NET 2023 Power Round

Advanced Division: Solutions

April 27, 2023

Instructions

This test consists of six questions. While you are free to attempt all six questions, we will only grade your four best-performing questions, regardless of how well you do on the other two. A question's point value is *not* informative of its difficulty; although questions have different point values, each question is weighted independently of its point value in your final cumulative score. After normalizing point-values of each question to the same weight, your cumulative score will be calculated as the sum of the scores of your four best-performing questions. You are encouraged to work together on these questions. Answer each question as clearly and succinctly as possible. You may write on a blank sheet of paper where you *clearly indicate* where your answer to each part is. If you are unsure of your answer, take your best guess: there is no penalty for incorrect answers. If you find yourself stuck on a question, skip it and return to it at the end if necessary. You will have two hours (120 minutes) to complete the exam. Remember, we do *not* share your answers or scores with Northwestern admissions, nor do we keep them for ourselves. You are not expected to know how to answer each question on the exam; rather, this test is designed to assess your economic and formal reasoning skills. Have fun, and good luck!

Problem 1: Comrade, to Infinity and Beyond!

In this question, we consider how game theory can be used to model cooperation in settings where naive applications of basic game theory may at first predict otherwise. This problem requires no mathematical prerequisites outside of the ability to sum infinite geometric series.

Part A *(6 points total)* Consider the game between two players, represented in the following payoff matrix where player 1's actions are represented in the rows and their payoffs are in the first coordinate.

	C	D
C	$(3, 3)$	$(-1, 4)$
D	$(4, -1)$	$(0, 0)$

Table 1: Stage Game

(1) *(1 point)* Suppose you could force the individuals to choose an outcome. Which one would maximize payoffs for both players?
The surplus-maximizing choice is (C, C); In particular, it gives 6 payoff to both players. *A correct identification was given both points.*

(2) *(3 point)* What is the *unique* Nash equilibrium of the game? Explain why this is a Nash equilibrium and why this is the only Nash equilibrium.
The unique Nash equilibrium is (D, D). We know that any action supported in a (potentially mixed) Nash equilibrium must not be strictly dominated; (0, 0) is the only undominated action, and thus it is the only pure strategy Nash equilibrium. Moreover, it is in fact an equilibrium because the best response for both players playing D is to play D. *(One point for identifying the NE. One point for justifying why it is an NE. One point for explaining why it is the unique NE).*

(3) *(2 points)* Comment on the friction between Part (A) and Part (B). Is this reasonable?
The Nash equilibrium is not Pareto efficient, and not the welfare-maximizing equilibrium. This seems a little weird, because if both players could agree and communicate, they might want to play (C, C) always. *(One point was given for recognizing the friction between equilibrium and efficiency. A second point was given for any explanation (either way) for why this was or wasn't reasonable).*

Part B *(2 points)* As concisely as possible, give one real-world example of a situation that we might want to model using a matrix like this. Is the prediction of the game reasonable in the context of economic intuition? What about in the context of observed behavior?
There are multiple examples. The tragedy of the commons, where each group can choose to use resources cooperatively or selfishly, is one. The prisoner's dilemma is another one, where individuals need to keep a secret together, but by doing so they will each get some small punishment. *Any reasonable example was awarded both points.*

Part C *(7 Points Total)* Now assume instead that this game is played twice in a row.

(1) *(2 points)* Assume we are in the second period. Without knowing anything about the first period, what will happen in equilibrium in the second period? *Hint: Think of the sunk cost fallacy.*
In the second period, we have to play a stage-game Nash equilibrium regardless of the history, as anything that happened before is a sunk cost. As a result, we definitely play (D, D). *One point for the correct prediction. One point for the justification.*

(2) *(2 points)* Given your answer in the second period, what will individuals do in the first period? Explain.
Because in the second period, they will always play (D, D), individuals have no intertemporal incentives. As a result, they will play (D, D) in the first period as well, since that is their stage-game equilibrium strategy. In particular, (D, D) is dominant across *both* periods. *One point for the correct prediction. One point for the justification.*

(3) *(3 points)* Extrapolating on your answer here, is it possible for (C, C) to be played in any period of this game in equilibrium? What if the game ends in $N > 2$ periods? Explain.
Of course, we can backwards induct the argument above arbitrarily many times; because of the presence of the "last period," any hope of cooperation unravels. As a result, it is impossible. Formally, in the last period, T, they will play (D, D). Knowing this, they will play the *One point was given for noting (C, C) cannot ever be played in the two period game. Ont point was given for realizing this was true for an arbitrarily long (but finite) number of periods. Finally, one point was given for an explanation in the second case.*

(Part D) *(14 points total)* Assume now that individuals repeat this game infinitely many times, and discount the future at a rate of δ. For example, if both players cooperate in every period, then their payoff for both players would be would be

$$\sum_{t=0}^{\infty} 3\delta^t$$

while if they played (C, C) in even periods and (C, D) in odd periods, the payoffs for players 1 and 2, respectively, would be

$$\sum_{t=0}^{\infty} 3\delta^{2t} + \sum_{t=0}^{\infty}(-1)\delta^{2t+1} \text{ and } \sum_{t=0}^{\infty} 3\delta^{2t} + \sum_{t=0}^{\infty} 4\delta^{2t+1}$$

(1) *(2 points)* Show that playing (D, D) in every period regardless of what the other person is doing is a Nash equilibrium in this game. *(Hint: One way to do this is by showing that if one player plays D in every period regardless of what the other player does, then they cannot do better than playing D in every period as well.)*
The game is symmetric, so we need only show playing D in every period is a best response to the other person playing D in every period. This is clearly true, because if the other person plays D regardless of our action, then our unique best response in each period is to play D. Hence this is a Nash equilibrium because each team is best responding to the other. *One point was given for checking the best response property. A second was given for a justification for why this is enough for it to be a Nash equilibrium.*

(2) *(1 point)* Assume instead that players play (C, D) in the second period, and (D, D) in every period after that. What is the payoff to player 2? Player 1?
If players play (C, D), then (D, D) in every period after, then they only get a nonzero payoff in the first period. Hence, player 2, who defects, gets a payoff of 4, while player 1, who cooperated, gets a payoff of -1. *One point was given for the correct answer.*

(3) *(4 points)* Consider now the following *(grim-trigger) strategy* for player 2: player 2 will play C if, in every period before, player 1 played C (and will play C in the first period). If player 1 players D in any period, player 2 will play D forever. Show that if player 2 is playing a grim trigger strategy, player 1 will want to play C in every period if $\delta = 0.9$.
Consider the first period. Player 2 will play C, so Player 1 has two choices: either play C, or play D. If they play D, then they will get a payoff of 4, since this is the situation in

Part (2), since they will have to play D in every period after. If they play C forever, then they will get a payoff of

$$\frac{3}{1-\delta} = 30 > 4$$

Hence, they would prefer to play C *forever* to deviating and playing D in period 1. Moreover, note the game has a "Markov" structure[1]: at any history, the only thing that matters is if C has been played forever or not. Thus, at *every* history where the player has played C forever, he will prefer to play C forever from that point onward. This implies that playing C forever is a best response to this strategy. *Two points were given for the computations of payoffs. Two points were given for a justification for why you would want to play C forever at every clean history, and why this implied playing C forever would be optimal.*

