Problem Set 1: Insurance

ECON 24450: Inequality and the Social Safety Net (Deshpande)

Due Tuesday, April 13, 8am on Canvas

1. **Model of adverse selection.** Consider the following adverse selection model based on Einav and Finkelstein (2011). Suppose an agent has income I and experiences a loss of L with probability θ. The agent can purchase full insurance in a competitive market for an amount p. The agent has a utility function u ($u' > 0$, $u'' < 0$).
   (a) Write down the condition for whether the individual insures. Solve for θ and interpret your result.
   (b) Suppose that the distribution of θ in the economy is given by the cdf $F(\theta)$. Write down how the competitive insurance company will set the price p in terms of θ. You do not need to solve explicitly for the equilibrium p.
   (c) Show graphically the AC, MC, and demand curves, and characterize the efficient outcome and equilibrium outcome for insurance coverage in the following cases:
      i. There is adverse selection and demand is everywhere above marginal cost.
      ii. There is adverse selection and demand is everywhere above average cost.
      iii. There is adverse selection and demand is everywhere above MC and everywhere below AC.
      iv. There is advantageous selection (i.e., those with highest demand have the lowest θ). Why might such a situation arise?

2. **Adverse selection.** The likelihood of getting breast cancer is determined by a combination of family history and lifestyle choices, especially smoking. Suppose the probability of getting cancer breaks down as follows:

	Smoker	Non-smoker
Good history	10%	5%
Bad history	20%	10%

   The expected cost for treating breast cancer is $10,000. Assume that in addition to the expected dollar value of the policy, all women (no matter their risk) value the security of having insurance for breast cancer for an additional $400. Assume that 25% of women fall into each category.
   (a) Suppose the insurance company cannot observe either smoking status or family history and offers the same insurance policy to all individuals. What price would it have to charge to break even if everyone participated in the insurance market? At the break-even price, which women would participate in the market? (Don't forget the risk premium.) Relate your answer to adverse selection.
   (b) Suppose the insurance company can observe whether an individual is a smoker and is allowed to price separately for smokers and non-smokers. Is there a price at which the insurer could make a profit and provide insurance for all smokers? Is there a price at which the insurer could make a profit and provide insurance for all non-smokers?
   (c) Given your answer to (b), if the insurer cannot serve all of the individuals in one of the smoking categories, what price must it charge to break-even while serving only the individuals with the "bad" family history in that smoking category?
   (d) Given your answers to (b) and (c), when the insurance company can discriminate on the basis of smoking behavior, what is the difference between the break-even prices offered to smokers vs. non-smokers? Which women participate in the insurance market in this case? Which groups are made better off or worse off than in the case with a single policy from part (a)?

3. **Unemployment insurance.** Ron Swanson works at a steel mill, and earns an amount E. He faces some risk of being unemployed, denoted by p. If he becomes unemployed, he loses all of his income. His utility is of the form $U = ln(C)$, where C is consumption and is equal to income.

   (a) Suppose initially that there is no public unemployment insurance, but there is a well-functioning private unemployment insurance market. In this market, you pay a premium (a) if you are employed, and in return you get some net benefits if unemployed (b). Suppose that the private unemployment insurance is actuarially fair (so that insurer makes zero expected profits on Ron). Ron is deciding how much insurance to buy, or how much income he will have if he is employed next period and how much he will have if he is unemployed next period. Derive mathematically how much insurance Ron will buy. Show whether this is more or less than full insurance.

   (b) Continue to assume that the private insurance is actuarially fair. Now the government introduces a public insurance system. The system replaces half of Ron's lost earnings ($E/2$) if he becomes unemployed. It is financed by an actuarially fair tax on Ron if Ron is employed—so the government expects to break even on this insurance policy. Now how much private insurance will Ron want? How does it relate to part (a)? Why? (The algebra is kind of a mess, but Ron would say that it builds character.)

   (c) What is the effect of providing this public unemployment insurance on Ron's well-being (as measured by expected utility)?

   (d) Now suppose that there is no private unemployment insurance at all. In this world, calculate the effect on Ron's well-being of introducing actuarially fair public unemployment insurance that replaces half of his lost income. Is this gain larger or smaller than in (c)?

   (e) Is the analysis of (c) or (d) more applicable to US unemployment insurance and why? What is the empirical evidence on how public unemployment insurance affects well-being? What is the empirical evidence on the costs of public unemployment insurance and does this question consider them?

4. **Risk aversion; moral hazard.** Jean-Ralphio can exert effort ($e = 1$) to avoid an accident at work or not exert any effort ($e = 0$). If $e = 1$, the probability of an accident is $1/2$. If $e = 0$, the probability of an accident is 1. Jean-Ralphio's income without the accident is \$100. In case of an accident, medical expenses will be \$64, so his net income in that case is \$36. Jean-Ralphio's utility of income is $\sqrt{I}$. The cost of effort, $C(e)$, is 0 if effort is $e = 0$ and 1 if effort is $e = 1$. Jean-Ralphio's utility function is $u(I, e) = \sqrt{I} - C(e)$.

   (a) Is Jean-Ralphio risk averse, risk neutral, or risk loving? Explain algebraically. Consider a world without insurance in which Jean-Ralphio, uncharacteristically, exerts effort. Calculate the certainty equivalent and risk premium associated with the risk of getting injured and explain what each term means in words.

   (b) Compare Jean-Ralphio's expected utility when he exerts effort versus when he doesn't. Will Jean-Ralphio choose to exert effort?

