The Heckman Selection Model

When Your Data Only Includes People Who Showed Up

Jake Anderson

March 3, 2026

Outline

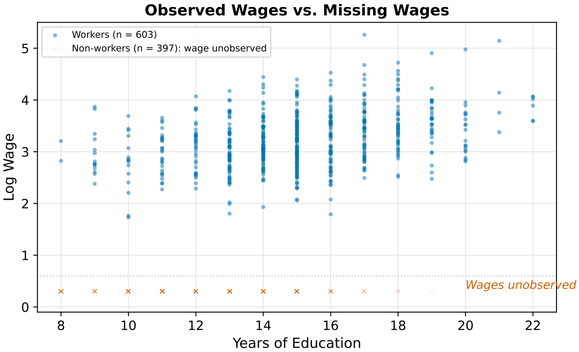
The Data: Wages and Education

A labor economist surveys **1,000 adults** and records their hourly wage, years of education, work experience, number of children, and spouse's income.

The Data: Wages and Education

A labor economist surveys **1,000 adults** and records their hourly wage, years of education, work experience, number of children, and spouse's income.

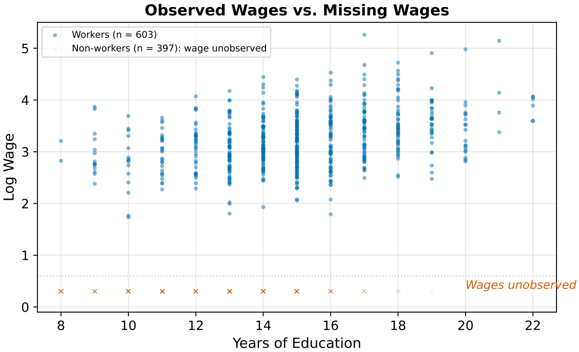
She wants to estimate the **return to education**: how much does one more year of schooling raise log wages? There is a catch: **only 603 people work**. The other 397 have no wage to observe.



The Data: Wages and Education

A labor economist surveys **1,000 adults** and records their hourly wage, years of education, work experience, number of children, and spouse's income.

She wants to estimate the **return to education**: how much does one more year of schooling raise log wages? There is a catch: **only 603 people work**. The other 397 have no wage to observe.



Blue dots: workers with observed wages. Orange crosses: non-workers with *no wage data*.

Who Are the Non-Workers?

The non-workers are not a random sample of the population:

Who Are the Non-Workers?

The non-workers are not a random sample of the population:

	Workers (n = 603)	**Non-Workers** (n = 397)
Mean education (years)	15.0	12.3
Mean children	0.9	1.6
Mean spouse income ($1000s)	36.2	44.7

Who Are the Non-Workers?

The non-workers are not a random sample of the population:

	Workers (n = 603)	**Non-Workers** (n = 397)
Mean education (years)	15.0	12.3
Mean children	0.9	1.6
Mean spouse income ($1000s)	36.2	44.7

Non-workers have **less education**, **more children**, and **higher-earning spouses**. The people with missing wages are systematically different from those with observed wages.

Who Are the Non-Workers?

The non-workers are not a random sample of the population:

	Workers ($n = 603$)	**Non-Workers** ($n = 397$)
Mean education (years)	15.0	12.3
Mean children	0.9	1.6
Mean spouse income ($1000s)	36.2	44.7

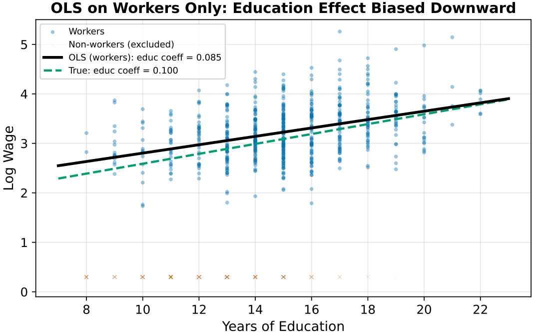
Non-workers have **less education**, **more children**, and **higher-earning spouses**. The people with missing wages are systematically different from those with observed wages.

$\Longrightarrow$ This is **sample selection**: the decision to work is not random, and it correlates with the outcome we care about (wages).

OLS on Workers Only: Biased

Ignoring the missing data and running OLS on the 603 workers:

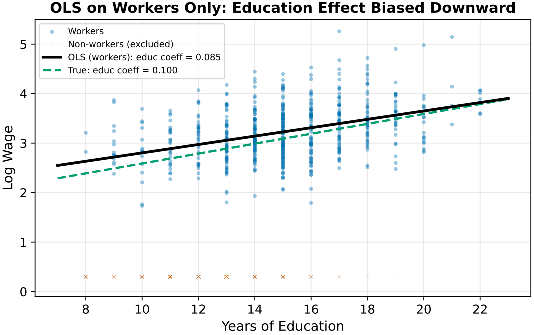
$$\widehat{\log(\text{wage})}_i = 1.373 + \underbrace{0.085}_{\substack{\text{biased downward} \\ \text{in this setting}}} \cdot \text{educ}_i + 0.040 \cdot \text{exper}_i$$



OLS on Workers Only: Biased

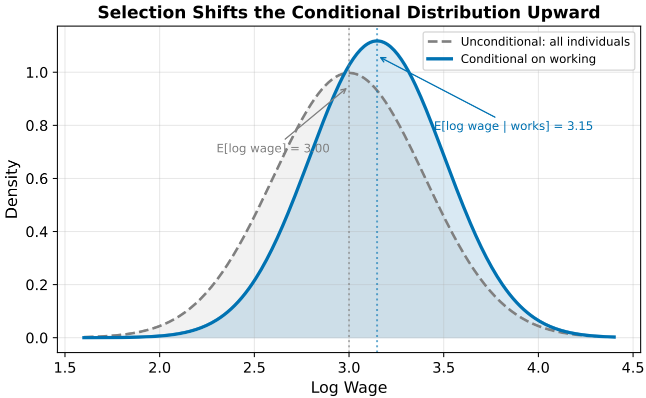
Ignoring the missing data and running OLS on the 603 workers:

$$\widehat{\log(\text{wage})}_i = 1.373 + \underbrace{0.085}_{\substack{\text{biased downward} \\ \text{in this setting}}} \cdot \text{educ}_i + 0.040 \cdot \text{exper}_i$$

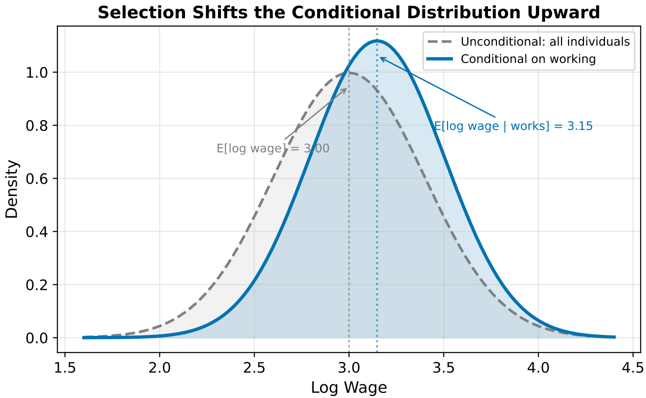


The OLS education coefficient is **0.085** (true $= 0.100$). That is a 15% underestimate. Why?

What OLS Misses: The Conditional Distribution

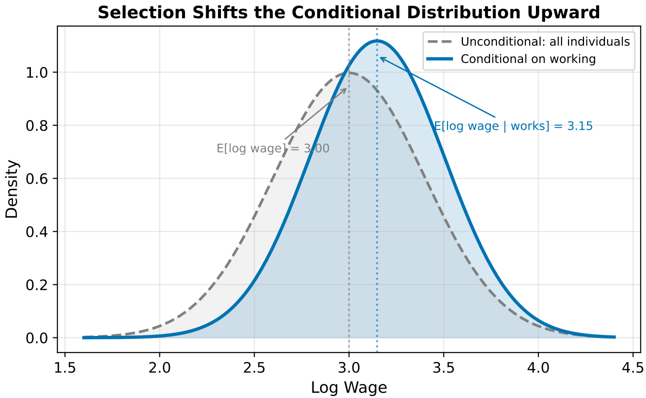


What OLS Misses: The Conditional Distribution



Gray: wage distribution for *all* individuals. Blue: *conditional on working*. Selection shifts the distribution to the right.

What OLS Misses: The Conditional Distribution



Gray: wage distribution for *all* individuals. Blue: *conditional on working*. Selection shifts the distribution to the right.

OLS on workers fits a line through the blue distribution, but the true regression line passes through the gray one. The shift gets absorbed into the intercept and correlated coefficients, biasing them.

What If We Include Non-Workers?

One natural idea: assign wage $= 0$ (or $\log(\text{wage}) = 0$) to non-workers and run OLS on all 1,000 people.

What If We Include Non-Workers?

One natural idea: assign wage $= 0$ (or $\log(\text{wage}) = 0$) to non-workers and run OLS on all 1,000 people.

