Methods Comprehensive Exam
September 5, 2025

Instructions: Read all questions before answering any of them. You have 6 hours to complete the exam. When you use substantive examples in your answers, we prefer to see examples from political science. Answer all questions in Part I. Answer 2 questions in Part II. Answer 2 questions in Part III. If you are completing the exam at GW, feel free to hand-write answers or parts of answers in a blue book, but carefully label them and note that you are using the blue book in your typed document. If you are completing the exam at home, you can similarly include photos/scans of hand-written material. Good luck!

This exam is open book (including articles from class) and notes. You may use a calculator or R for certain calculations. But you are not allowed to consult any other resource on the Internet, including chatGPT or other AI-based chat-bots.

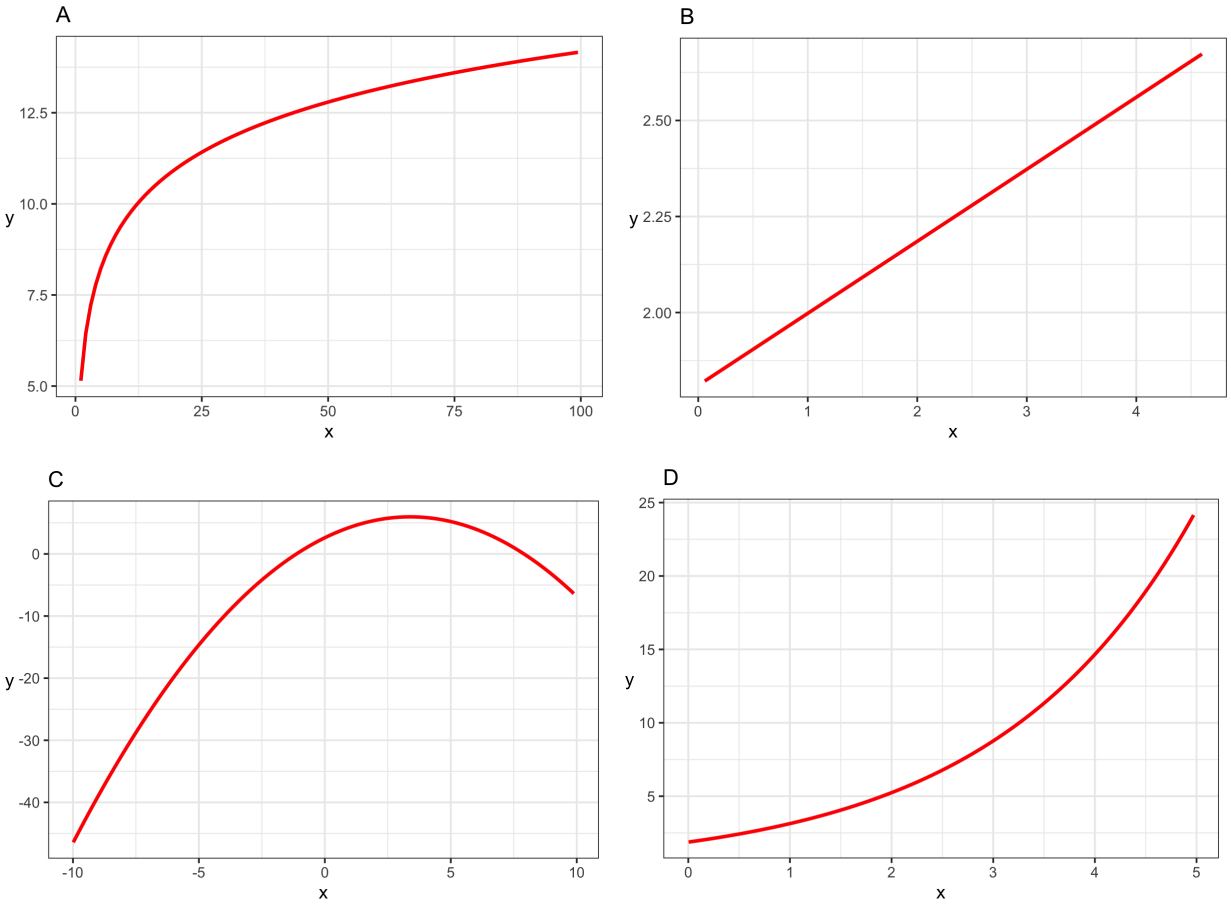
Part I

In this section, please answer ALL questions.

