

Exploring Variation and Interaction Terms

ECON 490

Taylor Mackay || Email: tmackay@fullerton.edu

Slides Overview

In these slides, we'll discuss:

- Interaction terms in regression analysis
- An application from an R in-class activity using the capstone data sets

Exploring Variation with Interaction Terms

Regression models typically estimate average relationships across all observations

- For example, in $Y = \beta_0 + \beta_1 X + u$, everyone shares the **same** β_1
- What if this relationship varies across groups, regions, or time periods, etc.?

Interaction terms allow us to capture and quantify these differences

Let's consider the following question: *"Does the relationship between unemployment and crime differ across regions of the United States?"*

Baseline Model: Unemployment and Property Crime

Let's start with the following:

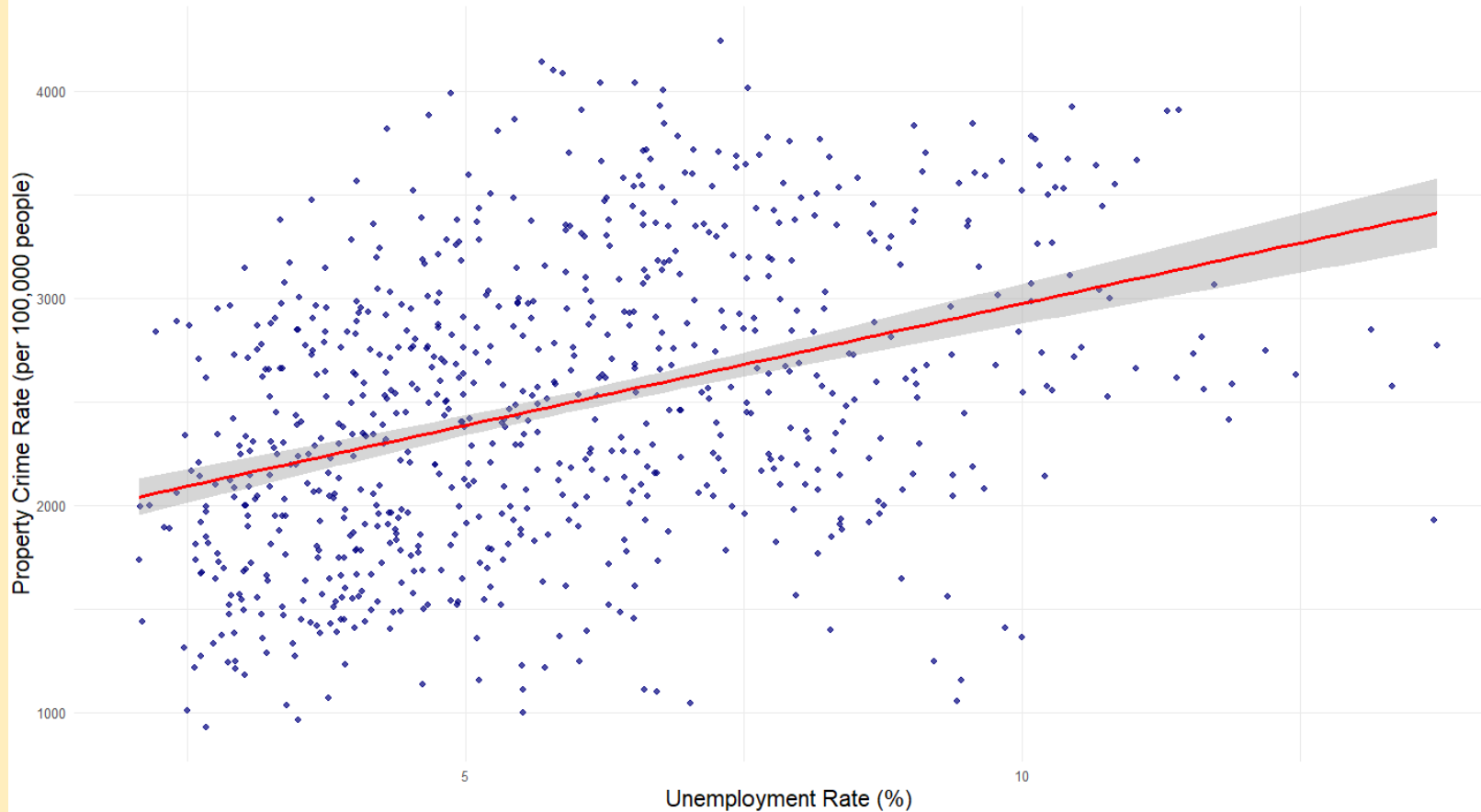
$$\textit{Property Crime Rate} = \beta_0 + \beta_1 \textit{Unemployment Rate (UR)} + u$$