(4) *(3 points)* Using your previous answers, explain why it is a Nash equilibrium for both players to play the grim trigger strategy when $\delta = 0.9$. What will the observed behavior be?
We know that, if 1 plays a grim-trigger strategy, 2 will want to play C forever in whatever strategy they end up playing, but play D as long as they knew the other person would play D. This is satisfied by 2 playing a grim-trigger strategy as well, in which case the argument in (3) implies that both sides are best-responding to one another, hence implying we are in a Nash equilibrium. *Two points were given for the justification grim trigger strategies best reply to one another. A third point was given for concluding this implies the grim-trigger strategy for both players constitutes a Nash equilibrium.*

(5) *(2 points)* Find the smallest value of δ that can sustain (C, C) in every period.
We know that, whenever someone deviates, the lowest payoff they can get comes from playing (D, D) in every period after. As a result, grim-trigger strategies will give us the "most slack" in decreasing δ while still incentivizing cooperation. The relevant equality is for

$$4 = \frac{3}{1-\delta} \iff 4 - 4\delta = 3 \iff \delta = \frac{1}{4}$$

Hence, any $\delta > \frac{1}{4}$ can incentivize players to play (C, C) in every period. *One point was given for the computation. A second point was given for the justification.*

(6) *(2 points)* Assume now that you are an economist defending the assumption individuals are playing this game infinitely times against a skeptic who objects that individuals only live for finitely many periods. Give one (potential) justification for this assumption.
There are a few potential answers to this. One answer is that, if the game has an end-point, then knowing that an endpoint exists messes with the entire reasoning of the game. Because this is not how we think that people naturally behave, an infinite horizon allows us to attach a "point at infinity" at which incentives are irrelevant to anchor the game instead. The second is that, for small enough time intervals, we can "approximate" finite behavior with behavior at infinitely anyway. So if the game is played essentially continuously (people have to continue to choose to cooperate), then this is a reasonable assumption. *One point was given for any attempt at a justification. A second was given based on the quality of the justification.*

(Part E) *(2 points)* Compare and contrast your answers in Parts (C) and (D) in the context of Part (B).
A lot of this is subsumed by the discussion in Part (D)(VI). The main contrast is the way in which

[1] Formally, this requires an appeal to the *one shot deviation principle* to rigorously prove, which we do not expect people to cite. one *bonus* point was given to anyone who invoked the result or gave an argument similar to it.

differing intertemporal incentives affect the possibility of cooperation. Because there is no "last stage" in which cooperation must necessarily break down in the infinite horizon, cooperation is possible. This is impossible in the finite-horizon case because the last period "looms large." This result highlights the need for continuous incentive provision to sustain cooperation, and highlights that so long as we can *continue* to bet on cooperation in the future, we will always be able to. *One point was given for pointing out the difference in the possibility of sustaining cooperation. A second point was giving for sufficient articulation of the justification for why this difference held.*

(Bonus) *(4 points) (The Folk Theorem under Perfect Monitoring)* Show that, for some sufficiently large δ, a modified version of the grim trigger strategy can sustain any potential *mixed strategy* in the stage game. Thus, for sufficiently large discount rates, *any* observed behavior can be rationalized for sufficiently patient individuals!

Answer for the bonus not provided. Email danielluo.pigmail.com for a discussion of the solution, if interested.

Problem 2: Exchange Rates and Arbitrage (25 points)

(Inconsistencies in exchange and interest rates between countries lead to arbitrage. This problem compares two ways to resolve these inconsistencies.)

Part A *(1 point)* An exchange rate is the price of one currency in terms of another. If a particular jacket costs 800 Canadian dollars, and the exchange rate is \$1 CAD = \$0.75 USD, how much does that jacket cost in USD?

$$800 \text{ CAD}(\frac{0.75}{1}) = 600 \text{ USD}$$

1 point is awarded for a correct answer.

We can differentiate between certain exchange rates based on periods of time. The *spot exchange rate* is the current exchange rate between currencies. In contrast, the *forward exchange rate* is the exchange rate at a point in time in the future. Sometimes, disparities between these two rates create arbitrage: the possibility to make risk-free profit after accounting for transaction costs.

Part B *(4 points)* Suppose the spot exchange rate between CAD and USD is 0.75 and the forward exchange rate (one year into the future) is 0.8. Furthermore, suppose the annual real interest rate in Canada is 10% while it is 2% in the United States. Demonstrate how this creates arbitrage by describing a possible investment strategy to make risk-free profit.
An example strategy would be to enter a forward contract with a payment date in 1 year in the US. (This entails borrowing some amount of money in USD and paying back the loan with interest a year later.) Then, convert the borrowed USD to CAD, invest in Canada, and then convert the earnings from one's investment back to USD and paying back the loan with interest.

1 point is awarded for identifying a strategy with a starting point of USD. 1 point is awarded for converting into CAD and investing in Canada. 1 point is awarded for converting the money back to USD. 1 point is awarded for using a loan (or forward contract.)

Part C *(2 points)* What is the rate of return of the strategy you described in Part B?
For the example strategy described above, the return rate is 15.33%. This can be calculated by supposing one borrows 1 and computing the total profit gained by the time one pays the investment back. This is given by: $(1/0.75)(1.1 * 0.8) - 1.02 = 0.1533$.

1 point is awarded for correctly calculating the value of the investment after 1 year in Canada. 1 point is awarded for subtracting the cost of the investment (which is the value of the loan with added interest).

Interest Rate Parity is an economic equilibrium in which arbitrage due to disparities in exchange and interest rates cannot occur. One way of achieving such an equilibrium is through the proper setting of forward exchange rates. This equilibrium is known as *covered interest rate parity.*

Part D *(2 points)* In the scenario described in Part B, what would the forward exchange rate have to be to achieve interest rate parity? To achieve covered interest rate parity, we need the ratio of the forward and spot exchange rates to match the ratio of the interest rates in the two countries. Therefore, we solve the following equation for x:

$$(1/0.75)(1.1 * x) = 1.02 \quad (1)$$

from which we find $x = 0.695$.

1 point is awarded for recognizing the relationship between the ratios of the forward and spot rates and the interest rates. 1 point is awarded for the correct answer.

Part E *(4 points)* Based on your work from the previous part, develop an equation for F_1, the covered interest rate parity forward exchange rate one year into the future. Your answer should be in terms of the interest rate in the domestic country, i_D, the interest rate in the foreign country, i_F, and the spot exchange rate, S.
Using our previous analysis, we have that $\frac{F}{S} = \frac{1+i_d}{1+i_f}$. Solving for F, we get $F = S\frac{1+i_d}{1+i_f}$.

1 point is awarded for an answer where F varies directly with S. 1 point is awarded for an answer where F varies directly with i_d. 1 point is awarded for an answer where F varies inversely with i_f. 1 point is awarded for a correct answer.

Part F *(3 points)* Now generalize your equation to any period of time. That is, develop an equation for F_t, the equilibrium forward exchange rate t years in the future.
Adding the time component only affects the rate of return on one's investment in each country. This is completely dependent on the interest rates of each country. Therefore,

$$F = S(\frac{1+i_d}{1+i_f})^t \tag{2}$$

2 points were awarded for recognizing that the time component affects interest rates. 1 point is awarded for the correct answer.