   (c) Now suppose there is a risk neutral insurance company. Suppose the insurance company cannot monitor Jean-Ralphio's behavior. The insurance company considers three contracts, labeled A, B, and C. Each contract specifies the price p and the amount of money d Jean-Ralphio gets in case of an accident. Given p and d, Jean-Ralphio's final income in case of no accident is $I_0 = 100 - p$ and his final income in case of an accident is $I_A = 36 - p + d$. The contracts are as follows:

Contract	Price p	Payment d
A	\$36	\$64
B	\$19	\$47
C	\$19	\$32

   For each of the two contracts, calculate his final incomes, I_0 and I_A, in the two outcomes, and list them in a table like this:

Contract	I_0	I_A
A		
B		
C		

Which of these contracts offers full insurance to JR?

(d) For each of these contracts, determine which of the two effort levels, $e = 0$ or $e = 1$, would be expected utility maximizing for Jean-Ralphio if he accepted that contract. Assume that Jean-Ralphio, if both effort levels yield the same expected utility, chooses $e = 1$. Relate your answer to the concept of moral hazard.

(e) Which of these contracts are such that JR would accept the contract rather than staying uninsured? Assume that he accepts a contract if indifferent between insuring and not insuring.

(f) Which of the three contracts gives the insurance company the highest expected profits? What are the expected profits of the insurance company if it offers this contract?

5. **Baily-Chetty derivation.** Using the lecture notes as a guide, set up the Social Planner's problem and derive the Baily-Chetty formula:

$$\frac{\epsilon_{1-e,b}}{e} = \frac{u'(c_L) - u'(c_H)}{u'(c_H)}$$

Some useful substitutions in deriving

$$e\frac{dt}{db} = (1-e)[1 + \epsilon_{1-e,b}\frac{1}{e}]$$

include the following:

$$\begin{aligned} \frac{de}{db} &= -\frac{d(1-e)}{db} \\ t &= \frac{1-e}{e}b \\ \epsilon_{1-e,b} &\equiv \frac{b}{1-e}\frac{d(1-e)}{db} \end{aligned}$$

Explain the intuition behind both sides of the equation and how they relate to each other.

6. **Data exercise: instrumental variables.** In a few weeks, we will discuss disability insurance in detail. This data exercise is based on the paper *The Impact of Economic Conditions on Participation in Disability Programs: Evidence from the Coal Boom and Bust* (Black, Daniel, and Sanders 2002), which you should read before starting this question.

(a) In one sentence, what is the causal relationship of interest in this paper? Write down the regression equation of interest ("structural" equation) as it appears in the paper.

(b) Using the "omitted variables bias" formula, show why OLS might produce a biased estimate of the causal relationship.

(c) Instruments

i. What instruments do the authors propose to solve the OVB problem?

ii. Explain the two conditions ("first stage" and "exclusion restriction") for a valid instrument and write down the first-stage equation as it appears in the paper.

iii. Explain why the instruments might satisfy these conditions. List one potential violation of the exclusion restriction.

(d) Download the data set from Canvas and import it into your preferred statistical analysis program (e.g., Stata, R). Please attach your code and output to your problem set, and CLEARLY LABEL the portion of the code that corresponds to each question. [Note: If you use R, these resources might be helpful for calculating standard errors: Basic Robust SEs in R (http://www.drewdimmery.com/robust-ses-in-r/); Function felm() in package 'lfe' (https://cran.r-project.org/web/packages/lfe/lfe.pdf).]

i. Construct the outcome variables.

A. Transform the following variables by taking natural logs: DI payments, SSI payments, county population (in thousands) and county earnings.

B. Construct log differences for the following variables: DI payments, SSI payments, county population (in thousands) and county earnings. (Hint: the log difference for variable x at time t is $ln(x_t) - ln(x_{t-1})$.)

C. Construct a variable that equals the fraction of county earnings from manufacturing in 1969.

ii. Construct the instruments.

A. Construct a variable that equals the annual Producer Price Index (PPI) for coal divided by the annual Consumer Price Index of Urban wage earners (CPIU). This variable measures the price of coal.

B. Construct a variable that equals the log difference of the price of coal multiplied by the log of coal reserves if coal reserves are greater than 0, and 0 otherwise. This variable measures the change in the value of coal reserves. Also construct two time lags of this variable.

iii. Construct estimates: because the authors suppress some confidential data in the public-use data set, your results will not exactly match the results in Table 3 from the paper.

A. Estimate the structural equation by OLS for the following outcome variables: log difference in county SSI payments, log difference in county DI payments. You should include a full set of state-by-year dummy variables in your regression equations. Obtain heteroskedastic-robust standard errors. Interpret the main estimate in one sentence.

B. The authors don't show first-stage estimates, but if they had taken this class, they would have known that they need to. Estimate the first-stage equation. Interpret the main estimates in one sentence.

C. Estimate the structural equation by 2SLS, using the change in the value of coal reserves and two time lags as instruments, for the following outcome variables: log difference in county SSI payments, log difference in county DI payments. You should include a full set of state-by-year dummy variables in your regression equations. Obtain heteroskedastic-robust standard errors. Interpret the main estimate in one sentence.

D. Construct a table that contains the main estimates from (A) and (C). Report standard errors in parentheses below each estimate. In your table, indicate whether each estimated coefficient is significant at the 1%, 5% or 10% level.