In words, we want to predict or generate conditional means of crime given UR

- Key point – our prediction does **not** depend on what state you consider
- I.e., given $UR = 4\%$, we'll get the same prediction for California and Arkansas

Relationship Between Unemployment and Property Crime

State-level data, 2008-2022

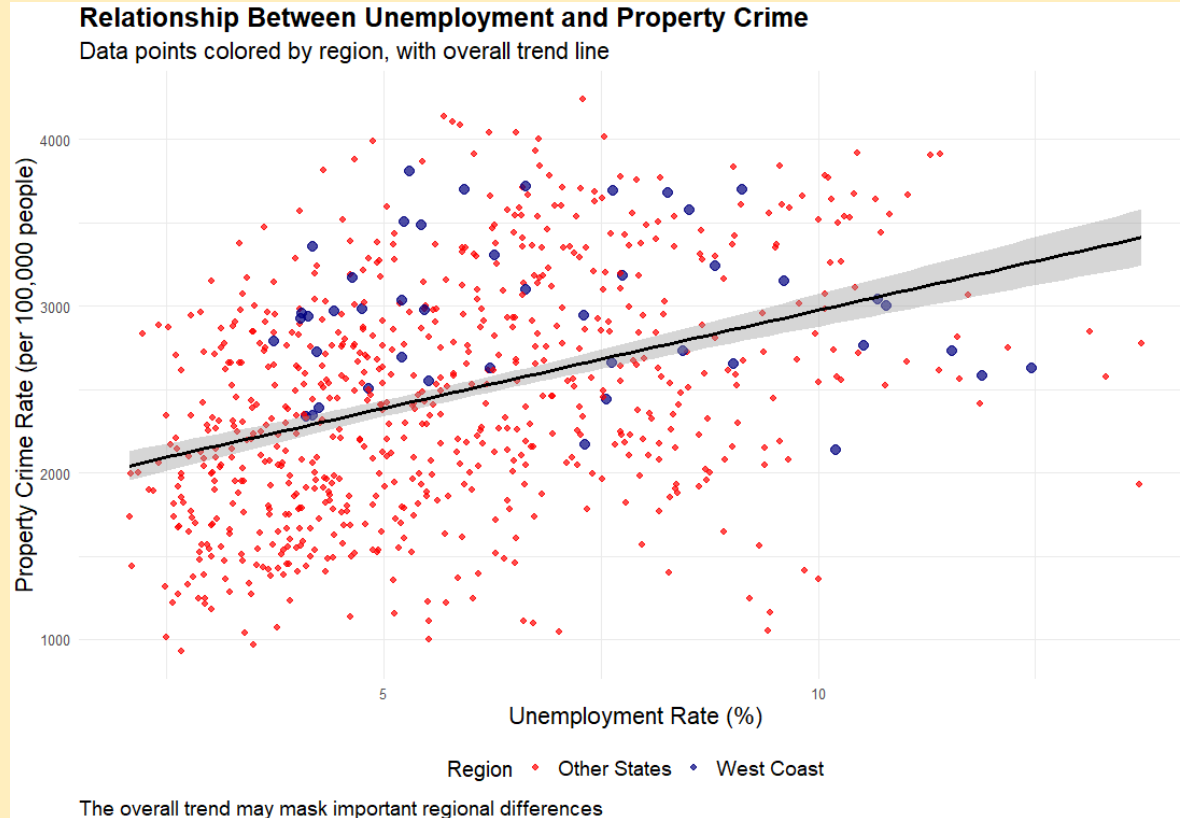


A 1 percentage point increase in unemployment is associated with
~118 more property crimes per 100,000 people

Does the Relationship Vary by Region?

Bolded points represent
CA + WA + OR

Notice they don't quite
follow the overall trend



Exploring Variation

We can use an interaction term to see how this relationship varies

- Create a binary indicator for being a West Coast (WC) state
- Include 1) this new variable and 2) the interaction of this variable and UR:

$$\text{Crime Rate} = \beta_0 + \beta_1 UR + \beta_2 I(\text{West Coast}) + \beta_3 UR \times I(\text{West Coast}) + u$$

Now, we get **two** predicted values for crime rates given a value for UR

- Different predictions for WC and non-WC states
- Make sure you can calculate both values by setting $WC = 0, 1$

