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.

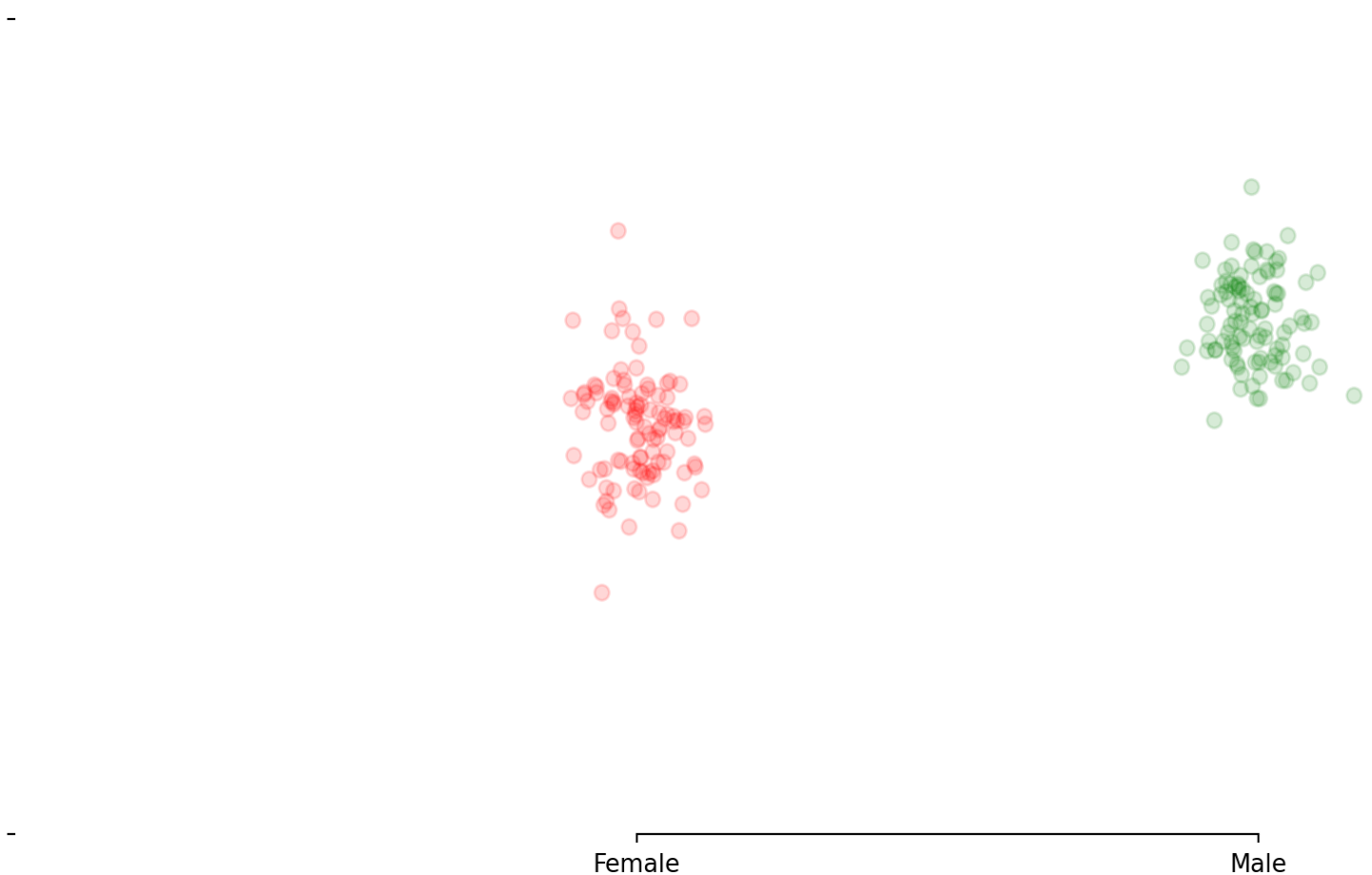
Key Questions:

