Problem Set 1 Solutions: Insurance

Professor Manasi Deshpande

Due Tuesday, October 11, at the beginning of class

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.

   (a) Write down the condition for whether the individual insures. Solve for θ and interpret your result.

   Solution

   Individual insures iff $u(I-p) \geq (1-\theta)u(I) + \theta u(I-L) \implies \theta^* \geq \frac{u(I)-u(I-p)}{u(I)-u(I-L)}$. This condition reflects adverse selection Those who insure have higher expected costs.

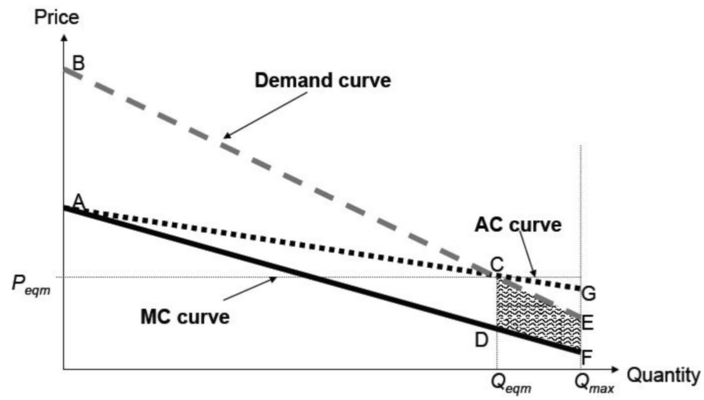
   (b) Suppose that the distribution of θ in the economy is given by $F(\theta)$. Write down how the competitive insurance company will set the price p in terms of θ.

   Solution

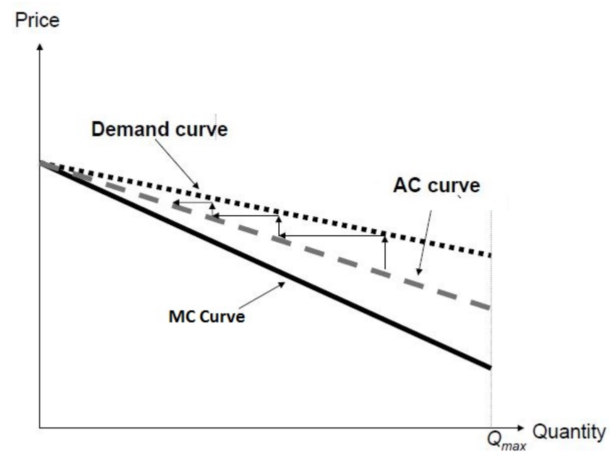
   The market is competitive, so the insurer must price at expected cost in equilibrium: $p = E[\theta \times L | \theta \geq \theta^*] = E[\theta \times L | 1 - F(\theta^*)]$.

   (c) Show graphically the AC, MC, and demand curves, and characterize the efficient outcome and equilibrium outcome for insurance coverage in the following cases:

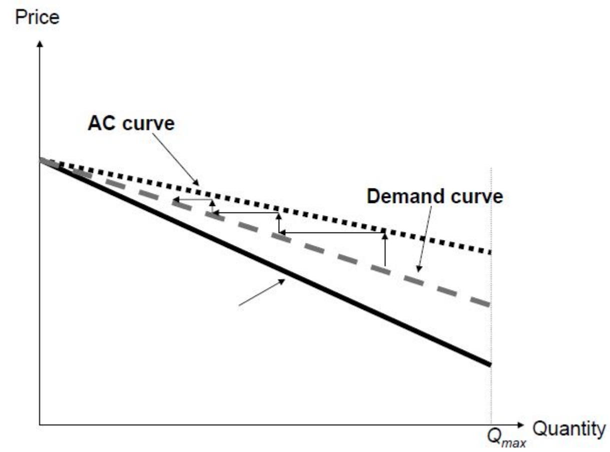
      i. There is adverse selection and demand is everywhere above marginal cost.