1. In a recent public opinion survey of Americans:
    - 51.43% of respondents were Democrats, 45.41% were Republicans.
    - 26.3% were liberal, and 36.2% were conservative.
    - 25.7% were both liberal and Democrat
    - 31.3% were both conservative and Republican
    - 0.3% were both liberal and Republican
    - 4.5% were both conservative and Democrat

    a. What is the probability of being a "partisan" (i.e., a Democrat or Republican)? Indicate what assumption you make.
    b. What is the probability of being an "ideologue" (i.e., liberal or conservative)? Again, indicate what assumption you make.
    c. What is the probability of being liberal or a Democrat? Indicate the assumption you make.
    d. What is the probability of being conservative or a Republican? Indicate the assumption you make.
    e. In sum, which "events" are disjoint and which are not disjoint?
    f. In the bullet-point description above, which are related to joint probabilities, and which are related to marginal probabilities?
    g. Estimate the following conditional probabilities: P(Democrat | Liberal) and P(Democrat | Conservative). Interpret what you find in comparing these two probabilities.
    h. Estimate the following conditional probabilities: P(Republican | Conservative) and P(Republican | Liberal). Interpret what you find in comparing these two probabilities.
    i. In general, how would you describe the connection between ideology and party in the U.S.?

2. Consider the following functional forms for the relationship between x and y:



a. Identify each functional form and describe the nature of each in substantive terms.
b. For each functional form, write out the OLS equation that can model each relationship.
c. For each plot, write out the equation for the marginal effects of x. Show your work. For each, describe what the marginal effects "are doing" as x increases from minimum to maximum.
d. For plot C, identify what the sign (positive or negative) of each regression coefficient would be and explain how you know.

3. Banks and Hicks (2019, *AJPS*) test whether political campaign ads that allude to race in the 2016 presidential campaign (between Trump and Clinton) decrease support for Trump among white voters. The authors run a survey experiment where all subjects watch a campaign ad described as follows (from Banks and Hicks, p. 309):

> berg 2001; Valentino, Hutchings, and White 2002). The ad makes no direct reference to race but includes visual images of blacks to communicate a racial message undercover. In the ad, Trump emphasizes the need to regain law and order in our nation's inner cities while showing images of young black men. In a voice-over, Trump states, "we have to have law and order. Hundreds of killings are in Baltimore. Hundreds of killings are in Chicago, and New York is not doing so great in terms of that front, and so many other cities," while an image of young black men rioting in the streets appears on the screen. Trump goes on to say, "we have to give strength and power back to the police," while the ad displays images of white police officers in riot gear holding nightsticks. He concludes by saying, "police have to regain some control of this tremendous crime wave and killing wave that's happening in this country," which is accompanied by an image of looting blacks.

Respondents are randomly assigned to one of two experimental groups: (1) the control group, which only watches the ad with *implicit* racial appeals, and (2) a treatment group, which watches the ad but also reads a response to this campaign ad by a politician who *explicitly* calls out Trump for race baiting:

> further quoted as saying that "Trump should be ashamed of letting an ad like this run.... We believe all people are created equal. This is fundamental, and if someone wants to be president of our great nation, they must understand this. This kind of race baiting is shameful and Trump knows better." As in the implicit condition, the

The core **independent variable** is the treatment variable: 1= "explicit politician condition" (the treatment, where a politician calls out Trump for race baiting), 0=implicit condition only (control). The authors posit "heterogenous treatment effects" conditional on two forms of individual (and measured) racial views:

(1) *Racial resentment*, which taps the degree of white resentment toward blacks on issues of equality; conservative responses reflect whites thinking blacks need to pull themselves up by the bootstraps (i.e., "do more") in order to achieve equality, while liberal responses acknowledge that blacks face systemic societal obstacles for achieving equality. The variable is based on a multi-item scale and ranges from 0 (racially liberal) to 1 (racially conservative).