Alternatively, equilibrium can be achieved without the use of forward exchange rates. Instead, it involves adjusting the spot exchange rate to adapt to predicted changes in interest rates between countries. This is known as *uncovered interest rate parity*.

Part G *(2 points)* Return to the scenario in Part B. To prevent arbitrage, what should a policymaker set the spot exchange rate to?
To achieve uncovered interest rate parity, we also need the ratio of the forward and spot exchange rates to match the ratio of the interest rates in the two countries. Therefore, we solve the following equation for x:

$$(1/x)(1.1 * 0.8) = 1.02 \tag{3}$$

from which we find $x = 0.862$.

1 point is awarded for recognizing the relationship between the ratios of the forward and spot rates and the interest rates. 1 point is awarded for the correct answer.

Part H *(4 points)* Based on your work from the previous part, develop an equation for the uncovered interest rate parity spot exchange rate, S. Your answer should be in terms of the interest rate in the domestic country i_D, the interest rate in the foreign country i_F, and the expected forward exchange rate F.
Using our previous analysis, we have that $\frac{F}{S} = \frac{1+i_d}{1+i_f}$. Solving for S, we get $S = F\frac{1+i_f}{1+i_d}$.

1 point is awarded for an answer where S varies directly with F. 1 point is awarded for an answer where S varies directly with i_f. 1 point is awarded for an answer where S varies inversely with i_d. 1 point is awarded for a correct answer.

Part I *(3 points)* When are covered interest rate parity and uncovered interest rate parity theoretically the same? Provide either an intuitive or mathematical justification. (Hint: your answer should involve forward and spot exchange rates.)

Covered interest rate parity and uncovered interest rate parity are theoretically the same when forward and (expected) spot exchange rates are equal. Full credit was given to an answer which recognized this and had an accompanying explanation. 1 point was awarded for identifying the correct relation between spot and forward exchange rates.

Problem 3: If Everyone Jumped Off a Bridge, You Would Too! (20 points)

In this problem, we consider how agents learn from one another and build a model (of recommendations) to examine why bad businesses may flourish in some conditions. This problem requires no mathematics besides an intuitive understanding of conditional probability.

Preliminaries Assume that there are two colleges, Northwestern and UChicago. One school is good, and one school is bad (pretend we do not know which is which). Sequentially, individuals need to choose a school without observing which school is good. However, each individual student receives a *signal*, $s \in \{N, C\}$. The signal is *accurate*, so that if Northwestern is the good school, then the signal realization will be N with probability $p > \frac{1}{2}$. Meanwhile, if UChicago is the good school, the signal realization will be C with probability $q > \frac{1}{2}$. Each student can observe both the actions of past students, and (sometimes) the signals those students received as well.

Part A *(4 points)* Assume there are $\{t\}_{t=0}^{\infty}$ many students. Thus, there will be $\{s_t\}$ many signals.

(1) *(1 point)* Suppose Northwestern is the good school and $p = \frac{3}{4}$. In the first one million signals, how many N signals do we expect to see?
We expect to see $1,000,000 * \mathbb{E}[\mathbf{1}_{\{N\}}] = 750,000$ by basic arithmetic. *One point was given for the correct answer.*

(2) *(2 points)* Suppose student 1 million sees exactly as many N signals as predicted before. Which school will they choose? What about students after that?
The law of large numbers tells us that the *limiting frequency* of N will be approximately equal to the true probability we expect to see it. Because we see $\frac{3}{4}$ many N from $1,000,000$ draws, we expect that Northwestern is the good school and as a result choose Northwestern. All students after that will choose Northwestern so long as this pattern continues. *One point was given for any correct justification (computation of Bayes' rule, law of large numbers, confidence intervals) etc. that gives a "reasonable" argument. A second point was given for concluding students after this one would choose Northwestern so long as the pattern continued to hold (without stating this caveat, half a point was taken off.*

(3) *(1 point)* Give an intuitive argument for why, after sufficiently many previous students receive signals, *every* student will choose the good school.
The law of large numbers says that, after enough signals, we converge to the correct frequency. As a result, because the patterns are statistically identifiable (the limiting frequency when Northwestern and Chicago are the good schools are different) we should be able to "perfectly identify" the good school with probability arbitrarily close to 1 after observing arbitrarily many signals. *Any intuitive explanation mirroring this one received full credit.*

Part B *(4 points)* Let $p = q$. Consider the individual who first observes a signal, before anyone else.

(1) *(1 point)* Suppose the individual sees signal N. What is the probability of seeing N when the good school is Northwestern? What about the probability of seeing N when the good school is UChicago?
p and $(1 - p)$, respectively. *One point was given for each correct answer.*

(2) *(1 point)* Suppose Northwestern and UChicago are equally likely to be the good school. Then what is the total probability of seeing N?

The total probability of seeing N is

$$\frac{1}{2}p + \frac{1}{2}(1-p) = \frac{1}{2}$$

That is, the total probability of seeing N given a uniform prior is exactly $\frac{1}{2}$, because there is a $\frac{1}{2}$ probability of seeing it with frequency p and $\frac{1}{2}$ probability of seeing it with frequency $1-p$. *One point was given for the correct answer and any reasonable work to justify it.*

(3) *(1 point)* Based on your answers in (1) and (2), what is the probability that Northwestern is the good school, conditional on seeing signal N?
The probability that Northwestern is the good school, conditional on seeing N is:

$$\mathbb{P}(\text{NU is good}|N) = \frac{\mathbb{P}(N|\text{NU is good })\mathbb{P}(\text{NU is good})}{\mathbb{P}(N)} = p$$

using Bayes' rule, where the unconditional probabilities are both $\frac{1}{2}$ and the conditional probability on the right hand side is p, given the information from the problem. *One point was given for the correct answer and any reasonable work to justify it.*

(4) *(1 point)* Using your answers above, show that if the individual sees signal N, they will pick Northwestern, and if they see C, they will pick UChicago.
Because $p > \frac{1}{2}$, if the individual sees N, Part (3) implies they believe Northwestern is more likely to be the good school. Because $p = q$, if they see C, a symmetric argument implies they will be more likely to pick UChicago. *Half a point was given for the justification of each school, with reference to Part (3) (or symmetry) necessary for both justifications.*

Part C *(6 points)* Consider now the point of view of the second individual who can see their own signal and the action (but not the signal) of the first individual.

