longrightarrow$ 2SLS isolates the variation in education that is driven by experience and union patterns, not by ability.

HT Procedure: Step 3 (FGLS Combination)

Combining within and between estimates optimally gives efficiency, just as RE does when its assumptions hold.

The full HT estimator uses feasible GLS to combine:

1. **Within information** (from FE): identifies time-varying coefficients
2. **Between information** (from IV): identifies time-invariant coefficients

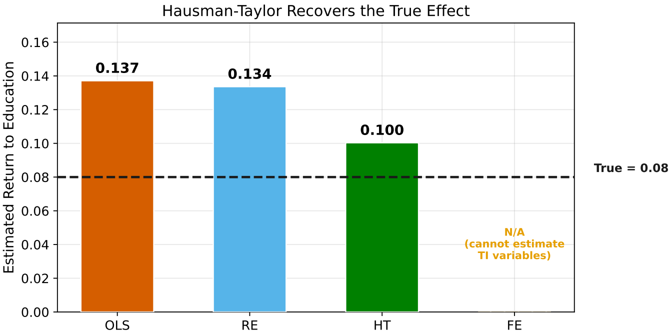
FGLS uses the variance components $(\hat{\sigma}^2_\alpha, \hat{\sigma}^2_\varepsilon)$ to weight observations: when σ^2_α is large relative to σ^2_ε, the weighting leans more heavily on the within variation (similar to FE); when it is small, the weighting uses more of the between variation.

In practice, software handles this. In Stata:

```
xthtaylor lwage exper union educ black, endog(educ) constant(educ black)
```

$\Longrightarrow$ You specify which variables are time-invariant and which are endogenous. The estimator does the rest.

The Result: HT Recovers the True Effect



OLS (≈ 0.14) and RE (≈ 0.13) are both biased upward. HT estimates ≈ 0.10, much closer to the true $\beta_{\text{educ}} = 0.08$.

$\implies$ By instrumenting education with within-group means of experience and union status, HT removes most of the ability bias.

Note: the HT estimate shown is from the IV-between step (Step 2b). The full FGLS estimator (Step 3) produces a slightly different number; Stata's `xthtaylor` implements the complete procedure.

How Each HT Component Solves a Problem

Problem	**HT Component**	**Result**
TV coefficients biased	FE (within estimation)	Consistent $\hat{\beta}_{\text{exper}}$, $\hat{\beta}_{\text{union}}$
TI endogenous variable	IV using $\bar{X}_{1,i}$ as instruments	Consistent $\hat{\beta}_{\text{educ}}$
Combine efficiently	GLS weighting	Efficient estimates

Each step addresses a specific failure of the standard estimators:

- FE handles TV coefficients (consistent) ✓
- IV handles TI endogenous variables (via internal instruments) ✓
- GLS combines them efficiently ✓

$\Longrightarrow$ HT fills the gap between FE (drops TI variables) and RE (requires full exogeneity).

Outline

When Does HT Work?

HT requires three conditions:

1. **Correct classification.** You must correctly identify which variables are exogenous and which are endogenous. If a "time-varying exogenous" variable is actually endogenous, the instruments are invalid.
2. **Instrument relevance.** The within-group means of TV exogenous variables must be correlated with the TI endogenous variables. If $\text{Corr}(\bar{X}_{1,i}, Z_{2,i}) \approx 0$, the instruments are weak and HT will perform poorly.
3. **Enough instruments.** You need at least as many TV exogenous variables as TI endogenous variables (order condition). More instruments enable overidentification testing.

In our example: 2 TV exogenous (exper, union) instrumenting 1 TI endogenous (educ) $\implies$ overidentified (good: we can test validity).

Formal Assumptions

Partition the variables:

- $X_{1,it}$: time-varying, exogenous (exper, union)
- $X_{2,it}$: time-varying, endogenous (none here)
- $Z_{1,i}$: time-invariant, exogenous (black)
- $Z_{2,i}$: time-invariant, endogenous (educ)

The Hausman-Taylor assumptions:

1. $\mathrm{E}[\varepsilon_{it} \mid X_1, X_2, Z_1, Z_2, \alpha_i] = 0$ (strict exogeneity of idiosyncratic error)

   The idiosyncratic error is pure noise.

2. $\mathrm{E}[\alpha_i \mid X_1, Z_1] = 0$ (α_i uncorrelated with exogenous variables)

   Ability is unrelated to experience, union status, and race.

3. $\mathrm{E}[\alpha_i \mid X_2, Z_2] \neq 0$ (endogeneity of X_2, Z_2)

   Ability IS related to education. This is why we need HT.