This is also wrong:

- Non-workers do not have a wage of zero. They have a **missing** wage. Imputing zero creates artificial data points that pull the regression line toward zero for people with low education

What If We Include Non-Workers?

One natural idea: assign wage $= 0$ (or $\log(\text{wage}) = 0$) to non-workers and run OLS on all 1,000 people.

This is also wrong:

- Non-workers do not have a wage of zero. They have a **missing** wage. Imputing zero creates artificial data points that pull the regression line toward zero for people with low education
- The resulting slope reflects a mix of two relationships: the true effect of education on wages *and* the effect of education on whether someone works at all

What If We Include Non-Workers?

One natural idea: assign wage $= 0$ (or $\log(\text{wage}) = 0$) to non-workers and run OLS on all 1,000 people.

This is also wrong:

- Non-workers do not have a wage of zero. They have a **missing** wage. Imputing zero creates artificial data points that pull the regression line toward zero for people with low education
- The resulting slope reflects a mix of two relationships: the true effect of education on wages *and* the effect of education on whether someone works at all
- If you set missing wages to any fixed number, you change the distribution of the dependent variable in a way that depends on the selection process

What If We Include Non-Workers?

One natural idea: assign wage $= 0$ (or $\log(\text{wage}) = 0$) to non-workers and run OLS on all 1,000 people.

This is also wrong:

- Non-workers do not have a wage of zero. They have a **missing** wage. Imputing zero creates artificial data points that pull the regression line toward zero for people with low education
- The resulting slope reflects a mix of two relationships: the true effect of education on wages *and* the effect of education on whether someone works at all
- If you set missing wages to any fixed number, you change the distribution of the dependent variable in a way that depends on the selection process

$\Longrightarrow$ Neither dropping non-workers (OLS on workers) nor imputing values fixes the problem. We need a model that **explicitly accounts for the selection process**.

The Basketball Player Analogy

Suppose you want to estimate the effect of **height on free-throw accuracy** in the general population.

The Basketball Player Analogy

Suppose you want to estimate the effect of **height on free-throw accuracy** in the general population.

But you only observe people who **play basketball**. Who plays? Mostly tall people. And among tall people, the ones who play are not randomly selected: they are the ones with basketball talent, which also helps free throws.

The Basketball Player Analogy

Suppose you want to estimate the effect of **height on free-throw accuracy** in the general population.

But you only observe people who **play basketball**. Who plays? Mostly tall people. And among tall people, the ones who play are not randomly selected: they are the ones with basketball talent, which also helps free throws.

The result: among basketball players, height looks less important for free throws, because everyone is already tall and talented. The height effect is attenuated by selection into the sample.

The Basketball Player Analogy

Suppose you want to estimate the effect of **height on free-throw accuracy** in the general population.

But you only observe people who **play basketball**. Who plays? Mostly tall people. And among tall people, the ones who play are not randomly selected: they are the ones with basketball talent, which also helps free throws.

The result: among basketball players, height looks less important for free throws, because everyone is already tall and talented. The height effect is attenuated by selection into the sample.

$\implies$ Our wage data has the same problem. Workers are not a random draw: they are the people whose unobserved characteristics (motivation, ability) pushed them into the labor force. These same characteristics also affect wages.

What Would a Better Model Need?

OLS on workers fails because it ignores the selection process. A better approach should:

What Would a Better Model Need?

OLS on workers fails because it ignores the selection process. A better approach should:

1. **Model the selection:** why some people work and others do not. The non-workers carry information about the selection process, even though they have no wages

What Would a Better Model Need?

OLS on workers fails because it ignores the selection process. A better approach should:

1. **Model the selection:** why some people work and others do not. The non-workers carry information about the selection process, even though they have no wages
2. **Correct the wage equation:** the workers we observe are not representative. Their unobserved characteristics are systematically different from the population average

What Would a Better Model Need?

OLS on workers fails because it ignores the selection process. A better approach should:

1. **Model the selection:** why some people work and others do not. The non-workers carry information about the selection process, even though they have no wages
2. **Correct the wage equation:** the workers we observe are not representative. Their unobserved characteristics are systematically different from the population average
3. **Recover the true return to education:** the structural relationship between education and wages, free from selection bias

What Would a Better Model Need?

OLS on workers fails because it ignores the selection process. A better approach should:

1. **Model the selection:** why some people work and others do not. The non-workers carry information about the selection process, even though they have no wages
2. **Correct the wage equation:** the workers we observe are not representative. Their unobserved characteristics are systematically different from the population average
3. **Recover the true return to education:** the structural relationship between education and wages, free from selection bias

$\Longrightarrow$ We need to jointly model the wage process and the selection process. This is different from Tobit (censoring): here the **decision to work is a separate equation** from the wage itself.

Outline

Two Equations, Two Error Terms

The Heckman model has two equations, each with its own error term:

Two Equations, Two Error Terms

The Heckman model has two equations, each with its own error term:

Wage equation (outcome, only observed for workers):

$$\log(\text{wage}_i) = \underbrace{1.0}_{\beta_0} + \underbrace{0.10}_{\beta_1} \cdot \text{educ}_i + \underbrace{0.04}_{\beta_2} \cdot \text{exper}_i + u_i$$

Two Equations, Two Error Terms

The Heckman model has two equations, each with its own error term:

Wage equation (outcome, only observed for workers):

$$\log(\text{wage}_i) = \underbrace{1.0}_{\beta_0} + \underbrace{0.10}_{\beta_1} \cdot \text{educ}_i + \underbrace{0.04}_{\beta_2} \cdot \text{exper}_i + u_i$$

Selection equation (determines who works):

$$\text{work}_i^* = \underbrace{-2.5}_{\gamma_0} + \underbrace{0.30}_{\gamma_1} \cdot \text{educ}_i \underbrace{-0.50}_{\gamma_2} \cdot \text{kids}_i \underbrace{-0.02}_{\gamma_3} \cdot \text{spouse\_inc}_i + v_i$$

Two Equations, Two Error Terms

The Heckman model has two equations, each with its own error term:

Wage equation (outcome, only observed for workers):

$$\log(\text{wage}_i) = \underbrace{1.0}_{\beta_0} + \underbrace{0.10}_{\beta_1} \cdot \text{educ}_i + \underbrace{0.04}_{\beta_2} \cdot \text{exper}_i + u_i$$

Selection equation (determines who works):

$$\text{work}_i^* = \underbrace{-2.5}_{\gamma_0} + \underbrace{0.30}_{\gamma_1} \cdot \text{educ}_i \underbrace{-0.50}_{\gamma_2} \cdot \text{kids}_i \underbrace{-0.02}_{\gamma_3} \cdot \text{spouse\_inc}_i + v_i$$

The variable work_i^* is an **unobserved latent variable**: we never see it directly, only whether it crosses zero. Person i works if $\text{work}_i^* > 0$. We only observe $\log(\text{wage}_i)$ when $\text{work}_i^* > 0$.

Two Equations, Two Error Terms

The Heckman model has two equations, each with its own error term:

Wage equation (outcome, only observed for workers):

$$\log(\text{wage}_i) = \underbrace{1.0}_{\beta_0} + \underbrace{0.10}_{\beta_1} \cdot \text{educ}_i + \underbrace{0.04}_{\beta_2} \cdot \text{exper}_i + u_i$$

Selection equation (determines who works):

$$\text{work}_i^* = \underbrace{-2.5}_{\gamma_0} + \underbrace{0.30}_{\gamma_1} \cdot \text{educ}_i \underbrace{-0.50}_{\gamma_2} \cdot \text{kids}_i \underbrace{-0.02}_{\gamma_3} \cdot \text{spouse\_inc}_i + v_i$$

The variable work_i^* is an **unobserved latent variable**: we never see it directly, only whether it crosses zero. Person i works if $\text{work}_i^* > 0$. We only observe $\log(\text{wage}_i)$ when $\text{work}_i^* > 0$.

$\Longrightarrow$ If u_i and v_i are correlated ($\rho \neq 0$), the workers are a **selected** subsample, and OLS on workers is biased.

Where the Bias Comes From: Correlated Errors

The Two-Equation Framework

Wage Equation (Outcome)

$$\log(\text{wage}_i^*) = 1.0 + 0.10 \cdot \text{educ}_i + 0.04 \cdot \text{exper}_i + u_i$$

Selection Equation

$$\text{work}_i^* = -2.5 + 0.30 \cdot \text{educ}_i$$
$$-0.50 \cdot \text{kids}_i - 0.02 \cdot \text{spouse\_inc}_i + v_i$$

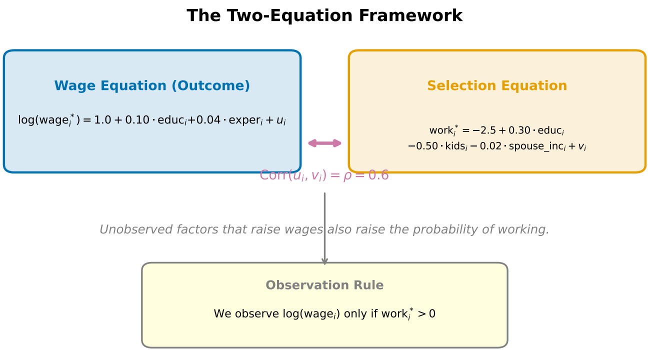
$\text{Corr}(u_i, v_i) = \rho = 0.6$

Unobserved factors that raise wages also raise the probability of working.