- *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

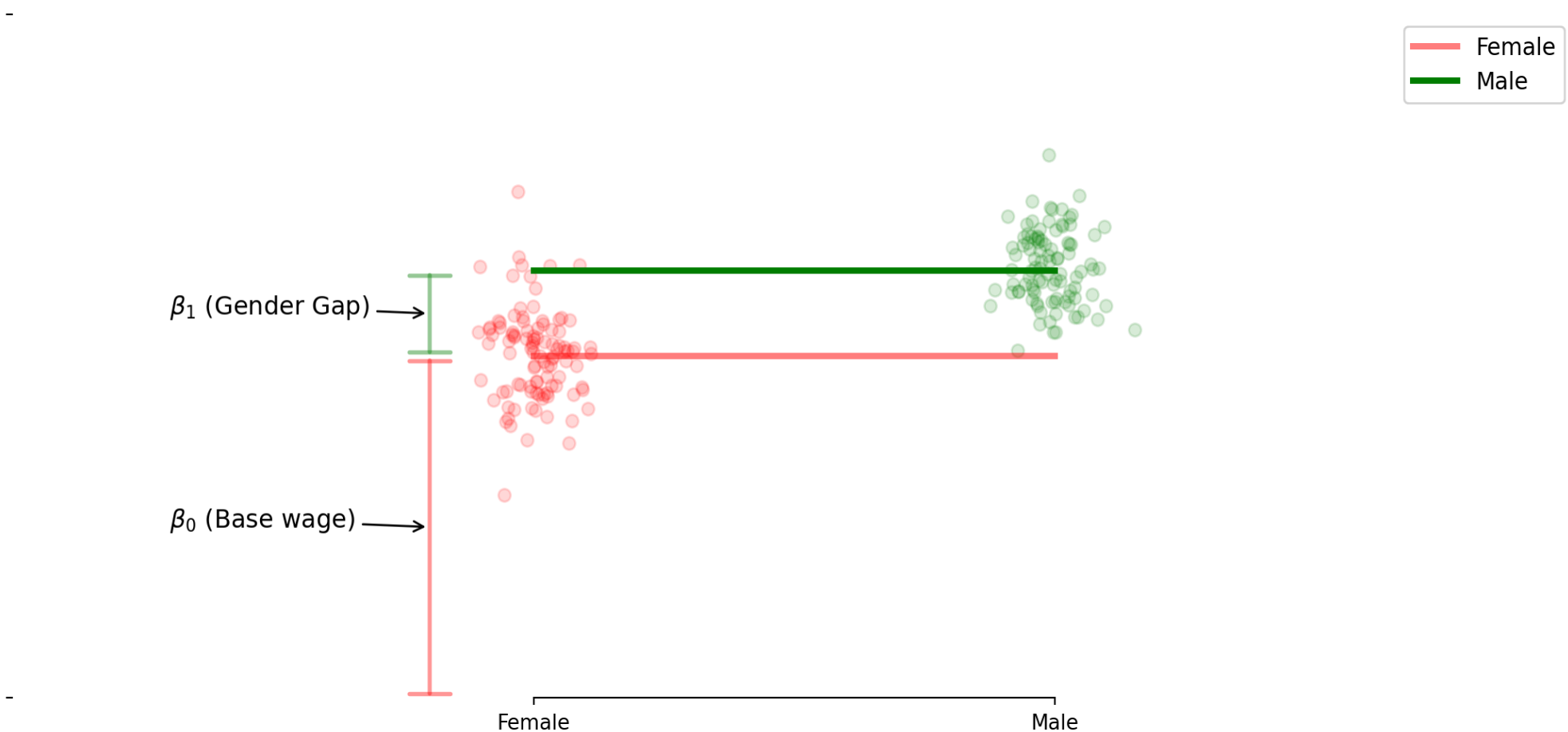
Model 1: Basic Gender Wage Gap

The simplest model with just a gender indicator.



Model 1: Basic Gender Wage Gap

The simplest model with just a gender indicator.



Model 1: Basic Gender Wage Gap

The simplest model with just a gender indicator.

