ECON 0150 | Economic Data Analysis

The economist's data analysis pipeline.

Part 5.1 | Fixed Effects and Interaction Models

Example: The Gender Wage Gap

Using the general linear model to understand wage differences.

Key Questions:

- *Is there a wage gap between male / female?*
- *Are returns to education different between male / female?*
- *How can we model these questions with a regression framework?*

> lets build this analysis step by step

Model 1: The Gender Wage Gap

We can use an indicator variable to capture level differences.

$$\text{Wage} = \beta_0 + \beta_1 \times \text{Education} + \beta_2 \times \text{Male} + \varepsilon$$

Model 1: The Gender Wage Gap

We can use an indicator variable to capture level differences.

$$\text{Wage} = \beta_0 + \beta_1 \times \text{Education} + \beta_2 \times I(\text{Male}) + \varepsilon$$

> β_0 *is the base wage for those with no post-middle school education*

> β_2 *represents the gender wage gap - added to the intercept for males only*

> *model assumes parallel lines - same returns to education (*β_1*) for everyone*

Model 1: The Code

Implementing the gender fixed effect model

```python
import statsmodels.formula.api as smf

# Fit the model with male indicator
model1 = smf.ols('INCLOG10 ~ EDU + MALE', data=df).fit()
print(model1.summary())
```

Interpretation:

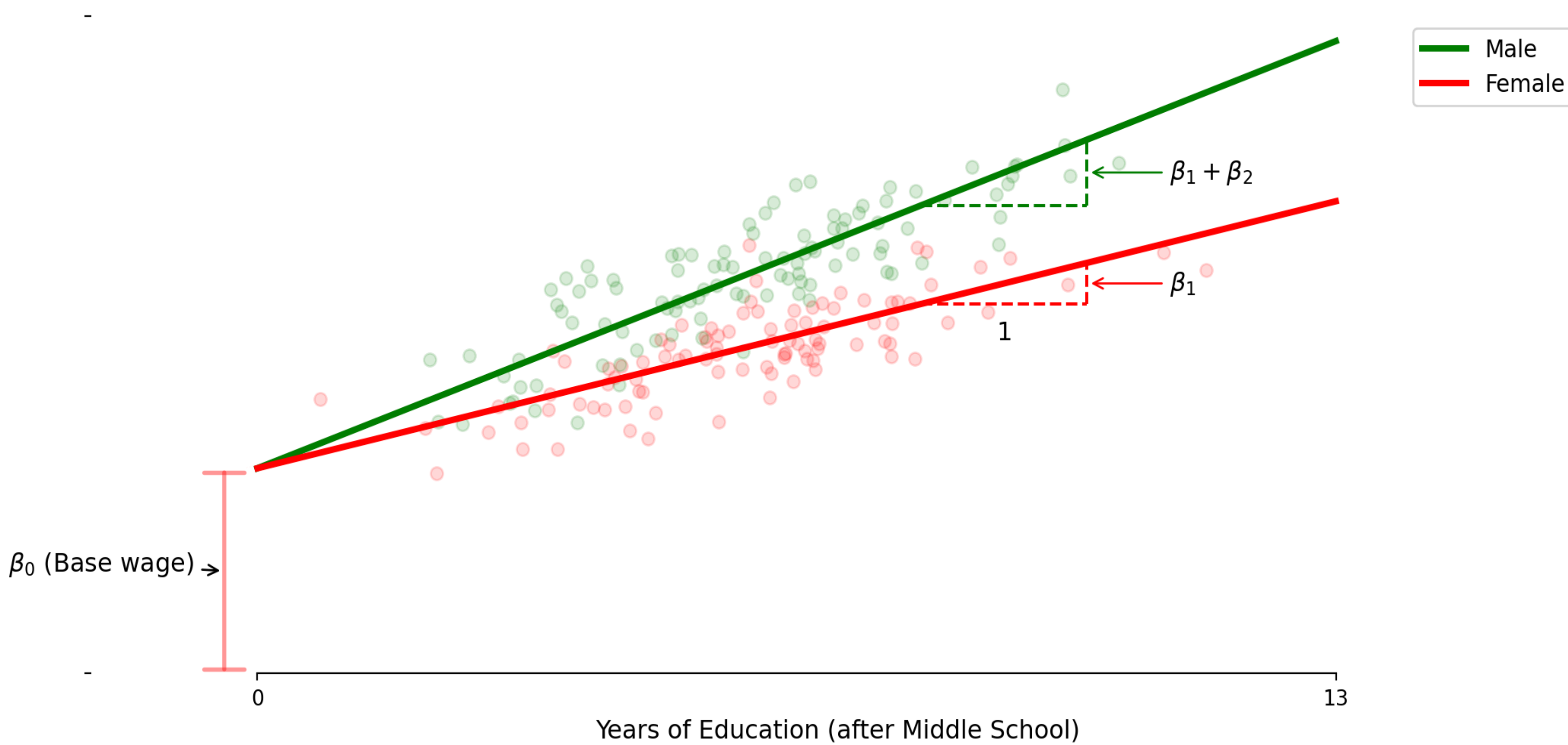
- β_0: *Base wage for females with zero education*
- β_1: *Return to each year of education (for everyone)*
- β_2: *Additional wage premium for males*