Observation Rule

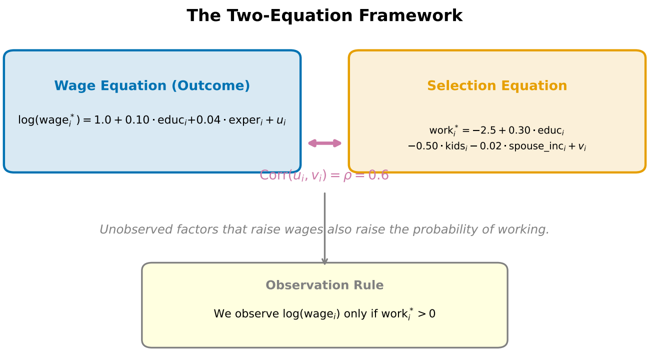
We observe $\log(\text{wage}_i)$ only if $\text{work}_i^* > 0$

Where the Bias Comes From: Correlated Errors



In our DGP, $\rho = \text{Corr}(u_i, v_i) = 0.6 > 0$. Unobserved factors that raise wages (ability, motivation) *also* make a person more likely to work. Among workers, the average u_i is **positive**, not zero.

Where the Bias Comes From: Correlated Errors



In our DGP, $\rho = \text{Corr}(u_i, v_i) = 0.6 > 0$. Unobserved factors that raise wages (ability, motivation) *also* make a person more likely to work. Among workers, the average u_i is **positive**, not zero.

$\implies$ OLS assumes $\text{E}[u_i \mid \text{works}] = 0$, but in reality $\text{E}[u_i \mid \text{works}] > 0$. This violates the zero conditional mean assumption.

The Conditional Expectation: What Does Selection Do?

What is the expected wage of a worker, given that she chose to work?

The Conditional Expectation: What Does Selection Do?

What is the expected wage of a worker, given that she chose to work?

Population regression (without conditioning on selection):

$\mathrm{E}[\log(\mathrm{wage}_i) \mid X_i] = \beta_0 + \beta_1\,\mathrm{educ}_i + \beta_2\,\mathrm{exper}_i$

The Conditional Expectation: What Does Selection Do?

What is the expected wage of a worker, given that she chose to work?

Population regression (without conditioning on selection):
$\mathrm{E}[\log(\mathrm{wage}_i) \mid X_i] = \beta_0 + \beta_1 \, \mathrm{educ}_i + \beta_2 \, \mathrm{exper}_i$

Selection means v_i was large enough for the person to work. Conditioning on this:

$$\mathrm{E}[\log(\mathrm{wage}_i) \mid \mathrm{works}] = \beta_0 + \beta_1 \, \mathrm{educ}_i + \beta_2 \, \mathrm{exper}_i + \underbrace{\mathrm{E}[u_i \mid v_i > -\gamma_0 - \gamma_1 \mathrm{educ}_i - \cdots]}_{\text{selection bias term}}$$

The Conditional Expectation: What Does Selection Do?

What is the expected wage of a worker, given that she chose to work?

Population regression (without conditioning on selection):
$\mathrm{E}[\log(\mathrm{wage}_i) \mid X_i] = \beta_0 + \beta_1\,\mathrm{educ}_i + \beta_2\,\mathrm{exper}_i$

Selection means v_i was large enough for the person to work. Conditioning on this:

$$\mathrm{E}[\log(\mathrm{wage}_i) \mid \mathrm{works}] = \beta_0 + \beta_1\,\mathrm{educ}_i + \beta_2\,\mathrm{exper}_i + \underbrace{\mathrm{E}[u_i \mid v_i > -\gamma_0 - \gamma_1 \mathrm{educ}_i - \cdots]}_{\text{selection bias term}}$$

Because u_i and v_i are correlated, knowing that someone works (v_i is large enough) tells us something about their wage error (u_i). This "something" is the selection bias term.

The Conditional Expectation: What Does Selection Do?

What is the expected wage of a worker, given that she chose to work?

Population regression (without conditioning on selection):
$E[\log(\text{wage}_i) \mid X_i] = \beta_0 + \beta_1 \, \text{educ}_i + \beta_2 \, \text{exper}_i$

Selection means v_i was large enough for the person to work. Conditioning on this:

$$E[\log(\text{wage}_i) \mid \text{works}] = \beta_0 + \beta_1 \, \text{educ}_i + \beta_2 \, \text{exper}_i + \underbrace{E[u_i \mid v_i > -\gamma_0 - \gamma_1 \text{educ}_i - \cdots]}_{\text{selection bias term}}$$

Because u_i and v_i are correlated, knowing that someone works (v_i is large enough) tells us something about their wage error (u_i). This "something" is the selection bias term.

$\Longrightarrow$ If we can calculate this term and include it in our regression, we remove the bias.

The Selection Index and Normal Distribution Notation

Before deriving the correction formula, we need two pieces of notation.

The Selection Index and Normal Distribution Notation

Before deriving the correction formula, we need two pieces of notation.

The selection index. Define the shorthand:

$$\mathsf{GZ}_i \equiv \gamma_0 + \gamma_1\,\mathsf{educ}_i + \gamma_2\,\mathsf{kids}_i + \gamma_3\,\mathsf{spouse\_inc}_i$$

This is the linear combination from the selection equation (everything except the error v_i). A larger GZ_i means person i is more likely to work.

The Selection Index and Normal Distribution Notation

Before deriving the correction formula, we need two pieces of notation.

The selection index. Define the shorthand:

$$\text{GZ}_i \equiv \gamma_0 + \gamma_1 \,\text{educ}_i + \gamma_2 \,\text{kids}_i + \gamma_3 \,\text{spouse\_inc}_i$$

This is the linear combination from the selection equation (everything except the error v_i). A larger GZ_i means person i is more likely to work.

Standard normal PDF and CDF. Recall from your statistics courses:

- $\phi(z)$: the standard normal **density** (PDF) evaluated at z. Bell-curve height
- $\Phi(z)$: the standard normal **cumulative distribution** (CDF) evaluated at z. Area to the left of z

The Selection Index and Normal Distribution Notation

Before deriving the correction formula, we need two pieces of notation.

The selection index. Define the shorthand:

$$\mathrm{GZ}_i \;\equiv\; \gamma_0 + \gamma_1\,\mathrm{educ}_i + \gamma_2\,\mathrm{kids}_i + \gamma_3\,\mathrm{spouse\_inc}_i$$

This is the linear combination from the selection equation (everything except the error v_i). A larger GZ_i means person i is more likely to work.

Standard normal PDF and CDF. Recall from your statistics courses:

- $\phi(z)$: the standard normal **density** (PDF) evaluated at z. Bell-curve height
- $\Phi(z)$: the standard normal **cumulative distribution** (CDF) evaluated at z. Area to the left of z

$\Longrightarrow$ In the probit model, $P(\mathrm{work}_i = 1) = \Phi(\mathrm{GZ}_i)$. Both ϕ and Φ will appear in the selection correction formula.

The Inverse Mills Ratio

For bivariate normal errors (u_i, v_i), the selection bias term has a closed form:

$$E[u_i \mid \text{works}] = \rho\,\sigma_u \cdot \underbrace{\frac{\phi(GZ_i)}{\Phi(GZ_i)}}_{\equiv \lambda_i \text{ (inverse Mills ratio)}}$$

The Inverse Mills Ratio

For bivariate normal errors (u_i, v_i), the selection bias term has a closed form:

$$E[u_i \mid \text{works}] = \rho\,\sigma_u \cdot \underbrace{\frac{\phi(GZ_i)}{\Phi(GZ_i)}}_{\equiv \lambda_i \text{ (inverse Mills ratio)}}$$

The IMR $\lambda_i = \phi(GZ_i)/\Phi(GZ_i)$ measures how strongly selection affects each individual. It depends only on the selection index, not on the wage equation.

