Department of Economics
MIT
14.32/14.320 Fall 2021

Final Exam
1.30 pm., Tuesday December 14, 2021

Part I: The Benefits of Attending Class

At a certain small state college, admitted students have the choice of taking a class in person or by online delivery only. This part examines results from a recent study at this college, which compares final exam performance of online and in-person students in an intermediate macroeconomics course.

The Web and in-person classes were taught by the same instructor using the same course materials (posted lecture notes, problem sets, textbook). All students had been admitted into the college's undergraduate degree program, and the students took the same exam at the same time and place (the final exam sitting was the only time that the online students showed up in person). The final exam consisted of questions chosen at random from the test bank accompanying the course textbook. The main difference between the two classes was that that the in-person class had traditional classroom delivery, whereas the Web class had access to the instructor only by email, telephone, and an electronic "discussion board" (chat room) that was for Web students only.

The unit of observation is the student, with n = 109; of these, 31 chose the Web version and 78 chose the traditional version. Regression results are given in Table I.

Table I. Determinants of Web Class Choice (*Web*) and Final Exam Grade (*Grade*)

		(1)	(2)	(3)	(4)
	Dependent variable	***Web***	***Web***	***Grade* (0-100 points)**	***Grade* (0-100 points)**
	Estimation Method	Probit	OLS	OLS	2SLS
Regressors	**Variable definition**				
Web	= 1 if Web delivery = 0 if traditional delivery			-2.323 (2.13)	-19.179** (6.300)
Prior economics	= 1 if took a prior economics course = 0 otherwise	.051 (.031)	.031 (.044)	.187 (1.89)	-.491 (1.83)
Male	= 1 if male = 0 if female	.121** (.030)	.072** (.021)	-.076 (1.67)	-2.688 (1.85)
White	= 1 if white = 0 otherwise	.064 (.055)	.041 (.031)	-1.555 (2.77)	3.208 (3.15)
Age	Age in years	-.021** (.007)	-.0050* (.0022)	.556* (.289)	1.151** (.35)
Prior course hours	Total course hours earned by the start of the semester		-.015 (.026)	.034 (.035)	.107** (.04)
GPA	Overall GPA at start of semester		-.066* (.031)	6.787** (1.88)	2.319 (2.41)
ACT	ACT score (college admission standardized test)		-.0012* (.0005)	1.086** (.283)	1.518** (.31)
Hours worked	Hours worked per week for pay		.016** (.003)	-.069 (.061)	.091 (.08)
Father college degree	= 1 if father has a college degree = 0 otherwise		.023 (.040)	3.713** (1.732)	2.978* (1.68)
Mother college degree	= 1 if mother has a college degree = 0 otherwise		-.005 (.042)	2.595 (1.709)	3.064* (1.64)
Business major	= 1 if business major = 0 otherwise		.021 (.038)	.586 (1.819)	.800 (1.75)
On campus	= 1 if student lives on campus = 0 otherwise		-.121** (.031)		
Web whiz	= 1 if student reports self as being proficient on Web, = 0 otherwise.		.092** (.031)		
Constant		-.100 (.102)	.710** (.050)	14.336 (9.94)	1.902 (10.52)
F-statistic (*p*-value) testing zero coefficients on *On campus* and *Web whiz.*			11.21 (.0001)		
R2			.130	.462	
N		109	109	109	109

Notes: The probit regression was estimated by maximum likelihood. The 2SLS regression uses as instruments *On campus* and *Web whiz*. Standard errors are heteroskedasticity-robust.
* significant at 5%; ** significant at 1% (using the standard errors in parentheses).

Questions for Part I (45 points)

Please answer these questions in Blue Book 1

The questions in Part I refer to Table I.

1) Consider a student – call him Joe – who has no prior economics courses, is a 20-year old white male with 10 prior course hours, has a GPA of 3, an ACT score of 20, who is not working for pay, whose parents did not go to college, who is not a business major, who lives off campus, and who is not a Web whiz.

(3 points) Using **regression 1**, compute the predicted probability that Joe will take the course by Web delivery.

For the probit model,
$Pr(Web=1|X) = \Phi(.051*0+.121*1+.064*1-.021*20-.100) = \Phi(-.335) = .369.$

(3 points) Using **regression 1**, estimate the difference in the probability of taking a course by Web delivery for Joe and for an otherwise identical student who is ten years older.