(1) *(1 point)* Suppose the second individual sees the first individual pick UChicago. Does the second person know what signal this person got, and if so, which one?
Yes, the second person knows what signal they got. We showed that, in the last part, that someone picks UChicago if and only if they saw C (in the world where they see only their own signal). Thus, person 2 knows that person 1 saw C. *Half a point was given for the correct answer, and another half point was given for the justification.*

(2) *(2 points)* Suppose this individual saw signal C. Which school will they choose? Why?
If this individual saw C, they would choose C; after knowing the first signal was C, their belief the good school is UChicago would be p. Thus, the total probability of seeing the second c is

$$p^2 + (1-p)(1-p) = 1 + 2p^2 - 2p = 1 - 2p(1-p)$$

Note for this value to be greater than $\frac{1}{2}$, we have

$$1 - 2p(1-p) \geq \frac{1}{2} \iff 2 - 4p(1-p) \geq 1 \iff 1 \geq 4p(1-p) \iff \frac{1}{4} \geq p(1-p)$$

Which holds, because $p(1-p)$ is maximized at $\frac{1}{4}$ when $p = \frac{1}{2}$ (the derivative of $p(1-p)$ is $1-2p$). Thus, it is more likely than not that UChicago is the good school. *One point was given for the correct answer, and any argument that UChicago was more likely to be the good school. One point was given for any rigorous justification (not necessarily the one given above) for why this was true.*

(3) *(2 points)* Suppose instead they see signal N after the first person picked UChicago. What is the probability of this happening if the good school is Northwestern? UChicago?
There are four outcomes possible: (Northwestern, C, C), (Northwestern, C, N), (UChicago, C, C) and (UChicago, C, N) where the coordinates denote (The Good School, First Signal, Second Signal). If Northwestern is the good school, the probability of seeing (C, C) is $(1-p)^2$, and the probability of seeing (C, N) is $(1-p)p$. If Chicago is the good school, the probability of seeing (C, C) is p^2 and the probability of seeign (C, N) is $p(1-p)$. Thus, the probability of seeing N after the first person picked UChicago is

$$\frac{(1-p)^2}{(1-p)^2 + p(1-p)} \text{ and } \frac{p(1-p)}{p^2 + p(1-p)}$$

if the good school is Northwestern and UChicago, respectively. *One point was given for each probability being correct, and the second for any derivation and work, e.g. through Bayes' rule.*

(4) *(1 point)* Explain why the second person will "follow their signal" regardless of the first person's action: they will choose Northwestern if they see N and UChicago if they see C.
If the person sees the signal C, then we know that they will pick UChicago. IF they see N, then they are indifferent between the two schools, so they will follow their signal. *Half a point was given for any justification that used arguments from the previous two problems. One bonus point was given for noting that indifference can rationalize either behavior and that the posterior belief after seeing (C, N) of the good schools is $(\frac{1}{2}, \frac{1}{2})$; that is, they are equally likely to be good.*

Part D *(7 points)* Consider now the third individual, and suppose both individuals beforehand had picked Northwestern.

(1) *(2 points)* What is the probability that both individuals picked Northwestern if it was the good school? What about if it was the bad school?
Assume Northwestern was the good school. From Part (C), we know the both individuals will pick Northwestern if (N, N) or (C, N) occurs. Both individuals pick C if (C, C) or (N, C) occurs. If Northwestern is the good school, the total probability of the two outcomes where they pick Northwestern is

$$p^2 + p(1-p)$$

Meanwhile, if Northwestern is the bad school, the total probability is

$$(1-p)^2 + p(1-p)$$

One point was given for each probability, and the accompanying justification.

(2) *(1 point)* Thus, conditional on seeing the two other individuals seeing Northwestern, what is the probability it is the good school?
By Bayes' rule and the previous part, we know that the probability Northwestern is the good school if it gets picked is

$$\frac{p^2 + p(1-p)}{p^2 + 2p(1-p) + (1-p)^2}$$

Because $p > \frac{1}{2}$, we know that $p^2 + p(1-p) > p(1-p) + (1-p)^2$ and in particular this implies that the above is greater than $\frac{1}{2}$. *One point was given for a correct application of Bayes' rule and hence correct answer.*

(3) *(1 point)* If the third person sees signal N, what school will they pick?
Let $r > \frac{1}{2}$ be an arbitrary value. If r is the belief Northwestern is the good school, then the posterior after seeing N that Northwestern is the good school is

$$\frac{rp}{rp + (1-r)(1-p)} > \frac{1}{2}$$

Because the probability in Part (D2) is the probability that Northwestern is the good school before the third person sees their signal, this implies they will choose Northwestern, as they belief it is more likely to be the good school. *One point was given for recognizing the posterior belief after (N, N, N) that Northwestern was the good school was greater than $\frac{1}{2}$, using any valid justification.*

(4) *(2 points)* Suppose now the third person sees the signal C. What school will they pick? Justify your answer probabilistically. *(Hint: this sub-question is computationally longer than the others. Compute the conditional probabilities of seeing C and the other two individuals picking Northwestern conditional on Northwestern and UChicago being the good school, respectively.)*
We can't get away with an abstract $r > \frac{1}{2}$ as before, anymore. We need to explicitly compute the probabilities. First, we know that the probability N is chosen given t he belief from Part (D2) that Northwestern is the good school is

$$\frac{p^3 + p(1-p)}{p^2 + 2p(1-p) + (1-p)^2}$$

This is the numerator in our final fraction. The total probability that N is seen is this plus the probability that we see N given that UChicago was the good school; that is,

$$\begin{aligned} &\frac{p^3 + p(1-p)}{p^2 + 2p(1-p) + (1-p)^2} + (1-p)\left(1 - \frac{p^2 + p(1-p)}{p^2 + 2p(1-p) + (1-p)^2}\right) \\ = &\frac{p^3 + p(1-p) + (1-p)\left(p(1-p) + (1-p)^2\right)}{p^2 + 2p(1-p) + (1-p)^2} = \frac{p^3 + 2p(1-p) + p^2(1-p) + (1-p)^3}{p^2 + 2p(1-p) + (1-p)^2} \end{aligned}$$

We will show that this is greater that $\frac{1}{2}$. To do this, note that the requirement is that

$$2p^3 + 4p(1-p) + 2p^2(1-p) + 2(1-p)^3 \geq p^2 + 2p(1-p) + (1-p)^2$$

Rearranging, it is sufficient to show

$$2p^3 + 4p(1-p) - 2p(1-p) + 2p^2 - p^2 - 2p^3 + 2(1-p)^3 - (1-p)^2 \geq 0$$

Which is in turn equivalent to asking that

$$2p(1-p) + p^2 + 2(1-p)^3 - (1-p)^2 \geq 2p(1-p) + 2(1-p)^3 > 0$$

where the middle inequality comes from the fact $p > \frac{1}{2}$ so $p^2 > (1-p)^2$. This implies that the third person thinks Northwestern is more likely, and picks that. *One point was given for setting up the computation, and a second for completing it and arriving at the correct conclusion.*

(5) *(1 point)* If both individuals one and two picked the same school (say Northwestern), how many individuals after them will also pick Northwestern?
All of them. Because, after (N, N), everyone will pick Northwestern, all information is lost, and everyone is at the same "informational point" as the third player, who we know will pick Northwestern. *One point was awarded for the correct answer and justification.*

Part E *(2 points)* Suppose Northwestern is the good school. Is it possible that students will go to UChicago (the equivalent of jumping off a bridge[2]) even though it is the bad school if (i) they can see the *signals* past students received, or (ii) they can only see whether past students chose Northwestern or UChicago? Interpret your answers using economic intuition. *(Hint: Use Parts (A) - (D))*

Yes. Suppose Northwestern is the good school. Every student will *definitely* choose UChicago even if this is true if the first two people see (C, C), which occurs with probability $(1-p)^2 > 0$. Thus there is positive probability that everyone picks the wrong school. *One point was given for recognizing that people would pick the wrong school with positive probability. One point was given for justifying why this was true.*

Part F *(2 points)* This behavior is often known as *herding*. Can you think of another example other than the one given in this problem where herding phenomena occurs?