(2) *Old-fashioned racism*, which draws on racial stereotypes and more explicit acknowledgements of white superiority over blacks. The variable is based on a multi-item scale and ranges from 0 (racially liberal) to 1 (racially conservative).

The **dependent variable** is a 101-point thermometer rating to evaluate how warmly or coldly voters feel about Trump. The variable ranges from 0 (very cold, or disapproving) to 100 (very warm, or approving); 50=moderate approval. Once again, all respondents are white. Model results are included below in Table 1. For this dependent variable, the authors report OLS estimates for model 3 only. But in addition, Table 1 below includes models 1 and 2:

Table 1: Model Results

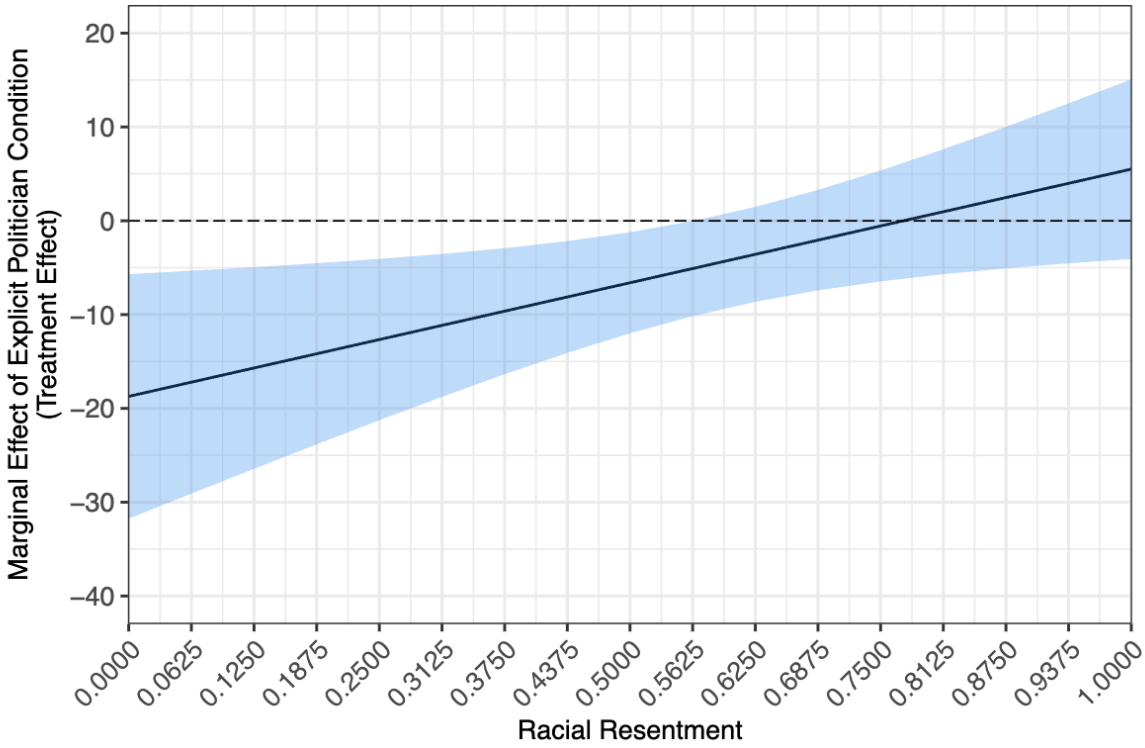
	(1)	(2)	(3)
Explicit Politician Condition (Treatment Variable)	−4.68	−18.72**	−19.97**
	(2.80)	(6.64)	(6.60)
Racial Resentment		37.11**	30.26**
		(9.22)	(9.48)
Explicit Politician Condition × Racial Resentment		24.23*	22.71*
		(10.28)	(10.58)
Old-Fashioned Racism			21.98*
			(9.21)
Explicit Politician Condition × Old-Fashioned Racism			4.56
			(10.28)
Intercept	47.65**	25.31**	23.05**
	(2.50)	(5.96)	(5.92)
Adj. R^2	0.00	0.17	0.20
N	967	961	958