For the student ten years older,
$Pr(Web=1|X) = \Phi(.051*0 + .121*1 + .064*1 - .021*30 - .100) =\Phi(-.545)=.293$
so the predicted difference is .293 - .369 = -.076. The older student is 7.6 percentage points less likely to take the course via the Web.

(3 points) Using **regression 2**, compute a 95% confidence interval for the difference in the probability of taking a course by Web delivery for Joe and for an otherwise identical student who is ten years older.

In the linear probability model, the change in the predicted probability is just the coefficient on Age, times 10, so the 95% confidence interval is

$$-.0050*10 \pm 1.96*.0022*10 = -.050 \pm .043 = (-.093, -.007)$$

2) (3 points) Consider **regression 3**. Suggest an omitted variable that would imply that the coefficient on *Web* would be a biased estimate of the treatment effect (the causal effect on the final exam grade) of taking a course by Web delivery.

*If the **Web** students are different from the regular students in ways that affect **Grade**, then this will result in omitted variable bias. For example, suppose that the **Web** students are more interested simply in taking the course to learn something, and less interested in grades, than the regular students (who might be considering applying to graduate school). If so, the regular students might study harder than the **Web** students, so that effort in the course would be omitted. Because effort in the course would be a determinant of **Grade,** the two criteria for omitted variable bias (correlated with X, and a determinant of Y) would be satisfied. Note that in this example, **Web** is negatively correlated with effort, so that the coefficient on **Web** would be biased downwards.*

3) (3 points) What is the estimated treatment effect on the final exam grade of Web class delivery based on regression 4? Is this effect statistically significant at the 5% level? Is this effect large or small in a real-world sense?

Web delivery is estimated to reduce course grade by 19.18 points out of 100, and this is statistically significant at the 5% level. This is an enormous effect.

4) (3 points) Suppose a 2SLS regression is estimated using weak instruments. What problem(s) might arise?

If the instruments are weak, the 2SLS estimator can be biased, and the estimator and t-statistic are not in general normally distributed. This means that the usual inferences (like the response to question 3) would not be reliable.

5) (3 points) Are the instruments used in regression 4 (*On Campus* and *Web whiz*) weak or strong instruments? Explain (be precise).

The first-stage F statistic is 11.21, which exceeds 10, the rule-of-thumb cutoff, so the instruments can be treated as strong.

6) Do you think that the instruments in regression 4 are exogenous? Why or why not? Make a case, in a concrete way (perhaps using an example) in favor of or against the exogeneity of:

Both arguments in favor and against may get a full score depending on the quality of argument. Here are some arguments in favor for example:

(3 points) ***On campus***

*The requirement is that **On campus** be uncorrelated with the error term in the **Grade** equation (that is what is meant by exogeneity of an instrument). The omitted factors that enter the error term would include diligence (beyond what is measured by GPA), luck on the exam, whether the student finds macro an intuitive course, and related factors that would determine good performance in a macro course, after controlling for all the other factors in the equation. If these factors, say diligence and proclivity for macro, are unrelated to whether the student lives on campus, then **On campus** – resident and nonresident students are equally diligent – would be exogenous.*

(3 points) ***Web whiz***.

*Similarly, **Web Whiz** would be exogenous if it is unrelated to these other factors, such as diligence and proclivity for macro, are unrelated to **Web Whiz**, then there would be no omitted variable bias. This is plausibly so – why would being good at navigating the Web make you particularly good at macro, or make you more or less diligent than those who don't use the Web as much? If so, then **Web Whiz** would be exogenous.*

7) (6 points) MIT faculty members have expressed concern about low attendance rates in many large undergraduate classes. Some undergraduates retort that they do just fine thank you: they get the material they need from the textbook, the course Web site, and occasional emails with TAs. Does the

estimate of *Web* treatment in regression 4 provide MIT administration with useful information to counter this undergraduate retort? Explain.

This question can be answered both yes and no. The Yes answer would address internal and external validity. Concerning internal validity, the instruments are strong, and if the answer to 6 is found to be compelling then they are plausibly exogenous. Thus the instruments are valid and the result would be an estimate of the causal effect of interest. (One could raise the question of whether there would be a difference between LATE and ATE here, but that is not needed.) Concerning external validity, one would need to believe that these results from Western Kentucky University can be generalized to MIT students. The students are, undoubtedly, very different, but that does not necessarily mean that they have different responses to Web delivery vs. in-person delivery. MIT students would be faster and better learners, but that does not logically imply that the way they learn is any different than how the WKU students learn, the MIT students are just better at it.