Multiple potential answers are possible. Some include network effects, where early adoption of an inferior product through advertising can cause that product to become mainstream, or e.g. career choices, where seeing a lot of people make the same choice (such as investment banking) leads to that industry having a surplus of workers. Many answers were given full credit for this part. *Any answer that had the features of 1) imperfect information aggregation and 2) cascade or herding behavior was given both points.*

[2]This is a joke.

Problem 4: Stimmies and the Permanent Income Hypothesis (25 points)

(In this problem, we rigorously microfound the permanent income hypothesis of Milton Friedman and consider how it interacts with the Keynesian hypothesis.)

Part A *(2 points)* Suppose we are in a Keynesian economy with a marginal propensity to consume of 0.8 and no taxes. What is the expenditure multiplier? If the government exogenously increases spending by 100 million dollars, by how much will GDP increase?

The expenditure multiplier is $\frac{1}{1-0.8} = 5$, so an increase in government spending of 100 million dollars would correspond to a \$500 million increase in GDP. *One point was given for the multiplier, and a second for a correct computation in the increase in GDP.*

Part B *(2 points)* Now consider a microfounded version of this problem. Assume there is a (single) representative agent with the following utility function from consumption, c_t:

$$u(c_t) = b_1 c_t - \frac{b_2}{2} c_t^2$$

The consumer is infinitely lived, and discounts the future at rate β. Suppose that the consumer consumes the sequence $\{c_t\}_{t=0}^{\infty}$. What is the lifetime utility of this stream to the consumer?

$$\sum_{t=0} \infty \beta^t \left(b_1 c_t - \frac{b_2}{2} c_t^2 \right)$$

One point was given for recognizing the additive separability of the per-period utility. A second point was given for the correct exponents on the discount rate.

Part C *(2 points)* Assume now that, instead of facing a consumption stream, the consumer faces a stream of *random* income[3], $y_t \in \mathcal{Y}$, which is independent across time. In period t, they can choose to split their income in two ways: they can either invest it in an asset, or they can choose to consume it. If they invest it in an asset, they will get a return of $(1+r) = \frac{1}{\beta}$ in the next period. Suppose a_t is the amount of savings that an individual has at time t. Write out a_{t+1}, the savings they will have at time $(t+1)$, as a function of (r, y_t, a_t, c_t). *(Hint: your function will require only addition and multiplication. Think economically!)*

Note that the savings they have in period $(t+1)$ is the amount they have in the bank at time t, less their consumption, augmented by per-period interest: in particular,

$$a_{t+1} = (1+r)(y_t + a_t - c_t)$$

One point was given for the expression, and one point for a justification of any kind.

Part D *(1 point)* Explain why, using your answers from Parts (B) and (C), along with the fact that the income is *random*, that the optimal sequence of consumption and savings $\{c_t, a_{t+1}\}$ can be found by solving the maximum expressed below:

$$\max_{\{a_{t+1}, c_t\}} \left\{ \sum_{t=0}^{\infty} \mathbb{E} \left(\beta^t \left(b_1 c_t - \frac{b_2}{2} c_t^2 \right) \right) \right\} \text{ s.t. } a_{t+1} = (1+r)(y_t + a_t - c_t)$$

We will also add the assumption the individual cannot gain utility by saving money forever:

$$\mathbb{E} \left[\lim_{T \to \infty} \beta^T a_T^2 \right]$$

This is to ensure the problem is well-defined; if this confuses you, feel free to ignore it. We are now ready to do some calculus to explicitly solve the problem.

[3]Note: If it helps, you may assume $y_t = 0$ with probability $\frac{1}{2}$ and 1 with probability $\frac{1}{2}$ and manipulate this distribution without loss of credit from here on out. For full credit, though, we ask you to clearly demarcate when this substitution is used.

We know that in each period, we want to maximize our consumption value, which is given by $b_1 c_t - \frac{b_2}{2} c_t^2$ so given (B), we want to maximize the totla gvalue of this income stream. However, (C) shows that our consumption choices today affect our savings decisions tomorrow, and as a result we are constrained by how we trade off against time: the consumption path must be *feasible* across time, which is exactly the argument in Part (C). *Any justification got one point.*

Part E *(7 points)* It is a fact of mathematical optimization that the problem (with constraints) defined above will have the same maximizers as the following unconstrained problem (called the *Lagrangian*):

$$\mathcal{L}(\{c_t\}, \{a_{t+1}\}) = \mathbb{E}_{\{y_t\}\sim\mathcal{Y}}\left[\beta^t\left(\left(b_1 c_t - \frac{b_2}{2}c_t^2\right) + \lambda_t((1+r)(y_t + a_t - c_t) - a_{t+1})\right)\right]$$

for some weakly positive sequence $\{\lambda_t\}$, which are called *Lagrange multipliers*. The second term follows by subtracting the constraint from itself, so it is negative if the constraint doesn't bind and zero otherwise.

(1) *(2 points)* Differentiate in c_t as a variable (noting this is different from c_{t+1}, c_{t-1}, and the other indices) and set it equal to 0. You can use without proof in this section that

$$\frac{d}{dx}\mathbb{E}[f(x)] = \mathbb{E}\left[\frac{d}{dx}f(x)\right] \text{ and } \mathbb{E}[0] = 0$$

(Hint: your answer in c_t should not have an expectation operator).

(2) *(1 point)* Rearrange the derivative so that you obtain c_t as a function of the parameters (λ_t, r, b_1, b_2).

(3) *(2 points)* Differentiate now instead in a_{t+1}, noting that in this case, a_{t+1} shows up twice: in the time t equation and the time $t+1$ equation. *(Hint: in the time $t+1$ portion of the equation, e.g. the part with β^{t+1}, there will be an expectation. Your answer should be of form $\beta^t\lambda_t = \mathbb{E}[something]$).*

(4) *(2 points)* Recall $\beta = (1+r)^{-1}$. Use the equations from the above computations to prove

$$c_t = \mathbb{E}[c_{t+1}]$$

Note we are given explicitly the Lagrangian. Differentiating, the first order condition in c_t is

$$\beta^t\left(b_1 - 2b_2 c_t - \lambda_t(1+r)\right) = 0 \implies b_1 - 2b_2 c_t = \lambda_t(1+r)$$

by distributing over the sum and the Lagrangian in a_{t+1} is

$$-\beta^t\lambda_t + \mathbb{E}[\beta^{t+1}\lambda_{t+1}(1+r)] = 0$$

Dividing through by β^t and rearranging yields the implication

$$\lambda_t = \beta(1+r)\mathbb{E}[\lambda_{t+1}]$$

where we have to be careful to keep the uncertainty on future, higher order term, since we are conditioning on the information set at time t. Using the fact that $\beta(1+r) = 1$, this implies that

$$\lambda_t = \mathbb{E}[\lambda_{t+1}]$$

From this, we know that

$$b_1 - 2b_2 c_t = \lambda_t(1+r) = \mathbb{E}[\lambda_{t+1}](1+r)$$

and by rearranging the first order condition on c_{t+1}, we have that

$$\frac{1}{1+r}(b_1 - b_2 c_{t+1}) = \lambda_{t+1}$$

Combining these two, we get that

$$b_1 - 2b_2 c_t = \mathbb{E}[b_1 - 2b_2 c_{t+1}] \implies b_1 - 2b_2 c_t = b_1 - 2b_2 \mathbb{E}[c_{t+1}]$$

using the fact b_1, b_2 are constants and hence can be pulled out of the expectation by linearity. Together, this implies that $c_t = \mathbb{E}[c_{t+1}]$, exactly as desired. *(Hint: Using the result from Part (III), get an expression for λ_{t+1}. Relate this to consumption using the first order conditions in c_t and c_{t+1}).*

Part F *(2 points)* In Part (E), you showed that $c_t = \mathbb{E}[c_{t+1}]$. Interpret this condition economically, and explain how the model predicts consumers predict their consumption today.