- β_0 *is the average wage for females*
- β_1 *represents the gender wage gap - the additional wage for males*

> *model ignores education entirely - just comparing average wages by gender*

Model 1: The Code

Implementing the basic gender gap model

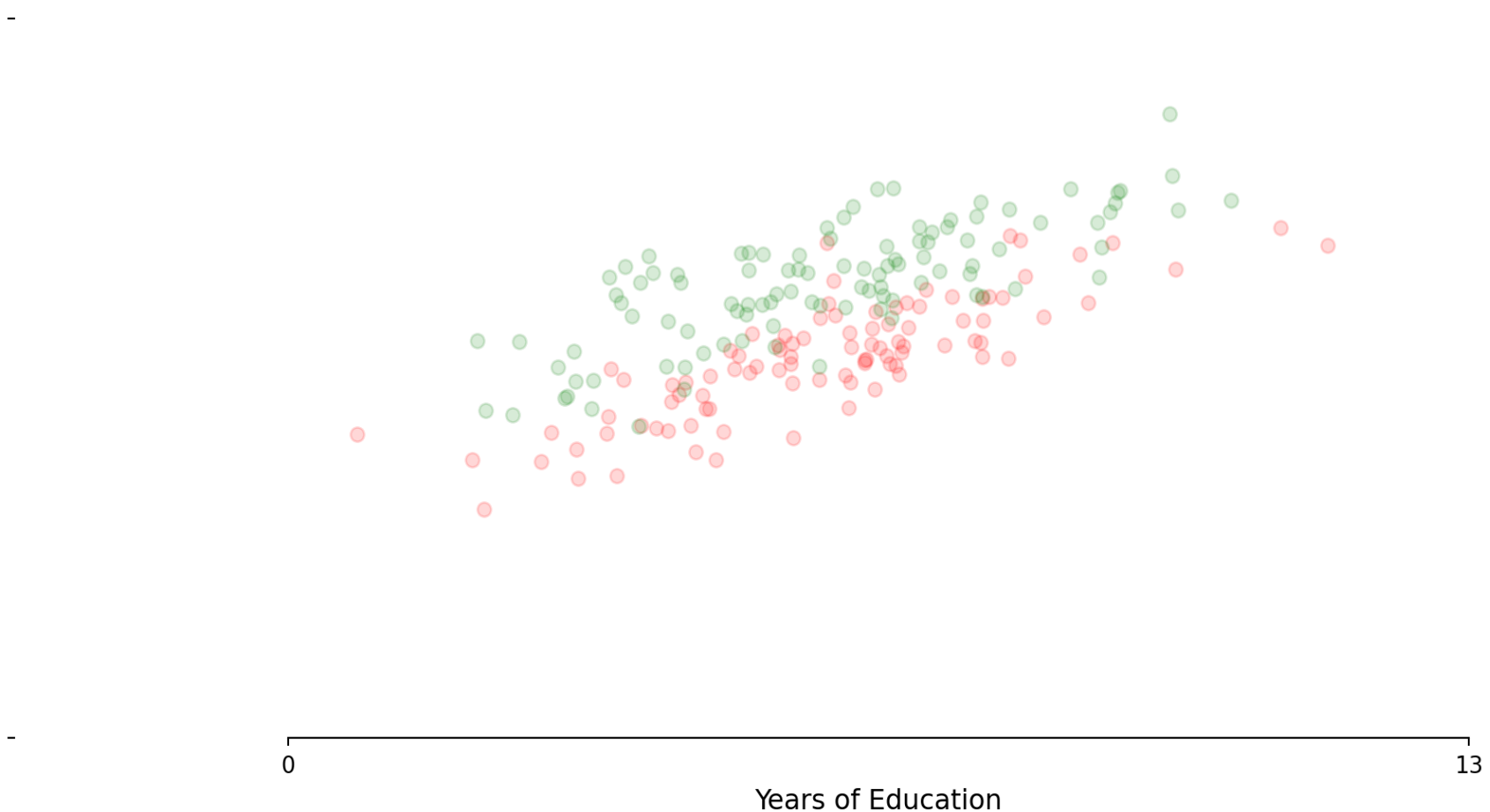
```python
import statsmodels.formula.api as smf