> *if* $\beta_2 > 0$ *and statistically significant, evidence of a gender wage gap*

Model 2: Different Returns to Education

What if education benefits genders differently?

$$\text{Wage} = \beta_0 + \beta_1 \times \text{Education} + \beta_2 \times \text{Education} \times \text{Male} + \varepsilon$$



Model 2: Different Returns to Education

What if education benefits genders differently?

$$\text{Wage} = \beta_0 + \beta_1 \times \text{Education} + \beta_2 \times \text{Education} \times \text{Male} + \varepsilon$$

> β_1 *represents the female return to education*

> β_2 *represents the additional male return to education - this changes the slope*

> *male education effect is* $\beta_1 + \beta_2$*, creating diverging wage paths*

Model 2: The Code

Implementing the education-gender interaction model

```
# Fit model with interaction between education and sex
model2 = smf.ols('INCLOG10 ~ EDU + EDU:MALE', data=df).fit()
print(model2.summary())
```

Interpretation:

- β_0: *Base wage with zero education*
- β_1: *Female return to education*
- β_2: *Additional male return to education*

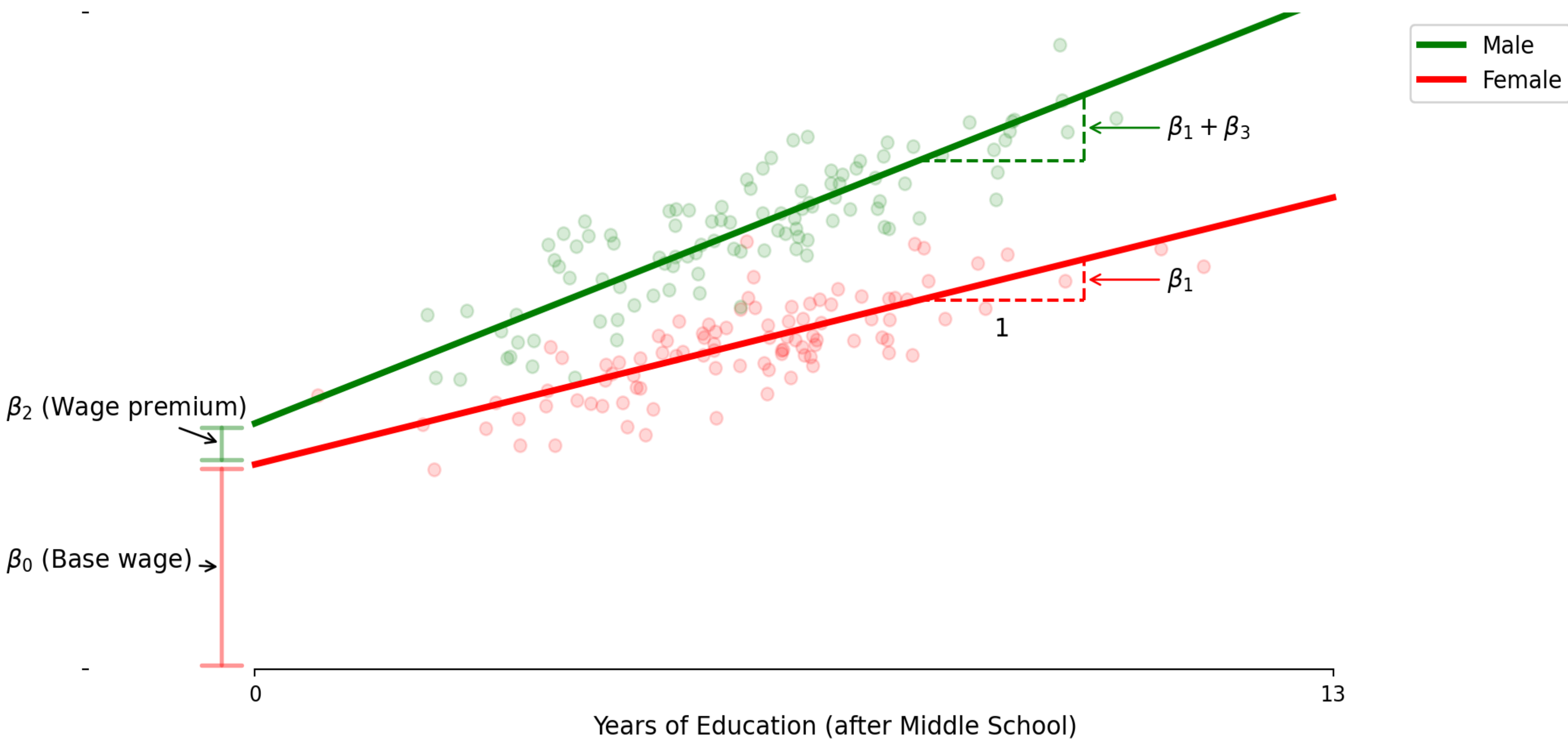
> *if* $\beta_2 > 0$ *and significant, male return to education is higher*