This implies that consumption follows a martingale, and in particular implies Friedman's *rational expectations* hypothesis: what I consume today is *equal* to what I will consume tomorrow (pending all idiosyncratic productivity shocks). As a result, a representative consumer's consumption path will make them equally as well-off in any period. *One point was given for recognizing that expected future consumption in period $t+1$ would equal consumption in period t. A second point was given for recognizing this implies the expected consumption stream would be completely constant over time.*

Part G *(2 points)* Using your answer in Parts (E) and (F), write c_t as a function of r and c_{t+j} for some $j \in \mathbb{N}$ periods in the future.

The budget constraint gives us that

$$c_t = y_t + a_t - \frac{a_{t+1}}{1+r}$$

Moreover, the time $t+1$ budget constraint tells us that

$$a_{t+1} = \frac{a_{t+2}}{1+r} - y_{t+1} + c_{t+1}$$

Substituting these two equations together gives us that

$$c_t = y_t + a_t + \frac{1}{1+r}(y_{t+1} - c_{t+1}) - \frac{1}{(1+r)^2} a_{t+2}$$

Though of course we can continue this pattern for an arbitrary amount of time in the future; as a result, we can induct on this to period j and get that

$$c_t = y_t + a_t + \left(\sum_{k=1}^{j} \frac{1}{(1+r)^k} (y_{t+k} - c_{t+k}) \right) - \frac{1}{(1+r)^{j+1}} a_{t+j}$$

which gives the desired argument. *One point was given for forward iterating in one period to obtain the expression in $(t+1)$. A second point was given for inducting on this to give the general expression.*

Part H *(3 points)* Define future expected wealth at time t to be

$$W_t = a_t + \sum_{j=0}^{\infty} \beta^j \mathbb{E}_t(y_{t+j})$$

Using your answer in Part (G), show that

$$c_t = \frac{rW_t}{1+r}$$

(Hint: sum a geometric series and use the fact $\beta = (1+r)^{-1}$)
The goal is to use Part (G) and the martingale properties of consumption to get the expression. Following the hint and using Part (G), note that we have

$$c_t = a_t + \left(\sum_{k=1}^{j} \beta^{t+k} (y_{t+k} - c_{t+k}) \right) - \beta^{j+1} a_{t+j}$$

Taking expectations at time t on both sides gives that

$$\mathbb{E}[c_t] = \mathbb{E}[a_t] + \mathbb{E}\left[\sum_{k=1}^{j} \beta^{t+k} (y_{t+k} - c_{t+k}) \right] - \mathbb{E}[\beta^{j+1} a_{t+j}]$$

Because this occurs at time t, $\mathbb{E}[c_t] = c_t$ and $\mathbb{E}[a_t] = a_t$. Taking the limit to ∞ and substituting these, we get that

$$c_t = y_t + a_t + \sum_{k=1}^{\infty} \beta^{t+k} \mathbb{E}[y_{t+k}] - \sum_{k=1}^{\infty} \mathbb{E}[\beta^{t+k} c_{t+k}] - \lim_{j\to\infty} \beta^{j+1} \mathbb{E}[a_{t+j}]$$

The last term is 0 by the transversality condition all the way back in Part (D) (in particular, also the fact that $\beta^{j+1} \to_j 0$). Moreover, we can interchange sums and expectations [4] Since $y_t = \mathbb{E}[y_t]$, we can reindex the term with the y_{t+j} to include $\beta^0 y_t$, which gives

$$c_t = a_t + \sum_{k=0}^{\infty} \beta^{t+k} \mathbb{E}[y_{t+k}] - \sum_{k=1}^{\infty} \beta^{t+k} c_{t+k} = W_t - \sum_{k=1}^{\infty} \beta^{t+k} \mathbb{E}[c_{t+k}]$$

Noting that we have not pulled the same reindexing trick with consumption. However, because consumption is a *martingale* (that is, $c_t = \mathbb{E}[c_{t+1}]$, we can induct to obtain the fact that $c_t = \mathbb{E}[c_{t+k}]$ (where again, all expectations are taken from the time t perspective). As a result, we can apply the result for geometric sums to obtain that

$$c_t = W_t - \sum_{k=1}^{\infty} \beta^{t+k} c_t = W_t - \frac{\beta}{1-\beta} c_t \implies \frac{1}{1-\beta} c_t = W_t$$

Recall that $\beta = (1+r)^{-1}$, and thus we have that

$$\frac{1}{1-\beta} = \frac{1+r}{r} \implies c_t = \frac{r}{1+r} W_t$$

which is exactly the desired equality. This finishes the derivation. *One point was given for taking expectations and substituting terms. A second point was using the martingale properties of c_t to write the consumption stream as $\frac{\beta}{1-\beta} c_t$. A final point was given for correctly finishing the derivation.*

Part I *(2 points)* Using your answers above, explain how you have just proved *Friedman's permanent income hypothesis*:

$$\Delta c_t = c_t - c_{t-1} = \frac{r}{1+r}(W_t - W_{t-1})$$

that is, the change in consumption today is a weighted fraction of the total change in lifetime total wealth induced by an exogenous shock.
The algebra follows from basic substitution: note that we have

$$c_t - c_{t-1} = \frac{r}{1+r} W_t - \frac{r}{1+r} W_{t-1}$$

[4] Formally, this is due to the dominated convergence theorem, but this is immaterial and the exchange was expected to happen without justification for full credit.

essentially mechanically. *This was a "freebie" question given the intensity for the problem. Both points were awarded for any attempt (however tautological) that yielded the correct response.*

Part J *(2 points)* Compare and contrast your answers from Parts (A) and (I). Use this contrast to explain why, even if the marginal propensity to consume is very high, stimulus may not be as effective as the baseline Keynesian model predicts.

Because spending creates a "one time" income shock, it may not change current consumption by a lot because individuals will smooth their consumption over their entire lifetime; as a result, while individuals may have high lifetime marginal propensities to consumer, they may have smaller per-period marginal propensities, which are ultimately what matter in the short-run stimulus that is necessary in Keynesian models. *One point was given for noting the difference between intertemporal and per-period MPC, and a second point was given for any synthesis of ideas from the previous question.*

Problem 5: An Orchestra of Tools (25 points)

(In this problem, we consider how to estimate a dependent effect when the independent variable is not easily measurable but correlated with another measurable event.)

Part A *(2 points)* Suppose we would like to study the effect of education on wages, and we have a large sample of individuals' education and wages. We can run the regression

$$w_i = \beta e_i + \varepsilon_i$$

where e_i is individual i's education (measured in years of schooling), w_i is individual i's wage, and ε_i is everything else that we cannot observe. Suppose that both $\mathbb{E}[\varepsilon_i] = 0$ and $\mathbb{E}[\varepsilon_i | w_i, e_i] = 0$; that is, our unobservables are independent of the data we can observe, and the unobservables will eventually "wash out." How should we interpret β?