On the No side, there must be a failure of internal or external validity or both. Internal validity would fail if the instruments are weak (they are not) or endogenous. On the latter, perhaps On Campus is actually related to other innate skills that matter for your macro grade, for example students on campus might have greater motivation, and if so On campus would be correlated with motivation which would appear in the error term of the Grade equation, even after controlling for all the observable variables in that equation. Or, WKU students and MIT students might learn differently, so that problems at WKU with the Web delivery would not apply to MIT. NOTE however that is it not enough simply to say that MIT students are smarter than WKU students – the MIT students must learn in a sufficiently different way that the responses to Web delivery are different between the two groups. It should be noted that the Web delivery at WKU is essentially the same as skipping class at MIT, so the program being studied does generalize to the MIT situation.

Now consider a hypothetical study that could be undertaken at MIT. At the start of the semester half the students in 14.32 are randomly selected and told they will receive a cash payment of $250 if they attend at least 90% of the lectures. Attendance is taken every class (assume this is possible), so there are data on *Attendance* (the number of lectures attended), as well as on final exam grades and student characteristics from the Institute's database. The objective is to estimate the causal effect of attendance on grades. The specific regression of interest is the grade on the final exam (*Grade*) as a function of *Attendance*, controlling for student characteristics (major, gender/socioeconomic, prior GPA, SAT scores, high school quality, etc. –collectively denote these variables by *W*).

8) (3 points) Provide a reason why the coefficient on *Attendance* in the OLS regression of *Grade* on *Attendance* and *W* could be a biased estimate of the causal effect on *Grade* of *Attendance*.

Even controlling for all these items, students with more outside opportunities, and more commitment to extracurricular activities, would be likely to have lower attendance. The lower attendance presumably reflects demands on time that will impact study time, in which case low attendance would be correlated with low study time which would enter the Grade equation, leading to OV bias. (An argument going the opposite way is that those not attending are particularly *talented and find the course particularly easy, so their omitted factor is native ability and proclivity for the course – this would also lead to OV bias, but in the opposite direction).*

9) (3 points) Let *Cash* denote a binary variable that equals 1 if the student is chosen to be eligible for a cash enticement and equals 0 otherwise. Is *Cash* plausibly a valid instrument for *Attendance* in the 2SLS regression of *Grade* on *Attendance* and *W*? Explain.

Yes. The amount of money is probably sufficient to persuade some of the students to attend, in which case it would be relevant. It is clearly exogenous because it is randomly assigned and therefore uncorrelated with any other student attributes. So it is a valid instrument.

10) (3 points) Are the control variables *W* needed or useful in this regression? If not, should they be dropped from this regression? Explain.

*Because **Cash** is randomly assigned, the variables **W** are not needed to ensure exogeneity (**Cash** is uncorrelated with the error term even if **W** is omitted. Still, **W** might be useful because if the **W** variables are sufficiently important that the SER goes down substantially when **W** is included, then the standard errors on the regression with the **W** variable will be smaller than those without the **W** variable.*

11) (3 points) Is the local average treatment effect estimated using this 2SLS regression arguably less than, greater than, or the same as the average treatment effect? Explain.

*The local average treatment effect equals the average treatment effect if (i) there is no heterogeneity in the effect of **Cash** on **Attendance**; or (ii) there is no heterogeneity in the effect of **Attendance** on **Grade**; or (iii) these two effects (coefficients) might have heterogeneity, but that heterogeneity is uncorrelated. In this case: (i) is implausible – different students have different parental incomes; (ii) is implausible, if only because the learning styles of different students differ and some might find class attendance more helpful than others; (iii) hard to say. An argument for uncorrelatedness, in which case LATE=ATE, is that differences in parental income are unrelated to differences in learning styles.*

One could alternatively argue that particularly self-motivated students won't find the class time useful and will have a high value of time, so those students would have a low β and a low π. Under this story, β and π would be positively correlated, in which case LATE would be greater than ATE. Said differently, students who are smarter than the instructor don't need to waste their time attending class and also don't need \$250, so the instrument (being offered \$250 to attend) has little effect on them, so their low β isn't given much weight in this IV estimator – and if so LATE exceeds ATE.

Part II: Economic determinants of voting

Policy debates and personalities get much attention in presidential elections. You will examine whether economic conditions also matter. The question of interest is: what is the effect of economic conditions on the willingness to vote for an incumbent party?