The Inverse Mills Ratio

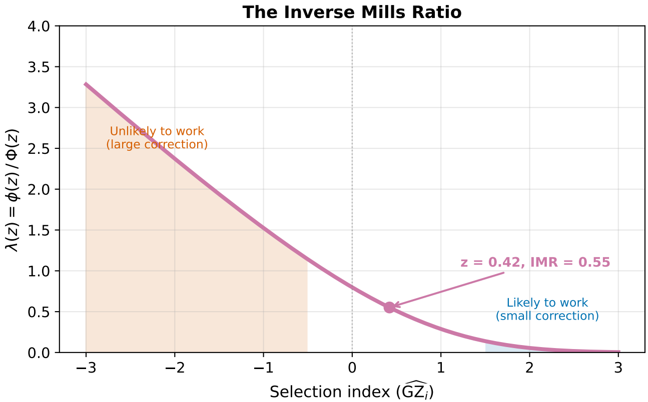
For bivariate normal errors (u_i, v_i), the selection bias term has a closed form:

$$E[u_i \mid \text{works}] = \rho\,\sigma_u \cdot \underbrace{\frac{\phi(GZ_i)}{\Phi(GZ_i)}}_{\equiv \lambda_i \text{ (inverse Mills ratio)}}$$

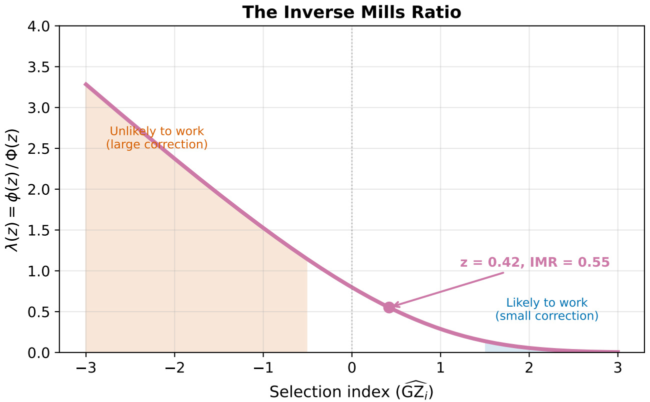
The IMR $\lambda_i = \phi(GZ_i)/\Phi(GZ_i)$ measures how strongly selection affects each individual. It depends only on the selection index, not on the wage equation.

$\Longrightarrow$ Connection to Tobit: you saw the IMR in the conditional expectation $E[y \mid y > 0]$. Same mathematical object, different context. In Tobit it corrects for censoring; here it corrects for selection.

How the IMR Works

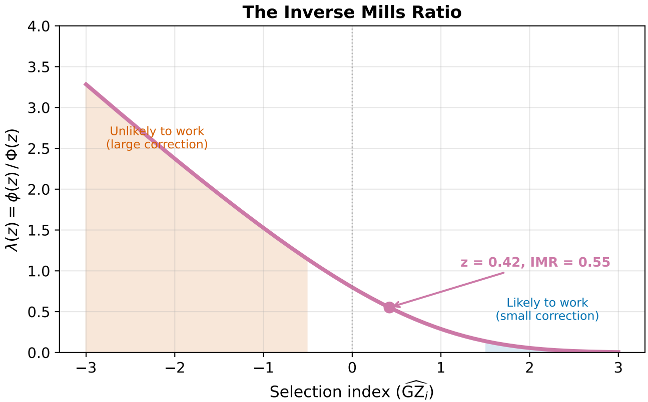


How the IMR Works



When GZ_i is small (unlikely to work), the IMR is **large**: if she works despite low predicted probability, her unobserved characteristics must be unusually favorable.

How the IMR Works



When GZ_i is small (unlikely to work), the IMR is **large**: if she works despite low predicted probability, her unobserved characteristics must be unusually favorable.

When GZ_i is large (very likely to work), the IMR is **small**: working tells us little about her unobservables. The correction is minimal.

The Corrected Wage Equation

Substituting the IMR into the conditional expectation:

$$\mathrm{E}[\log(\mathrm{wage}_i) \mid \mathrm{works}] = \beta_0 + \beta_1\,\mathrm{educ}_i + \beta_2\,\mathrm{exper}_i + \underbrace{\rho\sigma_u}_{\equiv\,\delta} \cdot \lambda_i$$

The Corrected Wage Equation

Substituting the IMR into the conditional expectation:

$$E[\log(\text{wage}_i) \mid \text{works}] = \beta_0 + \beta_1 \, \text{educ}_i + \beta_2 \, \text{exper}_i + \underbrace{\rho\sigma_u}_{\equiv \delta} \cdot \lambda_i$$

We combine ρ and σ_u into a single parameter $\delta = \rho\sigma_u$ because the two-step procedure cannot separately identify them: the second-stage OLS only estimates the product, not the individual components. Then:

$$\log(\text{wage}_i) = \beta_0 + \beta_1 \, \text{educ}_i + \beta_2 \, \text{exper}_i + \delta \cdot \lambda_i + \text{error}$$

The Corrected Wage Equation

Substituting the IMR into the conditional expectation:

$$\mathrm{E}[\log(\mathrm{wage}_i) \mid \mathrm{works}] = \beta_0 + \beta_1 \,\mathrm{educ}_i + \beta_2 \,\mathrm{exper}_i + \underbrace{\rho\sigma_u}_{\equiv \delta} \cdot \lambda_i$$

We combine ρ and σ_u into a single parameter $\delta = \rho\sigma_u$ because the two-step procedure cannot separately identify them: the second-stage OLS only estimates the product, not the individual components. Then:

$$\log(\mathrm{wage}_i) = \beta_0 + \beta_1 \,\mathrm{educ}_i + \beta_2 \,\mathrm{exper}_i + \delta \cdot \lambda_i + \mathrm{error}$$

The error in this equation is heteroskedastic (its variance depends on GZ_i). This is why the Step 2 standard errors need correction.

The Corrected Wage Equation

Substituting the IMR into the conditional expectation:

$$\text{E}[\log(\text{wage}_i) \mid \text{works}] = \beta_0 + \beta_1 \,\text{educ}_i + \beta_2 \,\text{exper}_i + \underbrace{\rho\sigma_u}_{\equiv \delta} \cdot \lambda_i$$

We combine ρ and σ_u into a single parameter $\delta = \rho\sigma_u$ because the two-step procedure cannot separately identify them: the second-stage OLS only estimates the product, not the individual components. Then:

$$\log(\text{wage}_i) = \beta_0 + \beta_1 \,\text{educ}_i + \beta_2 \,\text{exper}_i + \delta \cdot \lambda_i + \text{error}$$

The error in this equation is heteroskedastic (its variance depends on GZ_i). This is why the Step 2 standard errors need correction.

This is an OLS regression with one additional variable: λ_i. If we knew the γ coefficients (from the selection equation), we could compute λ_i for each worker and run this regression.

The Corrected Wage Equation

Substituting the IMR into the conditional expectation:

$$E[\log(\text{wage}_i) \mid \text{works}] = \beta_0 + \beta_1 \, \text{educ}_i + \beta_2 \, \text{exper}_i + \underbrace{\rho\sigma_u}_{\equiv \delta} \cdot \lambda_i$$

We combine ρ and σ_u into a single parameter $\delta = \rho\sigma_u$ because the two-step procedure cannot separately identify them: the second-stage OLS only estimates the product, not the individual components. Then:

$$\log(\text{wage}_i) = \beta_0 + \beta_1 \, \text{educ}_i + \beta_2 \, \text{exper}_i + \delta \cdot \lambda_i + \text{error}$$

The error in this equation is heteroskedastic (its variance depends on GZ_i). This is why the Step 2 standard errors need correction.

This is an OLS regression with one additional variable: λ_i. If we knew the γ coefficients (from the selection equation), we could compute λ_i for each worker and run this regression.

$\Longrightarrow$ This is the logic behind the Heckman two-step procedure.

Roadmap: From Formula to Estimation

Where we stand:

Roadmap: From Formula to Estimation

Where we stand:

1. We showed that OLS on workers is biased because $\mathrm{E}[u_i \mid \text{works}] \neq 0$

Roadmap: From Formula to Estimation

Where we stand:

1. We showed that OLS on workers is biased because $\mathrm{E}[u_i \mid \text{works}] \neq 0$
2. We derived that the bias equals $\rho\sigma_u \cdot \lambda_i$, where λ_i is the inverse Mills ratio

Roadmap: From Formula to Estimation

Where we stand:

1. We showed that OLS on workers is biased because $\mathrm{E}[u_i \mid \text{works}] \neq 0$
2. We derived that the bias equals $\rho\sigma_u \cdot \lambda_i$, where λ_i is the inverse Mills ratio
3. We showed that if we add λ_i to the wage regression, we can recover the true β coefficients

Roadmap: From Formula to Estimation

Where we stand:

1. We showed that OLS on workers is biased because $\mathrm{E}[u_i \mid \text{works}] \neq 0$
2. We derived that the bias equals $\rho\sigma_u \cdot \lambda_i$, where λ_i is the inverse Mills ratio
3. We showed that if we add λ_i to the wage regression, we can recover the true β coefficients

The remaining problem: computing λ_i requires the selection coefficients $\gamma_0, \gamma_1, \gamma_2, \gamma_3$, which we do not know.

Roadmap: From Formula to Estimation

Where we stand:

1. We showed that OLS on workers is biased because $E[u_i \mid \text{works}] \neq 0$
2. We derived that the bias equals $\rho\sigma_u \cdot \lambda_i$, where λ_i is the inverse Mills ratio
3. We showed that if we add λ_i to the wage regression, we can recover the true β coefficients

The remaining problem: computing λ_i requires the selection coefficients $\gamma_0, \gamma_1, \gamma_2, \gamma_3$, which we do not know.