β is the *expected treatment effect*: in particular, note that we get

$$\mathbb{E}[w_i] = \mathbb{E}[\beta e_i + \varepsilon] = \beta \mathbb{E}[e_i] + \mathbb{E}[\varepsilon_i] = \beta \mathbb{E}[e_i]$$

That is, β tells us how much, given a marginal change in education, we expect wage to increase. *One point was given for using the expectation to "wash out" the mean zero noise. A second point was given for the correct interpretation that β was the marginal treatment effect.*

Part B *(3 points)* Suppose now that $\mathbb{E}[\varepsilon_i] = c \neq 0$, so that there is (constant) systematic bias in the errors. Show that this problem can be fixed by subtracting a constant term from all of your observed data, and will recover the same β as if $\mathbb{E}[\varepsilon_i] = 0$.

Suppose we subtract a constant from the right hand side: this gives

$$w_i = \beta e_i + \varepsilon - c \implies \mathbb{E}[w_i] = \mathbb{E}[\beta e_i + \varepsilon - c]$$

Because the expectation is linear, we have that

$$\mathbb{E}[w_i] = \beta \mathbb{E}[e_i] + \mathbb{E}[\varepsilon_i] - -\mathbb{E}[c] = \beta \mathbb{E}[e_i] + c - c = \beta \mathbb{E}[e_i]$$

so we can use this to debias any systematic noise that shows up. *One point was given for correctly debiasing in the functional form. A second point was given for using linearity of expectation to show the bias vanishes. A final point was given for the correct conclusion.*

Part C *(3 points)* Suppose now, however, that one unobservable is ability, and higher-ability individuals are likely to get higher wages even with the same amount of education (though we do not know by how much). Will this violate the assumption $\mathbb{E}[\varepsilon_i | w_i, e_i] = 0$? How will this affect our estimate of β?

This does violate our assumption; in particular, we expect that, conditional on a higher w but with the same e, the "unobserved" part is higher; this would cause $\mathbb{E}[\varepsilon_i | w_i, e_i]$ to be increasing in w_i, and in particular mean it is not globally 0. This will change our derivation. In particular, if the *conditional independence* assumption is violated, then we expect different noises for each (w_i, e_i) pair; in particular, this implies that we may have a higher β value than the truly observed value, because we will also be including the effect of wages. *One point was given for noting this violates the conditional independence assumption. A second point was given for noting this will change our estimate of β, and a final point was given for noting this will cause β to be overstated in our estimates.*

Part D *(2 points)* Explain intuitively why this problem is much harder to fix, and cannot be done in the same way as we did for the problem in Part (B).

Note that before, we didn't need any information about the noise and how it depended on the observation, because we knew that there was just a constant bias, so we could add a fixed effect that "anchored" our estimates. However, in order to do the same correction here, we would need to know how much "ability" affected each individual observation, which is very difficult to do without directly observing ability because it varies with our observables in some mysterious, unknown way. *One point was given for a contrast between the constant and varying biases. A second point was given for an articulation of why controlling for this would be difficult.*

Part E *(3 points)* Suppose now that we can have information about each individual's SAT scores as well. For each of the following, explain if you think that the SATs will be very correlated or only a little correlated with the listed variable, and (briefly) explain why.

(1) Wages at a company.

(2) The education a student receives.

(3) Other unobservables (e.g. gender, location, height, etc.)

(1) Positively correlated, a little bit, because higher SAT scores might measure for some stuff that affect wages such as education, propensity to work hard, etc. (but only a little bit!) They may also measure childhood wealth and class, which affect future wages via things like networks and connections.

(2) Positive correlated, a lot, since education affects networking abilities, job access, employer-relevant skills, and many other things.

(3) Not correlated, since we generally think most of these are idiosyncratic (NB: an answer such as correlated with gender due to the wage gap, or correlated with location due to geographic heterogeneity in the labor market, were accepted as well).

One point was given for each answer; with an emphasis on justifying the direction more than the direction of change itself.

Part F *(2 points)* Let c_i denote individual i's college exam score and suppose that c_i is uncorrelated with the error term ε_i (so that $\mathbb{E}[\varepsilon_i|w_i, c_i] = 0$). Consider the estimate

$$w_i = \delta c_i + \varepsilon_i$$

Will our estimate of δ be accurate? What will δ measure?
Yes, our estimate of δ will be accurate, because we have the necessary exclusion restriction on the noise to apply the argument from Part (A). Here, δ measures the marginal effect of college exam score on future wage. *One point was given for noting δ would be unbiased, and the second for correctly interpreting it.*

Part G *(6 points)* The variable c_i is called an *instrument*. Ultimately, we do not care too much about δ, but want to estimate β. How can we do that?

(1) *(2 points)* Suppose we can estimate the effect of college test scores on education. Using the noise term η_i and coefficient α, write a regression with e_i as the dependent variable and c_i as the independent variable, under the assumption $\mathbb{E}[\eta_i|e_i, c_i] = 0$.
We have the exclusion restriction on e_i and c_i, so we can write

$$e_i = \alpha c_i + \eta_i$$

as the reduced form expression we are interested in estimating. *Two points were given for the correct functional form. Partial-credit was given based on the degree to which the expression was correct.*

(2) *(2 points)* Substitute your equation for w_i into the regression from Part (A). Interpreting the new error term as $\delta\eta_i + \varepsilon_i$, and under the assumptions we have made, is it true that $\mathbb{E}[\delta\eta_i + \varepsilon_i | w_i, c_i] = 0$?
Following instructions, we get that

$$w_i = \beta(\alpha c_i + \eta_i) + \varepsilon_i$$

The new error term is $\beta\eta_i + \varepsilon_i$, and as a result, under the assumptions from Part (A), it is not necessarily true.[5] *One point was given for the functional form, and one point was given for the conclusion. See the footnote for more accurate argument and the correct form of the question.*

(3) *(2 points)* Write β as a function of δ and α. Will this be an accurate estimator?
We know that $w_i = \beta\alpha c_i + \beta\eta_i + \varepsilon_i$ and $w_i = \delta c_i + \varepsilon_i$. Under our exclusion restriction, in *expectation*, we can expect $\delta = \beta\alpha$, and so if we know α from the regression form in Part (I), and δ from the regression in Part (F), we can find $\beta = \frac{\delta}{\alpha}$. Because our estimators for δ and α are accurate, this will be accurate as well. *This question could not be answered given the way Part (II) was written, and all students attempting it were given both points. If, however, the notation error in Part (II) was caught, and the above answer was given, students were given two points of extra credit.*

Part H *(4 points)* This regression method in Part (G) is called *two-stage least squares* regression. Can you think of situations where the estimation strategy may fail? Give two different, concrete examples or situations where the instrument may not be "valid." There were a list of potential reasons why two stage least squares may fail, even in the case where we had the conditional independence on (W_i, c_i) holding that we did not have in Part (C) on e_i. In particular:

(1) The assumption $\mathbb{E}[\varepsilon_i | w_i, c_i]$ may fail, because test scores are correlated with wage through for example intergenerational wealth and prep classes.