$^{**}p < 0.01$; $^{*}p < 0.05$. OLS regression coefficients, standard errors in parentheses.

a. What is the "average treatment effect" (for the explicit politician condition variable)? Explain where you got your answer. Interpret this effect in terms of direction, size, and statistical significance.
b. From model 2, write out the equation for the marginal effect of the treatment variable ("explicit politician condition"). Use the results to answer whether racial resentment moderates the treatment effect. Make sure to discuss the direction and magnitude of the moderation and whether it is statistically significant.
c. From model 3, write out the marginal effect of the treatment variable. Note that it is interacted with both racial resentment and old-fashioned racism. What is the difference between your answer here and your answer in part b. Discuss the difference between models 2 and 3 in how one is modeling heterogeneous treatment effects.
d. What does the constituent term for the treatment variable mean in models 2 and 3? Be specific and compare and contrast.

e. Figure 1 includes a post-estimation plot from model 2. Describe what this model is testing and interpret all aspects of the results in sufficient detail.

Figure 1: Heterogeneous Treatment Effects



f. Let's assume that the treatment effects are (approximately) normally distributed across the racial resentment scale and that the mean treatment effect is centered on a racial resentment value of .55, with a standard deviation of .26. Based on this information, about what percentage of respondents have *negative* treatment effects? Describe/show how you got your answer.

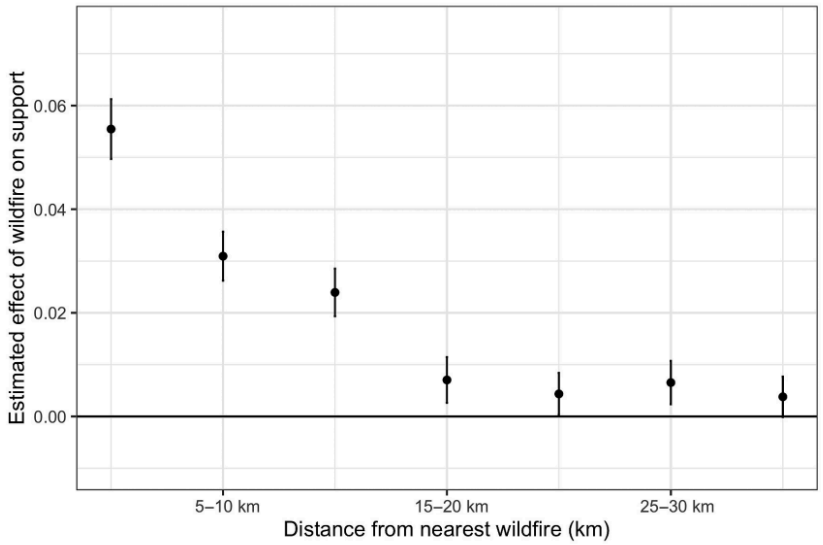
g. Discuss how this regression specification relates to OLS assumptions. Do you think it violates any assumptions? Does it particularly satisfy others?

Part II

In this section, please answer TWO of the following questions.