$\implies$ We need to estimate the γ coefficients first. This suggests a **two-step procedure**: (1) estimate the selection equation, (2) use the estimated $\hat{\lambda}_i$ in the wage regression.

Outline

Step 1: Estimate the Selection Equation (Probit)

Run a **probit** on all 1,000 observations (workers and non-workers):

$$P(\text{work}_i = 1) = \Phi(\gamma_0 + \gamma_1 \,\text{educ}_i + \gamma_2 \,\text{kids}_i + \gamma_3 \,\text{spouse\_inc}_i)$$

Step 1: Estimate the Selection Equation (Probit)

Run a **probit** on all 1,000 observations (workers and non-workers):

$$P(\text{work}_i = 1) = \Phi(\gamma_0 + \gamma_1\,\text{educ}_i + \gamma_2\,\text{kids}_i + \gamma_3\,\text{spouse\_inc}_i)$$

Parameter	**Probit estimate**	**True value**
$\hat{\gamma}_0$ (intercept)	-2.629	-2.5
$\hat{\gamma}_1$ (education)	0.314	0.3
$\hat{\gamma}_2$ (children)	-0.512	-0.5
$\hat{\gamma}_3$ (spouse income)	-0.019	-0.02

Step 1: Estimate the Selection Equation (Probit)

Run a **probit** on all 1,000 observations (workers and non-workers):

$$P(\text{work}_i = 1) = \Phi(\gamma_0 + \gamma_1 \,\text{educ}_i + \gamma_2 \,\text{kids}_i + \gamma_3 \,\text{spouse\_inc}_i)$$

Parameter	**Probit estimate**	**True value**
$\hat{\gamma}_0$ (intercept)	-2.629	-2.5
$\hat{\gamma}_1$ (education)	0.314	0.3
$\hat{\gamma}_2$ (children)	-0.512	-0.5
$\hat{\gamma}_3$ (spouse income)	-0.019	-0.02

More education $\implies$ more likely to work. More children, higher spouse income $\implies$ less likely to work.

Step 1: Estimate the Selection Equation (Probit)

Run a **probit** on all 1,000 observations (workers and non-workers):

$$P(\text{work}_i = 1) = \Phi(\gamma_0 + \gamma_1\,\text{educ}_i + \gamma_2\,\text{kids}_i + \gamma_3\,\text{spouse\_inc}_i)$$

Parameter	**Probit estimate**	**True value**
$\hat{\gamma}_0$ (intercept)	-2.629	-2.5
$\hat{\gamma}_1$ (education)	0.314	0.3
$\hat{\gamma}_2$ (children)	-0.512	-0.5
$\hat{\gamma}_3$ (spouse income)	-0.019	-0.02

More education $\Longrightarrow$ more likely to work. More children, higher spouse income $\Longrightarrow$ less likely to work.

From these estimates, compute the estimated selection index $\widehat{GZ}_i$ for every individual.

Step 1 Continued: Compute the IMR

For each worker i, compute the inverse Mills ratio using the estimated selection index $\widehat{GZ}_i$:

$$\hat{\lambda}_i = \frac{\phi(\widehat{GZ}_i)}{\Phi(\widehat{GZ}_i)}$$

Step 1 Continued: Compute the IMR

For each worker i, compute the inverse Mills ratio using the estimated selection index $\widehat{\mathrm{GZ}}_i$:

$$\hat{\lambda}_i = \frac{\phi(\widehat{\mathrm{GZ}}_i)}{\Phi(\widehat{\mathrm{GZ}}_i)}$$

Numeric example: a woman with 16 years of education, 2 children, spouse earning \$50k:

$$\begin{aligned}\widehat{\mathrm{GZ}} &= -2.629 + 0.314 \times 16 + (-0.512) \times 2 + (-0.019) \times 50 \\ &= -2.629 + 5.024 - 1.024 - 0.950 = 0.421\end{aligned}$$

Step 1 Continued: Compute the IMR

For each worker i, compute the inverse Mills ratio using the estimated selection index $\widehat{\mathrm{GZ}}_i$:

$$\hat{\lambda}_i = \frac{\phi(\widehat{\mathrm{GZ}}_i)}{\Phi(\widehat{\mathrm{GZ}}_i)}$$

Numeric example: a woman with 16 years of education, 2 children, spouse earning $50k:

$$\begin{aligned}\widehat{\mathrm{GZ}} &= -2.629 + 0.314 \times 16 + (-0.512) \times 2 + (-0.019) \times 50 \\ &= -2.629 + 5.024 - 1.024 - 0.950 = 0.421\end{aligned}$$

- $P(\text{works}) = \Phi(0.421) = 0.663$ (66% chance of working)
- $\hat{\lambda} = \phi(0.421)/\Phi(0.421) = 0.365/0.663 = 0.551$

Step 1 Continued: Compute the IMR

For each worker i, compute the inverse Mills ratio using the estimated selection index $\widehat{GZ}_i$:

$$\hat{\lambda}_i = \frac{\phi(\widehat{GZ}_i)}{\Phi(\widehat{GZ}_i)}$$

Numeric example: a woman with 16 years of education, 2 children, spouse earning $50k:

$$\begin{aligned}\widehat{GZ} &= -2.629 + 0.314 \times 16 + (-0.512) \times 2 + (-0.019) \times 50 \\ &= -2.629 + 5.024 - 1.024 - 0.950 = 0.421\end{aligned}$$

- $P(\text{works}) = \Phi(0.421) = 0.663$ (66% chance of working)
- $\hat{\lambda} = \phi(0.421)/\Phi(0.421) = 0.365/0.663 = 0.551$

$\Longrightarrow$ This worker has a moderate selection correction. If she were nearly certain to work, $\hat{\lambda}$ would be close to zero.

Step 2: OLS with the IMR as an Extra Regressor

Run OLS on the **603 workers**, adding $\hat{\lambda}_i$ as an additional regressor:

$$\log(\text{wage}_i) = \beta_0 + \beta_1 \, \text{educ}_i + \beta_2 \, \text{exper}_i + \delta \, \hat{\lambda}_i + \text{error}$$

Step 2: OLS with the IMR as an Extra Regressor

Run OLS on the **603 workers**, adding $\hat{\lambda}_i$ as an additional regressor:

$$\log(\text{wage}_i) = \beta_0 + \beta_1 \, \text{educ}_i + \beta_2 \, \text{exper}_i + \delta \, \hat{\lambda}_i + \text{error}$$

Parameter	**OLS (workers)**	**Heckman**	**True**
$\hat{\beta}_0$ (intercept)	1.373	0.874	1.0
$\hat{\beta}_1$ (education)	0.085	**0.111**	0.100
$\hat{\beta}_2$ (experience)	0.040	0.039	0.040
$\hat{\delta}$ (IMR)	–	0.260	0.240

Step 2: OLS with the IMR as an Extra Regressor

Run OLS on the **603 workers**, adding $\hat{\lambda}_i$ as an additional regressor:

$$\log(\text{wage}_i) = \beta_0 + \beta_1\,\text{educ}_i + \beta_2\,\text{exper}_i + \delta\,\hat{\lambda}_i + \text{error}$$

Parameter	**OLS (workers)**	**Heckman**	**True**
$\hat{\beta}_0$ (intercept)	1.373	0.874	1.0
$\hat{\beta}_1$ (education)	0.085	**0.111**	0.100
$\hat{\beta}_2$ (experience)	0.040	0.039	0.040
$\hat{\delta}$ (IMR)	–	0.260	0.240

The Heckman education coefficient (**0.111**) is much closer to the true value (0.100) than OLS on workers (0.085).

Step 2: OLS with the IMR as an Extra Regressor

Run OLS on the **603 workers**, adding $\hat{\lambda}_i$ as an additional regressor:

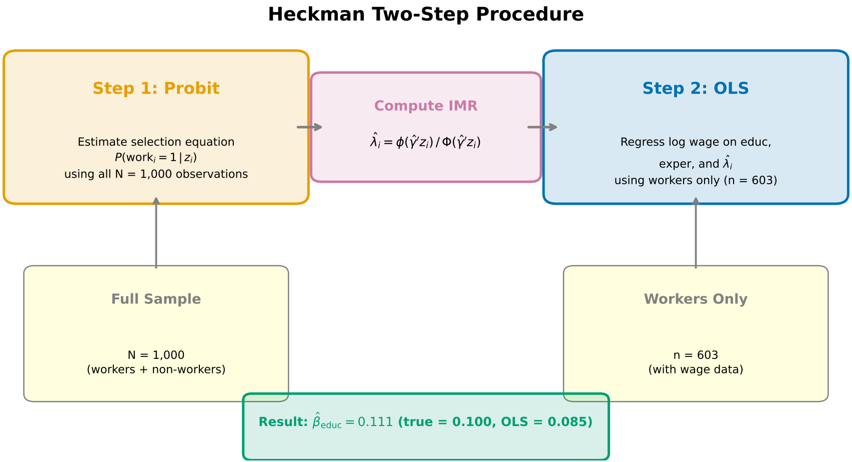
$$\log(\text{wage}_i) = \beta_0 + \beta_1\,\text{educ}_i + \beta_2\,\text{exper}_i + \delta\,\hat{\lambda}_i + \text{error}$$