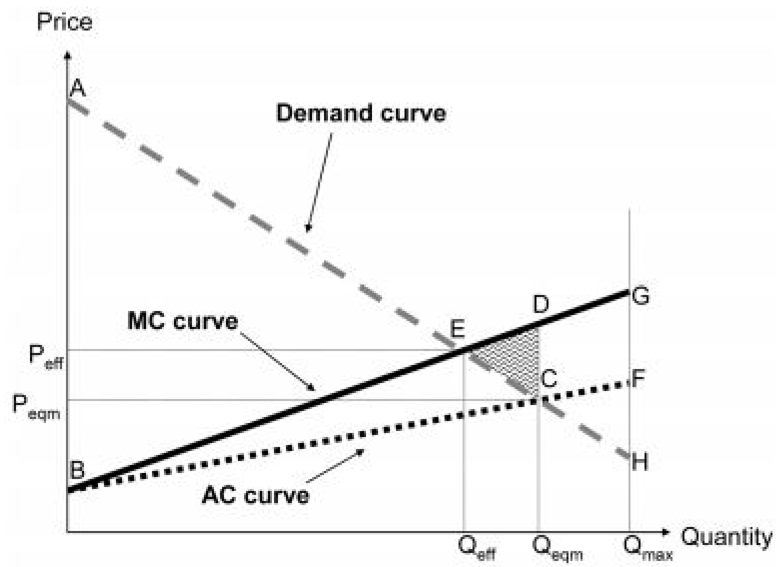
i. There is adverse selection and demand is everywhere above average cost.



i. There is adverse selection and demand is everywhere above MC and everywhere below AC.



i. There is advantageous selection (i.e., those with highest demand have the lowest θ). Why might such a situation arise?



(Graph from Einav and Finkelstein 2011)

Advantageous selection could arise if individuals with the lowest risk are also the most risk averse.

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.

Solution

The break-even price balances revenues with expected costs. Expected costs for an additional user can be written as follows:

$$.25 * (.10 + .05 + .20 + .10) * \$10,000 = \$1,125$$

At a break-even price of $1,125, women will buy the plan if their expected beneft exceeds the price of the plan. Define p to be the probability of getting cancer for a given individual. Individuals will buy the plan as long as:

$$10,000 * p + 400 > 1,125 \iff p > \frac{1,125 - 400}{10,000} = .0725$$

Thus, everyone except non-smokers with a good family history will buy insurance.

In this case, individuals have private information about their smoking status and family history. Private information creates adverse selection because insurers cannot observe risk and must offer a single insurance policy to all individuals. At the break-even price, only individuals with the highest risk (e.g., smokers or individuals with a poor family history) are willing to purchase the insurance policy.

(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?

Solution

For the smokers, the break-even price is:

$$.5 * (.10 + .20) * 10,000 = \$1,500$$

Thus, the firm must charge a price above this in order to have a positive profit. At a price of $1,500, individuals will buy insurance as long as:

$$p > \frac{1500 - 400}{10,000} = 0.11$$

But, smokers with a good family history have a probability of 0.10, and consequently will opt out of the market. Thus, no price allows the firm to sell insurance to all smokers while making a profit.

Now consider non-smokers. Here the break even price is:

$$.5 * (.10 + .05) * 10,000 = \$750$$

At this price, the probability cutoff for participation is given by:

$$p > \frac{750 - 400}{10,000} = 0.035$$

All non-smokers could buy insurance at a price at which the insurance company turns a profit.

(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?

Solution

If only smokers with bad histories buy insurance, the company must charge at least $.2 * 10,000 = \$2,000$.

(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)?

Solution

Non-smokers pay $750, while smokers with bad family histories pay $2,000. The only group that does not participate comprises smokers with good family histories.

Relative to the case with a single policy from part (a),

i. Non-smokers with a good family history are better off, since they receive insurance at a price they are willing to pay:

$$1,125 > 10,000 * .05 + 400 = 900 > 750$$

ii. Non-smokers with a bad family history are better off, since they receive insurance at a price less than the price they paid in (a):

$$10,000 * .10 + 400 = 1,400 > 1,125 > 750$$

iii. Smokers with a good family history are worse off, since they no longer receive insurance:

$$\underbrace{1,500}_{\text{Break-Even Price from (b)}} > 10,000 * .10 + 400 = 1,400 > 1,125$$

iv. Smokers with a bad family history are worse off, since they receive insurance at a price greater than the price they paid in (a):

$$10,000 * .20 + 400 = 2,400 > 2,000 > 1,125$$

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 [Ron approves], 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.