(2) The assumption $\mathbb{E}[\eta_i | e_i . c_i]$ may fail, because the same confounder (ability) may affect the regression.

One point was given for the validity of each claim.

[5] Note that there was a typo in this question, where δ should be replaced with β, so that the problem should be interpreted as $\beta\eta_i + \varepsilon_i$, where in this case we do in fact have the necessary conditions by linearity of conditional expectation along with the assumptions that we already had. If students picked on this typo and answered the version of the question with the correct algebra, they were rewarded an additional two bonus points.

Problem 6: Supply Chains and Con-shoe-mer Welfare (25 points)

(This problem uses a game theoretic approach to derive conclusions about economic welfare in a simple model of a vertically integrated supply chain.)

Preliminaries Suppose there are only two firms in the market for shoes. Firm 1 produces shoelaces. Firm 2 then buys shoelaces from Firm 1 and uses them to produce shoes. Firm 2 sells these finished shoes to consumers. Assume Firm 1 and Firm 2 are two independent firms. That is, they do not collude in order to maximize their collective profits. Suppose that the following two-stage game is played:

1. Firm 1 chooses to announce a positive quantity of shoelaces (denoted q_1) that they will produce. This automatically determines the price of each shoelace, which is, for positive constants a and b, given by the equation:

$$p = a - bq_1$$

2. Firm 2 announces the quantity of shoes (denoted q_2) they will produce. They then buy enough shoelaces (one shoelace for each shoe) to produce that many finished shoes. For simplicity, we assume that consumers will buy all shoes produced by Firm 2.

We express the market quantity of shoes sold, Q, as $Q = \min(q_1, q_2)$. The market price, P, of a shoe is expressed by the following linear inverse demand function:

$$P = \alpha - \beta Q$$

where α and β are positive constants.

Part A *(1 point)* Let π_1 represent the profit for Firm 1. Suppose that the marginal cost for the production of each shoelace is c_1. Determine an expression for π_1.
Profit is total revenue minus total costs. So, $\pi_1 = Q(a - bq_1) - q_1 c$. 1 point is awarded for having a correct answer.

Part B *(2 points)* Let π_2 represent the profit for Firm 2. Suppose that the only costs incurred for Firm 2 is the cost of buying shoelaces. Determine an expression for π_2.
Profit is still total revenue minus total costs. So, $\pi_2 = Q(\alpha - \beta Q) - Q(a - bq_1)$. 2 points are awarded for having a correct answer.

Part C *(4 points)* Suppose that we fix the quantity of shoelaces produced by Firm 1 (that is, treat p_1 and q_1 as constants). Determine the optimal quantity q_2^* of shoes that Firm 2 ought to produce. Your answer should be a function of q_1.
First, note that to have an optimal q_2^* we must have $q_2 \leq q_1$ since if Firm 2 chooses $q_2 > q_1$, Firm 2 receives the same amount of revenue as choosing $q_2 = q_1$ but with at least as much cost. Thus, in the equation from part B we can set $Q = q_2$. We solve for the profit-maximizing quantity for Firm 2 using first-order conditions. This yields $q_2^* = \frac{\alpha - (a - bq_1)}{2\beta}$.
1 point is awarded for recognizing that $q_2 \leq q_1$. 1 point is awarded for taking the correct first-order condition. Two points are awarded for a final correct answer.

Part D *(4 points)* Using the previous part, Firm 1 can reasonably predict the number of shoes Firm 2 will produce in response to what Firm 1 announces as their q_1. Knowing this, determine the optimal quantity of shoelaces that Firm 1 should announce.
To solve for the optimum quantity q_1^*, we substitute in what Firm 1 expects Firm 2's quantity as a function of q_1. Thus, in the equation from part A we can substitute the answer from part C

for $q_2 = Q$. We solve for the profit-maximizing quantity for Firm 1 using first-order conditions. This yields $q_1^* = \frac{2ab-\alpha b-\beta c}{b^2}$.
1 point is awarded for substituting for q_1. 1 point is awarded for taking the correct first-order condition. Two points are awarded for a final correct answer.

Part E *(4 points)* Based on the optimal q_1 and q_2 announced by Firms 1 and 2, determine the resulting profits for each firm, π_1 and π_2, the market price, P, and market quantity, Q.
We observed earlier that $Q = q_2$ since it is never optimal for $q_2 > q_1$. To compute pi_1, pi_2 and P we plug in the values of q_1^* and q_2^* obtained from parts C and D. First we substitute q_1^* into our expression for q_2^*. This yields $q_2^* = \frac{ab-\beta c}{2b\beta}$. Then, we have:

$$\pi_1 = \frac{\alpha ab^2 - 2ab\beta c - a^2b^2 + \beta^2c^2 + \alpha b\beta c}{2b^2\beta} \tag{4}$$

$$\pi_2 = \frac{(ab-\beta c)^2}{2b^2\beta} \tag{5}$$

$$P = \frac{2\alpha b - ab + \beta c}{2b} \tag{6}$$

One point was awarded for getting each of $P, Q, \pi_1 and \pi_2$ correct.

We now compare the results of this situation to that of a monopoly. Suppose Firms 1 and 2 combine into a single firm which has a monopoly over the shoe market. Now, the combined firm announces a single market quantity, q_M, of shoes which it will produce. Price is determined by the same market inverse demand function.

Part F *(2 points)* Determine an expression for π_M, the total profit of the monopoly.
Profit is total revenue minus total costs. So, $pi_M = q_M(\alpha - \beta q_M) - q_M c$. 1 point is awarded for having a correct answer.

Part G *(2 points)* Assuming consumers will buy every shoe produced by the monopoly, what is the optimal quantity, q_M^*, that the monopoly will choose to announce?
We solve for the profit-maximizing quantity for the monopoly firm using first-order conditions. This yields $q_M^* = \frac{\alpha-c}{2\beta}$.
1 point is awarded for taking the correct first-order condition. 1 point is awarded for a final correct answer.

Part H *(3 points)* Suppose the two combined firms split monopoly profits equally. In which market structure is each firm better off? Use your previous work to justify your answer.
Each firm makes half of the total profit, which is $\frac{\alpha-c}{4\beta}$. 1 point is awarded for correctly computing the profit of each individual firm under monopoly. 1 point is awarded for a correct comparison of Firm 1's profits under each market structure. 1 point is awarded for a correct comparison of Firm 2's profits under each market structure.

Part I *(4 points)* Calculate the monopoly price and monopoly quantity for this market. Are consumers better off under a monopoly?
When $\alpha = a$ and $\beta = b$, the market price for each market situation are equal. Similarly, we only need $\beta = b$ for the market quantity of each situation to be equal. Thus, if $a > \alpha$ consumers are worse off since prices are higher under monopoly. Likewise, if $a < \alpha$ consumers are better off since prices are lower under monopoly. Furthermore, if $b > \beta$ consumers are worse off since market price is higher and market quantity is lower under monopoly. If $b < \beta$ consumers are better off since market price is lower and market quantity is higher under monopoly. If either

$a > \alpha$ and $b < \beta$, or $a < \alpha$ and $b > \beta$, the effect is unclear since it depends on the magnitudes of the differences. Two points are awarded for identifying when the situations are equal. One point is awarded for a correct comparison of price in the two market situations. One point is awarded for a correct comparison of quantity in the two market situations.