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

> if $\beta_1 > 0$ and statistically significant, evidence of a raw gender wage gap

> this model doesn't control for any other factors that might explain wage differences

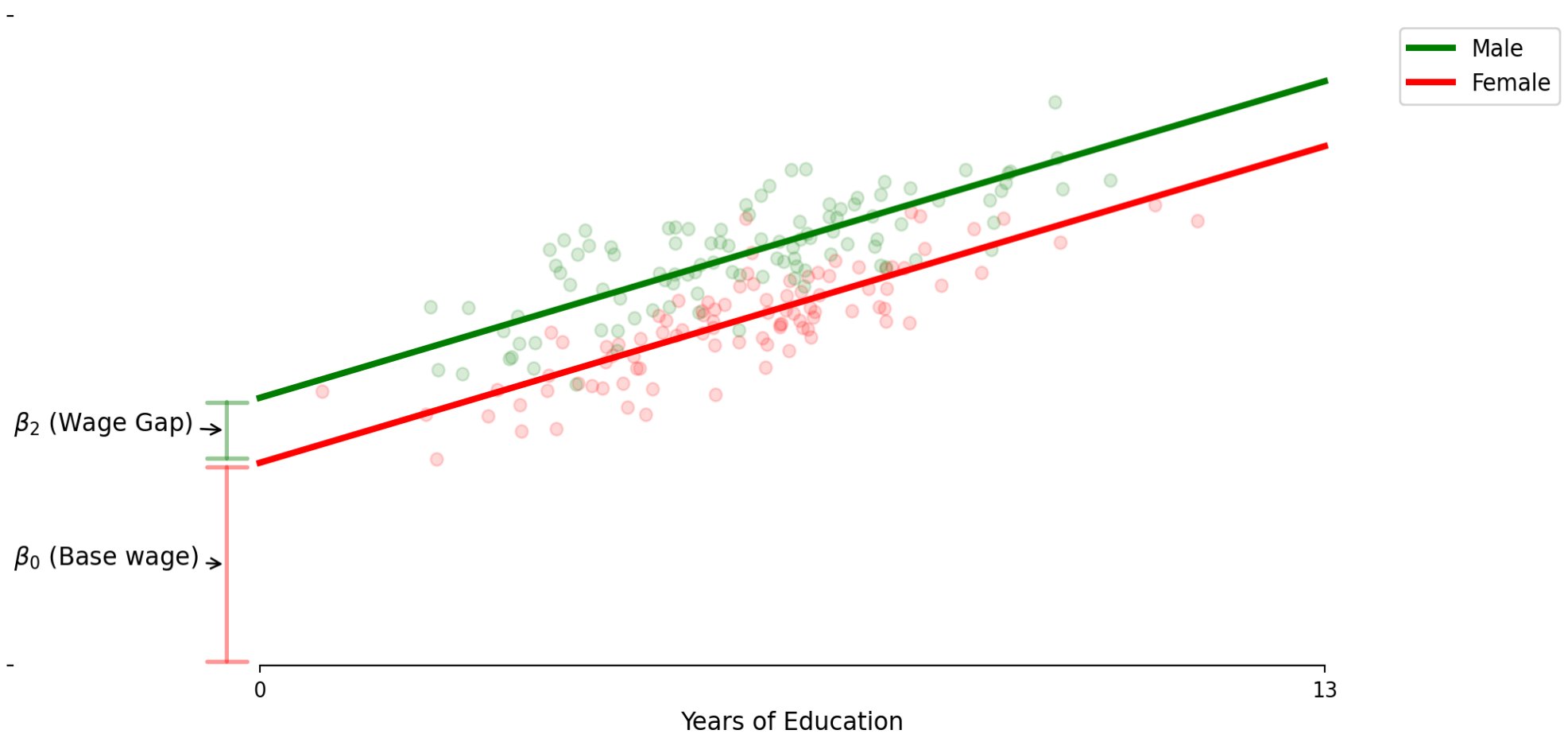
Model 2: Education + Gender Wage Gap

Adding education as a control variable.



