Practice Final Exam 2
PACKET 2

Instructions

1. **DO NOT READ THIS PACKET UNTIL YOU ARE TOLD.**

2. You have three hours for the final exam.

3. The exam has **four parts**.

4. **Please write your answers directly in the space provided.** Short answers should be very succinct. Show your work and intuition clearly: credit is given for explanations and not just having the correct answer.

5. **Write your answers using a pen (not pencil).** This is because we will scan the exam, so it is necessary. If you do not have a pen, please ask us for one. **Put your Harvard ID number on the top of every sheet in Packet 2. This is also required for the scanning.**

6. You are permitted two double-sided 8½" x 11" sheet of notes, plus a calculator. No computers, wireless, or other electronic devices without prior permission. Cell phones must be turned off for the duration of the exam. You may not share resources with anyone else. No collaboration of any kind is allowed on this exam.

7. During the exam, you may take breaks to use the restroom. To ensure that all students are treated equally and your work is assessed accurately, we ask that you sign the bathroom log before leaving the room and leave your cell phone with the teaching staff while you are out of the room. Your phone will be returned when you come back to the room.

8. **Please return both parts of this exam.**

The Harvard University Honor Code

Members of the Harvard University community commit themselves to producing academic work of integrity – that is, work that adheres to the scholarly and intellectual standards of accurate attribution of sources, appropriate collection and use of data, and transparent acknowledgement of the contribution of others to their ideas, discoveries, interpretations, and conclusions. Cheating on exams or problem sets, plagiarizing or misrepresenting the ideas or language of someone else as one's own, falsifying data, or any other instance of academic dishonesty violates the standards of our community, as well as the standards of the wider world of learning and affairs.

I understand and agree to take the exam by myself under good test conditions. I will not provide help to anyone or seek help from anyone, including classmates, family members, tutors, teaching fellows, and course assistants.

_____   _____
Printed Name of Student   Date

_____   _____
 Signature of Student   HUID

Questions for Part A

The questions in Part A refer to Table A2.

1) Interpret the coefficient on *After Medicaid expansion* in regression (1).

2) Use the omitted variable bias formula to explain why the coefficient on *After Medicaid expansion* changes in regression (2) compared with regression (1).

3) All the regressions in Table A2 use clustered standard errors, where the clustering is at the state level. What is the advantage of using standard errors clustered by state instead of each of the following:
   a) Heteroskedasticity-robust standard errors
   b) Standard errors clustered at the county level

4) State bankruptcy laws, such as asset exemptions, differ significantly across states, but have not changed since the early 1900s. For example, Texas has the most debtor-friendly asset exemption laws in the U.S., allowing households to protect the full value of their homes from creditors in bankruptcy proceedings. Delaware has the least debtor-friendly laws, with no homestead exemption. Discuss if state bankruptcy laws are arguably a source of omitted variable bias (i) in regression (2)? (ii) in regression (4)?

5) Column 6 presents estimates of the dynamic causal effect of the Medicaid expansions, by replacing *After Medicaid expansion* with three separate variables: $\Delta Expanded_t$, $\Delta Expanded_{t-1}$, and $Expanded_{t-2}$. Note that $Expanded_t$ is an indicator that equals 1 in the first year that a state expanded Medicaid (which is 2014 in our sample) and 0 otherwise, and that $\Delta Expanded_t$ equals $Expanded_t - Expanded_{t-1}$.
   a) What is the long run cumulative effect of Medicaid expansion?
   b) Graph the cumulative impulse response function (i.e. cumulative dynamic multipliers) in response to expanding Medicaid, by year, from year t through $t + 2$.
   c) Do you have enough information in the table to report the 95% confidence intervals for the cumulative impulse response function? Explain why or why not.

6) Regression (5) adds county specific linear time trends, as suggested in Chapter 5 of Angrist and Pischke (2015). Explain the motivation for adding these terms to the regression. What do you conclude from regression (5) about the internal validity of regression (4)?

7) Explain briefly how measurement error in bankruptcy rates would affect the estimates reported in Table A2.

Questions for Part B

The questions in Part B refer to Table B3.

8) Using <u>regression (1)</u>, can you reject the hypothesis that Hajj visas are randomly assigned? Report your <u>F statistic</u>, its <u>p-value</u>, the <u>5% critical value</u> for the test, and <u>number of restrictions</u> for this test.