You will examine the question using panel data on 50 states for the ten presidential elections from 1976 through 2012. The data are also described in Table 1 (the means are for 2008). Figure 1 shows 10 scatterplots, one for each election year, of the percentage vote for the incumbent party vs. the state unemployment rate.

Note

(i) The incumbent party is the party of the sitting president. For example, in 2020, the incumbent party was Republican (Trump) and the candidates were Trump (Republican, lost) and Biden (Democrat, won). But in 2016, the incumbent party was Democrat (Obama) and the candidates were Clinton (Democrat, lost) and Trump (Republican, won).

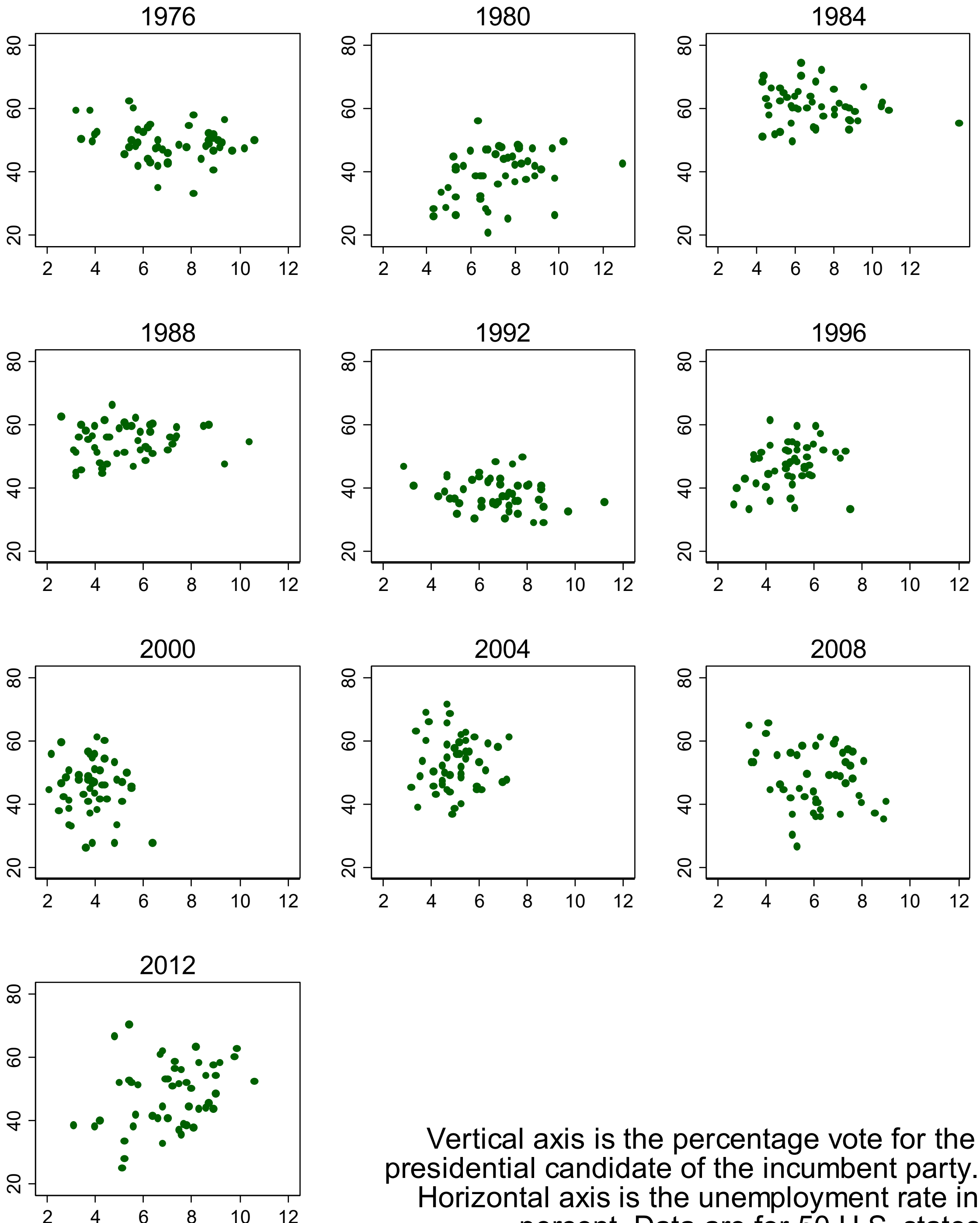
(ii) The panel data set also includes a binary variable indicating whether or not a given year was a non-recessionary year. Recessions are times of declining national economic conditions. For this exam, elections that overlapped with recession years were: 1976, 1984, 1992, and 2008. The remaining elections were in non-recession years.