Solution

We need to maximize expected utility taking account of the zero-profit constraint. The zero profit constraint means that expected payments by the insurance firm to unemployed workers have to be equal to expected premiums collected from unemployed workers. This can be expressed as follows:

$$pb = (1-p)a$$
$$b = \frac{1-p}{p}a$$

Ron maximizes expected utility subject to this zero-profit condition by picking the insurance contract (a, b) that satisfies:

$$\max_{a,b} EU = pln(b) + (1-p)ln(E-a) \text{ s.t. } b = \frac{1-p}{p}a$$

We can substitute the zero-profit constraint into the utility function:

$$\max_{a,b} EU = pln(\frac{1-p}{p}a) + (1-p)ln(E-a)$$

We take the derivative with respect to a and set that equal to 0:

$$\frac{\partial EU}{\partial a} = \frac{p}{a} - \frac{1-p}{E-a} = 0$$
$$\implies a(1-p) = p(E-a)$$

The optimal insurance contract is $a^* = pE$ and $b^* = \frac{a(1-p)}{p} = (1-p)E$.

Is this full insurance? To find out, we can compare Ron's consumption with and without employment:

$$C_{unemployed} = b = (1-p)E$$
$$C_{employed} = E - a = (1-p)E$$

Ron has purchased full insurance. Ron is risk averse and wants to fully insure against the risk of unemployment. He buys insurance at the actuarially fair rate from the insurer so that he can consume the same amount whether or not he becomes unemployed.

(b) Continue to assume that the private insurance is actuarially fair. Now the government introduces a public insurance system [Ron doesn't much like this]. 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.)

Solution

The government guarantees the payment of $E/2$ if Ron becomes unemployed. This means that the expected payout is $p(E/2)$. Given that the payouts are actuarially fair, the government will just break even by levying a tax. Such a tax has to satisfy the following budget constraint:

$$(1-p)T = p\frac{E}{2}$$
$$\implies T^* = \frac{pE}{2(1-p)}$$

Ron will now have an income of $b + E/2$ if he is unemployed and an income of $E - T - a$ if he is employed. The new maximization problem is:

$$\max_{a,b} EU = pln\Big(b + \frac{E}{2}\Big) + (1-p)ln\Big(E - T - a\Big)$$
$$\text{s.t. } b = \frac{(1-p)a}{p}, T = \frac{pE}{2(1-p)}$$

Substituting the expressions for T and the private market equilibrium constraint (for b) into the expected utility function yields:

$$\max_{a} pln\Big(\frac{(1-p)a}{p} + \frac{E}{2}\Big) + (1-p)ln\Big(E - \frac{pE}{2(1-p)} - a\Big)$$

We now maximize this with respect to a as before. We do this by taking the derivative with respect to a and setting that equal to 0:

$$\frac{\partial EU}{\partial a} = \frac{(1-p)}{\frac{(1-p)}{p}a + \frac{E}{2}} - \frac{(1-p)}{E - \frac{pE}{2(1-p)} - a} = 0$$

The denominators have to be the same for this to hold, so we can set them equal:

$$\frac{1-p}{a} + \frac{E}{2} = E - \frac{pE}{2(1-p)} - a$$
$$\implies \frac{a}{p} = \frac{E(1-p) - pE}{2(1-p)}$$
$$\implies a^* = \frac{p(1-2p)E}{2(1-p)}, b^* = \frac{(1-2p)E}{2}$$

Because Ron has full access to actuarially fair private markets, nothing changes with respect to part (a). He complements the government policy with a private policy that together provide him full insurance at an actuarially fair price. Combining the two policies shows that this is true:

$$\text{"premium"} = a + T = pE$$

$$\text{"benefit"} = \frac{E}{2} + b = (1-p)E$$

These combined benefits and premiums are both the same as the benefit and premium under the single private policy in part (a).

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

Solution

Expected utility is the appropriate measure of well-being. Both (a) and (b) yield the same expeted utility. Thus, there is no gain or loss to Joe of the government stepping in to provide public unemployment insurance. There is no missing market here and the public and private insurance are perfect substitutes for one another. There is really no reason for the government to intervene in this case.