9) Provide a reason why the coefficient on *Hajj* in <u>regression (2)</u> is plausibly biased. Use the omitted variable bias formula to give the direction of the bias.

10) Is *success* likely to be a valid instrument for *Hajj* in <u>regression (5)</u>? Explain.

11) All the regressions in Table B3 include the same set of controls (listed below). Which of these controls are needed in order for *success* to be a valid instrument in <u>regression (5)</u>? Explain.
    a) Place of departure × Party size fixed effects
    b) Demographic controls (listed below Table B3)

12) <u>Regression (5)</u> estimates a local average treatment effect (LATE). In your judgment is the LATE estimated in (5) greater than, less than, or equal to the average treatment effect? Explain.

13) One hypothesis is that the Hajj affects participants' beliefs by increasing their exposure to people from different countries and sects, and to members of the opposite gender outside their family. If this hypothesis is correct, pilgrims who travel in smaller parties, and thus have more opportunity to interact with non-Pakistanis, may experience larger treatment effects.
    a) Using <u>Column 6 of Table B3</u>, what is the effect of the Hajj for a pilgrim in a small party?
    b) What additional information would you need in order to compute a 95% confidence interval for the predicted effect in part (a)?
    c) Can you reject the hypothesis that the effect of the Hajj pilgrimage is the same for large parties and small parties?

Questions for Part C

The questions in Part C refer to Table C2.

14) Consider the regression in <u>column 1 of Table C2</u>
    a) Interpret the coefficient on *UnionWin* in column 1 of Table C2.
    b) As a general matter, provide a definition <u>in words</u> for a 95% confidence interval.
    c) Calculate the 95% confidence interval for the coefficient on *UnionWin* in <u>column 1 of Table C2</u>.

d) Are the end points on your 95% confidence interval large in a real world sense? In particular, do they allow you to rule out (at the 5% level) the estimated 15% union wage premium described in introduction to Part C?

15) The regression in column 1 of Table C2 is of the following form:

$$Y_i = \beta_0 + \beta_1 UnionWin_i + \beta_2(VoteShare_i - 50\%) + \beta_3(VoteShare_i - 50\%) \times UnionWin_i + v_i$$

Consider the following modified version of this regression that replaces $VoteShare_i - 50\%$ with $VoteShare_i$:

$$Y_i = \alpha_0 + \alpha_1 UnionWin_i + \alpha_2 VoteShare_i + \alpha_3 VoteShare_i \times UnionWin_i + u_i$$

What is the relationship between β_1 and α_1?

16) Explain what <u>Figure C1</u> shows and how it relates to the key assumption needed for the regressions in Table C2 to measure the causal effect of unionization on workers' wages.

Questions for Part D

The questions for Part D refer to Table D2.

17) Table D2 reports two types of standard errors below the estimated regression coefficients: heteroskedasticity-robust standard errors in parentheses, and Newey-West standard errors in square brackets.
   a) Briefly explain the difference between Newey-West standard errors and heteroskedasticity-robust standard errors.
   b) Explain which set of standard errors is preferred, and why.

18) Choose a preferred forecasting model among <u>regressions (1)-(5)</u> in Table D2. Explain your reasoning.

19) The bottom panel of Table D2 reports both the root mean squared error (RMSE) and a GARCH(1,1) forecast standard deviation for 2018w18. Consider the following statement:

"The forecast intervals constructed using the GARCH(1,1) estimate of the root mean squared forecast error (RMSFE) are narrower than the intervals constructed with the RMSE. Therefore, the GARCH(1,1) estimate of the RMSFE is preferred."

Do you agree with the logic in the statement? Explain.

20) The recent values of ILI outpatient visits are given below Table D1.
   a) Using <u>model (6)</u>, compute the forecast of ILI outpatient visit growth (percent per week) in the U.S. for 2018w18, which will be released tomorrow.
   b) Using <u>model (7)</u>, compute the forecast of ILI outpatient visit growth (percent per week) in the Taiwan for 2018w18, which will be released tomorrow.
   c) Using <u>model (6)</u> compute the forecast of ILI outpatient visit growth (percent per week) in the U.S. for 2018w19, which will be released next week.

21) <u>Figure D2</u> plots the Quandt Likelihood Ratio (QLR) statistic, computed for <u>regression (6)</u> from Table D2. The maximum value from the figure is 4.461.
   a) What are the critical value at the 5% significance level and number of restrictions for this test?
   b) What do you conclude from the QLR test?