Table 1. Variable Definitions and Summary Statistics for 2008

Unit of observation: U.S. state (n = 50) in a presidential election year, 1976-2012 (T = 10).

Variable	Definition	Mean (2008)	Std. Dev. (2008)
vote share for incumbent party	percent of the state vote going to the incumbent party	47.8	9.4
income	per capita income, in thousands of 2008 dollars	40.0	5.9
log income	logarithm of *income*	3.68	0.14
college	percent of the state population with a college degree or higher	27.2	4.8
white	percent of the state population that is white	69.1	16.0
UR	unemployment rate in the state in October of the election year	6.1	1.5
non-recession year	$= \begin{cases} 1 \text{ if the election is in a non-recession year} \\ 0 \text{ if the election is in a recession year (1976, 1984, 1992, 2008)} \end{cases}$	n/a	n/a

Figure 3
Incumbent vote percent vs. state unemployment rate



Vertical axis is the percentage vote for the presidential candidate of the incumbent party. Horizontal axis is the unemployment rate in percent. Data are for 50 U.S. states

Table 2. Regression Results: Determinants of Voting for Incumbent Party, 1976-2012

Data are for U.S. states (n = 50) for T = 10 presidential election years (1976, 1980,…, 2012)

	(1)	(2)	(3)	(4)	(5)
Dependent variable	*vote share for incumbent party*	*vote share for incumbent party*	*vote share for incumbent party*	*vote share for incumbent party*	*vote share for incumbent party*
Regression method	Fixed effects	Fixed effects	Fixed effects	Fixed effects	Fixed effects
***Regressors*:**					
UR (Unemployment rate)	0.13 (0.18)	0.13 (0.25)	-0.73* (0.31)	-0.66* (0.30)	-2.83* (1.36)
UR $\times$ non-recession year			1.91** (0.64)	1.98** (0.64)	0.15^{+} (0.08)
Income (thousands dollars)				0.22* (0.09)	0.20* (0.10)
UR^2					2.65 (1.89)
UR^2 $\times$ non-recession year					-0.19 (0.12)
State fixed effects?	yes	yes	yes	yes	yes
Year effects?	yes	yes	yes	yes	yes
Standard errors	HR	cluster	cluster	cluster	cluster
Number of observations	500	500	500	500	500
***F*-statistics testing that the coefficients on groups of variables are all zero (*p*-values in parentheses):**					
UR, UR$\times$Non-rec year			4.58 (.015)	4.83 (.012)	5.35 (.001)
UR^2, $UR^2$$\times$Non-rec year					1.97 (.151)
UR, UR$\times$Non-rec year, UR^2, $UR^2$$\times$Non-rec year					2.74 (.039)
Year dummy variables	241.3 (.000)	241.3 (.000)	272.0 (.000)	290.1 (.000)	219.5 (.000)

Notes: All regressions include an intercept. Standard errors are given in parentheses under estimated coefficients, and p-values are given in parentheses under F- statistics. Standard errors and F-statistics are heteroskedasticity-robust (HR) for regression (1) and are clustered for regressions (2)-(5), where clustering is at the state level. Coefficients are individually statistically significant at the $^{+}$10%, *5%, **1% significance level.

Panel data regressions in Table 2 (30 points)

Please answer these questions in Blue Book 2

1) (3 points) Regressions (1) and (2) in Table 2 are identical, except that regression (1) uses heteroskedasticity-robust standard errors and regression (2) uses clustered standard errors, where the standard errors are clustered by state. Which is preferable? Why? Explain.

Heteroskedasticity robust standard errors allow for the error term to be heteroskedastic, but require that the error term be uncorrelated across observations. In panel data such as we have in Table 2 this latter requirement is unreasonable because, for a given entity (here, state), the omitted variables that constitute the error term are plausibly correlated over time. Clustered standard errors allow for these error terms to be correlated over time within a state, as long as the errors are uncorrelated across states. Therefore, clustered standard errors are preferable.