(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)?

Solution

Once again, use expected utility as the indicator of well-being. In the case of no insurance, expected utility is minus infinity. We have the expression *ln(0)* in the case when Joe becomes unemployed. Joe is risk averse and gets a huge negative utility shock from the possibility of being unemployed with no earnings and no insurance benefits. Thus, the gain is certainly much larger than in (d) where the public program made no difference.

$$\text{Without insurance} : EU = pln(0) + (1-p)ln(E) = -\infty$$

$$\text{With insurance} : pln\Big(\frac{E}{2}\Big) + (1-p)ln\Big(E - \frac{(1-p)E}{2p}\Big)$$

Thus, the improvement in utility is infinite.

(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?

Solution

The analysis of (d) is more relevant, though its conclusions are too extreme. There are no private unemployment insurance policies in the real world. There are too many adverse selection and imperfect information problems for that to be feasible. However, individuals do have some recourses since they can save to insure themselves against unemployment.

Gruber (AER 1997) finds that unemployment insurance does improve the welfare of the unemployed. Consumption falls less than without benefit receipt, though also increases less than if unemployment insurance were the only form of support unemployed workers have.

However, the analysis in this problem has ignored the fact that unemployment insurance can actually affect the value of p through moral hazard. Though there are benefits to providing workers with this form insurance, there are also potentially high moral hazard costs. The cost of these distortions may be greater than the benefits.

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.

   Solution

   Jean-Ralphio is risk averse. His utility function is concave in income, which means that he prefers a certain payout to an uncertain lottery that has the same expected value as the certain payout. We can prove this mathematically in two steps:

   i. Jean-Ralphio's utility function is concave in income because:
   $$\frac{\partial u(I,e)}{\partial I} = -\frac{1}{4}(I)^{\frac{-3}{2}} < 0$$

   i. Consider a lottery l that delivers payoff I_0 with probability p and payoff I_1 with probability$1 - p$. Because Jean-Ralphio's utility function is concave in income, Jensen's inequality tell us that:
   $$pu(I_0, e) + (1-p)u(I_1, e) < u(pI_0 + (1-p)I_1, e)$$

   The certainty equivalent is the amount of "certain" income that would give Jean-Ralphio the same utility as the uncertain lottery associated with the risk of getting injured. The certainty equivalent is given by:

   $$\begin{aligned} &\frac{1}{2} * \sqrt{100} + \frac{1}{2} * \sqrt{36} - 1 = \sqrt{CE} - 1 \\ \implies &8 = \sqrt{CE} \\ \implies &CE = \$64 \end{aligned}$$

   The risk premium is the difference between the expected income from the uncertain lottery associated with the risk of getting injured and the certainty equivalent. The risk premium is given by:

   $$\frac{1}{2} * 100 + \frac{1}{2} * 36 - 64 = \$4$$

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

Solution

Jean-Ralphio's expected utility when he exerts effort is:

$$\frac{1}{2} * \sqrt{100} + \frac{1}{2} * \sqrt{36} - 1 = 7$$

Jean-Ralphio's expected utility when he does not exert effort is:

$$0 * \sqrt{100} + 1 * \sqrt{36} - 0 = 6$$

Thus, Jean-Ralphio will 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?

Solution

Contract	I_0	I_A
A	100-36=64	36-36+64=64
B	100-19=81	36-19+47=64
C	100-19=81	36-19+32=49

Contract A offers full insurance to Jean-Ralphio, since this contract smooths his consumption over the two possible states of the world.

(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.

Solution

For contract A, Jean-Ralphio chooses $e = 0$ and receives expected utility equal to 8:

$$EU|_{e=1} = \frac{1}{2}\sqrt{64} + \frac{1}{2}\sqrt{64} - 1 = 7$$
$$EU|_{e=0} = \sqrt{64} = 8$$

For contract B, Jean-Ralphio chooses $e = 0$ and receives expected utility equal to 8:

$$EU|_{e=1} = \frac{1}{2}\sqrt{81} + \frac{1}{2}\sqrt{64} - 1 = 7.5$$
$$EU|_{e=0} = \sqrt{64} = 8$$