Parameter	**OLS (workers)**	**Heckman**	**True**
$\hat{\beta}_0$ (intercept)	1.373	0.874	1.0
$\hat{\beta}_1$ (education)	0.085	**0.111**	0.100
$\hat{\beta}_2$ (experience)	0.040	0.039	0.040
$\hat{\delta}$ (IMR)	–	0.260	0.240

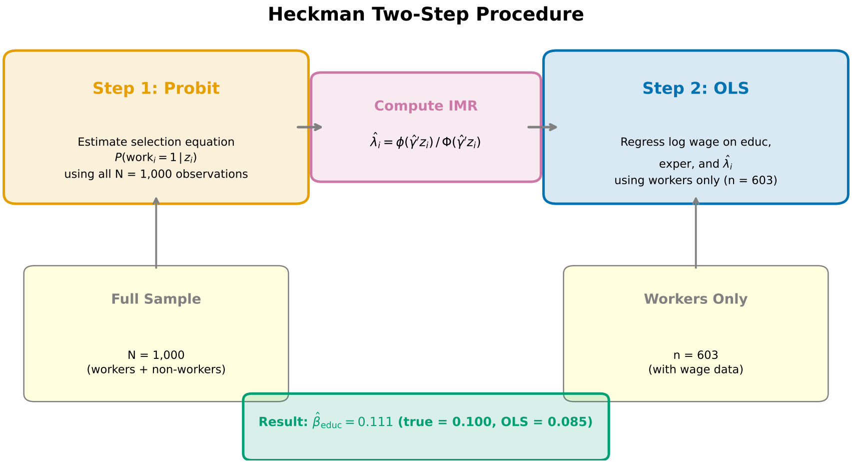
The Heckman education coefficient (**0.111**) is much closer to the true value (0.100) than OLS on workers (0.085).

$\Longrightarrow$ The selection correction removes the downward bias in the education coefficient.

The Procedure in One Picture

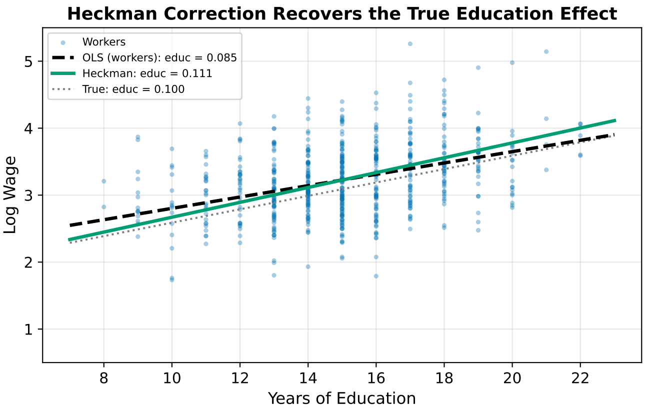


The Procedure in One Picture

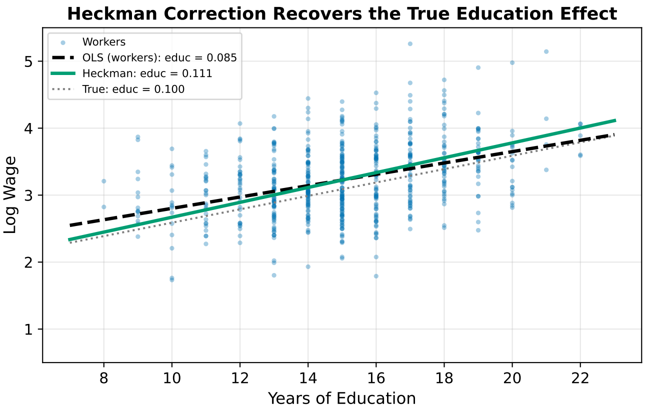


Step 1 uses **everyone** to learn who works. Step 2 uses **workers only** to learn how wages relate to education, after accounting for selection.

The Correction on the Scatter Plot



The Correction on the Scatter Plot



The dashed OLS line is too flat. The solid Heckman line is steeper, matching the true slope (gray dotted). Adding the IMR absorbs the selection effect that was biasing the education coefficient downward.

Interpreting the IMR Coefficient ($\hat{\delta}$)

The estimated IMR coefficient is $\hat{\delta} = 0.260$. What does it tell us?

Interpreting the IMR Coefficient ($\hat{\delta}$)

The estimated IMR coefficient is $\hat{\delta} = 0.260$. What does it tell us?

Recall: $\delta = \rho \cdot \sigma_u$, where $\rho = \text{Corr}(u, v)$ and σ_u is the wage error SD.

Interpreting the IMR Coefficient ($\hat{\delta}$)

The estimated IMR coefficient is $\hat{\delta} = 0.260$. What does it tell us?

Recall: $\delta = \rho \cdot \sigma_u$, where $\rho = \text{Corr}(u, v)$ and σ_u is the wage error SD.

- $\hat{\delta} > 0 \implies \hat{\rho} > 0$: unobserved factors that raise wages also raise the probability of working. Workers have higher-than-average wage errors

Interpreting the IMR Coefficient ($\hat{\delta}$)

The estimated IMR coefficient is $\hat{\delta} = 0.260$. What does it tell us?

Recall: $\delta = \rho \cdot \sigma_u$, where $\rho = \text{Corr}(u, v)$ and σ_u is the wage error SD.

- $\hat{\delta} > 0 \implies \hat{\rho} > 0$: unobserved factors that raise wages also raise the probability of working. Workers have higher-than-average wage errors
- $\hat{\delta} < 0$ would mean workers have *lower*-than-average unobserved wage determinants (e.g., people with high non-labor income stay home, regardless of their potential wage)

Interpreting the IMR Coefficient ($\hat{\delta}$)

The estimated IMR coefficient is $\hat{\delta} = 0.260$. What does it tell us?

Recall: $\delta = \rho \cdot \sigma_u$, where $\rho = \text{Corr}(u, v)$ and σ_u is the wage error SD.

- $\hat{\delta} > 0 \implies \hat{\rho} > 0$: unobserved factors that raise wages also raise the probability of working. Workers have higher-than-average wage errors
- $\hat{\delta} < 0$ would mean workers have *lower*-than-average unobserved wage determinants (e.g., people with high non-labor income stay home, regardless of their potential wage)
- $\hat{\delta} = 0$ would mean no selection bias: who works is unrelated to unobserved wage factors, and OLS on workers would be consistent

Interpreting the IMR Coefficient ($\hat{\delta}$)

The estimated IMR coefficient is $\hat{\delta} = 0.260$. What does it tell us?

Recall: $\delta = \rho \cdot \sigma_u$, where $\rho = \text{Corr}(u, v)$ and σ_u is the wage error SD.

- $\hat{\delta} > 0 \implies \hat{\rho} > 0$: unobserved factors that raise wages also raise the probability of working. Workers have higher-than-average wage errors
- $\hat{\delta} < 0$ would mean workers have *lower*-than-average unobserved wage determinants (e.g., people with high non-labor income stay home, regardless of their potential wage)
- $\hat{\delta} = 0$ would mean no selection bias: who works is unrelated to unobserved wage factors, and OLS on workers would be consistent

$\implies$ In our data, $\hat{\delta} = 0.260 > 0$: positive selection. The non-random sample of workers overrepresents high-ability individuals.

Where Are We? A Recap

1. **Problem:** OLS on workers underestimates the return to education (0.085 vs. 0.100) because high-ability people self-select into work

Where Are We? A Recap

1. **Problem:** OLS on workers underestimates the return to education (0.085 vs. 0.100) because high-ability people self-select into work
2. **Fix:** the Heckman two-step adds the inverse Mills ratio $\hat{\lambda}_i$ to the wage regression. This absorbs the selection effect and recovers a coefficient (0.111) much closer to the truth

Where Are We? A Recap

1. **Problem:** OLS on workers underestimates the return to education (0.085 vs. 0.100) because high-ability people self-select into work
2. **Fix:** the Heckman two-step adds the inverse Mills ratio $\hat{\lambda}_i$ to the wage regression. This absorbs the selection effect and recovers a coefficient (0.111) much closer to the truth
3. **The IMR coefficient** $\hat{\delta} = 0.260 > 0$ confirms positive selection: workers have above-average unobserved wage determinants

Where Are We? A Recap

1. **Problem:** OLS on workers underestimates the return to education (0.085 vs. 0.100) because high-ability people self-select into work
2. **Fix:** the Heckman two-step adds the inverse Mills ratio $\hat{\lambda}_i$ to the wage regression. This absorbs the selection effect and recovers a coefficient (0.111) much closer to the truth
3. **The IMR coefficient** $\hat{\delta} = 0.260 > 0$ confirms positive selection: workers have above-average unobserved wage determinants

What remains:

- What makes the Heckman model **credible**? (The exclusion restriction)
- How do we **test** whether selection bias is present?
- When does the model **fail**?