2) Consider regression (4).

a) (3 points) What is the effect on the incumbent vote share of a one percentage point increase in the unemployment rate in a recession year, holding income constant?

y = (. 0.66 *1 + 1.98 * 0)*unemployment rate =. 0.66*unemployment rate ,
So, an increase in one percentage point in the unemployment rate in a recession year reduces the incumbent vote share by 0.66 percentage points

b) (3 points) What is the effect on the incumbent vote share of a one point percentage point increase in the unemployment rate in a non-recession year, holding income constant?

y = (. 0.66 * 1 + 1.98 * 1)unemployment rate =. 0.66 + 1.98 = 1.32
So, an increase in one percentage point in the unemployment rate in a non-recession year increases the incumbent vote share by 1.32 percentage points.

c) (3 points) Compute the standard error for the effect you estimated in part (b). If additional information is required to do that, state precisely what you would need.

We need to compute the SE of β1 + β2 .To do this we need to apply the following formula:

$$\mathrm{SE}(\beta 1 + \beta 2) = \sqrt{Var(\beta 1 + \beta 2)} = \sqrt{Var(\beta 1) + Var(\beta 2) + 2Cov(\beta 1;\ \beta 2)}.$$

From Table 2 we know that the estimates of the two variances are,

$$Var(\beta 1) = (0.30)^2 = 0.09$$
$$Var(\beta 2) = (0.64)^2 = 0.41$$

However we need to know the value of Cov(β1; β2) , and this is not provided in the table.

3) (3 points) In regression (3), test the null hypothesis that the incumbent vote share does not depend on the unemployment rate (whether in a recession year or not), against the alternative that it does, at least in some years.

The null hypothesis that the unemployment rate does not enter in either type of year implies that we test:
H 0 : β1 = β2 = 0 vs H 1 : at least one coefficient is different from zero.
To perform this test we use the F-statistic. Since the p-value is 0.015, we reject H 0

4) Consider regressions (4) and (5).

a) (3 points) What does regression (5) capture that regression (4) does not?

Regression (5) includes two variables not in (4): $UR^2$and (UR×non recession year)^2. These two variables permit a nonlinear (quadratic) effect of the Unemployment rate in both recession and non-recession years, and allows that nonlinear effect to differ between the two types of years.

b) (2 points) Based on the results in the table, would you prefer specification (4) or (5)? Explain.

The F- statistic testing the coefficients on $UR^2$and (UR×non recession year)^2 is 1.97 with a p-value of .151, so the test does not reject at the 10% significance level. Moreover, the individual coefficients on the two squared terms are not individually significant at the 10% level. Therefore, there is no evidence against the hypothesis that the nonlinearities do not enter the expression. Therefore regression (4) is preferred.

5) Consider regression (4). Provide justification for each of your examples.

a) (2 points) Provide an example of a variable not explicitly in regression (4) which could cause omitted variable bias, but which is controlled for by including time fixed effects.

Time fixed effects capture factors changing over time but are shared equally across all states. And to cause OVB, the omitted variable W would have to be a determinant of Y=incumbent vote share and be correlated with either X=unemployment rate or X=average income. Examples of W include federal taxes (e.g. if they both affected the economy and repulsed voters), or the latest season of House of Cards (e.g. if everyone stopped going to work to watch it and watching it influenced voters' political leanings). War or international conflict could also affect both economic conditions at home and the incumbent party's popularity.

b) (5 points) Provide an example of a variable not explicitly in regression (4) which could cause omitted variable bias, but which is controlled for by including state fixed effects.

State fixed effects capture factors particular to a state but constant over time. Further, the omitted variable W would only cause OVB if it were both a determinant of Y=incumbent vote share and were correlated with unemployment rate or average income. Examples of W could include the state's historical voting patterns if states traditionally more likely to vote for the incumbent also had higher/lower unemployment or incomes. The industry makeup of a state could also cause OVB (if state FE were omitted) since industries vote differently and have different salaries and unemployment rates. [Note : it is actually rather hard to think of an omitted variable that would vary across states but not over time, which would be a determinant of voting for the incumbent, and which would be correlated with the unemployment rate. It is easier to think of examples that would be a determinant of voting Republican or Democrat and satisfy these conditions, such as whether the state is rural or urban; but harder to think of variables that would be related to generic anti-incumbent sentiment.]

6) (3 points) (One paragraph.) Explain, in everyday terms, what conclusions you draw from the regressions in Table 2 about the effect of economic conditions on the willingness of voters to support (or reject) the incumbent party.