For contract C, Jean-Ralphio chooses $e = 1$ and receives expected utility equal to 7:

$$EU|_{e=1} = \frac{1}{2}\sqrt{81} + \frac{1}{2}\sqrt{49} - 1 = 7$$
$$EU|_{e=0} = \sqrt{49} = 7$$

Under contract A (the full insurance case), Jean-Ralphio does not exert effort because he receives the same income regardless of his effort level. It is optimal for Jean-Ralphio to choose $e = 0$ under contract A because (1) it is costly for Jean-Ralphio to exert effort in order to reduce the likelihood of an accident and (2) Jean-Ralphio does not bear the cost of an accident since he receives the same payoff from the insurance company across all states of the world. This is the moral hazard problem.

(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.

Solution

From part (b), it is optimal for Jean-Ralphio to exert effort and receive expected utility equal to 7 when he is uninsured. Thus, Jean-Ralphio will accept any of these contracts rather than staying uninsured because each contract delivers expected utility greater than or equal to 7..

(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?

Solution

For contract A, JR chooses $e = 0$ and the insurance company's expected profits $E[\pi]$ are:

$$E[\pi] = 36 - 64 = -28$$

For contract B, JR chooses $e = 0$ and the insurance company's expected profits are:

$$E[\pi] = 19 - 47 = -28$$

For contract C, JR chooses $e = 1$ and the insurance company's expected profits are:

$$E[\pi] = 19 - \frac{1}{2} * 32 = 3$$

Thus, the insurance company receives the highest expected profits from contract C.

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.

Solution:

Agent's Problem:

$$\begin{aligned}\tilde{W} &= \max_{e} eu(c_H) + (1-e)u(c_L) - \Psi(e) \\ \text{FOC: } u(c_H) - u(c_L) &= \Psi'(e)\end{aligned}$$

Social Planner's Problem:

$$\max_{b,t} \tilde{W}(b,t) \text{ s.t. } te = (1-e)b$$

Rewrite in terms of b:

$$\begin{aligned}\max_{b} \tilde{W}(b, t^*(b)) &= eu(A + \omega_H - t(b)) + (1-e)u(A + \omega_L + b) - \Psi(e) \\ \text{FOC: } 0 &= \tilde{W}_b + \tilde{W}_t \frac{dt}{db} \\ &= (1-e)u'(c_L) - eu'(c_H)\frac{dt}{db}\end{aligned}$$

Find $e\frac{dt}{db}$ by totally differentiating the balanced budget constraint:

$$\begin{aligned}(de \times t) + (e \times dt) &= (db \times (1-e)) + (b \times d(1-e)) \\ \frac{de}{db}t + e\frac{dt}{db} &= (1-e) + b\frac{d(1-e)}{db} \\ \frac{-d(1-e)}{db} \times (\frac{1-e}{e} \times b) + e\frac{dt}{db} &= (1-e) + b\frac{d(1-e)}{db} \\ e\frac{dt}{db} &= 1 - e + b\frac{d(1-e)}{db}[1 + \frac{1-e}{e}] \\ &= 1 - e + b\frac{d(1-e)}{db}\frac{1}{e} \\ &= (1-e)[1 + \frac{b}{1-e}\frac{d(1-e)}{db}\frac{1}{e}] \\ &= (1-e)[1 + \epsilon_{1-e,b}\frac{1}{e}]\end{aligned}$$

where the third line follows because $\frac{de}{db} = \frac{-d(1-e)}{db}$ and $t = \frac{1-e}{e}b$ and $\epsilon_{1-e,b} \equiv \frac{b}{1-e}\frac{d(1-e)}{db} < 0$ is the elasticity of hazard (finding a job) to benefits.