Model 2: Education + Gender Wage Gap

Adding education as a control variable.



Model 2: Education + Gender Wage Gap

Adding education as a control variable.

> *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 2: The Code

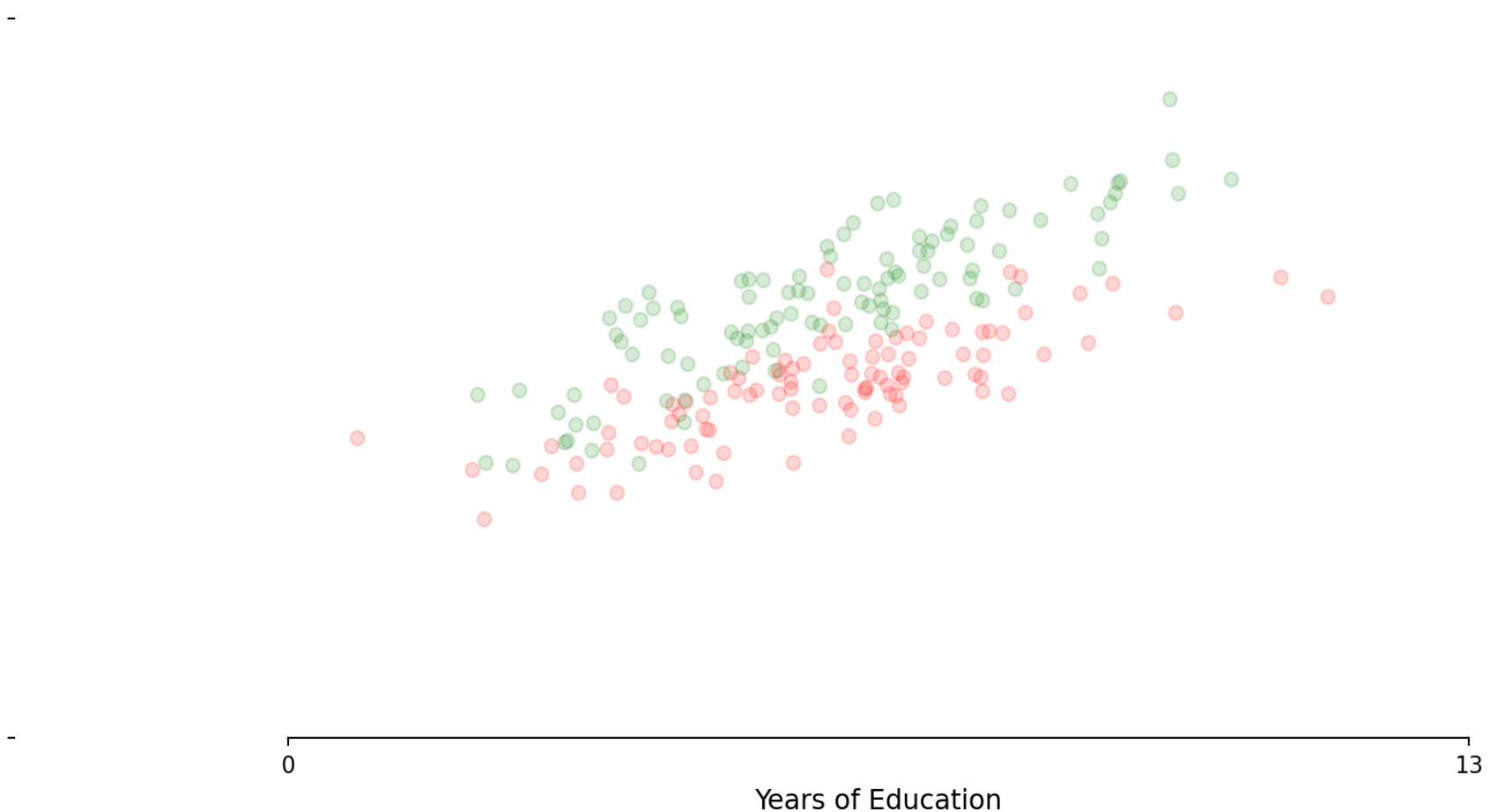
Implementing the gender fixed effect model