1. Hazlett and Mildenberger (2020) consider the question of whether direct exposure to climate-related threats affects voting behavior related to environmental issues. To do so, they use a two-way fixed effects design, regressing census block-group-level support of various California propositions related to the environment on a binary indicator of wildfire exposure (e.g. whether a fire occurred within a given distance from the block-group boundaries) in the last two years, including unit (block group) and time (year) fixed effects and controls for potential time-varying confounders (lagged Democratic vote-share and 2-year precipitation trends). They plot the resulting effects at different distances, shown in Figure 2a below.

a. What would be the threat to causal identification without the unit and time fixed effects? E.g. why not just run the naive regression of proposition support on wildfire exposure?
b. If the authors already included unit and time fixed effects, why did they also include additional time-varying, block-group covariates? (Hint: they didn't do so incorrectly. What confounding would we be worried about in this design?)
c. Given this design, provide 1 additional threat to causal identification. You may give a general threat to TWFE designs, but explain how it may occur in this context.
d. Given the evidence in Figure 2a, does wildfire exposure change pro-environmental voting? Explain why or why not.



Note: Estimates compared with response at the median distance (35–40 kilometers). All estimates derived from a linear model with block-group and year fixed effects and controlling for Democratic vote share in Congressional elections four years prior. Error bars show 99% confidence intervals, using standard errors clustered on block group.

Figure 2a

2. Weather (e.g. rainfall, heatwaves) is popularly used as an instrumental variable in social science research. For example, a famous paper by Acharya, Blackwell, and Sen (2016) contends contemporary racial attitudes in the American South are legacies of slavery, for which they use climatic suitability for cotton growing as an instrument.

a. Consider 2 research designs each using an instrumental variable approach: $Z_i \rightarrow X_i \rightarrow Y_i$, where *i* indicates the research design, Z the instrument, X the treatment, and Y the outcome. Say Z_i is the same across all studies: rainfall ($Z_1 = Z_2$). Is there an identification issue if $X_1 \rightarrow Y_2$? Why or why not?
b. A graduate student suggests an instrumental variable design to causally identify the effect of election-day individual sadness on vote choice for the incumbent party. They suggest using rainfall as an instrument, arguing rain makes people sad and is plausibly exogenous. Another student, however, notes that some people like the rain; it makes them happy, not sad. Say that this student is correct, and rainfall can both increase or decrease mood. Is this a threat to the causal identification of the proposed research design? Explain.
c. Considering the same proposed design as in 3(b), another student raises a concern that the researcher is only considering voting behavior in the area around Seattle, Washington, an

area known for frequent rainfall. They argue rainfall might fail to affect the mood of those who are accustomed to it. Is this a threat to identification? Why or why not?

d. Finally, another researcher is interested in the effect of soldiers' morale on battlefield performance (measured as acquisition of enemy-occupied territory) in civil wars. They argue rainfall could be used as an instrument for such morale. However, an objection is raised that mountainous terrain has been shown to affect both weather patterns and battlefield performance in previous studies. Is this a threat to identification? Why or why not?

3. Suppose you are studying whether increases in the minimum wage reduce crime rates. In 2015, one U.S. state enacted a large minimum wage hike, while most other states did not. You propose using a synthetic control method (SCM), where the "treated" unit is the state that raised its minimum wage and the "donor pool" consists of other U.S. states that did not.

a. Briefly explain how the synthetic control method works and how it differs from difference-in-differences.
b. What key assumptions are necessary for SCM to yield valid causal inferences in this application?
c. Describe two diagnostic checks or robustness tests you would conduct to evaluate whether the SCM results are credible.
d. Suppose you find that the treatment state shows a clear decline in crime rates after the minimum wage increase, but the results are highly sensitive to the inclusion of one large donor state in the synthetic control. How would you interpret this sensitivity? What steps could you take to strengthen the robustness of your findings?