Substituting into the Planner's Problem:

$$\begin{aligned}0 &= (1-e)u'(c_L) - (1-e)[1 + \epsilon_{1-e,b}\frac{1}{e}]u'(c_H) \\ \frac{\epsilon_{1-e,b}}{e} &= \frac{u'(c_L) - u'(c_H)}{u'(c_H)}\end{aligned}$$

Intuition:

- Suppose there is no moral hazard ($\epsilon = 0$). Then $u'(c_L) = u'(c_H) \implies c_L = c_H$. We equalize the MU of consumption (full insurance).
- RHS is quantifies the gain to society in transferring a dollar from employed to the unemployed state. When ϵ is larger, the RHS will be larger and we will be further from full insurance.
- LHS quantifies the cost to society of transferring a dollar from the employed to the unemployed state, since moral hazard means that the agent does not consider the government budget constraint. Moral hazard leads to higher unemployment rates and longer unemployment durations, which must be financed with taxes.
- The equation gives us the optimal benefit level b^* by equating the social gain from transferring another dollar to the unemployed state (which depends on the difference in MU between unemployed and employed state) with the social cost of the transfer (which is the taxes required to finance the additional unemployment and unemployment benefits created by the behavioral response of moral hazard).

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.

   Solution

   The authors want to measure the effect of local economic conditions (as measured by county-level earnings) on disability program participation. The structural equation is given by:

   $\Delta y_{ist} = \beta_0 + year_{st}\beta_{1st} + x_{ist}\beta_2 + \beta_3 \Delta earnings_{ist} + \epsilon_{ist}$

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

   Solution

$$\begin{aligned}
\beta_3^{OLS} &= \frac{cov(\Delta y_{ist}, \Delta earnings_{ist}|year_{st}, x_{ist})}{var(\Delta earnings_{ist}|year_{st}, x_{ist})} \\
&= \frac{cov(\beta_0 + year_{st}\beta_{1st} + x_{ist}\beta_2 + \beta_3 \Delta earnings_{ist} + \epsilon_{ist}, \Delta earnings_{ist}|year_{st}, x_{ist})}{var(\Delta earnings_{ist}|year_{st}, x_{ist})} \\
&= \beta_3 + \underbrace{\frac{cov(\epsilon_{ist}, \Delta earnings_{ist}|year_{st}, x_{ist})}{var(\Delta earnings_{ist}|year_{st}, x_{ist})}}_{\text{"omitted variable bias" term}}
\end{aligned}$$

If the omitted variable bias term does not equal zero because their exists some unobservable variable that (1) is a determinant of the outcome variable Δy_{ist} and (2) is correlated with $\Delta earnings_{ist}$, then OLS does not "identify" the causal parameter of interest β_3 and OLS will produce a biased estimate of β_3.

(c) Instruments

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

Solution

The authors propose using the log change in the value of a county's coal reserves and two time lags of this variable as instruments for the log change in the county's earnings.

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.

Solution

The "first stage" or "relevance" condition requires that the instrument be correlated with the endogenous variable. In this setting, the log change in the value of a county's coal reserves (and two time lags) must be correlated with the log change in the county's earnings.

The "exclusion restriction" states that the only way the instrument is correlated with the outcome variable is through the endogenous variable. In this setting, the log change in the value of a county's coal reserves (and two time lags) must be correlated with the log change in the county's disability receipt only through its effect on the log change in the county's earnings.

The first stage equation is given by $\Delta earnings_{ist} = \alpha_0 + year_{st}\alpha_{1st} + x_{ist}\alpha_{ist} + \Delta(valueofreserves)_{ist}\alpha_s + u_{ist}$

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

Solution

The instruments probably satisfy the first stage condition because the value of coal reserves is an important determinant of local economic conditions and employment in the coal-producing regions that the authors study. The instruments might satisfy the exclusion restriction if the change in the value of coal reserves primarily affects the residents of coal-producing regions through changes in the labor market (changes in earnings or employment opportunities in the coal sector).

The exclusion restriction would be violated if, for example, the increased coal-mining activity associated with the increase in the value of a county's coal reserves results in increased exposure to environmental pollutants. If this increased exposure to pollutants exacerbates chronic conditions like severe asthma that qualify an individual to receive disability benefits, then the exclusion restriction is violated: the change in the value of the county's coal reserves is correlated with the change in the county's disability receipt through a channel other than the change in the county's earnings.

(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).]

See log file and .do file on Canvas for detailed solutions.

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.

Solution

A 10 percent increase in earnings within a county reduces the amount of DI payments by approximately 3.44 percent.

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.
Solution
A 10 percent increase in the value of coal reserves within a county reduces county earnings by approximately 0.225 percent.

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.
Solution
A 10 percent increase in earnings within a county reduces the amount of SSI payments by approximately 8.62 percent.

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.