$\Longrightarrow$ Assumption 2 is what makes $\bar{X}_{1,i}$ a valid instrument: it inherits exogeneity from $X_{1,it}$.

Testing HT: Overidentification

When you have more TV exogenous variables than TI endogenous variables, the model is **overidentified**.

This allows a **Sargan/Hansen test** of instrument validity:

- H_0: instruments are valid (uncorrelated with α_i)
- H_1: at least one instrument is invalid

In our example:

- 2 instruments ($\overline{\text{exper}}_i$, $\overline{\text{union}}_i$) for 1 endogenous variable (educ_i)
- Degrees of overidentification $= 2 - 1 = 1$
- Test statistic $\sim \chi^2(1)$ under H_0

$\Longrightarrow$ Rejection suggests that at least one of experience or union is not truly exogenous to ability. This would invalidate the HT approach.

Hausman Test: HT vs. FE

For **time-varying** coefficients, both FE and HT are consistent (under their respective assumptions). We can compare them:

$$H_0 : \mathrm{Cov}(\alpha_i, X_{1,it}) = 0 \quad \text{(necessary for HT consistency)}$$

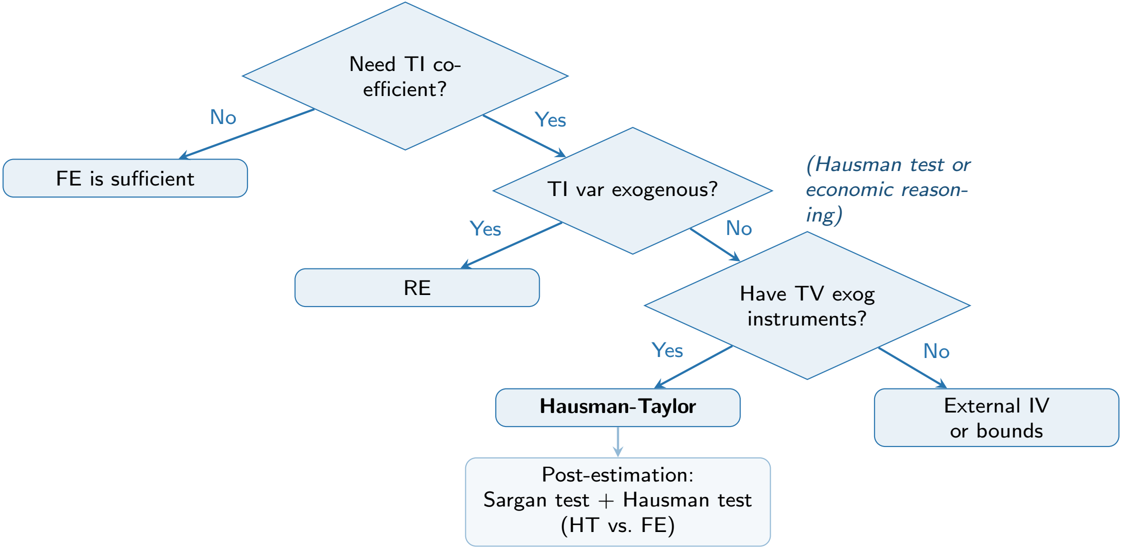
The Hausman test compares $\hat{\beta}^{FE}_{\mathrm{TV}}$ and $\hat{\beta}^{HT}_{\mathrm{TV}}$:

- If they are "close," the HT assumptions are not rejected
- If they diverge, the classification of exogenous variables may be wrong

$\Longrightarrow$ This tests whether the TV variables you classified as exogenous truly are. It does *not* test the validity of the TI classification.

Outline

Decision Flowchart



Outline

Summary

We started with a panel of 200 workers and wanted to estimate the return to education. OLS and RE both overestimated it because unobserved ability is correlated with education. FE eliminates ability but also eliminates education.

1. **FE** removes time-invariant unobservables but **cannot estimate** coefficients on time-invariant regressors.
2. **RE** can estimate time-invariant coefficients but requires all regressors to be uncorrelated with α_i. When this fails, RE is biased.
3. **Hausman-Taylor** uses within-group means of TV exogenous variables as instruments for TI endogenous variables. It fills the gap between FE and RE.
4. HT requires **correct variable classification**, **relevant instruments**, and at least as many TV exogenous vars as TI endogenous vars.
5. Use the **Sargan/Hansen test** (overidentification) and the **Hausman test** (HT vs. FE) to check the assumptions.

$\Longrightarrow$ When you need a time-invariant coefficient but suspect endogeneity, check whether your TV exogenous variables can serve as internal instruments.

Thank you!

jakeanderson@g.ucla.edu