Interactions in R

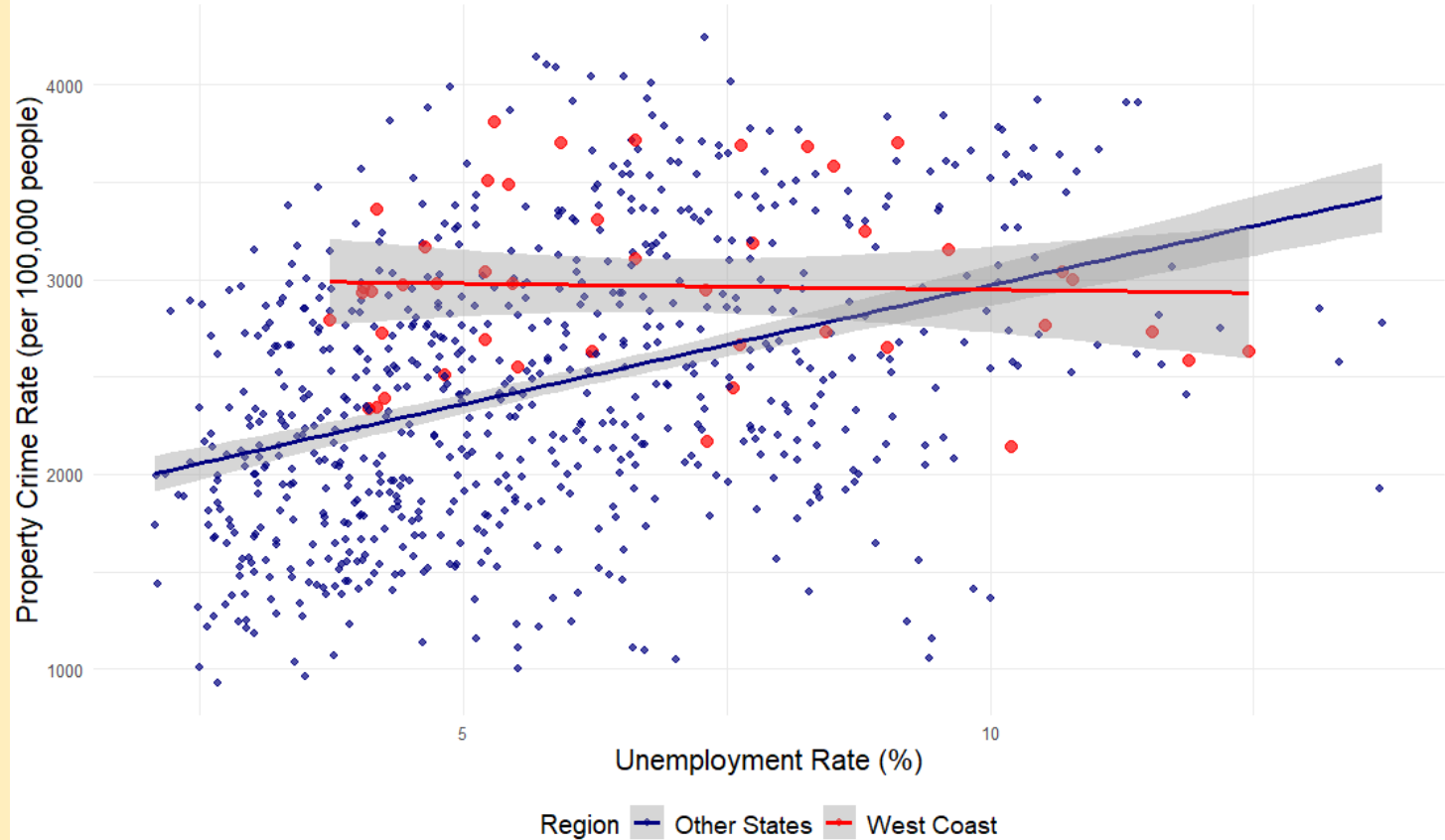
Run this regression using “:” in R: `lm(crime ~ UR + WC + UR:WC, data)`

- This tells R to create the interaction of *UR* and *WC*
- Intuitively, imagine R adding two new columns to our data:

| State | Year | Crime Rate | Unemployment Rate (%) | $I(WC)$ | $UR \times I(WC)$ |
|------------|------|------------|-----------------------|---------|-------------------|
| Arizona | 2010 | 1,700 | 4 | 0 | 0 |
| Alabama | 2015 | 2,100 | 3 | 0 | 0 |
| California | 2017 | 2,000 | 6 | 1 | 6 |

Regional Differences in Unemployment-Crime Relationship

Separate trend lines for West Coast vs. Other States



Note the different slopes between regions, suggesting varying relationships between unemployment and crime

Things to Remember

Interactions allow you to flexibly explore variation across groups

- Technically, you can interact two continuous variables
- However, for this class, we want groups to be defined by factor variables

Ideally, this factor variable is binary or has a “limited” number of levels

- How many levels is too many? Context matters – think about interpretation

SOMETHING TO REMEMBER: Interactions increase the amount of data you need

- Too little data and we’ll lose precision relative to “vanilla” OLS
- Key point – need data across groups (i.e., observations in both WC & non-WC)