Regression (4) seems the most credible. It clusters standard errors correctly at the entity level, reduces OVB by including time and state fixed effects. It also includes state income and the interaction term between unemployment rate and non-recession, which are significant predictors of incumbent vote share. Regression (4) excludes the quadratic terms, since they were found to be statistically insignificant.

Based on regression (4), states richer on average are more likely to vote for the incumbent party. States with higher unemployment rates are slightly less likely vote for the incumbent party, except in non-recession years, where they are actually more likely to vote for the party in power, holding constant income. Specifically, a 1 percentage point increase in a state's unemployment rate during a non-recession year is associated with a 1.32pp increase in the incumbent's vote share. Although insufficient information was provided to test the significance of this coefficient for non-recession years, it is in fact significant.

One explanation for the nonintuitive result that higher unemployment rates make the incumbent party more popular during non-recession years could be salience. When the economy is in the news, we get the expected relationship that higher unemployment leads to lower incumbent vote share. When the economy is fine, voters pay less attention to the unemployment and thus, consider it less in their decisions. However, that the effect of unemployment rate on vote share is actually positive and significant during non-recession years is a mystery and suggests that regression (4) suffers from some form of OVB, and more work is warranted.

Questions for Part III (25 points)

Please answer these questions in Blue Book 3

1) (5 points) Briefly explain the difference between Newey-West standard errors and heteroskedasticity-robust standard errors.

   HR standard errors allow for heteroskedasticity but not autocorrelation of the error. NW standard errors adjust HR standard errors for autocorrelation of the error terms.

2) (5 points) TRUE, FALSE, UNCERTAIN Imagine that you have a regression with one endogenous regressor. You decide to use IVs to estimate this. You have four potential instrument variables that you're not so sure about. You come up with an idea to pick out the right set of instruments. You run J-tests for every possible combination of instruments and choose a maximal set that doesn't get rejected by the test. This is a valid way of choosing instruments. Explain.

   False. There are at least two good explanations, either one of them gets full credit. (1) rejection of J-test suggests that at least one of the instruments disagrees with some other, but it does not suggest which one of them is correct. (2) In the heterogeneous effect case J-test may reject even when all instruments are exogenous due to difference in LATEs.

3) (6 points) Assume Y_t is a *change* in monthly inflation, and you estimated the OLS regression (using past data):

$$Y_t = 0.02 + 0.2Y_{t-1}$$

   Monthly inflation in September 2021 was 0.4%, while monthly inflation in October 2021 was 0.9%. What would be your forecast of the monthly inflation for November 2021?

   Let T=October of 2021, then $Y_T = 0.9 - 0.4 = 0.5$, and the forecast suggested by the model for November 2021 is

$$Y_{T+1|T} = 0.02 + 0.2 * 0.5 = 0.12$$

   This suggests that the forecasted inflation is 0.9+0.12=1.02%

4) (6 points) Assume that one runs two-stage least square regression of outcome variable *Y* on treatment *X* using binary outcome *Z*. Derive that the TSLS estimator in this case has the following form:

$$\hat{\beta} = \frac{(\bar{Y}|Z=1) - (\bar{Y}|Z=0)}{(\bar{X}|Z=1) - (\bar{X}|Z=0)}$$

   Where $(\bar{Y}|Z=1)$ is the average of the outcome variable over those observations with *Z*=1.

$$\hat{\beta} = \frac{S_{YZ}}{S_{XZ}} = \frac{S_{YZ}/S_{ZZ}}{S_{XZ}/S_{ZZ}} = \frac{\hat{\delta}}{\hat{\pi}}$$

   Here $\hat{\delta}$ is the OLS estimator from regressing Y on Z, while $\hat{\pi}$ is the OLS estimator from regressing X on Z. In both cases the regressor is a dummy variable, and as such the OLS estimates are equal to difference of averages, when Z=1 and Z=0.

5) (3 points) Assume one estimates the following probit model:

$$P(person\ is\ employed) = \Phi(\alpha + \beta\ age + \gamma\ age^2 + \delta education).$$

Derive the formula of the marginal effect of age on probability of being employed.

Denote φ the probability density function of a standard normal.

$$\frac{\partial\ P(employed)}{\partial age} = \varphi(\alpha + \beta age + \gamma age^2 + \delta education)(\beta + 2\gamma age)$$