Outline

The Exclusion Restriction: Why Kids and Spouse Income

Notice that **kids** and **spouse income** appear in the selection equation but *not* in the wage equation.

The Exclusion Restriction: Why Kids and Spouse Income

Notice that **kids** and **spouse income** appear in the selection equation but *not* in the wage equation.

This is the **exclusion restriction**: we need at least one variable that affects *whether* someone works but does not directly affect their wage:

- **Number of children:** affects labor force participation (childcare responsibilities) but does not directly determine hourly productivity
- **Spouse income:** affects the financial need to work (reservation wage) but does not directly affect the wage an employer offers

The Exclusion Restriction: Why Kids and Spouse Income

Notice that **kids** and **spouse income** appear in the selection equation but *not* in the wage equation.

This is the **exclusion restriction**: we need at least one variable that affects *whether* someone works but does not directly affect their wage:

- **Number of children:** affects labor force participation (childcare responsibilities) but does not directly determine hourly productivity
- **Spouse income:** affects the financial need to work (reservation wage) but does not directly affect the wage an employer offers

Without an exclusion restriction, the IMR $\hat{\lambda}_i$ would be a nonlinear function of the same variables in the wage equation. The model is "technically identified" by functional form alone, but estimates become very unstable.

The Exclusion Restriction: Why Kids and Spouse Income

Notice that **kids** and **spouse income** appear in the selection equation but *not* in the wage equation.

This is the **exclusion restriction**: we need at least one variable that affects *whether* someone works but does not directly affect their wage:

- **Number of children:** affects labor force participation (childcare responsibilities) but does not directly determine hourly productivity
- **Spouse income:** affects the financial need to work (reservation wage) but does not directly affect the wage an employer offers

Without an exclusion restriction, the IMR $\hat{\lambda}_i$ would be a nonlinear function of the same variables in the wage equation. The model is "technically identified" by functional form alone, but estimates become very unstable.

$\Longrightarrow$ A credible exclusion restriction is what separates a convincing Heckman model from one that is essentially relying on the normality assumption.

Full MLE: An Alternative to Two-Step

The two-step procedure is intuitive and easy to implement, but there is an alternative: estimate both equations simultaneously by **maximum likelihood**.

Full MLE: An Alternative to Two-Step

The two-step procedure is intuitive and easy to implement, but there is an alternative: estimate both equations simultaneously by **maximum likelihood**.

The full MLE maximizes the joint likelihood of the wage data (for workers) and the selection data (for everyone) together, estimating β, γ, σ_u, and ρ in one step.

Full MLE: An Alternative to Two-Step

The two-step procedure is intuitive and easy to implement, but there is an alternative: estimate both equations simultaneously by **maximum likelihood**.

The full MLE maximizes the joint likelihood of the wage data (for workers) and the selection data (for everyone) together, estimating β, γ, σ_u, and ρ in one step.

Advantages of full MLE:

- More efficient (smaller standard errors) than two-step
- Directly estimates ρ and σ_u separately

Full MLE: An Alternative to Two-Step

The two-step procedure is intuitive and easy to implement, but there is an alternative: estimate both equations simultaneously by **maximum likelihood**.

The full MLE maximizes the joint likelihood of the wage data (for workers) and the selection data (for everyone) together, estimating β, γ, σ_u, and ρ in one step.

Advantages of full MLE:

- More efficient (smaller standard errors) than two-step
- Directly estimates ρ and σ_u separately

Advantages of two-step:

- Less sensitive to distributional misspecification: the two-step does not impose the full joint likelihood structure, so it degrades more gracefully when normality is approximate
- Easier to diagnose: you can examine the probit and OLS stages separately
- The IMR coefficient test (next slide) provides a simple check for selection bias

Full MLE: An Alternative to Two-Step

The two-step procedure is intuitive and easy to implement, but there is an alternative: estimate both equations simultaneously by **maximum likelihood**.

The full MLE maximizes the joint likelihood of the wage data (for workers) and the selection data (for everyone) together, estimating β, γ, σ_u, and ρ in one step.

Advantages of full MLE:

- More efficient (smaller standard errors) than two-step
- Directly estimates ρ and σ_u separately

Advantages of two-step:

- Less sensitive to distributional misspecification: the two-step does not impose the full joint likelihood structure, so it degrades more gracefully when normality is approximate
- Easier to diagnose: you can examine the probit and OLS stages separately
- The IMR coefficient test (next slide) provides a simple check for selection bias

$\Longrightarrow$ In practice, many researchers run both and compare results. If they agree, the findings are more credible.

Testing for Selection Bias: Is the IMR Significant?

A simple test for selection bias: in the Heckman second stage, check whether $\hat{\delta}$ is statistically significant.

Testing for Selection Bias: Is the IMR Significant?

A simple test for selection bias: in the Heckman second stage, check whether $\hat{\delta}$ is statistically significant.

$$H_0\colon\ \delta = 0 \quad \iff \quad \text{no selection bias (OLS on workers is consistent)}$$
$$H_1\colon\ \delta \neq 0 \quad \iff \quad \text{selection bias present}$$

Testing for Selection Bias: Is the IMR Significant?

A simple test for selection bias: in the Heckman second stage, check whether $\hat{\delta}$ is statistically significant.

$$H_0\colon\ \delta = 0 \quad \iff \quad \text{no selection bias (OLS on workers is consistent)}$$

$$H_1\colon\ \delta \neq 0 \quad \iff \quad \text{selection bias present}$$

If you cannot reject H_0, selection bias may not be a problem, and OLS on workers is adequate.

Testing for Selection Bias: Is the IMR Significant?

A simple test for selection bias: in the Heckman second stage, check whether $\hat{\delta}$ is statistically significant.

$$H_0\colon\ \delta = 0 \quad \iff \quad \text{no selection bias (OLS on workers is consistent)}$$
$$H_1\colon\ \delta \neq 0 \quad \iff \quad \text{selection bias present}$$

If you cannot reject H_0, selection bias may not be a problem, and OLS on workers is adequate.

If you reject H_0, the Heckman correction is needed.

Testing for Selection Bias: Is the IMR Significant?

A simple test for selection bias: in the Heckman second stage, check whether $\hat{\delta}$ is statistically significant.

$$H_0\colon\ \delta = 0 \quad \Longleftrightarrow \quad \text{no selection bias (OLS on workers is consistent)}$$
$$H_1\colon\ \delta \neq 0 \quad \Longleftrightarrow \quad \text{selection bias present}$$

If you cannot reject H_0, selection bias may not be a problem, and OLS on workers is adequate.

If you reject H_0, the Heckman correction is needed.

Caveat: the standard errors from the naive Step 2 OLS are *incorrect* because $\hat{\lambda}_i$ is a generated regressor (estimated in Step 1). Software adjusts for this automatically; if computing by hand, you need a correction.

Testing for Selection Bias: Is the IMR Significant?

A simple test for selection bias: in the Heckman second stage, check whether $\hat{\delta}$ is statistically significant.

$$H_0\colon\ \delta = 0 \quad \Longleftrightarrow \quad \text{no selection bias (OLS on workers is consistent)}$$
$$H_1\colon\ \delta \neq 0 \quad \Longleftrightarrow \quad \text{selection bias present}$$

If you cannot reject H_0, selection bias may not be a problem, and OLS on workers is adequate.

If you reject H_0, the Heckman correction is needed.

Caveat: the standard errors from the naive Step 2 OLS are *incorrect* because $\hat{\lambda}_i$ is a generated regressor (estimated in Step 1). Software adjusts for this automatically; if computing by hand, you need a correction.

$\Longrightarrow$ In our data, $\hat{\delta} = 0.260$ is positive and statistically significant, confirming that selection bias is present and the correction is needed.