4. In a recently published paper, Carles Boix examines how political emancipation shaped modern Jewish national identity in Eastern Europe. To identify causal effects, he employs a regression discontinuity design (RDD) based on the boundary of the Pale of Settlement in the Russian Empire, treating communities just inside the Pale as "treated" (limited emancipation) and those just outside as "control" (greater emancipation).

a. What are the key assumptions required for a geographic RDD to produce valid causal inferences?
b. Identify at least two potential threats to the validity of the RDD in this setting.
c. What diagnostic tests could you conduct to evaluate the plausibility of the RDD assumptions?
d. If the RDD assumptions are not fully satisfied, what alternative research designs could you use to study the relationship between political emancipation and Jewish national identity?

Part III

In this section, please answer ONE of the following questions.

1. Consider a researcher interested in the homogenization of political debates across geographic subdivisions. The researcher contends that the decline of local media and heightening of national polarization has resulted in politics being functionally identical regardless of where you might be in the country. The researcher considers measuring how elected officials across geographic subdivisions discuss politics. They propose using a

topic modeling approach to descriptively show rates of topic discussion across the geographic subdivisions.

a. In detail, discuss the features of a topic model design. That is, what quantities do topic models generate, and how can they be interpreted?
b. Part of this project requires acquiring raw textual data. Describe a potential way the researcher may acquire relevant data, and discuss how that process may influence the inferences made from the topic model.
c. How might one implement such a design? Specifically, data pre-processing, formatting (e.g. what is an observation, what is a column, what is a token, etc.), and any other relevant technical details.
d. How might you validate the output of the topic model? Give an example of a situation where you may be skeptical of the model's results.
e. Propose an alternative to a topic-modeling approach for textual data. Give one reason why this alternative may be superior, and one for why it may be inferior.

2. Imagine that you conduct a field experiment to test whether providing information about community safety increases participation in local policing meetings. You randomly assign 1,500 residents to receive a text message with detailed information about recent crime rates and police initiatives (treatment) and 1,500 residents to receive a neutral reminder about the date and location of the meeting (control). On the day of the meeting, you record actual attendance from sign-in sheets (the main outcome). However, when you attempt to collect a short follow-up survey one week later (measuring secondary outcomes such as trust in police and perceptions of safety), you find that 35% of the treatment group and 20% of the control group fail to complete the survey.
    a. Why is differential attrition in the follow-up survey problematic for drawing inferences about the secondary outcomes? Under what conditions would attrition bias your causal estimates?
    b. What diagnostic tests or checks could you perform to evaluate whether attrition is random or systematically related to treatment assignment or covariates?
    c. Suppose the differential attrition rates vary significantly across participants' characteristics. For example, you discover that college-educated participants show higher attrition than those without college degrees. How does this pattern of heterogeneous attrition complicate your analysis?
    d. Discuss at least two methodological strategies for addressing this attrition problem (e.g., bounding approaches, weighting, imputation, re-contact efforts). For each, explain the assumptions, strengths, and limitations.
    e. Suppose administrative records confirm meeting attendance for all 3,000 participants, regardless of survey response. How does this change the severity of the attrition problem? How would you leverage this fact in your analysis?

3. Suppose you want to estimate crime rates for small geographic units (e.g., neighborhoods or census tracts), but official crime data are only available at the county level. You propose using multilevel regression and poststratification (MRP), combining (1) county-level crime statistics, (2) demographic and socioeconomic data from the census at the tract level, and (3) a multilevel model that relates tract-level demographics to county-level crime.

a. Explain how the two stages of MRP would work in this setting. What types of predictors would you want to include in the regression model?
b. What assumptions are necessary for the MRP estimates of neighborhood-level crime rates to be valid? Discuss at least two potential threats to validity in this application.
c. Suppose you estimate crime rates for all census tracts in a metropolitan area. How would you assess whether the MRP predictions are credible, given that you only observe crime data at the county level?
d. One potential concern is that crime rates often exhibit spatial autocorrelation (neighboring areas tend to have similar crime levels). How would you modify the MRP approach to account for this?
e. Imagine you also gain access to limited administrative crime data for a random sample of neighborhoods. How could you use these data to evaluate and potentially improve the MRP estimates?