```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())
```

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

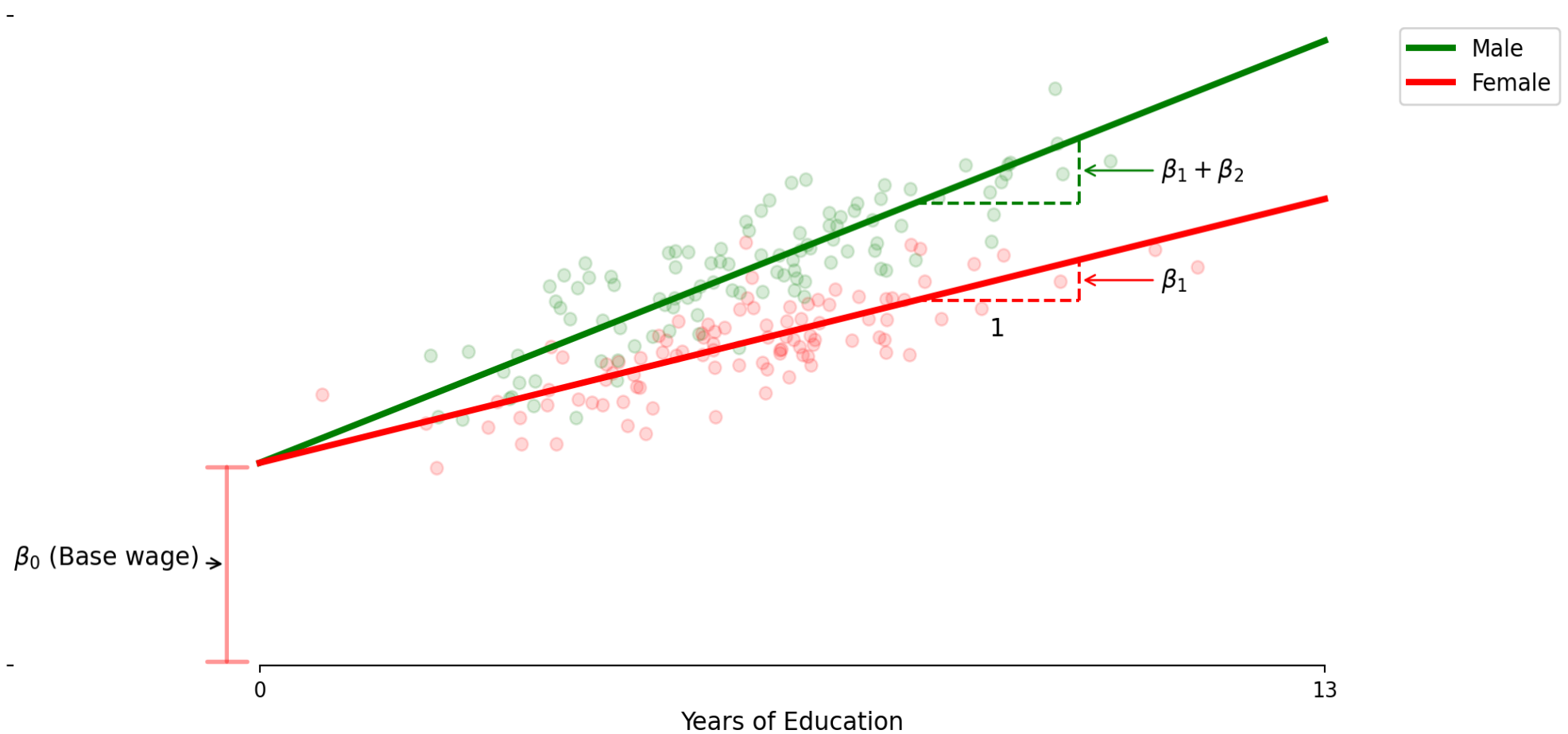
Model 3: Different Returns to Education

What if education benefits genders differently?



Model 3: Different Returns to Education

What if education benefits genders differently?



Model 3: Different Returns to Education

What if education benefits genders differently?

> *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 3: The Code

Implementing the education-gender interaction model

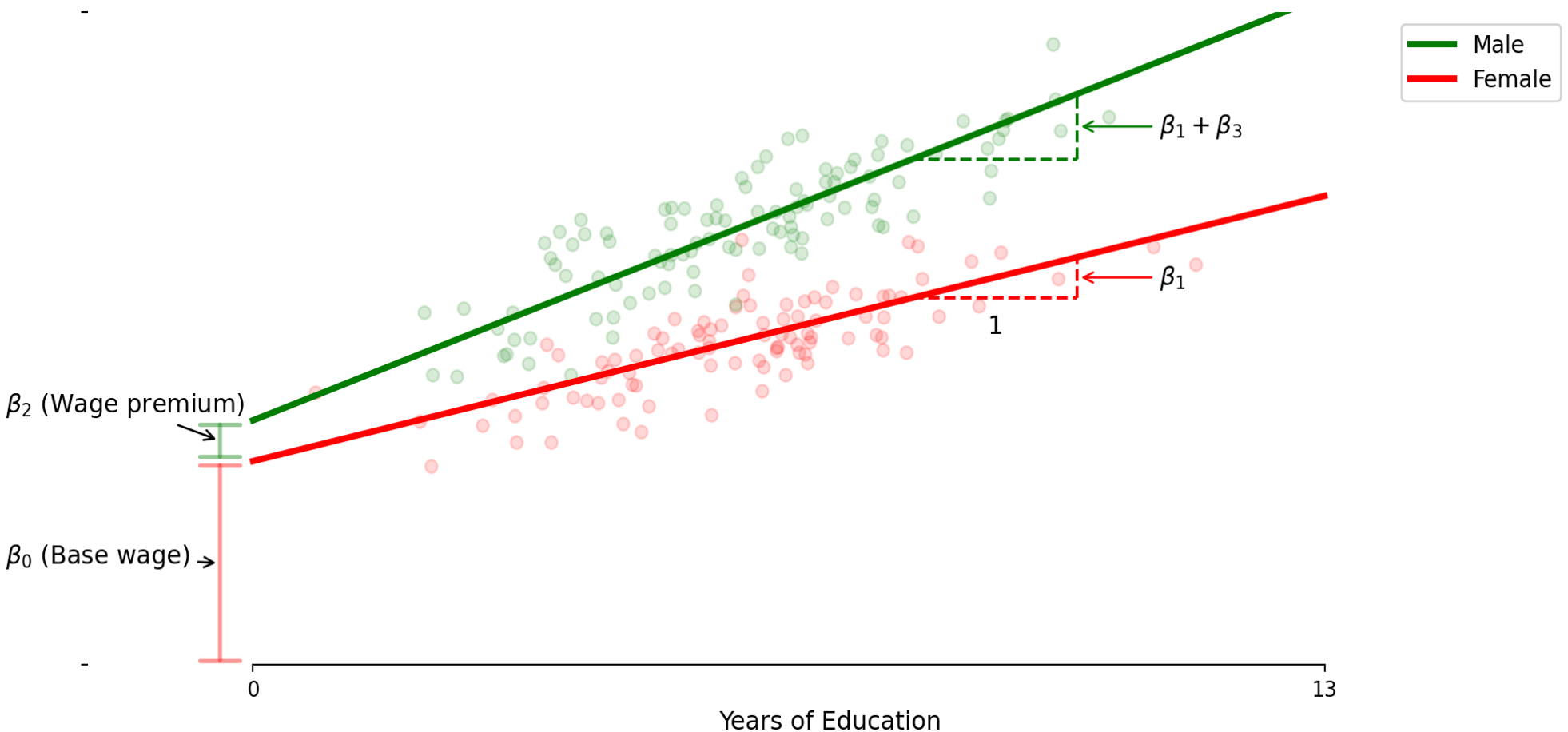
```
# Fit model with interaction between education and sex
model2 = smf.ols('INCLOG10 ~ EDU + EDU:MALE', data=df).fit()
print(model2.summary())
```

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

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

Model 4: Full Gender Difference Model

Combining fixed effects and interactions



Model 4: Full Gender Difference Model

Combining fixed effects and interactions

> 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 4: The Code

Implementing the full gender difference model

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

> *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?"*

2. ***Model 2: Fixed Effect with Control***
    - *Question: "Is there a gender wage gap controling for education?"*

3. ***Model 3: Interaction Only***
    - *Question: "Are there differences in returns to education?"*

4. ***Model 4: Full Model***
    - *Question: "Does the gender wage gap vary with education level?"*

> choose the model that best addresses your research question

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