When the Heckman Model Fails

The Heckman model relies on several assumptions. It can fail when:

When the Heckman Model Fails

The Heckman model relies on several assumptions. It can fail when:

1. **No valid exclusion restriction:** if every variable that affects selection also affects wages, the IMR is identified only by functional form. Small deviations from normality produce large changes in estimates

When the Heckman Model Fails

The Heckman model relies on several assumptions. It can fail when:

1. **No valid exclusion restriction:** if every variable that affects selection also affects wages, the IMR is identified only by functional form. Small deviations from normality produce large changes in estimates
2. **Non-normal errors:** both the probit in Step 1 and the IMR formula assume joint normality of (u_i, v_i). Heavy tails or skewness invalidate the correction

When the Heckman Model Fails

The Heckman model relies on several assumptions. It can fail when:

1. **No valid exclusion restriction:** if every variable that affects selection also affects wages, the IMR is identified only by functional form. Small deviations from normality produce large changes in estimates
2. **Non-normal errors:** both the probit in Step 1 and the IMR formula assume joint normality of (u_i, v_i). Heavy tails or skewness invalidate the correction
3. **Misspecified selection equation:** if the probit model is wrong (missing variables, wrong functional form), the estimated IMR is wrong, and the correction introduces bias rather than removing it

When the Heckman Model Fails

The Heckman model relies on several assumptions. It can fail when:

1. **No valid exclusion restriction:** if every variable that affects selection also affects wages, the IMR is identified only by functional form. Small deviations from normality produce large changes in estimates
2. **Non-normal errors:** both the probit in Step 1 and the IMR formula assume joint normality of (u_i, v_i). Heavy tails or skewness invalidate the correction
3. **Misspecified selection equation:** if the probit model is wrong (missing variables, wrong functional form), the estimated IMR is wrong, and the correction introduces bias rather than removing it
4. **Weak selection:** if almost everyone works (or almost no one does), the IMR has very little variation across individuals, making it hard to identify δ

When the Heckman Model Fails

The Heckman model relies on several assumptions. It can fail when:

1. **No valid exclusion restriction:** if every variable that affects selection also affects wages, the IMR is identified only by functional form. Small deviations from normality produce large changes in estimates
2. **Non-normal errors:** both the probit in Step 1 and the IMR formula assume joint normality of (u_i, v_i). Heavy tails or skewness invalidate the correction
3. **Misspecified selection equation:** if the probit model is wrong (missing variables, wrong functional form), the estimated IMR is wrong, and the correction introduces bias rather than removing it
4. **Weak selection:** if almost everyone works (or almost no one does), the IMR has very little variation across individuals, making it hard to identify δ

$\Longrightarrow$ The Heckman model is powerful but not a magic fix. A credible exclusion restriction and reasonable normality are essential.

Decision Flowchart: Heckman vs. Tobit vs. OLS

1. Is your outcome missing for a non-random subset of observations?

Decision Flowchart: Heckman vs. Tobit vs. OLS

1. Is your outcome missing for a non-random subset of observations?
   - **Yes:** the missing values come from a *separate selection process* (e.g., wages unobserved because the person does not work)
     - Do you have a valid exclusion restriction? $\implies$ **Heckman Selection Model**
     - No exclusion restriction? $\implies$ Consider **bounds** or **sensitivity analysis**

Decision Flowchart: Heckman vs. Tobit vs. OLS

1. Is your outcome missing for a non-random subset of observations?
   - **Yes:** the missing values come from a *separate selection process* (e.g., wages unobserved because the person does not work)
     - Do you have a valid exclusion restriction? $\Longrightarrow$ **Heckman Selection Model**
     - No exclusion restriction? $\Longrightarrow$ Consider **bounds** or **sensitivity analysis**
   - **No:** the zeros are *corner solutions* (the person would choose a negative value but is constrained)
     - Same mechanism for participation and amount? $\Longrightarrow$ **Tobit**
     - Different mechanisms? $\Longrightarrow$ **Two-Part Model**

Decision Flowchart: Heckman vs. Tobit vs. OLS

1. Is your outcome missing for a non-random subset of observations?
   - **Yes:** the missing values come from a *separate selection process* (e.g., wages unobserved because the person does not work)
     - Do you have a valid exclusion restriction? $\Longrightarrow$ **Heckman Selection Model**
     - No exclusion restriction? $\Longrightarrow$ Consider **bounds** or **sensitivity analysis**
   - **No:** the zeros are *corner solutions* (the person would choose a negative value but is constrained)
     - Same mechanism for participation and amount? $\Longrightarrow$ **Tobit**
     - Different mechanisms? $\Longrightarrow$ **Two-Part Model**
2. Is there no selection or censoring at all?
   - $\Longrightarrow$ **OLS** is fine

Decision Flowchart: Heckman vs. Tobit vs. OLS

1. Is your outcome missing for a non-random subset of observations?
   - **Yes:** the missing values come from a *separate selection process* (e.g., wages unobserved because the person does not work)
     - Do you have a valid exclusion restriction? $\Longrightarrow$ **Heckman Selection Model**
     - No exclusion restriction? $\Longrightarrow$ Consider **bounds** or **sensitivity analysis**
   - **No:** the zeros are *corner solutions* (the person would choose a negative value but is constrained)
     - Same mechanism for participation and amount? $\Longrightarrow$ **Tobit**
     - Different mechanisms? $\Longrightarrow$ **Two-Part Model**
2. Is there no selection or censoring at all?
   - $\Longrightarrow$ **OLS** is fine

$\Longrightarrow$ Selection (Heckman) and censoring (Tobit) address different problems. The distinction is economic: does the outcome *exist but go unobserved*, or is it *constrained to a boundary*?

Outline

Summary: Back to Wages and Education

1. **The data problem:** 40% of individuals do not work, so their wages are **missing**. Workers have more education, fewer kids, and lower spouse income than non-workers

Summary: Back to Wages and Education

1. **The data problem:** 40% of individuals do not work, so their wages are **missing**. Workers have more education, fewer kids, and lower spouse income than non-workers
2. **OLS on workers is biased:** the education coefficient is 0.085 instead of the true 0.100. Unobserved factors that raise wages also raise the probability of working ($\rho = 0.6$)

Summary: Back to Wages and Education

1. **The data problem:** 40% of individuals do not work, so their wages are **missing**. Workers have more education, fewer kids, and lower spouse income than non-workers
2. **OLS on workers is biased:** the education coefficient is 0.085 instead of the true 0.100. Unobserved factors that raise wages also raise the probability of working ($\rho = 0.6$)
3. **The Heckman model** adds a selection equation (probit) and corrects the wage equation with the inverse Mills ratio. The corrected education coefficient is 0.111, close to the true 0.100

Summary: Back to Wages and Education

1. **The data problem:** 40% of individuals do not work, so their wages are **missing**. Workers have more education, fewer kids, and lower spouse income than non-workers
2. **OLS on workers is biased:** the education coefficient is 0.085 instead of the true 0.100. Unobserved factors that raise wages also raise the probability of working ($\rho = 0.6$)
3. **The Heckman model** adds a selection equation (probit) and corrects the wage equation with the inverse Mills ratio. The corrected education coefficient is 0.111, close to the true 0.100
4. **The exclusion restriction** (children, spouse income affect selection but not wages) is what makes the model credible

Summary: Back to Wages and Education

1. **The data problem:** 40% of individuals do not work, so their wages are **missing**. Workers have more education, fewer kids, and lower spouse income than non-workers
2. **OLS on workers is biased:** the education coefficient is 0.085 instead of the true 0.100. Unobserved factors that raise wages also raise the probability of working ($\rho = 0.6$)
3. **The Heckman model** adds a selection equation (probit) and corrects the wage equation with the inverse Mills ratio. The corrected education coefficient is 0.111, close to the true 0.100
4. **The exclusion restriction** (children, spouse income affect selection but not wages) is what makes the model credible
5. **Testing:** if the IMR coefficient is significant, selection bias is present and the correction is needed

Summary: Back to Wages and Education

1. **The data problem:** 40% of individuals do not work, so their wages are **missing**. Workers have more education, fewer kids, and lower spouse income than non-workers
2. **OLS on workers is biased:** the education coefficient is 0.085 instead of the true 0.100. Unobserved factors that raise wages also raise the probability of working ($\rho = 0.6$)
3. **The Heckman model** adds a selection equation (probit) and corrects the wage equation with the inverse Mills ratio. The corrected education coefficient is 0.111, close to the true 0.100
4. **The exclusion restriction** (children, spouse income affect selection but not wages) is what makes the model credible
5. **Testing:** if the IMR coefficient is significant, selection bias is present and the correction is needed
6. **Heckman vs. Tobit:** Tobit is for censoring (corner solutions). Heckman is for selection (missing data from a separate decision process)

James Heckman received the Nobel Prize in Economics in 2000, in part for developing this model.

Comparison: OLS vs. Heckman on Our Data

	OLS (workers)	**Heckman**	**True**
Education	0.085	0.111	0.100
Experience	0.040	0.039	0.040
IMR ($\hat{\delta}$)	–	0.260	0.240
Bias in educ	-15%	$+11\%$	–

Comparison: OLS vs. Heckman on Our Data

	OLS (workers)	Heckman	True
Education	0.085	0.111	0.100
Experience	0.040	0.039	0.040
IMR ($\hat{\delta}$)	–	0.260	0.240
Bias in educ	-15%	$+11\%$	–

The Heckman two-step does not recover the true parameter perfectly (0.111 vs. 0.100), but the bias is smaller and in the opposite direction. With larger samples, both converge to the true values.

Comparison: OLS vs. Heckman on Our Data

	OLS (workers)	**Heckman**	**True**
Education	0.085	0.111	0.100
Experience	0.040	0.039	0.040
IMR ($\hat{\delta}$)	–	0.260	0.240
Bias in educ	-15%	$+11\%$	–

The Heckman two-step does not recover the true parameter perfectly (0.111 vs. 0.100), but the bias is smaller and in the opposite direction. With larger samples, both converge to the true values.

$\implies$ The selection correction works because it accounts for the fact that workers are not a random sample. Ignoring selection systematically underestimates the return to education in this setting.

Thank you!

jakeanderson@g.ucla.edu