> *this model assumes same baseline (intercept) for both sexes*

Model 3: Full Gender Difference Model

Combining fixed effects and interactions

$$\text{Wage} = \beta_0 + \beta_1 \times \text{Education} + \beta_2 \times \text{Male} + \beta_3 \times \text{Education} \times \text{Male} + \varepsilon$$



Model 3: Full Gender Difference Model

Combining fixed effects and interactions

$$\text{Wage} = \beta_0 + \beta_1 \times \text{Education} + \beta_2 \times \text{Male} + \beta_3 \times \text{Education} \times \text{Male} + \varepsilon$$

> β_0 = *base wage*

> β_2 = *initial wage gap (at zero education)*

> β_1 = *female returns to education*

> β_3 = *male education return premium*

Model 3: The Code

Implementing the full gender difference model

```python
# Fit full model with both sex indicator and interaction
model3 = smf.ols('INCLOG10 ~ EDU + MALE + EDU:MALE', data=df).fit()
print(model3.summary())
```

Interpretation:

- β_0: *Base wage for females with zero education*
- β_1: *Return to education for females*
- β_2: *Additional base wage for males (at zero education)*
- β_3: *Additional return to education for males*

> *allows for differences in both baseline wages and educational returns*

Comparison of Models

Different models answer different questions

1. ***Model 1: Fixed Effect***
   - *Question: "Is there a gender wage gap?"*
   - *Focus: Level differences in wages*
2. ***Model 2: Interaction Only***
   - *Question: "Are there differences in returns to education?"*
   - *Focus: Slope differences in education effects*
3. ***Model 3: Full Model***
   - *Question: "Does the gender wage gap vary with education level?"*
   - *Focus: Comprehensive gender differences pattern*

> choose the model that best addresses your research question

When to Use Each Model?

Choosing the right model for your research question

1. ***Use Model 1 (Fixed Effect) when:***
   - *You want to estimate the average wage gap across all education levels*
   - *You believe returns to education are similar for all groups*
2. ***Use Model 2 (Interaction Only) when:***
   - *You're specifically interested in differential returns to education*
   - *You believe there's no baseline difference between groups*
3. ***Use Model 3 (Full Model) when:***
   - *You want to explore how the wage gap varies with education*
   - *You're testing for both baseline differences and differential returns*

> *always let your research question drive model selection*

Key Takeaways

General linear model for analyzing group differences

> the general linear model is a versatile tool for inequality research

*1. **Fixed effects** capture level differences between groups*

*2. **Interactions** capture slope differences (differential returns)*

*3. **Combining both** gives a complete picture of how relationships vary by group*

*4. **Model choice** should be guided by your research question*

> these tools are essential for analyzing disparities in economics